基礎論文 Paper

Novel View Synthesis Based on View-dependent Texture Mapping with Geometry-aware Color Continuity

Keita Katagiri[*1], Yuta Nakashima[*1], Tomokazu Sato[*1], and Naokazu Yokoya[*1]

Abstract – **Novel view synthesis (NVS) is renowned for its variety of applications, and recent techniques are based on an image-based approach, which uses multiple input images to synthesize novel views, with the help of 3D geometry. View-dependent texture mapping is one of the most well-known techniques of this kind. However, it sometimes suffers from blurry or duplicated edges when the geometry is inaccurate because it blends several input images for alleviating inconsistent lighting conditions among input images. This paper presents a view-dependent texture mapping-based method for NVS that leverages geometry-aware color continuity in texture selection. Our color continuity constraint suppresses noticeable texture boundaries in resulting novel views without blending, while our method relaxes the constraint on steep geometric changes. Since such geometric changes are often followed by surface property changes, our new color continuity constraint can lead to more natural novel views. We formulate our texture selection by an energy minimization problem and solve it with the graph cut algorithm. Our experimental results have demonstrated the superiority of our method.**

Keywords : **Novel view synthesis, view-dependent texture mapping, geometry-aware color continuity, graph cut**

1 Introduction

Novel view synthesis (NVS) is a technique to synthesize an image of a scene/object from an arbitrary viewpoint using a set of multiple images. This technique has received much attention for its various applications, including free-viewpoint television [1]. A conventional approach for NVS uses a 3D geometry of the scene/object and possibly applies textures, which are extracted from the image set, to the geometry. Since handcrafting such 3D geometry is demanding and is almost impossible for ordinary users, automatic techniques have been adopted for 3D reconstruction, such as structure from motion [2, 3] together with multi-view stereo (MVS) [4]. Unfortunately, these techniques may suffer from inaccurate 3D geometry, resulting in the loss of details. Another well-known approach is image-based rendering, which smoothly interpolates different images or uses light fields obtained from the image set. This approach can synthesize a faithful view but requires a vast number of images.

A series of techniques have been proposed in between the geometry- and image-based approaches for alleviating their drawbacks. View-dependent texture mapping (VDTM) [5] is one of such techniques, which applies some images to a 3D geometry as its textures. These images are adaptively selected from the image set based on the spatial relationship among the geometry, the viewpoint specified by the user (hereinafter, the novel viewpoint), and the viewpoints of cameras used for capturing the images. It can reconstruct the details of the target while the number of required images can reduce thanks to the geometry. Due to this desirable features, VDTM has been widely studied, and a variety of its extensions have been proposed [6, 7].

However, the VDTM technique may still suffer from visual artifacts. To alleviate the influence of lighting condition change while capturing the images, it smoothly blends multiple images and uses the blended image as a texture [5]. This may lead duplicated or blurry patterns on the target surface as shown in Fig. 1(top) if automatic technique for 3D reconstruction are not sufficiently accurate.

This paper presents a new method for NVS, which is also an extension of VDTM to remedy the problem of blurry or duplicated patterns. We cast the problem of selecting an appropriate image to be applied as texture to an energy minimization problem. In addi-

[*1] Nara Institute of Science and Technology

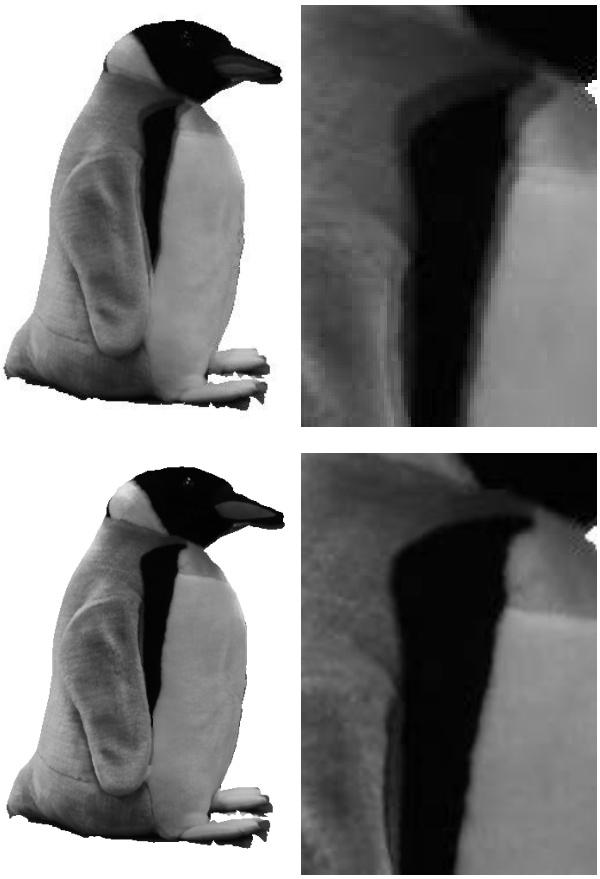

Fig. 1 Examples of original VDTM [5] and its closeup view (top) and our method and its closeup view (bottom).

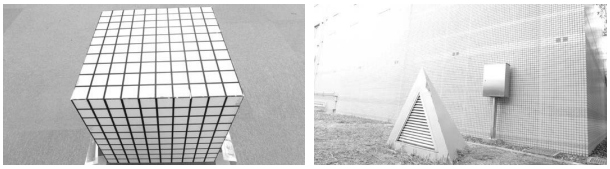

Fig. 2 Different lighting conditions for different surfaces (left) and different surface properties on the grassy ground and the concrete wall (right).

tion to the spatial relationship among the viewpoints and the geometry, our method takes into account the color continuity in a resulting view, as Kopf et al. [7] do. Moreover, observing that steep changes in the geometry imply different surface properties and lighting conditions as shown in Fig. 2, we introduce a novel geometry-aware weighing to the energy function so that steep changes in the geometry coincide the texture boundaries. Since this texture selection may not fully remove discontinuities, our method also modifies the luminance of the pixels.

