

This ECCV 2018 workshop paper, provided here by the Computer Vision Foundation, is the author-created version. The content of this paper is identical to the content of the officially published ECCV 2018 LNCS version of the paper as available on SpringerLink: https://link.springer.com/conference/eccv

GreenWarps: A Two-Stage Warping Model for Stitching Images using Diffeomorphic Meshes and Green Coordinates

Geethu Miriam Jacob¹[0000-1111-2222-3333]</sup> and Sukhendu Das¹[0000-0002-2823-9211]</sup>

Visualization and Perception Lab, Department of Computer Science and Engineering, Indian Institute of Technology, Madras, India geethumj@cse.iitm.ac.in, sdas@iitm.ac.in

Abstract. Image Stitching is a hard task to solve in the presence of large parallax in the images. Specifically, for a sequence of frames from unconstrained videos which are considerably shaky, recent works fail to align such a sequence of images accurately. The proposed method "GreenWarps" aims to accurately align frames/images with large parallax. The method consists of two novel stages, namely, Prewarping and Diffeomorphic Mesh warping. The first stage warps unaligned image to the reference image using Green Coordinates. The second stage of the model refines the alignment by using a demon-based diffeomorphic warping method for mesh deformation termed "DiffeoMeshes". The warping is performed using Green Coordinates in both the stages without the assumption of any motion model. The combination of the two stages provide accurate alignment of the images. Experiments were performed on two standard image stitching datasets and one dataset consisting of images created from unconstrained videos. The results show superior performance of our method compared to the state-of-the-art methods.

Keywords: Green Coordinates · Diffeomorphic Registration · Content Preserving Warps · Image Stitching

1 Introduction

Image Stitching is a widely studied problem in the field of computer vision and graphics, which generates a single wide field-of-view image from a set of narrow field-of-view images. Several warping models, including homography-based warps [1,2] spatially varying warping models [3–5], hybrid models [6–9], parallax tolerant models [10–12] and image stitching softwares such as Adobe Photoshop and Autostitch, fail to perform well when non-ideal data is provided as input. The main challenges of any stitching algorithm are parallax error, occlusions, motion blur and presence of moving objects. Specifically, for stitching frames of an unconstrained video (e.g shaky/jittery videos), the state-of-the-art techniques fail to provide satisfactory results. The reason is that image stitching assume

specific underlying motion models, thus making the task highly challenging in presence of large parallax.

Common approaches to the image stitching algorithms follow the pipeline of estimating transformations between the images, aligning the images using a warping model and stitching them using seam techniques or blending techniques. We present a novel mesh-based warping model termed "GreenWarps", utilizing Green coordinates [13] and a demon-based diffeomorphic warping model [14] to align the images. "GreenWarps" warping model consists of two stages, namely, pre-warping and "DiffeoMeshes". The first stage produces a global conformal mapping between the images to be stitched. The conformal mappings induce no shear at all, thereby providing shape-preserving and distortion-free deformations. The second stage of the proposed method, termed "DiffeoMeshes", provides a mesh deformation based on semi-dense correspondences of the two images and refines the alignment obtained from the first stage. Both the stages utilize Green coordinates for warping the deformed meshes, instead of warping the images based on computed transformations as in previous approaches. Since our method does not assume any motion model, it is immune to large parallax error.

2 Proposed Framework

The steps of the proposed "GreenWarps" method are: (i) estimate SIFT correspondences, (ii) pre-warping based on Green coordinates, (iii) mesh deformation based on DiffeoMeshes and warping based on Green coordinates, (iv) blend the images to obtain stitched image. Similar to spatially varying warps, GreenWarps perform a shape preserving deformation of the mesh for aligning images to the reference image. Interestingly, our approach does not compute any transformation matrix during the process of alignment or warping. This ensures that our method does not assume any motion model. Warping in both the stages (Prewarping and DiffeoMeshes) is performed based on Green coordinates.

