Code here
https://github.com/IntelVCL/DCFlow

Accurate Optical Flow via Direct Cost Volume Processing

Jia Xu René Ranftl Vladlen Koltun
Intel Labs

Introduction

- Optical Flow: dense motion of pixels between two images
- Key building block for many computer vision systems
- Challenges: large displacement, computational complexity

We show that direct cost volume processing is feasible:

- With moderate amount of downsampling
- Incorporate best practices from stereo estimation [1]
- Combine with modern interpolation schemes

Stereo

Left Image Right Image

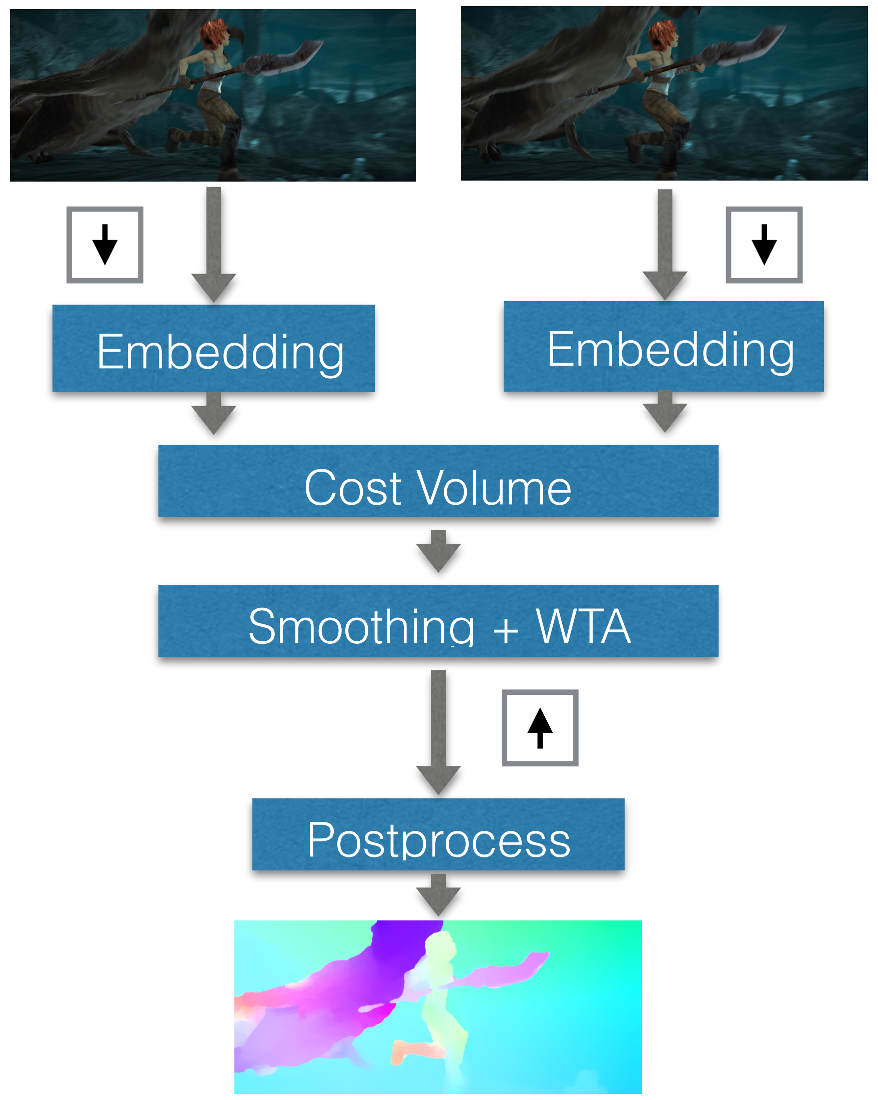
Optical Flow

First Image Second Image

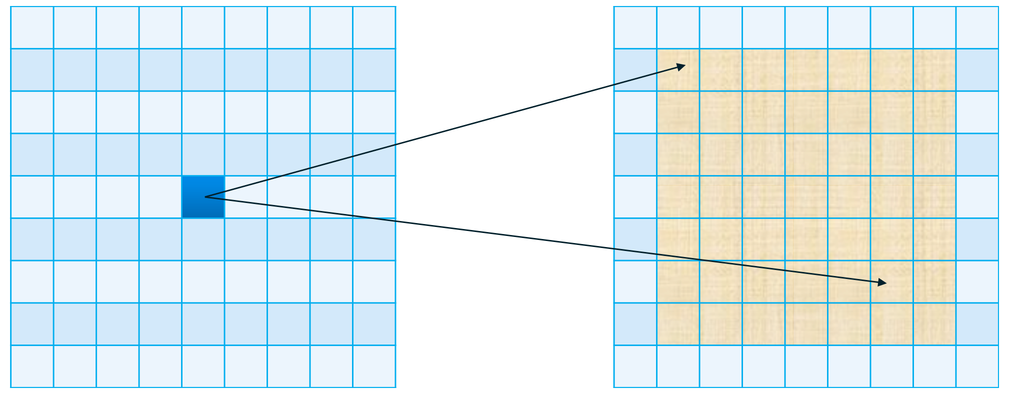