Our main contributions are summarized as follows: (i) Being different from Kopf et al.'s method [7], our method leverages the geometry in the texture selection through an energy function. This energy function allows our method to relax the color continuity constraint around steep geometric changes. (ii) We introduce smooth luminance modification to reduce discontinuities. Compared with image blending, our luminance modification is more robust to misalignment of texture patterns. (iii) We experimentally demonstrate the effectiveness of our new energy function by comparing synthesized views with various parameters. We also show that luminance modification reduces discontinuities in the views.

2 Related Work

In recent decades, NVS has been widely studied, and various approaches have been proposed. One of the major approaches is geometry-based one. This approach uses 3D geometries of objects to synthesize their novel views, possibly with textures on the surfaces. The 3D geometries can be handcrafted or automatically generated using a 3D reconstruction technique, such as shape from silhouette [8, 9] and MVS [10, 11, 4]. Depth cameras are also used for acquiring geometries [12]. The quality of novel views depends much on the quality of the 3D geometries, and an automatic 3D reconstruction technique is often insufficient. Another approach for NVS is image-based rendering, which interpolates input images to synthesize a novel view among them [13, 8, 9] or obtains light fields from input images [14, 15, 16, 17, 18]. This type of technique is sometimes criticized for its requirement of a large number of input images.

Recent research efforts try to improve image-based rendering by making use of the 3D geometries of target objects. The unstructured light field is one of such techniques proposed by Davis et al. [18]. In order to reduce the number of input images over the original light field technique, they use a rough proxy of the object geometry. Debevec et al. proposed VDTM [5], which much relies on the 3D geometry. It applies the input images to the 3D geometry as its textures. VDTM adaptively changes the images to be applied as texture in accordance with the novel viewpoint such that the angles formed by the novel viewpoint, the input image's viewpoint, and the object are the smallest among the other input images. Since it can use different input images for a single surface, the texture boundary can be noticeable and

degrade the visual quality. In order to alleviate the texture boundary problem, they blend multiple input images so as to smoothly change the texture. Nobuhara et al. proposed to find plausible geometry given a viewpoint by minimizing an energy function with a photo-consistency term [19].

An interesting extension of VDTM is the technique proposed by Kopf et al. [7]. They defined an energy function that counts the continuity of the adjacent textures as well as the original VDTM's angle-based criterion, which determines the input images to be used as texture by minimizing the energy function.

In this paper, we propose a method for NVS similar to Kopf et al.'s. The main difference of our method from theirs is our new energy function that relaxes the continuity constraint in texture if the 3D geometry changes rapidly. Since the surface property of the object/scene often changes coincidently with the geometry changes, our new energy function can synthesize a more visually natural novel view.

3 Novel View Synthesis Considering Geometry-aware Color Continuity

Figure 3 shows an overview of the proposed method, which consists of geometry reconstruction (GR) and NVS stages. In the GR stage, we estimate the intrinsic and extrinsic camera parameters for each image in an input image set $S = \{I_n | n = 1, \ldots, N\}$ using structure from motion [2, 3]. It then reconstructs the geometry of the target with MVS [4]. The geometry is represented in a 3D mesh model M. In the NVS stage, provided a viewpoint of a novel view as the virtual camera's intrinsic and extrinsic parameters, our method generates a depth map for the view from M with a standard rendering pipeline and selects an image in S used as the texture for each pixel in the view by energy minimization. Based on the texture selection result, it colors the pixels in the novel view and smoothly modifies luminance to reduce color discontinuity. In the following sections, we describe our texture selection and smooth luminance modification.

3.1 Texture selection

We consider that the requirements for texture selection can be summarized as follows: (1) For pixel i in the novel view, texture selection must find the image that captures its corresponding 3D point from a direction close to that of the novel viewpoint. (2) The 3D point corresponding to the pixel i is potentially occluded in image I_n. Texture selection must verify the visibility of the 3D point in I_n (visibility check). (3) Texture selection must take into account the color continuity of adjacent pixels in the novel view. (4) Steep changes in the geometry may imply changes in the target's surface properties. For such regions, requirement (3) must be relaxed.

As mentioned above, our method formulates texture selection as an energy minimization problem with respect to labels $x_i \in \{1, \ldots, N\}$, which represents the image used as the texture for pixel i in the novel view. Encoding above four requirements, our energy function is defined by

$$E(X) = \sum_{i \in A} \alpha_i(x_i) f_i(x_i) + K \sum_{(i,j) \in B} \beta_{ij} g_{ij}(x_i, x_j), \quad (1)$$

where A is the set of pixels to which depth values are assigned, B is the set of pairs of adjacent pixels, and $X = \{x_i | i \in A\}$. Requirements (1)–(4) are encoded in f_i, α_i, g_{ij}, and β_{ij}, respectively. K is a parameter to determine the contribution of the pairwise term. This energy function can be minimized by the graph cut algorithm [20].

Data term $f_i(x_i)$ penalizes the label that represents an image in S captured from a direction different from that of the novel viewpoint, as in [5]. Let $\mathbf{t}_\mathrm{v}$ be the 3D position of the novel viewpoint and $\mathbf{t}_{x_i}$ be that of image I_{x_i}'s viewpoint. $\mathbf{t}_{x_i}$ can be calculated from the extrinsic camera parameters of I_{x_i}. From the depth value d_i for pixel i in the novel view, we can regain its corresponding 3D position $\mathbf{p}_i$ on geometry M using the intrinsic and extrinsic camera parameters for the novel viewpoint. The data term is defined with respect to $\mathbf{p}_i$ by the angle, i.e.,

$$f_i(x_i) = \cos^{-1} \frac{(\mathbf{t}_\mathrm{v} - \mathbf{p}_i) \cdot (\mathbf{t}_{x_i} - \mathbf{p}_i)}{\|\mathbf{t}_\mathrm{v} - \mathbf{p}_i)\| \, \|\mathbf{t}_{x_i} - \mathbf{p}_i\|}. \quad (2)$$

Data term weight $\alpha_i(x_i)$ is concerned with visibility check, which gives a large value when $\mathbf{p}_i$ is occluded in image I_{x_i}. Let d'_{ix_i} denote the depth value of $\mathbf{p}_i$ in I_{x_i}, which is calculated using $\mathbf{p}_i$ and I_{x_i}'s camera parameters, and $\mathbf{p}_i$ is projected to pixel k in I_{x_i}. After the GR stage, we can generate the depth map for each image in S. We denote the depth value at pixel k in I_{x_i} by e_{kx_i}. The depth values d'_{ix_i} and e_{kx_i} are close to each other if $\mathbf{p}_i$ is not occluded but not otherwise. Using a certain threshold θ_d for the difference between these depth values, we define

