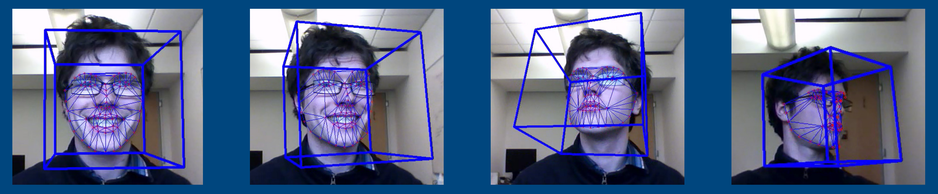
3D Constrained Local Model for Facial Tracking

Tadas Baltrušaitis, Peter Robinson & Louis-Philippe Morency



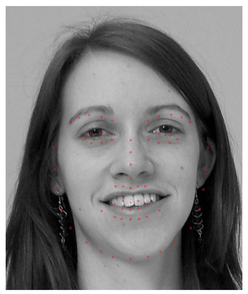
Motivation

Facial expression and head pose are rich sources of information which provide an important communication channel for interaction. A crucial initial step in many affect sensing, face recognition and human behaviour understanding systems is the estimation of head pose and tracking of facial feature points. The availability of accurate and affordable depth cameras leads us to explore this new channel for facial tracking.

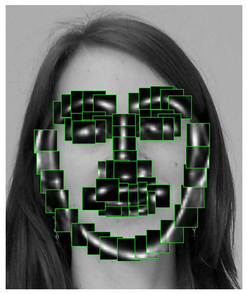
Contributions

- 3D Constrained Local Model (CLM-Z) formulation that takes full advantage of both depth and intensity information in a unified framework.
- Robust normalisation function for depth patches
- Combining a rigid and non-rigid facial trackers

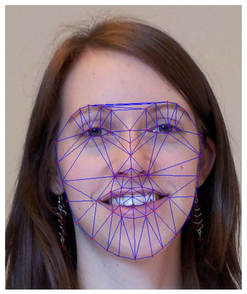
Overview



First, an initial location of feature points is estimated from a region provided by a face detector or Generalised Adaptive View-Based Model [1].



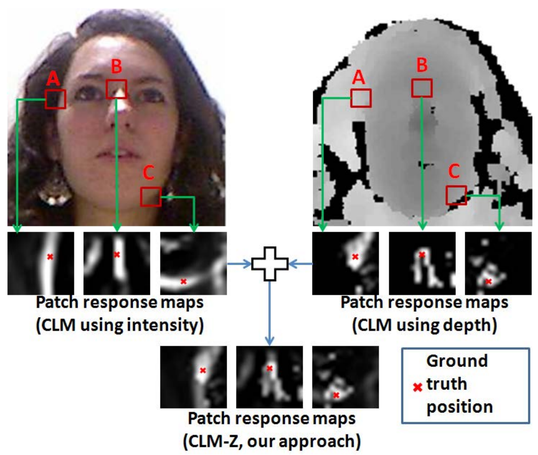
Secondly, patch experts around the initial estimates are evaluated to provide response maps for each facial feature points. This is done on both intensity and depth images.



Finally, an optimisation step which maximises the *a posteriori* probability estimate is performed. It finds the locations of feature points that satisfy both the statistical face shape model and local patch experts best.

Patch experts

To create a patch response we sum both intensity and depth patch responses resulting in a combined response. As patch experts we use linear Support Vector Machines in combination with logistic regressors.



Response maps of three patch experts. Intensity response maps (left) contain strong responses along the edges, making it hard to find the actual feature position when using intensity alone. Adding depth allows us create more accurate and robust patch experts.

Values may be missing from the depth data and these must be discounted. We ignore them when calculating the mean and standard deviation used for normalisation, and afterwards set the missing entries to zero.

Model fitting

For CLM-Z fitting we use Regularised Landmark Mean Shift [2] to find the maximum *a posteriori* estimate of our model.

$$p(\mathbf{p}|\{l_i=1\}_{i=1}^n,\mathcal{I},\mathcal{Z}) \propto p(\mathbf{p}) \prod_{i=1}^{n} p(l_i=1|\mathbf{x}_i,\mathcal{I},\mathcal{Z})$$

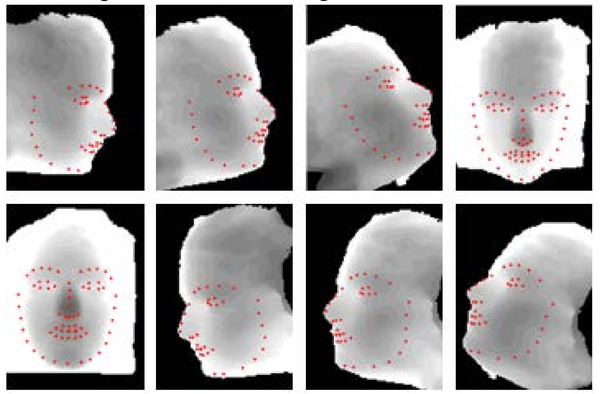
Our model formulation allows us to incorporate both the point distribution model of a face together with our intensity and depth patch experts.

