

1+N FUSION: CASCADED SELF-PORTRAIT ENHANCEMENT

Shuai Yang[†], Jiaying Liu[†], Sifeng Xia[†] and Zongming Guo^{†‡*}

[†]Institute of Computer Science and Technology, Peking University, Beijing, China

[‡]Cooperative Medianet Innovation Center, Shanghai, China

ABSTRACT

In this paper, we present a novel cascaded framework to solve a self-portrait enhancement problem we call “1+N” problem, in which a self-portrait is enhanced with the help of N supporting photos that share the same scene and similar shooting time. The key idea is to exploit the extra information of these N photos to expand the field of view of the self-portrait and improve its lighting style. We achieve this by alternately optimizing two complementary tasks, namely illumination unification and photo registration. Based on the correspondences extracted in the input 1+N photos, our method estimates and updates the illumination and registration coefficients in a cascaded manner. Then a Markov Random Field formulation is proposed to globally fuse the aligned photos. Experimental results demonstrate the proposed method achieves high-quality results in this novel application scenario.

Index Terms— Image enhancement, field of view expansion, illumination consistency, image registration

1. INTRODUCTION

Sharing self-portraits starts trending nowadays with the boom in social networks and the rise of smartphones. However, self-portraits have particularly limited Field Of View (FOV), making it hard to create the experience of re-visiting the wonderful scenes. Indeed, the photographer can occupy as much as one fourth of the self-portrait, resulting in poor composition balance between the foreground and background. The background scene in self-portraits can be either incomplete or too small. Fortunately the photographer can easily take some additional photos of the same scene after taking the self-portrait, and these photos (we call supporting photos) can help solve the FOV problems. Specifically, we fuse the self-portrait and supporting photos into an enhanced self-portrait.

Extrapolating image contents to expand FOV has recently attracted attentions. Data-driven image extrapolation [1–5] combines multiple images that share similar scenes to extend image boundaries. The pioneering work in [2] enables navigation in photo collections by extrapolating an original image in one direction using the gradient descent alignment. Shan *et al.* [4] employ multi-view stereo under a Markov Random Field (MRF) framework to warp and fuse images into a wide

FOV image. A graph-based image representation which balances high-level structures and low-level features is proposed in [5] to achieve boundary consistency. These methods carefully search illumination-consistent reference images for extrapolation. However, in our “1+N” problem, the self-portrait and supporting photos usually have very different lighting styles. This inconsistency is caused by different shooting parameters between the normal camera and the front-facing camera. And it is aggravated by the fact that the photographer occupying large area of the scene can greatly affect the camera focus and exposure time. In this case, the aforementioned methods may yield illumination variance artifacts.

To unify lighting styles, illumination correction methods have been studied. Single image methods [6] mainly focus on statistical-based white balance correction. For multi-photo illumination correction, modern techniques [7, 8] seek correspondences and assume a certain color model to robustly optimize the illumination consistency. In this work, we combine illumination unification with photo registration using correspondences in a cascaded framework to improve each other.

In this paper, we raise the “1+N” problem where one self-portrait is enhanced with the help of N supporting photos. To solve this problem, two issues of illumination unification and photo registration must be addressed. These two tasks are complementary - photos in good illumination consistency achieve accurate alignment while illumination unification is better guided if photos are well aligned. We seamlessly bridge these two complementary tasks using pixel correspondences and optimize them progressively in our cascaded framework. To our knowledge, this is the first work to integrate these two closely related tasks into a unified cascaded framework. Specifically, in each iteration, our method leverages SIFT correspondences to estimate and update the illumination and registration coefficients. Then the photos are deformed with their lighting styles adjusted using these coefficients, enabling more accurate correspondences extracted. Finally the aligned illumination-consistent photos are globally fused using an MRF formulation. Experimental results show that our method can obtain appealing results that provide excellent experience of re-visiting the scene.

The rest of this paper is organized as follows. Sec. 2 describes the proposed self-portrait enhancement method. Experimental results are shown in Sec. 3 and concluding remarks are given in Sec. 4.

* Corresponding author

This work was supported by National Natural Science Foundation of China under contract No. 61471009.

