DASC: Robust Dense Descriptor for Multi-modal and Multi-spectral Correspondence Estimation -Supplementary Materials-

Seungryong Kim, *Student Member, IEEE,* Dongbo Min, *Senior Member, IEEE,*
Bumsub Ham, *Member, IEEE,* Minh N. Do, *Fellow, IEEE,* and Kwanghoon Sohn, *Senior Member, IEEE*

In this supplemental materials, we provide more detailed analysis and results for the DASC descriptor.

- In section 1, we describe how Eq. (10) is derived from Eq. (9) in our paper.
- In section 2, we provide additional experimental results to evaluate the accuracy and runtime efficiency of the DASC descriptor when using the symmetric weight and the asymmetric weight, respectively.
- In section 3, we show the multi-modal and multi-spectral dataset used in the sampling pattern learning for the patch-wise receptive field pooling, and visualize the estimated sampling pattern.
- In section 4, we visualize non-maximal suppression in WMSD detector.
- In section 5, we analyze the effect of four parameters (support window size, descriptor dimension, patch size, and the number of log-point circular point) used in the DASC descriptor, and provide more results in three datasets; Middlebury stereo benchmark, multi-modal and multi-spectral image pairs, DIML multi-modal benchmark, and MPI SINTEL optical flow benchmark.

- *S. Kim and K. Sohn are with the School of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, Korea.*
E-mail: {srkim89, khsohn}@yonsei.ac.kr
- *D. Min is with the Department of Computer Science and Engineering, Chungnam National University, Daejeon 305-764, Korea.*
E-mail: dbmin@cnu.ac.kr
- *B. Ham is with Willow Team, INRIA, Paris 75013, France.*
E-mail: bumsub.ham@inria.fr
- *M. N. Do is with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. E-mail: minhdo@illinois.edu*

Manuscript received Apr. 27, 2016

1 Derivation of Decomposition (10) from (9)

In this section, we describe the derivation of Eq. (10) from Eq. (9) in our paper. For efficient description, we also re-arrange the sampling pattern $(s_{i,l}, t_{i,l})$ to referenced-biased pairs $(i, j) = (i, i + t_{i,l} - s_{i,l})$. $\Psi(i, j)$ is then approximated as follows:

$$\tilde{\Psi}(i, j) = \frac{\sum\limits_{i',j'} \omega_{i,i'}(f_{i'} - \mathcal{G}_i)(f_{j'} - \mathcal{G}_{i,j})}{\sqrt{\sum\limits_{i'} \omega_{i,i'}(f_{i'} - \mathcal{G}_i)^2} \sqrt{\sum\limits_{i',j'} \omega_{i,i'}(f_{j'} - \mathcal{G}_{i,j})^2}},$$

where $\mathcal{G}_i = \sum_{i'} \omega_{i,i'} f_{i'}$. Furthermore, $\mathcal{G}_{i,j} = \sum_{i',j'} \omega_{i,i'} f_{j'}$ which means weighted average of $f_{j'} \in \mathcal{F}_j$ with a guidance image $f_{i'} \in \mathcal{F}_i$. It is worth noting that the robustness of $\Psi(s, t)$ can be still applied to $\tilde{\Psi}(i, j)$ since their difference is just weight factors.

We then decompose numerator and denominator in $\tilde{\Psi}(i, j)$ after some arithmetic derivations. Firstly, the numerator in $\tilde{\Psi}(i, j)$ can be decomposed as

$$\sum_{i',j'} \omega_{i,i'} f_{i'} f_{j'} - \mathcal{G}_{i,j} \sum_{i'} \omega_{i,i'} f_{i'} - \mathcal{G}_i \sum_{i',j'} \omega_{i,i'} f_{j'} + \mathcal{G}_i \mathcal{G}_{i,j}.$$

Secondly, the denominator in $\tilde{\Psi}(i, j)$ can be decomposed as

$$\sqrt{\sum\nolimits_{i'} \omega_{i,i'} f_{i'}^2 - 2\mathcal{G}_i \sum\nolimits_{i'} \omega_{i,i'} f_{i'} + \mathcal{G}_i^2} \sqrt{\sum\nolimits_{i',j'} \omega_{i,i'} f_{j'}^2 - 2\mathcal{G}_{i,j} \sum\nolimits_{i',j'} \omega_{i,i'} f_{j'} + \mathcal{G}_{i,j}^2}.$$

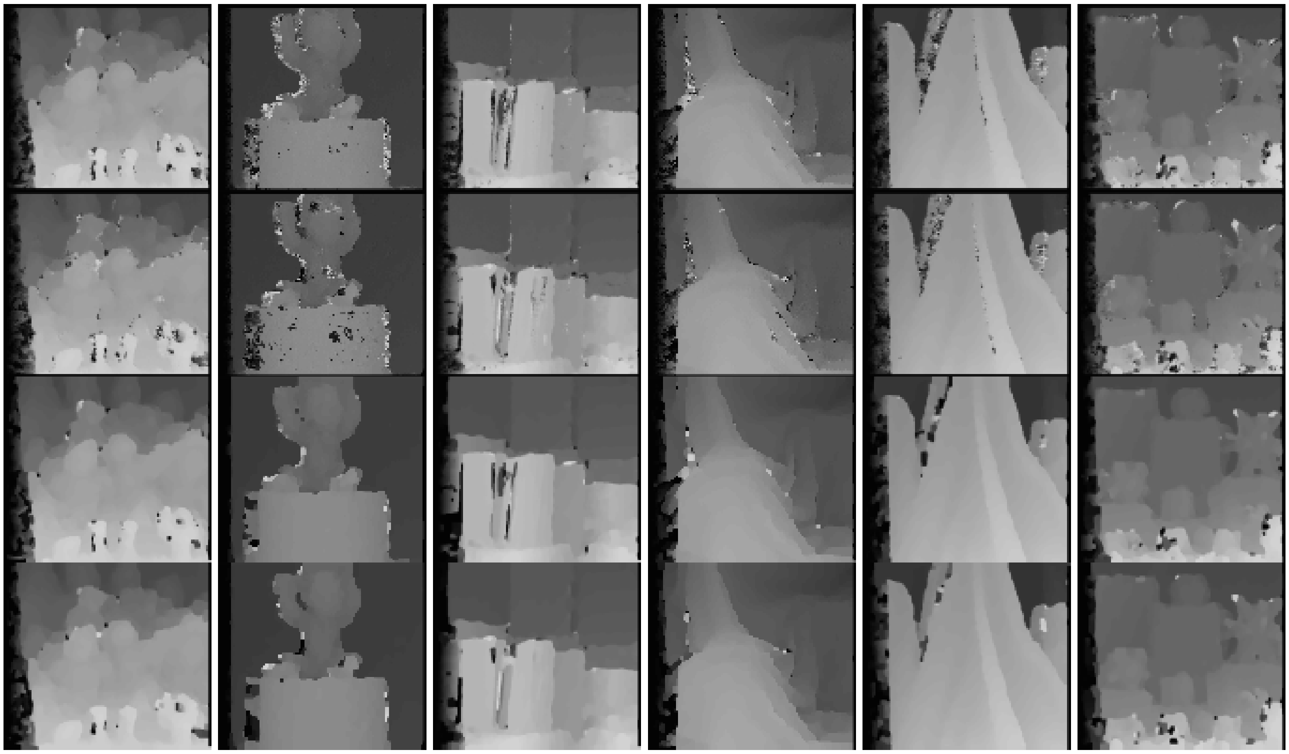
With these derivations, $\tilde{\Psi}(i, j)$ can be decomposed as

$$\frac{\mathcal{G}_{i,ij} - \mathcal{G}_i \cdot \mathcal{G}_{i,j}}{\sqrt{\mathcal{G}_{i^2} - \mathcal{G}_i^2} \cdot \sqrt{\mathcal{G}_{i,j^2} - \mathcal{G}_{i,j}^2}},$$

where $\mathcal{G}_{i^2} = \sum_{i'} \omega_{i,i'} f_{i'}^2$, $\mathcal{G}_{i,ij} = \sum_{i',j'} \omega_{i,i'} f_{i'} f_{j'}$, and $\mathcal{G}_{i,j^2} = \sum_{i',j'} \omega_{i,i'} f_{j'}^2$.

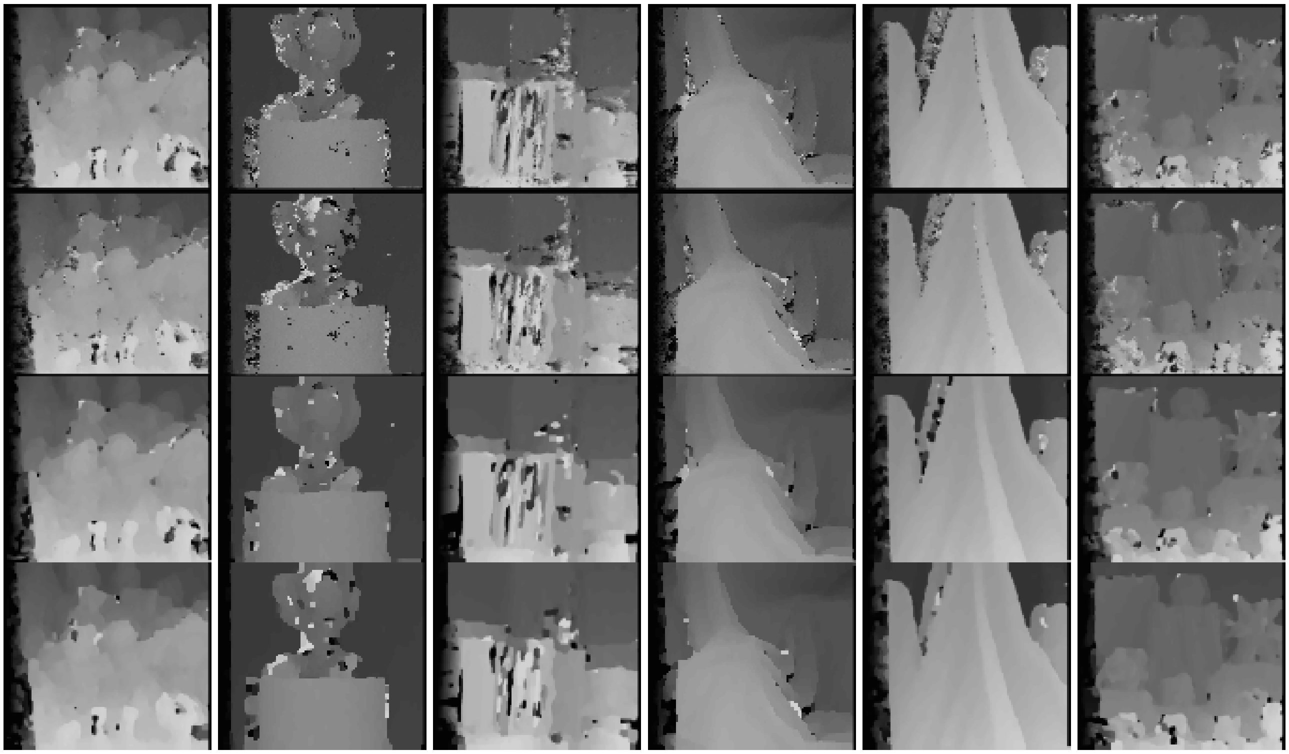
2 Performance evaluation between symmetric measure $\Psi(i, j)$ and asymmetric measure $\tilde{\Psi}(i, j)$ in the DASC descriptor

This section provides qualitative performance evaluation between symmetric measure $\Psi(i, j)$ and asymmetric measure $\tilde{\Psi}(i, j)$ in the DASC descriptor described in Sec. 6.2.5 of the paper. Fig. 1 and Fig. 2 show the comparison of the disparity estimation for *Dolls*, *Baby1*, *Books*, *Cloth3*, *Cloth4*, and *Moebius* image pairs taken under exposure combination '0/0' and '0/2', respectively. As shown in Fig. 1 and Fig. 2, a performance gap between using the asymmetric measure $\tilde{\Psi}(i, j)$ and the symmetric measure $\Psi(i, j)$ in the DASC descriptor is negligible, while using the asymmetric measure is much faster.



