A Fully Connected Layered Model of Foreground and Background Flow: Supplemental Material

Deqing Sun[1] Jonas Wulff[2] Erik B. Sudderth[3] Hanspeter Pfister[1] Michael J. Black[2]
[1]Harvard University [2]MPI for Intelligent Systems [3]Brown University

Sec. 1 provides the detailed formulas for the mean field approximation algorithm and discusses alternative temporal update schemes. Sec. 2 provides screen shots of the evaluation tables on Middlebury and MPI Sintel datasets.

1. Detailed Derivation Using Mean Field Approximations

1.1. Energy Function

Given the flow fields for each layer, the distribution for the binary masks is

$$P(\mathbf{g}) = \frac{1}{Z}\exp\left\{-E(\mathbf{g})\right\}, \tag{1}$$

in which the energy function is

$$E(\mathbf{g}) = \sum_{t=1}^{T-1}\sum_{(p,q)\in\mathcal{E}_{tk}}\left\{\sum_{k=1}^{2}\phi^{k}_{\text{data}}(g_t^p, g_{t+1}^q) + \lambda_c\phi_{\text{time}}(g_t^p, g_{t+1}^q)\right\} + \lambda_b\sum_{t=1}^{T}\sum_{p}\sum_{q\neq p}\phi_{\text{space}}(g_t^p, g_t^q). \tag{2}$$

The potential function for the spatial term is

$$\phi_{\text{space}}(g_t^p, g_t^q) = w_q^p\delta(g_t^p \neq g_t^q). \tag{3}$$

The potential function for the temporal term is

$$\phi_{\text{time}}(g_t^p, g_{t+1}^q) = \delta(g_t^p \neq g_{t+1}^q). \tag{4}$$

The data term is

$$\phi^1_{\text{data}}(g_t^p, g_{t+1}^q) = \big(\rho_D(I_t^p - I_t^q) - \lambda_D\big)g_t^p g_{t+1}^q, \tag{5}$$
$$\phi^2_{\text{data}}(g_t^p, g_{t+1}^q) = \big(\rho_D(I_t^p - I_t^q) - \lambda_D\big)\bar{g}_t^p \bar{g}_{t+1}^q. \tag{6}$$

where ρ_D is a robust penalty function.

1.2. Mean Field Approximation

The mean field approximation solves for an approximate distribution that minimizes the K-L divergence

$$\begin{aligned} D(Q||P) &= \sum_{\mathbf{g}} Q(\mathbf{g})\log\left(\frac{Q(\mathbf{g})}{P(\mathbf{g})}\right) & (7)\\ &= -\sum_{\mathbf{g}} Q(\mathbf{g})\log(P(\mathbf{g})) + \sum_{\mathbf{g}} Q(\mathbf{g})\log(Q(\mathbf{g})) & (8)\\ &= \sum_{\mathbf{g}} Q(\mathbf{g})E(\mathbf{g}) + \sum_{\mathbf{g}} Q(\mathbf{g})\log(Q(\mathbf{g})) + \log(Z) & (9)\\ &= \mathbb{E}_Q[E(\mathbf{g}] + \mathbb{E}_Q[\log(Q(\mathbf{g}))] + \log(Z) & (10)\end{aligned}$$

The mean field approximation assumes that the approximate distribution can be factorized as

$$Q(\mathbf{g}) = \prod_{t=1}^{T} \prod_{p} Q_t^p(g_t^p). \tag{11}$$

The general mean field iteration update formula is [1]

$$Q_t^p(g_t^p) = \frac{1}{Z_t^p} \exp\left\{ - \sum_{\phi: g_t^p \in \text{scope}[\phi]} \mathbb{E}_{(g_\phi - \{g_t^p\}) \sim Q} [\ln \phi(g_\phi, g_t^p)] \right\}, \tag{12}$$

where scope$[\phi]$ contains all the pixels that are affected by the potential function ϕ and Z_t^p is a normalization constant.

We can obtain the detailed mean field update equation as

$$\begin{aligned} Q_t^p(g_t^p) = \frac{1}{Z_t^p} \exp\Bigg\{ & - \lambda_b \sum_{q \neq p} w_q^p \mathbb{E}_Q[\delta(g_t^p \neq g_t^q) | g_t^p] - \lambda_c \sum_{(p,q) \in \mathcal{E}_{tk}} \mathbb{E}_Q[\delta(g_t^p \neq g_{t+1}^q) | g_t^p] \\ - \sum_{k=1}^{2} \sum_{(p,q) \in \mathcal{E}_{tk}} \mathbb{E}_Q[\phi_{\text{data}}^k(g_t^p, g_{t+1}^q) | g_t^p] & - \lambda_c \sum_{(q,p) \in \mathcal{E}_{t-1,k}} \mathbb{E}_Q[\delta(g_{t-1}^q \neq g_t^p) | g_t^p] - \sum_{k=1}^{2} \sum_{(q,p) \in \mathcal{E}_{t-1,k}} \mathbb{E}_Q[\phi_{\text{data}}^k(g_{t-1}^q, g_t^p) | g_t^p] \Bigg\}. \end{aligned} \tag{13}$$

$$\mathbb{E}_Q[\phi_{\text{data}}^1(g_t^p, g_{t+1}^q) | g_t^p] = \left[\rho_D\left(I_t^p - I_{t+1}^q\right) - \lambda_d\right] g_t^p Q(g_{t+1}^q = 1), \tag{14}$$

$$\mathbb{E}_Q[\phi_{\text{data}}^2(g_t^p, g_{t+1}^q) | g_t^p] = \left[\rho_D\left(I_t^p - I_{t+1}^q\right) - \lambda_d\right] \bar{g}_t^p Q(g_{t+1}^q = 0), \tag{15}$$

$$\mathbb{E}_Q[\phi_{\text{data}}^1(g_{t-1}^q, g_t^p) | g_t^p] = \left[\rho_D\left(I_{t-1}^q - I_t^p\right) - \lambda_d\right] Q(g_{t-1}^q = 1) g_t^p, \tag{16}$$

$$\mathbb{E}_Q[\phi_{\text{data}}^2(g_{t-1}^q, g_t^p) | g_t^p] = \left[\rho_D\left(I_{t-1}^q - I_t^p\right) - \lambda_d\right] Q(g_{t-1}^q = 0) \bar{g}_t^p. \tag{17}$$

$$\mathbb{E}_Q[\phi_{\text{space}}(g_t^p, g_t^q) | g_t^p] = w_q^p \sum_{l=0}^{1} \delta(g_t^p \neq l) Q\left(g_t^q = l\right) = w_q^p Q\left(g_t^q = 1 - g_t^p\right), \tag{18}$$

$$\mathbb{E}_Q[\phi_{\text{time}}(g_t^p, g_{t+1,k}^q) | g_t^p] = \sum_{l=0}^{1} \delta(g_t^p \neq l) Q\left(g_{t+1,k}^q = l\right) = Q\left(g_{t+1}^q = 1 - g_t^p\right), \tag{19}$$

$$\mathbb{E}_Q[\phi_{\text{time}}(g_{t-1}^q, g_t^p) | g_t^p] = \sum_{l=0}^{1} \delta(g_t^p \neq l) Q\left(g_{t-1,k}^q = l\right) = Q\left(g_{t-1,k}^q = 1 - g_t^p\right). \tag{20}$$

Table 1.2 provides the detailed algorithm for the mean field algorithm for two-layer case.

