A general dense image matching framework combining direct and feature-based costs

FROM RESEARCH TO INDUSTRY cea tech

list

UdA Project FLEXABLE

Jim Braux-Zin[1,2], **Romain Dupont**[1] and **Adrien Bartoli**[2]

[1]CEA, LIST, Gif-sur-Yvette, France [2]ISIT, CNRS and Université d'Auvergne, Clermont-Ferrand, France

Contact: `jim.braux-zin@cea.fr`

ALCoV ISIT

Our goal: a unified approach for narrow- and wide-baseline dense image matching

Target applications

- Optical flow with small or large displacements
- Stereo depth estimation from narrow/wide-baseline pairs
- Small/large deformation estimation of a non-rigid surface

Narrow-baseline example

KITTI benchmark optical flow pair 14

Wide-baseline example

Herzjesu stereo dataset, frames 2 and 5

Notations

Inputs: target image $\mathcal{I}_0$, source image $\mathcal{I}_1$, feature matches $\mathcal{F}$, match composed of two features $\mathbf{f} = (\mathbf{f}_0, \mathbf{f}_1) \in \mathcal{F}$

Output: displacement field $\mathbf{u}$ such as $\mathcal{I}_0(\mathbf{q}) \approx \mathcal{I}_1(\mathbf{q} + \mathbf{u}(\mathbf{q}))$

$$\text{Minimize } \mathcal{C}(\mathbf{u}, \mathcal{F}, \mathcal{I}_1, \mathcal{I}_0) = \lambda \mathcal{C}_{\text{direct}}(\mathbf{u}, \mathcal{I}_1, \mathcal{I}_0) + \text{TGV}^2_{\alpha_0,\alpha_1}(\mathbf{u}) + \beta \sum_{\mathbf{f} \in \mathcal{F}} \mathcal{C}_{\text{feat}}(\mathbf{u}, \mathbf{f})$$

State-of-the-art variational optical flow

Data term

Different smooth data term $\mathcal{C}_{\text{direct}}(\mathbf{u}, \mathcal{I}_1, \mathcal{I}_0)$ can be used, for example:

- **Absolute Difference:** $\mathcal{C}_{\text{AD}}(\mathbf{u}, \mathcal{I}_1, \mathcal{I}_0) = \int_\Omega |\mathcal{I}_0(\mathbf{q}) - \mathcal{I}_1(\mathbf{q} + \mathbf{u}(\mathbf{q}))|_1 \mathrm{d}\mathbf{q}$
- **Census distance:** $\mathcal{C}_{\text{Census}}(\mathbf{u}, \mathcal{I}_1, \mathcal{I}_0) = \int_\Omega |\Delta\mathcal{I}_0(\mathbf{q}) - \Delta\mathcal{I}_1(\mathbf{q} + \mathbf{u}(\mathbf{q}))|_{\text{Hamming}} \mathrm{d}\mathbf{q}$
  $\Delta\mathcal{I}(\mathbf{q})$: Census transform, binary string encoding for each neighboring pixel if it is brighter or darker in $\mathcal{I}$. **Encodes the local image structure.**
- **AD-Census:** $\mathcal{C}_{\text{ADC}} = 2 - \exp(-\mathcal{C}_{\text{AD}}/\mu_{\text{AD}}) - \exp(-\mathcal{C}_{\text{Census}}/\mu_{\text{Census}})$
  More robust than AD, more accurate than Census.

Self-occlusions can be handled as in [4].

Regularization: Second-order Total Generalized Variation [1]

The two channels of the optical flow are regularized independently, $u \in \{u_x, u_y\}$.

$$\text{TGV}^2_{\alpha_0,\alpha_1}(u, \alpha_0, \alpha_1) = \min_{\mathbf{w} \in \Omega \to \mathbb{R}^2} \left\{ \alpha_1 \int_\Omega |\nabla u - \mathbf{w}| \, \mathrm{d}\mathbf{q} + \alpha_0 \int_\Omega |\nabla \mathbf{w}| \, \mathrm{d}\mathbf{q} \right\}$$

Favors piecewise-affine solutions.

Optimization

Alternate **minimization** (iterative linearization of the data term) and **regularization**. Chambolle & Pock primal-dual algorithm [3]. **Coarse-to-fine** approach.

✓ **High accuracy, fast with parallel implementations**

✗ **Restricted to small displacements and subject to local minima**

Conclusion

Feature matches **guide variational flow estimation out of local minima**.

