L-Università ta' Malta

Sequential Non-Rigid Factorisation for Head Pose Estimation

Stefania Cristina, Kenneth P. Camilleri

Department of Systems and Control Engineering,
University of Malta

stefania.cristina@um.edu.mt, kenneth.camilleri@um.edu.mt

Introduction

- Within the context of eye-gaze tracking, the capability of permitting the user to move naturally is an important step towards allowing for more natural user interaction in less constrained scenarios.

- Non-rigid facial expressions are a common occurrence during tracking, and it has also been shown that these introduce pose estimation errors if they are not catered for [1].

- Nonetheless, this challenge has not been widely addressed by both the eye-gaze tracking and the head pose estimation communities.

- The few methods that factor the challenge of handling face deformations into the head pose estimation problem, often require the availability of a pre-defined face model or a considerable amount of training data [2,3].

[1] S. Cristina and K. P. Camilleri, “Model-free non-rigid head pose tracking by joint shape and pose estimation,” *Machine Vision and Applications*, vol. 27, pp. 1229–1242, 2016.

[2] Y. Wu, C. Gou, and Q. Ji, “Simultaneous facial landmark detection, pose and deformation estimation under facial occlusion,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5719–5728.

[3] K. Oka and Y. Sato, “Real-time modeling of face deformation for 3d head pose estimation,” in *International Workshop on Analysis and Modeling of Faces and Gestures*, 2005, pp. 308–320.

Introduction

- In this paper, we direct our attention towards the application of shape-and-motion factorisation for head pose estimation, since this does not generally rely on the availability of an initial face model.

- Over the years, various shape-and-motion factorisation methods have been proposed to address the challenges of rigid and non-rigid shape and motion recovery, in a batch or sequential manner. However, the real-time recovery of non-rigid shape and motion by factorisation remains, in general, an open problem.

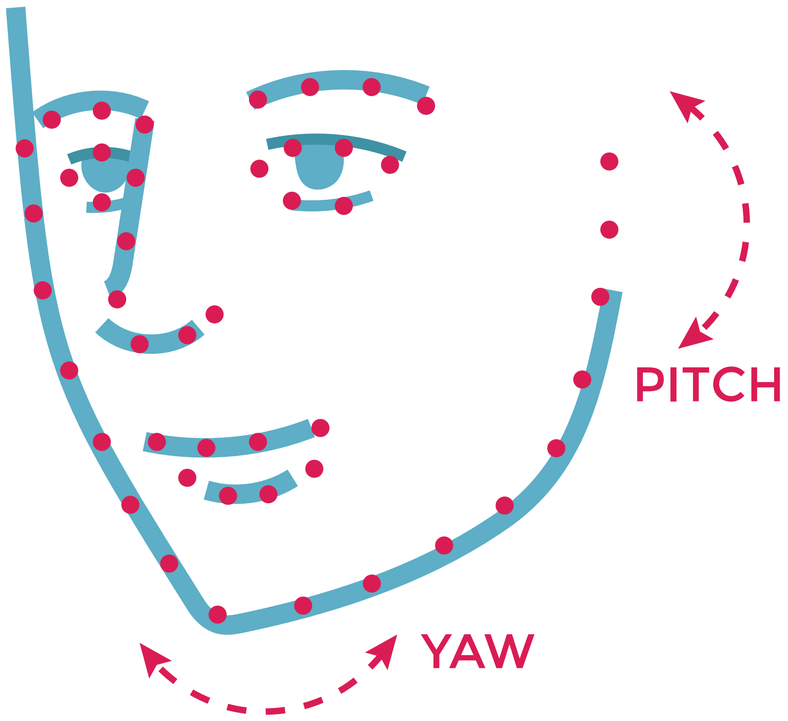
- Our work combines the sequential rigid method of Morita and Kanade [4] together with the non-rigid batch-type method of Bregler et al. [5] into a sequential factorisation method for non-rigid shape and motion recovery.

[4] T. Morita and T. Kanade, “A sequential factorization method for recov- ering shape and motion from image streams,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 858–867, 1997.
[5] C. Bregler, A. Hertzmann, and H. Biermann, “Recovering non-rigid 3d shape from image streams,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000, pp. 690–696.

Method

- P face landmark points are first localised by the method of Kazemi and Sullivan [6] and stored in a covariance-type matrix, $\mathbf{Z}_f$, of size $P \times P$ [4]:

$$\mathbf{Z}_f = \mathbf{Z}_{f-1} + \mathbf{x}_f \mathbf{x}_f^T + \mathbf{y}_f \mathbf{y}_f^T$$



[6] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874.

