A Long-Range Vision System for Projection Mapping of Stereoscopic Content in Outdoor Areas

Behnam Maneshgar[1], Leila Sujir[2], Sudhir P. Mudur[1], and Charalambos Poullis[1]
[1]*Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada*
[2]*Department of Studio Arts, Concordia University, Montreal, Canada*
{*behnam.maneshgar, leila.sujir, sudhir.mudur, charalambos.poullis*}*@concordia.ca*

Keywords: stereoscopic projection, outdoor projection mapping, long-range projection

Abstract: Spatial Augmented Reality, or its more commonly known name Projection Mapping (PM), is a projection technique which transforms a real-life object or scene into a surface for video projection (Raskar et al., 1998b). Although this technique has been pioneered and used by Disney since the seventies, it is in recent years that it has gained significant popularity due to the availability of specialized software which simplifies the otherwise cumbersome calibration process (Raskar et al., 1998a). Currently, PM is being widely used in advertising, marketing, cultural events, live performances, theater, etc as a way of enhancing an object/scene by superimposing visual content (Ridel et al., 2014). However, despite the wide availability of specialized software, several restrictions are still imposed on the type of objects/scenes on which PM can be applied. Most limitations are due to problems in handling objects/scenes with (a) complex reflectance properties and (b) low intensity or distinct colors. In this work, we address these limitations and present solutions for mitigating these problems. We present a complete framework for calibration, geometry acquisition and reconstruction, estimation of reflectance properties, and finally color compensation; all within the context of outdoor long-range PM of stereoscopic content. Using the proposed technique, the observed projections are as close as possible [constrained by hardware limitations] to the actual content being projected; therefore ensuring the perception of depth and immersion when viewed with stereo glasses. We have performed extensive experiments and the results are reported.

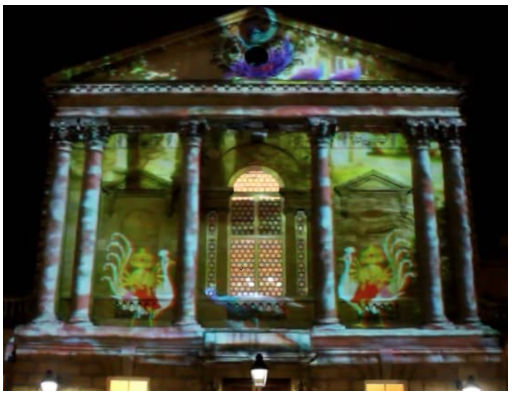
1 INTRODUCTION

Many successful techniques have been developed for capturing and modeling the shape and reflectance properties of objects/scenes (Torrance and Sparrow, 1967), (Phong, 1975), (Oren and Nayar, 1995), (Lafortune et al., 1997). These have been particularly successful in cases where the acquisition is performed under controlled lab conditions without the presence of any dynamic elements. Although PM can also be used indoors in a similar fashion, the majority of its applications involve large-scale and/or outdoor objects/scenes. Perhaps the only work reported in the literature to address the capture of complex geometry and the estimation of reflectance properties of outdoor objects is by Debevec et al. (Debevec et al., 2004).

In an outdoor setting there are several challenges: (a) there is no control over the lighting, (b) there is no control over other dynamic elements which may be present in the scene e.g. people walking, cars passing by, clouds, rain, etc, (c) most often limited time is available to perform the capture because of the aforementioned challenges. An example of one of our outdoor projection mapping experiments is shown in Figure 1(a). The viewers were wearing stereo glasses in order to perceive depth since the projected content was stereoscopic, as shown in Figure 1(b). Objects with complex reflectance properties such as the windows and columns are not specifically handled during the projection and resulted in color distortions to the red-cyan stereo content which further caused the loss of depth perception and some noticeable visual artifacts.

