UNCALIBRATED VIEW SYNTHESIS USING PLANAR SEGMENTATION OF IMAGES

Jae-Ho Yun and Rae-Hong Park

Department of Electronic Engineering, Sogang University, Seoul, Korea
yastoil@sogang.ac.kr, rhpark@sogang.ac.kr

ABSTRACT

This paper presents an uncalibrated view synthesis method using piecewise planar regions that are extracted from a given set of image pairs through planar segmentation. Our work concentrates on a view synthesis method that does not need estimation of camera parameters and scene structure. For our goal, we simply assume that images of real world are composed of piecewise planar regions. Then, we perform view synthesis simply with planar regions and homographies between them. Here, for accurate extraction of planar homographies and piecewise planar regions in images, the proposed method employs iterative homography estimation and color segmentation-based planar region extraction. The proposed method synthesizes the virtual view image using a set of planar regions as well as a set of corresponding homographies. Experimental comparisons with the images synthesized using the actual three-dimensional (3-D) scene structure and camera poses show that the proposed method effectively describes scene changes by viewpoint movements without estimation of 3-D and camera information.

KEYWORDS
Uncalibrated View Synthesis, Planar Segmentation, Planar Region-based View Synthesis, Homography

1. INTRODUCTION

The goal of view synthesis is generating a virtual image at an arbitrary viewpoint using multiple images taken from a camera [1–8]. Generally, view synthesis methods are divided into two categories depending on whether or not camera calibration is used: calibrated [2–4] and uncalibrated [5–8]. Camera calibration [9,10] represents the estimation procedure of camera parameters, which is a fundamental step for estimation of scene structures and definition of arbitrary viewpoints. Generally, view synthesis methods are performed with camera calibration. However, inconveniences arise in performing camera calibration, which requires a precise calibration object and off-line processing. On the contrary, uncalibrated view synthesis methods do not perform camera calibration. They do not estimate calibration parameters and thus they could synthesize virtual view images at a limited viewpoint. Recently, researches on uncalibrated view synthesis have been done to improve the performance [5–8].

Besides camera calibration, depth estimation [11–14] is needed to synthesize a virtual view regardless of calibrated/uncalibrated view synthesis methods. Depth estimation represents the estimation procedure of the distance of a point from a camera, from which the scene structure is reconstructed. Given more than two images, depth estimation is performed mostly using stereo algorithms [13] or triangulation [14]. Both approaches have difficulties in ensuring the accuracy of the estimated depth, because depth estimation needs to establish image correspondence at all points or feature points in an image, requiring costly and time-consuming operations.

After estimation of the scene structure, virtual view images are synthesized by warping images with transformation matrices that specify virtual viewpoints. In the case of calibrated view synthesis, transformation matrices for virtual viewpoints are arbitrarily defined by specifying the change of the camera geometry with camera parameters [15]. However, in uncalibrated view synthesis methods, transformation matrices are defined by recovering camera parameters using a

self-calibration method or by utilizing the relation matrix that specifies the relationship of image correspondences from an image pair, such as a homography or fundamental matrix [7,16]. However, incorrectly estimated image correspondences produce large calibration error. In general, there exists a trade-off between the precision and complexity in generating a virtual view image. When the high precision is needed, both procedures (camera calibration and depth estimation) are necessary, requiring high computational complexity.

This paper focuses on developing an uncalibrated view synthesis method that accurately produces synthesized images at virtual viewpoints without any three-dimensional (3-D) information. 3-D reconstruction approach with a single image [17] also could accurately synthesize a desired image from an arbitrary viewpoint. However, it needs to assign vanishing points and to separate object boundaries interactively. Although 3-D reconstruction with multiple images [1–8] could automatically synthesize images, essentially it requires to perform additional stereo matching or camera calibration. Thus, the proposed method aims at developing a view synthesis method that does not require estimation procedures of camera parameters and scene structure. For our goal, the proposed method employs a planar region based processing [18–20]. Planar region based processing is based on the assumption that scene structure could be simplified by composition of a number of piecewise planar regions [19,20]. Also, planar region based processing offers an desirable advantage that a planar region is transformed exactly by homography projection from one image to the other without any depth information [15]. According to this concept, the proposed method extracts planar regions from an image with a color segmentation-based approach. Color segmentation [21–23] is introduced for boundary preservation of planar regions. In [20], an interactive planar segmentation method is proposed by assigning user-defined sample regions. Interactive approach shows competitive results for complex scenes. Since the proposed method excludes user interaction, it iteratively finds initial planar regions and extracts planar regions. After extraction of planar regions, virtual viewpoints are defined with composition of virtual homographies of planar regions, in which virtual homographies are determined by scalar multiples of homographies that represent interpolated or extrapolated viewpoints [5,8]. At last, the proposed method synthesizes virtual view images by homography projection of planar regions with virtual homographies for interpolated or extrapolated viewpoints.

The rest of the paper is structured as follows. Section 2 describes an overview of uncalibrated view synthesis. Section 3 proposes an uncalibrated view synthesis using planar segmentation of images. Section 4 gives experimental results and discussions. Section 5 concludes the paper.

2. RELATED WORK

Uncalibrated view synthesis methods consist of several procedures to estimate scene structure from images. Whether input images are rectified or not is an important factor for adapting these procedures. With a rectified image pair, uncalibrated view synthesis [13] was presented, in which a stereo algorithm estimates the horizontal displacement, called disparity, between two images. Disparity is inversely proportional to depth, i.e., a large disparity represents that the point is close to a camera. Gradual increment of disparity corresponds to the pixel movement caused by the viewpoint change from one image to the other. However, view synthesis with a rectified image pair has limited ability to deal with scene changes caused by camera rotation.

With an unrectified input image pair, various methods, which employ additional procedures such as self-calibration [16], image rectification [24], or motion parallax [5,8], have been developed for view synthesis. Because a self-calibration technique represents the estimation process of camera parameters using image correspondences and a fundamental matrix, an uncalibrated view synthesis method using self-calibration could estimate depth by triangulation. After estimating camera parameters and scene structure, virtual view images are generated by the same procedures

as in the calibrated view synthesis method. However, it is hard to guarantee the accuracy of camera parameters that are estimated by self-calibration.

Another approach to virtual view synthesis is an uncalibrated view synthesis method with image rectification [24]. Image rectification is the transformation process that converts an unrectified image into a rectified one. After performing rectification of input images, a stereo algorithm estimates disparities between the rectified input images. Then, a virtual view image is synthesized in the same way as in the view synthesis method with a rectified image pair.

The last approach to virtual view synthesis is an uncalibrated view synthesis technique with motion parallax [5,8]. Unlike disparity, motion parallax is the vertical and horizontal displacements of a pixel between two images with respect to the estimated homography. Then, motion parallax is derived by measuring the distance between the detected point and its corresponding point estimated by homography projection. After estimation of motion parallaxes at all points, virtual view images are synthesized by homography projection of given points with a new homography and with compensation of motion parallax at each point.

All the uncalibrated view synthesis methods need to estimate scene structure by estimating disparity or motion parallax. In the case of the uncalibrated view synthesis using disparity, it is difficult to estimate the camera motion and disparity. In the case of using motion parallax, it is difficult to estimate the motion parallax. This paper presents an uncalibrated view synthesis method that accurately generates synthesized images at virtual viewpoints without estimating any 3-D information such as disparity or motion parallax.

