

CSRN 2503

(Ed.)

- **Vaclav Skala**
University of West Bohemia, Czech Republic

Computer Science Research Notes

**23rd International Conference in Central Europe on
Computer Graphics, Visualization and Computer Vision
WSCG 2015
Plzen, Czech Republic
June 8 - 12, 2015**

Proceedings

WSCG 2015

Posters Proceedings

ISSN 2464-4617 (print)

ISSN 2464-4625 (CD-ROM)

CSRN 2503

(Ed.)

- **Vaclav Skala**
University of West Bohemia, Czech Republic

Computer Science Research Notes

**23rd International Conference in Central Europe on
Computer Graphics, Visualization and Computer Vision
WSCG 2015
Plzen, Czech Republic
June 8 - 12, 2015**

Proceedings

WSCG 2015

Posters Proceedings

© Vaclav Skala – UNION Agency

This work is copyrighted; however all the material can be freely used for educational and research purposes if publication properly cited. The publisher, the authors and the editors believe that the content is correct and accurate at the publication date. The editor, the authors and the editors cannot take any responsibility for errors and mistakes that may have been taken.

Computer Science Research Notes CSRN 2503

Editor-in-Chief: Vaclav Skala
c/o University of West Bohemia
Univerzitni 8
CZ 306 14 Plzen
Czech Republic
skala@kiv.zcu.cz <http://www.VaclavSkala.eu>

Managing Editor: Vaclav Skala

Publisher & Author Service Department & Distribution:
Vaclav Skala - UNION Agency
Na Mazinach 9
CZ 322 00 Plzen
Czech Republic
Reg.No. (ICO) 416 82 459

ISSN 2464-4617 (Print)
ISBN 978-80-86943-67-1 (Print)

ISSN 2464-4625 (CD/DVD)
ISBN 978-80-86943-63-3 (CD/-ROM)

WSCG 2015

International Program Committee

Andrew, Glassner (United States)	Patow, Gustavo (Spain)
Baranoski, Gladimir (Canada)	Pedrini, Helio (Brazil)
Benes, Bedrich (United States)	Platis, Nikos (Greece)
Benger, Werner (Austria)	Renaud, Christophe (France)
Bengtsson, Ewert (Sweden)	Richardson, John (United States)
Bourke, Paul (Australia)	Rojas-Sola, Jose Ignacio (Spain)
Dachsbacher, Carsten (Germany)	Ruyam, Acar (Turkey)
Daniel, Marc (France)	Segura, Rafael (Spain)
Daniels, Karen (United States)	Semwal, Sudhanshu (United States)
Debelov, Victor (Russia)	Schultz, Thomas (Germany)
Feito, Francisco (Spain)	Schulz, Hans-Jorg (Germany)
Ferguson, Stuart (United Kingdom)	Sousa, A. Augusto (Portugal)
Gavrilova, Marina (Canada)	Stroud, Ian (Switzerland)
Guthe, Michael (Germany)	Szececi, Laszlo (Hungary)
Jung, Soon Ki (Korea)	Teschner, Matthias (Germany)
Kalra, Prem K. (India)	Tevfik, Akgun (Turkey)
Klosowski, James (United States)	Tokuta, Alade (United States)
Kraus, Martin (Denmark)	Ugur, Gudukbay (Turkey)
Linsen, Lars (Germany)	Wu, Shin-Ting (Brazil)
Lu, Aidong (United States)	Wuensche, Burkhard, C. (New Zealand)
Mark, Finch (United States)	Wuethrich, Charles (Germany)
Molla, Ramon (Spain)	Zemcik, Pavel (Czech Republic)
Muller, Heinrich (Germany)	Zwettler, Gerald (Austria)
Murtagh, Fionn (United Kingdom)	
Oyarzun Laura, Cristina (Germany)	
Pan, Rongjiang (China)	
Paquette, Eric (Canada)	

Board of Reviewers

Agathos, Alexander (Greece)	Horain, Patrick (France)
Aires, Kelson (Brazil)	Hu, Xianlin (United States)
Aliaga-Badal, Carlos (Spain)	Hua, Binh-Son (Singapore)
Apolinario Junior, Antonio Lopes (Brazil)	Chajdas, Matthaeus (Germany)
Assarsson, Ulf (Sweden)	Chen, Ding (Japan)
Ayala, Dolors (Spain)	Chen, Weiya (France)
Bae, Juhee (United States)	Iwasaki, Kei (Japan)
Birra, Fernando (Portugal)	Jarabo, Adrian (Spain)
Bourke, Paul (Australia)	Jeschke, Stefan (Austria)
Brandao, Andre (Brazil)	Jones, Mark (United Kingdom)
Bucak, Serhat (United States)	Jones, Ben (United States)
Cakmak, Hueseyin Kemal (Germany)	Jung, Soon Ki (Korea)
Carozza, Ludovico (United Kingdom)	Kahler, Olaf (United Kingdom)
Cline, David (United States)	Kasprzak, Wlodzimierz (Poland)
Didandeh, Arman (Canada)	Kerkeni, asma (Tunisia)
Djado, Khalid (Canada)	Klosowski, James (United States)
dos Santos, Jefersson Alex (Brazil)	Kolcun, Alexej (Czech Republic)
Drechsler, Klaus (Germany)	Kraus, Martin (Denmark)
Durikovic, Roman (Slovakia)	Kriglstein, Simone (Austria)
Eisemann, Martin (Germany)	Kumar, Subodh (India)
El Shafey, Laurent (Switzerland)	Kurillo, Gregorij (United States)
Emile, Bruno (France)	Kurt, Murat (Turkey)
Fabio, Pellacini (Italy)	Lange, Benoit (France)
Facon, Jacques (Brazil)	Last, Mubbasir (United States)
Frejlichowski, Dariusz (Poland)	Lee, Jong Kwan (United States)
Fuenfzig, Christoph (Germany)	Lee, YT (Singapore)
Galo, Mauricio (Brazil)	Leite, Neucimar (Brazil)
Gao, Zhi (Singapore)	Leon, Jean-Claude (France)
Garcia Hernandez, Ruben Jesus (Germany)	Lessig, Christian (Germany)
Garcia-Alonso, Alejandro (Spain)	Li, Bo (United States)
Gobron, Stephane (Switzerland)	Lin, Yuewei (United States)
Gois, Joao Paulo (Brazil)	Linsen, Lars (Germany)
Gomez-Nieto, Erick (Brazil)	Little, James (Canada)
Griffin , Amy (Australia)	Livesu, Marco (Italy)
Grottel, Sebastian (Germany)	Loscov, Celine (France)
Hast, Anders (Sweden)	Lu, Aidong (United States)
Hernandez, Benjamin (United States)	Maciel, Anderson (Brazil)
Hinkenjann, Andre (Germany)	Mantiuk, Radoslaw (Poland)
Hitomi, Yasunobu (Japan)	Marques, Ricardo (France)
Hlawatsch, Marcel (Germany)	Masia, Belen (Spain)
	Meiguins, Bianchi (Brazil)

Meng, Weiliang (China)
Menotti, David (Brazil)
Mestre, Daniel,R. (France)
Meyer, Alexandre (France)
Michael, Despina (Cyprus)
Michels, Dominik (United States)
Monti, Marina (Italy)
Montrucchio, Bartolomeo (Italy)
Movania, Muhammad Mobeen (Pakistan)
Mukai, Tomohiko (Japan)
Mura, Claudio (Switzerland)
Nagai, Yukie (Japan)
Nah, Jae-Ho (Korea)
Nanni, Loris (Italy)
Nogueira, Keiller (Brazil)
Nurzynska, Karolina (Poland)
Nyul, Laszlo (Hungary)
Oliveira, Joao Fradinho (Portugal)
Oztimur Karadag, Ozge (Turkey)
Paiva, Jose Gustavo (Brazil)
Parsons, Paul (Canada)
Patane, Giuseppe (Italy)
Paul, Padma Polash (Canada)
Peethambaran, Jiju (India)
Penedo, Manuel (Spain)
Pina, Jose Luis (Spain)
Pobegailo, Alexander (Belarus)
Puig, Anna (Spain)
Ramos, Sebastian (Germany)
Rasool, Shahzad (Singapore)
Reddy, Pradyumna (India)
Rehfeld, Stephan (Germany)
Rind, Alexander (Austria)
Rupprecht, Christian (Germany)
Sadlo, Filip (Germany)
Saito, Shunsuke (United States)
Santagati, Cettina (Italy)
Saraiji, MHD Yamen (Japan)
Saru, Dhir (India)
Seipel, Stefan (Sweden)
Shesh, Amit (United States)
Shi, Xin (China)
Shimshoni, Ilan (Israel)
Schaefer, Gerald (United Kingdom)

Schmidt, Johanna (Austria)
Schultz, Thomas (Germany)
Schwarz, Michael (Switzerland)
Silva, Romuere (Brazil)
Silva, Samuel (Portugal)
Singh, Rajiv (India)
Solis, Ana Luisa (Mexico)
Soriano, Aurea (Brazil)
Souza e Silva, Lucas (Brazil)
Spiclin, Ziga (Slovenia)
Svoboda, Tomas (Czech Republic)
Tavares, Joao Manuel (Portugal)
Teixeira, Raoni (Brazil)
Theussl, Thomas (Saudi Arabia)
Tomas Sanahuja, Josep Maria (Mexico)
Torrens, Francisco (Spain)
Tytkowski, Krzysztof (Poland)
Umlauf, Georg (Germany)
Vasseur, Pascal (France)
Vazquez, David (Spain)
Veras, Rodrigo (Brazil)
Walczak, Krzysztof (Poland)
Wanat, Robert (United Kingdom)
Wang, Lili (China)
Wang, Ruizhe (United States)
Wang, Lisheng (China)
Wenger, Rephael (United States)
Wijewickrema, Sudanthi (Australia)
Wu, YuTing (Taiwan)
Wu, Jieting (United States)
Wuensch, Burkhard,C. (New Zealand)
Xiong, Ying (United States)
Xu, Tianchen (Hong Kong SAR)
Xu, Chang (China)
Yang, Shuang (China)
Yasmin, Shamima (United States)
Yoshizawa, Shin (Japan)
Yu, Hongfeng (United States)
Zheng, Jianping (United States)
Zhong, Li (China)

WSCG 2015

Posters Proceedings

Contents

	Page
Gashnikov,M., Glumov,N.: Hyperspectral images repository using a hierarchical compression	1
Thomsen,K., Kraus,M.: Simulating Small-Scale Object Stacking Using Stack Stability	5
Maher,J., Mohamed,J.: Removed paper	9
Ganguly,S., Bhattacharjee,D., Nasipuri,M.: Efficient Representation of Range Face Images Using Vectorfaces	15
Novotortsev,L., Voloboy,A.: Automated Detection of Buildings on Aero Images	23
Taertulakarn,S., Tosranon,P., Pintavirooj,C.: Using Intrinsic Surface Geometry Invariant for 3D Ear Alignment	29
Kim,E.S., Choi,S.I., Park,S.Y.: A new 6D ICP algorithm with color segmentation-based adaptive sampling	35
Maas,S., Overhoff,H.M.: A COM-based Toolkit for Real Time Volume Visualization	41
Petrova,Y.: Virtual Museum as an Environment for Visual Representation of Urban Planning Concepts and Lost Architectural Objects	45
Chadimova,L.: The creation of serious games intended for historically oriented subjects at the 1st level of at the 1st level of primary school and such as a part of exhibition chosen historical buildings	49
Mitaritonna,A., Abásolo,M.: Improving Situational Awareness in Military Operations using Augmented Reality	53
Elbahi,A., Omri,M.N.: Conditional Random Fields For Web User Task Recognition Based On Human Computer Interaction	59
Kopenkov,V., Myasnikov,V.: Detection and tracking of vehicles based on the videoregistration information	65
Choi,J., Kim,Y., Choe,Y.: A Novel Retinex Model Based on Sparse Source Separation	69
Gordeev,D.: Reasoning about Graph Algorithm Visualization	75
Bilinskas,M.J., Dzemyda,G., Trakymas,M.: Computed Tomography Image Analysis: the Model of Ribs-Bounded Contour	81
Marek,J., Rak,J., Jetensky,P.: Statistical solution of 3D transformation problem	85
Perechesova,A.D., Soloveva,G.A., Kalapyshina,I.I.: Hough Transform for the Calculation of Twist Angle of Aramid Torsion	91

Priyadarshi,S., Jain,P.K., Roy,J.J., Samal,M.K., Roy,D., Tandon,P.: Feature based assessment of forming force behavior in Incremental Sheet Forming	95
Jaszuk,M., Szostek,G., Starzyk,J.A.: Building 3D Object Representation Using SURF Local Features	105
Ferrer,C., Taboada-Crispi,A.; Lorenzo-Ginori,J.V.: Signal and Image Processing in the Center of Cuba: Center for Studies on Electronics and Information Technologies (CEETI)	113
De Giusti,A., Abásolo,J.M, Naiouf,M., Castro,S., Guerrero,R.: Computer Graphics and Vision Labs in Argentina	117

Hyperspectral images repository using a hierarchical compression

Gashnikov M.V.

Samara State Aerospace University
mgash@smr.ru

Glumov N.I.

Samara State Aerospace University
nglu@smr.ru

ABSTRACT

The possibilities of hierarchical compression in hyperspectral images repository are investigated. The image analysis of «SpecTIR» and «AVIRIS» hyperspectrometers is carried out. In order to increase the compression ratio, the spectral bands approximation algorithms are proposed to provide fast access to individual bands. The effectiveness of the developed algorithms is investigated through computational experiments using real 16-bit hyperspectral images.

Keywords

Hierarchical compression, image approximation, hyperspectral image, image repository, maximum deviation, spectral band.

1. INTRODUCTION

An important problem of developing a hyperspectral [Che13a], [Bor04a] images repository is the large number of spectral bands that reaches several hundreds. For example, the results of «SpecTIR» hyperspectrometer shooting [Spe14a] contain more than 300 spectral bands, the results of «AVIRIS» shooting [Avi13a] contain more than 200 bands. The bit rate of each spectral band is 16 bits.

The result is a non-trivial task to store such images in the repository: an extremely large amount of hyperspectral data entails excessive demands on storage capacity and an unacceptably low rate of access to these data. Therefore, when creating a hyperspectral images repository, a compression should be applied.

In this paper, for the compression of hyperspectral images in the repository, we suggest using a modified method, based on a hierarchical grid [Gas10a] interpolation (HGI). This modification includes new algorithms for spectral bands approximation that reduce access time to the individual bands.

In addition, in this paper we present the results of computational experiments to investigate the effectiveness of the proposed algorithms on real 16-bit hyperspectral images.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

2. RECENT SOLUTIONS

To store hyperspectral images one of these data formats is commonly used: hdf4, hdf5, lan, img+hdr, dat+hdr [Che13a], [Avi13a], [Spe14a]. These formats are not sufficiently effective for hyperspectral images repository, as do not support data compression.

An important feature of hyperspectral images is that common compression methods [Sal07a] for them are ineffective or even inapplicable. Transform coding techniques (primarily JPEG [Wal91a] and JPEG-2000 [Mar00a]), fractal methods [Sal07a] and video compression methods (including MPEG methods) [Has97a] are computationally too complex and do not allow a strict control of compression error (in particular, the maximum deviation, see below), which must be in the organization of a repository of unique hyperspectral data. Differential methods and 1-D approximation methods have not sufficiently high compression ratio even with compression of one-component images [Sal07a]. 2-D approximation methods based on segmentation [Pra07a] are overly complicated.

In [Gas14a] the HGI-method [Gas10a] was proposed for the compression of hyperspectral images and the results of numerical experiments demonstrating the benefits of HGI over JPEG method in this situation are given.

However, proposed in the above-mentioned article version of HGI method can not be used to store images in hyperspectral images repository, because it does not provide access to the individual spectral bands without decompressing a large part of the compressed image.

3. PROPOSED SOLUTION

In order to formulate the requirements for the compression method in the repository, analysis of the hyperspectral image characteristics was performed in this paper.

3.1. Analysis of hyperspectral images

Analysis of hyperspectral images features was carried out by the following data publicly available:

- five image of spectrometer «SpecTIR» [Spe14a] (360 bands of 16-bit pixels 320x600 size);
- five image of spectrometer «AVIRIS» [Avi13a] (224 bands of 16-bit pixels 614x1086 size).

Examples of spectral bands of these images are shown in Fig. 1.

Statistical characteristics of spectral bands were evaluated for analysis of the hyperspectral images features. Features of bands interdependencies were also evaluated. Some results are shown in the result section in Fig. 4-5. The following conclusions were formulated based on the analysis of these results.

- The difference between the maximum and minimum of the brightness gradations reaches thousands or tens of thousands. These images can not be converted to a "byte".
- Spectral bands are strongly interdependent, as the correlation [Pra07a] between the bands is extremely high.
- Most spectral bands have a high correlation within the band. It can be expected that the use of compression will have a significant effect.

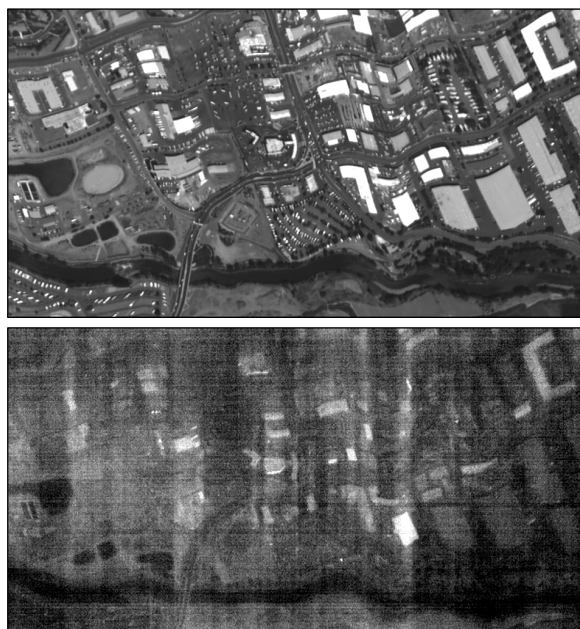


Figure 1. Spectral bands №27, №269 of image «Urban and Mixed Environment» of the spectrometer «SpecTIR»

3.2. Compression based on HGI

The method HGI [Gas10a] is proposed for hyperspectral image compression in [Gas14a]. The method uses the decimated image for interpolation of «less decimated» image. Interpolated pixels are subtracted from original pixels; the calculated differences are encoded to reduce data size.

The method provides control of the maximum deviation. Moreover, access time to an image fragment does not depend on the fragment scale, as the block-hierarchical [Gas10a] format of compressed data used in this method.

3.3. Independent bands compression based on HGI

Independent processing of spectral bands is the easiest way to use the HGI-method to compress hyperspectral images. This approach is very ineffective, since it does not use the correlation between the spectral bands.

3.4. Compression based on a «sliding approximation of bands»

To use the correlations between the spectral bands, the approximation of the spectral bands is proposed for use. Each spectral band is approximated on the basis of other bands that have already been compressed and restored. High correlation between the bands should provide good approximation accuracy. Compression of the difference between the original and approximated spectral bands instead of compression of the original spectral band should significantly improve the compression ratio.

Let X^s be a spectral band number s (matrix of numbers). Let S be a number of spectral bands X^s in the hyperspectral image $\{X^s, 0 \leq s < S\}$.

We will sequentially compress the spectral bands with the numbers $0, 1, 2 \dots (S-1)$. When compressing each spectral band X^s we will first calculate approximating band:

$$\hat{X}^s = \sum_{i=0}^{N-1} k_i \overline{X}^{s-i-1}, 0 \leq s < S,$$

where $\overline{X}^i, i \geq 0$ are previous spectral bands that have already been compressed and decompressed,

N is the number of previous spectral bands that are used for approximation (algorithm parameter), $\{k_i, 0 \leq i < N\}$ are approximation coefficients, which are calculated on the basis of minimizing the mean square error of approximation. At the same time control of the maximum deviation is provided.

Thus, a set of base bands at the approximation is a «sliding window» in the spectral dimension.

3.5. Compression based on the «non-overlapping sets of bands»

The above-described compression algorithm based on the «sliding approximation of bands» is not suitable for storage of hyperspectral images in the repository. The reason is the sequential compression of the spectral bands, wherein the previous bands are used for approximation of the next bands. As a result for decompression of any spectral band all previous bands should be decompressed. This is a serious obstruction in providing quick access to images.

In the repository decompression of any band should entail decompression as small as possible number of other bands, which may not be needed. To this end, we propose to use an approximation of the spectral bands within the «non-overlapping sets of bands».

While compressing the entire set of spectral bands is divided into «non-overlapping sets of bands». Each set contains N bands, see Fig. 2. Within each set the described algorithm of the «sliding approximation of bands» is used. Then for decompression of any spectral band we do not need to decompress all previous bands. Only previous bands of the corresponding set must be decompressed.

3.6. Compression based on the «shared base bands»

The compression ratio of the algorithm based on the «non-overlapping sets of bands» is significantly smaller than the compression ratio of the algorithm based on «sliding approximation of bands». The reason is that a much smaller number of base bands used in the algorithm based on the «non-overlapping sets of bands» for approximation.

To solve this problem, the approximation algorithm based on «shared base bands» (see Fig. 3) is proposed in this paper. «Shared base bands» are selected from the complete set of spectral bands with the step N . Also integer parameter C is specified.

Shared bands are compressed by the algorithm based on the «sliding approximation of bands». The number of base bands used for the approximation is equal to $N + C - 2$. Then, after the compression of shared bands, «other» bands are also compressed. Set of the «other» spectral bands is separated into subsets

$$\mathbf{X}_m = \{X^s, mN < s < (m+1)N\}, 0 \leq m \leq S/N,$$

located between the shared bands.

Each subset \mathbf{X}_m is compressed by the algorithm based on the «sliding approximation of bands». As the base bands for approximation we use the bands of the same subset \mathbf{X}_m and the nearest of the shared bands. The total number of bands used for the approximation is also equal to $N + C - 2$.

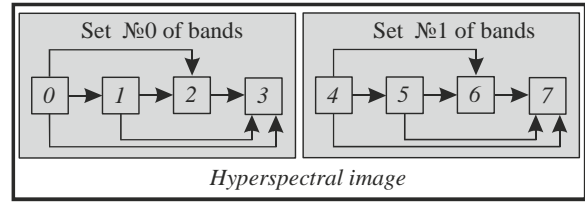


Figure 2. Bands approximation algorithm based on the «non-overlapping sets of bands» ($N = 4$)

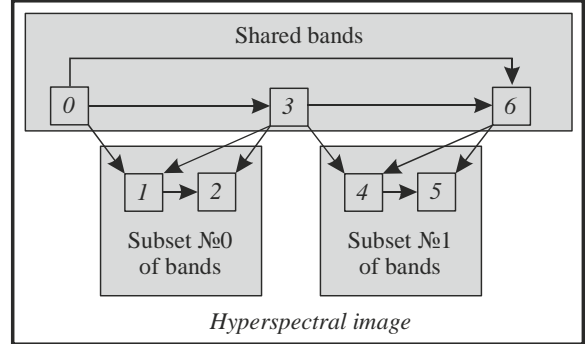


Figure 3. Bands approximation algorithm based on the «shared bands» ($N = 3, C = 1$)

4. EXPERIMENTAL RESULTS

4.1. Statistical characteristics of hyperspectral images

Statistical characteristics of spectral bands were evaluated for analysis of the hyperspectral images features. Some results are shown in Fig. 4-5.

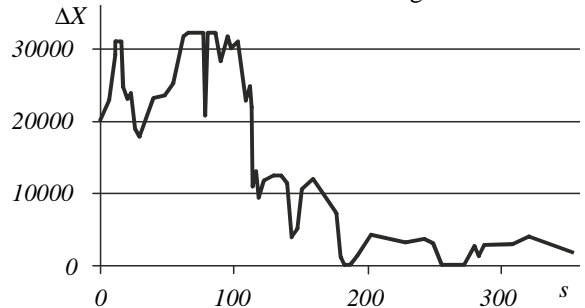


Figure 4. The difference ΔX between the maximum and the minimum depending on the band number s («SpecTIR»)

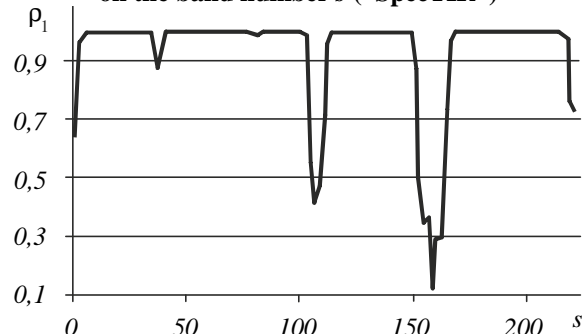


Figure 5. Evaluation of the correlation coefficient ρ_1 between the neighboring bands, depending on the number of band s («AVIRIS»)

4.2. Efficiency of the HGI-compression in image repository

The dependency of compression ratio from error was calculated to evaluate the effectiveness of compression (see Fig. 6). Computational experiment was conducted on real 16-bit images of hyperspectrometers «SpecTIR» and «AVIRIS».

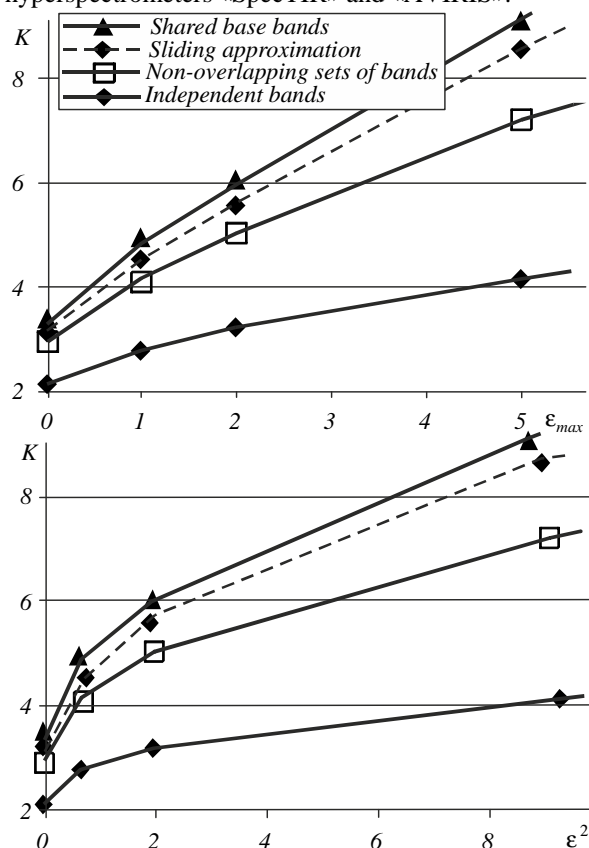


Figure 6. Averaged over five «SpecTIR» images compression ratio K depending on the maximum deviation ϵ_{\max} and MSE ϵ^2 ($L = 5, N = 7, C = 4, V_b \times H_b = 112 \times 614$)

The following conclusions were formulated on the basis of the obtained results.

1. All the developed algorithms have a high compression ratio. Therefore, these algorithms can be effectively used in systems of hyperspectral images storage.
2. Algorithms based on the "non-overlapping sets of bands" and "shared bands" provide quick access to compressed data. These algorithms can be effectively used in hyperspectral image repository.
3. The use of any bands approximation can significantly improve the compression ratio.
4. The advantage of using the bands approximation increases with compression ratio increasing.
5. In a small error algorithms based on the "sliding approximation of bands" and "shared bands"

demonstrate better results. The algorithm based on "shared bands" is preferred as it provides fast access to data.

5. CONCLUSION

New algorithms for approximate of the spectral bands developed for the compression method based on HGI. These algorithms are well adapted for use in the hyperspectral images repository. The effectiveness of the HGI-compression evaluated on real images of 16-bit hyperspectrometers. The prospects for using the HGI-compression in the hyperspectral images repository demonstrated.

6. ACKNOWLEDGEMENTS

This work was financially supported by the Russian Scientific Foundation (RSF), grant no. 14-31-00014 "Establishment of a Laboratory of Advanced Technology for Earth Remote Sensing".

7. REFERENCES

- [Avi13a] AVIRIS Data – Ordering Free AVIRIS Standard Data Products. Jet Propulsion Laboratory: http://aviris.jpl.nasa.gov/data/free_data.html
- [Che13a] Chein-I, C. Hyperspectral Data Processing: Algorithm Design and Analysis. Wiley Press, 2013.
- [Bor04a] Borengasser, M. Hyperspectral Remote Sensing – Principles and Applications. CRC Press, 128 p, 2004.
- [Gas14a] Gashnikov, M.V., Glumov, N.I. Hierarchical grid interpolation for hyperspectral image compression. Computer optics, V. 38, No.1, pp. 87-93, 2014. (In Russian)
- [Gas10a] Gashnikov, M.V., Glumov, N.I., Sergeev, V.V. A hierarchical compression method for space images. Automation and Remote Control, V. 71, No.3, pp. 501-513, 2010.
- [Has97a] Haskel, B., Puri, A., Netravaly A., Digital Video An Introduction to MPEG-2. NY: Chapman&Hall, 1997.
- [Mar00a] Marcellin, M., Gormish, M., Bilgin A., Boliek, M. An Overview of JPEG-2000. Proceedings of the 2000 IEEE Data Compression Conference, Snowbird, Utah, pp. 523-541, 2000.
- [Pra07a] Pratt, W. Digital image processing. Wiley, 4ed, 2007.
- [Sal07a] Salomon, D. Data Compression. The Complete Reference. Springer-Verlag, 4ed, 2007.
- [Spe14a] SpecTIR Data – Advanced Hyperspectral and Geospatial Solutions. Corporate Headquarters SpecTIR Remote Sensing Division: <http://www.spectir.com/free-data-samples>.
- [Wal91a] Wallace, G. The JPEG Still Picture Compression Standard. Communications of the ACM. V. 34, No. 4, pp. 30-44, 1991.

Simulating Small-Scale Object Stacking Using Stack Stability

Kasper Kronborg Thomsen

Aalborg University

Department AD:MT

Rendsburggade 14

Denmark, 9000 Aalborg

kasperkronborgthomsen@gmail.com

Martin Kraus

Aalborg University

Department AD:MT

Rendsburggade 14

Denmark, 9000 Aalborg

martin@create.aau.dk

ABSTRACT

This paper presents an extension system to a closed-source, real-time physics engine for improving structured stacking behavior with small-scale objects such as wooden toy bricks. The proposed system was implemented and evaluated. The tests showed that the system is able to simulate several common stacking scenarios, which the base physics engine cannot simulate.

Keywords

Simulation, object stacking, engine extension, game physics.

1 INTRODUCTION

The real-time simulation of dense structure stacking in game physics engines is a challenging research topic with many applications in video games and virtual reality simulators. In this paper, stacking refers to dense structured stacking as seen when creating deliberate structures. Our main focus is on the simulation of stacking of small-scale objects, which is particularly challenging because gravitational acceleration is relatively large compared to the size of these objects. To achieve a real-time simulation of such stacking scenarios, we propose a correction system for a closed-source physics engine, namely the Nvidia physics engine that is integrated in the Unity 4 game engine.

2 PREVIOUS WORK

A widely cited paper on stacking objects in a physics engine was presented by Erleben [Erleben, 2007]. His approach uses a standard collision detection algorithm but includes its own complementarity formulation, solver and error correction algorithms. The paper discusses an error correction scheme that is partially based on stack layers derived from the contacts of objects. The method proposed by Erleben was compared to several other physics engines and performs better on all

stacking scenarios while still running at near real-time speeds. His publication showed that using the structure of a stack has to some extent been considered before and is a viable approach.

A slightly more recent paper by Kaufman et al. [Kaufman et al., 2008] focuses on accurately simulating friction and enabling friction-dependent behavior. Their method is usable in both rigid and deformable bodies' simulation. Although not the main focus, their system can simulate stacking scenarios that are friction-dependent in near real time. However, it does not construct stable structures during simulation. This paper demonstrates that simulating correct friction behavior supports a system's ability to simulate more stacking scenarios, in particular a cardhouse stack. This scenario is interesting because it is mainly dependent on friction to remain stable.

Hsu and Keyser [Hsu and Keyser, 2010] describe an approach to improve the performance of random stacks in simulation. The contribution of this paper is an advanced object sleep paradigm. By forcing objects into an object sleep state, depending on pile specific conditions, more realistic pile behavior is achieved. The algorithm runs at real time and can even provide slight performance improvements. Although the authors have implemented the algorithm inside a physics engine, this approach should be able to function independently of the underlying physics engine. The approach is similar to the one describe in the present paper. The system proposed by Hsu and Keyser is, however, only meant to simulate random stacks.

A more recent paper by the same authors [Hsu and Keyser, 2012] presents a method to sta-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

belise stacks by adding constraints between objects based on local equilibrium. The algorithm periodically revises the grouping and removes objects from groups if objects are no longer in equilibrium or are impacted by fast objects. The system is designed to allow for art-directed stacking, but not necessarily realistic stacking; however, the system cannot create stacks that are massively unstable. It was tested on both structured stack and random stack scenarios. The system is able to maintain stacks created before simulation, but it is not clear if the system can handle stacks that are created during simulation as this scenario was not shown. The system runs at real-time to near real-time speed. It is the closest prior work to what is proposed in the present paper. However, the system uses the Bullet physics library, an open source engine, thus the authors have access to the source code even though it is claimed to work as a general extension. The method presented by the present paper uses a closed-source engine; thus, only API calls are possible. The system described by Hsu and Keyser [Hsu and Keyser, 2012] uses grouping of objects as the main method of generating stable stacks whereas the method proposed by the present paper uses object sleep state to create stacks. The focus of Hsu and Keyser [Hsu and Keyser, 2012] is artistic control over stacks, whereas the focus of the present paper is stack simulations under user control.

A paper by Han et al. [Han et al., 2013] sets out to test two commonly used hypotheses in physics animation. The first is that users are unable to perceive distortions in the simulation that are due to approximated simulation methods. The second hypothesis is that freezing transformations of objects in a random pile does not affect the visual plausibility of a simulation. Both hypotheses are confirmed in a user study. This paper represents one of the rare instances where alternate simulation methods are tested on users for visual plausibility.

One common feature of all the papers proposing new simulation methods is that these are not demonstrated to work on small object scale while our system is designed specifically for small scales as they are characteristic for toy bricks.

3 METHOD

As the proposed method is an extension of a physics engine, objects will start under the control of the physics engine. If an object is detected to be in a stable configuration, the object stability system will freeze it in the physics engine and take over stability calculations. If unstable configurations are detected, the object stability system will wake up the object and the physics engine will then simulate the object until it again returns to a stable configuration.

For every frame, new contacts must be found and any potential new additions to the stack of frozen objects

must be tested, the state of the stack must be updated and the overall stability of the stack must be tested. Firstly, all contacts are added to the data structure of the system, then all contacts between two objects that are both in unstable positions are removed as they cannot be resting on a part of the stack. Contacts coming from impacts above a certain speed are also disregarded. Contact areas between stable and potentially stable objects are then calculated based on the remaining contact points. Using these contact areas potential objects are tested for stability, and if they are in a stable configuration, they are added to the stack. The system then updates stability data through the stack. Lastly, the stack is tested for overall stability. This is done by calculating the centre of mass for all objects above each object, and testing if the vertical projection of this centre of mass is outside of the main contact area of the object. If it is then the object and everything above it is woken up and flagged as unstable.

The contact area between two objects is found by generating the convex hull of the projections of the contact points between the two objects onto the ground plane. To test the stability of a single object, a vertical ray is cast from the centre of mass of the object. If the ray hits the contact area, the object is considered stable.

For more than one object stacked above another object, different types of stack branching have to be considered. The simplest form of branching is a single branch, where objects are stacked on top of each other one at a time, thus only one object rests on another object. To calculate the stability of this stack, objects above a given object must be taken into consideration when calculating stability. To this end, the centre of mass of an object and all objects stacked on top of it is defined as C_{stack} , which is calculated using Equation 1 where C_{obj} is the centre of mass for the object, m_{obj} is the mass of the object, $C_{\text{stack,above}}$ is C_{stack} for the stack above the current object and $m_{\text{stack,above}}$ is the mass of the stack above the current object. Analogously, m_{stack} is the mass of the stack calculated by Equation 2.

$$C_{\text{stack}} = \frac{(C_{\text{obj}}m_{\text{obj}}) + (C_{\text{stack,above}}m_{\text{stack,above}})}{m_{\text{obj}} + m_{\text{stack,above}}} \quad (1)$$

$$m_{\text{stack}} = m_{\text{obj}} + m_{\text{stack,above}} \quad (2)$$

Two types of multi-branch stacking are distinguished. One is split branching, where multiple objects rest on one object. This case is handled similarly to the single branch case, except that Equations 1 and 2 are used for every object resting on the current object. The second type of branching is merge branching. In this case an object rests on more than one object. For a more clear explanation, the object above the one being updated will be denoted object A. For merge branching

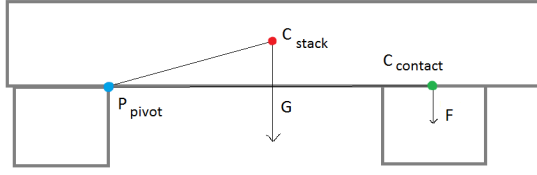


Figure 1: The two object merge-branching case. The bottom right object is being updated.

there are three different cases. The first case is that the entire weight of the stack and object A can be supported by a single object. This is true if the C_{stack} of object A is within one of the individual contact areas. This is tested similarly to the single object stability, but using the individual contact areas instead of the combined contact area. If this is true, the entire weight of object A will be on this supporting object. If the supporting object is the object being updated, it will be updated similarly to the single case. If it is not, no update of data will be made (if there are also other objects resting on the object, these will update normally). The second case is that object A is resting on more than two objects. In this case the update is done similarly to the single branch case but the mass m_{stack} of object A for the calculations is reduced to a fraction matching the number of objects it rests on.

The last case is that object A is resting on two objects. In this case, the distribution between the two supports is calculated. The calculation is based on the setup depicted in Figure 1. The assumption is that gravity would cause a torque around the pivot point at the other support object as if the supporting object that is being updated was not present. The equivalent force of this torque would be exerted on the supporting object that is being updated and can be recalculated as a weight. This assumption goes both ways, and generates the distribution of mass.

The pivot point is the contact point closest to C_{stack} . The torque exerted on this point by gravity is calculated using Equation 3.

$$\mathbf{T} = (\mathbf{C}_{stack} - \mathbf{P}_{pivot}) \times \mathbf{G} \quad (3)$$

Where \mathbf{T} is the torque, \mathbf{P}_{pivot} is the pivot point, and \mathbf{G} is the gravitational force.

From the torque a force \mathbf{F} on the other contact area can be calculated. Using Equation 4, and rearranging it, the downwards force equating the torque can be calculated, and the equation used in the implementation is shown in Equation 5.

$$|\mathbf{T}| = |\mathbf{C}_{contact} - \mathbf{P}_{pivot}| |\mathbf{F}| \sin \theta \quad (4)$$

Where θ is the angle between the lever arm and the force.

$$|\mathbf{F}| = \frac{|\mathbf{T}|}{|\mathbf{C}_{contact} - \mathbf{P}_{pivot}| \sin \theta} \quad (5)$$

The direction of the force is parallel to the gravity vector. The equivalent mass, m_{equiv} , is calculated using the equation for the force of gravity, i.e., $m_{equiv} = |\mathbf{F}|/g$.

The object can then be updated using the same base formula as in the single object case Equation 1 but using the equivalent mass instead of the $m_{stack,above}$.

The algorithm for leaning stability uses some of the same principles as the weight distribution function seen before, but friction calculations are included as well. First, the system identifies the number of objects in leaning contact. If the number is one or less, leaning stability is not possible. If the number is over 2 the system also returns false, as this scenario is too complicated for the current model to handle and this case was not observed in the test how users stack with toy bricks. Flat stacking is tested first, so if a object can be stable using the above version, more than two contact bricks are possible. The system handles the case where an object is in contact with two other bricks (or ground objects), and at least one of these is part of a stack.

First the lowest contact area (blue dot in Figure 2) is found. The next step is to find the contact point, on the horizontal plane, closest to the centre of mass of the object. This point will function as the pivot point. The torque from gravity is calculated in the same way that it is calculated in the weight distribution function previously described with Equation 3. To find the direction in which the torque acts on the second contact (green in Figure 2), the cross product between the torque vector and the line from the pivot to the contact point is normalized. Using an approach similar to the one in Equation 5 the magnitude of this torque-induced force can now be computed. By multiplying the magnitude of the torque induced force with the force vector found before, the force vector is determined. In order to calculate the amount of friction supporting the object from the contact point, the force perpendicular to the contact surface must be calculated. This force is found by projecting the force vector onto the inverted normal of the contact area. Using this normal force, the Coulomb friction can be calculated using Equation 6 [Erleben et al., 2005].

$$|\mathbf{F}_{friction}| = \mu |\mathbf{F}_{normal}| \quad (6)$$

Next the sliding force, the force the object wants to move with, in the same direction as the friction force, must be calculated in order to find the direction for the friction force. Usually this would be done with a projection, but as the normal force has already been calculated, the sliding force can be found by subtracting the normal force from the force the object is generating on the contact. Now the force on the pivot point can

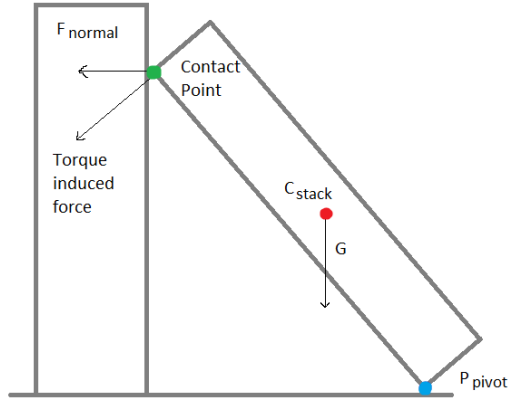


Figure 2: Leaning object

be calculated using Equation 7. The normal force is inverted as this is the force the other object is assumed to be exerting in a stable configuration to avoid object penetration. It is assumed that the friction at the contact point holds as much weight of the object as the friction force allows.

$$\mathbf{F}_{\text{pivot}} = -\mathbf{F}_{\text{normal}} + \mathbf{G} - \mathbf{F}_{\text{friction}} \quad (7)$$

Where $\mathbf{F}_{\text{pivot}}$ is the total force on the pivot point, $\mathbf{F}_{\text{normal}}$ is the normal force on the contact point and $\mathbf{F}_{\text{friction}}$ is the friction force on the contact point. \mathbf{G} is the gravitational force on the pivot point.

The normal force for the pivot point then can be found in the same way as before, the friction and sliding force as well. If the friction force is larger than the sliding force, then the object is considered stable.

4 RESULTS

The tests were designed to investigate the capabilities of the proposed system compared to the base physics system. The tests included three stacking scenarios (see Figure 3) which are based on common stacking behavior observed in a preliminary user test. The simulations were compared to each other and to recordings of the



Figure 3: The three stack types, from left to right: 4-stack, card house and leaning tower.

behaviour of real bricks in similar stacking configurations. The simulation stacks were constructed during the simulation, with one brick being added to the stack at approximately 1 second intervals until the structure is completed or collapses. For better comparison with real-world objects, KAPLA wooden bricks were used as the base for all scenarios.

The results of testing the 4-stack structure showed that the proposed system is capable of generating towers over 25 levels high. The base physics system is only capable of getting this structure to a height of four levels high.

The card house stack can be simulated by the proposed system and behaves stable. The base physics system is not able to generate a stable inverted V which is the base of the structure.

As the leaning tower scenario is to some extent similar to the 4-stack, the base physics engine is not able to simulate the scenario. The proposed system makes the tower collapse when the structure reaches 9 levels. The collapse appears to occur at the correct time, however, it is too vertical since real towers tend to pivot on the lowest level and remain mostly intact before impact with the ground.

5 CONCLUSION

We presented an extension system to a physics engine, which is able to simulate common stacking scenarios that the base physics engine cannot simulate. These results encourage further research on improvements of the proposed system, in particular related to the simulation of leaning bricks, which is currently not robust and can in some cases allow incorrect stability.

6 REFERENCES

- [Erleben, 2007] Erleben, K. (2007). Velocity-based shock propagation for multibody dynamics animation. *ACM Trans. Graph.*, 26(2):article no. 12.
- [Erleben et al., 2005] Erleben, K., Spöring, J., Henriksen, K., and Dohlmann, H. (2005). *Physics-Based Animation*. Charles River Media.
- [Han et al., 2013] Han, D., Hsu, S.-W., McNamara, A., and Keyser, J. (2013). Believability in simplifications of large scale physically based simulation. *ACM Symposium on Applied Perception*, pages 99 – 106.
- [Hsu and Keyser, 2010] Hsu, S.-W. and Keyser, J. (2010). Piles of objects. *Proc. SIGGRAPH Asia*, 29(6):article no. 155.
- [Hsu and Keyser, 2012] Hsu, S.-W. and Keyser, J. (2012). Automated constraint placement to maintain pile shape. *ACM Trans. Graph.*, 31(6):article no. 150.
- [Kaufman et al., 2008] Kaufman, D. M., Sueda, S., James, D. L., and Pai, D. K. (2008). Staggered projections for frictional contact in multibody systems. *Proc. SIGGRAPH Asia*, 27(5):article no. 164.

This page is empty – paper removed

This page is empty – paper removed

This page is empty – paper removed

This page is empty – paper removed

This page is empty – paper removed

This page is empty – paper removed

Efficient Representation of Range Face Images Using Vectorfaces

Suranjan Ganguly
Computer Science and
Engineering Department,
Jadavpur University
188 Raja S.C. Mullik Road,
Kolkata-700 032
suranjanganguly@gmail.com

Debotosh Bhattacharjee
Computer Science and
Engineering Department,
Jadavpur University
188 Raja S.C. Mullik Road,
Kolkata-700 032
debotoshb@hotmail.com

Mita Nasipuri
Computer Science and
Engineering Department,
Jadavpur University
188 Raja S.C. Mullik Road,
Kolkata-700 032
mnasipuri@cse.jdvu.ac.in

ABSTRACT

Advancement in scientific representation should accelerate the processing of images if it is more relevant and worthy with the experiment. Scientific visualizing of data (here, face images) has an enormous impact on exploring detailed inner content of images. Hence, the quality of processing depends on the quantity and informative data that might be accumulated, preserved as well as visualized in a particular image. In this paper, authors have described a novel technique for representation of range face image by 'Vectorfaces', which is proved to be more effective towards better recognition purpose in terms of recognition rate. Range face image is particularly important for 2D visual images for accomplishing depth data from 3D images. Other than an efficient representation of 'Vectorfaces' images, authors have also emphasized its significance for selecting better features compared to conventional range images. The major goal of the present work reported in this article is to evaluate, visualize and compare the role of 'Vectorfaces' over range face images. Change of tracks for different mathematical notations to visualize the images are noted. Moreover, Mean-Maximum curvature image pair is accumulated from range image as well as 'Vectorfaces' for extraction of features. SVD, followed by a feed-forward backpropagation neural network have been used for recognition purpose. In this work, 3D face images from Frav3D database have been considered. A statistical evaluation of this investigation is also given in the case study section.

Keywords

Scientific representation, 3D face image, range face image, Vectorfaces, Surface extraction, Mean-Maximum curvature, Face recognition.

1. INTRODUCTION

In recent years, 3D images have gained much attention of the researchers for its underlying crucial information. Unlike 2D images, 3D images are used to accumulate data points along X, Y and Z axes. Now, to process 3D images more effectively, it is important to accumulate required information that can be analyzed by quantitative as well qualitative fashion. Before quantitative analysis, it is very much required to know how to process the information that are already belonging to it. Hence, image representation (i.e. qualitative observation) methods are becoming more important to explore the hidden information.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

The idea to visualize the depth-oriented technique is popular and well accepted in terms of range image [1]. Other than range face image, authors have developed a novel face representation technique for depth data, i.e. 'Vectorfaces'. In general, range image preserve normalized depth data. Therefore, Vectorfaces is another approach to visualizing a range face image that exhibits depth data more than 2.5 D range image [2-3].

In practice, it is easy to learn and analyze any image by different appearance and selecting appropriate representation mechanism, which encourages the researchers to apply suitable techniques for further processing. In this article, various representation mechanisms of range image and Vectorfaces illustrate its suitable applications in the domain of computer vision especially in face recognition.

In general, any three-dimensional face image contains the data in the form of points rather than intensity of conventional 2D face image. The depth data (i.e. Z) in X-Y plane renders the valuable

surface information from a face that is not supported by 2D. In addition, due to the presence of depth in the 3D image, the external illumination and shades do not affect its content. Further, facial pose registration for rotated face images by the available data across three axes is an additional feature of 3D scanned face image. In figure 1, the phenomenon of intensity vs. depth data of a randomly selected person from Frav3D database is illustrated.

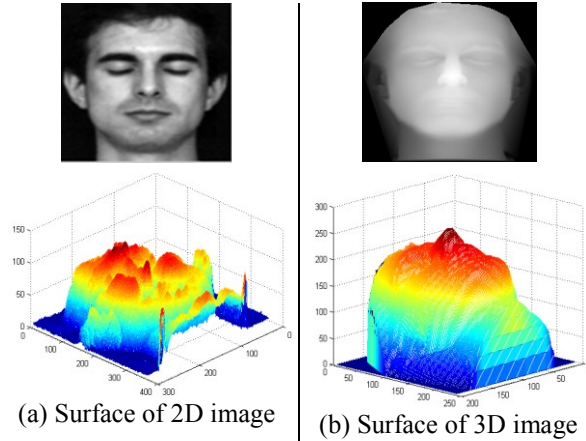


Figure 1. Intensity and Depth Comparison

From figure 1, it is observed that due to the variation of illuminated light source, the values that are having maximum intensity are represented by number of peaks in surface representation form of 2D visual image. Whereas for 3D face image (for frontal face) nose region (especially tip of the nose or 'pronasal') is having maximum depth value having a single peak. For establishing the meaningful comparison, authors have normalized depth data from 3D image and intensity data from the 2D image. The available data from 2D and 3D images have been normalized between 0 to 255. All the color 2D face images have been converted to grayscale by following the standard formula as shown in equation 1.

$$G_r = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B \quad (1)$$

where, G_r is the output gray image and R, G and B are the three color channels of the input color image

However, representation of range face images in Vectorfaces is aimed towards processing the points that are available using the computation of normal vector on facial surface preserved by different depth values. Thus, in Vectorfaces surface of a 3D face is well represented in comparison to corresponding range face images and hence that will certainly improve the recognition task significantly. The principal contributions of this investigation are summarized below:

- A novel face image, 'Vectorfaces', is introduced to describe a face image, by computing normal component and tangent component at each point of a depth image

(especially 2.5D range image). From every pair of normal and tangent components, surface normal is computed that has been again arranged together to form an intermediate matrix, processed by the bi-cubic interpolation. The resultant matrix is Vectorfaces for that input image.

- It uncovers the prime structure and emphasizes on the overall depth information of a 3D image.
- Visual representation of output images obtained from different mathematical formulations are also shown in this paper.
- Other than qualitative analysis by different representation mechanisms, authors have also applied curvature based face recognition technique from both the range and Vectorfaces to accomplish quantitative measurement.

The remaining discussion in this paper is arranged as follows. In section 2, the detailing of 'Vectorfaces' is done. Various representation techniques and their significance have been explained in section 3. A case study for this representation i.e. faces recognition using Vectorfaces has been discussed in section 4. Conclusion and future scope are in section 5.

2. VECTORFACES

On the face surface, ' p ' is a point having two components, namely a tangent component and a normal component to derive surface normal ' V '. The vector ' V ' is the summation of these two components as shown in equation 2.

$$V = V_t + V_n \quad (2)$$

To calculate the normal components for surface normal, unit normals on the face surface, let ' F ', is considered. The unit normal is actually the unit vector (\hat{n}) that is perpendicular to ' F ' at any point say ' p '. Now, V_n has been calculated as described in equation 3.

$$V_n = (V \cdot \hat{n}) \cdot \hat{n} \quad (3)$$

here ' \cdot ' is used to denote dot product operation

The remaining component i.e. tangential component has been calculated by following equation 4.

$$V_t = -\hat{n} \times (\hat{n} \times V) \quad (4)$$

here ' \times ' is used to denote cross product operation

Now, for accomplishing Vectorfaces from surface normal, at first it has been processed by bicubic interpolation [4] method. During interpolation mechanism, some missing links have been restored. Thus, some points have been restored, and a visual representation is accomplished. For the desired output, authors have chosen the value of the parameter (σ), with an iterative approach. Here, ' σ ' is

considered as 3. Finally, the smoothed matrix grid is termed as 'Vectorfaces' as shown in figure 2. Hence, the basis of such face image is surface normal vector, and every mathematical formulation are done following a vector operation, authors have termed the ultimate resultant matrix as 'Vectorfaces'.