1.3. Temporal Message Update Schemes

There are various possible schedules for updating temporal messages. Our experiments use parallel updates based on all frames in the preceding iteration, as illustrated in Figure 1. We also tested a forward-backward schedule inspired by optimal temporal filtering algorithms, but found it performed slightly worse in practice.

2. Additional Results

2.1. Results on the MPI Sintel Data Set

Figure 2 shows the screen shot of the evaluation table at the time of writing (April, 2013). Note that **MDP-Flow2** was the previous best published method on the data set. **Deep-Matching-Flow** and **Complex-Flow** are anonymous submissions. Figures 3-6 show per-sequence results on the Sintel test set. For each sequence, the image is given in the top left, the ground truth flow (obtained from the website) in the top right, the segmentation in the bottom left, and the estimated flow in the bottom right. Our method performs well if the two-layer assumption holds (see *market_1*, top in Fig. 5, and *wall*, bottom in Fig. 6. It fails, on the other hand, if objects are moving very fast (*ambush_1*, top in Fig. 3), or are very small (the dragon in *temple_1*, Fig. 6 top). However, these are problems that are common to all current methods for optical flow computation.

Algorithm 1 Mean field for non-local layers

Compute $C^p_{tk} = \left[\rho_D\left(I^p_t - I^q_{t+1}\right) - \lambda_d\right]$, $(p,q) \in \mathcal{E}_{tk}$
Initialize $Q^p_t(l) \propto \exp\{-C^p_{t,2-l}\}$
while not converged **do**
  $Q^{\text{prev}} \leftarrow Q$
  Change weight on temporal term λ_c as scheduled
  Spatial message passing
    $\tilde{Q}^p_t(l) \leftarrow \lambda_b \sum_{q \neq p} w^p_q Q^q_t(\bar{l})$
  Temporal message passing from next frame
    $\tilde{Q}^p_t(l) \leftarrow \tilde{Q}^p_t(l) + \lambda_c \sum_{(p,q)\in\mathcal{E}_{t1}} Q^q_{t+1}(\bar{l})$
    $\tilde{Q}^p_t(1) \leftarrow \tilde{Q}^p_t(1) + \sum_{(p,q)\in\mathcal{E}_{t1}} C^p_{t1} Q^q_{t+1}(1)$
    $\tilde{Q}^p_t(0) \leftarrow \tilde{Q}^p_t(0) + \sum_{(p,q)\in\mathcal{E}_{t2}} C^p_{t2} Q^q_{t+1}(0)$
  Temporal message passing from previous frame
    $\tilde{Q}^p_t(l) \leftarrow \tilde{Q}^p_t(l) + \lambda_c \sum_{(q,p)\in\mathcal{E}_{t-1,1}} Q^q_{t-1}(\bar{l})$
    $\tilde{Q}^p_t(1) \leftarrow \tilde{Q}^p_t(1) + \sum_{(q,p)\in\mathcal{E}_{t-1,1}} C^q_{t-1,1} Q^q_{t-1}(1)$
    $\tilde{Q}^p_t(0) \leftarrow \tilde{Q}^p_t(0) + \sum_{(q,p)\in\mathcal{E}_{t-1,2}} C^q_{t-1,2} Q^q_{t-1}(0)$
  Exp and normalize
    $Q^p_t(l) \leftarrow \frac{\exp\{-\tilde{Q}^p_t(l)\}}{\exp\{-\tilde{Q}^p_t(0)\}+\exp\{-\tilde{Q}^p_t(1)\}}$
  Damping
    $Q \leftarrow \mu Q + (1-\mu) Q^{\text{prev}}$
  Median filter Q when λ_c changes
end while

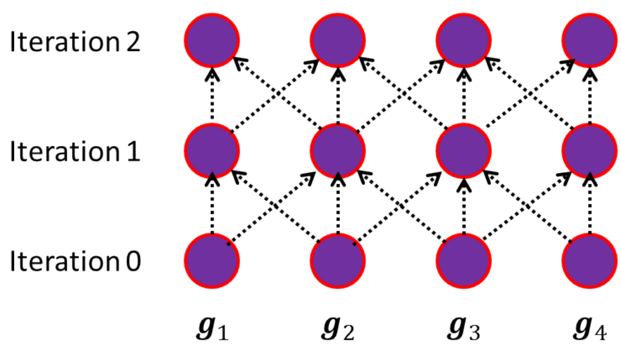


Figure 1. Temporal update schedule for the mean field algorithm. Each frame uses the marginal estimates for its temporal and spatial neighbors at the previous iteration.

Table 1. Average end-point error (EPE) on the Middlebury *training* set. The fast version uses a faster flow computation method but achieves very closer performance.

	Avg.	Venus	Dimetrodon	Hydrangea	RubberWhale	Grove2	Grove3	Urban2	Urban3
FC-2Layers-FF	0.205	0.228	0.143	0.155	0.072	0.094	0.362	0.199	0.391
Fast version	0.212	0.227	0.139	0.159	0.077	0.095	0.383	0.214	0.405

2.2. Results on the Middlebury Data set

Figure 7 shows the top 15 methods on EPE and AAE from the Middlebury hidden table at the time of submission (November 2012). With FlowFusion, the proposed method (**FC-2Layers-FF**) is ranked 8^{th} in EPE and 5^{th} in AAE. Without FlowFusion, the proposed method (**FC-2Layers-FF**) is ranked 9^{th} in EPE and 11^{th} in AAE.

FC-2Layers-FF achieves similar performance as the local layered model **nLayers**, but more than 10 times faster (about 45 minutes vs 10 hours). The main computational bottleneck is in computing the initial flow field by **Classic+NL** in MATLAB. We have developed a fast version of **Classic+NL** by using preconditioned conjugate gradient and reduced the total computational time to about 10 minutes, with very slight loss in accuracy, as shown in Table 1.

	EPE all	EPE matched	EPE unmatched	d0-10	d10-60	d60-140	s0-10	s10-40	s40+	
GroundTruth [1]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	Visualize Results
Deep-Matching-Flow [2]	7.223	3.338	38.868	5.651	3.146	2.210	1.293	4.109	44.159	Visualize Results
Complex-Flow [3]	7.249	2.973	42.088	4.896	2.817	2.218	1.159	4.183	44.866	Visualize Results
FC-Layers-FF [4]	8.137	4.261	39.723	6.537	4.257	2.946	1.034	4.835	51.349	Visualize Results
MDP-Flow2 [5]	8.445	4.150	43.430	5.703	3.925	3.406	1.420	5.449	50.507	Visualize Results
EP-PM [6]	8.499	4.369	42.139	5.993	3.977	3.691	1.534	5.551	50.150	Visualize Results
LDOF [7]	9.116	5.037	42.344	6.849	4.928	4.003	1.485	4.839	57.296	Visualize Results
Classic+NL [8]	9.153	4.814	44.509	7.215	4.822	3.427	1.113	4.496	60.291	Visualize Results
Horn+Schunck [9]	9.610	5.419	43.734	7.950	5.658	3.976	1.882	5.335	58.274	Visualize Results
Classic++ [10]	9.959	5.410	47.000	8.072	5.554	3.750	1.403	5.098	64.135	Visualize Results
Classic+NL-fast [11]	10.088	5.659	46.145	8.010	5.738	4.160	1.092	4.666	67.801	Visualize Results
AnisoHuber.L1 [12]	11.927	7.323	49.366	9.464	7.692	5.929	1.155	7.966	74.796	Visualize Results
AtrousFlow [13]	14.173	9.573	51.548	11.511	10.027	8.092	2.011	12.052	79.484	Visualize Results

