Dense Disparity Estimation from Omnidirectional Images

Zafer Arican   Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute
Lausanne, Switzerland

Abstract

This paper addresses the problem of dense estimation of disparities between omnidirectional images, in a spherical framework. Omnidirectional imaging certainly represents important advantages for the representation and processing of the plenoptic function in 3D scenes for applications in localization, or depth estimation for example. In this context, we propose to perform disparity estimation directly in a spherical framework, in order to avoid discrepancies due to inexact projections of omnidirectional images onto planes. We first perform rectification of the omnidirectional images in the spherical domain. Then we develop a global energy minimization algorithm based on the graph-cut algorithm, in order to perform disparity estimation on the sphere. Experimental results show that the proposed algorithm outperforms typical methods as the ones based on block matching, for both a simple synthetic scene, and complex natural scenes. The proposed method shows promising performances for dense disparity estimation and can be extended efficiently to networks of several camera sensors.

1 Introduction

Dense disparity estimation on omnidirectional images has become a part of localization, navigation and obstacle avoidance research. Omnidirectional imaging certainly represents a great advantage compared to classical projective camera models in terms of accuracy and efficiency in the representation of scenes. When several cameras capture the same scene, the geometry of the scene can be estimated by comparing the images from the different sensors. The differences between the respective positions of a 3D point on multiple 2D images represent disparities that can be estimated by stereo matching methods. Disparities computed for each pixel form dense disparity maps, which are in general well studied for standard cameras where fast algorithms have been developed. These fast estimation methods are generally based on block matching and pixel correlation due to their simplicity and speed (e.g., [1, 2]). Alternatively, several works have proposed the use of global energy minimization methods for dense disparity estimation on standard images, and such approaches have become quite popular with fast and accurate algorithms based on graph-cut [3] or belief propagation [4]. These methods converge either to global minimum or to strong local minima in polynomial time, and become therefore suitable for realtime applications.

The strategies that have been developed for dense disparity estimation from standard camera images are also applicable to omnidirectional images. However, dense stereo matching on omnidirectional camera images is mainly limited to planar omnidirectional images. The algorithms are generally based on re-projection of omnidirectional images on simpler manifolds. For example, Takiguchi *et al* [1] re-project omnidirectional images onto cylinders, while Gonzalez-Barbosa *et al* [2] and Geyer *et al* [5] rectify omnidirectional images on the rectangular grid. Both cylindrical and rectangular projections are not sufficient to represent spatial neighborhood and correlations among the pixels. However, the equiangular grid on the sphere has a better representation for the omnidirectional images and it is possible to map omnidirectional images onto the 2D sphere by inverse stereographic projection, as shown in [6].

In this paper, we propose to address the problems of stereo and dense disparity estimation from omnidirectional images into the spherical framework. We exploit the advantages of the spherical framework for the representation of 3D scenes, and the processing of the plenoptic information. We first perform rectification of omnidirectional images in the spherical domain. Then we extend a global energy minimization method based on the graph-cut algorithm to spherical images by considering the typical characteristics of images on the 2D sphere. In that respect, the work that is maybe the closest to the approach proposed in this paper has been performed by Fleck *et al* [7], where 3D modeling is computed from omnidirectional images using a graph-cut algorithm. However, the graph-cut algorithm is applied on planar omnidirectional images, along with post-processing for refinement of the disparity estimation. Projections on

planar images certainly introduce discrepancies that we propose to avoid by working directly in the spherical domain. We process disparities directly on the 2D sphere in order to better exploit the geometry of omnidirectional images and to improve the accuracy of the estimation. Experimental results show that the proposed method provides efficient disparity estimation, for synthetic and natural scenes. In addition, the proposed algorithm can be extended to networks of multiple camera sensors, where the accuracy of the scene modeling can yet be improved.

The paper is structured as follows. Section 2 describes the rectification of omnidirectional images in the spherical domain. Section 3 formulates the dense disparity estimation problem, and describes the solution based on the graph-cut algorithm. Experimental results are presented in Section 4, and the extension of the algorithm to networks of three cameras is presented in Section 5.

