An Opto-Mechanical Apparatus for Binocular Helmholtz Stereopsis in Static and Dynamic Scenes

Wagner Barros, Rodrigo Carceroni, Lara Coelho
Universidade Federal de Minas Gerais
Departamento de Ciência da Computação
Av. Antônio Carlos, 6627, Belo Horizonte, MG
{wbarros,carceron,lara}@dcc.ufmg.br

José Queiroz-Neto
Centro Federal de Educação Tecnológica
Departamento de Informática
Av. Danilo Areosa, 138, Manaus, AM
pinheiro@cefetam.edu.br

Abstract

This paper presents an opto-mechanical apparatus for binocular stereopsis based on the Helmholtz reciprocity principle. The proposed system makes it possible to obtain three-dimensional models of static and dynamic scenes with arbitrary reflectance properties including highly specular surfaces and holograms. We also propose an empirical methodology for the treatment of occlusions that improves the description of half-occluded areas. Experimental results with real images validate the proposed system.

1. Introduction

The use of computer vision to obtain three-dimensional models from images is a promising approach in several applications, such as preservation of historical and cultural heritage [7] and analysis of underwater scenes [11]. However, image analysis for 3D reconstruction is a hard task. Stereo vision requires handling technical difficulties such as: half occluded regions (regions that can be seen from just one viewpoint); view-dependence of specular highlight positions; and surfaces with non-parametric reflectance, such as holograms. Except for half-occluded regions, all difficulties listed above are due to image appearance variations as a function of observation position, illumination and orientation of each observed scene point.

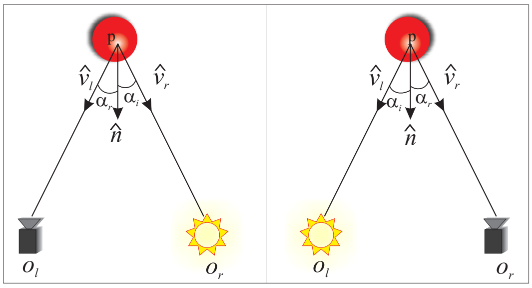
The majority of 3D reconstruction methods based on images makes use of simplified surface reflectance models. It is common to assume that the surface reflectance of scene objects is either approximately diffuse (*Lambertian*) can be modeled by simplified parametric functions such as Phong [10], Oren-Nayar [9] or Torrance-Sparrow [12] models. However, assuming such simplifications results in restricting the set of possible scenes to be reconstructed. For instance, with such models it is often impossible to reconstruct scenes formed by sets of objects with distinct and heterogeneous reflectance properties.

Other methods in the literature such as photometric stereo (*Shape-from-Shading*) assume knowledge about the environment illumination conditions. In such methods, several images are acquired from a fixed camera position with varying light source positions. Among the imposed simplifications, the most important is that these methods assume that the scene surfaces *Bidirectional Reflectance Distribution Functions* (BRDFs) are known or are measured *a priori* in a calibration step [14, 5]. When such restrictions are not satisfied the quality of the generated model is compromised.

Recently, new methodologies for 3D reconstruction have been developed to overcome these limitations imposed by conventional methods. In particular, one of these methods relies on a new system to acquire images exploring the Helmholtz reciprocity principle [13]. Such reconstruction methodology is mentioned on literature as *Helmholtz stereopsis* or *reciprocal stereo*. The idea behind the Helmholtz stereopsis is to exploit the symmetry of the surface reflectance function (BRDF). As an example, consider two images acquired as shown in Figure 1. Because of Helmholtz reciprocity these images have an important property: for corresponding pixels, the ratio of incident irradiance (onto the object) to emitted radiance (from the object) will be the same. In a simple way, we can derive a relationship between the intensities of corresponding pixels that does not depend on the physical properties of the surface such as its reflectance or its transmittance [15].

From Figure 1 we can observe that a system for *reciprocal pair* (stereo pairs acquired based on Helmholtz reciprocity principle) acquisition must be able to swap camera and light source positions as precisely as possible, in order to assure that Helmholtz reciprocity is useful.

A system for acquisition of reciprocal pairs employing a single camera and light source mounted on a wheel (Figure



(a) (b)

Figure 1. Reciprocal stereo pair acquisition. (a) An image is acquired by a camera at position o_l, and the scene is illuminated by a single source light at position o_r. (b) Camera and light source positions are swapped and a new image is acquired.