	Stereo	Optical Flow
Search space	64-256 locations	4,000-65,000 locations
Previous work	Direct cost volume processing, high accuracy	Continuous optimization methods, NN search, coarse-to-fine approximation, low accuracy

References
[1] Zbontar and LeCun, Stereo matching by training a convolutional neural network to compare image patches. *JMLR* 2016
[2] Revaud et. al., EpicFlow: Edge-preserving interpolation of correspondences for optical flow. *CVPR* 2015

Our Approach



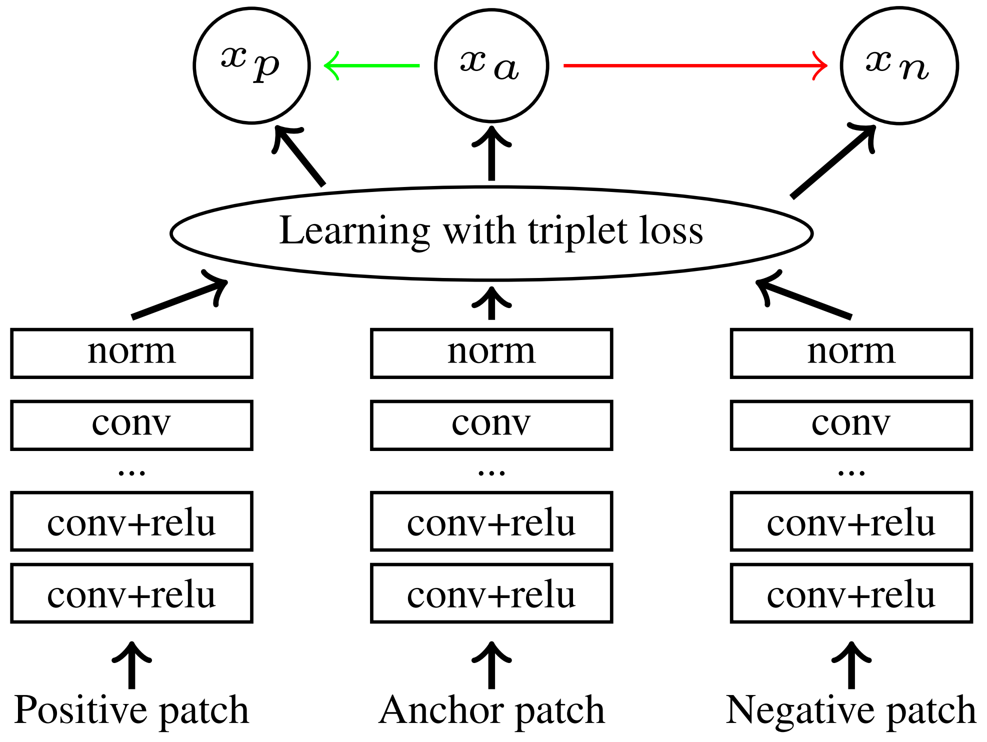
4-D Cost Volume



Regular structure of size MxNxRxR

- 3x downsampling
- ~25,000 labels per pixel
- Embedding and regularity enable efficient construction