Our algorithm has the unique combination of the following properties:

⇒ **Flexible:** loosely coupled data-term, regularizer and features can be easily swapped

⇒ **Accurate:** among the top performing methods for small and large displacements

⇒ **Robust and versatile:** thanks to the large convergence basin

Our feature-based cost

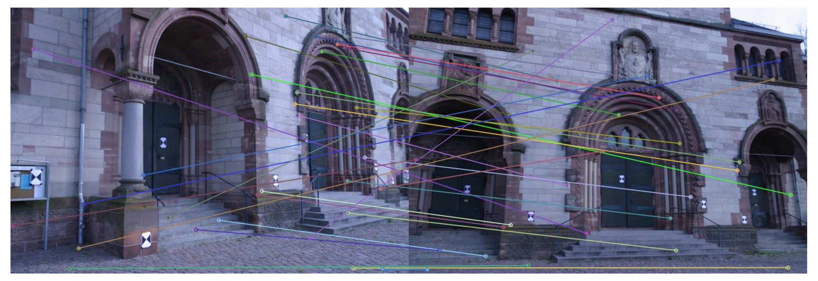
Feature cost is more convex than direct cost for large flows

Point features for small baseline

- Usually detector (Harris, FAST...) and descriptor (SIFT...)
- Matching by nearest-neighbor search in the descriptor space **over the whole image**.
- Based on photometric properties such as Histogram of Gradients, not robust to large distortions.

Segments for wide baseline

- Segment clusters provide a **semi-global description** of the scene.
- Complex matching [6] by comparison of clusters geometric properties (angles, length ratios...)



FAST-SIFT matching, many outliers

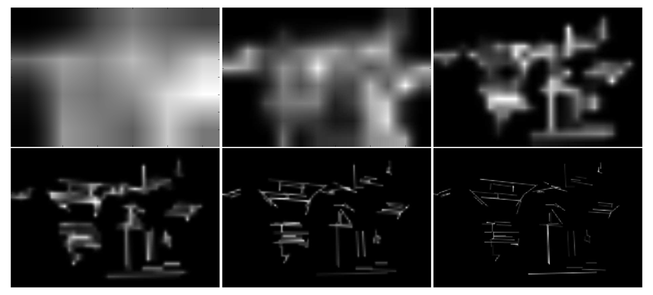
Wide-baseline segment matching

Features can increase the convergence basin of variational optical flow

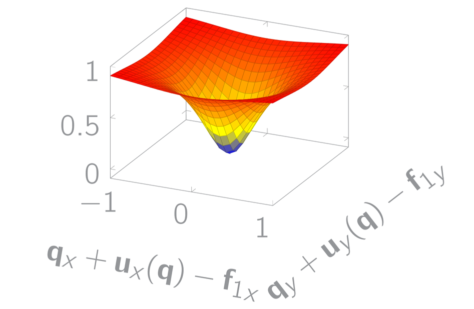
$$\mathcal{C}_{\text{feat}}(\mathbf{u}, \mathbf{f}) = \int_\Omega \rho(\mathbf{q}, \mathbf{f}_0) D(\mathbf{q}, \mathbf{u}(\mathbf{q}), \mathbf{f}_1) \, \mathrm{d}\mathbf{q}$$

$\rho(\mathbf{q}, \mathbf{f}_0)$: 2×2 bilinear influence in $\mathcal{I}_0$. Smaller relative area at finer resolution. **Features' influence decreases in an annealing-like manner [2].**

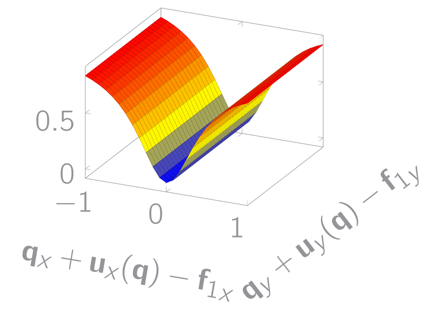
$D(\mathbf{q}, \mathbf{u}(\mathbf{q}), \mathbf{f}_1)$: cost in $\mathcal{I}_1$ penalizing the distance between $\mathbf{q} + \mathbf{u}(\mathbf{q})$ and the feature $\mathbf{f}_1$. Geman McClure robust estimator $\Psi_\sigma(x) = x^2/(\sigma + x^2)$ for **implicit feature filtering**, robust to outliers.



