

3D VISION TECHNIQUES AND APPLICATIONS TO PHOTO-REALISTIC SCENE RECONSTRUCTION

Jun-Sik Kim, Ho-Won Kim and In-So Kweon

Department of Electrical Engineering and Computer Science, KAIST, Daejeon, Korea
jskim@rcv.kaist.ac.kr, imp97@kaist.ac.kr, iskweon@kaist.ac.kr

Abstract: We present some novel concepts for the scene geometry recovery by a multi-camera system. A new global appearance measure is introduced for a novel generalized scene recovery methodology, called “the appearance-cloning.”

For the efficient calibration of the multi-camera system, we investigate the projective properties of the concentric circle pattern. Specifically, two algebraic constraints from concentric circles are defined to calibrate the intrinsic parameters of the multi-camera system. Through a variety of scene reconstruction experiments, we demonstrate that the use of concentric circles greatly simplifies the calibration problem for the photo-realistic recovery of dynamic 3D scene.

Keywords: photorealistic 3D reconstruction, volume based approach, camera calibration

1. INTRODUCTION

Recent advances in image-based rendering approaches (view morphing, light field rendering) make it possible generate a photo-realistic novel view only from photographed images without scene geometry information [1-5]. Many applications, however, still require the correct scene geometry information in image based modeling and rendering (IBMR).

Computer vision techniques to extract the geometric information of real-world scene for IBMR can be categorized according to the sensing technology into either the active or passive sensing approaches.

Active sensing technologies (e.g., optical, laser, and time-of-flight range finder) normally provide high accuracy and good quality of 3D measurements. Many commercial solutions have been widely applied to various IBMR systems that use the scene geometry information. However, in most cases, the geometry information acquired from the active sensing technology should be combined with photographs for the photo-realistic 3D model generation and rendering. Moreover, the 3D scanner is still very expensive for the general consumers compared with the passive sensors such as a digital camera and cam-coder.

On the other hand, the passive sensing technology directly recovers the scene geometry information from a set of photographs, so no additional sensor is needed for IBMR. Moreover, digital cameras are available at a reasonable price. Therefore, 3D scene recovery by the passive sensing technology has attracted a great deal of attentions during the last several decades, and many elegant methodologies have been introduced and demonstrated those viabilities at each specialized application area and conditions. However, those methodologies are still hard to be a solution for our generalized scene

recovery problem because of many hard-to-set ad hoc factors such as scene dependent control parameters, assumptions, and constraints [6-15]

For a methodology to be a practical solution for the generalized scene recovery problem, it should not rely on any ad hoc parameters to be set.

In this paper, we introduce a novel “global appearance measure” instead of local measures, and then propose a novel methodology, called “the appearance-cloning”, as a solution to the generalized scene recovery problem.

For the scene recovery by a multi-camera system, an effective camera calibration method is necessary to deal with the partial views of calibration pattern by the multi-camera system. We develop a novel calibration pattern of concentric circles that can be extracted even with small partial view of the circle pattern. A two-step camera calibration method estimates the intrinsic camera parameters and the camera pose parameters. The proposed constraint properties of concentric circle patterns have concrete physical meanings, such as the position of the projected circle center and the projected radius-0 imaginary circle. Based on the constraints of the concentric circles, the intrinsic parameter is estimated from the image of the concentric circles without any assumption about a camera with three or more views.

2. APPEARANCE CLONING

Since the recovery of real-world scene from a set of photographs is generally ill-posed problem [6], the best that we can recover from photographs is the maximal photo-consistent 3D model that is re-projected into all the original photographs and reproduces those photographs [7]. Many previous researches have tried to find this maximal photo-consistent 3D model from photographs, and recent several volumetric approaches have introduced

promising methodologies by representing the recovered shape information in a common volumetric space defined by user.

The use of this common 3D space elegantly solves the view-dependent self-occlusion problem of arbitrarily-shaped scene, and makes the traditional “correspondence search problem” in 2D image spaces to be reformulated as the “occupancy decision problem” of each 3D primitive in the volumetric space by using the normalized cross-correlation measure of oriented 3D patch on a grid point [8] or the photo-consistency measure in a voxel [9].

