Image Registration for Multi-exposure High Dynamic Range Image Acquisition

Anna Tomaszewska
Szczecin University of Technology
atomaszewska@wi.ps.pl

Radoslaw Mantiuk
Szczecin University of Technology
rmantiuk@wi.ps.pl

ABSTRACT

We present a fully automatic method for eliminating misalignments between a sequence of hand-held photographs taken at different exposures. The key component of the technique is the SIFT method that is employed to search for key-points (or feature-points) in consecutive images. The key-points are used to find matrices, that transform a set of images to a single coordinate system and eliminate any global misalignments (including general planar homography). We employ this technique to capture high dynamic range images from a set of photographs taken at different exposures, where misalignments can cause blurring and artifacts, and prevent achieving high quality HDR images. The proposed alignment technique works well for over- and under-exposed images and is not sensitive to an image content. We present our implementation of the technique and the results of tests made for variety of photographs.

Keywords: image alignment, image registration, contrast domain, HDR capture, High Dynamic Range Imaging.

1 INTRODUCTION

There is tremendous progress in the development and accessibility of high dynamic range (HDR) imaging technology [RWPD05]. Modern image processing and graphics software becomes HDR enabled. Also HDR digital photography replaces low dynamic range (LDR) technologies. HDR photographs are of much better quality and easier to be processed in a digital darkroom. Unfortunately, HDR cameras are still very expensive and not available for average users. On the other hand, taking HDR photographs seems to be legitimate and crucial. In the near future LDR images may become almost obsolete due to the progress in LCD technology [SHS+04] and it will not be easy to display LDR image correctly. LDR photographs will look pale and not interesting on HDR LCD displays.

The multi-exposure HDR capture technique [MP95] seems to be a good alternative to HDR cameras, which can be used to create an HDR image from photographs taken with a conventional LDR camera. The technique uses differently exposed photographs to recover the response function of a camera. From the response function, the algorithm creates an HDR image whose pixel values are proportional to the true radiance (or luminance) value of a scene. Because this technique requires multiple input photographs, there is a high likelihood of misalignments between pixels in the sequence of exposures due to moving of a hand-held camera (global motion) or dynamic object in a scene (local motion causing ghosting). It is crucial that misalignments between input photographs should be removed before fusing an HDR image.

Our work addresses the problem of global alignment (registration) of hand-held photographs taken with different exposures. The method compensates misalignments caused by any movement of a camera. In particular, general planar homography (or linear planar projective transform) is considered so any linear transformation of a point position between planes on images is recovered (see [HZ06] for details on the properties and limitations of planar homography). The key component of the technique is the scale-invariant feature transform (SIFT) [Low04] that is employed to search for key-points (often called feature-points) in consecutive images. We modified the algorithm by automatizing contrast threshold value calculation. The key-point correspondences are used to find matrices, that map a set of images to a single coordinate system. The bilinear interpolation of the output images based on the calculated matrices completes the algorithm. The registered photographs are fused using our implementation of multi-exposure HDR capture technique proposed by Mitsunaga and Nayar [MN99]. Our alignment technique is fully automatic and is not sensitive to image content. It works correctly even for highly under- and over-exposed images. We modified registration algorithm so it can be applied not only for standard LDR images but also for HDR images. We can use extended dynamic range photographs stored in RAW format as an input to alignment algorithm. This way, less input photographs can be used to create high dynamic range image. Moreover, we noticed that the accuracy of image registration is much better for input RAWs than for standard LDR photographs.

In Section 2 we present a review of previous work related to image alignment for multi-exposure HDR capture techniques. A detailed description of our SIFT based HDR alignment technique is presented in Section 3. In Section 4 we show implementation of the method followed by the results of the tests run on hand-captured photographs (Section 5). Section 6 presents conclusions and discusses future work directions.

2 RELATED WORK

In recent years significant progress has been made in the development of algorithms that allow to capture HDR images using low dynamic range sensors (standard LDR cameras) [MP95, DM97, MN99, RBS99]. These algorithms retrieve high dynamic range information from a sequence of photographs. The authors suggest using tripod to avoid camera movement and they do not address the problem of eliminating misalignments.

