3D Modelling of Static Environments Using Multiple Spherical Stereo

Hansung Kim and Adrian Hilton

Centre for Vision, Speech and Signal Processing, University of Surrey
Guildford, GU2 7XH, Surrey, UK
{h.kim,a.hilton}@surrey.ac.uk

Abstract. We propose a 3D modelling method from multiple pairs of spherical stereo images. A static environment is captured as a vertical stereo pair with a rotating line scan camera at multiple locations and depth fields are extracted for each pair using spherical stereo geometry. We propose a new PDE-based stereo matching method which handles occlusion and over-segmentation problem in highly textured regions. In order to avoid cumbersome camera calibration steps, we extract a 3D rigid transform using feature matching between views and fuse all models into one complete mesh. A reliable surface selection algorithm for overlapped surfaces is proposed for merging multiple meshes in order to keep surface details while removing outliers. The performances of the proposed algorithms are evaluated against ground-truth from LIDAR scans.

Keywords: Environment modelling, Spherical stereo, PDE-based disparity estimation, Multiple stereo reconstruction.

1 Introduction

In recent years generating accurate graphical models of environments has been addressed through computer vision techniques. Approaches to the environment modelling can be classified into active methods using range sensors and passive methods using normal camera images. Light Detection and Ranging (LIDAR) is one of the most popular depth ranging techniques [1]. However, there are problems with respect to hardware cost, materials in the environment and temporal/spatial consistency with an imaging sensor. Therefore active sensing methods are used as reference for image-based modelling [2] or manual computer graphics modelling.

On the other hand, passive approaches require a simpler and less expensive setup. They are temporally and spatially consistent with images because they extract depth information from the captured images. There have been many researches into accurate outdoor scene reconstruction from multi-view images [3-5]. Strecha et al. created a benchmarking site for the quantitative evaluation of algorithms against ground-truth by LIDAR scanning [2]. However, the biggest problem of multi-view stereo is the fact that normal cameras provide only a partial description of the surrounding environment. Agarwal et al. reconstructed full 3D model of streets from 150,000 photos on internet using grid computing with 500 cores for 24 hours [6]. This is very

impressive work but it requires higher costs for renting parallel compute resources and data transfer. The second problem is calibration of multiple cameras. Strecha et al. provided accurate calibration data calculated using attached markers on buildings and LIDAR scanning for the data sets [2], but it is sometimes hard to accurately calibrate all cameras in advance.

Instead of using fixed multiple cameras, structure from motion (SfM) uses video sequences from a moving camera [7][8]. The basic idea of the SfM is similar as the multi-view reconstruction but it reconstructs 3D positions and the cameras motions simultaneously by feature tracking. However, the limitation from narrow field of view (FOV) of normal cameras still remains. Pollefeys et al. used 3,000 frames to reconstruct one building and 170,000 frames for a small town [9].

Another way to capture the full 3D space is to use a catadioptric omnidirectional camera or fisheye lens [10, 11]. These approaches only use one CCD to capture the full 3D space so that the resolution of partial images from the full view is too low to recover details of the environment. Instead of using original spherical images, Feldman and Weinshall [12] used a cross-slits projection with a rotating fisheye camera to generate a high quality spherical image. Kim and Hilton extended this to spherical stereo for reconstructing a 3D environment from a stereo pair of high resolution images [13].

There have been researches on combining active and passive sensors for outdoor environment modelling. Boström et al. reconstructed large urban environments using a wide angle laser range finder (LRF) and a calibrated colour CCD camera [14]. Asai et al. also reconstructed wide outdoor areas using an omnidirectional LRF and an omnidirectional multi-camera system which can capture high-resolution images [15].