Now, how can we recover the photo-consistent 3D model by using a global appearance measure? As shown in Figure 1, the global appearance measure is the photograph itself that is projected from the real-world scene to a known camera viewpoint. We can make the equivalent system in the digital domain in which we can control full operations. We first define a candidate 3D model in our digital domain, and then render this model to the known camera viewpoints. Through this process, we can get another global appearance which is composed of view-dependent appearances of candidate 3D model and has one-to-one correspondence with the given global appearance.

Then, we can get the photo-consistency of candidate 3D model about given photographs by directly comparing the appearance similarity between the given appearances and recovered appearances. And then, we can control our 3D model by using the appearance similarity error because we exactly know the relationship between the recovered appearance and our 3D model. Through these iterative operations, we can find the optimal photo-consistent 3D model by cloning the global appearance of given photographs. We call this methodology as “the appearance-cloning.” If our equivalent system can correctly resemble the actual projection relationship from the real-world scene to each camera viewpoint and enough appearances are given to us, then the recovered 3D model can perfectly regenerate given photographs and can be a photo-consistent clone of the real-world scene according to the chain-rule.

2.1 Modeling of digital 3D model space

To represent the real-world scene in our digital domain, we use the voxel as a basic primitive. We define the voxel as a single-colored opaque cubic volume. Consequently, the state of each voxel can be either opaque or transparent in the voxel space.

Now, define a voxel space $\mathcal{V} \in \mathbb{R}^3$ that encloses the scene with a comfortable margin and is composed of (L, H, W) voxels. Since we define the voxel as a cubical volume, each voxel $v \in \mathcal{V}$ projects to several pixels per visible camera (i.e., follows the volume-projection). Then, the scene can be approximately represented as a combination of opaque and colored surface voxels. Let such a

combination of opaque surface voxels be a candidate 3D model \mathcal{S} . Then, there are $2^{L \times H \times W}$ candidate 3D models. The optimal photo-consistent 3D model \mathcal{S}^* that we want to find is one of those candidate 3D models. Let $2^{\mathcal{V}}$ be the set of those candidate 3D models. Then, $\mathcal{S}^* \in 2^{\mathcal{V}}$. We call $2^{\mathcal{V}}$ digital 3D model space.

These complete candidate 3D models can be rendered to each camera viewpoint, which is used to capture given appearance, and then we can get the recovered appearances from those candidate 3D models.

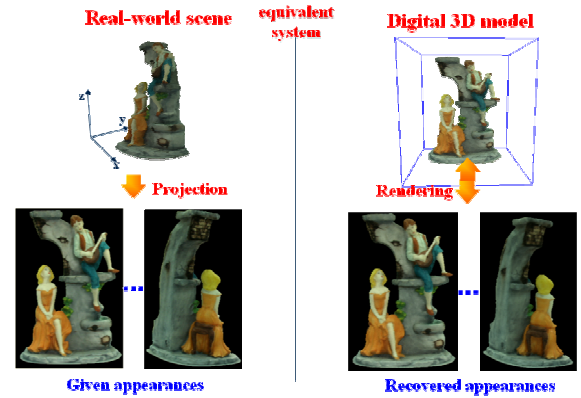


Figure 1. Appearance cloning: Left) the real-world projection relationship. Right) the equivalent digital-world rendering relationship.

2.2 Appearance similarity measure

Now, we can directly compare the appearance similarity between given appearances and recovered appearances. However, in most of scene recovery from multi-view images, the real-world scene is not observed in whole image region. We adopt the region of interest (ROI) to exclude those projected regions of uninteresting scenes. If our interesting scene is a single real-world object, ROI of each image is exactly the same with the inside silhouette region of projected object. However, since we can directly recover multiple real-world objects or landscape, we use the notion of ROI rather than silhouette. In actual scene recovery, this ROI can be automatically extracted by using the blue screen technique, or manually extracted by using a tool like the intelligent scissor [16] or its variants.

From now on, suppose that the ROI information of the given input images is known as a prior knowledge. Before defining appearance similarity measure of a candidate 3D model, we use several notations. Let \mathcal{I} be the given appearances within ROI of input photographs and $\mathcal{R}(\mathcal{S})$ be the one-to-one corresponding recovered appearances rendered from a candidate 3D model \mathcal{S} . Let \mathcal{P} be the set of all ROI pixels in the given input photographs. Then, we can derive the uniqueness constraint from the ROI information of input photographs:

[Uniqueness constraint]

If a candidate 3D model \mathcal{S} is to be a candidate of photo-consistent 3D model, every pixel $p \in \mathcal{P}$ on the recovered appearances $\mathcal{R}(\mathcal{S})$ must have its own occupying voxel $v \in \mathcal{S}$ and each surface voxel $v \in \mathcal{S}$ should not project to the outside of ROI at any camera viewpoint.