In this paper, we address the problem of long range projection mapping of stereoscopic content on outdoor areas and propose a complete framework which automates the following processes: (a) system calibration, (b) structure and appearance information acquisition, (c) approximating model of projection surface's reflectance properties, (d) color compensation to the extent possible with the given projection surface. The result is compensated image/video con-

(a)



(b)

Figure 1: (a) The facade of the Roman Baths, UK. (b) Stereoscopic content being projected onto the facade.

tent such that its projection onto the particular surface will produce an image/video which, when viewed, will seem as close to the original as possible. This is, of course, limited by the projection surface properties, as it may not always be possible to completely compensate for the surface reflectance behavior.

The paper is organized as follows: Section 2 gives an overview of the state-of-the-art in the area and Section 3 presents a technical overview of the proposed framework. In Section 4 we discuss the calibration of the cameras and the projector with respect to each other, and compute both extrinsic and intrinsic parameters for them. Next, Section 5 describes the geometry acquisition. The reflectance properties are modelled to their best approximation as described in Section 6. Lastly, Section 7 presents the color compensation and illustrative experimental results are provided in Section 8.

2 Related Work

R. R. Garcia and A. Zakhor (Garcia and Zakhor, 2013) calibrated a multi-camera-projector system to set up a multi-view structure light. By taking the advantage of binary codes they encode each projector pixel. By projecting the patterns on a screen and capturing the images sequence by the cameras and decoding the binary codes, they generate dense correspondences between the cameras and projector. Bundle adjustment is performed to calibrate the system. Svoboda et al. (Svoboda et al., 2005) also proposed a fully multi-camera self-calibration method for virtual environments. They wave a detectable bright spot in the working volume and capture the volume with at least three synchronised cameras. These points are validated through pairwise epipolar constraints. They calibrate the system by using these correspondences.

3D reconstruction is a well-studied area in computer graphics. Many techniques are proposed and many commercial products are available on the market. Microsoft Kinect is one of the most popular devices and comes with game consoles. It is able to scan the scene/object with high accuracy in short range. Y. Furukawa and J. Ponce (Furukawa and Ponce, 2010) proposed an algorithm to reconstruct an object/scene by using calibrated multi-view stereo. The algorithm detects feature points of each image and then finds the matches between each pose and outputs a dense set of patches covering the surface of the object/scene.

The diffuse component and specularity or reflection component of the scene is a very important issue in computer graphics, computer vision and in general vision systems. Typically, techniques try to first detect the specularity of the scene and then cancel or reject them as an outlier. Several methods with different viewpoints proposed to tackle this problem, Single image techniques and multiple image techniques have been proposed to extract diffuse map and specular map of the scene. Lin et al. (Lin et al., 2002) proposed a color-based method to identify and separate specular component from an input image sequence by using a multi-baseline stereo system. Fries et al. (Feris et al., 2004) proposed a multi-flash method to achieve this separation. They used a fixed camera and a varying flashlight to capture images of the scene with different light source positions. Seitz et al.(Seitz et al., 2005) proposed a method for the cancellation of n-bounced inter-reflection light. In this, they first proved the existence of a set of linear operations by applying the inverse light transportation theory. Then, by probing the scene using a very narrow beam of the light, can compute the operators. Michael D. Grossberg et al. (Grossberg et al., 2004) introduced an im-

proved radiometric model to control the appearance of an object in short-range. By using a camera and a projector, they try to make the captured image similar to the projected image.

The above techniques have been largely applied to short-range capture and monocular projection. The work reported in this paper is mainly concerned with long range and stereoscopic projection. We present our investigations in extending the above techniques and/or development of new techniques, as may be needed, due to the new problems posed by long-range capture, calibration and projection.

3 System Overview

An overview of our system is shown in Figure 2. In the first stage, the system is calibrated. This includes calibration of the individual cameras, calibration of each camera with respect to other cameras, and the calibration of the projector with respect to the cameras. Next, the geometry of the projection surface is captured using a structured-light scanning technique. Using the images and geometry of the surface, the reflectance properties at each point are estimated. Surface points with complex reflectance properties i.e. transparent or translucent are identified. Finally, the original stereoscopic content is compensated to account for the reflectance properties of the projection surface prior to projecting it on the surface.