(a) Dolls (b) Baby1 (c) Books (d) Cloth3 (e) Cloth4 (f) Moebius

Fig. 1. Comparison of the disparity estimation for *Dolls*, *Baby1*, *Books*, *Cloth3*, *Cloth4*, and *Moebius* image pairs taken under exposure combination '0/0'. The first two rows shows the disparity maps obtained using the DASC descriptor with asymmetric and symmetric weights, where the winner-takes-all (WTA) method is used for optimization. The third and fourth rows shows the disparity maps, where the Graph Cuts (GC) method is used for optimization.



(a) Dolls (b) Baby1 (c) Books (d) Cloth3 (e) Cloth4 (f) Moebius

Fig. 2. Comparison of the disparity estimation for *Dolls*, *Baby1*, *Books*, *Cloth3*, *Cloth4*, and *Moebius* image pairs taken under exposure combination '0/2'. The first two rows shows the disparity maps obtained using the DASC descriptor with asymmetric and symmetric weights, where the winner-takes-all (WTA) method is used for optimization. The third and fourth rows shows the disparity maps, where the Graph Cuts (GC) method is used for optimization.

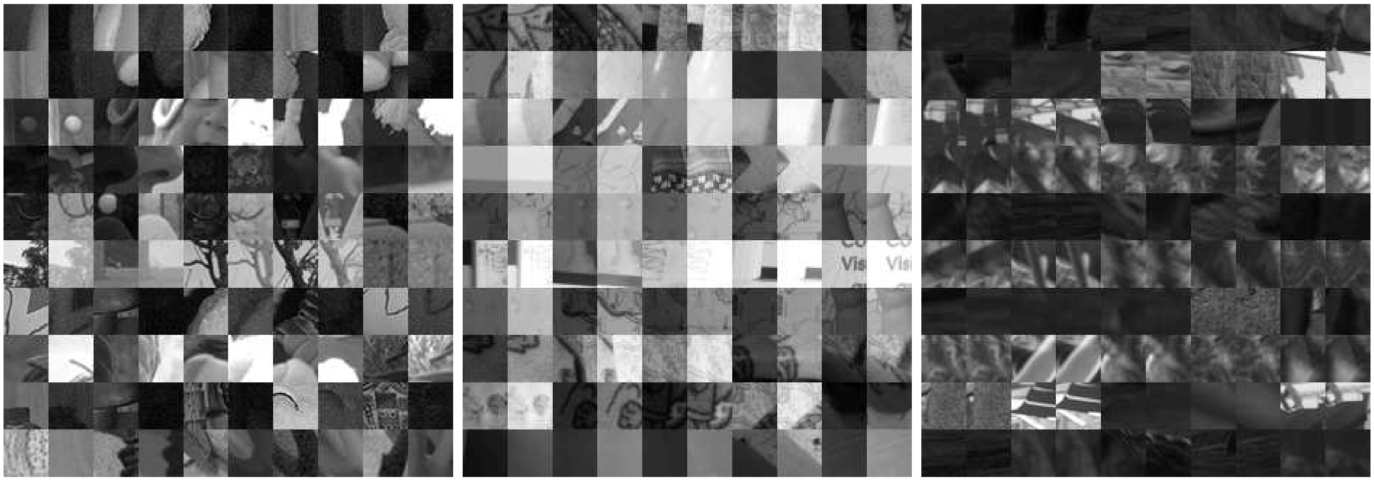
Table 1 reports the computational complexity of the DASC descriptor for the brute-force implementation and the proposed efficient implementation when using the symmetric and asymmetric weights. The DASC descriptor with asymmetric weights provides a low computational complexity thanks to its efficient computational framework.

image size	SIFT [1]	DAISY [2]	LSS [3]	DASC* w/ sym.	DASC†w/ sym.	DASC* w/ asym.	DASC†w/ asym.
463×370	$130.3s$	$2.5s$	$31s$	$197.2s$	$9s$	$128s$	$5s$

TABLE 1
Evaluation of the computational complexity. The brute-force and efficient implementation of the DASC is denoted as * and †, respectively. However, the DASC descriptor with symmetric weights need more computational load compared to that of asymmetric weights.

3 Multi-modal and Multi-spectral Feature Learning

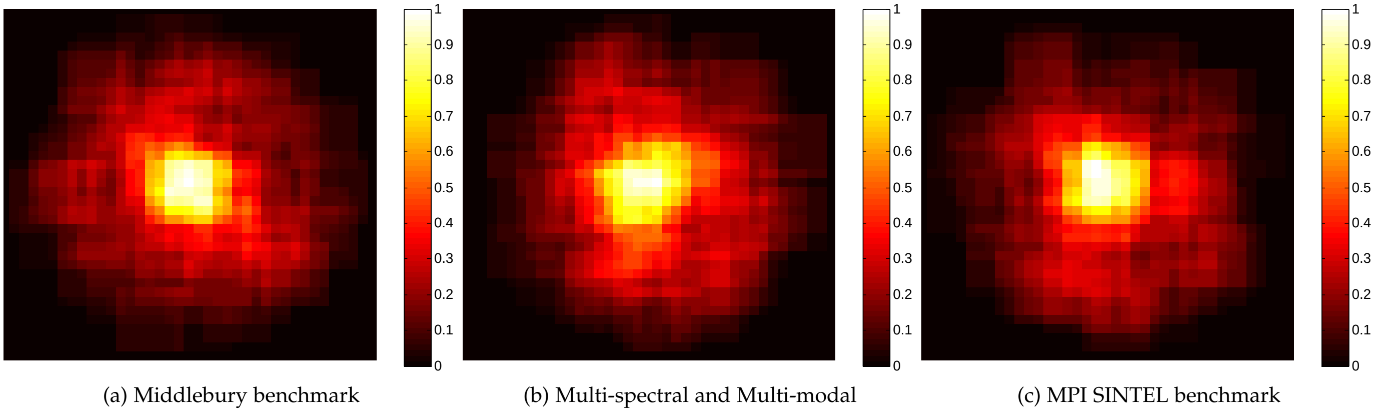
In this section we provide an example of training pairs, denoted as $\mathcal{P} = \{(\mathcal{R}^1_m, \mathcal{R}^2_m, y_m)|m = 1, ..., N_t\}$, used in the sampling pattern learning where $(\mathcal{R}^1, \mathcal{R}^2)$ are support window pairs, and N_t is the number of training samples. y is a binary label that becomes 1 if two patches are matched or 0 otherwise. The training data set $\mathcal{P}$ was built from ground truth correspondence maps for images captured under varying illumination conditions and/or with imaging devices [4], [5]. It should be noted that since multi-modal and multi-spectral pairs do not have a ground truth dense correspondence, we manually obtained ground truth displacement vectors [6]. In our experiments, we first established $50,000$ multi-spectral and multi-modal support window pairs, as shown in Fig. 3. Among them, $5,000$ matching support window pairs (positive samples, *i.e.*, $y_m = 1$) were randomly selected from true matching pairs, while $5,000$ non-matching support window pairs (negative samples, *i.e.*, $y_m = 0$) were made by randomly selecting two support windows from different matching pairs. Thus, in total, $N_t = 10,000$ training support window pairs were built. In experiments, each training set is mutually used to learn a sampling pattern. Specifically, the sampling pattern for Middlebury benchmark data set is learned from the multi-spectral and multi-modal benchmark. In a similar way, the sampling patterns for multi-modal and multi-spectral benchmark and MPI SINTEL benchmark are learned from MPI SINTEL benchmark and multi-modal and multi-spectral benchmark, respectively.



(a) Middlebury benchmark (b) Multi-spectral and Multi-modal (c) MPI SINTEL benchmark

Fig. 3. Some examples of 50,000 support window training pairs built from Middlebury stereo benchmark, multi-spectral and multi-modal benchmark, and MPI optical flow benchmark.

Fig. 4 shows patch-wise receptive fields on learned sampling patterns used in our DASC descriptor. For an effective visualization, we followed the practice used in [7]. We stacked all patch-wise receptive fields learnt from the Middlebury stereo benchmark [4], the multi-modal and multi-spectral benchmark [6], [8], [9], [10], [11], and the MPI SINTEL benchmark [5], respectively. A set of histogram bins corresponding to the patch of each patch-wise receptive field are incremented by one, and they are finally normalized with the maximum value. The density of patch-wise receptive fields tends to be concentrated on the center. In many literature, it has been shown that such a center-biased density distribution pooling in the local feature provides the robustness [7], [12].



(a) Middlebury benchmark

(b) Multi-spectral and Multi-modal

(c) MPI SINTEL benchmark

Fig. 4. Visualization of patch-wise receptive fields of the DASC descriptor which are learned from Middlebury benchmark, multi-spectral and multi-modal benchmark, and MPI SINTEL benchmark.

4 Non-maximal Suppression in WMSD

In this section, we visualize the non-maximal suppression scheme in weighted maximal self-dissimilarity (WMSD). For feature response maps $\mathbf{\Omega}_i = \{\Omega_i^k\}$, the local maxima are obtained by the non maximal suppression as shown in Fig. 6, which compares Ω_i^k to its 8 neighbors on the current scale and 18 neighbors on the $(k+1)^{th}$ and $(k-1)^{th}$ scales. Similar to SIFT [1], a feature point $i \in \mathcal{I}'$ is detected only if $\{\Omega_i^k\}$ has an extreme value compared to all of these neighbors, and its scale ρ_i is defined with ρ_k. $\mathcal{I}' \subset \mathcal{I}$ is a sparse discrete image domain.

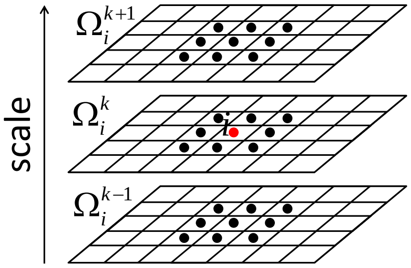


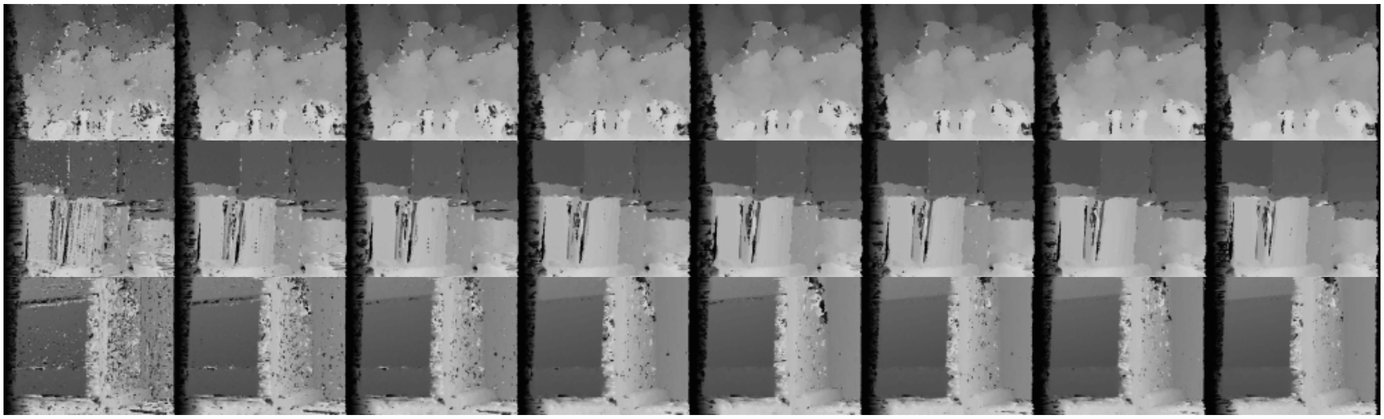
Fig. 5. Feature detection with corresponding scales using the WMSD detector. For feature response maps $\mathbf{\Omega} = \{\Omega^k\}$, the extreme point is detected by comparing to its 26 neighbours in 3×3 for each pixel.

