21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Web Image-based Super-resolution

Jongho Lee, Sang Chul Ahn, Hwasup Lim, Ig-Jae Kim, Jaewon Kim, Hyoung-Gon Kim
Imaging Media Research Center
Korea Institute of Science and Technology (KIST)
{purmod, asc, hslim, kij, jaewonk, hgk}@imrc.kist.re.kr

Abstract

Super-resolution using web images is a challenging problem because accurate motion estimation is difficult due to their viewpoint change and photometric variation. To overcome such difficulties, this paper formulates a unified framework where local color correction is incorporated into fuzzy motion estimation after global registration. Thus, both geometric and photometric variations are simultaneously taken into account for implicit motion estimation.

As demonstrated in the results using web images under various conditions, the proposed approach magnifies user-selected region of interest with visually appealing details.

1. Introduction

Multi-image super-resolution (SR) is, in general, the reconstruction of a high-resolution (HR) target image from several low-resolution (LR) reference images with the assumptions of small motion and illumination change [1].

When comes to using web images by search engines where large viewpoint change and photometric variation are observed, classical SR methods with simple motion estimation might fail to find the most appropriate rather than the best matched patch such that the quality of the reconstructed SR image is even worse than that of the upsampled image with a simple interpolation method.

There have been many attempts to overcome such difficulties with geometric and photometric variations. Global transformations such as translation or affine warping are introduced to find the simple global motion between the LR images [2][3]. SR algorithms without explicit motion estimation are also proposed to provide solutions to complex motion between images [4][5]. However, they are only applicable to the images under the same illumination condition. SR algorithms considering photometric diversity of the LR images are also presented in [3][6]. However, they consider only the global photometric transformation between images, which is not common in general images.

In this paper, we propose a unified SR framework that overcomes difficulties with large viewpoint change and photometric variation in the web images. To this end, the local color correction is incorporated into fuzzy motion estimation after global registration using keypoints. Thus the weighted correspondences between pixels in LR images are accurately estimated compensating both geometric and photometric variations. The SR image is then reconstructed using the multi-image SR procedure with weighted correspondences and corrected color values.

The remaining of the paper is organized as follows. Section 2 presents the overall framework. Section 3 explains the details of the proposed algorithm. Section 4 demonstrates the experimental results. Finally, we conclude our work in Section 5.

2. Overall framework

Fig. 1 shows the overall framework of our algorithm. First, a user collects the web images of the scene of interest using image search engines such as Google or Flickr. The region of interest (ROI), namely the target image, to be magnified is assigned by the user on a certain LR image. Good reference images are then selected from the collected web image set if the images can be aligned with the target image by global transformation. Finally, high-quality result of ROI is generated through local reconstruction considering photometric diversity.

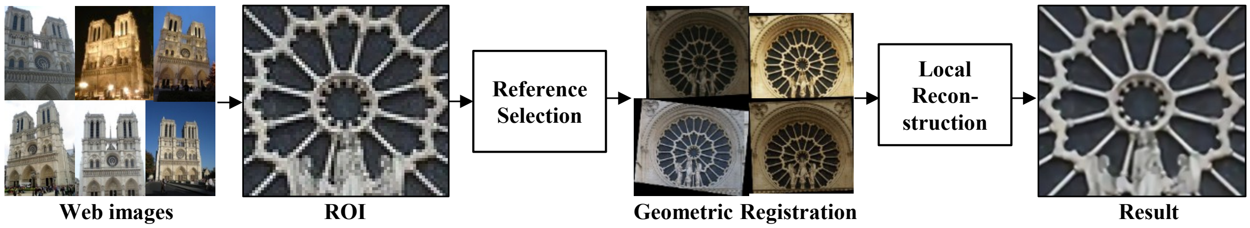

Figure. 1. Overall framework.

3. Proposed algorithm

3.1. Reference selection

Most of all, good reference images need to be selected among various web images to recover the high frequency component of the target image well. The following conditions are used in our framework.

First, the reference images include enough corresponding feature points to the target image. These correspondences not only identify the matching region, but enable robust estimation of homography between the target image and the reference image.

Second, the reference images contain enough regions matched with the target image. It is more likely that the larger matching region contains the more information for SR.

Based on the above conditions, we choose the reference images with more than T_c, the number of corresponding SIFT feature points and above R_c, the relative ratio of the areas of two bounding boxes including all the corresponding feature points. Here T_c, and R_c are experimentally set to 10 and 1 respectively.

