Supplementary Material for Slow Flow: Exploiting High-Speed Cameras for Accurate and Diverse Optical Flow Reference Data

Joel Janai[1] Fatma Güney[1] Jonas Wulff[2] Michael Black[2] Andreas Geiger[1,3]
[1]Autonomous Vision Group, MPI for Intelligent Systems Tübingen
[2]Perceiving Systems Department, MPI for Intelligent Systems Tübingen
[3]Computer Vision and Geometry Group, ETH Zürich
{joel.janai,fatma.guney,jonas.wulff,michael.black,andreas.geiger}@tue.mpg.de

Abstract

In this supplementary document, we first give some details on the normalization of the data terms used in our Flowlets in Section 1 and the derivation of the Markov random field used for the dense tracking problem in Section 2. In addition, we show with an additional experiment the impact of the frame rate on the estimation error in Section 3, we give more details on the 3D reconstruction dataset (Section 4) and on the comparison of the state-of-the-art methods with our reference data (Section 5). Finally, we show some qualitative results of our dense tracking approach, of the motion blur synthesized with the help of our method and of three state-of-the-art methods in Section 6.

1. Normalization of the Flowlets Data Term

In [5,9] the authors showed that the first-order Taylor approximation leads to a weighting of the data term according to the image gradient. This results in high weights when the linear assumption is violated and they propose to use a normalization of the data term to alleviate this problem. The normalization terms of the data terms in our Flowlets

$$\varphi_1^t(\mathbf{F}(\mathbf{p})) = \rho(\mathbf{J}_t(\mathbf{p}+t\mathbf{F}(\mathbf{p})) - \mathbf{J}_{t+1}(\mathbf{p}+(t+1)\mathbf{F}(\mathbf{p})))$$
$$\varphi_2^t(\mathbf{F}(\mathbf{p})) = \rho(\mathbf{J}_t(\mathbf{p}+t\mathbf{F}(\mathbf{p})) - \mathbf{J}_0(\mathbf{p}))$$

are derived as

$$\theta_i^t = \left(t^2 \cdot \left\|\nabla \mathbf{J}_t^i\right\|_2^2 + \epsilon^2\right)^{-1}$$
$$\theta_i^{t,t+1} = \left(\left\|t \cdot \nabla \mathbf{J}_t^i - (t+1) \cdot \nabla \mathbf{J}_{t+1}^i\right\|_2^2 + \epsilon^2\right)^{-1}$$

with θ^t, $\theta^{t,t+1}$ denoting vectors of dimension c, θ_i^t the i'th column of the vectors and $\mathbf{J}_t^i$ the i'th channel of frame $\mathbf{J}_t$. Using these normalization factors we obtain the normalized data terms

$$\varphi_1^t(\mathbf{F}(\mathbf{p})) = \rho((\theta^{t,t+1})^T \cdot (\mathbf{J}_t(\mathbf{p}+t\mathbf{F}(\mathbf{p})) - \mathbf{J}_{t+1}(\mathbf{p}+(t+1)\mathbf{F}(\mathbf{p}))))$$
$$\varphi_2^t(\mathbf{F}(\mathbf{p})) = \rho((\theta^t)^T \cdot (\mathbf{J}_t(\mathbf{p}+t\mathbf{F}(\mathbf{p})) - \mathbf{J}_0(\mathbf{p})))$$

2. Derivation of the Markov random field

For the dense tracking problem we consider the location $\mathcal{H} = \{\mathbf{H}_1, \ldots, \mathbf{H}_N\}$ and visibility state $\mathcal{V} = \{\mathbf{V}_1, \ldots, \mathbf{V}_N\}$ of each pixel in the reference image $\mathbf{I}_1$ and each frame of the sequence. Furthermore, $\mathbf{h_p} = \{\mathbf{H}_1(\mathbf{p}), \ldots, \mathbf{H}_N(\mathbf{p})\}$ is the location and $\mathbf{v_p} = \{\mathbf{V}_1(\mathbf{p}), \ldots, \mathbf{V}_N(\mathbf{p})\}$ indicates the visibility state of a reference pixel $\mathbf{p}$ in all frames.

Given the location and visibility we defined our objective

$$
\begin{aligned}
E(\mathcal{H}_*, \mathcal{V}_*) &= \lambda^{\mathcal{D}_A} \sum_{t<s} \underbrace{\psi_{ts}^{\mathcal{D}_A}(\mathbf{H}_t, \mathbf{V}_t, \mathbf{H}_s, \mathbf{V}_s)}_{\text{Appearance Data Term}} \\
&+ \lambda^{\mathcal{D}_F} \sum_{s=t+1} \underbrace{\psi_{ts}^{\mathcal{D}_F}(\mathbf{H}_t, \mathbf{V}_t, \mathbf{H}_s, \mathbf{V}_s)}_{\text{Flow Data Term}} \\
&+ \lambda^{\mathcal{F}_T} \sum_{\mathbf{p}\in\Omega} \underbrace{\psi_{\mathbf{p}}^{\mathcal{F}_T}(\mathbf{h}_\mathbf{p})}_{\text{Temporal Flow}} + \lambda^{\mathcal{F}_S} \sum_{\mathbf{p}\sim\mathbf{q}} \underbrace{\psi_{\mathbf{pq}}^{\mathcal{F}_S}(\mathbf{h}_\mathbf{p}, \mathbf{h}_\mathbf{q})}_{\text{Spatial Flow}} \\
&+ \lambda^{\mathcal{V}_T} \sum_{\mathbf{p}\in\Omega} \underbrace{\psi_{\mathbf{p}}^{\mathcal{V}_T}(\mathbf{v}_\mathbf{p})}_{\text{Temporal Vis.}} + \lambda^{\mathcal{V}_S} \sum_{\mathbf{p}\sim\mathbf{q}} \underbrace{\psi_{\mathbf{pq}}^{\mathcal{V}_S}(\mathbf{v}_\mathbf{p}, \mathbf{v}_\mathbf{q})}_{\text{Spatial Vis.}}
\end{aligned} \tag{1}
$$

Optimizing this objective is very difficult and we therefore discretize the variables as $\{(\mathbf{h}_\mathbf{p}^{(1)}, \mathbf{v}_\mathbf{p}^{(1)}), \ldots, (\mathbf{h}_\mathbf{p}^{(M)}, \mathbf{v}_\mathbf{p}^{(M)})\}$ for approximate inference. Estimating $\mathbf{X} = \{x_\mathbf{p} | \mathbf{p} \in \Omega\}$ with $x_\mathbf{p} = (\mathbf{h}_\mathbf{p}, \mathbf{v}_\mathbf{p}) \in \{(\mathbf{h}_\mathbf{p}^{(1)}, \mathbf{v}_\mathbf{p}^{(1)}), \ldots, (\mathbf{h}_\mathbf{p}^{(M)}, \mathbf{v}_\mathbf{p}^{(M)})\}$ can be phrased as inference in a simpler Markov random field:

By inserting our definitions of the data and smoothness terms we obtain

$$
\begin{aligned}
E(\mathcal{H}_*, \mathcal{V}_*) &= \lambda^{\mathcal{D}_A} \sum_{\mathbf{p}\in\Omega} \sum_{t<s} \mathbf{v}_\mathbf{p}^t \, \mathbf{v}_\mathbf{p}^s \, \|\mathbf{I}_t(\mathbf{h}_\mathbf{p}^t) - \mathbf{I}_s(\mathbf{h}_\mathbf{p}^s)\|_1 \\
&+ \lambda^{\mathcal{D}_F} \sum_{\mathbf{p}\in\Omega} \sum_{s=t+1} \mathbf{v}_\mathbf{p}^t \, \mathbf{v}_\mathbf{p}^s \, \|\mathbf{h}_\mathbf{p}^s - \mathbf{h}_\mathbf{p}^t - \mathbf{F}_{t\to s}(\mathbf{h}_\mathbf{p}^t)\|_1 \\
&+ \lambda^{\mathcal{F}_T} \sum_{\mathbf{p}\in\Omega} \sum_{t=2}^{N-1} \|\mathbf{h}_\mathbf{p}^{t-1} - 2\,\mathbf{h}_\mathbf{p}^t + \mathbf{h}_\mathbf{p}^{t+1}\|_1 + \lambda^{\mathcal{F}_S} \sum_{\mathbf{p}\sim\mathbf{q}} \xi(\mathbf{p},\mathbf{q}) \sum_{t=1}^{N} \|(\mathbf{h}_\mathbf{p}^t - \mathbf{h}_\mathbf{p}^1) - (\mathbf{h}_\mathbf{p}^t - \mathbf{h}_\mathbf{p}^1)\|_2 \\
&+ \lambda^{\mathcal{V}_T} \sum_{\mathbf{p}\in\Omega} \sum_{t=1}^{N-1} [\mathbf{v}_\mathbf{p}^t \neq \mathbf{v}_\mathbf{p}^{t+1}(\mathbf{p})] - \lambda^{\mathcal{V}} \sum_{t=1}^{N} \mathbf{v}_\mathbf{p}^t + \lambda^{\mathcal{V}_S} \sum_{\mathbf{p}\sim\mathbf{q}} \xi(\mathbf{p},\mathbf{q}) \sum_{t=1}^{N} [\mathbf{v}_\mathbf{p}^t \neq \mathbf{v}_\mathbf{q}^t]
\end{aligned}
$$