	EPE all	EPE matched	EPE unmatched	d0-10	d10-60	d60-140	s0-10	s10-40	s40+	
GroundTruth [1]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	Visualize Results
Deep-Matching-Flow [2]	5.385	1.771	34.823	4.518	1.534	0.836	0.963	2.729	33.752	Visualize Results
Complex-Flow [3]	5.386	1.397	37.896	2.722	1.341	1.004	0.683	2.245	36.342	Visualize Results
EP-PM [4]	5.573	1.949	35.099	3.804	1.907	1.390	0.661	2.643	37.043	Visualize Results
MDP-Flow2 [5]	5.837	1.869	38.158	3.210	1.913	1.441	0.640	2.603	39.459	Visualize Results
FC-Layers-FF [6]	6.781	3.053	37.144	5.841	3.390	1.688	0.580	3.308	45.962	Visualize Results
LDOF [7]	7.563	3.432	41.170	5.353	3.284	2.454	0.936	2.908	51.696	Visualize Results
Classic+NL [8]	7.961	3.770	42.079	6.191	3.911	2.509	0.573	2.694	57.374	Visualize Results
Classic++ [9]	8.721	4.259	45.047	6.983	4.494	2.753	0.902	3.295	60.645	Visualize Results
Horn+Schunck [10]	8.739	4.525	43.032	7.542	5.045	2.891	1.141	3.860	58.243	Visualize Results
Classic+NL-fast [11]	9.129	4.725	44.956	7.157	4.974	3.331	0.558	2.812	66.935	Visualize Results
AnisoHuber.L1 [12]	12.642	7.983	50.472	10.457	8.675	6.320	0.753	9.976	77.835	Visualize Results
AtrousFlow [13]	14.200	9.584	51.758	11.964	10.338	7.926	1.702	12.440	80.185	Visualize Results

Figure 2. **Screen shots of the methods from the MPI Sintel evaluation table (April 2013).** Top: Final set; bottom: clean set. The proposed method is **FC-Layers-FF** in the table.

Figures 8, 9, 10, and 11 show the estimated flow and segmentation on the Middlebury training and test sets. Note the sharp motion boundaries recovered by **FC-2Layers-FF**, such as in "Schefflera" (third row in Figure 8), "Teddy" (bottom row in Figure 9), and "Grove3" (second row in Figure 11).

Table 2 shows the KL divergence (up to a constant) between the approximate distribution and the true distribution, for the proposed method with and without FlowFusion. Using FlowFusion consistently produces results with lower K-L divergence to the actual distribution.

References

[1] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

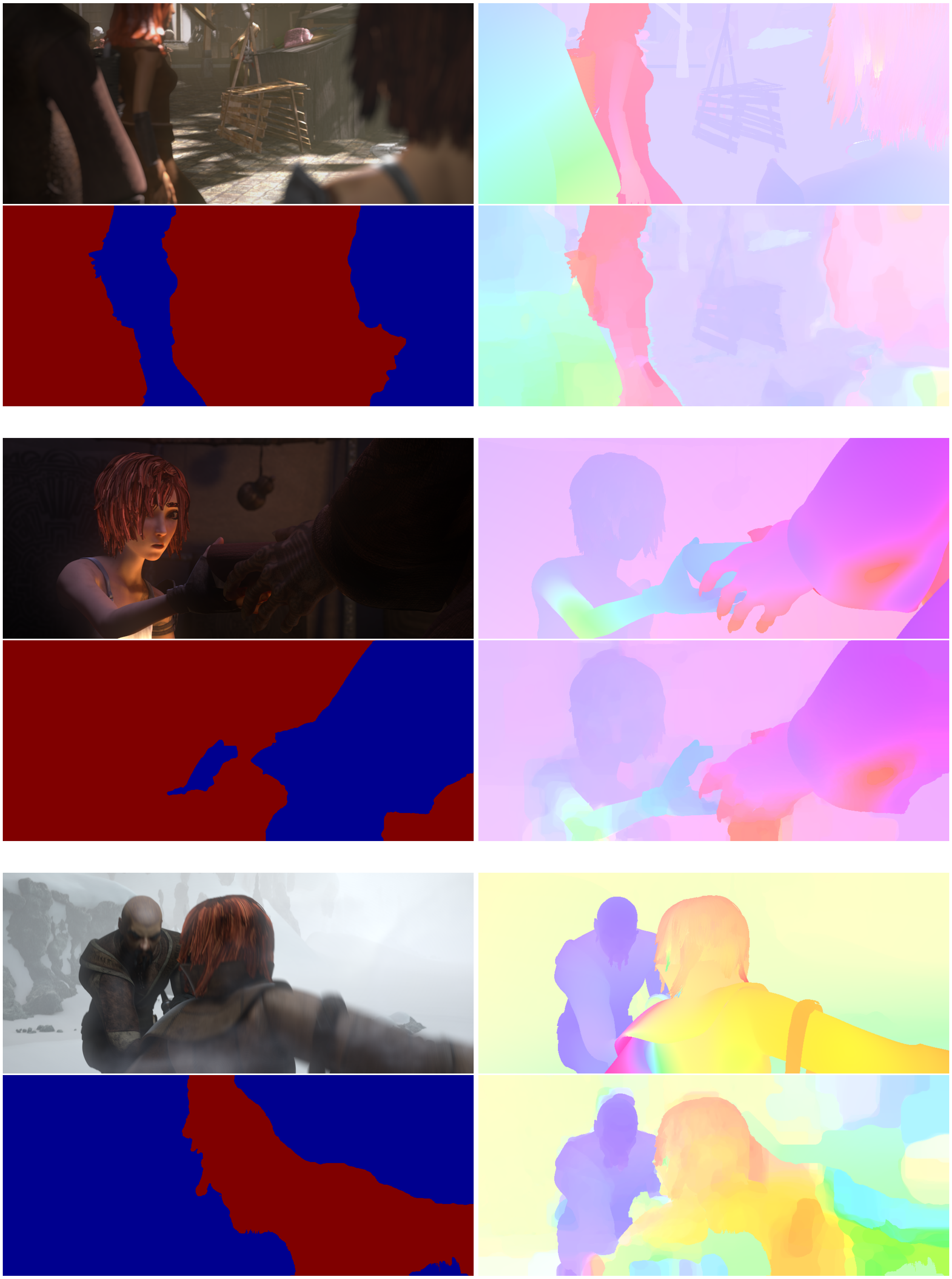


Figure 3. MPI-Sintel test results I. From top to bottom, the groups of images show the sequences *PERTURBED_market_3*, *PERTURBED_shaman_1*, and *ambush_1*. Within each group, the top left image shows a frame from the sequence, the top right image shows the ground truth optical flow, the bottom left image shows the segmentation, and the bottom right image shows the estimated flow.