3. PROPOSED UNCALIBRATED VIEW SYNTHESIS METHOD
3.1. Overview of the Proposed Method

This paper proposes an uncalibrated view synthesis method using planar segmentation of images, in which both the camera information and scene structure are not required. Unlike the previous uncalibrated view synthesis methods [5–8], the proposed method does not require any 3-D scene structure information such as disparity or motion parallax. Instead, it utilizes planar regions in an image because planar regions are transformed by the homography projection from one image to the other without any depth information. In addition, it employs a color segmentation-based method for accurate extraction of planar regions from an image pair, which preserves well the boundary of planar regions. If there are some homogeneous color regions in an image pair, pixel-based processing shows a limited performance in estimating depth. Through region-based processing, it reduces incorrect displacement caused by incorrect estimation of depth, which often occurs in synthesizing images by pixel-based processing.

The proposed method consists of two parts: planar segmentation and view synthesis by warping planar regions. Fig. 1 shows the overall block diagram of the proposed uncalibrated view synthesis method and Table 1 lists parameters used in this paper. In Fig. 1, the planar segmentation part consists of color segmentation, image correspondence estimation, iterative homography estimation, and segment-based matching. The view synthesis part consists of virtual homography estimation and virtual image generation.

Given an image pair, planar homographies are initially estimated by iterative homography estimation using image correspondences. Along with homography estimation, color segmentation splits an image into a number of homogeneous color segments. With the given
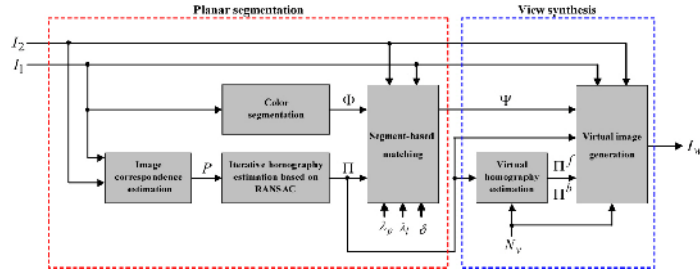
Fig. 1. Block diagram of the proposed uncalibrated view synthesis method.

Table 1. List of parameters used in the proposed method.

Symbol	Description
I_1, I_2	Input images
I_w	Synthesized warped image
I_1', I_2'	Intensities of I_1 and I_2 that are normalized to one
Φ	Set of color segments
φ_j	jth color segment in Φ
Ψ	Set of planar regions
ψ_j	jth planar region in Ψ
K	Set of disoccluded regions
κ_j	jth disoccluded region in K
X	Set of grouped disoccluded regions
χ_j	jth disoccluded region in X
ξ_j	Set of neighboring pixels of φ_j
V_j	Set of indices that indicate the neighboring segments of φ_j
S	Index set of color segments
ω_j	warped area of φ_j by homography projection
υ_j	warped area of κ_j by homography projection
Π	Set of planar homographies
Π^f, Π^b	Set of virtual homographies (*f*: forward, *b*: backward)
L	Set of homography labels for Φ
l_j	Homography label for φ_j
A	Set of homography labels for K
α_j	Homography label for κ_j
N_v	Number of virtual views
N_h	Number of homographies
P	Set of image correspondences
$\lambda_p, \lambda_l, \delta$	Parameters for segment-based matching
$E^i(\cdot), E^p(\cdot), E^l(\cdot)$	Measurement functions for all the color segments (*i*: intensity, *p*: position, *l*: label)
$e_{L,\Phi}^i(\cdot), e_{L,\Phi}^p(\cdot), e_{L,\Phi}^l(\cdot)$	Matching cost functions for each color segment (*i*: intensity, *p*: position, *l*: label)
$u_{L,\Phi}^p(\cdot), u_{L,\Phi}^l(\cdot)$	Matching cost functions for each pixel in a color segment (*p*: position, *l*: label)
$C(\cdot)$	Operator that counts the number of pixels in a segment

μ_j	Intensity mean of I_1' in the area of φ_j
$e_{\kappa_j}^o(\cdot)$	Matching cost function for each occlusion
$d_\Psi^i(\cdot)$	Measurement function for the degree of identity in the intersection area

homographies and color segments, planar regions in an image are extracted by classifying all the color segments according to the corresponding homographies. If some color segments correspond to the same homography, these segments are assumed to be in the same planar region. Then, the homography corresponding to each planar region is determined by warping the first (left) image of an image pair with the homography and by measuring the difference between the warped image and the second (right) image of the image pair.

After planar segmentation, virtual views are synthesized by homography projection of all the planar regions with the newly generated homographies that represent virtual viewpoints. Scalar multiples of homographies are employed to define virtual homographies at interpolated or extrapolated viewpoints. In addition, the proposed method includes occlusion handling and image inpainting to improve the quality of the synthesize images. Detailed descriptions of each procedure are given in the following.

3.2. Planar Segmentation: First Part
In this part, planar segmentation is described in detail. The proposed planar segmentation is organized in two subsections: homography estimation and segment classification.

3.2.1. Iterative Homography Estimation Based on Random Sample Consensus (RANSAC)
Homographies in an image pair contain the information of planar regions in the scene because the homography is defined by the image correspondences of the same planar region. Thus, with given image correspondences, the proposed method uses iterative homography estimation to detect the appropriate number of planar homographies that exist in the scene, after which a RANSAC based method is employed for homography estimation [15]. The RANSAC based homography method estimates a planar homography with image correspondences. Then, it classifies the image correspondences into two sets: inlier and outlier sets. The inlier (outlier) set represents the set of image correspondences that satisfy (do not satisfy) the homographic relation. Because RANSAC homography estimation classifies the initial image correspondences into inlier and outlier sets, planar homographies in the scene are repeatedly estimated by iterative homography estimation using the remaining outlier set, which is depicted in Fig. 2.

The proposed method is described as follows. First, a RANSAC based homography method is used with image correspondences. Then, image correspondences are initially classified as inliers and outliers. RANSAC based homography estimation is again applied to the point correspondences that are initially classified as outliers, i.e., the point correspondences in the outlier set are classified again as inliers and outliers by the re-estimated homography. These procedures are performed repeatedly until a few point correspondences are left. At last, we can find the appropriate number of planar homographies that exist in the scene.
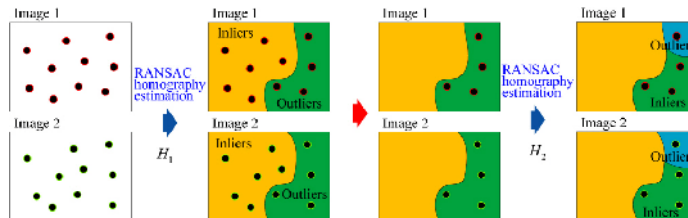

Fig. 2. Iterative homography estimation based on RANSAC.

3.3.3. Segment-based Matching for Planar Segmentation

Given planar homographies in the scene, planar regions are extracted by splitting an image into several image segments and by assigning the corresponding homography to the image segment. Then, the proposed method employs a color segment-based approach because color segmentation methods preserve the object shape or boundary by splitting an image into a number of homogeneous color segments [23]. Thus, accurate planar regions are derived by classifying the color segments into the corresponding planar homographies. Fig. 3 shows a schematic example of segment based matching for easy understanding. The procedure of the proposed segment-based matching framework is described as follows. With a pair of input images I_1 and I_2, let I_1 and I_2 be a target image and a reference image, respectively. Then, a set of color segments is obtained by applying color segmentation to the target image I_1.
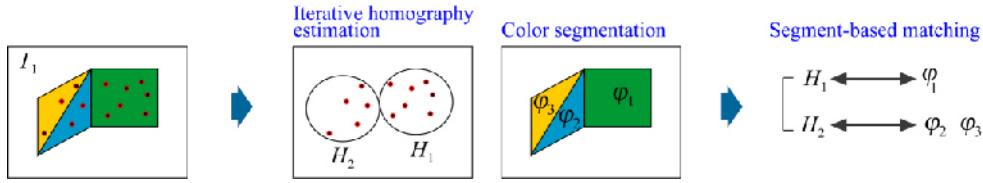

Fig. 3. Segment-based matching framework.

