HDRi2013 - First International Conference and SME Workshop on HDR imaging (2013), paper no.9
Bessa M., and Mantiuk R. (Editors)

Patch-based registration for auto-stereoscopic HDR content creation

R. Ramirez Orozco [1], C. Loscos [2], I. Martin [1] and A. Artusi [3]

[1] Group of Geometry and Graphics. Universitat de Girona (UdG), Spain
[3] Graphics & Imaging Laboratory (GILab). Universitat de Girona (UdG), Spain
[2] CReSTIC SIC, Universite du Reims Champagne-Ardenes (URCA), France

Abstract

Creating High Dynamic Range (HDR) images of static scenes is a common procedure nowadays that combines several Low Dynamic Range (LDR) images. However, HDR video and 3D content creation and management is an active, unsolved research topic. This work proposes a method to build HDR images from Low Dynamic Range (LDR) input images taken with multi-view cameras. We propose an image registration method to produce 3D HDR content for auto-stereoscopic displays. This method is based on the Patch Match algorithm which has been adapted to take advantage of epipolar geometry constraints of multi-view cameras. Different image similarity measures are used to improve the accuracy of the matching process. In our case we use an 8-view LDR camera from which we generate an 8-view HDR output. Partial results show accurate registration and HDR reconstruction for each LDR input image.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Computational Photography—3D High Dynamic Range

1. Introduction

The creation of high dynamic range (HDR) content has been an intense field of research lately. Several approaches to combine multi-exposed images either of static or dynamic scenes have been presented with different degrees of success. In cases where the pixels of source images are not perfectly aligned, ghosting effects may appear in the output HDR image. There are two main causes for ghosting: camera movement during acquisition and dynamic objects in the scene. The first one can be easily solved using image alignment techniques like the bitmap threshold method presented by Ward [War03]. Several methods to deal with dynamic scenes have been presented but there is not a standard solution for this, specially when both camera and objects in the scene move. The lack of robust methods for multi-exposed image registration of dynamic scenes is one of the main drawbacks for HDR video acquisition. The main goal of this work is to propose a method for image registration of image sequences for auto-stereoscopic 3D HDR content creation.

Bonnard *et al.* [BLV*12] have proposed a methodology to create content that combines depth (3D) and high dynamic range video for auto-stereoscopic displays. The input for such displays is a set of images taken with a fix exposure on each objective. The Octo-cam [PcPD*10] has eight objectives synchronized in time. The different exposures are achieved by mean of neutral density filters placed in each objective that fix the percentage of light that arrives to each sensor. Each filter divides by two the amount of light received by each sensor. The fact that the eight sensors are synchronized prevents of capturing movement in the scene. The ghosting problem due to dynamic objects does not exist because all the objectives capture the scene at same time and same shutter speed. However, images are different because they are taken from different points of view and at four different exposures like in Figure 1. Images need to be perfectly aligned to avoid artifacts in the construction of an HDR frame for each view of a video sequence. Bonnard *et al.* used reconstructed depth information from epipolar geometry to lead the pixel match procedure. This method lack of robustness especially on under and over exposed areas. To avoid this

Figure 1: Set of LDR multi-view images taken with an eight view camera using natural color filters to control exposures.

we propose a solution to combine stereoscopic LDR images into HDR using image registration based on the Patch Match algorithm [BSFG09]. This algorithm has been used recently by Sen et al. [SKY*12] to build HDR images. The results were promising for multi-exposure sequences where the reference image is moderately under or over exposed, but it fails otherwise. We propose to overcome this drawback using other color difference measure than SSD and take advantage of geometric constraints in the set of images to help the matching process. Figure 2 shows an overview of the process. We iterate over the set of multi-exposed images from the Octo-cam selecting a reference image each time, then all the remaining images are registered using the modified patch match and finally they are merged into one HDR per view.

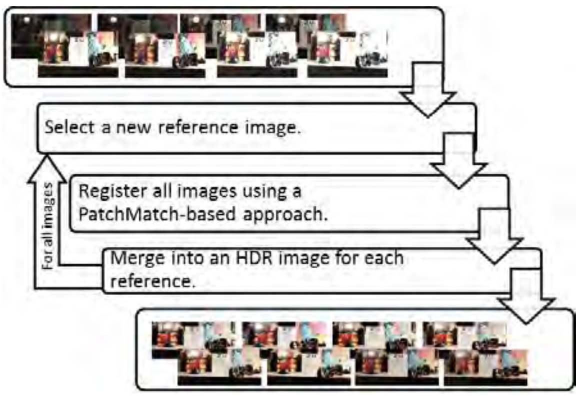

Figure 2: Proposed workflow to generate HDR images for auto-stereoscopic display.

