反復復元法に基づいた仮想視点画像の生成

久保田 彰[†] 齊藤隆弘[††]

[†] 東京工業大学 大学院総合理工学研究科
[††] 神奈川大学 工学部電子情報フロンティア学科
E-mail: †kubota@ip.titech.ac.jp

あらまし 多眼カメラで取得された多視点画像を用いて任意視点からの画像を合成する問題は，被写体の3次元構造が未知である場合，悪設定問題となる．そのため，多視点画像から3次元構造の推定を行い，その結果に基づき合成を行う手法が主流である．本稿では，画像の鮮鋭化手法を適用することによって，3次元構造の推定を行うことなく，高品質な任意視点画像を生成する手法を提案する．提案手法は二つの段階からなる．第一段階では，対象シーン中に複数枚の平面 (仮定平面) を想定し，それぞれの平面に基づいて light field rendering (LFR) 法により複数枚の任意視点画像を合成する．この段階で合成された画像では，奥行きが仮定平面に近い領域は鮮鋭に生成されるが，仮定平面から離れた領域には劣化が生じる．第二段階では，これらの画像にエネルギー関数最小化に基づいた反復復元法を適用し，劣化を除去した高品質な画像を生成する．

キーワード イメージ・ベースド・レンダリング，画像合成，画像復元，変分法，Total Variation

Virtual View Synthesis by a Variational Recovery Method with a Multiple Depth-Layer Model

Akira KUBOTA[†] and Takahiro SAITO[††]

[†] Tokyo Institute of Technology, Interdisciplinary Graduate School of Science and Technology
[††] Dept. of Electronics and Informatics Frontiers, Kanagawa University
E-mail: †kubota@ip.titech.ac.jp

Abstract This paper presents a novel method for virtual view synthesis using an image recovery scheme based on energy minimization. The presented method first synthesizes multiple virtual views at the same position based on multiple depth layers by using the conventional view interpolation method. The interpolated views suffer from blurring and ghosting artifacts due to the pixel mis-correspondence. Secondly, the multiple views are integrated into a novel view in which all regions are focused. In this paper, we formulate the integration problem as the image recovery problem of minimizing energy that consists of data-fidelity and regularization terms. This integration method effectively recovers the all-focused view, avoiding problems of feature correspondence and scene geometry reconstruction.

Key words image-based rendering, view synthesis, image recovery, variational method, total variation

1. Introduction

View synthesis problem using multiple views has recently attracted further interest in the fields of signal/image processing, computer vision and graphics. The problem is to synthesize a novel view from arbitrary position using a set of images (views) taken from different positions for an unknown scene. Two main approaches to solving the problem have been studied; one is image-based modeling and rendering (IBMR) [14], the other is image-based rendering (IBR) [12], [13].

IBMR approach reconstructs the scene structure or estimate the information about the scene structure such as feature correspondence. Once the scene structure is obtained, it is easy to synthesize a novel view. This approach is natural in the sense that the view synthesis problem is originally ill-posed because no information about the scene structure is available beforehand. It is however generally hard to estimate scene geometry in precise. This is an essential problem in IBMR.

In contrast to IBMR, IBR approach treats the view synthesis problem as a sampling problem [16]: to sample light rays flowing in the viewing space such that one can resample new light rays (i.e., a novel view) from the sampled rays (i.e., given multiple views) without artifacts. Therefore sampling theorem gives us the answer to how many samples are needed [19]. The required number of samples is, however, quite many in practical. This is an essential problem in IBR.

In this paper, we formulate the view synthesis problem as a problem of image recovery. The presented method consists of two steps. In the first step, multiple views at a given virtual point are generated by the traditional view interpolation method based on multiple planes we assume in the scene. The interpolated views suffer from blurring and ghosting artifacts due to the pixel mis-correspondence. In the second step, the multiple views are integrated into a novel view in which all regions are focused. In this paper, we formulate the integration problem as the image recovery problem of minimizing energy that consists of data-fidelity and regularization terms. This energy minimization method effectively recovers the all-focused view, avoiding problems of feature correspondence and scene geometry reconstruction. We show that our approach is effective for the case when there are not huge occlusions. Basic strategy is the same as that presented in [21], but the method ignores the regularization in the second step. In the presented method in this paper, we use regularization using a smooth constraint that has an effect to suppress the artifacts caused in occluded boundaries.

2. Problem formulation

We set a XYZ world coordinate system in a 3D space. In this paper, we assume that Z axis represents depth in the space and all the cameras including a virtual camera are directed to Z axis.

