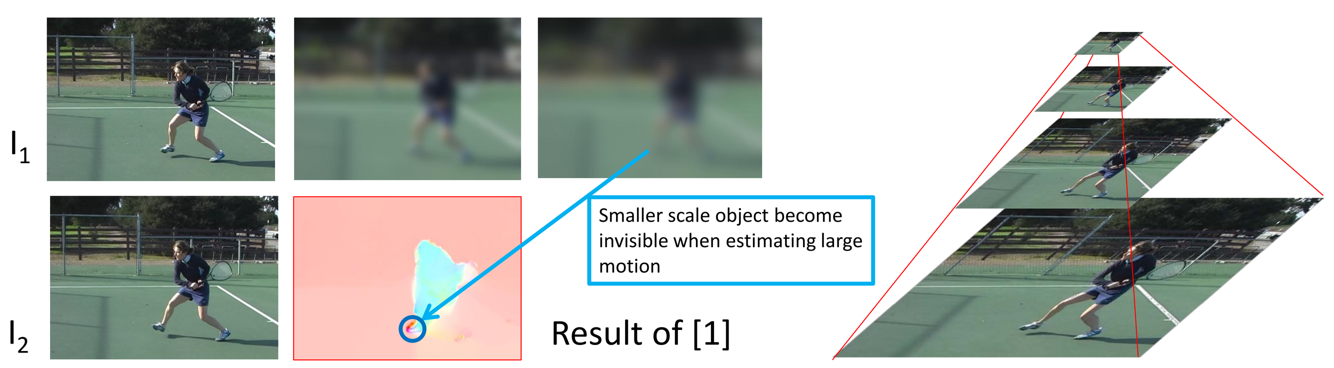
Superpixel-Based Large Displacement Optical Flow

Haw-Shiuan Chang and Yu-Chiang Frank Wang

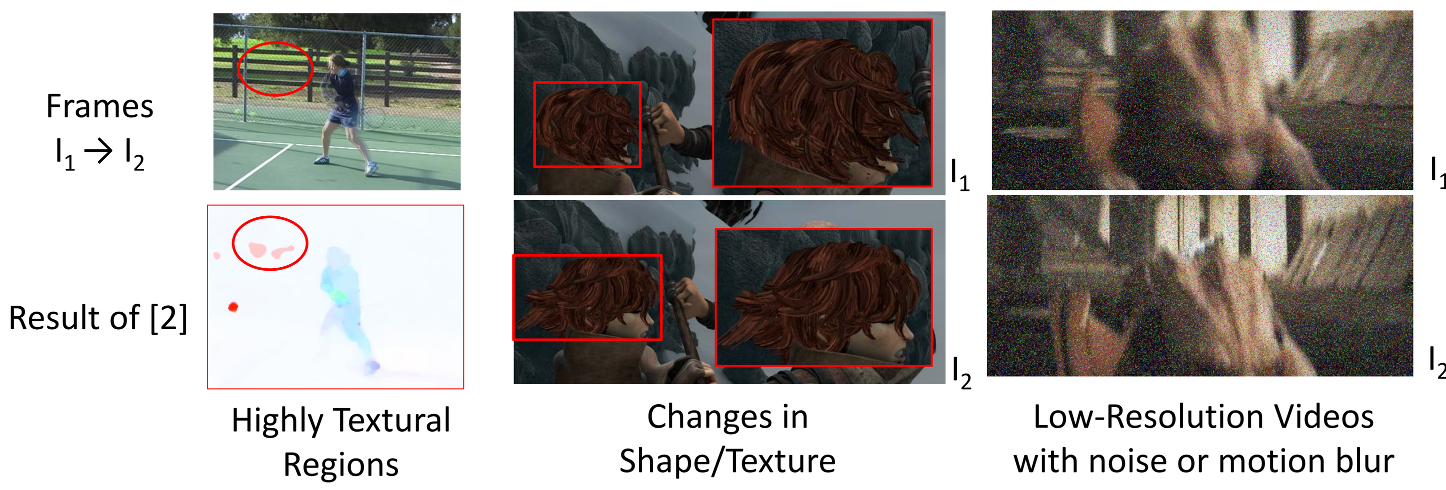
Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