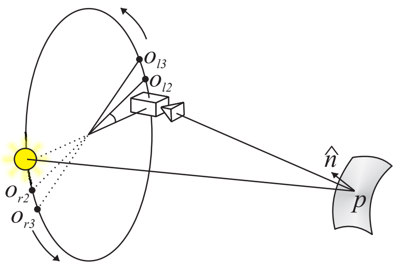


Figure 2. Reciprocal pairs acquisition system in which camera and light source are placed in diametrally opposite positions on a wheel. The reciprocal pair acquisition involves a rotation of the wheel by 180^o after each image acquisition.

2) has been implemented by Zickler *et al.* [15, 16, 8]. However, such acquisition system has a few drawbacks: the lack of freedom in the system *baseline* size; the requirement of an accurate system to rotate the wheel by exact 180^o; and the time spent to obtain reciprocal pairs, which is too high to allow reconstruction of dynamic scenes.

Moreover, Zickler *et al* [15] assumed some restrictions about the geometric properties of the scene objects, such as the fact that the observed surfaces were predominantly smooth. As result, low resolution depth maps have been obtained. Moreover, that work did not exploit to a full extent the presence of half-occluded areas, which can substantially improve the results on reciprocal stereo algorithms.

In this paper we propose a method to recover 3D information based on a new opto-mechanical apparatus. The proposed method makes it possible to reconstruct scenes with complex reflectance properties and geometric discontinuities, as well as highly specular scenes and even scenes including holograms. The method also exploits information about half-occluded regions that can be directly inferred from the visible shadows in the reciprocal pairs. The new opto-mechanical apparatus has the potential to allow acquisition of reciprocal pairs in real-time, enabling real-time acquisition of reciprocal videos.

2. Helmholtz Reciprocity

According to the Helmholtz reciprocity principle [13], the reflectance of a point remains unchanged when we swap the incidence and reflection directions. Thus, any BRDF f is such that:

$$f(\hat{v}_l, \hat{v}_r) = f(\hat{v}_r, \hat{v}_l)$$

where $\hat{v}_l$ is the reflected light direction and $\hat{v}_r$ is the incident light direction, as shown in Figure 1(a).

To apply this principle to 3D reconstruction, we consider an image pair obtained by a stereo capture system [15], as shown in Figure 1. We acquire the first image placing the camera center in the o_l position and a single light source in the o_r position, as we show in the Figure 1(a); we obtain the second image swapping the camera center and the light source positions to o_r and o_l positions, respectively, as in Figure 1(b). Thus, when a scene surface point p with a normal vector $\hat{n}$ is projected to the images acquired as shown in Figure 1, any pair of corresponding radiance values i_a and i_b from the images acquired as in Figures 1(a) and (b), respectively, satisfy:

$$i_a = f(\hat{v}_r, \hat{v}_l)\frac{\hat{n}\cdot\hat{v}_r}{|o_r - p|^2}, \quad (1)$$

$$i_b = f(\hat{v}_l, \hat{v}_r)\frac{\hat{n}\cdot\hat{v}_l}{|o_l - p|^2}, \quad (2)$$

where $\hat{v}_r$ and $\hat{v}_l$ are the directions from point p to the camera center and the light source, respectively, and $\frac{1}{|o_r-p|^2}$ and $\frac{1}{|o_l-p|^2}$ are the light attenuation factors due the distances between the two light source positions and point p.

Defining d as the distance between a scene point p and the baseline, (*i*.e., the line specified by o_l and o_r) and applying the Helmholtz reciprocity principle we rewrite the Equations (1) and (2) as:

$$\left(i_a\frac{\hat{v}_l}{|o_l - p|^2} - i_b\frac{\hat{v}_r}{|o_r - p|^2}\right)\cdot\hat{n} = \mathbf{w}(d)\cdot\hat{n} = 0. \quad (3)$$

In Equation (3) only the normal $\hat{n}$ and the depth d are unknown. The i_a and i_b values are obtained directly from the images by applying the camera's inverse photometric transfer function (photometric calibration). We obtain the o_l and

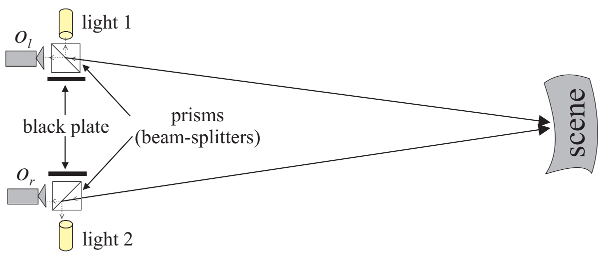


Figure 3. The proposed reciprocal pair acquisition system. We use two cameras (at o_l and o_r positions), two light sources and two beam-splitters.