The view synthesis problem that we deal with in this paper is formulated as follows. We are given a set of reference images $\{l_{s,t}(x,y)\}$ taken with a 2-D camera array, each camera $C_{s,t}$ of which is located at a 2-D grid position (X_s, Y_t) regularly sampled on the XY plane with equal interval of Δs (i.e., $\Delta s = X_{s+1} - X_s = Y_{t+1} - Y_t$) in both X and Y direction, where $(s,t) \in [0,1,...,S-1] \otimes [0,1,...,T-1]$ is index for both the reference images and the capturing cameras. $S \times T$ represents the total number of the capturing cameras. The problem is, given a virtual camera C_v at an arbitrary position (X_v, Y_v, Z_v), to reconstruct a virtual view l_v at the virtual camera C_v by processing the given set of reference images $\{l_{s,t}(x,y)\}$ for an unknown scene. In our problem setting, the scene geometry is not known, but we assume that the depth range $[Z_{min}, Z_{max}]$ is known.

3. Plane-sweeping view interpolation

In the first step of our method, given a virtual camera position (X_v, Y_v, Z_v), we interpolate multiple virtual views from the same position based on multiple planes at different depths. This processing is achieved by performing the conventional view interpolation method based on different plane. At every time, the plane is swept along Z axis. An efficient arrangement of multiple planes in terms of both computation and quality was discussed in [19], [?] and we follow the presented arrangement. We call this processing *Plane-Sweeping View Interpolation* in this paper. We also call the assumed plane *focal plane*.

The idea of using multiple planes, alternative to the traditional stereo matching [15], is mainly used in a number of methods [5]~[11] for depth estimation or its application to view synthesis. In those methods, a color consistency is evaluated among corresponding pixels (or voxels) with respect to the multiple planes or depths; thus the multiple views are not usually generated. In our method, we use the interpolated multiple views as the inputs in the second step, not computing a color consistency.

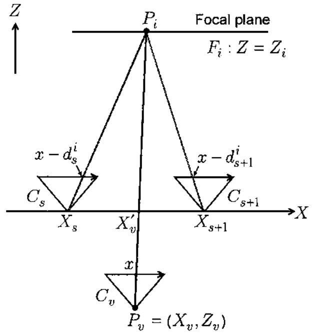


図 1: View interpolation based on the focal plane at Z_i

The multiple interpolated views can be obtained by plane-sweeping view interpolation as follows. The algorithm is essentially the same as the one presented in [17], [18]. For simplicity, consider the case where Y and y are fixed, as shown in fig. 1. Let L be the number of the virtual views, which is the same number of the focal planes we assume. Let $g_i(x)$ $(i = 1, ..., L)$ be a virtual view that we interpolate from a point $P_v = (X_v, Z_v)$ based on the focal plane at Z_i (the plane is referred to as F_i). The virtual view $g_i(x)$ is computed as the weighted average of the shifted two reference images:

$$g_i(x) = w_s \cdot l_s(x - d_s^i) + w_{s+1} \cdot l_{s+1}(x - d_{s+1}^i). \qquad (1)$$

The three pixel coordinates in the above equation, x, $x - d_s^i$ and $x - d_{s+1}^i$, are the corresponding pixel positions with respect to the point P_i on the focal plane F_i (see fig. 1). The displacement d_s^i is determined to be

$$d_s^i(x) = \frac{1}{Z_i}(X_s - X_v + Z_v x). \qquad (2)$$

We use the two reference images l_s and l_{s+1} at the two cameras C_s and C_{s+1} near the position X_v' that is the intersection of the line $P_v P_i$ with the X axis. Coefficients w_s and w_{s+1} are weighting values that are determined from $|X_s - X_v'|$ and $|X_{s+1} - X_v'|$. w_s is represented by

$$w_s = 1 - \frac{|X_v' - X_s|}{\Delta s}. \qquad (3)$$

Note that $w_s + w_{s+1} = 1$ holds.

Obviously, none of the obtained multiple views $\{g_i\}$ is the desired view l_v since we assume the scene geometry is just a plane that is different from the correct one. The regions appear in focus when the depth of the focal plane is on their corresponding depth. Aliasing artifacts are observed in the regions not in the plane. (For example, see images in fig. 3 (a)-(c).) In the second step of the presented method, we try to integrate the obtained multiple views into an all-focused view l_v where artifacts are significantly suppressed.

4. Multiple depth-layer model

Before describing the second step of our method, we introduce a multiple depth-layer model for the multiple views obtained in the first step and model the artifacts caused in these views. This modeling was first presented in [21], but we derive the model of the artifacts in detail.

