Facial Feature Detection and Tracking with a 3D Constrained Local Model

Meng Yu
University of St. Andrews, U.K.
yumeng@cs.st-andrwes.ac.uk

Bernard P. Tiddeman
University of St. Andrews, U.K.
bpt@cs.st-andrews.ac.uk

ABSTRACT

In this paper, we describe a system for facial feature detection and tracking using a 3D extension of the Constrained Local Model (CLM) [Cris 06, Cris 08] algorithm. The use of a 3D shape model allows improved tracking through large head rotations. CLM uses a joint shape and texture appearance model to generate a set of region template detectors. A search is then performed in the global pose / shape space using these detectors. The proposed extension uses multiple appearance models from different viewpoints and a single 3D shape model. During fitting or tracking the current estimate of pose is used to select the appropriate appearance model. We demonstrate our results by fitting the model to image sequences with large head rotations. The results show that the proposed 3D constrained local model algorithm improves the performance of the original CLM algorithm for videos with large out-of-plane head rotations.

Keywords: Active appearance models, Multi-view face models, Constrained local model, Facial feature tracking, Facial feature detection

1 INTRODUCTION

This paper describes a method for tracking human face features using a 3D shape model and view-dependent feature templates. We match the 3D face model to previously unseen 2D video sequences of human faces by applying a shape constrained search method, using an extension of the constrained local model algorithm.

The original CLM algorithm [Cris 06] works with limited rotations from the front face view. The extension to the algorithm proposed here works not only on the front face view but also on the face with large head rotations in videos. The proposed multi-view CLM consists of a 3D shape model and several 2D texture models from multiple views.

In our implementation, the shape model is first given some suitable initialisation (approximate rigid body alignment, scaling). In each subsequent iteration square region are sampled around each feature point and projected into the allowed appearance model space. The shape and pose parameters are then found that maximise the correlation between the synthesised appearance template patches and patches extracted around the current estimates of the feature point locations in image space. The proposed algorithm is a view based, in that we switch appearance models depending on the current estimate of the face orientation.

After a brief review of face and face feature detection, we will describe the model building and fitting methods in more detail, followed by experimental results demonstrating the performance of the proposed multi-view CLM method.

2 RELATED WORK

The problems of facial feature detection and tracking have received a great deal of attention in the literature, here we only cover the more immediately relevant work. Active Shape Models (ASM) [Coot 95] use Principal Component Analysis (PCA) to learn the main axes of variation from a training set of labelled examples. Fitting the shape model to a new image involves local searches for matching features alternated with projection of the shape estimate back into the allowed model space.

Active Appearance Models(AAMs) [Coot 98, Coot 01a] use the same PCA based shape model as ASMs together with a PCA based model of appearance (i.e. shape normalised texture). It has been used for face modelling and recognising objects [Lani 97, Jone 98], fitting unseen images [Gros 05, Peyr 07], tracking objects[Ahlb 01, Steg 01] and medical image processing [Coot 01b, Mitc 01]. The original implementation [Coot 01a] learnt a linear model relating the error image (between the model and the image) and the required parameter updated at each time step. Following the forwards additive algorithm [Luca 81], the inverse additive algorithm [Hage 98], and the forwards compositional algorithm [Shum 01], Mathews and Baker [Bake 01, Bake 02, Matt 04] derived more mathematically elegant methods in

WSCG 2010 conference proceedings,
WSCG'2010, February 1 – February 4, 2010
Plzen, Czech Republic.

which the updates are always calculated in the average shape and then concatenated with the current guess. This inverse compositional method allows the pre-computation of the gradient images and inverse Hessian matrix for greater efficiency. Later work demonstrated that the inverse compositional algorithm is only really suitable for person-specific fitting and tracking, and that simultaneous estimation of the shape and appearance parameters was required for robust face fitting [Gros 05].

Cristancce et al. [Cris 06, Cris 08] proposed a patch based algorithm to model a deformable object. Their Constrained Local Model (CLM) algorithm is another method with the similar appearance model to that used in the AAMs [Coot 98]. It learns the variation in appearance of a set of template regions surrounding individual features instead of triangulated patches. The fitting algorithm first finds the best match of the combined shape-appearance model to the current guess, then searches locally using a non-linear optimiser to find the best match to the model. Further study on patch based appearance models have been carried out – exhaustive local search (ELS) algorithm [Wang 07], generic convex quadratic fitting (CQF) approach [Wang 08] and Bayesian constrained local models (BCLM) [Paqu 09]. The approach has been proven to outperform the active appearance models (AAMs) [Coot 01a] as it is more robust to occlusion and changes in appearance and no texture warps are required. ELS, CQF and BCLM are all gained some improvements over CLM fitting to certain databases. In this work, we will use the original normalised cross correlation error metric and concentrate on the extensions to large head rotations.