4 System Calibration

Accurate calibration is of imperative importance when dealing with long-range vision-projection systems, such as in this case. A small error in image space i.e. pixels, can lead to vast displacements in the projected space. In this section we describe our system calibration process, which involves:

- the calibration of the cameras
- the pose recovery with respect to the cameras and intrinsic parameter calibration of the projector

4.1 Camera Calibration

Perhaps the most popular technique for calibrating a camera is the one proposed by Tsai et al. (Tsai, 1987) and Zhang et al. (Zhang, 2000). Given a set of points in world space and their corresponding image points, one can recover both the intrinsic and extrinsic parameters of the camera. The pinhole camera model is used to describe these parameters which are specified by the camera matrix C in equation 1,

$$C = \underbrace{\begin{bmatrix} \alpha & -\alpha cot(\theta) & u_0 \\ 0 & \frac{\beta}{sin(\theta)} & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{intrinsic} \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}}_{extrinsic} \tag{1}$$

where $\alpha = kf_x$, $\beta = kf_y$, (f_x, f_y) is the focal length on the x and y axis respectively, θ is the skew angle, u_0, v_0 is the principal point on the x and y axis respectively and r_{1-3}, t_{x-z} determine the camera's rotation and translation relative to the world. Lens distortion is also taken into account and is modeled by:

$$D = \begin{bmatrix} k_1 & k_3 & P_1 & P_2 & k_3 \end{bmatrix} \tag{2}$$

which Bell et al. (Bell et al., 2016) explained in their work.An inherent assumption during camera calibration using the above method is that the captured images of 3D objects with known calibration, say, a checkerboard, are *in-focus*. This is indeed the case for many applications. However, when dealing with projection mapping in outdoor areas this is not the case. The cameras and projector are focused on the projection surface, which is far away. Calibrating the system using a traditional technique means that the checkerboard must be also positioned at the same far distance. Although this does not pose a physical limitation, it most often leads to inaccurate estimation of the parameters. The reason being that at greater distances the checkerboard will only occupy a very small area of the captured image, therefore making the distribution of world-image correspondences degenerate. This leads to erroneous calculations. In particular, distortion parameters cannot be accurately recovered in the case where the captured images of the checkerboard do not provide good coverage across the entire area covered by the camera. In order to overcome this problem, we follow an approach similar to Tyler Bell et al. (Bell et al., 2016) which, instead of a calibration object, uses projected patterns which are by-design robust to out-of-focus cameras.

This method encodes feature points into phase shifted patterns being displayed on a monitor visible to the cameras. The feature points can then be accurately decoded even when blurred because this does not affect the phase of the pattern sequence. One vertical and one horizontal phase map are required where each vertical/horizontal line has a unique phase value. Thus, each pixel appearing on the monitor has a unique pair of (Φ_v, Φ_h) to identify the feature. These phase maps are carried by the phase shifting patterns. Equation 3 is used to generate N equally phase-shifted

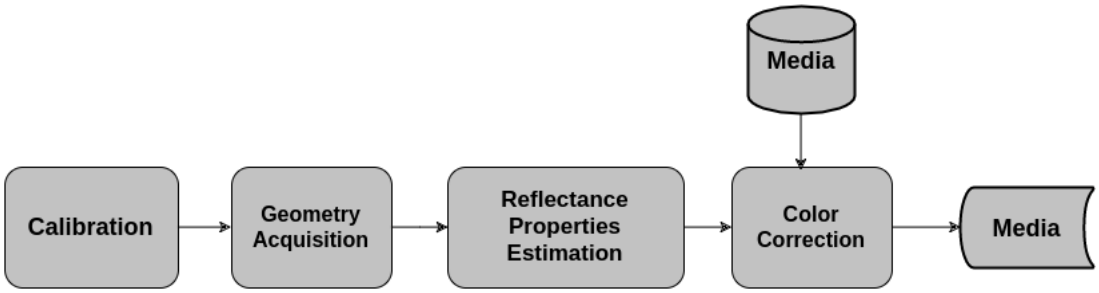


Figure 2: Pipeline

vertical and horizontal fringe patterns and is given by,

$$I_v^i(u,v) = 0.5\left[1+\cos\left(\Phi_v + 2i\pi/N\right)\right], \\ I_h^i(u,v) = 0.5\left[1+\cos\left(\Phi_h + 2i\pi/N\right)\right] \quad (3)$$