The problem of image alignment and matching was intensively studied during last years [ZF03, Bro92] but not for registration of images of different exposures. The only solution that addressed exactly the problem of capture of HDR photographs, was proposed in [War03, RWPD05]. The technique employs conversion of input photographs into percentile threshold bitmaps. The bitmaps are analyzed and then aligned horizontally and vertically using shift and difference operations over each image. In chapter 5 we compare the results achieved with this algorithm to our method.

Kang et al. [KUWS03] described a technique for creating high dynamic range video from a sequence of altering light and dark exposures. A part of the technique is a HDR stitching process, which includes global and local alignment to compensate for pixel motion. The stitching process can be also used to compensate for camera movement when creating an HDR still from a series of bracketed still photographs. However, the presented technique seems to be suitable for video where there are no large differences between consecutive frames.

In [ST04] Sand and Teller present a global and local matching algorithm, which is robust to changes in exposure of photographs. The key idea behind this technique is to identify which parts of the image can be matched without being confused by parts that are difficult to match. Such assumption seems to be not valid for images with large differences in exposures, where there is usually not enough information for correct matching. The technique was designed for matching two video sequences and was not tested on still photographs.

Recently, Cerman and Hlavac [CH06] presented an alignment method based on unconstrained nonlinear optimization. In this method, each image is linearized using the estimated camera response function and multiplied by the exposure ratio. Then, a normalized difference summed across all corresponding pixels is used to estimate misalignments. The method can compensate global rotation and horizontal and vertical shifts.

There are a few techniques which compute camera response function based on misaligned photographs [GN03, KP04]. However, these methods are not meant to create HDR images. The problem of removing ghosting artifacts in a multi-exposure sequence of photographs was also investigated [EAKR06, RWPD05] but proposed algorithms do not take into consideration a compensation of camera movements.

In the next section, we present our technique of removing global misalignments between photographs in a multi-exposure sequence. This technique can be applied to still images and it compensates any camera movement.

3 MISALIGNMENT COMPENSATION

Our modification of multi-exposure HDR capture method is presented in Figure 1. We align all input photographs to a selected reference (the image with the best exposure) before fusing an HDR image rather than use a tripod as suggested in the previous methods.

The algorithm of our alignment technique is shown in Figure 2. The technique establishes correspondences between points, lines or other geometrical entities in a set of images. We use a modified SIFT (Scale Invariant Feature Transform) [Low04] algorithm to extract local features called key-points. They are located at scale-space maxima/minima of a difference of Gaussian function. The key-points descriptors accumulated in orientation histograms form the invariant descriptors.

In the next step, these descriptors are used to find correspondences between key-points in the reference image and remaining images. The number of the correspondences is reduced by RANSAC [FB81] algorithm because only four pairs of key-points (8 coefficients) are needed to calculate transformation matrix. The RANSAC selects a set of key-points correspondences that are compatible with a homography between the images. It uses the singular value composition (SVD) method to solve over-determined system of linear equations required by the optimization algorithm.

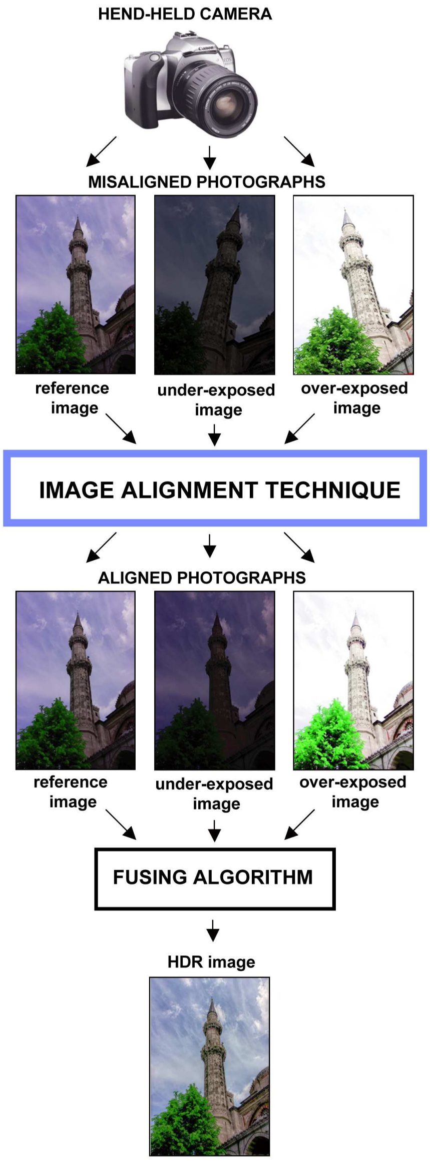

Figure 1: A technique of capturing HDR photographs with a hand-held camera based on the image alignment and image fusing techniques.