5 More Results

In this section, we first analyze the effects of the support window size and the feature dimension in our DASC descriptor in Sec. 5.1. Then, we provide the additional results for our DASC descriptor and state-of-the-art descriptor-based methods and area-based methods using the Middlebury stereo benchmark in Sec. 5.2, the multi-modal and multi-spectral image pair benchmark in Sec. 5.3, and the MPI SINTEL optical flow benchmark in Sec. 5.5. Furthermore, we provide the additional qualitative results for our DASC and GI-DASC descriptor and state-of-the-art geometry robust methods on DIML multi-modal benchmark in Sec. 5.4. Finally, we conduct the additional results for multi-spectral RGB-NIR image pairs under geometric variations in Sec. 5.6.

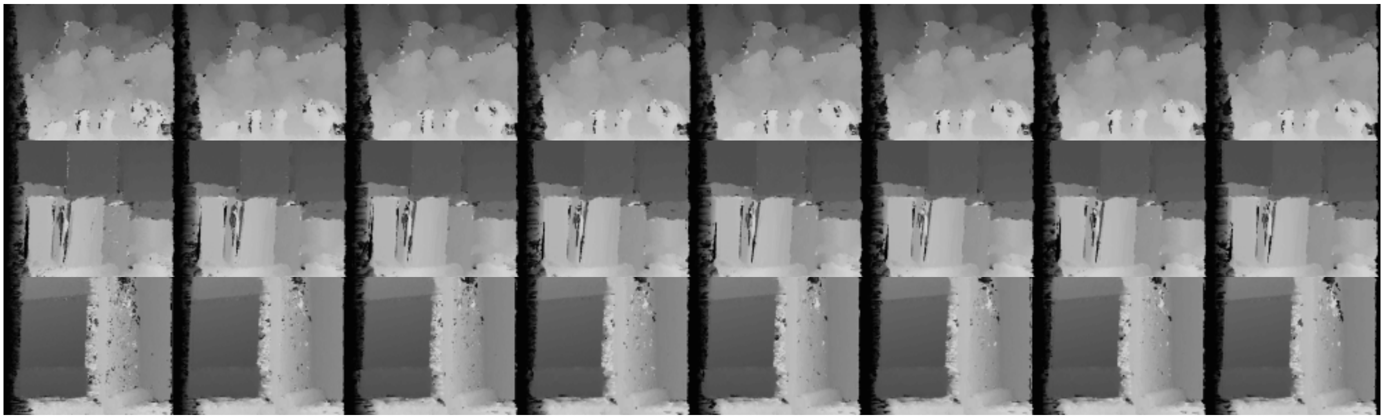
5.1 Parameter Sensitivity Analysis

Fig. 6, Fig. 7, Fig. 8, and Fig. 9 intensively analyzed the performance of the DASC descriptor as varying associated parameters, including support window size M, descriptor dimension L^{dasc}, patch size N, and the number of log-point circular point N_c.



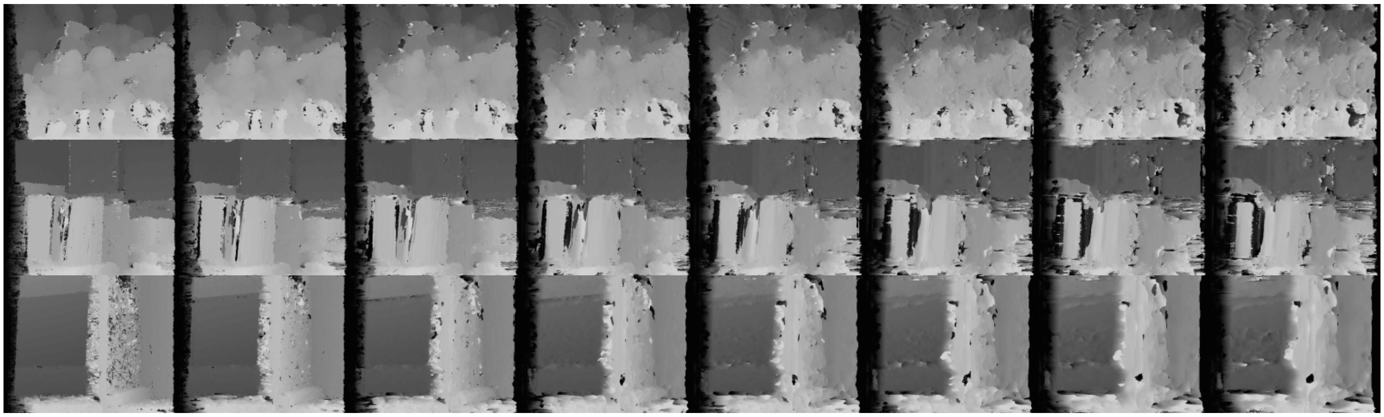
(a) 5×5 (b) 9×9 (c) 13×13 (d) 17×17 (e) 21×21 (g) 25×25 (e) 29×29 (g) 33×33

Fig. 6. Results of disparity estimation for *Dolls* and *Wood1* image pairs taken under exposure combination ‘0/1’ by varying the support window size M in the DASC descriptor. In our work, we used $M = 31 \times 31$ as the size of support window.



(a) 50 dim. (b) 100 dim. (c) 150 dim. (d) 200 dim. (e) 250 dim. (g) 300 dim. (e) 350 dim. (g) 400 dim.

Fig. 7. Results of disparity estimation for *Dolls* and *Wood1* image pairs taken under exposure combination ‘0/1’ by varying the descriptor dimension L^{dasc} in the DASC descriptor. In our work, we used $L^{\mathrm{dasc}} = 128$ as the length of descriptor dimension.



(a) 3×3 (b) 5×5 (c) 7×7 (d) 9×9 (e) 11×11 (g) 13×13 (e) 15×15 (g) 17×17

Fig. 8. Results of disparity estimation for *Dolls* and *Wood1* image pairs taken under exposure combination ‘0/1’ by varying the descriptor dimension N in the DASC descriptor. In our work, we used $N = 5 \times 5$ as the length of descriptor dimension.

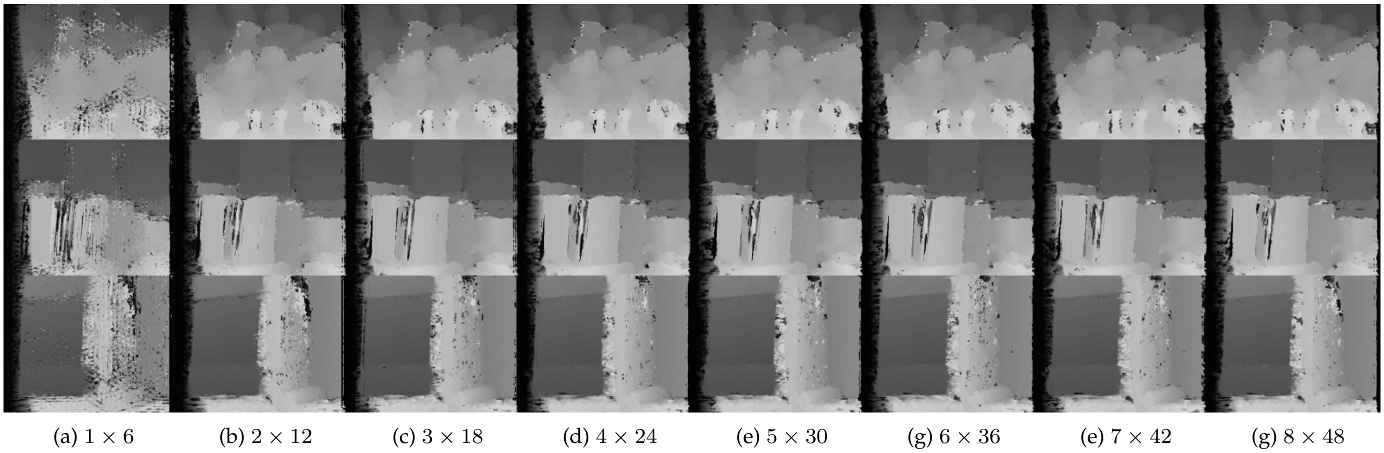


Fig. 9. Results of disparity estimation for *Dolls* and *Wood1* image pairs taken under exposure combination ‘0/1’ by varying the descriptor dimension N_c in the DASC descriptor. In our work, we used $N_c = 4 \times 36$ as the length of descriptor dimension.

5.2 Middlebury Stereo Benchmark

In Middlebury stereo benchmark, we used the *Art*, ***Baby1***, ***Books***, *Bowling2*, *Cloth3*, *Cloth4*, ***Dolls***, ***Moebius***, ***Reindeer***, and *Wood1*. In this supplementary materials, the results for **bold** image pairs are shown. Fig. 10 compare the disparity maps estimated for stereo image pairs taken with an exposure combination '0/2'.

