Outdoor Gallery and Its Photometric Issues

Katsushi Ikeuchi[*]
Takeshi Oishi
Graduate School of Interdisciplinary Information Studies
the University of Tokyo

Masataka Kagesawa[†]
Atsuhiko Banno
Rei Kawakami
Institute of Industrial Science
the University of Tokyo

Tetsuya Kakuta[‡]
Yasuhide Okamoto
AsukaLab Inc.

Boun Vinh Lu[§]
Graduate School of Information Science and Technology
the University of Tokyo

Abstract

We have been developing an outdoor gallery system in Asukakyo. Asukakyo is one of the ancient capitals, which is well-known to has lots of temples, palaces and buildings. Nevertheless, most of the assets have been deteriorated after more than fourteen centuries. The outdoor gallery system introduces the virtual appearance of ancient Asukakyo to visitors at the original site with the help of Mixed Reality (MR). To reconstruct the virtual Asukakyo in the outdoor gallery system, it is necessary to handle occlusion problem in synthesizing virtual objects correctly into the real scene with respect to existing foregrounds and shadows. Furthermore, outdoor environment makes the task more difficult due to the unpredictable illumination changes. This paper proposes novel outdoor illumination constraints for resolving the foreground occlusion problem in outdoor environment for the outdoor gallery system. The constraints can be also integrated into a probabilistic model of multiple cues for a better segmentation of the foreground. In addition, we introduce an effective method to resolve the shadow occlusion problem by using shadow detection and recasting with a spherical vision camera. We have applied the method in our outdoor gallery system in Asukakyo and verified the effectiveness of the method.

CR Categories: I.3.7 [Computer Graphics]: Three-dimensional Graphics and Realism—Virtual Reality; I.4.6 [Image Processing and Computer Vision]: Segmentation—Pixel Classification

Keywords: Virtual Reality, Augmented Reality, Geometric Consistency, Occlusion Handling, Photometric Consistency, Shadow Recasting

1 Introduction

Cultural heritage sites, such as Roma or Kyoto, have many interesting historic spots for the tourists. However, on visiting such historic spots, they often find that original cultural assets have been damaged or, in the worst case, vanished. The upper image in Fig. 1 shows such an example in Asukakyo. Asukakyo is one of the ancient capitals, which used to has lots of temples, palaces and buildings. Nevertheless, most of the assets have gone after more than fourteen centuries, and current appearances are those of simply plain rice fields.

Figure 1: *The upper image shows Asuka village nowadays which it is said to have existed a temple named Kawaradera. The lower one illustrates Virtual Asukakyo at the original site.*

In order to entertain the tourists and/or deepen their understanding of such important heritage spots, we have been developing the system, which we name as the outdoor gallery system. The system enables the tourists to see the virtual images of lost assets superimposed on the real background scenes through their goggles as shown in the lower image in Fig. 1. Here, the outdoor gallery system displays the original appearances of the cultural assets at the exact spot, where such assets used to exist. This is an important difference from a usual virtual reality (VR) system, which displays those images on the indoor theater screen at different place from the original historic spot. The outdoor gallery provides the understanding of the value of the original historic spot.

Asukakyo provides an excellent experimental field for developing the outdoor gallery system. Asukakyo indicates the generic name

[*]e-mail:{ki, oishi}@cvl.iis.u-tokyo.ac.jp
[†]e-mail:{kagesawa, vanno, rei}@cvl.iis.u-tokyo.ac.jp
[‡]e-mail:{kakuta, okamoto}@asukalab.co.jp
[§]e-mail:lbvinh@cvl.iis.u-tokyo.ac.jp

VRCAI 2010, Seoul, South Korea, December 12 – 13, 2010.
© 2010 ACM 978-1-4503-0459-7/10/0012 $10.00

of the capital where the palaces of various emperors were said to be in the Asuka area from the late sixth to the late seventh centuries. The period is well-known for its significant transformation in art, culture and politics in Japanese society. Though many tourists visit the historic spots of Asuka village, most of them cannot visualize the original assets and there is no way other than to rely on imagination to visualize the spectacle of Asuka at that time. Since Asuka village is a part of the Law Concerning Special Measures for Preservation of Historic Natural Features in Ancient Cities and there is strict development control, it is difficult to build a new museum or replica for the tourists. Moreover, excavation still continues and there is a possibility that the restoration plan will change with a newly-discovered archaeological fact. Against such a background, the outdoor gallery system is one of the most effective ways to depict those assets by overlaying virtual appearances of those assets on the real spot without any side effects such as damaging the historic spot or converting the rice field into man-made replica.