This uniqueness constraint greatly reduces the number of meaningful candidate 3D models in our 3D model space $2^{\mathcal{V}}$. Let $2^{\mathcal{CV}}$ be the set of candidate 3D models satisfying the uniqueness constraint. Then, $2^{\mathcal{CV}} \subset 2^{\mathcal{V}}$ and $\mathcal{S}^* \in 2^{\mathcal{CV}}$. From now on, we consider a candidate 3D model $\mathcal{S} \in 2^{\mathcal{CV}}$, except where noted.

To define the appearance similarity measure $\Pr(\mathcal{S} | \mathcal{I})$ of a candidate 3D model \mathcal{S} , we use the Bayes' theorem

$$\Pr(\mathcal{S} | \mathcal{I}) = \frac{\Pr(\mathcal{I} | \mathcal{S}) \Pr(\mathcal{S})}{\sum_{\mathcal{S}^k \in 2^{\mathcal{CVH}}} \Pr(\mathcal{I} | \mathcal{S}^k) \Pr(\mathcal{S}^k)}. \quad (1)$$

Here, if the prior probability is uniform for all candidate shapes, then $\Pr(\mathcal{S} | \mathcal{I}) \propto \Pr(\mathcal{I} | \mathcal{S})$.

Let $\mathbf{c}_I(p)$ be the pixel color of given appearance image at a pixel $p \in \mathcal{P}$ and $\mathbf{c}_R(p | \mathcal{S})$ be the pixel color of corresponding recovered appearance image under the candidate 3D model \mathcal{S} . Then, the appearance similarity measure of a candidate 3D model \mathcal{S} can be defined as

$$\Pr(\mathcal{I} | \mathcal{S}) = \prod_{p \in \mathcal{P}} p[\mathbf{c}_I(p), \mathbf{c}_R(p | \mathcal{S}) | \boldsymbol{\theta}] \quad (2)$$

where $p[\circ | \boldsymbol{\theta}]$ is a probability model of each pixel $p \in \mathcal{P}$, which has a control parameter vector $\boldsymbol{\theta}$, that returns the reliability of $\mathbf{c}_R(p | \mathcal{S})$ about $\mathbf{c}_I(p)$.

Now, the remainder for the appearance similarity measure is how to define $p[\circ | \boldsymbol{\theta}]$. Although we have to use the probability model because of volumetric approximation of scene surface, we can inversely use the properties of volume-projection to model $p[\circ | \boldsymbol{\theta}]$.

2.3 Per-pixel statistical modeling

Since $\mathbf{c}_I(p)$ is given as an evidence, we have to model the probabilistic characteristics of $\mathbf{c}_R(p | \mathcal{S})$ in $p[\circ | \boldsymbol{\theta}]$. In practice, $\mathbf{c}_R(p | \mathcal{S})$ is the color of which a voxel ($v \in \mathcal{S}$) occupy that pixel ($p \in \mathcal{P}$). Therefore, $\mathbf{c}_R(p | \mathcal{S})$ is the function of candidate 3D model and we cannot directly estimate its characteristics in the 3D model space. On the other hand, if we approach inversely from the image space, we can estimate its characteristics from the physical

volume-projection relationship between a pixel and its occupying voxels.

The evidence that is given to use is only the view-dependent appearances of real-world scene. Fortunately, we have a solution, which can estimate candidate $\mathbf{c}_R(p | \mathcal{S})$ (i.e., voxel color) from neighborhood color information centered on each pixel $p \in \mathcal{P}$, by using the physical volume-projection relationship.

To derive per-pixel probability model, we use the Euclidean color distance as a color similarity measure

$$x(p | \mathcal{S}) = \|\mathbf{c}_R(p | \mathcal{S}) - \mathbf{c}_I(p)\| \quad (3)$$

and assume that the intensity difference of each color-channel follows the zero-mean Gaussian distribution.