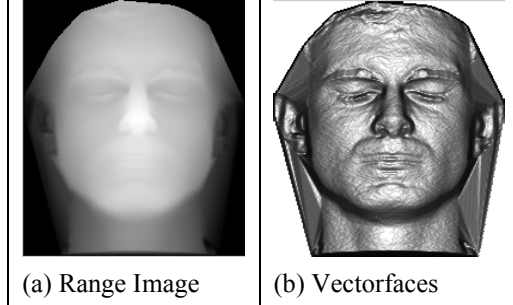


Figure 2. Representation of range image and Vectorfaces.

It can be noticed that, from the Vectorfaces facial shapes, as well as different facial objects (like: eyes, nose, lips etc.), are accurately detectable than range images. This quality has proven to be more significant for face recognition than range images. It has been explained later in section 4 of the paper. Moreover, Vectorfaces preserves more information content that has been observed from two different mathematical measures, like entropy and mean of standard deviations. Entropy [5] is used to compute the amount of information it preserves. The more data it contains, the more entropy can be noted. Whereas, evaluation of mean of standard deviations is used to measure the average of standard deviations that has been calculated from each column of the face image. Standard deviation defines the amount of variations in the data. The mean of standard deviation over all the columns of the image is the measurement of average variation from the image. It has been observed that, the entropy of range image shown in figure 2(a) is 0.7069 whereas from Vectorfaces (displayed in figure 2(b)) it is 4.9790. Besides this, the mean of standard deviations of a range image is 38.4320 and for Vectorfaces the value is 0.2556.

From this analysis, it can be concluded that, although the entropy of the Vectorfaces is much higher than range image, it is having very less average variations among data it preserves. It signifies that Vectorfaces accumulates images in much informative and correlated way and an excellent way of representation for preserving facial features.

3. VECTORFACES VS. RANGE IMAGE

In this section, authors are concerned to explain the efficient methods and their representation that explains the significance of Vectorfaces of range face images. Representation of 3D face image cannot be

usually seen. Here, in figure 3, authors have displayed a 3D face image in a reference plane. There are data points for X-Y grid to represent Z as depth values.

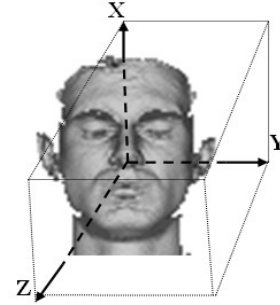


Figure 3. A 3D is a referential image.

Now, to emphasize the implication of Vectorfaces, authors have considered all the data points that are along the side of three axes. Rendering three volumes of datasets from three axes, authors have determined various surface metrics, like line, area, curves, etc. from face surface using depth data. Hence, to process and visualize these metrics authors have applied some methods, such as curvature analysis, shape index, and curvedness index.

3.1. Curvature analysis

In general curvature is used to define the amount by which the surface of any object deviates from a line or plane. Among two types of curvatures, like: extrinsic and intrinsic, Mean curvature belongs to the extrinsic property whereas Gaussian curvature is having the intrinsic property of face surface. Mean curvature (H) describes the curvatures of the local face region. It has also been noticed that the mean curvature values are the trace of second fundamental form, and that has been computed by equation 5. It can also be computed from principal curvatures by simple averaging them.

$$H = \frac{(1 + f_x^2)f_{yy} - 2h_x h_y h_{xy} + (1 + f_y^2)f_{xx}}{(1 + f_x^2 + f_y^2)^{3/2}} \quad (5)$$

Gaussian curvature (K) is computed by taking the product of principal curvatures. The alternative definition that the authors have followed here is shown in equation 6.

$$K = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2} \quad (6)$$

The principal curvature at any point on face surface is the minimum and maximum values of the curvature that can be denoted by P_{\max} and P_{\min} and they are perpendicular to each other if they are not equal. Now, it has been computed from Gaussian and mean curvature values using equation 7 and 8.

$$P_{\max} = H + \sqrt{H^2 - K} \quad (7)$$

$$P_{min} = H - \sqrt{H^2 - K} \quad (8)$$

In these equations, f_x and f_y is the first derivative of f , such that $Z = f(x, y)$, with respect to X and Y axes. f_{xx} and f_{yy} are the second derivative of f with respect to X and Y axes. f_{xy} represent the mixed derivative.

In this investigation process authors have also implemented these mathematical derivations for Vectorfaces for suitable comparison with corresponding range face image. In Table 1, the curvature points for both the images from the various analysis are displayed with quantification using number of points it preserves. The images in table 1 are of binary type. The white points are the representation of the detected curve points from individual curvature analysis.

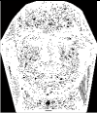

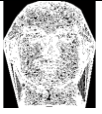


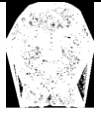


(i) H	(ii) K	(iii) P_{max}	(iv) P_{min}
			
No. of points: 34190	No. of points: 25223	No. of points: 30705	No. of points: 29708
(a) From range face image			
			
No. of points: 38085	No. of points: 35472	No. of points: 36938	No. of points: 36624
(b) From Vectorfaces			

Table 1. Observation of curvature points on range and Vectorfaces image

From this type of analysis and corresponding representation highlights that the number of curvature points are accumulated by various curvature maps from Vectorfaces than range images. Hence, it proves that Vectorfaces would be more useful for further processing.

3.2. Shape index and Curvedness index

The curve points are mainly focused on global curvedness whereas shape index (SI) is used to calculate local features that might also be useful during the face recognition task. The curvedness index (CI) is also computed to measure the magnitude of the curvedness at any point on face surface that are particularly dependent on principal curvatures of the surface. The SI and CI are followed by equation 9 and 10. These particular descriptors from both the face images are displayed in figure 4.

$$CI = \sqrt{\frac{P_{max} + P_{min}}{2}} \quad (9)$$

$$SI = -\frac{2}{\pi} \tan^{-1} \frac{P_{max} + P_{min}}{P_{max} - P_{min}} \quad (10)$$

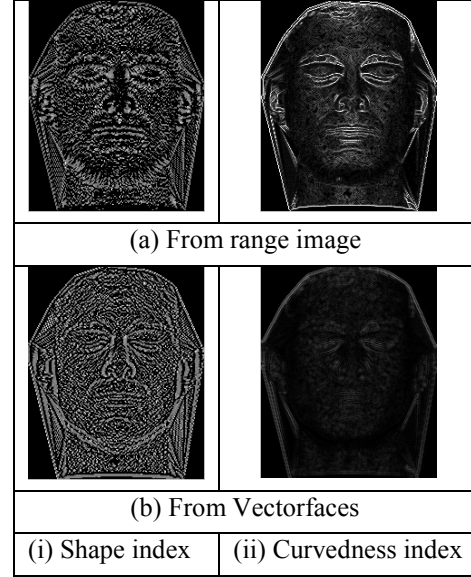


Figure 4. SI and CI from face images.

3.3. H-K classification and SI range

Here, authors have represented various facial surface's regions from H-K classification [6] and SI range [6]. H-K table is useful to describe various surface information for instance, concave and convex cylinder, planer, valley ridge, rut, dome etc. following a conditional parameters of H and K . Though much work has been progressed using curvatures, authors have planned to explain the significance of Vectorfaces with suitable representations. In Table 2, various surface information from range face image as well as Vectorfaces using H-K table is described. In addition, the values of SI (shown in equation 10) can be interpreted between -1 to 1 that specifically describes more local information than H-K. In Table 3, facial surfaces from SI are also described for both the images.

From such analysis, described in table 1, 2 and 3, it is proved that Vectorfaces is capable to accomplish more global as well as local facial features than conventional range face images. The efficient representation of Vectorfaces implies that it contains more number of effective data points than range face images and as a matter of fact it accomplishes other tasks in the sequence, typically face recognition with much higher success rate.

$K < 0 \text{ \& } H < 0$	$K = 0 \text{ \& } H < 0$	$K > 0 \text{ \& } H < 0$	$K < 0 \text{ \& } H = 0$	$K = 0 \text{ \& } H = 0$	$K < 0 \text{ \& } H > 0$	$K = 0 \text{ \& } H > 0$	$K > 0 \text{ \& } H > 0$
(a) From range image							
(b) From Vectorfaces							

Table 2. Representation of facial surfaces using H-K table

$[-1, -0.625)$	$[-0.375, -0.125)$	$[-0.125, 0.125)$	$[0.125, 0.375)$	$[0.375, 0.625)$	$[0.625, 1)$
Spherical cup	Saddle Rut	Saddle	Saddle Ridge	Ridge	Dome
(a) Various surface description from SI					
(b) Extraction of various face surfaces from range image					
(c) Extraction of various face surfaces from Vectorfaces					

Table 3. Facial surfaces from SI range

Various representations in these tables of face images from both the image types are binary images. Here also the white points used to visualize the validate points from H-K table and SI. Here, the term 'validate' means the valid points that are within the range of H-K table and SI.

4. EXPERIMENTAL STUDY

In this section, authors have presented a case study to underline the significant role of Vectorfaces over range face images in performing face recognition. Face recognition is an interesting and challenging research domain that has been studied during the past decades. In addition, there is numerous holistic as well as feature-based techniques for robust and accurate recognition purpose. Therefore, an effective visualization and representation would be one of the key contributions for implementing the successful face recognition system. Hence, to select the suitable feature set from input face images it is very much required to discover the underlying information for more robust recognition purpose. In figure 5, the role of

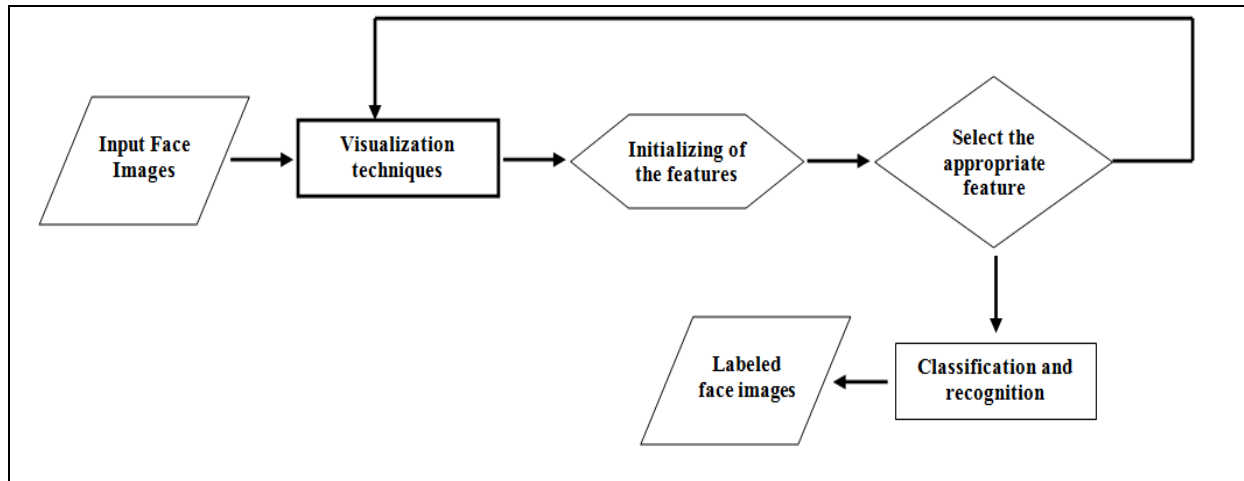


Figure 5. Importance of Representation techniques in face recognition scheme.

Representation techniques for executing an efficient recognition algorithm has been illustrated. In this figure, a feedback loop between feature selection and representation techniques blocks implies the cohesive relationship. A better representation of the data points is very much advantageous for selecting better features that ultimately results in better recognition rate.

In connection with the qualitative assessment, in terms of representation or representation, the said case study have been followed by the authors. In figure 6, an overall sketch of the recognition algorithm for quantitative evaluation of these images is illustrated.

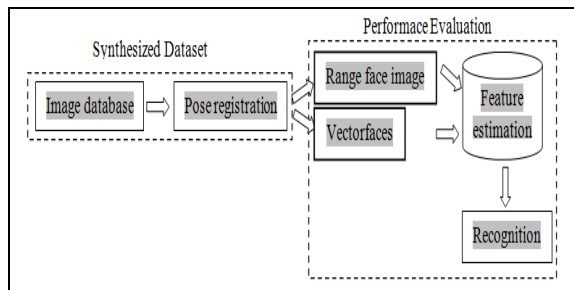


Figure 6. Sketch of the quantitative estimation.

4.1. State-of-the-art

For the motivation of such quantitative evaluation after qualitative discussion, authors have described the similar type of approaches that have already been considered by the researchers. In [7], Ganguly et al. researchers have compared the recognition performance of two paired curvature maps, such as Maximum Principal curvature-Mean and Gaussian-Mean. It is investigated that the features from the set of feature vector, better recognition is achieved by Maximum Principal curvature-Mean curvature pair. The key attributes from face image have been extracted using SVD based feature extraction mechanism. Later these features have been classified by five layers feed-forward

backpropagation neural network. Lin et al. [8] also proposed curvature based 3D face recognition algorithm which are invariant of expression. Here, authors have mainly focused on Gaussian curvature maps. To create feature matrix, curvature points from various facial properties, like: nose, mouth eyes, etc. are considered. Later the feature vector is classified by distances between query image and gallery images. In [9], researchers have proposed PCA based face identification and verification process after successful detection and normalization of input 3D scan face image. Here, authors have used curvature values to detect whether the input image is facing for not. Mahoor et al. [10], proposed Hausdorff distance based face matching technique from surface ridge data. Surface ridge data has also been obtained from curvature analysis.

4.2. Discussion

Here, authors have followed and modified the recognition algorithm [7] for evaluation of Vectorfaces as well as range face image in terms of face recognition [14] rate. Since the algorithm proposed in [7], directly considers the curvature data as feature to discriminate the individuals, authors have followed this technique. Moreover, the data points accomplished by the curvature maps are more for Vectorfaces than corresponding range image. Therefore, aiming to compare the qualitative assessment, as well as quantitative measurement of same representation, has also inspired the authors to investigate the reorganization scheme from curvature maps.

In particular, authors have implemented and validated the algorithm using three layer feed-forward backpropagation neural network on synthesized face dataset from Frav3D database. In addition, unlike the algorithm [7], the concatenated features from paired curvature maps has not been sorted for this examination. Therefore, individual feature's position as well as its magnitude holds its

important characteristics in the final feature vector. The synthesized dataset is accomplished by registering the rotated face images using ERFI model [11]. Other than registered face images, it contains face images with neutral as well as an expression.

In Table 4, the perceptible improvement of face recognition rate due to Vectorfaces is presented with a comparison of corresponding range face images. Here, rank based selection of features from SVD feature set is carried out for both the image types for recognizing 3D face images.

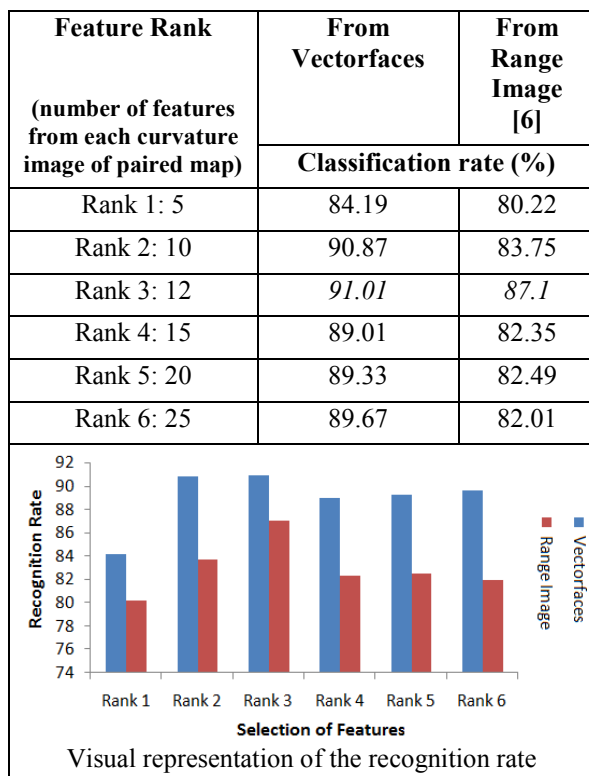


Table 4. Quantitative evaluation of Vectorfaces and range face image

From this tabular representation of the evaluation (in terms of recognition rate), it has been proved that the data points that are available in Vectorfaces is very much informative and useful for better recognition purpose. In Table 1, it has already been described that number of detailed points in the considered pair for Vectorfaces is much higher than range images. The recognition rates from the proposed mechanism also exhibit the same and support the qualitative measurement.

Although the said recognition rate is comparable with other's [12-13], authors have only tried to prove the importance of Vectorfaces by visualizing (i.e. qualitative fashion) and quantitative measurements (in terms of recognition rate).

5. CONCLUSION

During this investigation process, authors have established the role of Vectorfaces in the domain of 3D technology by qualitative as well quantitative assessment. Representation of the Vectorfaces in various fashion and relative importance establish that it would be effective for recognizing 3D face images like 2.5D range image.

Individual components of Vectorfaces, namely tangent and normal components can eventually be useful for further processing. Other than these investigations, Vectorfaces can be used to extract different facial properties, like nose, eyes, lips, etc. which might be eventually helpful for recognition purpose. Hence, it has certainly many useful applications in computer vision domain.

6. ACKNOWLEDGMENTS

Authors are thankful to a project supported by DeitY (Letter No.: 12(12)/2012-ESD), MCIT, Govt. of India, at Department of Computer Science and Engineering, Jadavpur University, India for providing the necessary infrastructure for this work.

7. REFERENCES

- [1] H. Cantzler, "An overview of range images," URL: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/CANTZLER2/range.html. Last visited: 20.02.2015.
- [2] S. Ganguly, D. Bhattacharjee and M. Nasipuri, "2.5D Face Images: Acquisition, Processing and Application", in ICC 2014- Computer Networks and Security, International Conference on Communication and Computing (ICC-2014), pp. 36-44 ISBN: 9789351072447, 2014.
- [3] P. Westfeld, R. Hempel, "Range Image Sequence Analysis By 2.5-D Least Squares Tracking With Variance Component Estimation And Robust Variance-Covariance Matrix Estimation," pp. 933-938. URL: http://www.isprs.org/proceedings/XXXVII/congress/5_pdf/162.pdf
- [4] R. C. Gonzalez, R. E. Woods, "Digital Image Processing," 3rd Edition, Prentice Hall publisher.
- [5] E. C. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal, vol. 27, issue 3, pp. 379-423, DOI:10.1002/j.1538-7305.1948.tb01338.

- [6] H. Cantzler and R. B. Fisher, "Comparison of HK and SC curvature description methods", URL: homepages.inf.ed.ac.uk/rbf/PAPERS/hc3dim.pdf.
- [7] S. Ganguly, D. Bhattacharjee, M. Nasipuri, "3D Face Recognition from Range Images Based on Curvature Analysis," in *ICTACT Journal on Image and Video Processing*, vol. 4, issue 3, pp. 748-753, 2014.
- [8] S. W. Lin, S. S. Hao, J. L. Chang, S. Y. Li, "3D Face Recognition Based on Curvature Feature Matching, with Expression Variation," in *Intelligent Autonomous Systems 12, Advances in Intelligent Systems and Computing Volume 193*, 2013, pp 289-299, 2013.
- [9] A. Colombo, C. Cusano, R. Schettini, "A 3D face recognition system using curvature-based detection and holistic multimodal classification," *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pp. 179 - 184, DOI: 10.1109/ISPA.2005.195406, 2005.
- [10] M. H. Mahoor, M. A. Mottaleb, "Face recognition based on 3D ridge images obtained from range data," *Pattern Recognition* 42 (2009), pp. 445 - 451.
- [11] S. Ganguly, D. Bhattacharjee, M. Nasipuri, "Range Face Image Registration Using ERFI from 3D Images," in *proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014, Advances in Intelligent Systems and Computing Volume 328*, 2015, pp 323-333.
- [12] B. Bornak, S. Rafiei, A. Sarikhani, A. Babaei, "3D Face Recognition by Used Region-Based with Facial Expression Variation," in *proceedings of 2nd International Conference on Signal Processing Systems (ICSPS)*, pp. V1-710- V1-713.
- [13] N. Belghini, A. Zarghili, J. Kharroubi, "3D Face Recognition using Gaussian Hermite Moments," *Special Issue of International Journal of Computer Applications (0975 – 8887) on Software Engineering, Databases and Expert Systems – SEDEXS*, pp. 1-4, 2012.
- [14] C. Zhong, Z. Sun, T. Tan and Z. He, "Robust 3D face recognition in uncontrolled environments," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, doi: 10.1109/CVPR.2008.4587646, 2008.

Automated Detection of Buildings on Aero Images

Leonid Novotortsev

Keldysh Institute of Applied Math RAS
Moscow, Russia
torets13@gmail.com

Alexey Voloboy

Keldysh Institute of Applied Math RAS
Moscow, Russia
voloboy@gin.keldysh.ru

ABSTRACT

One of the challenging problems in photogrammetry is extracting of three-dimensional objects from aero images, in particular, extraction of different kinds of buildings. All methods that provide satisfactory results are rather time consuming and process data quite long. In the paper we propose a method that detects areas on aero images that might contain a building behind them. Our method allows reducing the amount of data which should be processed by more complex algorithms. This leads to reducing of the total time spent on extraction process.

Keywords

computer vision, line extraction, building extraction.

1. INTRODUCTION

Photogrammetry deals with making measurements from photographs, especially for recovering the exact positions of surface points and reconstruction of the whole scene or specific object types. At this point there are a lot of methods that provide satisfactory results in most cases. But still there are problems that cannot be solved by general methods with required accuracy or speed. One of them is the detection and reconstruction of buildings from aero images.

There are several approaches to that problem. There is an approach, which proposes to use special data for building recognition such as LIDAR [Rot02a], [Soh07a]. Results, provided by those methods, have high detection and low miss percentage, but in many cases LIDAR data is not available.

Another approach is to use structural, contextual, and spectral information from satellite photos [Jin05a] and aero images. There are several modern methods, which based on this approach that provide good results (recognition with probability ~85%) [Ok13a], [Gha14a], [Sin14a]. But there is a general limitation to this approach. In some cases it is impossible to detect buildings, which have the same colour and have no visible shadow.

To overcome this problem it is possible to use photos of the same scene, which have been shot from a bit different views. That way the object that is non distinct in one image should be clearly visible in the other. By using multiple images it is also possible to

extract and utilize three-dimensional information about the scene. In order to process multiple images more efficiently special methods are needed.

The first approach is to build DEM (Digital Elevation Model) and process it in order to extract buildings. General methods of extracting buildings [Bru97a], [Cor97a] and [Gir98a] are rather rough. For that reason, methods that use prior information about building shape are applied. For example there is a method, which detects only flat roofs [Bre95a]. Such methods provide accurate results but are too limited.

The other approach based on extracting linear objects from images. The first step is to detect and extract lines, then match them and apply plane sweep strategy [Bai00a]. But this method rather complex, since it requires matching lines on several images (it is recommended to use 5 or 6 images). Processing such amount of data takes quite long. This problem could be solved by preprocessing each image and limiting the area to which this method is applied.

Since the main purpose of this paper is to propose preprocessing algorithm, we will refer to method, described above as “complex algorithm” or “complex method” (because of its comparable complexity to proposed method).

In order to get regions which might contain buildings it is possible to use methods, which were designed to detect buildings on single image [Jin05a], [Ok13a], [Gha14a], [Sin14a]. But these methods are rather complex themselves and they utilize completely different approaches. This way the advantage in computation time which we have got from preprocessing is decreased. In order to avoid this it is better to use more simple method for preprocessing. It is even better if some information acquired by preprocessing can be used by complex method.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

The general idea of this paper is to detect straight lines on image, then unite them into contours and select those which might contain a building within.

It is important to notice, that our goal is not to detect building edges themselves but to find areas, which contain buildings. Then a complex algorithm will process much less data. Also that complex method can reuse straight lines and segments detected by proposed method. This way the amount of additional computations (compared to computations made by complex algorithm) on preprocessing step will be rather small.

2. GENERAL IDEA

In this paper we propose the usage of the fact that most buildings have linear edges. Therefore, it is convenient to use line extraction methods in order to achieve our goal, because such limitation allows to easily and accurately analyze the data. The whole algorithm can be divided into following steps:

1. Edge filter
2. Line extraction
3. Combine lines into contours
4. Analyze contours and leave only those, which might belong to the buildings.

In that scheme the first and the second steps are also done by complex algorithm and only the last two steps are additional. This way it is possible to pass lines, extracted on the second step to complex method in order to prevent repeated computations. It is also noticeable that the last two steps are rather simple since they process mostly vector data rather than raster data

3. EDGE FILTER

There are few methods that detect edges on images. Among them, Sobel filter [Sob73a] and Canny filter [Can86a] are noticeable.

Sobel Filter

The Sobel filter was designed for edge detection algorithms and creates an image which emphasizes edges. Technically, it is a discrete operator, computing an approximation of the gradient of image channel or intensity. The operator uses 3x3 kernels for horizontal and vertical directions, which are applied to each point of the image:

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \times A \quad (1)$$

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \times A \quad (2)$$

Where A is the source image, G_x and G_y are images which at each point contain the horizontal and vertical derivative approximations, \times is 2-dimensional convolution operation.

From these values the gradient magnitude is calculated:

$$G = \sqrt{G_x^2 + G_y^2} \quad (3)$$

The result of the Sobel operator is a 2-dimensional map of gradient at each point. The higher the gradient value the more likely there is an edge at this point of the image.

But in our task it is necessary to strictly determinate, if there is an edge for all points in image. Just using a threshold for values from Sobel filter is not enough.

Canny filter was proposed to overcome such problems as thinning edge and removal of false edges, created by image noise.

Canny Filter

Canny filter can be broken down to a four steps.

The first step is to apply a Gauss filter. On the second step the Sobel filter is applied and the edge direction angle is calculated from equations (1), (2) and rounded to one of the 4 values: 0, 45, 90, or 135 degrees.

$$\theta = \text{atan2}(G_y, G_x) \quad (4)$$

On the third step non-maximum suppression is applied to “thin” the edge. The algorithm for each pixel in the gradient image is:

1. Compare the gradient magnitude of the current pixel with the edge strength of the pixel in the positive and negative gradient directions
2. If the edge strength of the current pixel is not the largest compared to the other pixels in the mask with the same direction the value will be suppressed.

After that, pixels with a gradient magnitude larger than certain threshold (often called high threshold) are marked as strong edge. A pixel is marked as weak edge if gradient magnitude is larger than the second threshold (low threshold) but smaller than high threshold. Pixel is suppressed if the gradient magnitude is lower than the low threshold.

On the last step all weak edge pixels that don't have strong edge pixel in their neighborhood are suppressed.

In this paper we propose to use improved Canny edge detector, proposed in [Bao05a], since this algorithm provides even more precise results.

4. LINE EXTRACTION

The next step of our method is to extract lines and segments form edge map we received on previous

step. And one of the most widely used techniques for locating straight lines is Hough transform [Hou62a]. The main idea of the Hough transform is to map image pixels, defined by their position (x, y) to the ρ - θ parameter space corresponding to all possible lines through the point, using the following formula:

$$r = x \cos \theta + y \sin \theta, \quad (5)$$

where ρ is the perpendicular distance of the line to the origin, and θ is the angle between a normal to the line and the positive x axis. Usually the parameter space is partitioned into adjoining rectangular cells using a predefined resolution in both the ρ and θ dimensions.

The accuracy of the Hough transform received much attention due to the errors resulting from discretization of parameters [Ngu08a, O'Rp81a, Li86a].

An interesting approach is suggested in [Du10a]. In this paper it is proposed not to process straight line segments but rather their neighborhood and search for intersections of these neighborhoods in a voting process. That approach provides much better results than a plane voting approach.

In this paper we propose to use the following approach to straight segment detection. First we divide the image into equal square cells (except for those, which are placed at the border of the image). In each cell the Hough transform is applied. Finally, straight line segments from neighbouring cell are merged if they are located on the same line.

As the result of the described algorithm we will get a set of straight line segments, corresponding parts of linear objects. But at this point there are no relations between these segments, so it is necessary to unite them.

5. COMBINING STRAIGHT LINE SEGMENTS

First we will define two parameters, which are be used in the merging algorithm. The first parameter we will call "expected distance between buildings" or EDBB for short. The second parameter we will call "expected building size" or EBS. These parameters will define how lines will be merged. The first parameter has the following meaning: if distance between objects is shorter than EDBB value, then these objects don't belong to the same building or don't belong to any building at all. Technically, this is average linear size of buildings in scene. EBS is set in such a way that objects that are further than this value, then they either belong to different building or other type of objects. Technically, this is the size of the smallest object which is classified as building. In this paper we propose to set EDBB and EBS manually. The possibility of automated definition of these parameters will be studied in future researches.

There are 3 different types of segment layout, which will be analyzed separately:

1. Segments are positioned on the same line
2. Segments are positioned on intersecting lines
3. Segments are positioned on parallel lines

After segment analysis and merging closed contour finalizing is performed. On this phase all lines, that were not included into closed contours or located inside one of them, are marked as rejected. But they might be reused again in methods, which

Segments on the Same Line

Sometimes a single straight line edge is divided into two or more separate edges. This might occur because of low image quality, shadows or low contrast. In this case these segments must be united.

Usually the task is solved by using a threshold. That approach is not correct since a single threshold is not enough to accurately determinate if segments should be merged.

In this paper we suggest to use two thresholds. If the distance between the segments is lower than the first threshold (EDBB) and shorter than one of the segments, then these segments should be merged. If the distance is larger than the second threshold (EBS) or longer than both segments, then segments shouldn't be merged. Otherwise additional analysis should be done.

First we define a region of interest. Coordinates of closest ends of segments are interpreted as left, top, right and bottom border of a rectangle, representing that region. Then the rectangle is expanded for 2 pixels in each direction in order to detect edges at the border of the region.

After that, two thresholds are selected for the Canny filter based on the histogram of that rectangle (the using absolute minimum or maximum is not effective, because of random deviations produced by noise). Then the Canny filter is applied. After that, the number of pixels, which were marked as edges by the Canny filter, are counted in the neighborhood of the "lost" segment (the one between two involved segments). If this number is close to the high threshold then these two segments should be merged.

Segments on Intersecting Lines

This case is similar to the previous one. Just as in the previous subsection, there are three cases to consider, based on distance between the nearest ends of segments. Also the same methods are applied to determinate if segments should be merged.

One of the differences is the fact, that segments (one or both) might contain a point of intersection. In this case, the segment that has an intersection point is divided into two segments at that point.

In some cases false angles between segments might appear after merging. This problem is solved in the following way. If the angle between segments is close to 180 degrees (e.g. they are nearly positioned on the same line) an addition check is performed. Then, if the length of one completed (extended to point of intersection) segment is much shorter than the length of the other one, they can be considered as single straight line segment.

Segments on Parallel Lines

This type of layout can represent four kinds of building layouts.

The first kind is parallelism of random edges that are not related in any way. This type is characterized by significant difference in edge length. In this case the edges are left as is.

The second and the third kinds of layout represent parallel edges of the same building or of different buildings accordingly. These two types are distinguished and processed with method used in previous subsections (two thresholds and the Canny filter).

The last one is duplication of building an edge (for example if there is a drainpipe at the roof of the building). If lines are positioned close to each other and only one of them is part of some contour, then they should be merged into one and the contour should be expanded accordingly.

6. FILTER CONTOURS

At this step of the described method we filter those closed contours which were created in previous steps. Also all straight line segments, which are not part of some contour, are ignored (but they will still be passed to complex algorithm).

In this method we propose to use two criteria for the filter. The first one is the limitation of closed contour area. It should be at least more, than $EBS \cdot EBS$. A second criterion is the shape of close contours. They should have a convex n-gon shape.

These two criteria are quite simple, so there is a lot of space for further research here.

7. RESULTS

In Figure 1 and Figure 2 the source images are presented. These images are shot from airplane when it flew in different directions (strips). We applied the described method to those images. The results of region detection are shown in Figure 3 and Figure 4 accordingly.

Besides buildings the selected regions include shadows from these buildings and some other objects. Two regions are in fact just a parking lot. But that inaccuracy is acceptable, since the described method is supposed to decrease the amount of data passed to much more complex algorithms.



Figure 1. Source image, the first strip



Figure 2. Source image, the second strip



Figure 3. Contours produced by proposed method from Figure 1.



Figure 4. Contours produced by proposed method from Figure 2.

The average detection quality of described method is rather low: around 71%. But that is compensated by the fact that if the building which have not been detected in one image, would be detected in the other. Since complex method processes all images, detection of each building at least in one image is enough. That way the detection rate will be nearly the same as if only complex method [Bai00a] is applied to set of images, but the amount of data processed is much smaller.

8. CONCLUSION

The problem of detection and extraction of buildings is quite a challenging problem in computer vision and photogrammetry in particular. But existing methods are either not accurate enough or have high computational complexity. Of course the last type of algorithms is preferable. But in photogrammetry it is common to use aero images which may be larger than 10000 by 10000 pixels. Complex algorithms will process such amount of data too long. So we present a method that will decrease the volume of input data for them.

On presented example our method selected all buildings without missing a single one. Besides, given the fact that in aero images buildings occupy approximately 20-60% of images (depending on territory, where images were shot), our method will give around 40-80% boost, which is a very good result. The described algorithm is original.

9. REFERENCES

- [Bao05a] Bao P., Zhang D., Wu X. Canny edge detection enhancement by scale multiplication. IEEE transactions on pattern analysis and machine intelligence, 27(9), pp. 1485-1490, 2005.
- [Du10a] Du S., van Wyk, B. J., Tu, C., & Zhang, X. An improved Hough transform neighborhood map for straight line segments. IEEE Transactions on Image Processing, 19(3), pp. 573-585, 2010.
- [Hou62a] Hough P. V. C. Method and Means for Recognizing Complex Patterns, U.S. Patent 3069654, Dec. 18, 1962.
- [Ngu08a] Nguyen T. T., Pham X. D., and Jeon J. An improvement of the Standard Hough Transform to detect line segments. IEEE International Conference on Industrial Technology, pp. 1-6, 2008.
- [O'R081a] O'Rourke J. Dynamically Quantized Spaces for Focusing the Hough Transform. IJCAI, 81, pp. 24-28, 1981.
- [Li86a] Li H., Lavin M. A., Le Master R. J. Fast Hough transform: A hierarchical approach. Computer Vision, Graphics, and Image Processing, 36(2), pp. 139-161, 1986.
- [Can86a] Canny J. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, pp. 679-698, 1986
- [Cor97a] Cord M., Paparoditis N., Jordan M. Dense, reliable and depth discontinuity preserving DEM computation from HRV urban stereopairs. International Archives of Photogrammetry and Remote Sensing, 32, pp. 49-56, 1997.
- [Bru97a] Brunn A., Weidner U. Extracting buildings from digital surface models. International Archives of Photogrammetry and Remote Sensing, 32(3) sect. 4W2, pp. 27-34, 1997.
- [Gir98a] Girard S. et al. Building detection from high-resolution color images. Remote Sensing – International Society for Optics and Photonics, pp 278-289, 1998.
- [Bre95a] Berthod M. et al. High-resolution stereo for the detection of buildings. Automatic Extraction of Man-Made Objects from Aerial and Space Images. – Birkhäuser Basel, pp. 135-144, 1995.
- [Bai00a] Baillard C., Zisserman A. A plane-sweep strategy for the 3D reconstruction of buildings from multiple images. International Archives of Photogrammetry and Remote Sensing, 33(B2) part 2, pp 56-62, 2000.
- [Rot02a] Rottensteiner F., Brieke C. A new method for building extraction in urban areas from high-resolution LIDAR data. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, 34(3/A), pp 295-301, 2002.
- [Soh07a] Sohn G., Dowman I. Data fusion of high-resolution satellite imagery and LIDAR data for automatic building extraction. ISPRS Journal of Photogrammetry and Remote Sensing, 62(1), pp. 43-63, 2007.
- [Jin05a] Jin X., Davis C. H. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. EURASIP Journal on Applied Signal Processing, 2005, pp 2196-2206, 2005.
- [Sob73a] Sobel I. and Feldman G. A 3×3 Isotropic Gradient Operator for Image Processing. R. Duda and P. Hart (Eds.), Pattern Classification and Scene Analysis, pp. 271-272, 1973
- [Ok13a] Ok A. O.. Automated Extraction of Buildings and Roads in a Graph Partitioning Framework. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W3, pp 79-84, 2013.
- [Gha14a] Ghaffarian S., Ghaffarian S. Automatic building detection based on supervised classification using high resolution Google Earth images. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(3), pp 101-106, 2014.
- [Sin14a] Singhal S., Radhika S. Automatic Detection of Buildings from Aerial Images Using Color Invariant Features and Canny Edge Detection. International Journal of Engineering Trends and Technology(IJETT) 11(8), pp 393-396, 2014.

Using Intrinsic Surface Geometry Invariant for 3D Ear Alignment

S.Taertulakarn^{1,2},

¹ Faculty of Engineering,
King Mongkut's Institute of
Technology Ladkrabang, Thailand
10520

² Faculty of Allied Health Sciences,
Thammasat University,
Pathumtani, Thailand 12121.
Email: somchatt@yahoo.com

P.Tosranon

Faculty of Applied Sciences, King
Mongkut's University of
Technology North Bangkok,
Thailand 10800
Email: ptn@kmutnb.ac.th

C.Pintavirooj

Faculty of Engineering, King
Mongkut's Institute of Technology
Ladkrabang, Thailand 10520
Email:
chuchartpintavirooj@gmail.com

ABSTRACT

In this study we derive novel surface fiducial point's detection that is computed from the differential surface geometry. The fiducial points are intrinsic, local, and relative invariants, i.e., they are preserved under similarity, affine, and nonlinear transformations that are piecewise affine. In our experiment, the fiducial points, computed from high order surface shape derivatives, are used in a non-iterative geometric-based method for 3D ear registration and alignment. The matching is achieved by establishing correspondences between fiducial points after a sorting based on a set of absolute local affine invariants derived from them. Experimental results showed that our purposed surface feature is suitable for further application to 3D ear identification because its robustness to geometric transformation.

Keywords— 3D ear registration, Surface geometric invariant, Zero torsion

1. INTRODUCTION

Biometrics is an emerging technique that involves the use of physiological and behavioral characteristics to determine the identity of an individual. At the present time, the physical biometrics, for example fingerprints, facial patterns and eye retinas, are developed and enhanced. In this work, we are interested to identify humans by ear structure. The ear anatomy has a lot of unique structures which do not change with changing event or age. In addition, compared with the other biometrics techniques, the ear data can be registered in a non-invasive way [Che07a, Bur00a].

Geometric Invariance is a central problem in visual information system, computer vision, pattern recognition image registration, and robotics. The term invariance is referred to the geometrical properties of the relative distance among a collection of static spatial features of an object [Gov99a-Pin13a].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Our study is focused on registration techniques for biomedical 3D images, Surface registration. Surface registration is a vital step in medical imaging literature. Various techniques have been proposed for surface registration from which the following general methodology can be defined. Firstly, the landmarks or specific structures are extracted from each image to be registered. Secondly, correspondence between the extracted landmarks is established. Thirdly, choice of geometric transformation, such as rigid transformation affine transformation or polynomial transformation, is entertained. Fourthly, the geometric transformation parameters are estimated. Lastly, the two surfaces are aligned. Surface registration methods can be categorized into polynomial transformation [Sin09a], similarity-based [Woo98a] surface-based [Bes92a], energy-based [Por94a] and fiducial point - based registration [Kan81a-Ibr98a].

In this paper, we consider the problem of 3D ear alignment in the different orientations. The data is 3D and obtained using a laser scanner. Our approach is based on the differential geometry of the surface, consists of two processes, start from computing intrinsic local fiducial points on the surface and on curves that reside on the surface. A fast non-iterative alignment is purposed in our study that establishes reliable correspondences between fiducial points

without any prior knowledge of the overall nonlinear global transformation that took place after the changes. This is achieved through the construction of a set of ordered novel absolute local affine invariants. With enough fiducial points collection as correspondents, the overall nonlinear transformation is computed and the ear before and after the transformation are aligned. Finally, this paper deals with the computation of the fiducial points that are based on high order surface shape and curve derivatives, which are honorable for their sensitivity to measurement errors, round off error and distortion.

This paper is organized as the following: - Section 2 introduces the intrinsic local geometric fiducial points on surfaces. Section 3 shows how to construct a set of absolute invariants derived from these fiducial points and how to manage the correspondences between two partial sets of fiducial points residing on two surfaces that are transformations of one another. The experimental results on robustness of surface geometric invariance applied compared with the 3D ear are described in section 4. The finally, discussions and conclusions are given in section 5.

2. GEOMETRIC INVARIANT SHAPE MEASURE

2.1 Parabolic Contour Points

Local and invariant intrinsic properties are presented by the Frenet frames [8], which states that for a curve $\mathbf{r}(s)$ parameterized by arc length s , the tangent $\mathbf{t}(s) = \mathbf{r}'(s)$, the curvature $\mathbf{k}(s) = \mathbf{r}''(s)$, the vector $\mathbf{b}(s) = \mathbf{t}(s) \times \mathbf{k}(s)$, and the torsion $\tau(s) = -\langle \mathbf{r}''(s), \mathbf{b}'(s) \rangle$ determines as a set of local coordinates on the curve at each point that completely characterizes the curve at that point, where $\mathbf{r}^{(k)}(s)$ stands the k^{th} order derivative of \mathbf{r} with respect to s , and \times is the cross product operation.

As we are interested in finding the relative and absolute invariant to the affine transformation, we observe that since arc length is not preserved under the affine transformation, neither $\mathbf{t}(s)$ nor $\mathbf{b}(s)$ cannot be used because they are not relative invariants. We seek to find geometric invariance on the surface which is intrinsic, local and affine invariant. When a surface undergoes an affine transformation, the parabolic contours are the affine transformed parabolic contours of the original curve, i.e., they are preserved. Similarly, the fiducial points residing on these contours are also preserved under the affine transformation. In this section, we briefly introduce theory related to the geometric invariance.

Parabolic contours are space curves that reside on a surface when either one of the two principal curvatures, k_1 or k_2 , is zero [Do76a, Mil97a]. In that case, the Gaussian curvature ($K_G = k_1 k_2$), which is

intrinsic, and vanishes at these points. For a surface, represented by the parameterization $\mathbf{r}(u, v): U \subset \mathbb{R}^2 \rightarrow S$, the Gaussian curvature is given by the determinant of the second fundamental form parameters [Do76a]

$$K_G = k_1 k_2 \propto \begin{vmatrix} e & f \\ f & g \end{vmatrix} \quad (1)$$

Where

$$e = \langle \mathbf{N}, \mathbf{r}^{(2,0)} \rangle, f = \langle \mathbf{N}, \mathbf{r}^{(1,1)} \rangle, g = \langle \mathbf{N}, \mathbf{r}^{(0,2)} \rangle$$

$$\mathbf{N} = \langle \mathbf{r}^{(1,0)}, \mathbf{r}^{(0,1)} \rangle, \mathbf{r}^{(1,0)} = \frac{\partial \mathbf{r}}{\partial u}, \mathbf{r}^{(0,1)} = \frac{\partial \mathbf{r}}{\partial v}$$

$$\mathbf{r}^{(2,0)} = \frac{\partial^2 \mathbf{r}}{\partial u^2}, \mathbf{r}^{(0,2)} = \frac{\partial^2 \mathbf{r}}{\partial v^2}, \mathbf{r}^{(1,1)} = \frac{\partial^2 \mathbf{r}}{\partial u \partial v}$$

The parabolic contours are given by solving

$$eg - f^2 = \langle \mathbf{N}, \mathbf{r}^{(2,0)} \rangle \langle \mathbf{N}, \mathbf{r}^{(0,2)} \rangle - \langle \mathbf{N}, \mathbf{r}^{(1,1)} \rangle^2 = 0 \quad (2)$$

The parabolic contours, which are based on the Gaussian curvature, are intrinsic [Do76a] and preserved under the affine transformation.

The mean curvature was computed from half of Gaussian curvature.

$$K_m = 1/2 (k_1 k_2) \quad (3)$$

2.2 Zero Volume and Zero torsion Points on Parabolic Contour Curves

Given a parabolic contour curve $\mathbf{r}(t)$ we can also obtain intrinsic curve points by creating volume relative invariants. One such relative affine invariant can be had by considering the volume of the parallelepiped spanned by the zero, first, and second curve derivatives given by the scalar triple product

$$v_1(t) = \langle \mathbf{r}^{(0)}(t) \times \mathbf{r}^{(1)}(t), \mathbf{r}^{(2)}(t) \rangle \quad (4)$$

Where $\mathbf{r}^{(k)}(t)$ is the k^{th} derivative of the curve with respect to parameter t . Equation (5) is a relative affine invariant when the affine transformation has zero translation, i.e., when it is a purely linear transformation. Another relative invariant that carries in the case of a nonzero translation, is the volume of the parallelepiped that is spanned by the first, second, and third curve derivatives given by the scalar triple product

$$v_2(t) = \langle \mathbf{r}^{(1)}(t) \times \mathbf{r}^{(2)}(t), \mathbf{r}^{(3)}(t) \rangle \quad (5)$$

3. ALIGNING THE SURFACES

To align two surfaces that we need to establish the corresponding fiducial points on the two set of

parabolic contours from which we can estimate the overall transformation using least squares. To establish these correspondences and in the absence of the knowledge of the transformation parameters a priori, we construct absolute invariants derived from the fiducial points residing on any contours. These are blind to the transformation and remain unchanged before and after the affine transformation.

3.1 Constructing Absolute Invariant from the Fiducial Points

The set of absolute invariants are constructed from the sequence of relative invariants formed by the volumes of the parallelepipeds spanned by set of four intrinsic surface fiducial points on observed contour. The volume spanned by a set of four points m, n, k and l , is given by $V(m, n, k, l) = |<r_m - r_l, (r_n - r_l), (r_k - r_l)>|$. When the surface is affine mapped, the volume $V_a(m, n, k, l)$ spanned by the mapped points m, n, k and l , relates to

$$V_a(m, n, k, l) = \det\{[L]\} V(m, n, k, l) \quad (6)$$

To facilitate the process of finding the correspondences by reducing to string matching, we rearrange the fiducial points in accordance with the order described below.

For a collection of n intrinsic fiducial points, we pick an intrinsic surface fiducial point, say point i . By using this point as well as the other three points selected from the combination $\binom{n-1}{3}$, we compute the volume spanned by these four vectors. The smallest volume out of the $\binom{n-1}{3}$ computed volumes is assigned as the point i relative invariant. We restart the process with the next point and the remaining $(n-2)$ points excluding the i^{th} point. The process repeats until the list is depleted. The order of the intrinsic surface fiducial point is then sorted according to increasing volume. If both surfaces have n fiducial points, the two sequences of volume patches on the two surfaces would be.

$$(V(1) < V(2) < \dots < V(n-3) \& V_a(1) < V_a(2) < \dots < V_a(n-3))$$

where there volume patches are related by

$$V_a(i) = \det\{[L]\} V(k), k = 1, 2, \dots, n-3 \quad (7)$$

The absolute invariants on the original and transformed surface are the defined as the ratio of the consecutive volume element in the ordered sequence, i.e.,

$$I(k) = \left\{ \frac{V(k)}{V((k+1) \bmod (n-3))} \right\}, k = 1, 2, \dots, n-3$$

$$I_a(k) = \left\{ \frac{V_a(k)}{V_a((k+1) \bmod (n-3))} \right\}, k = 1, 2, \dots, n-3 \quad (8)$$

3.2 Establishing Correspondence of the Fiducial Points

In the absence of noise and occlusion, each of $I_a(k)$ will have a counterpart $I(k)$ with $I_a(k) = I(k)$, with that counterpart easily determined through a circular shift involving n comparison where n is the number of invariants. To allow for noise and distortion, a smaller error percentage between counterpart invariants is tolerated. The lower the error percentage, the stricter the matching is. In this our experiment, an error percent of 5% is used. A run length method is applied to decide on the correspondences between the two ordered set of zero-torsion points. For every starting point on the transformed set, this run length method computes a sequence of consecutive invariants that satisfies $|I(k) - I_a(k)| < 0.05 |I(k)|$ and declare a match based on the longest string. Once this correspondence is found, these matched fiducial points are used to estimate the polynomial transformation.

4. EXPERIMENT AND RESULTS

The experiments are divided into two parts. In part one; we would like to test the robustness of the purposed method to ear feature extraction. In this test, 3D ear cloud point data are subjected. In the second part, we would like to test the aligning curvature feature of our method.

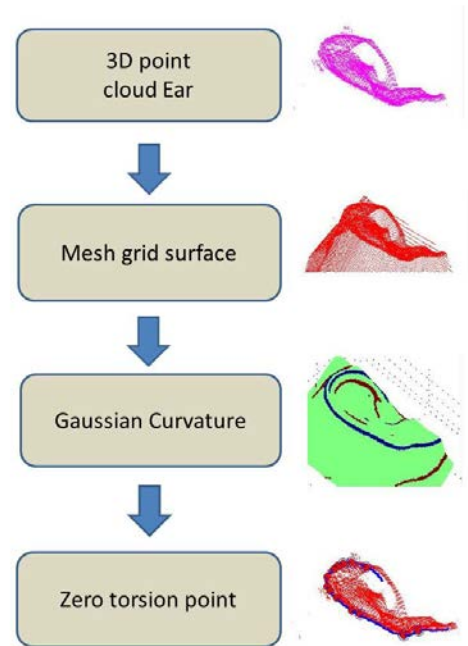


Figure 1. Process of extracting color-mapped images of zero torsion points of surfaces computed from Gaussian curvature.

A. Feature Extraction

The experiment of the robustness of the surface geometric invariant feature on the 3D ear surface extracted from cross-sectional contours to certain class of geometric transformation is described in this section. The process which shows in Fig.1, is as follows:

- (i) Extract 3D coordinate from a set of dense 3D data.
- (ii) Numerically compute to create the grid on the surface
- (iii) Compute Gaussian curvature. The parabolic contours are then derived by approximately solving by Equation (2).
- (iv) Compute Zero torsion points. We obtain intrinsic curve points by creating volume relative invariants which following by Equation (4, 5).

B. Aligning the Surfaces

In this process, the experiment is based on real 3D scans of the same person taken under different orientations. The ear point cloud data is collected in 5 positions. We elect to use the distance map that displays the distance between any point of one surface and the closest point on the other surface after undoing the transformation to the second surface. The two ears taken of the same person at different orientations and with two different positions. Corresponding fiducial points are found, the affine transformation estimated using LSE fitting and the two surfaces are then aligned. The alignment is shown in Fig. 4 whereas we show the two surfaces before and after the alignment in Fig. 2 and 3 respectively. The average distance map error after alignment is shown in table 1.

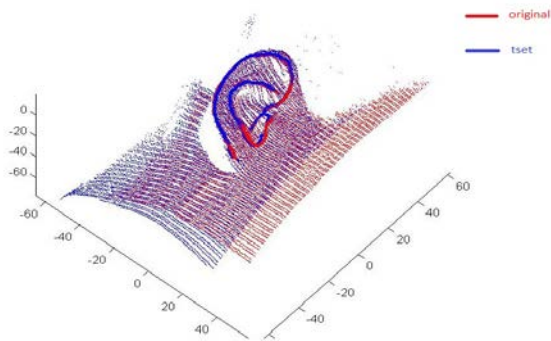


Figure 2. The pre-process of ear alignment

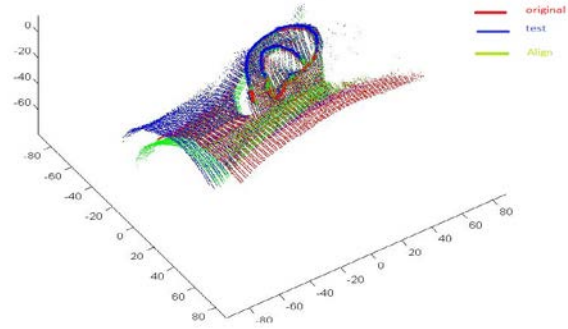


Figure 3. The result of ear alignment



Figure 4. The ear alignment in the presence of occlusion

Error	
Mean	STD
1.908	0.723

Table 1. Distance map error after alignment in the case of occlusion (Unit in mm)

5. DISCUSSIONS AND CONCLUSIONS

In this study, we introduced geometric-based methods to perform shape matching by aligning 3D surfaces. For the 3D-to-3D alignment, a novel collection of surface fiducial points, which are the points on the affine-invariant contours, e.g. parabolic contours, where the volume of parallelepiped spanned by two derivative vectors is zero, are computed. In addition, the fiducial points are preserved under affine transformations. To establish correspondences between the fiducial points on the two shapes, a set of absolute invariants were derived based on the volumes confined between parallelepipeds spanned by sets of the fiducial point quadruplets. Once the correspondences were established, the parameters of a relevant transformation were estimated and the two

surfaces were aligned. The performance of our method is demonstrated by the ability to register the 3D ear data scanned under a host of shape transformations, including ones that arise from change in ear position. Alignment errors, which were found to be within the 3D scanner resolution of 0.8 mm. This will be particularly relevant to applications where there is intra-class variability in the 3D, or due to the use of different modalities.