Finally, re-arranging the terms yields

$$
\begin{aligned}
E(\mathcal{H}_*, \mathcal{V}_*) &= \sum_{\mathbf{p}\in\Omega} \Bigg[\lambda^{\mathcal{D}_A} \sum_{t<s} \mathbf{v}_\mathbf{p}^t \, \mathbf{v}_\mathbf{p}^s \, \|\mathbf{I}_t(\mathbf{h}_\mathbf{p}^t) - \mathbf{I}_s(\mathbf{h}_\mathbf{p}^s)\|_1 + \lambda^{\mathcal{D}_F} \sum_{s=t+1} \mathbf{v}_\mathbf{p}^t \, \mathbf{v}_\mathbf{p}^s \, \|\mathbf{h}_\mathbf{p}^s - \mathbf{h}_\mathbf{p}^t - \mathbf{F}_{t\to s}(\mathbf{h}_\mathbf{p}^t)\|_1 \\
&\qquad + \lambda^{\mathcal{F}_T} \sum_{t=2}^{N-1} \|\mathbf{h}_\mathbf{p}^{t-1}(\mathbf{p}) - 2\,\mathbf{h}_\mathbf{p}^t + \mathbf{h}_\mathbf{p}^{t+1}\|_1 + \lambda^{\mathcal{V}_T} \sum_{t=1}^{N-1} [\mathbf{v}_\mathbf{p}^t \neq \mathbf{v}_\mathbf{p}^{t+1}] - \lambda^{\mathcal{V}} \sum_{t=1}^{N} \mathbf{v}_\mathbf{p}^t \Bigg] \\
&+ \sum_{\mathbf{p}\sim\mathbf{q}} \xi(\mathbf{p},\mathbf{q}) \sum_{t=1}^{N} \left[\lambda^{\mathcal{F}_S} \xi(\mathbf{p},\mathbf{q}) \|(\mathbf{h}_\mathbf{p}^t - \mathbf{h}_\mathbf{p}^1) - (\mathbf{h}_\mathbf{p}^t - \mathbf{h}_\mathbf{p}^1)\|_2 + \lambda^{\mathcal{V}_S} [\mathbf{v}_\mathbf{p}^t \neq \mathbf{v}_\mathbf{q}^t] \right]
\end{aligned}
$$

which can be written as

$$
E(\mathbf{X}) = \sum_{\mathbf{p}} \psi_\mathbf{p}^{\mathcal{U}}(x_\mathbf{p}) + \sum_{\mathbf{p}\sim\mathbf{q}} \psi_{\mathbf{pq}}^{\mathcal{P}}(x_\mathbf{p}, x_\mathbf{q}) \tag{2}
$$

where $x_\mathbf{p} = (\mathbf{h}_\mathbf{p}, \mathbf{v}_\mathbf{p})$ and

$$
\begin{aligned}
\psi_\mathbf{p}^{\mathcal{U}}(x_\mathbf{p}) &= \lambda^{\mathcal{D}_A} \sum_{t<s} \mathbf{v}_\mathbf{p}^t \, \mathbf{v}_\mathbf{p}^s \, \|\mathbf{I}_t(\mathbf{h}_\mathbf{p}^t) - \mathbf{I}_s(\mathbf{h}_\mathbf{p}^s)\|_1 + \lambda^{\mathcal{D}_F} \sum_{s=t+1} \mathbf{v}_\mathbf{p}^t \, \mathbf{v}_\mathbf{p}^s \, \|\mathbf{h}_\mathbf{p}^s - \mathbf{h}_\mathbf{p}^t - \mathbf{F}_{t\to s}(\mathbf{h}_\mathbf{p}^t)\|_1 \\
&+ \lambda^{\mathcal{F}_T} \sum_{t=2}^{N-1} \|\mathbf{h}_\mathbf{p}^{t-1}(\mathbf{p}) - 2\,\mathbf{h}_\mathbf{p}^t + \mathbf{h}_\mathbf{p}^{t+1}\|_1 + \lambda^{\mathcal{V}_T} \sum_{t=1}^{N-1} [\mathbf{v}_\mathbf{p}^t \neq \mathbf{v}_\mathbf{p}^{t+1}] - \lambda^{\mathcal{V}} \sum_{t=1}^{N} \mathbf{v}_\mathbf{p}^t \\
\psi_{\mathbf{pq}}^{\mathcal{P}}(x_\mathbf{p}, x_\mathbf{q}) &= \xi(\mathbf{p},\mathbf{q}) \sum_{t=1}^{N} \left[\lambda^{\mathcal{F}_S} \xi(\mathbf{p},\mathbf{q}) \|(\mathbf{h}_\mathbf{p}^t - \mathbf{h}_\mathbf{p}^1) - (\mathbf{h}_\mathbf{p}^t - \mathbf{h}_\mathbf{p}^1)\|_2 + \lambda^{\mathcal{V}_S} [\mathbf{v}_\mathbf{p}^t \neq \mathbf{v}_\mathbf{q}^t] \right]
\end{aligned}
$$

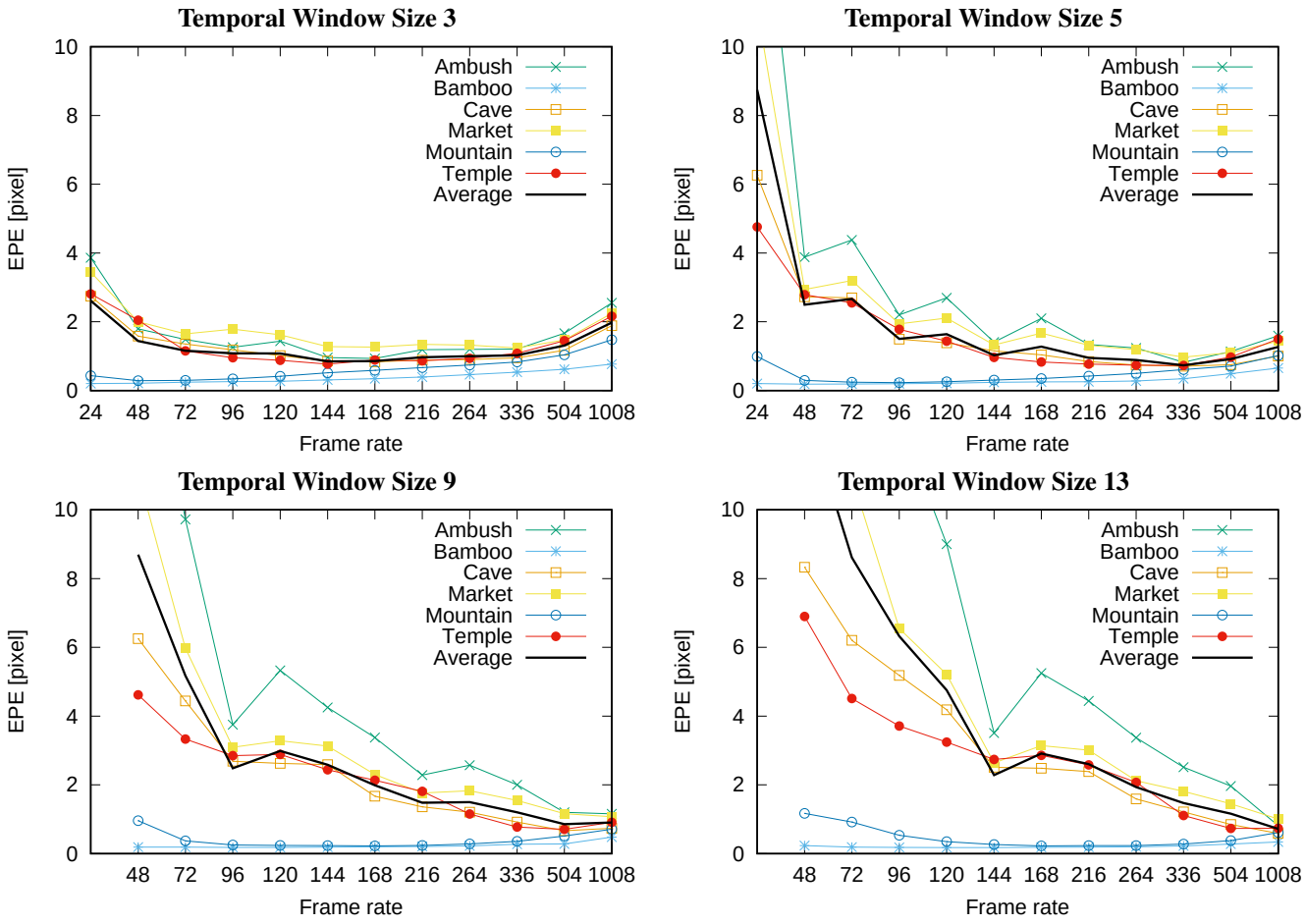

Figure 1: EPE (MPI Sintel) of non occluded pixels for different frame rates (x-axis) using temporal window size 3,5,9 and 13 for the Flowlets and the naive accumulation.