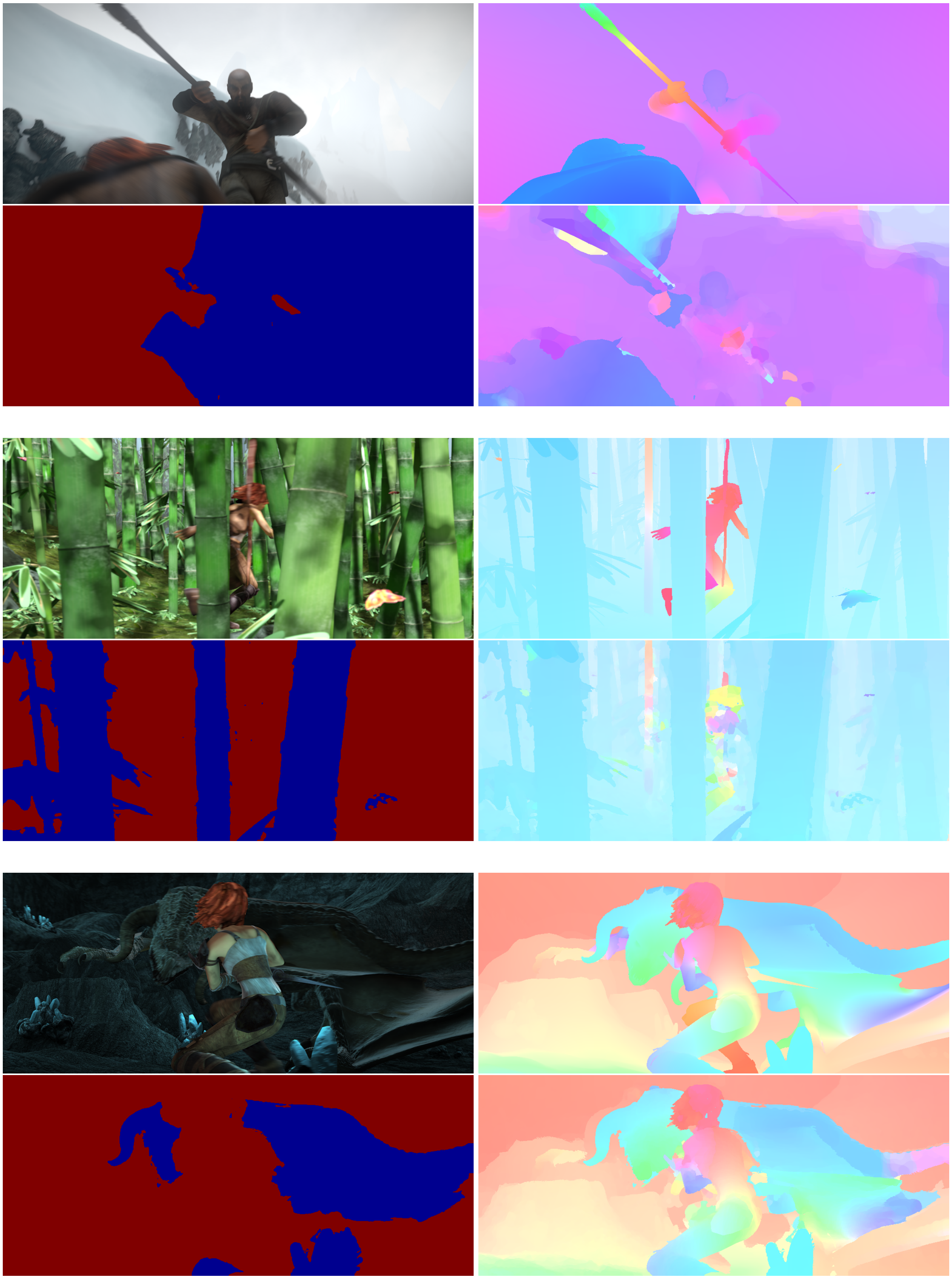


Figure 4. MPI-Sintel test results II. From top to bottom, the groups of images show the sequences *ambush_3*, *bamboo_3*, and *cave_3*. Within each group, the top left image shows a frame from the sequence, the top right image shows the ground truth optical flow, the bottom left image shows the segmentation, and the bottom right image shows the estimated flow.

