

Subpixel Semantic Flow

Berk Sevilmis

berk_sevilmis@brown.edu

Benjamin B. Kimia

benjamin_kimia@brown.edu

LEMS

Brown University

Providence, RI 02912 USA

Abstract

Dense semantic correspondence is usually cast as a variational optimization problem. Current methods generally focus on obtaining more discriminative features (to improve the data/correspondence term), and adopt a message-passing algorithm for inference, which is generally a variant of loopy belief propagation. One drawback of such optimization is that the flow vectors are constrained to be discrete variables resulting in pixel resolution, “blocky” flow fields. The main hinderance to formulating the problem in continuous space, and hence solving the problem at subpixel resolution, is the use of histogram based descriptors such as SIFT, HOG, etc. Such sparse feature descriptors are distinctive but linearize poorly. In this paper, we revisit a classic dense descriptor, namely Geometric Blur, which is, in contrast, extracted from a linear filter (spatially varying Gaussian) response that can be linearized and therefore interpolated at subpixel values. In addition to the data and smoothness terms used in variational models, we also add a term promoting bidirectional flow consistency. As there is no longer a finite set of values a flow vector can take, we use gradient descent based continuous optimization. We present promising results encouraging the use of gradient based continuous optimization in establishing dense semantic correspondences. The proposed subpixel approach is applicable to any dense descriptor, including the mid-level convolutional layer outputs of convolutional neural networks (CNNs).

1 Introduction

Image alignment deals with finding the optimal geometric transformation that maps corresponding pixels across images. It is one of the core problems in computer vision and is used in many applications such as semantic image segmentation [13], scene parsing [14], video depth estimation [12], image enhancement [8], etc. The fundamental difficulty in aligning a pair of images is not only due to the vast range of photometric and geometric variations images might exhibit, but also due to the variations in the meaning of *correspondence*. In stereo matching and optical flow, correspondence means projections of the same 3D world coordinates, while in semantic alignment of objects, correspondence is anatomical equivalence, e.g., the eyes, beak and tail of one type of bird should align with the eyes, beak, and tail of a different instance of a bird, respectively. In other words, in some problems correspondence is faced with projection variation while in others it is faced with intra-class variation.

Recently, many approaches have targeted the semantic image alignment problem. Most of these methods adopt a variational model combining correspondence and transformation



Figure 1: **Subpixel semantic flow.** (a) A pair of car images (source on top and target at the bottom). From (b) to (d): The top rows show the results obtained using SIFT flow [18], one of the state-of-the-art methods in dense semantic image alignment, and the bottom rows show the results obtained using the proposed method. (b) Warped target images. (c) Overlay of source and target images after alignment. (d) Flow fields. Notice the “block” effect in SIFT flow [18] due to flow fields obtained at pixel resolution.

constraints in data and regularizer terms, respectively. In order to accomplish correspondences across different instances of objects/scenes while being robust to small affine deformations, popular sparse feature representations such as SIFT [20], HOG [8], etc., have been used. The general approach to optimization has been graph-based, especially message-passing algorithms, such as belief propagation, to infer flow fields relating the images. The limitation in such a message-passing based algorithm is that the search problem is cast as a labelling problem, thus constraining the flow vectors to be discrete variables taking values from a finite set of labels. This results in obtaining pixel resolution flow fields. Even when robust functions, such as truncated L^1 norms, have been utilized to eliminate outliers and account for flow discontinuities, “blocky” flow fields are obtained. Alternatively, gradient descent type of optimization techniques have generally been overlooked in providing local minimizers, mainly because it is not straightforward to approximate the descent direction [9] and hence the linearization of discriminative but sparse feature representations. Subpixel resolution, being a superset of pixel resolution, based flow fields are preferable as they can better localize flow discontinuities and can better keep the integrity of objects after warp transformation.

The contribution of this paper is to present a framework to address this problem that achieves dense semantic correspondences at subpixel resolution. We revisit a dense descriptor, specifically Geometric Blur [9] in this paper, and reformulate the semantic image alignment problem in continuous space. We also introduce a new term to the variational objective function related to bidirectional flow consistency and show how to optimize the new objective function. Experiments have been carried out on a recently introduced dataset having different object categories, showing promising results.