3. Importance of Frame rate

Our high frame rate version of the Sintel dataset allows us to analyse the impact of frame rate on the performance. In the following experiment we use the naive accumulation of Flowlets on different frame rates to compare them on the original frame rate of 24fps. In addition, we exclude occluded regions since the naive accumulation ignores occlusions. We obtain the different frame rates by skipping frames at the highest frame rate of 1008fps. *Note:* For some frame rates linear interpolation is necessary to obtain the flow in the original frame rate of 24fps. In the sequences Ambush and Market the linear interpolation of non-linear motions cause errors that are not completely following the expected trend.

In Fig. 1 we show the estimation error (EPE) with a temporal window size of 3, 5, 9 and 13 frames using different frame rates (x-axis). Overall, we observe decreasing errors with higher frame rates. Interestingly, however, this holds true only until a certain frame rate for a temporal window size of 3 and 5 frames. The reason for the error to increase after the optimal frame rate, is the accumulation of small estimation errors that causes a significant drift for higher frame rates. Larger temporal windows perform weaker on lower frame rates but at the same time show a smaller drift over time. Using a larger temporal window can be considered as using a lower frame rate with additional temporal information since we impose the hard constant velocity constraint over a longer time period. Therefore, the optimal frame rate is higher with larger temporal frame rates than with smaller ones, e.g. 144fps using 3 frames and 504 fps using 9 frames temporal windows. The optimal frame rate also highly depends on the scene. Whereas higher frame rates perform better for Ambush, Cave, Market and Temple, the lowest frame rate of 24fps is optimal for Bamboo and Mountain. In Bamboo and Mountain we have already very small motions in the original frame rate thus higher frame rates only lead to a larger drift.

Figure 2: The reconstructed point clouds for 4 different scenes used for evaluating the reference flow fields.

We can therefore conclude as in [6] that while very high frame rates help in general, the optimal frame rate depends not only on the available resources, but also on the imaging modalities and the scenario at hand. Thus, it is impossible to choose a single optimal frame rate across all sequences. We therefore use an adaptive frame rate according to the 90% quantile in our dense tracking approach as described in the paper.

4. 3D Reconstruction Dataset

Our reconstruction dataset consists of 4 point clouds with in total 20 sequences for the evaluation. The point clouds, after removing outliers manually, are shown in Fig. 2. In the 20 sequences we used different camera motions and viewpoints for a diverse dataset. In Fig. 3 we show the generated flow fields from our approach for different flow magnitudes and in Fig. 4 we compare our flow fields to the Epic Flow baseline. All flow illustrations are generated using the Middlebury [1] color scheme and for the comparison of two flow fields we normalize by the maximum flow of both.

For 100px magnitudes we observed similar performance of our approach and Epic Flow. Therefore we omitted these flow fields here. In case of occlusions, Epic Flow is having trouble to estimate the correct flow whereas our motion boundaries are much better (first and second row). At the same time, we observe that repetitive patterns as the bench moving out of the image are very troublesome for Epic Flow but can easily be handled by our approach. Furthermore, fine details are better maintained with our approach than Epic Flow as can be seen in the third and forth row.

5. Real-World Benchmark

In tables 1, 2, 3 we show the performance of several state-of-the-art optical flow methods (evaluated against our reference flow fields) on different scene types (columns), and using different levels of motion blur (rows). We grouped our sequences in different scenes with similar objects and motions. In the scenes Motocross, Skatepark, Rally, Kids and Ball we have only a few objects moving in contrast to Marathon, Town, Road and Animals. Furthermore, the objects in Kids, Skatepark, Marathon, Animals and Town are mostly non-rigid which causes complex non-linear motion and self occlusions.

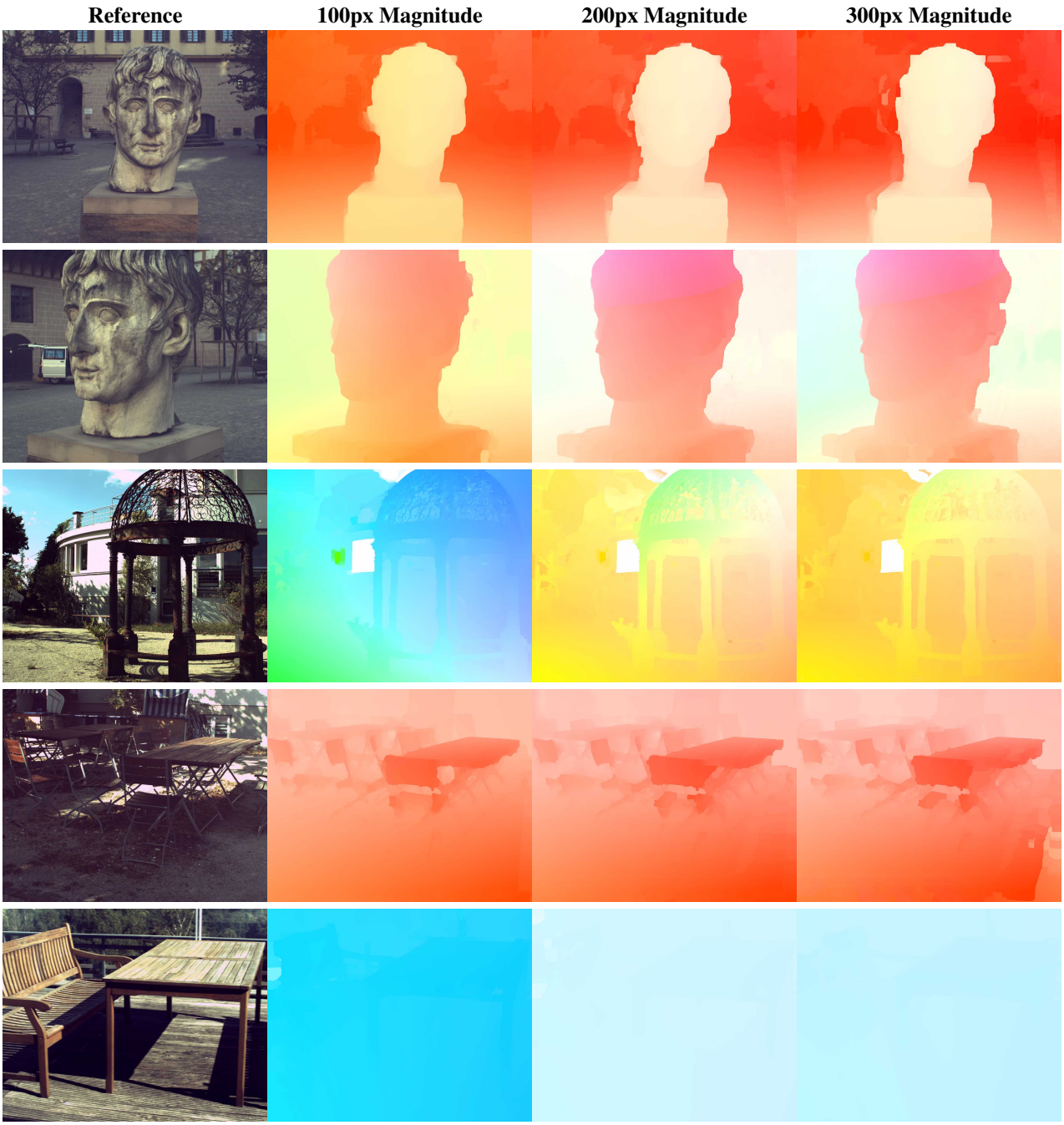

Figure 3: Slow Flow estimation examples with 100px, 200px, 300px motion magnitudes for the reconstruction dataset.

We observe some methods to have particular difficulties in the scenes Motocross, Town, Rally and Road. In Motocross for instance Full Flow [3], ClassicNL [10], LDOF [2] and FlowNet [4] achieve around 5 to 6 pixel EPE whereas the others achieve 2 to 3 pixel EPE in the simple case of 100px motion magnitude. This gap still remains with larger motion magnitudes. In the scene Town the methods Full Flow [3], FlowNet [4] and in Rally the methods Full Flow [3], LDOF [2] have this kind of difficulties. Besides these difficulties with some scenes, we observe the strongest impact of larger motion magnitudes in scenes with non-rigid objects. For all methods the performance decreases strongly for larger (200 and 300 pixel) motion magnitudes. In Skatepark, for example, the EPE of the best performing method DiscreteFlow [7] increases from 1.71 to

Figure 4: Comparison of Slow Flow and Epic Flow estimations on the reconstruction dataset.