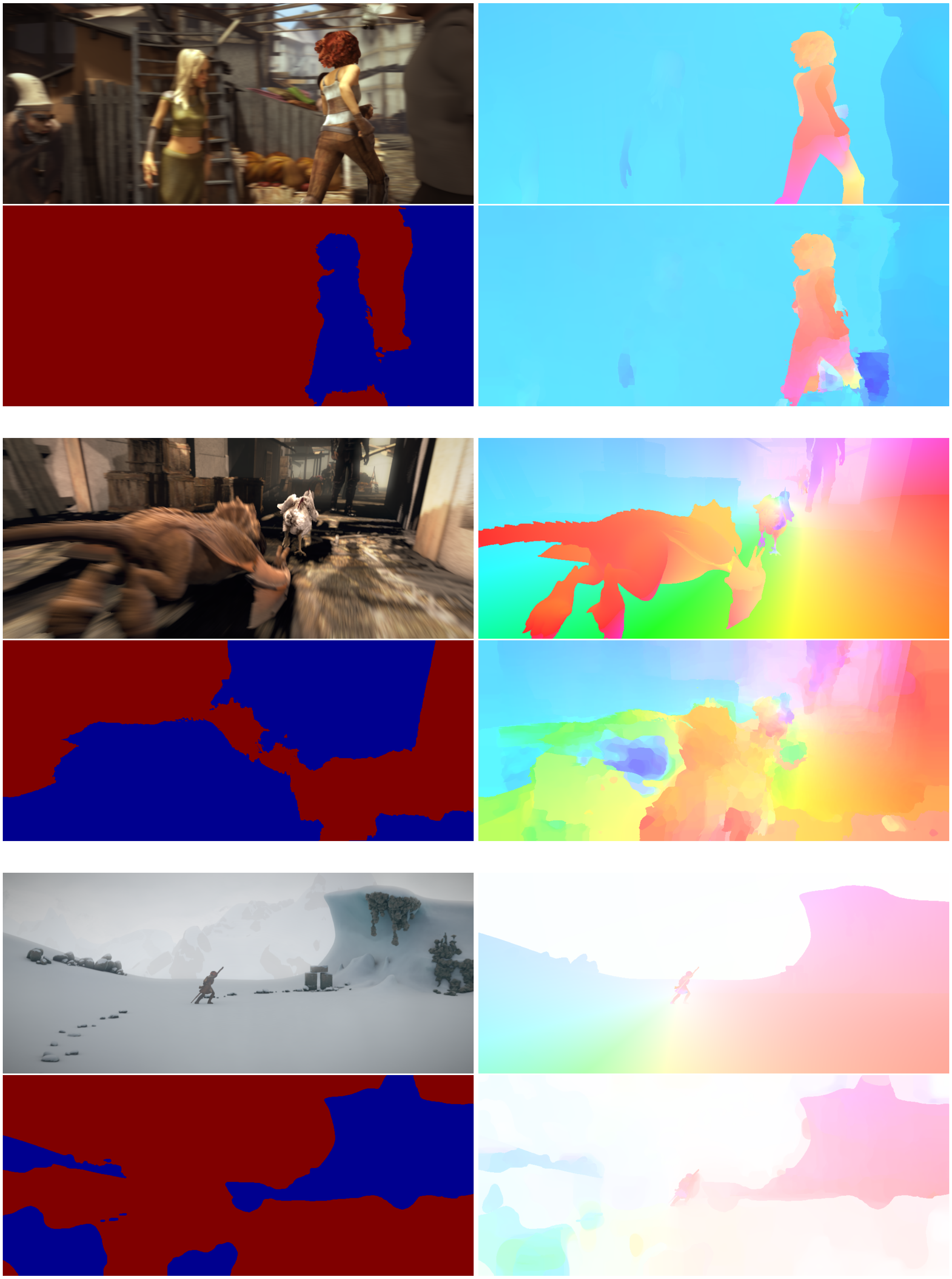


Figure 5. MPI-Sintel test results III. From top to bottom, the groups of images show the sequences *market_1*, *market_4*, and *mountain_2*. Within each group, the top left image shows a frame from the sequence, the top right image shows the ground truth optical flow, the bottom left image shows the segmentation, and the bottom right image shows the estimated flow.

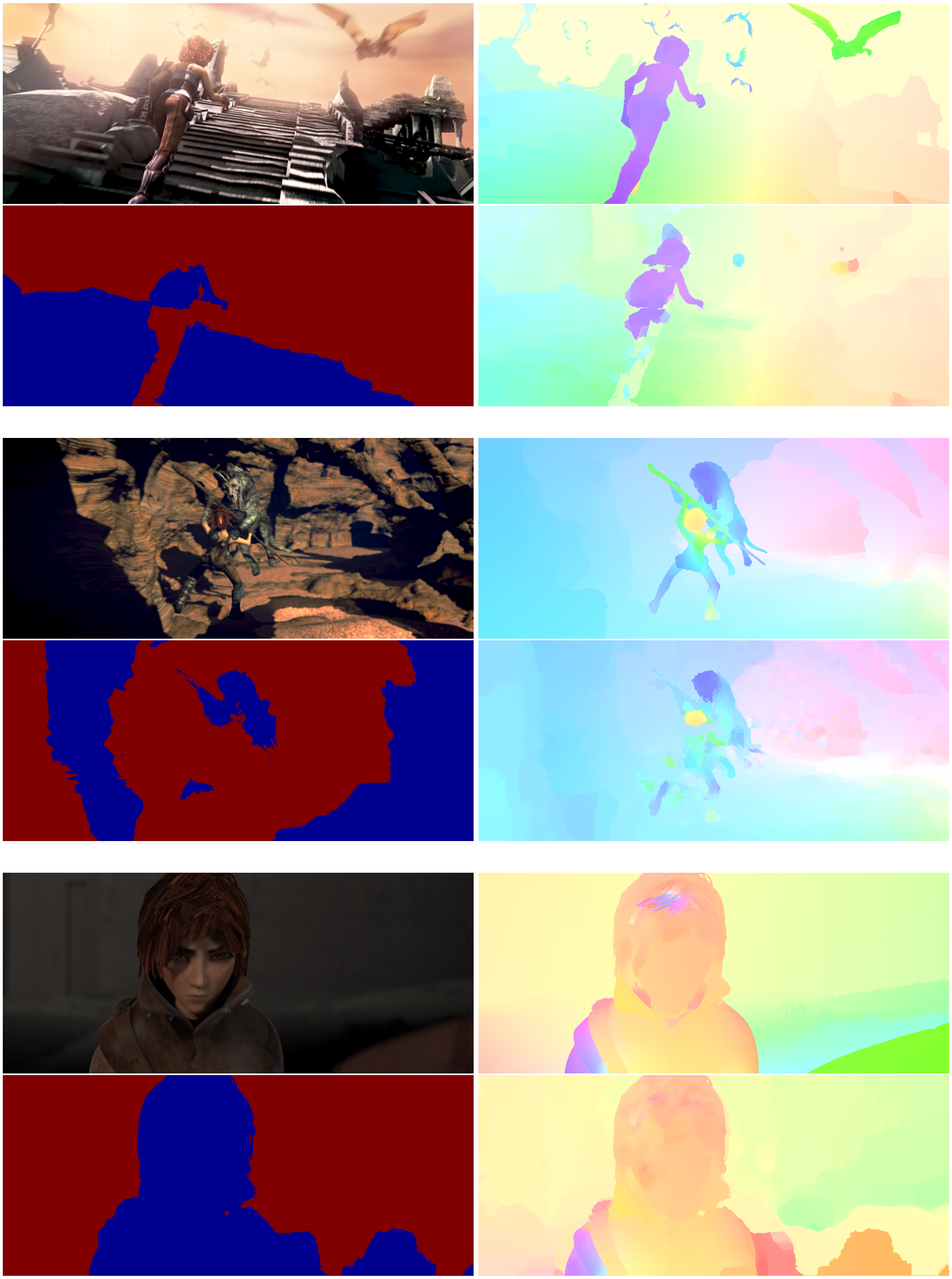


Figure 6. MPI-Sintel test results IV. From top to bottom, the groups of images show the sequences *temple_1*, *tiger*, and *wall*. Within each group, the top left image shows a frame from the sequence, the top right image shows the ground truth optical flow, the bottom left image shows the segmentation, and the bottom right image shows the estimated flow.