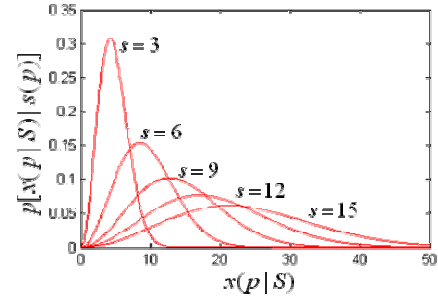


Figure 2. Maxwell distributions according to the control parameter s

Then, the random variable $x(p | \mathcal{S})$ follows the Maxwell distribution [17]

$$p[x(p | \mathcal{S}) | s(p)] = \sqrt{\frac{2}{\pi}} \frac{x^2}{s^3} \exp\left(-\frac{x^2}{2s^2}\right) \quad (4)$$

where $x \in [0, \infty)$. $s(p)$ is the control parameter of Maxwell model.

In this Maxwell model, the control parameter $s(p)$ guides $x(p | \mathcal{S})$ to be the statistically optimal distance $\sqrt{2}s(p)$, and strongly constrains the allowable $x(p | \mathcal{S})$, as shown in Figure 2. Therefore, we can substitute this self-constrained per-pixel Maxwell model $p[x(p | \mathcal{S}) | s(p)]$ for $p[\circ | \boldsymbol{\theta}]$ in Eq. (3). Then we can get the reliability of recovered appearance color $\mathbf{c}_R(p | \mathcal{S})$ about $\mathbf{c}_I(p)$ from the per-pixel Maxwell model.

Now, we can estimate per-pixel Maxwell model for each pixel $p \in \mathcal{P}$ by fitting the control parameter $s(p)$ to the color statistics of candidate voxel colors.

Fortunately, the control parameter $s(p)$ of Maxwell model has following relationship with $x(p | \mathcal{S})$

$$s(p) = \frac{1}{2} \sqrt{\frac{\pi}{2}} E[x(p | \mathcal{S})] \quad (5)$$

where $E[x(p | \mathcal{S})]$ is the expectation of $x(p | \mathcal{S}) = \|\mathbf{c}_R(p | \mathcal{S}) - \mathbf{c}_I(p)\|$ centered on $\mathbf{c}_I(p)$.

Since we can substitute $\mathbf{c}_R(p | \mathcal{S})$ with our candidate voxel color, the per-pixel Maxwell model can be estimated with the closed-form solution. Let $\tilde{\mathbf{c}}(p)$ be the candidate voxel color that is estimated from two constraints of volume-projection. Then, we can estimate this $\tilde{\mathbf{c}}(p)$ with very simple shifted average footprint operation.

3. MULTI-CAMERA CALIBRATION USING CONCENTRIC CIRCLES

3.1 Algebraic constraints: rank deficiency

Without loss of generality, we can assume that the concentric circle center coincide with the origin of the world coordinate system. The circle \mathbf{Q} is a diagonal matrix whose elements are $[1, 1, -\rho^2]$. From these algebraic properties, we can derive some algebraic constraints of the projected concentric circle matrices.

Two ellipses from the same homography $\mathbf{P} = \mathbf{C}[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{T}]$ satisfy

$$\lambda_1 \mathbf{A}_1 = \mathbf{P}^T \mathbf{Q}_1 \mathbf{P}^{-1}, \quad \lambda_2 \mathbf{A}_2 = \mathbf{P}^T \mathbf{Q}_2 \mathbf{P}^{-1} \quad (6)$$

where \mathbf{C} is a camera matrix and \mathbf{r}_n is a rotation vector.

Taking the inverse on Eq (6) gives

$$\frac{1}{\lambda_1} (\mathbf{A}_1)^{-1} = \mathbf{P} (\mathbf{Q}_1)^{-1} \mathbf{P}^T, \quad \frac{1}{\lambda_2} (\mathbf{A}_2)^{-1} = \mathbf{P} (\mathbf{Q}_2)^{-1} \mathbf{P}^T \quad (7)$$

The circle matrices \mathbf{Q}_1 and \mathbf{Q}_2 are simply diagonal as

$$\mathbf{Q}_1 = \text{diag}(1, 1, -\rho_1^2), \mathbf{Q}_2 = \text{diag}(1, 1, -\rho_2^2) \quad (8)$$

