Journées de l'Association Française d'Informatique Graphique, Reims Image 2014

Génération de séquences d'images multivues HDR: vers la vidéo HDR

HDR multiview image sequence generation: toward 3D HDR video

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

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

Résumé

La génération d'images " high-dynamic range " (HDR) des scènes statiques en combinant plusieurs images " low dynamic range " (LDR) est une procédure maintenant acceptée. Pourtant, la recherche sur la création et la manipulation de contenu vidéo HDR 3D est très active car c'est encore un problème non résolu. Ce travail analyse les avancées récentes en imagerie 3D-HDR et propose une méthode pour construire des séquences 3D HDR à partir de données LDR acquises par une caméra multivue. Notre méthode est basée sur un algorithme basé "patch" (zones) adapté pour s'appuyer sur les contraintes épipolaires des caméras multivues pour obtenir l'alignement des pixels nécessaire à la génération d'images HDR. Elle ne s'appuie pas sur l'alignement stéréo traditionnel basé disparité et n'a pas non plus besoin d'une calibration géométrique exacte des objectifs des caméras. Pour des données d'entrée de 8 images LDR acquises par un équipement dédié, nos résultats expérimentaux produisent un alignement suffisant pour la génération de 8 images HDR, sur lesquels un algorithme de tone mapping peut être appliqué pour une sortie sur un écran autostéréoscopique.

Creating High Dynamic Range (HDR) images of static scenes combining several Low Dynamic Range (LDR) images is a common procedure nowadays. However, 3D HDR video content creation and management is an active, unsolved research topic. This work analyze the latest advances in 3D HDR imaging and proposes a method to build Stereo HDR images from LDR input image sequences acquired with a multi-view camera. Our method is based on the Patch Match algorithm which has been adapted to take advantage of epipolar geometry constraints of multi-view cameras. Our method does not require the traditional stereo matching for disparity map calculation to obtain accurate matching between the stereo images. Geometric calibration is not required either. We use an 8-view LDR camera from which we generate an 8-view HDR output. The eight tonnemapped HDR images are used to feed an auto-stereoscopic display. Experimental results show accurate registration and HDR reconstruction for each LDR view.

Mots clé : Computational Photography, 3D High Dynamic Range, Multi-Exposure Stereo Matching

1. Introduction

Most of modern digital cameras provide some kind of auto exposure control to determine the correct parameters of aperture, ISO value and shutter speed to cover the widest light intensity range possible for a given scene. But still the majority of cameras available at consumer level have limited sensors, capable of recording only a very small part of light intensities. Camera sensors produce saturated pixels where the amount of light exceeds the range of values allowed and under-exposed pixels when not enough light reach the camera sensor. One of the goals in HDR imaging is to enlarge the range of light recorded of a scene.

The most extended way to acquire HDR images is combining images with different exposures of the same scene. There are specific HDR sensors already in the market but too expensive for their expansion at user level. There are different approaches to combine information from a sequence of Low Dynamic Range (LDR) exposures [MP95, DM97, IEE99, MN99, RBS03]. Ghosting effects may appear in the output HDR image when the pixels of source images are not perfectly aligned. There are two main causes for ghosting : camera movement during acquisition and the objects moving in the scene. There are several solutions for image alignment, many of them explained in a detailed survey Zitová [ZF03]. However, these methods were created for images captured under the same lighting conditions and the transition to differently exposed images is not straightforward.

Some 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. One of the main drawbacks for HDR video acquisition is the lack of robust methods for deghosting HDR images resulting from multi-exposed sequences of dynamic scenes. Both Hadziabdic *et al.* [HTM13] and Srikantha *et al.* [SS12] provide good reviews and comparisons between up to date methods for HDR deghosting.

The creation of high dynamic range (HDR) content has been an intense field of research lately, not only for digital photography but for a wide range of imaging applications. Stereoscopic 3D images and video are also important goals for the HDR research community. Most work on 3D HDR are based on combining two stereo exposures.

Bonnard *et al.* [BLV*12] propose a methodology to create content that combines depth (3D) and high dynamic range video for auto-stereoscopic displays. They use an Octo-cam [PcPD*10] which is a camera prototype with eight objectives synchronized in time that produces pairs of four different exposures of the same scene (Figure 1). 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.