The difference between the traditional MR system and the outdoor gallery system exists in the former having the wider concept for providing a total service of tourism and education of histories. We assume that it is important to make a cyclic framework for tourism in which the users visit the historical site again and again for the revitalization of local regions. To realize this framework, we have to use network and transportation systems efficiently. Tourists are motivated to visit the historical site by the Internet website and learn about the historical facts concerning to the site in the traditional museums. They move within the historical site by using transportation systems like Personal Mobility Vehicles (PMV). MR technologies can virtually reconstruct lost cultural heritage assets on the real scene and provides a deep impression to the users. They upload their experiences by movies and photographs etc. on the Internet, which helps to motivate them to visit the historical site again.

The outdoor gallery system needs to handle both geometric and photometric consistency. Geometric consistency provides the coordinates consistency between the virtual objects and background scenes so that the virtual objects are consistently rendered into the background scene. In our outdoor gallery system, this has been achieved through the hardware [Kakuta et al. 2005].

Another geometric consistency issue exists in foreground/background consistency. In operating the outdoor gallery system, it is inevitable for many tourists around the system to appear in the background images. We conducted a preliminary experiment to superimpose virtual objects over the scenes with tourists. From this experiment, we have found that the observer of the outdoor gallery system felt no inconsistency, when a tourist farther with respect to the virtual objects was occluded. On the other hand, they felt strong inconsistency, when a nearby tourist was occluded; they felt that the virtual objects are floating in the air. Thus, we concluded that it was necessary to decide which tourist was nearer and which tourist was farther with respect to the virtual objects. We will refer to this issue as the foreground/background consistency problem.

Regarding the foreground/background consistency problem, this paper sets up foreground segmentation with complicated illumination conditions, decides which regions are near based on the ground-plane assumption, and finally pastes nearer regions as is, and far regions occluded. In Fig. 2, the spherical vision camera, depicted as VIDEO, provides background images, and then extracts foregrounds (tourist regions) with shadow regions surrounding them. For this foreground segmentation, we introduce a set of illumination constraints to handle both the sudden illumination change and the gradual one. The multiple cue-based segmentation in [Kakuta et al. 2008] is extended by adding the illumination and the motion cue with background attenuation. Then, the depth of the region is estimated by assuming the ground plane and the relative positions of the regions as proposed in [Kakuta et al. 2008].

Photometric consistency provides the similar appearances of the virtual objects as those of surrounding background objects. The brightness of the virtual objects is adjusted by measuring the illumination condition from the sky part of the spherical image. This is achieved using the spherical image from the VIDEO module. In addition, the natural shadows are re-casted around the virtual objects with respect to the coherence of the location and the illumination conditions [Kakuta et al. 2005].

Related to the foreground/background consistency problem of the sightseers in the background images, shadows around them are also necessary to be re-casted properly. There are three cases. Shadows around a person farther with respect to the virtual objects are simply occluded by them. In case shadows around a nearer person are not overlapping to the virtual object, the shadow region should be pasted as is the case of the nearer person's image. In case that those shadows are overlap the virtual object, those not occluded shadow regions are pasted as is, while occluded regions are pasted differently on the virtual objects.

For this shadow occlusion issue, this paper introduces an effective method to extract those shadow regions. First, with the camera sensitivity, shadow region is detected by using the illumination invariant constraint with energy minimization depicted as shadow region in Fig. 2. The detected shadow area is then used to estimate light direction from the estimated height of the corresponding foreground. Finally, real-time shading and shadow in [Kakuta et al. 2005] is applied to render virtual objects.

The remainder of this paper explains the details of the system. Section 2 describes two new constraints for segmenting foregrounds and detecting shadow regions under varying illumination conditions. Section 3.1 explains how to use those constraints for shadow detection, and Section 3.2 handles how to re-cast those detected shadows on the background images. Section 4 evaluates the system performance, and Section 5 concludes this paper.

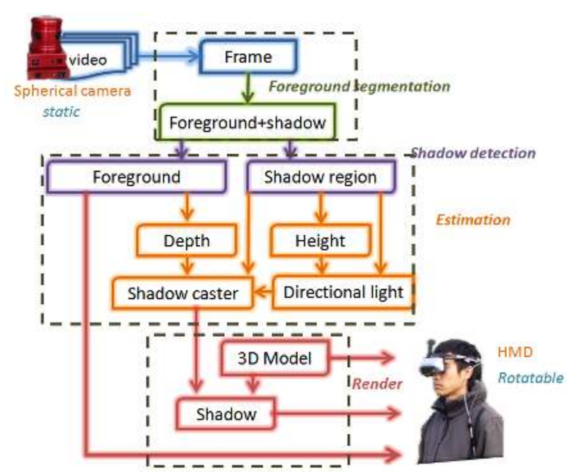

Figure 2: *An overview of the whole system.*

2 Foreground occlusion handling

Regarding the foreground segmentation process in occlusion handling, besides illumination changes, the outdoor scene challenges the task due to the occupation of moving foreground and moving background such as trees, leaves and clouds. Therefore, it is wise to combine different available cues together to segment the foreground online because a single cue is not reliable. We extend the work of Kakuta *et al.* [Kakuta et al. 2008] by adding the illumination cue and the motion cue with background attenuation which are explained in detail in [Lu et al. 2010].