Among the images to be stitched, we take one of them as Reference image (R) and the other as Unaligned image (U). The unaligned image is first divided into image grids, where each grid has 4 vertices. The pre-warping stage takes a 2×2 mesh grid of U. Every point X_k of the unaligned image is defined in terms of the Green coordinates [13] of its corresponding mesh grid as $X_k = \phi_k (X_k)^T V_k + \psi_k (X_k)^T N_k$, where $\phi_k (X_k), \psi_k (X_k)$ are the Green coordinate vectors associated with the 4 coordinates and edges of the mesh grid containing the point X_k , V_k is a vector of 4 vertices and $N_k = [n(t_k^1) \ n(t_k^2) \ n(t_k^3) \ n(t_k^4)]$ is a vector of normals of edges t_k^i of the grid containing the point X_k . An as-similar-as-possible mesh deformation [3] is performed generating the deformed vertices \hat{V} based on the corresponding SIFT features. The Green coordinates (for every pixel in the image) are first estimated from the initial mesh as derived in [15]. The warping of the image based on the deformed vertices are performed using the computed Green coordinates. The corresponding position of any point of the unaligned image in the pre-warped image (\hat{X}_k) , with the deformed vertices \hat{V} and updated normals \hat{N} is obtained as follows: $\hat{X}_k = \phi_k (X_k)^T \hat{V}_k + \psi_k (X_k)^T m_k \hat{N}_k$. Here, m_k is

the normalized edge length [13]. Warping based on Green coordinates, as in [13] provides a conformal mapping, preserving the shape of the structures. Thus, Green coordinates provide a natural transformation of the image for alignment without assuming any motion model. Perspective distortion, a problem in many previous approaches [10, 12, 16] is absent in our approach.

The second stage of our approach, termed DiffeoMeshes, refines the alignment by obtaining a per-pixel displacement (spatial transformation) of the region of the overlap of the pre-warped and reference images. Let the overlap regions of the pre-warped and reference images be M_U and M_R respectively. A mesh deformation is performed based on the spatial transformation obtained. The demon-based diffeomorphic transformation, s is estimated using the following optimization function [14]:

$$E_{diff}(s) = Sim(M_U, M_R \circ s) + Reg(s) \tag{1}$$

The similarity (correspondence) term is $Sim(M_U, M_R \circ s) = \sum_{p=1}^{L} ||M_U(p) - M_R(p) \circ s(p)||_2^2$, and the second regularization term is defined as $\sum_{p=1}^{L} ||\nabla s(p)||_2$, where, \circ indicates the per-pixel spatial warping function and $L = |M_U| = |M_R|$, where |.| is the cardinality function. All the demon-based diffeomorphic registrations [17, 14, 18] uses Gaussian smoothing for the purpose of regularization. Our proposed method utilizes TV-based regularization [19] and this helps in preserving the edges while updating the transformation. An iterative alternating minimization of the correspondence energy and the regularization energy is performed to obtain the diffeomorphic transformation.

Let the mesh grid vertices at the second stage before and after deformation be \mathcal{V} and $\hat{\mathcal{V}}$. DiffeoMeshes minimizes the optimization function $E(\hat{\mathcal{V}}) = E_d(\hat{\mathcal{V}}) + w_s E_s(\hat{\mathcal{V}})$ where E_d is the data term and E_s is the smoothness term. The data term minimizes the distance between the measured point and the interpolated location in the mesh using diffeomorphic transformation. The data term of DiffeoMeshes is: $E_d(\hat{\mathcal{V}}) = \sum_{p=1}^{N_d} ||s(p)||_2^2$, where N_d is the number of pixels selected from the overlap region of the pre-warped and reference images and s(p) is the diffeomorphic transformation in pixel p. Only those pixels belonging



Fig. 1. Comparison of our method with (a) SPHP [8], (b) Parallax tolerant [10], (c) APAP [4], (d) SEAGULL [12]. In each example, the first column shows the input images, the second column shows the output of the comparing method and the third column shows the output of the proposed method.

G. M. Jacob and S. Das Author

4

E_{mg}					E_{corr}				
Average	[4]	[8]	[3]	Ours	Average	[4]	[8]	[3]	Ours
Dataset 1 [10]	4.28	6.44	7.37	2.31	Dataset 1 [10]	0.36	0.34	0.35	0.32
Dataset 2 [12]	4.74	6.34	6.99	3.14	Dataset 2 [12]	0.36	0.33	0.35	0.32
Dataset 3	25.49	11.91	11.47	9.03	Dataset 3	0.33	0.31	0.32	0.28

Table 1. Comparison of the performances for two standard datasets [10, 12] for large parallax and one dataset consisting of frames from unconstrained videos.

to the edges, with exact match are taken for obtaining the mesh (semi-dense correspondences). $E_s(\hat{\mathcal{V}})$ is the same as that used in [3]. The smoothness term minimizes the deviation of each deformed mesh grid from a similarity transformation of its input mesh. The solution of the problem is obtained using a Jacobi based linear solver. Once the deformed mesh vertices are obtained, the refined alignment is obtained by warping using Green coordinates similar to the first stage. The aligned images are then blended using the multi-band blending method of [20].