We propose a solution to combine stereoscopic LDR images into HDR using image correspondences 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 take advantage of geometric constraints in the set of images to help the matching 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. No geometric calibration or disparity map calculation is needed while using Patch Match.

The rest of the paper is organized as follow. Section 2 focuses on giving a background about the state of the art on 3D HDR video, section 3 describes the solution we are developing and finally section 4 shows some experimental results and comparisons.

2. Background

2.1. Stereo HDR Acquisition

Several prototypes have been proposed to acquire HDR videos from multi-exposure sequence of images. Most of approaches [TKS06, LC09, SMW10, Ruf11, THM13, BRG*14, TCJHM14] are based on a rig of two cameras placed like a conventional stereo configuration. They are focused on finding accurate correspondences to generate HDR images per each view. Troccoli *et al.* [TKS06] proposed to use cross correlation stereo matching to get a primary disparity match. The correspondences are used to calculate the camera response function (CRF) to convert pixels values to radiance space. Stereo matching is executed again, this time in radiance space to extract one depth maps per view.

Lin and Chang [LC09] use SIFT descriptors to find correspondences. Afterward they select the best correspondences using epipolar constrains and use them calculate the CRF. On radiance space they perform a stereo matching algorithm based on belief propagation to derive the disparity map. A ghost removal technique is used to avoid artifacts due to noise or stereo mismatches. Rüfenacht [Ruf11] presents two different approaches to obtain stereoscopic HDR video content. The first is the temporal approach, where different exposures are captured by temporally changing the exposure times of both cameras at a time, recording two frames of the same exposure in each shot alternating the exposure time between consecutive pairs of frames. The second is called '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.

Akhavan *et al.* [THM13, TCJHM14] study different ways to obtain disparity maps from HDR, LDR and tonemapped images and compare them. Selmanovic *et al.* [SDBRC14] propose to generate Stereo HDR video from a pair HDR-LDR, using an HDR camera and a traditional digital camera. In this case, an HDR view needs to be reconstructed. Three methods were proposed : (1) generate the HDR frame by warping the existing one using a disparity map,(2) increase the range of the LDR view using an expansion operator and (3) an hybrid of the two which, according to their comparison, provides the best results. Bätz *et al.* [BRG*14] present a framework with two LDR cameras. The CRF calculation is performed offline. The input images are rectified before the disparity estimation. The stereo matching exposure invariant used Zero-Mean Normalized Cross Correlation (ZNCC) like matching cost. The matching is performed on the gray-scale radiance space image followed by a local optimization and disparities refinement.

Stereo HDR acquisition has been addressed with many different approaches in the last decade. However, multi-view HDR video (more than two views) has not yet been addressed except by Bonnard *et al.* [BLV*12]. All existing methods rely on accurate stereo matching to obtain the disparity maps.They require accurate geometric calibration to ensure that the epipolar geometry conditions are satisfied.

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 transfor-

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

mations that takes place are slightly 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 adn Gibson [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 *et al.* [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 artifacts may appears when there are large under or over exposed areas in the reference image.

3. Patch-based Stereo HDR Generation

3.1. The Patch Match algorithm

Stereo matching is a mature research field ; very accurate algorithms are available for images taken under the same lighting conditions and same exposure. However, most of algorithms are not fully accurate for images with important radiometric variations. We propose a framework based on a variation of the work from Barnes*et al.* [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. Our method does not requires stereo matching for disparity maps calculation. Calibration can be also avoided since the matching process does not requires a perfect epipolar geometry. 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 adjacent 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 the original paper). After NNFs are calculated, a distance metric proposed by Simakov [SCSI08] is used to guarantee both the coherence and completeness of the output image.

3.2. Stereo HDR Generation

Most of current auto-stereoscopic displays accept from five up to nine different stereo views [LLR13]. The camera used to capture the input sequence is the Octo-cam [PcPD*10], a multiview camera prototype composed by eight objectives disposed horizontally (Figure 2). The eight sensors are synchronized in time and they produce four pairs of different exposures of the same scene like shown in the Figure 1. The different exposures are achieved by mean of neutral density filters in each objective [BLV*12]. Filters fix the percentage of light arriving to each sensor dividing by two the amount of light received by each sensor. This is equivalent to keep constant the aperture and divide by two the exposure time (1 F-stop) in a conventional digital camera. 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.