The rest of the paper is organized as follows. In Section 2 we review the relevant work on semantic image alignment. In Section 3 we introduce our subpixel semantic flow formulation. Section 4 presents experimental results leading to conclusions and insights on future work, which are then discussed in Section 5.

2 Related Work

One of the earliest works, SIFT flow [18], leaped into densely aligning similar scene contents from densely aligning same 3D scene content using SIFT [20] features, as low level

image cues such as observed pixel brightness values is not of much help in the former case. Extracting SIFT signatures at a predefined scale and adopting a graph based optimization, loopy belief propagation is used to obtain flow fields in a hierarchical manner. Scale-Space SIFT flow (SSF) [20] additionally extracts SIFT [20] descriptors at different scales in order to robustly match similar image content when viewed at different scales. Deformable Spatial Pyramid (DSP) [20] matching reconsiders the graph representation, introducing deformable grid cells connected across a pyramid. Generalized Deformable Spatial Pyramid (GDSP) [20] extends the search space of feature correspondences by considering various scales and rotations. The work of [20] combines flow field outputs obtained using different feature descriptors so that discriminative powers of each can be efficiently exploited.

There has also been interest in obtaining fast dense correspondences. The Generalized PatchMatch [2] algorithm performs a fast nearest neighbor search across translations, scales and rotations, however, sacrifices geometric coherency of matches. The DAISY filter flow (DFF) [25] uses the PatchMatch search strategy and combines filter-based inference to obtain smoother flow fields.

When the background clutter in which the different instances of the same object category is seen is not of a simple type or when only some portions of a pair of images depict similar scene content, object proposals have been exploited to detect the commonality across images. The work of [2] uses object proposals to avoid cross-talk in the correspondence between the common objects of interest and background clutter. ProposalFlow [20] uses object proposal matching to anchor and guide the geometric transformation relating the images. The work of [2] aims to jointly cosegment and align the common object of interest.

FlowWeb [26] aims to densely align a set of images, starting from pairwise flow fields initialized with DSP and iteratively improving them in order to achieve cycle consistency.

Last but not least, convolutional neural networks have also been used to either train diagnostic feature descriptors, which are then used in graph based optimization of the previous approaches stated above, or to train an end-to-end network outputting a flow field [2, 20, 27].

3 The Subpixel Flow Approach

Given two images I_1 and I_2 , our goal is to find a geometric transformation that maps a point $\mathbf{p} = (x, y)$ in I_1 to a point $\bar{\mathbf{p}} = \mathcal{T}(\mathbf{p})$ in I_2 . Alternatively, we can write $\mathcal{T}(\mathbf{p}) = \mathbf{p} + \mathbf{w}(\mathbf{p})$, where $\mathbf{w}(\mathbf{p})$ is called a flow vector and is comprised of a horizontal and a vertical component ($\mathbf{w}(\mathbf{p}) = (u(\mathbf{p}), v(\mathbf{p}))$). Flow vectors have to be specified everywhere, since the geometric transformation is nonparametric. We consider the following objective function:

$$E(u, v) = \int \psi \left(\sum_k |C_2(\mathbf{p} + \mathbf{w}(\mathbf{p}), k) - C_1(\mathbf{p}, k)|^2 \right) d\mathbf{p} + \alpha \int \psi \left(|\nabla u(\mathbf{p})|^2 + |\nabla v(\mathbf{p})|^2 \right) d\mathbf{p} \quad (1)$$

where $\psi(\cdot)$ is a robust function and $C_1(\cdot)$ and $C_2(\cdot)$ are densely extracted, normalized (zero mean, unit variance) Geometric Blur descriptors treated as multi-channel images indexed by k . Geometric Blur descriptors have been compared using normalized cross-correlation as the similarity function [2]. It can be easily shown that normalized cross-correlation of any two descriptors is proportional to the mean of the sum of squared differences between the normalized descriptors, hence being the choice used in Eq. 1. The total energy is the weighted sum of the data term, responsible for local descriptor matching, and smoothness term which penalizes the total variation of the flow field under a piecewise smooth flow field

assumption. We choose $\psi(x^2) = \sqrt{x^2 + \epsilon^2}$ with $\epsilon = 0.001$, which results in a differentiable, convex approximation of the L^1 norm. The objective function in Eq. 1 can be minimized using iterative reweighted least squares.