Their inverse matrices are

$$(\mathbf{Q}_1)^{-1} = \text{diag}(1, 1, -\frac{1}{\rho_1^2}), (\mathbf{Q}_2)^{-1} = \text{diag}(1, 1, -\frac{1}{\rho_2^2}) \quad (9)$$

Subtracting Eq. (9) from each other gives

$$\mathbf{A}_\Delta^{-1} \sim \mathbf{P} \mathbf{Q}_\Delta^{-1} \mathbf{P}^T \quad (10)$$

where

$$\mathbf{A}_\Delta^{-1} = (\mathbf{A}_1)^{-1} - \frac{\lambda_1}{\lambda_2} (\mathbf{A}_2)^{-1} = (\mathbf{A}_1)^{-1} - \alpha (\mathbf{A}_2)^{-1} \quad (11)$$

$$\mathbf{Q}_\Delta^{-1} = (\mathbf{Q}_1)^{-1} - (\mathbf{Q}_2)^{-1} = \text{diag}\left(0, 0, \frac{1}{\rho_2^2} - \frac{1}{\rho_1^2}\right) \quad (12)$$

From Eq. (12), the rank of both sides of Eq. (10) must be one.

We can rewrite Eq. (10) as

$$\mathbf{P} \mathbf{Q}_\Delta^{-1} \mathbf{P}^T = \left(\frac{1}{\rho_2^2} - \frac{1}{\rho_1^2}\right) (\mathbf{C} \mathbf{T})(\mathbf{C} \mathbf{T})^T \quad (11)$$

Because we assume that the concentric circle center is equal to the origin of the world coordinate system, $\mathbf{C} \mathbf{T}$ describes the position of the projected

circle center. The resulting homogeneous coordinate removes a scale factor relating to radii.

Moreover, the matrices of the imaged concentric circles have the other constraint – rank 2 constraint. This is because the circle matrices \mathbf{Q}_1 and \mathbf{Q}_2 have two repeated elements as in Eq. (9).

The rank-2 subtractive matrix of \mathbf{Q}_1 and \mathbf{Q}_2 is intuitively given by

$$\begin{aligned} \mathbf{Q}_{\Delta 2} &= (\rho_2^2 / \rho_1^2) \mathbf{Q}_1 - \mathbf{Q}_2 \\ &= \text{diag}(\rho_2^2 / \rho_1^2 - 1, \rho_2^2 / \rho_1^2 - 1, 0) \\ &\sim \text{diag}(1, 1, 0) \end{aligned} \quad (12)$$

This property holds if we assume the non-degenerate case of the projections. Once we find the suitable ratio of scale factor β , there exists the rank-2 subtractive matrix $\mathbf{A}_{\Delta 2}$ defined by

$$\mathbf{A}_{\Delta 2} = \beta \mathbf{A}_1 - \mathbf{A}_2 \quad (13)$$

The resulting $\mathbf{A}_{\Delta 2}$ has an interesting physical meaning. Because of the homogeneity of the coordinates, the resulting matrix is expressed as

$$\mathbf{A}_{\Delta 2} \sim \mathbf{P}^{-T} \text{diag}(1, 1, 0) \mathbf{P}^{-1} \quad (14)$$

From Eq. (14), $\mathbf{A}_{\Delta 2}$ is the projected imaginary circle whose radius is zero.

A final remaining problem is how to determine the relative scale factors α and β . There are various ways to determine them, and we suggest a method to find it by setting the scale factor with the generalized eigenvalues of \mathbf{A}_1 and \mathbf{A}_2 [18]. Due to the special form of the concentric circle matrices, the generalized eigenvalues must have two repeated ones and single distinct one. For rank one case, the repeated one is selected as a scale factor. Selecting single distinct one as a scale factor makes rank two subtractive matrix.

3.2 Intrinsic parameters from concentric circles

As described in [19], once the IAC is determined, the intrinsic parameters are obtained directly. Furthermore, the IAC can be estimated with six or more imaged circular points. In this section, we show that one set of the image of the concentric circles is sufficient to find a pair of imaged circular points.

The imaged circular points are on the vanishing line, as well as on the image of an imaginary conic whose radius is 0. The image of the radius-0 circle can be extracted up to scale factor using the rank-2 constraint. The projected imaginary conic is expressed as Eq. (14).