3.79 and 6.31 pixel. The reasons are complex non-linear motions, appearance changes and self occlusions that become more problematic with larger motion magnitudes. With stronger motion blur we observe the strongest loss in performance for all methods in the scenes Town and Animals. This is primarily caused by a complex non-linear camera motion that makes it hard to find good matches in the large background region. Only the two approaches FlowNet [4] and ClassicNL [10], not using feature matches, are not affected as strongly as the others.

In conclusion, the performance on the different scenes are giving us important insights into the strengths and weaknesses of the different methods. Whereas the methods Full Flow [3], ClassicNL [10], LDOF [2] and FlowNet [4] have difficulties

Blur	Method	Kids	Motocross	Skatepark	Marathon	Town	Rally	Road	Animals	Ball	Avg
0	Discrete Flow	1.16	2.00	1.71	2.22	0.66	0.63	3.28	0.76	0.21	1.62
	Full Flow	1.22	6.68	1.99	2.61	4.30	6.84	5.33	2.62	2.07	3.99
	ClassicNL	1.67	6.87	3.35	3.82	0.88	1.57	5.51	1.34	1.66	3.12
	Epic Flow	1.57	3.35	2.61	2.84	1.19	1.24	3.97	0.99	0.36	2.26
	Flow Fields	1.19	2.71	1.92	2.69	0.68	0.71	3.63	0.78	0.23	1.83
	LDOF	1.59	5.60	3.27	3.02	1.30	5.85	6.81	1.88	0.72	3.67
	PCA Flow	2.15	2.75	3.17	3.49	1.33	1.70	4.45	1.87	0.66	2.67
	FlowNetS	2.38	5.33	3.75	4.47	3.31	2.44	6.43	1.88	1.91	3.93
	SPyNet	1.70	6.62	3.28	3.83	2.01	1.68	5.80	1.11	2.67	3.36
1	Discrete Flow	1.24	2.43	2.02	2.50	3.10	0.99	3.37	1.76	0.32	2.37
	Full Flow	1.29	6.90	2.42	2.93	5.34	7.10	5.32	2.95	1.38	4.34
	ClassicNL	1.73	6.98	3.54	4.03	1.92	1.63	5.49	1.84	1.34	3.43
	Epic Flow	1.63	4.17	2.83	3.14	2.72	1.43	4.40	1.47	0.60	2.85
	Flow Fields	1.28	2.90	2.18	2.91	1.93	0.93	3.54	1.28	0.31	2.21
	LDOF	1.67	6.29	3.71	3.57	4.04	2.31	7.17	3.43	0.76	4.35
	PCA Flow	2.24	3.07	3.50	3.70	5.64	1.85	4.61	3.23	0.70	3.79
	FlowNetS	2.41	5.00	3.82	4.45	3.30	2.22	6.46	1.95	1.51	3.89
	SPyNet	1.77	6.88	3.47	4.05	2.41	1.94	5.77	1.49	3.13	3.58
3	Discrete Flow	1.32	3.24	2.30	3.26	7.42	1.22	3.54	3.28	0.54	3.60
	Full Flow	1.40	7.27	2.68	3.02	6.35	6.18	5.13	2.89	1.22	4.48
	ClassicNL	1.79	6.82	3.59	4.22	2.88	1.70	5.38	2.05	1.37	3.63
	Epic Flow	1.66	4.95	3.25	3.29	3.48	1.59	4.58	2.03	2.21	3.29
	Flow Fields	1.32	3.20	2.41	3.32	3.11	1.12	3.71	1.74	0.57	2.64
	LDOF	1.87	6.67	4.01	3.58	5.36	2.28	7.42	4.45	1.04	4.87
	PCA Flow	2.70	4.26	4.06	3.89	9.10	1.96	4.93	4.93	1.15	4.97
	FlowNetS	2.47	5.92	4.03	4.57	3.56	2.23	6.52	2.11	1.62	4.10
	SPyNet	1.85	7.03	3.62	4.11	2.73	2.16	5.77	1.75	3.41	3.74
5	Discrete Flow	1.72	4.49	3.55	4.40	23.42	1.79	3.97	7.73	1.00	7.70
	Full Flow	1.73	7.57	3.43	3.88	9.20	4.78	5.50	4.34	1.65	5.40
	ClassicNL	2.12	7.25	4.31	4.84	5.22	2.15	5.50	3.67	1.52	4.56
	Epic Flow	2.03	5.84	3.91	3.99	6.63	1.98	4.86	3.75	3.68	4.48
	Flow Fields	1.68	4.75	3.59	4.07	6.54	1.65	4.29	3.70	1.20	4.10
	LDOF	2.24	7.73	5.06	4.10	7.94	2.57	7.80	6.31	1.63	6.02
	PCA Flow	3.34	5.32	5.44	5.20	18.08	2.51	5.20	10.38	1.72	7.91
	FlowNetS	2.64	6.36	4.31	4.99	5.07	2.49	6.54	2.67	1.82	4.61
	SPyNet	2.10	7.61	4.10	4.50	4.07	2.68	5.98	2.53	3.57	4.35
7	Discrete Flow	2.33	5.99	5.12	4.92	14.71	2.42	4.60	9.87	1.58	7.01
	Full Flow	2.17	8.37	4.61	4.77	11.95	5.46	5.88	6.86	2.07	6.73
	ClassicNL	2.56	8.13	5.28	5.55	7.90	2.85	5.73	5.69	2.00	5.74
	Epic Flow	2.58	7.16	4.98	5.03	15.38	2.62	5.21	6.43	4.95	6.99
	Flow Fields	2.29	6.03	4.82	6.04	10.91	2.51	4.69	7.16	1.82	6.02
	LDOF	2.83	8.79	6.12	4.94	11.03	3.20	8.22	8.36	2.56	7.38
	PCA Flow	4.54	7.99	7.84	8.24	23.86	3.36	5.75	18.57	2.90	11.20
	FlowNetS	2.92	7.17	4.66	5.56	6.89	3.03	6.67	3.55	2.15	5.32
	SPyNet	2.48	8.33	4.70	5.05	5.99	3.24	6.26	3.50	3.76	5.16

Table 1: State-of-the-art comparison on the generated reference data with 100 pixel motion magnitude wrt. motion blur.

with some scenes in general, the motion magnitude has an adversarial effect on all methods. The motion magnitude affects the performance in particular when dealing with non-rigid objects and the motion blur is problematic for feature matching methods, especially with complex camera motion.

6. Qualitative Results

In Figures 5 and 6, we show some qualitative results for our proposed approach. In the scenes Ball, Marathon, Motocross, Rally, Road and Town, the resulting flow fields look almost perfect but there are some errors in Animals, Kids and Skatepark. On the one hand, details are missing, e.g. parts of the wheel in Skatepark, and on the other hand errors occur in occluded regions, e.g. next to the head of the right kid with 300px magnitude and under the head of the horse. In these cases, a confidence measure would allow us to exclude these problematic areas. Therefore, we are interested in a probabilistic version of our approach for future work.

The different level of motion blur for the corresponding reference frames are shown in Figures 7 and 8. In scenes Animals,