Average endpoint error	avg.	Army (Hidden texture) GT im0 im1			Mequon (Hidden texture) GT im0 im1			Schefflera (Hidden texture) GT im0 im1			Wooden (Hidden texture) GT im0 im1			Grove (Synthetic) GT im0 im1			Urban (Synthetic) GT im0 im1			Yosemite (Synthetic) GT im0 im1			Teddy (Stereo) GT im0 im1		
	rank	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
MDP-Flow2 [70]	5.4	0.08 4	0.21 2	0.07 10	**0.15** 1	**0.48** 1	**0.11** 1	0.20 2	0.40 2	**0.14** 1	0.15 10	0.80 19	0.08 6	0.63 11	0.93 12	0.43 11	**0.26** 1	0.76 2	0.23 3	0.11 9	0.12 6	0.17 10	0.38 2	0.79 2	0.44 2
NN-field [73]	6.0	0.08 4	0.22 8	**0.05** 1	0.17 4	0.55 4	0.13 5	**0.19** 1	**0.39** 1	0.15 3	**0.09** 1	**0.48** 1	**0.05** 1	**0.41** 1	**0.61** 1	**0.20** 1	0.52 33	**0.64** 1	0.26 5	0.13 23	0.13 20	0.20 19	**0.35** 1	0.83 3	**0.21** 1
ADF [67]	11.5	0.08 4	0.22 8	0.06 2	0.18 5	0.62 10	0.14 9	0.29 18	0.71 22	0.17 8	0.16 21	0.91 31	0.07 2	0.69 19	1.03 19	0.47 14	0.43 12	0.91 5	0.28 7	0.12 15	0.12 6	0.20 19	0.43 4	0.88 6	0.63 10
Layers++ [37]	11.5	0.08 4	0.21 2	0.07 10	0.19 9	0.56 5	0.17 19	0.20 2	0.40 2	0.18 12	0.13 2	0.58 3	0.07 2	0.48 2	0.70 2	0.33 4	0.47 21	1.01 8	0.33 22	0.15 40	0.14 35	0.24 34	0.46 9	0.88 6	0.72 22
LME [72]	12.0	0.08 4	0.22 8	0.06 2	**0.15** 1	0.49 2	**0.11** 1	0.30 21	0.64 15	0.31 54	0.15 10	0.78 17	0.09 17	0.66 14	0.96 14	0.53 21	0.33 3	1.18 21	0.28 7	0.12 15	0.12 6	0.18 13	0.44 5	0.91 8	0.61 8
IROF++ [58]	12.2	0.08 4	0.23 13	0.07 10	0.21 18	0.68 18	0.17 19	0.28 15	0.63 14	0.19 21	0.15 10	0.73 11	0.09 17	0.60 8	0.89 8	0.42 9	0.43 12	1.08 14	0.31 15	0.10 4	0.12 6	0.12 4	0.47 11	0.98 15	0.68 18
nLayers [57]	12.5	**0.07** 1	**0.19** 1	0.06 2	0.22 24	0.59 6	0.19 33	0.25 10	0.54 9	0.20 30	0.15 10	0.84 22	0.08 6	0.53 3	0.78 4	0.34 5	0.44 16	0.84 3	0.30 13	0.13 23	0.13 20	0.20 19	0.47 11	0.97 14	0.67 15
FC-Layers-FF [77]	13.9	0.08 4	0.21 2	0.07 10	0.21 18	0.70 21	0.17 19	0.20 2	0.40 2	0.18 12	0.15 10	0.76 15	0.08 6	0.53 3	0.77 3	0.37 6	0.49 26	1.02 9	0.33 22	0.16 48	0.13 20	0.29 53	0.44 5	0.87 5	0.64 12
FC-Layers [78]	16.5	0.08 4	0.22 8	0.07 10	0.21 18	0.70 21	0.17 19	0.21 5	0.43 5	0.18 12	0.15 10	0.75 14	0.08 6	0.58 5	0.84 5	0.42 9	0.51 30	1.12 18	0.34 28	0.16 48	0.13 20	0.30 58	0.48 14	0.94 10	0.68 18
ALD-Flow [68]	16.7	**0.07** 1	0.21 2	0.06 2	0.19 9	0.64 15	0.13 5	0.30 21	0.73 23	0.15 3	0.17 26	0.92 35	0.07 2	0.78 26	1.14 26	0.59 25	0.33 3	1.30 28	**0.21** 1	0.12 15	0.12 6	0.28 48	0.54 26	1.19 28	0.73 25
FESL [75]	16.8	0.08 4	0.21 2	0.07 10	0.25 40	0.75 32	0.19 33	0.27 11	0.61 12	0.18 12	0.14 4	0.68 4	0.08 6	0.61 10	0.89 8	0.44 12	0.47 21	1.03 11	0.32 18	0.14 31	0.15 44	0.25 37	0.50 20	0.96 12	0.63 10
COFM [59]	16.9	0.08 4	0.26 29	0.06 2	0.18 5	0.62 10	0.14 9	0.30 21	0.74 25	0.19 21	0.15 10	0.86 24	0.07 2	0.79 27	1.14 26	0.74 41	0.35 6	0.87 4	0.28 7	0.14 31	0.12 6	0.28 48	0.49 16	0.94 10	0.71 21
SCR [74]	16.9	0.08 4	0.23 13	0.07 10	0.22 24	0.71 24	0.17 19	0.27 11	0.60 11	0.19 21	0.14 4	0.73 11	0.08 6	0.63 11	0.92 11	0.44 12	0.51 30	1.08 14	0.33 22	0.15 40	0.13 20	0.29 53	0.47 11	0.93 9	0.67 15
TC-Flow [46]	17.0	**0.07** 1	0.21 2	0.06 2	**0.15** 1	0.59 6	**0.11** 1	0.31 25	0.78 28	**0.14** 1	0.16 21	0.86 24	0.08 6	0.75 24	1.11 25	0.54 22	0.42 11	1.40 35	0.25 4	0.11 9	0.12 6	0.29 53	0.62 30	1.35 30	0.93 41
Efficient-NL [60]	17.3	0.08 4	0.22 8	0.06 2	0.21 18	0.67 17	0.17 19	0.31 25	0.73 23	0.18 12	0.14 4	0.71 8	0.08 6	0.59 7	0.88 7	0.39 7	1.30 57	1.35 31	0.67 53	0.14 31	0.13 20	0.26 39	0.45 8	0.85 4	0.55 5