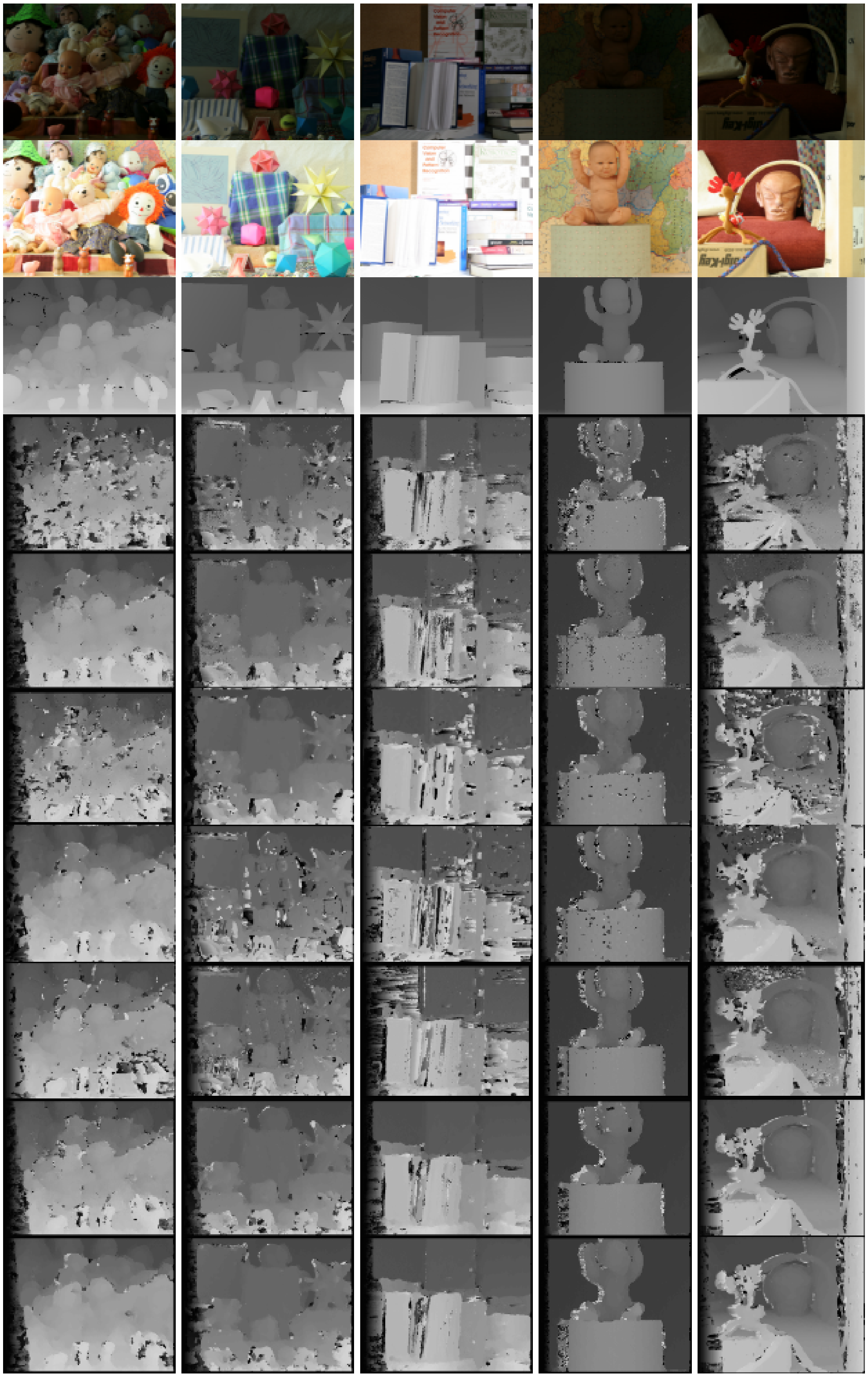


Fig. 10. Comparison of disparity estimation for *Dolls*, *Moebius*, *Books*, *Baby1*, and *Reindeer* image pairs taken under exposure combination 0/2. (from left to right, top and bottom) Left color image, right color image, and disparity maps for the ground truth, ANCC [13], BRIEF [14], DAISY [2], SIFT [1], LSS [3], DASC+RP, and DASC+LRP.

5.3 Multi-modal and Multi-spectral Image Pairs

In experiments, the multi-modal and multi-spectral image pairs consist of RGB-NIR images, flash-noflash images, images taken under different exposures, and blurred-clean images.

- RGB-NIR image pairs: *epfl1*, *epfl2*, *epfl3*, *epfl4*, *epfl5*, ***epfl6***, ***lion***, *myrgbnir*, ***orchid***, *stereo3*, and *stereo4*.
- Flash-noflash image pairs: ***Dolls1***, ***Dolls2***, and ***Dolls3***.
- Image pairs taken under different exposures: *altar*, *BabyAtWindow*, *BabyOnGrass*, ***balcony***, *books*, *ChristmasRider*, *clouds*, *FeedingTime*, *flower*, *HighChair*, *LadyEating*, ***lantern***, *mpi*, *PianoMan*, ***room***, *SantasLittleHelper*, *street*, and *window*.
- Blurred-clean image pairs: *avisar*, ***books1***, *books2*, ***cars1***, *cars2*, *children*, ***face1***, *face2*, *flowers*, , *numbers*, and *yemin*.

In this supplementary materials, the results for **bold** image pairs are shown. Fig. 11, Fig. 12, Fig. 13, and Fig. 14 show the warped color image and its corresponding 2-D flow fields for multi-modal and multi-spectral image pairs. For the results of objective comparison, please refer to Table 2 in our paper.

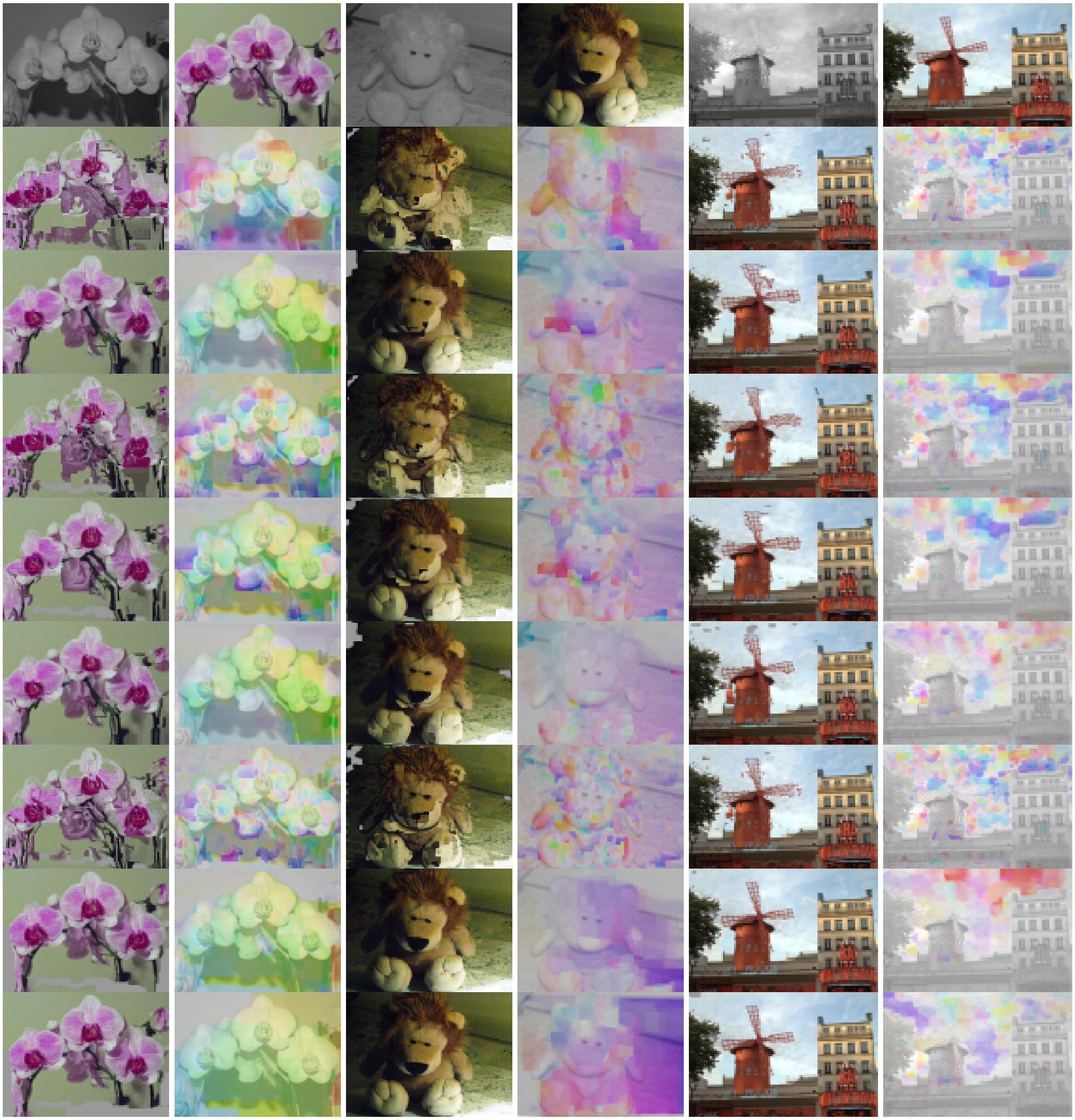


Fig. 11. Comparison of dense correspondence for RGB-NIR images including *orchid*, *lion*, and *epfl6*. (from top to bottom) Input image pairs, MI+SIFT [13], RSNCC [6], BRIEF [14], DAISY [2], SIFT [1], LSS [3], DASC+RP, and DASC+LRP.

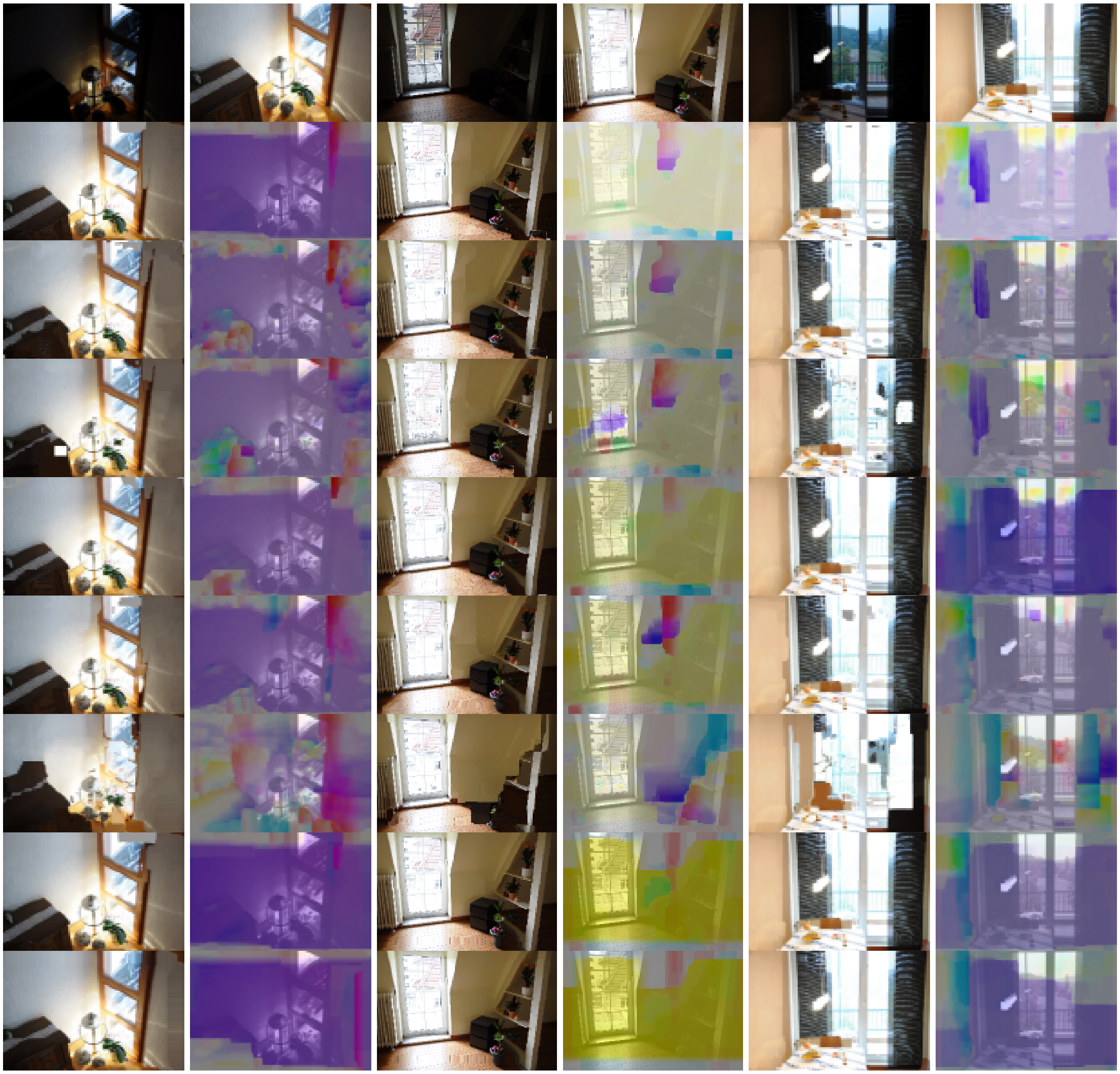


Fig. 12. Comparison of dense correspondence for different exposure images including *lantern*, *balcony*, and *room*. (from top to bottom) Input image pairs, MI+SIFT [13], RSNCC [6], BRIEF [14], DAISY [2], SIFT [1], LSS [3], DASC+RP, and DASC+LRP.