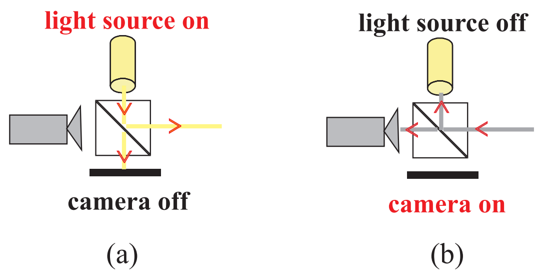


Figure 4. One light/camera pair of the acquisition system. When the light source is on, the camera must be off and vice-versa. The beam-splitters have the role of making the light emission center and the camera projection center optically coincident.

o_r values from the camera's geometric calibration. If we calculate the normal ($\widehat{n}$) and the depth (d) values, we obtain $\widehat{v}_l$ and $\widehat{v}_r$. Therefore, by the correspondence between pixels intensities of the two images, Equation (3) allows us to obtain the distance d and the surface normal $\widehat{n}$, even without knowing the surface BRDF. But since the Equation (3) has three degrees of freedom (one for d and two for $\widehat{n}$), we need more than a single image pair to recover that information, in general.

3. Methodology

The proposed methodology uses an opto-mechanical system in order to simplify the general case described in Section 2. In this simpler case, the corresponding points in the reciprocal pair must have the same intensity independently on the observed surface. Moreover, we propose to solve the half-occlusion problem using only the pixel intensity information available in the images.

3.1. Fronto-Parallel Reciprocal Pairs

In the Helmholtz Reciprocity it is necessary more than one reciprocal pair to establish correspondences through Equation (3). However, we use a special case where the distance between the projection centers of the cameras (baseline) is small relative to the scene depth. In this case,

$$|o_l - p|^2 \approx |o_r - p|^2. \tag{4}$$

Moreover, from the fact that the system has a small baseline we obtain:

$$\widehat{n} \cdot \widehat{v}_l \approx \widehat{n} \cdot \widehat{v}_r. \tag{5}$$

By substituting the Equations (4) and (5) in Equation (3), we get

$$i_a = i_b. \tag{6}$$

Thus, we reduce the correspondence problem to a simple comparison between pixel intensities in the images without imposing any restriction on surface reflectance, as in [15]. The main advantage of this approach is that a single fronto-parallel reciprocal pair can be used as input to any stereo algorithm, providing good results with almost no need to write additional code.

3.2. The Opto-Mechanical System

The reciprocal pair acquisition system must be capable switching the positions of camera and light source precisely This is a critical task in the process, because the reconstruction methodology uses the stereo system extrinsic parameters, which are extremely sensitive to movement carried through the opto-mechanical system. Because of this, it is very important to avoid a human intervention while acquiring the reciprocal pair.

To mitigate the disadvantages of the acquisition systems used in [15, 16, 8], we present a novel reciprocal pair acquisition system that uses the following equipment: two CCD cameras, two light sources and two beam-splitters. The system design is shown in Figure 3.

The optical system is used to simulate that each camera has a light source at its projection center. With the proposed system it becomes easy to acquire the reciprocal pair. First, turn on the left light source and acquire an image with the right camera. Then, turn off the left light and turn on the right light and then acquire another image with the left camera (Figure 4).

The beam-splitter used in the system is composed by two joined prisms forming a cube. It divides the light rays reflected by the scene so that about 50% of the light is transmitted without suffering reflection and reaches the camera, while about another 50% is reflected through an angle of 90^o in direction to the light source. In the same way, the

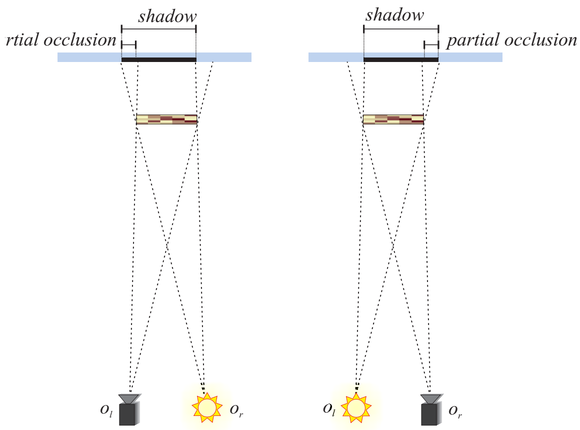


Figure 5. Partially-occluded regions in Helmholtz Stereopsis. Shadows in each image correspond to occlusions in the other.