Given two sets of planar homographies and color segments, Π and Φ, respectively, we formulate the minimization problem to determine the homography label that corresponds to each color segment. Finding the homography label for each segment is written as

$$L^* = \arg\min_L \left(E^i(L) + \lambda_p E^p(L) + \lambda_l E^l(L) \right), \tag{1}$$

where L signifies a set of homography labels assigned to all color segments, with l_j representing the label for jth color segment, i.e., jth element of L. After optimization using Eq. (1), the optimum label set L^* is obtained and each l_j in L^* indicates a homography in Π. λ_p and λ_l denote the weight factors for $E^p(\cdot)$ and $E^l(\cdot)$, respectively. In the optimization procedure, $E^i(\cdot)$, $E^p(\cdot)$, and $E^l(\cdot)$ measure how closely, according to the intensity, position, and label, respectively, all the color segments correspond to their assigned homographies after warping. If all the segments have been correctly assigned to homography labels, $E^i(\cdot)$, $E^p(\cdot)$, and $E^l(\cdot)$ have small values that are close to zero.

Given the index set of all the color segments, S, two sets of labels and segments are expressed as $L = \{l_j \mid j \in S\}$ and $\Phi = \{\varphi_j \mid j \in S\}$, respectively. Then, we can compute

$$E^i(L) = \sum_{j \in S} e^i_{L,\Phi}(j), \tag{2}$$

$$E^p(L) = \sum_{j \in S} e^p_{L,\Phi}(j), \tag{3}$$

$$E^l(L) = \sum_{j \in S} e^l_{L,\Phi}(j), \tag{4}$$

where $e^i_{L,\Phi}(j)$, $e^p_{L,\Phi}(j)$, and $e^l_{L,\Phi}(j)$ measure the matching costs of jth segment, with the superscripts i, p, and l representing the intensity, position, and label, respectively. Three matching costs are described in detail in the following. In Eq. (2), $e^i_{L,\Phi}(j)$ is defined as

$$e^i_{L,\Phi}(j) = \frac{1}{C(\varphi_j)} \sum_{\mathbf{x} \in \varphi_j} \left| I'_1(\mathbf{x}) - I'_{2,\Pi^{-1}(l_j)}(\mathbf{x}) \right|, \tag{5}$$

where $I_1'(\mathbf{x})$ and $I_2'(\mathbf{x})$ are intensities of I_1 and I_2 at $\mathbf{x}$ that are normalized to one, respectively, with $\mathbf{x}$ denoting the image coordinate vector. $I_{2,\Pi^{-1}(l_j)}'$ represents the warped image of I_2 with the homography that is the l_j th element in the set of Π^{-1}, in which Π^{-1} is a set that contains inverse homographies of all the homographies in Π. $C(\cdot)$ represents the operator that counts the number of pixels in a segment.

In Eq. (3), $e_{L,\Phi}^p(j)$ is a measurement function of the positional accuracy by homography projection of jth segment, which is defined as

$$e_{L,\Phi}^p(j) = \frac{1}{C(\omega_j)} \sum_{\mathbf{x} \in \omega_j} u_{L,\Phi}^p(\mathbf{x},\ j), \tag{6}$$

where ω_j is warped area of φ_j by homography projection, with the homography selected by l_j in Π. $u_{L,\Phi}^p(\mathbf{x},\ j)$ is defined as

$$u_{L,\Phi}^p(\mathbf{x},\ j) = \begin{cases} 1, & \text{if } \mathbf{x} \text{ is overlapped by other segments} \\ \delta, & \text{if } \mathbf{x} \text{ is outside the image plane} \\ 0, & \text{otherwise} \end{cases}, \tag{7}$$

with δ representing the penalty value between zero and one. Eq. (3) represents that there are no overlap regions if all the segments are correctly warped.

In Eq. (4), $e_{L,\Phi}^l(j)$ measures how correctly the label is assigned to jth segment in comparison with the neighboring segments, which is defined as

$$e_{L,\Phi}^l(j) = \frac{1}{C(\xi_j)} \sum_{k \in V_j} u_{L,\Phi}^l(j,\ k), \tag{8}$$

where ξ_j and V_j are the set of neighboring pixels and the set of indices that indicate the neighboring segments of jth segment, respectively. $u_{L,\Phi}^l(j,\ k)$ is defined as

$$u_{L,\Phi}^l(j,k) = \begin{cases} C(\xi_j \cap \varphi_k) \cdot \left(1 - 0.5 \cdot |\mu_j - \mu_k|\right), & \text{if } l_j \neq l_k \\ 0, & \text{otherwise} \end{cases}, \tag{9}$$

where μ_j represents the intensity mean of I_1' in the area of jth segment. For easy understanding of the relation of ξ_j and V_j, Figs. 4(b) and 4(c) show examples for ξ_j and V_j, respectively, when φ_j is given as shown in Fig. 4(a). Eq. (8) represents that there is high possibility that neighboring segments are warped with the same homography. Then, the proposed method employs a greedy search algorithm [25] to find the optimum label set.
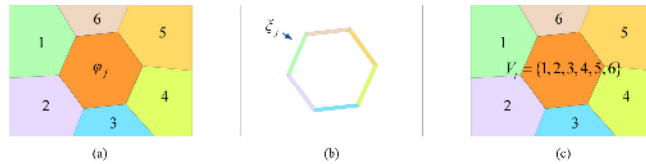

Fig. 4. Symbol definitions. (a) Given segment, φ_j. (b) Neighboring pixels, ξ_j. (c) Set of indices of neighboring segments, V_j.

Segment-based matching is related to finding correspondences betwee homographies and color segments. When there are n homographies and m color segments, we estimate n^m times to find the correct relationship between homographies and color segments, which requires high computational complexity. Thus, we use an iterated local maxima search procedure, greedy

search algorithm, to find the optimum solution. This problem was also considered in segment-based stereo matching [21–23], in which each segment corresponds to a plane equation for estimating disparity at each pixel. Then, to solve this correspondence problem, segment-based stereo matching algorithms employ a greedy search and derive comparative results. Segment-based matching for planar segmentation and segment-based stereo matching for disparity estimation are based on the same correspondence problem, and the proposed method employs a greedy search. After finding the optimum label, planar regions are extracted by grouping all the segments according to their homography labels. As a result, planar regions are derived as a set, Ψ, and each planar region is represented as ψ_j that corresponds to jth homography in Π.

3.3. View Synthesis Using Planar Regions: Second Part

When synthesizing virtual view images, the transformation matrices that specify virtual viewpoints are needed. Then, the proposed method uses scalar multiples of homographies that represent the homographies at the interpolated or extrapolated viewpoints [5,8]. After estimation of new homographies for virtual views, virtual view images are synthesized by warping all the planar regions with their new homographies. At this point, the warping order, i.e., which planar region will be warped beforehand, must be determined. For this reason, the proposed method verifies the visibility between planar regions after homography projection to determine the warping order. Lastly, occlusion handling and inpainting are sequentially employed to fill the unfilled pixels in virtual view images to increase the quality of the synthesized images.