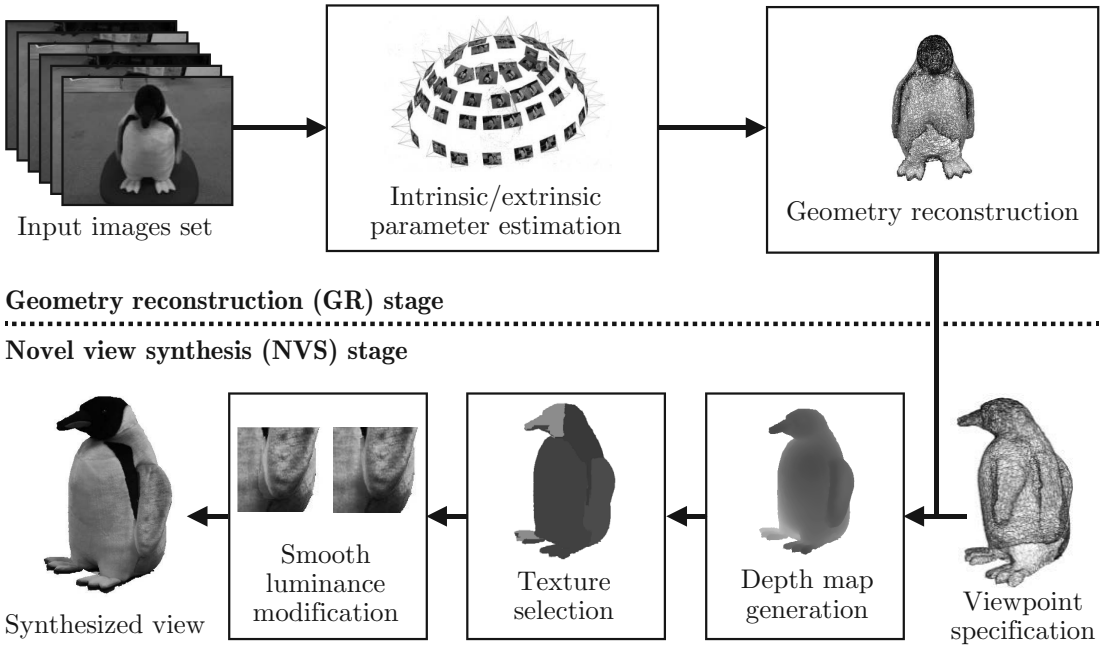

Fig. 3 Overview of the proposed method.

data term weight $\alpha_i(x_i)$ as

$$\alpha_i(x_i) = \begin{cases} 1 & \text{if } |d'_{ix_i} - e_{kx_i}| < \theta_{\mathrm{d}} \\ L & \text{otherwise} \end{cases}, \qquad (3)$$

where L is a sufficiently large value.

Pairwise term $g_{ij}(x_i, x_j)$ encodes requirement (3), where a pair of pixels (i, j) are in the set B of adjacent pixels. This term gives a large value when the labels x_i and x_j disagree and the corresponding colors $\mathbf{c}_{ix_i}$ and $\mathbf{c}_{jx_j}$ are different, where $\mathbf{c}_{ix_i}$ represents the color (in the RGB color space) at the pixel corresponding to $\mathbf{p}_i$ in I_{x_i}. We implement this by

$$g_{ij}(x_i, x_j) = \max(\|\mathbf{c}_{ix_i} - \mathbf{c}_{ix_j}\|_\infty, \|\mathbf{c}_{jx_i} - \mathbf{c}_{jx_j}\|_\infty), \qquad (4)$$

where $\|\cdot\|_\infty$ is the maximum norm. g_{ij} is 0 if $x_i = x_j$; otherwise, it gives the maximum difference of RGB components in adjacent pixels i and j.

pairwise term weight β_{ij} relaxes the influence of g_{ij} by reducing its value when the geometry changes steeply. This weight can be viewed as a geometry-aware color continuity together with g_{ij}. Observing that changes in the geometry can be detected using its normals, we first generate a normal map, as shown in Fig. 4 for the novel viewpoint by calculating the normal for each face in mesh model M, and assign the normal to corresponding pixels. Since inaccurate M causes noisy normal map, we apply a Gaussian filter to the map and normalize each pixel's normal to length 1. We denote the normal for pixel i by $\mathbf{n}_i$ and define the pairwise term weight by

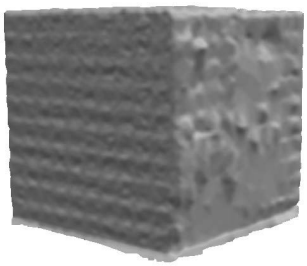

Fig. 4 Examples of normal maps (Best viewed in color).

$$\beta_{ij} = \begin{cases} 1 & \text{if } \cos^{-1} \frac{\mathbf{n}_i \cdot \mathbf{n}_j}{\|\mathbf{n}_i\| \|\mathbf{n}_j\|} < \theta_{\mathrm{n}} \\ \epsilon & \text{otherwise,} \end{cases} \qquad (5)$$

where θ_{n} is a threshold and $\epsilon \leq 1$ a factor to reduce g_{ij}.