6. REFERENCES

- [Che07a] Chen, H., and Bhanu, B. Contour Matching for 3D Ear Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 29, no.4, pp.718-737, 2007.
- [Bur00a] Burge, M., and Burger, W. Ear Biometrics in Computer Vision. *Proc. 15th International Conf. of Pattern Recognition*, vol. 2, pp. 822–826, 2000.
- [Gov99a] Govindu, V., and Shekhar, C. Alignment using Distributions of Local Geometric Properties. *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 21, no. 3, pp.1031-1043, 1999.
- [Bes92a] Besl, P., and McKay, N. D. A method of registration of 3-D shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol 14, no.2, pp.239–256, 1992.
- [Pin13a] Pintavirooj, C., Cohen, F. S., and Tosranon, P. 3D face alignment and registration in the presence of facial expression differences. *IEEE Trans. Electrical and Electronic Engineering.*, vol 8, no.4, pp.395-407, 2013.
- [Sin09a] Singh, M., R. Brechner, R. R., and Henderson, W. V. Neuromagnetic Localization using Magnetic Resonance Imaging. *IEEE Trans. Med. Imag.*, vol 11, no. 1, pp: 129-134, 1992.
- [Woo98a] Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R. and Mazziotta, J. C. Automated Image Registration: I, General Methods and Intrasubject, Intramodality Validation. *Journal of Computer Assisted Tomography*, Vol. 22, No. 1, pp. 139-152, 1998.
- [Bes92a] Besl, P. J., and McKay, N. D. A Method for Registration of 3D Shapes. *IEEE Trans. Pattern. Analysis Mach. Intell.*, vol 14, no.2, pp:239-256, 1992.
- [Por94a] Porrill, J., and Ivins, J. A Semiautomatic tool for 3-D Medical Image Analysis using Active Contour Models. *Med. Inform.*, vol. 19, no.1, pp:81-90, 1994.
- [Kan81a] Kanal, L. N., Lambird, B. A., Lavine, D. and Stockman, G. C. Digital Registration of Images from Similar and Dissimilar sensors. *Proceedings of the International Conference on Cybernetics and Society*, pp: 347-351, 1981.
- [Gos86a] Goshtasby, A. Piecewise Linear Mapping Functions for Image Registration. *Pattern recog.*, vol 19, no.6, pp: 459-466, 1986.
- [Ibr98a] Ibrahim, W. S., and Cohen, F.S. Registration Coronal Histological 2D Sections of a Rat Brain with Coronal Sections of a 3D Brain Atlas Using Geometric Curve Invariants. *IEEE. Trans. on Medical Imaging*, TMI, vol 17, no. 6, 1998.
- [Do76a] Do Carmo, M. P. *Differential geometry of curves and surfaces*, Prentice hall, Englewood Cliffs, NJ. 1976.
- [Mil97a] Millman, R. S., and Parker, G. D. *Elements of differential geometry*. Prentice Hall, Englewood Cliffs, NJ. 1977.

A new 6D ICP algorithm with color segmentation-based adaptive sampling

Eung-Su Kim

School of Computer Science and
Engineering , Kyungpook
National University
80, Daehak-ro
Buk-gu
KS002 ,Daegu, Republic of
Korea
jsm80607@gmail.com

Sung-In Choi

School of Computer Science and
Engineering , Kyungpook
National University
80, Daehak-ro
Buk-gu
KS002 ,Daegu, Republic of
Korea
ellim5th@naver.com

Soon-Yong Park

School of Computer Science and
Engineering , Kyungpook
National University
80, Daehak-ro
Buk-gu
KS002 ,Daegu, Republic of
Korea
syPark@knu.ac.kr

ABSTRACT

In ICP-based algorithms, the closest points are considered as the corresponding points. However, this method fails to find matching points accurately when the initial position of the point clouds is not sufficiently close. In this paper, we propose a new method to solve this problem using six-dimensional (6D) distance, which consists of color information and three-dimensional (3D) distance, and color distribution matching. First, before finding the corresponding points using this method, a Gaussian filter is applied on the input color image. A color based image segmentation is done on that image and then n number of samples are randomly chosen from each segment. This process is applied in order to improve the computational time and performance. Second, corresponding point candidates are searched by solving a local minima problem using 6D distance. Then the color distribution matching is applied on these candidates to find the final corresponding point. Several experiments are conducted to evaluate the proposed method and the experimental results prove it has improved over the conventional methods.

Keywords

Iterative Closest Point(ICP), Point-to-Plane ICP, 3D registration, Color segmentation

1. INTRODUCTION

3D registration is a computer vision technique of aligning multi-view range images with respect to a reference coordinate system. Various 3D registration algorithms have been introduced in the past few decades. Iterative Closest Point (ICP) algorithm introduced by Besl and McKay [Besl92a] is one of the vastly used 3D registration algorithm, which got various modification later on. But ICP like algorithms have the local minima problem which is hard to solve. Many alternative algorithms have been introduced throughout the past few years to solve this local minima problem such as probabilistic and color ICP

methods. Among these methods, Point-to-Plane method, which was introduced by Chen and Medioni [Chen91a], minimizes the cost function value between the estimated point cloud and improves the performance by combining 3D surface information. However, Point-to-Plane algorithms also consider the closest points as the corresponding points and, as a result, there could exist some performance drawbacks as well as repeated iterations.

Considering these facts, in this research, we have introduced a new 6D ICP algorithm which manages to increase the matching performance by using the color information to find more accurate corresponding

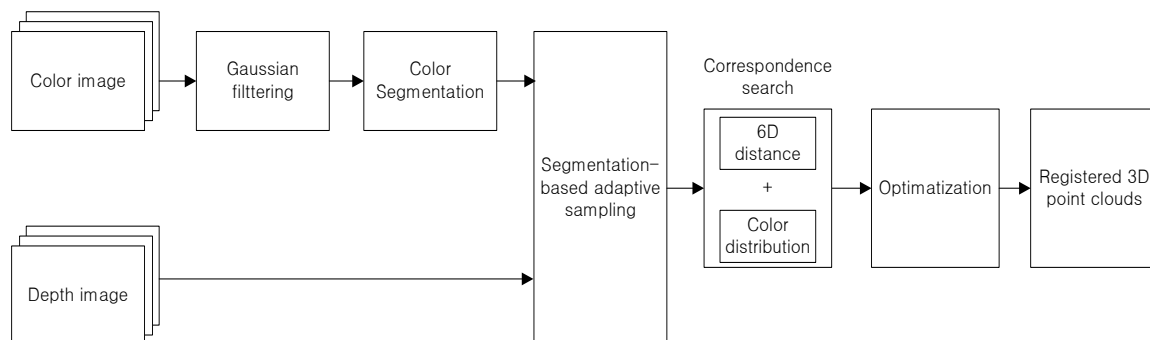


Figure 1. Flow chart of the proposed algorithm.

points. Johnson and Kang introduced a 6D color ICP algorithm [Johnson99a], which manages to find the matching information by implementing location information as well as color information. The introduction of finding corresponding points using color information contributes to solving the local minima problem that cannot be solved only using 3D information. However, the performance of this method deteriorates when the repeated identical color patterns exist.

In order to overcome this drawback and also to improve the performance of the algorithm, this paper uses 6D distance information as well as color distribution similarities while searching for corresponding points. Section 2 describes how to use 6D distance and color distribution similarity to find corresponding points, whereas Section 3 describes the use of the color segmentation-based adaptive sampling to improve the computation time and performance. Section 4 verifies the performance of the proposed method through experimental results and Section 5 gives the conclusion of the paper.

2. 6D DISTANCE AND COLOR DISTRIBUTION SIMILARITY

Algorithm 1 represents the pseudo code of the standard ICP algorithm. Modified ICP algorithms, such as Point-to-Plane ICP or Generalized ICP [Segal09a], increase the performance by improving the error minimization function represented in line 11 of the Algorithm 1. Even though the error minimization function is different in these methods, most of them use the same nearest neighbor algorithm

input: Two point clouds: $A=\{a_1, \dots, a_n\}$,

$B=\{b_1, \dots, b_n\}$

An initial transformation: T_0

output: The Correct transformation: T

```

1:   $T \leftarrow T_0$ 
2:  while not converged do
3:    for  $i \leftarrow 1$  to  $N$  do
4:       $m_i \leftarrow \text{FindClosestPointInA}(T \cdot b_i)$ ;
5:      if  $\|m_i - T \cdot b_i\| \leq d_{max}$  then
6:         $w_i \leftarrow 1$ ;
7:      else
8:         $w_i \leftarrow 0$ ;
9:      end if
10:   end for
11:    $T \leftarrow \operatorname{argmin}_T [\sum_i w_i \|T \cdot b_i - m_i\|^2]$ 
12: end while
    
```

Algorithm 1. Standard ICP

in line 4 to search the corresponding points. This nearest neighbor algorithm consumes less computational time for correspondence search as the closest point is considered the corresponding point. However, there is a limitation on solving the local minima problem as a reason of finding the corresponding point only using the 3D distance [Rusinkiewicz01a].

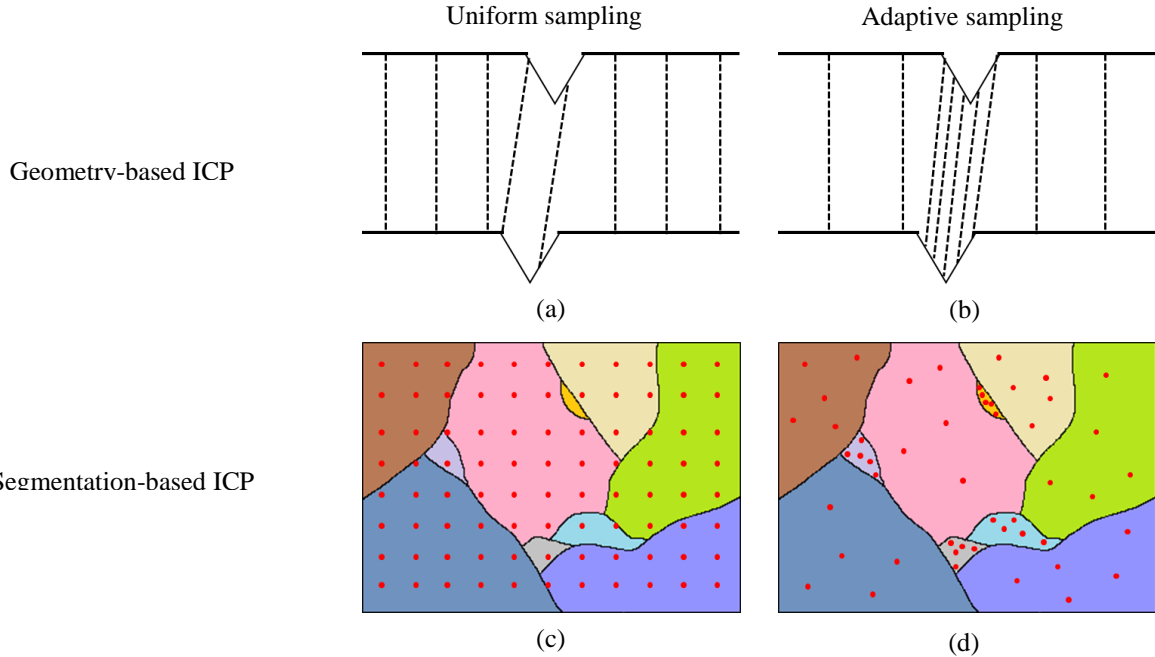


Figure 2. Two different sampling method used for corresponding point search in Geometry-based ICP and Segmentation-based ICP

In this paper, we improve the corresponding point searching method using the three-dimensional distance with the color information. Corresponding point search method proposed in this paper can be summarized in two steps. In the first step, corresponding point candidates are searched using 6D distance. In the second step, the color distribution of each candidate's neighboring points is matched with the query point's color distribution. The best matching candidate is selected as the final corresponding point.

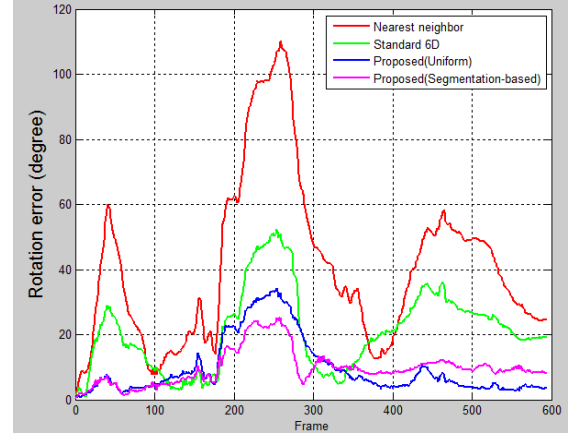
First, we assume there are two point clouds $A = \{a_i\}_{i=1,\dots,n}$ and $B = \{b_i\}_{i=1,\dots,m}$. Then, for each query point (a_i) on the cloud A, we search k number of candidate corresponding points (b_{i0}, \dots, b_{ik}) from point cloud B using the 6D distance defined in equation (1).

$$d_6 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 + \alpha_1(r_1 - r_2)^2 + \alpha_2(g_1 - g_2)^2 + \alpha_3(b_1 - b_2)^2} \quad (1)$$

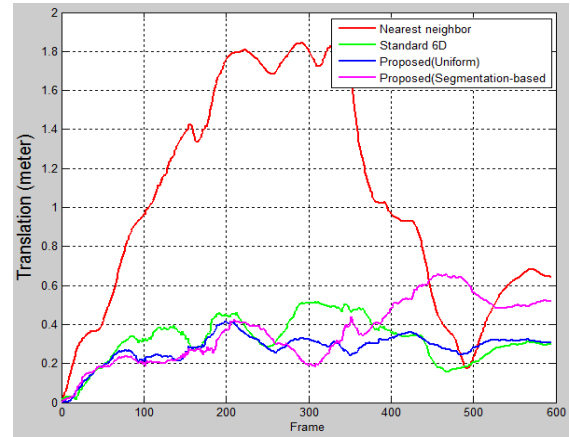
Here, $x_i, y_i, z_i (i = 1, 2)$ represents the 3D position of the selected two points while $r_i, g_i, b_i (i = 1, 2)$ represents the corresponding RGB color values. $\alpha_i (i = 1, 2, 3)$ is a experimentally determined weight coefficient for each RGB component.

In order to select the final corresponding point out of all the candidate points, the color distribution of the neighboring area of the candidate point (b_{ij}) and query point (a_i) is compared. Comparison of the color distribution is done by finding the first three eigenvectors and eigenvalues through the PCA (Principal Component Analysis) for each candidate's neighbors. If one of the calculated eigenvalues is too high compared to other two, that candidate point is considered as an outlier and excluded from the candidate corresponding points list. The reason is that, if one eigenvalue is excessively higher compared to others then the color of that point is very similar to surrounding area. Deciding the final corresponding point in such areas is difficult.

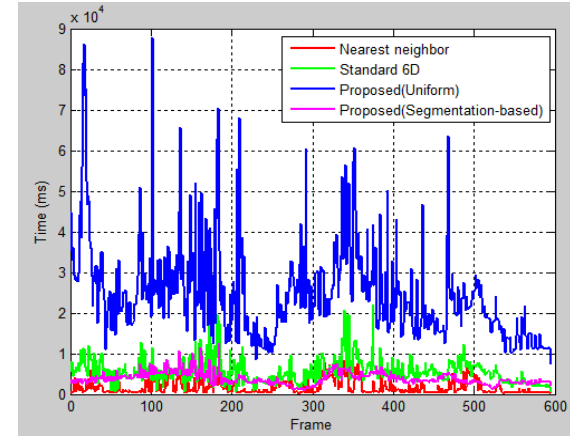
After excluding the outliers, remaining candidate points and a_i is converted to quaternion in order to calculate the error through the vector calculus. If the dot product between a corresponding point candidate b_{ij} and the reference point a_i is closed to 1, then the color distribution of these two points are considered to be similar. The candidate point that has the highest dot product value is considered as the final corresponding point.



(a)



(b)



(c)

Figure 3. Estimated position between the ground truth and each algorithm of rotation error(a), translation error(b) and computational time(c) in the world coordinate.

3. COLOR SEGMENTATION-BASED ADAPTIVE SAMPLING

The proposed method requires a lot of computational time because it calculates the PCA for all points of the point cloud. Uniform sampling is generally used in

order to reduce the computational time. However, uniform sampling is not suitable for point clouds, which have small features that are vital to determine correct alignment, such as in Fig 2. (a) and (b). Uniform sampling techniques (Fig. 2(a)) generally select only a few samples in these small feature areas. Computational time can be improved by doing dense sampling on feature areas and sparse sampling on the rest of the area as in Fig.2 (b).

A similar approach can be applied on the RGB-D image, which have both color and depth information. First, Gaussian filter is applied on the RGB image for minimizing the effect of motion blur. Then, a graph-based segmentation method [Felzenszwalb04a], which shows good performance and consume less time, is applied on the color image to do the color segmentation. Instead of doing uniform sampling such as in Fig.2 (c), an adaptive sampling technique, which selects same number of samples from each color segment as shown in Fig.2 (d), is used in our proposed method. This method makes dense sampling on small color segments, which are most like to be feature points, and sparse sampling on larger color segments, which are possibly not feature points. As a result, this method improves the computational time by reducing the total number of sample points and performance by increasing total number of samples on feature areas.

4. EXPERIMENT

In this section, we evaluate the performance of the proposed method using Freiburg dataset [Sturm12a] taken inside a general office environment. In the experiments, we evaluate for different corresponding point search methods; nearest neighbor, standard 6D, proposed method with uniform sampling, and proposed method with color segmentation-based adaptive sampling. Inside all these four methods, we use the same error minimization function, which is used in the Point-to-Plane ICP.

	Average rotation error (deg)	Average translation error (cm)	Time (sec/frame)
Nearest neighbor	1.12	1.45	1.74
Standard 6D	0.64	0.68	5.65
Proposed (uniform)	0.53	0.59	24.66
Proposed (segmentation)	0.54	0.71	3.89

Table 1. Average error of relative rotation and translation and computational time per frame

Fig.3 shows a comparison of the translation and rotation error of the proposed method with the conventional method in the world coordinate. Ground truth data provided with the Freiburg dataset is used to find the translation and rotation error of each four methods. Rotation error is calculated using the equation (2) and translation error is calculated by equation (3).

$$E_r = \arccos\left(\frac{\text{trace}(\Delta R)}{2}\right) \quad (2)$$

$$E_t = \|t_e - t_g\| \quad (3)$$

Here, ΔR can be represented by equation $\Delta R = R_e R_g^{-1}$ where R_g is the ground truth rotation and R_e is the estimated rotation by each method. t_g represents the ground truth translation matrix and t_e represents the estimated translation matrix using the four methods.

Fig.3 and Table 1 shows that the rotation and translation error of the proposed method is less than the conventional nearest neighbor method and standard 6D method. The proposed method with uniform sampling takes considerably high computational time compared to other conventional methods. However, using the proposed segmentation based adaptive sampling, we were able to reduce the computational time up to a comparable level with other methods.

Fig.4 shows the registration results of the four methods on the 200th frame. This figure concludes that the proposed method shows more accurate aligning results than the nearest neighbor or standard 6D method.

5. CONCLUSION

This paper proposed a new corresponding point search method using 6D distance and color distribution matching. And also this method is improved using color segmentation-based adaptive sampling. In order to evaluate the performance of the proposed method, we compared it with conventional methods using verified data set and the results prove it has improved over the conventional methods. As a future work, we are planning to improve the performance of the proposed method to achieve more accurate results in less time.

6. ACKNOWLEDGMENT

This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) support program (NIPA-2014-H0401-14-1004) supervised by the NIPA(National IT Industry Promotion Agency.) and also by the National Research Foundation of Korea funded by the Korean Government (NRF-331-2007-1-D00423).

7. REFERENCES

- [Chen91a] Chen, Y., and Medioni, G., Object Modeling by Registration of Multiple Range Image. Proc. of the IEEE Intl. Conf. on Robotics and Automation, pp. 2724-2729, 1991.
- [Besl92a] Besl, P. J., and McKay, N. D. Method for registration of 3-D shapes. Robotics-DL tentative. International Society for Optics and Photonics, 1992.
- [Felzenszwalb04a] Felzenszwalb, P. F., and Huttenlocher, D. P. Efficient graph-based image segmentation. International Journal of Computer Vision, 59(2), pp. 167-181, 2004.
- [Johnson99] Johnson, A., and Kang, S. B. Registration and integration of textured 3-d data. Image and vision computing, 17(2), pp. 135-147, 1999.
- [Segal09a] Segal, A. V., Haehnel, D., and Thrun, S. Generalized-ICP. in Robotics: Science and Systems, 2009.
- [Sturm12a] Sturm, Jürgen, et al. A benchmark for the evaluation of RGB-D SLAM systems. Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on. IEEE, pp. 573-580, 2012.
- [Rusinkiewicz01a] Rusinkiewicz S., and Levoy M. Efficient Variants of the ICP Algorithm. Proc. Third International Conference on 3-D Digital Imaging and Modeling IEEE, pp. 145-152, 2001.

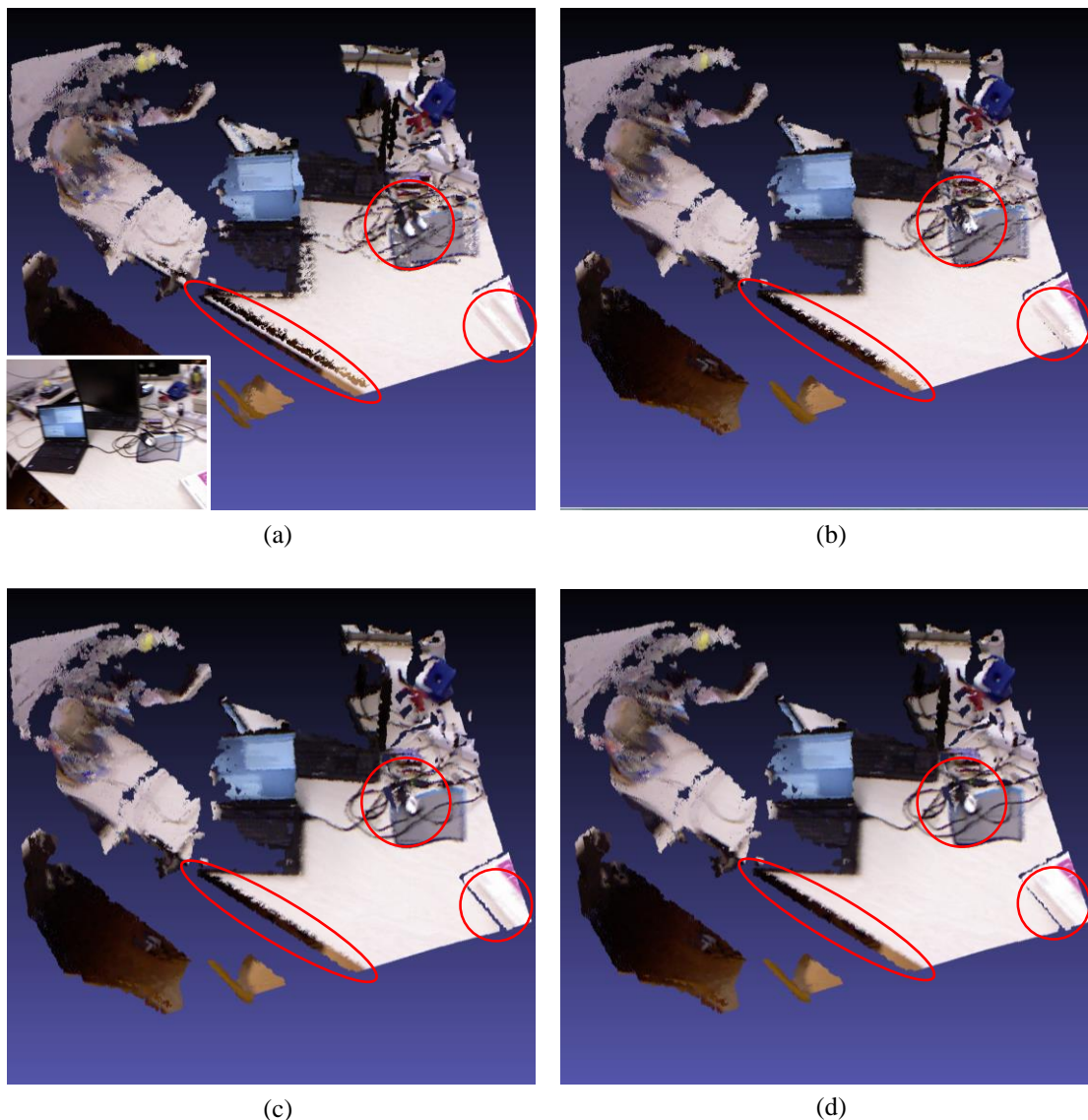


Figure 4. Registration results of the four method on the 200th frame sample. (a) Nearest neighbor (Reference image in bottom left corner). (b) Standard 6D. (c) Proposed method with uniform sampling. (d) Proposed method with segmentation-based adaptive sampling.

A COM-based Toolkit for Real Time Visualization

Stefan Maas

Westfälische Hochschule Gelsenkirchen
Medical Engineering Laboratory
Neidenburger Straße 43
45877 Gelsenkirchen, Germany
stefan.maas@w-hs.de

Heinrich Martin Overhoff

Westfälische Hochschule Gelsenkirchen
Medical Engineering Laboratory
Neidenburger Straße 43
45877 Gelsenkirchen, Germany
heinrich-martin.overhoff@w-hs.de

ABSTRACT

Collaborative software development in different languages is not unusual, but leads to minor resource utilization during collaboration as a result of porting or reprogramming needs. Additionally in cooperative projects, frequently legal and market economic issues prohibit an exchange of source code between the project partners. Combining modules from different languages is possible using the Component Object Model (COM). Additionally COM offers an efficient way to combine modules from several development teams.

To solve the common issues of collaborative software development and to fulfil the needs of a real time visualization toolkit, “RTVCOM” was designed and realized. To demonstrate the capability of this approach an example client was developed that combines COM components written in OpenCL C, OpenGLSL, C++ and C#. It that processes 3D+t ultrasound data at 45.2 MB/s reconstructs the associated volume data and visualizes them in real time. The visualization is fully interactive, and different pre- and post-processing filters can be applied.

Keywords

Collaborative software development, COM, GPU, real time rendering

1. INTRODUCTION

Collaborative software development in different languages is not unusual. But this leads to minor resource utilization as a result of porting or reprogramming needs. Additionally in cooperative projects frequently legal and market economy issues prohibit an exchange of source code between the project partners. The Component Object Model (COM) [Mic15a] was designed to support development of software using different languages and offers an efficient way to combine modules from several development teams. COM enables inter-process communication and applications with loose coupling and high cohesion.

Using graphics processing units (GPUs) in real time visualization applications is state of the art. While GPU host programs are implemented in languages like OpenGLSL or OpenCL C, clients on central processing units (CPUs) are written in languages like C++ or C#.

To facilitate collaborative development of real time visualization software in cooperative projects the

toolkit “RTVCOM” (“Real Time Visualization using the Component Object Model”) was designed and realized. RTVCOM allows the implementation of clients in different CPU- and GPU-languages to gain a great variety of application possibilities.

To evaluate the capabilities of RTVCOM an example client was created that processes 3D+t ultrasound data at 45.2 MB/s in real time. Volume rendering including pre and post-processing filtering contribute to a complex functionality of this client.

2. RECENT SOLUTIONS

The open source “Visualization Toolkit” (VTK) [Kit15a] and the “Insight Segmentation and Registration Toolkit” (ITK) [Kit15b] are widely spread in the medical visualization community. Other toolkits like the “Medical Interaction Toolkit” (MITK) [Ger15a] and frameworks like “MeVisLab” [Mev15a] are built upon VTK and ITK or integrate them. The advantage of these toolkits and frameworks is the immense range of functions and the acceptance in the community. But these toolkits are not fully realized for GPU execution and hence the algorithms are too slow for 3D+t volume visualization in real time.

[Gob08a] created a pure GPU ray casting framework for massive volumetric datasets. The algorithms of that framework would be fast enough in principle. But it is only laid out for handling static volumes.

Some GPU frameworks as presented in [Sch11a], [Mem11a] or [Chu10a] are dealing with medical

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

image segmentation or reconstruction, but not with continuous volume data.

Therefore, all of these toolkits and frameworks are incapable of visualizing volume data at rates mentioned above.

Simulator X [Lat12] is an example for software that is based on the actor model. It is fast enough to allow real time visualization with low coupling and high cohesion. But as the toolkits/frameworks above the actor model is not laid out to fulfil the needs of collaborative software development in different languages. Also the mentioned legal and market issues in cooperative projects are not regarded.

3. PROPOSED SOLUTIONS

RTVCOM was designed and realized to solve the issues mentioned in the previous section. While RTVCOM was primary implemented for visualizing medical image data it was laid out to include all kind of real time data in principle. RTVCOM was developed as a modular system consisting of COM components as “In-process-Servers” with defined interfaces. Clients are not part of the toolkit and can be implemented arbitrary.

3.1 Component Categories

RTVCOM includes components from six categories (Table 1), which were implemented by the authors and three different project partners in two collaborative projects:

- Reader

The category “Reader” contains interfaces to (raw) data providers like ultrasound devices or other devices with continuous 2D+t or 3D+t raw or image data streaming.

For prior recorded data support this category also contains readers for different file formats like DICOM or Insight Meta Image (.mha, .mhd). Components in this category are usually written in CPU languages.

- Preprocessor

COM-components for raw data processing or for processing image data before visualization belong to the category “Preprocessor”. This category consists of two subcategories:

- “GPU”

Real time processing filters can be found in subcategory “GPU”. These filters are written in OpenCL C or OpenGLSL and are used before

the visualization process. For example the conversion of raw data to image data can be found in this category.

- “CPU”

During development time it can be reasonable to write filters using CPU languages first. For example to examine the quality of a serial filter in C++ before parallelizing it using OpenCL C. These filters belong to subcategory “CPU”.

- Viewer

“Viewer” contains all graphical COM-components that were built to visualize data. This category is divided in two subcategories: “2D(+t)” and “3D(+t)”. Currently all viewer-components use OpenGLSL vertex and fragment shaders.

- Postprocessor

It is often useful to enhance visualizations with post processing algorithms (“filters”). This category is divided in the same subcategories that are used in “Preprocessors” for the same reasons.

- Helper

This category contains all components that support the visualization indirectly like e.g. a GUI element for manipulating the opacity transfer function of volume visualizations. (The opacity transfer function itself is part of “Viewer” components.) Components in this category are usually written in CPU languages.

- Import/Export

Import/Export components are used to include functionality from third party products like MATLAB. The “MLApp”-COM server allows external applications to use MATLAB-functions. Since MATLAB is too slow for real time processing, this component is only used during the first phase of filter development. In the second phase promising filters are usually ported to OpenCL C or OpenGLSL.

3.2 Interfaces

The interfaces of the components are standardized. This simplifies the data exchange between the components on the client side. It is possible to convert data from OpenCL C to OpenGL shaders and vice versa without leaving the host. To minimize memory consumption and to enable fast data transfer, generally memory addresses or pointers are exchanged over the interfaces.

Category	Reader	Preprocessor	Viewer	Postprocessor	Helper	Import/Export
CPU	X	X		X	X	X
GPU		X	X	X		

Table 1. Overview of RTVCOM categories

3.3 Sharing Components

Sharing RTVCOM components is currently realized by a Microsoft Team Foundation Server 2010. This server is part of a Microsoft SharePoint Server 2010. These servers are the base of the quality management system used for developing RTVCOM.

Depending on the settings of an integrated rights management system the developed components can be shared, for example, with project partners without revealing internal knowledge. Also jointly created clients are possible, where each partner can use the COM components of the other partners. Source code access is only possible for developers with rights in accordance to the rights management system.

3.4 Client Creation

Clients can be built by assembling components from the given categories. While the usage of “Reader”- and “Viewer” components is mandatory, components

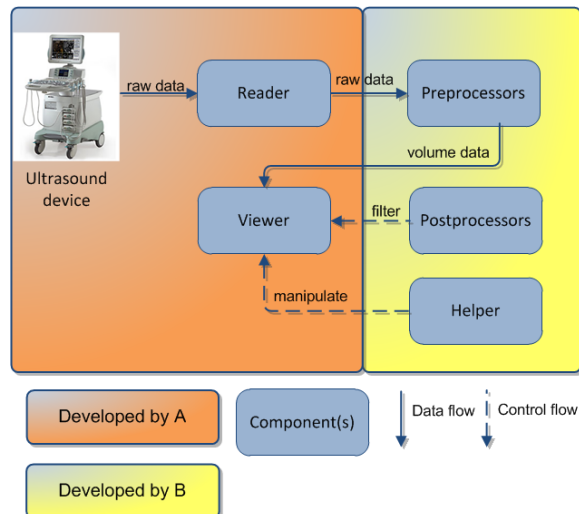


Figure 1. Example for a jointly created client regarding developers

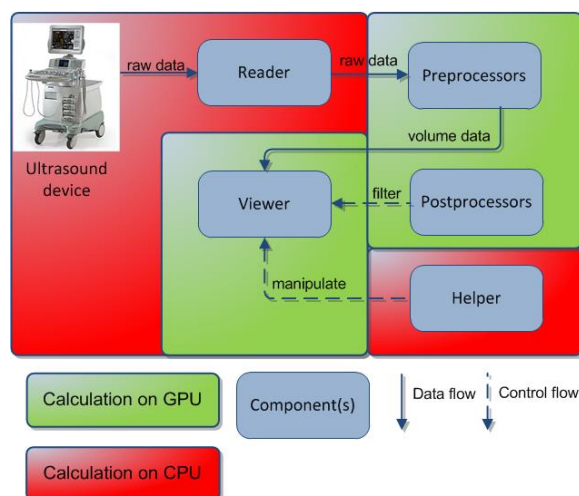


Figure 2. Example for a jointly created client regarding processing location

from other categories can be used optionally in type and numbers.

A client, that was created by the authors and one industrial partner during a jointly project, is shown in Figure 1. In this client the industry partner (“A”) provides an ultrasound device including software components of category “Reader” and “Viewer”. This software allows only a conventional view on ultrasound images. The authors “B” integrated components from category “Preprocessor”, “Postprocessor” and “Helper”. During development “B” is able to use the components from “A” but cannot access their source code due to missing access rights. After the project ends “B” can provide “A” with the developed components without revealing any knowledge.

A different view on this example client (see Figure 2) illustrates that the use of RTVCOM enables both partners to create GPU and CPU components and merge them in a joint client.

4. EXPERIMENTAL RESULTS

To demonstrate the capability of RTVCOM an example client was developed that combines COM components written in OpenCL C, OpenGLSL, C++ and C#. It acquires raw data streams at 45.2 MB/s and 50 Hz from an ultrasound device, reconstructs the associated volume data and visualizes the data in real time.

The used hardware consists of an Intel Core2Duo, 2.6 MHz CPU, PC with an NVIDIA Geforce 760 GTX. On this system the visualization is fully interactive, and different post processing filters can be applied without losing the real time capability.

4.1 Client Details

The data acquisition, data flow and visualization process of this C#-based client (see also Figure 1) in detail:

1. The first component (“Reader”; written in C++) starts the data acquisition from the ultrasound device. Raw data (short values, 8 bit) are transferred at 50 Hz as 2D+t slices from the ultrasound device via a network card to the visualization computer.
2. Within the second component (“Preprocessor”; written in C++ and OpenCL C) the short values are uploaded to the GPU. Due to a better float value support of GPUs the short values are converted to float values in the first OpenCL kernel. In the next step the volume data are recalculated from the raw data using four different OpenCL kernels.

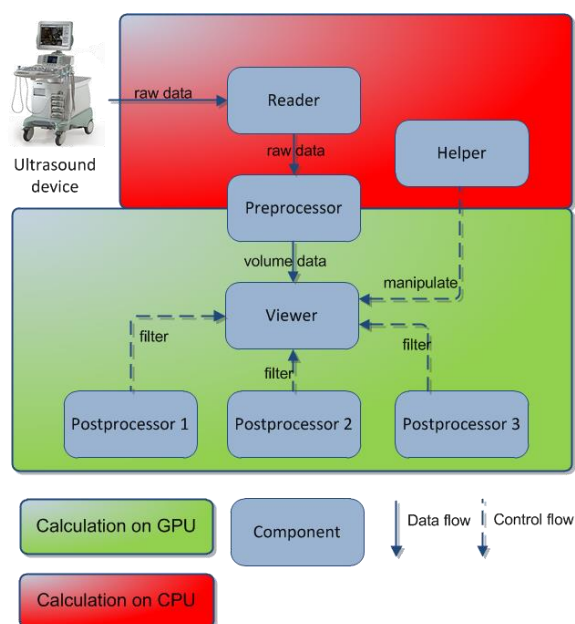


Figure 1. Schematic overview of example client

3. The next component (“Viewer”; written in C# and OpenGLSL vertex and fragment shaders) visualize the reconstructed volume using a raycasting technique. This 3D+t volume is updated every incoming slice resulting in an update rate of 50 Hz.
4. Three filter components (“Postprocessor”; written in C# and OpenGLSL compute shaders) were added for enhancing image quality. These filters can be activated/deactivated interactively.
5. Additionally the visualization of the volume can be optimized by manipulating the opacity and grayscale transfer functions using another component (“Helper”; written in C#).

4.2 Comparison

Before the example client was implemented, the visualization was realized by a single main program. This data and control flow was exactly like in the example client but without any COM components. Because COM is said to be slow regarding data transfer, a performance comparison between both implementations were implemented. It revealed no measurable difference. Both implementations were able to visualize the 3D+t data at 45.2 MB/s.

5. CONCLUSION

It has been demonstrated that RTVCOM and RTVCOM-based clients can solve the common issues of collaborative software development and to fulfil the needs of a real time visualization toolkit:

- Support of collaborative software development in different CPU and GPU languages

- Enabling development of jointly created clients in cooperative projects regarding legal and market economy issues
- Visualization in real time

The advantage as well as disadvantage is the fact that RTVCOM is not laid out to reveal source code. So this approach cannot directly be used for open source projects.

Future works will extend RTVCOM to support the Distributed Component Object Model (DCOM). This will simplify the exchange and replacement of COM objects and make the Team Foundation Server obsolete for this purpose. Additionally the number of filters will be enlarged to increase visualization quality for specific applications. Furthermore readers for other types of data like real time image data from magnetic resonance tomography (MRT) will be implemented.

6. REFERENCES

- [Chu10a] Chunlan, X., and Anyuan Z., and Liu, D.C. Optimized GPU Framework for Ultrasound B-Mode Imaging. Bioinformatics and Biomedical Engineering (iCBBE) 2010, pp. 1-4, 2010.
- [Ger15a] German Cancer Research Center. MITK. <http://mitk.org/wiki/MITK>.
- [Gob08a] Gobbetti, E., and Marton F., and Iglesias Guitián, J.A. A single-pass GPU ray casting framework. The Visual Computer, pp. 797-806, 2008.
- [Kit15a] Kitware, Inc. VTK. <http://www.vtk.org>.
- [Kit15b] Kitware, Inc. ITK. <http://www.itk.org>.
- [Lat12a] Latoschik, M.E. and Tramberend, H. A Scala-Based Actor-Entity Architecture for Intelligent Interactive Simulations. Software Engineering and Architectures for Realtime Interactive System (SEARIS) 2012. 5th Workshop on, pp. 9-17, 2012.
- [Mem11a] Membarth, R., and Hannig F., and Teich J., and Körner M., and Eckert W. Frameworks for GPU Accelerators: A Comprehensive Evaluation using 2D/3D Image Registration. 2011 IEEE 9th Symposium on Application Specific Processors (SASP), pp. 78-81, 2011.
- [Mev15a] MeVis Medical Solutions AG. MeVisLab. <http://www.mevislab.de>.
- [Mic15a] Microsoft Corporation. Microsoft COM. <https://www.microsoft.com/com/default.msp>.
- [Sch11a] Schmid, J and Iglesias Guitián, J.A., and Gobbetti, E., and Magnenat-Thalmann, N. A GPU framework for parallel segmentation of volumetric images using discrete deformable models, The Visual Computer, pp 85-95, 2011.

Virtual Museum as an Environment for Visual Representation of Urban Planning Concepts and Lost Architectural Objects

Yulia Petrova

Saint-Petersburg State University
7-9, Universitetskaya nab.
199034, St.Petersburg, Russia
design5566@gmail.com

ABSTRACT

Intensive development of the Web has significant influence on social communication processes. New trends of information distribution demand new approach as being involved in on-line communication. Social institutions, including museums and exhibition centers, aim to develop their presence on the Web, searching for ways to engage bigger number of visitors by offering them new experience.

A phenomenon of the virtual museum is studied in the paper as an online representation of a tangible museum collections and as an information resource that has no physical analogue. The target is to analyze the potential and relative advantages and disadvantages of virtual museums.

The main subject of the research is effectiveness of online tools for representation of large architectural and urban objects, unimplemented projects and lost monuments.

Virtual exhibition is often based on the products of virtual reconstruction. It is very important to formulate strict methods of this approach, to develop a methodology and establish a practice of visual distinction between true (and/or relatively true) parts of a virtual model and its authorial parts.

Keywords

Virtual, museum, urban, architecture, heritage, culture, reconstruction, exhibition.

1. INTRODUCTION

Intensive development of information technologies brings influence on all aspects of human being life, including art and culture. From a perspective of museology computerization and informatization of society have led to development of new methods of working with collections and exhibits as well as communication with museum audience.

On-line presence of museums obviously becomes a requirement for effective communications. Museums are interested to put their collections online not only for the purpose of preserving the cultural heritage, but also to make the information content accessible for potential visitors.

Recently, a number of surveys have provided compelling evidence that online museums actually drive physical museum attendance instead of discouraging physical visits [Mar07].

The information on the museum's website, is available to a wide audience (including professionals working in various museums), which brings museums possibility to associate with each other, to identify similarity and find partners. Virtual museums are beginning to form a new type of global virtual museum environment, accessible to everyone and customizable to the needs of each user. Therefore, it is important to study the methods of presentation of the museum collections on-line, the factors forming quality information content and the effectiveness of the instruments exhibiting objects in virtual reality.

2. VIRTUAL MUSEUM DEFINITION

According to ICOM definition, "museum is a non-profit, permanent institution in the service of society and its development, open to the public, which acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

humanity and its environment for the purposes of education, study and enjoyment” [Ico15]. Today museum is a social and cultural center for edutainment, pedagogic etc. purposes.

Definition and above-mentioned functions of physical museums are partly applicable to on-line museums.

The virtual presence of museums may involve a variety of forms, differing in their technology immersion but similar in their appeal to potential visitors. In general there are three main types of online museums:

1. Physical museum representation (digital museum) – the internet-based version of a collection that exists in the real world.
2. Virtual museum (online museum, electronic museum, web museum) – museum that can present either digital copies of works or works that have no tangible analog.
3. Virtual museum that shows works that exist only in the virtual space.

The first type of on-line representation supports and accompanies collections of a physical museum, implements representative, advertising and promotional functions. These web resources are actually data bases, where digital copies of exhibits, photo-, audio-, video materials and other corresponding multimedia products are stored and displayed.

There is no standard definition prevailing for the term ‘virtual museum’ [Syl09]. However, it is evident that virtual museum greatly differs from digital library, catalogue or data base. Following the research in this field, “only those websites, which are implemented in compliance with principles of museum exhibition design, can be called virtual museums. It means that virtual museum as a physical one should have scientific concept, architectural and art concept and exhibition scenery [Nic99]. Virtual museum shall be based on well-developed scientific concept and design solution.

Main criteria of virtual museums are as follows:

1. Representativeness and content completeness precluding distortion of facts, which causes misconception of history or a particular epoch;
2. Multiplicity and multi-layer structure of presented information to make it apprehensible for different audience;
3. Clear and user-friendly interface, strict structure, efficient navigation and quick search.

3. OPPORTUNITIES AND ADVANTAGES OF VIRTUAL ENVIRONMENT

Virtual museum with thought-out conception has certain features, which make it different from representative web site of a real museum. These features bring new opportunities for exhibits’ displaying and for communication with museum audience. Virtual technologies enable qualitative development of museums as educational and social centers. Audience of virtual museums gets access to collections on-line, however, it does not decrease visitation of real museums – quite the contrary – it helps to attract new audience, especially young people: schoolchildren and students.

Today there are quite many examples of virtual museums on the Web, such as on-line collections of Vatican museum, the multimedia guide of the Louvre-Lens museum, on-line collections of the State Hermitage Museum and others. Comparing to real museum, a virtual one has several significant advantages, especially when we deal with large objects (for example, urban or architectural), which cannot be displayed in museum rooms at all, or it would be very complicated to place them indoors. It is also a challenge to exhibit decrepit objects, which require certain temperature and humidity conditions.

Virtual museums can provide access to exhibition from any space to all end-user groups including people with special needs (visual, acoustic, speech and other disabilities). In virtual environment exhibit can be observed from all angles, so user can obtain information that often cannot be shown in the physical museum [Syl09].

Virtual museums do not face a common problem of traditional museums - preservation of exhibits (reducing of natural dilapidation, protection from consequences of force majeure circumstances including natural disasters and disorders). Costs of making on-line collection or exhibition do not include expensive transportation and overheads. Virtual exposition can include exhibits from private collections and museums located in different parts of the world [Kis08].

Worldwide images posted on the Internet are protected by copyright laws. In recent years, the use of licensing terms and conditions to control access to and use of art images has been subject to increasing criticism, primarily when considering restrictions imposed on works that are in the public domain [Cre12]. Many institutions have changed common museum practice and have begun to make their collections available digitally, with the option to download a high-quality image suitable for

publishing (The Victoria & Albert Museum, the British Museum, the National Portrait Gallery, etc.).

Despite these developments, many libraries and museums continue to impose restrictive terms and conditions on their digital art collections, as well as assert their ability to control the use of digital reproductions of public domain artwork through copyright [Cre12].

Museum use different methods to protect their digital collections: exhibition of lower quality images or images with watermarks, protection against direct downloading of images, requirement to accept the terms of use for downloading high quality images.

4. VIRTUAL RECONSTRUCTIONS

It is a common situation when pieces of art and culture are partially or completely lost or significantly modified. In this case virtual reconstruction is an efficient tool. Results of reconstruction can be used as a basis for a new virtual museum.

The term virtual reconstruction implies that the representation takes place in a three dimensional space, which is usually called virtual environment and the final product is usually called a 3D virtual model. Virtual reconstruction is a technology for making models of different complexity. These can be small exhibits, pieces of arts and crafts, monuments, historical plans and maps, buildings and large architectural ensembles, city areas and suburban landscapes, etc.

3D virtual reconstructions significantly support studies for the eventual real reconstruction of the monument in the future. A virtual reconstruction would also enable the examination of various alternative solutions and help making decisions for the suitable restoration or reconstruction methods [Geo14]. Using this technology we handle digital copy of an object. It means that when we work with a piece of art or craft, or architectural object in virtual space, we do not bring any harm to its original prototype.

3D modeling has been originally used for buildings' designing. Up to the present moment it is especially efficient for handling architectural objects and their fragments. The result is a 3D model, which can be presented in different ways:

- 2d and 3d graphics;
- Interactive plans and maps;
- Interactive panoramas;
- Virtual excursions;
- Virtual game-tours;
- Mobile application using augmented reality technology (AR);
- Virtual reality (VR).

Interactive maps and panoramic views are widely used on web sites creating virtual tours for city exploration, tourism and sightseeing. A map as an element of interface makes user's interaction with website much easier and more efficient, especially if the content includes large architectural and urban objects.

Panoramic images and videos are a 360° representation of a certain scene. Video objects, dynamic 3D computer models, or spatial audio can be embedded in order to vitalize the scene. Interactive panoramas are developed to the extent of an excursion with a variety of routes and exhibits, which full descriptions are available for user just by clicking the mouse. Thus, large high quality environments can be created that enable the user to immerse into the virtual world.



Figure 2,3. Interactive maps of St.-Petersburg (1720th, 2010th). VM of the Neva river delta



Figure 4. Example of an interactive panorama. Palace embankment n 1720th in St.-

Multimedia products based on AR technology are supported by mobile devices and as well as the portable VR headsets can efficiently enhance and enrich the 2D and 3D representation of artefacts.

An example of AR application is Multimedia Information Resource «The Church of the Savior on Ilyina Street in Novgorod the Great», which represents lost painting of the church. In this project a basis for interactive multimedia product is a completed 3d reconstruction [Las13]. One of the biggest concerns is how [Den12].

5. ISSUES AND CHALLENGES IN VIRTUAL RECONSTRUCTION OF ARCHITECTURAL MONUMENTS

Despite broad experience in applying virtual technologies, methods of reconstruction are still not

sufficiently developed from a perspective of standardization and scientific approach [Kis08]. The main problem is that all elements of reconstructed space are usually done in the same manner: with same detail design and of same hues.

Difficulties in virtual reconstruction arise from quality of documents, which are used as a basis for modeling, and interpretation of these documents. [Kis08]. The main principle of reconstruction is maximum conformance of a model with text description, illustrations and other information regarding original object. Available information from archive sources is often controversial and almost always incomplete. Researchers and designers have thus to make a decision whether to leave the model uncompleted as well or to finish the job using not only relatively accurate data from archives, but also analogues. Virtual reconstruction based on analogues allows several virtual models of one object, and all these versions are adequate, if they conform with any of available analogues. However, these models are not true copies of an object. For the purpose of supporting scientific approach, it is important to develop a method and establish a practice of visual distinction between true and/or relatively true parts of a model and authorial parts.



Figure 5. Winter Palace n 1720th in St.-Petersburg. Using different hue and references



Figure 6. Le Blond's master plan for St. Petersburg. Unimplemented project. 1716-1717. Using different hue and references

Solutions to the mentioned problems are:

1. Distinction between visualization of authentic and fabricated parts of a virtual model:
 - By different design technics;
 - By different level of detail;
 - By different hue;
2. Using a system of references to archival materials.
3. Using a video showing step-by-step sequence of developing the model.

Therefore methods of virtual reconstruction should be developed closer to methods of classic restoration, when particular attention is given to visual distinctions between evidence and hypothesis, and between different levels of probability.

6. CONCLUSION

Nowadays virtual museums have become very popular due to many reasons and their number is on the rapid increase. Virtual museum can act as a complementary and auxiliary service for physical museum, or it might show reconstructions that exist only in the virtual space. Distinctive feature of virtual museum is that it provides unimpeded access to digital exhibitions and gives almost unlimited room for large architectural exhibits, which cannot be placed indoors. Virtual museums do not face a problem of exhibits' preservation like protection from consequences of force majeure circumstances.

Virtual reconstruction is an effective tool for representing modified or lost objects in their original view, but methods and rules of implementation are still not fully scientifically unstructured and standardized. It is still unclear what is in priority for end users: photorealism of a model, its authenticity and accuracy or just general visual impression. But nevertheless 3d reconstructions and interactive tools are very popular among visitors and can perform traditional museum functions online such as: acquisition, preservation, research, communication and education.

7. REFERENCES

- [Ico15] <http://icom.museum/the-vision/museum-definition/>.
- [Nic99] Nikishin N.A. 1999. Museum and New Technologies. On the way to the Museum of the XXI century. Moscow: Progress-Tradition.
- [Syl09] Sylaiou Styliani, Liarakapis Fotis, Kotsakis Kostas, Patias Petros. Virtual museums, a survey and some issues for consideration. Journal of Cultural Heritage 10, pp. 520–528, 2009.
- [Mar07] Marty P.F. Museum Websites and Museum Visitors: Before and After the Museum Visit. Museum Management and Curatorship Volume 22, Issue 4, pp. 337–360, 2007.
- [Kis08] Kissel O.M., Potapenko N.V. Virtual Reconstructions as Interpretation of Art Heritage. Theses of the 11th Annual International Conference "EVA 2008 Moscow", 2008.
- [Geo14] Georgopoulos A. 3d Virtual Reconstruction of Archaeological Monuments. Mediterranean Archaeology and Archaeometry, Vol.14, No4, pp. 155–164, 2014.
- [Cre12] Crews, Kenneth D., Museum Policies and Art Images: Conflicting Objectives and Copyright Overreaching. Fordham Intellectual Property, Media & Entertainment Law Journal, Vol. 22, p. 795, 2012.
- [Las13] Tatyana V. Laska, Irina V. Tcymbal, Sergey V. Golubkov Multimedia Information Resource "The Church of the Savior on Ilyina Street in Novgorod the Great" Procedia Computer Science, Volume 25, 2013.
- [Den12] Denard, Hugh. A New Introduction to the London Charter. Paradata and Transparency in Virtual Heritage. Digital Research in the Arts and Humanities Series. Ashgate, 2012, 57–71.

POSTER

The creating serious games for historical subjects at the 1st level of primary school and such as a part of exhibition chosen historical buildings

Lenka Chadimova

Univerzita Hradec Kralove

Rokitanského 62

500 05, Hradec Kralove,

Czech Republic

Lenkachadimova@gmail.com

ABSTRACT

The aim of the study was create serious games for historical buildings which are important for each region of Czech Republic. These games are making guide to historical building and give to pupils the information about architecture, history of building and eventually about surroundings of historical building and the like.