In this paper, we propose a 3D environment modelling method from multiple pairs of spherical stereo images. We capture a static environment as a vertical stereo pair with a rotating line scan camera at multiple locations. Dense floating-point disparity fields are estimated using a novel PDE-based stereo method giving a 3D reconstruction of the scene for the surfaces visible from each stereo pair. A 3D rigid transform is calculated between views using SURF feature matching [16] and RANSAC algorithm to register the reconstructed models from multiple viewpoints. Finally a complete 3D model of the environment is generated as a single mesh by selecting the most reliable surfaces.

The main contributions of this paper are:

- (1) We propose to use vertical stereo with a rotating line scan camera for scene modelling which does not require calibration steps for stereo reconstruction and model registration. (Section 2.1 and 3.1)
- (2) We extend an existing stereo PDE formulation to handle the occlusion problem in stereo and over-segmentation problem in highly textured regions. (Section 2.2)
- (3) We propose a simple and efficient approach to merge multiple stereo reconstructions into a single model based on selection of the best viewpoint for unoccluded surface regions by considering visibility, surface normal vectors and distance. (Section 3.2 and 3.3)

Fig. 1. Spherical geometry

- (4) We evaluate the accuracy of reconstruction against ground-truth model scanned by a LIDAR sensor. (Section 4)

2 Depth Reconstruction from Spherical Stereo

2.1 Capture System and Spherical Stereo Geometry

We use a line scan camera system which captures a full spherical view from a rotating lens around a vertical axis [13][17]. A spherical image is generated by mosaicing rays from a vertical slit at the centre of a rotating fisheye lens. The maximum resolution of the image is 12574x5658. The scene is captured with the camera at two different heights to recover depth information of the scene through stereo geometry. There are two advantages of using this line scan cameras for stereo imaging as well as acquiring high resolution images. First, the stereo matching can be simplified to a 1D search along the scan line if the two capture points are vertically aligned as shown in Fig. 1 (a), while normal spherical images require complex search along conic curves or rectification of images. Second, calibration for depth reconstruction only requires knowledge of the baseline distance between the stereo image pair. Radial distortion is rectified using a 1D table to evenly map pixels on the vertical central line to $0 \sim 180^\circ$ range. Internal lens distortion parameters are fixed so it can be calculated for the lens in advance.

If we assume the angles of the projection of the point p onto the spherical image pair displaced along the y -axis are θ_t and θ_b , respectively, the angle disparity d of point p can be defined as the difference of the angles as $d = \theta_t - \theta_b$, and the distances of the point p from the two cameras are calculated as follows from the relationship between two cameras in Fig. 1 (b).

$$\begin{aligned}
 r_t &= Bf \left(\frac{\sin \theta_t}{\tan(\theta_t + d)} - \cos \theta_t \right) \\
 r_b &= Bf \left(\cos \theta_b - \frac{\sin \theta_b}{\tan(\theta_b - d)} \right)
 \end{aligned} \tag{1}$$

Fig. 2. Precision in surface reconstruction

2.2 PDE-Based Disparity Estimation

There have been a large number of algorithms proposed to solve the stereo correspondence problem over the last four decades. Scharstein and Szeliski present a test bed for the quantitative evaluation of stereo algorithms [18]. However, most disparity estimation algorithms including graph-cut (GC) and belief-propagation (BP) methods solve the correspondence problem on a discrete domain such as integer, half- or quarter-pixel levels which are not sufficient to recover a smooth surface. Spherical stereo image pairs have relatively small variations in disparity and serious radial distortion because of wide FOV of the fisheye lens. Figure 2 shows the difference in surface reconstructions from integer and floating-point disparity fields. We can see that all surface details have disappeared and it shows stepwise artefact in Fig. 2 (a).