Active appearance models (AAMs) [Coot 98, Coot 01a] were originally formulated as 2D and most of the algorithms for AAM fitting have been a single-view [Coot 02]. Automatically locating detailed facial landmarks across different subjects and viewpoints, i.e. 3D alignment of a face, is a challenging problem. Previous approaches can be divided into three categories: view (2D) based, 3D based and combined 2D+3D based. View based methods [Coot 00, Zhou 05, Fagg 05, Peyr 08], train a set of 2D models, each of which is designed to cope with shape or texture variation within a small range of viewpoints. We have found for some applications that switching between 2D views can cause notable artefacts (e.g. in face reanimation). 3D based methods [Blan 99, Romd 02, Bran 01, Jone 98, Vett 97, Zhan 04], in contrast, deal with all views by a single 3D model. 3D Morphable model fitting is an expensive search problem in a high dimensional space with many local minima, which often fails to converge on real data. 2D+3D based methods [Xiao 04, Hu 04, Kote 05, Ramn 08] used AAMs and

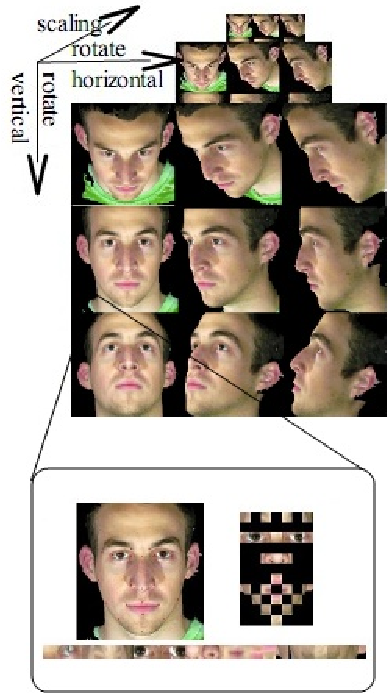

Figure 1: The Mult-view CLM consists of a shape model and several appearance models from different views. There are 15 views and 3 scales used to cover all the likely circumstances in the application. (There are only 9 views in the figure since the views from the right side are approximately mirrorring copies of the ones from the left side.)

estimated 3D shape models to track faces in videos, but these algorithms are generally most suitable in the person specific context. Our proposed multi-view CLM algorithm is a 3D extension of the CLM algorithm [Cris 06, Cris 08] which could be more useful for fitting to unseen face images or tracking.

3 ALGORITHM

3.1 An Overview

The model (*Figure* 1) consists of a model of 3D shape variation and 15 models of the appearance variations in a shape-normalised frame. A training set of labelled images, where key landmark points are marked on each example object, is required. We use landmark points placed on a set of 3D face models to generate the 3D shape model. The appearance model for each view is found by rendering the face model from the appropriate viewpoint and sampling square patches from the rendered image about the projected location of the feature point.

We use 14 subjects (8 males, 6 females) performing 7 posed expressions (neutral, happy, sad, disgust, surprise, fear, anger) and 7 posed vise mes (/ah/, /ch/, /ee/, /k/, /oo/, /p/, /th/) captured using a stereophotogrametric system (www.3dMD.com). From the set of landmark

Figure 2: Example of training images

points, a statistical model of shape variation can be generated using Principal Component Analysis (PCA). We extract a 20x20 block of pixels around each feature point at each of 3 spatial scales. (*Figure* 2) These patches are vectorised and used to build the appearance model. All the features are formed into a 500x20 block of pixel strip before the PCA analysis is applied.

In the original CLM work a combined shape and appearance models was created by performing PCA on the combined shape and appearance parameter vectors, and the search was carried out in this space. The use of multiple appearance models in out algorithm would require the use of multiple combined models. In order to simplify switching of the appearance model with a single shape model in this work we use separate models of shape and appearance instead of using a combined model. In future work we will experiment with switching with a combined model.

3.2 Shape Model

The shape model is built from normalised shape coordinates $s(x,y,z)$. To calculate the principal components, the templates are aligned to the average template (in a standard position) by performing a rigid body and scale transform. The covariance matrix of the vectorised points (templates), X, is created using the formula:

$$Cov = \frac{1}{N-1}\sum_{i=0,j=0}^{m,n}(X_i-\overline{X})(X_j-\overline{X}) \qquad (1)$$

where $\bar{X}$ is the mean of the vectorised points, N is the number of the templates. The model is then obtained by applying Jacobi's method to the covariance matrix to find the eigenvectors and eigenvalue, which represent the principal components and their distributions.

$$s = \bar{s} + P_s b_s \qquad (2)$$

where $\bar{s}$ is the mean shape, P_s is a set of orthogonal modes of variation derived from the shape templates training set and b_s is a set of shape parameters. The equation can then be used to reconstruct new shapes by varying the given shape parameters.