For creating serious games are using tools of 3D graphic software 3Ds MAX in combination with developing software UNITY 3D. In the modelling software 3Ds MAX were creation 3D models of historical buildings. The gaming environment was creating in UNITY 3D with programming language C#.

Keywords

Serious games, education, 3D, UNITY 3D, interactive.

1. INTRODUCTION

For approaching a cultural heritage, especially historically important buildings, to students of the primary school with attractive way due to which is easier to remember their subject matter, is the best way to use elements of gamification. Children are confronted with a virtual environment from an early age, primarily through an entertainment industry, whether they are films (mainly animated cartoons), or computer games. Due to this fact students get up close to this environment and they are well-versed in this field.

As the main historical building in this series of serious game was chosen a medieval castle in Hradec Kralove for several reasons. Both because of close cooperation with primary schools in Hradec Kralove, secondly because it is burned-out and now non-existent castle.

Given the fact that the castle is no longer existed in any form (not even as a ruin) there is an opportunity of sub-objective research – a comparison of children's way of learning, respectively they remember an information about historical buildings, which they are able to see both live (and to look at them indeed) and virtually and which they only learn about.

The castle complex of Hradec Kralove, a castle and Minorite monastery, was destroyed and burned out during The Hussite Wars (a civil war in the first half of 15 century), the monastery in 1419 and the castle in 1423 (Richter, Vokolek, 1995). The castle complex have never been rebuilt and later in its place were built town houses. An appearance of the castle is captured only on several vedutas, the other documentation does not exist. Due to this fact was a virtual 3D reconstruction of castle complex made by analogy with till now non-existent medieval castles, along with consultations with experts from Museum of Eastern Bohemia in Hradec Kralove. These information were supplemented by graphic designs, sketches and other documents, which originally served for a creation of paper model, which is the part of permanent exposition of Museum of Eastern Bohemia in Hradec Kralove.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

The gamification required is achieved by insertion of 3D model of historical building into game environment and by interactivity. In this environment is able to browse through building, take a look at interiors and exteriors, get information about the specific architectural, historical and other elements and get acquainted with local legends and with history of everyday life. To verify the acquired knowledge there is a test as a part of this educational game.

2. The creation of serious games

The preparatory works

The preparatory work with proposed ground plans and cuts was carried in vector editor Inkscape. From proposals there were created concrete documents for the creation of 3D model of castle complex.

Textures of outdoor plasters used in 3D model of castle complex were created from real outdoor plasters' photos of historical buildings situated in the centre of Hradec Kralove. Other textures was taken from photos of random objects, which had required colour and surface. A photo processing for creation of textures was carried in raster editor Gimp 2.8.

The 3D modelling

All modeling work were carried in 3Ds MAX 2011 and 3Ds MAX 2014 softwares.

3D virtual reconstruction of castle complex buildings were created in three ways:

- a) The basic objects – *“Standard Primitives”*,

mainly tip *“Box”* and *“Cylinder”* and with their deformations and modifications and further with combination with Boolean operations (*“Boolean”* or *“ProBoolean”*) if it was necessary.

- b) The objects *“AEC Extended”* and their modifications.
- c) Sketching an outline with tool *„Line“* a its spatial rendering with command *„Extrude“*.

A selection of tools for creation of particular part of castle was dependent on complexity of castle's element. If it was a simple object with no need of other adjustments, for example to add windows only, then a castle part was created using objects as *„Door“* from offer *„AEC Extended“*. If there was a castle element of irregular floor plan, a command *„Convert to*

Editable Poly“ was used. By modifications of positions as *„Vertex“*, *„Edge“*, *„Border“* or *„Polygon“* was achieved an effective distortion of object and an irregular floor plan. With use of these deformations there were created roofs, church support system, battlements etc.

The Boolean operations were used mostly for creation of passages, passes (f.e. an entrance gate) and monastic arcades.

A tool *„Line“* from offer *„Splines“* and his spatial rendering with help of commands *„Extrude“* was used for creation of railings, galleries etc.

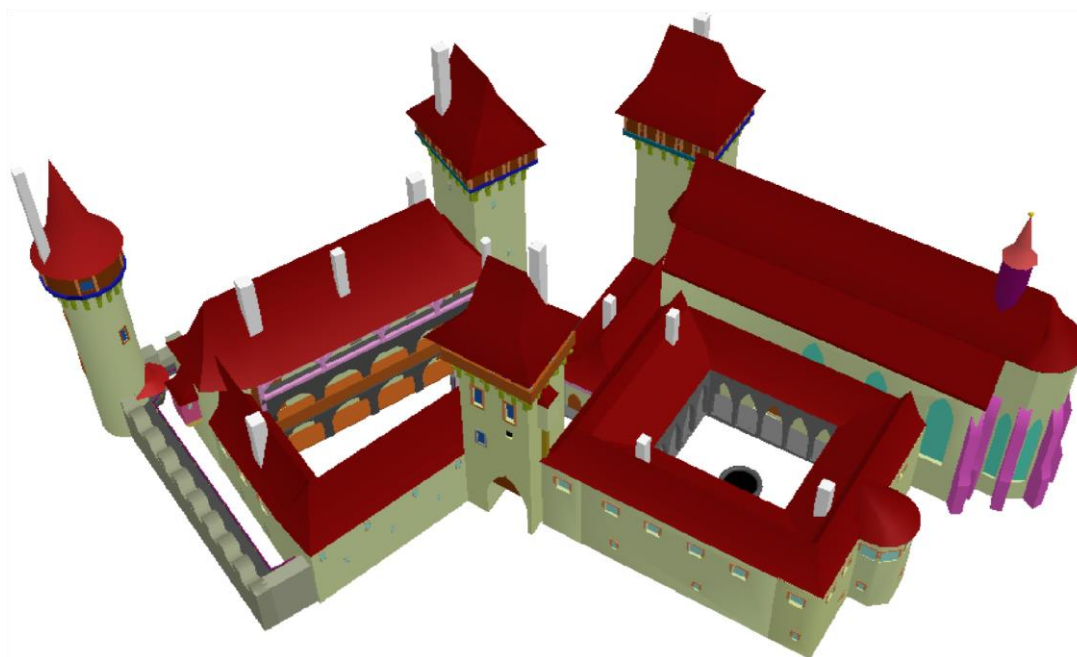


Fig. 1. Completed 3D model of the castle complex in Hradec Kralove. Source: The author.

After a finish of modelling works there were assigned to particular structural elements in advance performed textures, with help of tools from „Material Editor“. The last step in software was an export of 3D model to FBX format.

The developing of gaming interface

The environment of interactive presentation enriched with elements of gamification was developed in software Unity 3D. After putting imported 3D model of medieval castle into scene, here was a completion of visual scene, as castle surroundings, trees, access road, green vegetation, sky etc. Here was set also a lightening of castle.

For creation of „walk-through“ presentation with mini map, i.e. watching a movement of user from above, there were put tools „*First Person Controller*“ to the scene, which form de facto a figure of user and at the same time scan the whole scene and „*Camera*“, which is situated above „*First Person Controller*“ scans its movement. A performance properties and linking of these tools was defined by statements of C# language, which are scripts „*Keyboard Camera Control.cs*“ and „*FollowCamera.cs*“. Due to the last named script there was created preview window Mini maps for monitoring a position of user.

The usage of a map is very important for secondary school children, because they have different spatial perception than adults. Without Mini map children would get lost in virtual 3D model of medieval castle.

This interactive presentation is followed by other elements of gamification. Among the most significant element is a menu, where a user chooses a thematic area (History, Architecture, Legends etc.) and dropdown „information tables“ with lessons, which will be unrolled after clicking on highlighted element (f.e. pointed window). All these interactive elements are defined in programming language C#.

Finished serious game is exported to the user environment on the platform „Web Player“, where is boot file in HTML format and a source file in

UNITY3D. This platform is chosen because of easy running. A school and institutions using these serious games, are not forced to buy a new technology.

3. The Gamification and pedagogical research

Gamification is a setting up new game elements to a different environment, in this case the environment of national history on a primary school. A term gamification is spread from 2010 (Deterding, 2011). The main goal of gamification is to motivate students (serious game users) by achieving rewards.

From a study from July 2012 follows, that gamification is approximately from 5 to 10 years before reaching the top of mass using, as the Garhnerov curve Hype cycle shows. (See Fig. 2) (Maněnová, Chadimová, 2014).

The aim of the pedagogical research was to test what influence had an implementation of serious games for historical buildings on communication between teacher and pupils and the whole interaction in the frame of subject matter. On the base of analysis of lessons, during which serious game was using, and inquiry questions (the interview was with teacher and with pupils) was by means of category system in modified scheme Flanders system of interaction analysis was detected the interaction is higher than 1, it means the pupils are more active than teacher during lessons. (Maněnová, Chadimová, 2014)

For basic quantitative processing and for obtaining demanded data program CodeNet was used. The program was developed at department of pedagogy and psychology of

Pedagogical Faculty, University of Hradec Králové (authors: T. Svatoš and V. Žák). The CodeNet defines up to 20 different followed categories of activity and allows them to assign a code interval and further it is possible to make time cuts in it. (Svatoš, Doležalová 2011).

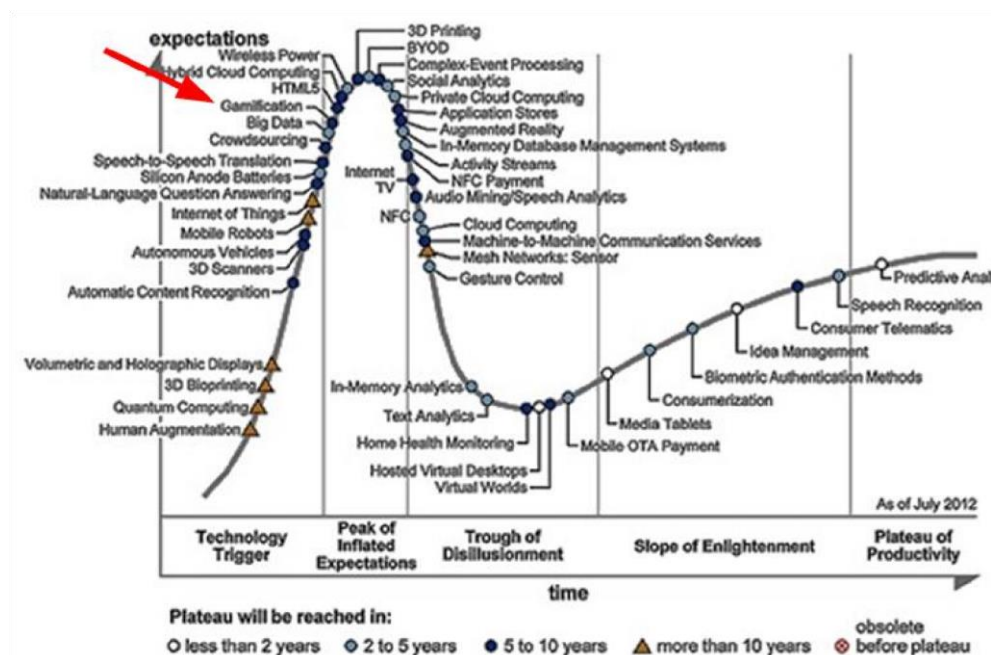


Fig. 2. Garthner curve of Hypo cycle
Source: Maněnová, Chadimová, 2014

4. REFERENCES

- [Det01a] DETERDING, Sebastian, et al. From game design elements to gamefulness: defining gamification. In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. ACM, 2011. p. 9-15.
- [Fla01a] Flanders, N. A. (1970). *Analyzing Teaching Behavior*. London : Addison-Wesley, 1970.
- [Sva01a] Svatoš, T., Doležalová, J. (2011). *Pedagogická interakce a komunikace pohledem vývoje kategoriálního systému*. Hradec Králové: Gaudeamus, 2011. ISBN 80-7041-176-7.

- [Jou01a] Maněnová, M and Chadimová, L. (2015) '3D Models of Historical Objects in Teaching at the 1 st Level of Primary School', *Procedia – Social and Behavioral Sciences - 5th ICEEPSY International Conference on Education & Education Psychology (ICEEPSY 2014)*, Vol. 171, 16 January 2015, pp. 830-836.

- [Ric01a] RICHTER, Miroslav a VOKOLEK, Vít. *HRADEC KRÁLOVĚ; Slovanské hradiště a počátky středověkého města*. Muzeum východních Čech v Hradci Králové a Archeologický ústav Akademie věd České republiky, 1995. ISBN 80-85031-14-0.

Improving Situational Awareness in Military Operations using Augmented Reality

Alejandro Mitaritonna¹

¹ Instituto de Investigaciones Científicas y
Técnicas para la Defensa (CITEDEF)
San Juan Bautista de La Salle 4397
(B1603ALO) Villa Martelli, Buenos Aires,
Argentina
amitaritonna@citedef.gob.ar

María José Abásolo^{2,3}

² Comisión de Investigaciones Científicas de la
Provincia de Buenos Aires (CICPBA)
³ Instituto de Investigación en Informática LIDI
(III-LIDI)
Facultad de Informática – Universidad
Nacional de La Plata (UNLP)
Calle 50 y 120 (1900) La Plata, Buenos Aires,
Argentina
mjabasolo@lidi.info.unlp.edu.ar

ABSTRACT

During military operations, the battlefields become fractured zones where the level of confusion, noise and ambiguity impact on achieving tactical objectives. Situational Awareness (SA) becomes a challenge because the unstable perception of the situation leads to a degraded understanding that disables the soldier in projecting the proper results. To meet this challenge various military projects have focused their efforts on designing integrated digital system to support decision-making for military personnel in unknown environments. This paper presents the state of art of military systems using Augmented Reality (AR) in the battlefield.

Keywords

Augmented Reality, Situational Awareness, Devices Mobile, Context Awareness, Network-Centric Warfare.

1. INTRODUCTION

According to [Bry09a] the fratricide remains a very real threat in the current battlefields. To address this, the military have put much effort in the development of combat identification technologies to improve the ability of soldiers to accurately identify the enemy.

[Saa10a] argue that future military operations will be based on tools of Command, Control, Communications, Computers, Information and Intelligence for optimal performance in their assigned tasks in versatile and hostile environments.

Situational Awareness (SA) is a mental representation and understanding of objects, events, interactions, environmental conditions and any other factors in a specific situation that may affect the development of human tasks. Many military operations take place in unknown environments. The SA solutions allow soldiers to make effective use of diverse information in a context of battle being one of the major goals the reduction of cognitive load in times of stress. New technologies offer innovative methods of getting contextual information and then this information is visually represented in a natural and non-invasive way without affecting the cognitive process of the soldier. This is the case of Augmented Reality (AR).

AR, defined by [Azu97a] refers to interactive applications in which 3-D virtual objects (3-D

objects, sounds, text, etc) are integrated into the real environment in real time and according to the position of the user. There are several projects that incorporate the use of AR in military applications, since its use could produce dramatic improvements in the soldier performance and provide a great advantage in combat.

The rest of the article is organized as follows: section 2 introduces definitions such as Situational Awareness, Context Aware, Augmented Cognition, Augmented Reality and Network-Centric Warfare. Section 3 presents the requirements of an AR software framework for a military application. Section 4 presents a review of various military projects using AR to improve the SA on the battlefield. Finally, Section 5 presents the conclusions and future work.

2. DEFINITIONS

Situational awareness

[Bro12a] defines SA as the perception, understanding and anticipation of the elements within an operational environment required to act effectively within that environment.

[Tre11a] mentions that SA is a prerequisite to timely and accurate decision-making in the fast and highly stressful context of infantry operational environments. The introduction of electronic support

technologies onto the battlefield is expected to improve SA by providing the right information, at the right time and in the right format.

[End88a] and [End95a] mention that SA is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future.

Figure 1 depicts the SA model in the dynamic decision making. Endsley mentions that the formal definition of SA is categorized into three hierarchical phases: perception of elements in current situation; Comprehension of current situation; and Projection of future status. The relationships between these phases and task/system and individual factors.

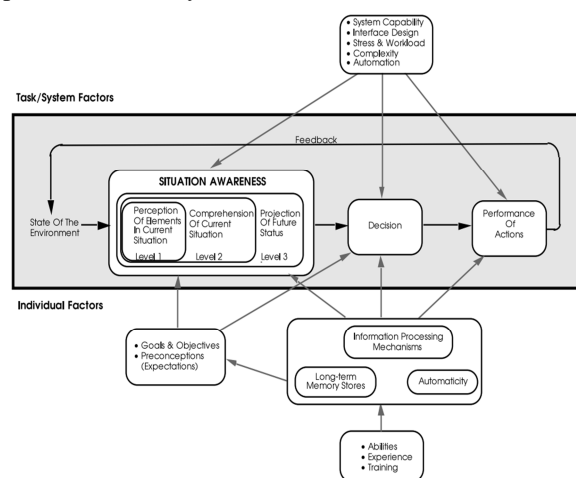


Figure 1 – Endsley's model of situation awareness (Endsley, 1995)

[End03a] determine that one of the most important factors which underlies in the development of a suitable SA is the presence of mental models and prototypical situations schemes. It provides a fundamental mental construction that leads the way to allocate attention and to highlight the critical issues.

Context aware

The formal definition of Context Aware most widely accepted is provided by [Dey00a]: "*Context is any information that can be used to characterize the situation of an entity. An entity can be a person, place or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves*".

According to [Dey99a] the context may be considered as a set of information including user activity, location, personal preferences and current status. The mobility creates situations in the user context. The context is dynamic and keeps on changing. The context is best defined as states or environmental settings, such as location, orientation,

time, nearby objects or people, ambient light level, noise and temperature.

According to [Sch94a] the access to the context increases the wealth of man-machine communication and the effectiveness of the elaboration of the task.

[Hul97a] define Context Aware as computer systems capable of sensing, interpreting and responding, according to the environment in which the user is located.

Augmented cognition

To develop an information display system the information needs must be examined and also the best way to present that information in order to make the system robust, usable and effective must be determined. The information processing capabilities of humans have quickly become a limiting factor in human-computer interaction. This problem has motivated the development of a new scientific discipline called Augmented Cognition (AC) [Kob06a]. The specific concerns of AC are the design of methods to detect and mitigate the limitations of human processing of information and the design of solutions to improve the exchange and use of information on man-machine systems.

Augmented reality

According to [Hic03a], AR provides the user with superimposed information that can be seen in the real world, that is, it complements the real world with virtual information. AR improves the perception of the natural world by adding information to the senses such as visual, sound, smell or tactile sensations. AR refers to the mix of the signals from the tridimensional real environment on the user perception. Particularly it denotes the fusion of virtual 3-D images on the users' natural vision of the world around them, using glasses or HMD (head-mounted display). Through the ability to present superimposed information integrated in the user environment, AR has the potential to provide significant benefits in many application areas. Many of these benefits arise from the fact that the virtual signals shown through AR system may go beyond what is physically visible.

Network-centric warfare

According to [Dod05a] network-centric warfare is a military doctrine that aims to turn an informational advantage in a competitive advantage through a strong network of forces, geographically dispersed, but well-connected and informed.

[Mof02a] describe that it is moving towards an organizational structure of network-centric warfare which is flat, fast and it is based on information, in contrast to the hierarchical structure of slow movement, based on the model of command and control. In the Network-Centric Warfare, computers

integrate information acquired from multiple sources creating an image that provides critical and relevant information to all levels of command and control including the soldier. The networks are formed by nodes with the information transmitted through command positions, vehicles and the soldiers' wearable computer.

3. REQUIREMENTS OF AUGMENTED REALITY MILITARY PROJECTS

In this section is detailed the minimum requirements to have in consideration when we are thinking about augmented reality software in the military area. Therefore, the section is organized as follows: the Information Model explains the information transformation process. Information filtering and representation explains the importance of filtering and the representation of information acquired from the context of the environment. Operational capabilities describe a set of specific needs of military infantry in order to improve the SA of soldiers in the battlefield.

Information Model

In the present study, the term Information Model (IM) is used to describe the information transformation process that starts when the information is acquired from the context and that ends when the information is visually displayed (Fig. 2). The IM describes how the information passes through different stages focused on Acquiring, Sending, Processing and Representing the information obtained from the environment context. In the Acquiring stage the information is obtained from the environment context mainly through sensors geographically dispersed. In the Sending stage the acquired information in the previous stage is sent through communication devices. In the Processing stage the received information is computed. The information is treated through processing techniques such as detection, extraction, classification, recognition, identification, etc. Representing is the last stage of the IM where the information that was processed in the previous step is represented. In this stage AR techniques could be used to enrich the user's perception of the real world.

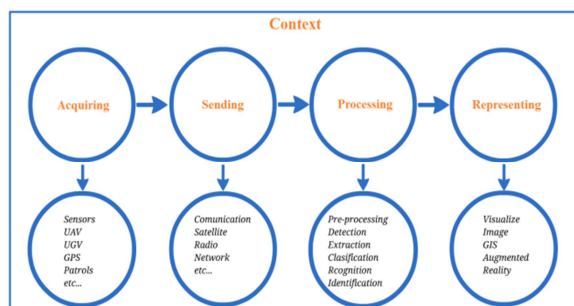


Figure 2 – IM stages. The information transformation process

Information filtering and representation

[Jul00a] have presented the idea of using a real-world context as a search cue in information retrieval and implemented a system which filters information based on physical location, for selecting what is displayed to the user by means of AR. The main purpose of information filtering is to prioritize and reduce the amount of information presented in order to show only what is relevant to the user.

[Ses00a] mention that because of the soldier's mobility through the environment, the context can change dramatically depending on their position. The amount of information that can be displayed to a user in a virtual world can be overwhelming. To fix this problem, the system must sort and prioritize the information so that it should show the features that are relevant to the soldier, such as threats.

[Liv02a] use an information filter to add objects to, or remove objects from, the user's display. They use a spatial filter to show only those objects that lie in a certain one around the user. This zone can be visualized as a cylinder whose main axis is parallel to the user's up vector, where objects that fall within the cylinder's walls are shown, and the user can vary the inner and outer diameters of the cylinder walls.

[Kei13a] propose a method for contextual information filtering based on the user's movement and location in order to enable the intuitive usage of an "internet of things" via AR without information overload. They are employing concepts of camera- and motion-based interaction techniques and use the metaphors of "investigation" and "exploration" to control the way augmented and visually superimposed elements are presented.

Operational capabilities

We have analyzed a set of specific needs of military infantry in order to improve the SA of soldiers on the battlefield by obtaining information from the context of the environment and then to represent it visually by using the AR in order to help the soldier to make decisions under stress. The following operational capabilities are based on an understanding of the current state of digital technology in warfare, the changing nature of combat, the changing role of the soldier and the growing importance of SA.

Operational Capabilities:

- Interactions with gesture and voice recognition for data collection
- Multiple tracking (GPS, sensors, vision, etc.)
- Detection and recognition of 3-D objects
- Face recognition

- Identification of allies and enemies
- Information filter system
- Implementation and integration of the prototype on mobile devices

Potential features:

- Stand-alone (just a few dependence on external network access)
- Omnidirectional (communication among members of the patrol and command and control center)
- Light-weight (gesture & voice recognition) and Low-power devices
- Security (data & communication channel encryption)
- Open source (Framework & OS)
- Mobile (Smartphone, tablets and goggles -see-through display-)

4. REVIEW OF MILITARY PROJECTS

In this section is presented a review of various military projects using small wearable computers and AR systems in order to improve the SA on the battlefield.

Background

[Zie02a] mention that in 1989, the U.S. Army used a small wearable computer to help soldiers on the battlefield tasks.

James Schoening, a research analyst working at the CECOM (Communications Electronics Command) of U.S. Army, is who started to use wearable computers. Working with Matt Zieniewicz, Schoening transformed his idea into system architecture with specific technologies, such as wireless data transmission, image capture and integrated Global Positioning System (GPS). In 1990, Schoening and Zieniewicz associated with John Flatt, Sal Barone and Almon Gillette to demonstrate the Soldier's Computer System. Later, based on the Soldier's Computer project it gave rise to project SIPE (Soldier Integrated Protective Ensemble). It was in the SIPE Project, directed by Carol Fitzgerald, that the U.S. Army treated the components of the combat devices as an integrated system for the first time.

Eyekon

[Hic03a] define the Eyekon project as a support system for decision-making based on intelligent agents installed on a wearable computer carried by the soldiers. The dismounted soldier visualizes the target information and other information on his weapon. The aim of the project is to develop the smart icons and notations that are superimposed on the video of soldier's weapon. The basic functions are on a wearable computer connected via a secure

wireless network to a local and remote database. The system incorporates sensors that provide real-time information (e.g. inertial sensor, GPS, IR, etc.). Eyekon is an intelligent agent-based decision support system hosted on a wearable computer with an available database. This system updates via radio links and on the weapon display, the information is superimposed using AR techniques.

BARS

The Naval Research Laboratory (NRL) developed a prototype augmented reality system known as BARS (Battlefield Augmented Reality System) [Liv02a]. BARS focused on developing a digital system to help address the increasing emphasis on Military Operations in Urban Terrain (MOUT). The BARS user interface includes a sophisticated but disjoint set of functions that assist the warfighter in understanding the surrounding environment, including information filter to annotate the most important or nearby objects, a set of representations of occluded objects, etc. BARS tracked the position and orientation of the user's head and superimposed graphics and annotations that aligned with real objects in the user's field of view. Multiple units shared a common database, where individuals could choose to join a given channel to access its graphic data.

iARM

Tanagram Partners was awarded a contract from the Defense Advanced Research Projects Agency (DARPA) to develop the Intelligent Augmented Reality Model project (iARM) [Juh10a]. The objective of iARM is to develop an integrated digital system that could significantly improve decision making of military personnel in complex contested environments via an integrated operating system, a data services model, and a digitally enhanced head-mounted display. The aim is that all these components work together in a seamless fashion which allows soldiers to perceive, comprehend and what is most important, project the best course of action for increased performance to achieve tactical objectives. iARM project covers many of the attributes of artificial intelligence. In Figure 3 is depicted the conceptual design of the HMD and the soldier's vision through glasses

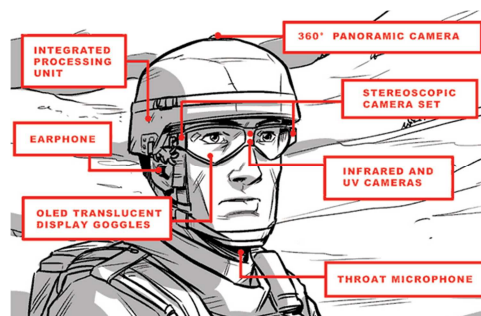


Figure 3 – Conceptual design of iARM (Tanagram Partners. Final Project Report, 2010)

ULTRA-Vis

In [Ult08a] is detailed the Urban Leader Tactical Response, Awareness and Visualization (ULTRA-Vis) project, supported by the Defense Advanced Research Projects Agency (DARPA), has developed an AR prototype system for soldiers on the battlefield. The ULTRA-Vis system overlays full-color graphical iconography onto the local scene observed by the soldier. To enable this capability, the program developed and integrated a light-weight, low-power holographic see-through display with a vision-enabled position and orientation tracking system. Using the ULTRA-Vis system, a soldier can visualize the location of other forces, vehicles, hazards and aircraft in the local environment even when they are not visible to the soldier. The prototype will be equipped for the gesture recognition using a glove. The system allows superimposed symbols in the 3-D battlefield, to locate enemy targets and locate the allied forces. ULTRA-Vis provides to squads a clear tactical advantage enabling collaboration among members of the squadron. ULTRA-Vis enables high SA and the ability to make decisions while on the move in the field of operations. In addition, the system can be used to communicate to the soldier a variety of tactically significant information including imagery, navigation routes and alerts. In Figure 4 is depicted the conceptual design of ULTRA-Vis.

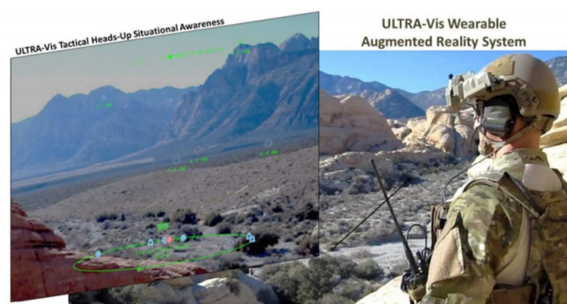


Figure 4 –ULTRA-Vis Prototype System (DARPA, 2013)

5. CONCLUSION AND FUTURE WORK

This article has described how to improve the Situational Awareness by using Augmented Reality as an advanced technique of information representation in a military context. For a right understanding was defined the meaning of terms such as Situational Awareness, Context Awareness, Augmented Reality, Augmented Cognition and Network Centric Warfare. The Information Model

represents how the information goes through different stages from the information acquisition from the context of the environment to the information representation by using AR techniques. The importance of filtering information is enhanced because it dictates what kind of information should be displayed to the soldier and when. The article provided the historical review of military digital systems using AR to improve SA.

6. REFERENCES

- [Azu97a] Azuma R. A survey of Augmented Reality. Presence: Teleoperators and Virtual Environments, vol. 6, no. 4, pp. 355-385, 1997
- [Bro12a] Brown, David Wm. A Survey of Mobile Augmented Reality Technologies for Combat Identification Applications. MSc thesis. Athabasca University, 2012
- [Bry09a] Bryant, D.; Smith, D. Comparison of Identify-Friend-Foe and Blue-Force Tracking Decision Support for Combat Identification. DRDC: Toronto, Rep. 2009
- [Dod05a] Department of Defense of USA –DoD-. The Implementation of Network-Centric Warfare. Washington, D.C.. p. 7, 2005
- [Dey99a] Dey, Anind K.; Abowd, Gregory D.; Brown, Peter J.; Davies, Nigel; Smith, Mark; Steggles, Pete. Towards a Better Understanding of Context and Context-Awareness. Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing. Pages 304-307, 1999
- [Dey00a] Dey, A.; Abowd G. D. Towards a better understanding of context and context-awareness. En: CHIA'00 workshop on Context-Awareness, 2000
- [End88a] Endsley, M. R. Design and evaluation for situation awareness enhancement. In Proceeding of the Human Factors Society 32nd Annual Meeting (pp. 97-101). Santa Mónica, CA: Human Factors Society, 1988
- [End95a] Endsley, M. R. A taxonomy of situation awareness errors. In R. Fuller, N. Johnston & N. McDonald (Eds.), Human Factors in Aviation Operations (pp. 287-292). Aldershot, England; Averbury Aviation, Ashgate Publishing Ltd., 1995
- [End03a] Endsley, M. R.; Bolstad, Cheryl A.; Jones, Debra G.; Riley, Jennifer M. Situation Awareness Oriented Design: From User's Cognitive Requirements to Creating Effective Supporting Technologies. Human Factors and Ergonomics 47th Annual Meeting, Denver, Colorado, EEUU., 2003
- [Hic03a] Hicks, Jeffrey; Flanagan, Richard; Dr. Petrov, Plamen; Dr. Stoyen, Alexander. Eyekon: Distributed Augmented Reality for Soldier Teams. © Copyright 21st Century Systems, Inc., 2003

- [Hol01a] Holmquist, J.; Barnett, J. Digitally Enhanced Situation Awareness: An Aid to Military Decision-Making. Proceedings of the Human Factors and Ergonomics Society Annual Meeting , vol. 45 no. 4 542-546, 2001
- [Hul97a] Hull, R.; Neaves, P.; Bedford-Roberts, J. Towards situated computing. En: 1st International Symposium on Wearable Computers, pp. 146–15, 1997
- [Jul00a] Julier, S.; Lanzagorta, M.; Baillot, Y.; Rosenblum, L.; Feiner, S.; Hollerer, T.; Sestito S. Information filtering for mobile augmented reality. In: Augmented Reality. (ISAR 2000).Proceedings.IEEE and ACM International Symposium, 2000
- [Juh10a] Juhnke, Joseph; Kallish, Adam; Delaney, Dan; Dziedzic, Kim; Chou, Rudy; Chapel, Tim. Tanagram Partners.Final Project Report.Aiding Complex Decision Making through Augmented Reality: iARM, an Intelligent Augmented Reality Model, 2010
- [Kei13a] Keil Jens, Zoellner Michael, EngelkeTimo, WientapperFolker, Schmitt Michael. Controlling and Filtering Information Density with Spatial Interaction Techniques via Handheld Augmented Reality.Virtual Augmented and Mixed Reality.Designing and Developing Augmented and Virtual EnvironmentsLecture Notes in Computer Science Volume 8021, 2013, pp 49-57, 2013
- [Kob06a] Kobus, D. A.; Brown C. M. DARPA Improving Warfighter Information Intake Under Stress—Augmented Cognition. Pacific Science & Engineering Group, Inc. SSC San Diego, 2006
- [Liv02a] Livingston, Mark A.; Rosenblum, Lawrence J.; Julier, Simon J.; Brown, Dennis; Baillot, Yohan; Swan II, J. Edward; Gabbard, Joseph L.; Hix, Deborah. An Augmented Reality System for Military Operations in Urban Terrain. Proceedings of Interservice / Industry Training, Simulation & Education Conference (IITSEC), December 2 -5, Orlando, Florida, page 89 (abstract only), 2002
- [Mof02a] Moffat, J.; Atkinson, S. R. Libro: The Agile Organization: From Informal Networks to Complex Effects & Agility, 2002
- [Moo10a] Moon, Yong-Woon; Jung, Hae-Sun; Jeong, Chang-Sung. Context-awareness in Battlefield using Ubiquitous Computing. Network Centric Warfare. 2010 10th IEEE International Conference on Computer and Information Technology, 2010
- [Saa10a] Saarelainen, Tapio; Jormakka, Jorma. C4I2-Tools for the Future Battlefield Warriors.IEEE - Fifth International Conference on Digital Telecommunications, 2010
- [Ses00a] Sestito, Sabrina; Julier, Simon, Lanzagorta, Marco; Rosenblum, Larry. Intelligent Filtering for Augmented Reality. In: Proceedings of SimTecT 2000, Sydney, Australia, 2000
- [Sch94a] Schilit, B.; Adams, N.; Want R. Context-aware computing applications. En: 1st International Workshop on Mobile Computing Systems and Applications, pp. 85-90, 1994
- [Tre11a] Tremblay, Sébastien; Jeuniaux, Patrick; Romano, Paul; Lowe, Jacques; Grenier, Richard. A Multi-Perspective Approach to the Evaluation of a Portable Situation Awareness Support System in a Simulator Infantry Operation.IEEE - International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), Miami Beach, FL., 2011
- [Ult08a] ULTRA-Vis. BAA 08-36. Broad Agency Announcement for Information Processing Techniques Office and Defense Advanced Research Projects Agency, 2008
- [Zie02a] Zieniewicz, Matthew J.; Johnson, Douglas C.; Wong, Douglas C.; Flatt, John D. The Evolution of Army Wearable Computers. PERVASIVE Computing, 2002

Conditional Random Fields for Web User Task Recognition based on Human Computer Interaction

Anis Elbahi
Research Unit MARS,
Department of computer sciences
FSM - Monastir 5019 Tunisia.
Elbahi.anis@gmail.com

Mohamed Nazih Omri
Research Unit MARS,
Department of computer sciences
FSM - Monastir 5019 Tunisia.
MohamedNazih.Omri@fsm.rnu.tn

ABSTRACT

In this paper we apply the Conditional Random Fields approach for modeling human navigational behavior based on mouse movements to recognize web user tasks. In fact, inferring activity of web users is an important topic of Human Computer Interaction. To improve the interaction process, many studies have been performed for understanding how users interact with web interfaces in order to perform a given activity. The Experimental evaluation and analysis of the results of the model we present in this paper demonstrate the efficiency of our model in human tasks recognition.

Keywords

Pattern recognition, Conditional Random Fields, E-learning Activity Recognition, Mouse movement tracking, Human Computer Interaction.

1. INTRODUCTION

The study of the activity of web users is an important topic of HCI. For years, various techniques have been used in this field, such as eye movements tracking [1], mouse tracking [7] and click-through analysis [10]. Understanding navigational behavior of users can improve interfaces usability, provide assistance for users with disabilities and others applications such as e-learning. On the one hand, the activity of mouse cursor can be easily captured and recorded. On the another hand, analysis of cursor behavior can provide high quality clues of a spontaneous, precise, direct and unbiased trace of user behavior. Such trace can be considered as a good indicator of the user reasoning strategy during a web activity. In this paper, we used the CRF approach [11] in order to recognize the tasks of web users, based on their navigational behavior using mouse movement.

2. ANALYSIS OF USER NAVIGATIONAL BEHAVIOR USING MOUSE MOVEMENT TRACKING

For each task (information searching, mail sending,

downloading), users perform basic operations such as keyboard events, moving a cursor, clicking and pressing a button.

Using a cursor pointing device during web activities, users “draw” their navigational behavior. Mouse movement tracking has been evaluated as an alternative to eye tracking for determining attention on the web page. Therefore, various studies have been achieved in this context such as the study of Chen et al. [3] who have found that mouse and eye movements are strongly related and that 75% of mouse saccades move to significant regions of the screen where eye gaze are moved and they have been confirmed that mouse data can be used to infer the intent of user. So, mouse movements are explored to infer the user tasks during e-learning activity [5] and to provide insights into the intention behind a web search query [7]. Authors of [12] presented a user re-authentication approach using behavioral biometrics provided by mouse dynamics and in reference [8] Heimgartner identify users only by analyzing their interaction behavior mainly based on mouse events.

Elbahi et al. [16] presented a new possibilistic approach based only on mouse behavior for user task identification. Many other researches [4,18] have proposed different models based on possibility theory, on bayesian and semantic networks to recognize the goal of the users.

Obviously mouse movement tracking is a very effective technique, easy to use, freely available and does not disturb user behavior.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

In this paper, we propose a new CRF model to automatically recognize web user tasks based on mouse trajectory recorded data.

3. THE USER TASK AS A SEQUENCE OF FIXED AREAS OF INTEREST

Each web interfaces can be described as a set of significant regions called Areas Of Interest (AOI) which can be manually specified or automatically discovered [9]. During a task, users move the cursor across the web interfaces and fix various AOI. Figure 1 presents an example of a sequence describing fixed AOI during “logging into Gmail account” task.

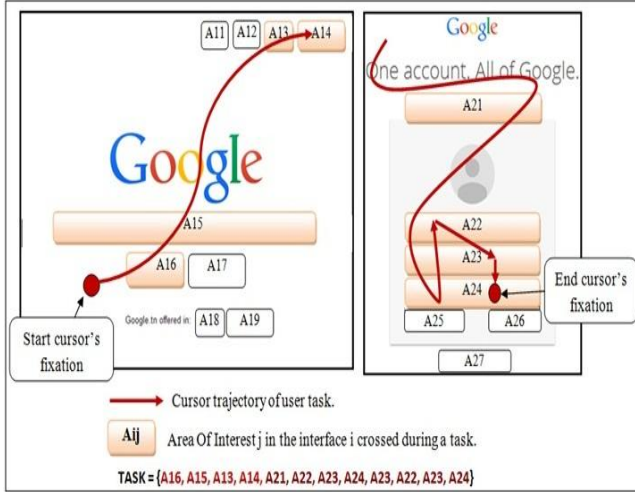


Figure 1. Example of user task defined as a sequence of fixed AOI.

Thus, each user task can be defined as a temporal sequence of fixed AOI during a period of time T . $TSK_i = \{AOI_1, AOI_2, \dots, AOI_T\}$. Despite this clear definition of task, their automatic recognition is very challenging to solve.

The automatic task identification can improve the general interaction process by giving help in real time to unfamiliar users, helping users with disabilities and improving interfaces usability.

4. CRF: A BRIEF PRESENTATION

Hidden Markov Models [13] have been widely used for modeling and labeling stochastic sequences. In spite of their efficiency, CRF theory [11], have been proposed to alleviate HMM assumptions. Therefore, various studies have been successfully achieved for modeling and labeling sequences using CRF [11, 17].

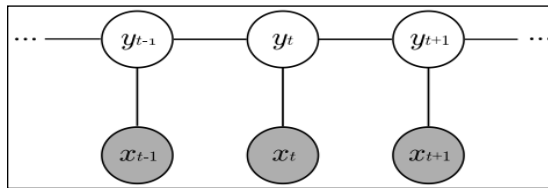


Figure 2. CRF linear-chain graphical representation

Figure 2 shows that CRF model involve hidden and observable variables at each time step and the edges between nodes are not oriented, making the CRF an undirected graphical model.

Due to the discriminative nature of CRF, it becomes possible to represent much more knowledge in the model using feature functions. With CRF we try to maximize the conditional probability distribution $P(Y|X)$ represented as follows:

$$P(Y|X) = \frac{1}{Z(X)} \exp \left(\sum_{t=1}^T \sum_{k=1}^N \lambda_k f_k(y_{t-1}, y_t, X) + \sum_{t=1}^T \sum_{k=1}^N \mu_k g_k(y_t, X) \right) \quad (1)$$

where:

- $Z(X)$ is a normalization factor used to ensure that outcome of $P(Y|X)$ is a probability,
- $f_k(y_{t-1}, y_t, X)$ and $g_k(y_t, X)$ are features functions that return a real value.
- λ_k and μ_k are weights of each feature function,
- T is the length of the sequence X ,
- N is the number of features functions,

CRF are designed to estimate the model parameters using an iterative gradient method such as BFGS algorithm and to perform the inference process using Viterbi algorithm. For CRF parameters estimation, we use a training set defined by: $D = \{x^{(i)}, y^{(i)}\}_{i=1}^{|D|}$, where each $x^{(i)}$ is a sequence of inputs and each $y^{(i)}$ is a sequence of desired predictions.

The estimation of weights of feature function (θ) is performed by maximizing the conditional log-likelihood of annotated sequences of D .

$$L(\theta) = \sum_{i=1}^{|D|} \log P(y^i | x^i)$$

For more details about CRF, reader can see [11, 15].

5. CRF FOR USER TASK MODELING

5.1 The user task modeling

As shown in figure 3 the “Equation Grapher” simulator¹ interface is described as a set of areas of interest $AOI = \{A, B, C, D, E, F, G, H\}$ judged by an expert as frequently pointed regions during users tasks.

Like presented previously, each task can be defined as a finite, temporal, stochastic sequence of AOI set by a user during a period of time.

¹ Phet available on : <http://phet.colorado.edu>

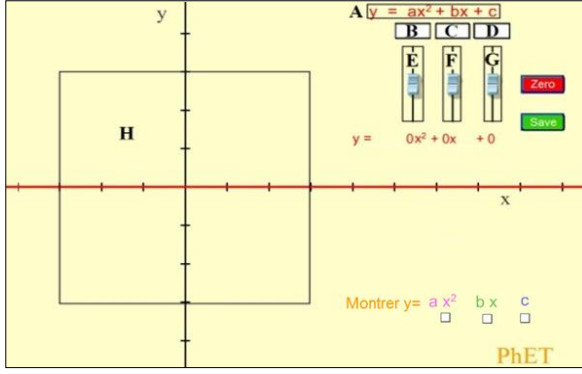


Figure 3. Areas Of Interest in “Equation Grapher” interface.

To define each user task (sequence of fixed AOI), coordinates of mouse cursor have been recorded at each time slice Δt .

5.2 The proposed model

The model structure is described by:

- $TASKs = \{tsk_1, tsk_2, \dots, tsk_M\}$: set of M labels concerning M tasks of users.
- $AOI = \{aoi_1, aoi_2, \dots, aoi_N\}$: set of N areas of interest of the web interface that can be pointed by users during tasks.
- $X = \{aoi_k_1, \dots, aoi_k_t, \dots, aoi_k_T\}$: the sequence of observations describing AOI fixed by mouse cursor during a task for a period of time T , with $1 \leq k \leq N$.
- $Y = \{tsk_i_1, \dots, tsk_i_t, \dots, tsk_i_T\}$: the label sequence, with $1 \leq i \leq M$.

Each sequence of observations (sequence of fixed AOI) given to CRF model must be entirely labeled using a single tag corresponding to performed task. Graphically, our model can be presented as follows:

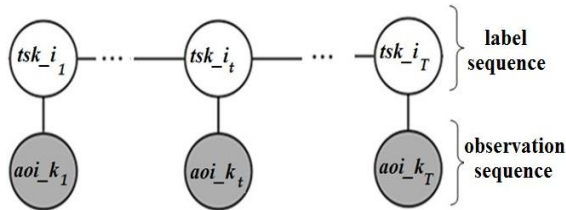


Figure 4. Graphical representation of proposed CRF model.

Let $F = \{f_1, f_2, \dots, f_n\}$ be a set of features functions. Each function $f_j(y_{t-1}, y_t, X_t)$ looks at a pair of adjacent labels (y_{t-1} and y_t) and all the observation sequence (X) at each time step (t).

In order to validate the proposed model, we used CRF++ tool, by which we can define templates to automatically generate a set of features functions. Next, we present some examples of used features functions generated using CRF++ templates.

Template1 : $U00 : \%x[0,0]$ generate a set of functions like:

$$f_1(y_{t-1}, y_t, X_t) = \begin{cases} 1 & \text{if } y_t = tsk_1 \text{ and } x_t = aoi_2; \\ 0 & \text{otherwise} \end{cases}$$

The function f_1 return 1 if the current label (y_t) is tsk_1 and the current observation (x_t) is aoi_2 else f_1 return 0.

Template2 : $U01 : \%x[-1,0]/\%x[0,0]/\%x[1,0]$ generate a set of functions like:

$$f_2(y_{t-1}, y_t, X_t) = \begin{cases} 1 & \text{if } y_t = tsk_2 \text{ and } x_t = aoi_1 \\ & \text{and } x_{t+1} = aoi_2 \text{ and } x_{t-1} = aoi_4; \\ 0 & \text{otherwise} \end{cases}$$

The function f_2 return 1 if the current label (y_t) is tsk_2 and the current observation (x_t) is aoi_1 and next observation (x_{t+1}) is aoi_2 and previous observation (x_{t-1}) is aoi_4 else f_2 return 0.

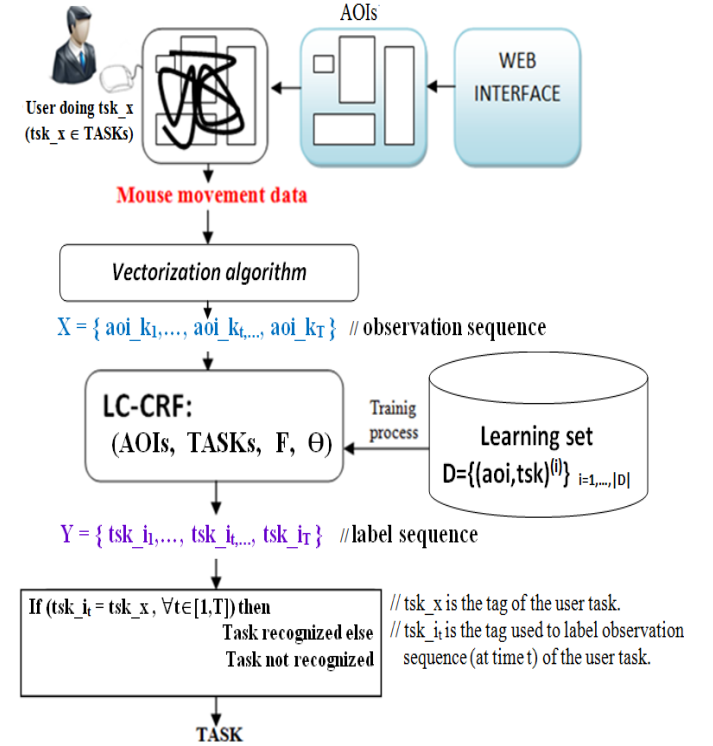


Figure 5. CRF model for task recognition using mouse movement data.

Once the model structure and features functions was defined, we train the model using labelled observation sequences. Each one corresponds to a single task and entirely labeled using a single tag. The figure 5 summarizes our task recognition approach.

6. EXPERIMENTS AND DISCUSSION

6.1 Experimental settings

We prepared a “training and test” set based on real manipulations, which consists of three tasks

performed by students using the “equation grapher” interface. During each task, the student is asked to perform a graphical representation and to keep in memory the shape of the drawn curve. The three tasks are:

Task1(DEG2): representation of a quadratic equation of the form $ax^2+bx+c=0$ (with a, b and c $\neq 0$).

Task2(DEG1): representation of a quadratic equation of the form $ax^2+bx+c=0$ (with a=0 and b, c $\neq 0$).

Task3(INT): a student is asked to discover (and keep in memory) intersection coordinates of a quadratic equation of the form $ax^2+bx+c=0$ (with a,b,c $\neq 0$) and a quadratic equation $ax^2+bx+c=0$ (with a,c=0).

So, the three tasks are very similar and complex to distinguish based only on cursor trajectory.

For sequence of observations preparation, each user perform only one task and we use OGAMA tool [14] for recording mouse cursor coordinates. Based on the obtained data we produce an observation sequence corresponding to the performed task. Once, the 51 observation sequences are prepared and labeled, we estimate the parameters of the CRF model. The used sampling technique is LOOCV (Leave One Out Cross Validation).

6.2 Experimental results and discussion

Although, three tasks are quite similar, experimental results, presented in table 1, showed that CRF model make out 88,23% as recognition rate. These results show that the proposed model have a good ability in user task recognition.

Task Type	Samples	Error	Recognition rate
Task1	17	3	82,35%
Task2	17	2	88,23%
Task	17	1	94,11%
Total	51	6	88,23%

Table 1. Recognition rate of the proposed model.

In order to explain obtained results, remember that each task is defined as a finite, temporal, stochastic sequence of AOI fixed during a period of time T. Thus, each task is described by a sequence of observations $X=\{aoi_{k_1}, \dots, aoi_{k_t}, \dots, aoi_{k_T}\}$. Therefore, in order to recognize a given task, it is necessary to take into consideration all focused AOI during a task.

Due to primary advantage of CRF approach which is the relaxation of the independence assumption, CRF model can take into account more complex dependencies between variables. So, all focused AOI during a task can be taken into consideration by CRF model.

For this reason, CRF presents high performance in user's tasks recognition.

TASK AOI	DEG1	DEG2	INT	TASKX	TASKY
A	1,02	1,81	0,00	0,00	0,00
B	2,57	15,18	10,60	6,91	18,86
C	14,74	6,33	9,14	16,31	6,93
D	14,91	6,47	10,03	6,01	21,24
E	3,61	28,84	13,64	9,85	29,51
F	18,57	11,05	6,83	4,93	2,84
G	21,84	4,57	7,82	6,33	0,53
H	3,71	7,93	24,51	34,04	4,47

Table 2. Average Mouse Fixations (AMF) rate per AOI during tasks.

Table 2 shows the average cursor fixations in some AOI for three tasks. The same table also presents average cursor fixations in each AOI during two tasks TASKX and TASKY. TASKX is INT task and correctly recognized by CRF model while TASKY is INT task and judged by the model as DEG2 task. Likewise, Table 2 shows that area A is rarely fixed during three tasks, therefore the area A can be considered as unimportant item of interest [6] which do not attract the user cursor during interaction.

Task DEG2 is too dependent to areas E (28.84%) and B(15.18%) and task DEG1 is too dependent to areas G(21.84%) , F(18.57%), D(14.91%) and C(14.74%) while most used AOI for task INT are H(24.51%) and E(13.64%). These results show that each type of task attracts user attention into well defined regions in the interface. So, during each task mouse movements can be used to describe the strategy of user. A deeper analysis of mouse movements can give insights of the cognitive processes of the user during a task [2].

During TASKX which is INT task was correctly recognized by CRF model, in fact, mouse fixations rate of TASKX show that user focuses on areas E(9.85%) and H(34.04%) which are more relevant for INT task than DEG1 and DEG2 tasks.

TASKY which is of type INT, but recognized by the model as DEG2, the user usually focuses on relevant AOI for task DEG2 E(29.51%) and B(18.86%) and ignores area H(4.47%) considered as important for the task INT.

Knowing that all users have successfully performed the required tasks, we can see that during TASKY, the user adopts a different strategy to perform an INT task. This explains the failure of the model in recognizing TASKY because the CRF model adjusts its configuration based on the strategy of the group. To perform a given task, a human may adopt a

strategy which is quite different of the one adopted by the majority of users; this task may be the cause of CRF failure. In fact a normal realization of a given task result in a normal use of important AOI which are relevant for this task and ignoring of important areas, or overusing of unimportant areas, should be considered as an indicator of different user strategy during task realization.

7. CONCLUSION

During interaction process, the analysis of mouse movement of the user can tell us about the user's task, AOI that have a user's attention high attraction and ignored AOI. Also, the analysis of cursor behavior can give insights about the strategy adopted by the majority of users and the particular user's strategy during a given task. In this work, we used CRF approach in order to recognize tasks performed by users. Experimental results show the good performance of the proposed model in user task recognition mainly based on mouse movements. Also, results show that each task type have a great impact on mouse behavior because the cursor is more attracted by some AOI than others according to each type of task.