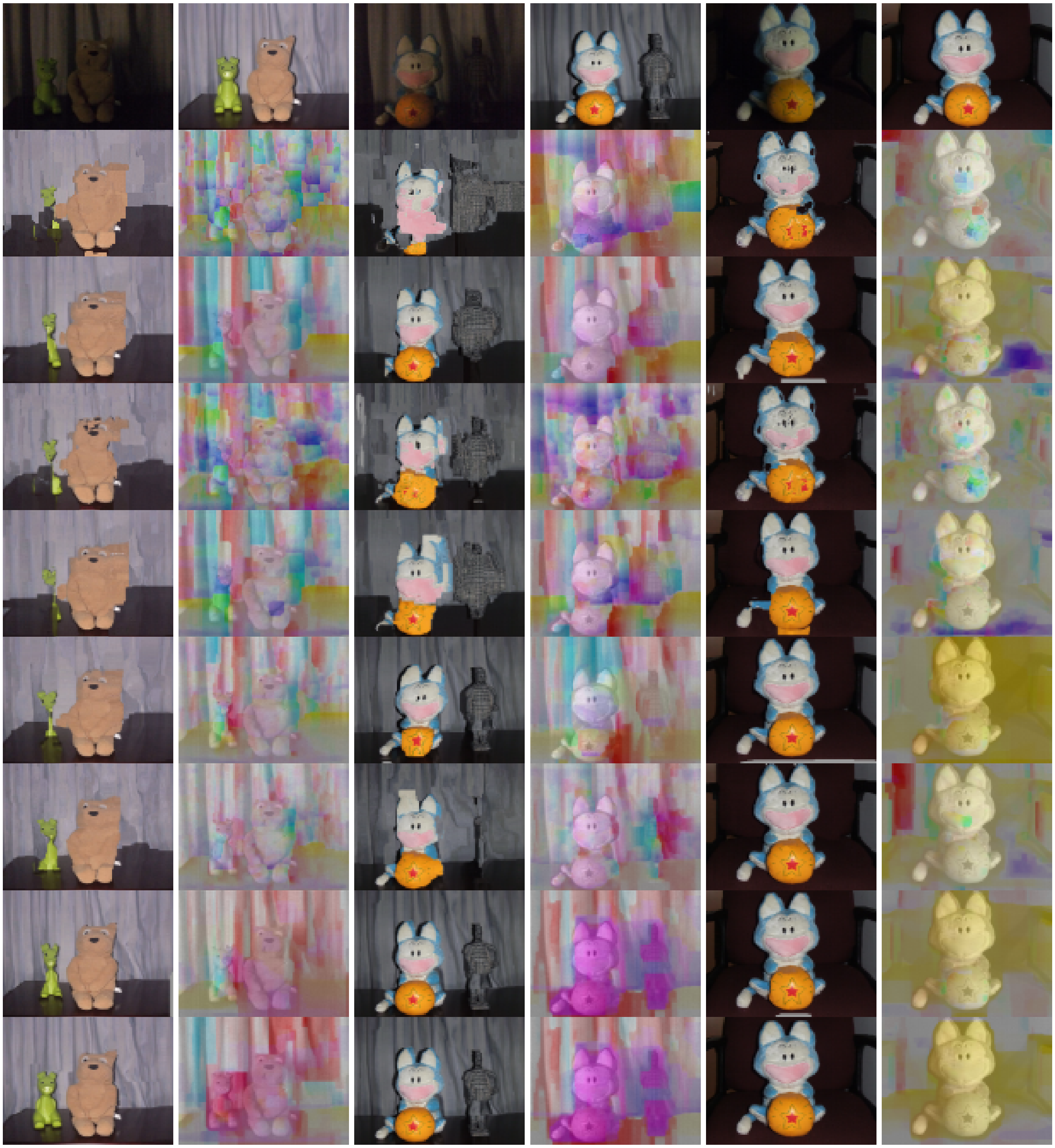


Fig. 13. Comparison of dense correspondence for flash-noflash images including *Dolls1*, *Dolls2*, and *Dolls3*. (from top to bottom) Input image pairs, MI+SIFT [13], RSNCC [6], BRIEF [14], DAISY [2], SIFT [1], LSS [3], DASC+RP, and DASC+LRP.

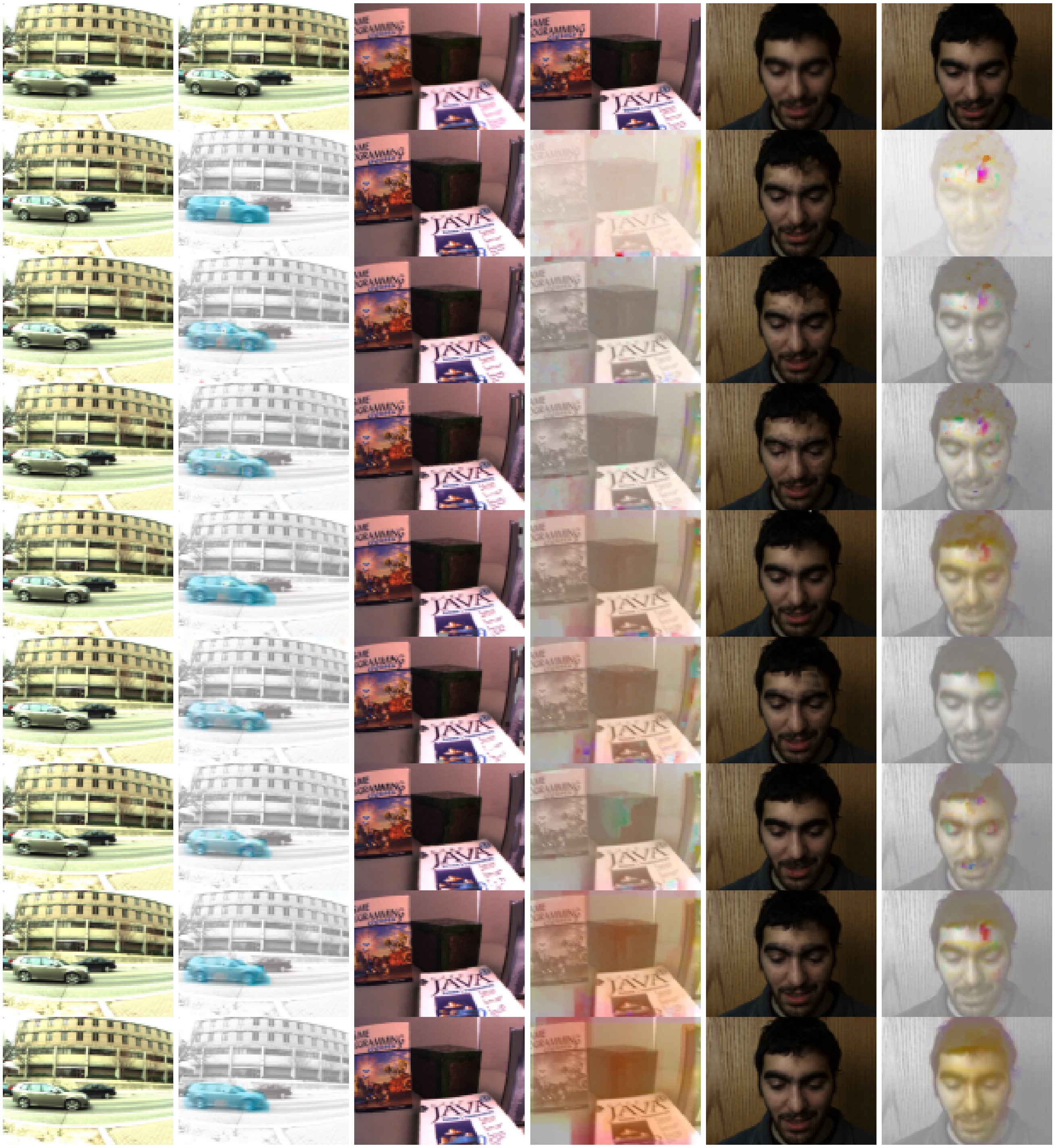


Fig. 14. Comparison of dense correspondence for blurred images *cars1*, *books1*, and *face1*. (from top to bottom) Input image pairs, MI+SIFT [13], RSNCC [6], BRIEF [14], DAISY [2], SIFT [1], LSS [3], DASC+RP, and DASC+LRP.

5.4 DIML Multi-modal Benchmark

Since there have been no database with both photometric and geometric variations, we built the DIML multi-modal benchmark. All databases were taken by SONY Cyber-Shot DSC-RX100 camera in a darkroom with the lighting booth GretagMacbeth SpectraLight III. In terms of geometric deformations, we captured 10 geometry image sets by combining geometric variations of viewpoint, scale, and rotation, and each image set consists of images taken under 5 different photometric variation pairs including illumination, exposure, flash-noflash, blur, and noise. Therefore, the DIML multi-modal benchmark consists of 100 images with the size of 1200×800. Furthermore, by following [15], we manually built ground truth object annotation maps to evaluate the performance quantitatively, and computed the label transfer accuracy (LTA) $\mathcal{A}^{\mathrm{LTA}}$ such that

$$\mathcal{A}^{\mathrm{LTA}} = \frac{1}{\mathcal{T}_a} \sum\nolimits_{i \in \mathcal{I}} 1(e_i \neq a_i, a_i > 0) \tag{1}$$

where the ground-truth annotation is a_i, estimated annotation is e_i, and $\mathcal{T}_a = \sum\nolimits_{i \in \mathcal{I}} 1(a_i > 0)$ is the number of labeled pixels. This metric has been widely used in wide-baseline matching tasks [16]. Though $\mathcal{A}^{\mathrm{LTA}}$ does not measure a matching performance in a pixel precision, it was shown in [15] that this metric is an excellent alternative that has a discriminative power enough to evaluate the performance of descriptors in case that there are no ground truth correspondence maps available. For one image from the reference geometry image set, we estimated visual correspondence maps with images from other geometry image set, and then computed the LTA. Furthermore, visual correspondence maps were estimated for each photometric pair. Here, matching results at occluded pixels should be excluded in the evaluation as they have no corresponding pixels. We hence warped an image taken from near into an image taken at a distance, when computing the LTA.

Fig. 15, Fig. 16, Fig. 17, and Fig. 18 show qualitative evaluation results on DIML multi-modal benchmark. Fig. 19 shows the LTA error rates as varying photometric and geometric deformations.