Average angle error	avg.	Army (Hidden texture) GT im0 im1			Mequon (Hidden texture) GT im0 im1			Schefflera (Hidden texture) GT im0 im1			Wooden (Hidden texture) GT im0 im1			Grove (Synthetic) GT im0 im1			Urban (Synthetic) GT im0 im1			Yosemite (Synthetic) GT im0 im1			Teddy (Stereo) GT im0 im1		
	rank	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
NN-field [73]	4.9	2.89 3	8.13 8	**2.11** 1	2.10 3	7.15 4	1.77 6	**2.27** 1	5.59 2	1.61 3	**1.58** 1	**8.52** 1	**0.79** 1	2.35 2	3.05 3	**1.60** 1	**1.89** 1	**5.20** 1	**1.37** 1	2.43 26	3.70 25	1.95 20	**1.01** 1	**2.25** 1	**0.53** 1
nLayers [57]	8.5	**2.80** 1	**7.42** 1	2.20 3	2.71 15	7.24 5	2.55 34	2.61 6	6.24 6	2.45 31	2.30 3	12.7 6	1.16 3	**2.30** 1	**3.02** 1	1.70 2	2.62 5	6.95 2	2.09 3	2.29 19	3.46 13	1.89 17	1.38 8	3.06 11	1.29 8
MDP-Flow2 [70]	9.6	3.23 20	7.93 5	2.60 12	**1.92** 1	**6.64** 1	**1.52** 1	2.46 4	5.91 4	1.56 2	3.05 22	15.8 25	1.51 22	2.77 13	3.50 10	2.16 12	2.86 7	8.58 7	2.70 14	2.00 9	3.50 18	1.59 8	1.28 5	2.67 5	0.89 3
ADF [67]	12.1	2.98 6	8.32 12	2.28 4	2.27 5	8.35 11	1.81 7	3.55 19	9.74 21	2.17 16	3.15 28	16.8 31	1.29 7	2.64 10	3.55 11	1.81 4	3.02 8	9.08 9	2.38 6	2.29 19	3.48 15	2.07 23	1.34 6	3.03 9	1.11 4
FC-Layers-FF [77]	12.8	3.02 8	7.87 4	2.61 13	2.72 16	9.35 18	2.29 21	2.36 2	**5.47** 1	2.15 15	2.48 4	12.6 5	1.28 5	2.49 4	3.19 4	2.03 10	3.39 19	8.92 8	2.83 21	2.83 41	3.92 37	2.80 38	1.25 4	2.57 4	1.20 6
Layers++ [37]	14.3	3.11 10	8.22 10	2.79 24	2.43 9	7.02 3	2.24 17	2.43 3	5.77 3	2.18 19	2.13 2	9.71 2	1.15 2	2.35 2	**3.02** 1	1.96 6	3.81 28	11.4 24	3.22 32	2.74 37	4.01 42	2.35 29	1.45 10	3.05 10	1.79 19
Efficient-NL [60]	14.9	2.99 7	8.23 11	2.28 4	2.72 16	8.95 15	2.25 19	3.81 23	9.87 23	2.07 13	2.77 16	14.3 15	1.46 16	2.61 7	3.48 9	1.96 6	3.31 15	8.33 5	2.59 10	2.60 31	3.75 26	2.54 34	1.60 15	3.02 8	1.66 14
LME [72]	15.3	3.15 14	8.04 7	2.31 7	1.95 2	6.65 2	1.59 3	4.03 28	9.31 19	4.57 54	2.69 12	13.6 9	1.42 11	2.85 19	3.61 14	2.42 24	3.47 22	12.8 29	3.17 29	2.12 13	3.53 21	1.73 11	1.34 6	2.75 6	1.18 5
FESL [75]	15.7	2.96 5	7.70 2	2.54 10	3.26 38	10.4 25	2.56 35	3.25 12	8.39 12	2.17 16	2.56 6	13.2 7	1.31 8	2.57 6	3.40 7	2.12 11	2.60 4	7.65 3	2.30 4	2.64 35	4.22 47	2.47 31	1.75 19	3.49 17	1.71 16
ALD-Flow [68]	15.8	2.82 2	7.86 3	2.16 2	2.84 22	10.1 23	1.86 10	3.73 21	10.4 24	1.67 6	3.10 24	16.8 31	1.28 5	2.69 11	3.60 13	1.85 5	2.79 6	11.3 23	2.32 5	2.07 10	3.25 5	3.10 54	2.03 26	5.11 27	1.94 21
IROF++ [58]	16.5	3.17 16	8.69 16	2.61 13	2.79 19	9.61 20	2.33 23	3.43 15	8.86 16	2.38 25	2.87 18	14.8 18	1.52 23	2.74 12	3.57 12	2.19 13	3.20 13	9.70 15	2.71 15	1.96 8	3.45 12	1.22 5	1.80 20	4.06 21	2.50 28
FC-Layers [78]	16.5	3.08 9	8.16 9	2.72 20	2.78 18	9.38 19	2.30 22	2.52 5	6.05 5	2.18 19	2.60 7	13.5 8	1.37 10	2.63 9	3.41 8	2.22 14	3.51 24	9.78 17	2.85 22	2.84 43	3.79 29	2.88 46	1.48 12	3.21 13	1.35 9
SCR [74]	16.7	3.12 11	8.48 13	2.59 11	2.95 28	10.4 25	2.35 24	3.19 10	8.09 10	2.43 29	2.63 8	13.9 12	1.35 9	2.81 15	3.64 15	2.30 16	3.02 8	8.29 4	2.39 8	2.77 40	3.79 29	2.89 47	1.39 9	2.85 7	1.60 13
TC-Flow [46]	18.0	2.91 4	8.00 6	2.34 8	2.18 4	8.77 12	**1.52** 1	3.84 25	10.7 28	**1.49** 1	3.13 25	16.6 30	1.46 16	2.78 14	3.73 19	1.96 6	3.08 10	11.4 24	2.66 11	1.94 6	3.43 11	3.20 58	3.06 32	7.04 31	4.08 51
Sparse-NonSparse [56]	18.0	3.14 13	8.75 18	2.76 23	3.02 31	10.6 28	2.43 28	3.45 17	8.96 17	2.36 23	2.66 10	13.7 11	1.42 11	2.85 19	3.75 20	2.33 18	3.28 14	9.40 12	2.73 16	2.42 25	3.31 6	2.69 36	1.47 11	3.07 12	1.66 14

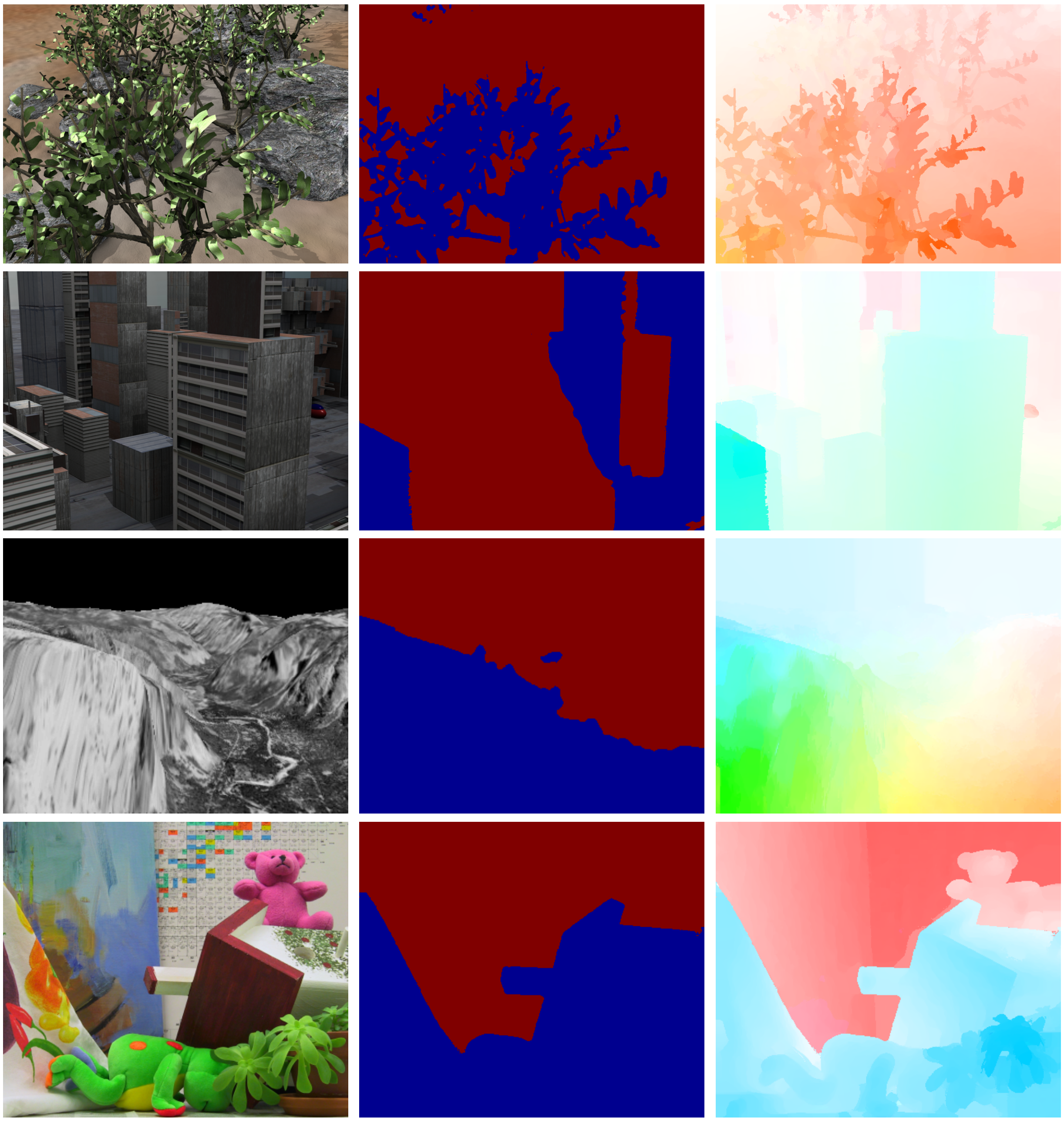
Figure 7. **Screen shots of the top 15 methods from the Middlebury hidden table (Nov. 2012).** Top: EPE; bottom: AAE. The proposed methods are **FC-Layers-FF** and **FC-Layers**.