Blur	Method	Kids	Motocross	Skatepark	Marathon	Town	Rally	Road	Animals	Ball	Avg
0	Discrete Flow	2.44	5.26	3.79	3.96	0.87	0.59	3.58	1.51	0.41	2.63
	Full Flow	4.56	15.12	8.99	8.42	4.69	14.94	9.95	15.15	10.40	9.93
	ClassicNL	4.12	12.18	8.27	8.99	1.44	2.55	8.57	3.55	9.33	6.14
	Epic Flow	3.30	6.55	5.42	5.14	1.52	0.96	5.41	1.78	1.49	3.71
	Flow Fields	2.62	4.77	4.39	5.90	0.93	0.61	5.39	1.67	0.89	3.26
	LDOF	3.67	9.10	6.57	5.60	1.78	6.41	10.96	2.90	8.48	6.08
	PCA Flow	3.78	6.18	6.36	5.28	1.66	1.68	6.18	2.79	1.97	4.24
	FlowNetS	4.84	10.77	7.21	8.67	3.88	3.48	8.84	3.85	6.09	6.44
	SPyNet	3.90	14.40	7.16	7.79	2.61	2.35	9.17	3.41	11.43	6.42
1	Discrete Flow	2.50	7.61	4.21	4.15	3.20	1.02	3.78	3.51	1.01	3.71
	Full Flow	4.59	13.94	8.90	8.82	5.73	14.52	10.06	15.49	11.27	10.10
	ClassicNL	4.13	11.84	8.60	9.37	2.46	2.66	8.44	4.06	8.36	6.41
	Epic Flow	3.47	7.93	5.84	7.04	3.10	1.17	5.45	2.34	5.09	4.53
	Flow Fields	2.52	5.36	4.26	5.53	2.08	0.95	5.45	2.37	1.43	3.62
	LDOF	3.78	10.56	7.35	6.42	4.61	2.85	11.12	6.90	7.83	7.21
	PCA Flow	4.16	6.57	6.61	6.80	6.03	1.78	6.73	4.62	2.90	5.66
	FlowNetS	4.84	10.81	7.37	8.82	3.86	3.40	8.73	4.04	5.39	6.45
	SPyNet	3.92	14.57	7.47	7.99	2.98	2.60	9.09	3.72	11.70	6.62
3	Discrete Flow	2.97	7.35	4.93	6.11	7.34	1.34	4.20	8.03	2.29	5.50
	Full Flow	4.62	14.20	8.76	8.65	6.71	12.40	10.00	14.84	10.33	9.99
	ClassicNL	4.23	11.80	8.59	9.51	3.43	2.74	8.18	4.23	7.94	6.56
	Epic Flow	3.65	8.62	6.10	6.58	3.93	1.42	5.85	3.36	9.04	5.08
	Flow Fields	2.87	6.28	4.91	7.03	3.40	1.05	5.21	2.78	2.50	4.21
	LDOF	4.29	12.93	8.87	6.61	6.09	2.89	11.79	10.53	7.90	8.60
	PCA Flow	5.12	7.85	10.61	7.67	11.36	1.91	6.69	7.26	4.02	7.84
	FlowNetS	4.91	11.20	7.67	9.07	4.12	3.50	8.63	4.50	5.46	6.65
	SPyNet	3.97	14.79	7.67	8.18	3.29	2.79	9.07	3.94	12.25	6.80
5	Discrete Flow	3.68	11.00	8.33	7.93	19.97	1.92	4.96	17.02	4.12	10.31
	Full Flow	4.64	15.19	9.92	8.77	9.62	9.67	9.80	14.76	9.43	10.51
	ClassicNL	4.42	12.59	9.14	9.95	5.70	3.40	8.17	5.57	7.78	7.41
	Epic Flow	4.18	12.39	6.69	7.11	7.69	1.68	6.31	6.40	12.24	6.88
	Flow Fields	3.08	8.52	6.76	7.58	6.45	1.66	5.74	6.70	4.33	6.03
	LDOF	5.00	14.92	11.63	7.30	8.71	3.35	11.81	12.92	7.72	10.11
	PCA Flow	6.59	9.04	11.92	9.85	19.58	3.41	7.57	16.50	6.71	11.55
	FlowNetS	5.18	11.49	8.05	9.65	5.59	3.68	8.59	4.80	5.74	7.12
	SPyNet	4.14	14.81	8.19	8.74	4.65	3.24	9.10	4.64	13.24	7.34
7	Discrete Flow	4.50	10.82	11.06	9.12	16.49	2.81	5.63	20.52	6.55	10.93
	Full Flow	5.01	15.43	10.00	8.73	12.18	8.62	10.24	14.19	9.68	10.97
	ClassicNL	4.78	13.55	10.06	10.55	8.21	4.47	8.14	7.19	8.17	8.46
	Epic Flow	4.74	14.66	7.80	7.62	14.28	2.46	6.76	15.12	13.01	9.90
	Flow Fields	3.86	10.45	7.50	8.96	11.34	3.04	6.08	10.25	6.13	8.07
	LDOF	5.63	16.44	13.22	8.39	11.61	4.61	12.32	15.49	8.63	11.70
	PCA Flow	7.69	12.93	15.59	12.65	28.13	4.03	8.31	20.76	9.18	15.12
	FlowNetS	5.59	12.02	8.52	10.46	7.40	4.06	8.74	5.67	6.54	7.86
	SPyNet	4.41	15.02	8.68	9.51	6.54	3.78	9.10	5.52	14.05	8.03

Table 2: State-of-the-art comparison on the generated reference data with 200 pixel motion magnitude wrt. motion blur.

Ball, Kids, Motocross and Rally, the foreground has the dominant motion which yields a stronger blur on the foreground whereas the scenes Skatepark, Road and Town also have a quite blurry background. All in all, the motion blur seems very realistic and creates new interesting challenges in our benchmark.

Finally, we show some qualitative results for DiscreteFlow [7], Epic Flow [8] and FlowNet [4] on the 300px magnitude reference data without blur (Fig. 9) and with blur length 7 (Fig. 10). The illustrations are normalized by the maximum flow of the reference data. The results discussed in Section 5 can also be observed in the visualization of the flow fields. Without blur, DiscreteFlow is the closest to the reference data in almost all sequences. Only in the Skatepark sequence the three methods are failing completely. Increasing the blur length to 7 frames strongly affects the performance of DiscreteFlow and Epic Flow whereas FlowNet still achieves quite good results.

Blur	Method	Kids	Motocross	Skatepark	Marathon	Town	Rally	Road	Animals	Ball	Avg
0	Discrete Flow	4.74	7.27	6.31	5.69	1.34	1.38	7.43	2.65	4.72	4.62
	Full Flow	9.43	29.38	21.68	14.34	14.28	25.11	22.05	43.26	11.68	22.71
	ClassicNL	7.60	14.84	13.55	16.44	4.99	3.27	16.73	8.19	10.59	10.97
	Epic Flow	6.11	8.02	8.80	7.45	2.23	1.62	11.38	3.85	5.48	6.44
	Flow Fields	4.84	7.63	7.59	9.47	1.46	0.89	12.07	2.66	5.06	6.10
	LDOF	7.44	14.91	12.92	13.28	4.41	6.84	21.15	31.44	10.50	14.92
	PCA Flow	6.93	9.11	9.18	7.47	2.47	2.20	13.40	4.38	6.00	7.24
	FlowNetS	8.56	17.11	12.83	15.07	6.34	5.31	16.29	18.86	8.75	12.76
	SPyNet	7.71	17.95	12.38	14.43	5.95	3.48	17.36	13.55	12.21	12.02
1	Discrete Flow	5.07	7.59	7.59	6.67	3.67	1.96	7.75	5.30	5.05	5.84
	Full Flow	9.75	28.73	21.80	14.37	15.22	25.82	22.23	43.83	11.97	23.05
	ClassicNL	7.58	14.81	13.89	16.62	6.06	3.50	16.78	8.33	9.94	11.25
	Epic Flow	6.37	9.60	8.96	9.12	3.92	1.84	12.18	4.25	7.61	7.35
	Flow Fields	4.97	7.33	7.73	10.69	2.86	1.39	11.54	3.45	5.34	6.50
	LDOF	7.68	15.97	14.28	13.94	8.99	3.60	21.37	32.49	9.93	16.03
	PCA Flow	7.79	9.25	9.74	10.42	8.85	3.07	13.63	11.85	6.63	9.95
	FlowNetS	8.70	17.40	12.68	15.21	6.00	4.95	16.08	19.79	8.16	12.77
	SPyNet	7.72	18.37	12.60	14.49	6.21	3.76	17.18	13.43	12.62	12.12
3	Discrete Flow	5.59	8.71	8.18	9.70	10.40	2.24	7.77	7.00	6.09	7.82
	Full Flow	10.03	29.65	21.44	14.78	15.84	24.35	22.42	44.49	11.37	23.25
	ClassicNL	7.56	14.71	13.93	15.98	7.30	3.60	16.57	7.49	9.55	11.27
	Epic Flow	6.95	10.32	8.55	8.26	4.30	2.06	11.59	4.57	9.59	7.38
	Flow Fields	5.70	9.69	8.18	10.99	4.20	1.50	11.61	4.00	6.19	7.20
	LDOF	8.42	18.86	15.70	14.49	11.54	3.95	21.93	34.19	9.92	17.41
	PCA Flow	10.12	10.24	11.48	11.75	18.95	3.18	13.85	16.51	7.51	13.11
	FlowNetS	9.02	17.96	13.10	15.94	7.13	5.14	15.95	21.27	8.25	13.36
	SPyNet	7.83	18.75	12.87	14.60	6.55	4.03	17.05	13.13	13.01	12.23
5	Discrete Flow	7.18	11.18	11.34	11.86	18.39	3.40	9.14	19.43	7.18	12.33
	Full Flow	10.35	30.34	22.48	15.60	17.93	18.85	22.30	45.90	10.90	23.64
	ClassicNL	7.79	15.40	14.75	16.37	9.64	4.47	16.25	8.32	9.45	12.03
	Epic Flow	7.89	14.27	10.01	9.82	8.03	2.64	12.25	7.27	13.03	9.43
	Flow Fields	6.33	13.63	9.82	13.15	7.77	3.79	12.53	7.66	7.45	9.54
	LDOF	9.83	19.75	19.61	15.24	14.22	5.32	22.13	34.62	9.65	18.87
	PCA Flow	10.91	12.53	18.00	22.38	31.03	4.72	15.06	27.50	9.24	19.26
	FlowNetS	9.56	18.09	13.84	16.32	10.86	5.30	15.79	21.24	8.41	14.20
	SPyNet	8.09	19.35	13.35	15.03	7.87	4.67	17.01	13.26	13.73	12.72
7	Discrete Flow	8.30	12.48	12.69	11.99	21.10	4.83	10.71	35.31	9.02	15.88
	Full Flow	10.83	29.48	22.88	16.45	19.06	16.31	22.64	45.52	11.36	23.75
	ClassicNL	8.20	16.22	15.44	17.20	11.70	5.47	15.87	8.87	9.81	12.75
	Epic Flow	8.32	16.49	10.75	11.53	19.46	3.49	12.83	12.49	13.86	12.91
	Flow Fields	8.46	12.93	11.94	14.15	13.71	4.47	12.72	10.45	8.81	11.59
	LDOF	10.54	21.29	22.36	15.33	16.18	8.13	22.26	34.78	10.28	20.09
	PCA Flow	12.41	13.80	23.17	16.25	33.66	5.23	16.33	39.06	11.17	22.11
	FlowNetS	10.18	18.59	14.31	16.78	12.69	5.60	15.77	17.72	9.00	14.27
	SPyNet	8.40	19.98	13.91	15.56	9.41	5.32	16.96	13.42	14.31	13.27