2 Rectification of Spherical Images

Rectification is an important step in stereo vision using standard camera images. It aims at reducing the stereo correspondence estimation to a one-dimensional search problem. It basically consists in image warping, which is computed such that epipolar lines coincide with the scan lines. It does not only facilitate the implementation of disparity estimation algorithms, but also makes computations faster.
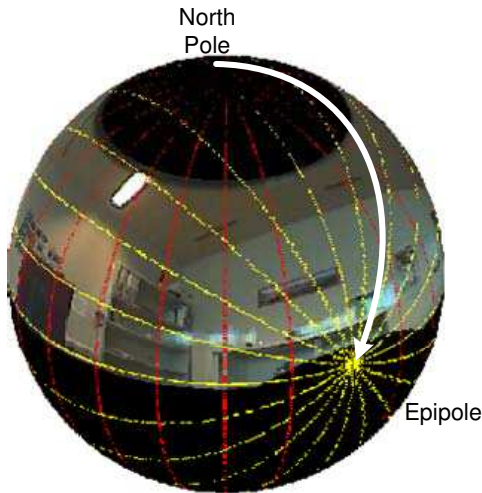

Figure 1: Relation between longitudes and epipolar great circles and corresponding rotation.

In the spherical framework, it is also possible to perform rectification on spherical images. We simply use the following observation about epipoles on spherical images: (i) epipoles resemble the coordinate poles and (ii) epipolar great circles intersecting on epipoles are like longitude circles. We therefore rotate spherical image pairs in the spherical domain such that epipoles coincide with the coordinate poles. In this way, epipolar great circles coincide with the longitudes and disparity estimation becomes a mono-dimensional problem.

Figure 1 illustrates the rectification strategy used in this paper. In order to be able to display the spherical images easily, as well as to use the computation algorithms that have been optimized for rectangular images, we represent the spherical images as rectangular images with latitude and longitude angles as axes. Figure 2 shows original and rectified images in this representation.

At the end of the rectification step, two rectified stereo spherical images are obtained in the form of rectangular images on a equiangular grid. It permits to extend the disparity estimation algorithms developed for standard images to spherical images, and thus perform fast computations.
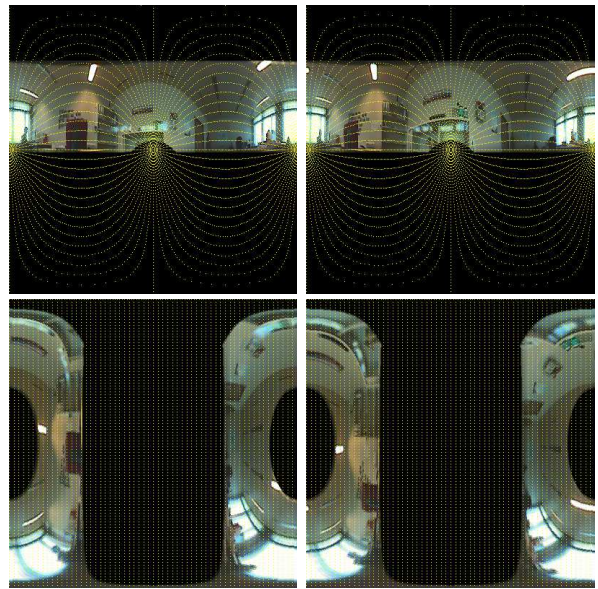

Figure 2: Original images (top), and rectified ones (bottom). Epipolar great circles turn into straight vertical lines in rectified images.