The rest of the paper is organized as follow. Section 2 focuses on giving a brief background about the state of the art on this topic, section 3 describes the solution we are developing and finally section 4 shows some partial results of the ongoing work.

2. Background

2.1. HDR Video Acquisition

Several prototypes have been proposed to acquire HDR videos from multi-exposure sequence of images. Lin and Chang [LC09] proposed a scheme based on a rig of two cameras placed like a conventional stereo configuration. Each stereo image pair was acquired with different shutter speed and was used for HDR image synthesis and stereo disparity computation. They use a stereo matching algorithm based on propagation to derive the disparity map and a ghost removal technique is used to avoid artifacts due to noise or stereo mismatches. Rüfenacht's master thesis [Ruf11] presents a prototype of rig with two cameras to capture stereoscopic video varying the exposure of each view. He presents two different approaches to obtain stereoscopic HDR content. The first is the temporal approach, where different exposures are captured by temporally changing the exposure times of both cameras, recording two frames of the same exposure in each shot alternating the exposure time between consecutive pairs of frames. The second is the spatial approach, here each camera have a different exposure time for all the shots, in this case each frame in the same shot have different exposures. Tocci *et al.* [TKTS11] present an optical architecture for HDR imaging that allows simultaneous capture of high, medium, and low-exposure images on three sensors at high fidelity with efficient use of the available light. Their work also includes an HDR merging algorithm to complement their prototype, which avoids undesired artifacts when there is a large exposure difference between the images. They use beam splitter inside the camera to split the light beam. The resulting HDR video is thus the combination of information from three different exposures taken simultaneously from a same viewpoint. Recently, Kronander *et al.* [KGBU13] presented a new approach for HDR assembly directly from raw sensor data in a single processing operation. This method was implemented using CUDA and show real time performance. They assume that geometric missalignments between sensors can be described by a

2D transformation matrix so that the images can be registered which is not always possible. HDR acquisition with dynamic content or video input has been addressed with many different approaches in the last decade. However, besides for [Ruf11] and [LC09] who proposed a method for stereo pairs, multi-view HDR video (more than two views) has not yet been addressed except by Bonnard *et al.* [BLV*12].

2.2. Image Alignment

In the HDR context most of methods for image alignment focus on movement between images caused by hand-held capture, small movement of tripods or moving pixels from dynamic objects in the scene. Most of them assume that all images are taken from the same viewpoint so they are not suitable for the multi-view images since the kind of transformations that takes place are different. Sand and Teller [ST04] proposed a combination of feature matching and optical for spatiotemporal alignment of different exposed videos. They align a pair of videos by searching for frames that best match according to an image registration process. This process uses locally weighted regression to interpolate and extrapolate image correspondences. This method is robust to changes in exposure and lighting, but if there are objects moving at high speed, artifacts may still appear.

Mangiat [MG10] propose to use a method of block-based motion estimation and refine the motion vectors in saturated regions using color similarity in the adjacent frames of an alternating multiexposed sequence. Niquin et al [NPR10] use the octocam to reconstruct a 3D scene using a pixel matching method based on graph cuts. This method is suitable for images with the same exposure but the precision of the matching is not good for multiexposed images. Sun *et al.* [SMW10] proposed an algorithm based on the assumption that the disparity map between two rectified images can be modeled as a Markov random field. The matching problem is then posed as a Bayesian labeling problem in which the optimal values are obtained minimizing an energy function. The energy function is composed of a pixel dissimilarity term (using NCC as similarity measure) and a smoothness term which correspond to the MRF likelihood and the MRF prior, respectively. Most of stereo matching algorithms perform energy minimization schemes, which imply high computational cost.

Sen [SKY*12] recently presented a method based on a patch-based energy-minimization formulation that integrates alignment and reconstruction in a joint optimization. This allows to produce an HDR result that is aligned to one of the exposures and contains information from all of them, but important artifacts appears when there are large under or over exposed areas in the reference image.

3. Patch-based approach

Our solution is based on the work of Barnes [BSFG09, BSGF10] and Sen's work [SKY*12]. We propose to adapt the matching process to the multi-view context resulting in a more robust and faster solution. To understand the basis of our approach we need first to introduce briefly the original patch-based algorithm [BSFG09, BSGF10].

