Conference contribution :

Fang, H., Deng, J., Xie, X. & Grant, P. (2013). *From clamped local shape models to global shape model.* 2013 IEEE International Conference on Image Processing, (pp. 3513-3517). Melbourne, VIC, Australia:
http://dx.doi.org/10.1109/ICIP.2013.6738725

FROM CLAMPED LOCAL SHAPE MODELS TO GLOBAL SHAPE MODEL

Hui Fang, Jingjing Deng, Xianghua Xie and Phil W. Grant

Computer Science Department,
College of Science, Swansea University,
Swansea, Wales, U.K., SA2 8PP

ABSTRACT

Facial fiducial point localization is a crucial step for most facial analysis applications, e.g., face recognition, expression recognition and facial aging simulation. Although state-of-art methods have the ability to provide good salient point location on frontal faces, finding a global solution under large variations caused by off-plane rotations and exaggerated expression changes is still a challenge. In this paper, we present a system with a two-level shape model to facilitate accurate facial fiducial point localization. In the first level, two local component models interact with each other in order to offer novel shape constraints. At the same time, the clamped local shape models provide constrained non-linear shape initialization for better convergence performance of the shape model as a whole. The experimental results confirm that the proposed method is capable of dealing with the face alignment under large shape variations.

***Index Terms*—** Facial fiducial point localization, Active shape models, Discriminate texture models.

1. INTRODUCTION

Facial fiducial point localization, an important facial analysis step, has been investigated extensively since the active shape model was first proposed in [1]. Statistical shape models provide a suitable generic framework for the task because the size of the search space can be reduced significantly by removing the redundancy in the point sets. On the one hand, real-time implementations of the application become possible optimizing the lower dimensional shape model. On the other hand, the feature points can be located more robustly by adding shape constraints to avoid outliers.

Many methods, such as Active Appearance Models (AAM) [2, 3], Constrained Local Models (CLM) [4, 5], hierachical models [6, 7] and linked statistical shape model (LSSM) [8] have achieved reasonable performance when the environment is controlled. The Active Appearance Model [2] is a typical method searching for facial salient points by building combined shape and appearance models and predicting point locations by fitting the models to the input image. In recent years, local texture models have become more attractive due to their better performance when non-linear texture relationships exist between facial components. Constrained Local Models [4] and cascade of combined shape models (c-CSM) [7] are methods using Bayesian frameworks to integrate local texture models with global shape model constraints. LSSM [8] links shape variations across multiple modalities to facilitate concurrent segmenation of MRI and CT images. Another popular approach is to improve the performance by using discriminate models instead of generic models as discriminate models have been proved more powerful for evaluating the fitness of parameters on images [9, 10, 11]. For example, SVM classifiers are used in [5] to give more distinctive scores compared to generic models.

However, the convergence performance of these methods relies heavily on the point initialization because the optimization process can easily terminate in a local minimum in the global shape model. In order to achieve real-time implementation of active shape models, gradient descent based methods [2, 12, 5] and simplex optimization [4] are the most popular for finding optimal solutions. Even though the texture fitness function is well defined in the above models, the optimization schemes are easily trapped into local minima because non-linear variations caused by large off-plan rotations and exaggerating expression changes (shown in Figure 1) introduce many additional minima into the search space. Although sampled-based optimization methods, such as particle filtering, have been successfully employed in the shape model framework to avoid the local minima problem [13], it is still difficult to implement in a real-time system due to the increasing computational complexity of the cost calculation. Thereby, facial fiducial point localization under large shape variations is still a challenging problem.

Non-linear correlation between different face components is the main reason for the problem since movements of the regions are controlled by different groups of facial muscles. Inspired by this observation, we present a system integrating local component shape models with a global shape model. Due to the independent movements of mouth region and eye region, the local component shape models decouple the linear correlations between upper and lower face regions as well as provide spatial shape constraints based on the overlapping fiducial facial points covered by both models (which we call

Fig. 1. Examples of face images with large variations

clamped local shape models). Furthermore, the initialization using these models enables the search in the global shape model space to escape the local minima and thus achieve a better convergence. Providing better convergence and ensuring computational efficiency is the main objective of the proposed system. The two-level configuration of the proposed framework is capable of handling large pose and expression variations with better convergence.

This paper is organized as follows. Section 2 provides details of the proposed method. The experimental results are presented in Section 3 demonstrating the performance of the system and future work and conclusions are drawn in Section 4.