3 Dense Disparity Estimation

In the spherical framework, disparity can be defined as the difference in angle values between the representation of the same 3D point, in two different omnidirectional images. Since pixel coordinates are defined with angles, we define the disparity γ as the difference between the angles corresponding to pixel coordinates on the two images, i.e., $\gamma = \theta - \phi$,
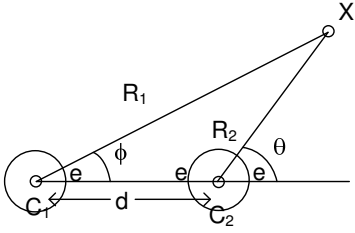
Figure 3: 2D representation of the geometry between cameras and the 3D point

Figure 3 shows the 2D representation of the geometry between cameras and the 3D point. The depth $R1$ is defined as the distance of the 3D point to the reference camera center. The relation between the disparity γ, the depth $R1$ and baseline distance d, is given as

$$\gamma = arcsin\frac{dsin\theta}{R1} \qquad (1)$$

This relation holds for all epipolar great circles on rectified stereo images.

We now describe the disparity estimation problem, as an energy minimization problem. Let $\mathcal{P}$ denote the set of all pixels on two rectified spherical images. Let further $\mathcal{L} = \{l_1, l_2, \ldots, l_{max}\}$ represent the finite set of discrete labels corresponding to disparity values. Recall that the disparity estimation problem is mono-dimensional, due to the image rectification step. A single value per pixel is therefore sufficient to represent the disparity, and the disparity values for each pixel together form the disparity map. If $f : \mathcal{P} \rightarrow \mathcal{L}$ is a mapping so that each pixel is assigned a disparity label, our aim is to find the optimum mapping f^* such that the disparity map is as accurate and smooth as possible.

The computation of the optimum mapping f^* can be formulated as an energy minimization problem, where the energy function $E(f)$ is built on two components E_d and E_σ, which respectively represent the data and smoothness functions :

$$E(f) = E_d(f) + E_\sigma(f) \qquad (2)$$

The data function first reports the photo-consistency between the omnidirectional images. It can be written as :

$$E_d(f) = \sum_{(p,q)\in\mathcal{P}^2} D(p, q), \qquad (3)$$

where p and q are corresponding pixels in two images under a mapping function f. $D(.,.)$ is non-positive cost function, which can be expressed as :

$$D(p, q) = \min\{0, (I(p) - I(q))^2 - K\}, \qquad (4)$$

where $I(i)$ represents the intensity or luminance of pixel i and K is a positive constant. In this work, we define the intensity $I(i)$ in a way that is similar to the function proposed in Birchfield and Tomasi's work [8]. It presents the advantage to be insensitive to image sampling, which is quite useful since the equiangular grid on the sphere causes non-uniform sampling.

The smoothness function then captures the variations of the disparity between neighboring labels. The goal of the smoothness function is to penalize the estimation of labels that are different from their neighborhood in order to obtain a smooth disparity field. The neighborhood $\mathcal{N}$ is generally represented by the 4 surrounding labels. The smoothness function under a mapping f can be expressed as :

$$E_\sigma(f) = \sum_{(p,q)\in\mathcal{N}} V_{p,q}(f(p), f(q)). \qquad (5)$$

The term $V_{p,q} = \min\{|l_p, l_q|, K\}$ is a distance metric that reports the difference between labels attributed to neighboring pixels in $\mathcal{N}$. We use in this work the distance function proposed in [9].

The dense disparity estimation problem now consists in minimizing the energy function $E(f)$, in order to obtain an accurate and smooth disparity map. Such global energy minimization problems typically find solutions in algorithms based on graph-cut, or belief propagation methods. In this paper, we choose to minimize the function $E(f)$ with a graph-cut algorithm, which has been adapted to the specificities of the spherical framework. Basically, the graph-cut algorithm converts the energy minimization problem into several minimum cut problems. A binary graph is constructed on the pixels and labels. Since the graph-cut algorithm is optimum in binary graphs, multiple labels are processed iteratively by the construction of successive binary graphs for each label.