The Patch Match solves the matching problem between two images A and B at patch level using Nearest Neighbor Fields (NNF) minimization. NNF is defined over all possible patch coordinates (locations of patch centers) in image A; for some distance function D between two patches of images A and B. Given a patch coordinate **a** in image A and its corresponding nearest neighbor **b** in image B, NNF(**a**) is simply **b**. The values of NNF for all coordinates are stored in an array with the same dimensions of A. It is a randomized algorithm that works iteratively improving the NNF until convergence. Initially it can be filled either with random values sampled across image B or with previous hint information. An iterative process is performed to improve the NNF for a fixed number of iterations or until a convergence criteria. The algorithm tries to improve NNF using two sets of candidates: propagation and random search. Propagation uses the known nearest neighbors patches to improve NNF, and converges very quickly but it may end in a local minimum. A second set of candidates is used to avoid local minimum by introducing random samples. This step is called random search, the candidates are sampled from a distribution of pixels located at an exponentially decreasing distance from each patch. For each pixel v_i, the candidates u_i are sampled at an exponentially decreasing distance $u_i = v_0 + w\alpha^i R_i$ where R_i is a uniform random in [-1,1]x[-1,1], w is the maximum search radius and α is a fixed ratio between window sizes (1/2 in our case). After NNFs are calculated, a distance metric proposed by Simakov [SCSI08] is used to guarantee both the coherence and completeness of the output image.

The camera we use to capture the input sequence is described in [PcPD*10]. There are some geometrical features that could help to reduce the number of potential corresponding patches. Since all the objectives of the camera are aligned in the vertical axis, the resulting images have a different perspective but pixels in different images of the sequence share the same epipolar line. This fact reduces the random search to 1D (only on the epipolar line). This constrain would reduce the search space making it more precise. The use of a random search function instead of a full energy minimization scheme contributes also to speed up the solution. We propose to modify the search function to look only in the same horizontal line adding a Δ value in the vertical axis to prevent errors in case of geometrical misscalibration of images.

Patch Match have been originally implemented using sum of squared differences (SSD) as a distance measure between patches in L*a*b color space. Despite SSD is one of the most popular measure for stereo matching, several works compare SSD with different similarity measures [VLS*06, OMLV12, HS09]. We compare the results with other measures based on pixels correlation and mutual information (MI).

4. Work in Progress Results

(a) Reference

(b) Target

(c) Difference 1st Iteration

(d) Reconstructed 1st Iteration

(e) Difference 5th Iteration

(f) Reconstructed 5th Iteration

Figure 3: Images d) and f) show the result of registering b) to a) for the first and the fifth iterations of our Patch Match implementation. Images c) and e) shows the normalized SSD difference between the result and the reference image

This section shows some of the results obtained up today to illustrate the feasibility of our proposal. Several more tests need to be done to present some statistic comparison of the improvements and show the final stereoscopic HDR reconstruction. We have tested our approach using different sequences of input images acquired with the camera described in [PcPD*10]. Figure 3 shows two images from the sequence of Figure 1, the result of the registration after five iterations and using a patch size of 2x2 and the normalized sum of squared difference (SSD) error between the result and the reference image.

Tests were made performing the matching step in two different stages of the procedure. For some test the matching step was performed directly on the LDR 8-bits images like the tests of Figure 3 and Figure 4, while for other tests we calculated the camera response curve using Debevec and Malik's method [DM97] and transformed all images to the radiance domain before doing the matching step. In the second approach the differences in exposure has less influence on the matching process and the convergence is much faster.

(a) Reference

(b) Target

(c) Difference 1st Iteration

(d) Reconstructed 1st Iteration

(e) Difference 5th Iteration

(f) Reconstructed 5th Iteration

Figure 4: Images d) and f) show the result of registering b) to a) for the first and the fifth iterations of our Patch Match implementation. Images c) and e) shows the normalized SSD difference between the result and the reference image

The two input images in Figure 4 are from the stereo dataset of the Caltech-256 (`http://vision.middlebury.edu/stereo/data/scene2006`) also taken with different exposures and from different points of view. In this test the registration was made on the original 8-bits LDR image with the same patch size than described above. These results were obtained with a number of iterations equal to 5, the normalized sum of squared difference (SSD) error between the result and the reference images is shown in Figure 4 (c). As shown in the difference image, the difference between our output and the reference images is small, making the results promising for future work. The biggest errors are located around the color boundaries, it is important to compare results using different similarity measures to avoid artifacts related with the use of SSD for comparing colors. The influence of the patch size and the computation time on the results need to be addressed.

Figure 5 shows an example of the matching process executed after transforming images to the same domain. The convergence in this case is faster than matching 8-bits images. The full test showed in Figure 6 was generated using only two iterations per image. The images shown as in Figure 5 d) and e) are mapped back to 8-bits using the camera response curve.

(a) Reference

(b) Target

(c) Difference 1st Iteration

(d) Reconstructed 1st Iteration

(e) Difference 2nd Iteration

(f) Reconstructed 2dn Iteration

Figure 5: Images d) and f) show the result of registering b) to a) for two first iterations of our Patch Match implementation. Images c) and e) shows the normalized SSD difference between the result and the reference image

Figure 6 shows a full test over a set of seven images from the Caltech-256 datased mentioned before. The resulting HDR images have no visible ghosting effects due to missalignment but they are different in terms of color because of large saturated areas in some of the reference images. Future work needs to find a solution for completing color information in saturated/under saturated areas.