Figure 3: The first strip is the CLM texture model for a particular view. The second strip holds the stencil channel which is used to exclude hidden or background pixels.

The two-dimensional coordinates of the shape model can be calculated with the following equation:

$$s_{2d} = M \cdot V \cdot (\bar{s} + P_s \cdot b_s) \qquad (3)$$

where V is a vector of the pose (translation, rotation, scaling) transforming parameters $T_x, T_y, T_z, S, \theta, \phi, \gamma$ and M is the opengl frustum projection matrix.

3.3 Appearance Models

To build a model of the appearance, we render each 3D face model in our training set from a particular viewpoint and sample a square patch around each feature point. By transforming the face with different scale, rotation, shift and lighting parameters, we build a set of texture patches. After vectorising all the patches, PCA analysis is applied to the textures from a particular viewpoint and scale to build an appearance model:

$$g = \bar{g} + P_g b_g \qquad (4)$$

where $\bar{g}$ is the mean normalized gray-level vector, P_g is a set of orthogonal modes of variation derived from appearance training sets, the texand b_g is a set of gray-level parameters. We build 45 appearance models, one for each of 15 viewpoints across 3 different scales.

3.4 Other Features

To increase the stability with varied backgrounds, we use visibility information from the rendered patches to estimate occluded pixels. We grab the stencil channel (*Figure* 3) from the rendering canvas when we extract the texture patches to mark out the edges between the face and the background. We also compare the depth of the projected landmark with the depth buffer to identify self occlusion, and set the grabbed stencil buffer values to zero for occluded points. In the current work, we use a fixed visibility model for each viewpoint based on the average for the model.

Multi-scale techniques are standard in computer vision and image processing. They allow short range models to extend over longer ranges and optimisation to be achieved in fewer steps. In our model, a Gaussian filter is used as the convolution function to build a multi-scale texture pyramid. The processing time is much shorter with lower resolution images. So we fit the unseen image with the lowest resolution image first to improve the performance. When fitting we also use a Gaussian pyramid built for each frame and then the CLM search is applied at each layer from the coarsest to the finest. The process can be illustrated in *Figure* 4.

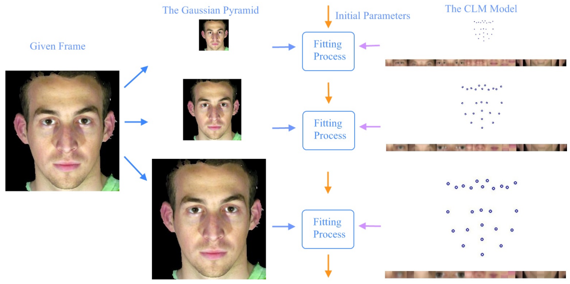

Figure 4: A skeleton of the three scales image searching technique.

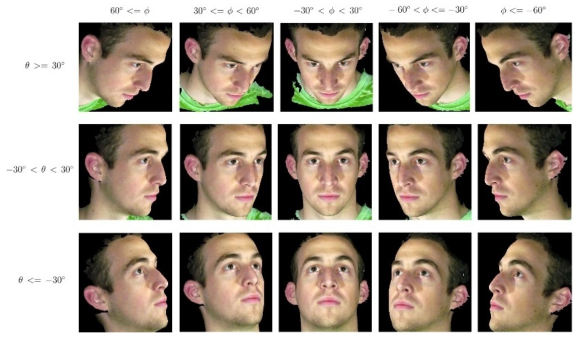

Figure 5: Multiple appearance models.

3.5 Texture Model Selection

In the proposed algorithm, there are a global three-dimensional shape model and fifteen texture models (three in vertical and five in horizontal). One additional step to the original algorithm is the selection of the texture model while searching with the multi-view CLM algorithm. For tracking face movements, the algorithm needs to select the correct texture model for the current pose automatically. As each texture model is built from a certain view, we can use the rotation parameters θ, ϕ to estimate the view by testing the criteria shown in *Figure* 5. θ and ϕ can be obtained from the current estimate of head rotation using one of the shape template updating methods.

The texture model selection process is given by the following steps repeatedly until the end of the tracking.

1. The multi-view CLM algorithm is applied to the given frame accompanied with the initial parameters.

2. A set of new parameters are obtained including θ and ϕ which is the estimated rotation angles for the current face pose.

3. To estimate the next frame, θ and ϕ is then passed into the texture model selection module to choose the proper appearance model.

3.6 Search Algorithm

With the texture model selection algorithm, we can extend the searching method [Cris 06] for use with a three-dimensional model.