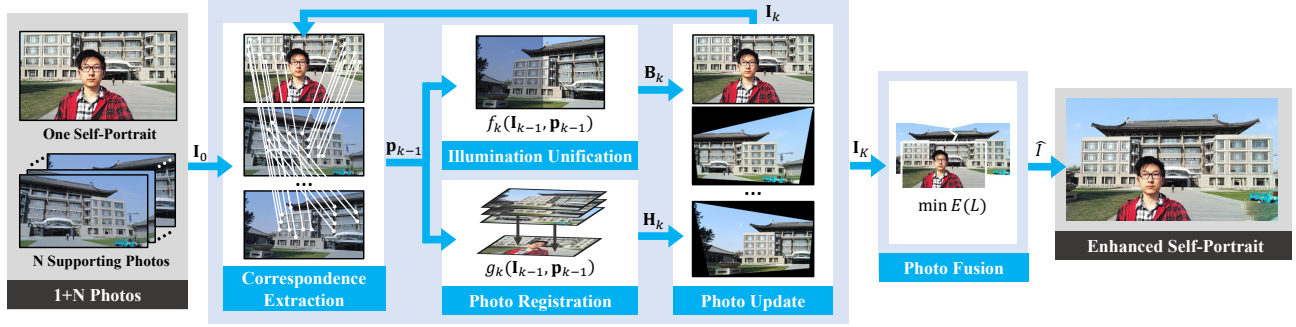


Fig. 1. Framework of the proposed cascaded self-portrait enhancement method.

2. SELF-PORTRAIT ENHANCEMENT APPROACH

Given a set of 1+N photos $\mathbf{I} = \{I_i\}_{i=0}^N$ (one self-portrait photo I_0 and N supporting photos $\{I_i\}_{i=1}^N$) that share the same scene and similar shooting time, our goal is to optimize their illumination consistency and registration to generate an enhanced self-portrait with much broad FOV. To accomplish this task, we propose a cascaded framework to alternatively refine the illumination and the registration. Let $\mathbf{B} = \{B_i\}_{i=0}^N$ and $\mathbf{H} = \{H_i\}_{i=0}^N$ denote the illumination coefficients and registration coefficients respectively. Our framework can be formulated as the following processes with K iterations,

$$\hat{\mathbf{B}}_k = f_k(\mathbf{I}_{k-1}, \mathbf{p}_{k-1}); \mathbf{B}_k = \mathbf{B}_{k-1} \otimes \hat{\mathbf{B}}_k; \quad (1)$$

$$\hat{\mathbf{H}}_k = g_k(\mathbf{I}_{k-1}, \mathbf{p}_{k-1}); \mathbf{H}_k = \mathbf{H}_{k-1} \otimes \hat{\mathbf{H}}_k; \quad (2)$$

where k iterates from 1 to K , \mathbf{p}_{k-1} are the SIFT correspondences extracted among \mathbf{I}_{k-1} . f_k represents the illumination coefficient estimation while g_k stands for the registration coefficient estimation. The operation \otimes updates the coefficients with the newly estimated ones. At the end of each iteration, we update the photos $\mathbf{I}_k = \mathbf{H}_k(\mathbf{B}_k(\mathbf{I}))$ and recount the correspondences \mathbf{p}_k . The iteration starts from $\mathbf{I}_0 = \mathbf{I}$ and ends with the aligned illumination-consistent output \mathbf{I}_K . Finally, photos in \mathbf{I}_K are fused into a wide FOV self-portrait \hat{I} . Fig. 1 shows the framework of our approach.

We now introduce the illumination coefficient estimation (f_k) and registration coefficient estimation (g_k) procedures. For simplicity, we will omit the iteration subscript k in the following unless otherwise specified.

2.1. Correspondence Extraction

For input photos $\mathbf{I} = \{I_i\}_{i=0}^N$, we first extract their Scale-Invariant Feature Transform (SIFT [9]) features. Then the nearest neighbor (NN) descriptor matching is performed on image pairs to generate the sparse SIFT correspondences $\mathbf{p} = \{p_j\}_{j=1}^M$ with their position $\{x_{ij}\}$, where M is the correspondence number and x_{ij} is the pixel corresponding to p_j in I_i .

2.2. Illumination Unification

This section presents the illumination coefficient estimation (f_k) to adjust the photos and make their lighting styles con-

sistent with one chosen visually satisfying photo I_t in \mathbf{I} . The matrix factorization based approach [8] is used, in which correspondences \mathbf{p} are stacked into a matrix based on a simple illumination correction model and a low-rank matrix factorization is optimized to estimate the illumination coefficients.