The transformation matrix is computed by the direct linear transform (DLT) algorithm. This matrix is used to transform a given image to the coordinates of the reference image:

$$x' \sim \mathbf{H} \cdot x, \quad (1)$$

where x' and x are points from destination and source images respectively, and H is a transformation matrix:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix},$$

h_{ij} are unknowns.

The algorithm computes the color values of consecutive output pixels using bilinear interpolation of appropriate input pixels. This solution ensures that there are no holes in an output image. Moreover a sub-pixel precision of calculation improves the quality of images (decreases aliasing).

We noticed that the accuracy of the transformation matrix can be decreased due to doubtful correspondences. The RANSAC algorithm chooses the best pairs of key-points based on a random technique, which can fail. We improve the accuracy of the correspondences selection by choosing only these key-points which occur in all photographs simultaneously.

The key-points searching algorithm works in contrast domain so it is intensity invariant. Therefore, we can use it to find the similarities between images of different exposures, where intensity values can change significantly. The algorithm uses a multi-scale pyramid of Difference-of-Gaussian images to improve the key-points detection (a larger neighborhood of a pixel is considered).

As has been mentioned above the key-searching algorithm operates also with RAW format as an input data. However, in order to improve the accuracy of image registration for such pictures we modified it. Instead of using constant contrast threshold value we calculate if for each image separate in relation its dynamics. This modification makes it possible to detect far more key-points in RAW images in comparison to LDR ones. It enhances accuracy for pictures taken in bad light conditions. The results are presented in section 5.

4 IMPLEMENTATION

The algorithm for multi-exposure image alignment has been implemented as a part of *pfstools* [1] package [MKM07]. The main components of *pfstools* are programs for reading and writing images in all major HDR and LDR formats (e.g. OpenEXR, Radiance's RGBE, etc.) and programs for basic image manipulation (rotation, scaling, cropping, etc.). The typical usage of *pfstools* involves executing several programs joined by UNIX pipes. The first program transmits the current frame or image to the next one in the chain. The final program should either display an image or write it to a disk. An example of command line is given below:

[1] *pfstools* is distributed as Open Source under the GPL license and the project web page can be found at `http://pfstools.sourceforge.net/`