4. 1 Linear image formation model

We assume that the desired all focused view l_v is composed of a sum of L components $\{\varphi_j\}$ $(j = 1, ..., L)$:

$$l_v = \sum_{j=1}^{L} \varphi_j. \tag{4}$$

The component φ_j is defined as texture lying on the focal plane F_i, but is not known. We also assume that the multiple views g_i $(i = 1, ..., L)$ obtained in the first step is modeled as

$$\begin{cases} g_1 &= h_{11} \circ \varphi_1 + h_{12} \circ \varphi_2 + \cdots + h_{1L} \circ \varphi_L \\ g_2 &= h_{21} \circ \varphi_1 + h_{22} \circ \varphi_2 + \cdots + h_{2L} \circ \varphi_L \\ &\vdots \\ g_L &= h_{L1} \circ \varphi_1 + h_{L2} \circ \varphi_2 + \cdots + h_{LL} \circ \varphi_L . \end{cases} \tag{5}$$

where h_{ij} denotes ghosting artifacts caused in texture φ_j in g_i. For $i = j$, h_{ij} becomes an identity operation. Model of ghosting artifact is described in the next subsection.

There are not any other assumptions on φ_j, if they simultaneously satisfy the above models in (4) and (5). This is because there are many possibilities on the existence of such $\{\varphi_j\}$. For instance, φ_j can be considered as a segment of each layer at depth Z_j. We may say that φ_j contains dominant features or edges lying at depth Z_j much more than other $\varphi_{i \neq j}$.

4. 2 Model of ghosting artifact

We analyze the aliasing artifacts and model them as spatially varying filters. Consider the case when the scene contains one plane object at depth Z_j whose surface is lambertian. In this case, the model in (5) is represented by

$$\begin{cases} g_j(x) &= \varphi_j(x), \\ g_i(x) &= h_{ij} \circ \varphi_j(x), \quad \text{for } i \neq j. \end{cases} \tag{6}$$

We derive the model of the operation of h_{ij}. It is found that the following relationship is held:

$$\varphi_j(x) = l_s(x - d_s^j) = l_{s+1}(x - d_{s+1}^j), \tag{7}$$

because surface property of the object plane is lambertian and because the focal plane F_j is the same as the scene geometry itself. From equations (1) and (7), the interpolated virtual view g_i can be represented by

$$\begin{aligned} g_i(x) &= w_s l_s(x - d_s^i) + w_{s+1} l_{s+1}(x - d_{s+1}^i), \\ &= w_s \varphi_j(x - d_s^i + d_s^j) + w_{s+1} \varphi_j(x - d_{s+1}^i + d_{s+1}^j). \end{aligned} \tag{8}$$

Comparison of the resultant equation in eq. (6) indicates that the operation of h_{ij} can be modeled as a filter whose coefficients are the weighting values w_s and w_{s+1}. This filter is also the model of aliasing artifacts. It should be noted that this filter is linear but shift varying (i.e., it changes depending on the pixel coordinate x), since d_s^i and d_s^j change with x, as shown in eq. (2). The filter varies with the virtual view point and the depth of the focal plane as well.

5. Variational method for recovering an all-focused virtual view

The reconstruction problem in the second step of our method is formulated as the problem of solving a set of linear equations in eq. (5) for $\{\varphi_j\}$ and reconstructing l_v in eq. (4), given $\{g_i\}$ and $\{h_{ij}\}$. One simple way to solve those equations is to use the conventional solution of simultaneous equation such as the substitution method and the iterative method. However, their solutions will be unstable, since the problem is ill-posed. This can be explained by the properties of the operators h_{ij}; they are linear but space-variant (in addition not commutative), a pseudo-inverse operation of the operator is generally unstable and sensitive to noise, which would cause undesirable ringing artifacts. Moreover, one can not also use Fourier transform based approach to the problem.

Therefore, a regularized spatial-domain approach is required for solving the eq. (5) as the well-posed problem. We present a regularized variational method in this paper. In this section, we define the energy functional to be minimized and present its minimization algorithm using time-evolution nonlinear-diffusion process.

5. 1 Energy functional

We employ the energy functional of textures $\{\varphi_j\}$ that consists of two terms, the data-fidelity energy term and the regularization energy term, as follows:

$$E[\varphi_1, ..., \varphi_L] = \iint_\Omega \Big(\|\nabla l_v\| + \frac{\lambda}{2} \textstyle\sum_{i=1}^{L} e_i^2 \Big) dxdy, \tag{9}$$

where Ω denotes the domain of the image space and λ is a positive parameter.

The first term in the energy functional is the regularization term that evaluates the total variation (TV) of the desired view l_v. $\|\nabla l_v\|$ is written by

$$\|\nabla l_v\| = \sqrt{(\partial_x \varphi_1 + \cdots + \partial_x \varphi_L)^2 + (\partial_y \varphi_1 + \cdots + \partial_y \varphi_L)^2},$$

where ∂_x and ∂_y is a differential operator in x and y direction, respectively. TV term imposes smoothness constraint on the desired view l_v. Other effects or advantages of using TV term was studied in both practical and mathematical; TV-based image restoration/enhancement approaches have been presented for many application such as image zooming [3] and noise removal [4].

