2—2 Enhanced Augmented Reality with Shadows in Naturally Illuminated Environments

Taeone Kim, Ki-Sang Hong*
Department of Electronic and Electrical Engineering
POSTECH

Imari Sato and Katsushi Ikeuchi†
Insititute of Industrial Science
University of Tokyo

Abstract

In this paper, we propose a method for generating graphic objects having realistic shadows inserted into video sequence for the enhanced augmented reality. Our purpose is to extend the work of [1], which is applicable to the case of a static camera, to video sequence. However, in case of video, there are a few challenging problems, including the camera calibration problem over video sequence, false shadows occurring when the video camera moves and so on. We solve these problems using the convenient calibration technique of [2] and the information from video sequence. We present the experimental results on real video sequences.

1 Introuduction

AR(Augmented Reality) in video sequences is in progress under intensive research([3, 4]). AR comes to occur midway between the computer vision and the computer graphics, and is considered to have various applications now and in near future such as computer-guided surgery, robot teleoperation, and special effects for the film industry. The purpose of AR is basically to insert the computer generated objects seamlessly into video sequence or an image, so that the generated images appear to be augmented like part of the original scene.

In this paper, we propose a method for generating graphic objects having realistic shadows inserted into video sequence for the enhanced augmented reality. The shadows of the inserted graphic objects are an important cue for human to feel an illusion that the augmented objects match the real scene well such that as if they exist there originally. Recently, in [1, 5], a new method was developed for estimating the illumination distribution of a real scene from shadows in an image captured by a static camera. By using the occlusion information of the incoming light which is assumed to be infinitely distant from an occluding object they estimated the illumination distribution and generated the shadows of synthetic objects in one image. Our work is to extend their shadow generation method to video sequence obtained by a video camera. Let us suppose that the occluding objects cast their shadows onto a planar surface which is assumed to be Lambertian. Then, from the image of the shadow region on the planar surface, we can estimate the illumination distribution.

The estimation technique of the illumination distribution is briefly explained in section 2. The camera calibration technique for video sequence, which is needed for the estimation of the illumination distribution, is explored in section 3. In section 5.1 is explained the limitation of the illumination distribution estimation with a static camera. We show that this limitation is overcome if we use the information from video sequence. Finally, the results of testing the proposed method on real video sequences are given in section 5.

2 Estimation of Illumination Distribution from Shadows

According to [1], the final image we observe is represented by

$$P(x_k, y_k) = K_d(x_k, y_k) \sum_{i=1}^{n} L(\theta_i, \phi_i) S(\theta_i, \phi_i) cos(\theta_i), \quad (1)$$

where (x_k, y_k) is the image pixel coordinate, K_d the Lambertian constant, L the real illumination radiance, S is the occlusion coefficient. Note that Eq. (1) is summed over the direction, $(\theta_i, \phi_i), i = 1, \cdots, n$, corresponding to the node directions of geodesic dome. For each pixel (x_k, y_k), $S(\theta_i, \phi_i) = 0$ if the corresponding illumination radiance $L(\theta_i, \phi_i)$ is occluded by the occluding object, otherwise $S(\theta_i, \phi_i) = 1$. Let an image which is obtained by removing $S(\theta_i, \phi_i)$ from Eq. (1) be $P'(x_k, y_k)$. Then,

*Address: San 31 Hyoja Dong, Pohang, Kyungbook, 790-784, Korea. E-mail: {kimm,hongks}@postech.ac.kr
†Address: 7-22-1 Roppongi, Minato-ku, Tokyo 106-8558 Japan. E-mail: {imarik,ki}@iis.u-tokyo.ac.jp

the ratio of two images at each pixel position is derived as

$$\frac{P(x_k, y_k)}{P'(x_k, y_k)} = \frac{K_d(x_k, y_k) \sum_{i=1}^{n} L(\theta_i, \phi_i) S(\theta_i, \phi_i) cos(\theta_i)}{K_d(x_k, y_k) \sum_{j=1}^{n} L(\theta_j, \phi_j) cos(\theta_j)}$$

$$= \sum_{i=1}^{n} \frac{L(\theta_i, \phi_i)}{\sum_{j=1}^{n} L(\theta_j, \phi_j) cos(\theta_j)} S(\theta_i, \phi_i) cos(\theta_i),$$