In an outdoor scene, changes in illumination are inevitable and challenging to most researchers in the computer vision field. In this section, we propose an effective solution to handle sudden changes in illumination in most cases of day light condition ranging from a sunny one to a heavy cloudy one. The illumination constraints will be explained in Sect. 2.1 and 2.2. Both constraints can be used as an illumination cue which can be integrated into a multiple-cue background model.

2.1 Illumination invariant constraint

Let's assume that a surface patch S_t at time t is Lambertian with normal $\vec{n}$ and the corresponding *surface reflectance* ρ. Our proposed outdoor illumination model with the visible portion of the sky reads

$$I_t = \rho E_t \tag{1}$$

where

$$E_t = \left(g E_t^{sun} \left(\vec{n} \vec{D}_t \right) + E_t^{sky} \cos^2 \frac{\beta}{2} \right)$$

and I_t denotes the corresponding *image irradiance* and E for the *irradiance* at the surface patch.

If we assume that the camera sensitivity is sufficiently narrow and that daylight is blackbody radiation, we can easily infer the constraint among the log-ratios of *image irradiances* $\ln R_{t_1,t_2}^{\lambda_R}$, $\ln R_{t_1,t_2}^{\lambda_G}$ and $\ln R_{t_1,t_2}^{\lambda_B}$ as in [Lu et al. 2010]:

$$d_c(I_1, I_2) = \| \ln R_{t_1,t_2}^{\lambda_R} - C \ln R_{t_1,t_2}^{\lambda_G} + (C-1) \ln R_{t_1,t_2}^{\lambda_B} \| = 0 \tag{2}$$

where $C = \left(\frac{1}{\lambda_R} - \frac{1}{\lambda_B}\right) / \left(\frac{1}{\lambda_G} - \frac{1}{\lambda_B}\right)$. We call Eq. (2) the illumination invariant constraint.

2.2 Illumination ratio constraint

Furthermore, we can also obtain

$$\lambda \ln R_{t_1,t_2}^{\lambda} = -c_2 \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \tag{3}$$

Since it is obvious that the right side of Eq. (3) is independent of the wavelength, another constraint which we call the illumination ratio constraint can be obtained as :

$$\lambda_R \ln R_{t_1,t_2}^{\lambda_R} = \lambda_G \ln R_{t_1,t_2}^{\lambda_G} = \lambda_B \ln R_{t_1,t_2}^{\lambda_B} \tag{4}$$

Furthermore, the flow of segmentation process which explains how the constraints are integrated into the robust segmentation to handle large changes in illumination can be found in [Lu et al. 2010].

3 Shadow occlusion handling

In order to handle the shadow occlusion problem, it is necessary to separate the shadow region from detected foreground and then to recast the shadow into virtual objects correctly. In addition, the foreground depth and height, and the light direction should also be estimated. Shadow detection comes next in section 3.1 and then it is followed by how we recast the detected shadow in section 3.2.

3.1 Shadow detection

Natural shadow in an outdoor scene is normally formed by the fact that only sky light reaches the surface area without the participation of the Sun light. With the assumption of blackbody radiation and the narrow-banded camera, we can apply the proposed illumination invariant constraint in Eq. (2) to detect shadow at per-pixel level with the observation that the brightness of shadow is lower than that under sunlight (*brightness constraint*). Here we introduce a region-based optimization by using energy minimization. The shadow energy can be represented as :

$$E_{shadow} = E_{data}(X^t, I^t, I_B^t) + \alpha E_{smooth}(X^t, I^t, I_B^t) \tag{5}$$

The distance from Eq. (2) is then used to form the shadow likelihood as :

$$E_{data}(X^t = B, I^t, I_B^t) = \sum_{i_r \in I} -\log(d_c(i_r^t, i_B^t)) \tag{6}$$

The smoothness term in (5) is defined as in [Sun et al. 2006]. Finally, the optimum label can also be optimized using energy minimization with graph cut as explained in [Lu et al. 2010].