Motivation

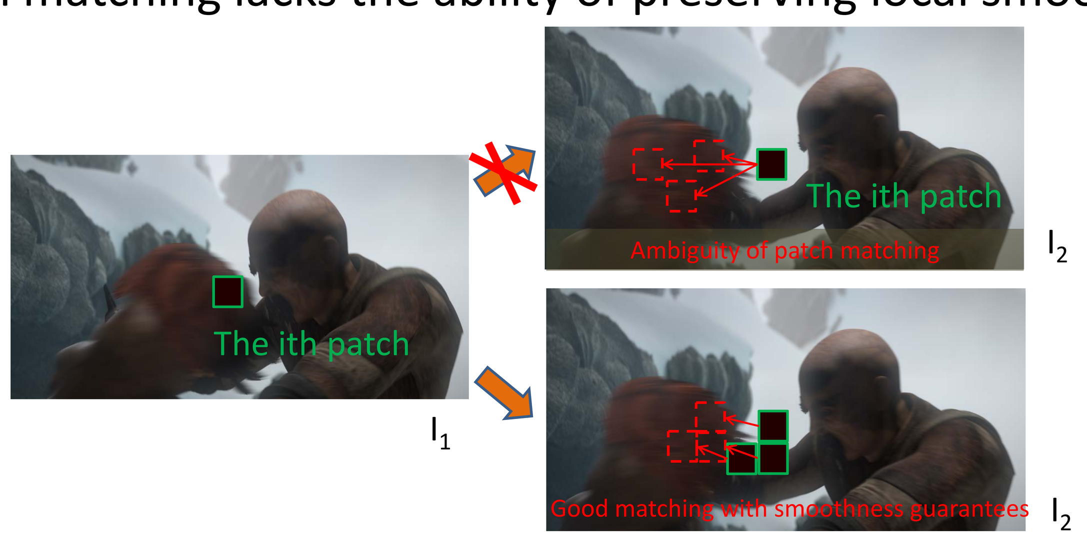
- Existing large displacement optical flow (LDOF) methods rely on coarse to fine warping (e.g. [1]), descriptor (e.g., [2]) or patch matching (e.g., [3]).
  - Coarse to fine warping can not find large motion on small scale object



  - Descriptor matching might not be preferable if:



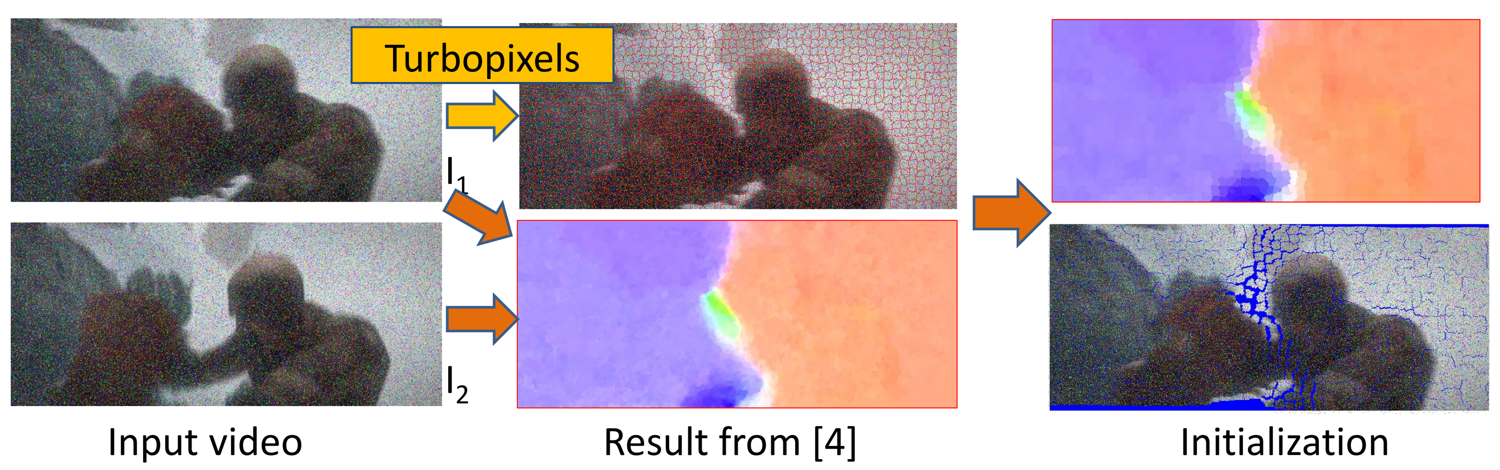
  - Patch matching lacks the ability of preserving local smoothness.