where i is the index of fringe pattern, Φ_v and Φ_h represent the vertical and horizontal phase maps, respectively. Each phase-map is then extracted as follows,

$$\phi(x,y) = \tan^{-1}\left[\frac{\sum_{i=1}^{N} I^i \sin\left(2i\pi/N\right)}{\sum_{i=1}^{N} I^i \cos\left(2i\pi/N\right)}\right] \quad (4)$$

where I^i is the intensity of a specific pixel in the i^{th} captured image. This equation generates a non-continuous wrapped phase map with values in the range of $[-\pi,\pi]$. Next, the phase-maps are unwrapped to produce a unique phase value for each of their columns/rows in the pattern. Adding an offset to each section of the phase-map generates an unwrapped phase-map which has a unique phase value at each horizontal/vertical pixel line as shown in the following equation,

$$\Phi(x,y) = \phi(x,y) + k \times 2\pi \quad (5)$$

The process can be summarized as shown in Algorithm 1. Figure 3 shows the encoding of the features using the vertical and horizontal phase maps.

After decoding of the correspondences, the camera can be calibrated using the traditional technique. The result is the intrinsic and extrinsic parameters for each camera. The extrinsic parameters are given with respect to the first (top-left) encoded feature point in the monitor.

4.2 Projector Calibration

The projector is treated as an inverted camera and is calibrated with the traditional method using a set of 2D to 3D correspondences resulting from the geometry acquisition. The explanation on the extraction of the 2D-3D correspondences is deferred to Section 5. Given a set of 3D world points and their corresponding 2D image locations, the projector's intrinsic and extrinsic parameters are recovered.

Algorithm 1: Camera calibration and pose estimation using phase shifting.

1. generate vertical and horizontal fringe patterns
2. display and capture images from monitor at different poses
3. compute wrapped phase-map
4. calculate k in Equation 5
5. calculate unwrapped phase map
6. match the encoded feature points with the decoded phase map
7. calibrate camera using the traditional technique

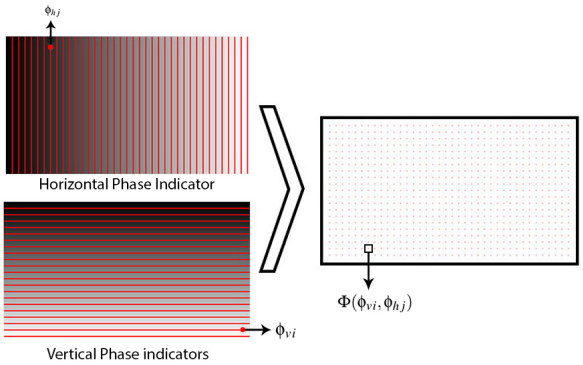


Figure 3: This figure shows the encoding process of the feature points.

5 Geometry Acquisition

The geometry of the projection surface directly impacts the appearance of the surface as well as the projected content. Identifying surface points with complex reflectance properties requires that the geometry (surface points and normals), and reflectance properties (response to light) of the projection surface are known.

Structured-light scanning (Herakleous and Poullis, 2014) is used to capture the geometry of the projection surface. Encoded patterns are projected onto the surface in sequence. The cameras capture one image per pattern. These images are decoded

to produce a dense correspondence between the projector's pixels and the camera's pixels. A render of the reconstructed geometry of the Roman Baths (Figure 1) is shown in Figure 4. This was generated from 44 images captured by three cameras.