Method

- The covariance-type matrix, $\mathbf{Z}_f$, is related to the measurement matrix, $\mathbf{W}_f$, as follows: $\mathbf{Z}_f = \mathbf{W}_f^T \mathbf{W}_f$

  where $\mathbf{W}_f$, of size $2F \times P$, collects the $\mathbf{x}_f$ and $\mathbf{y}_f$ landmark coordinates over $f = 1, \ldots, F$ image frames.

- If matrix, $\mathbf{W}_f$, can be decomposed into unitary matrices, $\mathbf{U}_f$ and $\mathbf{V}_f$, and diagonal matrix, $\mathbf{\Lambda}_f$: $\hat{\mathbf{W}}_f = \mathbf{U}_f \mathbf{\Lambda}_f \mathbf{V}_f^T$

  where $\hat{\mathbf{W}}_f$ is the best estimate of $\mathbf{W}_f$ following its decomposition by singular value decomposition (SVD) in the presence of noise.

- Then it follows that $\hat{\mathbf{Z}}_f$ may also be similarly decomposed [4]: $\hat{\mathbf{Z}}_f = (\mathbf{U}_f \mathbf{\Lambda}_f \mathbf{V}_f^T)^T \mathbf{U}_f \mathbf{\Lambda}_f \mathbf{V}_f^T = \mathbf{V}_f \mathbf{\Lambda}_f^2 \mathbf{V}_f^T$

Method

- This is compared to the formulation of the measurement matrix proposed by the batch-type method of Bregler et al. [5]:

$$\hat{\mathbf{W}} = \begin{bmatrix} l_{11}\hat{\mathbf{M}}_1 & \dots & l_{K1}\hat{\mathbf{M}}_1 \\ l_{12}\hat{\mathbf{M}}_2 & \dots & l_{K2}\hat{\mathbf{M}}_2 \\ & \vdots & \\ l_{1F}\hat{\mathbf{M}}_F & \dots & l_{KF}\hat{\mathbf{M}}_F \end{bmatrix} \cdot \begin{bmatrix} \hat{\mathbf{S}}_1 \\ \hat{\mathbf{S}}_2 \\ \vdots \\ \hat{\mathbf{S}}_K \end{bmatrix} = \hat{\mathbf{Q}}\,\hat{\mathbf{B}}$$

- where matrix $\hat{\mathbf{Q}}$ contains the motion matrices, $\hat{M}_f$, and configuration weights, $l_{1f}, \dots, l_{Kf}$ for K basis shapes, while matrix $\hat{\mathbf{B}}$ contains the key-frame basis shapes.

- It is seen that the eigenvectors $\mathbf{V}_f$ of $\hat{\mathbf{Z}}_f$ capture the non-rigid face deformations in matrix, $\hat{\mathbf{B}}_f$.

Method

- The vectors comprising matrix, $\hat{\mathbf{Q}}_f$, are subsequently recovered [1]: $\mathbf{q}_f^{(1)} = \mathbf{x}_f^T \hat{\mathbf{B}}_f \qquad \mathbf{q}_f^{(2)} = \mathbf{y}_f^T \hat{\mathbf{B}}_f$

  where $\mathbf{q}_f^{(1)}$ and $\mathbf{q}_f^{(2)}$ of size $1 \times 3K$ correspond to the two rows that form matrix, $\hat{\mathbf{Q}}_f$.

- A rank-1 factorisation is applied on $\hat{\mathbf{Q}}_f$ to recover the motion matrix, $\hat{\mathbf{M}}_f$ and the configuration weights, l_{Kf} [4].

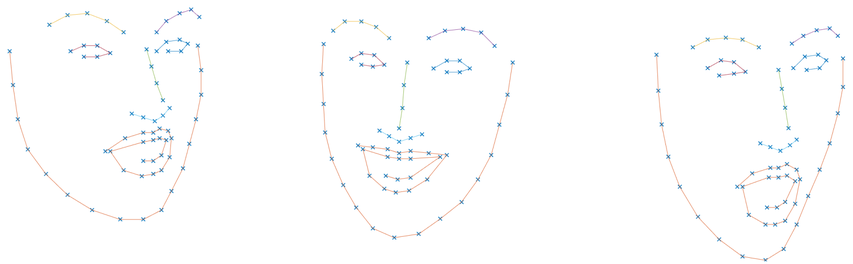
- The final step enforces orthonormality constraints on the motion matrix, $\hat{\mathbf{M}}_f$, resulting in the computation of a 3×3 transformation matrix producing a unique decomposition of the measurement matrix: $\mathbf{A}_f = \mathbf{U}_f^{(M)} \, \mathbf{V}_f^{(M)T}$

- Matrices, $\mathbf{U}_f^{(M)}$ and $\mathbf{V}_f^{(M)}$ are the unitary matrices resulting from an SVD operation on matrix, $\hat{\mathbf{M}}_f$.