3.3.1. Definition of Virtual Viewpoints

Scalar multiples of a transformation matrix is used to synthesize the virtual view images [8]. Whereas the method [8] employed scalar multiples of a transformation matrix that is composed of homography and motion parallax, the proposed method derives only the scalar multiples of homography without estimating motion parallax because the planar region lies in the same 3-D plane with zero parallax. Because motion parallax estimation requires high computational complexity, the proposed method without motion parallax is effective. Given jth homography, H_j in Π, two virtual homographies for kth interpolated view are defined along forward and backward directions as [26]

$$H_j^{f_k} = \mathrm{expm}\left(\frac{k}{N_v} \mathrm{logm}(H_j) \right) \tag{10}$$

$$H_j^{b_k} = \mathrm{expm}\left(\frac{N_v - k}{N_v} \mathrm{logm}(H_j^{-1}) \right) \tag{11}$$

respectively, where N_v represents the number of virtual views. Operators $\mathrm{expm}(\cdot)$ and $\mathrm{logm}(\cdot)$ compute matrix exponential and logarithm, respectively. The matrix exponential is defined as

$$\mathrm{expm}(M) = \sum_{k=0}^{\infty} \frac{M^k}{k!}, \tag{12}$$

where M denotes a square matrix. Also, the matrix logarithm is similarly defined. Then, virtual homographies along forward and backward directions are used to synthesize the virtual view images from I_1 and I_2, respectively.

3.3.2. Synthesis of Virtual Views

The proposed method decides warping order and handles occlusion to correctly synthesize virtual view images with planar regions and virtual homographies. Because the warped planar region could be occluded by the planar region warped afterwards, determination of correct warping order is important to accurately deal with scene changes by viewpoint movement. It is needed to process disoccluded areas when viewpoint is changed from one image to the other.

The proposed method determines the warping order of all the planar regions by checking the visibility among planar regions. Although the warping order must be determined according to the distance from a camera, the proposed method determines the correct warping order by comparing the warping orders between all possible pairs of two planar regions and by combining them. In detail, given two planar regions, ψ_j and ψ_k, the warping order is determined by comparing the visibility in two cases: when ψ_j is warped before ψ_k and vice versa. If intersection of warped ψ_j and ψ_k is the same as the region in the reference image when ψ_j is warped before ψ_k, ψ_k is in front of ψ_j. Given two planar regions as shown in Fig. 5(a), for easy understanding of two cases, Figs. 5(b) and 5(c) show two examples: when ψ_j is warped before ψ_k and ψ_k is warped before ψ_j, respectively. Note that each intersection in Figs. 5(b) and 5(c) has different color. Then, the degree of identity in the intersection area, which is compared with the reference image, is defined as

$$d_{\Psi}^{i}(j, \omega_j \cap \omega_k) = \frac{1}{C(\omega_j \cap \omega_k)} \sum_{\mathbf{x} \in \omega_j \cap \omega_k} \left| I'_{1,\Pi(j)}(\mathbf{x}) - I'_2(\mathbf{x}) \right|, \tag{13}$$

$$d_{\Psi}^{i}(k, \omega_j \cap \omega_k) = \frac{1}{C(\omega_j \cap \omega_k)} \sum_{\mathbf{x} \in \omega_j \cap \omega_k} \left| I'_{1,\Pi(k)}(\mathbf{x}) - I'_2(\mathbf{x}) \right|, \tag{14}$$

where $\omega_j \cap \omega_k$ represents the intersection of warped ψ_j and ψ_k, and the superscript i denotes intersection. After determination of the warping orders for all possible pairs of two planar regions, the sequence that satisfies all the ordering conditions of all possible pairs of two planar regions is the optimal warping order of all the planar regions.
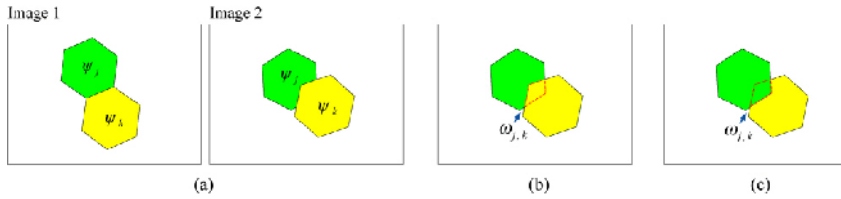

Fig. 5. Decision of the warping order between two planar regions. (a) Given two planar regions, ψ_j and ψ_k in images 1 and 2. (b) When ψ_j is warped before ψ_k. (c) When ψ_k is warped before ψ_j.

By determining the warping order, the proposed method efficiently deals with occlusion in planar regions because a previously warped planar region could be occluded by planar regions that are warped afterwards. However, disoccluded regions cannot be dealt with planar regions because planar regions that are extracted from a target image do not have disoccluded regions. Thus, after detecting the disoccluded regions in the reference image, the proposed method assigns a homography to each disoccluded region to compensate for the quality degradation caused by disoccluded regions in a synthesized image.

In detail, disoccluded regions are detected by subtracting the warped planar regions from the reference image because disoccluded regions exist only in the reference image. After detecting the disoccluded regions, the proposed method simply employs region growing to group the disoccluded regions. By the results of region grouping, κ_j denotes jth region, an element of the set of K. Then, corresponding homography for each κ_j is determined by measuring how many pixels in κ_j cover unfilled pixels after homography projection in the virtual views, where the measurement function is defined as

$$\alpha_j^* = \arg\max_{\alpha_j} e_{\kappa_j}^o(\alpha_j), \tag{15}$$

where α_j has a homography label corresponding to κ_j in Π, α_j^* has its optimum homography label that satisfies Eq. (15), and the superscript o denotes occlusion. $e_{\kappa_j}^o(\cdot)$ represents the measurement function that counts how many pixels in κ_j cover unfilled pixels after homography projection in the virtual views. $e_{\kappa_j}^o(\cdot)$ is defined as

$$e_{\kappa_j}^o(\alpha_j) = \sum_{\mathbf{x} \in \upsilon_j} u_{\kappa_j}^o(\mathbf{x}), \tag{16}$$

where υ_j represents the warped region of κ_j with the homography that is chosen by α_j, with $u_{\kappa_j}^o(\mathbf{x})$ defined as

$$u_{\kappa_j}^o(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \text{ is in unfilled regions} \\ 0, & \text{otherwise} \end{cases}. \tag{17}$$

After finding the optimum label for each disoccluded region, disoccluded regions are grouped again according to their homography labels. Then, grouped disoccluded regions are represented as a set, X. In addition, each disoccluded region is represented as χ_j that corresponds to jth homography in Π at virtual views.

Given planar regions, disoccluded regions, and virtual homographies, virtual view images are synthesized by warping disoccluded regions and planar regions with their corresponding homographies sequentially. Then, when warping the disoccluded regions, backward homographies are employed because disoccluded regions are extracted from the reference image, in which backward homographies are computed by applying the homographies to Eq. (11). However, when warping the planar regions, forward homographies are employed because planar regions are extracted from the target image, in which forward homographies are computed by applying the homographies to Eq. (10). After warping all the regions, remaining unfilled pixels are filled by image inpainting [27].

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed method assumes that scene is composed of piecewise planar regions. Curved objects in images cannot be faithfully extracted with the proposed method. However, some artificial objects are composed of or approximated by piecewise planar regions. Then, the proposed method is applicable to the image of objects with piecewise planar regions. Note that in this paper all of scenes used in experiments are assumed to be composed of the piecewise planar objects. To show the effectiveness of the proposed method, we give the experimental results with the images that are synthesized using both the actual 3-D scene structure and camera positions. Also, we provide experimental results with three real image pairs to show that the proposed method accurately deals with the scene changes by viewpoint movements without any 3-D information.