Table 3: State-of-the-art comparison on the generated reference data with 300 pixel motion magnitude wrt. motion blur.

References

[1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision (IJCV)*, 92:1–31, 2011. 4

[2] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 33:500–513, March 2011. 5, 6

[3] Q. Chen and V. Koltun. Full flow: Optical flow estimation by global optimization over regular grids. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016. 5, 6

[4] A. Dosovitskiy, P. Fischer, E. Ilg, P. Haeusser, C. Hazirbas, V. Golkov, P. v.d. Smagt, D. Cremers, and T. Brox. Flownet: Learning optical flow with convolutional networks. In *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*, 2015. 5, 6, 8

[5] S. Lai and B. C. Vemuri. Reliable and efficient computation of optical flow. *International Journal of Computer Vision (IJCV)*, 29(2):87–105, 1998. 1

[6] S. Lim, J. G. Apostolopoulos, and A. E. Gamal. Optical flow estimation using temporally oversampled video. *IEEE Trans. on Image Processing (TIP)*, 14(8):1074–1087, 2005. 4

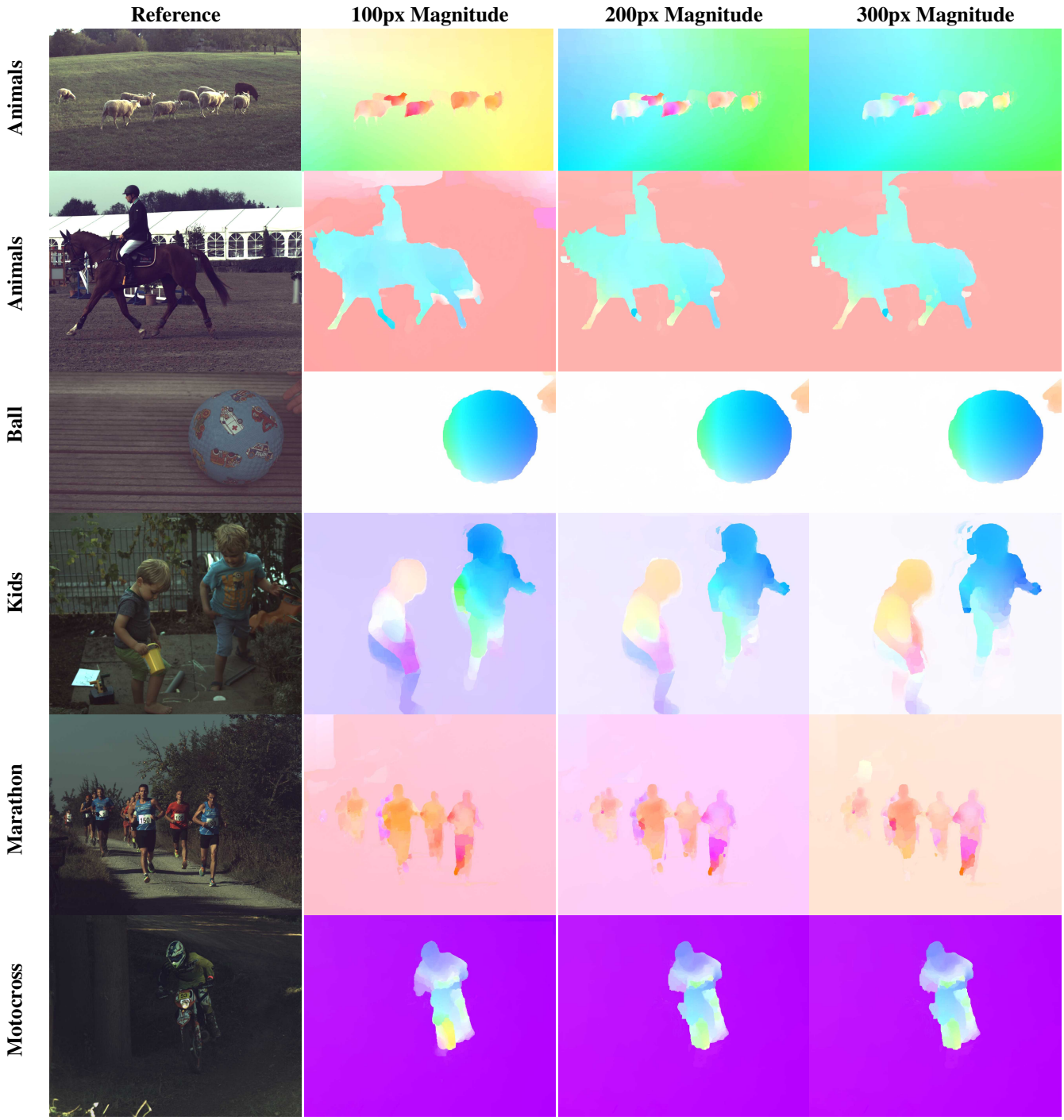

Figure 5: Reference data of different flow magnitudes for some real-world sequences.

[7] M. Menze, C. Heipke, and A. Geiger. Discrete optimization for optical flow. In *Proc. of the German Conference on Pattern Recognition (GCPR)*, 2015. 5, 8

[8] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid. EpicFlow: Edge-preserving interpolation of correspondences for optical flow. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015. 8