The illumination correction model is as follows:

$$I = (cI')^\gamma, \quad (3)$$

where I is the input photo, I' is the desired photo, c is the white balance coefficient and γ is the gamma mapping coefficient. The illumination coefficients of I_i are therefore defined as $B_i = [c_i, \gamma_i]$. And its illumination is adjusted by

$$B_i(I_i) = I_i^{1/\gamma_i} / c_i. \quad (4)$$

Based on (3), given a correspondence p_j , its pixel value at x_{ij} in I_i can be written as

$$I_i(x_{ij}) = (c_i a_j e_{ij})^{\gamma_i}, \quad (5)$$

where a_j is the constant albedo of p_j and e_{ij} represents the pixel error of x_{ij} that cannot be modeled with (3). By taking logarithms on both sides of (5) and rewriting it in matrix form, we obtain the illumination equation,

$$\mathbf{O} = \mathbf{C} + \mathbf{A} + \mathbf{E}, \quad (6)$$

where \mathbf{O} is the observation matrix derived from the correspondences \mathbf{p} . \mathbf{C} , \mathbf{A} and \mathbf{E} are the illumination coefficient matrix, the albedo matrix and the residual matrix respectively. It can be proved that the rank of $\mathbf{C} + \mathbf{A}$ is equal to 2 ([8]). Therefore, a low-rank matrix factorization optimization can be constructed to solve c and γ . Specifically, two augmented matrices are defined: $\mathbf{P} = [\mathbf{c} \odot \mathbf{g}, \mathbf{g}] \in \mathbb{R}^{(N+1) \times 2}$ with $c_i = \log c_i$, $g_i = \gamma_i$, \odot denotes element-wise multiplication operator, and $\mathbf{Q} = [\mathbf{1}, \mathbf{a}] \in \mathbb{R}^{M \times 2}$ with $\mathbf{a}_j = \log a_j$. Then \mathbf{P} and \mathbf{Q} satisfy $\mathbf{P}\mathbf{Q}^\top = \mathbf{C} + \mathbf{A}$ and can be solved by

$$\mathbf{P}^*, \mathbf{Q}^* = \arg \min_{\mathbf{P}, \mathbf{Q}} \|\mathbf{W} \odot (\mathbf{O} - \mathbf{P}\mathbf{Q}^\top)\|_1 + \lambda_1 (\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2) + \lambda_2 (\|\mathbf{Q} - \mathbf{Q}'\|_F^2), \quad (7)$$

where \mathbf{W} is the mask matrix indicating the missing entries, $\mathbf{Q}' = [\mathbf{1}, \mathbf{a}']$ is an approximate solution of surface albedo with \mathbf{a}'_j as the median pixel value of p_j to regularize \mathbf{Q} . We refer to [8] for details of formula derivation and optimization.

After obtaining c_i and γ_i , we can solve the constant albedo of the scene in I_i as $B_i(I_i)$. Sometimes, the lighting style of a certain photo $I_t \in \mathbf{I}$ is more preferable than the albedo image and the user may want to appoint I_t as the reference for other photos to adjust their illumination. In this case, c_t and γ_t are applied to the albedo image following (3): $B_t^{-1}(B_i(I_i)) = I_i^{1/(\gamma_i/\gamma_t)} / (c_i/c_t)^{\gamma_t}$. This illumination transfer operation can be simply achieved by modifying the illumination coefficients as $c_i = (c_i/c_t)^{\gamma_t}$ and $\gamma_i = \gamma_i/\gamma_t$. An example of the illumination transfer is shown in Fig. 2(b).

In the iteration phase, $c = 1$ and $\gamma = 1$ for \mathbf{B}_0 . Supposing the newly estimated illumination coefficients for I_i are $\hat{B}_i = [\hat{c}_i, \hat{\gamma}_i] \in \hat{\mathbf{B}}_k$, then the illumination correction result is $\hat{B}_i(B_i(I_i)) = I_i^{1/(\gamma_i \hat{\gamma}_i)} / (\hat{c}_i c_i^{1/\hat{\gamma}_i})$. Thus \mathbf{B}_k is updated as

$$\mathbf{B}_{k-1} \otimes \hat{\mathbf{B}}_k := \{[\hat{c}_i c_i^{1/\hat{\gamma}_i}, \gamma_i \hat{\gamma}_i]_{i=0}^N\}. \quad (8)$$

2.3. Homography-Based Photo Registration

We apply a simple yet effective strategy to align each supporting photo I_i to the self-portrait I_0 leveraging homography chains. We take the homography transformation model for registration and define the registration coefficient H_i as a 3×3 homography matrix which maps each pixel in I_i to its corresponding pixel in I_0 (e.g. $H_i(x_{ij}) = x_{0j}$). For registration coefficient estimation (g_k), we have found in practice that directly computing SIFT feature-based homographies between some I_i and I_0 is unreliable due to their minor overlaps or great deformations. Thus we propose to use minimum spanning tree along homography chains to bridge I_i and I_0 .