A variational approach which theoretically works on a continuous domain can be a solution for accurate floating-point disparity estimation. In this approach, the disparity vector fields are extracted by minimizing an energy functional involving a fidelity term and a smoothing term such as:

$$E(d) = E_f(d) + E_s(d) = \int_{\Omega} (I_t(x) - I_b(x+d))^2 dx + \lambda \int_{\Omega} \psi(\nabla d, \nabla I_t) dx, \quad (2)$$

where $x \in \Omega$ is an open bounded set of R^2 and d is a 2D disparity vector. The minimization problem can be solved by solving the associated partial differential equation (PDE) in Eq. (3) and (4) with Neumann boundary conditions [19][20].

$$-\nabla E(d) = \lambda \operatorname{div}(g(\nabla I_t) \nabla d) + (I_t(x) - I_b(x+d)) \frac{\partial I_b(x+d)}{\partial x} = 0 \quad (3)$$

$$\frac{\partial d}{\partial t} = \lambda \operatorname{div}(g(\nabla I_t) \nabla d(x)) + (I_t(x) - I_b(x+d)) \frac{\partial I_b(x+d)}{\partial x} \quad (4)$$

This method produces accurate depth fields across most regions, but it has several limitations related to stereo occlusion around depth discontinuities and over-segmentation in highly textured regions.

In Eq. (4), $g(\bullet)$ is a regularisation function which controls the direction and amount of smoothing of the disparity field. Traditional image-driven functions preserve sharp object boundaries but results in over-segmentation in highly textured regions [13][19]. Zimmer et al. proposed a disparity-driven method to avoid this problem [21], but this method tends to blur object boundaries. Sun et al. recently proposed joint image-/flow-driven optical flow based on steerable random fields to obtain sharp object boundaries without over-segmentation [22].

Fig. 3. Disparity fields produced by various diffusivity functions

In terms of the occlusion problem, Ben-Ari and Sochen proposed an iterative method consisting of occlusion detection, disparity estimation and anisotropic filtering [23]. Alvarez et al. proposed a symmetrical dense optical flow energy functional which includes a bi-directional disparity checking term [24]. Ince and Konrad also proposed similar bi-directional disparity checking method, but they put it into the data term [25].

We propose a new PDE which handles occlusions and over-segmentation while preserving sharp object boundaries as follows.

$$\frac{\partial d_t}{\partial t} = \lambda_1 \text{div}(g(\nabla I_t, \nabla d_t) \nabla d_t(x, y)) + H(1 - O(x))(I_t - I_b(x + d_t)) \frac{\partial I_b(x + d_t)}{\partial x} \quad (5)$$

$$g(\nabla I, \nabla d) = \frac{1}{(1 + s(\nabla I, \nabla d))^2} \cdot K_{2\sigma} * (\nabla I_\sigma \nabla I_\sigma^T) \quad (5-1)$$

$$s(\nabla I, \nabla d) = -\ln(0.35 + 0.65e^{-|\nabla d|}) \cdot |\nabla I|^2 \quad (5-2)$$

$$O(x) = |d_t(x) + d_b(x + d_t(x))| \quad (5-3)$$

In the above equations, $I_\sigma = K_\sigma * I$, K_σ denotes a Gaussian kernel with standard deviation σ , $*$ is a convolution operator and $H(\bullet)$ is a unit step function.

The term $K_{2\sigma} * (\nabla I_\sigma \nabla I_\sigma^T)$ in Eq. (5-1) works as a structure tensor for anisotropic diffusion filtering [20] and the first term $-\ln(0.35 + 0.65e^{-|\nabla d|})$ in Eq. (5-2) is a monotonically increasing function and scales the diffusivity according to the gradient of the disparity field. As a result, the diffusivity function of Eq. (5-1) for regularisation is mainly controlled by image gradient but scaled by disparity gradient to avoid over-segmentation in highly textured region. This is simpler and more intuitive than Sun's method [22]. Figure 3 shows a comparison of diffusivity functions. The scene is composed of three slanted planes and the image-driven method cannot regularise the fields enough due to strong textures. The disparity-driven method produces smooth surfaces, but object boundaries are blurred because of diffusion of the field across discontinuities. Compared with the other two methods, the proposed method produces a very smooth field on each plane while keeping sharp object boundaries.

