Research Article

Panorama completion for street views

Zhe Zhu¹, Ralph R. Martin², and Shi-Min Hu¹ (🖂)

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract This paper considers panorama images used for street views. Their viewing angle of 360° causes pixels at the top and bottom to appear stretched and warped. Although current image completion algorithms work well, they cannot be directly used in the presence of such distortions found in panoramas of street views. We thus propose a novel approach to complete such 360° panoramas using optimization-based projection to deal with distortions. Experimental results show that our approach is efficient and provides an improvement over standard image completion algorithms.

Keywords image completion; panorama; street views; structure-rectifying warp

1 Introduction

A panorama is an image with a wide angle of view. Panoramas are widely used for such purposes as landscape photos, group photos, and street view photos. In street view photos, panoramas often have a view angle of 360°. When presented as a rectangular image, such panoramas have large distortion, and as a result panorama editing requires different editing techniques different from traditional image editing. There are many tools to create panoramas from regular images. AutoStitch [1, 2] is a popular and robust panorama composition tool that stitches together a set of automatically aligned images. Summa et al. [3] proposed a method

Manuscript received: 2014-10-31; accepted: 2015-02-16

called panorama weaving, a fast technique for seam creation and editing when stitching images to make a panorama. However, when a panorama is stitched, artifacts like boundary mismatch may occur necessitating further post processing. Gao and Brown [4] devised an interactive editing tool for panorama correction based on a local warping strategy; it can deal with registration errors. In street view applications such as Google Street View, Tencent Street View, etc., completion is a basic requirement since pedestrian removal, sensitive text elimination, and many similar issues lead to frequent gaps which need a completion step. Holes may also occur due to missing data.

The main requirements for image completion methods are visually plausible completion results, and speed. Image completion methods generally fall into two categories. Earlier work [5–9] typically used diffusion-based methods while more recent work [10– 13 has focused on exemplar-based methods. Adobe content-aware fill [14], implemented in a commercial system, is a state-of-the-art image completion tool of the latter type providing a good balance between speed and quality; it is based on ideas from Refs. [10, 11]. In exemplar-based methods, structural information plays an important role. Sun et al. [15] required the user to mark missing lines and use them to guide completion. He and Sun [13] matched similar patches in the image, and they found that the statistics of these offsets (relative positions) are sparsely distributed. They used the dominant offsets as structure cues to guide the completing of the image. As well as seeking infill information within a single image, structures can also be considered between sloppily pasted image pieces [16]. In Ref. [16], Huang et al. extracted salience curves that approach the



¹ TNLIST, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. E-mail: ajex1988@gmail.com, shimin@tsinghua.edu. cn (☒).

² School of Computer Science & Informatics, Cardiff University, UK. E-mail: martinrr@cardiff.ac.uk.

gaps from non-tangential directions. The likely correspondences between pairs of such curves can be regarded as structural information to guide further completion. Huang et al. [17] first estimated planar projection parameters, approximated the 3D structure of an image by several softly segments planes, and discovered the translational regularity within these planes, which are important structural cues to guide low-level completion algorithms. In panoramas, structures are always distorted, as lines become bent, so straightforward image completion techniques do not work. A structure rectifying warp is needed to overcome this problem.

state-of-the-art image completion Although techniques achieve good completion results, directly using these techniques on street view panoramas usually leads to unacceptable results or even failure. Figure 1 shows a panorama; all panoramas used in this paper are actual Tencent Street Views. Due to limitations of the capture device, bottom regions of these panoramas are missing; we draw them in black. In order to allow the users a view with 360°, completion should be done first in the missing region. The middle image shows the completion result using Adobe content-aware fill while the bottom image shows our completion result. Content-aware fill produces a poor result in this case because its algorithm has a patch-matching step: the most similar patches in the source set are selected as matches to patches in the target set. Here source set is composed of patches in no-hole regions while target set is composed of patches in hole regions. In an image there always exist similar patches especially when there are linear structures (straight lines, rectangles, etc.), and we refer this as patch coherence. However, the non-linear distortion in panoramas destroys the patch coherence within a single image, so the patch matching step is unable to find suitable matches. The distortion means that data redundancy within a single image [18] becomes unreliable so that it is inappropriate to be utilized by exemplar-based image completion methods when the input is a street view panorama.

While panorama completion results were presented [13]. However, the vertical viewing-angles of their panoramas are small, and distortion is not obvious. In contrast, our street panoramas are 360° and the holes are typically located in







Fig. 1 Panorama completion result. Top: source panorama; the black region at the bottom is the hole to complete. Middle: panorama completed using Adobe content-aware fill. Bottom: panorama completed using our approach.

highly stretched and warped regions, making the completion task much more challenging.