Bidirectional flow consistency: Minimizing the above objective function does not necessarily yield symmetric flow fields when the roles of images are interchanged, namely when finding a geometric transformation mapping points from I_2 to I_1 . Specifically, if $\mathbf{p} \in I_1$ is mapped to $\bar{\mathbf{p}} = \mathbf{p} + \mathbf{w}_1(\mathbf{p}) \in I_2$ for the (I_1, I_2) alignment, and $\bar{\mathbf{p}}$ is mapped to $\hat{\mathbf{p}} = \bar{\mathbf{p}} + \mathbf{w}_2(\bar{\mathbf{p}})$ for the (I_2, I_1) alignment, we may not have $\hat{\mathbf{p}} = \mathbf{p}$. Equivalently, it may not be that $\mathbf{w}_1(\mathbf{p}) + \mathbf{w}_2(\mathbf{p} + \mathbf{w}_1(\mathbf{p})) = 0$. Except for the many points in one of the images mapping to one point in the other image, the one-to-one correspondence between the points in images is desired, so that forward and backward flows remain consistent. We extend the above objective function to simultaneously consider flow fields from both directions and introduce a consistency term as follows:

$$\begin{aligned} E(u_1, v_1, u_2, v_2) = & \int \psi \left(\sum_k |C_2(\mathbf{p}_1 + \mathbf{w}_1(\mathbf{p}_1), k) - C_1(\mathbf{p}_1, k)|^2 \right) d\mathbf{p}_1 + \alpha \int \psi \left(|\nabla u_1(\mathbf{p}_1)|^2 + |\nabla v_1(\mathbf{p}_1)|^2 \right) d\mathbf{p}_1 + \\ & \int \psi \left(\sum_k |C_1(\mathbf{p}_2 + \mathbf{w}_2(\mathbf{p}_2), k) - C_2(\mathbf{p}_2, k)|^2 \right) d\mathbf{p}_2 + \alpha \int \psi \left(|\nabla u_2(\mathbf{p}_2)|^2 + |\nabla v_2(\mathbf{p}_2)|^2 \right) d\mathbf{p}_2 + \\ & \beta \int \phi \left(|\mathbf{w}_1(\mathbf{p}_1) + \mathbf{w}_2(\mathbf{p}_1 + \mathbf{w}_1(\mathbf{p}_1))|^2 \right) d\mathbf{p}_1 + \beta \int \phi \left(|\mathbf{w}_2(\mathbf{p}_2) + \mathbf{w}_1(\mathbf{p}_2 + \mathbf{w}_2(\mathbf{p}_2))|^2 \right) d\mathbf{p}_2. \end{aligned} \quad (2)$$

First two terms correspond to variational formulation of the forward flow, introduced in Eq. 1, similarly the next two terms are responsible for the variational formulation of the backward flow, and the last two terms are responsible for establishing symmetry between forward and backward flow fields. We chose L^2 norm in order to measure bidirectional consistency, *i.e.* $\phi(x^2) = x^2$. As the flow vectors and consistency errors are unknown at the same time, it is not straightforward to minimize the objective function in Eq. 2. Let us represent the objective function in Eq. 2 as a sum of two energy functions, one focused on forward, and one focused on backward flow

$$E(u_1, v_1, u_2, v_2) = E_1(u_1, v_1, u_2, v_2) + E_2(u_1, v_1, u_2, v_2) \quad (3)$$

where

$$\begin{aligned} E_1(u_1, v_1, u_2, v_2) = & \int \psi \left(\sum_k |C_2(\mathbf{p}_1 + \mathbf{w}_1(\mathbf{p}_1), k) - C_1(\mathbf{p}_1, k)|^2 \right) d\mathbf{p}_1 + \alpha \int \psi \left(|\nabla u_1(\mathbf{p}_1)|^2 + |\nabla v_1(\mathbf{p}_1)|^2 \right) d\mathbf{p}_1 + \\ & \beta \int \phi \left(|\mathbf{w}_1(\mathbf{p}_1) + \mathbf{w}_2(\mathbf{p}_1 + \mathbf{w}_1(\mathbf{p}_1))|^2 \right) d\mathbf{p}_1 \end{aligned} \quad (4)$$