Experimental Procedure

- A subject was cued to perform various facial expressions: neutral, happiness, sadness, fear, surprise, anger, disgust and contempt.

- Each facial expression lasts 100 image frames, and ground truth head yaw, pitch and roll angles were recorded by an inertial measurement unit positioned on top of the head.



Experimental Procedure

- A further evaluation was carried out on image sequences from the Extended Cohn-Kanade (CK+) dataset to widen the expressions and subject appearances under consideration. Each image sequence starts with a neutral facial pose and ends with the peak formation of the facial expression.

- In absence of ground truth head rotation angles, subjects maintaining a stationary frontal and upright head pose were manually selected.

Results - Cued Dataset

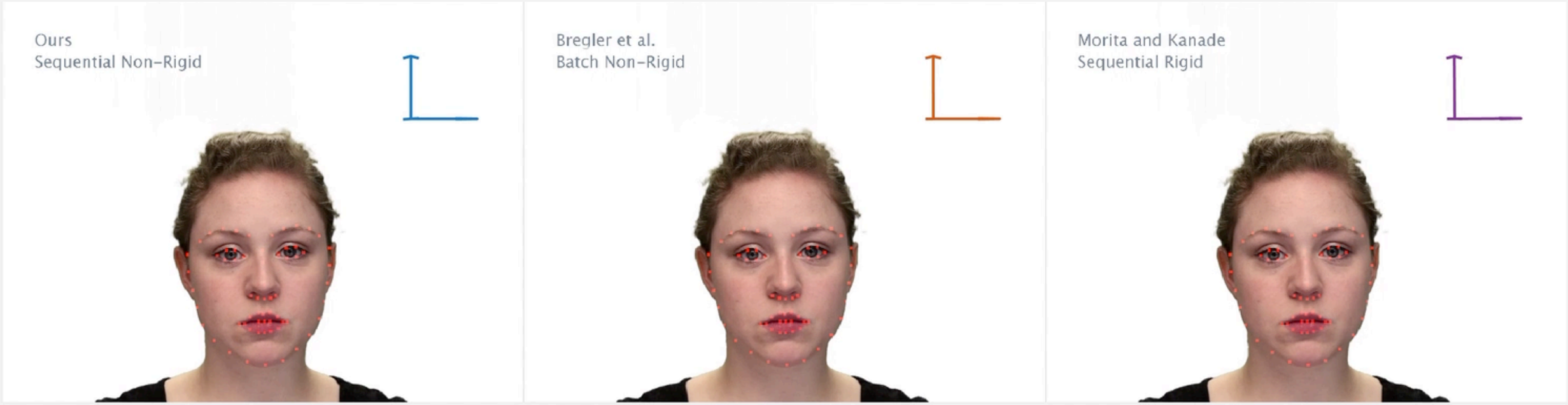
Test case	Method	MAE/° (SD/°) Yaw	Pitch	Roll	Range of head rotation angles/[Min/°, Max/°)] Yaw	Pitch	Roll
1	Proposed	0.22 (0.30)	0.35 (0.45)	0.38 (0.54)	[-0.19, 0.92]	[-0.48, 0.24]	[-0.23, 2.58]
	[5]	0.86 (1.25)	1.62 (2.06)	0.57 (1.38)			
	[4]	0.37 (0.46)	2.27 (2.73)	0.30 (0.41)			
2	Proposed	2.29 (3.33)	2.87 (3.46)	1.70 (2.50)	[-18.38, 18.60]	[-19.39, 16.82]	[-4.49, 6.74]
	[5]	2.95 (4.31)	2.16 (3.56)	1.48 (2.19)			
	[4]	7.56 (8.18)	4.86 (5.95)	4.55 (6.00)			
3	Proposed	0.24 (0.35)	2.26 (2.68)	0.48 (0.67)	[-1.49, 0.53]	[-0.40, 1.36]	[-1.37, 0.22]
	[5]	0.58 (0.96)	6.45 (7.80)	0.87 (1.41)			
	[4]	8.06 (12.00)	10.16 (18.46)	6.30 (8.07)			
4	Proposed	2.83 (4.33)	2.97 (3.71)	1.82 (2.40)	[-15.27, 24.54]	[-15.01, 12.91]	[-5.45, 6.19]
	[5]	6.58 (9.52)	4.17 (6.51)	2.47 (3.08)			
	[4]	3.91 (5.21)	24.90 (30.57)	25.07 (28.70)			
5	Proposed	1.50 (1.86)	1.83 (2.34)	1.04 (1.29)	[-15.10, 21.93]	[-14.70, 13.36]	[-4.45, 5.53]
	[5]	1.76 (2.29)	2.70 (4.74)	1.30 (1.81)			
	[4]	4.88 (5.25)	13.52 (15.32)	3.42 (4.10)			
6	Proposed	2.13 (2.46)	1.24 (1.62)	1.35 (1.87)	[-18.83, 14.29]	[-7.01, 11.17]	[-10.79, 10.06]
	[5]	3.16 (4.33)	1.39 (1.91)	1.36 (2.26)			
	[4]	2.32 (2.80)	3.77 (4.96)	3.30 (3.88)			