The vanishing line \mathbf{l}_p extracted from a conic is

$$\mathbf{l}_p = \mathbf{A} \mathbf{x}_c \quad (15)$$

where \mathbf{A} is the ellipse matrix and \mathbf{x}_c is the projected circle centers [20].

In this case, the imaged circular points are the intersection of the projected radius-0 conic $\mathbf{A}_{\Delta 2}$ and the vanishing line of the plane \mathbf{l}_p .

The position of the projected circle center can be estimated from the algebraic rank-1 decomposition property.

However, comparing to the conventional point correspondence based algorithm, the proposed algorithm can estimate the intrinsic parameters only with partial views of concentric circles due to occlusion or limitation of field of view. Also, we don't need any kind of Euclidean measurements like lengths or even radii of the circles.

3.3 Experiments

The noise characteristic of the proposed calibration algorithm has been tested using synthetic data with respect to the numbers of views and the number of concentric circles. In practical case, the error is mainly caused by the quantization error of the edge points. The performance of this algorithm is analyzed under practical edge quantization noise that is in one pixel. The results are shown in Table 1. As expected, the performance of the proposed calibration algorithm improves as the numbers of views and circles increase.

Table 1. RMS error in focal length estimation (relative error) with edge quantization noise

	Error in au (%)	Error in av (%)
3 views with 2 circles (minimal case)	2.2952	1.5385
4 views with 2 circles	1.192	0.743
5 views with 2 circles	0.9264	0.5258
3 views with 3 circles	0.6514	0.3609
4 views with 3 circles	0.5064	0.3301
5 views with 3 circles	0.4427	0.2830

4. EXPERIMENTAL RESULTS

To show the practicality of the proposed algorithm, we applied it to multi-camera calibration. Figure 3 shows a camera setup, consisting of ten SONY DFW-V500 cameras, for the reconstruction of 3-D scene. In this case, the captured images suffer from FOV limitation. Figure 4 shows some captured images which are used for the calibration of intrinsic parameters. For camera calibration, we used nine images selected for each camera. Figures 5 and 6 show some of captured images and synthesized views from the 3-D reconstruction result using 'the appearance-cloning', respectively. Figure 5 shows one set of dynamic image sequences and full sequences can be reconstructed using the same framework.

Figure 7 shows some of 'Marble dataset' and their corresponding image and re-synthesized views using various reconstruction methods. To verify the accuracy of the reconstructed scene, the depth map is also shown. Note that the result from the proposed method is much similar to the ground truth.

Figure 8 shows one of 'Junglimsaji pagoda dataset' which are taken from one of Korean National Heritages by a digital hand-held camcorder. It shows that our optimization-based scene recovery can refine the scene geometry relatively well with pictures taken outdoor.



Figure 3. Multi-camera setup



Figure 4. Examples of images for calibration in multi-camera system

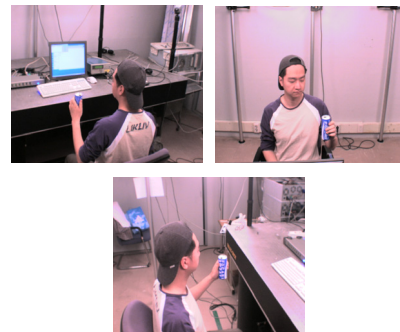


Figure 5. Examples of captured images



Figure 6. Examples of synthesized images

5. CONCLUSION

We have interpreted the photograph as a view-dependent appearance instead of combination of projected surface colors that is commonly adopted interpretation for the scene recovery. Based on this novel interpretation, we have introduced a new methodology, called "the appearance-cloning", in which we simulate the photographing mechanism of real world in our digital domain. Through our resembled mechanism, we can generate recovered appearances of our candidate 3D model that are one-

to-one corresponding to given appearances acquired from the real-world scene. Then, we have controlled our candidate 3D model by using our appearance similarity measure so that the recovered appearances resemble the given appearances.

For the practical implementation of multi-camera system, we propose a two-step camera calibration method using concentric circles. First, we calibrate each camera intrinsically, and the camera pose are subsequently estimated. The proposed constraint properties of concentric circles provide concrete physical meanings, such as the position of the projected circle center and the projected radius-0 imaginary circle. Based on the constraints of the concentric circles, we develop a new camera calibration algorithm.