4.1. Experiments with Synthetic Image Pairs

We apply the proposed uncalibrated view synthesis method to the test image pair that is synthesized with the 3-D model information and camera geometry. Then, we generate the virtual view images using the proposed method and compare them with the images that are synthesized by gradually changing the 3-D camera geometry. Figs. 6(a) and 6(b) show the synthetic Box image pair and its planar regions in the image pair, respectively. In Fig. 6(b), regions within the solid line represent the planar regions and those within the dotted line represent the occlusions. Similarly, Figs. 7(a) and 7(b) show the synthetic Pyramid image pair and its planar regions in the image pair, respectively.
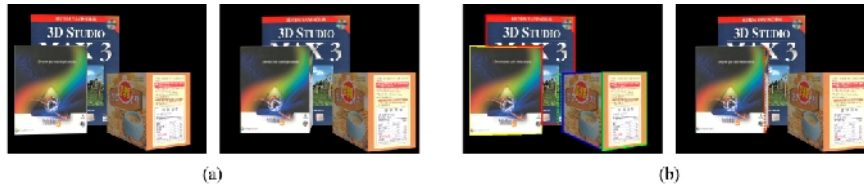
Fig. 6. Synthetic Box image pair. (a) Test image pair (left: target image, right: reference image). (b) Planar regions and occlusions in an image pair.

Fig. 7. Synthetic Pyramid image pair. (a) Test image pair (left: target image, right: reference image). (b) Planar regions in an image pair.

The scale invariant feature transform (SIFT) [28] is applied to two test image pairs to establish image correspondences. After that, iterative homography estimation is performed to estimate all the planar homographies that exist in the image pair. Figs. 8(a) and 8(b) show the estimation results of image correspondences and classification results of image correspondences from the synthetic Box image pair according to the planar homography, respectively. Figs. 9(a) and 9(b) also show the estimation results of image correspondences and classification results of image correspondences from the synthetic Pyramid image pair, respectively. Using the SIFT and iterative homography estimation, a total of 462 image correspondences and five planar homographies are found from the synthetic Box image pair. Although there are actually planar homographies in the image pair, the number of the estimated homographies is larger than that of the actual homographies. From the synthetic Pyramid image pair, a total of 427 image correspondences and five planar homographies are found.

Fig. 8. Image correspondences and their classification results of the synthetic Box image pair. (a) Estimation result of image correspondences. (b) Classification result of image correspondences.

Fig. 9. Image correspondences and their classification results of the synthetic Pyramid image pair. (a) Estimation result of image correspondences. (b) Classification result of image correspondences.

Note that the accuracy of homography estimation depends on the selection of the minimum number of image correspondences to determine a homography. In this experiment, we set the minimum number of image correspondences to five. Then, among all image correspondences, most of image correspondences are classified as an inlier set according to each planar homography. The remaining few image correspondences among all image correspondences are classified as an outlier set because the minimum number of image correspondences to estimate a

homography is set to five in this experiment. If we set the minimum number of image correspondences to a value more than five, we could remove the last planar homography that just has only few image correspondences in an inlier set because it is not an actual homography that exists in the image pair. However, the final synthesis results slightly depend on the minimum number of image correspondences because the color segment-based matching in the next procedure could decide that there is no color segment that corresponds to the last homography.

Before segment-based matching, we apply a mean-shift segmentation [29] for color segmentation to split the target image into a number of homogeneous color segments. Figs. 10(a) and 11(a) show the target images in Figs. 6(a) and 7(a), respectively. Figs. 10(b) and 11(b) show the segmentation results of Figs. 10(a) and 11(a), respectively. With two sets of homographies and color segments, we perform segment-based matching to classify the color segments according to their corresponding planar homographies, in which we set the constants to $\lambda_p = 0.8$, $\lambda_l = 0.4$, and

$\delta = 0.1$. First, when it is difficult to determine the corresponding homography for each color segment in an image because of similar intensities among color segments, an increase in λ_p will help to determine the correct homography for each color segment by measuring the positional accuracy after homography projection. Second, when adjacent color segments have similar homographies but not the same, an increase in λ_l will help to determine the correct homography for each color segment by comparing the homography labels of neighboring segments. Lastly, when some color segments are outside the image area after homography projection because of incorrectly assigned homography, an increase in δ will help to correct incorrect homography assignment. Figs. 10(c) and 11(c) show the results of planar segmentation of Figs. 10(a) and 11(a), respectively. Note that the target image is correctly classified into four and five planar regions for each synthetic image pair, respectively. Then, a background area (black colored area) in the target image of the synthetic Box image pair is combined into one of the four planar regions because the background area does not have any image correspondences.
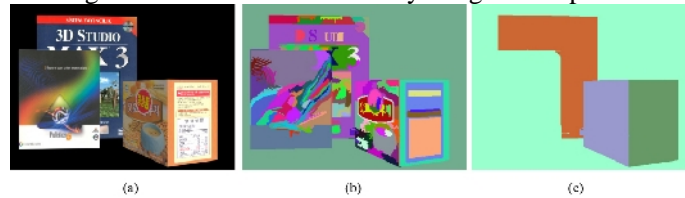

Fig. 10. Results of color segmentation and planar segmentation of the target image in the synthetic Box image pair. (a) Target image. (b) Color segmentation of (a). (c) Planar segmentation of (a).
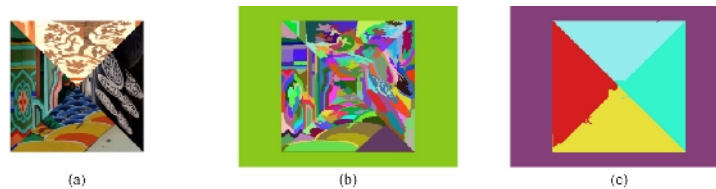

Fig. 11. Results of color segmentation and planar segmentation of the target image in the synthetic Pyramid image pair. (a) Target image. (b) Color segmentation of (a). (c) Planar segmentation of (a).

After planar segmentation, Fusiello *et al.*'s virtual homography estimation method [5] is employed to generate virtual homographies at extrapolated viewpoints. Then, Colombari *et al.*'s virtual homography estimation method [8] that is an extended work of [5] is employed to generate virtual homographies at interpolated viewpoints. Also, the warping order among planar regions is determined using the information of overlap regions for each synthetic Box and Pyramid image pair. Then, because the proposed method determines the warping order using overlap regions, if there is no overlap region among planar regions, the proposed method could not determine the accurate warping order. However, if there are no overlap regions, incorrectly estimated warp order does not effect on the result of image synthesis. After that, the proposed method extracts

occlusion from the reference image to fill the disoccluded regions in the synthesized images. Figs. 12(a) and 13(a) show the target images in Figs. 6(a) and 7(a), respectively. Figs. 12(b) and 13(b) show the estimation results of occlusions that are extracted from Figs. 12(a) and 13(a), respectively. Then, we also need to know the corresponding homography for each occlusion to fill the disoccluded regions in the virtual view images. Thus, the proposed method determines the corresponding homography for each occlusion by measuring how many pixels in each occlusion cover unfilled pixels in the virtual view image. Figs. 12(c) and 13(c) show the classification results of occlusions for the synthetic Box and Pyramid image pairs according to each corresponding homography, respectively.
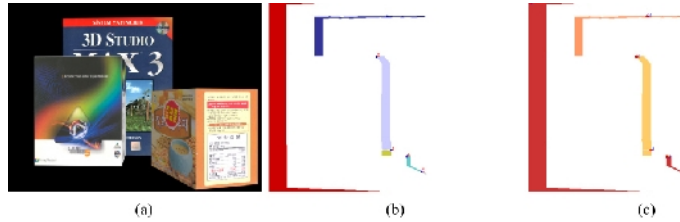
Fig. 12. Occlusion extraction results of the reference image in the synthetic Box image pair. (a) Reference image. (b) Extraction of occlusions. (c) Classification of occlusions.
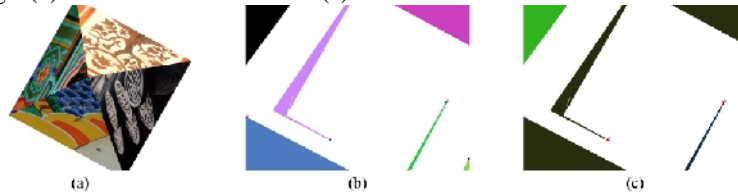

Fig. 13. Occlusion extraction results of the reference image in the synthetic Pyramid image pair. (a) Reference image. (b) Extraction of occlusions. (c) Classification of occlusions.

