Imperial College London

ICCV15

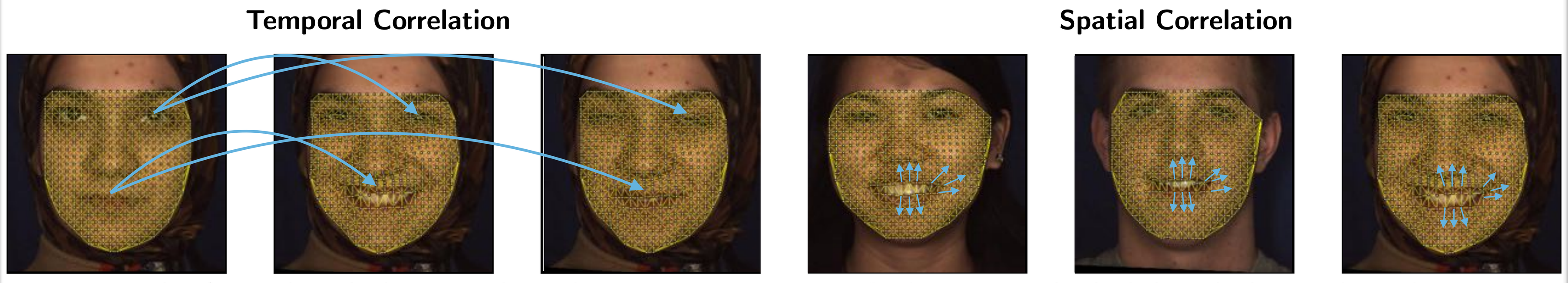
FACE FLOW

PATRICK SNAPE
YANNIS PANAGAKIS
ANASTASIOS ROUSSOS
STEFANOS ZAFEIRIOU

VISIT THE MENPO DEMO!
Wednesday December 16, 2015
0945 - 1215 (Multi Purpose Area B)

