Automatic Video Generation Using Floor Segmentation from a Single 2D Image

Geetha Kiran A

Malnad College of Engineering

India (ZIP) 573202, Hassan, Karnataka

geethaamk@gmail.com

Murali S

Maharaja Institute of Technology

India (ZIP) 570001, Mysore, Karnataka

murali@mitmysore.in

ABSTRACT

Image based video generation paradigms have recently emerged as an interesting problem in the field of robotics. This paper focuses on the problem of automatic video generation of indoor scenes that mainly consist of orthogonal planes. The algorithm infers frontier information directly from the images using a geometric context-based segmentation scheme that uses the natural scene structure in indoor environments. The presence of floor is a major cue for obtaining the termination point for the video generation. First, we perform floor segmentation using dilation and erosion methods. Second, compute the length of the floor using distance method which is used as the termination point for video generation. Finally, video is generated by cropping the image. Our approach needs no human interventions, hence it is fully automatic. We demonstrate the technique to a variety of applications, including virtual walk through ancient time images, in forensics and in architectural sites. The algorithm is tested on nearly 100 images obtained from different buildings, all of them are fairly different in interior decoration themes from each other.

Keywords

dilation, erosion, floor segmentation, floor length, video generation

1. INTRODUCTION

Video generation from a single image is inherently a challenging problem. In Imaging devices, there is a trade-off between the images (snapshots) and video because of the limitation in storage capacity. Video clips need more storage space compared to images. This motivated to generate the video from a single 2D image rather than storing video clips. Humans analyze variety of single image cues and act accordingly, unlike robots. The work is an attempt to make robots analyze similar to humans using single 2D image.

The task of generating video from photographs is receiving increased attention to generate video of architectural sites. We are addressing here the key case where dimension of the real world object or measurement of object dimension in 2D plane is unknown. However generating video using above methods is very difficult because of perspective view. Alternatively, video could be generated using proper ground known i.e., floor segmentation. In the absence of accurate measurements, we wish to exploit geometric characteristics (windows/doors) along with the color variations. Such relationships are plentiful in man-made structures and often provide sufficient information to our work. The work is well-suited for navigation on Personal Digital assistants(PDA's) and personal computers, includes cases where buildings are destroyed and only the archieve images are available. The work here is mainly carried out using indoor images. We describe a unified framework for generating video from a single 2D image.

This paper focuses on the problem of automatic video generation of indoor scenes that mainly consist of orthogonal planes. The presence of floor is a major cue for obtaining the termination point for the video generation.

Video generation using single image finds applications, including virtual walk through ancient time images, in forensics and as waiters in restaurants. It not only helps the users to enjoy the important details of the image but also provides a vivid viewing manner.

In the next section, a review on the related works is highlighted.

2. RELATED WORK

It is observed that some methods have been developed for segmentation on a single image, few which are directly relevant to the work are highlighted here. Erick Delage *et. al*. have used a graph based segmentation algorithm to generate a partition of the image and assigned a unique identifier to each partition output by the segmentation algorithm in [Con00a]. Erick Delage *et. al*.[Con00c] have built a probabilistic model that incorporates a number of local image features and tries to reason about the chroma of the floor in each column of the image. Ma Ling *et. al.* [Con00e] have segmented the floor region automatically

by adopting clustering analysis and also have proposed a PCA based improved version of the algorithm to remove negative effect of shadow for segmented results. Xue – Nan Cui *et. al.* [Con00f] have proposed detecting and segmenting the floor by computing plane normals from motion fields in image sequences. A geometric characteristic that objects are placed perpendicular to the ground floor can be utilized to find the floor in 2D images. Surfaces often have fairly uniform appearances in texture and color and thus image segmentation algorithms provide another set of useful features which can be used in many other applications, including video generation.

A very few Researchers have proposed different methods for video generation from a single 2D image. Shuqiang jiang *et. al.* [Con00g] have proposed a method to automatically transform static images to dynamic video clips in mobile devices. Xian-sheng Hua *et. al.* [Con00h] developed a system named photo2video to convert a photographic series into a video by simulating camera motions. The camera motion pattern (both the key-frame sequencing scheme and trajectory/ speed control strategy) is selected for each photograph to generate a corresponding motion photograph clip. A region based method to generate a multiview video from a conventional 2-dimensional video using color information to segment an image has been proposed by Yun-Ki-Baek *et. al.* [Con00i]. Na-Eun Yang *et.al.*[Con00j] have proposed method to generate depth map using local depth hypothesis and grouped regions for 2D-to-3D conversion. The various methods of converting 2D to stereoscopic 3D images involves the fundamental, underlying principle of horizontal shifting of pixels to create a new image so that there are horizontal disparities between the original image and the new version. The extent of horizontal shift depends on the distance of the feature of an object to the stereoscopic camera that the pixel represents. It also depends on the inter-lens separation to determine the new image viewpoint.

The methods proposed by the authors for floor segmentation is time consuming and have made certain assumptions specific to the application. These artifacts are not of much importance in our work, this made us to propose a simple method for floor segmentation in lesser time. Using the segmented image, length of the floor could be computed by distance method. This helps in video generation.