1. Stationary head and rigid face.
2. Free head movement and rigid face.
3. Stationary head pose and cued non-rigid face deformations.
4. Free head movement and cued non-rigid face deformations.
5. Free head movement and cued non-rigid face deformations, followed by free head movement and rigid face, followed by stationary head pose and rigid face.
6. Free head movement and natural non-rigid face deformations during conversation.

Results - CK+ Dataset

Subject - Sequence number	Method	MAE/° (SD/°)			Method	MAE/° (SD/°)		
		Yaw	Pitch	Roll		Yaw	Pitch	Roll
S034 - 004	Proposed	0.23 (0.28)	1.58 (1.78)	0.48 (0.53)	[5]	0.42 (0.53)	4.79 (5.48)	0.15 (0.21)
S037 - 004		0.10 (0.14)	1.18 (1.81)	0.06 (0.09)		0.07 (0.13)	1.97 (2.99)	0.10 (0.13)
S057 - 005		0.10 (0.14)	0.91 (1.13)	1.03 (1.15)		0.33 (0.38)	4.59 (5.45)	0.06 (0.07)
S074 - 005		0.28 (0.35)	0.78 (0.99)	0.47 (0.65)		1.46 (1.62)	4.29 (4.56)	0.62 (0.66)
S089 - 003		1.49 (1.61)	0.40 (0.56)	0.48 (0.55)		1.42 (1.50)	0.08 (0.10)	0.62 (0.70)
S097 - 003		0.15 (0.17)	0.95 (1.01)	0.13 (0.15)		0.30 (0.31)	4.19 (4.41)	0.10 (0.15)
S113 - 007		0.14 (0.17)	0.69 (0.89)	0.32 (0.36)		0.55 (0.64)	3.35 (3.95)	0.20 (0.22)
S124 - 005		0.14 (0.16)	0.59 (0.72)	0.38 (0.44)		0.13 (0.16)	0.46 (0.66)	0.07 (0.11)
S126 - 001		0.11 (0.21)	0.70 (0.85)	0.50 (0.55)		0.46 (0.52)	5.25 (5.91)	0.06 (0.06)
S133 - 003		0.18 (0.23)	1.31 (1.41)	1.68 (1.93)		0.34 (0.49)	7.00 (8.06)	0.23 (0.27)
S135 - 001		0.11 (0.14)	1.37 (1.59)	0.40 (0.44)		0.35 (0.45)	5.45 (6.72)	0.32 (0.43)
S503 - 001		0.59 (0.65)	0.11 (0.12)	0.47 (0.60)		1.71 (1.78)	0.30 (0.31)	0.12 (0.16)
S999 - 003		0.21 (0.29)	1.13 (1.43)	0.89 (1.05)		0.04 (0.04)	0.06 (0.07)	0.35 (0.51)
Mean		0.29 (0.35)	0.90 (1.10)	0.56 (0.65)		0.58 (0.66)	3.21 (3.74)	0.23 (0.28)

Results


Ours
Sequential Non-Rigid
Bregler et al.
Batch Non-Rigid
Morita and Kanade
Sequential Rigid

Conclusion

- In this work, we have proposed a method that combines the sequential rigid method of Morita and Kanade [4] together with the non-rigid batch-type method of Bregler et al. [5] into a sequential factorisation method for non-rigid shape and motion recovery.

- The results revealed that the proposed method performed better than the batch-type method of Bregler et al. [5], with the important advantage of running in real-time rather than in batch mode.

- The improvement in accuracy over the rigid factorisation method of Morita and Kanade [4] confirms the importance of compensating for non-rigid face deformations.