- We propose superpixel-based matching for LDOF with improved performance.

Initialization

- Segment the input image I_1 by Turbopixels[4].
- For each superpixel, we consider the averaged standard optical flow outputs[1] as initialization of our LDOF process.



Our Proposed Formulation

- Estimate the motion vector $\mathbf{u}_i$ for superpixel i at $\mathbf{x}_i$ by

$$\min.\ E=\sum_i\left(E_i^C(\boldsymbol{u}_i)+\sum_{j\in N_i^G}w_{(i,j)}^G E_{(i,j)}^G(\boldsymbol{u}_i)+\sum_{j\in N_i^S}w_{(i,j)}^S E_{(i,j)}^S(\boldsymbol{u}_i,\boldsymbol{u}_j)\right)$$

Color term

$$\sum_{p\in B_i^C} g(\|\boldsymbol{x}_p-(\boldsymbol{x}_i+\boldsymbol{u}_i)\|_2)\rho(\|\mathbf{I}_2(\boldsymbol{x}_p)-\boldsymbol{\mu}_i^C\|_2)$$

New location in I_2 of the superpixel i

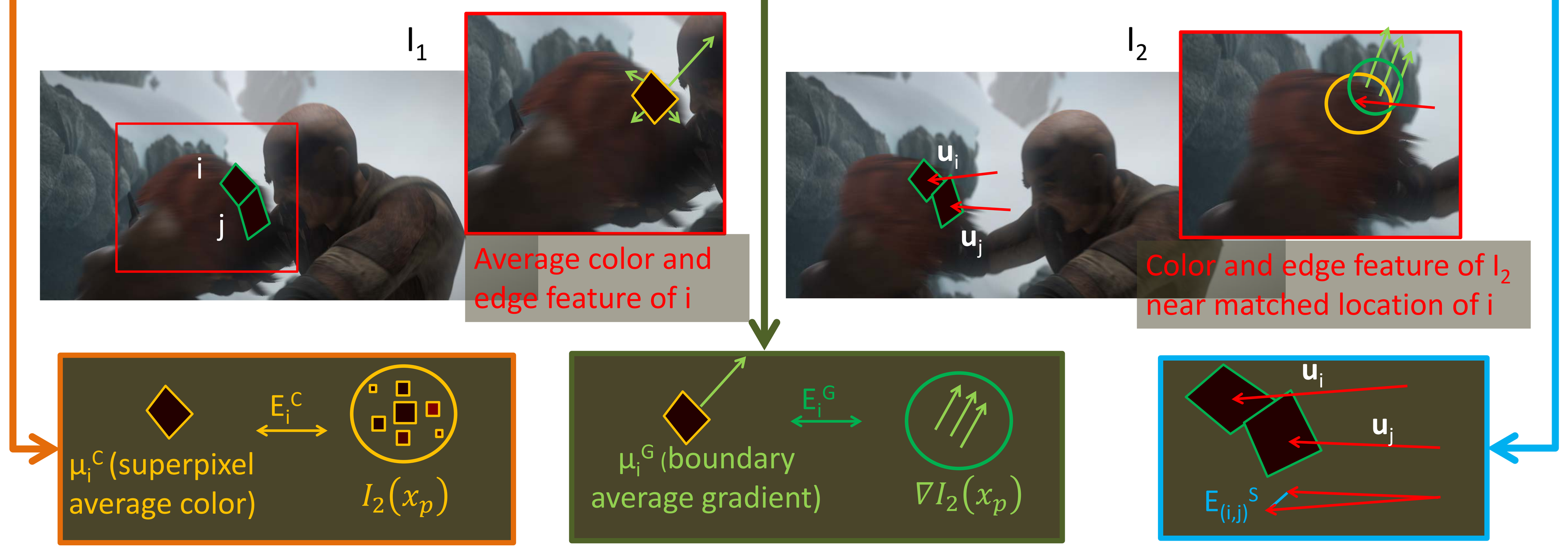
Gradient term

$$\sum_{p\in B_{(i,j)}^G} g(\|\boldsymbol{x}_p-(\boldsymbol{x}_{(i,j)}+\boldsymbol{u}_i)\|_2)\rho(\|\nabla\mathbf{I}_2(\boldsymbol{x}_p)-\boldsymbol{\mu}_{(i,j)}^G\|_2)$$