The second term in the energy functional is the data-fidelity term that evaluates the square error of e_i on the difference between the interpolated view g_i and its formation model in eq. (5):

$$e_i = (h_{i1} \circ \varphi_1 + \cdots + \varphi_i + \cdots + h_{iL} \circ \varphi_L) - g_i.$$

5. 2 Energy minimization

Euler-Lagrange equation minimizing the energy functional $E[\varphi_1, ..., \varphi_L]$ with respect to φ_j is given as a partial differential equation (PDE) as follows:

$$\text{div}\left[\frac{\nabla \varphi_j}{\|\nabla l_v\|}\right] - \lambda \sum_{i=1}^{L} (h_{ij}^* \circ e_i) = 0, \tag{10}$$

where operator h_{ij}^* denotes an adjoint operator of h_{ij}.

We solve the solution of the PDE as the steady-state solution of the following time-evolution PDE:

$$\partial_\tau \varphi_j = \mathrm{div}\left[c(x,y;\tau)\nabla\varphi_j\right] - \lambda\sum_{j=1}^{L}\left(h_{ji}^{*}\circ e_j\right), \qquad (11)$$
$$c(x,y;\tau) = \min(1, 1/\|\nabla l_v\|),$$
$$\varphi_j(x,y;0) = g_j/L,$$

where τ is an artificial time-variable and the Neumann boundary condition is applied. The final solution of the desired view l_v is given as the sum of the obtained solutions of $\varphi_1,...,\varphi_L$.

The time-evolution PDE in eq. (11) acts as a nonlinear diffusion process [1] with the conduction coefficient of c if $\lambda = 0$. In addition, if c is a constant, it reduces to the isotropic heat diffusion, which equals to the blurring process by Gaussian kernel. To prevent a much large diffusion that causes blur in the solution, we use the conduction coefficient c as $\min(1, 1/\|\nabla l_v\|)$ such that the maximum is not larger than 1, instead of use of $1/\|\nabla l_v\|$. This is a similar idea used in robust anisotropic diffusion [2]. The second term in eq. (11) acts as a pseudo-inverse process, i.e., a back projection image recovery.

Notably, in the presented recovering process, neither depth map estimation nor feature matching is explicitly performed pixel by pixel. In this paper, we try to find the combination of textures φ_j that gives the minimum of the energy functional we define.

6. Experimental results on synthesized images and discussion

6.1 Numerical implementation

From the discrete version of the PDE in (11), we derive the iterative numerical scheme as the following up-date equations:

$$\varphi_{j(m,n)}^{(\tau+1)} = \varphi_{j(m,n)}^{(\tau)}$$
$$+\epsilon\Bigg\{\left(c_{(m,n)}^{(\tau)}\cdot\Delta^x\varphi_{j(m,n)}^{(\tau)} - c_{(m-1,n)}^{(\tau)}\cdot\Delta^x\varphi_{j(m-1,n)}^{(\tau)}\right)$$
$$+\left(c_{(m,n)}^{(\tau)}\cdot\Delta^y\varphi_{j(m,n)}^{(\tau)} - c_{(m,n-1)}^{(\tau)}\cdot\Delta^y\varphi_{j(m,n-1)}^{(\tau)}\right)$$
$$-\lambda\sum_{i=1}^{L}\left(\sum_{(k',l')\in\Phi} h_{ij(m-k',n-l';k',l')}\; e_{i(m-k',n-l')}^{(\tau)}\right)\Bigg\}, \qquad (12)$$

$$\Delta^x\varphi_{i(m,n)}^{(\tau)} = \varphi_{i(m+1,n)}^{(\tau)} - \varphi_{i(m,n)}^{(\tau)},$$
$$\Delta^y\varphi_{i(m,n)}^{(\tau)} = \varphi_{i(m,n+1)}^{(\tau)} - \varphi_{i(m,n)}^{(\tau)},$$
$$c_{(m,n)}^{(\tau)} = \min(1,\ 1/|\nabla l_v|_{(m,n)}^{(\tau)}),$$
$$|\nabla l_v|_{(m,n)}^{(\tau)} = \sqrt{\left(\textstyle\sum_{i=1}^{L}\Delta^x\varphi_{i(m,n)}^{(\tau)}\right)^2 + \left(\textstyle\sum_{i=1}^{L}\Delta^y\varphi_{i(m,n)}^{(\tau)}\right)^2},$$
$$e_{i(m,n)}^{(\tau)} = \left(\sum_{(k,l)\in\Phi} h_{i1(m,n;k,l)}\varphi_{1(m+k,n+l)}^{(\tau)} + \cdots\right.$$
$$\left. + \sum_{(k,l)\in\Phi} h_{iL(m,n;k,l)}\varphi_{L(m+k,n+l)}^{(\tau)}\right) - g_{i(m,n)},$$

where (m,n) is the sampled pixel position, τ is now time steps (iterations), a constant ϵ is a positive parameter that determines the rate of update. In all the experiments, we set $\epsilon = 0.05$. The operator h_{ij} is expressed by space-variant filter $h_{ij(m,n;k,l)}$ where $(k,l) \in \Phi$, its kernel support in the discrete form. The all-focused view at iteration τ is given by the sum of the obtained solutions $\{\varphi_j^{(\tau)}\}$: $l_v^{(\tau)} = \varphi_1^{(\tau)} + \cdots + \varphi_L^{(\tau)}$.