For occlusion handling, we take a compromise between Ben-Ari's work [23] and Ince's work [25]. Ince used a bi-directional disparity matching as a scaling factor for the data term in Eq. (4), but the data term still causes blurred and distorted fields in occluded regions because there is no correspondence for the regions. Ben-Ari detected occlusions by a level-set method and performed disparity estimation only for

Fig. 4. Reconstruction from single spherical stereo pair (Top Captured spherical image, Middle: Reconstructed mesh model, Bottom: Model with texture mapping)

visible regions. They filled up the occlusions by anisotropic filtering as a separate step. This approach produces very good results but it is inefficient to run all steps iteratively. Therefore, we use the bidirectional matching as a switch to turn on/off the data term in Eq. (5). The normal balanced diffusion equation with data term is run to find the optimised solution in visible regions while only pure anisotropic diffusion filtering is performed to propagate correct depth information from visible regions to occluded regions.

Equation (5) can be solved explicitly or semi-implicitly with an iterative method by updating the timely discretised field [19][20]. One mathematical problem is the convergence of the proposed PDE. As the Ince's method could not guarantee the mathematical convergence of an iterative solver with the bi-directional disparity matching term, the switch $H(1-O(x))$ in Eq. (5) can make the solver stuck and resonant between visible and occlusion modes. We set a maximum number of iterations in the solver to avoid being trapped into an eternal loop. However, we have not found this problem in our experiments.

Another problem is a local minimum problem. Alvarez et al. used a scale-space approach [19] and Brox et al. used a warping method [26] to avoid the local minimum problem for large displacements. We also use a similar coarse-to-fine structure which starts from the lowest resolution images and recursively refines the result at higher levels. For expanding multi-resolution images, we utilize a pyramidal approach to construct the L-level hierarchical image structure, which involves low-pass filtering and down-sampling the image by a factor of 2. At each level, the input disparity field from a previous level is up-sampled and used as an initial field for calculating disparity field at that level. At the earliest level, initial disparity is calculated by block matching with a region-dividing technique [27]. This hierarchical approach has another merit of reducing computation time for large images.

3 3D Model Reconstruction

3.1 Single-View Reconstruction

The estimated dense disparity fields can be converted into depth information by camera geometry as described in Section 2.1. A mesh model of the scene is obtained by sampling vertices and triangulating adjacent vertices from the original texture and depth fields. The original images are described in spherical coordinates, so we convert them into the Cartesian coordinate system, and then project them to 3D space to generate a 3D mesh.

Figure 4 shows examples of reconstruction from a spherical stereo image pair. The results show a natural-looking environment around the captured location, but changes in viewpoints cause distortion because of self-occlusion from a spherical image as seen in the circled regions in Fig. 4. There is no way to get information about invisible regions behind any object from a single input image. This occlusion problem occurs not only between objects but even on the same object due to the wide FOV of the fisheye lens. The faces on occluded regions can be removed by thresholding with angles of their normal vectors, but it can damage other parts such as the ground or details on the surfaces. It produces noisy surface with small isolated faces. We found that it looks more natural to remain them as long as they do not conflict with other surfaces in multi-view reconstruction.

In order to overcome the occlusion problem, we need more information of the scene structure, shape and appearance from multiple viewpoints. Merging multiple stereo pairs into a common 3D scene structure is a possible solution.

3.2 Registration of Multiple Stereo Reconstructions

As listed on Strecha's multi-view benchmarking site [2] and the Middlebury benchmarking site [28], a number of multi-view reconstruction algorithms have been developed. However, most of them are not applicable for our data sets because they are focused on optimising surfaces with accurate calibration parameters of normal cameras, while we are using spherical line scan cameras with very simple calibration.