Two terminal vertices called source and sink represent the binary values for a label. That is, the source vertex represents the label α and the sink one represents the label α^-. The pixels represent the other vertices. There are edges between terminal vertices and pixel vertices to represent the data cost E_d and edges between neighboring pixels to represent the smoothness cost E_σ. These edges are weighted by the value of the corresponding cost functions. A cut is then performed through the edges and the total cost is the sum of the weights of the edges affected by the cut. The optimization problem is to find the minimum cut with minimum total cost, which corresponds to the minimum of the energy $E(f)$. Finally, we use the α-expansion method to find the minimal cut [9, 10]. In the experiments, we use the graph-cut algorithm based on the implementation proposed in [9]. We have adapted it to the spherical framework by taking into account the particular connectivity of the image boundaries.

3

4 Experimental Results

This section evaluates the performance of the dense disparity estimation algorithm for both synthetic and real images. We have first generated two synthetic spherical images with the Blender[11] and Yafray[12] programs. The constructed scene consists of 4 planes with distances 7,7,8 and 10 units from the cameras. Figure 4 shows the left and right images, and the corresponding synthetic scene with a top view. Cameras are placed 1 unit of distance apart, with no rotation. With these parameters, epipoles are located at $90°$ of latitude, $0°$ and $180°$ longitude angles for both cameras.
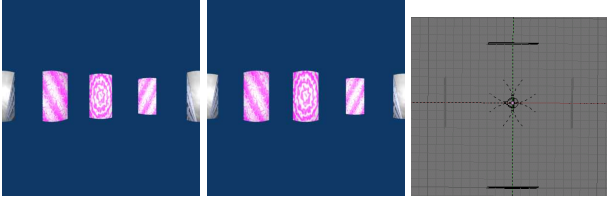

Figure 4: Left and right images, and top view of the synthetic scene.

For real images, we used a catadioptric system with a paraboloidal mirror. The two images of a room captured from two different positions are shown in Figure 5. We map the images on the 2D sphere via inverse stereographic projection [6]. Images mapped on the sphere and rectified images are shown in Figure 2.
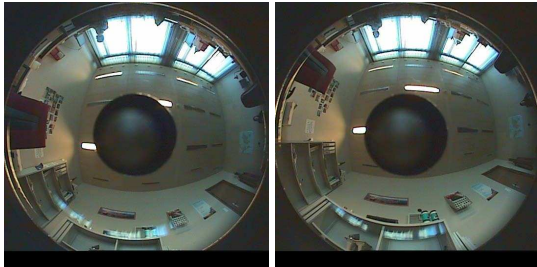

Figure 5: Captured omnidirectional images.

We compare the performance of the proposed algorithm (GC) with a block matching algorithm (WTA) based on the winner-gets-all principle. It consists in a local optimization method that tries to find the best match between blocks of pixels, based on the highest correlation score. Figure 6 shows the disparity maps computed by the proposed algorithm, and the disparity map obtained by the block matching algorithm. It can be seen that the disparity estimation is more precise in the GC method than in the WTA algorithm, which is not able to provide a very dense disparity map.

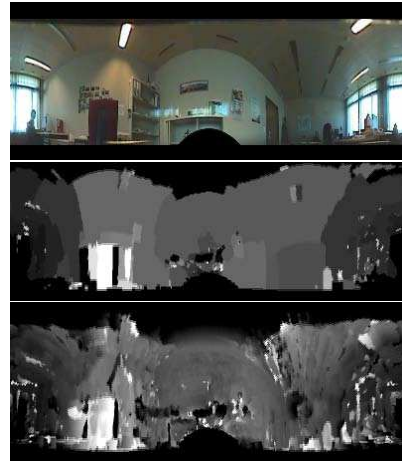In order to analyze further the performance of the proposed algorithm, we warp the left image using the disparity maps to get the right image. Figure 7 shows the ground truth right image, along with the results of warping based on the disparity maps obtained respectively with WTA and GC. The images are back projected onto a plane for better representation. We can see that GC clearly allows to obtain a better estimate of the right image. This is confirmed by the MSE distortion computed on the right image. The WTA and GC algorithm respectively results in a distortion of 1973 and 1122 MSE points. Overall, the GC algorithm gives better performance, both visually and quantitatively.