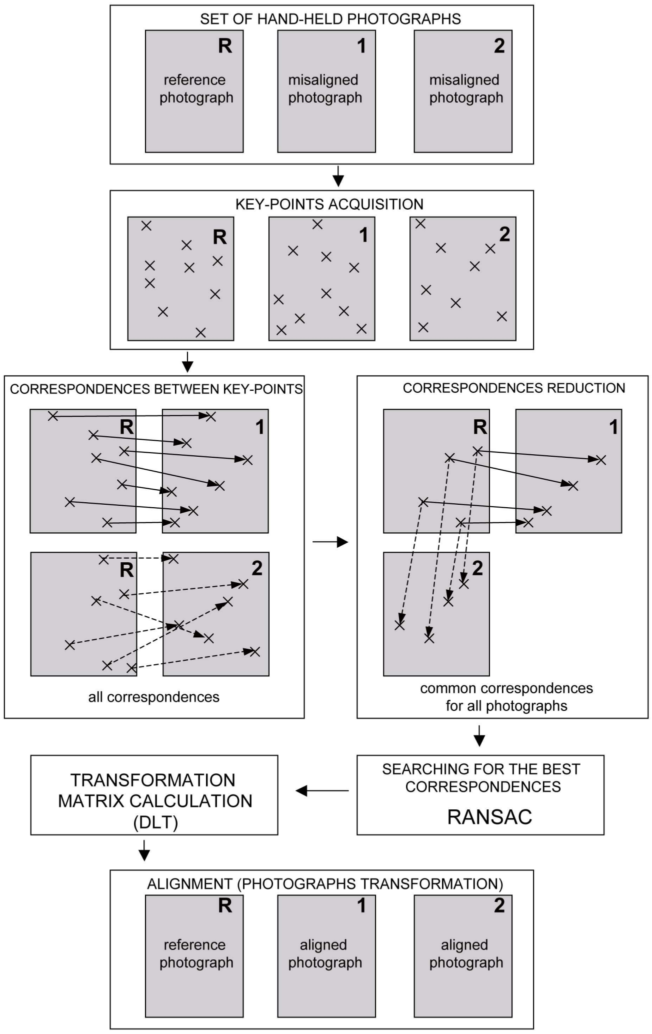

Figure 2: The algorithm of the image alignment technique.

```
pfsin␣input1.jpg␣input2.jpg␣|␣pfsalign␣/
␣|␣pfshdrcalibrate␣|␣pfsout␣output.exr
```

Read the images `input1.jpg` and `input2.jpg`, align the images (`pfsalign`), create HDR image (`pfshdrcalibrate`) and write the output to `output.exr`.

The *pfsalign* program (see the example above) is responsible for recovering misalignments and transforming input images to a single coordinate space. This program was implemented in C++ and, like the *pfstools* package, can be run under Mac OS, Linux and Windows (with cygwin support).

The *pfsalign* consists of two main modules. The first one looks for key-points and correspondences between them, and then calculates transformation matrix. We used SVD solver implementation from numerical recipes book [PTVF92] to aid the calculation of the transformation matrix. The second module transforms images to single coordinate space with sub-pixel precision so that there are no holes in resulting images.

The *pfshdrcalibrate* program (see the example above) fuses images of different exposures and calculates a HDR radiance map. We implemented the fusing algorithm proposed by Mitsunaga and Nayar [MN99] which is based on polynomial approximation of the camera response function. We find this algorithm very suitable, especially in case of the input sequence of a limited number of photographs (e.g. 2-3). The polynomial function fills gaps in a response function so that it is possible to calculate correct radiance value even for pixels without good representation in the input photographs (e.g. pixels with maximum value (255) in all photographs).

5 RESULTS

To test the usability of our multi-exposure image alignment method, we have executed an extensive set of tests on photograph sequences. The main goal of the tests was to check the accuracy of alignment for images of short and long exposures (very light and very dark) and of different image content. All photographs were taken with hand-held digital camera (Canon 10D with Sigma 18-125 mm F3.5-5.6 DC lens) and encompassed any possible combination of camera movements (shifts, rotations and zooming). To evaluate the accuracy of the image alignment, we fuse a set of aligned images into a HDR image and subjectively estimate the quality of that image.

In the top row of Figure 3 a set of misaligned photographs of the same scene is depicted (the set consists of three photographs of resolution 1536x1024 pixels stored in 8-bits JPEG format). The photographs were aligned using our algorithm (the second row) and then a HDR image was created (bottom row). The photographs were aligned with sufficient accuracy so that there are no visible artifacts in the final HDR image. There are also no visible holes in the aligned images (from the middle row), as they were filled with bilinearly interpolated pixel values. The algorithm requires about 450 seconds to align three presented photographs on the CPU. The time of registration depends on an image content. For comparison, we also show a false HDR (the third row - left) which was fussed based on misaligned images from the top row. In the third row (right), the HDR image created by *hdrgen* (available at `http://www.anyhere.com/`) program is presented. It can be seen in this figure that simple transformations implemented in *hdrgen* are not enough for accurate image registration.

In Figure 4 we show a few additional results of image alignment. In these examples the capabilities of removing shifts, rotations, zooms and a combination of these transformations are presented. The algorithm correctly recognizes even large movements of a camera.

Figure 5 presents another example of input image set with very light (over-exposed) and very dark (under-exposed) photographs. One can see that despite a large difference in exposure, sufficient number of key-points was found to establish a correspondence between images and calculate transformation matrices. Moreover, there are no key-points in unsteady too dark or too light areas.