Rigid body tracking

CLM-Z and CLM trackers are not very accurate at estimating the head pose, have no indication of successful convergence. Therefore, we combine our approach with Generalised Adaptive View-Based Model. Both trackers help each other during tracking by providing each other with priors.

Synthetic data generation

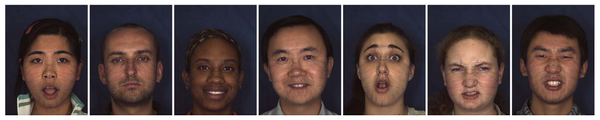
To generate depth images for patch expert training we used the BU-4DFE dataset of range scans of facial expressions. This allowed us to generate training examples at different poses from a single labeled texture image.



Examples of synthetic depth images used for training depth patch experts

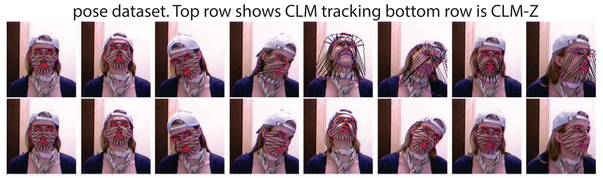
Evaluation

We evaluated our CLM-Z approach on 4 publicly available datasets, 3 of which had head pose ground truth data. We also hand labeled subsets of Biwi Kinect head pose dataset and BU-4DFE for facial feature points.



Examples of tracked feature points on BU-4DFE dataset using CLM-Z

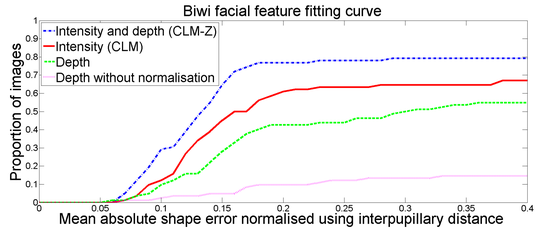
Examples of tracked feature points on a sequence from Biwi Kinect head pose dataset. Top row shows CLM tracking bottom row is CLM-Z

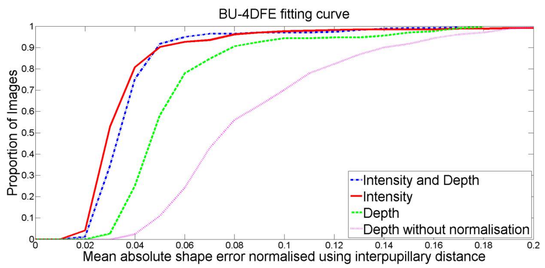


Examples of tracked feature points on a sequence from ICT-3DHP dataset. Top row shows CLM tracking bottom row is CLM-Z

Examples of tracked feature points and head pose of a sequence from Boston University head pose dataset

Results for Non-rigid facial tracking





Results for Rigid facial tracking

All pose errors are measured in mean absolute degree error.

Method	Yaw	Pitch	Roll	Mean
Regression forests	9.2	8.5	**8.0**	8.6
CLM	28.85	18.30	28.49	25.21
CLM-Z	14.80	12.03	23.26	16.69
CLM-Z with GAVAM	**6.29**	**5.10**	11.29	**7.56**

Head pose tracking on the Biwi Kinect head pose dataset

Method	Yaw	Pitch	Roll	Mean
GAVAM	3.79	4.45	2.15	3.47
CLM	5.23	4.46	2.55	4.08
CLM with GAVAM	**3.00**	**3.81**	**2.08**	**2.97**

Head pose tracking on the Boston University head pose dataset

Method	Yaw	Pitch	Roll	Mean
Regression forests [14]	7.17	9.40	7.53	8.03
GAVAM [19]	3.00	3.50	3.50	3.34
CLM [23]	11.10	9.92	7.30	9.44
CLM-Z	6.90	7.06	10.48	8.15
CLM-Z with GAVAM	**2.90**	**3.14**	**3.17**	**3.07**

Head pose tracking on the ICT-3DHP dataset

The results demonstrate the benefits of our contributions: using a novel CLM-Z model, our normalisation function, and our combination of trackers.

References

[1] L.-P. Morency, J. Whitehill, and J. R. Movellan. Generalized Adaptive View-based Appearance Model: Integrated Framework for Monocular Head Pose Estimation. Face and Gesture, 2008

[2] J. Saragih, S. Lucey, and J. Cohn. Deformable Model Fitting by Regularized Landmark Mean-Shift. International Journal of Computer Vision, 2011

THALES

http://www.cl.cam.ac.uk/emotions/facetrack

UNIVERSITY OF CAMBRIDGE Computer Laboratory

RAINBOW RESEARCH GROUP