In this paper, a method for systematically exploring an unknown bounded indoor workspace is presented.

3. FLOOR SEGMENTATION

The goal is to obtain floor segmentation of a given single 2D indoor image. The crucial part of the work is detecting the pixels belonging to the floor. There are methods available for floor segmentation with known camera parameters. Requirements is to segment floor without having knowledge of camera parameters. There is possibility to find the geometric relationship, may be using color. The primary steps involves converting the given color image to gray, further convert the gray image to binary image by computing a global threshold. Finally, segment the floor by applying the dilation and erosion methods.

Segmentation:

The floor path is the major cue to generate video from a single 2D image of indoor scenes. To segment the floor from the remaining parts of the indoor image scenes, dilation and erosion techniques using the structuring elements are used.

Structuring element is used for probing and expanding the shapes contained in the input image yielding to floor segmentation. The output of the above steps is given in Figure 1.

(a)
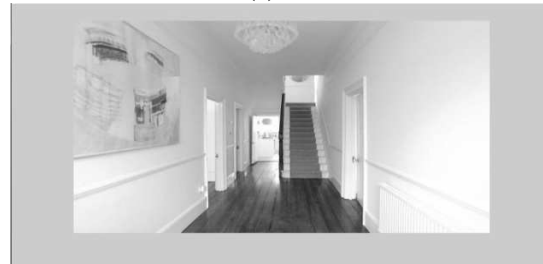
(b)

(c)

(d)

Figure 1. (a) Original Image (b) Gray Image (c) Binary image using Otsu's method (d) Segmented Image

The segmented image obtained in Figure 1(d) is used to find the length of the floor. The distance between the start and end of the white pixel (row wise) from the floor segmented image is found by using the Euclidean distance method. This length of the floor identified could directly be used to decide the number of frames to be generated, generally 1:2 depends on the length and it can be varied with requirements. These frames are incorporated in the video generation.

4. Video Generation

The information obtained in the floor segmentation is used to generate the video. The input for the video generation are - single 2D image, computed termination point based on the distance calculated using floor segmentation, the size of the rectangle based on which cropping takes place.

The input image is considered as the first frame and the image is cropped based on the size of the predefined rectangle. Then the cropped image is resized to the original image and stored in an array of images. An appropriate set of key-frames are determined for each image based on the distance computed by using floor segmentation. The images obtained after cropping is given in Figure 2.

(a)

(b)

(c)

(d)

Figure 2. (a) Frame 1 (b) Frame 40

(c) Frame 80 (d) Frame 200

Further video is generated, using the key frames stored in the array by writing the frames to the video file. This method provides vivid dynamic effect from global view to local details.

5. EXPERIMENTAL RESULTS

The algorithm is applied to a test set of 97 images obtained from different buildings, all of them are fairly different in interior decoration themes from each other. Since the indoor images contained a diverse range of orthogonal geometries (wall posters, doors, windows, boxes, cabinets etc.), we have observed that the results presented are indicative of the algorithm performance on images of new buildings (interior) and scenes.

We also have evaluated the algorithm by manually detecting the floor path of a set of images and compared it with the floor path generated by our method. The overall accuracy obtained from the result is 91.46% as given in Figure 3.
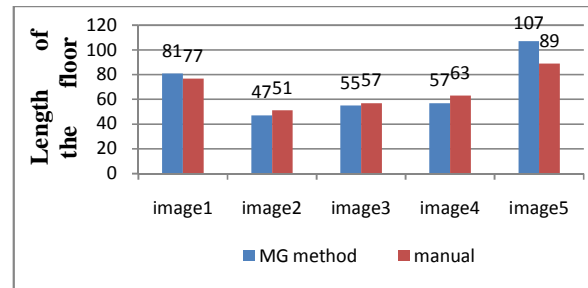

Figure 3. Comparing the length of the floor computed manually and by our method *(MG method)*

The first, intermediate and final frame generated by the method after floor segmentation is shown in Figure 4. We can observe the finer details in the intermediate and final frames that could be used in various applications including virtual walk through ancient time images, in forensics and in architectural sites.

The painting faithfully follows the geometric rules and also have color variations and therefore we can apply the methods developed here to have a virtual walk in the imaginary world.

6. CONCLUSION

An algorithm for automatic video generation from a single 2D image is proposed and experimented for only indoor images. This paper provides a solution to transform static single 2D image into video clips. It not only helps the users to enjoy the important details of the image but also provides a vivid viewing manner. The experimental results show that the algorithm is performing well on a number of indoor scenes.

Further work can be extended to produce videos including side view, working at planar level. This requires maintenance of perspective view of the scene.

REFERENCES

[Con00a] Erick Delage Honglak Lee Andrew Y. Ng: A dynamic Bayesian network model for autonomous 3d reconstruction from a single indoor image, CVPR 2006.

[Con00b] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Analysis and Machine Intelligence, pp.603–619, 2002.