3.2 Smooth Luminance Modification

Our energy minimization-based texture selection reduces discontinuities in the synthesized view. However, it may not fully remove such visual artifacts if the lighting condition changes while capturing images in S. One approach to alleviating such artifacts can be blending as in [5], but this suffers from blurring or duplicated patterns caused by inaccurate camera parameter estimation and 3D geometry reconstruction. Another approach can be globally ad-

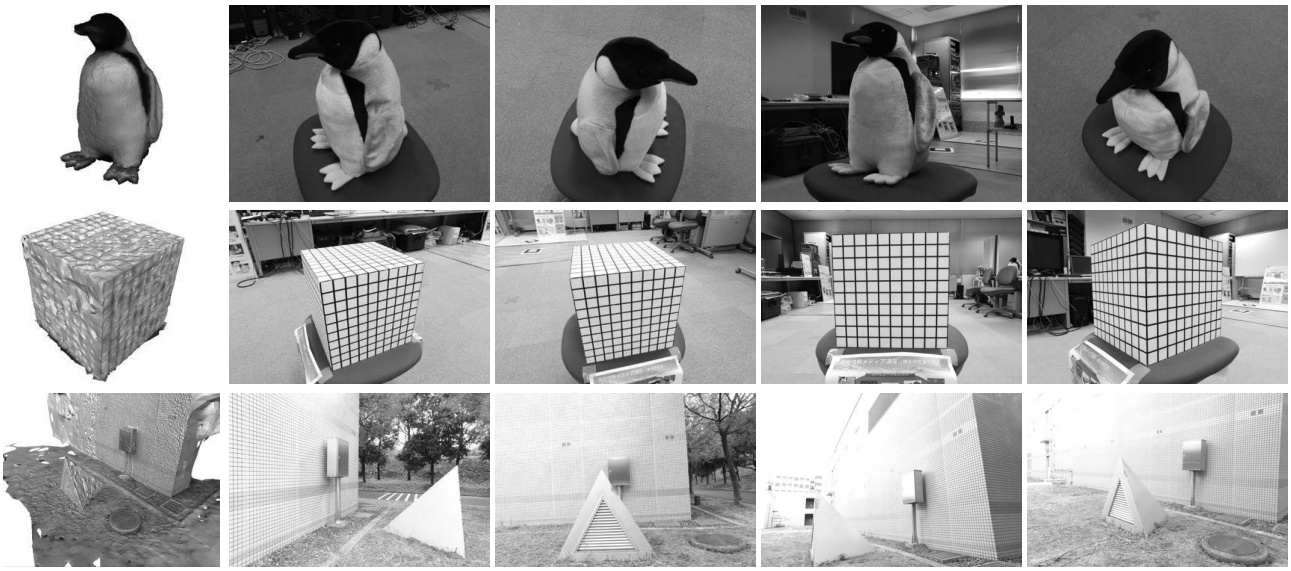

Fig. 5 Datasets. From top to bottom: datasets (A), (B), and (C). The leftmost column shows reconstructed 3D mesh models and others show examples of input images.

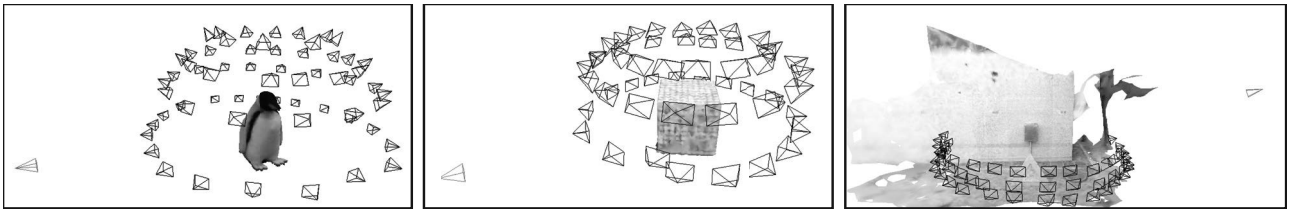

Fig. 6 Cameras that captured the input images (black pyramids) and viewpoints for novel view synthesis in our experiments (red pyramids).

justing luminance of each input image beforehand; however, it cannot handle, for example, specular reflection and local luminance changes caused by the shadow of the photographer. We instead address this problem by locally and smoothly modifying the luminance.

For pixel i, we set a window W_i centered at that pixel. Let $Q_i = \{x_j | j \in W_i\}$ be the set of labels in W_i and thus the elements in Q_i are not unique because the same label can be assigned to different pixels. Also $\bar{I}_{in}$ denotes a local luminance value around i calculated from a novel view whose pixels' labels are all n. In other words, $\bar{I}_{in}$ is calculated from a novel view that is synthesized only from I_n. For a smooth change in luminance, we modify the novel view's color $\mathbf{c}_i$ for pixel i whose label is x_i as follows:

$$\mathbf{c}'_i = \frac{\mathbf{c}_i}{\bar{I}_{ix_i}|Q_i|} \sum_{n \in Q_i} \bar{I}_{in}. \tag{6}$$

This luminance modification mixes the luminance values $\bar{I}_{in}$ using a weight proportional to the number of pixels in the window, to which label n is assigned. Misalignment of texture patterns does not cause blurring or duplicated patterns.

Table 1 Specification of datasets used in experiments.

Dataset	Image size	Number of images
(A)	1696×1280	61
(B)	1920×1280	51
(C)	1920×1280	51

4 Experimental Results

We experimentally demonstrate the effects of our pairwise term by comparing synthesized views with different parameter values of K in Eq. (1) and ϵ in Eq. (5). We also show the advantage of smooth luminance modification.

In our implementation, we used VisualSFM [2, 3] for camera parameter estimation and CMPMVS [4] for 3D geometry reconstruction. The parameter values were heuristically fixed to $L = 10000$, $\theta_{\mathrm{d}} = 0.1$,

and $\theta_{\mathrm{n}} = 5°$. For smooth luminance modification, we used a 51×51 window. Table 1 shows the specification of the datasets used in our experiments and Fig. 5 shows some examples of input images and 3D mesh models by CMPMVS. In Fig. 6, the black pyramids represent the camera parameters for input images. Dataset (A) contains a stuffed penguin, and CMPMVS was able to reconstruct it accurately. Dataset (B) captures a box with grid- and dot-pattern texture. Estimated camera poses and the 3D mesh model were erroneous. Dataset (C) is a scene with a wall, grassy ground, and a small structure. CMPMVS worked relatively well with this dataset. K and ϵ were empirically determined for all the datasets in Fig. 5. With a larger K, our method only uses a smaller number of input images because the pairwise term dominates the energy function. K may need to be adjusted for each dataset.

Figures 7, 9, and 10 show some examples of synthesized novel views with different parameters (K, ϵ), as well as novel views by blending [5], for datasets (A), (B), and (C), respectively. Novel view's camera parameters are represented in the red pyramids in Fig. 6. For datasets (A) and (C), closeup views are provided in Figs. 8 and 11. In these results, luminance modification was disabled to demonstrate how our energy function remedies discontinuities. The pairwise term is not used with $K = 0$ (and thus no ϵ), which corresponds to VDTM [5], except for blending. A non-zero K with $\epsilon = 1$ roughly mimics Kopf et al's method [7].