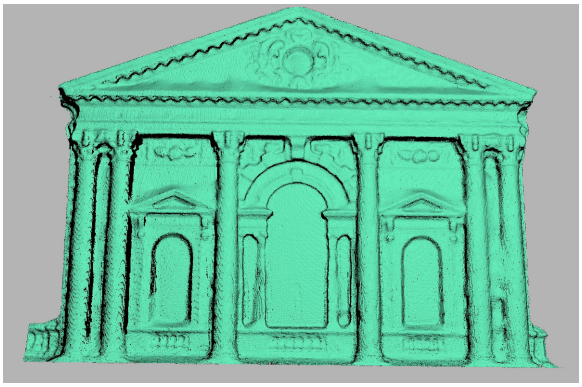


Figure 4: A render of the reconstructed geometry of the Roman Bath. 44 images captured by three cameras.

6 Estimation of Surface Reflectance Properties

As part of the geometry acquisition images are captured from each camera from different viewpoints. Each image records the brightness at every visible surface point from that particular viewpoint. Given three such measurements per surface point, we approximate the local reflectance properties.

We use the Phong illumination model, which is a local illumination model that is easy to compute and use, to describe the local interactions between the material and the light and is given by:

$$I = L\kappa_d \cos\theta + L\kappa_s \cos\phi^{\alpha} \tag{6}$$

where I is the brightness of the reflected light as recorded by the camera, L is the incident radiance emitted from the projector, K_d is the 3-vector diffuse reflection coefficient, K_s is the 3-vector specular reflection coefficient and α is the shininess coefficient of the material.

A non-linear optimization (Levenberg, 1944) is used to compute the optimal values for the material parameters such that the energy function $E(f)$ is minimized,

$$E(f) = E_{data}(f) + E_{smooth}(f) \tag{7}$$

where $E_{data}(f)$, the energy data term and $E_{smooth}(f)$, the energy smoothness term are as defined below.

Energy data term $E_{data}(f)$: This term is a measure of how appropriate the optimized material parameters are given the observed data and is defined as,

$$E_{data}(f) = \sum_{i=0}^{n} |I_r^i - I_m^i|^2 \tag{8}$$

where n is the number of cameras, I_r^i is the rendered image as viewed from camera i using the acquired geometry and the material parameters being optimized, and I_m^i is the observed image captured by camera i.

Energy smoothness term $E_{smooth}(f)$: This term is a measure of the smoothness between brightness values in neighbouring pixels of the rendered image and is given by,

$$E_{smooth}(f) = \sum_{i=0}^{n} \Big[\sum_{j=0}^{w \times h} \Big[\sum_{m}^{8} |B_j - B_m|^2 \Big] \Big] \tag{9}$$

where n is the number of cameras, j is the number of pixels within the rendered image from camera i, m is the 8-neighbourhood around the pixel j, B_j is the brightness at pixel j, and B_m is the brightness at pixel m. This term ensures that the optimal values will provide smooth results. This is illustrated in Figure 5 which shows a comparison between the smoothed/non-smoothed procedures for computing material coefficients. For the statue in Figure 5a the specular map without smoothing can be seen in Figure 5b and with smoothing in Figure 5c. As evidenced, without the smoothness term there is noise between neighbouring pixels which is removed when the smoothness term is introduced.

Figure 6 shows the results of the energy minimizations for three synthetic test/validation cases shown in Figure 6a: a perfectly specular, a diffuse/specular and a perfectly diffuse sphere, respectively. Figures 6b, 6c, 6d show the progress of error minimization corresponding to the shininess, diffuse and specular coefficient respectively.

7 Color Compensation

Changes in the appearance of the projected content caused by the reflectance properties of the projection surface need to be compensated. We use the estimated geometry and reflectance properties at each surface point to compensate the content prior to its projection. For each frame of the stereoscopic content to-be-projected we calculate the compensated projector's brightness L_p for each pixel p as follows,

$$\begin{aligned} L_o &= L_p \left[\kappa_d cos(\theta) + \kappa_s cos(\phi)^{\alpha} \right] \\ L_p &= \frac{L_o}{\left[\kappa_d cos(\theta) + \kappa_s cos(\phi)^{\alpha} \right]} \end{aligned} \tag{10}$$

(a) (b) (c)

Figure 5: (a) The object/projection surface. (b) The estimation of the specular parameters without the smoothness term and (c) the estimation with the smoothness term.

where L_o is the original image's brightness value at pixel p, L_p is the compensated image which, when projected on the surface, ideally yields L_o.