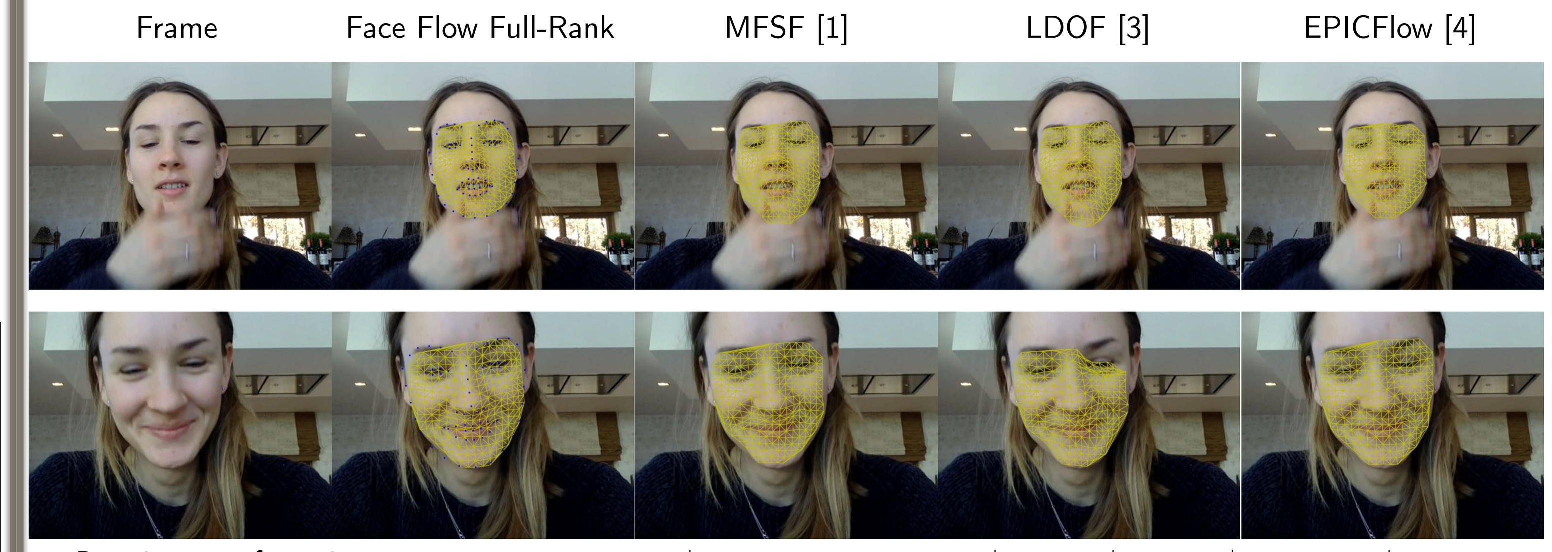
Motivation

Dense correspondence is one of the most important problems in Computer Vision - many problems are trivially solved once correspondence has been computed. In this work, we exploit both the high temporal correlation of a single facial sequence as well as the low-rank properties of modelling a single well-defined object (human faces).



Video frames have high temporal correlation — Smiles across individuals have high spatial correlation