Given planar regions, homographies for planar regions at virtual views, and occlusions, the proposed method synthesizes virtual view images at interpolated and extrapolated views. Figs. 14 and 15 show the results of synthesized images of the synthetic Box and Pyramid image pairs, respectively. First, Figs. 14(a) and 15(a) represent the initial synthesized images without occlusion handling and inpainting. Second, Figs. 14(b) and 15(b) show the synthesized images with occlusion handling. Lastly, Figs. 14(c) and 15(c) show the synthesized images with occlusion handling and inpainting. In Figs. 14(a), 14(b), 15(a), and 15(b), green colored areas show unfilled pixels. We observe that the occlusion handling and inpainting make the virtual view images more natural by minimizing the number of unfilled pixels. In Figs. 14(a), 14(b), and 14(c), the virtual view images in the first and last columns are synthesized at the extrapolated views whereas those in the second, third, and fourth columns are synthesized at the interpolated views. Figs. 15(a), 15(b), and 15(c) are similarly defined. As you see in Figs. 14(c) and 15(c), pixel displacement by viewpoint movement is smoothly made because planar regions are correctly positioned by homography projection. Also, addition of occlusion handling and inpainting makes the virtual view images look more natural. However, in producing extrapolated view images with occlusions, some pixels in occlusions are incorrectly placed. It is because exact homography for occlusion cannot be estimated. That problem will be solved by using another reference image that has the same occlusion in it. For easy understanding of interpolated and extrapolated views, Fig. 16 shows the position of each viewpoint with the representation of virtual homography. In Fig. 16, the numbers on circles represent relative viewpoint distances between image 1 and image 2. Thus, in the case of the interpolated view, if we try to synthesize a virtual view image at the middle of viewpoints between image 1 and image 2, relative viewpoint distance could be 0.5. Then, virtual homography estimated using Eq. (10) along the forward direction with $k/N_v = 0.5$. On the contrary, in the case of the extrapolated view, virtual

homography is estimated using Eq. (10) with H^{-1} or HH instead of H. Relative viewpoint distances are similarly defined.

Fig. 14. Synthesized images of the synthetic Box image pair at virtual viewpoints. (a) Initially synthesized images. (b) Synthesized images with occlusion handling. (c) Synthesized images with occlusion handling and inpainting.
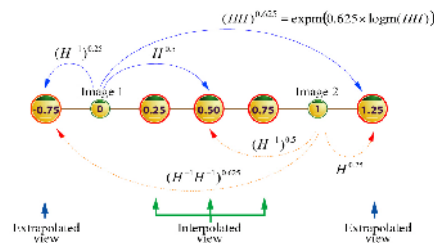
Fig. 15. Synthesized images of the synthetic Pyramid image pair at virtual viewpoints. (a) Initially synthesized images. (b) Synthesized images with occlusion handling. (c) Synthesized images with occlusion handling and inpainting.

Fig. 16. Interpolated and extrapolated viewpoint representation with virtual homography.

To show the effectiveness of the proposed method, we compare the images synthesized by the proposed method with those that are synthesized with the exact 3-D models and camera parameters. Figs. 17 and 18 show the comparison results of two sets of images for the synthetic Box and Pyramid image pairs, respectively. Figs. 17(a) and 18(a) show the synthesized images with the exact 3-D models of the synthetic Box and Pyramid image pairs, respectively, whereas Figs. 17(b) and 18(b) show those with the proposed method. Fig. 17(c) shows inverted images of the differences between Figs. 17(a) and 17(b). Fig. 18(c) is similarly defined. From the difference image depicted in Figs. 17(c) and 18(c), we can see that the proposed method effectively deals

with scene changes by viewpoint movements. Although there are some more remarkable differences near object boundaries, it is not unpleasant to the eye to represent the scene changes naturally by viewpoint movements.

Fig. 17. Comparison results of synthesized images of the synthetic Box image. (a) Synthesized images with actual 3-D model. (b) Synthesized images by the proposed method. (c) Inverted images of the differences between (a) and (b).
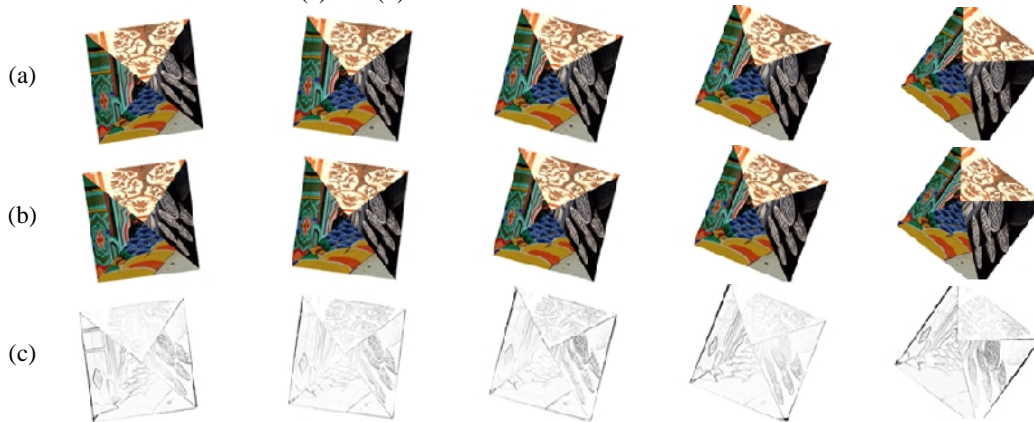

Fig. 18. Comparison results of synthesized images of the synthetic Pyramid image. (a) Synthesized images with actual 3-D model. (b) Synthesized images by the proposed method. (c) Inverted images of the differences between (a) and (b).

4.2. Experiments with Real Image pairs

To show the effectiveness of the proposed method, we employ three real image pairs that consist of three or four planar regions. Figs. 19(a), 19(b), and 19(c) show the Box, Book, and Building image pairs, respectively, in which target images are shown in the left column whereas reference images in the right column. First, Fig. 19(a) has no occlusion within viewpoint movement and planar regions are perpendicular to each other. Second, Fig. 19(b) has overlap areas between two planar regions. Lastly, Fig. 19(c) has some occlusions. To estimate image correspondences for each image pair, the SIFT is performed on the Box, Book, and Building image pairs, in which 88, 586, and 150 image correspondences are established, respectively. Figs. 20(a), 20(b), and 20(c) show the estimation results of image correspondences (left) and their classification result (right) for the Box, Book, and Building image pairs, respectively. Both target and reference images are shown.