or simply

$$\boldsymbol{Al} = \boldsymbol{b}, \tag{2}$$

where $\mathbf{A}$ is a $m \times n$ matrix having $S(\theta_i, \phi_i)cos(\theta_i)$ as it's components, $\mathbf{l}$ is a $n \times 1$ vector composed of $\frac{L(\theta_i, \phi_i)}{\sum_{j=1}^{n} L(\theta_j, \phi_j) cos(\theta_j)}$ (*illumination radiance ratio*), and $\mathbf{b}$ of the ratios of pixel values. This system is solved linearly by SVD(Singular Value Decomposition) method([6]). The computation of the illumination radiance ratio enables us to do the shading and shadow of synthetic objects in section 4. In fact, the image $P'(x_k, y_k)$ is obtained without the occluding object as described in section 5.

3 Extention to video sequence

The camera calibration is prerequisite for estimating the illumination distribution. The camera calibration for video sequence is, in general, a difficult problem. For this we introduce *virtual camera*([2]) which moves matching closely the motion of the real video camera. It plays a role not only of a calibrated camera for the estimation of the illumination radiance ratio but of a graphic camera for the shading and shadow of synthetic objects.

3.1 Projective Motion and Structure

Given image matching points ($\boldsymbol{x}_{ki}$) over video sequence, we can compute the projective motion($\boldsymbol{P}_k$) and structure($\boldsymbol{X}_i$) that satisfies ([7])

$$\lambda_{ki} \boldsymbol{x}_{ki} = \boldsymbol{P}_k \boldsymbol{X}_i, \quad \text{for} \quad i = 1, \ldots, M, \quad k = 1, \ldots, N, \tag{3}$$

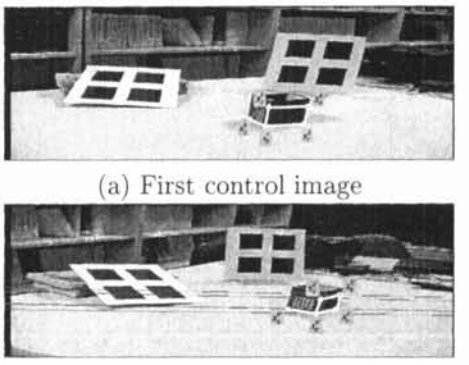
where λ_{ki} is scale factor. Using this reconstruction we can insert efficiently a graphic world coordinate system over video sequence, from which the virtual camera is computed as in the following section.

3.2 Embedding Procedure

To insert a graphic world coordinate system over video sequence, we first insert 5 basis points of the coordinate system into the two selected control images, $\boldsymbol{I}$ and $\boldsymbol{I}'$([2, 8]). The embedding procedure consists of two steps:

1. Insert the graphic world coordinate system into the first control image $\boldsymbol{I}$ by specifying the image locations $\{\boldsymbol{x}_i^b\}_{i=0}^4$ of the Euclidean five basis coordinates $\{\boldsymbol{E}_0, \boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3, \boldsymbol{E}_4\}$ of the coordinate frame.(Fig. 1(a))

2. With the help of the epipolar geometry, we choose four image locations $\{\boldsymbol{x}_i^{b'}\}_{i=0}^3$ of the coordinates $\{\boldsymbol{E}_0, \boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3\}$ in the second control image $\boldsymbol{I}'$.(Fig. 1 (b))



(a) First control image

b) Second control image

Figure 1: Embedding graphic coordinate frames into two selected control images via epipolar constraint

We have illustrated the embedding procedure in Fig. 1. The two images are selected from our experimental sequence. We need not specify the fifth basis position $\boldsymbol{x}_4^{b'}$ in the second control image, because it's position is automatically determined by a 2D homography. Finally, the projective 3D coordinates $\{\boldsymbol{X}_i^b\}_{i=0}^4$ of the inserted image basis points are computed and then the positions of the basis points in the other images $\boldsymbol{I}_k$ are determined by projection with the already reconstructed projective cameras $\boldsymbol{P}_k$ in section 3.1. Note that if the graphic world coordinate system is inserted *correctly*, there is no difference between the graphic world coordinate system and world coordinate system. However, there may be projective distortion as discussed in [2].