For a given set of initial points, $X = (x_0, y_0, z_0, x_1, y_1, z_1 \cdots, x_{n-1}, y_{n-1}, z_{n-1})$, the initial pose parameters V are estimated for the shape model built previously. Then the multi-view appearance CLM algorithm shown in *Figure* 6 is applied.

1. Initialise with the global face detector.

2. Estimate the initial pose parameters V.

3. Repeat

   (a) Repeat

      i. Compute the feature coordinates, s, and extract the feature patches, g.

      ii. Estimate the texture model from the pose parameters V.

      iii. Synthesise the feature patches from the updated coordinates and the selected texture model.

      iv. Apply the alpha channel feature, the hidden points feature to the extracted and synthetic feature patches.

      v. Optimise the error metrics with the shape template updating methods to get a new set of pose and shape parameters, V, b_s.

   (b) Until converged.

4. Until converged for all selected scales.

3.7 Shape updating methods

The original CLM algorithm [Cris 06] used the Nelder-Meade simplex algorithm [Neld 65] to optimize the Cross Correlation. This algorithm works by using N+1 samples in the N dimensional parameter space. Each iteration the worst sample is discarded and a new sample is added based on a set of simple heuristics. In this work, we instead use Powell's method [Pres 07] as this is supposed typically to require fewer function evaluations than the Nelder-Meade algorithm.

Optimisation techniques based on off-the-shelf non-linear optimisers like those described above are typically slow to converge. We compare optimisation using Powell's method with a direct method for optimising the global NCC using an estimate of the Jacobean and Hessian matrices and solving a linear system and a quadratic equation [Tidd 07].

We also compare the techniques described above with minimisation of the sum of squared errors (SSE) as an error metric. This is similar to the above, requiring the Jacobean and inverse Hessian matrices and solution of a linear system. This method is essentially equivalent to the additive inverse compositional AAM alignment, [Hage 98, Matt 04] wrapped in a slightly different fitting algorithm.

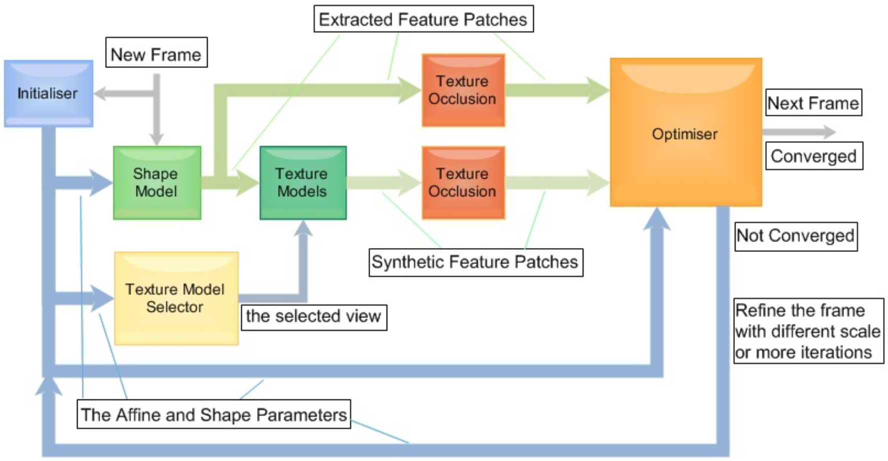

Figure 6: Multi-view CLM tracking process

4 RESULTS

We have evaluated the proposed algorithms using a mixture of synthetic and real data. Synthetic data is generated by rendering multiple 3D face scans from different viewpoints, and is useful because the 3D models provide accurate ground-truth data. We also test the algorithms on real video data with hand-labelled feature points.

We have designed two experiments to evaluate the proposed multi-view appearance 3D CLM. The first to compare single-view CLM to the proposed multi-view CLM. The second set of experiments compare the various optimisation algorithms within the multi-view CLM framework.

4.1 Synthetic Data Experiments

This experiment aims to compare the performance of the proposed multi-view 3D CLM algorithm to the single-view CLM algorithm. A set of face sequences with fixed expression and head rotation of over 40 ° were synthesised using rendered models captured from the 3dMD system. We use 10 sequences comprising over 700 images in the experiment. We applied both the 2D single view and the 3D multi-view methods to the same set of face sequences using the FastNCC algorithm [Tidd 07] as the optimisation method. An illustration can be seen in *Figure* 7.

The statistical results of fitting are shown in *Figure* 8 and *Table* 1. We can see that the fitting with the multi-view CLM algorithm converges better. A Student t-test (p=0.05) has been carried out for the the errors d_e from both methods. The hypothesis is $\mu_0 < \mu_1$, where μ_0 is the error d_e with single-view CLM algorithm and μ_1 is with multi-view CLM algorithm. The sig. is $8.16e-5$. The fitting speed is also improved. Therefore, the fitting process is more stable with multi-view CLM algorithm.