[9] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1991. 1

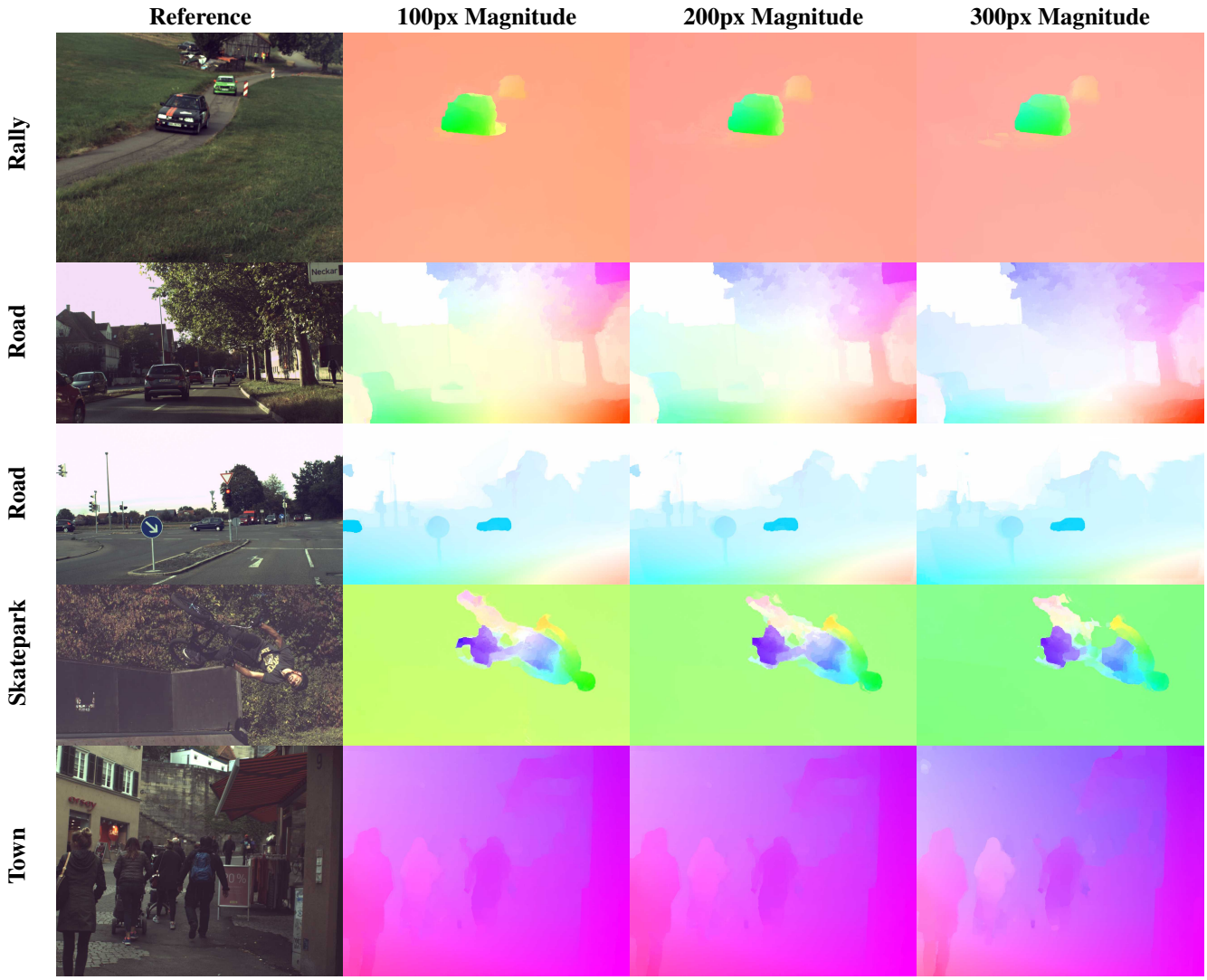

Figure 6: Reference data of different flow magnitudes for some real-world sequences.

[10] D. Sun, S. Roth, and M. J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision (IJCV)*, 106(2):115–137, 2014. 5, 6

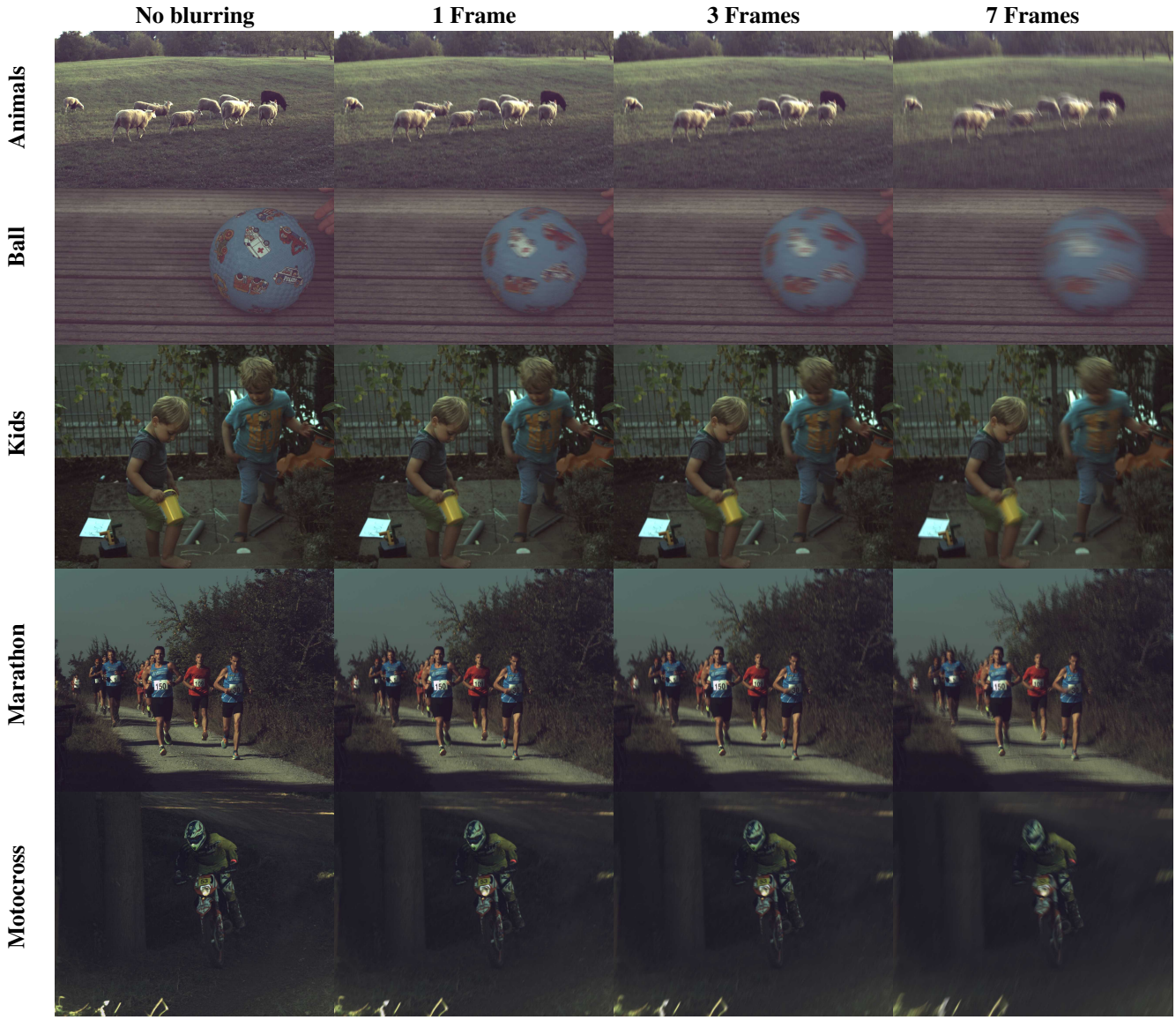

Figure 7: Different levels of blurring of the reference frame.

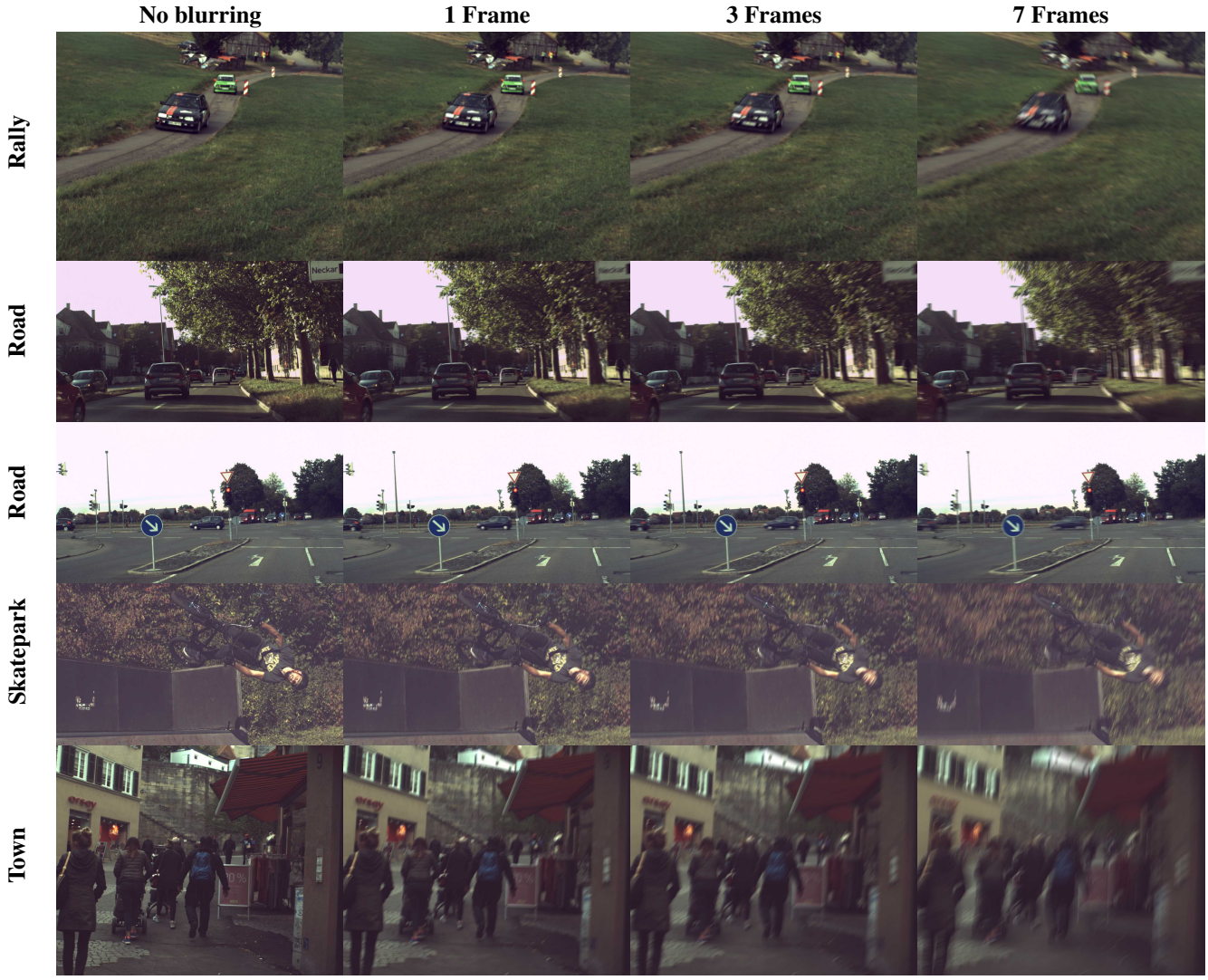

Figure 8: Different levels of blurring of the reference frame.