2. METHODOLOGY

2.1. Two-level shape models

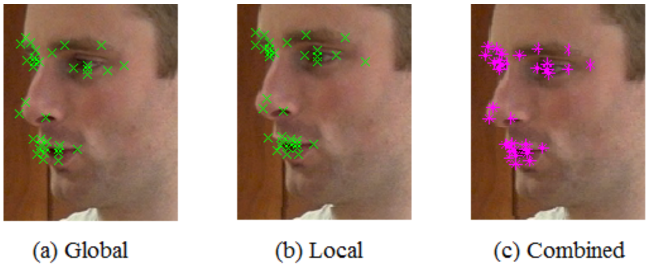
The Point Distribution Model (PDM) [1] has been widely used to model the shape of deformable objects, such as in [4, 10, 12]. Given a set of N points $\vec{x}_i = (x_i, y_i)$, they are represented as one $2N$ dimensional vector $\vec{S} = (\vec{x}_1, ..., \vec{x}_i, ..., \vec{x}_N)$ which in turn can be represented as the linear combination:

$$\vec{S} = s_0 + \sum_{j=1}^{n} p_j v_j \tag{1}$$

where s_0 is the average shape, v_j are the eigenvectors of the n largest eigenvalues derived from PCA and p_j represent the parameters in the subspace. PDM reduces the optimization dimension significantly by considering correlations between the points. Although the subspace has the ability to represent a large number of deformable shape variations, the linear constraints in the lower dimensional space make it possible for the search to converge to local minimum.

In this paper, a two-level shape model consisting of clamped local shape models and a global shape model is proposed to overcome the optimization problem. An intuitive solution to the problem is to represent the point set by using two shape models with overlapping members. This configuration breaks the link between the points which have weak correlations while it still has the strong spatial shape constraints by keeping the overlapping points as close as possible. This is illustrated in Figure 2 (a), the search using the global shape model converges to the local minimum and both the feature points on the eye and mouth regions fail to converge to a global optimal solution.

However, the search stages for the clamped local shape models, where two PDM models are trained based on feature points from upper and lower faces, provide a better solution which is shown in Figure 2 (b). Although the convergence on both models may not be perfect, due to the lack of global shape constraints, it provides an initial parameter vector which is much closer to the global optimal solution. With the parameter vector set as the initial position for the search in the second level global shape model space, the method converges to the global optimal solution shown in Figure 2 (c).



(a) Global

(b) Local

(c) Combined

Fig. 2. Search results for the global shape model, clamped local shape models and two-level shape models

2.2. Optimization function design

The optimization function for fiducial point localization is composed of a texture cost function and a shape cost function. The texture cost function makes sure that the patches extracted from the feature points match well with the standard face component templates, while the shape cost function gaurantees that the algorithm converges avoiding illegal shape varation.

2.2.1. *Texture cost function*

Local texture representation, based on Haar-like rectangular features [14, 10], is extracted in a 24×24 rectangular patch around each fiducial facial point. After the features are extracted, the GentleBoost algorithm [15] is used to learn a discrimination function for providing a response score. The patches around the manually labelled fiducial facial points in the training dataset are set as the positive training samples for each boosting classifier. At the same time, a number of negative samples are collected by random perturbation for the classifier where the Euclidean distances between the ground truth positions and the perturbed positions are greater than 5 pixels. Examples of positive patches and negative patches sampled around the ground truth location of the left eye center and perturbed positions are shown in Figure 3. Each classifier provides a confidence value for how well each feature point (in the point set defined by the shape model) is coded by the face shape parameter vector.

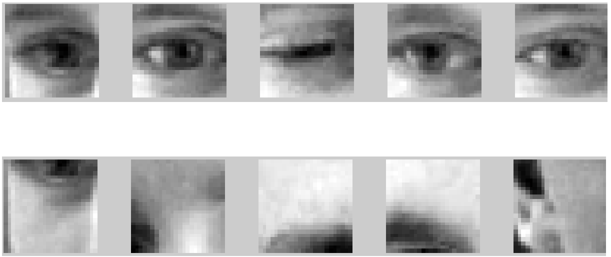


Fig. 3. Positive patches (upper row) and negative patches (lower row) sampled around the ground truth location of the left eye center

For the evaluation stage, given the shape parameters p, the texture fitness function is:

$$TF(p) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} f_m(I(x_i(p))) \tag{2}$$

where I represents the patch sampled from the i^{th} point x_i given parameters p from either a local shape model or a holistic model. f_m represents the response value obtained from the m^{th} weak classifier trained from the boosting algorithm.

2.2.2. *Shape cost function*

The shape cost function provides shape constraints on the optimization in order to avoid illegal shape variations and outliers. We follow the work in [4] to define the shape constraints in the global shape model as follows:

$$SF(p) = -\frac{1}{k} \sum_{i=1}^{k} \frac{p_i^2}{\lambda_i} \tag{3}$$

where k is the number of dimensions in the shape model space, p_i represent the i^{th} parameter and λ_i represent the i^{th} eigenvalue.