One way to get a complete model is to merge partial meshes into one complete mesh by registration. Iterative closest point (ICP) algorithm is widely used for mesh registration [29]. The ICP algorithm iteratively find the optimized transform matrix $A(R, t)$ minimizing the energy:

$$E(R, t) = \sum_i \sum_j^{N_m, N_d} w_{i,j} \|m_i - (Rp_j + t)\|^2 \quad (6)$$

where N_m and N_d are the number of points in the model set M and data set P , respectively, and $w_{i,j}$ are the weights for a point match.

In order to automate the registration, we use SURF feature matching [16] between views on captured images and used them as 3D matching references by projecting them into 3D space with the estimated depth field. However, these points are not reliable enough to be used for references of ICP algorithms because two possibilities

of errors exist: one from SURF matching error between image pairs with radial distortion and the other from depth error (this is more serious because many features are extracted around depth discontinuity regions which induce stereo occlusions). Therefore we use a RANSAC algorithm to calculate an optimised 3D rigid transform between two meshes excluding outliers. If the SURF matching cannot extract enough features, we can still use the ICP algorithm with manual feature matches.

3.3 Reliable Surface Extraction

The final step is to merge registered meshes into one complete mesh structure with surface refinement because the registered meshes have many overlapped parts and false surfaces from self-occlusion. Poisson reconstruction [30] or range image merging algorithm [31] can be ways to produce single mesh structure from set of oriented points or multiple depth fields, but they may lose details on the original surface because the algorithms generate combined surface from overlapped surfaces. False surfaces from self-occlusion also induce errors in optimisation. Furukawa et al. proposed an optimized surface boundary extraction for self-occlusion using axis alignment for Manhattan-world scenes [32] but it also has limitations on keeping surface details. Therefore we propose a dominant surface selection algorithm to choose the most reliable surface among overlapped surfaces.

Fig. 5. Reliable surface extraction

Figure 5 (a) shows an illustration of real surface and overlapped surfaces reconstructed from three camera pairs. The overlapped surfaces include false surfaces from self-occlusion and less reliable (secondary) surfaces. We assume that the reconstructed surface is more reliable when the surface normal vector and camera viewing direction are aligned, and the distance to the camera is closer. Our purpose is to choose dominant surfaces from the set of false and secondary surfaces.

Figure 5 (b) shows notations used for evaluating reliabilities of conflicted surfaces. Let V_s is a vertex on a surface reconstructed from camera C_i . A projection vector c_s , a surface normal vector n_s and a facing angle θ_s are expressed as follows:

$$c_s = C_i - V_s \quad (7)$$

$$n_s = \frac{1}{N^2} \sum_{k \in N(s)} n_k \quad (8)$$

$$\theta_s = \arccos\left(\frac{c_s \cdot n_s}{|c_s| |n_s|}\right) \quad (9)$$

In Eq. (8), we use an average of normal vectors of neighbouring $N \times N$ vertices to calculate a global normal of the vertex. First, we remove all vertices whose facing angle θ_s is larger than an angle threshold Th_θ because they have high possibility to be a false surface. Then we search conflicted surface from the vertex V_s along the normal vector n_s in the range of Th_R . If confliction with other vertex V_t from camera C_j is detected, we calculate c_t , n_t , θ_t for the vertex V_t and camera C_j as Eq. (7)-(9), and also calculate unreliability $U(V)$ of each vertex based on the facing angle and the distance as Eq. (10). Finally the vertex with higher unreliability is removed.

$$U(V_s) = U_d(V_s) + \lambda_2 U_\theta(V_s) \quad (10)$$

$$U_d(V_s) = \begin{cases} |c_s|/d_{\max}, & \text{if } |c_s| < d_{\max} \\ 1, & \text{else} \end{cases} \quad (10-1)$$

$$U_\theta(V_s) = \left| \frac{\theta_s}{\pi/2} \right|, \quad \left(-\pi/2 < \theta_s < \pi/2 \right) \quad (10-2)$$

Application of the algorithm for all vertices simultaneously is time-consuming process and produces erroneous results with small isolated surfaces. For efficient computation we segment the surface into $N/2 \times N/2$ vertex patches and perform the merging in patches. In our experiment, we set the thresholds as: $N=10$, $Th_\theta=1.48$ (85°), $d_{\max}=30$ and $Th_R = d_{\max}/50$.