The alignment accuracy is measured with the following equation:

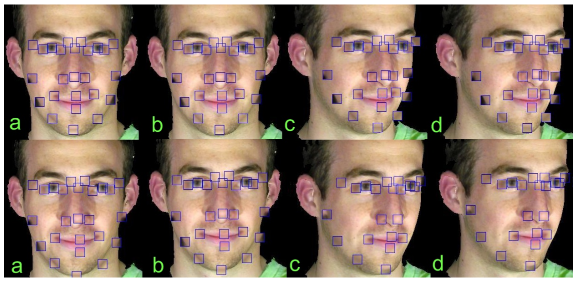

Figure 7: Each row consists of a set of selected frames from a tracking sequence with the synthetic texture patches drawn on which indicates the location of the features. The results in the first row is from the single-view approach and the second row is from the multi-view approach. When the rotating angle reaches certain degrees (b,c), the algorithm continues tracking the face well by auto-switching the appearance model to a side view model while the patches start getting off the position with single-view model.

Time (*ms*)	Multi-view CLM	Single-view CLM
mean	445.4	457.5
variance	46.3	65.1

Table 1: The speed comparison between the multi-view CLM algorithm and the Single-view CLM algorithm. The experiments are applied to synthetic images.

$$d_e = \frac{\sum \sqrt{(x_{std}-x_c)^2+(y_{std}-y_c)^2}}{Nd} \tag{5}$$

where x_{std}, y_{std} represent the manually placed "ground truth" feature points locations, x_c, y_c represent the tracked feature points locations, d represents the distance between the center of the eyes and N is the number of features.

To make the further investigation, we compare the average errors's frame by frame as we can see in *Figure* 9. In the experiments, the first frame is the frontal face image and the face rotates one degree per frame. For

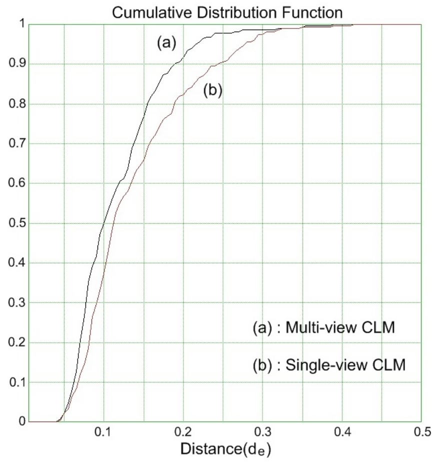

Figure 8: The error d_e comparison between the multi-view CLM algorithm and the Single-view CLM algorithm. The experiments are applied to synthetic images.

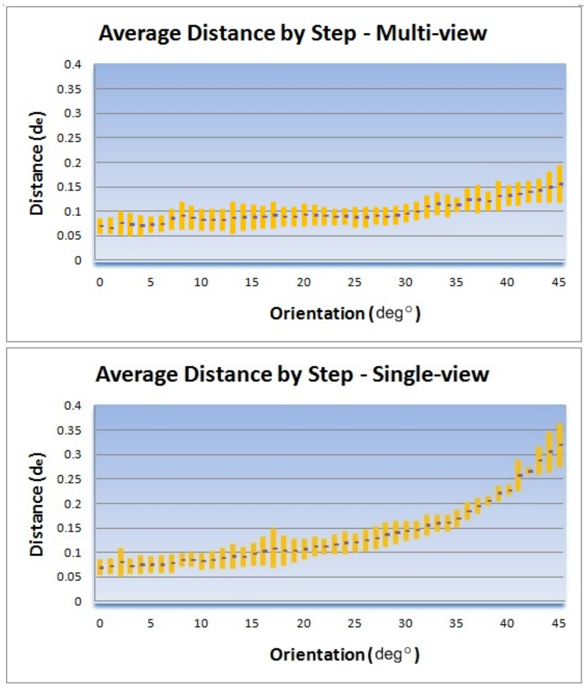

Figure 9: The figure contains the mean error d_e for each step (1 degree). The upper figure is for the Multi-view 3D algorithm and the lower is for the Single-view 2D algorithm.

the multi-view 3D CLM method the fitting errors are approximately constant across different views. For the Single-view CLM algorithm, we can see that the tracking is robust within the first 15 frames but the fitting becomes increasingly inaccurate as the rotation angle increases, as one would expect. The fitting is fairly stable up to about 20°–30°, which gives an indication of the optimal range of each appearance model.