Figure 6 depicts the arrangement and number of key-points detected for LDR and RAW images respectively. The pictures were taken in the same light conditions. In RAW images, in contrast to LDR pictures, features were detected also in poorly lit fragments of images. Due to that we have a more regular arrangement of key-points in a picture compared with LDR images, and what follows we can calculate a more accurate transformation matrix.

We successfully tested our alignment technique for many photographs of different content. The technique requires only four pairs of corresponding key-points to calculate transformation matrix so it will give correct results also for images with large non-textured areas and parallax effects.

The algorithm fails for images that contain many recurrent textures occupying a large portion of an image. In such cases, many key-points have the same descriptors and RANSAC (more precisely SVD) cannot compute good (it means falling below an error threshold) key-point correspondences. We find this problem difficult to solve based on the feature based alignment method (our algorithm belongs to this group). Fortunately, such kinds of photographs are rather rare in practice.

6 CONCLUSION AND FUTURE WORK

The major outcome of this work is an image registration method that can be used to remove misalignments between images in a sequence of differently exposed photographs (in RAW and/or JPEG format). The method is fully automatic and compensates misalignments caused by any movement of a camera. It works in contrast domain so is less sensitive to image content and changes in exposure. The method employs SWIFT algorithm to find key-points, RANSAC algorithm to choose the best key-point correspondences and DLT technique to calculate transformation matrix. The input photographs are transformed to a single coordinate system with sub-pixel precision to avoid holes and aliasing artifacts. The results of this work facilitate capturing HDR images based on the multi-exposure techniques. Our method simplifies acquisition of HDR images by removing the main disadvantage of the multi exposure techniques, which is a necessity of using a tripod.

We implemented and tested our method for a variety of photographs. We find it effective in most cases but several difficult images cause the method to fail. The solution of this problem is further optimization of SIFT and RANSAC algorithms for adjusting them to high dynamic range input images. We also plan to estimate the transformation matrix not only for planar homography but also for more general 3D case with depth reconstruction. Since our current implementation does not deal with local changes, we intend to implement a de-ghosting technique in the future. We plan to speed-up the implementation based on the GPU programming to facilitate the practical usage of our method.

ACKNOWLEDGMENTS

We would like to thank Piotr Ostiak for basic implementation of the image matching algorithm. Special thanks go to Rafal Mantiuk and Grzegorz Krawczyk