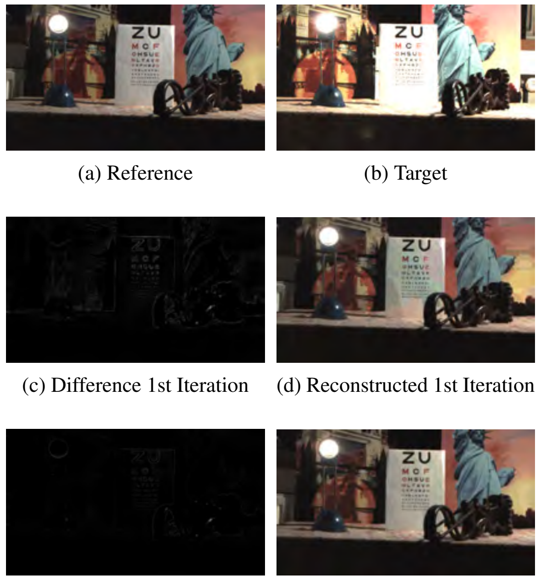
There are some geometrical features that could help to reduce the number of potential corresponding patches. Since all the objectives of the camera are aligned horizontally, the resulting images have a different perspective but pixels in different images of the sequence share the same epipolar line. This fact reduces the random search to 1D (only on the epipolar line). 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 miscalibration of images.

Figure 2: The Octo-cam multiview Camera prototype.

4. Experimental Results

This section shows some experimental results of our proposal. We have tested our approach using different sequences of input images acquired both with the camera described by Prevoteau *et al.* [PcPD*10] and images from the Middlebury Stereo Dataset [vis06]. Images from Middlebury are 8-bits .png files while from the octocam we get linear raw 10 bit per pixel images. From Middlebury sets we have only 7 different views and from the Octo-cam, 8 views.

Our tests execute matching at two different stages of the procedure : in LDR or in radiance domain. Figure 3 and Figure 4 show the result of matching directly the images with different exposures. For tests of matching the radiance domain we calculated the camera response curve using RANSAC's method [MN99] to transform 8 bits images from Middlebury. Despite that the matching is successful in both cases, matching radiance domain converges much faster. Figure 3 shows the convergence of the PatchMatch used on two different exposed images from the sequence of Figure 1



(a) Reference (b) Target

(c) Difference 1st Iteration (d) Reconstructed 1st Iteration

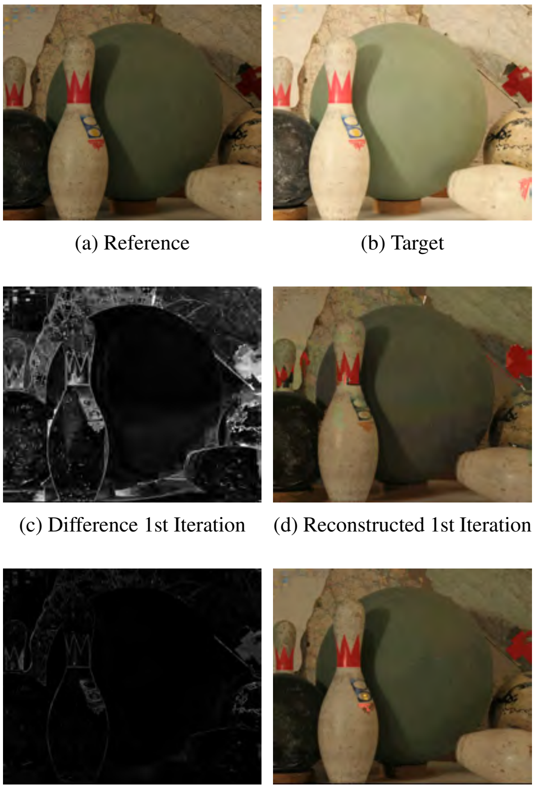
(e) Difference 5th Iteration (f) Reconstructed 5th Iteration

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

The two input images in Figure 4 are from the Middlebury stereo dataset. In this test the matching was made on the original 8-bits LDR image with the same patch size than described above. This 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). 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.