For dataset (A), as shown in Figs. 7 and 8, the pairwise term successfully reduced the visual artifacts around label boundaries, compared with the $K = 0$ case. Different parameter values for the pairwise term's weight ϵ did not change the synthesized view much. The novel view by blending exhibits a duplicated boundary between the black and gray regions around the penguin's neck, although the discontinuity around label boundaries are reduced.

The novel view from dataset (B) has discontinuities for $K = 0$ (horizontal lines are broken) as in Fig. 9. For $K = 0.4$ and $\epsilon = 1$, the horizontal lines are still disconnected, but the labels change on a certain horizontal line to maintain color continuity. When $K = 0.4$ and $\epsilon = 0.5$, the labels' change coincides with steep change in the geometry. Due to relatively large errors in the geometry and cam-

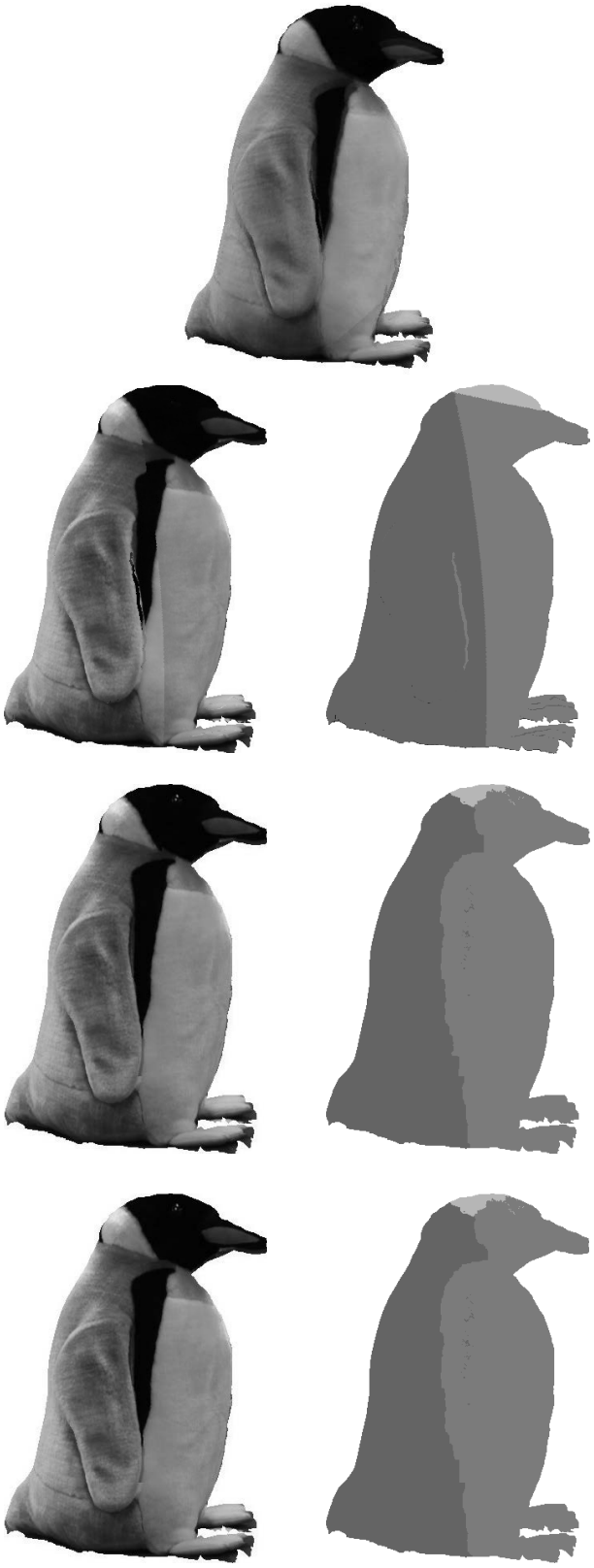

Fig. 7 Synthesized novel views for dataset (A) (left) and label maps representing x_i (right). From top to bottom: blending, $K = 0$, $(K, \epsilon) = (0.4, 1)$, $(K, \epsilon) = (0.4, 0.5)$. White pixels indicate that they are not on the 3D mesh model.

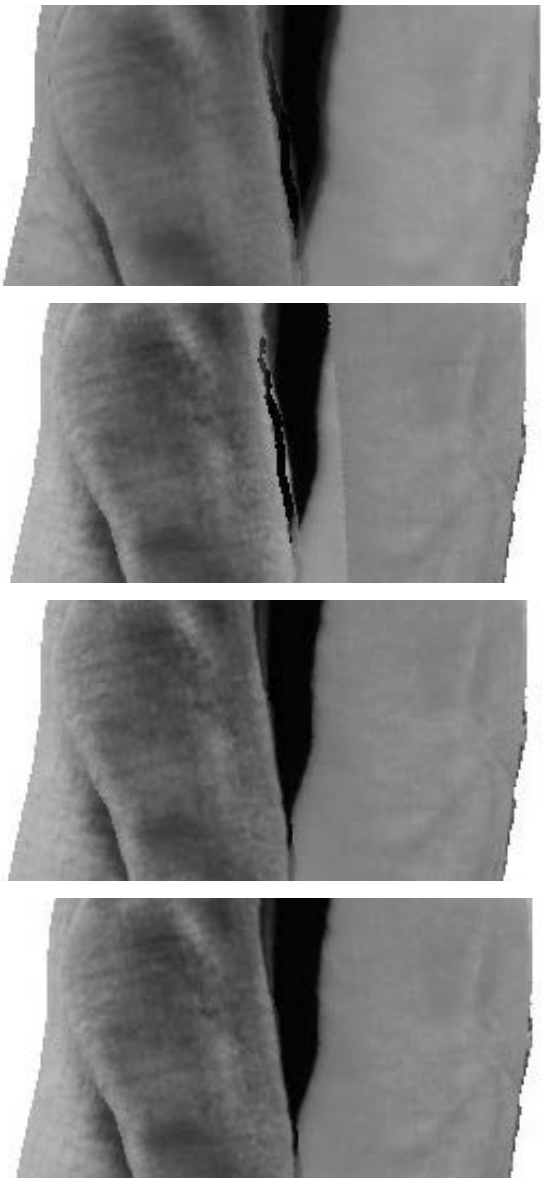

Fig. 8 Closeups of novel views in Fig. 7.

