

Automated Face Pose Estimation Using Elastic Energy Models

Sanqiang Zhao and Yongsheng Gao

School of Engineering, Griffith University, QLD, Australia 4111

{s.zhao, yongsheng.gao}@griffith.edu.au

Abstract

Face pose estimation forms an important part in a face recognition system. However, fully automated and accurate pose determination still remains an unsolved problem in the research community. In this paper, we propose a novel Elastic Energy Model to automatically estimate face poses. Our method employs statistical energy contributions of a set of feature points, which can avoid over-trusting selected anchor points. It provides a robust solution to the feature localisation inaccuracy problem, which is inevitable in practical applications with cluttered backgrounds. As a general configuration, our model can be easily implemented and extended to other non-rigid objects. Its effectiveness and robustness are revealed in our experiments.

1. Introduction

3D pose variation has long been one of the bottlenecks underlying the field of computerised face recognition [1]. Conventionally, methods for face pose estimation are based on either feature detection [2], [3] or intensity distribution [4], [5] of face images. Gee and Cipolla [2] chose a few relatively stable feature points, or anchor points, to estimate the gaze direction under weak perspective, with an assumption that the ratios of these points did not change significantly for different facial expressions. Their key points, however, were manually labelled. Gao et al. [3] presented an efficient pose recovery approach using locations of the two eyes and the symmetric axis of the face, but their method was very sensitive to the location error of the facial features.

A statistical 3D morphable model [4] was proposed to deal with pose variations as well as illuminations, which requires textured 3D scans of heads and involves expensive computations. In View-based Active Appearance Models [5], the relationship between the pose angle and the shape parameters of three distinct models was assumed to determine the pose. This assumption, however, is not quite convincing, for the shape parameters in their approach are representations

with both *interface* and *intraface* shape variations (Fig. 1).

To separate these two different variations, Dias and Buxton [6] proposed an integrated model by combining Cootes et al.'s Active Shape Models [10] with Ullman and Basri's Linear Combination of Views (LCV) [7]. This approach could represent the interface and intraface shape variations independently, but the LCV technique required accurately locating at least three corresponding feature points in each of the 2D images of different views.

Research in this paper differs from most existing methods. In 2D image context, Elastic Energy Models (EEMs) are proposed to represent the pose-induced intraface shape variations. We argue that the pose angle could be estimated straightforwardly from the information of 2D shape variations through building at least two EEMs. In stead of over-trusting some selected anchor points, our model relies on statistical energy contributions of a group of elastic landmarks to handle the problem of feature localisation error, which is inevitable in practical applications with cluttered backgrounds. The effectiveness of our method is demonstrated by its full automation and robustness.

The paper is organised as follows: the next section introduces the theory of EEMs in detail, and Section 3 describes the implementation and experiments, with brief conclusions in Section 4.

2. Elastic Energy Models (EEMs)

2.1 Motivation

According to Hooke's Law, the restoring force of a spring is proportional to the negative displacement from the equilibrium position of the spring:

$$\mathbf{F} = -k\mathbf{x}, \quad (1)$$

where \mathbf{F} is the force that tends to restore a spring to its equilibrium position, \mathbf{x} is the displacement of the spring from its equilibrium position, and k is the spring constant. In physics, this law describes the elastic property of a spring. Also, the corresponding potential energy can be stated as

$$E = \frac{1}{2} k \| \mathbf{x} \|^2. \quad (2)$$

If we consider each feature point constituting one face shape as a special “elastic” landmark, other shapes of different poses could be decomposed as a group of elastic landmarks’ displacements from equilibrium positions (red arrows in Fig. 1). This motivates us to develop an energy associated elastic shape model in 2D image context, namely *Elastic Energy Model*, to describe the pose variations. Although physics-based models have been employed to analyse facial images in the early work [8], they were mainly developed to model 3D human facial tissues.