Pixel-level Feature Embedding



- Compact network (4 layers,112K parameters)
- Can be trained from ~200 ground truth images
- Euclidean embedding

Cost Volume Processing

- Smooth cost volume to propagate information to textureless regions
- Modified SGM energy:

$$E(\mathbf{V}) = \sum_p \Bigg(\sum_{q \in \mathcal{N}(p)} P_1 [\|\mathbf{V}_p - \mathbf{V}_q\|_1 = 1] + \sum_{q \in \mathcal{N}(p)} P_2^{p,q} [\|\mathbf{V}_p - \mathbf{V}_q\|_1 > 1] + \mathbf{C}(p, \mathbf{V}_p) \Bigg)$$

- Highly efficient implementation

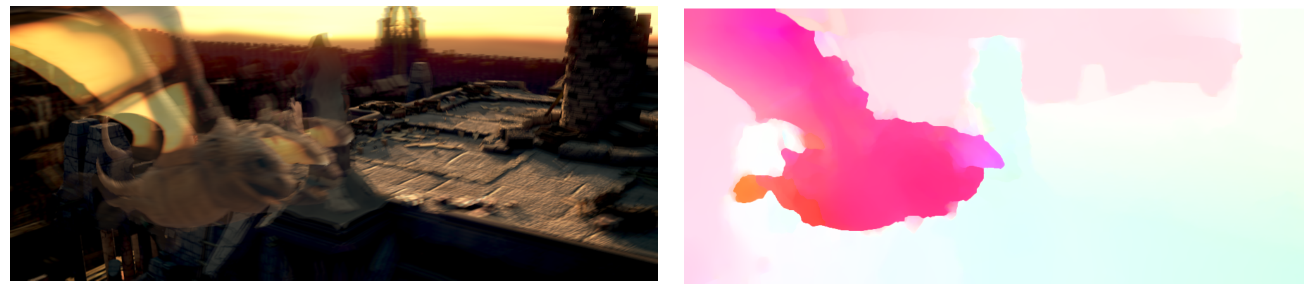
Post-processing

- Forward-backward consistency check
- Edge-preserving interpolation using EpicFlow [2]
- Local homography fitting

Runtime breakdown	fast	accurate
Feature extraction	0.02	0.02
Cost volume (fwd + bwd)	0.06	0.24
SGM (fwd + bwd)	0.45	2.59
EpicFlow	2.87	2.87
Homography inpainting	–	2.91
Total	3.40	8.63