In this paper, we propose a novel approach to complete holes in 360° street view panoramas. Our approach comprises three steps, as indicated in Fig. 2. First we perform a structure-rectifying warp around the holes. Then in the completion step we combine the ideas used in the approaches in Refs. [10] and [13]. Finally we warp the completed image back.

The remainder of the paper is organized as follows: in Section 2, we present our structure-rectifying warping method, while in Section 3 we introduce our completion scheme. Details of experiments are provided in Section 4. Section 5 concludes our work.

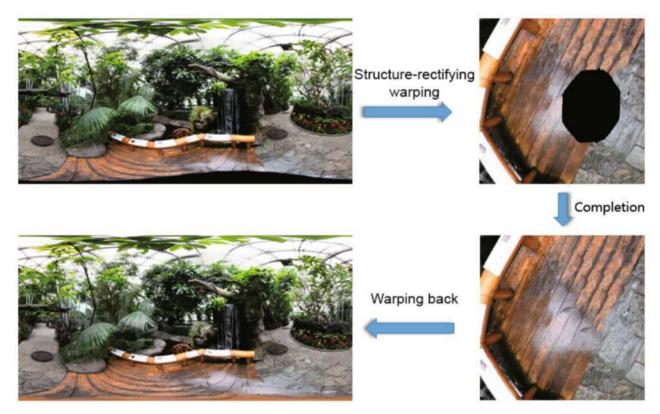


Fig. 2 Pipeline.

2 Structure-rectifying warp

Our input is a 360° panorama with known projection. The projection maps each pixel to a sphere, giving it a unique longitude and latitude. We sort all the pixels in the hole by longitude and latitude separately. We note the minimum and maximum longitude and latitude values in the hole, and extend these ranges by a proportion α , then project this extended region to the 2D plane by a warp. We call this region the target region. In our experiments we set α to 5. In practice, α depends on the area of the hole region. If the area of the hole is large, α should be set smaller to ensure that the extended region does not exceed 180°. Besides, α can be set larger to afford sufficient source regions to complete the hole. It is straightforward to use a perspective projection, but as the field of view increases, regions far away from the center of the projection become extremely stretched. No global projection from the 3D sphere to the 2D plane can simultaneously preserve straight lines and local shapes. Many different projections such as Mercator, stereographic, and Pannini projections

are available, providing different balances between the two properties. These properties are important not only because human eyes are sensitive to them, but also because structural information provides an important basis for image completion (e.g., straight lines can be used to guide image completion [15]). Inspired by Ref. [19], we use a content-aware warp that automatically rectifies structures during the projection from 3D to 2D.

We first perspectively project the candidate region to a plane. The projected image area increases dramatically as the viewing angle approaches 180°, so we limit the value allowed to less than 150° to constrain the filling problem (in our cases hole regions typically encompass less than 60° range in longitude direction). Then we use the approach of Ref. [20] to detect line segments, and filter out those that are shorter than 20 pixels. We know that straight lines in the real world lies in the great circles on the viewing sphere. When perspectively projected, they remain straight.

These line segments are projected back to the sphere. Next we adopt a mesh-based warp that projects the candidate region from the sphere to a plane. We uniformly sample a grid (every



9°) in longitude and latitude on the surface of the sphere (see Fig. 3(a)). This forms a mesh of 3D quads each related to a 2D quad in the original panorama. Optimisation is used to find the corresponding position in the 2D plane $V_{i,j}$ of each spherical vertex $U_{i,j}$ of a 3D quad; the goal is to both preserve the straight lines and local shape as well as possible. This is because both straight lines and local shapes provide important structural cues for completion. The position of each pixel is then determined by bilinear interpolation of the four quad vertexes surrounding it. Due to the size of the sampling of the mesh, only a few hundred spherical vertices fall into the candidate region, so the optimisation problem can be solved quickly. We express the goals to preserve both straight lines and local shape as energy terms in the optimisation problem, as we now explain.

