Robust Image Alignment with Multiple Feature Descriptors and Matching-Guided Neighborhoods

Kuang-Jui Hsu[1,2], Yen-Yu Lin[1], Yung-Yu Chuang[2]
[1]Academia Sinica, Taiwan. [2]National Taiwan University, Taiwan.

Image alignment aims to densely identify pixel correspondences across images. It is an active and fundamental topic in computer vision. The major challenge that image alignment techniques must face is the large photometric and geometric variations between images to be aligned. Such unfavorable variations significantly degrade the performance of conventional optical flow approaches. Descriptor-based methods, such as [2, 6], address this challenge via adopting more powerful features instead of raw pixel features. Despite their effectiveness, these methods still have two main limitations. First, the performance of descriptor-based approaches to image alignment relies on the chosen descriptor, but the optimal descriptor typically varies from image to image, or even pixel to pixel. Second, the neighborhood structure for smoothness enforcement is usually predefined before alignment. However, object boundaries are often better discovered during alignment. The proposed approach tackles the two issues by using adaptive descriptor selection and dynamic neighborhood construction, respectively.

To align two given images, I_1 and I_2, we integrate the learning of pixel-specific affine matrices, $A = \{A_i\}_{i=1}^N$, and neighborhoods, $E = \{\mathbf{e}_i\}_{i=1}^N$, into the process of alignment, where N is the number of pixels in I_1. Multiple descriptors are applied to better characterize each pixel in I_1 and I_2. A set of flow map proposals can be generated by using any descriptor-based algorithm. We in this work use SIFT flow [2]. Suppose M descriptors are used. The generated flow proposals would be $\{W^m\}_{m=1}^M$, where $W^m = \{\mathbf{w}_i^m\}_{i=1}^N$ is the flow map produced by using SIFT flow with the substituted descriptor m. Our method yields the flow map $W = \{\mathbf{w}_i\}_{i=1}^N$ by referring to only proposals $\{W^m\}_{m=1}^M$, i.e., $\mathbf{w}_i \leftarrow \mathbf{w}_i^{\ell_i}$, where $\ell_i \in \{1,2,...,M\}$. Hence, it is formulated as a labeling problem over $L = \{\ell_i\}_{i=1}^N$.

Specifically, our approach is cast as the following constrained optimization problem:

$$\min_{E,A,L} \quad \sum_{i=1}^N J(\ell_i, A_i, \mathbf{e}_i) \tag{1}$$

$$\text{s.t.} \quad \mathbf{e}_i \succeq 0,\ \mathbf{e}_i^\top \mathbf{1} = 1, \text{ for } i = 1,2,...,N, \tag{2}$$

where $\mathbf{1}$ is a column vector whose elements are one. The constraints in Eq. (2) ensure that the sought neighborhood for each pixel i is non-negative and normalized. $J(\mathbf{e}_i, A_i, \ell_i)$ is the energy function regarding pixel i, and is defined below:

$$\begin{aligned} J(\ell_i, A_i, \mathbf{e}_i) =& \gamma\|\mathbf{p}_i' - A_i\mathbf{p}_i\|^2 + \sum_{j\in\mathcal{N}_i} e_{ij}\|\mathbf{p}_j' - A_i\mathbf{p}_j\|^2 \\ &+ \alpha \sum_{j\in\mathcal{N}_i}(e_{ij} - s_{ij})^2 + \beta \sum_{j\in\mathcal{N}_i} e_{ij}[\ell_i \neq \ell_j], \end{aligned} \tag{3}$$

where α, β, and γ are three non-negative constants, $\mathcal{N}_i$ is the index set of the neighbors of pixel i, and $\mathbf{p}_i' = \mathbf{p}_i + \mathbf{w}_i^{\ell_i}$. s_{ij} measures the geometric consistency beween correspondences at pixels i and j. Three optimization variables present in the objective function pertaining to pixel i, including its proposal selector ℓ_i, affine matrix A_i, and neighborhood weight vector $\mathbf{e}_i = \{e_{ij}\}_{j\in\mathcal{N}_i}$. Four terms in Eq. (3) are employed to enforce the compatibility among the three variables.