A texture model from one view could possibly cover the rotation range around 20°–25°. However, it has to be cautious about the criterion area because the fitting

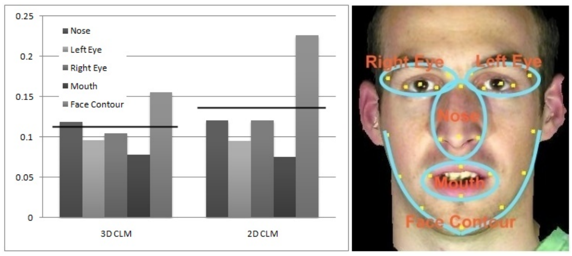

Figure 10: The average errors between the tracked feature points and the corrected feature points of the faces divided into different regions. The horizontal line across each group is the average error of all the face feature points.

could fail at that area referring to *Figure* 7. In order to do that, we choose a different texture model every 30° which makes each texture model cover 15° for each direction. This requires a fifteen texture model system to cover about 100° in the vertical direction and 160° in the horizontal direction.

We also investigate the performance of the algorithms for different facial regions. We split the face into jaw, mouth, nose and two eye regions and calculate the errors separately (*Figure* 10). The proposed algorithm improves mostly the performance on the face contour and the right eye area, which is somehow hidden during rotation. The performance in these areas benefits from the fixed visibility model applied to the multiple appearance models.

4.2 Real Data Experiment

Another experiment is carried out to investigate the accuracy and stability of the proposed algorithm on real video data. The video data consists of four different subjects showing expression, speech and some head rotation (1280 frames in total) (*Figure* 11) These images and subjects are independent of the CLM training sets.

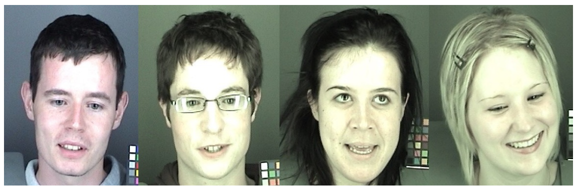

Figure 11: Example images from the text video clips fitting to.

The image sequences are roughly initialised with the face detector described in [Chen 08] before a CLM search is applied to the frame and the following frames while tracking. Three optimising methods are used for the experiments: Powell's method [Pres 07], FastNCC algorithm [Tidd 07] and the Gauss-Newton method [Hage 98, Matt 04], a maximum of 4 iterations are used per frame while tracking.

The results are shown in *Figure* 12 and *Figure* 13. For Powell's method and Gauss-Newton algorithm, nearly

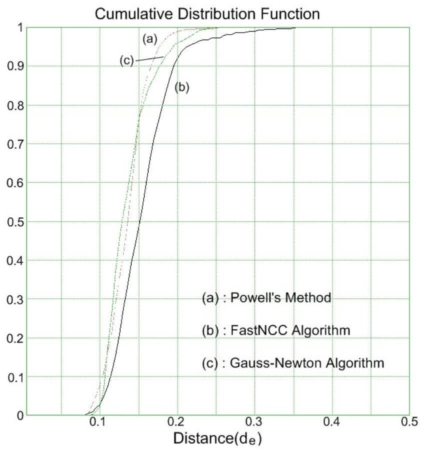

Figure 12: Results of fitting performance with three optimising algorithms to sets of real images.

Sig.	Powell's Method	FastNCC
Gauss-Newton	0.4658	2.8e-10

Table 2: Results from two samples Student T-Test (p=0.05) for Gauss-Newton algorithm against Powell's method and Gauss-Newton Algorithm against FastNCC algorithm. The hypothesis is that $\mu_0 < \mu_1$, where μ_0 is the error d_e for Gauss-Newton algoithm, μ_1 is the error d_e for the other two methods.

80 % of the points are within 0.15 d_e. For FastNCC algorithm, nearly 80 % of the points are within 0.17 d_e. The proposed multi-view CLM algorithm gives robust performance for all three optimising methods. Collaborating with the anova test shown in (*Table* 2), the Gauss-Newton algorithm and the Powell's method are more stable and accuracy.

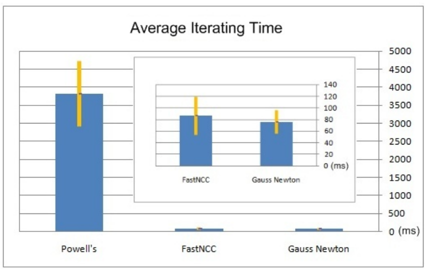

Figure 13: Average iteration time estimated on the fitting using three optimising algorithms on the set of real images.

The iteration time is measured roughly and could be further optimised. The results are shown in *Figure* 13. The Gauss-Newton algorithm is the fastest method. The FastNCC algorithm performs at the same level as those two methods. Powell's method takes much more time than the other two methods.

5 CONCLUSION AND FUTURE WORK

The presented multi-view 3D CLM algorithm is based on the original Constrained Local Model search [Cris 06, Cris 08]. There are 15 texture models built from different views of faces and a 3D shape model in the algorithm. For each view, the selected texture model and the shape model forms a combined model and applies a constrained local search on the given image. This algorithm can be used in locate and track human face features in sequences with large head rotations.