3.3 Virtual Camera and its Decomposition

With the help of the inserted basis points of the coordinate system we can compute the zero-skew *virtual camera* $\boldsymbol{P}_k^V$ corresponding to each image using the 3D-2D correspondence, $(\boldsymbol{E}_i, \boldsymbol{x}_i^b)$([2]). To use this virtual camera matrix for a graphic camera, we should decompose it into three parts for all images:

$$\boldsymbol{P}_k^V \sim \boldsymbol{K}_k[\boldsymbol{R}_k|\boldsymbol{t}_k], \tag{4}$$

where $\boldsymbol{K}_k$ is 3×3 zero-skew calibration matrix, $\boldsymbol{R}_k$ rotation matrix and $\boldsymbol{t}_k$ translation.

Finally, we have obtained all virtual cameras for a video sequence. Notice also that the inserted graphic world coordinate system makes it possible to determine the size of a occluding object which has a simple shape such as, for example, the box in Fig. 1.

4 Shading and Shadow

The shading and shadow is done as in [1]. The difference is in that we perform it for each image with the computed virtual camera. We implemented it in OpenGL environment using soft shadows generation technique for multiple light sources.(See appendix of [9]). The shadow generation equation is written as

$$P(x_k, y_k) = \frac{E(x_k, y_k)}{E'(x_k, y_k)} P'(x_k, y_k), \qquad (5)$$

where $E(x_k, y_k)$ is the total image irradiance at a pixel position with occluding objects, $E'(x_k, y_k)$ the one without occluding objects and $P'(x_k, y_k)$ is an image into which synthetic objects are inserted. Finally, $P(x_k, y_k)$ is overlaid with shaded objects.

5 Experimental Results

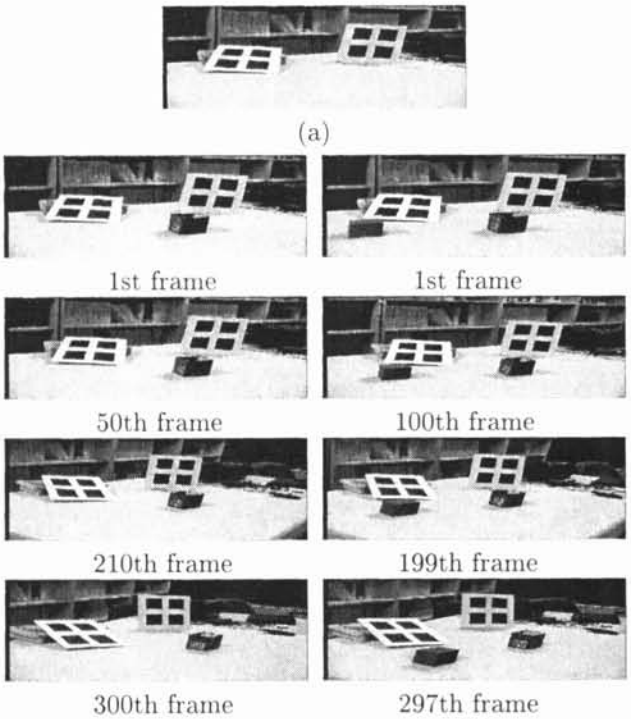


Figure 2: (a) Background image. (b) Exemplary frames of experimental video sequence. (c) Superimposing one synthetic object.

In Fig. 2 is shown our experimental video sequence (b), including a background image without the occluding object (a). The video sequence is captured in an indoor environment under the lighting by two paralleled fluorescent tubes on the ceiling. The background image is captured without the occluding object by a fixed video camera at the first moment, the 1st frame is captured with the occluding object inserted with the same fixed video camera, and then the other frames are obtained with the camera freely moving. Note that we have inserted the planar pattern to facilitate feature tracking. The background image and 1st frame(with a foreground object) is shown in Fig. 3 with an overlaid rectangle region which is used for the estimation of illumination radiance ratio in Eq. (2). The result of estimation is visualized in Fig. 4(a) and the result of superimposing one synthetic object is shown in Fig. 2(c). There are two illuminating regions in Fig. 4(a), each placed in opposite side, which correspond to the real fluorescent tubes on the ceiling. However, notice that there is also a large bright region which is enclosed in white rectangle for the purpose of illustration. This seems to occur due to the limitation of using single image(foreground image) to compute the illumination radiance ratio. That is, the invisible planar surface region in the 1st frame, behind the occluding object, make it not obvious whether there are shadows or not. This problem occurs inherently when we use only single image as in [1], and generates false shadows as shown in Fig. 5 when the camera moves.