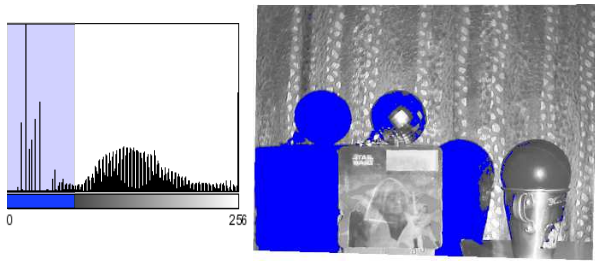


Figure 6. Analysis of intensity histogram for shadow detection. Right: only intensities in the range highlighted in blue are regarded as possible shadow intensities. Left: only pixels in highlighted regions are regarded as being possible shadows.

light emitted by the source is divided in two portions that go toward the black plate and the scene.

The role of the black plate is to cover the other side of the beam-splitter to avoid interference of light from another scenes. For example, if we didn't have the black plate, the left camera would receive light from the scene to the right side of the system.

The proposed system allows us to eliminate the main problems found in the current systems, providing more flexibility to change the baseline; more precision to acquire the reciprocal pair; and the possibility of using an automatic on/off control to make the system useful for acquisition of reciprocal video.

3.3. Occlusion Handling

In a reciprocal pair, there is a one-to-one correspondence between regions covered by shadows in each image and regions occluded in the other image. We use this property to improve the disparity maps obtained with dense stereo by reasoning explicitly about half-occluded areas. Figure 5 shows that we can infer the localization of occluded regions just by detecting regions covered by shadows in the images.

To exploit this fact, we use a simple heuristic based on the analysis of intensity histograms to compute adaptive thresholds for shadow intensity. The computed thresholds are used to binarize the images in the reciprocal pair into regions that are likely to be in shadow and regions that are not. This segmentation is then fed as additional constraints to a dense stereo algorithm, allowing it to determine the discontinuities in the observed surfaces more accurately.

More specifically, since we are looking for dark pixels, we limit our analysis to histogram bins that correspond to intensities lower than an empirically set threshold L_{prior}, as shown in Figure 6. Within this selected low-end of the dynamic range, we choose the histogram peak, *i.e.*, the histogram bin that received most votes from the image pixels. Let L_{peak} be the image intensity represented by the selected histogram bin. We simply multiply L_{peak} by a constant α greater than one, also determined empirically, to obtain a final shadow threshold L_{shadow}. This final step is necessary to account for the inherent variance in shadow intensities, which depend on the properties of the surfaces where shadows are projected, among other factors.

In practice, we have found that it is generally better to detect false shadows than to miss real shadows, which can be accomplished by choosing a relatively large value for α. In order to overcome the difficulties created by the resulting false-positive shadows, we use a symmetric stereo method [6]. That method uses color restrictions to determine if the disparity calculated for one pixel in an image is equivalent to the disparity of its corresponding pixel in the other image (*cross-check*). Only when the calculated disparities for corresponding pixels in the two images are different, the shadow threshold L_{shadow} is used to verify if one of the pixels is occluded or not.

While the simple-minded heuristic described in this section has been sufficient to handle occlusion properly in most datasets that we have done experiments with, it can fail dramatically in some cases. Since this work was originally developed [1], more robust schemes for detection of half-occluded regions have appeared in the literature [3].

3.4. Automatic Control for Dynamic Acquisition

To use the proposed opto-mechanical system (Figure 3) for reconstruction of dynamic scenes, we designed an automatic on/off control between each camera/light module, as shown in Figure 7.

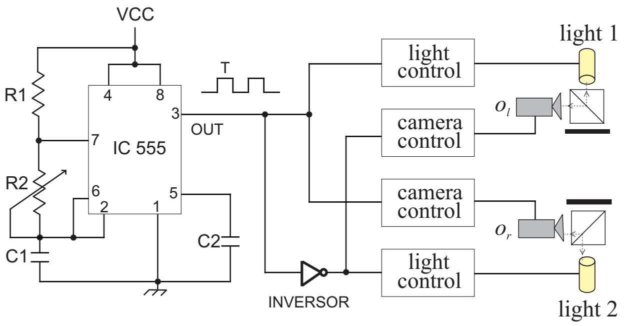


Figure 7. Automatic system to acquire reciprocal video. It uses an oscillator circuit to guarantee the synchronism between the on/off control circuits of cameras and lights, within the optical system of the Figure 8.