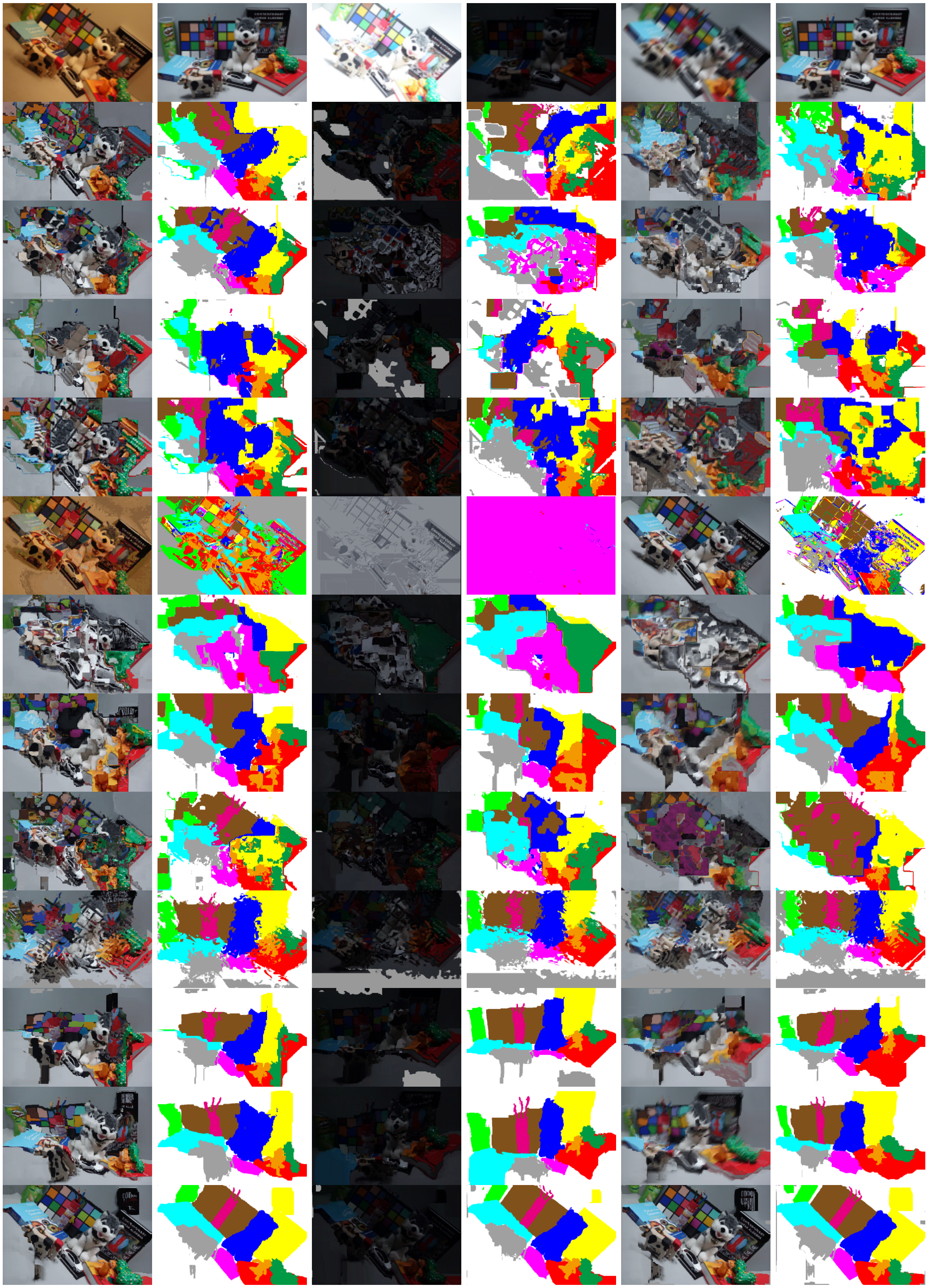


Fig. 15. Comparison of qualitative evaluation for photometric and geometric variations on DIML multi-modal benchmark [23] (img s0r1v0). The results consist of warped color images and warped ground truth annotations.

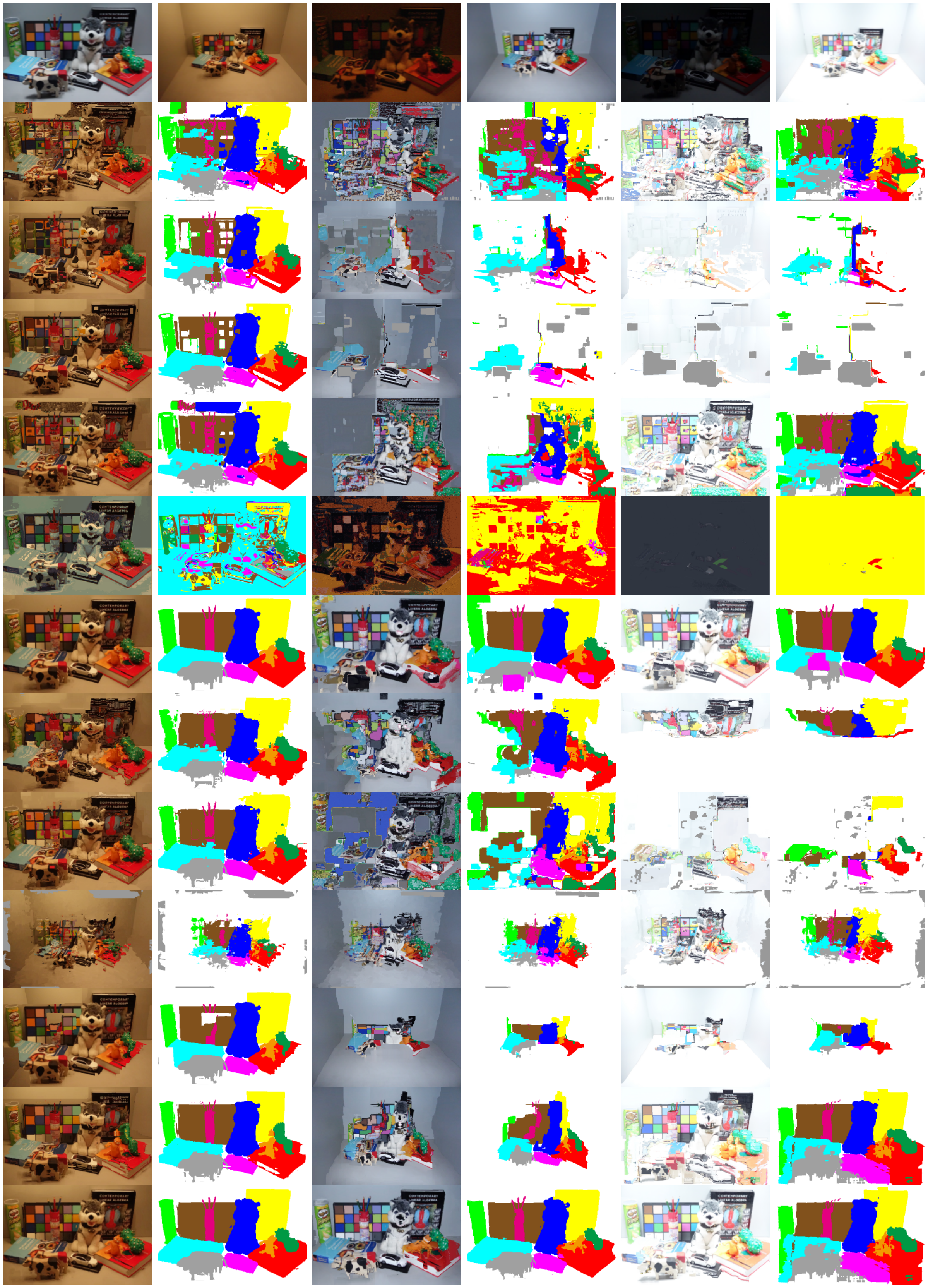


Fig. 16. Comparison of qualitative evaluation for photometric and geometric variations on DIML multi-modal benchmark [23] (img s1r0v0). The results consist of warped color images and warped ground truth annotations.

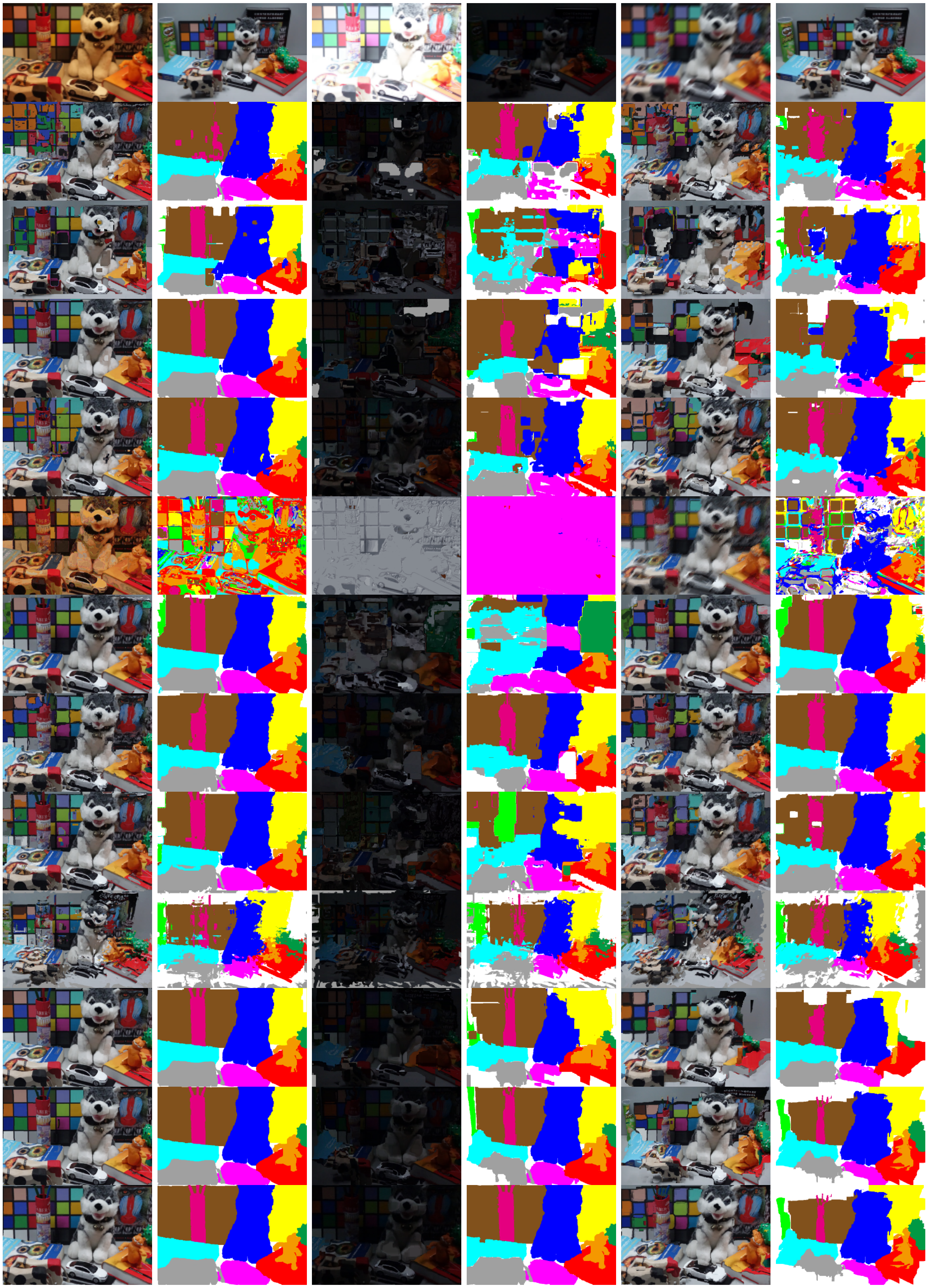


Fig. 17. Comparison of qualitative evaluation for photometric and geometric variations on DIML multi-modal benchmark [23] (img s2r0v0). The results consist of warped color images and warped ground truth annotations.