With estimated image correspondences, we perform the iterative homography estimation to derive planar homographies for each image pair. Four, five, and three planar homographies are estimated from the Box, Book, and Building image pairs, respectively. In the Box and Book

image pairs, one of the four planar homographies and two of the five planar homographies are spurious estimates of homography, respectively. In the Building image pair, three planar homographies are correctly estimated. After estimating planar homographies for each image pair, by applying a mean-shift segmentation on the target image, sets of color segments are derived. With two sets of homographies and color segments, we perform segment-based matching. The results are depicted in Figs. 21(a), 21(b), and 21(c) from the Box, Book, and Building image pairs, respectively, in which color segments of target images are shown in the left column whereas planar regions of target images in the right column. We set the constants to $\lambda_p = 0.8$, $\lambda_l = 0.4$, and $\delta = 0.1$ in the cases of the Book and Building image pairs. Whereas, we set the constants to $\lambda_p = 0.8$, $\lambda_l = 0.7$, and $\delta = 0.3$ in the case of the Box image pair because a gray pattern on the box is difficult to distinguish from gray colored background plane.

After planar segmentation, we generate virtual homographies at interpolated and extrapolated viewpoints with the estimated homographies. Then, determination of warping order and extraction of occlusions are performed sequentially. Table 2 shows the determination results of the warping order. In Table 2, the number of elements in the warping order and the number of planar regions in the image are different. It is because the warping order is determined with all

Fig. 19. Test image pairs for view synthesis (target images in the top row and reference images in the bottom row). (a) Box image pair. (b) Book image pair. (c) Building image pair.

Fig. 20. Estimated image correspondences and their classification results. (a) Box image pair. (b) Book image pair. (c) Building image pair.
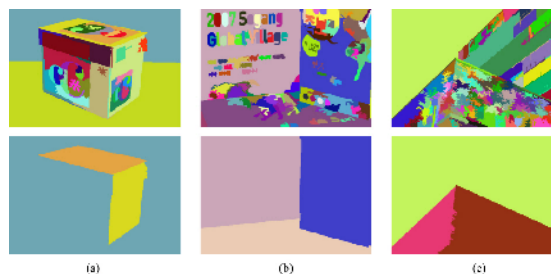

Fig. 21. Results of color segmentation and planar region extraction. (a) Box image pair. (b) Book image pair. (c) Building image pair.

the homographies estimated by iterative homography estimation. However, it does not effect on the result of image synthesis because some elements in the warping order not existing in the image pair actually do not have corresponding planar regions. Among three test image pairs, determined warping orders are only effective for the Book and Building image pairs because the Box image pair does not have any occlusion. In the Book image pair, the planar region that is indexed by the circled number 2 in Table 2 is in front of two other planar regions. Thus, the planar region that is indexed by the circled number 2 has to be warped lastly and the determined warping order, $[1, 3, 2]$, represents this condition. In the Building image pair, two planar regions that are indexed by the circled numbers 2 and 3 are in front of the planar region that is indexed by the circled number 1. Thus, these planar regions have to be warped after warping the planar region that is indexed by the circled number 1. The determined warping order, $[1, 3, 2, 4, 5]$, represents this conditions. The warping order between two planar regions indexed by the circled numbers 2 and 3 is not important because they do not have any occlusions.

Table 2. Determination of the warping order

	Box image	Book image	Building image
Planar regions			
Warping order	$[1, 2, 3, 4]$	$[1, 3, 2]$	$[1, 3, 2, 4, 5]$
Non-planar regions	4		4, 5

After warping order determination, we extract occlusions from the three reference images to fill the disoccluded regions in the synthesized images. Fig. 22 shows the estimation and classification results of occlusions. Figs. 22(a), 22(b), and 22(c) show the estimation result of occlusions and classification results of occlusions according to the corresponding homography from the Box, Book, and Building image pairs, respectively, by measuring how many pixels in each occlusion cover unfilled pixels in the virtual view image. In Fig. 22, extracted occlusions in reference images are shown in the top row whereas grouped/classified occlusions in the bottom row. In Figs. 22(a), 22(b), and 22(c), although some occlusions are classified as inexact homographies, incorrectly warped occlusions will be occluded by warping other planar regions.
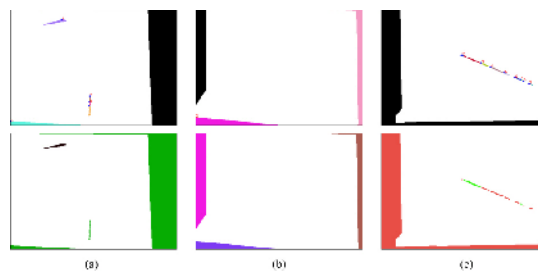

Fig. 22. Extraction and classification results of occlusions. (a) Box image pair. (b) Book image pair. (c) Building image pair.

Given planar regions, homographies for planar regions at virtual views, and occlusions, the proposed method synthesizes virtual view images at interpolated and extrapolated views. Fig. 23 shows the results of synthesized images using the Box image pair. Fig. 23(a) shows the initial synthesized image without occlusion handling and inpainting. Fig. 23(b) shows the synthesized images with occlusion handling whereas Fig. 23(c) shows those with occlusion handling and inpainting. In Figs. 23(a) and 23(b), green colored areas reflect unfilled pixels. We observe that the occlusion handling and inpainting make the virtual view images look more natural by minimizing the number of unfilled pixels. In Figs. 23(a), 23(b), and 23(c), the virtual view images in the left and right columns are synthesized at the extrapolated views and the virtual view images

in the second, third, and fourth columns are synthesized at the interpolated views, respectively. Then, each viewpoint is defined with homography that is shown in Fig. 16.
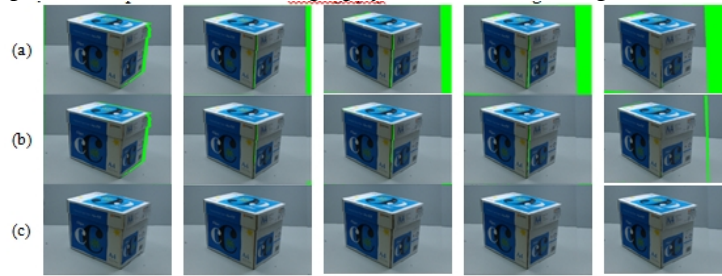

Fig. 23. Synthesized images (Box). (a) Initially synthesized images. (b) Synthesized images with occlusion handling. (c) Synthesized images with occlusion handling and inpainting.

Similarly, Figs. 24 and 25 show synthesized Book and Building image pairs, respectively, as in Fig. 23. In the case of the Book image pair, all the occluded regions are warped with the exact homographies because occluded regions exist in one of three planar regions. Thus, the synthesized images with occlusion handling are correctly done. Also, we can see that the vertical intersection area between two planar regions is exactly occluded by the viewpoint movements. In the case of the Building image pair, synthesized images accurately deal with the partial disappearance and appearance of the planar regions on the right and left sides of the brick wall, respectively, by the viewpoint movements.

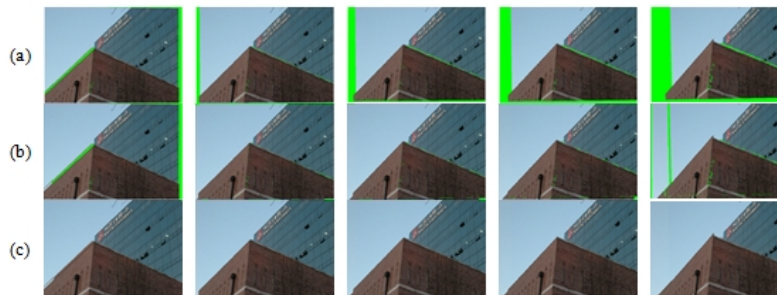Fig. 24. Synthesized images (Book). (a) Initially synthesized images. (b) Synthesized images with

Fig. 25. Synthesized images (Building). (a) Initially synthesized images. (b) Synthesized images with occlusion handling. (c) Synthesized images with occlusion processing and inpainting.