We demonstrate that the photo-realistic scene reconstruction for indoor, outdoor and dynamic environments is feasible using some real-image sets by a multi-camera system or a single moving camera.

Acknowledgments

This research has been supported by NRL (code number M1-0302-00-0064) of MIST, Korea.

REFERENCES

- [1] McMillan, L. and Bishop, G. 1995. Plenoptic modeling: An image-based rendering system. In *Computer Graphics (SIGGRAPH'95)*, pp 39-46
- [2] Manning, R. A. and Dyer, C. R. 1999. Interpolating view and scene motion by dynamic view morphing. In *Proc. International Conference on Computer Vision and Pattern Recognition*. pp.388-394.
- [3] Seitz, S. M. and Dyer, C. R. 1996. View morphing. In *Computer Graphics (SIGGRAPH'96)*, pp. 21-30.
- [4] Levoy, M. and Hanrahan, P. 1996. Light field rendering. In *Computer Graphics (SIGGRAPH'96)*. pp.31-42.
- [5] Gorthler, S. J. Grzeszczuk, R. and Szeliski, R. 1996. The Lumigraph. In *Computer Graphics (SIGGRAPH'96)*. pp.43-54.
- [6] Poggio, T., Torre, V., and Koch, C. 1985. Computational Vision and Regularization Theory. *Nature*. 317(26). pp:314-319
- [7] Kutulakos, K. N. and Seitz, M. 1998. What Do N Photographs Tell Us about 3D Shape?. Technical Report TR680, Computer Science Dept., U. Rochester, January.
- [8] Faugeras, O. and Keriven, R. 1998. Complete dense stereovision using level set methods. *IEEE Trans. on Image Processing*, 7(3):336-344.
- [9] Kutulakos, K. N. and Seitz, S. 2000. A Theory of Shape by Space Carving. *International Journal of Computer Vision*, 38(3):199-218.
- [10] Seitz, S. and Dyer, C. 1997. Photorealistic Scene Reconstruction by Voxel Coloring. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1067-1073.
- [11] Kutulakos, K. N. and Seitz, S. 1999. A Theory of Shape by Space Carving. In *International Conference on Computer Vision*, pp. 307-314.
- [12] Culbertson, W. B., Malzbender, T., and Slabaugh, G. 1999. Generalized Voxel Coloring. In *International Conference on Computer Vision Workshop on Vision Algorithms, Theory and Practice*, Springer-Verlag Lecture notes in Computer Science 1883, pp. 100-115.
- [13] Kutulakos, K. N. 2000. Approximate N-View Stereo. In *Proc. 6th European Conference on Computer Vision*, pp.67-83.
- [14] Broadhurst, A. and Cipolla, R. 2000. A Statistical Consistency Check for the Space Carving Algorithm. In *British Machine Vision Conference*, pp. 282-291.
- [15] Slabaugh, G., Culbertson, W. B., Malzbender, T., Stevens, M. R., and Schafer, R. W. 2004. Methods for Volumetric Reconstruction of Visual Scenes. *International Journal of Computer Vision*, 57(3):179-199
- [16] Mortensen, E. and Barrett, W. 1995. Intelligent scissors for image composition. In *Computer Graphics (SIGGRAPH'95)*. pp:191-198.
- [17] Papoulis, A. 1984. Probability, Random Variables, and Stochastic Processes, New York, McGraw-Hill.
- [18] G. Golub and C. van Loan, *Matrix Computation*, The Johns Hopkins University Press, 1996
- [19] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge press, 2000
- [20] K. Kanatani and W. Liu, "3D Interpretation of Conics and Orthogonality," *CVGIP: Image Understanding*, 58(3):286-301, 1993

