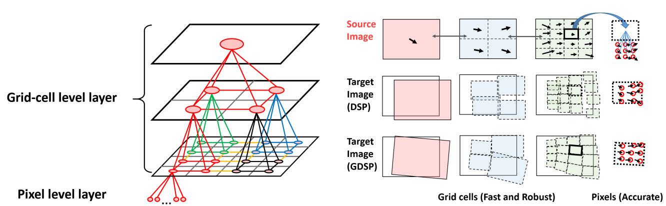
Generalized Deformable Spatial Pyramid: Geometry-Preserving Dense Correspondence Estimation

Junhwa Hur[1], Hwasup Lim[1,2], Changsoo Park[1], and Sang Chul Ahn[1,2]
[1]Center for Imaging Media Research, Robot & Media Institute, KIST, Seoul, Korea
[2]HCI & Robotics Dept., University of Science & Technology, Korea

Introduction: Densely matching two correlated images at the pixel level is one of the most fundamental tasks in computer vision applications. Specifically, for general dense correspondence algorithms, there mainly exist two principal challenges: (1) photometric variations due to different camera settings and illumination conditions and (2) geometric variations due to viewpoint changes, object pose changes, and the non-rigid deformation of objects between the images. These various factors are projected onto the 2D space; thus, it is challenging to decompose these factors from the images.

In this paper, we propose the Generalized Deformable Spatial Pyramid (GDSP) model to resolve the challenges and extend the capability of matching images under versatile forms of geometric variations. We reformulate the existing DSP [1] model by imposing rotation and scale invariant properties and considering the spatial relationship in the high dimensional search space through the pyramid structure. This high dimensional regularization directly links to our main contribution: we can effectively preserve the meaningful inherent geometry and texture in images while allowing a broad range of geometric variations such as affine, perspective and even non-rigid deformation. We provide an optimization method of our high dimensional objective functions by modifying loopy belief propagation to our formulation, which is the second contribution of our work.

Generalized Deformable Spatial Pyramid (GDSP): We propose a Generalized Deformable Spatial Pyramid (GDSP) model, which incorporates a rotation and scale term into the original DSP model [1] in Fig. 1(a). Our model allows each grid cell to rotate and increase or decrease itself, which gives it more flexibility to find its correspondence, as in Fig. 1(b).



(a) The graphical structure of DSP model (b) Comparison of matching methods between DSP and our GDSP model

Figure 1: The original DSP model and our GDSP model

Let I_S and I_T denote a source image and a target image to match, respectively. Our generalized objective function becomes

$$E(\mathbf{t},r,s) = \sum_i D_i(\mathbf{t}_i,r_i,s_i) + \sum_{\{i,j\}\in E} V_{ij}(\mathbf{t}_i,r_i,s_i,\mathbf{t}_j,r_j,s_j). \quad (1)$$

Each node i takes three states: t_i, r_i and s_i, which denote the translation, rotation and scale in the image coordinate, respectively.

In Eq. (1), data term $D_i(\mathbf{t}_i,r_i,s_i)$ calculates the SIFT matching cost of node i given its state $(\mathbf{t}_i,r_i,s_i)$ for all sampling pixels $\mathbf{p}$ in the node. The pairwise term V_{ij} penalizes the state discrepancy of two nodes that are connected by an edge. To simultaneously regulating multiple states (scale, rotation, translation) that have dependencies, we reflect the influence of rotation and scale variation on measuring the translation discrepancies by reasoning in the local spatial coordinate. This spatial reasoning provides a reasonable smoothness regularization when scale and rotation vary.

In the optimization process, we adopt loopy belief propagation with modified four-dimensional distance transform. Our optimization decouples high dimensional correlated states and allows for sequential message update of such states. This optimization scheme reduces the complexity of our optimization problem from $O(n^2)$ to $O(n)$.