Since directly optimizing Eq. (1) is difficult, we instead adopt an iterative, alternating strategy. At each iteration, one of the three variables is optimized while keeping the others fixed, and then their roles are switched sequentially. It turns out that Eq. (1) is respectively reduced to a labeling problem, a weighted least square problem, and a convex quadratic programming problem. There exist off-the-shelf solvers to effectively optimize each of the three problems. As shown in the paper, the object-aware neighborhoods are gradually revealed by the learned affine matrices through the iterative optimization, while the alignment results would be progressively improved owing to the better neighborhoods.

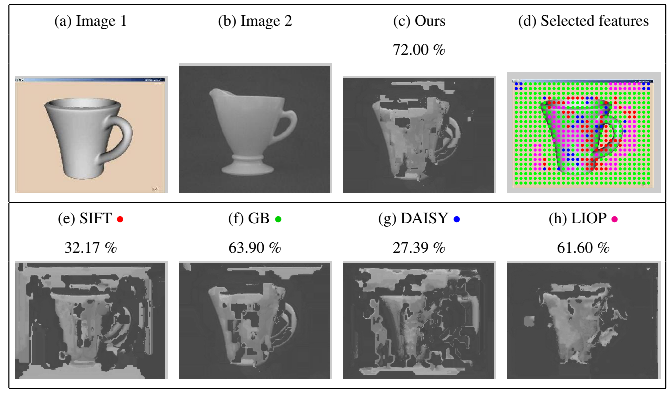

Figure 1: Example image alignment results using the proposed method and SIFT-flow variants with four different feature descriptors. On the top, we show its IoU score. (a) & (b) Images to be aligned. (c) Our result. (d) Feature selection visualization. (e) ∼ (h) SIFT flow results with different features, (e) SIFT, (f) GB, (g) DAISY, (h) LIOP.

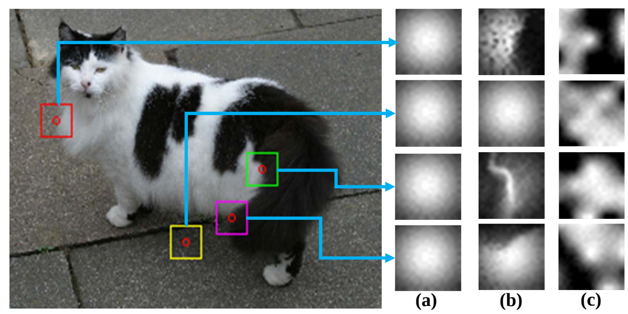

Figure 2: Comparison of three different weighting schemes. (a) Gaussian weight. (b) Bilateral weight. (c) The proposed method.

In the experiments, our approach is compared with the state-of-the-art methods on four benchmark datasets, which cover the tasks of dense correspondence, object and scene matching. The experimental results demonstrate that our approach can effectively produce object-aware neighborhoods and select a proper descriptor for each pixel, thus leading to a remarkable performance boost. Figure 1 shows that our method can pick the most plausible descriptor for each pixel, and the result is better than those by the SIFT-flow variants with four different descriptors, including *SIFT* [3] , *GB (geometric blur)* [1], *DAISY* [4], and *LIOP* [5]. Figure 2 gives an example for comparing the proposed method to Gaussian and bilateral weights. It can be observed that the neighborhoods by our method are more object-aware.

[1] A. C. Berg and J. Malik. Geometric blur for template matching. In *CVPR*, 2001.

[2] C. Liu, J. Yuen, and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. *TPAMI*, 2011.

[3] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.

[4] E. Tola, V. Lepetit, and P. Fua. DAISY: An efficient dense descriptor applied to wide-baseline stereo. *TPAMI*, 2010.

[5] Z. Wang, B. Fan, and F. Wu. Local intensity order pattern for feature description. In *ICCV*, 2011.

[6] H. Yang, W.-Y. Lin, and J. Lu. Daisy filter flow: A generalized discrete approach to dense correspondences. In *CVPR*, 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.