Figure 6: Reference image and disparity images calculated with GC and WTA methods respectively for the real image set.
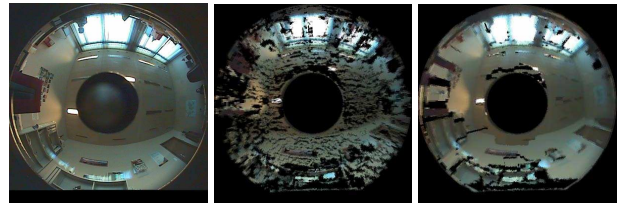

Figure 7: Right image, and warped images using WTA and GC respectively

The lack of precision around the epipoles prevents the GC algorithm to achieve better results due to the global optimization characteristics of the graph-cut algorithm. We therefore run a partial GC algorithm for the central regions of the omnidirectional images. Figure 8 shows the ground truth image cropped to the central region, and the same image after warping of the left image based on the disparity maps computed by GC, and Partial GC. The distortion in the central region corresponds to 486 and 401 MSE points respectively, for the GC and Partial GC algorithms. This shows that removing the epipolar regions prior to disparity estimation, gives better results than post-processing the

image to remove epipolar regions as performed in [7].

Figure 8: Right image cropped to the central region, warped images for partial GC and cropped warped image for full GC

Finally, we analyze the performance of the disparity estimation method by computing the distortion on depth values. As the ground truth depth values are not available for the real images, we calculate the depth values from the disparity maps on the synthetic images, and we compute the MSE with respect to the ground truth. Figure 9 shows the ground truth depth, along with the error on depth when the WTA and GC methods are used. The distortion respectively corresponds to 10.21 and 7.13 MSE points. The GC algorithm is therefore more accurate than the WTA solution for depth estimation, as expected from the above results. Finally, when the distortion is calculated only on the central regions to avoid the inaccuracies around epipoles, the performance is respectively of 50.97 and 1.25 MSE points for WTA and GC algorithms.
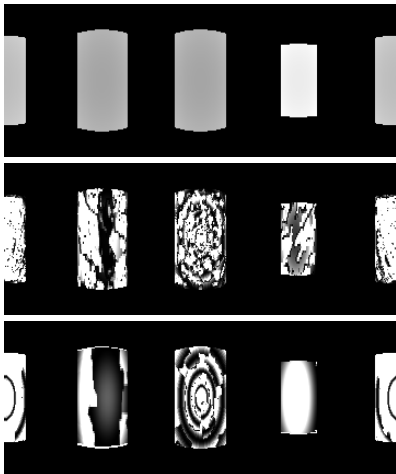

Figure 9: Ground truth depth values and scaled depth estimation error with WTA and GC algorithms. Note that all values bigger than 1 are marked with white. Max error is 76.82 for WTA while 16.38 for GC.

5   Depth Estimation with 3 cameras

This section extends the proposed algorithm to scenarios with three cameras. It can be noted that the addition of one extra camera is expected to overcome the problem of inaccuracies around the epipolar regions that appears when only two images are used for disparity estimation. Unless the three cameras are co-linear, an area that lies in the epipolar region for one camera pair, stays inside the central region for an other camera pair.
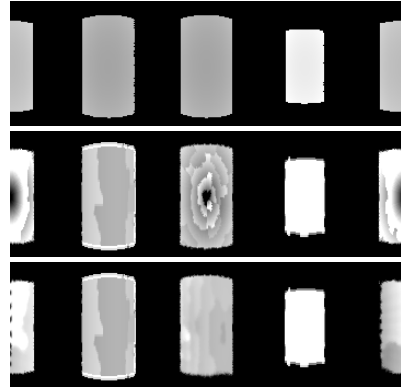

Figure 10: Ground truth distances, and distances obtained with GC on two and three cameras.