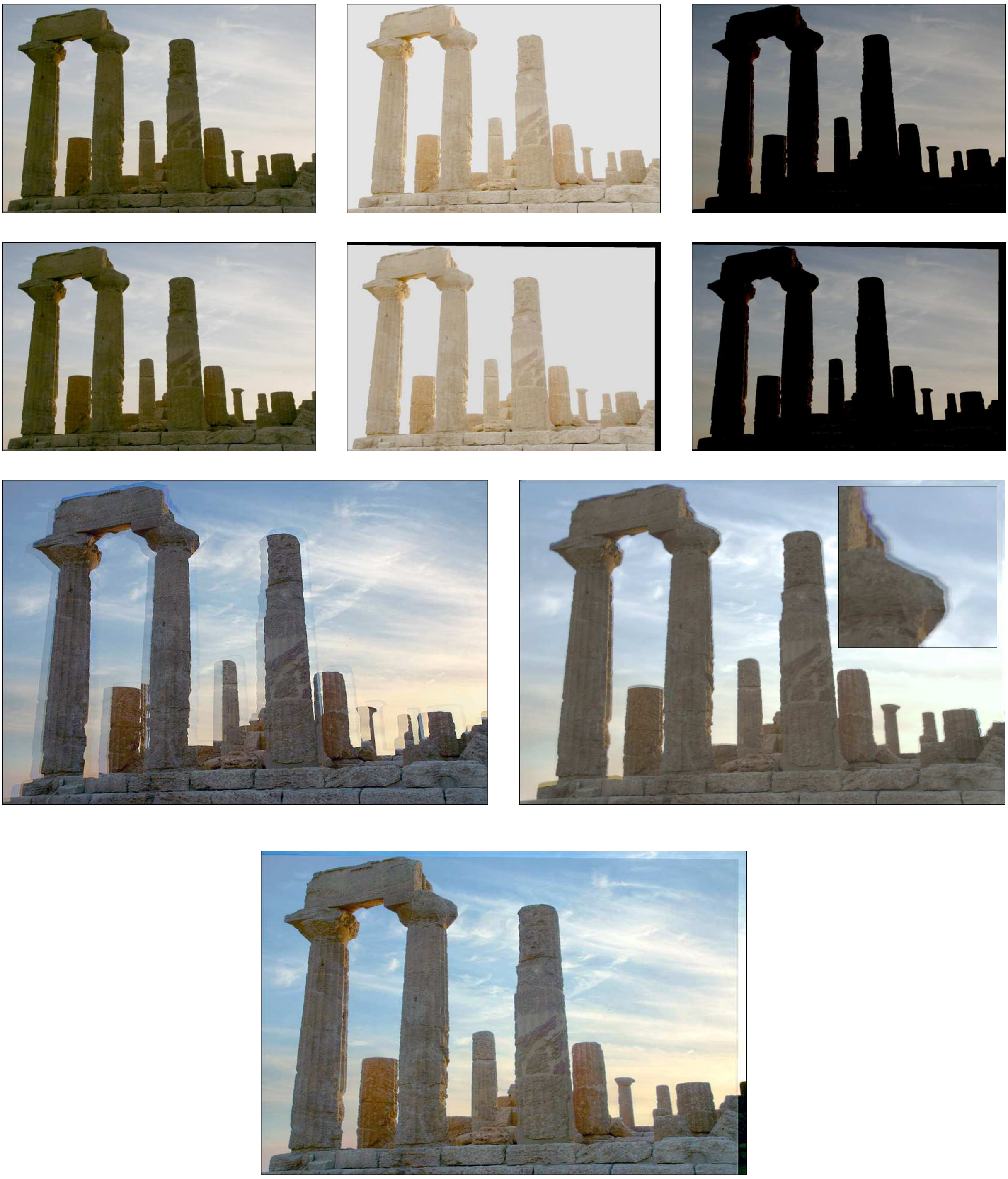

Figure 3: The HDR image captured with a hand-held camera. Top row: misaligned photographs. The second row from top: aligned photographs, the first photograph is a reference image. The third row: HDR image created by fusing the misaligned photographs (left) and by *hdrgen* software (right) (the inset depicts misalignment artifacts). Bottom row: HDR image created from aligned photographs. All HDR images were tone mapped using contrast domain operator. [MMS05])

Figure 4: The results of image alignment. Left column: reference photographs. Middle column: photographs before alignment. Right column: photographs after alignment (transformed to the coordinate system of the reference photograph). From the top to the bottom row, the examples show shift, rotation, zoom, and complex camera movement (general homography).

for their helpful comments concerning this work and for delivering the *pfstools* software.

REFERENCES

[Bro92] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Comput. Surv.*, 24(4):325–376, 1992.

[CH06] Lukas Cerman and Vaclav Hlavac. Exposure time estimation for high dynamic range imaging with hand held camera. In *Proc. of Computer Vision Winter Workshop, Czech Republic*, February 2006.

[DM97] P.E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM SIGGRAPH 1997*, pages 369–378, 1997.

[EAKR06] Ahmet Oguz Akyuz Erum Arif Khan and Erik Reinhard. Ghost removal in high dynamic range images. In *Proc. of IEEE International Conference on Image Processing*, October 2006.

[FB81] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381Ð395, 1981.

[GN03] M.D. Grossberg and S.K. Nayar. Determining the camera response from images: What is knowable? *PAMI*, 25(11):1455–1467, November 2003.

[HZ06] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2006.

[KP04] S.J. Kim and M. Pollefeys. Radiometric self-alignment of image sequences. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.

[KUWS03] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High dynamic range video. *ACM Trans. Graph.*, 22(3):319–325, 2003.

[Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[MKM07] Rafal Mantiuk, Grzegorz Krawczyk, and Radoslaw Mantiuk. High dynamic range imaging pipeline: Perception-motivated representation of visual content. In *To be published in Proc. of Human Vision and Electronic Imaging XII, IS&T/SPIE's Annual Symposium on Electronic Imaging (2007)*, SPIE Proceedings Series, San Jose, California USA, January 2007.