We first perform a RANSAC-based voting algorithm over the correspondences \mathbf{p} between each I_i/I_j pair to find the best fitting transformation matrix $H_{i,j}$ that maps I_i to I_j and compute the number of its inlier matches $m_{i,j}$. Then we construct a fully connected undirected match graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} represents 1+N photos, and \mathcal{E} denotes the confidence for our homography estimation with $\mathcal{E}_{i,j} = 1/m_{i,j}$. The minimum spanning tree is next calculated. For each I_i , it reaches I_0 through a path (for instance, $I_i - I_{i_1} - I_{i_2} - \dots - I_{i_{n_i}} - I_0$) along the tree. According to the chain rule, the registration coefficient is estimated as: $\hat{H}_i = H_{i_{n_i}, 0} \dots \cdot H_{i_1, i_2} \cdot H_{i, i_1} \in \hat{\mathbf{H}}_k$.

In the iteration phase, initially $H_i \in \mathbf{H}_0$ is set as the unit matrix. Once $\hat{\mathbf{H}}_k = \{\hat{H}_i\}_{i=0}^N$ is estimated, \mathbf{H}_k is updated as

$$\mathbf{H}_{k-1} \otimes \hat{\mathbf{H}}_k := \{H_i \cdot \hat{H}_i\}_{i=0}^N. \quad (9)$$

After the transformation, all the supporting photos are aligned to I_0 , as shown in Fig. 2(c). We further provide the alignment results at different iteration stages in Fig. 2(d)(e) to verify the improvement brought by our cascaded strategy.

2.4. MRF-Based Photo Fusion

In this section, we propose to fuse aligned illumination-consistent photos in \mathbf{I}_K into a single self-portrait \hat{I} using MRF optimization. Specifically, the aligned photos overlap



(a) Input photo collection \mathbf{I} . From left to right: I_0, I_1, I_2



(b) Illumination unification result $\mathbf{B}_1(\mathbf{I})$ with $I_t = I_0$



(c) Registration result $\mathbf{I}_1 = \mathbf{H}_1(\mathbf{B}_1(\mathbf{I}))$



(d) Mean image of \mathbf{I}_1

(e) Mean image of \mathbf{I}_3

Fig. 2. Illumination unification and photo registration example: (a) Input photo collection. (b) Our illumination unification result. (c) Our registration result. (d)(e) The mean images generated from the output at the first and the third iterations. See the improvement in the red cropped zoomed regions.

and our fusion problem is to decide which photo to use in the overlapped regions. We formulate it as a labelling problem, in which each label represents a photo and each pixel in \hat{I} is assigned with a label. The value of pixel x in \hat{I} with label i is set to $I_i(x)$. Our goal is to find the optimal labelling scheme.

We first initialize \hat{I} as $I_0 \in \mathbf{I}_K$ so that the photographer will not be covered by the supporting photos. Let Ω denote the region where pixel values are to be determined (for example, the black region of I_0 in Fig. 2(c)) and $\delta\Omega$ be its boundary. Given N labels, we define the MRF formulation to evaluate the labelling scheme:

$$E(L) = \sum_{(x,y)} E_s(L(x), L(y)) + \lambda_3 \sum_{x \in \Omega} E_d(L(x)), \quad (10)$$

where $L(x) = i$ means assigning pixel x with label i , (x, y) are two adjacent pixels and λ_3 is a weight that compromises between two terms:

Smoothness term: $E_s(L(x), L(y))$ penalizes the discontinuity between adjacent pixels. It is defined as (for simplicity, we assume $L(x) = i, L(y) = j$)

$$E_s(i, j) = \|I_i(x) - I_j(x)\|_1 + \|I_i(y) - I_j(y)\|_1. \quad (11)$$

Data term: $E_d(L(x))$ is defined as

$$E_d(i) = \begin{cases} +\infty, & \text{if } I_i(x) \text{ has no value} \\ 0, & \text{if } I_i(x) \text{ has value } \wedge x \in \Omega \setminus \delta\Omega, \\ \|\Psi_i(x) - \hat{\Psi}(x)\|_1, & \text{other} \end{cases} \quad (12)$$



Fig. 3. Comparisons with [1] and [10]. (a) The input photo collections \mathbf{I} . The lighting style references I_t are marked with red boxes. (b) Results of Scene Collage [1]. (c) Results of Photoshop PhotoMerge [10]. (d) Our results. Best viewed in color.

where $\Psi_i(x)$ and $\hat{\Psi}(x)$ are the patches centered at pixel x in I_i and \hat{I} respectively. $\|\Psi_i(x) - \hat{\Psi}(x)\|_1$ is the patch difference measuring the boundary consistency of the self-portrait.