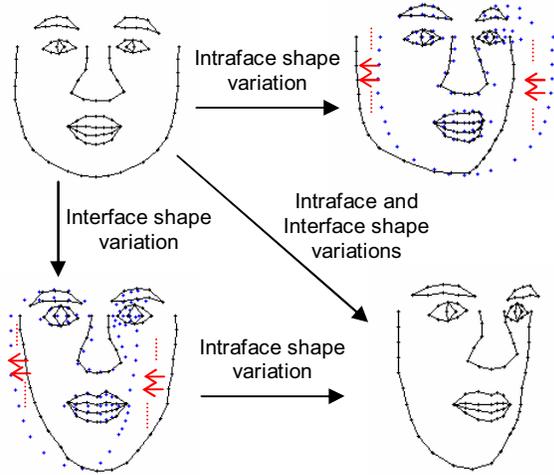


Fig. 1. Interface and Intraface shape variations. Red arrows denote intraface shape deformation.

Let $\mathbf{x}_{i\theta} = (x_{i\theta}, y_i)$ denote the 2D position of the i -th landmark corresponding to a yaw angle θ (only yaw rotation is considered here for simplicity). When this landmark “moves” from equilibrium position $\mathbf{x}_{i\theta_0} = (x_{i\theta_0}, y_i)$ to a different position $\mathbf{x}_{i\theta_1} = (x_{i\theta_1}, y_i)$ due to the yaw rotation from angle θ_0 to θ_1 , this movement is similar to a spring stretch/compression (corresponding to left/right rotation) exerted by an external force

$$\mathbf{F}_{i(\theta_0 \rightarrow \theta_1)} = -k_i (\mathbf{x}_{i\theta_1} - \mathbf{x}_{i\theta_0}), \quad (3)$$

where k_i is the “spring constant” of the i -th landmark, and we name it as *elastic constant*. Similarly, the potential energy associated to this external force is

$$E_{i(\theta_0 \rightarrow \theta_1)} = \frac{1}{2} k_i \| \mathbf{x}_{i\theta_1} - \mathbf{x}_{i\theta_0} \|^2 = \frac{1}{2} k_i (x_{i\theta_1} - x_{i\theta_0})^2. \quad (4)$$

Considering the entire face shape, all landmarks’ potential energies are added together to represent a global deformation of the model caused by the in-

traface variation. However, the simple summation is appropriate only when both two face shapes of angle θ_0 and θ_1 refer to the same person. Otherwise, the calculated summation will include confounded energies from both interface and intraface shape deformations (Fig. 1).

Based on the symmetric property of the human face [9], in this research, all the landmarks constituting one face shape are first divided into two symmetric groups (Fig. 2a). The potential energy caused by the interface shape deformation could then be eliminated through considering only every pair-wise energy disparity (Fig. 2b). With the yaw rotation only, the global potential energy resulted from intraface shape deformation E^I is the summation of all pair-wise energy disparities of a set of landmarks’ horizontal displacements:

$$\begin{aligned} E^I_{(\theta_0 \rightarrow \theta_1)} &= \sum_{i=1}^{N/2} E^I_{i(\theta_0 \rightarrow \theta_1)} \\ &= \frac{1}{2} \sum_{i=1}^{N/2} (k_i (x_{i\theta_1} - x_{i\theta_0})^2 - k_j (x_{j\theta_1} - x_{j\theta_0})^2) \end{aligned}, \quad (5)$$

where N is the total number of landmarks constituting the face shape, and the i -th landmark symmetrically corresponds to the j -th landmark. Since the elastic constant k_i is a constant, the potential energy is therefore a quadratic function of 2D shape displacements.

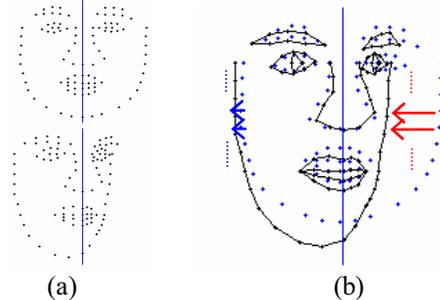


Fig. 2. Energy separation between two types of variations. (a) Landmark grouping (the blue line divides landmarks into two symmetric groups). (b) Pair-wise energy disparity (displacements of red arrows symmetrically correspond to those of blue arrows).