8. REFERENCES

- [1] A. Bulling, J. A. Ward, H. Gellersen and G. Troster, "Eye movement analysis for activity recognition using electrooculography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4), 2011, pp. 741-753.
- [2] A. Nijholt and D. Tan "Brain-Computer Interfaces and human-computer interaction," In *Brain-Computer Interfaces-Appling our Minds to Human-Computer Interaction*, Springer - London, 2010 , pp. 3-19.
- [3] M. C. Chen, J. R. Anderson and M. H. Sohn "What can a mouse cursor tell us more?: correlation of eye / mouse movements on web browsing," In *CHI'01 extended abstracts on Human factors in computing systems*, March 2001, pp. 281-282.
- [4] M. N. Omri, C. Tijus, S. Poitrenaud, and B. Bouchon-Meunier, "Fuzzy sets and semantic nets for on line assistance," *11th Conference on Artificial Intelligence for Applications*, 1995, pp. 374-379.
- [5] A. Elbahi, M. A. Mahjoub and M. N. Omri, "Hidden markov model for inferring user task using mouse movement," In *Information and Communication Technology and Accessibility (ICTA)*, IEEE, October 2013, pp. 1-7.
- [6] A. Poole, and L. J. Ball, "Eye tracking in HCI and usability research," *Encyclopedia of human computer interaction*, 1, pp. 211-219. 2006.
- [7] Q. Guo and E. Agichtein, "Exploring mouse movements for inferring query intent," In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, July 2008, pp. 707-708.
- [8] R. Heimgärtner, "Identification of the User by Analyzing Human Computer Interaction," In *Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction*. Springer - Berlin Heidelberg, 2009, pp. 275-283.
- [9] S. Mathe, and C. Sminchisescu, "Action from still image dataset and inverse optimal control to learn task specific visual scanpaths," In *Advances in Neural Information Processing Systems*, 2013, pp. 1923-1931.
- [10] J. Huang, R. W. White and S. Dumais, "No clicks, no problem: using cursor movements to understand and improve search," In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, May 2011, pp. 1225-1234.
- [11] J. Lafferty, A. McCallum and F. C. Pereira "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.
- [12] N. Zheng, A. Paloski and H. Wang, "An efficient user verification system via mouse movements," In *Proceedings of the 18th ACM conference on Computer and communications security*, 2011, pp. 139-150.
- [13] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," In *Proceedings of the IEEE*, 77(2), 1989, pp. 257-286.
- [14] A. VOBKÜHLER, V. Nordmeier, L. Kuchinke, and A. M. Jacobs, "OGAMA (Open Gaze and Mouse Analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs," In *Behavior research methods*, 40(4), 2008, pp. 1150-1162.
- [15] C. Sutton and A. McCallum, "An introduction to conditional random fields for relational learning," *Introduction to statistical relational learning*, 2006, pp. 93-128.
- [16] A. Elbahi, M. N. Omri and M. A. Mahjoub, "Possibilistic reasoning effects on hidden markov models effectiveness," In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, August, 2015. (accepted).
- [17] F. Fkih and M. N. Omri. "Complex terminology extraction model from unstructured web text based linguistic and statistical knowledge," *International Journal of Information Retrieval Research (IJIRR)*, 2(3), 2012, pp. 1-18.
- [18] M. N. Omri, "Fuzzy knowledge representation, learning and optimization with bayesian analysis in fuzzy semantic networks," In *6th International Conference on Neural Information Processing*, 1999. *Proceedings. (ICONIP'99)*, 1, 1999, pp. 412-417.

DETECTION AND TRACKING OF VEHICLES BASED ON THE VIDEOREGISTRATION INFORMATION

V.N. Kopenkov

IPSI RAS, Samara State Aerospace University
Lukacheva 39B, (443086) Samara, Russia
vkop@geosamara.ru

V.V. Myasnikov

IPSI RAS, Samara State Aerospace University
Lukacheva 39B, (443086) Samara, Russia
vmyas@geosamara.ru

ABSTRACT

This paper describes the technology of detection and tracing vehicles on the sequence of images based on digital video stream analysis in the real-time mode. Such technology can be used for car identification, license plates recognition, and for defining of car traffic parameters: the cars speed and average speed of car traffic, the traffic density, et cetera. One of the main advantages of the presented technology is the simplicity of realization and the possibility of making a cheap solution based on it (cheap camera and a ordinary PC), in contrast to most of the specialized solutions with a high price, and requirements of high-performance hardware components as usual.

Keywords

objects detection and recognition, speed estimation.

1. INTRODUCTION

The tasks of image processing and signal analysis need to be solved in different fields of human activity [Luk01a]. Due to evolution of computer equipment we can realize the tasks which could not be implemented in practice some years ago. One of them - is the processing of video streams in real-time [Sta00a, Mya01a]. Such a process is usually required in teleconferencing systems, security, tracking and other video-analysis systems. Another trend of research in video stream processing technology evolving today, is the analysis of traffic flows based on videoregistration data [Luk01a, Cas00a].

The simple transfer of videos to the operator monitor is ineffective for videostream analysis in real-time. Initial videodata have to be preprocessed to improve the efficiency and quality of decisions of such tasks. It requires the use of algorithms of automatic track changes appropriate for finding, recognizing and analyzing various objects on the images and finally to inform the operator about any unusual situations.

Most of the traditional methods of traffic video information analysis are based on the selection and detection of some specific areas on the image and further tracking of these selected areas on the following frames. Such decision can often be functional but it's not effective due to computational

complexity, not always high precision, and strong dependence on the characteristic regions. Therefore it is unacceptable for the described task solution.

2. TASK DEFINITION AND THE VIDEO RECEIVING MODEL

The hardware

The hardware complex, which provides the function of the offered technology, should consists of the following components: data source (video camera); device of transferring an input signal to a PC - special video capture card or video card with a digital/analog input interface; PC for data processing and storage.

Special database (DB) containing images of all detected vehicle's were created for the testing of proposed information technology, as well as the appropriate software tools to work with this database. The database consist of 1041 images of vehicles detected in the video stream, were filmed on the road. DB store the following information about vehicles:

- the centre of the vehicle license plate (LP),
- the size of the vehicle LP - vertical/ horizontal,
- the LP tilt angle is relative to the horizontal,
- the existence of the vehicle LP,
- the visibility parameter of the LP on the image,
- the type of the vehicle LP,
- the content of the LP - registration number, the BMP filename, with the image of the vehicle.

Video recording model

As it is shown in Fig.1 – the video camera is mounted at a height h , and L_2 – is a distance to the nearest visible point (i.e., the end point of the area of tracking vehicle). L_1 – is a distance to a farthest visible point D . The optical axis of the camera is OB .

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

The plane of the video recording is perpendicular to the optical axis of the camera OB . Let it be located on the point EC . The number of rows in matrix - N_1 . In this case the following expressions are correct.

$$t = |OC| = \sqrt{h^2 + L_2^2}, \beta_1 = \arctan\left(\frac{h}{L_1}\right), \beta_2 = \arctan\left(\frac{h}{L_2}\right) \quad (1)$$

$$\gamma = \beta_1 - \beta_2, d = |EC| = 2t \sin(0,5 \cdot \gamma).$$

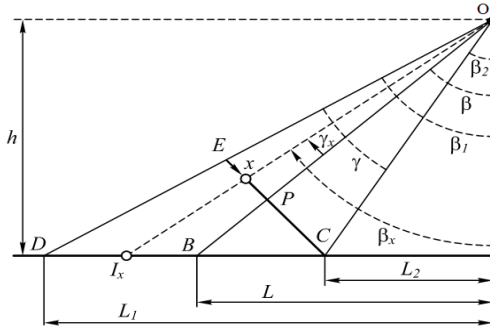


Figure 1. Geometrical model of video recording.

The video processing model can be represented as it follows: an video input module generates digital images which are based on the incoming video stream. After then those images are used as an input information for the vehicle detection module. According to the dynamic changes on a digital image a decision about presence or absence of the vehicle is taken. The image with the detected vehicle is used as an input for the tracking and speed determination module. Data about vehicle speed and the image of detected vehicle, have to be presented to the operator.

3. AN ALGORITHM OF VEHICLES DETECTION AND TRACKING

To determine whether the vehicle is presented on a digital image, there are two images used: current image and the current background component. The background component image is calculated as:

$$S_i = \alpha X_{i-1} + (1 - \alpha) S_{i-1}, \quad (2)$$

S_i – current background picture; X_{i-1} – image on the previous step; S_{i-1} – background picture on the previous step; α – coefficient (small enough), which determines the rate of change of the background component and it depends on the presence or absence of the vehicle on the current frame.

Such representation of the background component allows to accumulate changes of the background component and the system can automatically adjust to change in the light of the scene (day, night) and weather conditions (snow, rain, cloudy, sunny).

The image X_i can be presented as:

$$X_i = U_i + V_i, \quad (3)$$

U_i – background component (stationary); V_i – figure of vehicle (background changes is assumed as noise).

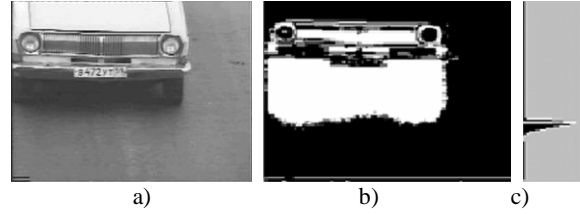


Figure 2. Vehicle detection.

To detect the vehicle on the image we can produce logical subtraction X_i (the current frame - fig. 2a) of S_i (current background). The result is the background component changing area. After the threshold processing obtain a mask representing the binarized region of changes - "trace" of the vehicle (Fig. 2b).

Parameter α determined the speed of background component changing depends on the mask:

- if the proportion of the mask more than 5% of the image size (there is vehicle in visible area), the coefficient α reduced by 50 times, so as not to accumulate noise on the background component;
- If the same proportion of the mask less than 5% of the image size, the coefficient α given back to the original value if the mask part more than 5%.

The exception of this rule is the "systems stabilization time" - time of setting up of background component and weather condition when α coefficient is constant regardless of the proportion of the mask on the image, until the difference between image and the background on the adjacent steps is less than a threshold.

When a proportion of the mask will be more 15% of the image size, the process of vehicle detecting and tracing is starting. And following algorithm is used:

1) calculate a mathematical expectations array in rows of the current mask; recalculate its values:

$$a_i^{new} = a_{i-k} - 10a_{i+k}, \quad k = 1 \dots 5;$$

2) detect the max of this new array. According to previous transformation the max positions corresponds to the maximum difference of brightness on the mask (area under the bumper) - see. Fig. 2c.;

3) the tracking of the vehicle on the image is finishing when there is a "jump of the max" (max is not found) or if the proportion of the mask < 15%.

4. AN ALGORITHM OF VEHICLES SPEED CALCULATION

To determine the vehicle speed the position of max value of the brightness difference array is used. These values received from two frames - at the beginning and at the end of vehicle tracking and used with the corresponded measurements of the frames time. If this frame is the first frame of the vehicle tracking process, then we save the coordinate i_1^{max} of the max value and the time t_1 as start parameters of the vehicle

tracking. If this frame is the last frame of the vehicle tracking, then we save the coordinate i_2^{\max} of the max value and the time t_2 as a vehicle tracking termination parameters. After than speed can be determined as:

$$v = \frac{l_x(i_1^{\max}) - l_x(i_2^{\max})}{t_2 - t_1}, \quad (3)$$

where $l_x(i_1^{\max})$, $l_x(i_2^{\max})$ – the actual distance to the vehicle from the recording device.

Considering that the brightness difference accounts to area under the bumper of the vehicle on the picture the speed estimation will be the most accurate in the case of precise observance of the receiving video system installation parameters (height, distance and camera position). In this case the camera mounting height in relation of the vehicle fixation point (brightness difference) is constant. At the same time when we use LP of the vehicle for the speed determining by finding and tracking the LP position on the pictures – the accuracy of this estimation can be considerably lower. This occurs because of different height of LP mounting towards to the camera position (bumper, radiator frame, bus, truck, etc.). Taking into account the geometric model of the video recording presented above the following algorithm is used for the vehicle speed estimation.

Denote by i the required row number on the recorded image and then corresponding to it point location on the segment EL , which is equal the distance from a point x to E , can be calculated as: $x = i \frac{d}{N_1 - 1}$.

At the same time actual distance l_x on the plane AD from the point A is defined as $l_x = h \cdot \tan(\beta_x)$, where the angle value β_x can be represented as:

$$\beta_x = \beta + \gamma_x, \quad \beta = \frac{\beta_1 + \beta_2}{2}, \quad \gamma_x = \arctan\left(\frac{(d/2) - x}{|OP|}\right).$$

$$|OP| = t \cos\left(\frac{\gamma}{2}\right), \quad \gamma_x = \arctan\left(\left(1 - \frac{2i}{N-1}\right) \tan\left(\frac{\gamma}{2}\right)\right).$$

The distance l_x can be executed in three steps:

1 step: calculate β_1 and β_2 by using (1):

2 step: calculate β_x :

$$\beta_x(i) = \frac{\beta_1 + \beta_2}{2} + \arctan\left(\left(1 - \frac{2i}{N-1}\right) \tan\left(\frac{\beta_1 - \beta_2}{2}\right)\right) \quad (5)$$

3 step: calculate $l_x(i)$: $l_x(i) = h \cdot \tan(\beta_x(i))$. (6)

To accelerate the calculating process the $l_x(i)$ values can be tabulated for all values of $i = 0, N-1$.

In a situation, when the analyzed point is located at some height h_0 (relative to the road), the expressions (1), (5), (6) can also be used for calculating of the actual distance to the object after replacing the values L_1, L_2 and h to the values L'_1, L'_2 and h' :

$$h' = h - h_0, \quad L'_1 = h' \cdot \tan(\beta_1), \quad L'_2 = h' \cdot \tan(\beta_2). \quad (7)$$

Obtaining information about the horizontal arrangement of the tracking object is only possible in case of precisely known distance to the camera. Suppose that this was performed, we know the actual shooting distance from the object to the camera – l_x (see. Fig. 3). In this case, we can obtain the equation:

$$w_x = l_x \frac{W_2}{L_2},$$

W_2 – width of the tracking area on the distance L_2 .

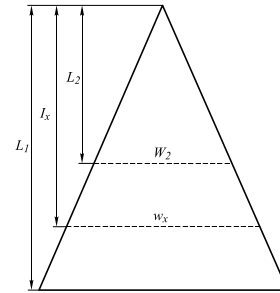


Figure 3. Geometrical model of vehicle tracking area.

Note that the following equality should be performed for the boundaries of digital image:

$$w_x = -\frac{W_2}{2}, \text{ if } \begin{cases} j_x = 0 \\ l_x = L_2 \end{cases}; w_x = +\frac{W_2}{2}, \text{ if } \begin{cases} j_x = N_2 - 1 \\ l_x = L_2 \end{cases},$$

N_2 – number of cols in matrix of registration, j_x – the desired pixel, the expression for the horizontal coordinate of the registered point is obtained:

$$w_x(j_x) = W_2 \cdot \left(\frac{j_x - (N_2 - 1) / 2}{N_2 - 1} \right). \quad (8)$$

In a situation, when the analyzed point is located at some height h_0 (in relation to the road surface), the expression (8) can be used after replacing the value W_2 to the value W'_2 , and note that $L_2 = h \cdot \tan(\beta_2)$,

$$\frac{W'_2}{W_2} = \frac{|OC'|}{|OC|} = \frac{\sqrt{h'^2 + (L'_2)^2}}{\sqrt{h^2 + (L_2)^2}}, \text{ the (8) can be presented:}$$

$$w_x(j_x, h_0) = W_2 \cdot \frac{h}{h - h_0} \left(\frac{j_x - (N_2 - 1) / 2}{N_2 - 1} \right). \quad (9)$$

5. EXPERIMENTAL RESEARCH

Vehicle detection on a videostream

Experimental researches were conducted on the basis

of the analysis of real traffic flow recorded movie, which represents a collection of video clips with different seasons, day-time and weather conditions.

Recording conditions	Vehicle missed	Detected too late	Multy-detection		
			2	3	4
Day (10 min.)	4	1	9	1	0
Night (20 min.)	8	22	16	1	1
Winter (9 min.)	10	0	13	2	1
All (39 min.)	22	23	38	4	2

Table 1. Statistics of detection of moving vehicles.

In case of the night video the number of missed and late detected vehicles is exaggerated, because this video consists of many small clips. The system has not enough time to tune on the background component changing. The probability of the vehicle detection is about 0.95 and the probability of missing - 0.05. There is a slight dependence on the shooting conditions of night, day, winter, etc. The probability of false detection of a non-existent vehicle is equal to 0. Some vehicle can be marked more the once.

Vehicle speed estimation

At the moment, there is no way to make an accurate assessment of the efficiency of vehicle speed estimation algorithm due to the fact that precise data of vehicles speed is not available (from the patented device for the velocity determination) for the experiments. The results of expert evaluation of the vehicle speed corresponds to the expected values. The error estimation is about 5-10%. Necessary to obtain a traffic flow video with accurate information of each vehicle speeds, the future investigation of the efficiency of the algorithm of vehicle speed determination, is as well as calibration of the parameters.

6. PARAMETERS OF TRAFFIC FLOW

The parameters are typically correlated to traffic flow [Gas00a, Cas01a] is a speed, density and the actual volume of traffic flow. The flow velocity v means the average speed of the vehicle (for the analyzed period). The density ρ - is the number of vehicles per area unit, the volume of the flow q is the number of vehicles per unit of time. Assume that, T - is the duration of the interval of an assessment accumulation, for which at a particular time t_0 K vehicles with the parameters $\{t_k, v_k\}_{k=0}^{K-1}$ are registered, where $t_k \in [t_0 - T, t_0]$ - the time of particular vehicle detecting, v_k - speed of this particular vehicle. In that case the estimations of the required parameters may be determined by formulas:

$$\hat{v}(t_0) = \frac{1}{K} \sum_{k=0}^{K-1} v_k, \quad \hat{q}(t_0) = \frac{K}{T}, \quad \hat{\rho}(t_0) = \frac{K}{\hat{v}(t_0) T W_2 \left(\frac{L_2 + L_1}{2L_2} \right)}.$$

7. CONCLUSIONS AND RESULTS

The paper presents an *efficient technology* that allows to solve the following tasks in a real-time mode:

- detect the vehicle on a videstream,
- evaluate the velocity of the vehicle on the image regardless of the presence of the vehicle LP,
- determine the density of traffic flow on the basis of the estimation of individual vehicle parameters.

The main advantages of this technology are:

1. A vehicle's detection and tracking (also as speed estimation) based on the detection of the vehicle itself but not of the vehicle LP (the problem of detecting the vehicle without LP can be solved).
2. The adaptive tuning of the algorithm parameters allows to detect the vehicle efficiently, regardless of changing weather or lighting conditions.
3. The computational efficiency of the developed algorithms in very high.

The perspectives of future works involve the specification of the model and to configure the algorithm of the speed estimation on the base of the real data on the vehicle speed. One more direction of the research is the development of techniques for the detection and recognition of LP on the vehicle.

8. ACKNOWLEDGMENTS

This work was financially supported by the Russian Scientific Foundation (RSF), grant no. 14-31-00014 "Establishment of a Laboratory of Advanced Technology for Earth Remote Sensing".

9. REFERENCES

- [Luk01a] Lukyanitsa A. // Digital processing of video images // A.A. Lukyanitsa, A.G. Shishkin AG - M.: ISS Press, 2009. - 518 p. - (In Russian)
- [Sta00a] Stauffer, Ch. Learning Patterns of Activity Using Real-Time Tracking Pattern Analysis and Machine Intelligence / Chris Stauffer, W. Eric L. Grimson // Pattern Analysis and Machine Intelligence, IEEE Transactions on. - 2000, Aug. - V. 22, Issue 8. - P. 747-757.
- [Mya01a] Myasnikov, V.V. Method for detection of vehicles in digital aerial and space remote sensed images // V.V. Myasnikov // Computer optics. - 2012. - №36(3). - P. 429-438 - (In Russian)
- [Gas00a] Gasnikov, A. Introduction to Mathematical Traffic Flow Modeling / A.V. Gasnikov, S.L. Klenov, E.A. Nurminski, Ya.A. Holodov, N.B. Shamrai // Moscow: MFTI, 2010. 362 p..
- [Cas01a] Cascetta, E. Transportation Systems Analysis: Models and Applications V.29. Springer Optimization and Its Applications, 2009. 760 p.
- [Kop01a] Kopenkov V.N. The estimation of the traffic flow parameters based on the videoregistration data analysis /Computer optics. 2014. №38(2). P. 351-355.

A Novel Retinex Model Based on Sparse Source Separation

Jangwon Choi

Department of Electrical and
Electronic Engineering
at Yonsei University
134 Shinchon-dong
Seodaemoon-gu
120-749, Seoul, Korea
garpield@yonsei.ac.kr

Yong-Goo Kim

Department of Newmedia
at Korea German Institute of
Technology
661 Deungchon-dong
Gangseo-gu
157-033, Seoul, Korea
ygkim@kgit.ac.kr

Yoonsik Choe

Department of Electrical and
Electronic Engineering
at Yonsei University
134 Shinchon-dong
Seodaemoon-gu
120-749, Seoul, Korea
yschoe@yonsei.ac.kr

ABSTRACT

Retinex was introduced by E.Land to explain and solve a problem of color constancy in human visual system (HVS). In this paper, we propose a novel Retinex model based on sparse source separation problem. Different from the existing models, we can explain a relation between the modeling and the effectiveness of Retinex decomposition with the proposed model. We demonstrate the performance of our model by experimental results.

Keywords

color constancy, Retinex theory, modeling, gradient domain, sparse source separation

1. INTRODUCTION

The color of object which is determined by machine visual system (MVS), such as a digital camera, is based on the amount of reflected light on the object. However, when human visual system (HVS) determines the object color, it also considers the amount of detail in the surrounding area and the variation of overall illumination. With this complex system, we can automatically discount the variation of illumination and so easily recognize the color of object which is same under varying illumination conditions [Pal09][Riz07]. This feature of the HVS is called color constancy [Ebn07], and it has been studied over the forty years.

Land's Retinex theory is the first computational model to simulate and explain the color constancy of HVS [Lan71]. He simulated and explained how the HVS perceives color, based on experiments using Mondrian patterns. With this result, he proposed path-based Retinex algorithm to solve the discrepancy problem between the MVS and HVS [Lan83]. This algorithm extracts the reflectance components from the original image which is obtained by MVS. The extracted reflectance image is so clear and quite similar to the image of HVS because it has no illumi-

nation components in the image. After Land's pioneering studies of Retinex theory, there have been many researches to interpret, improve, and implement the Retinex algorithm.

The Retinex algorithm is usually categorized into four areas: path-based algorithms, recursive algorithms, center/surround algorithms and physics-based algorithms[Mor09]. Among these areas, the physics-based algorithms are widely studied in recent years because they can efficiently remove the global illumination from the images. Based on the main assumptions that the illumination varies smoothly and the reflectance is piecewise constant, the physics-based Retinex algorithms set models of the reflectance and the illumination firstly, and then decompose image intensity as a product of the reflectance and the illumination.

In [Kim03], Kimmel et al. proposed a variation model for Retinex which set the illumination as the variational framework. This algorithm can simply extract the illumination using the steepest descent method, but the reflectance model is not considered. To improve this algorithm, Michael et al. proposed a total-variation model for Retinex [Mic11]. This algorithm uses the total-variation model for the reflectance[MaW10], and also considers the illumination model using the assumption of spatially smoothness. This algorithm makes the decomposition of image more appropriate and reasonable than the previous works. However, there still remains a problem in the total-variation model that the mathematical relation between the modeling and the effectiveness of decomposition is not clear.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

In this paper, we propose a sparse source separation model of Retinex theory. Using the properties that the gradient of reflectance is spatially sparse and the gradient of illumination is sparse in frequency domain, we apply the sparse source separation algorithm to the Retinex model. With our model, we can explain the effectiveness of Retinex decomposition and can decompose the reflectance and illumination more accurately. Some experimental results are presented to show the effectiveness of the proposed model.

2. VARIATION/TOTAL-VARIATION MODEL FOR RETINEX

As mentioned above, the variation and total-variation model for Retinex are categorized in the physics-based Retinex algorithm. The algorithms in this category decompose the image intensity S as a product of the reflectance R and of the incident illumination intensity L as follows [Mor09] :

$$S = R \cdot L \quad (1)$$

,where $0 < R < 1$ and $0 < L < \infty$. In order to handle the product form, they are converted into the logarithmic domain, i.e.,

$$s = r + l \quad (2)$$

where $s = \log S$, $r = \log R$, and $l = \log L$. Based on the assumption that the illumination is spatially smooth, Kimmel et al. proposed the variation modeling for Retinex [Kim03], i.e.,

$$\argmin_l F[l] = \int_{\Omega} (|\nabla l|^2 + \alpha(l - s)^2 + \beta|\nabla(l - s)|^2) dx dy$$

$$\text{subject to : } l \geq s, \text{ and } \langle \nabla l, \vec{n} \rangle = 0 \text{ on } \partial\Omega \quad (3)$$

where Ω is the support of the image, $\partial\Omega$ its boundary, and \vec{n} is the normal to the boundary. α and β are free non-negative real parameters. This model is a quadratic programming problem which can be solved by many methods such project normalized steepest descent method as in [Kim03]. However, this model has some limitations because the reflectance piecewise constant assumption is not considered in its model.

To improve the variation model of Retinex, Michael et al. proposed the total-variation model for Retinex [Mic11], i.e.,

$$\begin{aligned} \argmin_{r,l} E(r,l) &= \int_{\Omega} |\nabla r| + \frac{\alpha}{2} \int_{\Omega} |\nabla l|^2 dx \\ &+ \frac{\beta}{2} \int_{\Omega} (l + r - s)^2 dx \\ &+ \frac{\mu}{2} \int_{\Omega} l^2 dx \end{aligned} \quad (4)$$

where α , β , and μ are positive numbers for regularization parameters, the term $\int_{\Omega} (l + r - s)^2 dx$ is used for the fidelity, and the term $\int_{\Omega} l^2 dx$ is used only for the theoretical setting. As shown in (4), they considered the both sides: the reflection function is piecewise constant, and the illumination function is spatially smooth. Using these properties, the total-variation model makes more appropriate and reasonable for the decomposition of Retinex than the previous variation model.

Nevertheless, there still remain some problems in the variation/total-variation model. The variation and total-variation model do not explain the mathematical relation between their modeling and the possibility of decomposition. So we do not know whether the modeling is suitable or not for the Retinex decomposition. They also have some halo effects in their decomposed reflectance images because they consider both the reflectance and illumination in the spatial domain. Due to the fact that the illumination was modeled as smooth in the spatial domain, the results of reflectance images lead to the creation of local halos, such as those around the letters.

3. PROPOSED RETINEX MODEL

In order to improve the previous model of Retinex, we present a sparse source separation model of Retinex. In this section, we introduce the sparse source separation algorithm firstly, and then we present a novel Retinex decomposition model using the sparse source separation algorithm.

Source separation algorithm in mixed two sparse signals

Suppose that there is a mixed signal $z \in \mathbb{R}^d$ of the superposition model, i.e.,

$$z = x + y \quad (5)$$

We assume that a vector $x \in \mathbb{R}^d$ is sparse if the most of its entries are equal to zero. Similarly, a vector $y \in \mathbb{R}^d$ is sparse in frequency if its discrete cosine transform (DCT) Dy is sparse, where $D \in \mathbb{R}^{d \times d}$ is the matrix that encodes the DCT. Using these properties, we can separate the z into x and y to search for the sparsest possible components [McC14], i.e.,

$$[\tilde{x}, \tilde{y}] = \argmin_{x,y \in \mathbb{R}^d} \{\|x\|_0 + \lambda \|Dy\|_0 : z = x + y\} \quad (6)$$

Where the ℓ_0 -norm measures the sparsity of its input, and $\lambda > 0$ is a regularization parameter that trades the relative sparsity of solutions.

In [Don01], Donoho et al. proved that if

$$\|x\|_0 + \|Dy\|_0 \leq \sqrt{d/2}, \quad (7)$$

then the solution to equation (6) is unique. In other words, we can perfectly separate x and y from z if x and Dy are sufficiently sparse.

Unfortunately, solving (6) involves an intractable computation problem. So we replace the ℓ_0 penalty with the convex ℓ_1 -norm to solve a classical sparse approximation program as follows :

$$[\hat{x}, \hat{y}] = \operatorname{argmin}_{x, y \in \mathbb{R}^d} \{ \|x\|_1 + \lambda \|Dy\|_1 : z = x + y \} \quad (8)$$

In [Don01], Donoho et al. also proved that if

$$\|x\|_0 + \|Dy\|_0 \leq \frac{1}{2} \sqrt{d/2}, \quad (9)$$

then the solution to equation (8) is also unique. In other words, we can also perfectly separate x_0 and y_0 from z using the ℓ_1 -norm if x and Dy are sufficiently sparse.

The change to the convex ℓ_1 -norm offers a benefit that we can use a number of highly efficient convex program algorithms for solving (8).

A sparse source separation model of Retinex theory

At first, we convert the equation (2) into a gradient domain, i.e.,

$$\nabla s = \nabla r + \nabla l \quad (10)$$

, because the reflectance r is piecewise constant, the gradient of reflectance ∇r is spatially sparse. And because the illumination l is spatially smooth, the gradient of illumination ∇l is piecewise constant in spatial domain, so it is sparse in frequency domain.

As we explained in the previous subsection, we can efficiently decompose the superposition signal which is mixed the spatially sparse signal and the sparse-in-frequency signal. So, applying this fact to the sparse source separation problem, we can decompose the ∇s into ∇r and ∇l , solving the follow problem :

$$[\hat{r}', \hat{l}'] = \operatorname{argmin}_{r', l' \in \mathbb{R}^N} \{ \|r'\|_0 + \lambda \|Dl'\|_0 : s' = r' + l' \} \quad (11)$$

where $r' = \nabla r$, $l' = \nabla l$, $s' = \nabla s$, and N is the size of image. Using the decomposed r' and l' , we can get the reflectance r and illumination l with the simple inverse process of equation (10).

Based on the equation (7), we can perfectly decompose s' into r' and l' if

$$\|r'\|_0 + \|Dl'\|_0 \leq \sqrt{N/2} \quad (12)$$

So, if r' and l' satisfy the above sparse condition, then we can accurately decompose the reflectance image from the original image.

However, as we mentioned before, it is difficult to solve the equation (11) because the ℓ_0 optimization problem is NP-hard. So we replace the ℓ_0 penalty of (11) with the ℓ_1 -norm, i.e.,

$$[\hat{r}', \hat{l}'] = \operatorname{argmin}_{r', l' \in \mathbb{R}^N} \{ \|r'\|_1 + \lambda \|Dl'\|_1 : s' = r' + l' \} \quad (13)$$

Based on the equation (9), we also see that the equation (13) is perfectly decomposed if

$$\|r'\|_0 + \|Dl'\|_0 \leq \frac{1}{2} \sqrt{N/2} \quad (14)$$

With our Retinex modeling, we can explain the relation between the modeling and the efficiency of Retinex decomposition that the sparsity of ∇r and $D\nabla l$ are proportional to the accuracy of Retinex decomposition, and if they satisfy the equation (14), then we can perfectly decompose the reflectance and illumination from the original image.

To solve the equation (13), we change this equation to the unconstrained form, i.e.,

$$[\hat{r}', \hat{l}'] = \operatorname{argmin}_{r', l' \in \mathbb{R}^N} \left\{ \|r'\|_1 + \lambda \|Dl'\|_1 + \frac{\beta}{2} \int (s' - r' - l')^2 dx \right\} \quad (15)$$

,where x is the axis of image, and β is the positive regularization parameter. To solve the equation (15), we use the alternating minimization scheme as

$$\begin{aligned} r'^{(k+1)} &= \operatorname{argmin}_{r' \in \mathbb{R}^N} \left\{ \|r'\|_1 + \frac{\beta}{2} \int (s' - r' - l'^{(k)})^2 dx \right\} \\ l'^{(k+1)} &= \operatorname{argmin}_{l' \in \mathbb{R}^N} \left\{ \lambda \|Dl'\|_1 + \frac{\beta}{2} \int (s' - r'^{(k+1)} - l')^2 dx \right\} \end{aligned} \quad (16)$$

In order to get the r and l directly without the inverse process of equation (10), we apply the split Bregman method [Gol09] in each subproblems in (16), i.e.,

$$\begin{aligned} r'^{(k+1)} &= \operatorname{argmin}_{r, d_1 \in \mathbb{R}^N} \left\{ \|d_1\|_1 + \frac{\beta}{2} \int (\nabla s - \nabla r - \nabla l^{(k)})^2 dx \right\} \\ l'^{(k+1)} &= \operatorname{argmin}_{l, d_2 \in \mathbb{R}^N} \left\{ \lambda \|d_2\|_1 + \frac{\beta}{2} \int (\nabla s - \nabla r^{(k+1)} - \nabla l)^2 dx \right\} \end{aligned} \quad (17)$$

,where $d_1 = \nabla r$ and $d_2 = D\nabla l$. With the iteration of each sub-problems, we can get the decomposed reflectance and illumination image.

4. EXPERIMENTAL RESULTS

In our implementation, we use the HSV Retinex which considers the intensity layer (V) only. We compare the results of our model with those of Kimmel's model [Kim03] and the Michael's model [Mic11]. The parameters of previous models set the same as their papers. In our model, we fix $\lambda = 4$ and $\beta = 0.1$ which apply to weight the illumination cost.

Figure 1 and 2 show the experimental results of the previous works and our proposed model with two Retinex test images. In these figures, we can see that the reflectance image of our model is the most clear in the edge of image.

In the next experiments, we apply the Retinex algorithms using the color circle images we made. As shown in figure 3, our proposed model restores the edge and intensity of reflectance more accurately than the other Retinex algorithms. Figure 4 shows the differences of S-CIELAB image [Zha97] between the

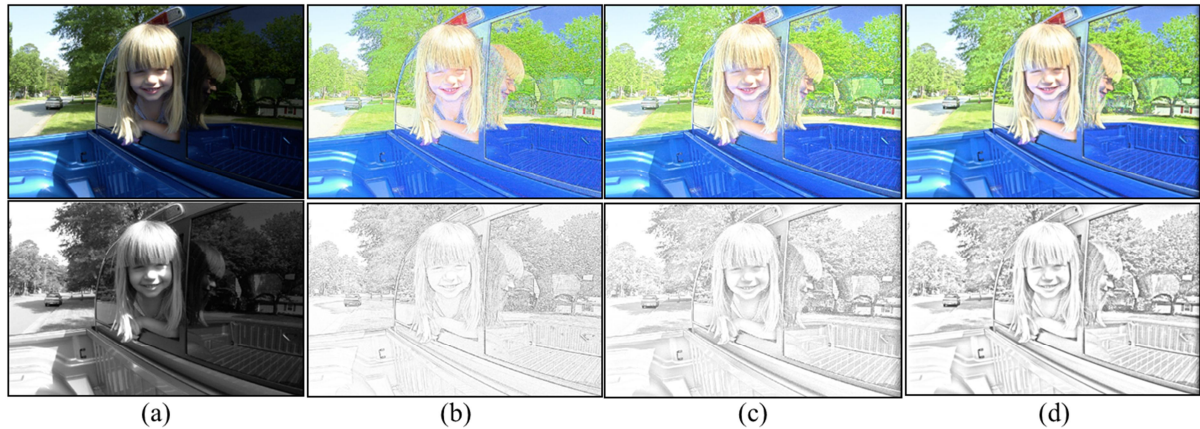


Fig. 1 The results of Retinex algorithms with ‘girl1’ test image. Top is the reconstructed color image, and bottom is the mono reflectance image of V layer. (a) The original image; (b) reflectance image by the variation model in [Kim03]; (c) reflectance image by the total-variation model in [Mic11]; (d) reflectance image by the proposed model.

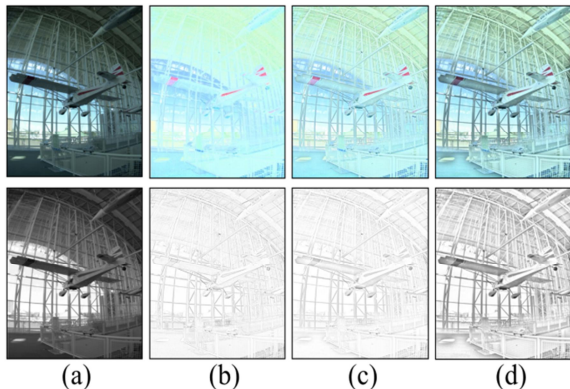


Fig 2 The results of Retinex algorithms with ‘interior1’ test image. Top is the reconstructed color image, and bottom is the mono reflectance image of V layer. (a) The original image; (b) reflectance image by the variation model in [Kim03]; (c) reflectance image by the total-variation model in [Mic11]; (d) reflectance image by the proposed model.

original color circle and the reflectance image which is obtained with Retinex algorithm. The brightness of S-CIELAB image means the amount of errors. As shown in figure 4, our Retinex model shows the less S-CIELAB differences than the other Retinex algorithms.

With the results in figure 5, we can also see that our model is almost free for the local halo effect, yet the results of previous model in figure 5(b)(c) lead to the creation of local halos.

With these results, we can say that our model is effective for the Retinex decomposition.

5. CONCLUSIONS

In this paper, we propose a sparse source separation model of Retinex algorithm. Based on the sparse source separation problem, we use the properties of illumination and reflectance in the source separation model that the gradient of illumination is sparse-in-frequency and the gradient of reflectance is sparse in spatial domain. Unlike the previous physics-based Retinex models, our model is able to explain the modeling and the efficiency of model for Retinex decomposition. Some experimental results show that our Retinex model is effective to decompose the reflectance components from the original images.

6. ACKNOWLEDGMENTS

"This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2015-H8501-15-1001) supervised by the IITP(Institute for Information & communications Technology Promotion)"

7. REFERENCES

- [Riz07] Rizzi A. and McCann J., “On the behavior of spatial models of color,” in *IST/SPIE Electronic Imaging*, vol. 649302, 2007.
- [Ebn07] Ebner M., *Color Constancy*. Wiley, 2007.
- [Pal09] Palma-Amestoy R., Provenzi E., Bertalmio M., and Caselles V., “A perceptually inspired variational framework for color enhancement,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 458–474, Mar 2009.

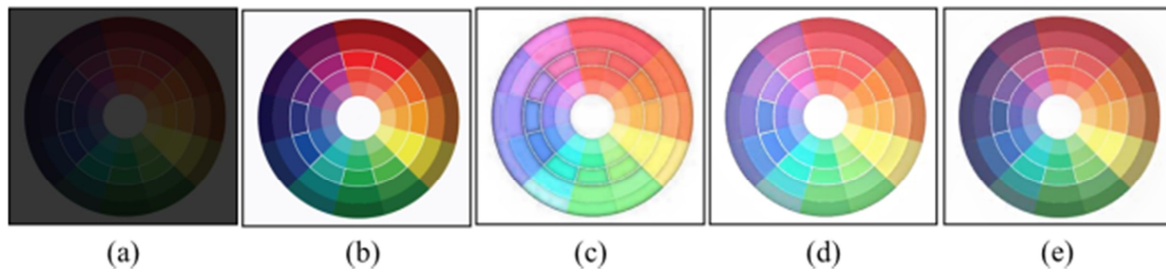


Fig 3 The results of Retinex algorithms with ‘color circle’ test image. (a) The illuminated image; (b) The original image; (c) reflectance image by the variation model in [Kim03]; (d) reflectance image by the total-variation model in [Mic11]; (e) reflectance image by the proposed model.

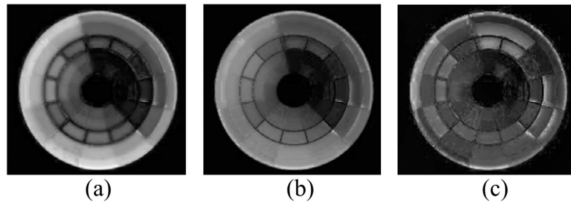


Fig 4 The S-CIELAB difference images between the original color circle image and the results of Retinex reflectance images. (a) With the reflectance image by the variation model in [Kim03]; (b) reflectance image by the total-variation model in [Mic11]; (c) reflectance image by the proposed model.

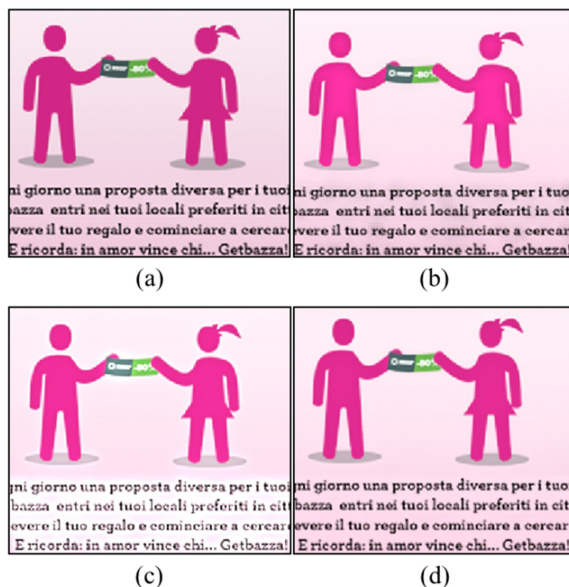


Fig 5 The results of Retinex algorithms with ‘text’ test image. Top is the reconstructed color image, and bottom is the mono reflectance image of V layer. (a) The original image; (b) reflectance image by the variation model in [Kim03]; (c) reflectance image by the total-variation model in [Mic11]; (d) reflectance image by the proposed model.

[Lan71] Land E. and McCann J., “Lightness and retinex theory,” *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, Jan 1971.

[Lan83] Land E., “Recent advances in the Retinex theory and some implications for cortical computations: Color vision and natural image”, *Proc. Natl. Acad. Sci. USA*, vol. 80, pp.5163-5179, 1983.

[Lan83] Land E., “Recent advances in the Retinex theory and some implications for cortical computations: Color vision and natural image”, *Proc. Natl. Acad. Sci. USA*, vol. 80, pp.5163-5179, 1983.

[Mor09] Morel J. M., Petro A. B., and Sbert C., “A PDE formalization of Retinex theory”, *IEEE Trans. Image Process.*, vol.19, pp. 2825–2837., 2010.

[Kim03] Kimmel R., Elad M., Shaked D., Keshet R., and Sobel I., “A variational framework for retinex”, *Int. J. Comput. Vis.*, vol. 52, no. 1, pp. 7–23, 2003.

[Mic11] Michael K. Ng and Wei W., “A Total Variation Model for Retinex”, *SIAM J. Imaging Sciences*, vol. 4, no. 1, pp. 345-365, 2011.

[McC14] McCoy M.B., Cevher V., Dinh Q.T., Asaei A., and Baldassarre L., “Convexity in Source Separation : Models, geometry, and algorithms”, *IEEE Signal Processing Magazine*, vol. 31, issue. 3, pp. 87-95, 2014.

[Don01] Donoho D.L., and Huo X., “Uncertainty principles and ideal atomic decomposition”, *IEEE Trans. Inform. Theory*, vol.47, no.7, pp.2845-2862, 2001.

[Gol09] Goldstein T., and Osher S., “The split Bregman method for L1 regularized problems”, *SIAM J. Imaging Sci.*, vol.2, issue. 2, pp.323-343, 2009.

[MaW10] Ma W., and Osher S., “A TV Bregman Iterative Model of Retinex Theory”, *UCLA CAM Report 10-13*, UCLA, 2010.

[Zha97] Zhang, Xuemei, and Brian A. Wandell. "A spatial extension of CIELAB for digital color-image reproduction." *Journal of the Society for Information Display* 5.1, pp.61-63, 1997.

Reasoning about graph algorithm visualization

Dmitry S. Gordeev

A.P. Ershov Institute of Informatics Systems,
Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
gds@iis.nsk.su

ABSTRACT

A method of graph algorithm visualization based on an implicit visual effect generation approach is described. The approach develops an idea to establish an algorithm as an input as well as input graph. Visualization of algorithms is carried out by means of a set of configurable visual effects. We consider a class of hierarchical graphs as a class of input graphs. This allows using wide set of input graphs and presenting additional data appearing during the algorithm work as part of a single visualized graph model. Described approach is be used both in research and education.

Keywords:

Graph, hierarchical graph, algorithm, visualization, visual effect.

INTRODUCTION

Algorithm visualization is a recurring subject in the field of data visualization. Researchers mostly use them in order to build some educational courses in computer science field [1, 2]. According to review papers a lot of algorithm visualization works are focused on construction new visualization for specific algorithm or for very limited number of ones, and also have different characteristics of effectiveness of visualization. It's remarked in reviews that it's quite difficult to create effective algorithm visualization. Thus most of works have low quality and often it's hard to find working samples for demonstration in education class.

Most of these visualizations cover simple cases of algorithms and data structures. The review paper [2] remarks that there is need for visualization working with more complex data structures like B-trees or NP-hardness. It does make sense in order to familiarize students with more complex concepts of computer science. Often authors of review makes conclusion that there is need for estimation of algorithm visualization effectiveness. Usually they estimate effectiveness in context of time required to understand concept, which visualization presents. Also review authors mark that effectiveness of visualizations is better if students are involved into a process of construction of algorithm visualization. If students have opportunity to interact with visualization then it produces more denotable effect in education.

This paper introduces the approach to construction of algorithm visualization constrained for graph algorithms. It's noted in paper [2] that constrains with domains allow to make corresponding assumption which lead to improvement of expressiveness and effectiveness of algorithm visualizations. We consider only graph algorithms. It means that a text of an algorithm is written on terms of the graph theory.

RELATED WORK

There are many tools to create algorithm visualization. Most of them focus on creation one-time visualization of particular algorithm which can be stored as video record or number of pictures, and later can be demonstrated in education class. Also there are several tools of software visualization field which show work of some program written with programming language. In this approach state of memory is been visualized [5,6,8,9,11]. Also there are approaches using static images in order to show dynamic process [4]. Also there are works which proceed with limited fix set of algorithms [12].

This does not allow user to change input algorithms what would be high degree of freedom in interaction with algorithm visualization. According to reviews [2] it would be useful to have opportunity to interact with visualization process for education purposes.

VISUALIZATION

In this work we consider approach in which we limit only graph algorithms instead of general algorithms. It means that we consider algorithm working only with graphs. It allows us to select events expressed in terms of graphs. For example adding of vertex, remove edge from subgraph or change attribute value of tree node or enumerate nodes incident to given one. And we can describe graph

Permission to make digital or hard copies of all part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

algorithms in such terms with using some programming language [10]. In this form algorithm can be compiled into executable program. The result of the program execution is information which is to be used in creation of the underlying algorithm visualization. An example of such instruction can be adding an edge or a change in the attributes of vertices. The following example shows the breadth-first search algorithm for any graph. In the given case, *Get(...)* and *Set(...)* instructions are used for reading and changing the graph element's attribute values. These instructions have formats *Get(vertex, attributename)* and *Set(vertex, attributename, attributevalue)*, respectively. Here is we have the first aspect of visualization interactivity. We change algorithm text and get a visualization with another properties without additional efforts. To construct a visualization of the breadth-first search algorithm, the state attribute is appointed to each graph vertex. The value of the state attribute reflects whether the vertex was visited during a traversal of a graph.

```
VertexQueue.Enqueue(Graph.Vertices[0]);
while (VertexQueue.Count > 0)
{
    Vertex v = VertexQueue.Dequeue();
    Set(v, "state", "visited");
    foreach (Edge e in v.Edges)
    {
        Vertex t = e.PortTo.Owner;
        string c = Get(t, "state");
        if (c != "visited")
        {
            Set(t, "state", "visited");
            VertexQueue.Enqueue(t);
        }
    }
}
VertexQueue.Clear();
```

Each instruction of the algorithm generates one or more images of the current state of the graph model. The graph model is a hierarchical marked graph. The hierarchical graph *H* is a tuple of two elements: the first is a graph *G* and the second is a tree of fragments. Each fragment is a subgraph of the graph *G*. For any two fragments *U* and *V*, only one of the following properties holds: *U* is a subgraph of *V*, *V* is a subgraph of *U*, or *U* equals *V* [3]. It is useful to highlight the current executing instruction in each image because it allows a user to keep attention on valuable events at this moment. To solve the problem of highlighting the current executing instruction in the image, the following approach is used. Each text line has a numeric index in all text lines. So that order value is added to arguments of the function corresponding to the text line. This additional parameter is the number of the current executing algorithm instruction. After this transformation, the text of the breadth-first search algorithm from the above example looks like this:

```
VertexQueue.Enqueue(Graph.Vertices[0]);
while (WhileCondition(2, VertexQueue.Count > 0))
{
```

```
    Vertex v = VertexQueue.Dequeue(4);
Poster Proceedings
```

```
    Set(5, v.ID, "state", "visited");
    foreach (Edge e in ForCollection(6, v.InEdges))
    {
        Vertex t = e.PortFrom.Owner;
        string c = Get(9, t, "state");
        if (IfCondition(10, c != "visited"))
        {
            Set(12, t, "state", "visited");
            VertexQueue.Enqueue(13, t);
        }
    }
}
VertexQueue.Clear();
```

The above example shows changes in the attributes of the graph elements, too. This is a typical situation for algorithms implementing only traversal of a graph – a method when all graph vertices are visited one by one. For example, the Pruefer encoding algorithm constructs a sequence of numbers by the given tree graph. During the coding process, the vertices of the graph are removed one by one. To perform this operation, the *RemoveVertex(...)* instruction should be used, which leads to generation of a visual effect of the corresponding vertex disappearing. Here is an example of the Pruefer encoding algorithm, how it can be formulated as a parameter of the graph algorithm visualization system:

```
int i=0;
List<Vertex> Leafs = new List<Vertex>();
int n = Graph.Vertices.Count;
while (i++ <= n-2)
{
    Leafs.Clear();
    foreach (Vertex v in Graph.Vertices)
    if (v.OutEdges.Count == 0) Leafs.Add(v);
    Vertex codeItem =
        Leafs[0].InEdges[0].PortFrom.Owner;
    Output.Add(codeItem);
    RemoveVertex(Leafs[0]);
}
```

Each algorithm instruction generates some information during execution of the transformed text of the original algorithm. This information describes the number of the current instruction, the name of an attribute of a graph element, the previous value of the attribute, a new value of the attribute and the identifier of the graph element. This information allows us to get the full history of operations executed over graph elements. This operation history contains the detailed information on the state of the graph model during the algorithm running. Further the history of operations, the input graph and the original text of the algorithm are used to generate the algorithm visualization. Each operation history element corresponds to some graphical effect over visual representation of graph elements. The simplest example of the visual effect for the breadth-first search algorithm is to change the color of the graph vertex representation when a state attribute of the vertex has been changed and to change the color of the text of the corresponding instruction. Here is the second aspect of visualization interactivity. We changes an assignment any particular visual effect and get changed.

EXPERIMENTS

A system of graph algorithm visualization has been constructed based on suggested approach. This system consists of several components: an algorithm execution module, a graph editor and a graph algorithm visualizer. It can be assumed without loss of generality that data are passed between components in a text form. This is useful if the components are implemented on different platforms and with different tools. The purpose of the algorithm execution module is to generate the execution history. The algorithm running is separated from its visualization. This allows performing the algorithm once and after that the operation history can be used to visualize and refine the visualization as many times as needed. This can be useful when computationally-intensive algorithms are visualized since the second cycle of execution of the algorithm is complex in such cases. To provide correct work of the algorithm execution module, it is necessary to meet a significant condition related to the algorithmic complexity. It is reasonable to visualize only efficient algorithms, because it will take much time to build the operation history of execution of an inefficient algorithm. Efficient algorithm means that an algorithm of polynomial complexity. We can use a small input graph for this case. This assumption allows constructing of visualization for a reasonable time.

The algorithm execution module takes the given algorithm text written down with a programming language, executes it and returns the log of operations generated during the algorithm run on a particular graph. The log of executed operations contains information about all changed attributes of graph elements and other events related to input graph elements during the execution. Further this information is used to generate graphical representation of process.

Another component of the visualization system is the visualizer itself. This component receives the algorithm text, the graph, the history of operations and additional graphical options. A history information item is added by special instructions created at the stage of preparation of the algorithm text. For example, these special instructions are the functions: *Set(...)*, *Get(...)*, *IfCondition(...)*, *WhileCondition(...)* and *ForeachCollection(...)*. Their first argument is the number of the corresponding text line. *IfCondition(...)* and *WhileCondition(...)* do not perform any changes in the graph model state but at least allow making a visual selection of the text line where it was inserted. *ForeachCollection(...)* is to be used to generate information which allows highlighting a set of vertices before they will be actually enumerated. To add these functions into appropriate places of the original text of the algorithm, it is sufficient to use a contextual replacement. The purpose of the preparation stage is to eliminate the need for declarative structures, which have no relation to the actual nature of the algorithm.