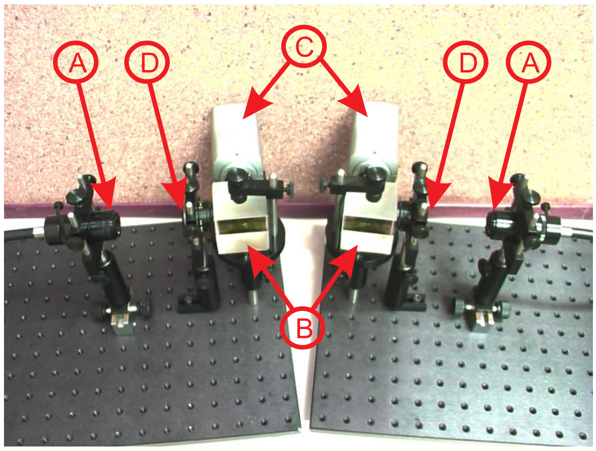


Figure 8. The reciprocal pair acquisition system. We used two approximately collimated light sources (A); two beam-splitters (B); two CCD cameras (C); and two 16mm lens (D).

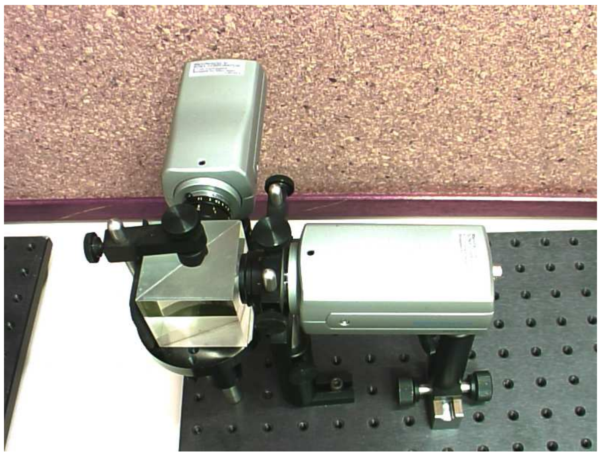


Figure 9. System configuration used in the camera/light module calibration.

The system uses an oscillator that generates a square waveform to synchronize the action of the two camera/light modules. The generated waveform is applied directly to the on/off control circuits of the right camera and left light. Then, it is inverted and applied to the on/off control circuits of the left camera and right light. This ensures that only the light reciprocal to a camera will be on when that camera acquires an image.

The oscillator circuit is composed of an integrated circuit 555, two resistors R_1 and R_2 and two capacitors C_1 and C_2. The period T of the waveform depends on the values of R_1, R_2 and C_2 ($T \cong 0.7(R_1 + 2\,R_2\,C_2)$), which must be calculated with $R_2 >> R_1$ to guarantee that the on/off periods will be identical. The resistance R_2 is used to adjust T because of the imprecision of the nominal values of the components used in the circuit. To have good results, it is important to set T to less than a half of the frame period in the movie that one wants to obtain.

The on/off light control block is equivalent to a stroboscopic light circuit, whose frequency will be given by the signal from the oscillator circuit. The light bulb must have a small latency to turn itself on and off, as a xenon light bulb. The camera control circuit is the only element in Figure 7 still not implemented.

It is important to say that the system requires limitations on the speed of the objects in the dynamic scene. A significant object displacement between the images in the reciprocal pair compromises the quality of disparity maps, much like the use of interlaced video compromises the quality of several Computer Vision techniques. Existing reciprocal-stereo acquisition systems, however, can not deal even with slowly-moving objects, because of the latency in the camera/light switching process.

4. Opto-mechanical system assembly

To assemble the opto-mechanical system described in Section 3.2, we used: two cameras *Sony DFW-X700*; two non-polarizing beam-splitters *Newport TS 50mm*; two manually-regulated 150-Watt light sources *Fiber-Lite PL900*; four high-resolution *Tamron 16mm* lenses; and two sets of mechanical supports to assemble the equipments rigidly. We show the assembled system in Figure 8.

