Unsupervised Visual Alignment with Similarity Graphs

Fatemeh Shokrollahi Yancheshmeh, Ke Chen, and Joni-Kristian Kämäräinen
Department of Signal Processing, Tampere University of Technology

Visual alignment of an image ensemble is to find the corresponding control points between them. This is an important pre-processing step for a number of high-level applications such as object detection and categorization [2]. In those applications, it can be made difficult due to the large pose variation of class examples in the images. The alignment remains hot yet challenging topic especially in large-scale visual recognition problems to i) avoid the manpower spent on annotating the control points or object landmarks in supervised object alignment [7], ii) the lack of moderately good initial alignments needed for image congealing algorithms [4], and iii) the manual seed selection in the feature-based alignment [5]. In this work, we adopt the feature-based approach, but our unsupervised visual alignment framework illustrated in Figure 1 can overcome the aforementioned drawbacks.

At the core of our method is the feature-based similarity between two images I_a and I_b based on local features, which can be measured in two ways: how similar feature points are to their corresponding feature points, and how much the spatial arrangement of the feature points is changed. The matching cost function can be divided to the feature match and feature geometric distortion costs:

$$C(I_a, I_b) = \lambda_1 C_{match}(I_a, I_b) + \lambda_2 C_{dist.}(I_a, I_b) \ .$$

where λ_1 and λ_2 are the trade-off parameters between the two terms.

In feature-based approaches, image representation consists of N feature descriptors $F_{i=1 \ldots N}$ (e.g., SIFT) and their spatial coordinates $x_{i=1 \ldots N}$. Since the feature matching $C_{match}(I_a, I_b)$ and geometric distortion $C_{dist.}(I_a, I_b)$ are dependent, the definition of cost can thus be formulated as:

$$\begin{aligned} & C(I_a, I_b) := C(I_a, I_b; T, A) = \\ & C\left(\left\{F^{(a)}, x^{(a)}\right\}, \left\{F^{(b)}, x^{(b)}\right\}; T, A\right) = \\ & \lambda_1 C_{match}\left(F^{(a)}, AF^{(b)}\right) + \lambda_2 C_{dist.}\left(X^{(a)}, AT(X^{(b)})\right) \end{aligned} \quad , \qquad (1)$$

where $A_{N_a \times N_b}$ is the assignment matrix and its element (i.e., a_{ij}) defines which feature $F_i^{(a)}$ of I_a correspond to which feature $F_j^{(b)}$ of I_b. T is geometric transformation such as a 3×3 linear homography matrix.

We use the similarity cost in (1) to find the pairwise similarity value of the images I_a and I_b:

$$C(I_a, I_b; T, A) = \min_{T, A} C(I_a, I_b) \ .$$

To avoid the dummy variables, we change the minimization of the similarity cost C to the maximization of the similarity S:

$$\begin{aligned} \text{maximize} \ & \sum_i \sum_j s_{ij} a_{ij} \\ \text{subject to} \ & \sum_j a_{ij} \leq 1 \quad i = 1, \ldots, N_a \\ & \sum_i a_{ij} \leq 1 \quad j = 1, \ldots, N_b \\ & a_{ij} \in \{0, 1\} \end{aligned} \quad . \qquad (2)$$

The maximization problem is known as *the generalized assignment problem*, which is NP-hard and even APX-hard to approximate [1].

We apply a fast approximation of the maximization problem with the computational complexity of $O(N)$ (see Algorithm 1 in the paper). Pair-wise image similarities of N images can thus be computed to construct a full $N \times N$ image similarity matrix $G(i,j) = S(I_i, I_j)$, which is the *weighted adjacency matrix* of a full connected graph G (illustrated in middle-right in Figure 1). Inspired by the random walk closeness centrality [6], we define a centrality measure which is used to identify good candidates ("alignment hubs") to which other images are accurately aligned. Such hubs correspond to manually selected "seeds" in [5] (see the middle-left block of Figure 1).

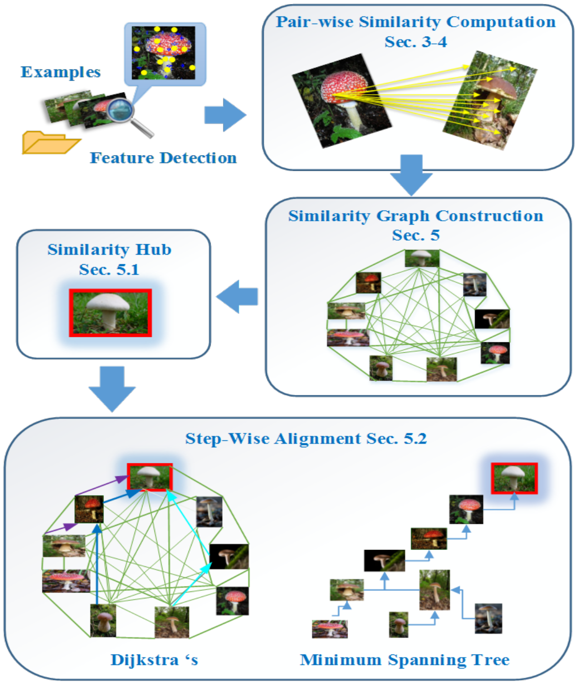

Figure 1: The workflow of our visual alignment approach.

Finally, compared to the baseline direct alignment [5], we employ two types of step-wise alignment strategies to exploit the graph G structure including: Minimum spanning tree (MST) and Djikstra's shortest path algorithm, which can step-wise align,"morph", an image to another within the path to the central hub.

In our experiments, our feature-based similarity graph driven step-wise alignment to the "similarity hubs" achieves superior results to the most recent alignment and congealing works [3, 5] with the same benchmarks.

[1] R. Cohen, L. Katzir, and D. Raz. An efficient approximation for the generalized assignment problem. *Information Processing Letters*, 100 (4), 2006.

[2] E. Gavves, B. Fernando, C. G. M. Snoek, A. W. M. Smeulders, and T. Tuytelaars. Fine-grained categorization by alignments. In *ICCV*, 2013.

[3] G. Huang, M. Mattar, H. Lee, and E. Learned-Miller. Learning to align from scratch. In *NIPS*, 2012.

[4] G.B. Huang, V. Jain, and E. Learned-Miller. Unsupervised joint alignment of complex images. In *ICCV*, 2007.

[5] J. Lankinen and J.-K. Kamarainen. Local feature based unsupervised alignment of object class images. In *BMVC*, 2011.

[6] J.D. Noh and H. Rieger. Random walks on complex networks. *Phys. Rev. Lett.*, 92(118701), 2004.

[7] Xuehan Xiong and Fernando De la Torre Frade. Supervised descent method and its applications to face alignment. In *CVPR*, 2013.

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