For the clamped local shape models, we introduce a new shape constraint term into the cost function. This term keeps the overlapping points in the local component shape models close to each other. As a result, the cost function can be defined as:

$$\begin{aligned} SF_{total}(p_{upper}, p_{lower}) = SF(p_{upper}) + SF(p_{lower}) \\ -\alpha D(p_{upper}, p_{lower}) \end{aligned} \tag{4}$$

where p_{upper} and p_{lower} represent the parameters from the local shape models, $D(p_{upper}, p_{lower})$ is the Euclidean distance between the generated overlapping points based on the two model parameters and α is a normalizing scale.

2.3. Optimization algorithm

The simplex algorithm [4] was used to optimize the fitness function which combined both the texture costs and shape costs balanced by another normalizing factor β, which is shown in Equation. 5.

$$CF(p) = TF(p) + \beta SF(p) \tag{5}$$

A K-simplex is a convex hull of its k+1 vertices and the algorithm drives the optimizing step away from the worst vertex in the space. The efficiency of this optimization scheme gaurantees that the system can be delivered in a real-time application.

3. EXPERIMENTAL RESULTS

The proposed two-level algorithm was applied to a dataset collected at Swansea University for evaluating the performance. There are 258 images under varying poses and expressions (examples from the dataset are shown in Figure 1). The reason for using this dataset is that the face instances demonstrate a high degree of non-rigid deformation which makes the localization task more difficult than usual. A set of 35 feature points, shown in Figure 4, were manually labelled to provide a training set and ground truth for quantitative evaluation.

We used CLM as a benchmark to evaluate the performance as it outperformed many state-of-the-art methods such as AAM and Pictorial Structure Matching algorithm [16]. The cumulative error distribution criterion [4], which refers to an accuracy curve by changing different tolerance point-to-point error levels, was used to compare the performance between the proposed two-level shape models and CLM[4]. The distance metric of point-to-point errors between the localization algorithm and the ground truth was defined as follow:

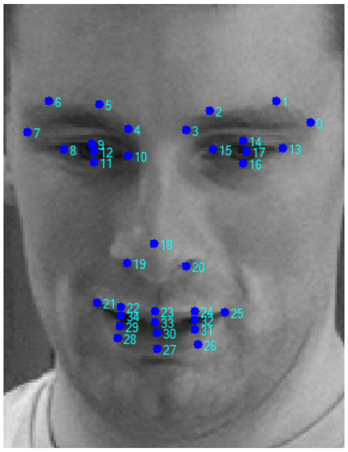


Fig. 4. Labelled ground truth 35 feature points

$$D_{avg} = \frac{1}{ns}\sum_{i=1}^{n} d_i \tag{6}$$

where d_i is the absolute Euclidean point to point error, n is the number of feature points and s is a scale factor being the ground truth distance between the left and right eye pupils. At each tolerance error level, a higher percentage of convergence means better performance. As shown in Figure 5, the proposed method provides more accurate facial fiducial point localization compared to the CLM algorithm. Given 6% displacement tolerance, the proposed method converged on 90% of images compared to 80% of images for CLM.

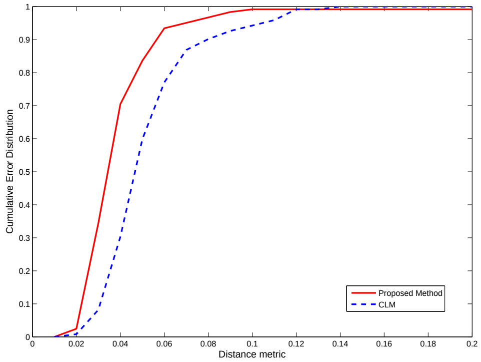


Fig. 5. Cumulative distribution of point to point error comparison between proposed method and CLM

Figure. 6 shows some visualized results of the facial fiducial points labelled by the proposed algorithm. The images demonstrate that the algorithm converges well even under large off-plane pose rotation, exaggerated expressions or wearing spectacles.