3.2. Geometric registration

The existence of non-planar 3D objects in the image hinders the accurate alignment with only homography, as it is well known, which performs well only with two cases [8]: images of a plane viewed under arbitrary camera motion and images of an arbitrary 3D scene viewed by camera rotation about its optic center and/or small zooming.

We focus on the web images with near-planar objects viewed under arbitrary camera motion. To handle inexact registration by homography, we use local reconstruction method with many reference images after homography transformation by RANSAC from correspondences.

Let t_{n0} be the target image, and r_n (n=1,…, N) the N initially warped reference images by homography for notation.

3.3. Local reconstruction on photometrically diverse reference images

Since the web images are captured in different positions and times with different cameras, specularity and shadowing caused by the position of the light source and the shape of the objects prevent photometric registration with simple global parameters. Thus, local color correction is known to be more robust than global one for photometric registration [9].

Ironically, local corresponding regions among images should be known for successful local color correction. Finding both good correspondence and color correction is a hard "chicken and egg" problem.

To tackle this problem, we incorporated local color correction into fuzzy motion estimation proposed by M. Protter [4]. The fuzzy motion estimation enables SR without explicit motion estimation even when the images have complex motion. Instead of direct searching the best matching point on the LR images to reconstruct the current pixel in HR grid, it considers many points on LR images with weights based on similarity. With these weights, all considered pixels in LR images are used to reconstruct one pixel in SR result.

The SR result using this fuzzy motion estimation is X that minimizes the following penalty function:

$$\eta_{SR}(X) = \sum_{(k,l)\in\Omega} \sum_{n\in[1,\ldots,N]} \sum_{(i,j)\in N^L(k,l)} w[k,l,i,j,n] \times \left\| D_p E^H_{k,l} HX - E^L_{i,j} y_n \right\|_2^2 + \lambda TV(X) \quad (1)$$

where Ω is the support of the entire image, (k, l) is the pixel location in the HR grid, (i, j) is the pixel location in the LR grid, N^L is the neighborhood of (k, l) in the LR image, [1,…,N] are the sequence numbers of the LR images, $w[k, l, i, j, n]$ is the weight, D_p is the patch decimation operator, $E^H_{k,l}$ and $E^L_{i,j}$ are the HR and LR patch extraction operators respectively, H is the blur operator, y_n is the LR image, and TV(X) is the Total-variation regularization term.

The weight w is computed using the similarity SSD between the two local patches in the interpolated target

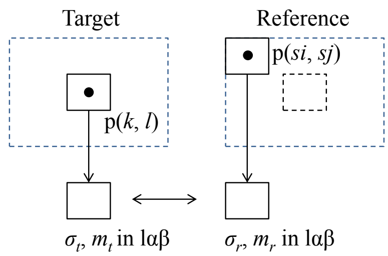

Figure. 2. Local color transfer between patches in target and reference images.

image and the reference image.

$$w[k,l,i,j,n] = \exp\left\{\left\|\hat{E}_{k,l} Z_{t0} - \hat{E}_{si,sj} Y_n\right\|_2^2 / 2\sigma^2\right\}, \quad (2)$$

where $\hat{E}_{k,l}$ and $\hat{E}_{si,sj}$ are the HR patch extraction operator at pixel (k, l) and (si, sj) respectively, s is the scaling factor, Z_{t0} is the interpolated target image, Y_n is the n-th interpolated reference image, and σ is the parameter that controls the shape of Gaussian function. For more information, please refer to [4].

In our method, colors of two local patches are corrected prior to calculating their similarity. As depicted in Fig. 2, the HR target patch centered at (k, l) is compared with the HR reference patch centered at (si, sj) in the neighborhood of (k, l) on the interpolated reference image. For the perceptional linearity in similarity measure, we use color transfer algorithm to match colors of two patches in lαβ space as follows.

$$\begin{aligned} l_c &= \frac{\sigma_t^l}{\sigma_r^l}\left(l_r - m_r^l\right) + m_t^l \\ \alpha_c &= \frac{\sigma_t^\alpha}{\sigma_r^\alpha}\left(\alpha_r - m_r^\alpha\right) + m_t^\alpha , \\ \beta_c &= \frac{\sigma_t^\beta}{\sigma_r^\beta}\left(\beta_r - m_r^\beta\right) + m_t^\beta \end{aligned} \quad (3)$$

where l, α, and β are the l, α, and β components of the patches, respectively, σ and m are the standard deviation and mean of patches in lαβ space, and subscript c, t, and r means corrected, target, and reference patches, respectively.