Figure 6 shows a full test over a set of seven images from the Middlebury datased mentioned before. The resulting HDR images have no visible ghosting effects. As we can see in Figure 7, matching is not correct in very saturated areas. This is a problem also for Sen *et al.* [SKY*12]. This happens because there is no information in the reference image to compare with. For instance, letters are not possible to recover for the lasts images in the sequence since the whole sheet of paper is totally saturated. In future approaches it is needed to address such problem.

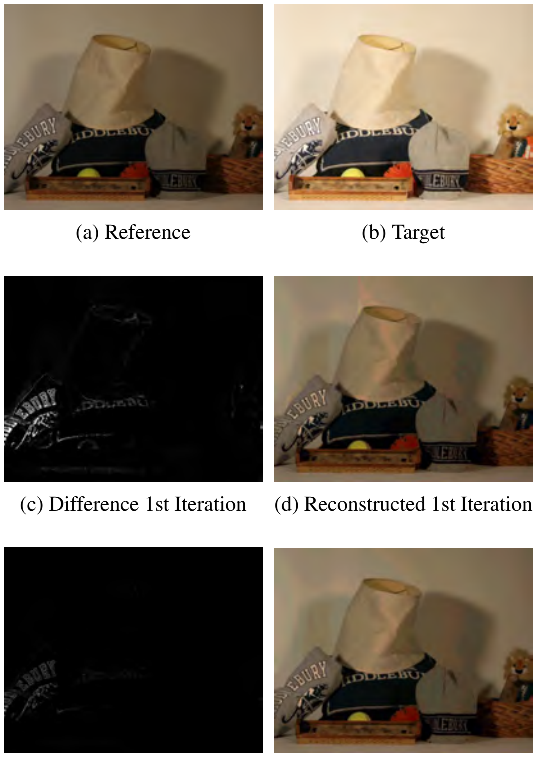


(a) Reference (b) Target

(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



(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

5. Conclusions

This paper presents work in progress 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 stereo matching. The random search function has been modified to take advantage of the epipolar geometry. Our method for stereo HDR generation does not require disparity map calculation or geometric calibration. We discussed about such modifications and presented some promising experimental results.

6. Acknowledgments

This work was partially financed by the Spanish Ministry of Science of Spain through the Innovation Sub-Program Ramón y Cajal RYC-2011-09372. It is also supported by the EU COST HDRi IC1005. Thanks a lot to Jennifer Bonnard for her continuous help and to those who have helped us in this work.

Références

[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.

[BRG*14] Bätz M., Richter T., Garbas J.-U., Papst A., Seiler J., Kaup A. : High dynamic range video reconstruction from a stereo camera setup. *Signal Processing : Image Communication. Vol. 29*, Num. 2 (2014), 191 – 202. Special Issue on Advances in High Dynamic Range Video Research.

[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). Vol. 28*, Num. 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.