A history item may also contain information about the

value of an attribute of a graph element. A graph element is a vertex, an edge or a port. If there is a vertex with its incident edge, then a port is a point where the edge enters the vertex. When rendering, it can be useful that the points are allocated for these additional objects. Ports simplify calculation of coordinates of graphical primitives which represent the edge elements. Strictly mathematically, it is possible to simulate a port with a labeled vertex. So the class of graphs with ports is isomorphic to the class of all graphs. An attribute of a vertex, an edge or a port can have a string name and a string value. The history of operations stores the previous value of the attribute for a particular graph element. This information is also useful for building the visualization, since it is possible to make a smooth visual effect from a previous value of an attribute to its new value. It is not obvious how to bind information from a history element to the visual effect. In this case, a user needs to interfere in order to set an explicit binding between the set of attributes in the text of the algorithm and the desired visual effects. For example, if the operation of a log item is about changing the coordinates of the graph element reflected with the use of the attribute "position", then it is reasonable to bind the attribute with the visual effect, which leads to a shift of the graph element. Another user example is to bind all log items to the effect of a color mark of a current graph element under processing. It can be a current vertex visited in the algorithm of deep-first search or in any other graph traversal. In this aspect the suggested approach is close to the interesting events approach, where an algorithm instruction is an interesting event. The figure below shows an example of visualization of the deep-first search algorithm on the graph, which is actually a binary tree graph. The figure is one of the screenshots taken during the process of visualization of the deep-first search algorithm. The left side of the figure displays the text of the algorithm formulated in terms of graphs. The attribute of a graph vertex state indicates the fact that the vertex has already been visited during the process of the graph traversal. A line of the algorithm text has one of the following states: dark thin, light thin and thick. The first state means that the instruction has been executed at least once. The second state means that the current image and the last shown visual effect is the result of this instruction. The last state means that the instruction has not been executed yet. The right part of the figure displays the graph model, which is a hierarchical graph with attributes. Only if this attribute is set, the corresponding attribute will be created during visualization. In this example, the visited vertices get the state attribute that changes the color of a vertex. Also, this attribute's value corresponds to the increase of line width showing the graph vertex circle. Vertices shown in a thin line has not been visited yet.

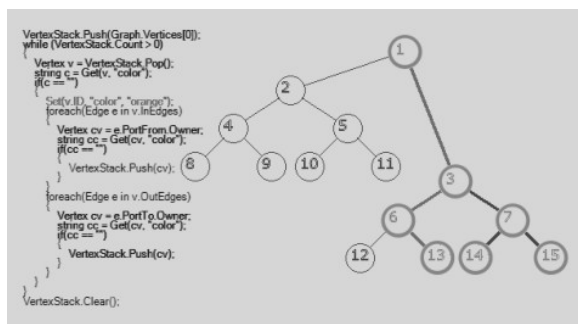


Fig. 1. Visualization of the deep-first search algorithm. This is one of intermediate images.

Displaying of additional data structures can be used to improve understanding of visualization of a graph algorithm. For example, the deep-first search algorithm uses a stack and the breadth-first search algorithm visualization uses a queue. The content of a stack or a queue can be represented as a graph. Since the visualization system allows us to use the hierarchical graphs, a stack graph or a queue graph can be included into a graph model for a particular visualization. So the working graph model consists of a graph with two vertices. The first vertex contains a stack graph and the second contains an input graph. Such graph model can be visualized with the created module of the system of graph algorithm visualization. The queue or stack size is changed during execution of the given algorithm and the corresponding vertices are added or removed from the stack graph. Hierarchical graphs are helpful for this purpose. If there is no stack or queue, then a tree of fragments only consists of one fragment, the input graph. For a stack the graph model consists of three fragments: a root and two children. The first child is the input graph and the second is a graph representation of the stack. So, if the given algorithm uses an input graph and N additional structures, then the tree of fragments contains $N+2$ elements. It is a root element and its $N+1$ children, one of which is the input graph and others are graph representations of additional data structures.

ACKNOWLEDGMENT

This work was supported in part by the Russian Foundation for Basic Research under grant N 15-07-02029.

CONCLUSIONS AND FUTURE WORK

This paper describes the approach of visualization of graph algorithms, providing the capability to build visualization with the help of a flexible system of visual effects and using the algorithm as an input parameter.

The system of graph algorithm visualization has been created in order to test proposed approach. The system uses runtime information of algorithm execution in order to construct corresponding visual effects. Visual effects are used to reflect intermediate states of graph model during algorithm execution.

Interactivity of visualization is supported by the capability to configure visual effects, to change the text of the algorithm and to build visualization once again. The implemented system allows us to observe the performed changes after restart of execution. Simple hierarchical graphs are considered to be a parameter for the visualization system. The class of algorithms admissible for the system is a subject for further research.

At present, the described method is used for construction of a visualization subsystem of a cloud parallel programming system being under development at Institute of Informatics Systems in Novosibirsk.

REFERENCES

- [1] Christopher D. Hundhausen, Sarah A. Douglas, John T. Stasko, A Meta-Study of Algorithm Visualization Effectiveness // *Journal of Visual Languages & Computing*, - Volume 13, Issue 3, - June 2002, - pp. 259-290.
- [2] Clifford A. Shaffer, Matthew L. Cooper, Alexander Joel D. Alon, Monika Akbar, Michael Stewart, Sean Ponce, and Stephen H. Edwards. 2010. // *Algorithm Visualization: The State of the Field*. *Trans. Comput. Educ.* - 10, 3, Article 9, - 22 pp.
- [3] Kasyanov V. N., Yevstigneyev V. A. *Graphs in programming: processing, visualization and application*. - SPb. BHV-Petersburg, 2003. - 1104 with. silt. ISBN 5- 94157-184-4.
- [4] Sorting algorithm visualizations [electronic source]. Available from <http://sortvis.org/visualisations.html> (accessed 01.05.2015).
- [5] Lisitsyn I.A., Kasyanov V.N. *Higres - visualization system for clustered graphs and graph algorithms* // *Proc. of Graph Drawing 99*. - Berlin a.o.: Springer Verlag, 1999. - P. 82-89. - (Lect. Notes in Comput. Sci.; Vol. 1731).
- [6] Higres graph drawing system [electronic source]. Available from <http://pcosrv.iis.nsk.su/higres/> (accessed 01.05.2015).
- [7] Demetrescu C., Finocchi I., Stasko J. T., Specifying Algorithm Visualizations: Interesting Events or State Mapping? // *In Proc. of Dagstuhl Seminar on Software Visualization - Lect. Notes in Comput. Sci.* - 2001. - P. 16-30.
- [8] C. Demetrescu and I. Finocchi, A general-purpose logic-based visualization framework, *Proceedings of the 7th International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media (WSCG'99)*, pp. 55-62, Plzen, Czech Republic, February 1999.
- [9] Leonardo programming environment [electronic source]. Available from <http://www.dis.uniroma1.it/~demetres/Leonardo/> (accessed 01.05.2015).
- [10] Gordeev D.S. Model of interactive visualization of graph algorithms. // *Works of KIS 2011* /

Workshop: The knowledge-intensive software. - Novosibirsk: IIS SB RAS, 2011. - pp. 58-62.

[11] Moreno A., Sutinen E., Joy M. Defining and evaluating conflictive animations for programming education: the case of Jeliot ConAn. // Proceedings of the 45th ACM technical symposium on Computer science education. - 2014. - pp. 629-634.

[12] Naps, T. L., Rößling G. HAVÉ–More Visualizers (and Visualizations) Needed. // Electronic Notes in Theoretical Computer Science. - 2007. - V.178, I. 0 – pp. 33-41.

Computed Tomography Image Analysis: The Model of Ribs-Bounded Contour

Mykolas J. Bilinskas
Institute of Informatics and
Mathematics, Vilnius University
Akademijos str. 4
LT-08663 Vilnius, Lithuania
mykolas.bilinskas@mii.vu.lt

Gintautas Dzemyda
Institute of Informatics and
Mathematics, Vilnius University
Akademijos str. 4
LT-08663 Vilnius, Lithuania
gintautas.dzemyda@mii.vu.lt

Mantas Trakymas
National Cancer Institute
Santariškių str. 1
LT-08660, Vilnius, Lithuania
mantas.trakymas@nvi.lt

ABSTRACT

In this paper, a method for analyzing transversal plane images from computer tomography scans is presented. This method allows not only the approximation of ribs-bounded contour but also the evaluation of patient rotation around the vertical axis during a scan. A mathematical model that describes the ribs-bounded contour was created and the problem of approximation is solved by finding out the optimal parameters of the mathematical model using least-squares.

Keywords

Ribs, ribs-bounded contour, mathematical model, computed tomography, cardioid, least squares, optimization.

1. INTRODUCTION

In medicine, more and more decisions are made via analysis of images. Image data comes from medical diagnostic techniques such as radiology, echoscopy, magnetic resonance, thermovision, tomography, etc.

Computed tomography (CT) is a technology allowing the inside of objects to be spatially viewed using computer-processed X-rays. It is very important in medical diagnostics because it shows human internal organs without cutting, e.g. brain, liver [Nug08], prostate [Che11]. CT scans are 3D images – a collection of 2D images (slices), representing slices by transversal plane. This paper deals with the finding the ribs-bounded contour. This is important for internal organ localization. In liver localization, the liver tissue looks very similar to intercostal muscles, which are separated by thin and almost invisible contours (where liver touches ribs). So, the ribs bounded contour restricts the region of location of internals in the slice. Internal organ localization is particularly important for the comparison of patient health state or to evaluate disease processes. Here the image registration problem arises (see e.g. [Tre08]).

CT scans show tissues of different electron density; bones have the highest density [Nug08] so bone tis-

sue has the highest intensity in the CT image. The goal of our paper is to define the ribs-bounded contour by some mathematical model and to propose a way to evaluate the parameters of this model from the CT image. In this paper, we are restricting the slices to ones where ribs are visible, but this does not lessen the significance of our work, because many important internal organs are located here: liver, heart, stomach, pancreas, lung, etc.

2. DATA TO BE ANALYZED

Example of CT scan slices is shown in Fig. 1. Depending on the slice, the heart, lungs, stomach, or liver can be seen. In all the cases of Fig. 1 internal organs are bounded by ribs. Fig. 1 was obtained after the patient was given a radiocontrast agent injection. In this image, the heart and aorta as well as all blood vessels are bright. We are investigating two dimensional images of size 512×512 .

3. METHOD

In this section a method of analysis of CT scan slice (through transversal plane) is proposed. This method allows us not only to find the ribs-bounded contour but also to evaluate patient rotation around the vertical axis during the scan.

We notice a symmetry of the ribs by the human sagittal plane, body rotation in respect to bed, aorta near spine, high grayscale intensity of bones, bed, heart and blood vessels (if the radiocontrast agent was injected to the patient). All these must be considered when building a method.

Our method consists of two steps: (a) extracting the bone tissue from image; (b) approximating the ribs-bounded contour with a mathematical function.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

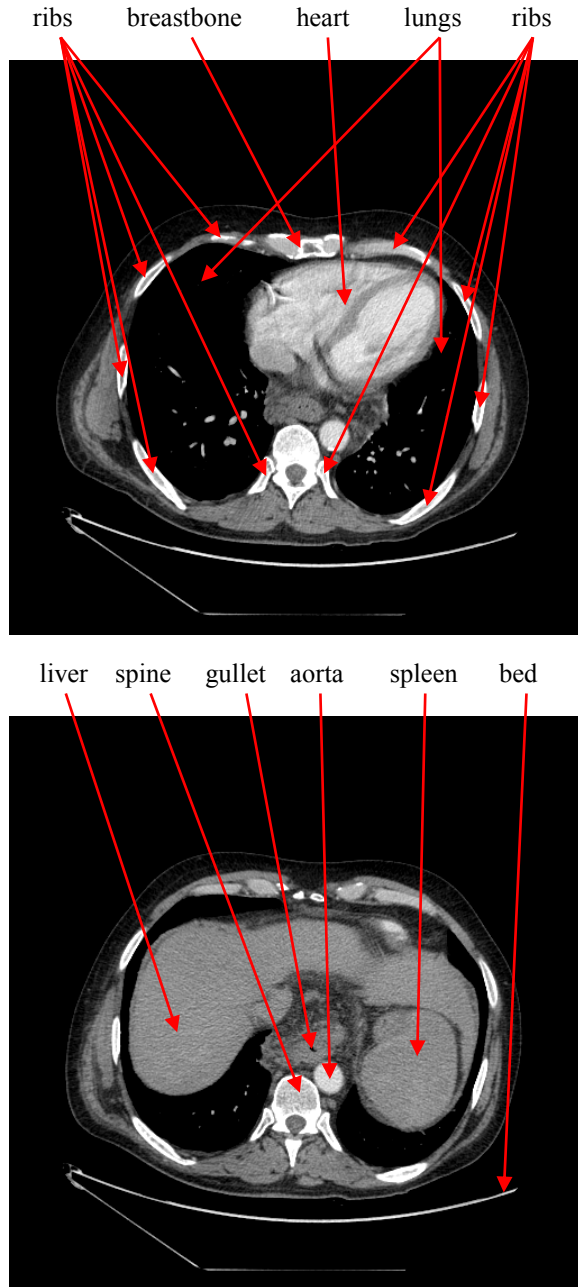


Figure 1. Examples of the CT image slices.

Body tissue extraction

Because bone tissue is naturally brighter than any other tissue, an image can be thresholded – extracting pixels with a value higher than 150 if the whole CT scan is linearly normalized between 0 and 255. This threshold is chosen after the discussion below.

In [Hen02], the electron density of internal organs is evaluated. The densities are given in Table 1. We see that the maximal electron density is of bone, at $5.29 \cdot 10^{23}$, and the minimal electron density of bone is $3.72 \cdot 10^{23}$. The lung has the minimal possible density among internal organs with $0.83 \cdot 10^{23}$. In CT scan slices, lung tissue is purely black. Therefore, the

problem is to transform the interval $[0.83 \cdot 10^{23}; 5.29 \cdot 10^{23}]$ into the interval $[0; 255]$ and to find the point of the last interval corresponding to $3.72 \cdot 10^{23}$. After such conversion, we get the brightness 165 corresponding to the minimal electron density of the bone. Leaving some reserve, we fixed the threshold at 150.

Tissue	Electron Density (electrons/cm ³)
Water	$3.35 \cdot 10^{23}$
Bone	$3.72 - 5.29 \cdot 10^{23}$
Spleen	$3.52 \cdot 10^{23}$
Liver	$3.52 \cdot 10^{23}$
Heart	$3.46 \cdot 10^{23}$
Muscle	$3.44 \cdot 10^{23}$
Kidney	$3.42 \cdot 10^{23}$
Pancreas	$3.40 \cdot 10^{23}$
Fat	$3.07 \cdot 10^{23}$
Lung	$0.83 \cdot 10^{23}$

Table 1. Electron density of body area [Hen02]

After thresholding, undesirable regions can be selected: lateral objects like the bed or metal implants, and everything that contains blood with radiocontrast agent – heart and vessels. These regions can be removed with morphological operations [Rus11].

To extract the bone tissue, we suggest the algorithm:

```

1. image=CTScan.Threshold(150);
2. for each slice in image
    slice=slice.FillIsolatedBlobs;
3. image=image.MorphologicalOpen(2);
4. spine=image.MorphologicalOpen(14);
5. for z from 0 to image.SliceCount
    image[z]=image[z]-
        ((spine[z].Dilate(4)-spine[z])&&
        Rectangle((0, 0)-
        (512,spine.GravityCenter.Y)));
6. convexHullPoints={};
7. for each slice in image
    convexHullPoints= convexHullPoints U
        slice.ConvexHull;
8. image=
    image.FilterConnectedComponents
        (convexHullPoints);

```

In the pseudo code, CTScan is the CT scan image, and the function Threshold(th) transforms the grayscale image into the binary one. Image is the 3D binary image. FillIsolatedBlobs fills with white the isolated black blobs that are on larger white ones, MorphologicalOpen(x) and Dilate(x) are morphological open and dilate operators respectively with round kernel of diameter $2x + 1$. Function Rectangle($(x_0, y_0) - (x_1, y_1)$) makes a binary image with a white filled rectangle at specified coordinates, and $\&\&$ is pixel-wise binary AND operator. The binary image subtraction $X - Y$ compares each pixel and returns the nonimplication. GravityCenter changes a binary blob to its center of gravity, FilterConnectedComponents(seeds) leaves bi-

nary blobs that are connected to seeds only. `ConvexHull` returns a set of points of the convex hull of all white pixels in the binary slice. A set `convexHullPoints` joins such convex hulls points of all slices of the CT scan. `FillIsolatedBlobs` is necessary because inner rib tissue may appear in black after the binarization using the threshold. `Spine[z].Dilate(4)-spine[z]` analyzes the slice z and gives the 4 pixels thickness ring around the spine. Its top part (intersection with rectangle) is subtracted from the image to separate aorta from spine where they touch each other. Then convex hull of slice is actually a convex hull of ribs – bone tissue is blobs that are connected to ribs.

As a result of the bone tissue extraction above, we get binary images – see the example in Fig. 2.



Figure 2. Pixels representing bone tissue.

Denote $B = \{(b_{1i}, b_{2i}), i = \overline{1, m}\}$ the set of coordinates of bone pixels obtained during analysis of CT image slices, m is the number of bone pixels.

Ribs-bounded contour approximation

The ribs form a shape similar to cardioid (see Fig. 1):

$$\rho = 1 + \cos(\varphi - \pi/2), \quad \varphi \in [-\pi/2; 3\pi/2] \quad (1)$$

Here ρ is the radius and φ is the polar angle. The shape of (1) is depicted in Fig. 3 (blue curve). It looks similar, because it features a cave which could be used to approximate ribs cave near spine. $\pi/2$ is introduced in (1) because the standard cardioid is rotated by 90° as compared with Fig. 3 and the ribs-bounded contour in the images should be oriented like ribs depicted in Fig. 1.

Our research showed that a ribs-bounded contour is more condensed vertically than the standard cardioid curve. Therefore, we suggest to add optimizable parameter – power s :

$$\rho = (1 + \cos(\varphi - \pi/2))^s \quad (2)$$

The (2) curve with different s is depicted in Fig. 3.

As we see in Fig. 1, the rib-bounded contour has some rotation in respect to the bed. Therefore, we should introduce the angle θ of such rotation. Parameter s influences not only the vertical scale of the curve (1), but the form of the curve, too (see Fig. 3 for curves with different values of s).

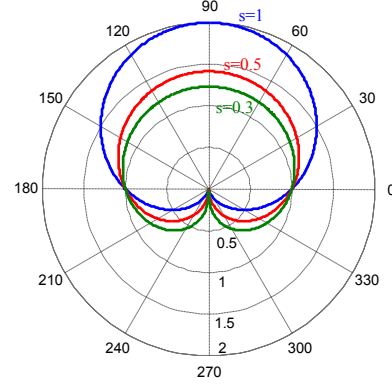


Figure 3. Blue – standard cardioid (1), red and green – (2) curve with $s = 0.5$ and $s = 0.3$.

In the CT scan slice (Figure 1), we see a cave influenced by the breastbone. Curve (2) is convex in this region. Therefore, we need to complement the model (2) redefining ρ with additional member ρ' whose form may vary depending on the cave:

$$\rho = (1 + \cos(\varphi - \pi/2))^s - \rho' \quad (3)$$

This member realizes the cave by subtraction of some value from the right side of (2) starting from $\varphi = \pi/2 - \beta$ till $\varphi = \pi/2 + \beta$. The member ρ' depends on φ and has special properties. It must (a) be unimodal non-negative function on φ , (b) achieve the maximal value as $\varphi = \pi/2$, (c) be symmetrical function in respect of $\varphi = \pi/2$, (d) be equal to 0 when $\varphi = \pi/2 - \beta$ and $\varphi = \pi/2 + \beta$, (e) have zero first and second derivatives on φ when $\varphi = \pi/2 - \beta$ and $\varphi = \pi/2 + \beta$. Function ρ' may be as follows:

$$\rho' = \begin{cases} c \sin^l \left(\frac{\pi(\varphi - \pi/2 + \beta)}{2\beta} \right), & \text{if } \beta \geq |\varphi - \pi/2| \\ 0, & \text{else} \end{cases} \quad (4)$$

In (4), we have three control parameters for which optimal values need to be found: β is an angle, defining the region of subtraction, c defines the maximal value of subtraction, l defines the steepness of curve describing the cave ($l \geq 2$). Example of ρ' is presented in Fig. 4 for different values of β , c , l . Moreover, we need some additional parameters a and b that define the horizontal and vertical scales of the curve that approximates the rib-bounded contour, respectively. The curve (3) should be fitted among ribs in the picture of bone tissue. For this reason, we need the optimal place of the point of (3) corresponding to $\rho = 0$ in the picture; denote coordinates of this point by (x_0, y_0) .

If the values of s , θ , a , b , x_0 , y_0 , β , c , l are fixed, we can draw some parametric curve $(x, y) = (x(\varphi), y(\varphi))$ approximating the rib-bounded contour:

$$\begin{aligned} x &= x_0 + a\rho \cos \varphi \cos \theta - b\rho \sin \varphi \sin \theta \\ y &= y_0 + a\rho \cos \varphi \sin \theta + b\rho \sin \varphi \cos \theta \end{aligned} \quad (5)$$

where ρ is defined by (3). If φ runs through the interval $[-\pi/2; 3\pi/2)$ with a step $2\pi/n$, we get from (5) a sequence of points $C_j = (x_j, y_j), j = \overline{1, n}$ of the curve. In our experiments, $n = 180$.

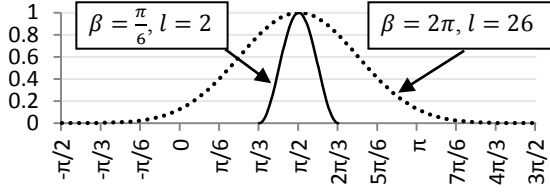


Figure 4. Dependencies of function ρ' on φ . $c = 1$.

Optimization problem

The model of ribs-bounded contour has nine parameters whose values can be varied seeking to find the best approximation of the contour: $s, \theta, a, b, x_0, y_0, \beta, c, l$. The optimal values of these parameters must be defined by the set B of coordinates of bone pixels obtained during the analysis of CT image slices. The optimization problem to find optimal $s, \theta, a, b, x_0, y_0, \beta, c, l$ is formulated as a least square one:

$$\min_{s, \theta, a, b, x_0, y_0, \beta, c, l} f(s, \theta, a, b, x_0, y_0, \beta, c, l) \quad (6)$$

$$f(\cdot) = \sum_i^m \|B_i - C_{k_i}\|^2, \quad k_i = \arg \min_j \|B_i - C_j\|$$

The optimization method to solve (6) may be any local minimization one. We use the Matlab realization of the quasi-Newton method [Unc15].

4. RESULTS

The performance and efficiency of the proposed method is illustrated in Figure 5. Approximation of the ribs-bounded contour is put on the image in red. The optimization results are as follows: $\min f(\cdot) = 212611$, $m = 5403$, $a = 160.51$, $b = 217.01$, $\theta = 0.076$, $s = 0.35$, $x_0 = 265.2$, $y_0 = 296.52$, $\beta = 6.25$, $c = 0.71$, and $l = 28.375$. The average distance from points from B to the curve approximating the ribs-bounded contour is 4.95 pixels. Both θ and β mean radians. Note, that the value of angle β is greater than π . This is acceptable because the values of function ρ' are close to zero at the boundaries of φ domain $[-\pi/2; 3\pi/2)$ when l is large (see Fig. 4).

5. CONCLUSIONS

In this paper, a method for the analysis of transversal plane images of computer tomography scans is presented. This method allows us not only to find the approximation of the ribs-bounded contour but also to evaluate the patient rotation around vertical axis during scans.

The experiments showed that the number m of bone pixels depends on the CT scan slice and is large. The angle θ of patient rotation around vertical axis during the scan is about 4° in the analyzed data. This angle

is not large, but its optimal estimate allowed more precise approximation of the rib-bounded contour. Power s in (2) and (3) is less than 1. The angle β , defining the region of subtraction, is greater than π . This means that the subtraction is performed for all ρ in entire domain of φ . The steepness l of curve describing the cave is very large. The combination of values of angle β and steepness l allows us to describe various forms of cave influenced by the breastbone.

The proposed approximation defines the rib-bounded contour exactly. The model may be applied to any 2D slice where the ribs are visible. Defining the area of internal organs from the ribs-bounded contour restricts essentially the search area, where these organs are located, and may serve as the effective start for the detailed localization of particular organ.

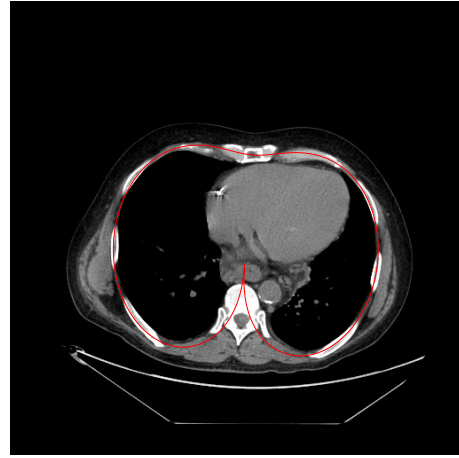


Figure 5. The ribs-bounded contour.

6. REFERENCES

- [Che11] Chen, S., Lovelock, D. M., Radke, R. J. Segmenting the prostate and rectum in CT imagery using anatomical constraints. *Medical Image Analysis* 15, No.1, pp.1-11, 2011.
- [Hen02] Hendee, W. R. and Ritenour, E.R. *Medical Imaging Physics*. Wiley, New York, 2002.
- [Nug08] Nugroho, H.A., Ihtatho, D., Nugroho, H. Contrast Enhancement for Liver Tumor Identification. *The MIDAS Journal: Grand Challenge Liver Tumor Segmentation (2008 MICCAI Workshop)*, 2008.
- [Rus11] Russ, C.J. *Image Processing Handbook*. CRC Press, North Carolina, 2011.
- [Tre08] Treigys, P., Saltenis, V., Dzemyda, G., Barzdziukas, V., Pauksnis A. Automated optic nerve disc parameterization. *Informatica* 19, No.3, pp.403-420, 2008.
- [Unc15] Unconstrained Nonlinear Optimization Algorithms. Mathworks. <http://se.mathworks.com/help/optim/ug/unconstrained-nonlinear-optimization-algorithms.html>.

Statistical solution of 3D transformation problem

Jaroslav Marek
University Pardubice,
Faculty of Electrical
Engineering and
Informatics,
532 10 Pardubice, Czech
Republic
jaroslav.marek@upce.cz

Josef Rak
University Pardubice,
Faculty of Electrical
Engineering and
Informatics,
532 10 Pardubice, Czech
Republic
josef.rak@upce.cz

Pavel Jetensky
University Pardubice,
Faculty of Electrical
Engineering and
Informatics,
532 10 Pardubice, Czech
Republic
pavel.jetensky@student.upce.cz

ABSTRACT

Obtaining the 3D model of an object is currently one of the most important issues that image processing is dealing with. Measurement of the points on 3D objects requires different scans from different positions in different coordinate systems. At our disposal are measured coordinates of an identical point, which can be obtained from a laser 3D scanner, depth sensor, or any motion input device as Microsoft Kinect. A point whose coordinates are known in both coordinate systems is called an identical point. Data transformation of identical points from one coordinate system to another coordinate system is therefore required. The aim of this contribution is to present a possible approach on how to estimate the unknown transformation parameters by regression models in a special transformation problem. This transformation in its standard version has been derived under the assumption that non-negligible random errors occur at points of that coordinate system into which the transformation is performed. Points of the inverse image coordinate system are assumed to be errorless.

Keywords

Transformation of coordinates, estimators of transformation parameters, Helmert transformation, nonlinear regression model, linearization

1 INTRODUCTION

The problem of obtaining a precise 3D digital model of some object today is a very actual theme. It is used for example in geodesy, civil engineering, 3D printing, etc. Measuring of points of 3D objects requires different scans from different sites. Measuring can be performed by terrestrial laserscanning, photogrammetry, using Microsoft Kinect etc. Coordinates of measured points on an object are given in different coordinate systems.

For a full object description, transformation of coordinates to common coordinate system is needed. Such a problem is connected with an estimation of parameters of transformation between different coordinate systems. Relatively large potential for the use of such models arise there, see [PPZN05],[SB05], [Sha06], [ZK03], [JSS15].

For simplicity, let us use case where we have two different systems of coordinates. We will assume that a

measurement error appears only in the second system. The transformation problem includes an estimation of translational parameters and an estimation of a rotation matrix between the two coordinate systems. Studied transformation problem leads to solving the overdetermined system of equations. These equations are often solved using numerical methods, rely on linearization, approximate starting values and iterations. The rotation matrix can be expressed by Cardan angles, Euler angles, angles yaw, pitch and roll, cf. [GA03], [LCSG13], [SMH14]. Large class of statistical problems and different approach for solutions of transformation problem arise there. Literature offer solutions based on Gauss-Jacobi algorithm, Procrustes algorithms, etc. In some cases, where the initial starting parameters are far from true values of unknown parameters, iteration processes may fail to converge. In geodesy an geoinformatics, is often successfully used constraints on orthogonality of rotation matrix. Therefore, the problem we describe by a model with constraints. A detailed inspection of regression model with constraints is provided in [Kub88], [Kub93], [KKV95], [KM04], [KM04].

2 TRANSFORMATION PROBLEM

Our problem is an estimation of the transformation coefficients $\beta_1, \dots, \beta_{12}$ specifying the transformation between the systems of coordinates.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Let us suppose that $i = 1, \dots, n$ are identical points at our disposal for the transformation. The formulated transformation problem can be written in the following form

$$\begin{aligned} \mu_i &= \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \mu_{i3} \end{pmatrix} = \phi(\beta, v) = \gamma + \mathbf{T}v_i \\ &= \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \beta_4 & \beta_5 & \beta_6 \\ \beta_7 & \beta_8 & \beta_9 \\ \beta_{10} & \beta_{11} & \beta_{12} \end{pmatrix} \begin{pmatrix} v_{i1} \\ v_{i2} \\ v_{i3} \end{pmatrix}, \\ i &= 1, \dots, n. \end{aligned} \quad (1)$$

Instead of vectors of actual coordinates $v_{I,i}$ (System I), $i = 1, \dots, n$, and $\mu_{II,i}$ (System II), $i = 1, \dots, n$, estimators of them only are at our disposal.

The estimator of μ has the form of random vectors $\mathbf{Y}_{II} \sim (\mu_{II}, \Sigma_{Y_{II}})$; the mean value of the random vector \mathbf{Y}_{II} is $E(\mathbf{Y}_{II}) = \mu_{II} = (\mu_{II,1}, \dots, \mu_{II,n})'$ and its covariance matrix $\text{var}(\mathbf{Y}_{II}) = \Sigma_{Y_{II}}$.

We will assume $\Sigma_{Y_{II}}$ is a positive definite (i.e. regular).

The problem in determining the optimum estimators of the unknown transformation parameters and of the transformed coordinates of the identical points simultaneously is caused by the fact that they are not errorless, which leads to application of the least squares method. Note that there are corrections of the coordinates within the System II.

2.1 Orthogonality condition

We assume that the systems of coordinates are orthogonal. The transformation coefficients should satisfy the condition

$$\begin{aligned} \mathbf{h}(\beta) &= (h_1(\beta), h_2(\beta), h_3(\beta), h_4(\beta), h_5(\beta), h_6(\beta))' \\ &= (1, 0, 0, 1, 0, 1)', \end{aligned} \quad (2)$$

where

$$\begin{aligned} h_1(\beta) &= \beta_4^2 + \beta_5^2 + \beta_6^2, \\ h_2(\beta) &= \beta_4\beta_7 + \beta_5\beta_8 + \beta_6\beta_9, \\ h_3(\beta) &= \beta_4\beta_{10} + \beta_5\beta_{11} + \beta_6\beta_{12}, \\ h_4(\beta) &= \beta_7^2 + \beta_8^2 + \beta_9^2, \\ h_5(\beta) &= \beta_7\beta_{10} + \beta_8\beta_{11} + \beta_9\beta_{12}, \\ h_6(\beta) &= \beta_{10}^2 + \beta_{11}^2 + \beta_{12}^2. \end{aligned}$$

3 REGRESSION MODEL WITH CONSTRAINTS AND ITS LINEARIZATION

If we know good approximations $\mu_{I,0}$ of the vector μ_I and β_0 of the parameters $\beta_1, \dots, \beta_{12}$, respectively,

such that the vectors $\delta\mu_I$ and $\delta\beta$ can be neglected ($\delta\mu_I = \mu_I - \mu_{I,0}$), then the linearized version of the transformation problem (1) with constraint (2) can be used. After linearization, we can write our model in form of linear model with linear constraint

$$\begin{pmatrix} \hat{\mu}_1 - \mathbf{X}_1\beta_0 \\ \vdots \\ \hat{\mu}_I - \mathbf{X}_I\beta_0 \end{pmatrix} \sim_{3I} \left[\begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_I \end{pmatrix} \delta\beta, \begin{pmatrix} \Sigma_1 & \dots & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \dots & \Sigma_I \end{pmatrix} \right], \quad (3)$$

$$\mathbf{b} + \mathbf{B}\delta\beta = 0 \quad (4)$$

Let consider $\Sigma_i = \sigma^2 \mathbf{V}$, the value of σ can be adapted from the documentation protocol of the measurement device. In our case, $\Sigma_i = (0.03 \text{ m})^2 \cdot \mathbf{I}_{3 \times 3}$.

The best linear unbiased estimator of unknown parameter is (see [KKV95], [K13])

$$\begin{aligned} \hat{\delta\beta} &= \left\{ \mathbf{I} - \mathbf{C}\mathbf{B}' [\mathbf{B}\mathbf{C}\mathbf{B}']^{-1} \mathbf{B} \right\} \times \\ &\times \mathbf{C}(\mathbf{X}_1'\Sigma_1^{-1}, \mathbf{X}_2'\Sigma_2^{-1}, \dots, \mathbf{X}_I'\Sigma_I^{-1}) \times \\ &\times \begin{pmatrix} \hat{\mu}_1 - \mathbf{X}_1\beta_0 \\ \vdots \\ \hat{\mu}_I - \mathbf{X}_I\beta_0 \end{pmatrix} - \mathbf{C}\mathbf{B}' [\mathbf{B}\mathbf{C}\mathbf{B}']^{-1} \mathbf{b}, \\ \mathbf{C} &= \left(\sum_{i=1}^I \mathbf{X}_i'\Sigma_i^{-1} \mathbf{X}_i \right)^{-1}, \end{aligned} \quad (5)$$

$$\mathbf{X}_i = \mathbf{1}_n \otimes \gamma + (\mathbf{I}_n \otimes \mathbf{T}) v_i, \quad (6)$$

$$\mathbf{I} \text{ is an identical matrix,} \quad (7)$$

$$\mathbf{1} \text{ is a matrix with entries equal to 0,} \quad (8)$$

$$\otimes \text{ is the Kronecker product.} \quad (9)$$

Matrix $\mathbf{B} = \frac{\partial \mathbf{h}}{\partial \beta}$ is of the form

$$\mathbf{B} = (\mathbf{0}_{6,3}, \mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3) \quad (10)$$

$$\mathbf{B}_1 = \begin{bmatrix} 2\beta_{0,4} & 2\beta_{0,5} & 2\beta_{0,6} \\ \beta_{0,7} & \beta_{0,8} & \beta_{0,9} \\ \beta_{0,10} & \beta_{0,11} & \beta_{0,12} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{B}_2 = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{0,4} & \beta_{0,5} & \beta_{0,6} \\ 0 & 0 & 0 \\ 2\beta_{0,7} & 2\beta_{0,8} & 2\beta_{0,9} \\ \beta_{0,10} & \beta_{0,11} & \beta_{0,12} \\ 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{B}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{0,4} & \beta_{0,5} & \beta_{0,6} \\ 0 & 0 & 0 \\ \beta_{0,7} & \beta_{0,8} & \beta_{0,9} \\ 2\beta_{0,10} & 2\beta_{0,11} & 2\beta_{0,12} \end{bmatrix}$$

and \mathbf{b} is of the form

$$\mathbf{b} = \begin{pmatrix} \beta_{0,4}^2 + \beta_{0,5}^2 + \beta_{0,6}^2 - 1 \\ \beta_{0,4}\beta_{0,7} + \beta_{0,5}\beta_{0,8} + \beta_{0,6}\beta_{0,9} \\ \beta_{0,4}\beta_{0,10} + \beta_{0,5}\beta_{0,11} + \beta_{0,6}\beta_{0,12} \\ \beta_{0,7}^2 + \beta_{0,8}^2 + \beta_{0,9}^2 - 1 \\ \beta_{0,7}\beta_{0,10} + \beta_{0,8}\beta_{0,11} + \beta_{0,9}\beta_{0,12} \\ \beta_{0,10}^2 + \beta_{0,11}^2 + \beta_{0,12}^2 - 1 \end{pmatrix}, \quad (11)$$

$$\mathbf{X}_i = (\mathbf{I}, \mathbf{X}_i^1, \mathbf{X}_i^2, \mathbf{X}_i^3), \quad (12)$$

$$\mathbf{X}_i^1 = \begin{pmatrix} v_{i,1} & v_{i,2} & v_{i,3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{X}_i^2 = \begin{pmatrix} v_{i,1} & v_{i,2} & v_{i,3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{X}_i^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ v_{i,1} & v_{i,2} & v_{i,3} \end{pmatrix}.$$

3.1 Conversion of coordinates

For conversion of the non-identical points to another system of coordinates we have the formula

$$\hat{\mu}_j = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \beta_4 & \beta_5 & \beta_6 \\ \beta_7 & \beta_8 & \beta_9 \\ \beta_{10} & \beta_{11} & \beta_{12} \end{pmatrix} \mathbf{v}_j, \quad (13)$$

or formula

$$\hat{\mathbf{v}}_j = - \begin{pmatrix} \beta_4 & \beta_5 & \beta_6 \\ \beta_7 & \beta_8 & \beta_9 \\ \beta_{10} & \beta_{11} & \beta_{12} \end{pmatrix} \left[\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} - \hat{\mu}_j \right]. \quad (14)$$

4 EXAMPLE

Let us think about the task in Figure 1.

The points in the left picture were measured in coordinate system S_1 from position 1. The number of measured points is 27. Let us sign them with symbol μ .

The points in the right picture were measured in coordinate system S_2 from position 2. The number of measured points is 24. Let us sign them with symbol \mathbf{v} .

In both scans I identical points were polarized. In our case on the picture, the number of identical points is 14.

The first three points were the corners of the building and the fourth was the top of the tower. The coordinates are given at Table 1 and 2.

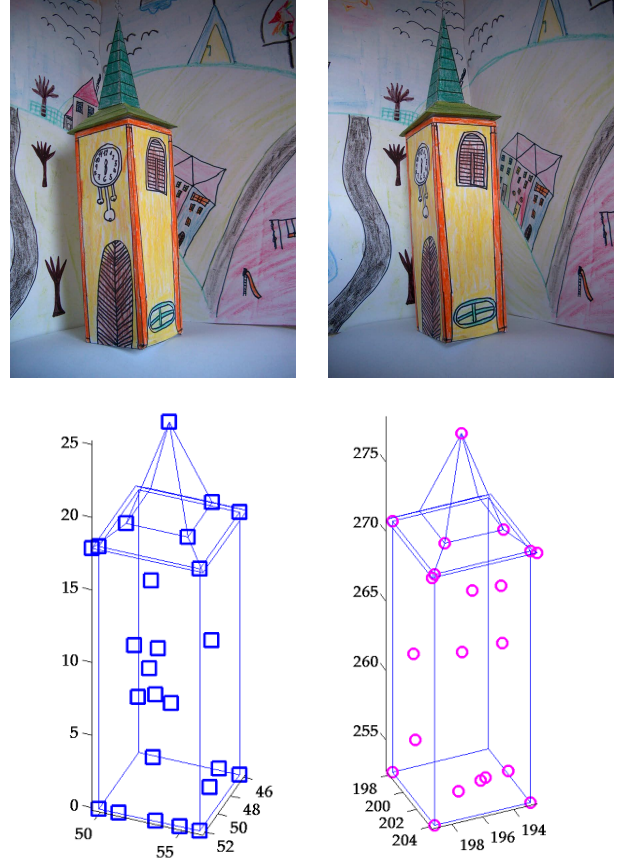


Figure 1: 3D object.

By transformation of formula (1) we get the non-consistent system of equations (in our case 42 equations with 12 unknowns). To get a solution, the method of least squares can be used. Another way is to get the solution using the first 4 identical points.

Using the method of least square we get the solution:

i	\mathbf{v}_i		
1	199.0011	197.7993	252.8010
2	198.9996	203.9991	252.8037
3	192.7959	204.0007	252.7993
4	195.8986	200.8992	278.2588
5	199.0003	197.8015	270.8577
6	199.0032	204.0042	270.8596
7	192.8020	203.9973	270.8624
8	197.7968	202.7980	272.0078
9	193.9998	202.8021	272.0113
10	198.9980	201.1369	257.1768
11	198.9975	200.9067	263.2470
12	195.9970	204.0031	255.2122
13	197.4546	204.0014	254.8240
14	197.2252	204.0039	264.8107

Table 1: Measured points — first scan

i	μ_i		
1	51.9913	49.7667	-0.1975
2	52.0058	55.9771	-0.1762
3	45.8238	55.9992	-0.1935
4	48.8862	52.9172	25.2851
5	52.0035	49.7963	17.8745
6	51.9882	56.0437	17.8573
7	45.8023	56.0213	17.8612
8	50.7981	54.7834	19.0159
9	46.9733	54.8143	19.0425
10	51.9681	53.1109	4.1871
11	51.9920	52.9199	10.2655
12	49.0106	56.0258	2.2254
13	50.4793	55.9760	1.8276
14	50.2220	55.9679	11.8168

Table 2: Measured points — second scan

$$\hat{\beta}^0 = \begin{pmatrix} \beta_1^0 & \beta_4^0 & \beta_5^0 & \beta_6^0 \\ \beta_2^0 & \beta_7^0 & \beta_8^0 & \beta_9^0 \\ \beta_3^0 & \beta_{10}^0 & \beta_{11}^0 & \beta_{12}^0 \end{pmatrix} = \begin{pmatrix} -146.7812 & 0.9984 & 0.0015 & -0.0008 \\ -148.4136 & -0.0020 & 1.0022 & 0.0013 \\ -252.6847 & -0.0008 & -0.0010 & 1.0002 \end{pmatrix}.$$

The problem is that our solution does not satisfy the orthogonality condition $\mathbf{h}(\hat{\beta}_0) = 0$ because

$$\mathbf{h}(\hat{\beta}_0) = \begin{pmatrix} 0.9969 \\ -0.0005 \\ -0.0016 \\ 1.0045 \\ 0.0003 \\ 1.0004 \end{pmatrix}.$$

In our case, $\Sigma_i = (0.03 \text{ m})^2 \cdot \mathbf{I}_{3 \times 3}$. Now using the formula (5), we get the estimate

$$\hat{\delta\beta} = \begin{pmatrix} \hat{\beta}_1 & \hat{\beta}_4 & \hat{\beta}_5 & \hat{\beta}_6 \\ \hat{\beta}_2 & \hat{\beta}_7 & \hat{\beta}_8 & \hat{\beta}_9 \\ \hat{\beta}_3 & \hat{\beta}_{10} & \hat{\beta}_{11} & \hat{\beta}_{12} \end{pmatrix} = \begin{pmatrix} -0.5992 & 0.0016 & 0.0012 & 0.0002 \\ 0.6167 & -0.0007 & -0.0022 & -0.0001 \\ -0.1496 & 0.0014 & -0.0001 & -0.0004 \end{pmatrix}.$$

The final result is obtained using the formula

$$\hat{\beta} = \hat{\beta}^0 + \hat{\delta\beta} = \begin{pmatrix} -147.3804 & 1.0000 & 0.0027 & -0.0006 \\ -147.7969 & -0.0027 & 1.0000 & 0.0012 \\ -252.8344 & 0.0006 & -0.0011 & 0.9998 \end{pmatrix}.$$

and the constraints (2) is now of the form

$$\mathbf{h}(\hat{\beta}) = \begin{pmatrix} 1.0000 \\ -0.0000 \\ 0.0000 \\ 1.0000 \\ 0.0001 \\ 0.9996 \end{pmatrix}.$$

Let us try to recalculate the coordinates of a point that matches the top of the clock, which can be found in Fig. 1. In the system S_2 , the point has coordinates

$$\mu = \begin{pmatrix} 52.0000 \\ 53.0000 \\ 16.2963 \end{pmatrix}.$$

By using the formula (14), where we put estimate $\hat{\beta}$ to

$$\beta_1, \dots, \beta_{12}, \text{ we get } \hat{\nu} = \begin{pmatrix} 199.7611 \\ 200.5815 \\ 268.9756 \end{pmatrix}.$$

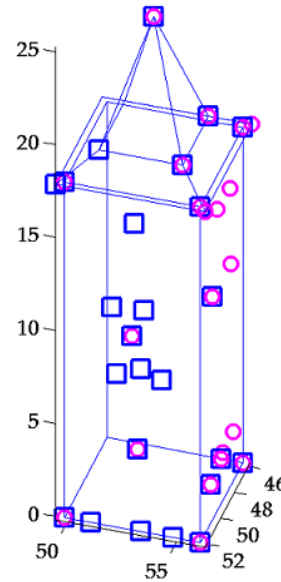


Figure 2: Points after transformation.

5 CONCLUDING REMARKS

In this paper, the problem to determine the optimum estimators of the unknown transformation parameters and of the transformed coordinates of the identical points simultaneously is solved.

A rotation transformation matrix and shift parameters can be estimated using a regression model, where rotation parameters fulfill the orthogonality constraints. This constraints is barely fulfilled for the initial solution/average solution.

To illustrate the whole process of estimation, we presented the results of an algorithm in an example.

6 ACKNOWLEDGMENTS

This research was supported by the Internal Grant Agency of University of Pardubice, the project SGFEI 08/2015.

7 REFERENCES

- [GA03] Grafarend, E.W., Awange, J.L. Nonlinear analysis of the three-dimensional datum transformation [conformal group C7(3)]. *Journal of Geodesy*, 77, pp. 66-76, 2003
- [JSS15] Jaszuk, M., Szostek, G., Starzyk, J.A. Building 3D Object Representation Using SURF Local Features. WSCG, International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision, Plzen, 2015.
- [KKV95] Kubacek, L., Kubackova, L. and Volaufova, J. Statistical models with linear structures. Veda, Publishing House of the Slovak Academy of Sciences, Bratislava, 1995.
- [K13] Kubacek, L. Statistical theory of geodetic networks. VUGTK, Zdiby, 2013. ISBN 978-80-85881-31-8.
- [KM04] Kubacek, L. and Marek, J. Partial optimum estimator in two stage regression model with constraints and a problem of equivalence. *Math. Slovaca* 54, 2004.
- [Kub88] Kubáček, L. Two stage regression model. *Math Slovaca*. 38, pp 383-393, 1998
- [Kub93] Kubáček, L. Two stage regression models with constraints. *Math Slovaca*. 43, pp 643-658, 1993
- [KM04] Kubáček, L. and Marek, J. Partial optimum estimator in two stage regression model with constraints and a problem of equivalence. *Math Slovaca* 55, pp 477-494, 2005
- [LCSG13] Liu, K., Da Costa, J.P.C.L., So, H.C., De A. Gadelha, L.F.A., Roemer, F., Haardt, M. 3-D unitary ESPRIT: curate attitude estimation for unmanned aerial vehicles with a hexagon-shaped ESPAR array. *Digital Signal Processing*, vol. 23, no. 3, pp. 701–711, 2013
- [PPZN05] Pavelka, K., Pazdera, J., Záleská, J., and Nečesánek, M. Using laser scanner in documentation of historical roof Optical 3-D Measurement Techniques VII, Vol I, Vienna University of Technology. pp. 45-50, 2005
- [SB05] Schuhmacher, S., Böhm, J. Georeferencing of Terrestrial Laserscanner Data for Applications in Architectural Modelling. 2005
- [Sha06] Shah T.R. Automatic Reconstruction of Industrial Installations Using Point Clouds and Images. *Publications on Geodesy* 62. NCG. Delft, The Netherlands, 2006
- [SMH14] Svoboda, M., Marek, J., and Heckenbergerova, J. Estimation of angles yaw, pitch and roll in a special transformation problem. *Nos-tradamus 2014: Prediction, Modeling and Analysis of Complex Systems*. Berlin: Springer, pp. 393-400, 2014. ISBN 978-3-319-07400-9.
- [ZK03] Zámečníková, J., Kopáček, A. Terrestrial Laser Systems. 11th Slovak Geodetic Days, Chamber of Surveyors and Cartographers, Bratislava, pp. 89-97, 2003

Hough Transform for the Calculation of Twist Angle of Aramid Torsion

Perechesova A.D.
SPbF IZMIRAN, ITMO University*
Universitetskaya emb., 5 B
Kronverkskiy pr., 49*
199034, 197101*, St. Petersburg
perechesova@gmail.com

Soloveva G.A.
ITMO University
Kronverkskiy pr., 49
197101, St. Petersburg
ymatem@rambler.ru

Kalapyshina I.I.
SPbF IZMIRAN, ITMO University*
Universitetskaya emb., 5 B
Kronverkskiy pr., 49*
199034, 197101*, St. Petersburg
weir521@yandex.ru

ABSTRACT

Aramid yarns are widely used for various technical applications. For example, they are used for creation of a magnetosensitive element of the torsion magnetometer, which is a part of the geophysical complex GI-MTS-1 (SPbF IZMIRAN). Such element is made of three microfilaments of aramid yarn. Study of such physical-mechanical characteristics of yarns as tensile, shear and torsion moduli, is necessary for improvement torsion magnetometer. This paper describes the preparatory phase of experimental study of yarn twist angle variation during tensile test. It includes equipment description, in particular, description of the scheme of video capture, and developed computer vision algorithm for determining an aramid torsion slope angles. Proposed scheme of video capture includes high-speed camera and allows precise focusing due to two-coordinate platform and avoiding of optical distortions. A special computer vision algorithm, which based on the Hough transform, was developed. This algorithm was tested on torsion images obtained with digital microscope. The algorithm works with different real yarns regardless of their color and twist type.

Keywords

Segmentation, Hough Transform, image recognition, microfilament aramid yarns, Helically Anisotropic Rod, torsion bar, tensile test, high-speed camera.

1. INTRODUCTION

Aramid yarns are widely used for various technical applications. For example, they are used for creation of a magnetosensitive element of the torsion magnetometer, which is a part of the geophysical complex GI-MTS-1 (SPbF IZMIRAN) [KSPLZ10a]. Such element is made of three microfilaments of aramid yarn, which have diameter of 0.016 mm, using special device [KSPL10a], [PKN14a]. Study of such physical-mechanical characteristics of yarns as tensile, shear and torsion moduli, is necessary for improvement torsion magnetometer.

The magnetosensitive element is a torsion bar, which has helical anisotropy elastic properties and can be modeled as Helically Anisotropic Rod (HAR).

Authors of the current paper propose an experimental analytical method of evaluating physical and

mechanical characteristics of the helical lines of HAR, videlicet, to fix lines slope angles versus axis of HAR during tensile test [Per12a]–[MMSS84a]. It is planned to use a high speed camera for getting yarn images during tensile. These images will be processed by computer vision algorithm to determine required slope angle.

There are several methods for determining a yarn twist angle (a slope angle of a coil). Cybulska [Cyb99a] used an algorithm based on least squares. Ozkaya et al. [OAJ2010a] used spatial and frequency domain methods combination. Kofnov and Shlyakhtenko [KS2014a] proposed to use twodimensional discrete fast Fourier transform. In the current paper authors propose an alternative method based on the Hough transform.

This paper describes the preparatory phase of experimental study of yarn twist angle variation during tensile test. It includes equipment description, in particular, description of the scheme of video capture and description of the developed computer vision algorithm for determining aramid torsion slope angles.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

2. OBJECT OF RESEARCH

An object of the current research is the elastic torsion suspension (Fig. 1), which is used to create magnet indicator rotation axis. This suspension is a «braid» of three strands (one thread in each strand). The envelope of torsion bar cross-section is the ellipse with longer axis of 0.046 mm and shorter axis of 0.033 mm. The total length is 100 mm. The density of braiding is 7 knots/mm.

An optical study of the elastic torsion structure was carried out in the laboratory of electron microscope and physical and technological researches of the Peter the Great St.Petersburg Polytechnic University.

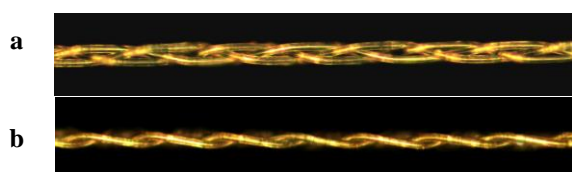


Figure 1. Photos of braided torsion bars: Two orthogonal projection: top view (a), side view (b)

3. EXPERIMENTAL EQUIPMENT

Experimental determination of physical and mechanical characteristics of torsion suspensions will be performed at the department of Mechatronics of ITMO University. Tensile tests will be carried out with the test machine AGS500NX SHIMADZU (SHIMADZU CORPORATION, Japan). Taking into account small size of samples special grips were designed. These grips are the bollard grips with a bollard diameter less than 20 mm. Tensile test data processing are produced with the special software «TRAPEZIUM LITE X» (SHIMADZU CORPORATION, Japan).