2.1 Line preservation

The first requirement is to keep straight lines straight after warping. We follow the approach used in Ref. [21]. Note that we warp the image via the meshes, so that each mesh only controls the line segments fallen into it. First of all we cut each line segment where it meets a mesh edge to give smaller segments, giving N_L segments in total. Then we quantize their orientations in the image plane into 50 bins. Our line preservation term encourages line segments in the same bin to share the same rotation angle θ_m . The line preservation energy $E_L(V, \{\theta_m\})$ is defined as

$$E_L(V, \{\theta_m\}) = \frac{1}{N_L} \sum_{i=1}^{N_L} \omega_i \|s_i \mathbf{R}_i \hat{\mathbf{e}}_i - \mathbf{e}_i\| \qquad (1)$$

On the left side, V represents the mesh vertexes that fall into the candidate region; each has an x and y

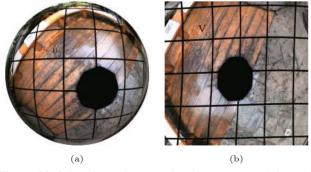


Fig. 3 Mesh-based warping optimizes the positions of the mesh vertices to simultaneously preserve line structure and local shape when mapping the image from the sphere to the plane. (a) Mesh on the sphere. (b) Output mesh in the plane.

position to be found. Each θ_m is also unknown and to be determined during optimization. On the right side, ω_i is a weight. A line segment belonging to a line crossing the hole is given a high weight, $\omega_i = 10$, while other line segments are given weights ranging from 1 to 5, inversely proportional to their distance from the center of the hole. \hat{e}_i is the orientation of the input line segment after perspective projection. e_i is the output orientation to be determined; it can be expressed in terms of the quad vertices by bilinear interpolation, so the right hand side is linearly dependent on V. R_i is a rotation matrix defined by

$$\boldsymbol{R}_i = \begin{bmatrix} \cos \theta_i & \sin \theta_i \\ -\sin \theta_i & \cos \theta_i \end{bmatrix}$$

 θ_i values come from $\{\theta_m\}$ and as noted above, line segments from the same bin share the same rotation. s_i is a scaling factor which can be determined directly from e_i by

$$s_i = (\widehat{\boldsymbol{e}}_i^{\mathrm{T}} \widehat{\boldsymbol{e}}_i)^{-1} \widehat{\boldsymbol{e}}_i^{\mathrm{T}} \boldsymbol{R}_i^{\mathrm{T}} \boldsymbol{e}_i$$

2.2 Shape preservation

To preserve local shape after warping, we encourage each quad to undergo a similarity transformation using the approach given in Ref. [22]. The shape preservation energy $E_S(V)$ is defined as

$$E_S(V) = \frac{1}{N_S} \sum_{i=1}^{N_S} \| (\boldsymbol{A}_i (\boldsymbol{A}_i^{\mathrm{T}} \boldsymbol{A}_i)^{-1} \boldsymbol{A}_i^{\mathrm{T}} - \boldsymbol{I}) \boldsymbol{V}_i \|^2$$
(2)

where N_S is the number of quads of the mesh, V is defined as before, and I is the identity matrix. A_i is defined as

$$\mathbf{A}_{i} = \begin{bmatrix} \hat{x}_{0} & -\hat{y}_{0} & 1 & 0\\ \hat{y}_{0} & \hat{x}_{0} & 0 & 1\\ \vdots & \vdots & \vdots & \vdots\\ \hat{x}_{3} & -\hat{y}_{3} & 1 & 0\\ \hat{y}_{3} & \hat{x}_{3} & 0 & 1 \end{bmatrix}$$
(3)

Here we denote the initial positions of the quad vertices by $(\hat{x}_i, \hat{y}_i)(i = 0, 1, 2, 3)$. As all these terms are known, A_i can be pre-computed. V_i represents the position of each quad vertex, and can be linearly determined from V. It is expressed as

$$\mathbf{V}_i = [x_0, y_0, x_1, y_1, \cdots, x_3, y_3]^{\mathrm{T}}$$
 (4)

2.3 Total energy and optimization

The total energy $E(V, \{\theta_m\})$ is now given by

$$E(V, \{\theta_m\}) = E_L + \lambda_S E_S \tag{5}$$

where λ_S controls the importance of the shape term. In our experiments we set λ_S to 10. Through our experiment we found that smaller λ leads straight lines better preserved but regions far from image center largely stretched. Though larger values of λ can better preserve local shapes, the straight lines can not be preserved well. We illustrate three different chooses of λ and their warping results in Fig. 4. Both V and θ_m are unknown. Minimizing the total energy to find both simultaneously is difficult, so we adopt an iterative approach which alternates two steps. The initialization is given by the local warping result (i.e., simply a regular grid). First we fix θ_m and solve for V. This is a quadratic optimization problem whose solutions can be found by solving a linear equation. As there are only a few hundred vertices in our system, this step is very fast. Then we fix V and update θ_m using Newton's method. Usually fewer than 10 iterations suffice. An illustration of a warped image produced by this process is shown in Fig. 3.