New location in I_2 of the edge between superpixel i and j

Smoothness term

$$\sum_{j\in N_i^S} w_{(i,j)}^S\|\boldsymbol{u}_i-\boldsymbol{u}_j\|_2^2$$



Energy caused by the color difference between i and I_2 near $\mathbf{x}_i+\mathbf{u}_i$

Energy caused by the gradient difference between edge of i and I_2 near $\mathbf{x}_{(i,j)}+\mathbf{u}_i$

Energy caused by the motion difference between i and j

Optimization (by Gradient Descend)

- At iteration t, the optical flow for the ith superpixel is:

$$\mathbf{u}_i^{t+1}=\mathbf{u}_i^t-\lambda\left(\frac{\partial E(\mathbf{u}_i)}{\partial\mathbf{u}_i}\right)=\mathbf{u}_i^t+\lambda'\left(\frac{\sum_{p\in B_i^C}\mathbf{x}_p^i\psi_p^{i^C}g(\|\mathbf{x}_p^i\|_2)}{\sum_{p\in B_i^C}\psi_p^{i^C}g(\|\mathbf{x}_p^i\|_2)}+\sum_{j\in N_i^G}\frac{w_{(i,j)}^{\prime G}\sum_{p\in B_{(i,j)}^G}\mathbf{x}_p^{(i,j)}\psi_p^{(i,j)^G}g(\|\mathbf{x}_p^{(i,j)}\|_2)}{\sum_{p\in B_{(i,j)}^G}\psi_p^{(i,j)^G}g(\|\mathbf{x}_p^{(i,j)}\|_2)}+\sum_{j\in N_i^S}w_{(i,j)}^{\prime S}(\mathbf{u}_j^t-\mathbf{u}_i^t)\right)$$

Take the derivative of our formulation

Data terms: Local search by mean shift

Smoothness term: Pass the information further

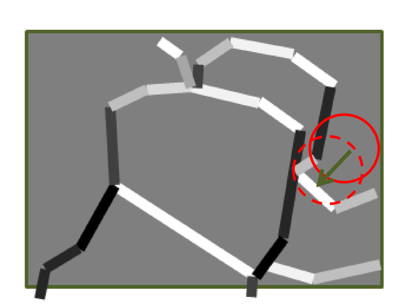
Color ($\partial E^C(u_i)/\partial u_i$)

Location of superpixel i of I_1

$$\psi_p^C=\exp\left(-\frac{\|\mathbf{I}_2(\mathbf{x}_p)-\boldsymbol{\mu}_i^C\|_2}{\beta^C}\right)$$

Color similarity map between superpixel i at I_1 and those at I_2

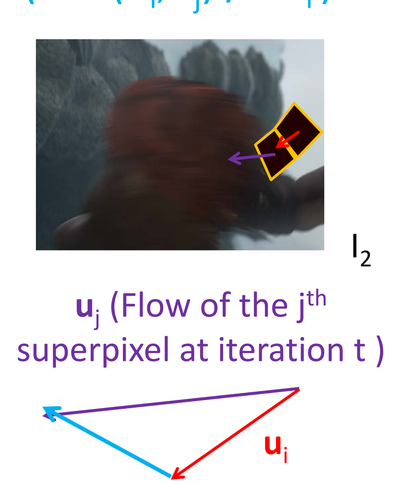
Gradient ($\partial E^G(u_i)/\partial u_i$)



$$\psi_p^{(i,j)^G}=\exp\left(-\frac{\|\nabla\mathbf{I}_2(\mathbf{x}_p)-\boldsymbol{\mu}_{(i,j)}^G\|_2}{\beta^G}\right)$$

Gradient similarity map between the edges of superpixel i of I_1 and those of I_2

Smoothness ($\partial E^S(u_i,u_j)/\partial u_i$)