$$\begin{aligned} E_2(u_1, v_1, u_2, v_2) = & \int \psi \left(\sum_k |C_1(\mathbf{p}_2 + \mathbf{w}_2(\mathbf{p}_2), k) - C_2(\mathbf{p}_2, k)|^2 \right) d\mathbf{p}_2 + \alpha \int \psi \left(|\nabla u_2(\mathbf{p}_2)|^2 + |\nabla v_2(\mathbf{p}_2)|^2 \right) d\mathbf{p}_2 + \\ & \beta \int \phi \left(|\mathbf{w}_2(\mathbf{p}_2) + \mathbf{w}_1(\mathbf{p}_2 + \mathbf{w}_2(\mathbf{p}_2))|^2 \right) d\mathbf{p}_2. \end{aligned} \quad (5)$$

This partitions the objective function into forward and backward flow field computations. Note that Eq. 4 and Eq. 5 are coupled due to the flow consistency terms. Similar to the work of [11], we adopt a coordinate descent approach to iteratively minimize it as follows. The initial flow fields, $\mathbf{w}_1^{(0)} = (u_1^{(0)}, v_1^{(0)})$ and $\mathbf{w}_2^{(0)} = (u_2^{(0)}, v_2^{(0)})$, are first obtained by

minimizing Eq. 1 separately for forward and backward directions. Denoting t as the iteration index, the flow fields are then updated as:

$$E \left(u_1^{(t+1)}, v_1^{(t+1)}, u_2^{(t+1)}, v_2^{(t+1)} \right) = E_1 \left(u_1^{(t+1)}, v_1^{(t+1)}, u_2^{(t)}, v_2^{(t)} \right) + E_2 \left(u_1^{(t)}, v_1^{(t)}, u_2^{(t+1)}, v_2^{(t+1)} \right), \quad (6)$$

where E_1 and E_2 are similarly minimized using iterative reweighted least squares. Please refer to the supplementary material for the full derivation and the algorithm.

Finally, we use the gPb edge detector [2] output to extract Geometric Blur [2] descriptors and use a coarse to fine optimization scheme on a Gaussian pyramid with downsampling rate of 0.5 similar to the work of [8].

4 Experimental Results

In the following experiments, we used $\alpha = 0.05$, $\beta = 0.5$ for the coarsest scale in pyramid, and $\alpha = 0.2$, $\beta = 0.25$ for the finer scales. The Taniai dataset [24], consisting of three groups FG3DCar, JODS, and PASCAL, which contains many instances of different object categories such as bikes, planes, horses etc. is used. We crop the images to their foreground bounding boxes. The algorithm performance, as measured in [22], is obtained as follows: First, the images are resized so that the larger dimension becomes of size 100 pixels. The flow error magnitude, $\|w_i - w_{GT,i}\|_2$, $i = 1, 2$ is calculated at every pixel and the average number of foreground pixels achieving flow error less than a varying threshold T is plotted with respect to T . The average area under this curve is obtained as a summary of matching accuracy. The matching accuracy when $T = 5$ pixels is also reported. We have experimented with two different versions of our algorithm. One, neglecting the flow consistency and obtaining forward and backward flow independently, named unidirectional flow (the solution to Eq. 1), and the other using bidirectional flow consistency detailed in the previous subsection, named bidirectional flow (the solution to Eq. 2). In addition to using Geometric Blur (GB) descriptors, we have also conducted a preliminary experiment on using deep learning features where GB descriptors are replaced by AlexNet [17] *conv1* (96 channels) features trained for the task of object classification. We evaluated our algorithm by comparing it to SIFT flow [18], SSF [20], DSP [19], and PatchMatch [8].