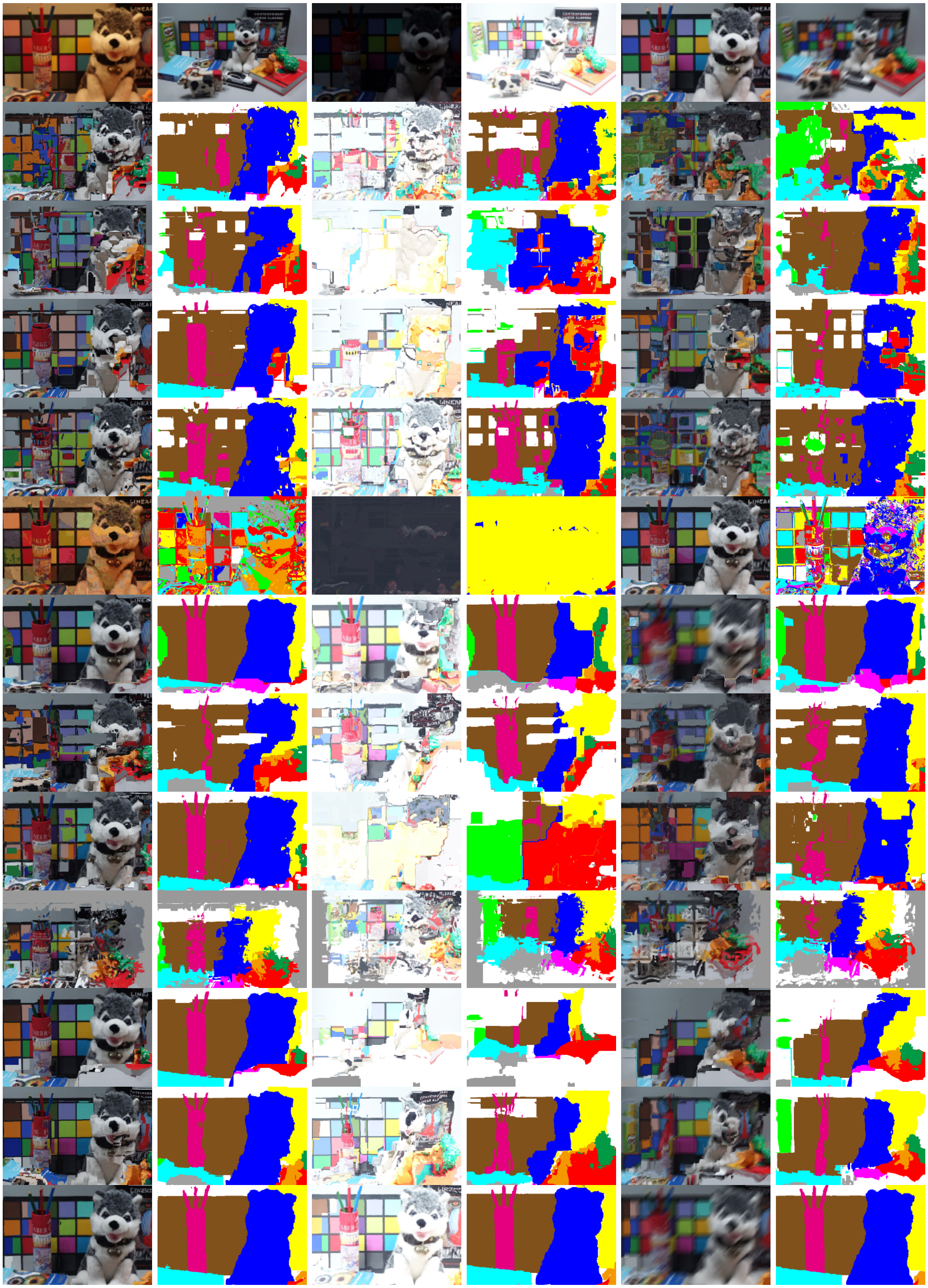
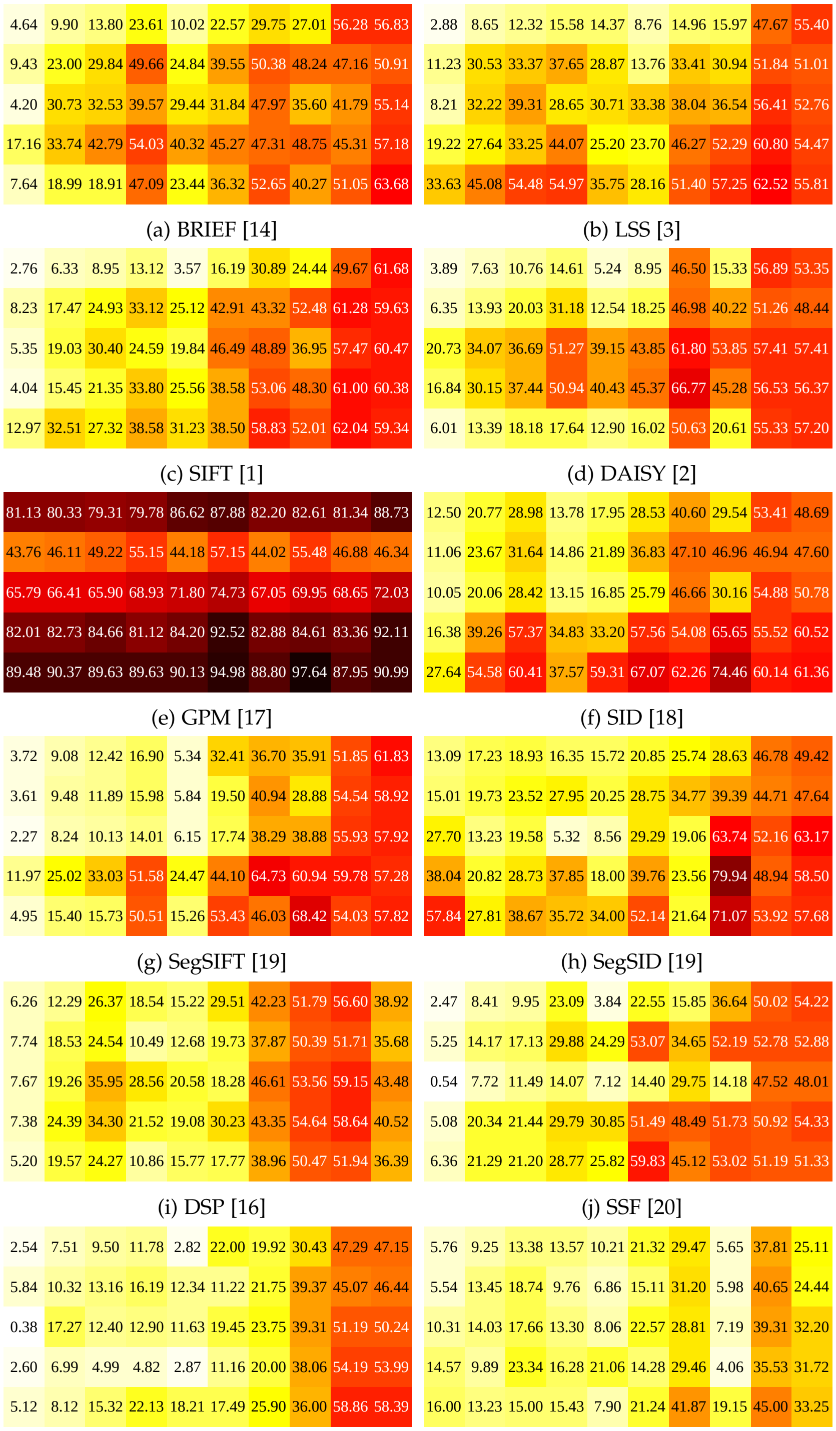


Fig. 18. Comparison of qualitative evaluation for photometric and geometric variations on DIML multi-modal benchmark [23] (img s3r0v0). The results consist of warped color images and warped ground truth annotations.



4.64	9.90	13.80	23.61	10.02	22.57	29.75	27.01	56.28	56.83
9.43	23.00	29.84	49.66	24.84	39.55	50.38	48.24	47.16	50.91
4.20	30.73	32.53	39.57	29.44	31.84	47.97	35.60	41.79	55.14
17.16	33.74	42.79	54.03	40.32	45.27	47.31	48.75	45.31	57.18
7.64	18.99	18.91	47.09	23.44	36.32	52.65	40.27	51.05	63.68

(a) BRIEF [14]

2.88	8.65	12.32	15.58	14.37	8.76	14.96	15.97	47.67	55.40
11.23	30.53	33.37	37.65	28.87	13.76	33.41	30.94	51.84	51.01
8.21	32.22	39.31	28.65	30.71	33.38	38.04	36.54	56.41	52.76
19.22	27.64	33.25	44.07	25.20	23.70	46.27	52.29	60.80	54.47
33.63	45.08	54.48	54.97	35.75	28.16	51.40	57.25	62.52	55.81

(b) LSS [3]

2.76	6.33	8.95	13.12	3.57	16.19	30.89	24.44	49.67	61.68
8.23	17.47	24.93	33.12	25.12	42.91	43.32	52.48	61.28	59.63
5.35	19.03	30.40	24.59	19.84	46.49	48.89	36.95	57.47	60.47
4.04	15.45	21.35	33.80	25.56	38.58	53.06	48.30	61.00	60.38
12.97	32.51	27.32	38.58	31.23	38.50	58.83	52.01	62.04	59.34

(c) SIFT [1]

3.89	7.63	10.76	14.61	5.24	8.95	46.50	15.33	56.89	53.35
6.35	13.93	20.03	31.18	12.54	18.25	46.98	40.22	51.26	48.44
20.73	34.07	36.69	51.27	39.15	43.85	61.80	53.85	57.41	57.41
16.84	30.15	37.44	50.94	40.43	45.37	66.77	45.28	56.53	56.37
6.01	13.39	18.18	17.64	12.90	16.02	50.63	20.61	55.33	57.20

(d) DAISY [2]

81.13	80.33	79.31	79.78	86.62	87.88	82.20	82.61	81.34	88.73
43.76	46.11	49.22	55.15	44.18	57.15	44.02	55.48	46.88	46.34
65.79	66.41	65.90	68.93	71.80	74.73	67.05	69.95	68.65	72.03
82.01	82.73	84.66	81.12	84.20	92.52	82.88	84.61	83.36	92.11
89.48	90.37	89.63	89.63	90.13	94.98	88.80	97.64	87.95	90.99

(e) GPM [17]

12.50	20.77	28.98	13.78	17.95	28.53	40.60	29.54	53.41	48.69
11.06	23.67	31.64	14.86	21.89	36.83	47.10	46.96	46.94	47.60
10.05	20.06	28.42	13.15	16.85	25.79	46.66	30.16	54.88	50.78
16.38	39.26	57.37	34.83	33.20	57.56	54.08	65.65	55.52	60.52
27.64	54.58	60.41	37.57	59.31	67.07	62.26	74.46	60.14	61.36

(f) SID [18]

3.72	9.08	12.42	16.90	5.34	32.41	36.70	35.91	51.85	61.83
3.61	9.48	11.89	15.98	5.84	19.50	40.94	28.88	54.54	58.92
2.27	8.24	10.13	14.01	6.15	17.74	38.29	38.88	55.93	57.92
11.97	25.02	33.03	51.58	24.47	44.10	64.73	60.94	59.78	57.28
4.95	15.40	15.73	50.51	15.26	53.43	46.03	68.42	54.03	57.82

(g) SegSIFT [19]

13.09	17.23	18.93	16.35	15.72	20.85	25.74	28.63	46.78	49.42
15.01	19.73	23.52	27.95	20.25	28.75	34.77	39.39	44.71	47.64
27.70	13.23	19.58	5.32	8.56	29.29	19.06	63.74	52.16	63.17
38.04	20.82	28.73	37.85	18.00	39.76	23.56	79.94	48.94	58.50
57.84	27.81	38.67	35.72	34.00	52.14	21.64	71.07	53.92	57.68

(h) SegSID [19]

6.26	12.29	26.37	18.54	15.22	29.51	42.23	51.79	56.60	38.92
7.74	18.53	24.54	10.49	12.68	19.73	37.87	50.39	51.71	35.68
7.67	19.26	35.95	28.56	20.58	18.28	46.61	53.56	59.15	43.48
7.38	24.39	34.30	21.52	19.08	30.23	43.35	54.64	58.64	40.52
5.20	19.57	24.27	10.86	15.77	17.77	38.96	50.47	51.94	36.39