4 Experimental Results

4.1 Evaluation of Disparity Estimation

In order to evaluate the general performance of the proposed disparity estimation algorithm, we used the Middlebury stereo benchmarking test bed [28]. Figure 6 shows subjective comparison of estimated disparity maps with state-of-the-art algorithms which also include occlusion handling. The proposed method produces smoother maps with sharp object boundaries even in occlusion regions. However, the proposed algorithm is ranked at 47 in the Bad Pixel Percentage (BPP) test among 78 algorithms on the test bed, while the Cost-Aggr [33] and Semi-Global [34] algorithms are ranked at 27th and 33rd, respectively. The BPP test calculates only the ratio of erroneous pixels and ignores the magnitude of errors. PDE-based methods tend to spread errors into neighbouring pixels to suppress prominent errors, and it caused the low ranks in the BPP test despite good subjective performance. Therefore we changed measuring method and compared root mean square error (RMSE) to the ground truth.

Figure 7 shows the result with two additional methods. Adapting-BP [35] is a improved BP method which is ranked at the top in the BPP test and Graph-cut is an alpha-expansion method with occlusion handling [36] which is ranked at 42nd. All algorithms compared in this test are ranked higher than the proposed algorithm in the BPP test. Comparison of RMSE shows that the proposed PDE-based method gives comparable to the best error rate for state-of-the-art methods.

Fig. 6. Estimated disparity maps (from left to right: Tsukuba, Venus, Teddy and Cones)

Fig. 7. Comparison of RMSE

However, good performance in these tests does not mean that it guarantees good surface reconstruction. The Semi-Global and Graph-cut algorithms produce discrete disparity maps which can cause stepwise artifact as discussed in section 2.2. The Adapting BP algorithm fits the disparity fields into segmented planes so it loses all surface details. The advantage of the proposed approach is to generate continuous depth map while preserving surface details.

4.2 Evaluation of Model Reconstruction

For objective evaluation of scene modelling, we chose two objects reconstructed from image pairs captured at three different locations and compared the models with ground-truth models scanned by a LIDAR sensor. Figure 8 shows the ground-truth models and the reconstructed models from single/multiple views by the proposed

Fig. 8. Reconstructed models (Left: Ground-truth by LIDAR scan, Middle: Reconstruction from single pair, Right: Reconstruction from multiple pairs)

Fig. 9. Depth errors (Left: Ground-truth depth in common regions, Middle: Depth error of single view reconstruction, Right: Depth error of multi view reconstruction)

algorithm. The "Gate" has width of 9m and height of 6m, and the "Cupola" has 6.2m x 3.8m. Both objects are around 6m apart from the central capture point and stereo pairs are captured with a baseline of 60cm. We can see that the reconstructed model shows very fine structure with details of the surface relief pattern. Especially the multi-view reconstruction recovers self-occluded regions while keeping surface details.

It is hard to compare the accuracy and completeness of reconstructed meshes because the reconstructed ranges and areas are different (even the model from the LIDAR scan does not have complete structure.). Therefore we produced depth maps

from arbitrarily chosen viewpoints and measured an average depth error of common regions. Table 1 shows evaluation results from two different viewpoints for each model and Fig. 9 shows examples of errors mapped into gray scale. In Table 1, we can see that the multi-view reconstructions have slightly better results than single-view reconstruction. It is a bit more obvious in the slanted view (Viewpoints 2) than the frontal view (Viewpoints 1). However, the differences are not remarkable because most of the errors in Table 1 are from the vertical self-occlusions as seen in Fig. 9. The vertical self-occlusions could not be recovered in this experiment because the test models were captured from horizontally scattered locations. Another point to be considered is the fact that this comparison was performed only for commonly reconstructed regions. As seen in Fig. 8 (a), the multi-view reconstruction could recover the third lower walls that the single-view reconstruction could not.