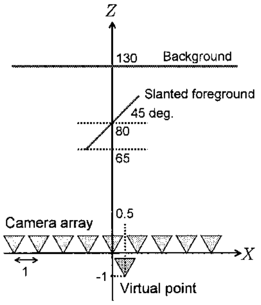


図 2: Configuration of the scene geometry and the camera array

6.2 Results

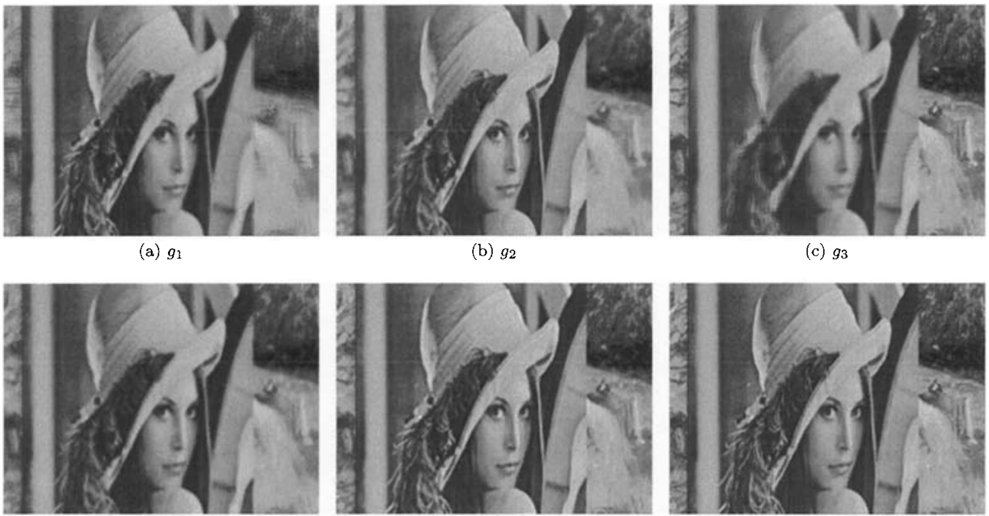
We test the performance of the presented algorithm on two different synthetic scenes. Using Pov-Ray, a ray tracing software, we synthetically created a set of 9x9 reference images $\{l_{s,t}\}$ (which are 24 bits color images of 320x240 pixels). The assumed capturing cameras are regularly arranged on the XY plane with equal spacing of $\Delta s = 1$ in parallel with Z axis.

The first test scene consists of two planes; a slanted foreground plane on which "lena" image is mapped, and a background plane on which a painting image is mapped. The depth range of the scene is [65, 130]. The configuration of the scene geometry and the camera array is illustrated in fig. 2. This scene is simple but contains depth discontinuity between the two planes.

We assumed three focal planes at 65, 87 and 130 in depth and generated three virtual views g_1, g_2 and g_3 from a position (0.5, 0.5, -1) by plane-sweeping view interpolation. The generated multiple views are shown in (a), (b) and (c) in fig. 3. In each generated view, the regions near the focal plane appear in-focus, while the regions not in the focal plane appear blurry and contain ghosting artifacts. Figure 3 (d) and (e) shows the initial solution $l_v^{(0)}$ and the all-focused view $l_v^{(300)}$ recovered from these three views by the presented method after 300 iterations. λ was set to 2 in this experiment. Although the initial solution suffers from blur in all the regions, the final solution was recovered in-focus. Figure 3 (f) shows the true virtual view that we synthetically created from the given virtual position with Pov-Ray as the ground truth. Comparison with the ground truth shows that the final solution was recovered with adequate quality.

We evaluate the performance of the presented method under various parameters settings of λ and L, and under conditions when Gaussian noise is added to the reference images, for the same test scene. Magnified region of the ground truth and the all-focused view recovered by the presented method after 300 iterations are shown in fig. 4. As a quantitative performance measure, peak signal-to-noise ratio (PNSR) is plotted for each case along with iteration as shown in fig. 5.