Based on the experiments carried out, we have shown that the proposed multi-view 3D algorithm gives better results when fitting to unseen images with large head rotations (the images are captured from 3dMD). The fitting is more robust (*Figure* 8, 9, 10) and efficient (*Table* 1) than the single-view CLM algorithm [Cris 06]. The effects of choice of the optimisation algorithm was marginal for the data tested here, but further experimentation with a wider range of video data would be needed to confirm this.

Future research will involve lighting and colour and more self occlusion factors, which could improve the matching rates under variant lighting, colour conditions and even with unexpected occlusions. To achieve that, more constrained searching should be applied while tracking and a more stable tracking failure recovery system should be introduced.

The fitting results (*Figure* 10) showed that the face contour part is not matched as well as internal facial features. We will investigate improved background modelling and contour alignment methods to improve facial contour alignment and tracking.

REFERENCES

[Ahlb 01] J. Ahlberg. "Using the active appearance algorithm for face and facial feature tracking". In: *International Conference on Computer Vision Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-time Systems*, pp. 68–72, 2001.

[Bake 01] S. Baker and I. Matthews. "Equivalence and efficiency of image alignment algorithms". In: *IEEE IEEE Transactions on Computer Vision and Pattern Recognition*, pp. 1090–1097, 2001.

[Bake 02] S. Baker and I. Matthews. "Lucas-Kanade 20 Years On: A Unifying Framework: Part 1". Tech. Rep. CMU-RI-TR-02-16, Robotics Institute, University of Carnegie Mellon, Pittsburgh, PA, July 2002.

[Blan 99] V. Blanz and T. Vetter. "A morphable model for the synthesis of 3D faces". In: *Computer graphics, annual conference series (SIG-GRAPH)*, pp. 187–194, 1999.

[Bran 01] M. Brand. "Morphable 3D models from video". In: *IEEE computer society conference on computer vision and pattern recognition*, pp. 456–463, 2001.

[Chen 08] J. Chen and B. P. Tiddeman. "Multi-cue Facial Feature Detection and Tracking". In: *International Conference on Image and Signal Processing*, pp. 356–367, 2008.

[Coot 00] T. F. Cootes, K. Walker, and C. Taylor. "View-Based Active Appearance Models". In: *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 227–232, IEEE Computer Society, Washington, DC, USA, 2000.

[Coot 01a] T. F. Cootes, G. J. Edwards, and C. J. Taylor. "Active appearance models". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 681–685, 2001.

[Coot 01b] T. F. Cootes and C. J. Taylor. "Statistical models of appearance for medical image analysis and computer vision". In: *SPIE Medical Imaging*, pp. 236–248, 2001.

[Coot 02] T. F. Cootes and P. Kittipanyangam. "Comparing Variations on the Active Appearance Model Algorithm". In: *British Machine Vision Conference*, pp. 837–846, 2002.

[Coot 95] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. "Active shape models - Their training and application". *Computer Vision and Image Understanding*, Vol. 61, pp. 38–59, 1995.

[Coot 98] T. F. Cootes, G. J. Edwards, and C. J. Taylor. "Active appearance models". *European Conference on Computer Vision*, Vol. 2, pp. 484–498, 1998.

[Cris 06] D. Cristinacce and T. Cootes. "Feature detection and tracking with Constrained Local Models". In: *British Machine Vision Conference*, pp. 929–938, 2006.

[Cris 08] D. Cristinacce and T. Cootes. "Automatic feature localisation with constrained local models". *Pattern Recognition*, Vol. 41, No. 10, pp. 3054–3067, 2008.

[Fagg 05] N. Faggian, S. Romdhani, J. Sherrah, and A. Paplinski. "Color Active Appearance Model Analysis Using a 3D Morphable Model". In: *Digital Image Computing on Techniques and Applications*, p. 59, IEEE Computer Society, Washington, DC, USA, 2005.

[Gros 05] R. Gross, I. Matthews, and S. Baker. "Generic vs. Person Specific Active Appearance Models". *Image and Vision Computing*, Vol. 23(11), pp. 1080–1093, November 2005.

[Hage 98] G. D. Hager and P. N. Belhumeur. "Efficient Region Tracking With Parametric Models of Geometry and Illumination". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, pp. 1025–1039, 1998.

[Hu 04] C. Hu, J. Xiao, I. Matthews, S. Baker, J. Cohn, and T. Kanade. "Fitting a Single Active Appearance Model Simultaneously to Multiple Images". In: *British Machine Vision Conference*, September 2004.