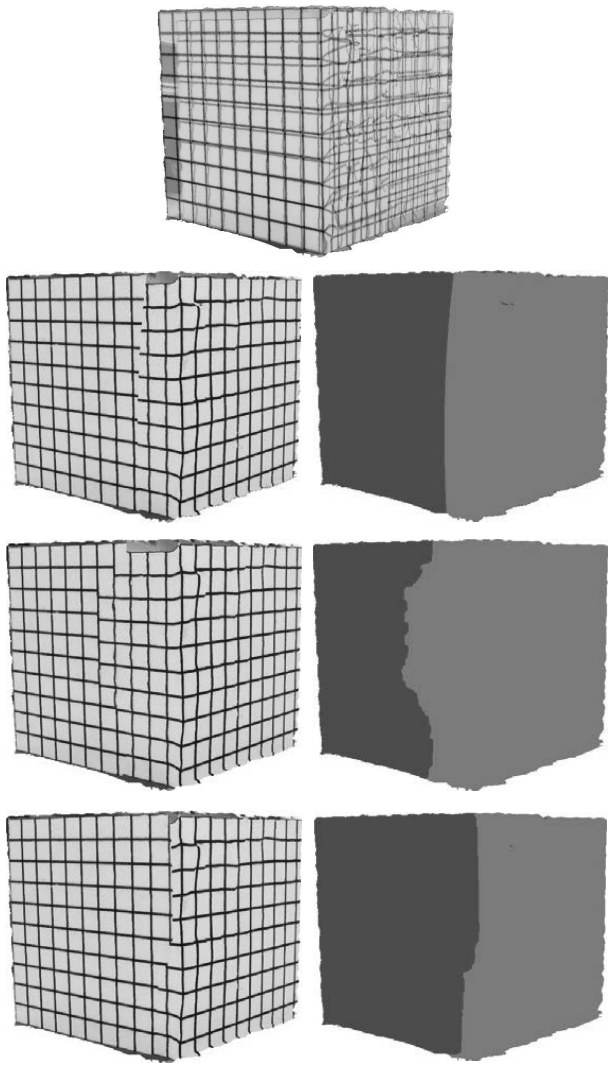

Fig. 9 Synthesized novel views for dataset (B) (left) and label maps representing x_i (right). From top to bottom: blending, $K = 0$, $(K, \epsilon) = (0.4, 1)$, $(K, \epsilon) = (0.4, 0.5)$.

era poses, all novel views were not satisfactory, but we confirmed that the proposed method, as well as VDTM considering color continuity ($K = 0.4$ and $\epsilon = 1$), reduced discontinuities without blending multiple input images. With blending, multiple black lines appear, especially in the face on the left hand side. This is because its geometry is not accurately compared to the other face, and the rough surface maps the texture at different positions.

For dataset (C), as shown in the closeup views in Figs. 10 and 11, the labels change in one of the faces of the trigonal pyramidal object for $K = 0$ and $(K, \epsilon) = (0.4, 1)$. The novel view by blending gave similar artifacts. On the other hand, the labels' boundary coincides the geometry change for $(K, \epsilon) = (0.4, 0.5)$, and the labels' boundary is no longer significant. For this dataset, the data term basically preferred the labels for $K = 0$ and $(K, \epsilon) = (0.4, 1)$ and the pairwise term was small enough to maintain the boundary since the difference in color was moderate yet noticeable for human viewers. For $(K, \epsilon) = (0.4, 0.5)$, the pairwise weight successfully worked as an incentive to change labels on the geometry change. The resulting views demonstrate the superiority of our method's visual quality.

To show the proposed method's sensitivity to K, we synthesized views from dataset (B) with various K values while fixing ϵ to 0.5 (Fig. 12). Even with $K = 0.2$, the pairwise term successfully reduces discontinuities. With $K = 1.6$, since too much penalty is imposed to different labels, the proposed method

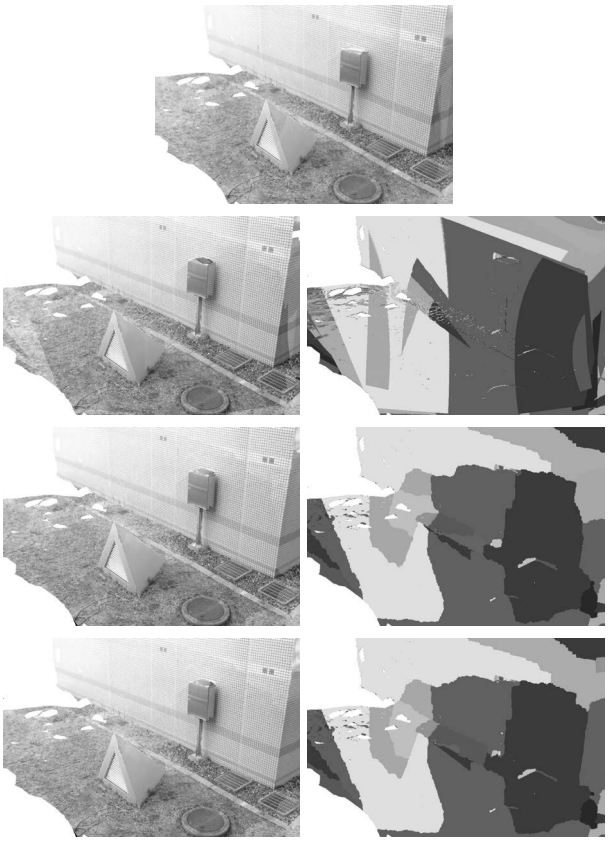

Fig. 10 Synthesized novel views for dataset (C) (left), and label maps representing x_i (right). From top to bottom: blending, $K = 0$, $(K, \epsilon) = (0.4, 1)$, $(K, \epsilon) = (0.4, 0.5)$.