3 Completion

Having now mapped the hole to a 2D image in such a way as to give it as natural appearance as possible, we may now complete it with a standard approach. He and Sun's method [13] is a state-of-the-art image completion algorithm which performs well on images with strong structural information, as it considers related structural information in other regions of the image. We combine it with Wexler et al.'s method [10]. He et al.'s method [23] uses Walsh-Hadamard transforms and k-d trees to compute approximate nearest-neighbor fields, which is about 10 times faster and more accurate than PatchMatch [11]. Wexler et al.'s method is used to fill the hole region. We combine them in such a way because He et al.'s method works well on

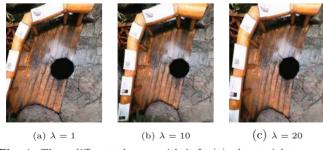


Fig. 4 Three different chooses with λ . In (a), the straight gaps in the ground are better preserved, but the main problem is that objects far from the image center are stretched. In (b), it is a good balance between local shape preservation and straight line preservation. In (c), straight gaps in the ground are curved.

structural images but always fails in natural images (images of grasses, trees, etc.), while Wexler et al.'s method can work on larger range of images that it has been embedded in Adobe content-aware fill. Our algorithm has the advantage of the speed of He et al.'s method and can generate visually pleasing completion results. We denote the image excluding the hole regions as source image and the image with holes completed as target image. In detail, we divide the source image into patches S_i , while the patches in the target image are T_i , $i = 1, 2, \dots, k$, where k is the number of patches in the image. We use patches which are 7×7 pixels; neighbouring patches overlap by 6 pixels.

Our strategy uses a hierarchical method, illustrated in Fig. 5. We create successively downsampled versions of the source image. We first fill the hole at the coarsest level, using hole pixels determined by simply interpolating the hole boundary. We then propagate the result to the next finer level as a basis for finding a suitable patch for filling the hole at that level. Let $N(S_i)$ be the approximate nearest-neighbor patch of patch S_i . The completion algorithm works follows:

- 1) Use the Walsh–Hadamard transform of the image in YCrCb format to build a k-d tree as in Ref. [24].
- 2) Build an image pyramid I_i , $i = 1, 2, \dots, l$, with l levels.
- 3) At the coarsest level I_l , interpolate the hole values from boundary pixel values.
- 4) Repeat the following for each finer level until the original resolution is reached.
 - a) For each patch T_i in the target image, find the approximate nearest-neighbor patch $N(T_i)$ in the source image.

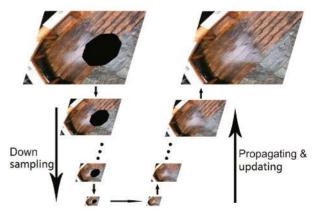


Fig. 5 Hierarchical hole filling.



- b) Update the pixel values in the hole of the target image using a voting scheme [25].
- c) Propagate the results to the next finer level.

After completing the hole in the warped image, we warp it back to the original panorama. As we have corresponding meshes for the original panorama and the warped image, this is straightforward.

4 Experiments

We have performed various experiments on Tencent Street View panoramas. These are 360° panoramas which include indoor and outdoor scenes. Due to limitations of the capture device, part of the ground is missing in the image, and must be completed. Other regions are less challenging and can often be completed by content-aware fill directly. The resolution of the original panoramas is 8192×4096 . We ran our algorithm on a PC with an Intel Core Q9400 2.66 GHz CPU and 8 GB RAM. The whole process take a few seconds to fill a hole.

If the area of the hole is not too large, our algorithm can generate good completion resultssee Fig. 6. The first row of Fig. 6 is a indoor scene. Due to the occlusion the ground in the middle of the panorama is darker than the ground in other parts. Though the completion result does not have a good transition in illuminance, the texture is visual pleasing. In the second row, the ground of the bridge has shadows due to the trees, and the completion result contains that kind of shadows; in the third row, the steps of the great wall are well completed; in the fourth row, the road is seamlessly filled; in the last row, the traffic lines and the shadow of the trees are bent in correct direction. Various cases where completion is less satisfactory are shown in Fig. 7 which are caused by illumination variance or the large size of the holes. No automatic tools can yet handle such cases.