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

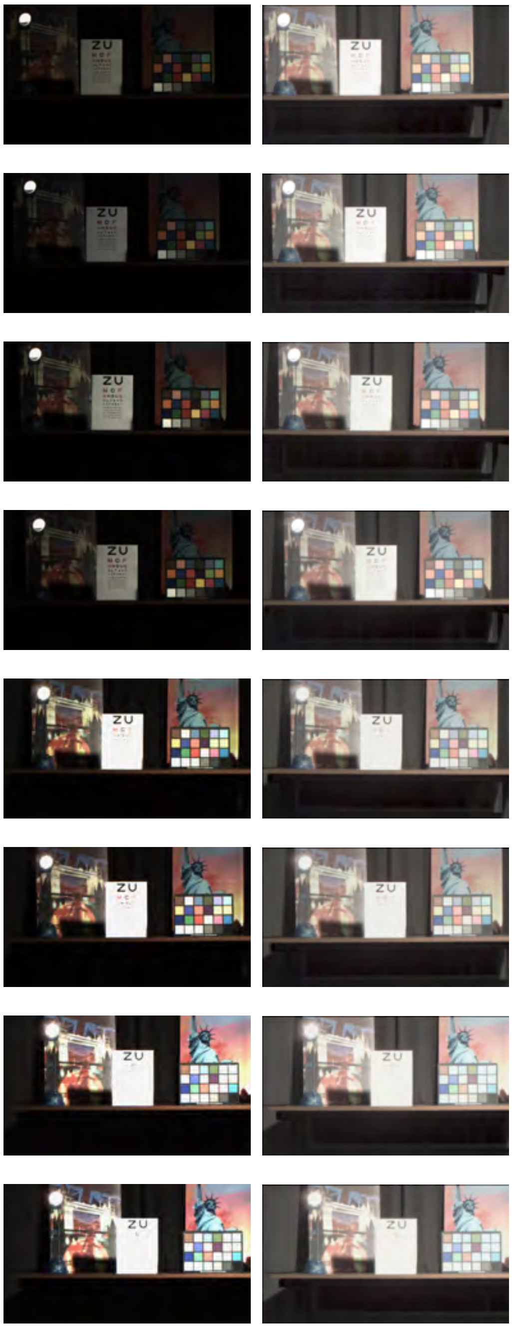


Figure 7: On the left hand side, a set of LDR 10 bits multi-view images obtained from the Octo-cam, on the right hand side the resulting tone mapped HDR.

[HTM13] HADZIABDIC K. K., TELALOVIC J. H., MANTIUK R. : Comparison of deghosting algorithms for multi-exposure high dynamic range imaging. In *Proceedings of the 29th Spring Conference on Computer Graphics* (New York, NY, USA, 2013), SCCG '13, ACM, pp. 021 :21–021 :28.

[IEE99] IEEE :. *Dynamic range improvement through multiple exposures* (1999), IEEE.

[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.

[LLR13] LUCAS L., LOSCOS C., REMION Y. : *3D Video from Capture to Diffusion*. Wiley-ISTE, octobre 2013.

[MG10] MANGIAT S., GIBSON J. : High dynamic range video with ghost removal. *Proc. SPIE. Vol. 7798* (2010), 779812–779812–8.

[MN99] MITSUNAGA T., NAYAR S. : Radiometric self calibration. *IEEE International Conf. Computer Vision and Pattern Recognition. Vol. 1* (1999), 374–380.

[MP95] MANN S., PICARD R. W. : *On Being undigital With Digital Cameras : Extending Dynamic Range By Combining Differently Exposed Pictures*. Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology, 1995.

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

[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. Vol. 2010* (mars 2010), 1–11.

[RBS03] ROBERTSON M. A., BORMAN S., STEVENSON R. L. : Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging. Vol. 12*, Num. 2 (2003), 219–228.

[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).

[SDBRC14] SELMANOVIC E., DEBATTISTA K., BASHFORD-ROGERS T., CHALMERS A. : Enabling stereoscopic high dynamic range video. *Signal Processing : Image Communication. Vol. 29*, Num. 2 (2014), 216 – 228. Special Issue on Advances in High Dynamic Range Video Research.

[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). Vol. 31*, Num. 6 (novembre 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).

[SS12] SRIKANTHA A., SIDIBÉ DÉSIRÉ D. : Ghost Detection and Removal for High Dynamic Range Images : Recent Advances. *Signal Processing : Image Communication* (février 2012), 10.1016/j.image.2012.02.001. 23 pages.

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

[TCJHM14] TARA A., CHRISTIAN K., JI-HO C., MARGRIT G. : Stereo hdr disparity map computation using structured light. In *HDRi2014 Second International Conference and SME Workshop on HDR imaging* (2014).

[THM13] TARA A., HYUNJIN Y., MARGRIT G. : A framework for hdr stereo matching using multi-exposed images. In *Proceedings of HDRi2013 First International Conference and SME Workshop on HDR imaging* (Paper no. 8, Oxford/Malden, 2013), The Eurographics Association and Blackwell Publishing Ltd.

[TKS06] TROCCOLI A., KANG S. B., SEITZ S. : Multi-view multi-exposure stereo. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on* (June 2006), pp. 861–868.

[vis06] VISION.MIDDLEBURY.EDU : Middlebury stereo datasets. http ://vision.middlebury.edu/stereo/data/, 2006.

[ZF03] ZITOVA B., FLUSSER J. : Image registration methods : a survey. *Image and Vision Computing. Vol. 21* (2003), 977–1000.