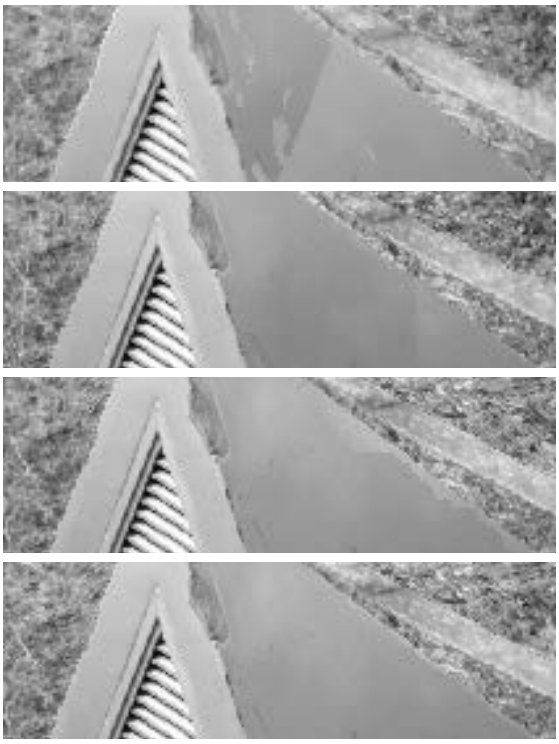

Fig. 11 Closeups of novel views in Fig. 10 (contrast is adjusted for visibility).

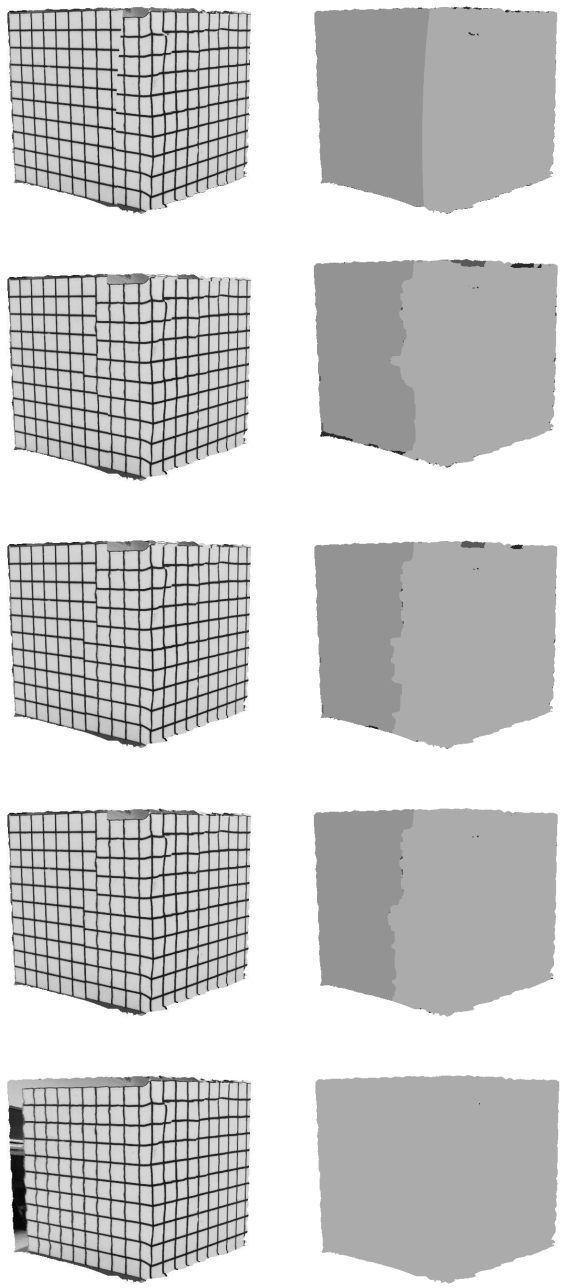

Fig. 12 Novel views (left) and their labels (right) for dataset (B) with various K values. From top to bottom: $K = 0$, 0.2, 0.4, 0.8, 1.6

assigned a single label to most pixels. From this result, the proposed method is not very sensitive to the choice of K.

Figure 13 compares the novel views before and after luminance modification. For dataset (A), we can see color discontinuity below the penguin's wing before luminance modification. Dataset (C) has a clear boundary around the center of the view (Fig. 13, top-left). It seems that, for these regions, both datasets