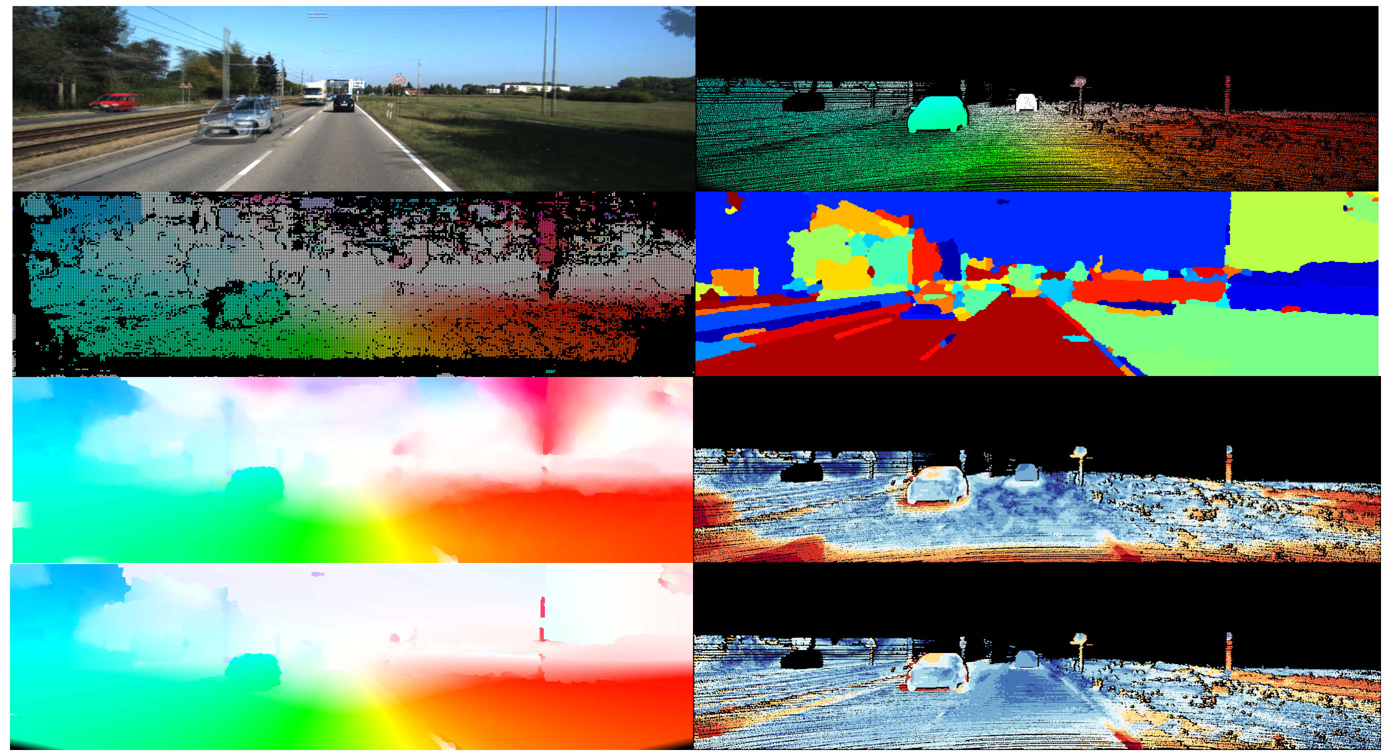
Sintel

	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
DCFlow [2]	5.119	2.283	28.228	4.665	2.108	1.440	1.052	3.434	29.351
FlowFieldsCNN [3]	5.363	2.303	30.313	4.718	2.020	1.399	1.032	3.065	32.422
MR-Flow [4]	5.376	2.818	26.235	5.109	2.395	1.755	0.908	3.443	32.221
FTFlow [5]	5.390	2.268	30.841	4.513	1.964	1.366	1.046	3.322	31.936
S2F-IF [6]	5.417	2.549	28.795	4.745	2.198	1.712	1.157	3.468	31.262
InterpoNet_ff [7]	5.535	2.372	31.296	4.720	2.018	1.532	1.064	3.496	32.633
RegionalFF [8]	5.562	2.595	29.741	4.921	2.393	1.639	1.122	3.477	32.625



KITTI 2015

Method	Domain-agnostic	Non-occluded pixels (%)			All pixels (%)			Runtime
		Fl-bg	Fl-fg	Fl-all	Fl-bg	Fl-fg	Fl-all	
SOF [31]	✗	8.11	18.16	9.93	14.63	22.83	15.99	6 min
JFS [19]	✗	7.85	14.97	9.14	15.90	19.31	16.47	13 min
SDF [2]	✗	5.75	18.38	8.04	8.61	23.01	11.01	–
EpicFlow [27]	✓	15.00	24.34	16.69	25.81	28.69	26.29	15 sec
FullFlow [7]	✓	12.97	20.58	14.35	23.09	24.79	23.57	4 min
CPM-Flow [18]	✓	12.77	18.71	13.85	22.32	22.81	22.40	4.2 sec
DiscreteFlow [25]	✓	9.96	**17.03**	11.25	21.53	**21.76**	21.57	3 min
DDF [13]	✓	10.44	21.32	12.41	20.36	25.19	21.17	1 min
PatchBatch [12]	✓	10.06	22.29	12.28	19.98	26.50	21.07	50 sec
DC Flow	✓	**8.04**	19.84	**10.18**	**13.10**	23.70	**14.86**	8.6 sec



Summary

- A step towards unifying optical flow and stereo
- Combines high accuracy with competitive runtimes