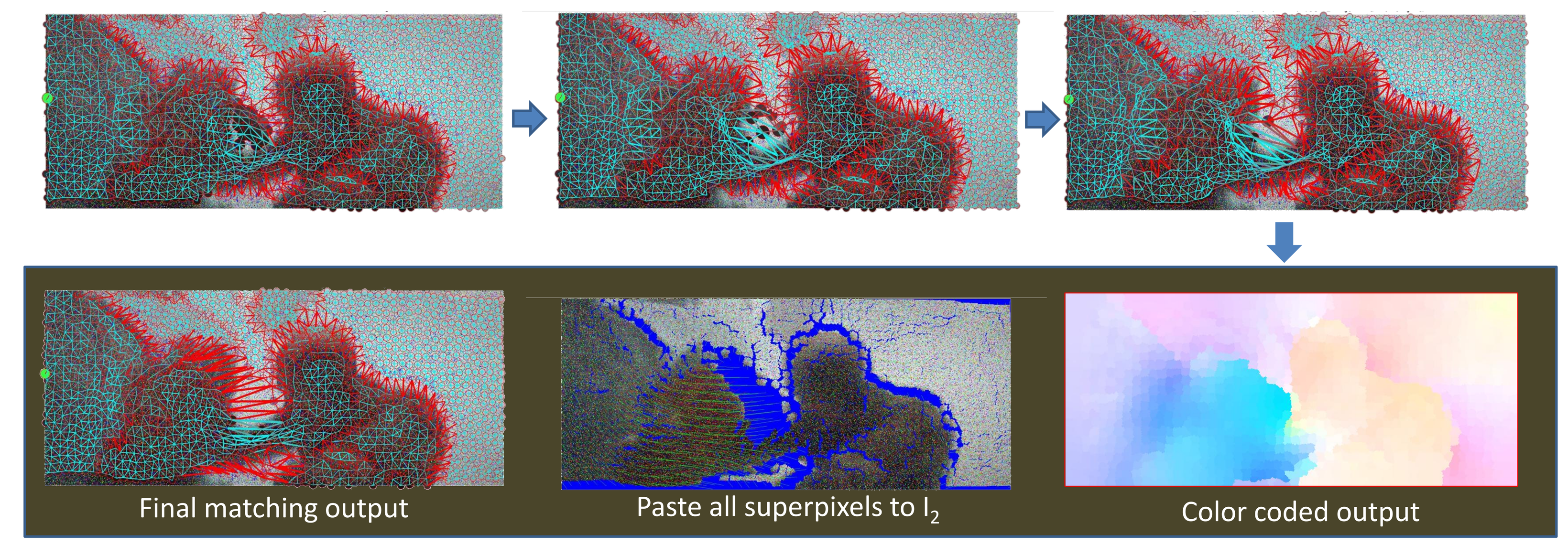


u_j (Flow of the jth superpixel at iteration t)

- Our method can be viewed as the tracking of superpixels by mean shift across frames, while local smoothness is preserved.

Experiments

- Example iterations and final outputs



- Dataset: MPI-Sintel[5] (with noise and motion blur)