Through experimental results, in the case of the interpolated views, the proposed method accurately synthesizes the scene changes by viewpoint movements. However, in the case of the extrapolated views, as the viewpoint gets farther from the given image pair, the error becomes larger in the synthesized image, which is expected. Also, when the unfilled region is large, inpainting result is inaccurate because of insufficient information of neighboring pixels for inpainting, which, however, could be solved with an enhanced inpainting algorithm.

5. CONCLUSIONS

This paper presents an uncalibrated view synthesis method using piecewise planar regions that does not need camera parameter and scene structure estimation. The proposed method divides an image into several planar regions and synthesizes the virtual view images with planar regions and their relative order that describes the front and rear relationship among planar regions in 3-D

space. Also, the proposed method defines the virtual viewpoints with planar homographies in an image pair. Experimental results using synthetic image pairs show that the proposed method effectively deals with scene changes by viewpoint movements without estimation of 3-D information. In addition, experimental results with three real image pairs show the effectiveness of the proposed method. Further research will focus on the development of a plane extraction method and of an arbitrary viewpoint definition method that can be used in various applications with more complex scenes.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Chen, S. E., Williams, L.: View interpolation for image synthesis. In: Proceedings of ACM SIGGRAPH 1993, vol. 27, pp. 279–288. Anaheim, CA (1993)

[2]   Scharstean, D.: Stereo vision for view synthesis. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 852–857, San Francisco, CA (1996)

[3]   Perez, M. M., Pagliari, C. L.: Multi-viewpoint synthesis from uncalibrated stereo cameras. In: Proceedings of IEEE International Conference on Image Processing, vol. 1, pp. 221–224, San Antonio, TX (2007)

[4]   Zitnick, C. L., Kang, S. B., Uyttendaele, M., Winder, S., Szeliski, R.: High-quality video interpolation using a layered representation. ACM Trans. Graph. 23(3), 600–608 (2004)

[5]   Fusiello, A., Caldrer, S., Ceglie, S., Mattern, N., Murino, V.: View synthesis from uncalibrated images using parallax. In: Proceedings of International Conference on Image Analysis and Processing, pp. 146–151, Mantova, Italy (2003)

[6]   Sharma, G., Kumar, A., Kamal, S., Chaudhury, S., Srivastava, J. B.: Novel view synthesis using a translating camera. Pattern Recognition Lett. 26(4), 483–492 (2005)

[7]   Pollard, S., Pilu, M., Hayes, S., Lorusso, A.: View synthesis by trinocular edge matching and transfer. Image and Vision Computing 18(9), 749–757 (2000)

[8]   Colombari, A., Fusiello, A., Murino, V.: Uncalibrated interpolation of rigid displacements for view synthesis. In: Proceedings of IEEE International Conference on Image Processing, vol. 1, pp. 1049–1052, Genova, Italy (2005)

[9]   Tsai, R. Y.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE Trans. Robotics and Automation 3(4), 323–344 (1987)

[10]   Zhang, Z.: Flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Machine Intell. 22(11), 1330–1334 (2000)

[11]   Seitz, S. M., Dyer, C. R.: Complete scene structure from four point correspondences. In: Proceedings of IEEE International Conference on Computer Vision, pp. 330–337, Cambridge, MA (1995)

[12]   Mohr, R., Boufama, B., Brand, P.: Understanding positioning from multiple images, J. Artif. Intell. 78 (1), 213–238 (1995)

[13]   Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Computer Vision 47(1), 7–42 (2002)

[14]   Hartley, R. I., Sturm, P.: Triangulation. Int. J. Computer Vision and Image Understanding 68(2), 146–157 (1997)

[15]   Hartley, R. Zisserman, A.: Multiple View Geometry in Computer Vision, Cambridge Press. Cambridge, UK (2003)

[16]   Fusiello, A.: Uncalibrated euclidean reconstruction: A review. Image and Vision Computing 18(6), 555–563 (2000)

[17]   Criminisi, A., Reid, I., Zisserman, A.: Single view metrology. Int. J. Computer Vision 40(2), 123–148 (2000)

[18]   Hengel, A. van den, Dick, A., Thormahlen, T., Torr, P. H. S.: Building models of regular scenes from structure and motion. In: Proceedings of 17th British Machine Vision Conference, pp. 197–206, Edinburgh, UK (2006)

[19]   Fraundorfer, F., Schindler, K., Bischor, H.: Piecewise planar scene reconstruction from sparse correspondences. Image and Vision Computing 24(4), 395–406 (2006)

[20] Vigueras, J. F., Rivera, M.: Registration and interactive planar segmentation for stereo images of polyhedral scenes. Pattern Recognition 43(2), 494–505 (2010)

[21] Wei, Y., Lhuillier, M., Quan, L.: Fast segmentation-based dense stereo from quasi-dense matching. In: Proceedings of Asian Conference on Computer Vision, pp. 360–365, Jeju, Korea (2004)

[22] Bleyer, M., Gelautz, M.: A layered stereo matching algorithm using image segmentation and global visibility constraints. Int. J. Photogrammetry and Remote Sensing 59(3), 128–150 (2005)

[23] Tao, H., Sawhney, H. S.: Global matching criterion and color segmentation based stereo. In: Proceedings of IEEE International Conference on Computer Vision, vol. 1, pp. 532–539, Vancouver, Canada (2001)

[24] Mallon, J., Whelan, P. F.: Projective rectification from the fundamental matrix. Image and Vision Computing 23(7), 643–650 (2005)

[25] Howard, R. A.: Dynamic Programming and Markov Processes, MIT Press, Cambridge, MA (1996)

[26] Alexa, M.: Linear combination of transformation. In: Proceedings of ACM SIGGRAPH 2002, pp. 380–387, San Antonio, TX (2002)

[27] Telea, A.: An image inpainting technique based on the fast marching method. J. Graphics Tools 9(1), 25–36 (2004)

[28] Lowe, D. G.: Distinctive image features from scale-invariant keypoints. Int. J. Computer Vision 62(2), 91–110 (2004)

[29] Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Machine Intell. 24(5), 603–619 (2002)

Authors

J. H. Yun received the M.S. degree in electronic engineering from Sogang University in 2007. His current research interests are computer vision and computer graphics.

Rae-Hong Park was born in Seoul, Korea, in 1954. He received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1981 and 1984, respectively. In 1984, he joined the faculty of the Department of Electronic Engineering, School of Engineering, Sogang University, Seoul, Korea, where he is currently a Professor. In 1990, he spent his sabbatical year as a Visiting Associate Professor with the Computer Vision Laboratory, Center for Automation Research, University of Maryland at College Park. In 2001 and 2004, he spent sabbatical semesters at Digital Media Research and Development Center, Samsung Electronics Co., Ltd. (DTV image/video enhancement). His current research interests are computer vision, pattern recognition, and video communication. He served as Editor for the Korea Institute of Telematics and Electronics (KITE) *Journal of Electronics Engineering* from 1995 to 1996. Dr. Park was the recipient of a 1990 Post-Doctoral Fellowship presented by the Korea Science and Engineering Foundation (KOSEF), the 1987 Academic Award presented by the KITE, and the 2000 Haedong Paper Award presented by the Institute of Electronics Engineers of Korea (IEEK), the 1997 First Sogang Academic Award, and the 1999 Professor Achievement Excellence Award presented by Sogang University.