To get images during the tensile the camera CamRecord CR450x3 Optronis (Optronis GmbH, German) will be used. This camera allows taking pictures at speeds up to 10 000 frames per second. Taking into account geometric parameters of the object of the research special scheme of video capture was designed (Fig. 2). This scheme includes: 1 – high-speed camera Optronis CR450x3; 2 – controlled positioning platform for focusing during microphotography; 3 – camera bellows; 4 – Nikon F to M42 mount adaptor; 5 – mechanical RMS–M42 adaptor; 6 – microscope lens [PK14a].

This scheme allows precise focusing due to two-coordinate platform and avoiding of optical distortions. Insignificant disadvantages of this scheme are absence of the diaphragm regulation, it is possible to solve with sufficient lighting; unwieldiness of the scheme, but it is not currently important in the laboratory.

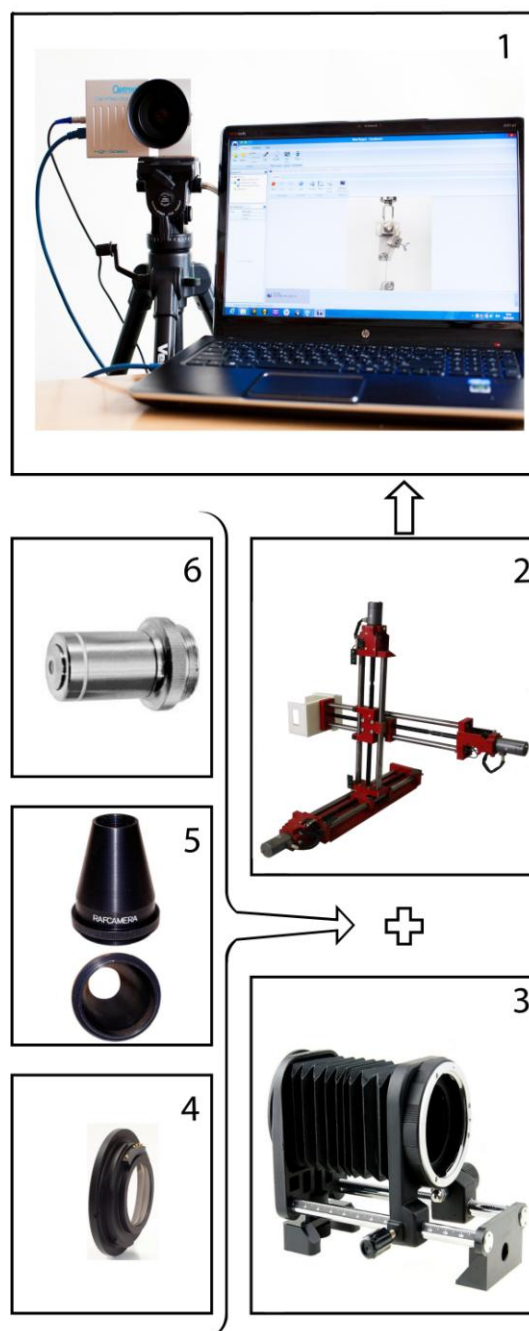


Figure 2. Schematic structure of the video capture organization

4. DEVELOPED ALGORITHM

A special computer vision algorithm was developed to determine aramid torsion slope angles. The main concept of this algorithm is based on image segmentation and application of the Hough transform. Image segmentation allowed identifying torsion edges. The Hough transform was used to detect lines that correspond to coils edges.

The program algorithm has the following steps:

- 1) Image loading and preparation;
- 2) Determination of torsion axis;
- 3) Edge detection;
- 4) Determination of coils vectors;
- 5) Calculation of angle of slope of coils;
- 6) Calculation of average angle
- 7) Output the results.

During the first stage image loading and its conversion to “double” format were conducted. After that the area with better lighting was selected. The purpose of these operations is the simplifying and changing a representation to make future analyze.

According to various researches [BL12a], [SC12a] Canny edge detection algorithm has the best performance especially for images without noise. This method was used in the current work. A threshold was selected using Otsu’s method. Therefore the image with torsion edges was obtained as a result of above operations.

An aramid torsion slope angle is determined as the inverse cosine of inner product of coils and torsion axe vectors. To find torsion axe a distance between torsion edges was analyzed, viz. a center of this distance was determined. Most of center points had the same horizontal coordinate describe torsion axe. To find coils vectors and a twist type the Hough transform [Hou62a] was used. After that torsion slope angle and average angle was determined.

Developed algorithm was tested for torsion images obtained with digital microscope. Figure 3 demonstrates the algorithm execution for images with 25X and 50X magnification. Also this algorithm was tested for images of staple yarn (Fig. 3c).

5. FUTURE WORK

Future work consists of testing developed algorithm on images obtained with high-speed camera during tensile test. After that it is planned to determine such elastic constants as modulus of elasticity (E_1), shear modulus (G_1), and Poisson ratio (ν_1) using following equations of HAR theory:

$$\left. \begin{aligned} \alpha_{11} &= G_1(9\phi_1 + 18\phi_2) - \frac{1}{2 - \nu_1} 9\phi_2 E_1 + E_1 - 3\phi_1 E_1 \\ \alpha_{12} &= -G_1(3\phi_1 + 12\phi_2) + \frac{1}{2 - \nu_1} 6\phi_2 E_1 + \phi_1 E_1 \\ \alpha_{22} &= G_1 \left(\frac{tg^2 \alpha_0}{2} + 8\phi_2 \right) - \frac{1}{2 - \nu_1} 4\phi_2 E_1 \end{aligned} \right\}$$

Where $\phi_1 = 1 - 2ctg^2 \alpha_0 \ln \sec \alpha_0$;

$$\phi_2 = \frac{1}{2} \sin^2 \alpha_0 - 1 + 2ctg^2 \alpha_0 \ln \sec \alpha_0 ;$$

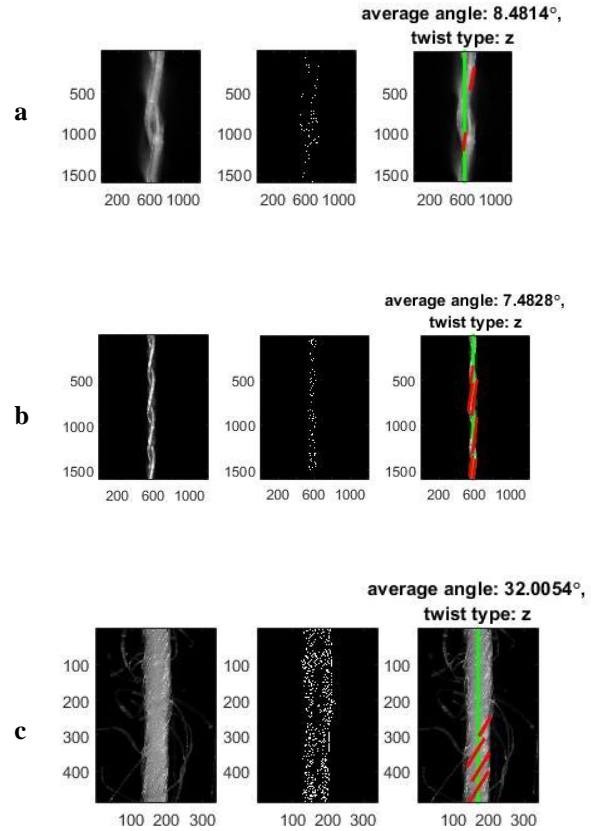


Figure 3. Algorithm execution: for torsion images with 50X(a) and 25X(b) magnification; for images of staple yarn (c)

α_0 – the determined angle of incline of the helical lines.

Within each interval find the minimum value of the functional for given values ν_1 :

$$F(\bar{x}) = [E_1 \cdot \bar{g}_1 + G_1 \cdot \bar{g}_2]^2 \longrightarrow \min$$

Finally, calculate the required data using optimization methods [Per12a]–[MMSS84a].

6. CONCLUSION

This paper describes the preparatory phase of analytical experimental study of aramid yarn elastic constants. It includes equipment description, in particular, description of the scheme of video capture, and developed computer vision algorithm for determining an aramid torsion slope angles. Proposed scheme of video capture allows precise focusing due to two-coordinate platform and avoiding of optical distortions. A special computer vision algorithm was developed to determine an aramid torsion slope angles. This algorithm works with different real yarns regardless of their color and twist type.

7. ACKNOWLEDGMENTS

Authors wish to thank Prof. Viktor. M. Musalimov (chair of Department of Mechatronics, ITMO University) and Prof. Yuri A. Kopytenko (Director of SPbF IZMIRAN)

8. REFERENCES

- [KSPLZ10a] Kopytenko Y.A., Sergushin P.A., Petrishchev M.S., Levanenko V.A., Zaytsev D.B. Torsion magnetic variometer with kevlar-hanger-based sensor // Key Engineering Materials Vol. 437 (2010). – Trans Tech Publications, Switzerland, 2010. P. 621-624. DOI:10.4028/www.scientific.net/KEM.437.621
- [KSPL10a] Kopytenko, Y.A., Sergushin, P.A., Petrishchev, M.S., Levanenko, V.A., Perechesova, A.D. Device for Manufacturing Torsion Bars with Helical Anisotropy UISAT-1 // Key Engineering Materials Vol. 437 (2010). – Trans Tech Publications, Switzerland, 2010. P. 625–628. DOI:10.4028/www.scientific.net/KEM.437.625
- [PKN14a] Perechesova A., Kalapyshina I., Nuzhdin K. Kinematics and dynamics analysis of different mechanisms using SIMMECHANICS / SIMULINK / MATLAB // Proceedings of the European Modeling and Simulation Symposium. - 2014. - P. 128-135
- [Per12a] Perechesova A. Calculation of Elastic Constants of the Torsion Bars with Helical Anisotropy using the methods of Optimization theory // Proceedings of the 23rd International Congress of Theoretical and Applied Mechanics. - Beijing (China), 2012. - SM04-050.
- [MZP12a] Musalimov Victor M., Zamoruev Georgy B., Perechesova Anna D. Calculation of physical and mechanical properties of Helical Anisotropic Rod Screw Elements // Scientific and Technical Journal «Priboorostroenie» 2012. Vol. 55, № 6. P. 24-30. (in Russian).
- [ZMP13a] Zamoruev Georgy B., Musalimov Victor M., Perechesova Anna D. The software "Parabola" for Calculation of physical and mechanical properties of Helical Anisotropic Rod Screw Elements (Federal Service for Intellectual Property RU – Rospatent, Certificate 2013616761 from 18.07.2013).
- [Per12b] Perechesova A.D. Analysis and synthesis of mechanism for manufacturing Torsion Bars of devices: the candidate dissertation of technical sciences: 05.02.18. – SPb, (2012, in Russian) 2012. – 125 p. (in Russian)
- [Mus05a] Musalimov V.M. Mechanics of deformable cables/Monograph. – St. Petersburg: IFMO, 2005, P. 203.
- [MMSS84a] Musalimov V.M., Mokryak S. Ya., Sokhaney B.V., Shiyanov V. D. Determination of elastic characteristics of flexible cables on basis of model of helically anisotropic solid// Translated from Mekhanika Kompozitnykh Materialov, No. 1, Tomsk (Russia), January-February, 1984. P. 136 -141. (in Russian)
- [Cyb99a] Cybulska M. Assessing Yarn Structure with Image Analysis Methods // Textile Research Journal 1999 vol. 69 no. 5 pp. 369-373 DOI: 10.1177/004051759906900511
- [OAJ10a] Ozkaya Y.A., Acar M. and Jackson M.R.,. Yarn twist measurement using digital imaging. // Journal of the Textile Industry, 2010 101 (2), pp. 91- 100.
- [KS14a] Kofnov O.V., Shlyakhtenko P.G. Analysis of Computed Diffraction Pattern Diagram for Measuring Yarn Twist Angle // Textiles and Light Industrial Science and Technology (TLIST) Volume 3, 2014 DOI: 10.14355/tlist.2014.0301.01
- [PK14a] Perechesova A.D., Kalapyshina I.I. Peculiarity of the actual Test Methods for tensile testing of aramid yarns // “Priboor”. - Moscow, Russian, 2014. - Vol. 172. - № 10. - P. 28-33. (in Russian)
- [BL12a] Bin L., Samieiyegeaneh M., Comparison for Image Edge Detection Algorithms, IOSR Journal of Computer Engineering (IOSRJCE), Volume 2, Issue 6, 2012, PP 01-04
- [SC12a] G.T. Shrivakshan, Dr.C. Chandrasekar A Comparison of various Edge Detection Techniques used in Image Processing IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012 269-276
- [Hou62a] Hough, P.V.C., “Method and Means for Recognizing Complex Patterns,” U.S. Patent No. 3069654, 1962.
- [Sol14a] Soloveva G.A. Application Laws’ texture energy measures for detection of surface macrodefects // Privolzhsky Scientific Bulletin № 9 (37) September 2014, pp 22-25

Feature based assessment of forming force behavior in Incremental Sheet Forming

Satwik Priyadarshi

Prashant K. Jain

J. J. Roy

PDPM Indian Institute
of Information Technology, Design
and Manufacturing Jabalpur,
Jabalpur-482005, Madhya Pradesh,
India

PDPM Indian Institute
of Information Technology, Design
and Manufacturing Jabalpur,
Jabalpur-482005, Madhya Pradesh,
India

Centre for Design & Manufacture,
Bhabha Atomic Research Centre,
Trombay-400085, Mumbai,
Maharashtra, India

satwik.p@iiitdmj.ac.in

pkjain@iiitdmj.ac.in

jjroy@barc.gov.in

M.K.Samal

Debanik Roy

Puneet Tandon

Reactor Safety Division
Bhabha Atomic Research Centre,
Trombay-400085, Mumbai,
Maharashtra, India

Board of Research in Nuclear
Sciences
Bhabha Atomic Research Centre,
Trombay-400085, Mumbai,
Maharashtra, India

PDPM Indian Institute
of Information Technology, Design
and Manufacturing Jabalpur,
Jabalpur-482005, Madhya Pradesh,
India

mksamal@barc.gov.in

deroy@barc.gov.in

ptandon@iiitdmj.ac.in

ABSTRACT

Forming forces and their behavior plays an important role in defining the mechanism of any sheet metal forming process. In Incremental sheet forming, due to inherent complexities, study of forming force behavior is a challenging task. In absence of geometry independent techniques, only studies specific to certain axisymmetric and planar symmetric geometries are available. Present work deals with development of a novel methodology for the study of behavior of forming forces in geometries with multiple features by mapping components of forming forces obtained with dynamometer readings to spatial coordinates of tool path coordinates. Techniques such as Nearest Neighbor Search, RANSAC and calculation of l_2 norm are employed for this very purpose.

Keywords

Incremental Sheet Forming, Forming Forces, Point cloud, Nearest Neighbor Search, RANSAC

1. INTRODUCTION

Incremental Sheet Forming (ISF) is a novel sheet forming process which allows user to produce sheet metal components directly from CAD models using a CNC milling center with nil or minimum part specific tooling. Unlike, conventional forming processes which requires costly part specific dies; it allows a single set-up without specialized tooling to produce a variety of geometries in different shapes and sizes. Immense flexibility in implementation and reduced requirement of part specific and process specific tools makes ISF an ideal choice for rapid prototyping of sheet metal components and

production of components in small batches [1].

1.1 Introduction to Incremental Sheet Forming

ISF is performed using a CNC controlled hemispherical tool following a tool path of normally finishing operation which when indented on a metal sheet clamped at periphery produces highly localized plastic deformation. This deformation when dragged along the path of CNC controlled tool progresses incrementally. In due course, this progression of incremental deformation results in formation of desired contour or geometry on the sheet. Most common and simple set-up for ISF is known as Single Point Incremental Forming (SPIF) as shown in figure 1. Here, sheet is deformed at a single point of contact by the tool.

Most significant of the parameters which governs the ISF process are wall angle (α), vertical step size (Δz), forming depth (h), tool diameter (d), etc. as shown in Figure 2.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

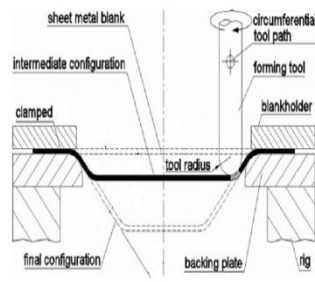


Figure 1. Incremental Sheet Forming [1]

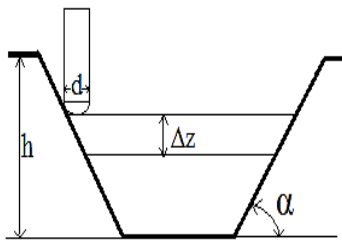


Figure 2. Schematics of ISF Parameters

In ISF, the angle subtended by the wall of the component with horizontal plane is referred as wall angle (α) as shown in Figure 2. Wall angle is formed by the tool while it follows a predefined tool path. Step size (Δz) is the distance moved by the tool, in vertical direction per revolution of the tool for helical tool path. Forming depth (h) means the maximum depth achievable through ISF for a given geometry with given material and process conditions. Tool diameter (d) is the diameter of the forming tool being used. Generally, hemispherical tipped tools are used in ISF. Tools transfer deformation force to sheet at tool-sheet contact to facilitate the desired local deformation.

1.2 Background

In the existing literature, study of forming force is limited to only simple axisymmetric geometries like truncated cone, truncated pyramid etc. [2]. These studies, utilizing the rate of force change obtained from measuring devices such as tool dynamometers and then correlated them with increasing forming depth [3].

Although, parameters like wall angle and step depth are sufficient to describe behavior of forming [4], [5], [6]; complex shaped geometries presents challenges in explaining the force behavior through traditional methods.

A complex geometry may comprise multiple features of different wall angles. Further, in case of adaptive tool path [7]; step depth may differ with change in depth.

In both the aforementioned cases, behavior of forces can't be explained by a single value of wall angle or step depth. Local geometry i.e. geometry at a given depth and plane or tool position must be involved in assessing and forecasting forming forces and their behavior for providing a more generalized view. Sometimes, due to constraints on machine, material, tooling or geometry; forming forces above certain level are undesirable. By incorporating the effect of tool position on forming force the effect of local geometry can be easily approximated. Here by visualizing forming forces in relation with tool positions one can plan a corrective strategy by using forming at elevated temperature, multi pass ISF [8] or any other technique to bring down the intensity of the forming forces to the desired level.

Several geometries have been studied for incremental sheet forming applications, but literature indicating force behavior for geometries involving multiple features are scarce. In the present work, force data from ISF was configured to obtain the spatial variations occurring during the forming process and the force change pattern was observed on a complex shaped geometry.

The present work discusses a methodology which can be used to map and visualize forming forces to tool positions and geometrical features.

2 METHODOLOGY

A computational tool is developed to establish mapping between spatial coordinates and components of forming force. A special purpose NC code, an In-house tool and an open source solution forms the basic constituent of the tool box utilized in the present investigation. The proposed methodology comprises of tool path generation, acquisition of forming force data, mapping between forming force components and coordinates of tool path points. The resulting point cloud of tool paths points with force components as scalar fields can be segmented in sub-clouds of different shapes and features to facilitate feature based assessment of forming forces.

2.1 Test geometry

Since, the proposed methodology is a geometry based technique, preparation of a 3D CAD model of the desired geometry is the first step required. As a case study, test geometry is necessary which comprises of multiple features and can be formed easily. Hence, an Irregular Hexagon was chosen as test geometry because of its unsymmetrical shape and presence of multiple features such as flat, concave and convex surfaces with fillets and sharp corners. A uniform wall angle was kept for all faces. Wall angle of the geometry was taken at 55° to ensure complete formation of geometry. 55° is a safe wall angle as most of the sheet metal alloys allows achievement of considerable forming depth with it.

Dimensions of the geometry are chosen according to the constraints of available machine tools and fixtures.

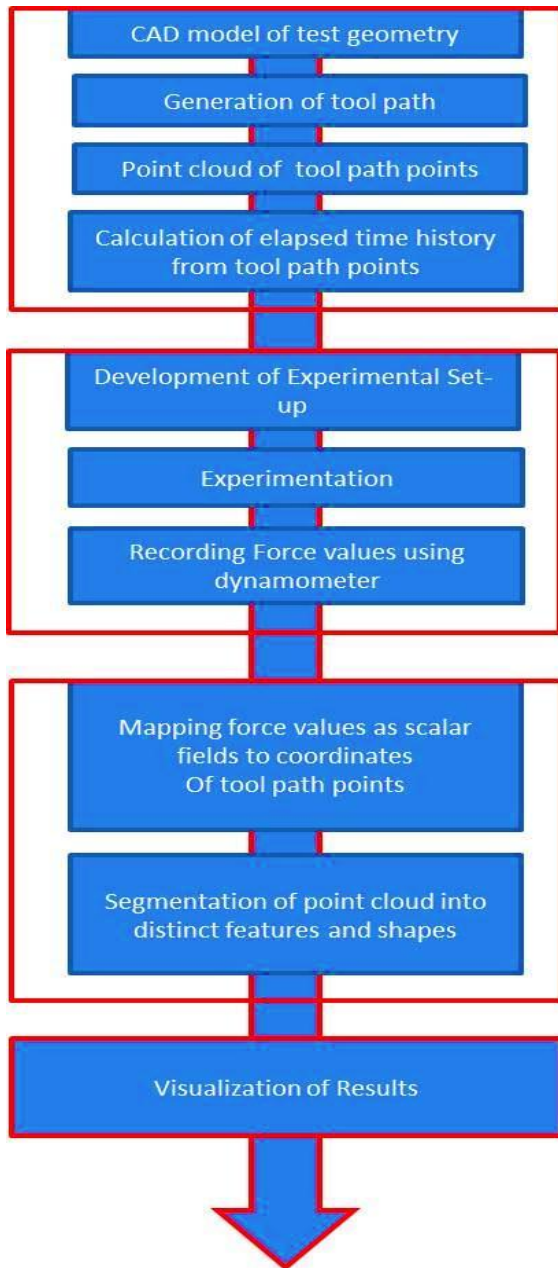


Figure 3.Methodology

Figure 4 shows the test geometry. Various faces of the test geometry are named as shown in Figure 4. Dimensions of the geometry is described in Figure 5 and Table 1.

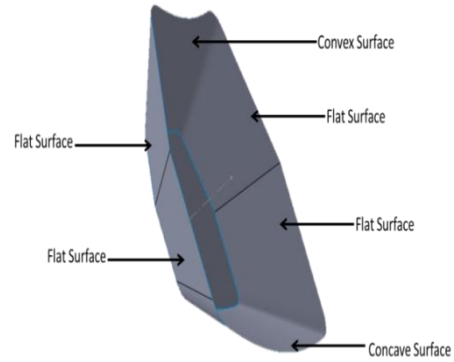


Figure 4.Test Geometry

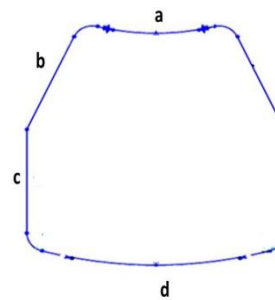


Figure 5.Variou Faces of Test Geometry

Dimensions	
Radius of Convex Surface (a)	10 mm
Length of flat Surface (b)	56.47 mm
Length of flat Surface (c)	57.49 mm
Radius of Concave Surface (d)	10 mm
Depth	30 mm
Wall Angle	55°

Table 1.Dimension of the test geometry

2.2 Generation of Tool path

The test geometry in STL format is sliced into planar contours which are used for producing helical tool path as discussed by Malhotra et al. [9] and Zhu et al[7]. Here, STL model when sliced in Z direction results in generation of “n” number of contours separated by incremental depth of tool path in negative Z direction. Number of slices from a STL model depends on geometry of the component and desired incremental depth. By interpolating a helical

curve on points of the contours; the tool path is generated.

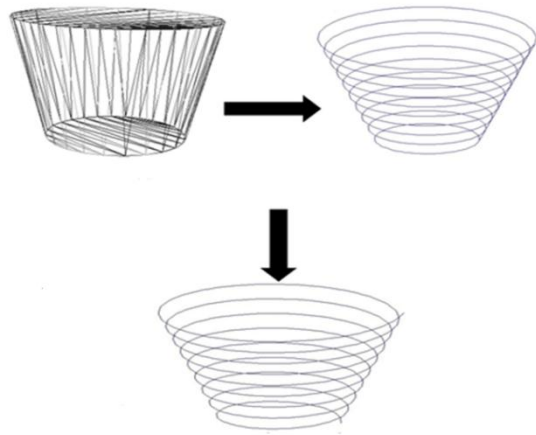


Figure 6. Generation of tool path

The set of points on the tool path can also be considered as a 3 dimensional point cloud. Figure 7 shows a point cloud formed using tool path points of the test geometry. Since, the point cloud being analyzed is synthetically produced; it is completely devoid of outliers and noise. Although, base of the geometry fails to get captured, it is not a matter of concern as the base lies un-deformed and detached from the forming process.

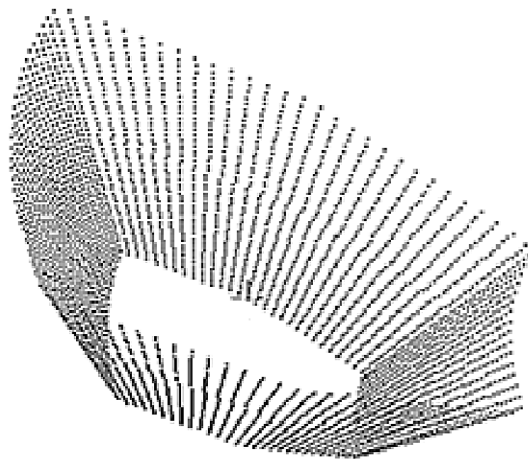


Figure 7. Point cloud formed using tool path points

2.3 Calculation of elapsed time between tool positions

As calculation of elapsed time between tool positions is required to facilitate mapping of tool position to

force data, the NC code was prepared with linear interpolation and circular interpolation was completely avoided. Feed rate was chosen as constant at 2000 mm/min.

Tool path generated only through linear interpolation at constant feed rate allows the tool path to be considered as a polyline comprising of multiple straight line segments. The coordinates of the points on tool path can also be considered as coordinates of tool's position while forming.

By taking L2 norm, distance between the two consecutive tool path points were calculated in 3 dimensional space which leads to calculation of elapsed time while tool moves between the given consecutive points. Therefore, complete time history of the tool movement can be calculated in this manner.

2.4 Experimental set-up & Experimentation

An experimental campaign was undertaken to form geometry through SPISF. An Experimental set-up was developed and experiments were performed to record force data using a tool dynamometer.

All experiments were performed using a 3-axis milling center. Fixtures and clamping mechanism were developed in-house. Fixtures were developed according to the test geometry and clamping plates were designed as per the external periphery of the test case. The design was realized by fabricating mild steel plates of 10 mm thickness. Provision for setting a tool dynamometer in the fixture was also included. Figure 8 shows the experimental set up with tool dynamometer.

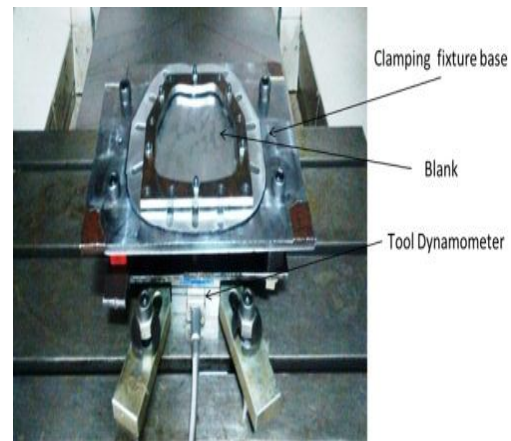


Figure 8. Experimental setup with dynamometer

Experiments were conducted using set-up described in the preceding section. Irregular Hexagons were formed through SPISF. Al 1050 sheets with 0.95 mm thickness were used as blank material. The process parameters selected for the experiment are described in Table 2. The process parameters were chosen

arbitrarily as their effects are not a point of concern for the present investigation.

Process Parameter	Selected Value
Feed Rate	2000 mm/min
Step Depth	0.5 mm
Tool speed	1000 rpm

Table 2.Process Parameters

2.5 Recording force values by a Tool Dynamometer

For measuring numerical values of forming in x, y and z axes a Kistler tool dynamometer has been used. It is a Piezoelectric Quartz based three-component dynamometer used for measuring the three orthogonal components of a force. The dynamometer consists of four three-component force sensors fitted under high preload between a baseplate and a top plate. Each sensor contains three pairs of quartz plates, one sensitive to pressure in the z direction and the other two responding to shear in the x and y directions respectively. The force components are measured practically without displacement. The outputs of the four built-in force sensors are connected inside the dynamometer in a way to allow multicomponent measurements of forces and moments to be performed [10]. Figure 9 shows a Kistler Multicomponent tool dynamometer.



Figure 9.Kistler Multicomponent tool dynamometer

Dynamometer was mounted below a base plate on the outer corner of which four holes were drilled to accommodate the desired fixture. Base plate rested

on the upper surface of the dynamometer while the fixture stood on the base plate by using four pegs of desired length. Set-up of dynamometer also consists of a Personal Computer or workstation and a controller to acquire and record data. Figure 4 shows arrangement of dynamometer in the experimental set-up.

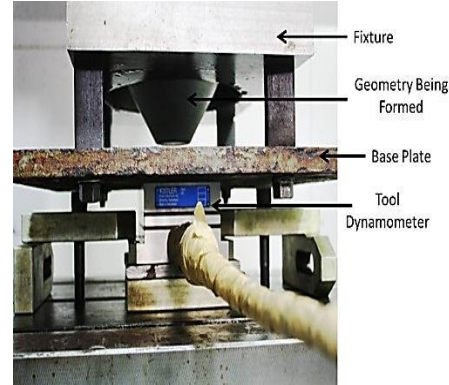


Figure 10.Arrangement of dynamometer in experimental set-up

Force components in x, y and z axes i.e. F_x , F_y and F_z were considered for the present investigation and were recorded in newtons. 10 Samples per second were recorded and saved in a csv file. Forming force vs. forming time plot with interval of 0.1 seconds is shown in Figure 11.

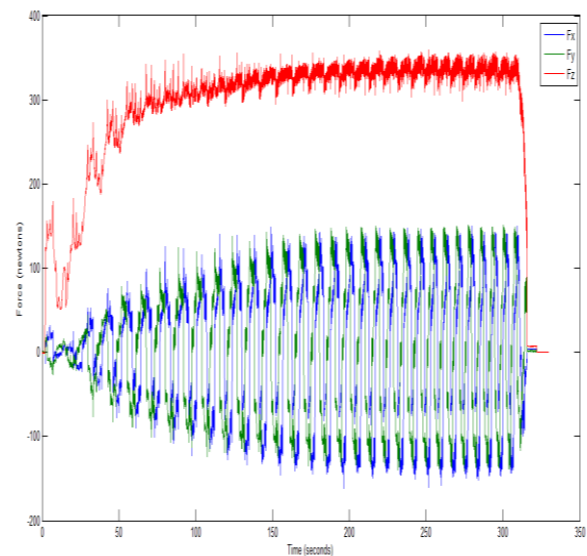


Figure 11.Plot of forming force vs. forming time

2.6 Mapping force values to coordinates of tool path points (tool positions)

A Nearest Neighbor Search (NNS) algorithm [11] was used to obtain the mapping of the time history of values to the spatial coordinates during the forming

operation. Initially, for every value of elapsed time in time history calculated from the tool path; a nearest neighbor was searched in the array of elapsed time values recorded by the dynamometer.

As it can be seen from Figure 12, a linear NNS algorithm was utilized for the very purpose. For any given data point in Time History calculated from the Tool path a nearest neighbor can be searched in the array of Elapsed time calculated using the dynamometer. Linear NNS returns the index value of the identified nearest neighbor. For example, as shown in Figure 12 at 17 second after the initiation of forming, i.e., at time = 17 second in time history calculated from the tool path the Nearest Neighbor in the array of Elapsed time calculated using the dynamometer is 17.82 sec with index value of 3.

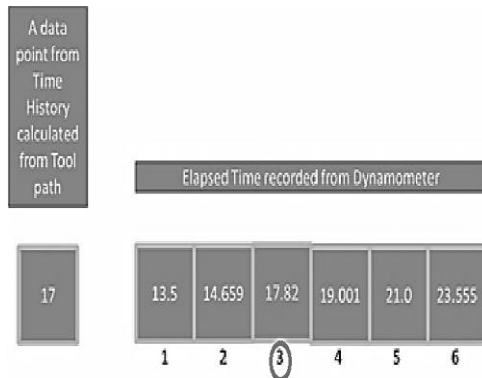


Figure 12. Linear NNS at T= 17 seconds

By calling all the values corresponding to index 3 in the array of Elapsed time calculated using the dynamometer and values corresponding to index of query point (i.e., 17); a new array can be generated with both spatial coordinates and force components. This is shown in Figure 13.

Force components can be considered as 3 dimensional scalar fields corresponding to the respective tool position points in the point cloud.

2.7 Segmentation of features from the point cloud

Collection of tool path points can also be considered as a 3 dimensional point cloud. Figure 7 showed a point cloud formed using tool path points of the test geometry. Force components can be considered as 3 dimensional scalar fields corresponding to the respective tool position points. Since, the point cloud being analyzed is synthetically produced it is completely devoid of outliers.

General point cloud segmentation techniques can easily be applied to such points to generate sub-clouds representing individual features. Sub-clouds can be generated from the original point cloud of tool path coordinates through implementation of a

commonly available technique known as Random Sample Consensus (RANSAC) Shape Detection.

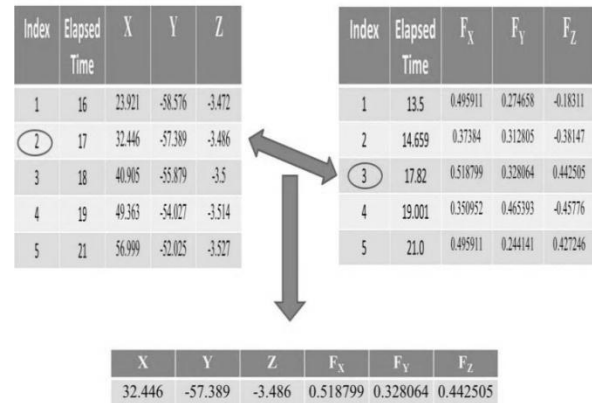


Figure 13. Mapped data

In the present work a variant of RANSAC developed by Schnabel et al. [12] is used. It was selected over other techniques such as Hough Transform [12] and MLESAC [12] as it is a simple and general technique with potential for wide applicability. Further, complex shaped geometries can easily be processed.

This operation can be performed using Cloud Compare (CC), an open source solution for point cloud analysis [13]. Cloud Compare uses the same implementation as introduced by Schnabel et al. and just provides a simple user interface on top of it.

Parameters selected for the segmentation operation are given as shown in Table 3. For the given test geometry, minimum number of support points per primitive was manually selected at 500. Even though minimum points required per primitive are 3, 4, 3 and 4 respectively for Plane, Sphere, Cylinder and Cone; a much larger set of points was used as the features inherent to test geometry were large themselves with high density of points. Furthermore, as it is visually evident that the test geometry has lack of toroid features; Plane, Sphere, Cylinder and Cone were the only primitives chosen. Number of maximum iterations is automatically computed by the Cloud Compare software.

Using data obtained from experimentation described in the previous section, the proposed methodology was implemented. Sub-clouds for various features were generated. The original geometry was divided into 5 different types of features. Figure 14 depicts the sub-clouds generated from the original point cloud using RANSAC.

Now, individual points with spatial co-ordinates of tool position and scalar values of forming force components corresponding to position are ready to be analyzed for drawing required inferences.

Parameters for segmentation	
Primitives	Plane, Sphere, Cylinder, Cone
Minimum support points per primitive	500

Table 3. Parameters for segmentation

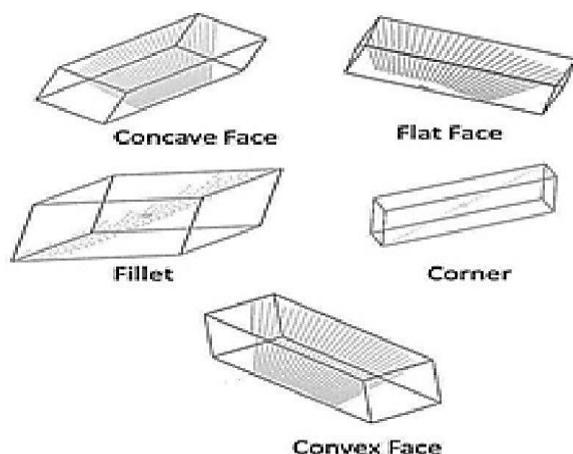


Figure 14. depicts the sub-clouds generated from the original point cloud

2. 3 ASSESSMENT OF FORCE BEHAVIOR FEATURE-BY-FEATURE

After the implementation of the proposed methodology, sub-clouds for various features were assessed for force behavior. After segmentation, Force Vs. time graphs for each feature was plotted separately, and the trends in all the three components of force values were observed. Histograms were used for quantitative description of the spatial force behavior.

3.1 Force vs. time graphs

Force vs. time graphs are shown in Figure 15 and Figure 16. Figure 15 shows graphs of force behavior at the faces. A common trend in force behavior was observed. Figure 16 shows graphs of force behavior at the corners and fillets. Graphs for corners and fillets too show similar trends.

It is the number of spatial points in features which results in differences visible in graphs; as the number of sample points varies across the features. Otherwise, force behavior is almost same across the features due to all features being at a constant wall angle.

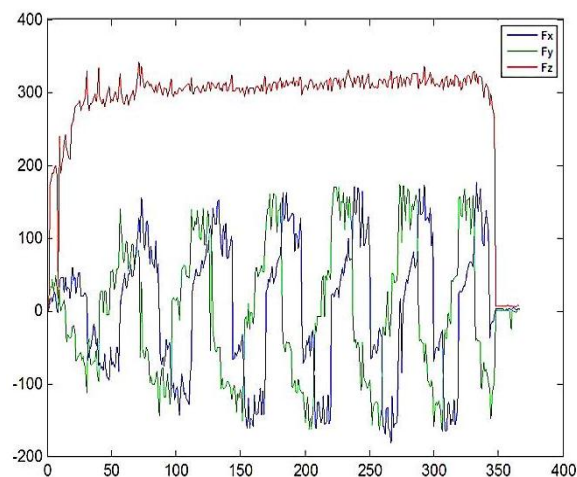


Figure 15 (a). Plot of forming force vs. forming time for flat surface

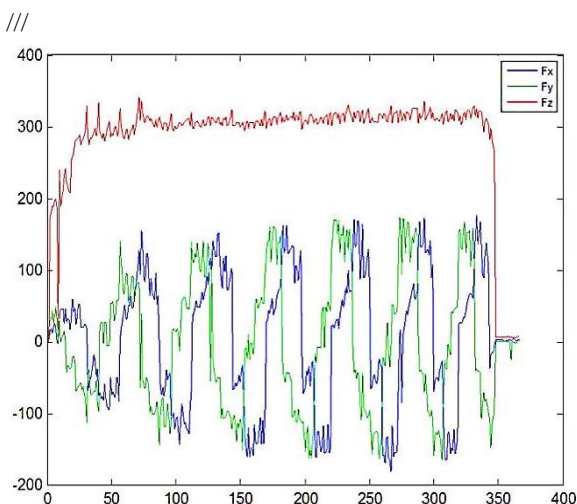


Figure 15 (b). Plot of forming force vs. forming time for convex surface

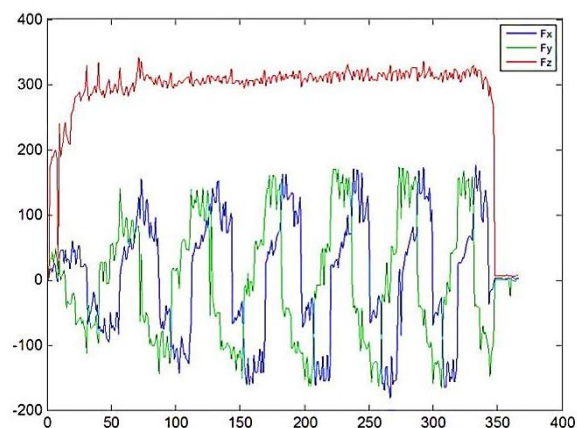


Figure 15 (c). Plot of forming force vs. forming time for concave surface

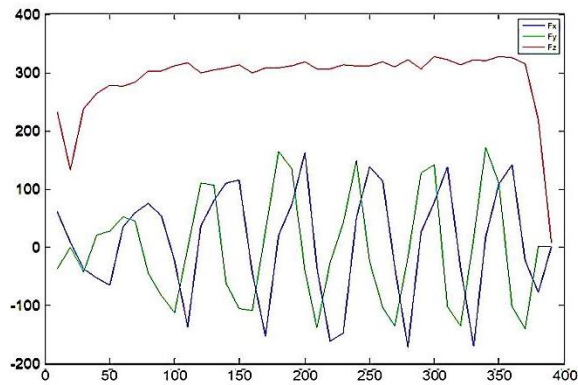


Figure 16 (a).Plot of forming force vs. forming time for a corner

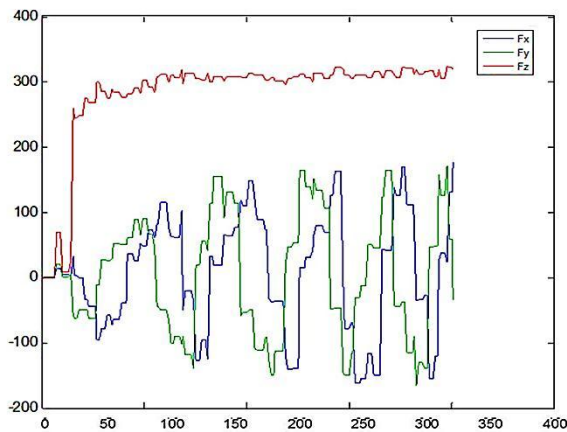


Figure 16 (b).Plot of forming force vs. forming time for a fillet

3.2 Quantitative analysis of the forces

Apart from the graphs shown in the preceding section a quantitative description of the force behavior is also required.

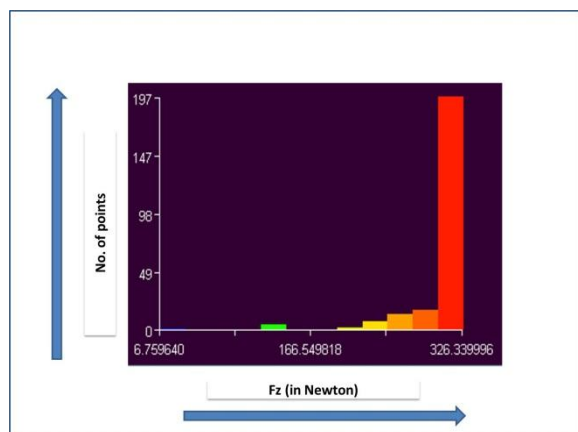


Figure 17 Histogram of Fz from point cloud of a fillet with values of Fz mapped as a scalar field

After mapping force components with spatial coordinates, quantitative variation in forming forces

is shown with the help of a Histogram as Shown in Figure 17.

4 RESULT & DISCUSSION

As presented in the previous sections, the methodology for spatial assessment of the forming forces in Incremental Sheet Forming was successfully developed and tested on the selected test geometry.

Force vs. time graphs were plotted for individual features by segmenting the point clouds mapped with force components as scalar fields. Histograms for individual features were plotted to gauge the spatial variation in forming forces in a quantitative manner.

Based on the visualizations and spatial assessment of forming forces it is now possible to take corrective actions to bring forming forces to the desired levels.

In future, present work can be extended by integrating the spatial assessment with corrective actions as a single solution.

5 ACKNOWLEDGEMENTS

We gratefully acknowledge the Board of Research in Nuclear Sciences (BRNS), India for their support and funding for this research work.

6 REFERENCES

- [1] Jeswiet, J., Micari, F., Hirt, G., Bramley, A., Duflou, J., Allwood, J. Asymmetric single point incremental forming of sheet metal. CIRP Annals - Manufacturing Technology 54, 88–114, 2005
- [2] Petek, Aleš, Karl Kuzman, and J. Kopac. "Deformations and forces analysis of single point incremental sheet metal forming." Archives of Materials science and Engineering 35.2 (2009): 35-42.
- [3] Li, Yanle, et al. "Efficient force prediction for incremental sheet forming and experimental validation." The International Journal of Advanced Manufacturing Technology 73.1-4 (2014): 571-587.
- [4] Duflou, J. and Tunçkol, Y. and Szekeres, A. and Vanherck, P., "Experimental study on force measurements for single point incremental forming", Journal of Materials Processing Technology, 2007.
- [5] Petek, A. and Kuzman, K. and Kopac, J., "Deformations and forces analysis of single point incremental sheet metal forming", Archives of Materials science and Engineering, 2009
- [6] Henrard, C. and Bouffioux, C. and Eyckens, P. and Sol, H. and Duflou, J.R. and Van Houtte, P. and Van Bael, A. and Duchêne, L. and Habraken, A.M.,

"Forming forces in single point incremental forming: prediction by finite element simulations, validation and sensitivity", Computational Mechanics, 2011

[7]Hu Zhu, Zhijun Liu, Jianhui Fu. Spiral tool-path generation with constant scallop height for sheet metal CNC incremental forming. Int J Adv Manuf Technol 2011, 54:911–919.

[8] Kim, T.J., Yang, D.Y. Improvement of formability for the incremental sheet metal forming process. International Journal of Mechanical Sciences 2000 42, 1271–1286.

[9] Rajiv Malhotra, N. V. Reddy, Jian Cao. Automatic 3D Spiral Toolpath Generation for Single Point Incremental Forming. Journal of Manufacturing Science and Engineering, 2010, Vol. 132 / 061003-1.

[10] Kistler Multicomponent tool dynamometer Manual

[11] Nene, Sameer A., and Shree K. Nayar. "A simple algorithm for nearest neighbor search in high dimensions." Pattern Analysis and Machine Intelligence, IEEE Transactions on 19.9 (1997): 989-1003.

[12] Schnabel, Ruwen, Roland Wahl, and Reinhard Klein. "Efficient RANSAC for Point Cloud Shape Detection." Computer graphics forum. Vol. 26.No.2. Blackwell Publishing Ltd, 2007.

[13] Girardeau-Montaut, D. Cloudcompare, a 3D point cloud and mesh processing free software. Technical Report, EDF Research and Development, Telecom ParisTech. <http://www.danielgm.net/cc/> (accessed 19.03. 2015).

Building 3D Object Representation Using SURF Local Features

Marek Jaszuk
University of Information
Technology and
Management
ul. Sucharskiego 2
Poland, 35-225, Rzeszów
marek.jaszuk@gmail.com

Grażyna Szostek
University of Information
Technology and
Management
ul. Sucharskiego 2
Poland, 35-225, Rzeszów
grazyna.szostek@gmail.com

Janusz A. Starzyk
School of Electrical
Engineering and
Computer Science
Ohio University
USA, OH 45701, Athens
starzykj@ohio.edu

ABSTRACT

The paper discusses an approach to create 3D representation of physical objects. The aim is creating a visual representation of an object, which allows for robust recognition, irrespectively of the distance and the direction of observation. The approach uses a set of rotational views of an object, which are transformed into a set of keypoints using the SURF visual feature detector. The key points are then collected to build a 3D model of the object. Such representation allows both for recognizing the objects based on local characteristics, and distinguishing different global geometry transformations that are needed to recognize the object in its 3D environment.

Keywords

visual reconstruction, visual memory, SURF descriptors

1 INTRODUCTION

Creating virtual models of physical objects using photographs, video records, or 3D scanning became very popular in computer vision. There is a growing number of techniques and devices serving this purpose. Most of the work done in this field focuses on reconstructing global geometry of objects. Such approaches usually lead to obtaining a set of characteristic points located on the surface of a reconstructed object. The points are then transformed into a polygon mesh, to use in a variety of applications, like 3D visualization. Although the precise geometry is highly desired and useful, such representation is not convenient for recognizing an object, because many objects may change their geometry, or part of the geometry might be invisible.

Object recognition typically uses 2D feature detection. Depending on the method used, the visual features can either be edges, corners, regions of interest, interesting points or ridges detected within an image. Detection of features starts from the pixel level, and transforms a local image contents into a set of low-level parametric objects. The collection of visual features

related to a particular object should allow for identifying the object within an image in which the object is visible. Local features are robust with respect to occlusions and changes in global geometry of the observed object. There is a number of approaches developed so far, designed for building recognition systems based on local features like [Low01, Rot04, Jun05]. However, the weakness of systems based exclusively on local features is their inability to distinguish changes in global geometry.

Various problems are faced, when one tries to build visual 3D recognition system. Such system should be able to recognize an object irrespectively of its translation, rotation, and scale in addition to changes in lighting conditions, shading, partial occlusions, and local deformations. The scale invariance is important, because it allows for recognizing objects irrespectively of the distance. An important achievement to overcome these difficulties, was introduction of the Speeded Up Robust Features (SURF) algorithm [Bay08]. This method first identifies a set of key points within an image, together with vectors of descriptors for each of the points. In this way, the local image contents around each of the key points is characterized using either 64 or 128 dimensional floating point vectors.

The problem of object recognition is, however, more complex when we consider 3D objects. In this case two possible approaches can be distinguished. Either the objects will still be described by 2D features, but transformed into 3D model, or we can develop 3D feature descriptors. The first approach is relatively simple to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

apply, because we can use the same well known feature descriptors that are used for recognizing objects in flat images. Moreover this approach requires nothing more than ordinary camera images. In the second approach the 3D object representation is assumed. This is more complicated, because we have to make 3D scans of the objects that we want to work with. There is a number of technologies available on the market designed for 3D scanning. They use laser rays, structured light, multiple camera views, or time-of-flight cameras [Pfe15, Tanb08] to obtain 3D geometry of an object. The literature is dominated with approaches describing such geometry in terms of global shape features, but there are also approaches to treat the problem with the local features, which are more appropriate for recognition purposes. An example of such an approach is the 3D extension of the SURF algorithm [Kno10].

The goal of this paper is to demonstrate an approach based on using SURF local feature descriptors, and sequences rotational views of objects to reconstruct the objects in 3D space. Using rotational views allows for recognizing the object irrespectively of the direction of observation. The SURF descriptors are scale invariant, which allows for recognizing the objects irrespectively of the distance. However, the scale invariance demonstrated by SURF is not sufficient for wide range of distances. Thus we have to extend the object representation to contain interest points identified in sequences of rotational views recorded from different distances. This allows to capture more effectively the features, which are not visible from large distance, or might be omitted, when seen from close distance. The same refers to the height from which the object is observed. To make the object representation complete we add additional sequences of views recorded from different vertical locations with respect to the object.

The described object representation is desired in many applications. An example could be a mobile robot memory, which would allow for recognizing objects and navigating 3D environment irrespectively of the temporary position of the robot. The additional advantage of placing the key points in 3D space is the possibility of identifying changes in geometry of objects. In this way objects can be analyzed in two stages. First is identification of the object as a loose collection of characteristic points found within an image, and then verification of the respective distances between the points. In this way we can find out, if there are any changes in the spatial configuration of the object, or identify missing or invisible parts of the object.

Despite building 3D representation of objects, we are still using 2D feature descriptors. This approach is computationally more efficient, than a truly 3D representation of features, like the one presented in [Kno10]. This is of particular importance, when we consider real

time image analysis in a real time system. Considering, that model is assumed to be used for processing individual camera images, we do not need a precise object geometry mapping.

The paper is organized as follows. In Sec. 2 we describe the turntable approach to reconstructing objects from a sequence of images. In Sec. 3 the way we use SURF local features to characterize objects is discussed. Sec. 4 describes how the points from 2D images are converted into 3D coordinates. Next, in Sec. 5 we describe the way of extending the basic approach in order to efficiently capture the features from different distances, and the height of the observation point. Finally in Sec. 6 we describe selected experimental results.

2 THE TURNTABLE APPROACH TO 3D OBJECT RECONSTRUCTION

Creating an object representation in our approach is based on registering sequences of rotational views of an object. We assume, that the object rotates on a turning table around a vertical axis within the field of view of a fixed camera. This solution has already been applied for creating 3D object reconstruction [Fre04, Zha09]. The method works as a camera based 3D scanners, which leads to creating a 3D mesh representing the object's surface. While useful in many applications, a mesh is of little use for recognition purposes. Our work is aimed at creating an object representation, which could be considered a visual object memory. We need to create this representation in a way, which allows for easy recognition of the object, irrespectively of the direction, and the distance of observation. Such a representation is useful in many applications, like a mobile robot memory system, which allows for object recognition, and environment navigation.