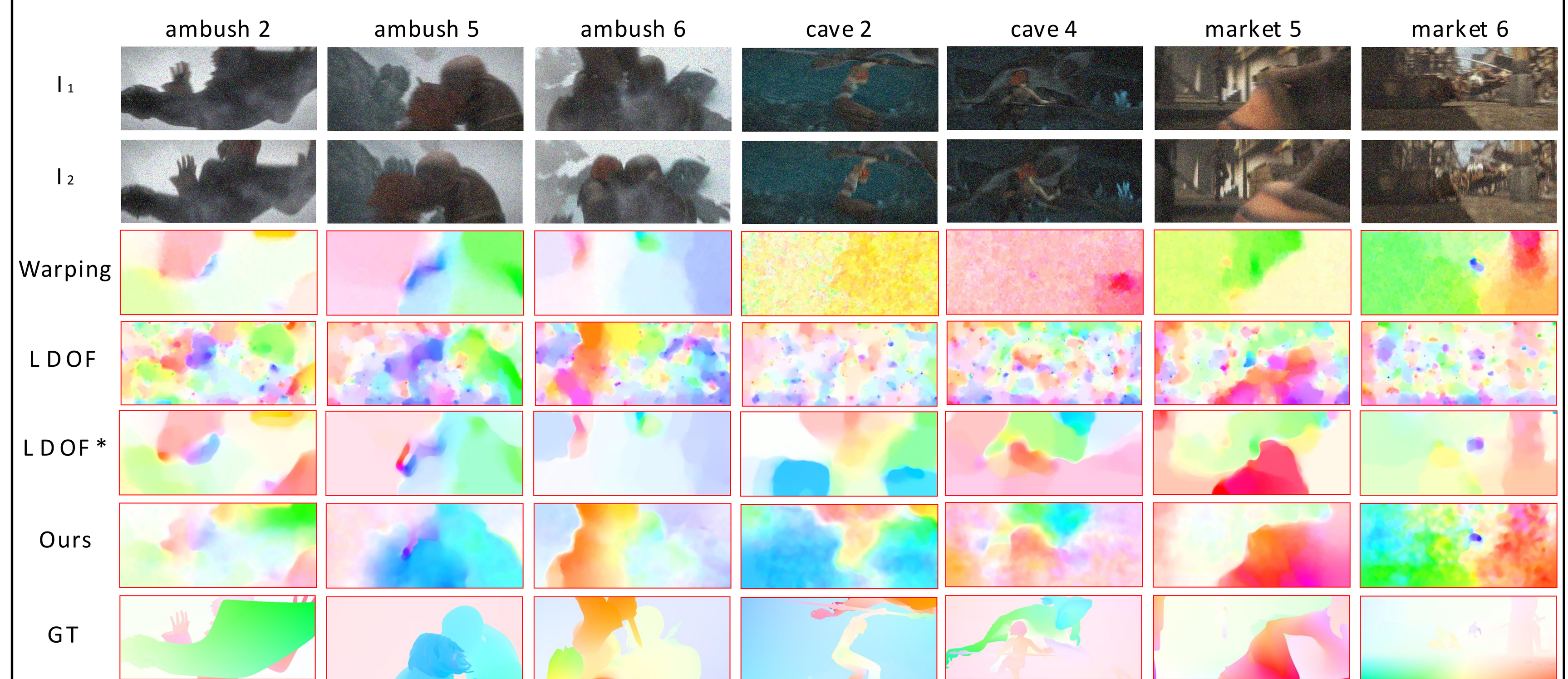


Table 1: Comparisons of end point error (EPE) for different methods. Note that * indicates the results of LDOF on manually-blurred videos, and the numbers in bold denote the best results for the corresponding videos.

Method \ Sequence	ambush2	ambush5	ambush6	cave2	cave4	market5	market6	Avg
Warping [1]	**65.45**	42.13	59.62	73.51	24.11	54.24	23.68	48.96
Horn+Schunck [6]	68.95	46.51	58.70	70.36	22.27	48.63	24.11	48.51
Classic+NL-fast [6]	69.02	45.61	57.31	70.21	21.06	51.05	24.48	48.39
LDOF [2]	73.24	49.17	60.71	72.03	29.66	53.80	33.02	53.09
LDOF* [2]	68.95	37.75	56.60	58.06	**19.03**	43.14	22.74	43.75
Ours	66.80	**35.63**	**45.09**	**45.63**	19.87	**33.34**	**19.34**	**37.96**

Conclusion

- We propose a novel LDOF method in which we
  - **Do not** linearize the data term.
  - **Do not** smooth the smaller object away by coarse to fine warping.
  - **Do not** perform global matching without considering smoothness term or decouple the data and smoothness term.

Reference

[1] T. Brox et al. , High accuracy optical flow estimation based on a theory for warping, ECCV 2004.
[2] T. Brox and J. Malik, Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation, TPAMI 2010.
[3] L. Xu et al. , Motion Detail Preserving Optical Flow Estimation, TPAMI 2012.
[4] A. Levinshtein et al. , Turbopixels: Fast superpixels using geometric flows, TPAMI 2009.
[5] D. J. Butler et al. , A naturalistic open source movie for optical flow evaluation, ECCV, 2012.
[6] D. Sun et al. , Secrets of optical flow estimation and their principles, CVPR, 2010.