[Jone 98] M. J. Jones and T. Poggio. "Multidimensional Morphable Models: A Framework for Representing and Matching Object Classes". In: *International Journal of Computer Vision*, pp. 107–131, Springer Netherlands, 1998.

[Kote 05] S. Koterba, S. Baker, I. Matthews, C. Hu, J. Xiao, J. Cohn, and T. Kanade. "Multi-View AAM Fitting and Camera Calibration". In: *IEEE International Conference on Computer Vision*, pp. 511–518, IEEE Computer Society, Washington, DC, USA, 2005.

[Lani 97] A. Lanitis, C. J. Taylor, and T. F. Cootes. "Automatic interpretation and codeing of face images using flexible models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19(7), pp. 742–756, 1997.

[Luca 81] B. D. Lucas and T. Kanade. "An Iterative Image Registration Technique with an Application to Stereo Vision". In: *International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

[Matt 04] I. Matthews and S. Baker. "Active appearance models revisited". *International Journal of Computer Vision*, Vol. 60, pp. 135–164, 2004.

[Mitc 01] S. C. Mitchell, B. P. F. Lelieveldt, R. J. Geest, J. G. Bosch, J. H. C. Reiber, and M. Sonka. "Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac MR images". In: *IEEE Transactions on Medical Image*, pp. 415–423, 2001.

[Neld 65] J. A. Nelder and R. Mead. "A simplex method for function minimization". *Computer Journal*, Vol. 7, pp. 308–313, 1965.

[Paqu 09] U. Paquet. "Convexity and Bayesian Constrained Local Models". *IEEE Transactions on Computer Vision and Pattern Recognition*, pp. 1193–1199, 2009.

[Peyr 07] J. Peyras, A. Bartoli, H. Mercier, and P. Dalle. "Segmented AAMs Improve Person-Independent Face Fitting". In: *British Machine Vision Conference*, 2007.

[Peyr 08] J. Peyras, A. J. Bartoli, and S. K. Khoualed. "Pools of AAMs: Towards Automatically Fitting any Face Image". In: *British Machine Vision Conference*, p. , 2008.

[Pres 07] W. H. Press., S. A. Teukolsky., W. T. Vetterling, and B. P. Flannery. *Numerical recipes - The art of scientific computing*. Cambridge University Press, 2007.

[Ramn 08] K. Ramnath, S. Koterba, J. Xiao, C. B. Hu, I. Matthews, S. Baker, J. F. Cohn, and T. Kanade. "Multi-View AAM Fitting and Construction". In: , Ed., *International Journal of Computer Vision*, pp. 183–204, 2008.

[Romd 02] S. Romdhani, V. Blanz, and T. Vetter. "Face identification by fitting a 3D morphable model using linear shape and texture error functions". In: *European Conference on Computer Vision*, pp. 3–19, 2002.

[Shum 01] H. Y. Shum and R. Szeliski. *Panoramic vision: sensors, theory, and applications*, Chap. Construction of panoramic image mosaics with global and local alignment, pp. 227–268. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.

[Steg 01] M. B. Stegmann. "Object tracking using active appearance models". In: *Danish Conference Pattern Recognition and Image Analysis*, pp. 54–60, 2001.

[Tidd 07] B. P. Tiddeman and J. Chen. "Correlated Active Appearance Models". In: *IEEE Transactions on Signal-Image Technology & Internet-Based Systems*, pp. 832–838, 2007.

[Vett 97] T. Vetter and T. Poggio. "Linear object classes and image synthesis from a single example image". In: *Pattern Analysis and Machine Intelligence*, pp. 733–742, 1997.

[Wang 07] Y. Wang, S. Lucey, J. Cohn, and J. M. Saragih. "Non-rigid Face Tracking with Local Appearance Consistency Constraint". In: *IEEE International Conference on Automatic Face and Gesture Recognition*, September 2007.

[Wang 08] Y. Wang, S. Lucey, and J. F. Cohn. "Enforcing convexity for improved alignment with constrained local models". In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2008.

[Xiao 04] J. Xiao, S. Baker, I. Matthews, and T. Kanade. "Real-time combined 2D+3D active appearance models". In: *the IEEE computer society conference on computer vision and pattern recognition*, pp. 535–542, 2004.

[Zhan 04] Z. Zhang, Z. Liu, D. Adler, M. F. Cohen, E. Hanson, and Y. Shan. "Robust and Rapid Generation of Animated Faces from Video Images: A Model-Based Modeling Approach". *International Journal of Computer Vision*, Vol. 58, No. 2, pp. 93–119, 2004.

[Zhou 05] Y. Zhou, W. Zhang, X. Tang, and H. Shum. "A Bayesian Mixture Model for Multi-View Face Alignment". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 741–746, IEEE Computer Society, Washington, DC, USA, 2005.