5 Conclusions

We have presented a novel approach to complete 360° panoramas of the kind use for street views, which ordinary image completion algorithms cannot handle well due to the large distortion present in such panoramas. Our approach can generate visually pleasing results, but also has some limitations. Some outdoor scenes have few obvious line structures

needed by our structure rectifying warping. Our approach also cannot handle illumination variances found in different regions. One solution may be to locally normalize the illuminance first.

Acknowledgements

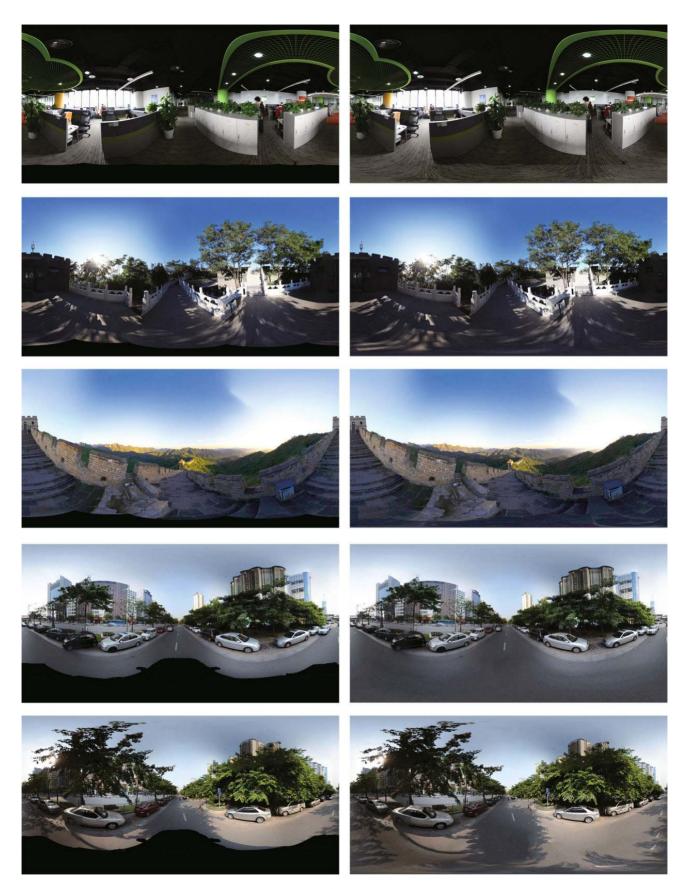
This work was supported by the National Basic Research Program of China (No. 2011CB302205), the National Natural Science Foundation of China (No. 61120106007), the National High-tech R&D Program of China (No. 2012AA011802), and EPSRC Travel Grant, a research grant of Beijing Higher Institution Engineering Research Center, and Tsinghua University Initiative Scientific Research Program.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Brown, M.; Lowe, D. G. Recognising panoramas. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, Vol. 2, 1218–1225, 2003.
- [2] Brown, M.; Lowe, D. G. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision* Vol. 74, No. 1, 59–73, 2007.
- [3] Summa, B.; Tierny, J.; Pascucci, V. Panorama weaving: Fast and flexible seam processing. ACM Transactions on Graphics Vol. 31, No. 4, Article No. 83, 2012.
- [4] Gao, J.; Brown, M. S. An interactive editing tool for correcting panoramas. In: SIGGRAPH Asia 2012 Technical Briefs, Article No. 31, 2012.
- [5] Ballester, C.; Bertalmio, M.; Caselles, V.; Sapiro, G.; Verdera, J. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Transactions on Image Processing* Vol. 10, No. 8, 1200–1211, 2001.
- [6] Bertalmio, M.; Sapiro, G.; Caselles, V.; Ballester, C. Image inpainting. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 417–424, 2000.
- [7] Bertalmio, M.; Vese, L.; Sapiro, G.; Osher, S. Simultaneous structure and texture image inpainting. In: Proceedings 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, II-707-12, 2003.
- [8] Levin, A.; Zomet, A.; Weiss, Y. Learning how to inpaint from global image statistics. In: Proceedings





 ${\bf Fig.~6} \quad {\rm Successful~completion~results}.$









Fig. 7 Unsatisfactory completion results. Top: the background is too complicated and patches can not be well matched. Bottom: illuminance changes are not taken into consideration.