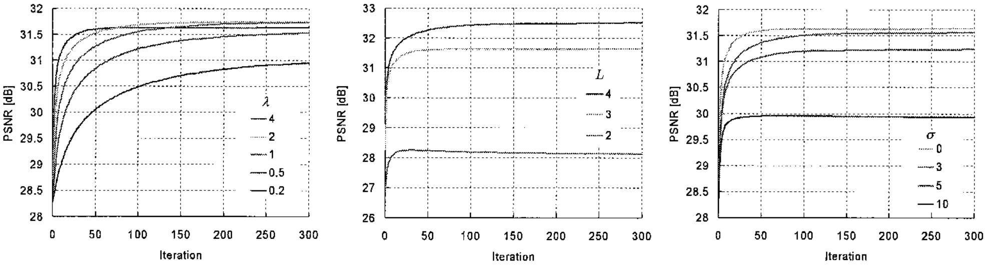
First, we tested the effect of choosing parameter λ. In fig. 4 (b), the recovered views when λ was set to 0.2, 2 and 4 are compared. In the case using λ=0.2, the fine textures are diffused and disappeared, although dominant edges still



(a) g_1 (b) g_2 (c) g_3

(d) $l_v^{(0)}$ (e) $l_v^{(300)}$ (f) The ground truth

図 3: Simulation on the synthesized test images for a scene that contains two planes; the foreground is a slanted plane on which "lena" image is mapped, the background plane on which a paint image is mapped. (a)-(c) show the views obtained through plane-sweeping view interpolation. (d) is the initial solution $l_v^{(0)}$, which is the averaged image of the multiple views in (a)-(c). (f) is the obtained solution after 300 iterations $l_v^{(300)}$. (e) is the ground truth synthetically created from the same view position.



(a) Effect of choosing parameter λ (L=3)

(b) Effect of the number of the focal planes, L (λ=2)

(c) Effect of Gaussian noise with S.D. σ (λ=2 and L=3)

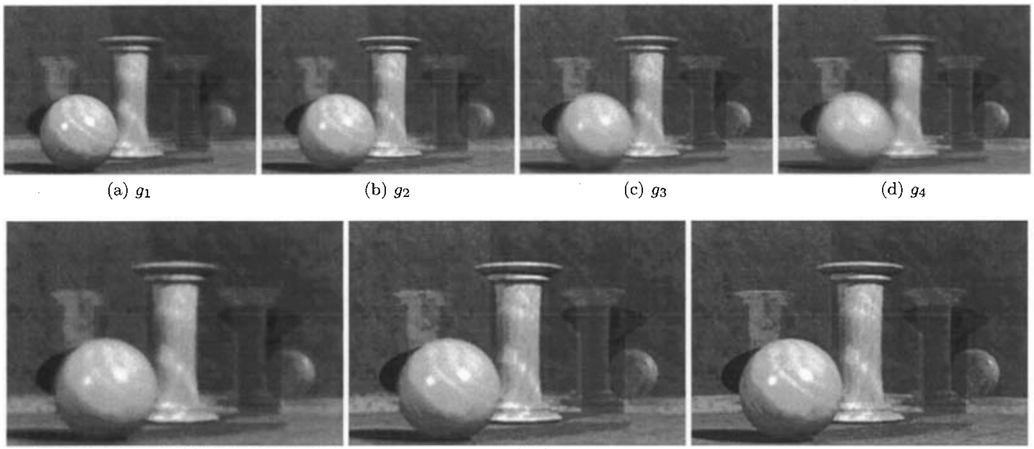
図 5: The update characteristics of the measured PSNR in green channel of the recovered view under various parameter settings and noise conditions.

remain preserved. The results using λ=2 and 4 are sharper than the result of λ=0.2, but look a little blurry compared with the ground truth in fig. 4 (a). Figure 5 (a) shows the update characteristics of the measured PSNR when various λ values was set. It indicates the tendency that the faster and the higher PSNR is increased, the larger λ is set. This is because the pseudo inverse process that acts de-blurring (or say de-ghosting) process is performed stronger than the diffusion process for the case of large λ. However, for the case of λ=4, 2, and 1, the values of PSNR reached almost the same value around 31.6 [dB] at 300 iterations; no visible difference in quality is observed between the recovered views in this case.

Second, we tested the effect of L, the number of focal planes, when we fixed λ at 2. Figure 4 (c) shows the recovered views when L is set to 2, 3 and 4. The result of L=4 is the most sharply recovered view and the result of L=2 appears blurry due to the lack of the number of focal planes, which observations on the image quality are quantitatively followed by the PSNR characteristics in fig. 5 (b).

Now two important issues are how many focal planes are needed and how the planes are arranged such that our method recovers the virtual view with the best quality. Based on the theory in [19], [20], we arranged the focal planes in such a way that their depth of fields [20] together cover the most efficiently as larger depth range as possible, given the scene range and the number of the planes. However, this arrangement may not be necessarily the best for the presented method to recover the virtual view with the best quality. It is important in both practical and theoretical to investigate those issues as a future work.