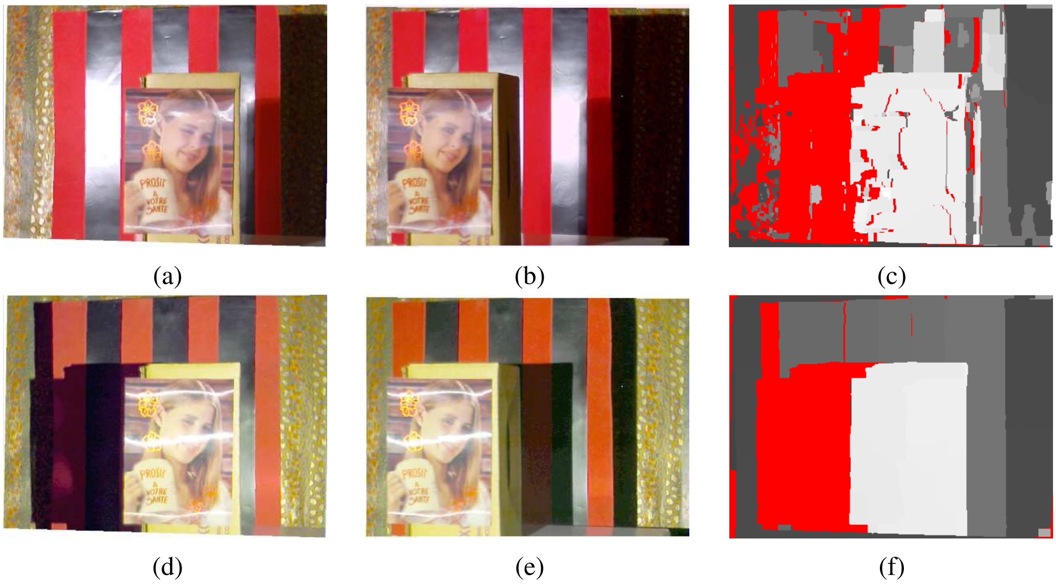


Figure 10. Comparison between typical conventional stereo and reciprocal stereo results: (a) and (b) stereo pair acquired under fixed illumination; (c) left-view depth map generated by conventional stereo, with half-occlusions shown in red; (d) and (e) stereo pair acquired with the proposed opto-mechanical system; (f) left-view depth map and occlusions detected by reciprocal stereo.

4.1. Camera/light module calibration

An important step in the assembly of the system is the calibration of the camera/light module, which corresponds to determining the light source position, given the camera position in the module. To calibrate each module, we use a *Tamron 16mm* lens positioned in front of the light source (see Figure 8). This lens is used as a reference for the light position and orientation. We adjust to the position of light beam by projecting it in the center of this lens and aligning the lens optically with the camera lens.

To do such optical calibration, we acquire images with a third camera connected to the Tamron lens in front of the light source, as shown in Figure 9. The light and camera lenses are aligned when the images from that third camera and from the fixed camera of the module become identical.

5. Results

In this section we show some experimental results that validate our proposed methodology. In all experiments camera/light source module calibration was performed as described in Section 4.1 and stereo parameters were calibrated according to [2]. The acquired images were then submitted to a chromatic compensation process (needed due to use of distinct cameras) and a geometric rectification process (as described in [4]), which are common steps in stereo vision reconstruction methods. Having all calibration parameters already computed (chromatic compensation parameters, intrinsic and extrinsic stereo system parameters) we proceed with the following steps:

1. acquire a reciprocal pair;
2. perform chromatic compensation;
3. rectify the stereo pair;
4. compute the shadow threshold L_{shadow};
5. execute a stereo matching algorithm;
6. execute a *cross-check* algorithm based on results of two previous steps.

More specifically, to perform the stereo correspondence we use an algorithm based on energy minimization via graph cuts [6], due to its ability to handle occlusions properly through an energy functional that imposes smoothness constraints but at the same time tries to preserve discontinuities. This algorithm deals with input images asymmetrically and thus it can be used on our method described in section 3.3 to locate half-occluded regions.

We use *depth maps* to exhibit the results obtained. In such maps, bright pixels are used to represent points near the camera, darker pixels represent more distant points and red pixels represent regions labeled as half-occluded. All experiments use the left image as the reference image for depth map computation. More details concerning these experiments are available in [1].