- of the Ninth IEEE International Conference on Computer Vision, Vol. 2, 305, 2003.
- [9] Roth, S.; Black, M. J. Fields of experts: A framework for learning image priors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, 860–867, 2005.
- [10] Wexler, Y.; Shechtman, E.; Irani, M. Space-time completion of video. *IEEE Transactions on Pattern* Analysis and Machine Intelligence Vol. 29, No. 3, 463– 476, 2007.
- [11] Barnes, C.; Shechtman, E.; Finkelstein, A.; Goldman, D. Patchmatch: A randomized correspondence algorithm for structural image editing. ACM

- Transactions on Graphics Vol. 28, No. 3, Article No. 24, 2009.
- [12] Darabi, S.; Shechtman, E.; Barnes, C.; Goldman, D. B.; Sen, P. Image melding: Combining inconsistent images using patch-based synthesis. ACM Transactions on Graphics Vol. 31, No. 4, Article No. 82, 2012.
- [13] He, K.; Sun, J. Image completion approaches using the statistics of similar patches. *IEEE Transactions* on Pattern Analysis and Machine Intelligence Vol. 36, No. 12, 2423–2435, 2014.
- [14] Information on http://www.adobe.com/technology/projects.html.
- [15] Sun, J.; Yuan, L.; Jia, J.; Shum, H.-Y. Image completion with structure propagation. ACM Transactions on Graphics Vol. 24, No. 3, 861–868, 2005.
- [16] Huang, H.; Yin, K.; Gong, M.; Lischinski, D.; Cohen-Or, D.; Ascher, U.; Chen, B. "Mind the gap": Tele-registration for structure-driven image completion. ACM Transactions on Graphics Vol. 32, No. 6, Article No. 174, 2013.
- [17] Huang, J.-B.; Kang, S. B.; Ahuja, N.; Kopf, J. Image completion using planar structure guidance. ACM Transactions on Graphics Vol. 33, No. 4, Article No. 129, 2014.
- [18] Zontak, M.; Irani, M. Internal statistics of a single natural image. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, 977–984, 2011.
- [19] Carroll, R.; Agrawal, M.; Agarwala, A. Optimizing content-preserving projections for wide-angle images. ACM Transactions on Graphics Vol. 28, No. 3, Article No. 43, 2009.
- [20] Von Gioi, R. G.; Jakubowicz, J.; Morel, J.-M.; Randall, G. LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern* Analysis and Machine Intelligence Vol. 32, No. 4, 722– 732, 2010.
- [21] Chang, C.-H.; Chuang, Y.-Y. A line-structurepreserving approach to image resizing. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 1075–1082, 2012.
- [22] Zhang, G.-X.; Cheng, M.-M.; Hu, S.-M.; Martin, R. R. A shape-preserving approach to image resizing. *Computer Graphics Forum* Vol. 28, No. 7, 1897–1906, 2009.
- [23] Hel-Or, Y.; Hel-Or, H. Real-time pattern matching using projection kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 27, No. 9, 1430–1445, 2005.
- [24] Sun, J. Computing nearest-neighbor fields via propagation-assisted KD-trees. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, 111–118, 2012.
- [25] Simakov, D.; Caspi, Y.; Shechtman, E.; Irani, M. Summarizing visual data using bidirectional similarity. In: IEEE Conference on Computer Vision and Pattern

Recognition 2008. Available at http://www.wisdom.weizmann.ac.il/~vision/VisualSummary/bidirectional_similarity_CVPR2008.pdf.



Zhe Zhu is a Ph.D. candidate in the Department of Computer Science and Technology, Tsinghua University. He received his bachelor's degree in Wuhan University in 2011. His research interests include computer vision and computer graphics.



Ralph R. Martin is currently a professor in Cardiff University. He obtained his Ph.D. degree in 1983 from Cambridge University. He has published more than 250 papers and 14 books, covering such topics as solid and surface modeling, intelligent sketch input, geometric reasoning, reverse

engineering, and various aspects of computer graphics. He is a Fellow of the Learned Society of Wales, the

Institute of Mathematics and its Applications, and the British Computer Society. He is on the editorial boards of Computer-Aided Design, Computer Aided Geometric Design, Geometric Models, the International Journal of Shape Modeling, CAD and Applications, and the International Journal of CAD/CAM. He was recently awarded a Friendship Award, China's highest honor for foreigners.



Shi-Min Hu received the Ph.D. degree from Zhejiang University in 1996. He is currently a professor in the Department of Computer Science and Technology, Tsinghua University, Beijing. His research interests include digital geometry processing, video processing, rendering, computer animation, and

computer-aided geometric design. He is on the editorial boards of several international journals, including *IEEE Transactions on Visualization and Computer Graphics*, The Visual Computer, Computer & Graphics, and Computer-Aided Design.