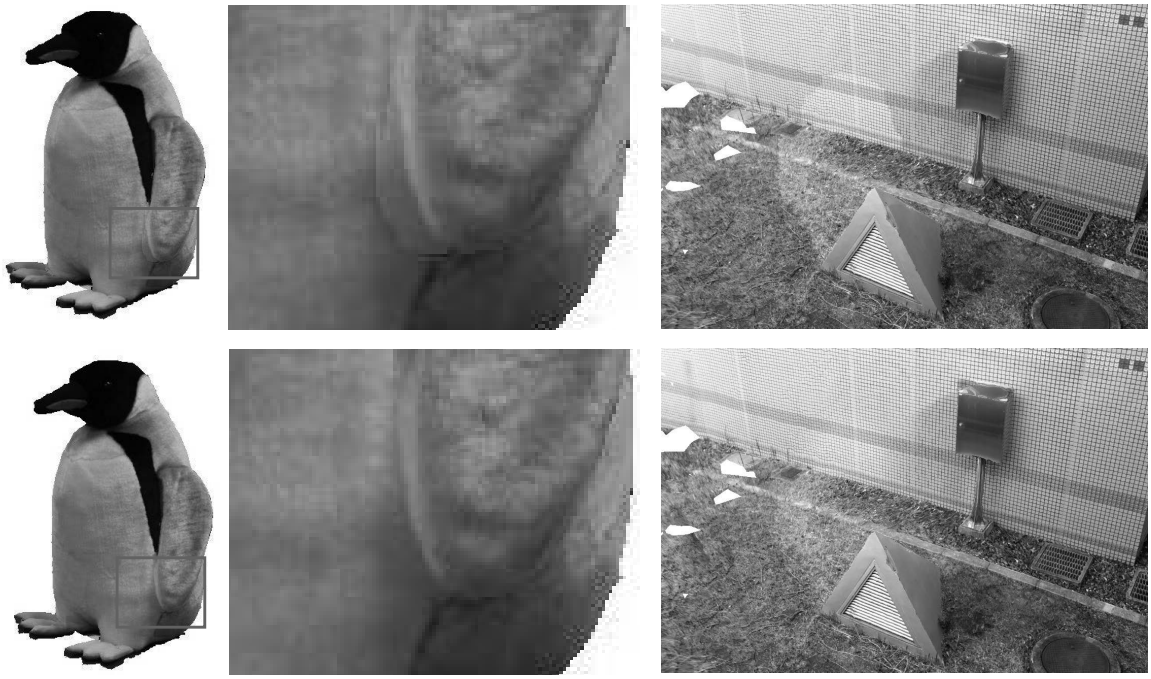

Fig. 13 Novel views (top) before and (bottom) after luminance modification for dataset (A) (left) and (C) (right). Best viewed in color.

Fig. 14 Original image (left) removed from dataset (A) and novel view (right) synthesized using the image's camera parameters.

have no better images than those used in Fig. 13, and thus our texture selection was not able to find smoother label transition than those in Fig. 13. After luminance modification, these discontinuities were successfully reduced.

To demonstrate our novel view's faithfulness and its quality when the input images are distant from the specified viewpoint, we removed one input image from the dataset and used the image's camera parameters to specify those of the novel view. Figure 14 shows the result. Since the original image, which is the most reasonable choice as the texture for this novel view, was not used, we can observe some artifacts. However, we consider that the proposed method reproduced the original image with convincing quality.

5 Conclusion

This paper has presented a new method for NVS, which leverages the geometry of the object to find suitable images for coloring pixels in novel views and alleviate visual artifacts due to texture changes. We have also introduced smooth luminance modification to further reduce discontinuities in texture. Experimental results have demonstrated the superiority of our method. Its main limitation is sensitivity to the geometry: If the geometrical change is even slightly above the threshold, the pairwise term is completely spoiled, which may cause extra discontinuities. We need a weighting scheme that deal with this problem.

Acknowledgments

This work was supported in part by JSPS KAKENHI Nos. 23240024 and 25540086.

References

[1] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo. Free-viewpoint TV. *IEEE Signal Processing Magazine*, 28(1):67–76, 2011.

[2] C. Wu. VisualSFM: A visual structure from motion system, http://ccwu.me/vsfm/ (last accessed: May 1, 2015). 2011.

[3] C. Wu, S. Agarwal, B. Curless, and S. M. Seitz. Multicore bundle adjustment. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 3057–3064, 2011.

[4] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 3121–3128, 2011.
[5] P. E. Debevec, Y. Yu, and G. Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. In *Proc. Eurographics Rendering Workshop*, pages 105–116, 1998.
[6] G. Chaurasia, S. Duchne, O. S. Hornung, and G. Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Trans. Graphics*, 32(3):30:1–30:12, 2013.
[7] J. Kopf, M. F. Cohen, and R. Szeliski. First-person hyper-lapse videos. *ACM Trans. Graphics*, 33(4):78:1–78:10, 2014.
[8] S. M. Seitz and C. R. Dyer. View morphing. In *Proc. ACM SIGGRAPH*, pages 21–30, 1996.
[9] J. Xiao and M. Shah. From images to video: view morphing of three images. In *Proc. Vision, Modeling and Visualization*, pages 495–502, 2003.
[10] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.
[11] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2009.
[12] Q.-Y. Zhou and V. Koltun. Color map optimization for 3D reconstruction with consumer depth cameras. *ACM Trans. Graphics*, 33(4):155:1–155:10, 2014.
[13] S. E. Chen and L. Williams. View interpolation for image synthesis. In *Proc. ACM SIGGRAPH*, pages 279–288, 1993.
[14] L. McMillan and G. Bishop. Plenoptic modeling: an image-based rendering system. In *Proc. ACM SIGGRAPH*, pages 39–46, 1995.
[15] T. Naemura, T. Takano, M. Kaneko, and H. Harashima. Ray-based creation of photo-realistic virtual world. In *Proc. Int. Conf. Virtual Systems and MultiMedia*, pages 59–68, 1997.
[16] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. ACM SIGGRAPH*, pages 31–42, 1996.
[17] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proc. ACM SIGGRAPH*, pages 43–54, 1996.
[18] A. Davis, M. Levoy, and F. Durand. Unstructured light fields. *Computer Graphics Forum*, 31(2):305–314, 2012.
[19] S. Nobuhara, W. Ning, and T. Matsuyama. A real-time view-dependent shape optimization for high quality free-viewpoint rendering of 3D video. In *Proc. Int. Conf. 3D Vision*, pages 665–672, 2014.
[20] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

(2015 年 7 月 15 日受付)

[Authors' Introduction]

Keita Katagiri

Keita Katagiri received his B.E. from Kyushu Institute of Technology in 2013 and the M.E. from Nara Institute of Science and Technology in 2015. Currently he is with Ricoh Company Ltd.

Yuta Nakashima

Yuta Nakashima received the B.E., M.E., Ph.D. degrees from Osaka University, in 2006, 2008, and 2012, respectively. He is currently an assistant professor at Nara Institute of Science and Technology and is also a Visiting Scholar at Carnegie Mellon University. He was a Visiting Scholar at the University of North Carolina at Charlotte in 2012.

Tomokazu Sato （正会員）

Tomokazu Sato received his B.E. degree in computer and system science from Osaka Prefecture University in 1999. He received his M.E. and Ph.D. degrees in information science from Nara Institute of Science and Technology in 2001 and 2003, respectively. He was appointed as an assistant professor at Nara Institute of Science and Technology in 2003 and became an associate professor in 2011.

Naokazu Yokoya （正会員）

Naokazu Yokoya received his B.E., M.E., and Ph.D. degrees in information and computer sciences from Osaka University in 1974, 1976, and 1979, respectively. He joined Electrotechnical Laboratory (ETL) in 1979. He was a visiting professor at McGill University in 1986-87 and has been a professor at Nara Institute of Science and Technology since 1992. He has also been a vice president at Nara Institute of Science and Technology since April 2013.