Due to the limited range of brightness values the projector can produce there is a restriction on the colors for which compensation will work.

Further, in this work, the content is stereoscopic which already uses a limited range of values because of the anaglyph processing. We have found that this works well for stereoscopic content being projected on outdoor surfaces such as building facades where the surface is primarily diffuse with strong specular components in the presence of windows, light-fixtures and other such objects.

As previously mentioned, the reflected color from the object depends on the material's reflectance properties and the emitted light. Knowing the reflectance properties of each surface point allows us to compensate up to a factor. First, using additive color mixing we calculate an image which when projected cancels out [if needed] the colors on the projection surface and makes the surface appear grayish. Next, we calculate the image which when projected on top of the 'grayish' surface will be as close to the original as possible. The result is the original image with increased brightness and reduced contrast depending on how bright the grayish image has to be.

8 Experimental results and Concluding Remarks

The proposed technique has been tested and the results are presented. All reported results were generated on an Intel-i7 PC with commodity hardware. The projector used was a Panasonic PT-VW435N projector with a native resolution of WXGA 1280×800. The statue of Alexander shown in the experiments has dimensions $25.5cm \times 19cm \times 14.5cm$ and was used for comparison purposes with (Herakleous and Poullis, 2014).

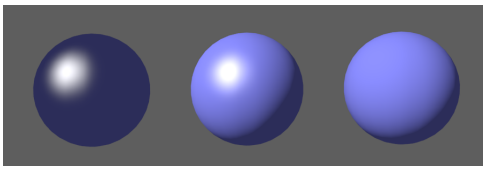
A 3D print of the Roman Baths was used for experimentation due to access restrictions on the real site. These experiments were conducted in relative scale. Figure 1 shows a projection on the real Roman Bath's building. Figure 7a shows a stereoscopic projection [without compensation] being projected onto the Roman Baths. Color distortions occurring on the anaglyph image due to the reflectance properties of the projection surface have a negative impact on the depth perception of the viewer. Figure 7b shows the stereoscopic projection after compensation using the proposed technique. Color distortions are minimized by taking into account the effect of the reflectance properties of the projection surface. The expected projection is shown in Figure 7c.

The expected projection cannot always be achieved [as in the above case] because of the limitations of additive color mixing and the hardware. Intuitively, a projection surface with a white-ish color can reflect a larger percentage of the projected light by the projector, therefore, more colors can be compensated. For example, a bright red color projected onto a white surface will appear as red. On the other hand, a projection surface with a darker color will absorb the projected light; a bright red color projected onto a dark surface will appear as dark red.

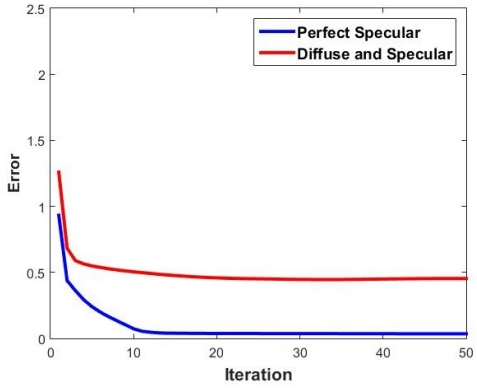
For the immediate future, we will investigate perception-driven compensation, by attempting to preserve major visual features. In addition, we will explore the application of different reflectance models.