Table 2. K-L divergence of the solutions by the proposed method with and without FlowFusion on the Middlebury optical flow benchmark *test* set. Using FlowFusion consistently produces results with lower K-L divergence to the actual distribution.

	Avg.	Army	Mequon	Schefflera	Wooden	Grove	Urban	Yosemite	Teddy
FC-Layers	-3.20	-6.46	-5.28	-3.56	-6.65	-5.12	3.34	-1.06	-0.83
FC-2Layers-FF	-3.42	-6.48	-5.44	-3.96	-6.74	-5.58	2.88	-1.04	-0.97

(a)First frame (b) Layer segmentation (c) Estimated flow field

Figure 8. Estimated flow fields and scene structure on the Middlebury *test* sequences. Left to right: first frame, layer segmentation, and estimated flow field. Top to bottom: "Army", "Mequon", "Schefflera", and "Wooden". Depth ordering: blue is foreground and red is background.

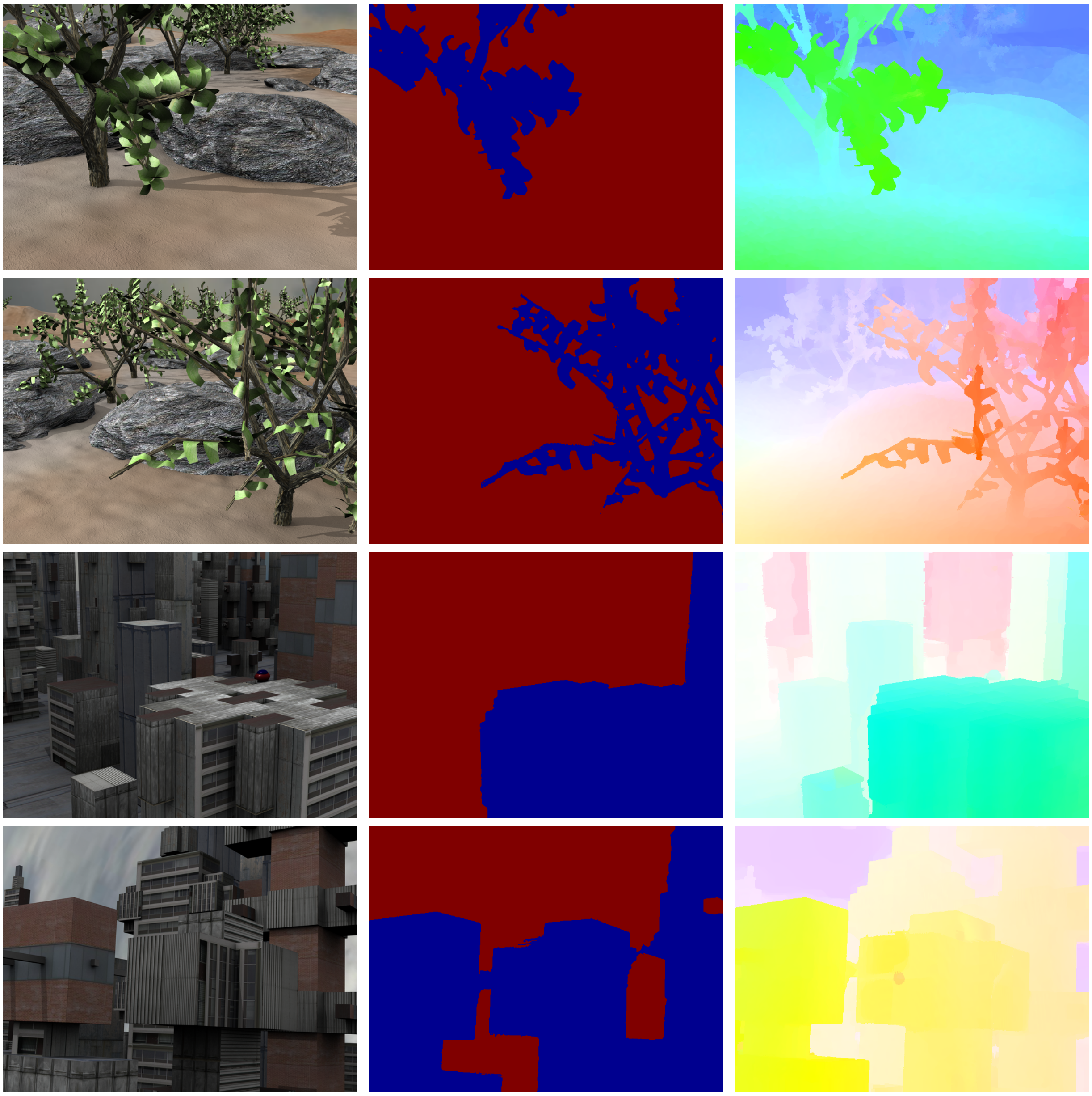


(a)First frame (b) Layer segmentation (c) Estimated flow field

Figure 9. Estimated flow fields and scene structure on the Middlebury *test* sequences. Left to right: first frame, layer segmentation, and estimated flow field. Top to bottom: “Grove”, “Urban”, “Yosemite”, and “Teddy”. Depth ordering: blue is foreground and red is background.

(a)First frame (b) Layer segmentation (c) Estimated flow field

Figure 10. Estimated flow fields and scene structure on the Middlebury *training* sequences. Left to right: first frame, layer segmentation, and estimated flow field. Top to bottom: “Venus”, “Dimetrodon”, “Hydrangea”, and “RubberWhale”. Depth ordering: blue is foreground and red is background.



(a)First frame (b) Layer segmentation (c) Estimated flow field

Figure 11. Estimated flow fields and scene structure on the Middlebury *training* sequences. Left to right: first frame, layer segmentation, and estimated flow field. Top to bottom: “Grove2”, “Grove3”, “Urban2”, and “Urban3”. Depth ordering: blue is foreground and red is background.