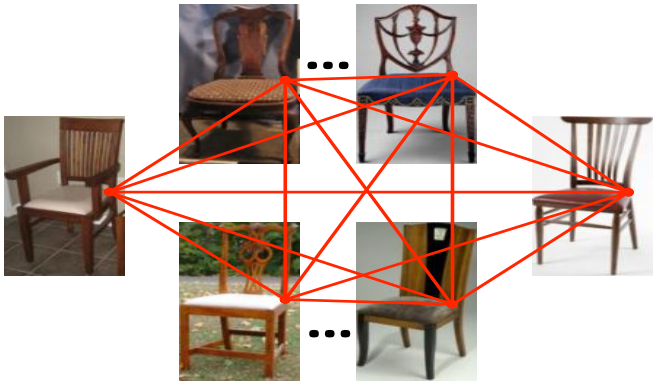
FlowWeb: Joint Image Set Alignment by Weaving Consistent, Pixel-wise Correspondences

Tinghui Zhou[1], Yong Jae Lee[2], Stella X. Yu[1,3], Alexei A. Efros[1],
[1]University of California, Berkeley. [2]University of California, Davis. [3]International Computer Science Institute (ICSI).

Consider a pair of chairs depicted in Fig. 1(a). While the chairs might look similar, locally their features (like the seat corner above) are very different in appearance, so even state-of-the-art image alignment approaches like SIFT Flow [4] have trouble finding correct correspondences. In this paper, we propose to "level the playing field" by starting with a *set of images* and computing correspondences *jointly* over this set in a globally-consistent way, as shown in Fig. 1(b).

(a) Example pixel correspondence using SIFT Flow [3]



(b) Example pixel correspondence using our FlowWeb

Figure 1: Finding pixel-wise correspondences between images is difficult even if they depict similar objects: (a) a typical correspondence error using SIFT Flow on a pair of images. (b) We propose computing correspondences jointly across an image collection in a globally-consistent way.

One can appreciate the power of joint correspondence by considering faces, a domain where correspondences are readily available, either via human annotation, or via domain-specific detectors. Large-scale face datasets, meticulously annotated with globally-consistent keypoint labels ("right mouth corner", "left ear lower tip", etc) were the catalyst for a plethora of methods in vision and graphics for the representation, analysis, 3D modeling, synthesis, morphing, browsing, etc. We believe that some of the same benefits of having large, jointly registered image collections should generalize beyond faces and apply more broadly to a range of visual entities, provided we have access to reliable correspondences.

FlowWeb Representation Given a collection of N images, we build a complete graph where a node denotes an image, and the edge between two nodes (i,j) is associated with flow field T_{ij} between images (I_i, I_j). For M pixels per image, T_{ij} is an $M \times 2$ matrix, each row containing the displacement vector between two matching pixels p and q in images I_i and I_j respectively: $T_{ij}^{pq} = x_q - x_p$, where x_p denotes the pixel coordinates of p.

Cycle Consistency Global correspondences in the image collection require the pairwise vector fields T to be consistent among different paths connecting two nodes in the flow graph. Cycle consistency criterion can be expressed as the net displacement along a cycle in the FlowWeb being zero, e.g. for three-image cycle (I_i, I_k, I_j), $T_{ik}^{pr} + T_{kj}^{rq} + T_{ji}^{qp} = 0$ if and only if pixels $< p, r, q >$ are cycle-consistent. The idea of utilizing consistency constraints within a global graph structure has also been applied to other vision and graphics problems, including co-segmentation [5], structure from motion [6], and shape matching [2].