2.2 Multi-EEM Based Pose Estimation

We develop a multi-EEM based method to estimate the yaw rotation of a human face. The reason of using multiple models is two folded. First, the unknown elastic constants cause that E^I could not be directly calculated using only one EEM. Second, it is difficult to give E^I an explicit function of the pose angle θ . Ideally we may interpolate this function using two or more EEMs. In this section, we deal exclusively with two models of yaw pose $\theta_1=0$ (frontal view) and $\theta_2=\theta^*$

(left/right yaw angle θ^*), though it is easy to extend to more views if necessary. Specifically, the equilibrium positions of these two models correspond to the frontal status and left/right yaw angle θ^* respectively. We assume that the energy functions associated with these two model views can be described as quadratic functions of yaw angle:

$$\begin{cases} E'_{(0 \rightarrow \theta)}(\theta) = \omega_1 \theta^2 \\ E'_{(\theta^* \rightarrow \theta)}(\theta) = \omega_2 (\theta - \theta^*)^2 \end{cases} \quad (6)$$

Let e_1 and e_2 denote the two potential energies spent on deforming the two EEMs to the target shape, and (6) can be further described as

$$\begin{cases} e_1 = \omega_1 \theta^2 \\ e_2 = \omega_2 (\theta - \theta^*)^2 \end{cases} \quad (7)$$

Note that the two models come from the same person, i.e. $\omega_1 = \omega_2$, the yaw pose can then be calculated directly (suppose $e_1, e_2 > 0$):

$$\theta = \begin{cases} \frac{\theta^*}{1 + \sqrt{e_2/e_1}}, & \theta \leq \theta^* \\ \frac{\theta^*}{1 - \sqrt{e_2/e_1}}, & \theta \geq \theta^* \end{cases} \quad (8)$$

Although the direct calculation of any of above potential energy is not available, we can find ways to estimate their ratios, e.g. it makes sense to assume that each elastic landmark contributes evenly to the overall energy.

3. Implementation and Experiments

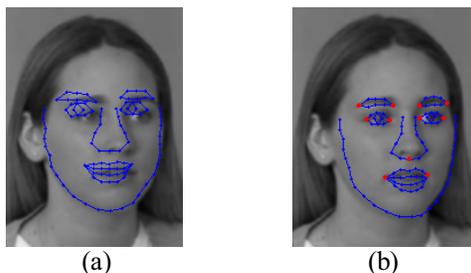


Fig. 3. Local landmarks refinement. (a) ASM can be easily stuck in local minima. (b) ASM combined with Gabor jet based searching algorithm (red points refer to control points).

To implement the Multi-EEM based pose estimation, we apply a variant of Active Shape Model (ASM) [10], considering its remarkable efficiency for shape-detecting task. Although ASM has been widely applied in various research fields, most approaches only care about its final matching results, ignoring the preceding searching steps that contains useful information. In our implementation, the searching and adjusting proce-

dures of ASM are signified as a deformation process of an EEM, and during the matching process, the potential energy corresponding to the EEM is accumulated at the end of each searching iteration.