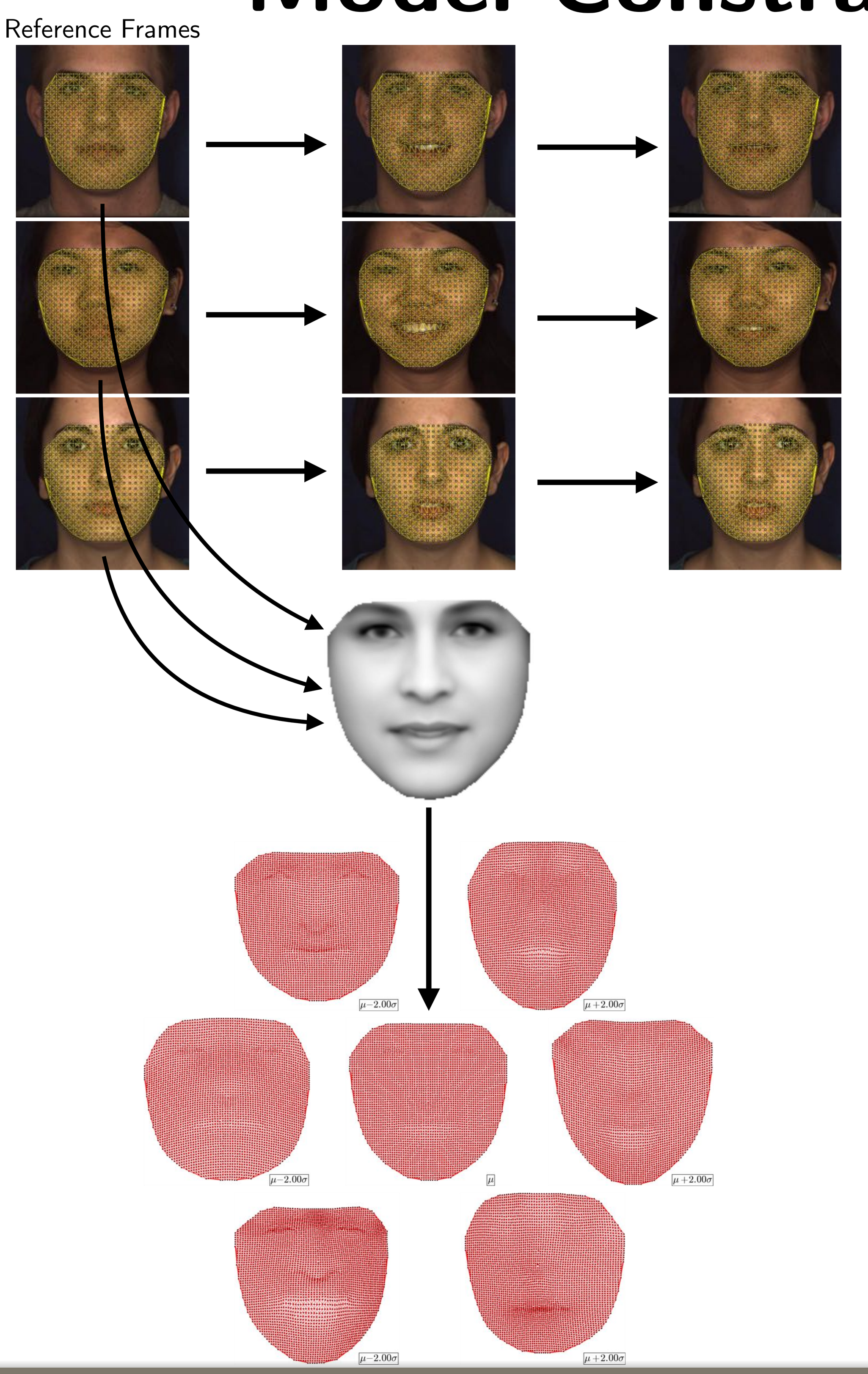
Challenging Test Sequence



Visit the QR code to see a Youtube Video of the results!

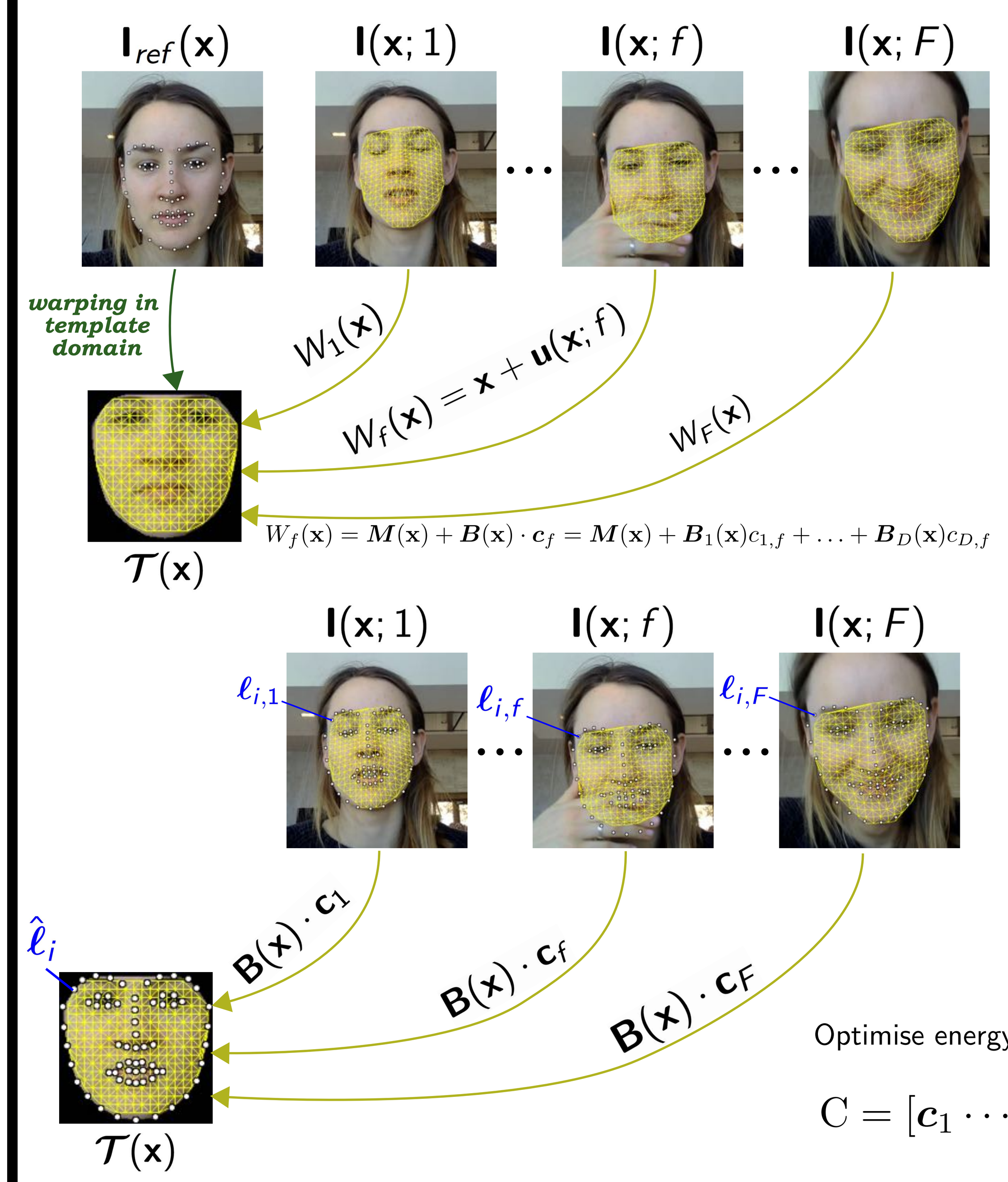
Run times per frame in seconds (640X480):	Face Flow Low-Rank	Face Flow Full-Rank	MFSF	LDOF	EpicFlow	SiftFlow
	1.4	0.5	20	15	50	15