9 ACKNOWLEDGMENT

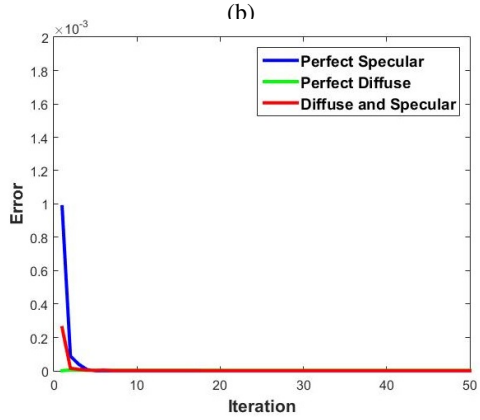
This research is based upon work supported by the Social Sciences and Humanities Research Council of



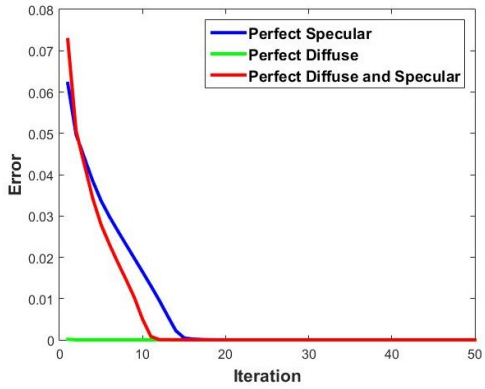
(a)



(b)



(c)



(d)

Figure 6: (a) Three synthetic test cases: perfectly specular, perfectly diffuse, and diffuse/specular. (b) Energy minimization for shininess. (c) Energy minimization for diffuse coefficients. (d) Energy minimization for specular coefficients. (e) Total energy minimization.

Canada under Grant No. SO1936, Concordia University's Faculty of Engineering and Computer Science under Grant No. VH0003, and the Concordia University CASA Research Grant No. CS1136.

REFERENCES

Anaglyph image. `https://goo.gl/9vSgkP`. Accessed: 2016-11-25.

Bell, T., Xu, J., and Zhang, S. (2016). Method for out-of-focus camera calibration. *Applied optics*, 55(9):2346–2352.

Debevec, P., Tchou, C., Gardner, A., Hawkins, T., Poullis, C., Stumpfel, J., Jones, A., Yun, N., Einarsson, P., Lundgren, T., et al. (2004). Estimating surface reflectance properties of a complex scene under captured natural illumination. *Conditionally Accepted to ACM Transactions on Graphics*, 19.

Feris, R., Raskar, R., Tan, K.-H., and Turk, M. (2004). Specular reflection reduction with multi-flash imaging. In *Computer Graphics and Image Processing, 2004. Proceedings. 17th Brazilian Symposium on*, pages 316–321. IEEE.

Furukawa, Y. and Ponce, J. (2010). Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8):1362–1376.