Table 1. Depth error evaluation (unit: cm, μ : mean, σ : standard deviation)

	Gate				Cupola			
	Point 1 (frontal)		Point 2 (Fig. 9)		Point 1 (frontal)		Point 2 (Fig. 9)	
	μ	σ	μ	σ	μ	σ	μ	σ
Single-view	5.68	15.29	1.83	22.24	3.84	19.65	5.64	28.85
Multi-view	4.59	16.68	1.05	20.13	3.37	17.10	4.58	24.19

Fig. 10. Full 3D scene rendering (Geometry and texture mapping)

Figure 10 shows the reconstructed full outdoor scene of Fig. 4 (c) from three spherical stereo pairs and results of texture mapping from the same viewpoints. The results show a natural-looking geometry and textures of the environment. Three

hole-like regions on the ground show the capture points which could not be reconstructed because of the divergence of spherical stereo in Eq. (1). Free-viewpoint videos of the reconstructed models are included in the supplemental video file.

5 Conclusion

In this paper, a 3D environment modelling method using multiple pairs of spherical stereo images was proposed. The environment is captured by spherical cameras at multiple locations and 3D mesh models for each pair are reconstructed by spherical stereo geometry. The proposed PDE-based stereo method reconstructs continuous depth fields with sharp object boundaries even in occluded regions and highly textured regions. Instead of cumbersome camera calibration for all cameras, 3D rigid transforms between views are calculated by SURF feature matching. A RANSAC algorithm is introduced to fuse incomplete models including self-occlusions. Finally a complete 3D model of the environment is generated as a single mesh by selecting the most reliable surfaces among overlapped surfaces by considering visibility, surface normals and distance. The biggest advantage of the proposed surface selection algorithm against other surface merging algorithms is to effectively eliminate outlier surfaces from occlusion. The performances of the proposed algorithms were evaluated against ground-truth from the stereo test bed and LIDAR scans. The final composite model can be rendered from any viewpoint with high quality textures.

Acknowledgments. This research was executed with the financial support of the EU IST FP7 project i3Dpost.

References

1. Lemmens, M.: Airborne LiDAR Sensor. GIM International 21 (2007)
2. Strecha, C., Hansen, W., Gool, L.V., Fua, P., Thoennessen, U.: On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery. In: Proc. CVPR, pp. 1–8 (2008)
3. Furukawa, Y., Ponce, J.: Accurate, Dense, and Robust Multi-View Stereopsis. In: Proc. CVPR (2007)
4. Vu, H., Keriven, R., Labatut, P., Pons, J.P.: Towards high-resolution large-scale multi-view stereo. In: Proc. CVPR, pp. 1430–1437 (2009)
5. Salman, N., Yvinec, M.: Surface Reconstruction from Multi-View Stereo. In: Proc. ACCV (2009)
6. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building Rome in a Day. In: Proc. ICCV (2009)
7. Pollefeys, M., Koch, R., Vergauwen, M., Gool, L.V.: Automated reconstruction of 3D scenes from sequences of images. ISPRS 55, 251–267 (2000)
8. Cornelis, N., Leibe, B., Cornelis, K., Gool, L.V.: 3D urban scene modeling integrating recognition and reconstruction. IJCV 78, 121–141 (2008)
9. Pollefeys, M., et al.: Detailed real-time urban 3D reconstruction from video. IJCV 78, 143–167 (2008)