The turntable approach presented here originates from the method presented in [Fre04]. However, the mentioned work was focused on reconstructing global geometry of an object, while the goal of our work is building the recognition system. The method deals with two 3D coordinate systems - one associated with the object and the other with the camera (Fig. 1). Moreover, the environment view is registered on a 2D image plane, with its own 2D coordinates. What we know is the position of particular scene elements $\mathbf{V}^I = (\mathbf{u}, \mathbf{v}, 1)$ in the 2D image plane (for convenience expressed in the homogeneous coordinates). This position results from location of the scene element with coordinates $\mathbf{V}^O = (\mathbf{x}^O, \mathbf{y}^O, \mathbf{z}^O, 1)$ in the 3D coordinate system associated with the object, as well as from the projection matrix \mathbf{P} , which transforms position of each point from the object coordinates \mathbf{V}^O into the image coordinates \mathbf{V}^I . This matrix results from the camera position and orientation with respect to the object coordinate system,

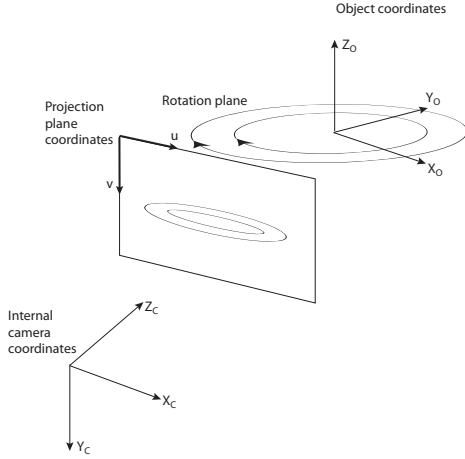


Figure 1: The coordinate systems considered in the turntable approach

and the camera focal length. We do not know the \mathbf{P} matrix *a priori*, but it can be found using a calibration procedure, like the one described in [Fre02], or any other suitable calibration method. So in further considerations we assume that the projection matrix is known.

The next point is the formula used to obtain the screen coordinates from the original object coordinates. In other words, this is the transformation from the 3D \mathbf{V}^O coordinates into the 2D image coordinates:

$$\lambda \mathbf{V}^I = \mathbf{P} \cdot \mathbf{V}^O, \quad (1)$$

where λ is a positive scalar value. Explicitly the above formula can be written as follows:

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} P_{1,1} & P_{1,2} & P_{1,3} & P_{1,4} \\ P_{2,1} & P_{2,2} & P_{2,3} & P_{2,4} \\ P_{3,1} & P_{3,2} & P_{3,3} & P_{3,4} \end{bmatrix} \cdot \begin{bmatrix} X^O \\ Y^O \\ Z^O \\ 1 \end{bmatrix}. \quad (2)$$

From the above we determine the λ constant:

$$\lambda = P_{3,1}X^O + P_{3,2}Y^O + P_{3,3}Z^O + P_{3,4}, \quad (3)$$

and the u, v image coordinates:

$$u = \frac{P_{1,1}X^O + P_{1,2}Y^O + P_{1,3}Z^O + P_{1,4}}{P_{3,1}X^O + P_{3,2}Y^O + P_{3,3}Z^O + P_{3,4}}, \quad (4)$$

$$v = \frac{P_{2,1}X^O + P_{2,2}Y^O + P_{2,3}Z^O + P_{2,4}}{P_{3,1}X^O + P_{3,2}Y^O + P_{3,3}Z^O + P_{3,4}}. \quad (5)$$

Now we have to note, that the Z^O axis is the rotation axis of the object. Thus the vertical position of every element of the object is constant, and will be denoted as h (height from the rotation plane 0 level). The coordinates that vary are X^O and Y^O . The trajectory of every element of the rotating object is a horizontal circle with

center in the $0, 0, h$ point. The equation of the circle can be written as follows:

$$O(X^O, Y^O, Z^O = h) = (X^O)^2 + (Y^O)^2 - R^2 = 0, \quad (6)$$

where R is the radius of the circle obtained separately for every interest point. Now we transform eq. (4) in order to extract X^O :

$$X^O = \frac{(P_{1,2} - P_{3,2}u)Y^O + (P_{1,3} - P_{3,3}u)h + (P_{1,4} - P_{3,4}u)}{P_{3,1}u - P_{1,1}}. \quad (7)$$

After rearrangement we get:

$$X^O = \frac{(P_{1,2} - P_{3,2}u)}{P_{3,1}u - P_{1,1}}Y^O + \frac{(P_{1,3} - P_{3,3}u)}{P_{3,1}u - P_{1,1}}h + \frac{(P_{1,4} - P_{3,4}u)}{P_{3,1}u - P_{1,1}}, \quad (8)$$

or in simpler form:

$$X^O = A_1^I(u, v)Y^O + B_1^I(u, v)h + C_1^I(u, v). \quad (9)$$

This after substituting to eq. (5) gives:

$$v = \frac{P_{2,1}(A_1^I Y^O + B_1^I h + C_1^I) + P_{2,2}Y^O + P_{2,3}Z^O + P_{2,4}}{P_{3,1}(A_1^I Y^O + B_1^I h + C_1^I) + P_{3,2}Y^O + P_{3,3}Z^O + P_{3,4}}. \quad (10)$$

The formula can be rearranged to get Y^O in the following way:

$$Y^O = \frac{(P_{2,1} - P_{3,1}v)B_1^I + P_{2,3} - P_{3,3}v}{(P_{3,1}v - P_{2,1})A_1^I + P_{3,2}v - P_{2,2}}h + \frac{(P_{2,1} - P_{3,1}v)C_1^I + P_{2,4} - P_{3,4}v}{(P_{3,1}v - P_{2,1})A_1^I + P_{3,2}v - P_{2,2}}, \quad (11)$$

or shorter:

$$Y^O = C_1(u, v)h + D_1(u, v). \quad (12)$$

The above can be substituted to eq. (9):

$$X^O = A_1^I(u, v)C_1(u, v)h + A_1^I(u, v)D_1(u, v) + B_1^I(u, v)h + C_1^I(u, v), \quad (13)$$

or shorter:

$$X^O = A_1(u, v)h + B_1(u, v). \quad (14)$$

X^O and Y^O written in this way can be substituted to the equation for the circle in 3D object coordinates (eq. (6)). As a result we get the equation specified by the radius and the height:

$$O(u, v) = (A_1(u, v)h + B_1(u, v))^2 + (C_1(u, v)h + D_1(u, v))^2 - R^2 = 0. \quad (15)$$

After reorganizing and grouping with respect to h we get the circle equation in the following form:

$$O(u, v) = A(u, v)h^2 + B(u, v)h + C(u, v) - R^2 = 0. \quad (16)$$

In the above equation we have two parameters h and R , which are unknown. To find them we have to fit the circle equation to at least 2 image points (u, v) representing the same interest point of the object seen at different angles of rotation. To make the fitting more reliable, it is desirable to collect larger number of the same point views, by following the point in images representing subsequent rotations.

3 CHARACTERIZING OBJECTS WITH SURF DESCRIPTORS

To characterize an object we use the SURF detector [Bay08]. It identifies a set of key points within an image, and then computes vectors of descriptors for each of the key points. The vectors are floating point number vectors of size 64 or 128 depending on the algorithm setting. Matching elements in two images is based on matching vectors of descriptors of particular key points. To build the object representation we need to extract the key points belonging to the object from all the key points identified in an image. Assuming that the object of interest is the only moving element in the scene, and the camera is static, we can easily distinguish the points belonging to the object from these of the background. It is enough to match the points between two images presenting different rotations of the object, then identify the points that changed their positions, and remove all the static background points. Fig. 2) shows an example of artificially rendered scene using the NeoAxis 3D rendering engine [Neo15]. The Girl object¹ is rotated by 10° between the left and right image. The operation repeated for full set of rotational views, delivers the collection of points representing the 360° panorama of the object. This is the basis for building the 3D object representation.

The number of matched points between images depends on the angle of the object rotation. It cannot be too large, because the number of matched points drops rapidly with the increase in the rotation angle value. It cannot be too small either, for efficiency reasons. In our experiments, we assumed, the fixed value of 10° angle of object rotation between subsequent images. It delivers sufficiently large number of matched points, and generates a complete rotational view of the object in 36 images.

All the subsequent images deliver new points, which are grouped into sequences matched between a number of subsequent views. We treat this sequence as the same key point observed in the subsequent images. Most of the sequences are limited to just two instances of a key point observed in two subsequent images. There is also a large number of longer sequences consisting of points

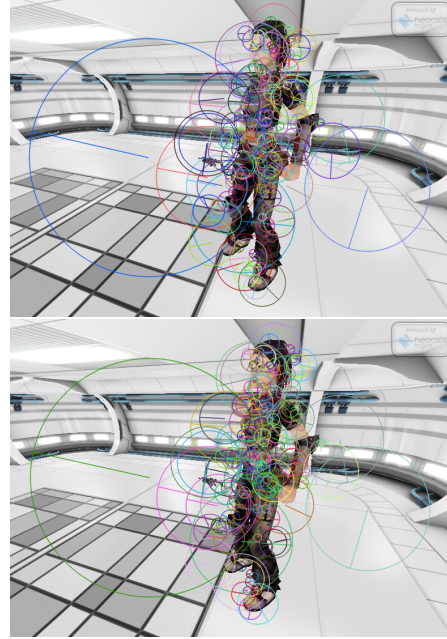


Figure 2: A pair of images with matched SURF key points between different object rotations

identified in up to 7 subsequent images. This corresponds to 60° rotation angle between the first, and the last point in the sequence. The larger number of points in a sequence, the more precisely the circular trajectory of the key point can be determined.

4 PLACING CHARACTERISTIC POINTS IN 3D MODEL

The sequences of characteristic points will be used to locate the point in the 3D model. To do that we start from fitting every sequence to the circle equation (16). The center of the circle is known, because this is the $(0,0)$ point in the (X^O, Y^O) coordinates. To position the key point in the 3D space, we need to find the h and R parameters. h is equivalent to the Z^O coordinate, and R is the radius of rotation of the key point around the Z^O axis. Finding a circle, that best matches N instances (in N subsequent images) of a particular key point is a minimization problem, with objective function written in the following form:

$$J(R^2, h) = \sum_{i=1}^N [A(u_i, v_i)h^2 + B(u_i, v_i)h + C(u_i, v_i) - R^2]^2 \rightarrow 0. \quad (17)$$

The minimum of that function is obtained by calculating its partial derivatives with respect to R^2 and h and finding their zeros. To reduce the size of expressions we will use a simplified notation for sums, i.e.

¹ One of the objects available by default in the NeoAxis Game Engine.

$\sum A = \sum_{i=1}^N [A(u_i, v_i)]$, etc. The equation for the partial derivative of J wrt R^2 is the following:

$$\frac{\partial J(R^2, h)}{\partial R^2} = 2 \sum [Ah^2 + Bh + C - R^2] \quad (18)$$

To simplify further transformations we can omit the constant before the sum, because it is irrelevant for the task of finding minimum of J . The resulting formula can be written as:

$$\frac{\partial J(R^2, h)}{\partial R^2} = -NR^2 + h^2 \sum A + h \sum B + \sum C = 0 \quad (19)$$

Similarly we get the derivative of J wrt h :

$$\begin{aligned} \frac{\partial J(R^2, h)}{\partial h} &= 2 \sum [(Ah^2 + Bh + C - R^2)(2Ah + B)] = \\ &= 2 \sum (2A^2h^3 + 3ABh^2 + B^2h + 2ACH + BC - \\ &= R^2 2Ah - R^2 B) = 2(2h^3 \sum A^2 + 3h^2 \sum AB + \\ &+ h \sum B^2 + 2h \sum AC + \sum BC - R^2(2h \sum A + \sum B)) \end{aligned} \quad (20)$$

After omitting the constant before the bracket, and small reorganization, we get

$$\begin{aligned} \frac{\partial J(R^2, h)}{\partial h} &= 2h^3 \sum A^2 + 3h^2 \sum AB + \\ &+ h(\sum B^2 + 2 \sum AC) + \sum BC - R^2(\sum B + 2h \sum A) = 0. \end{aligned} \quad (21)$$

We can easily isolate R^2 from eq. (19):

$$R^2 = 1/N(h^2 \sum A + h \sum B + \sum C) \quad (22)$$

After injecting it to eq. (21) and reorganizing the equation, the following expression of degree 3 with respect to h is obtained:

$$ah^3 + bh^2 + ch + d = 0, \quad (23)$$

where

$$\begin{cases} a &= 2N \sum A^2 - 2(\sum A)^2 \\ b &= 3(N \sum AB - \sum A \cdot \sum B) \\ c &= N(\sum B^2 + 2 \sum AC) - 2 \sum A \cdot \sum C - (\sum B)^2 \\ d &= N \sum BC - \sum B \cdot \sum C \end{cases} \quad (24)$$

When $4p^3 + 27q^2 < 0$ with $p = c/a - b^2/(3a^2)$ and $q = b/27a(2b^2/a^2 - 9c/a) + d/a$ the eq. (23) gives three real solutions. To find them the Cardano formula can be used. Then the value of the radius R can be found using eq. (19), for each of the found values of h .

Of course, not every solution, which minimizes the sum (17) is the desired solution. We have no clue, other than verifying the correctness of the solution. Thus we have to take the first of the found values for h . Compute for

it the value of R , and the rotation angle β , reconstruct the X^O and Y^O coordinates, and check if the computed 3D coordinates of the key point, multiplied by the transformation matrix (2), reproduce the u, v image coordinates. If yes, the key point has been placed correctly. If not, we have to repeat the checking procedure for the remaining solutions.

We already know how to compute the radius and height of the 3D circle. Reconstruction of the temporary location of the considered key point can be obtained in the following way:

$$\begin{cases} X &= R \cos(\beta \pm i\Delta\theta) \\ Y &= R \sin(\beta \pm i\Delta\theta) \\ Z &= h \end{cases} \quad (25)$$

where $\Delta\theta$ is the object rotation step between subsequent images, i is the image number in the sequence of all images, the $+$ sign is for counterclockwise rotations, the $-$ sign is for clockwise rotations, β is the unknown angle of rotation of the key point in the object coordinates. We assume that the object's coordinate system is stationary (not rotating with the object). As a consequence the key point changes its temporary angle of rotation along with rotations of the object, and β is the rotation angle in reference to the initial rotation of the object ($i = 0$).

The method to identify β is based on comparing the u, v key point coordinates from the image with coordinates obtained using eq. (25) multiplied by the transformation matrix P . The estimated image coordinates computed on the basis of reconstructed 3D coordinates (eqs. (4) and (5)):

$$\begin{cases} \hat{u}_i = \frac{P_{1,1}X^O + P_{1,2}Y^O + P_{1,3}Z^O + P_{1,4}}{P_{3,1}X^O + P_{3,2}Y^O + P_{3,3}Z^O + P_{3,4}} \\ \hat{v}_i = \frac{P_{2,1}X^O + P_{2,2}Y^O + P_{2,3}Z^O + P_{2,4}}{P_{3,1}X^O + P_{3,2}Y^O + P_{3,3}Z^O + P_{3,4}} \end{cases} \quad (26)$$

should fit the original image coordinates u_i, v_i .

To solve eqs. (26) we start from computing the X^O and Y^O coordinates by using the previously derived formulas (see eq. (14) and (12)). Then we divide the second by the first equation from eqs. (25), which gives:

$$\frac{Y^O}{X^O} = \tan(\beta + i\Delta\theta) \quad (27)$$

from the above, we get the formula for the β angle:

$$\beta = \begin{cases} (\arctan(\frac{Y}{X}) - i\Delta\theta) \bmod \pi - \text{counter-} \\ \text{clockwise rotations} \\ (\arctan(\frac{Y}{X}) + i\Delta\theta) \bmod \pi - \text{clockwise} \\ \text{rotations} \end{cases} \quad (28)$$

This formula is, however, limited to the range of $(0, \pi)$ (if we add π for negative angles). For full reconstruction the whole range of 2π has to be considered. If

β was from the range $(0, \pi)$ then the angle given by eq. (28) is correct. Otherwise (the angle from the range $(\pi, 2\pi)$) we need to add additional π . But the problem is that we do not know what the actual range of the angle is. We can find this by verifying the results produced by eqs. (26), with X^O and Y^O computed using (25) and the identified angle. If the results reproduce original u_i, v_i , then the angle is correct, otherwise we should add π to the angle. In this way we get the β angle in the range $[0, 2\pi]$, which allows for complete reconstruction of the position of the key point.

An additional issue, that has to be mentioned here, is how does a single key point is recorded in the object model. When the object is rotated, the key points are matched, between subsequent images, but this does not mean that their respective vectors of descriptors are identical. In fact, the vectors gradually change, along with the changing look of the key point neighborhood. In consequence, the memorized key point, is represented by a sequence of vectors of descriptors. And this sequence has to be memorized in order to increase the efficiency of object recognition. Choosing just one of the vectors, would reduce the ability of recognizing the point, when the angle of object rotation would significantly differ, from the angle for which the vector of descriptors has been recorded.

5 EXTENDING THE MODEL FOR LARGER ROBUSTNESS WITH RESPECT TO DISTANCE AND DIRECTION OF OBSERVATION

The presented approach assumed creating an object representation as seen by a camera from a fixed position (the same distance from the object and the same angle of observation). This allows for recognizing the object when the distance and angle of observation does not differ significantly from the one that was used while making photos. It is true that SURF descriptors are robust with respect to the scale transformation, which accompanies changes of the distance between the camera and the object. However, this robustness has its limits. When the difference of distances between the memory representation and the actual object view is too large, the number of matched key points drops too much. Thus we have to extend the model to allow it for incorporating the features visible from both large distance, and from close neighborhood.

This also refers to different directions of observation. We collect all rotational views of of an object, but with the camera located at a fixed height. However, when we look at the same object from a point located higher or lower, the number of matched key points will be reduced. Thus for having a complete object representation, we also have to consider this aspect in the memory model.

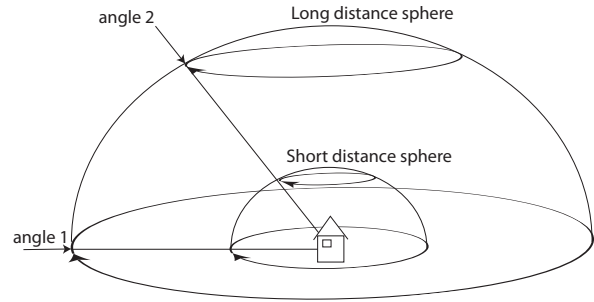


Figure 3: Illustration for placing the camera in different positions on the sphere during recording rotational views

The general extended model that we consider can include an arbitrary number of angles of observation and arbitrary number of distances from the observed object. This is equivalent to the camera motion along a number of arcs, each positioned at different distance from the object. The camera can look at any direction at the object as long as it does not change its position during recording particular sequence of views. Thus the points of observation are located on spheres with different radius along their meridians and parallels. For practical reasons we have to keep the balance of the number of spheres, and number of angles of observation. Fig. 3 shows the simplest case, where an object is observed, from two distances, and two angles (equivalent to two parallels on respective sphere). The product of the number of spheres and the number of angles gives the number of all registered sequences of rotational views, which in this case amounts to 4.

6 EXPERIMENTAL RESULTS

The Sec. 4 described all the computational steps needed to build a 3D model of key points. As already mentioned, we identify the key points using the SURF algorithm implementation available in the OpenCV library [Ope15]. The experiments presented in this work are based on images artificially generated using the NeoAxis game engine [Neo15]. Using artificial data allows for easy elimination of all kinds of noises persistent in photos of real objects, and focusing on the results generated by the algorithm itself. Moreover it is easy to set up any kind of experiments with different object and camera settings. The already presented Girl character (see Fig. 2) is transformed into the model, which is visualized in Fig. 4. It is easy to note, that the model does not reproduce the object's geometry precisely. Instead it can be considered a cloud of points around the object of interest. This is also visible in images presented in Fig. 2. However, in our case this is not a problem, because the model is designed for recognition purposes, not for precise shape reproduction. The model from Fig. 4 consist of about 1300 key points. This is more than enough, when we only expect to spot

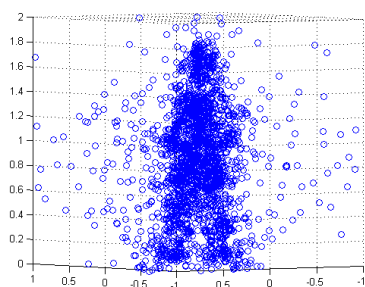


Figure 4: 3D reconstruction of the Girl object

the object within the field of view. However, for more precise analysis of the details, and possible deformation of the object geometry, a larger number of points is required. Thus for such a purpose the presented model complexity would be accurate. The number of points in the model results from the particular choice of parameters of the SURF feature detector, and the assumed precision of point fitting. It is easy to regulate the model complexity by choosing appropriate values of the parameters.

We tested the ability of perceiving objects from different distances using the extended representation as described in Sec. 5. The points from the images are identified by finding the best match to the points from the memorized object, with an assumed accuracy (distance between the respective vectors of descriptors of particular key points). This is not very advanced recognition mechanism, but allows for verification, if the model is properly constructed, and the respective points can be identified.

The object representation was extended, by making two sequences of rotational views from two different distances. The first sequence was taken from a large distance, where the object features are barely visible. The second sequence was taken from a small distance, where the object size is comparable to the size of the field of view, and a large number of object details can be perceived. The Fig. 5 demonstrates, how the memorized object is perceived in the environment for an angle observation close to horizontal direction. The number of identified key points differs significantly between the two images. When the object is seen from close distance, the number of key points typically exceeds 100. This number drops to no more than a few key points, when the object is seen from large distance. The demonstrated range in which the object can be spotted, would not be possible, without using representations coming from different distances.

We also registered the object representation from different angles of observation. Sample results are presented in Fig. 6 for the angle of about to 45° . Comparing the images those from Fig. 5, indicate that the number of key points is smaller, than in case of horizontal view.

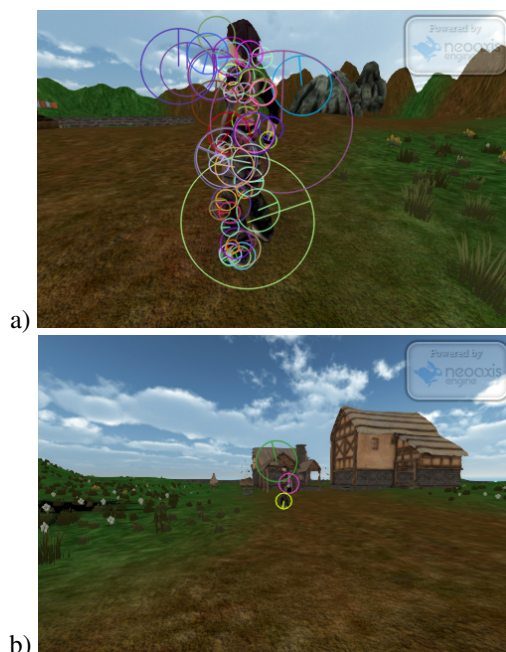


Figure 5: Key points identified within an object seen from horizontal direction (a) the Girl object seen from close distance, (b) the girl object seen from large distance

In fact our observations indicate, that for a fixed distance, the number of registered key points decreases with the growing angle of observation. For the demonstrated case of 45° the number of registered points is on average 60% of the respective number for horizontal direction of observation. This is not surprising, considering that typically from higher located observation point less details are visible. This of course depends also on the object geometry. A flat object like a carpet exhibits much more features, when observed from the top.

7 CONCLUSIONS

The presented method allows for creating 3D representations of objects based on local SURF features. The advantage of this approach is the ability to reproduce local object characteristics, which are robust with respect to occlusions, changing distance and direction of observation, as well as object geometry deformations. The method is extended, to allow for efficient recognition of objects from distant and close locations. We tested this approach in a virtual environment, where we create representations of selected objects, stored in the memory of a virtual agent. In this way, the agent is able to recognize the objects, irrespectively of the mutual location of the agent and the objects.

In addition to local features, the global object geometry is recorded in the 3D point locations. As a consequence the two aspects - local and global - can be analyzed separately. In this way the visual memory based on such a



Figure 6: Key points identified within an object seen from the direction of 45°(a) the Girl object seen from close distance, (b) the girl object seen from large distance

representation has a potential to maintain high recognition ability even if the object undergoes global geometry transformations, like changing body positions of a character. This gives the potential ability to distinguish different parts of an object, on the basis of their movements with respect to the remaining parts of the object. Further this can be used for creating internal characteristics of the objects. This issue will be investigated in the future. Our ultimate goal is creating the visual memory, which will be able to recognize different objects, create representation of the perceived scene on the basis of recognized objects' locations, and recognize different states of the objects, on the basis of their geometry transformations.

8 ACKNOWLEDGMENTS

The research is supported by The Polish National Science Centre, grant No. 2011/03/B/ST7/02518.

9 REFERENCES

- [Bay08] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L. Speeded Up Robust Features (SURF), Computer vision and image understanding, Vol. 110(3), pp. 346-359, 2008.
- [Fre02] Fremont, V., and Chellali, R. Direct Camera Calibration using Two Concentric Circles from a Single View, in Conf. Proc. ICART'02, Tokyo, pp. 93-98, 2002.

- [Fre04] Fremont, V., Chellali, R.: Turntable-based 3D object reconstruction, Conf. on Cybernetics and Intelligent Systems, 2004 IEEE, pp. 1277 - 1282
- [Jun05] Yokono, J.J., Poggio, T. Boosting a Biologically Inspired Local Descriptor for Geometry-free Face and Full Multi-view 3D Object Recognition, DSpace@MIT: Massachusetts Institute of Technology, [Online] <http://hdl.handle.net/1721.1/30557>, 2005.
- [Kno10] Knopp, J., and Prasad, M., and Willems, G., and Timofte, R., and Van Gool, L. Hough Transform and 3D SURF for robust three dimensional classification, Lecture Notes in Computer Science, Vol. 6316, pp 589-602, 2010.
- [Low01] Lowe, D.G. Local feature view clustering for 3D object recognition, IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, Hawaii, pp. 682-688, 2001.
- [Neo15] The NeoAxis Game Engine, <http://www.neoaxis.com/>, 2015.
- [Ope15] The OpenCV library <http://opencv.org/>, 2015.
- [Pfe15] Pfeifer, N., Briese, C., Laser Scanning - Principles and Applications, [Online] http://publik.tuwien.ac.at/files/pub-geo_1951.pdf, 2015.
- [Rot04] Rothganger, F., Lazebnik, S., Schmid, C., and Ponce J. 3D Object Modeling and Recognition Using Local Affine-Invariant Image Descriptors and Multi-View Spatial Constraints, Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Vol. 2, pp. 272-277, 2003.
- [Tanb08] Tangelder, J.W.H., Remco, C., Velkamp, R.C. A survey of content based 3D shape retrieval methods, Multimedia Tools and Applications, Vol. 39(3), pp 441-471, 2008.
- [Zha09] Zhang, J., Mai, F., Hung, Y.S., Chesi, G. 3D Model Reconstruction from Turntable Sequence with Multiple View Triangulation, Advances in Visual Computing, Vol. 5876, pp 470-479, 2009.

Signal and Image Processing in the Center of Cuba: Center for Studies on Electronics and Information Technologies (CEETI)

Carlos A. Ferrer
CEETI, Universidad Central
"Marta Abreu" de Las Villas
C. Camajuaní, Km 5½
54830, Santa Clara, Cuba
cferrer@uclv.edu.cu

Alberto Taboada-Crispí
CEETI, Universidad Central
"Marta Abreu" de Las Villas
C. Camajuaní, Km 5½
54830, Santa Clara, Cuba
ataboada@uclv.edu.cu

Juan V. Lorenzo-Ginori
CEETI, Universidad Central
"Marta Abreu" de Las Villas
C. Camajuaní, Km 5½
54830, Santa Clara, Cuba
juanl@uclv.edu.cu

ABSTRACT

A brief description of CEETI is given, addressing its history, research lines, main scientific results and current projects.

Keywords

Image Processing, Signal Processing, Cuba.

1. INTRODUCTION

The Center for Studies in Electronics and Information Technologies (CEETI) was founded in 1996 with the purpose of developing research projects in the fields of digital image and signal processing (DIP and DSP, respectively). Research activity at CEETI has led to various results, published in refereed scientific journals, a selection of which follows.

2. IMAGE PROCESSING AT CEETI

DIP at CEETI has been performed in response to both practical and theoretical problems. The main areas addressed are shown and commented below.

Factors affecting image diagnostic value.

The influence of radiologic dosage on image quality was addressed in [Per02] [Per03]. The reduction of the dosage while maintaining image diagnostic value is highly desirable to preserve patient's health. Expert's opinions together with ROC analysis were used to establish lower dosage limits without reaching a knee-like point in the ROC curve.

The effects of compression on image quality was the topic of [Paz09] in radiologic images, and of [Fal10] for leukocyte images. Lossy compression *codecs* like JPEG 2000, included in DICOM standard, introduce increasing image distortions as compression rates (CR) gets higher. The effect on a given image depends on its characteristics (color, contrast, noise levels, etc.). For this reason each image type is affected in a different way, and compression limits must be determined on a per-type basis. Fig. 1 illustrates the effect of compression in leukocyte images, as described in [Fal10].

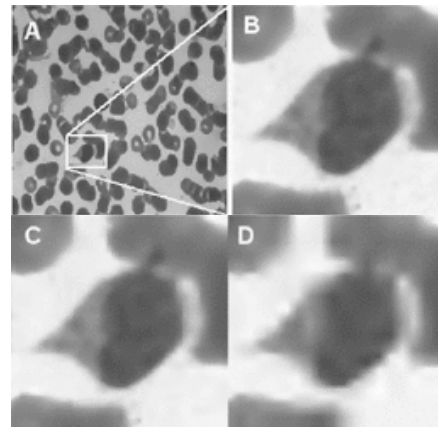


Figure 1: Effect of JPEG 2000 compression on leukocyte images. A: Original 1536H*2048W image. B, C and D: details of a monocyte with CR of 50:1, 100:1, and 200:1, respectively.

Again, expert's opinions and ROC curves were used, this time to determine CR limits.

Anomaly detection

The complex problem of anomaly detection in images was the content of [Tab09], and an application to cervical cancer detection was described in [Oro13].

In the latter work [Oro13], a preprocessing of the image is performed, based on the saliency concept from biological vision systems. Smear images show a high variability in content, as shown in Fig. 2. A new method for identifying salient regions in pap smear images is proposed and compared to previously reported approaches.

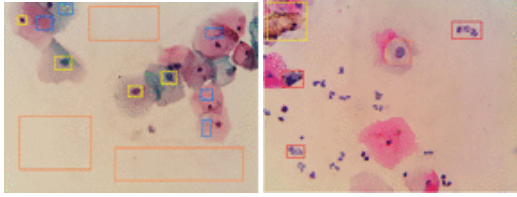


Figure 2: Left, illustration of background (orange rectangles), cytoplasm (blue) and nuclei (yellow) in Papanicolaou test images. Right, difficulties found: touching cells (yellow rectangle) and noise (red rectangles).

How such machine-saliency methods can be used to improve human performance in a realistic anomaly detection task is also described.

Noise reduction in phase images

Noise filtering of phase images was addressed in [Lor02] [Lor07] [Cru09]. An advanced application is reported in [Car02] using quantum-dot cellular automata. Phase images appear in several applications, like magnetic resonance images (MRI) or interferometric synthetic aperture radar (IFSAR).

Phase unwrapping in 2D is not uniquely defined, and noise greatly affects its estimation. In [Lor02] several methods and models were proposed and evaluated in terms of the quality of the reconstructed unwrapped phase. In Fig. 3 an example of the results obtained for the best denoising method is shown, for a spherical phantom.

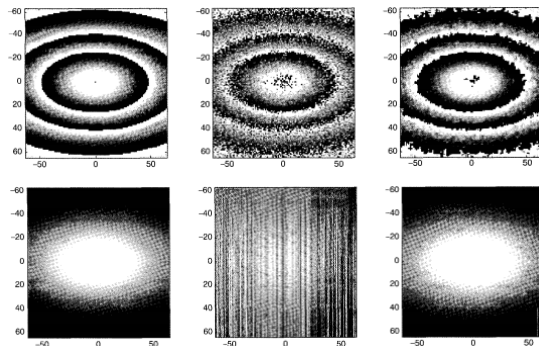


Figure 3 Wrapped (top) and unwrapped (bottom) phase images of a spherical phantom. Left: original, center: noisy, right: filtered.

Later, other denoising techniques were evaluated (e.g. wavelets in [Cru09]).

Cellular Microscopy

Research on DIP of cellular microscopy images has led to results reported in various papers, among them in [Oro11] [Lor13] and [Coc14].

In [Oro11], the segmentation of cell nuclei and cytoplasm is required, since the ratio between both areas in a cell is an important feature for cancer diagnosis. Cell nuclei in Papanicolaou tests are highly

variable in terms of shape and texture, as shown in Fig. 4.

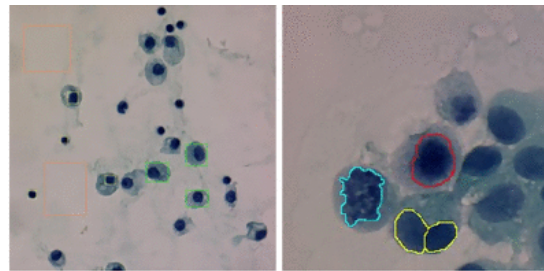


Figure 4: Left, illustration of background, cytoplasm and nuclei in Papanicolaou test images. Right: nuclei variability (cyan, red and yellow).

This work used a two phase approach for nuclei segmentation, validated against manually segmented images and compared to several state-of-the-art methods.

3. OTHER RELEVANT RESULTS

Through its history, CEETI has also reached visibility in research areas other than DIP.

Compression, filtering and classification of ECG signals was the main topic in CEETI's early years [Mor97] [Car98] [Tab99]. EEG signals have also been addressed, either for its compression [Car04] [Baz12a] [Baz12b] or analysis [Rui09] [Rui10] [Rui12].

Speech processing has been an active area of research at CEETI, where most significant results are associated to the measurement of perturbations in pathological speech [Fer05] [Fer09] [Fer10].

Some results have also been obtained in the processing of genomic sequences [Fue07] [Lor09].

General-purpose DSP algorithms and techniques have also been addressed, like noise reduction using wavelets [Can02] or bionic models in classification tasks [Gon10].

4. CURRENT PROJECTS

The line of DIP at CEETI is currently centered in two main topics, one in the detection of anomalies in cell microscopy, and other in neuronal (and arteries/veins) tracing methods. Biomedical applications are still a priority at CEETI.

As for other research topics outside DIP (i.e. DSP), pathological speech processing is the strongest line. The usefulness of speech to diagnose a pathology or reflect its severity is being studied on dysphonias, cognitive impairing diseases and sleep apnea, among others.

5. ACKNOWLEDGMENTS

The authors wish to acknowledge Dr. Vaclav Skala, for kindly inviting us to WSCG-2015.

6. REFERENCES

- [Baz12a] Bazán-Prieto, C., Blanco-Velasco, M., Cárdenas-Barrera, J.L. and Cruz-Roldán, F. Analysis of tractable distortion metrics for EEG compression applications. *Physiological measurement* 33 No.7, pp. 1237-1247, 2012.
- [Baz12b] Bazán-Prieto, C., Blanco-Velasco, M., Cárdenas-Barrera, J.L. and Cruz-Roldán, F. Retained energy-based coding for EEG signals. *Medical engineering & physics* 34 No.7, pp. 892-899, 2012.
- [Can02] Cancino-de-Greiff, H.F., Ramos-García, R., and Lorenzo-Ginori, J.V. Signal de-noising in magnetic resonance spectroscopy using wavelet transforms, *Concepts in Magnetic Resonance* 14 No. 6, pp. 388-401, 2002.
- [Car98] Cárdenas-Barrera, J.L. and Lorenzo-Ginori, J.V. Mean-shape vector quantizer for ECG signal compression. *IEEE Transactions on Biomedical Engineering* 46 No.1, pp. 62-70, 1998.
- [Car02] Cárdenas-Barrera, J.L., Plataniotis, K.N. and Venetsanopoulos, A.N. QCA implementation of a multichannel filter for image processing. *Mathematical Problems in Engineering* 8 No.1 pp. 87-99, 2002.
- [Car04] Cárdenas-Barrera, J.L., Lorenzo-Ginori, J. V. and Rodríguez-Valdivia, E. A wavelet-packets based algorithm for EEG signal compression. *Informatics for Health and Social Care* 29 No.1, pp. 15-27, 2004.
- [Coc14] Coca-Rodríguez, A. and Lorenzo-Ginori, J V. Effects of Interpolation on Segmentation in Cell Imaging. *Computación y Sistemas* 18 No.1, pp. 97-109, 2014.
- [Cru09] Cruz-Enríquez, H. and Lorenzo-Ginori, J. V. Combined Wavelet and Nonlinear Filtering for MRI Phase Images. *Lecture Notes in Computer Science LNCS 5627*, pp. 83-92, 2009.
- [Fal10] Falcón-Ruiz, A., Paz-Viera, J. and H Sahli Estimating Quality Bounds of JPEG 2000 Compressed Leukocytes Images. *Lecture Notes in Computer Science LNCS 6256*, pp. 107-114, 2010.
- [Fer05] Ferrer, C., González, E. and Hernández-Díaz, M. E. Correcting the use of ensemble averages in the calculation of Harmonics to Noise Ratios in voice signals. *Journal of the Acoustical Society of America* 118, No.2, pp. 605-608, 2005.
- [Fer09] Ferrer, C., González, E., Hernández-Díaz, M. E., Torres, D. and del-Toro, A. Removing the Influence of Shimmer in the calculation of Harmonics-to-Noise Ratios using Ensemble-Averages in Voice Signals. *EURASIP Journal on Advances in Signal Processing*, 2009:784379, 2009.
- [Fer10] Ferrer, C., Torres, D., and Hernández-Díaz, M. E. Using dynamic time warping of T0 contours in the evaluation of cycle-to-cycle Pitch Detection Algorithms, *Pattern Recognition Letters* 31 No.6, pp 517-522, 2010.
- [Fue07] Fuentes, A., Ginori, J. and Abalo, R. A new predictor of coding regions in genomic sequences using a combination of different approaches. *International Journal of Biomedical and Life Sciences* 3 No.2, pp. 1-5. 2007.
- [Gon10] Gonzalez, E., Liljenström, H., Ruiz, Y., and Li, G. A biologically inspired model for pattern recognition. *Journal of Zhejiang University SCIENCE B* 11 No.2, pp. 115-126, 2010.
- [Lor02] Lorenzo-Ginori, J.V., Plataniotis, K.N. and Venetsanopoulos, A.N. Nonlinear filtering for phase image denoising. *IEE Proceedings on Vision, Image and Signal Processing* 149 No.5, 290-296, 2002.
- [Lor07] Lorenzo-Ginori, J. V. An Approach to the 2D Hilbert Transform for Image Processing Applications. *Lecture Notes in Computer Science LNCS 4633*, pp. 157-165, 2007.
- [Lor09] Lorenzo-Ginori, J.V., Rodríguez-Fuentes, A., Abalo, R.G., and Rodríguez, R.S. Digital signal processing in the analysis of genomic sequences. *Current Bioinformatics* 4 No.1, pp. 28-40, 2009.
- [Lor13] Lorenzo-Ginori, J.V., Curbelo-Jardines, W., López-Cabrera, D. and Huergo-Suárez, S.B. Cervical Cell Classification Using Features Related to Morphometry and Texture of Nuclei. *Lecture Notes in Computer Science, LNCS 8259*, pp. 222-229, 2013.
- [Mor97] Morales, R.O., Sánchez, M.A.P, Ginori, J.V.L., Abalo, R.G. and Ramírez, R.R. Evaluation of QRS morphological classifiers in the presence of noise. *Computers and Biomedical Research* 30 No.3, pp. 200-210, 1997.
- [Oro11] Orozco-Montegudo, M. Sahli, H., Mihai, C. and Taboada-Crispi, A. A hybrid approach for Pap-Smear cell nucleus extraction. *Pattern Recognition, Lecture Notes in Computer Science LNCS 6718*, 2011, pp 174-183, 2011.
- [Oro13] Orozco-Montegudo, M., Taboada-Crispi, A., and Sahli, H. Biologically Inspired Anomaly Detection in Pap-Smear Images. *Lecture Notes in Computer Science LNCS 8259*, pp.17-24, 2013.
- [Paz09] Paz, J., Pérez, M., Schelkens, P. and Rodríguez, J. Impact of JPEG 2000 compression on lesion detection in MR imaging. *Medical Physics* 36 No.11, pp. 4967-4976, 2009.

- [Per02] Díaz, M.P., García, J.Q., Vicente, F.P. and Rizo, O.D. Administered activity optimization in patients studied by equilibrium gated radionuclide ventriculography using pyrophosphate and technetium-99m. Nuclear medicine communications 23 No.4, pp. 347-353, 2002.
- [Per03] Díaz, M.P., Aparicio, E.E., Rizo, O.D., Díaz, R.R. and Rodríguez, C.H. Administered activity optimization in ^{99m}Tc-MAG3 renography for adults. Journal of nuclear medicine technology 31 No.4, pp. 216-221, 2003.
- [Rui09] Ruiz, Y., Li, G., Freeman, W.J. and Gonzalez, E. Detecting stable phase structures in EEG signals to classify brain activity amplitude patterns. Journal of Zhejiang University SCIENCE A 10 No.10, pp. 1483-1491, 2009.
- [Rui10] Ruiz, Y., Pockett, S., Freeman, W.J., Gonzalez, E. and Li, G. A method to study global spatial patterns related to sensory perception in scalp EEG. Journal of neuroscience methods 191 No.1, pp. 110-118, 2010.
- [Rui12] Ruiz, Y., Li, G., Gonzalez, E. and Freeman, W.J. A new approach to detect and study spatial-temporal intracranial EEG frames. Digital Signal Processing 22 No.1, pp. 133-139, 2012.
- [Tab99] Taboada-Crispi, A., Lorenzo-Ginori, J.V. and Lovely, D.F. Adaptive line enhancing plus modified signal averaging for ventricular late potential detection. Electronics Letters 35 No.16, 1293-1295, 1999.
- [Tab09] Taboada-Crispi, A., Sahli, H., Orozco-Monteagudo, M., Hernandez-Pacheco, D. and Falcon-Ruiz, A. Anomaly Detection in Medical Image Analysis, pp 426-446, in: Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications, 2009.

Computer Graphics and Vision Labs in Argentina

Armando De Giusti	María José Abásolo	Marcelo Naiouf
III-LIDI Facultad de Informática Universidad Nacional de La Plata calle 50 y 120 Argentina (1900) La Plata degusti@lidi.info.unlp.edu.ar	III-LIDI Facultad de Informática Universidad Nacional de La Plata calle 50 y 120 Argentina (1900) La Plata mjabasolo@lidi.info.unlp.edu.ar	III-LIDI Facultad de Informática Universidad Nacional de La Plata calle 50 y 120 Argentina (1900) La Plata mnaiouf@lidi.info.unlp.edu.ar
Silvia Castro	Roberto Guerrero	
Laboratorio de Visualización y Computación Gráfica Dpto. de Cs. e Ingeniería de la Computación Universidad Nacional del Sur Alem 1253 Argentina (8000) Bahía Blanca vyglab@cs.uns.edu.ar	Departamento de Informática Facultad de Cs. Físico Matemáticas y Naturales Universidad Nacional de San Luis Ejército de los Andes 950 Argentina (5700) San Luis rag@unsl.edu.ar	

ABSTRACT

This article describes the Computer Graphics and Vision research that are being developed by three universities' laboratories in Argentina: III-LIDI Institute of Computing Research at National University of La Plata, VyGLab Visualization and Computer Graphics Laboratory at South National University, and CGLab Computer Graphics Laboratory at National University of San Luis.

Keywords

Computer Graphics, Computer Vision, Argentina

1. III LIDI INSTITUTE OF COMPUTING RESEARCH

The III-LIDI Institute of Computing Research¹ belongs to the Informatic Faculty at National University of La Plata – *Universidad Nacional de La Plata (UNLP)* (Figure 1).



Figure 1. Informatic Faculty at UNLP (Argentina)

Current research topics include the design and implementation of virtual and augmented reality, and

interactive TV educational applications. Besides there is a great effort in human resource training so the group works in consolidating a posgraduate career in Image Processing, Computer Graphics and Vision.

Augmented Reality in Education

There are educational opportunities for Augmented Reality (AR) and Geolocation to enrich Teaching and Learning. One of the goals of the group is to study the incorporation of this technology in educational contexts of different educational levels, from kindergarten to university. The design of educational applications focusing on active learning involves the creation of scenarios to explore and interact with their real environment related with their interests and situations related to real life. Also it is needed to develop tools that allow teachers to use new technologies.

Virtual Reality in Education

One of the main applications of Virtual Reality (VR) are simulators, which provide the user with a realistic experience in the management of vehicles or machinery. Ship simulator was developed for educational and community outreach purposes. The virtual reality equipment consists in three LED TVs at an angle of 120 ° between adjacent screens, and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

virtual reality glasses with tracking. Its main characteristic is the low-cost and its portability. This allows to move the equipment to sciences exhibitions which are visited by students and general public (Figure 2).



Figure 2. Ship simulator in a science exhibition

The group is also focused in the design and implementation of multi-resolution visualization of 3D terrain to improve realism taking advantage of the processing capabilities of the GPU and memory constantly growing.

Interactive Television

The group coordinates the “RedAUTI Thematic Network in Applications and Usability of Interactive Digital Television (IDTV)” - *ReAUTI Red de Aplicaciones y Usabilidad de la Televisión Digital Interactiva*ⁱⁱ- which is conformed of 39 iberoamerican research and development groups from Spain, Portugal and ten latinoamerican countries (Figure 3). The main goal is the design and implementation of applications, services and content production for IDTV, in its multiple platforms, open source to solve problems of the Latin American context.



Figure 3. Participants of RedAUTI Thematic Network in Applications and Usability of IDTV

Postgraduate Teaching

The Informatic Faculty creates in 2012 a new postgraduate career called “Specialization in Computer Graphics, Images and Computer Vision” with the aim of strengthen human resources in this areas. This carrer can be taken as the main step for a Phd student to route thesis work in this direction. The Teachers are both from domestic and also foreign universities.

2. VyGLab VISUALIZATION AND COMPUTER GRAPHICS LABORATORY

The Visualization and Computer Graphics Laboratory (VyGLab)ⁱⁱⁱ belongs to the Computer Science and Engineering Department at South National University – *Universidad Nacional del Sur (UNS)* (Figure 4).



Figure 4. VyGLab at UNS (Argentina)

This group performs core and applied research in visualization and computer graphics. The main areas of expertise are volume modeling, augmented reality and visualization. Current research topics include volume modeling with wavelets, augmented reality books, outdoors augmented reality, semantic based visualization and information visualization.

Volume modeling

The group is working on the definition of wavelets over non nested tetrahedral grids, allowing the representation of functions defined on the irregular tetrahedrization of a volume. In this way, it is possible to represent different attributes of the volume object and its interior such as color, brightness, density, etc.

Outdoors Augmented Reality

The group is working on visualization of geological information in the field by means of tablets PC. This allows geologists to visualize existing geological data sets superimposed to the ground terrain as though they were part of the environment and to explore various above and under-ground phenomena associated with these datasets. This is a new

application area with strong potential for a wide spectrum of applications.

Augmented Reality Books

The group have developed an interactive and collaborative system for traditional books augmentation that allows the addition of AR content to any pre-existent traditional book (Figure 5). It is working on interactions with the book and it is experimenting with users to validate the proposals.

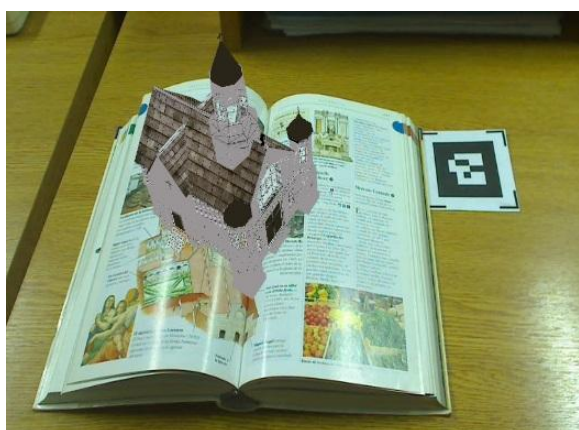


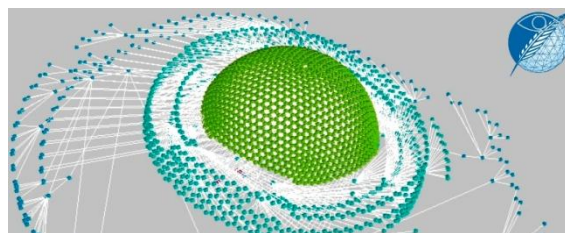
Figure 5. Augmented Reality Books

Semantic Based Visualization

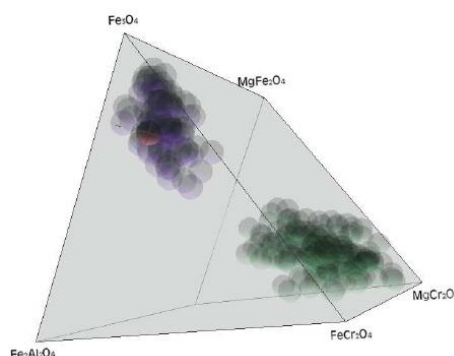
Taking into account that data visualization process is a very complex exploration activity and, even for skilled users, the group is working on *semantic based visualization* in order to figure out how to assist users and designers throughout the stages of the visualization process. At this time our goal is to obtain a formal representation of the visualization field to establish a common visualization vocabulary, including the underlying semantics, and enable the definition of visualization specifications that can be executed by a visualization engine with ontological support.

Large Data Visualization

Related to *Information visualization* we are mainly dedicated to multidimensional visualization of large volume of data sets originating from the natural sciences, and in particular from the Geological Sciences (Figures 6-a and 6-b). This poses serious challenges with regard to analysis and interpretation without visual support. Visualization contributes significantly to the exploration and understanding of the meaning of these datasets; currently we are mainly dedicated to the visualization of mineral compositions and to the automatic visual analysis of geological samples by point counting.



a)



b)

Figure 6. Information Visualization

Computer Graphics and Visualization Teaching

In addition to the research activities the members of the group dedicate to computer graphics and visualization teaching. The teaching program includes several undergraduate courses: a required course in Computer Graphics included in the Computer Engineering Curricula and optional courses on Advanced Computer Graphics, Real Time Rendering, Information Visualization and Big Data Visualization that can be taken by students from different degree programs. It also includes advanced postgraduate courses in the same subject. These courses are included in the postgraduate curricula offered in several Universities across the country.

3. CGLab COMPUTER GRAPHICS LABORATORY

The Computer Graphics Laboratory (CGLab) is a research and development group at National University of San Luis – *Universidad Nacional de San Luis (UNSL)* (Figure 7). All researchers work in the Informatic Department corresponding to the Physics, Mathematics and Nature Sciences Faculty.

The former group started computer graphics research in 2001 with the study of Non-photorealistic Computer Graphics applications by image processing techniques. Through the years basic tools have been enlarged to include powerful computer graphics theory concepts and techniques. Later in 2012 the group was formally constituted.



Figure 7. National University of San Luis (Argentina)

CGLab's research lines are: Non-photorealism, Virtual Reality, Real Life Simulation, Video games, Training. Most representative projects are *ReTTrAc*, *CAVE-Vox* and *Be Civic*.

Real Time Traffic Accidents Simulation

ReTTrAc, acronym of *Real Time Traffic Accidents*, is part of the project called Alfa III GAVIOTA financed by the European Community Programme^{iv}. It has brought together European and Latin American Computer Graphics, Engineering and Architecture research groups. The developed work is a general Virtual Reality computing platform that enables real time visualization of 3D scenarios for manufacturing, training and forensic simulations. The platform is able to treat static and dynamic 3D environments, allowing to share the experience of navigation in the scene among the users, even geographically distributed. The platform proposed was validated through real time 3D models manipulations and interaction in simulated car crashes.

Conversational Agents

CAVE-Vox is a conversational character as a question-answering assistant for task-generic applications into a Cave-like environment (Figures 8-9). This work develops software to design a virtual character and provides it with verbal skills interaction and locomotion. The proposed virtual character is based on open-source components and runs entirely into a Cave-like environment. The system design was validated by conducting human tests to measure the real time, realism, interaction ability and technical brilliance of conversation.



Figure 8. Cave environment at CGLab UNSL



Figure 9. Virtual reality application running in a cave environment at CGLab UNSL

Serious games

Be Civic project uses a serious game in a immersive learning environment for teaching civics. In order to improve the user's experience the game was developed to work in a cave-like immersive environment, using a conversational character for interaction. The game includes static and dynamic 3D environments, allowing to share the experience of navigation in the scene among the users, even geographically distributed.

ⁱ III-LIDI <http://weblidi.info.unlp.edu.ar>

ⁱⁱ RedAUTI <http://redauti.net>

ⁱⁱⁱ VyGLab <http://vyglab.cs.uns.edu.ar>

^{iv} ALFA GAVIOTA <http://www.alfagaviota.eu>