(i) DSP [16]

2.47	8.41	9.95	23.09	3.84	22.55	15.85	36.64	50.02	54.22
5.25	14.17	17.13	29.88	24.29	53.07	34.65	52.19	52.78	52.88
0.54	7.72	11.49	14.07	7.12	14.40	29.75	14.18	47.52	48.01
5.08	20.34	21.44	29.79	30.85	51.49	48.49	51.73	50.92	54.33
6.36	21.29	21.20	28.77	25.82	59.83	45.12	53.02	51.19	51.33

(j) SSF [20]

2.54	7.51	9.50	11.78	2.82	22.00	19.92	30.43	47.29	47.15
5.84	10.32	13.16	16.19	12.34	11.22	21.75	39.37	45.07	46.44
0.38	17.27	12.40	12.90	11.63	19.45	23.75	39.31	51.19	50.24
2.60	6.99	4.99	4.82	2.87	11.16	20.00	38.06	54.19	53.99
5.12	8.12	15.32	22.13	18.21	17.49	25.90	36.00	58.86	58.39

(k) DASC

5.76	9.25	13.38	13.57	10.21	21.32	29.47	5.65	37.81	25.11
5.54	13.45	18.74	9.76	6.86	15.11	31.20	5.98	40.65	24.44
10.31	14.03	17.66	13.30	8.06	22.57	28.81	7.19	39.31	32.20
14.57	9.89	23.34	16.28	21.06	14.28	29.46	4.06	35.53	31.72
16.00	13.23	15.00	15.43	7.90	21.24	41.87	19.15	45.00	33.25

(l) GI-DASC

Fig. 19. Comparison of quantitative evaluation on DIML multi-modal benchmark. Each result represents the LTA in (1) for geometric (x-axis) and photometric (y-axis) variations, respectively. The DASC outperforms conventional descriptors such as SIFT [1], DAISY [2], BRIEF [14], and LSS [3]. Interestingly, its accuracy is also higher than those of state-of-the-art geometry-invariant approaches including SID [18], SegSIFT [19], SegSID [19], GPM [17], DSP [16], and SSF [20]. The GI-DASC descriptor shows the best performance under varying photometric and geometric conditions.

5.5 MPI SINTEL Optical Flow Benchmark

In MPI SINTEL optical flow benchmark, the dataset consists of two kind of rendering frames, namely *clean pass* and *final pass*, each containing 12 sequences with over 500 frames in total [5]. Fig. 20 shows visual comparison on the MPI SINTEL benchmark, where the warped color image and its corresponding 2-D flow fields are depicted.

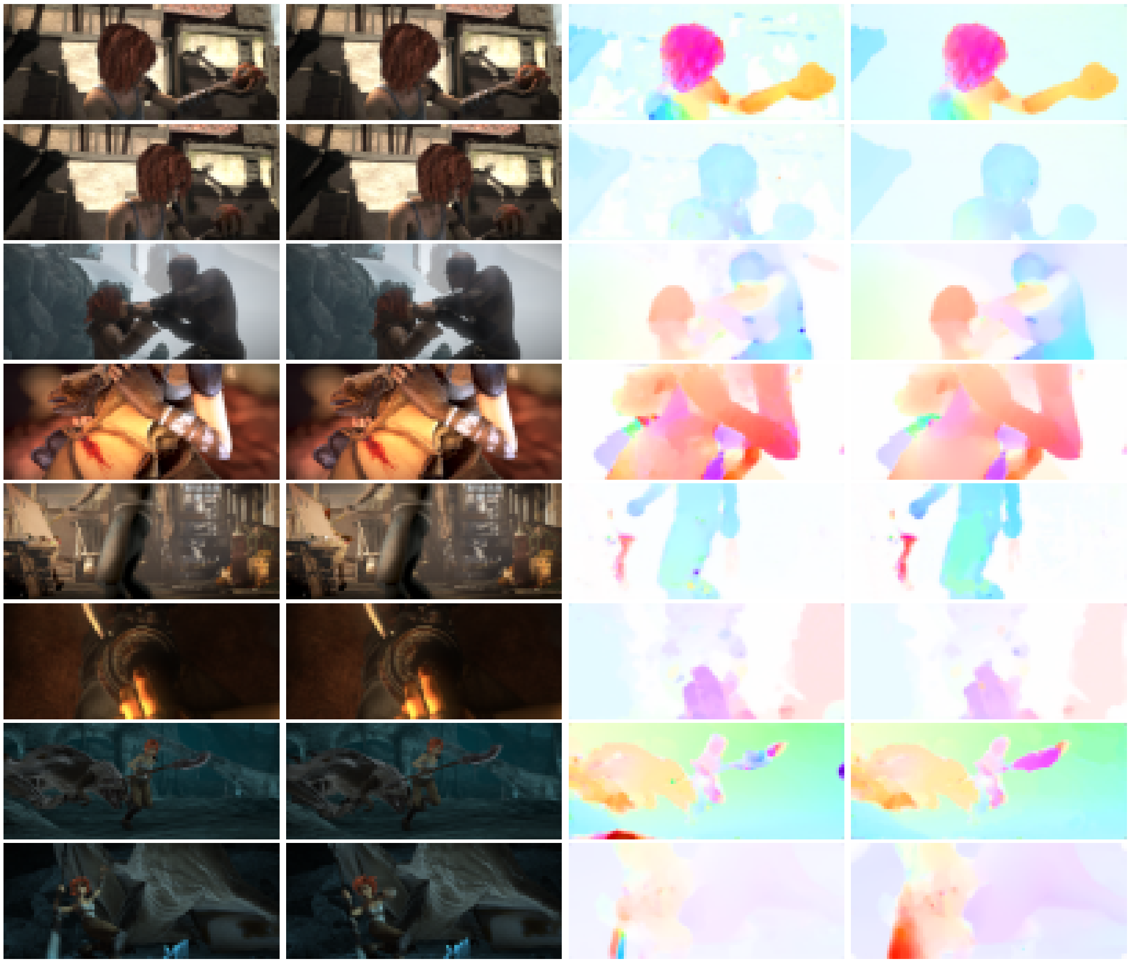
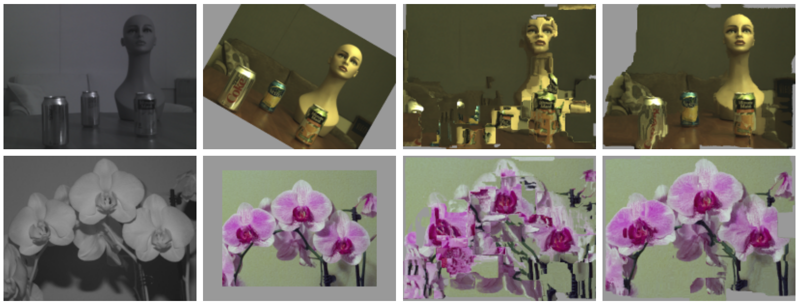


Fig. 20. Visual comparison on the MPI Sintel benchmark. (from left to right) Input image 1 and 2, flow field estimation results of LDOF [21] and LDOF with the DASC+LRP descriptor. Note that the histogram of oriented gradient (HOG) [22] is used in the original LDOF [21].

5.6 Dense Correspondence for Multi-spectral RGB-NIR Images Under Geometric Variations

Similar to [20], [23], [24], our GI-DASC approximately determines a relative scale using successive Gaussian smoothing scheme, which might work in only a limited range of scale variation. As shown in Fig. 21, similar to an existing method, GI-DASC also cannot deal with dramatically severe geometric variations. By leveraging an octave structure based on sub-sampling scheme like SIFT [1], a wider range of scale may be covered.



(a) image 1 (b) image 2 (c) SSF [20] (d) GI-DASC

Fig. 21. Limitations for images under severe geometric variations.

REFERENCES

[1] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[2] E. Tola, V. Lepetit, and P. Fua, “Daisy: An efficient dense descriptor applied to wide-baseline stereo,” *IEEE Trans. PAMI*, vol. 32, no. 5, pp. 815–830, 2010.

[3] E. Schechtman and M. Irani, “Matching local self-similarities across images and videos,” *In Proc. of CVPR*, 2007.

[4] http://vision.middlebury.edu/stereo/.

[5] D. Butler, J. Wulff, G. Stanley, and M. Black, “A naturalistic open source movie for optical flow evaluation,” *In Proc. of ECCV*, 2012.

[6] X. Shen, L. Xu, Q. Zhang, and J. Jia, “Multi-modal and multi-spectral registration for natural images,” *In Proc. of ECCV*, 2014.

[7] B. Fan, Q. Kong, T. Trzcinski, and Z. Wang, “Receptive fields selection for binary feature description,” *IEEE Trans. IP*, vol. 23, no. 6, pp. 2583–2595, 2014.

[8] M. Brown and S. Susstrunk, “Multispectral sift for scene category recognition,” *In Proc. of CVPR*, 2011.

[9] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, “Robust patch-based hdr reconstruction of dynamic scenes,” *In Proc. of ACM SIGGRAGH*, 2012.

[10] Y. HaCohen, E. Shechtman, and E. Lishchinski, “Deblurring by example using dense correspondence,” *In Proc. of ICCV*, 2013.

[11] H. Lee and K. Lee, “Dense 3d reconstruction from severely blurred images using a single moving camera,” *In Proc. of CVPR*, 2013.

[12] A. Alahi, R. Ortiz, and P. Vandergheynst, “Freak : Fast retina keypoint,” *In Proc. of CVPR*, 2012.

[13] Y. Heo, K. Lee, and S. Lee, “Joint depth map and color consistency estimation for stereo images with different illuminations and cameras,” *IEEE Trans. PAMI*, vol. 35, no. 5, pp. 1094–1106, 2013.

[14] M. Calonder, “Brief : Computing a local binary descriptor very fast,” *IEEE Trans. PAMI*, vol. 34, no. 7, pp. 1281–1298, 2011.

[15] C. Liu, J. Yuen, and A. Torralba, “Nonparametric scene parsing via label transfer,” *IEEE Trans. PAMI*, vol. 33, no. 12, pp. 2368–2382, 2011.

[16] J. Kim, C. Liu, F. Sha, and K. Grauman, “Deformable spatial pyramid matching for fast dense correspondences,” *In Proc. of CVPR*, 2013.

[17] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, “The generalized patchmatch correspondence algorithm,” *In Proc. of ECCV*, 2010.

[18] I. Kokkinos and A. Yuille, “Scale invariance without scale selection,” *In Proc. of CVPR*, 2008.

[19] E. Trulls, I. Kokkinos, A. Sanfeliu, and F. M. Noguer, “Dense segmentation-aware descriptors,” *In Proc. of CVPR*, 2013.

[20] W. Qiu, X. Wang, X. Bai, A. Yuille, and Z. Tu, “Scale-space sift flow,” *In Proc. of WACV*, 2014.

[21] T. Brox and J. Malik, “Large displacement optical flow: Descriptor matching in variational motion estimation,” *IEEE Trans. PAMI*, vol. 33, no. 3, pp. 500–513, 2011.

[22] N. Dalal and B. Trigg, “Histograms of oriented gradients for human detection,” *In Proc. of CVPR*, 2005.

[23] T. Hassner, V. Mayzels, and L. Zelnik-Manor, “On sifts and their scales,” *In Proc. of CVPR*, 2012.

[24] H. Yang, W. Lin, and J. Lu, “Daisy filter flow: A generalized discrete approach to dense correspondences,” *In Proc. of CVPR*, 2014.