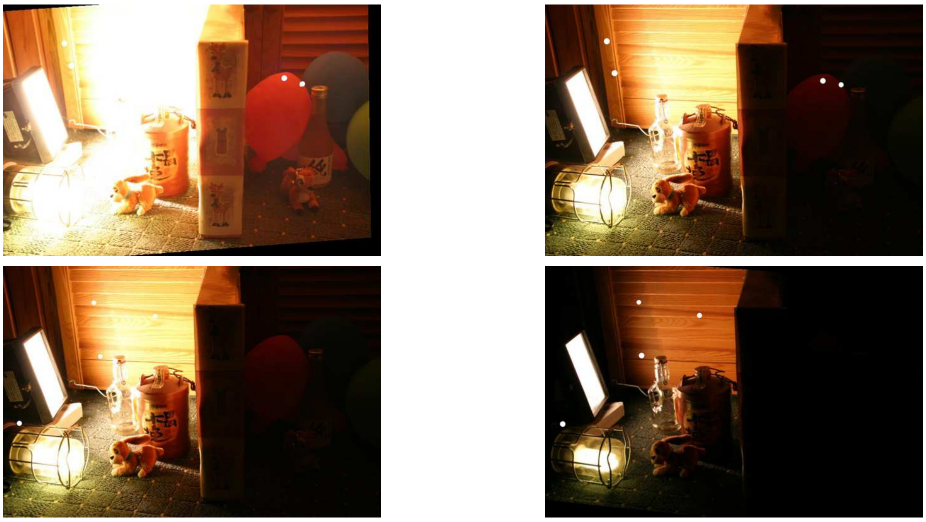

Figure 5: Example photographs taken with a very short and very long exposures. Valid key-points are marked as white dots. There are no key-points in too dark and too light areas.

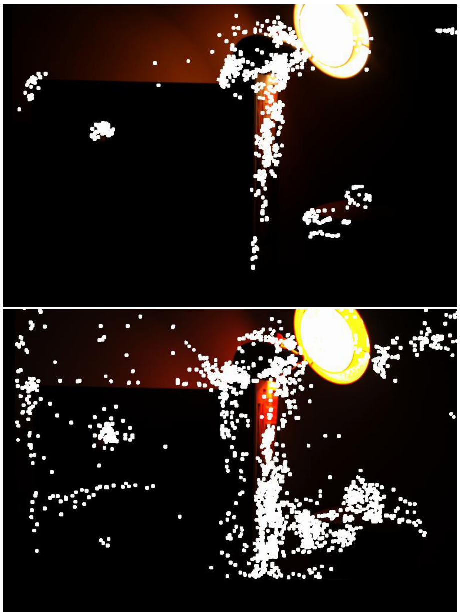

Figure 6: An arrangement and number of key-points detected for LDR (top) and RAW (bottom) images taken in the same light conditions.

SPIE.

[MMS05] Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. A perceptual framework for contrast processing of high dynamic range images. In *APGV '05: 2nd Symposium on Appied Perception in Graphics and Visualization*, pages 87–94, 2005.

[MN99] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. pages 374–380, 1999.

[MP95] S. Mann and R. Picard. Being 'Undigital' with Digital Cameras: Extending Dynamic Range by Combining Differently Exposed Pictures. *Proc. of IST's 48th Annual Conference*, pages 442–448, May 1995.

[PTVF92] William Press, Saul Teukolsky, William Vetterling, and Brian Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

[RBS99] M. Robertson, S. Borman, and R. Stevenson. Dynamic range improvement through multiple exposures. In *Proceedings of the 1999 International Conference on Image Processing (ICIP-99)*, pages 159–163, Los Alamitos, CA, October 24–28 1999.

[RWPD05] Erik Reinhard, Greg Ward, Sumanta Pattanaik, and Paul Debevec. *High Dynamic Range Imaging. Acquisition, Manipulation, and Display*. Morgan Kaufmann, 2005.

[SHS+04] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. *ACM Trans. on Graph.*, 23(3):757–765, 2004.

[ST04] Peter Sand and Seth Teller. Video matching. *ACM Trans. Graph.*, 23(3):592–599, 2004.

[War03] Greg Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *journal of graphics tools*, 8(2):17–30, 2003.

[ZF03] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.