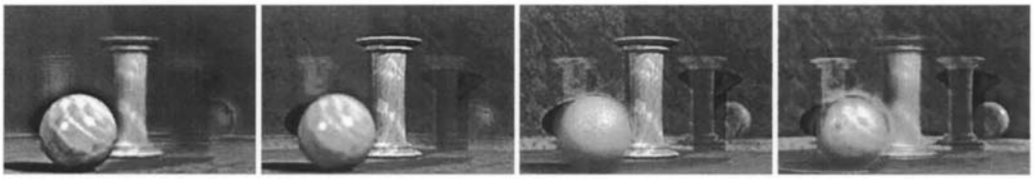
Third, the robustness of the presented method against noise was tested using the reference images to which Gaussian noise with zero mean and standard deviation σ is added.



(a) g_1 (b) g_2 (c) g_3 (d) g_4

(e) $l_v^{(0)}$ (f) $l_v^{(300)}$ (g) The ground truth

図 6: Virtual views synthesized by plane-sweeping view interpolation from a view point (0.5, 0.5,-1) for a scene whose depth range of [65, 100]. (e)-(g) show comparison between the initial solution $l_v^{(0)}$, the obtained solution after 300 iterations $l_v^{(300)}$, and the ground truth.



(a) φ_1 (b) φ_2 (c) φ_3 (d) φ_4

図 7: Obtained texture components φ_j after 300 iterations. Intensity values are multiplied by 4 for visibility.

Figure 4 (d) shows the recovered virtual view when σ is 3, 5 and 10 for fixed parameters of λ=2 and L=3. The PSNR characteristics are measured as shown in fig. 5 (c). Those results clearly shows that the presented method is robust to noise. This is one of advantages of the presented method. Almost no visible artifacts can be recognized in the recovered view for the case of $\sigma = 3$ and 5. Moreover, even in the most noisy case of $\sigma = 10$, which is a huge noise level, the noise amplification is not so visible and the loss of PSNR compared with the noise free case ($\sigma = 0$) was measured to be 1.7 [dB].

Finally, we show the result for the other scene that is more complex than the first scene. The scene consists of a specular sphere and a pole in front of reflective two walls, involving reflections, shadows and occlusions, as you can see fig. 6. Although the depth range of the scene is [65, 100], taking into account the reflected textures on the back wall, we used four focal planes at 65, 78, 98, and 130 in depth. Four virtual views generated by plane-sweeping view interpolation from a view point (0.5, 0.5,-1) are shown in (a)-(d) fig. 6. Note that the reflected textures of the sphere is in-focus in the view g_4.

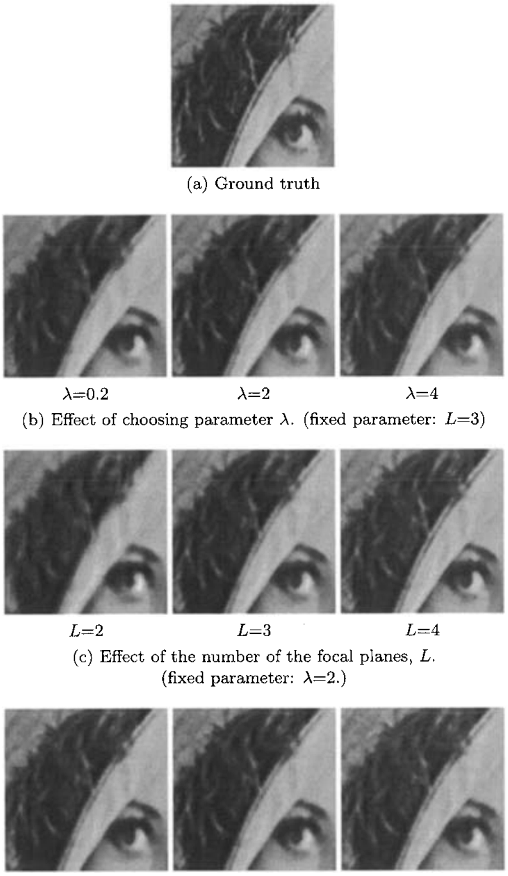
The initial solution $l_v^{(0)}$ and the finally recovered view $l_v^{(300)}$ is shown in fig. 6 (d) and (e), respectively. Parameter λ was set to 2. From comparison of the recovered view $l_v^{(300)}$ and the ground truth in fig. 6 (f), it can be seen that the presented method effectively integrates the focused regions of four views into the all-focused view. The recovered view seems to be slightly blurry, but it does not contain visible artifacts and reconstructed with sufficient quality.

Figure 7 shows obtained texture components φ_j. The results indicates that no precise segmentation is needed for the case when there are not so huge occlutions in a scene. The depth estimation in IBMR approaches corresponds to finding φ_j as a segmented layer. From the results in fig 7, it is not necessary for φ_j to be segmented region at the corresponding layer. The obtained φ_j contains dominant texture that exits at the corresponding depth.