In order to demonstrate the benefit of additional cameras, we generate three images of the synthetic scene described above, by placing the cameras $C1$, $C2$, and $C3$ on vertices of an equilateral triangle. The graph-cut algorithm is applied on two pairs of cameras, respectively, $(C1, C2)$ and $(C2, C3)$. After rectification, the GC algorithm is run in parallel to compute the disparities and distances to cameras $C1$ and $C2$. Low precision regions are filled with the distance values obtained for the camera $C2$. Figure 10 shows that the artifacts on the epipolar regions are effectively removed by the introduction of the third camera. In terms of distortion, the distance estimation for two and three cameras respectively results in 5.11 and 2.78 MSE points, which shows that the distortion is reduced by half due to the presence of a third view. Finally, we have tested the same procedure for a more realistic scene, as illustrated in Figure 11. A clear improvement of the precision around the epipolar regions is observed, and GC outperforms WTA in such a scene too. The distortion on estimated depth values corresponds to 20.72, 16.11 and 36.02 MSE points for GC algorithm with two and three cameras, and the WTA algorithm with two cameras, respectively. For fair comparisons, erroneous excessive distance values are clipped to the maximum ground truth distance. Even if the GC algorithm provides promising results when the number of cameras increases, we have observed that low texture areas still prevent the algorithm to converge to better results.

Figure 11: Synthetic images for $C1$, $C2$ and $C3$ respectively, in the Room scene.

6 Conclusions

We have addressed the problem of dense disparity estimation from omnidirectional images, in the spherical framework. We have adapted a strong energy minimization algorithm based on graph-cut for estimation of disparities on spherical images. A simple yet efficient rectification process based on simple rotation in the spherical framework, combined with an adapted graph-cut algorithm offers promising disparity estimation results. It outperforms a block matching algorithm based on a Winner-takes-all strategy. Problems however appear in the regions close to the epipoles, due to lack of precision in disparity estimation around these regions. The extension of the proposed algorithm with the introduction of additional cameras allows to overcome these problems. Overall, the proposed approach offers promising performance for dense disparity estimation in networks of omnidirectional cameras. We plan to extend the study to natural scenes with multiple cameras, at arbitrary positions. Finally, we will apply the proposed method to localization, or detection problems, in order to show the advantage of omnidirectional imaging in popular applications.

References

[1] J. Takiguchi, M. Yoshida, A. Takeya, J. Eino, and T. Hashizume, "High precision range estimation from an omnidirectional stereo system," *Intelligent Robots and System, 2002. IEEE/RSJ International Conference on*, vol. 1, pp. 263–268.

[2] J. Gonzalez-Barbosa and S. Lacroix, "Fast Dense Panoramic Stereovision," *Robotics and Automation, 2005. Proceedings of the 2005 IEEE International Conference on*, pp. 1210–1215, 2005.

[3] Y. Boykov, O. Veksler, and R. Zabih, "Markov random fields with efficient approximations," *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pp. 648–655, 1998.

[4] J. Sun, H. Shum, and N. Zheng, "Stereo matching using belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 787–800, 2003.

[5] C. Geyer and K. Daniilidis, "Conformal Rectification of Omnidirectional Stereo Pairs," *Omnivis 2003: Workshop on Omnidirectional Vision and Camera Networks*, 2003.

[6] C. Geyer and K. Daniilidis, "Catadioptric Projective Geometry," *International Journal of Computer Vision*, vol. 45, no. 3, pp. 223–243, 2001.

[7] S. Fleck, F. Busch, P. Biber, W. Strasser, and H. Andreasson, "Omnidirectional 3D Modeling on a Mobile Robot using Graph Cuts," *Robotics and Automation, 2005. Proceedings of the 2005 IEEE International Conference on*, pp. 1748–1754, 2005.

[8] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 4, pp. 401–406, 1998.

[9] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," *European Conference on Computer Vision*, vol. 3, pp. 82–96, 2002.

[10] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.

[11] "Blender 2.43." `http://www.blender.org`.

[12] "Yafray 0.09." `http://www.yafray.org`.