Model Construction



1. Perform Optical Flow [1] on each sequence in BU4D [2] - each with a separate reference frame.
2. Warp each reference frame, using a Piecewise Affine warp, into a common reference frame - mean of the sparse 68 points face of all the reference frames.
3. Compute PCA on the dense grids (automatic dense landmarks) to build a generative model of dense facial shape.

Fitting



At testing time, a neutral reference frame is provided and is either manually or automatically annotated with 68 sparse landmarks. These landmarks provide correspondence between the reference frame and the learnt deformation basis. To compute correspondence, the *image data term* is optimised, subject to a low-rank constraint on the model coefficients. This low-rank term attempts to model the temporal correlation, as well as regularising the spatial deformations. Optionally, as is the case in the real sequence, the fitting can be further constrained according to a set of previously estimated sparse landmarks.

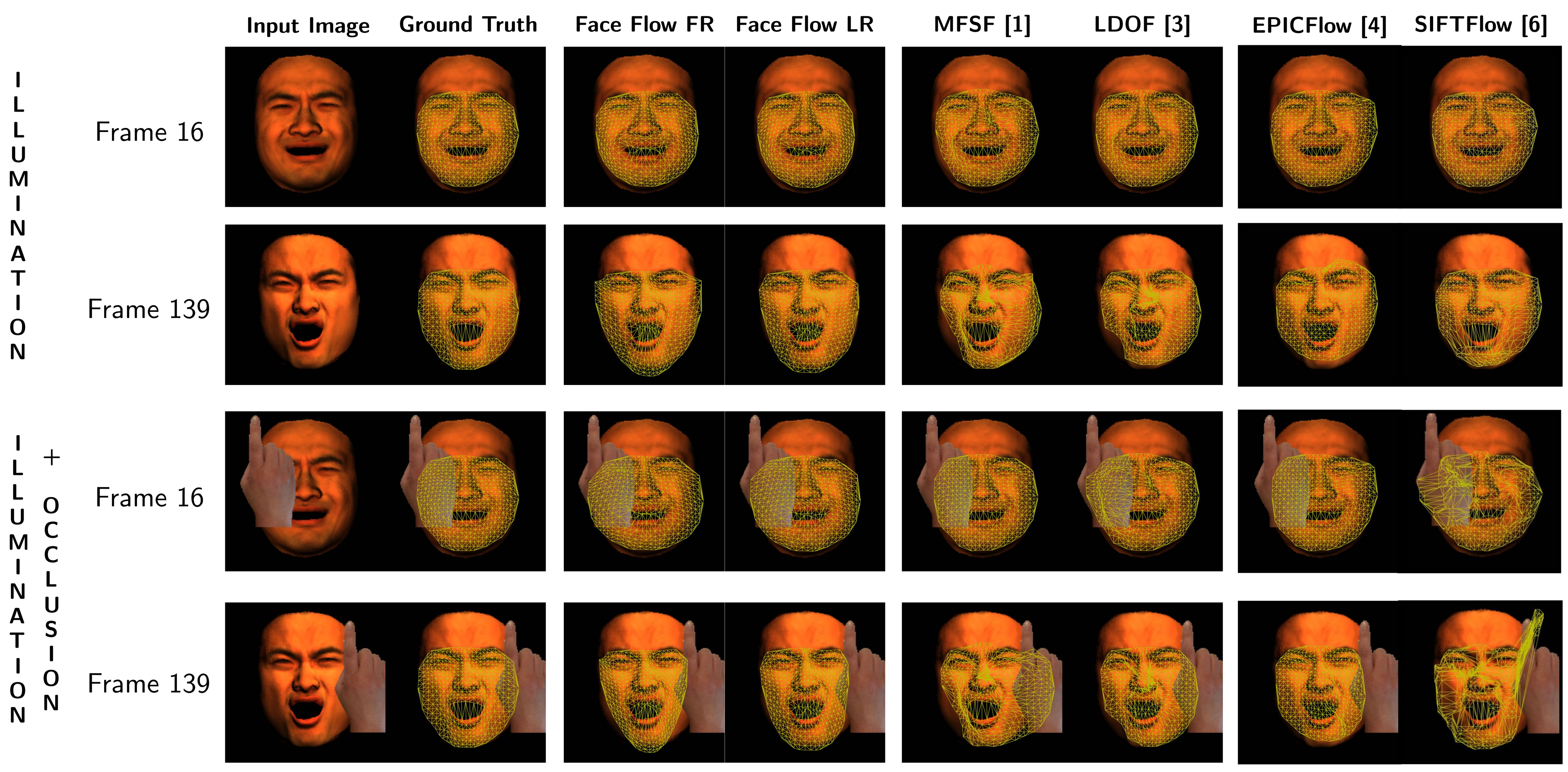
Optimise energy in terms of the model coefficients:

$$\mathrm{C} = [\boldsymbol{c}_1 \cdots \boldsymbol{c}_f \cdots \boldsymbol{c}_F] = \begin{bmatrix} \mathrm{C}_s \\ \mathrm{C}_{nr} \end{bmatrix} \begin{matrix} \} \text{ similarity (first 4 rows)} \\ \} \text{ non-rigid deformations} \end{matrix}$$

Minimise w.r.t. **C**:

$$\underbrace{\sum_{f=1}^{F} \int_M \|\boldsymbol{\mathcal{T}}(\mathbf{x}) - \boldsymbol{I}(\boldsymbol{B}(\mathbf{x}) \cdot \boldsymbol{c}_f\,;\, f)\|^2 \,\mathrm{d}\mathbf{x}}_{\textit{image data term}} + \beta \underbrace{\sum_{f=1}^{F} \sum_{i=1}^{L} \left\| \boldsymbol{B}(\hat{\boldsymbol{\ell}}_i) \cdot \boldsymbol{c}_f - \boldsymbol{\ell}_{i,f} \right\|^2}_{\textit{landmarks term}}, \text{ s.t. } \underbrace{\mathrm{rank}(\mathrm{C}_{nr}) \le \lambda}_{\textit{temporal coherency}}$$

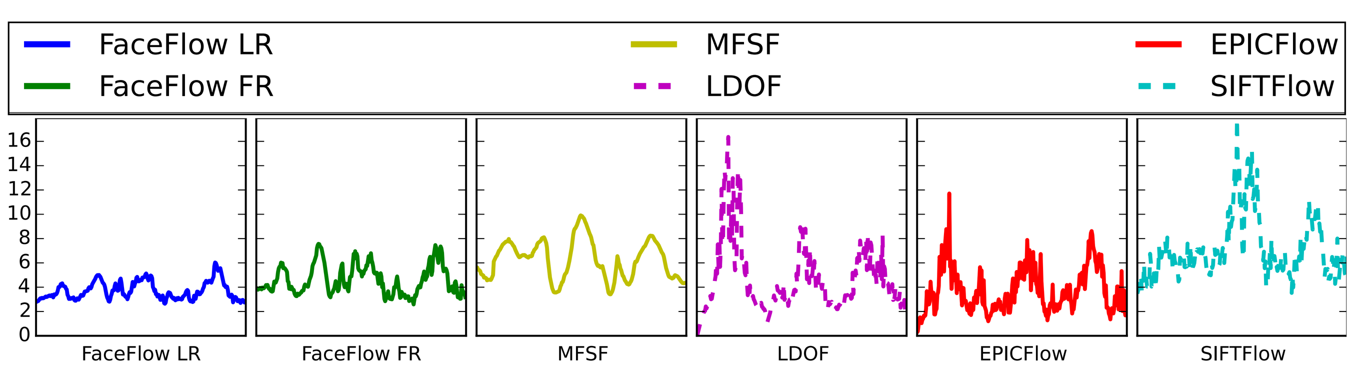
Synthetic Results



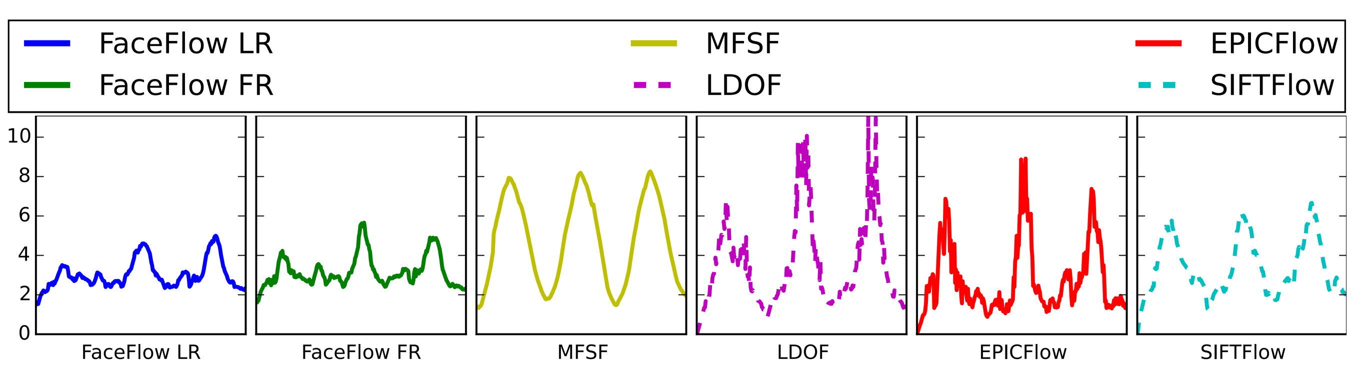
We generated a synthetic sequence using motion capture data [5]. We tested a number of state of the art optical flow methods on this data under 2 challenging scenarios - rendered under illumination variation and rendered under illumination plus an artificial occlusion.

This table shows the Endpoint Error, in terms of the Root-Mean Squared Error (RMSE) and the 95% percentile of the Average Endpoint Error (AE95).

	Original		Illum.		Ilum.+Occ.	
	RMSE	AE95	RMSE	AE95	RMSE	AE95
Face Flow Low-Rank	2.95	5.52	**3.56**	**6.63**	**4.48**	**8.47**
Face Flow Full-Rank	3.24	6.01	3.76	7.02	5.83	11.50
MFSF	1.73	3.20	6.33	13.68	8.25	17.30
LDOF	**1.56**	**2.79**	4.84	9.98	6.54	11.44
EPICFlow	1.66	3.25	4.02	9.61	5.15	11.61
SIFTFlow	2.65	5.15	4.89	11.81	11.82	23.05



The average endpoint error calculated for each frame of the illumination + occlusion variation mocap sequence. Vertical axis is average endpoint error, horizontal is frame number.



The average endpoint error calculated for each frame of the illumination variation mocap sequence. Vertical axis is average endpoint error, horizontal is frame number.

Visit the QR code to see a Youtube Video of the results!

References

[1] Garg, Ravi, Anastasios Roussos, and Lourdes Agapito. "A variational approach to video registration with subspace constraints." IJCV 104.3 (2013): 286-314.
Code available at **https://bitbucket.org/troussos/mfsf/**

[2] Yin, Lijun, et al. "A 3D facial expression database for facial behavior research." FG, 2006.

[3] Brox, Thomas, and Jitendra Malik. "Large displacement optical flow: descriptor matching in variational motion estimation." T-PAMI 33.3 (2011): 500-513.

[4] Revaud, Jerome, et al. "EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow." CVPR 2015

[5] Zhang, Li, et al. "Spacetime faces: High-resolution capture for~ modeling and animation." Data-Driven 3D Facial Animation. Springer London, 2008. 248-276.

[6] Liu, Ce, Jenny Yuen, and Antonio Torralba. "Sift flow: Dense correspondence across scenes and its applications." T-PAMI 33.5 (2011): 978-994.