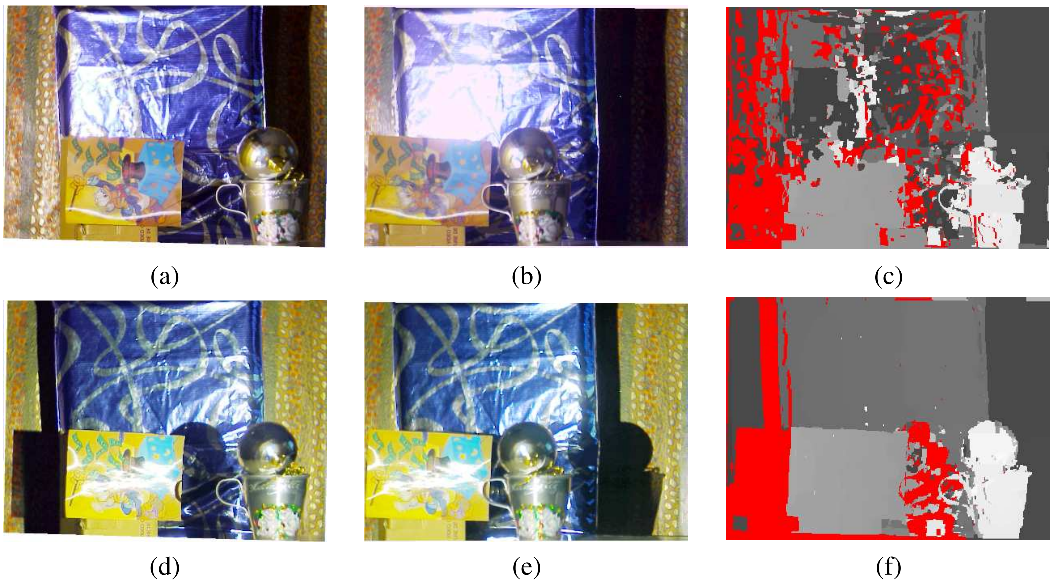


Figure 11. Reconstruction of a more challenging scene: (a) and (b) stereo pair acquired under fixed illumination; (c) left-view depth map generated by conventional stereo, with half-occlusions shown in red; (d) and (e) stereo pair acquired with the proposed opto-mechanical system; (f) left-view depth map and occlusions detected by reciprocal stereo.

5.1. Static Scenes

In Figure 10 we present the typical results obtained by applying the methodology described in this paper to a scene with highly specular objects. In comparison, we also present the results obtained using the same stereo matching algorithm (except by the treatment of half-occluded areas) when the same observed scenes were illuminated by a fixed light source (conventional stereo).

It is clear from Figure 10 that scenes containing highly specular objects pose a serious challenge to conventional stereo matching algorithms, due to variations of the highlight positions (Figures 10(a)-(b)). The depth maps recovered with conventional stereo for such scenes tend to be excessively fragmented and to contain some gross artifacts. On the other hand, in Figures 10(d)-(e) we can see that all specular reflections are in corresponding points in the two images. This allows our algorithm to work properly. In its result, shown in Figure 10(f), all major half-occluded regions were determined correctly. More results like this are described in [1].

To document the failure modes of the proposed methodology, we display more challenging cases in Figure 11 and 12. In both cases, reciprocal stereo still leads to obvious improvements over fixed-illumination stereo. But the specific algorithm described in this paper fails to detect a significant fraction of the half-occlusion caused by the metallic cup in Figure 11 and it fails to detect almost entirely a similar region in Figure 12. It is likely that such difficulties can be mitigated by the use of more principled occlusion handling [3].

5.2. Dynamic Scene

To illustrate what our system could do on dynamic scenes, we simulated the acquisition of a car moving over a planar surface. To do it, we used a manual camera/light control, since the procedure proposed in Section 3.4 has still not been fully implemented. The obtained result can be seen at `http://www.dcc.ufmg.br/~wbarros/Helmholtz/`.

6. Conclusions

We presented a new opto-mechanical apparatus for acquisition of reciprocal pairs using the Helmholtz Reciprocity Principle. In comparison to existing reciprocal-stereo systems, ours is easier to use, due to the flexible baseline size and the automatic camera/light switch, and it potentially allows acquisition of reciprocal video using an automatic control module of camera/light sets, still in development. We have shown that images of scenes with complex reflectance acquired with our apparatus yield qualitatively better depth maps than conventional stereo.