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

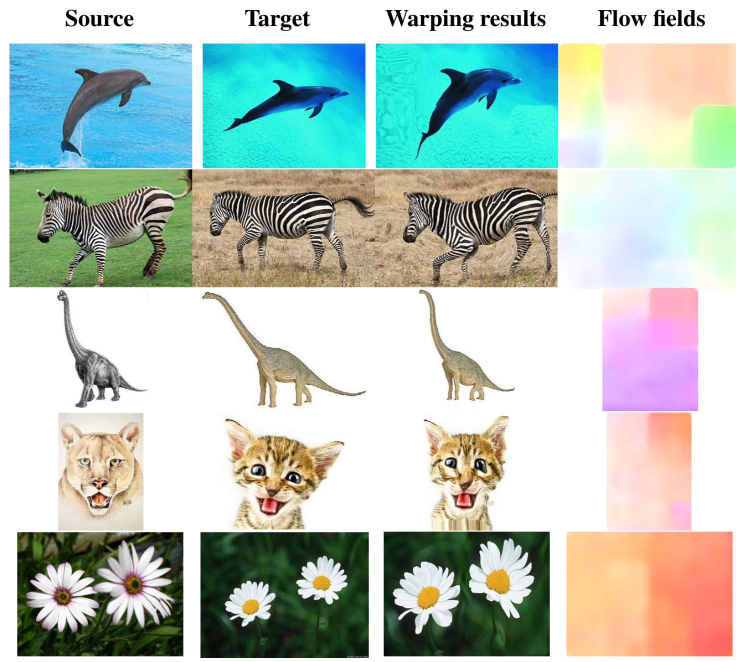


Figure 2: Backward warping results on the source images based on the obtained dense correspondence when non-rigid deformation exists. The more similar the warping result is to the source image, the more accurate the obtained dense correspondence field is.

Results: Experimental results on the public datasets and our own image collection indicate that our geometry-preserving smoothness shows its superiority when two images specifically share similar contents and lie under non-rigid deformation, as in Fig. 2.

Results on the Mikolajczyk *et al.* dataset in Table 1, which evaluates matching performances on scene alignment, reveals that our model best estimates dense correspondence fields under planar scale changes, rotation changes, and perspective transformation, comparing to state-of-the-arts. Results on the Moseg dataset and challenging non-rigid pairs from Caltech 101 dataset also show our better performance than other benchmarking algorithms in label-transfer metrics. The strength of our model comes from the high dimensional search, which includes rotation and scale variation while preserving the internal topology in images through the pyramid structure.

Scene	characteristic	**GDSP (Ours)**	DSP [1]	SIFT Flow [2]	DFF [5]	SSF [4]
Bikes	Blur	0.979	0.941	0.994	0.766	1.000
Trees	Blur	0.953	0.951	0.946	0.567	0.969
Graffiti	Viewpoint	0.503	0.033	0.238	0.242	0.521
Bricks	Viewpoint	0.771	0.230	0.491	0.465	0.829
Bark	rotation + scale	0.168	0.007	0.011	0.018	0.021
Boat	rotation + scale	0.312	0.003	0.006	0.150	0.002
Cars	Illumination	0.995	0.858	0.992	0.437	0.994
UBC	JPEG compression	0.998	0.969	0.897	0.753	0.980
Average Rank		**1.625**	**4.125**	**3.375**	**4.000**	**1.875**

Table 1: Percentages of correct match on Mikolajczyk *et al.* dataset [3].

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

[2] Ce Liu, Jenny Yuen, and Antonio Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE Trans. PAMI*, 33(5), 2011.

[3] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *IJCV*, 65(1-2), 2005.

[4] Weichao Qiu, Xinggang Wang, Xiang Bai, A Yuille, and Zhuowen Tu. Scale-space sift flow. In *WACV*, 2014.

[5] Hongsheng Yang, Wen-Yan Lin, and Jiangbo Lu. Daisy filter flow: A generalized discrete approach to dense correspondences. In *CVPR*, 2014.