Shape-based Image Correspondence

Berk Sevilmis
berk_sevilmis@brown.edu
Benjamin B. Kimia
benjamin_kimia@brown.edu

LEMS
Brown University
Providence, RI 02912 USA

Many of the computer vision tasks such as stereo correspondence, optical flow, biometric user verification, and object recognition require the establishment of dense pixel correspondences between pair of images which can differ in image acquisition setting, *i.e.,* scene content and scene configuration. On the one end of the spectrum is the narrow-baseline stereo correspondence, where these variations are at a minimum since the same 3D scene is captured from slightly different viewpoints. On the other extreme is the *semantic image alignment*, involving images captured from different 3D scenes sharing similar characteristics such as containing same but different instances of objects.

Recent state-of-the-art approaches [2, 5, 7, 8, 10] attempt to compute correspondences between pair of images by matching image signatures, *e.g.,* color histograms, SIFT descriptor [9], CNN features [6], extracted locally from pixels and enforce smoothness on the correspondence field by enforcing spatial regularity. This type of a variational approach is challenged by semantically related images featuring large visual variations given that the variation measure does not capture any semantic aspect of the scene beyond a local histogram over a neighborhood.

Our approach is to introduce certain *semantic concepts* into the correspondence process. Specifically, in this paper, we explore the effect of shape as an additional guideline to the variational correspondence process. We ask whether specifying a pair of corresponding shapes can influence the correspondence process significantly and under what scenarios. We also ask whether shape should be specified in the form of a contour fragment or in the form of a closed curve bounding a region. Finally, when such corresponding shape constraints are not available, we ask whether object proposals can serve this purpose and under what conditions.

In our experiments, we consider three types of tasks: *(i)* optical flow, *(ii)* wide-baseline stereo correspondence, *(iii)* semantic image alignment and use four publicly available datasets, *i.e.,* the MPI Sintel Flow dataset [3], the DTU Robot Image datasets [1], the CUB-200-2011 dataset [11], and the PASCAL-Part dataset [4].

The qualitative and quantitative experiments reveal that *(i)* for datasets depicting slight visual variations, traditional methods are effective and do not benefit from the introduction of a shape correspondence constraint; *(ii)* for datasets depicting large visual variation with the same scene context, the shape correspondence constraints improve the correspondence in the range of 7%; and *(iii)* for datasets depicting instance and configuration variation, there are significant improvements up to 170%. Shape seems to help bring pixels into proper registration. The experiments on the form of the shape constraining the correspondence show that closed curves generally perform better than contour fragments which in turn perform better than shape presented as an unorganized cloud of points. Moreover, the use of object proposals to automatically obtain a shape constraint is also very promising. Compared to the performance obtained when ground truth segmentations are used, a ~26% drop in segmentation accuracy in terms of Jaccard index leads to a ~10% drop in performance.

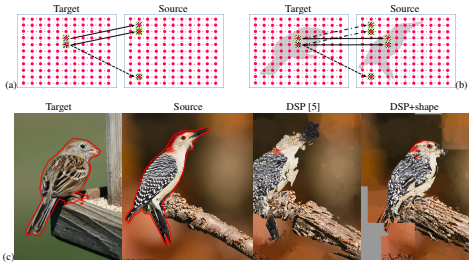

Figure 1: **Shape aligned dense correspondence.** (a) Spatial regularity in current state-of-the-art methods only disambiguate matches which are not locally consistent, *i.e.,* preferring the solid line correspondence to dashed one. (b) Shape alignment can reduce the ambiguity further by ruling out correspondences which violate inside-outside consistency. (c) A visual result. The warped source using shape alignment constraint is clearly superior.

[1] H. Aanæs, A.L. Dahl, and K. Steenstrup Pedersen. Interesting interest points. *IJCV*, 97:18–35, 2012.

[2] Connelly Barnes, Eli Shechtman, Dan B. Goldman, and Adam Finkelstein. The generalized patchmatch correspondence algorithm. In *ECCV*, pages 29–43, 2010.

[3] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *ECCV*, pages 611–625, 2012.

[4] 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 *CVPR*, 2014.

[5] Jaechul Kim, Ce Liu, Fei Sha, and Kristen Grauman. Deformable spatial pyramid matching for fast dense correspondences. In *CVPR*, pages 2307–2314, 2013.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1106–1114, 2012.

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

[8] Jonathan Long, Ning Zhang, and Trevor Darrell. Do convnets learn correspondence? In *NIPS*, pages 1601–1609, 2014.

[9] David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[10] Weichao Qiu, Xinggang Wang, Xiang Bai, Alan L. Yuille, and Zhuowen Tu. Scale-space SIFT flow. In *WACV*, pages 1112–1119, 2014.

[11] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.