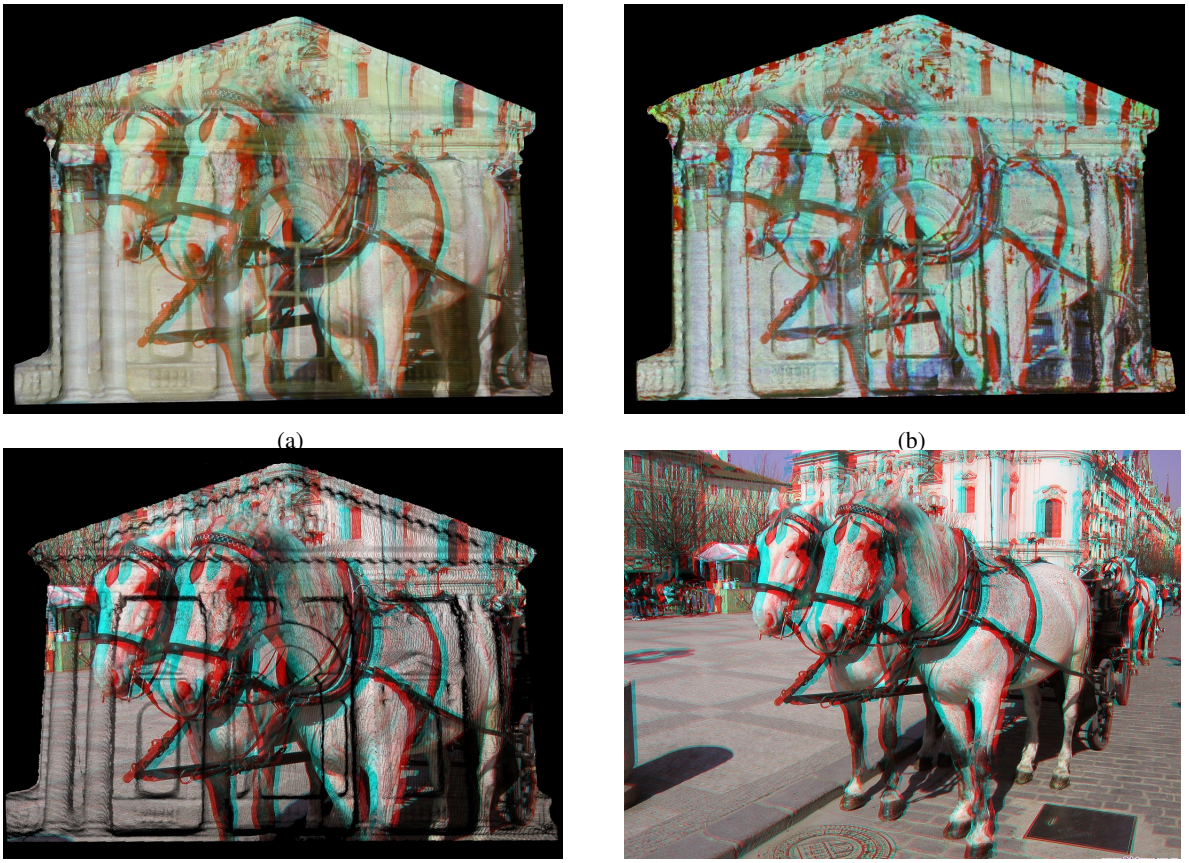
Garcia, R. R. and Zakhor, A. (2013). Geometric calibration for a multi-camera-projector system. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pages 467–474. IEEE.

Grossberg, M. D., Peri, H., Nayar, S. K., and Belhumeur, P. N. (2004). Making one object look like another: Controlling appearance using a projector-camera system. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–452. IEEE.

Herakleous, K. and Poullis, C. (2014). 3dunderworld-sls: An open-source structured-light scanning system for rapid geometry acquisition. *arXiv preprint arXiv:1406.6595*.

Lafortune, E. P. F., Foo, S.-C., Torrance, K. E., and Greenberg, D. P. (1997). Non-linear approximation of reflectance functions. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 117–126, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Levenberg, K. (1944). A method for the solution of certain non–linear problems in least squares.



(a) (b) (c) (d)

Figure 7: (a) A stereoscopic projection without compensation. Color distortions due to the reflectance properties of the projection surface negatively affect the depth perception of the viewer. (b) The stereoscopic projection after compensation using the proposed technique. Color distortions are minimized by taking into account the effect of the reflectance properties of the projection surface. (c) The expected projection. This cannot always be achieved due to the limitations of additive color mixing and the hardware.(d) The original stereoscopic image taken from (ste,)

Lin, S., Li, Y., Kang, S. B., Tong, X., and Shum, H.-Y. (2002). Diffuse-specular separation and depth recovery from image sequences. In *European conference on computer vision*, pages 210–224. Springer.

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

Phong, B. T. (1975). Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317.

Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. (1998a). The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188. ACM.

Raskar, R., Welch, G., and Fuchs, H. (1998b). Spatially augmented reality. In *First IEEE Workshop on Augmented Reality (IWAR98)*, pages 11–20.

Ridel, B., Reuter, P., Laviole, J., Mellado, N., Couture, N., and Granier, X. (2014). The revealing flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts. *Journal on Computing and Cultural Heritage (JOCCH)*, 7(2):6.

Seitz, S. M., Matsushita, Y., and Kutulakos, K. N. (2005). A theory of inverse light transport. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1440–1447. IEEE.

Svoboda, T., Martinec, D., and Pajdla, T. (2005). A convenient multicamera self-calibration for virtual environments. *Presence*, 14(4):407–422.

Torrance, K. E. and Sparrow, E. M. (1967). Theory for off-specular reflection from roughened surfaces. *JOSA*, 57(9):1105–1112.

Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334.