3 Experimental Results

Experiments were performed on two parallax-tolerant image stitching datasets [10, 12] and a new dataset consisting of 2-3 frames of unconstrained videos. Parallax error and presence of moving objects are the main challenges of the images in the dataset. Our method is evaluated with the state-of-the-art methods [3, 8, 4, 4]21]. The error measures used for determining the alignment quality of the images are mean geometric error (E_{mq}) and correlation error (E_{corr}) . E_{mq} measures the average distance between the corresponding feature points after alignment and E_{corr} is defined as the average of one minus Normalized Cross Correlation(NCC) over a neighborhood in the overlapped region. Lower values of the measure indicates better performance. Table 1 shows the average (over the whole dataset) alignment errors of all 3 datasets in comparison to the state-of-the-art methods. As seen in the table, our method outperforms the state-of-the-art methods for every dataset. Some qualitative results are also shown in figure 1. The comparison with the methods [8, 10, 4, 12] are shown in the figure. The red boxes indicate the erroneous regions of alignment, whereas the blue boxes shows the corresponding regions accurately aligned. The superiority of the method can be seen from the qualitative and quantitative results.

References

- 1. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. International Journal of Computer Vision **74**(1) (2007) 59–73
- Szeliski, R., Shum, H.Y.: Creating full view panoramic image mosaics and environment maps. In: SIGGRAPH. (1997)

- 3. Liu, F., Gleicher, M., Jin, H., Agarwala, A.: Content-preserving warps for 3d video stabilization. ACM Transactions on Graphics (TOG) **28**(3) (2009) 44
- Zaragoza, J., Chin, T.J., Brown, M.S., Suter, D.: As-projective-as-possible image stitching with moving DLT. In: CVPR. (2013)
- Lin, C.C., Pankanti, S.U., Natesan Ramamurthy, K., Aravkin, A.Y.: Adaptive as-natural-as-possible image stitching. In: CVPR. (2015)
- Yan, W., Hou, C., Lei, J., Fang, Y., Gu, Z., Ling, N.: Stereoscopic image stitching based on a hybrid warping model. IEEE Transactions on Circuits and Systems for Video Technology 27(9) (2017) 1934–1946
- Gao, J., Kim, S.J., Brown, M.S.: Constructing image panoramas using dualhomography warping. In: CVPR. (2011)
- Chang, C.H., Sato, Y., Chuang, Y.Y.: Shape-preserving half-projective warps for image stitching. In: CVPR. (2014)
- 9. Lin, K., Jiang, N., Liu, S., Cheong, L.F., Lu, M.D.J.: Direct photometric alignment by mesh deformation. In: CVPR. (2017)
- 10. Zhang, F., Liu, F.: Parallax-tolerant image stitching. In: CVPR. (2014)
- Chen, Y.S., Chuang, Y.Y.: Natural image stitching with the global similarity prior. In: ECCV. (2016)
- Lin, K., Jiang, N., Cheong, L.F., Do, M., Lu, J.: Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In: ECCV. (2016)
- Lipman, Y., Levin, D., Cohen-Or, D.: Green coordinates. ACM Transactions on Graphics (TOG) 27(3) (2008) 78
- Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1) (2009) S61–S72
- Lipman, Y., Levin, D.: Derivation and analysis of green coordinates. Computational Methods and Function Theory 10(1) (2010) 167–188
- Li, J., Wang, Z., Lai, S., Zhai, Y., Zhang, M.: Parallax-tolerant image stitching based on robust elastic warping. IEEE Transactions on Multimedia (2017)
- Thirion, J.P.: Image matching as a diffusion process: an analogy with maxwell's demons. Medical Image Analysis 2(3) (1998) 243–260
- Santos-Ribeiro, A., Nutt, D.J., McGonigle, J.: Inertial demons: A momentumbased diffeomorphic registration framework. In: MICCAI. (2016)
- Chambolle, A.: An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision 20(1) (2004) 89–97
- Burt, P.J., Adelson, E.H.: A multiresolution spline with application to image mosaics. ACM Transactions on Graphics (TOG) 2(4) (1983) 217–236
- Li, N., Xu, Y., Wang, C.: Quasi-homography warps in image stitching. arXiv preprint arXiv:1701.08006 (2017)