[Con00c] Erick Delage, Honglak Lee, and Andrew Y. Ng: Automatic Single-Image 3d Reconstructions of Indoor Manhattan World Scenes, ISRR,2005

[Con00d] D. Hoiem, A. A. Efros, and M. Hebert: Geometric context from a single image. 10th IEEE International Conference on Computer Vision, 17-21 Oct, 2005.

[Con00e] Ma Ling , Wang Jianming ; Zhang Bo ; Wang Shengbei :Automatic floor segmentation for indoor robot navigation, 2nd International Conference on Signal Processing Systems (ICSPS),pp. 684 - 689, 5-7 July 2010.

[Con00f] Xue-Nan Cui, Young-Geun Kim, and Hakil Kim: Floor Segmentation by Computing Plane Normals from Image Motion Fields for Visual Navigation,International Journal of Control, Automation, and Systems, pp.788-798, 2009.

[Con00g] Shuqiang Jiang and Huiying Liu and Zhao Zhao and Qingming Huang and Wen Gao: Generating video sequence from photo image for mobile screens by content analysis, ICME, pp.1475-1478, 2007.

[Con00h] Xian-sheng Hua and Lie Lu and Hong-jiang Zhang: Automatically Converting Photographic Series into Video, 12th ACM International Conference on Multimedia,pp.708-715, 2004.

[Con00i] Yun-Ki Baek, Young-Ho Seo,Dong-Wook Kim and Ji-Sang Yoo: Multiview Video Generation from 2-Dimensional video, International Journal of Innovative Computing, Information and Control, Vol 8, Number 5(A), pp. 3135-3148, May 2012.

[Con00j] Na-Eun Yang, Ji Won Lee, Rae-Hong Park: Depth Map Generation from a Single ImageUsing Local Depth Hypothesis, 2012 IEEE ICCE,pp.311-312,13-16 Jan 2012.

[Con00k] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. International Journal of Computer Vision, 59, 2004.

[Con00l] Young geun Kim and Hakil Kim:. Layered ground floor detection for vision based mobile robot navigation. In IEEE Robotics and Automation (ICRA), volume 1, pp 13 – 18, 2004.

[Con00m] Y. J. Jung, A. Baik, J. Kim, and D. Park, A novel 2D-to-3D conversion technique based on relative height depth cue, in Proc. Stereoscopic Displays and Applications XX, vol. 7237, Jan. 2009.
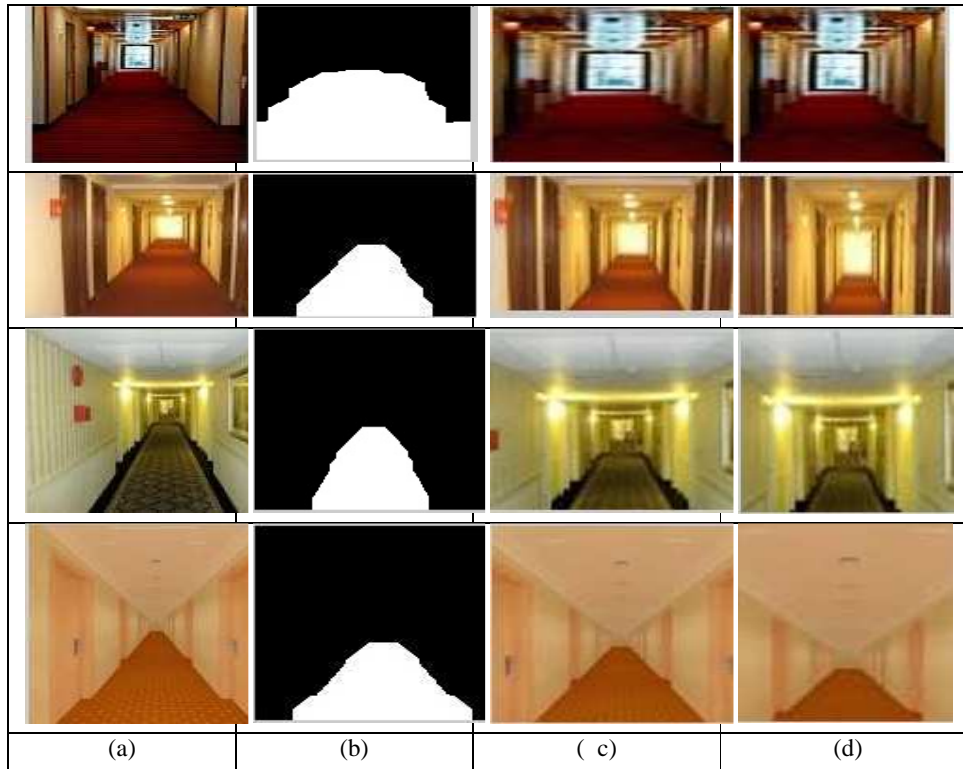
| (a) | (b) | (c) | (d) |

Figure 4. (a) Input Image (b) Floor Segmentation (c) Intermediate frame (d) Final Frame