The weight for each RGB channel is calculated based on similarity of these color-corrected patches:

$$w[k,l,i,j,n] = \exp\left\{-\left\|\hat{E}_{k,l} T_{n0} - \hat{E}_{si,sj} R_n\right\|_2^2 / 2\sigma^2\right\}, \quad (4)$$

where T_{n0} is the interpolation of LR target image, t_{n0} and R_n is the interpolation of LR warped reference image, r_n. σ is the parameter that controls the shape of the difference statistics. The final weight is determined with the average of 3 weights in RGB channels. Once the weight is calculated on the HR interpolated images, the corresponding LR patches in LR reference image should be corrected. It is also converted in lαβ space using the mean and standard deviations used in eq. (3).

Let $c_{i,j,n}$ be the corrected LR patch at (i, j) in n-th LR image, then the final result is X that minimizes eq. (5).

$$\begin{aligned} \eta_{SR}(X) = &\sum_{(k,l)\in\Omega} \sum_{n\in[1,\ldots,N]} \sum_{(i,j)\in N^L(k,l)} w_c[k,l,i,j,n] \\ &\times \left\|D_p E_{k,l}^H H X - c_{i,j,n}\right\|_2^2 + \lambda TV(X) \end{aligned} \quad (5)$$

4. Experimental results

We tested our algorithm with two web image sets; Notre Dame Cathedral and Starbucks, collected by Google image search. Fig. 3 and 4 show the results of Notre Dame Cathedral and Starbucks image set respectively. The quality of SR image by our algorithm significantly outperforms that by Lanczos interpolation in terms of sharp edges without noises.

Fig. 5 shows the comparison of SR results using the conventional multi-image SR methods and our method. Fig. 5 (a) is the result of global color correction using affine transformation as in [3] followed by SR with global motion estimation. Fig. 5 (b) is the result of SR with fuzzy motion estimation [4]. All input images are geometrically registered in advance. As demonstrated in the results our approach shows better performance for web images.

5. Conclusion

In this paper, we present a unified SR framework applicable to web images which have complex motion and photometric diversity. We argue that the incorporation of color correction into fuzzy motion significantly increases the robustness to geometric and photometric variation. Thus it leads to better visual quality than the conventional methods.

Acknowledgements

This work was supported by the Global Frontier R&D Program on <Human-centered Interaction for Coexistence> funded by the National Research Foundation of Korea grant funded by the Korean Government(MEST) (NRF-M1AXA003-2011-0028362)

References

[1] S. C Park, M. K. Park, and M. G. Kang, "Super-Resolution Image Reconstruction: A Technical Overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21-36, May 2003.

[2] S. Farsiu, and M. D. Robinson, "Fast and Robust Multiframe Super Resolution," *IEEE TIP*, vol. 13, no. 10, pp. 1327-1344, Oct. 2004.

[3] D. Capel, and A. Zisserman, "Computer Vision Applied to Super Resolution," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 75-86, May 2003.

[4] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the Nonlocal-Means to Super-Resolution Reconstruction," *IEEE TIP*, vol. 18, no. 1, pp. 36-51, Jan. 2009.

[5] H. Takeda, P. Milanfar, M. Protter, and M. Elad, "Super-Resolution without Explicit Subpixel Motion Estimation," *IEEE TIP*, vol. 18, no. 9, pp. 1958-1975, Sep. 2009.

[6] M. Gevrekci, and B. K. Gunturk, "Superresolution under Photometric Diversity of Images," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 1-12, April 2007.

[7] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color Transfer between Images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34-41, Sep./Oct. 2001.

[8] R. I. Hartley and A. Zisserman, "*Multiple View Geometry in Computer Vision*," London, U.K: Cambridge Univ. Press, 2000

[9] Y.-W. Tai, J. Jia, and C.-K. Tang, "Local Color Transfer via Probabilistic Segmentation by Expectation-Maximization," in *Proc. of CVPR*, vol. 1, pp. 747-754, 2005.

(a) (b) (c) (d)

Figure. 3. Notre Dame Cathedral. (a) Selected input images. (b) ROI. (c) Lanczos interpolation. (d) Result of proposed algorithm.

(a) (b) (c) (d)

Figure. 4. Starbucks. (a) Selected input images. (b) ROI. (c) Lanczos interpolation. (d) Result of proposed algorithm.

(a) (b) (c)

Figure. 5. SR results comparison. (a) SR with global color correction and motion estimation (b) SR with fuzzy motion estimation (c) SR with proposed algorithm