The energy optimization is achieved using multi-label graph-cuts algorithm. The Poisson fusion [11] is used to further hide seams and the content-aware fill [12, 13] is used to fill the unknown region if exists.

3. EXPERIMENTAL RESULTS AND ANALYSIS

We evaluated the performance of the proposed method and compared with Scene Collage [1] and Photoshop PhotoMerge tool [10] on eight photo collections taken by iPhone 6 Plus and Honor 7. Part of the experimental results are shown in this section. The whole photos and experimental results have been released on our website¹. In the experiment, we set $\lambda_1 = 2/\sqrt{\min(M, N + 1)}$, $\lambda_2 = 10\lambda_1$ and $\lambda_3 = 2$. $K = 3$ cascades can yield satisfying results.

The self-portrait enhancement results are shown in Fig. 3. Please enlarge and view these figures on the screen for better comparison. All the methods expand the FOV of the self-portrait. However, a common problem for Scene Collage [1] is the inability to handle complex photo registration and illumination variation. Thus great color discontinuities and background scene distortions can be found in the results of Scene Collage [1]. The photos are well aligned by Photoshop PhotoMerge [10]. The problem is the visible illu-

mination changes near the photographer in Photoshop’s results, which can be found in the first row and third row of Fig. 3(c). This is because the input photos suffer great inconsistency in the lighting style and Photoshop only tried to smooth the local brightness changes, failing to achieve global illumination consistency. Another problem for Photoshop is that, although different settings have been tried, the girl is always partly covered by the background scenes in the second row of Fig. 3(c). By comparison, thanks to the cascaded framework, our method achieves global illumination consistency and accurate registration simultaneously. We also gives self-portraits higher priority by initializing \hat{I} as I_0 , which effectively prevents the photographer from being obscured. In addition, by appointing one visually satisfying photo as reference for lighting style transfer, our results stand out from [1] and [10] in white balance and aesthetics, especially for the blue sky in the first row and the white statue in the third row.

4. CONCLUSION

In this study, we raise the “1+N” self-portrait enhancement problem and present a novel cascaded framework to solve it. We employ accurate correspondences to obtain illumination-consistent aligned photos and propose a MRF-based formulation to fuse them into a wide FOV self-portrait. Owing to the specific capability to progressively refine the illumination consistency and multi-photo registration in an alternating manner, we obtain visually appealing results that create excellent experience of re-visiting the scene.

¹<http://www.icst.pku.edu.cn/struct/Projects/IplusN.html>

5. REFERENCES

- [1] Y. Nomura, L. Zhang, and S. K. Nayar, "Scene collages and flexible camera arrays," in *Proc. Eurographics Conf. Rendering Techniques*, 2007, pp. 127–138, http://www1.cs.columbia.edu/CAVE/projects/scene_collage/scene_collage.php.
- [2] J. Sivic, B. Kaneva, A. Torralba, S. Avidan, and W. T. Freeman, "Creating and exploring a large photorealistic virtual space," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition Workshops*, June 2008, pp. 1–8.
- [3] Y. Zhang, J. Xiao, J. Hays, and P. Tan, "Framebreak: Dramatic image extrapolation by guided shift-maps," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2013.
- [4] Q. Shan, B. Curless, Y. Furukawa, C. Hernandez, and S. M. Seitz, "Photo uncrop," in *Proc. European Conf. Computer Vision*, September 2014, pp. 16–31.
- [5] M. Wang, Y. Lai, Y. Liang, R. R. Martin, and S. Hu, "Biggerpicture: data-driven image extrapolation using graph matching," *ACM Transactions on Graphics*, vol. 33, no. 6, pp. 1–13, 2014.
- [6] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics and scene semantics.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 687–698, 2011.
- [7] Y. Hacoheh, E. Shechtman, D. B. Goldman, and D. Lischinski, "Optimizing color consistency in photo collections," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 96–96, 2013.
- [8] J. Park, Y. W. Tai, N. S. Sinha, and I. S. Kweon, "Efficient and robust color consistency for community photo collections," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2016.
- [9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int'l Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, November 2004.
- [10] Adobe Photoshop PhotoMerge: <http://helpx.adobe.com/en/photoshop/using/create-panoramic-images-photomerge.html>.
- [11] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Transactions on Graphics*, vol. 22, pp. 313–318, July 2003.
- [12] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 463–476, March 2007.
- [13] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 341–352, August 2009.