5. Conclusions

This paper presents an ongoing work for image registration of image sequences for auto-stereoscopic 3D HDR content creation. We propose to modify the PatchMatch algorithm to make it more robust and suitable for our problem. The random search function has been modified and some similarity measures are used. We discussed about such modifications and presented some promising partial results.

Figure 6: On the left, a set of LDR multi-view images, on the right the resulting tone mapped HDR taking each LDR as reference respectively.

6. Acknowledgments

This work was funded through several projects. It is supported by the Spanish Ministry of Science and Innovation Sub-Program Ramón y Cajal RYC-2011-09372 and the TIN2010-20590-C02-02 project from the Ministry Science and Innovation (MCI) of Spain. Collaboration was permitted through the short term scientific mission (STSM) grant awarded by the Cost Action IC1005. Thanks a lot to Jennifer Bonnard for her continuous help and to those who have helped us in this work.

References

[BLV*12] BONNARD J., LOSCOS C., VALETTE G., NOURRIT J.-M., LUCAS L.: High-dynamic range video acquisition with a multiview camera. *Optics, Photonics, and Digital Technologies for Multimedia Applications II* (2012), 84360A–84360A–11.

[BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A., GOLDMAN D. B.: Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. SIGGRAPH) 28*, 3 (aug 2009).

[BSGF10] BARNES C., SHECHTMAN E., GOLDMAN D. B., FINKELSTEIN A.: The generalized PatchMatch correspondence algorithm. In *European Conference on Computer Vision* (sep 2010).

[DM97] DEBEVEC P., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *In proceedings of ACM SIGGRAPH (Computer Graphics)* (1997), vol. 31, pp. 369–378.

[HS09] HIRSCHMULLER H., SCHARSTEIN D.: Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence 31*, 9 (2009), 1582–1599.

[KGBU13] KRONANDER J., GUSTAVSON S., BONNET G., UNGER J.: Unified hdr reconstruction from raw cfa data. *IEEE International Conference on Computational Photography (ICCP)* (2013).

[LC09] LIN H.-Y., CHANG W.-Z.: High dynamic range imaging for stereoscopic scene representation. In *Image Processing (ICIP), 2009 16th IEEE International Conference on* (2009), pp. 4305–4308.

[MG10] MANGIAT S., GIBSON J.: High dynamic range video with ghost removal. 779812–779812–8.

[NPR10] NIQUIN C., PÈVOST S., REMION Y.: An occlusion approach with consistency constraint for multiscopic depth extraction. *International Journal of Digital Multimedia Broadcasting* (2010).

[OMLV12] OROZCO R. R., MARTIN I., LOSCOS C., VASQUEZ P.-P.: Full high-dynamic range images for dynamic scenes. *Optics, Photonics, and Digital Technologies for Multimedia Applications II 8436*, 1 (2012), 843609.

[PcPD*10] PRÉVOTEAU J., CON PIOTIN S. C., DEBONS D., LUCAS L., REMION Y.: Multi-view shooting geometry for multiscopic rendering with controlled distortion. *International Journal of Digital Multimedia Broadcasting (IJDMB), special issue Advances in 3DTV: Theory and Practice 2010* (Mar. 2010), 1–11.

[Ruf11] RUFENACHT D.: *Stereoscopic High Dynamic Range Video*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, Agost 2011.

[SCSI08] SIMAKOV D., CASPI Y., SHECHTMAN E., IRANI M.: Summarizing visual data using bidirectional similarity. *IEEE Conference on Computer Vision and Pattern Recognition 2008 (CVPR'08)* (2008).

[SKY*12] SEN P., KALANTARI N. K., YAESOUBI M., DARABI S., GOLDMAN D. B., SHECHTMAN E.: Robust patch-based HDR reconstruction of dynamic scenes. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2012) 31*, 6 (Nov. 2012), 203:1–203:11.

[SMW10] SUN N., MANSOUR H., WARD R.: Hdr image construction from multi-exposed stereo ldr images. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (Hong Kong, 2010).

[ST04] SAND P., TELLER S.: Video matching. *ACM Transactions on Graphics 23*, 3 (Aug. 2004), 592–599.

[TKTS11] TOCCI M. D., KISER C., TOCCI N., SEN P.: A versatile HDR video production system. *ACM Transactions on Graphics 30*, 4 (July 2011), 41:1–41:10.

[VLS*06] VAISH V., LEVOY M., SZELISKI R., ZITNICK C. L., KANG S. B.: Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2* (Washington, DC, USA, 2006), CVPR '06, IEEE Computer Society, pp. 2331–2338.

[War03] WARD G.: Fast, robust image registration for compositing high dynamic range photographs from handheld exposures. *Journal of Graphics Tools 8* (2003), 17–30.