Figure 3: (a) Background image. (b) Foreground image.

5.1 Improvements over a static camera

To remove the false bright region in Fig. 4(a), it is utilized that, in video sequence, the occluded planar surface region in the foreground image may be uncovered over video sequence because of camera movement. Using the 2D plane homography, we can find out the pixel values of the covered surface positions in the other images in video sequence. To reduce noise effects we average the pixel value of a covered surface point which is uncovered over multiple images. The homography between two images are easily obtained from virtual camera. Fig. 6 shows the recovered foreground image. This foreground image is used in Eq. (2) instead of the usual foreground image, thus removing the false bright region indicated by the rectangle in Fig. 4(a) as shown in Fig. 4(b).

The SVD solution of Eq. (2) causes the negative radiance ratios because of inherent noise in images. However, the light sources in natural environments should have positive radiance values. To impose this

positiveness on the estimated radiance ratio, we convert Eq. (2) into a non-linear constrained optimization form as

$$\min_{\mathbf{l}}(\mathbf{Al}-\mathbf{b})^2, \quad \text{constrained with} \quad \mathbf{l} \geq \mathbf{0} \qquad (6)$$

Fig. 4(b) shows the result of the estimated radiance ratio using the recovered foreground image and Eq. (6). Note that the illuminating region at the north pole compensates for ambient light. Finally, we shows some frames of augmented video in Fig. 7

(a) Linear solution by SVD (b) Constrained optimization

Figure 4: (a)Visualization of the illumination radiance ratio: upper view of geodesic dome.

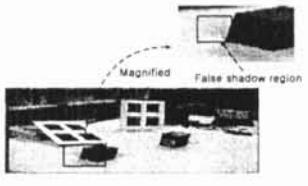


Figure 5: Maginification of 297th frame in Fig. 2 (c)

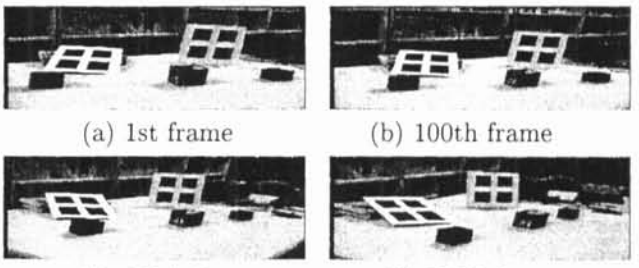
Figure 6: Recovered foreground image

6 Conclusion

In this paper, we suggested a method to extend the work of [1] to video sequences using virtual camera and also improved the limitation of a static camera approach. We remark that further works are needed to solve jittering phenomenon in augmented video.

7 Acknowledgement

We acknowledge this work is funded by Institute of Information Technology Assessment and BK21(Brain Korea) project.



(a) 1st frame (b) 100th frame

(c) 199th frame (d) 297th frame

Figure 7: Results of superimposing two synthetic objects.

References

[1] I. Sato, Y. Sato, and K. Ikeuchi. Ilumination distribution from shadows. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 306–312, 1999.

[2] Yongduek Seo and Ki-Sang Hong. Calibration-free augmented reality in perspective. in print for IEEE Trans. Visualization and Computer Graphics.

[3] Ronald T. Azuma. A survey of augmented reality. *PRESENCE: Teleoperations and Virtual Environments*, 6(4):355–385, August 1997.

[4] K. Kutulakos and J. Vallino. Affine object representations for calibration-free augmented reality. In *IEEE Virtual Reality Annual International Symposium*, pages 25–36, 1996.

[5] I. Sato, Y. Sato, and K. Ikeuchi. Illumination distribution from brightness in shadows: adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions. In *Proc. Int. Conf. on Computer Vision*, pages 875–882, 1999.

[6] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1992.

[7] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Computer Vision - ECCV92*, volume 588 of *Lecture Notes in Computer Science*, pages 563–578. Springer Verlag, 1992.

[8] Yongduek Seo. *Non-Metric Augmented Reality and Flexible Auto-Calibration.* PhD thesis, Department of Electronic and Electrical Engineering, POSTECH, Korea, 2000.

[9] OpenGL ARB. *OpenGL Programming Guide.* Addison Wesley, 1993.