Let $\triangle_{ij}^{pq}$ denote the set of image nodes that complete a consistent cycle with flow T_{ij}^{pq}. We define the *single flow cycle consistency* (SFCC) score as the cardinality of $\triangle$:

$$\mathcal{C}(T_{ij}^{pq}) = |\triangle_{ij}^{pq}|_{card} = \sum_{k=1, k\notin\{i,j\}}^{N} [T_{ij}^{pq} = T_{ik}^{pr} + T_{kj}^{rq}], \quad (1)$$

where $[\cdot]$ is the indicator function. We generalize the SFCC concept to the whole flow set $\mathbf{T}$ and define *all flow cycle consistency* (AFCC) which counts the total number of *unique* closed triangles in $\mathbf{T}$: $\mathcal{C}(\mathbf{T}) = \frac{1}{3}\sum_{i,j=1,i\neq j}^{N}\sum_{p\in I_i}\mathcal{C}(T_{ij}^{pq})$. Our overall objective function combines the cycle consistency criterion with the quality of match provided by the initial pairwise flow algorithm such as SIFT Flow[4] or DSP [3]:

$$\hat{\mathbf{T}} = \arg\max_{\mathbf{T}} \mathcal{C}(\mathbf{T}) - \lambda\mathcal{R}(\mathbf{T}, \mathbf{T}_0), \quad (2)$$

where $\lambda > 0$ can be chosen based on the initialization quality, $\mathbf{T}_0 = \{S_{ij}\}$ is the initial flow set obtained via some pairwise flow algorithm, and $\mathcal{R}(\cdot)$ measures the discrepancy between two flow sets in terms of Euclidean norm.

Our optimization procedure builds on the following intuition: even when pixels p and q do not have sufficient feature similarity to be matched directly, they should still be matched if there is *sufficient indirect evidence* from 1) their similarity to other images supporting the match (inter-image) and/or 2) proximity to neighboring pixels that have a good match (intra-image). Both are provided by the cycle consistency measure, and exploited alternately at each iteration. Please refer to the full paper for more details.

Fig. 2 shows some sample results on PASCAL-Parts dataset [1] for unsupervised part segment matching. DSP [3] is a state-of-the-art pairwise flow algorithm, and its output is used for initializing our method. Notice that many of the mistakes made by DSP are corrected by our joint alignment procedure.

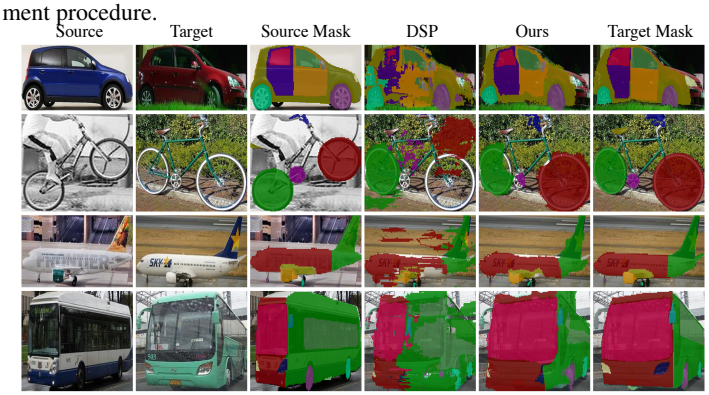


Figure 2: Correspondence visualization with color-coded part segments. Note that although the correspondences are shown as pairwise matching, our joint alignment is done across an image set.

[1] Xianjie Chen, Roozbeh Mottaghi, Xiaobai Liu, Sanja Fidler, Raquel Urtasun, and Alan Yuille. Detect what you can: Detecting and representing objects using holistic models and body parts. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[2] Q. Huang and L. Guibas. Consistent shape maps via semidefinite programming. In *SGP*, 2013.

[3] Jaechul Kim, Ce Liu, Fei Sha, and Kristen Grauman. Deformable spatial pyramid matching for fast dense correspondences. In *CVPR*, 2013.

[4] Ce Liu, Jenny Yuen, and Antonio Torralba. Sift flow: Dense correspondence across scenes and its applications. *TPAMI*, 33(5):978–994, 2011.

[5] F. Wang, Q. Huang, M. Ovsjanikov, and L. Guibas. Unsupervised multi-class joint image segmentation. In *CVPR*, 2014.

[6] Christopher Zach, Manfred Klopschitz, and Manfred Pollefeys. Disambiguating visual relations using loop constraints. In *CVPR*, 2010.

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