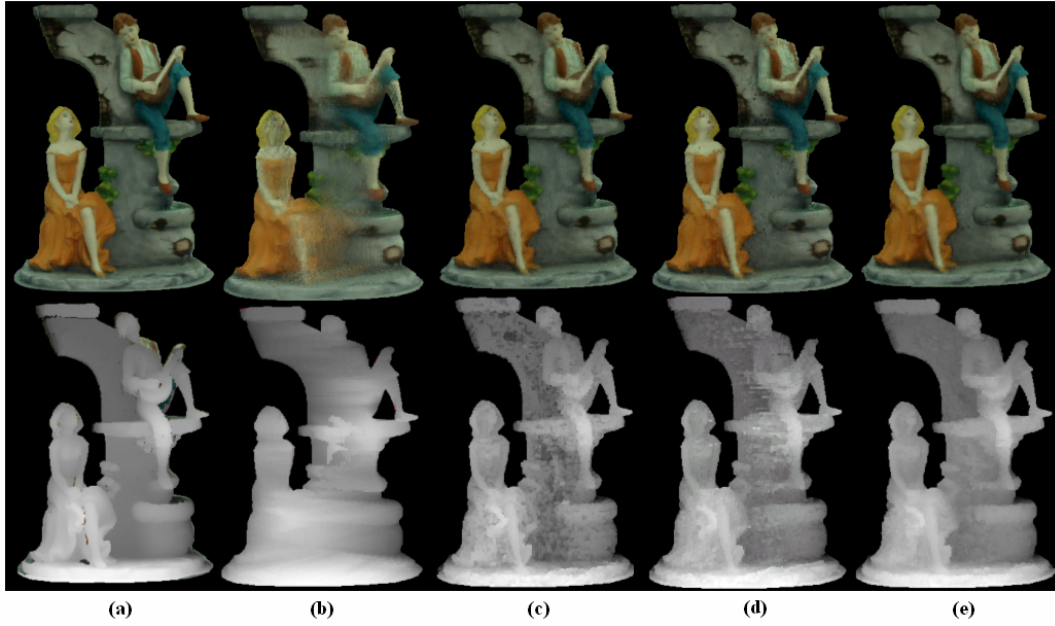


Figure 7 Qualitative comparison: First row shows the photometric recovery. Second row shows the geometric recovery. (a) Active 3D scanner. (b) Volume intersection (VI) - initial. (c) Probabilistic voxel Coloring (PVC) [14]. (d) Error optimization with the generalized voxel coloring (GVC/EO) [12]. (e) Appearance-cloning (Proposed).

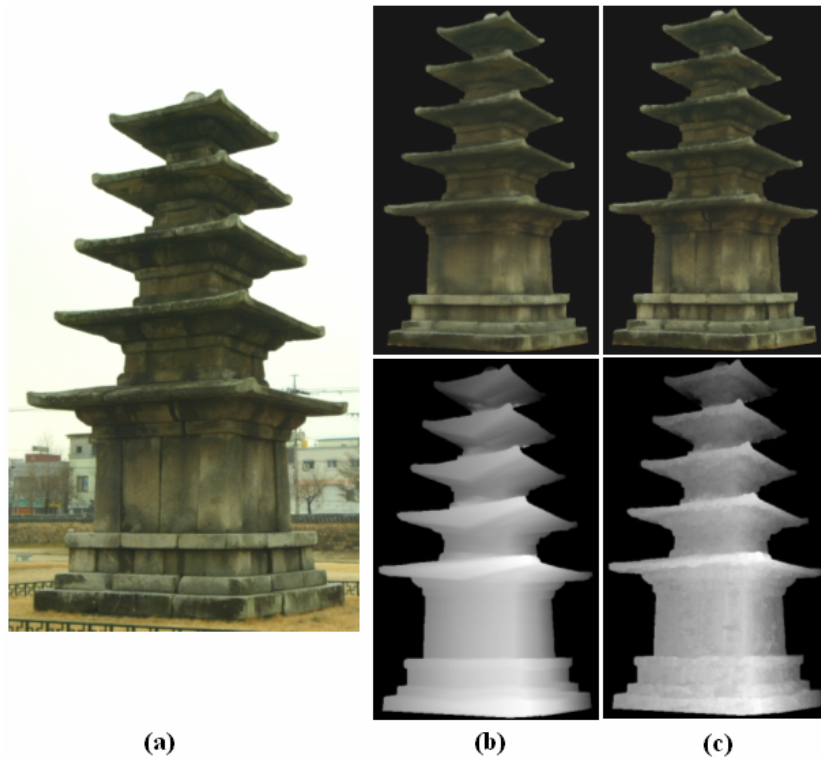


Figure 8. Scene recovery of Junglimsaji Pagoda dataset: (a) shows two of input images. (b) shows the recovered appearance image and depth map of initial visual hull. (c) shows those of recovered optimal 3D model using our appearance-cloning algorithm.