Fig. 2 shows the mean flow accuracy plots when varying thresholds are used, and Table 1 summarizes the matching accuracies of the algorithms. As can be seen in Fig. 2 and Table 1, our algorithm, when used with GB descriptors, performs better than other competing methods on JODS, and PASCAL subsets but not on FG3DCar. As the authors state in their paper, the ground truth flows have been obtained by marking some sparse keypoint correspondences and applying natural neighbor interpolation [23] to densify it. Visual inspection of the sparse correspondences and densified flow maps of FG3DCar reveals some poor flow maps, one of them shown in Fig. 3. Notice that the plates and front wheels of cars being some of the areas where misalignment occurs. Warped images using the densified flow fields do not adhere to the foreground object boundaries correctly. Compared to unidirectional flow, bidirectional flow, in general, achieves slightly better performances hence we opted for using bidirectional flow version of our algorithm in obtaining the rest of the quantitative and qualitative results. Interestingly, when handcrafted GB descriptors are replaced by *conv1* features, a drop in performance is seen. This might be due to the fact that these features are extracted with a stride of 4 pixels, sacrificing spatial resolution and hence precise localization which is crucial for correspondence. Hypercolumns [21], where outputs from multiple layers are stacked to cap-

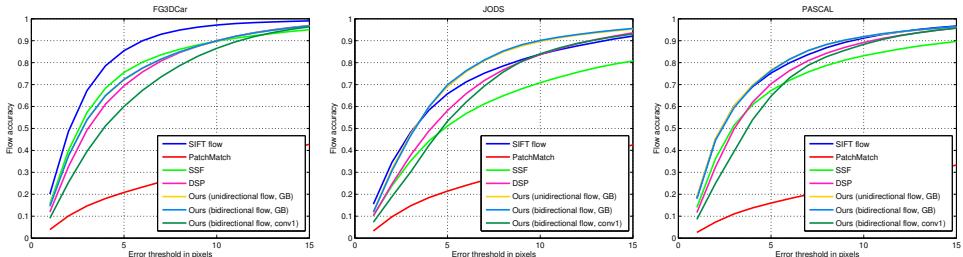


Figure 2: Average flow accuracies with varying error thresholds on Taniai dataset [24]. Note that the performances of the competing algorithms are different than the results reported in [24] as we crop the images to their foreground bounding boxes. Zoom in for better readability.

Methods	Descriptors	AUC			Match accuracy ($T=5$ pixels)		
		FG3DCar	JODS	PASCAL	FG3DCar	JODS	PASCAL
SIFT flow [18]	SIFT	80.27	66.96	74.41	85.48	65.82	75.28
PatchMatch [8]	SIFT	25.13	25.44	19.39	20.81	21.43	15.98
SSF [21]	SIFT	73.02	55.19	67.04	75.54	51.13	67.29
DSP [10]	SIFT	70.80	63.84	70.53	69.49	58.18	70.37
Our Method							
unidirectional flow	GB	72.09	70.16	74.96	72.13	69.13	76.62
bidirectional flow	GB	72.15	70.36	74.83	72.40	69.92	76.31
bidirectional flow	<i>conv1</i>	65.88	61.62	67.69	60.03	53.32	64.76

Table 1: Area under curve (AUC), and matching accuracy values at $T = 5$ pixels on Taniai dataset [24]. Our algorithm with GB descriptors performs better than other competing methods on JODS, and PASCAL subsets. Compared to unidirectional flow, in general, bidirectional flow achieves slightly better performances. Note that the performances of the competing algorithms are different than the results reported in [24] as we crop the images to their foreground bounding boxes.

ture both low level features and higher level semantics could also be used in our algorithm. We leave this as a future work.

Figs. 4, 5, 6, and 7 demonstrate qualitative results from the FG3DCar, JODS and the PASCAL subsets using bidirectional flow with GB descriptors. Notice the continuity of textures and how the images are smoothly warped, owing to subpixel flow fields.

5 Conclusion

In this paper we formulated the dense semantic correspondence problem in continuous space leading to subpixel flow fields. In addition to the data and smoothness terms, we incorporated bidirectional flow consistency terms to the variational objective function in order to promote correspondence symmetry. We tested our algorithm on a recently established dataset and showed promising results in this direction both qualitatively and quantitatively. We believe that such a subpixel formulation is very flexible, and given the differentiability of a cost term, the objective function can be easily extended to include complex constraints unlike graph