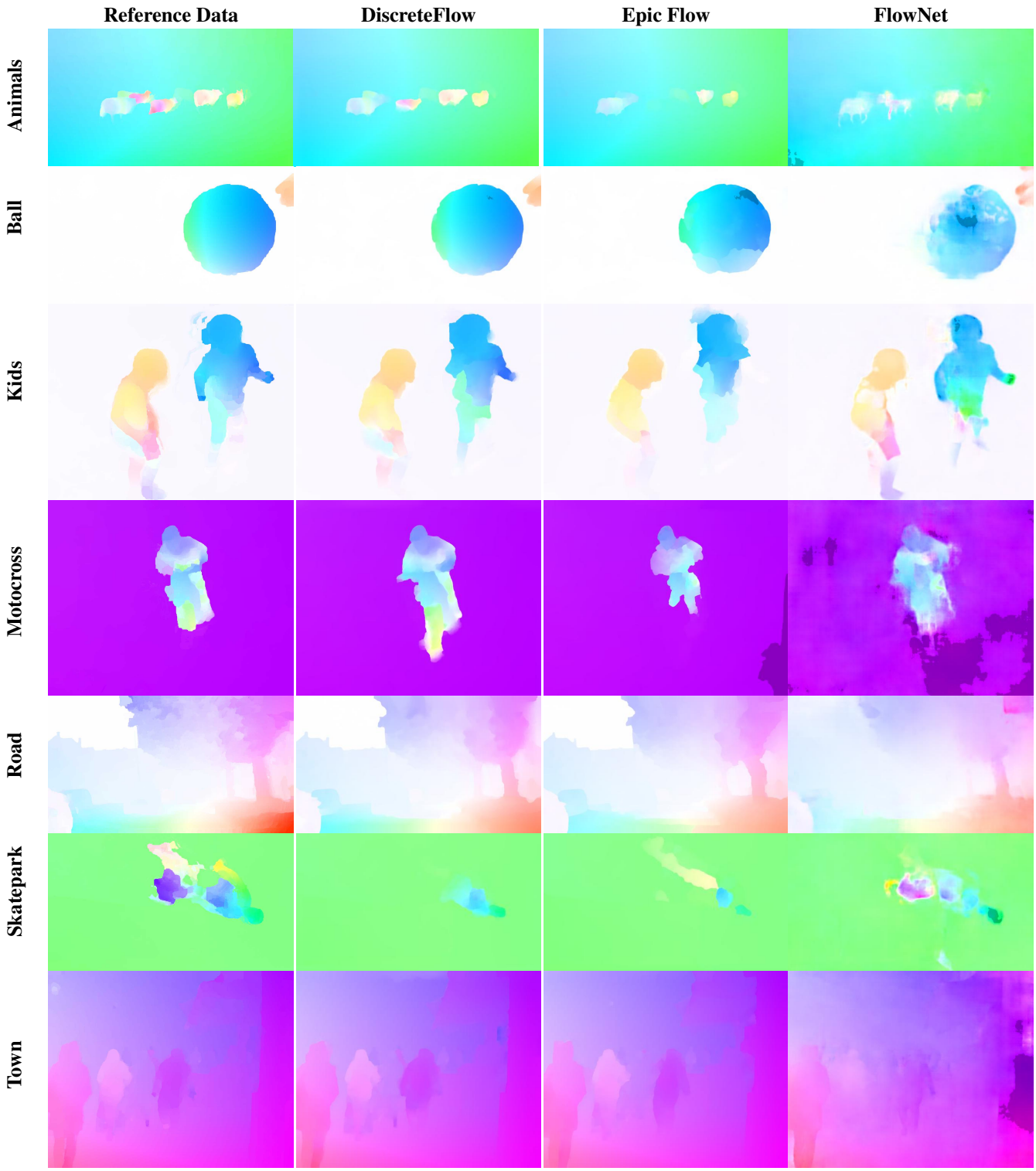

Figure 9: Comparison of DiscreteFlow, Epic Flow and FlowNet without blur to the reference data of 300px magnitude.

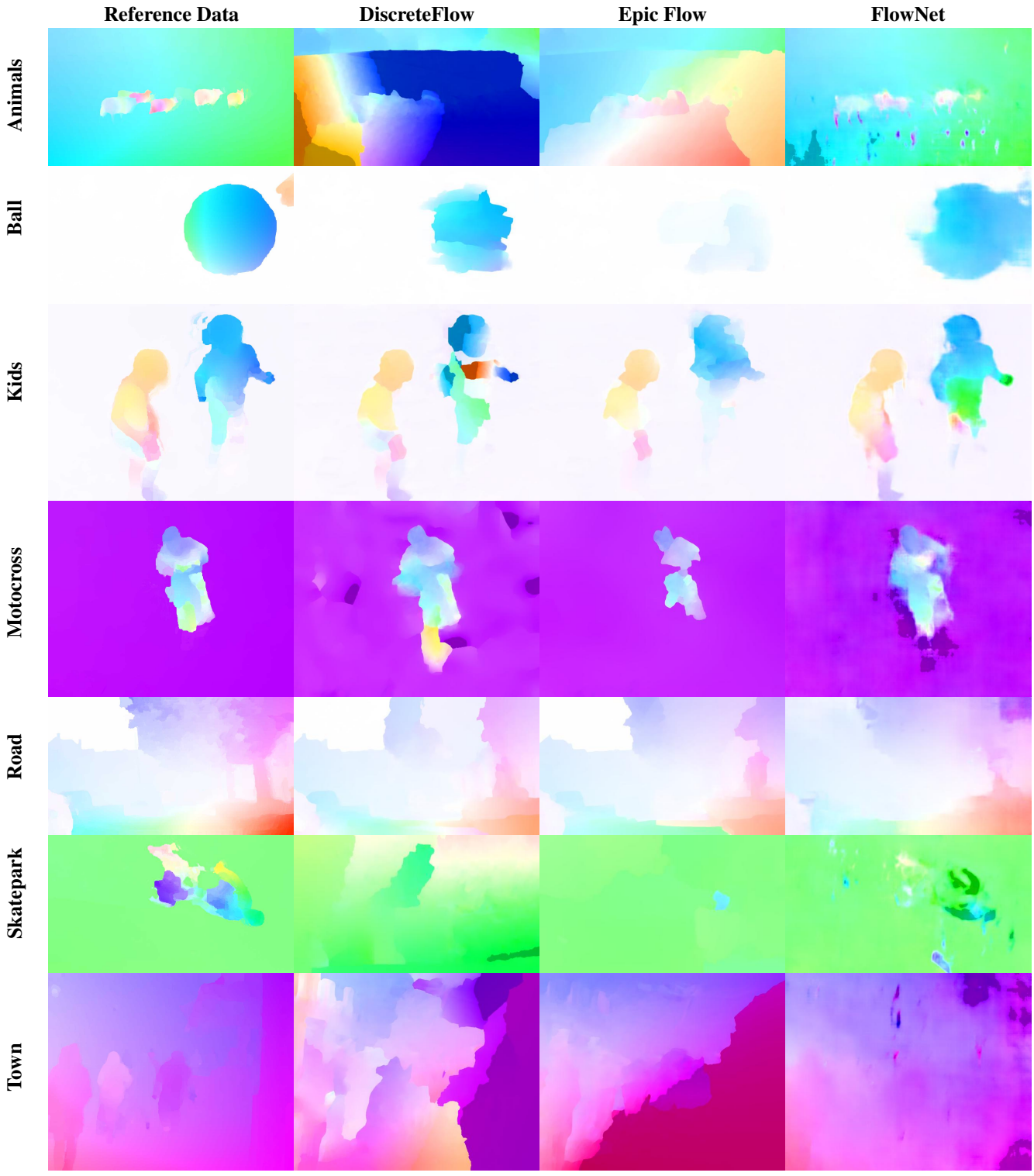

Figure 10: Comparison of DiscreteFlow, Epic Flow and FlowNet with 7 frames blur length to the reference data of 300px magnitude.