As future work, we plan to conclude the development of the automatic control module for dynamic scenes, to im-

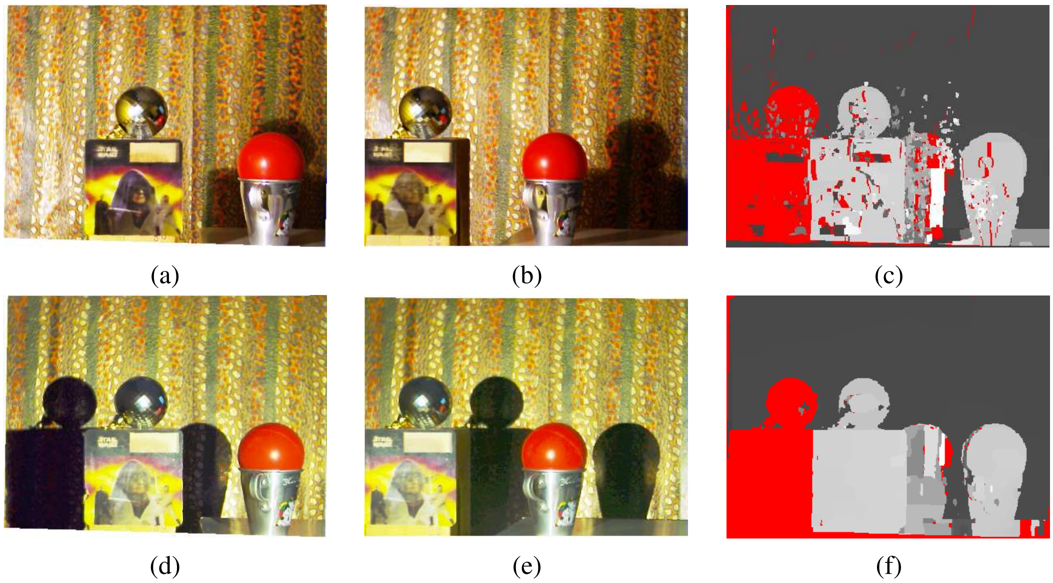


Figure 12. Reconstruction of an even more challenging scene: (a) and (b) stereo pair acquired under fixed illumination; (c) left-view depth map generated by conventional stereo, with half-occlusions shown in red; (d) and (e) stereo pair acquired with the proposed opto-mechanical system; (f) left-view depth map and occlusions detected by reciprocal stereo.

prove our shadow detection algorithm and to perform new experiments to establish the system limits.

Acknowledgments

We thank the support of Fapemig under Proc. No. CEX-227-04, of CNPq under multiple grants and of PRPq-UFMG (Fundo Fundep RD).

References

[1] W. F. Barros. Reconstrução 3D de cenas com propriedades de reflectância arbitrárias usando o Princípio da Reciprocidade de Helmholtz. Master's thesis, UFMG, 2004.

[2] J.-Y. Bouguet. Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/, 1999.

[3] R. Feris, R. Raskar, L. Chen, K.-H. Tan, and M. Turk. Discontinuity preserving stereo with small baseline multi-flash illumination. In *ICCV*, pages 412–419, 2005.

[4] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Journal of Machine Vision and Applications*, 12(1):16–22, 2000.

[5] A. Hertzmann and S. M. Seitz. Shape and materials by example: A photometric stereo approach. In *CVPR*, volume 1, pages 533–540, 2003.

[6] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *ECCV, Part III*, pages 82–96, 2002.

[7] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project. http://graphics.stanford.edu/projects/mich/, 1998.

[8] S. Magda, T. Zickler, D. J. Kriegman, and P. N. Belhumeur. Beyond lambert: Reconstructing surfaces with arbitrary brdfs. In *ICCV*, pages 391–398, 2001.

[9] M. Oren and S. K. Nayar. Generalization of the lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14(3):227–251, 1995.

[10] B. T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(8):311–317, 1975.

[11] J. P. Queiroz-Neto, R. L. Carceroni, W. F. Barros, and M. F. M. Campos. Underwater stereo. In *SIBGRAPI*, pages 170–177, 2004.

[12] K. E. Torrance and E. M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9):1105–1114, 1967.

[13] H. von Helmholtz. *Treatise on Physiological Optics*. Dover, 1925. Translated from the 1856 German edition.

[14] R. J. Woodham. Gradient and curvature from photometric stereo including local confidence estimation. *Journal of the Optical Society of America*, A(11):3050–3068, 1994.

[15] T. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. In *ECCV*, pages 869–884, 2002.

[16] T. Zickler, J. Ho, D. J. Kriegman, J. Ponce, and P. N. Belhumeur. Binocular helmholtz stereopsis. In *ICCV*, pages 1411–1417, 2003.