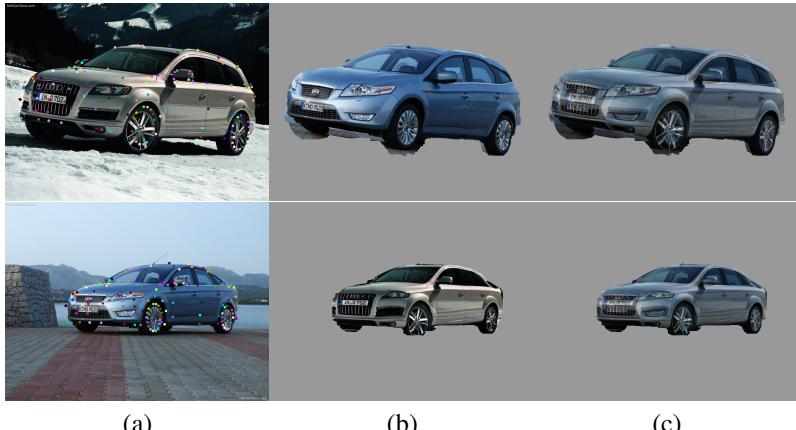


Figure 3: A demonstration of a ground truth flow error in FG3DCar dataset. (a) Two images with corresponding keypoints annotated as indicated by matching colors. The dense correspondence is then interpolated from these sparse correspondences. (b) Each image is warped into the other using the ground truth dense correspondence. Observe the misalignment around the front wheels and the front grill. (c) The blend of warped image and original image in each case highlights the misalignment further. This clearly shows that a veridical alignment is punished at these misalignments of the ground truth, which are specifically worse for the FG3DCar dataset. Zoom in for better visibility.



Figure 4: A qualitative result from the FG3DCar dataset using bidirectional flow with GB descriptors. The first row shows the source and target images and their blend before registration. The second row shows the warped ground truth result followed by results of SIFT flow [red box], PatchMatch [green box], SSF [blue box], DSP [orange box] and our method. The third row shows the ground truth blend of source and warped target images followed by results of SIFT flow [red box], PatchMatch [green box], SSF [blue box], DSP [orange box] and our method. The results are clearly better for our method. Zoom in for better visibility.



Figure 5: A qualitative result from the JODS dataset using bidirectional flow with GB descriptors. The first row shows the source and target images and their blend before registration. The second row shows the warped ground truth result followed by results of SIFT flow [18], PatchMatch [8], SSF [20], DSP [15] and our method. The third row shows the ground truth blend of source and warped target images followed by results of SIFT flow [18], PatchMatch [8], SSF [20], DSP [15] and our method. The results are clearly better for our method. Zoom in for better visibility.



Figure 6: A qualitative result from the JODS dataset using bidirectional flow with GB descriptors. The first row shows the source and target images and their blend before registration. The second row shows the warped ground truth result followed by results of SIFT flow [18], PatchMatch [8], SSF [20], DSP [15] and our method. The third row shows the ground truth blend of source and warped target images followed by results of SIFT flow [18], PatchMatch [8], SSF [20], DSP [15] and our method. The results are clearly better for our method. Zoom in for better visibility.



Figure 7: A qualitative result from the PASCAL dataset using bidirectional flow with GB descriptors. The first row shows the source and target images and their blend before registration. The second row shows the warped ground truth result followed by results of SIFT flow [X], PatchMatch [■], SSF [○], DSP [△] and our method. The third row shows the ground truth blend of source and warped target images followed by results of SIFT flow [X], PatchMatch [■], SSF [○], DSP [△] and our method. The results are clearly better for our method. Zoom in for better visibility.

based techniques where efficient minimizers exist only for simple higher order interactions. Although we made use of a classic feature descriptor, Geometric Flow [■], and conducted preliminary experiments using deep learning features, the method presented here is of a generic one and any descriptor that can be efficiently linearized can be used. We believe such a formulation can benefit from the recent advances in CNN features, and leave this direction as a future work.