Our method does not use any explicit constraints on the depth so there is the limitation of the depth range in which the presented method can recover the virtual view with sufficient quality. The limitation will be analyzed in future.

7. Summary

In this paper, we have presented a novel view synthesis method that consists of two steps. We interpolate multiple views by plane-sweeping view interpolation, then use them as initial estimate and recover the desired view by regularized variational method. The presented method does not require depth estimation or feature matching. For the case when there are not huge occlusions, the method can generate virtual view with adequate quality.



(a) Ground truth

λ=0.2 λ=2 λ=4

(b) Effect of choosing parameter λ. (fixed parameter: L=3)

L=2 L=3 L=4

(c) Effect of the number of the focal planes, L.
(fixed parameter: λ=2.)

σ=3 σ=5 σ=10

(d) Effect of Gaussian noise with standard deviation of σ.
(fixed parameters: λ=2 and L=3)

図 4: Comparison of magnified regions between the ground truth and the view recovered after 300 iterations under various parameters settings and noise conditions.

References

[1] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern. Anal. & Mach. Intell., 12, 7, pp.629-639, 1990.

[2] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," IEEE Trans. on Image Processing, 7, 3, pp.421-432, 1998.

[3] F. Malgouyres and G. Guichard, "Edge direction preserving image zooming: a mathematical and numerical analysis," J. Num. Anal., 39, 1, pp.1-37, 2001.

[4] L. Rudin, S. Osher, and E. Fetami, "Nonlinear total variation based noise removal algorithm," Physica D, 60, pp.259-268, 1992.

[5] R. T. Collins, "Space-Sweep Approach to True Multi-Image Matching," *CVPR96*, pp. 358-363, 1996.

[6] Szeliski, R. and Golland, P., "Stereo Matching with Transparency and Matting," *International Journal of Computer Vision*, Vol. 32, No. 1, pp. 45-61, Aug. 1999.

[7] I. Geys, T. P. Koninckx, and L. V. Gool, "Fast interpolated cameras by combining a GPU based plane sweep with a max-flow regularisation algorithm," Proc. of Intl Symp. 3D Data Processing, Visualization and Transmission, pp. 534-541, 2004

[8] K. Takahashi, A. Kubota, T. Naemura, "A Focus Measure for Light Field Rendering," *ICIP2004*, Vol. 3, pp. 2475-2478, 2004.

[9] S. M. Seitz and C. R. Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring," *CVPR97*, pp. 1067-1073, 1997

[10] K. Kutulakos and S. M. Seitz, "A theory of shape by space carving," Int. Journal of Computer Vision, 38(3):198218, July 2000.

[11] A. Broadhurst, T.W. Drummond and R. Cipolla, "A Probabilistic Framework for Space Carving," ICCV2001, Vol. 1, pp. 388-393, 2001.

[12] C. Zhang and T. Chen, "A survey on image-based rendering - representation, sampling and compression," *EURASIP Signal Processing: Image Communication*, Vol. 19, pp. 1-28, Jan. 2004.

[13] H-Y. Shum, S. B. He, and S-C. Chan, "Survey of Image-Based Representations and Compression Techniques", *IEEE Trans. on CSVT*, Vol. 13, No. 11, pp. 1020 - 1037, 2003.

[14] M. Oliveira, "Image-Based Modeling and Rendering Techniques: A Survey," *RITA - Revista de Informatica Teorica e Aplicada*, Volume IX, pp. 37-66, 2002.

[15] U. R. Dhond and J. K. Aggarwal "Structure from stereo: a review," *IEEE Trans. on System, Man, and Cybernetics*, vol. 19, no. 6, pp. 1489-1510, 1989.

[16] E. H. Adelson, J. R. Bergen, "The plenoptic function and the elements of early vision," *Computational Models of Visual Processing The MIT Press, Cambridge, Mass.* 1991.

[17] M. Levoy, P. Hanrahan, "Light field rendering," *SIGGRAPH96*, pp.31-42, 1996.

[18] A. Isaksen, M. Leonard, S. J. Gortler "Dynamically Reparameterized Light Fields," MIT-LCS-TR-778, 1999.

[19] J-X. Chai, X. Tong, S.-C. Chany, H.-Y. Shum, "Plenoptic Sampling," *SIGGRAPH2000*, pp. 307-318, 2000.

[20] K. Takahashi, T. Naemura, H. Harashima, "Depth of Field in Light Field Rendering," *ICIP2003*, pp. 409-412, 2003.

[21] A. Kubota, K. Aizawa, T. Chen, "Virtual View Synthesis through Linear Processing without Geometry," *ICIP2004*, pp. 3009-3012, 2004