3.2 Shadow recasting

With the help of a spherical vision camera, the estimation of the foreground depth and height as well as the directional light can be done with simple spherical geometry. Subsequently, the shadow mapping method is applied.

With the assumption that the ground is relatively flat and that the camera height is known, the foreground depth can be easily estimated by using simple spherical geometry calculation.Additionally, in order to recast the detected shadows, the light direction should be known; hence, the foreground height should be taken into account. Inheriting from the depth estimation, the foreground height is estimated by using spherical geometry.

Next, we assume that the detected foreground shadows are all casting shadows caused by the Sun light. Thus, the light source can be safely assumed to be directional. In order to estimate the directional light vector, foreground and its corresponding shadow should be projected onto the ground from the spherical image by using the spherical transformation. Traditional shadow mapping is used to correct the shadows on virtual objects.

4 Experimental results

In our experiment, the spherical vision camera Ladybug2 by Point Grey Research Inc. is used and fixed on a tripod at the height of around 160 *cm*. First, we experimented on video sequences captured in our campus and then applied in our project. In the system, we use a notebook, the spec of which is, CPU: Core2Quad 2.0 GHz, RAM: 4GB, GPU: nVIDIA GTS 160M 1024MB. As introduced in Fig. 2, we fix the spherical vision camera during the experiment while the viewers can move the head-mounted display (HMD).

Regarding the experiments on illumination change, we compare our foreground segmentation results with respect to the previous method in [Kakuta et al. 2008] which is a combination of [Criminisi et al. 2006] and [Sun et al. 2006]. Figure 3 shows that previous methods fail when the outdoor illumination changes. Meanwhile, our proposed outdoor illumination handling achieves correct foreground segments. The Table 1 shows that our proposed method achieves stable results regardless of illumination changes.

Figure 3: *Campus sequence. The images in the first row are original ones. The second row shows how the previous approach fails. The third row indicates probability of foreground from illumination constraints. The bottom row shows the results by our approach.*

Table 1: *Asuka and campus sequences.*

Sequence	*Method*	*Precision(%)*	*Recall(%)*
Asuka	Proposed method	92	97
	[Kakuta et al. 2008]	7	99
Campus	Proposed method	94	97
	[Kakuta et al. 2008]	56	99

The result in Fig. 4 illustrates how the shadow is recast. The recast shadow and the self-shadows on virtual objects match well with the natural ones of the foregrounds, from which we estimate the directional light. It is necessary to consider the stability of the directional light which changes very slowly in reality. A possible solution is the temporal consistency for the light direction.

Finally, with the above spec, our segmentation system can work at 5 frames/sec with the image resolution at 2048 by 1024 and at 2 frames/sec with shadow recasting. Improvement for the faster implementation is required to handle the foreground and shadow occlusion problem in outdoor augmented reality in real-time.

5 Conclusion

This paper describes the outdoor gallery system developed for Asukakyo. The outdoor gallery system introduces the virtual appearance of ancient cultural assets to visitors at the original site with the help of MR techniques. In this outdoor gallery system, we required to address foreground/background consistency issues. Our proposed illumination constraints and foreground segmentation strategy for the consistency handle well the cases of outdoor illumination changes ranging from the gradual change to the sudden one. Thus, foreground occlusion problem is resolved for the outdoor gallery system. In addition, shadow recasting results show that our solution to the shadow occlusion works effectively. Real experiments on campus and in the outdoor gallery system in Asuka project prove the effectiveness of the introduced method.

Acknowledgements

This research was supported, in part, by MEXT digital museum project.

References

CRIMINISI, A., CROSS, G., BLAKE, A., AND KOLMOGOROV, V. 2006. Bilayer segmentation of live video. In *CVPR (1)*, 53–60.

KAKUTA, T., OISHI, T., AND IKEUCHI, K. 2005. Shading and shadowing of architecture in mixed reality. In *ISMAR*, 200–201.

KAKUTA, T., VINH, L. B., KAWAKAMI, R., OISHI, T., AND IKEUCHI, K. 2008. Detection of moving objects and cast shadows using a spherical vision camera for outdoor mixed reality. In *VRST*, 219–222.

LU, B. V., KAKUTA, T., KAWAKAMI, R., OISHI, T., AND IKEUCHI, K. 2010. Foreground and shadow occlusion in outdoor augmented reality. In *ISMAR (To be appeared)*.

SUN, J., ZHANG, W., TANG, X., AND SHUM, H.-Y. 2006. Background cut. In *ECCV (2)*, 628–641.

Figure 4: *The left column indicates the original frames in the sequences. The right one shows the final results of occlusion handling in our system.*