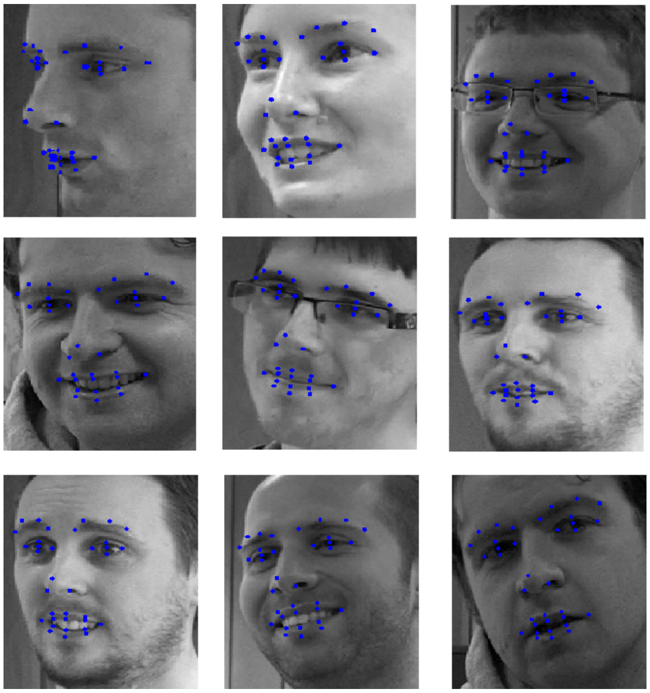


Fig. 6. Qualitative evaluations of the localization method

4. CONCLUSIONS AND FUTURE WORK

In this paper, a novel algorithm is presented to model face shape deformations in order to locate fiducial facial points automatically in unlabelled face images. A sequential system with two-level shape models which combine the advantages of local component shape models and a global model is implemented for the task of efficient points localization as well as avoiding convergence to local minima. The local component shape models provide flexibility in searching with a new spatial shape constraint. With the initalization from clamped local shape models, the system achieved a better convergence on a global shape model.

In future work, this algorithm will be used to track facial deformations in human facial interaction sequences in order to understand high-level semantic social conversational behaviors. It is believed that the better localization performance is able to improve the analysis of human social interactions and provides a sound base for the next generation human-computer-interaction techniques.

5. REFERENCES

[1] T. Cootes and C. Taylor, "Active shape models - 'smart snakes'," in *The proceedings of British Machine Vision Conference*, 1992, pp. 266–275.

[2] T. Cootes, G. Edward, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis*

and Machine Intelligence, vol. 23, no. 6, pp. 681–685, 2001.

[3] I. Matthews and S. Baker, “Active appearance models revisited,” *International Journal of Computer Vision*, vol. 26, no. 10, pp. 135–164, 2004.

[4] D. Cristinacce and T. Cootes, “Automatic feature localisation with constrained local models,” *Pattern Recognition*, vol. 41, pp. 3054–3067, 2008.

[5] S. Lucey, Y. Wang, M. Cox, S. Sridharan, and J. Cohn, “Efficient constrained local model fitting for non-rigid face alignment,” *Image and vision computing*, vol. 27, pp. 1804–1813, 2009.

[6] C. Liu, H. Shum, and C. Zhang, “Hierarchical shape modeling for automatic face localization,” in *The proceedings of European Conference on Computer Vision*, 2002, pp. 687 – 703.

[7] P. Tresadern, H. Bhaskar, S. Adeshina, C. Taylor, and T. Cootes, “Combining local and global shape models for deformable object matching,” in *The proceedings of BMVC*, 2009, pp. 1 – 12.

[8] N. Chowdhury, R. Toth, J. Chappelow, S. Kim, S. Motwani, S. Punekar, H. Lin, S. Both, N. Vapiwala, and S. Hahn, “Concurrent segmentation of the prostate on mri and ct via linked statistical shape models for radiotherapy planning,” *Medical Physics*, vol. 39, pp. 2214–2228, 2012.

[9] L. Zhang, H. Ai, S. Xin, C. Huang, S. Tsukiji, and S. Lao, “Robust face alignment based on local texture classifiers,” in *Proceedings of International Conferece of Image Processing*, 2005, vol. II, pp. 354–357.

[10] X. Liu, “Discriminative face alignment,” *IEEE Transactions on Pattern Analysis and Machine Intellignence*, vol. 31, pp. 1941–1954, 2009.

[11] D. Cristinacce and T. Cootes, “Boosted regression active shape models,” in *The proceedings of BMVC*, 2007, pp. 880 – 889.

[12] P. Tresadern, M. Ionita, and T. Cootes, “Real-time facial feature tracking on a mobile device,” *International Journal of Computer Vision*, vol. 96, pp. 280–289, 2012.

[13] W. Qu, X. Huang, and Y. Jia, “Segmentation in noisy medical images using pca model based particle ltering,” in *In SPIE Conf. on Medical Imaging*, 2008, pp. 687 – 703.

[14] P. Viola and M. Jones, “Robust real-time face detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[15] J. Friedman, T. Hastie, and R. Tibshirani, “Addictive logistic regression: a statistical view of boosting,” *The annals of statistics*, vol. 28, pp. 337–407, 2000.

[16] P. Felzenszwalb and D. Huttenlocher, “Pictorial structures for object recognition,” *International Journal of Computer Vision*, vol. 61, pp. 55–79, 2005.