10. Nayar, S.: Catadioptric Omnidirectional Camera. In: Proc. CVPR, pp. 482–488 (1997)
11. Li, S.: Full-View Spherical Image Camera. In: Proc. ICPR, pp. 386–390 (2006)
12. Feldman, D., Weinshall, D.: Realtime IBR with Omnidirectional Crossed-Slits Projection. In: Proc. ICCV, pp. 839–845 (2005)
13. Kim, H., Hilton, A.: 3D Environment Modelling Using Spherical Stereo Imaging. In: Proc. 3DIM Workshop in ICCV (2009)
14. Boström, G., Fiocco, M., Puig, D., Rossini, A., Gonçalves, J.G.M., Sequeira, V.: Acquisition, Modelling and Rendering of Very Large Urban Environments. In: Proc. 3DPVT, pp. 191–198 (2004)
15. Asai, T., Kanbara, M., Yokoya, N.: 3D Modeling of Outdoor Environments by Integrating Omnidirectional Range and Color Images. In: Proc. 3DIM, pp. 447–454 (2005)
16. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: SURF: Speeded Up Robust Features. CVIU 110, 346–359 (2008)
17. Li, S., Fukumori, K.: Spherical Stereo for the Construction of Immersive VR Environment. In: Proc. IEEE VR, pp. 217–222 (2005)
18. Scharstein, D., Szeliski, R.: A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms. IJCV 47, 7–42 (2002)
19. Alvarez, L., Deriche, R., Sánchez, J., Weickert, J.: Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-Space Based Approach. Journal of Visual Comm. and Image Repr. 13, 3–21 (2002)
20. Weickert, J.: A Review of Nonlinear Diffusion Filtering. In: ter Haar Romeny, B.M., Florack, L.M.J., Viergever, M.A. (eds.) Scale-Space 1997. LNCS, vol. 1252, pp. 3–28. Springer, Heidelberg (1997)
21. Zimmer, H., Bruhn, A., Valgaerts, L., Breuß, M., Weickert, J., Rosenhahn, B., Seidel, H.P.: PDE-based anisotropic disparity-driven stereo vision. In: Proc. VMV, pp. 263–272 (2008)
22. Sun, D., Roth, S., Lewis, J., Black, M.: Learning Optical Flow. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 83–97. Springer, Heidelberg (2008)
23. Ben-Ari, R., Sochen, N.: Variational Stereo Vision with Sharp Discontinuities and Occlusion Handling. In: Proc. ICCV, pp. 1–7 (2007)
24. Alvarez, L., Deriche, R., Papadopoulos, T., Sánchez, J.: Symmetrical Dense Optical Flow Estimation with Occlusions Detection. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part I. LNCS, vol. 2350, pp. 721–735. Springer, Heidelberg (2002)
25. Ince, S., Konrad, J.: Occlusion-aware optical flow estimation. IEEE Trans. Image Proc. 17, 1443–1451 (2008)
26. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
27. Kim, H., Sohn, K.: Hierarchical Depth Estimation for Image Synthesis in Mixed Reality. In: Proc. SPIE Electronic Imaging, pp. 544–553 (2003)
28. Middlebury stereo and multi-view benchmarking site,
<http://vision.middlebury.edu/>
29. Besl, P., McKay, N.: A method for Registration of 3-D Shapes. IEEE Trans. PAMI 14, 239–256 (1992)
30. Kazhdan, M., et al.: Poisson Surface Reconstruction. In: Proc. SGP, pp. 61–70 (2006)
31. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: Proc. SIGGRAPH, pp. 303–312 (1996)

32. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Reconstructing Building Interiors from Images. In: Proc. ICCV (2009)
33. Min, D., Sohn, K.: Cost aggregation and occlusion handling with WLS in stereo matching. IEEE Trans. Image Proc., 1431–1442 (2008)
34. Hirschmüller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. IEEE Trans. PAMI 30, 328–341 (2008)
35. Klaus, A., Sormann, M., Karner, K.: Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: Proc. ICPR, pp. 15–18 (2006)
36. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions using graph cuts. In: Proc. ICCV (2001)