Influence of features from coarse to fine resolution



Cost for points (based on Euclidean distance)



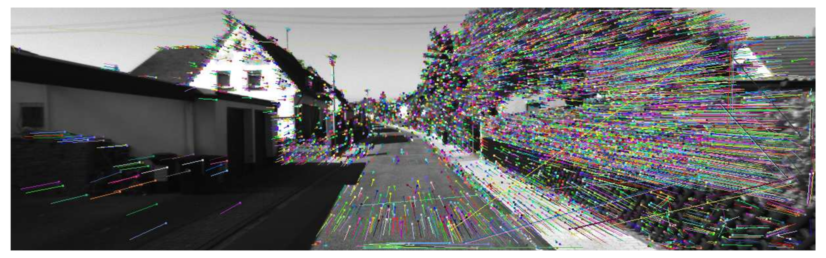
Cost for segments (orthogonal distance)

Main example

Census data term, TGV²

Errors due to local minima
Outliers (3px): 15%

FAST-SIFT matches



≈ 1% of outliers

Using keypoint matches in the variational optimization

Influence of local minima greatly reduced
Outliers (3px): 9.3%

One algorithm, many applications

Optical flow (KITTI benchmark)

Using FAST-SIFT matches

Rank	Method	Setting	Code	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime
1	SceneFlow			2.93 %	5.71 %	0.8 px	1.3 px	100.00 %	6 min
Anonymous submission									
2	PR-Sf+E			3.57 %	7.07 %	0.9 px	1.6 px	100.00 %	200 s
C. Vogel, S. Roth and K. Schindler: Piecewise Rigid Scene Flow. International Conference on Computer Vision (ICCV) 2013.									
3	PCBP-Flow			3.64 %	8.28 %	0.9 px	2.2 px	100.00 %	3 min
K. Yamaguchi, D. McAllester and R. Urtasun: Robust Monocular Epipolar Flow Estimation. CVPR 2013.									
4	PR-Sceneflow			3.76 %	7.39 %	1.2 px	2.8 px	100.00 %	150 sec
C. Vogel, S. Roth and K. Schindler: Piecewise Rigid Scene Flow. International Conference on Computer Vision (ICCV) 2013.									
5	MotionSLIC			3.91 %	10.56 %	0.9 px	2.7 px	100.00 %	11 s
K. Yamaguchi, D. McAllester and R. Urtasun: Robust Monocular Epipolar Flow Estimation. CVPR 2013.									
6	gtRF-DF			6.03 %	13.08 %	1.6 px	4.2 px	100.00 %	1 min
Anonymous submission									
7	TGV2ADCSIFT			6.20 %	15.15 %	1.5 px	4.5 px	100.00 %	12s
J. Braux-Zin, R. Dupont and A. Bartoli: A General Dense Image Matching Framework Combining Direct and Feature-based Co									

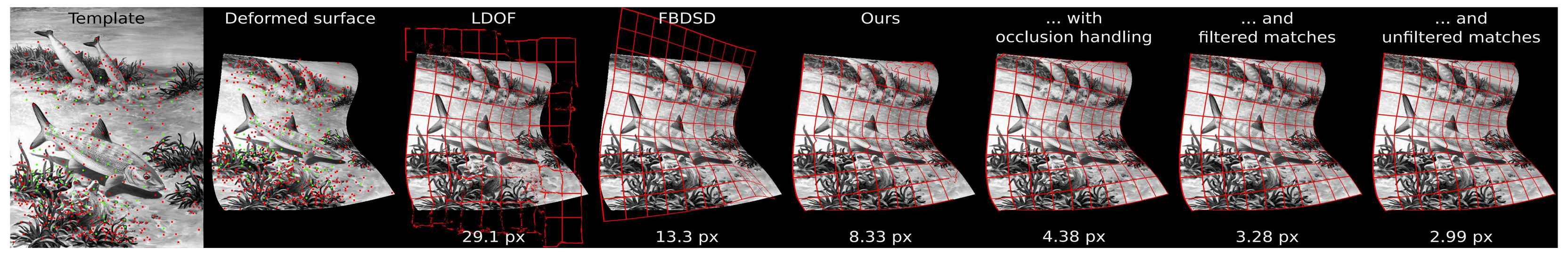
Wide-baseline stereo

Using segment matches

Average amount of correct depth values on the Herzjesu dataset:
DAISY [5] + Graph Cuts: **91.23%**
Optical flow with direct data-term only: **68%**
With our feature-based cost: **81.5%**

Non-rigid surface registration

Synthetic image pair. Using SIFT matches. Comparison with [2] and [4].



[1] K. Bredies. Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty. *SFB Report*, 6, 2012.

[2] T. Brox et al. Large displacement optical flow: descriptor matching in variational motion estimation. *PAMI*, 2011.

[3] A. Chambolle et al. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 2011.

[4] D. Pizarro et al. Feature-based deformable surface detection with self-occlusion reasoning. *IJCV*, 2012.

[5] E. Tola, et al. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *PAMI*, 2010.

[6] L. Wang, et al. Wide-baseline image matching using line signatures. In *ICCV*. 2009.