References

- [1] Luis Álvarez, Rachid Deriche, Théodore Papadopoulo, and Javier Sánchez Pérez. Symmetrical dense optical flow estimation with occlusions detection. In *Computer Vision - ECCV 2002, 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28-31, 2002, Proceedings, Part I*, pages 721–735, 2002.
- [2] Pablo Arbelaez, Michael Maire, Charless C. Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, 2011.
- [3] Connelly Barnes, Eli Shechtman, Dan B. Goldman, and Adam Finkelstein. The generalized patchmatch correspondence algorithm. In *Computer Vision - ECCV 2010, 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III*, pages 29–43, 2010.
- [4] Alexander C. Berg and Jitendra Malik. Geometric blur for template matching. In *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*

- (*CVPR 2001*), with CD-ROM, 8-14 December 2001, Kauai, HI, USA, pages 607–614, 2001.
- [5] Hilton Bristow and Simon Lucey. *In Defense of Gradient-Based Alignment on Densely Sampled Sparse Features*, pages 135–152. Springer International Publishing, Cham, 2016.
 - [6] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *Computer Vision - ECCV 2004, 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV*, pages 25–36, 2004.
 - [7] Christopher Bongsoo Choy, JunYoung Gwak, Silvio Savarese, and Manmohan Krishna Chandraker. Universal correspondence network. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 2406–2414, 2016.
 - [8] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*, pages 886–893, 2005.
 - [9] Yoav HaCohen, Eli Shechtman, Dan B. Goldman, and Dani Lischinski. Non-rigid dense correspondence with applications for image enhancement. *ACM Trans. Graph.*, 30(4):70, 2011.
 - [10] Bumsub Ham, Minsu Cho, Cordelia Schmid, and Jean Ponce. Proposal flow. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 3475–3484, 2016.
 - [11] Bharath Hariharan, Pablo Andrés Arbeláez, Ross B. Girshick, and Jitendra Malik. Hypercolumns for object segmentation and fine-grained localization. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 447–456, 2015.
 - [12] Kuang-Jui Hsu, Yen-Yu Lin, and Yung-Yu Chuang. Robust image alignment with multiple feature descriptors and matching-guided neighborhoods. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 1921–1930, 2015.
 - [13] Junhwa Hur, Hwasup Lim, Changsoo Park, and Sang Chul Ahn. Generalized deformable spatial pyramid: Geometry-preserving dense correspondence estimation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 1392–1400, 2015.
 - [14] Kevin Karsch, Ce Liu, and Sing Bing Kang. Depth extraction from video using non-parametric sampling. In *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part V*, pages 775–788, 2012.
 - [15] Jaechul Kim, Ce Liu, Fei Sha, and Kristen Grauman. Deformable spatial pyramid matching for fast dense correspondences. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, pages 2307–2314, 2013.

- [16] Seungryong Kim, Dongbo Min, Bumsub Ham, Sangryul Jeon, Stephen Lin, and Kwanghoon Sohn. FCSS: fully convolutional self-similarity for dense semantic correspondence. *CoRR*, abs/1702.00926, 2017.
- [17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pages 1106–1114, 2012.
- [18] Ce Liu, Jenny Yuen, and Antonio Torralba. SIFT flow: Dense correspondence across scenes and its applications. *PAMI*, 33(5):978–994, 2011.
- [19] Ce Liu, Jenny Yuen, and Antonio Torralba. Nonparametric scene parsing via label transfer. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(12):2368–2382, 2011.
- [20] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [21] Weichao Qiu, Xinggang Wang, Xiang Bai, Alan L. Yuille, and Zhuowen Tu. Scale-space SIFT flow. In *IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, March 24-26, 2014*, pages 1112–1119, 2014.
- [22] Berk Sevilmis and Benjamin B. Kimia. Shape-based image correspondence. In *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*, 2016.
- [23] R. Sibson. A brief description of natural neighbour interpolation. *Interpreting multivariate data, chapter 2*, pages 21–36, John Wiley & Sons, 1981.
- [24] Tatsunori Taniai, Sudipta N. Sinha, and Yoichi Sato. Joint recovery of dense correspondence and cosegmentation in two images. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 4246–4255, 2016.
- [25] Hongsheng Yang, Wen-Yan Lin, and Jiangbo Lu. DAISY filter flow: A generalized discrete approach to dense correspondences. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, pages 3406–3413, 2014.
- [26] Tinghui Zhou, Yong Jae Lee, Stella X. Yu, and Alexei A. Efros. Flowweb: Joint image set alignment by weaving consistent, pixel-wise correspondences. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 1191–1200, 2015.
- [27] Tinghui Zhou, Philipp Krähenbühl, Mathieu Aubry, Qi-Xing Huang, and Alexei A. Efros. Learning dense correspondence via 3d-guided cycle consistency. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 117–126, 2016.