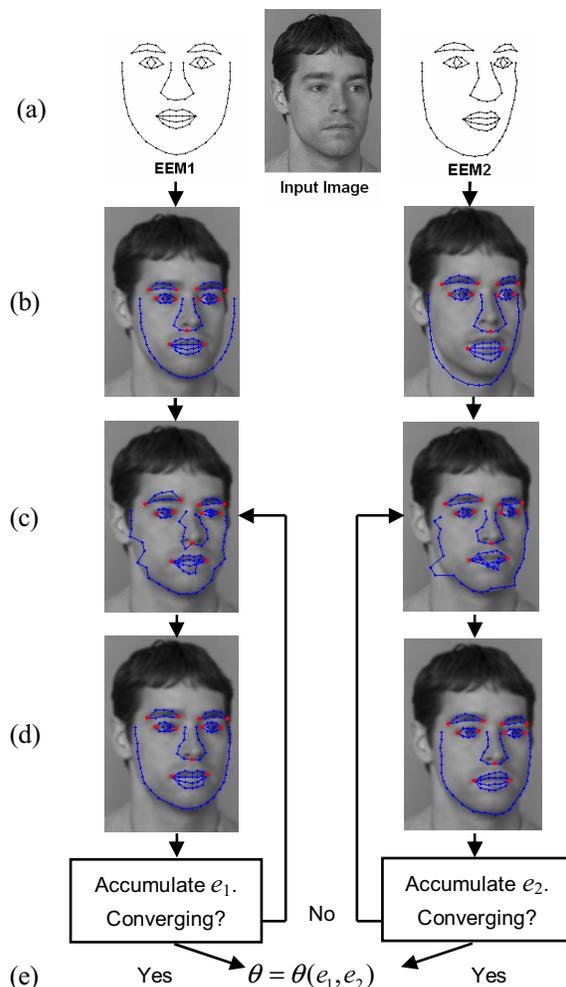


Fig. 4. Multi-EEM based pose estimation. (a) Input image (smoothed before matching) and two model shapes. (b) Initialisation. (c) Searching. (d) Adjusting. (e) Pose estimation.

Using Point Distribution Models (PDMs) [10], original ASM deforms the mean shape model by searching each landmark in the direction perpendicular to the contour within a certain range. It depends heavily on the initialisation and may easily be stuck in local minima (Fig. 3a). To alleviate the alignment error problem, in our application, the landmarks are separated into several groups according to the distribution of salient features in a face region. In each group, two control points are chosen, and a Gabor jet based searching algorithm [11] is utilised to adjust them. When the new positions of the control points are iden-

tified, an affine transformation is applied to adjust other points in every group. This process is combined into the ASM searching iteration and the alignment accuracy can be greatly improved (Fig. 3b).

From the FERET database [12], we select three subgroups (*ba*, *bc* and *bd*) that contain three poses (0° , $+40^\circ$ and $+25^\circ$) per individual, with a total of 200 people and 600 images. For each of the *ba* and *bc* subgroups, we manually label 200 images as the training data to build two PDMs, which are then used to do the shape alignment on subgroup *bd*. The number of the total landmarks in one face shape is 96, among them are 11 control points (Fig. 3b). For every control point, the Gabor feature bunch size is 70, through randomly selecting 35 from each training subgroup.

Table 1. Algorithm procedures.

(1)	Based on a face detection result, initialise two starting shapes by the two model shapes.
(2)	For the non-control points, use the standard normal-based searching method to find the required movements. For the control points, use Gabor jet based algorithm to search.
(3)	Affine transformation is used to adjust other points in the same group and two new shapes will be obtained. Calculate additional pose and shape parameter changes required to move the model shape as close as possible to the two new shapes respectively.
(4)	Update the shape and pose parameters by the above changes, and act on the model shape. The corresponding potential energy relating to this iteration step is calculated and accumulated to the overall potential energy.
(5)	Go to (2) for another iteration, until no considerable change is observed. In this case, two overall potential energy e_1 and e_2 are finalised to estimate the pose.

Table 2. Experimental results.

Pose Estimates	EEM (200 in FERET)	Reports in [4] (68 in CMU-PIE)	
Average Estimate	24.53°	18.1°	45.2°
Standard Deviation	4.09°	2.4°	4.5°
True Angle	25°	16.5°	45.6°

Considering two irises can be located automatically and quickly in most existing face detection systems [9], we manually provide their positions in our experiment. Table 1 describes the detailed procedures of the proposed algorithm, and Fig. 4 gives an example illustrating the estimation process. Tested on 200 images in subgroup *bd*, our system achieves an average estimated pose of 24.53°, within the error of 0.5° from the real angle (Table 2). Even with a greater number of subjects, the performance of our method is still comparable to the results in [4]. Moreover, EEM is more computationally efficient and also fully automated, which is different from [4].

4. Conclusions

This paper demonstrated a new Elastic Energy Model to address the issue of automatically estimating the human face poses. From the perspective of 2D image context, we obtained the stable and reliable result of the face pose successfully by accumulating global landmarks' energy contributions. The experimental results were promising and showed the effectiveness of the proposed model. With this method, one can automatically get a robust pose determination of human face, and even other non-rigid objects. We anticipate the approach will be applied on HCI-related applications in the future.

Acknowledgements

This research was supported by the Australian Research Council (ARC) Discovery Grant DP0451091.

References

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, 35(4): 399-459, 2003.
- [2] A. Gee and R. Cipolla, "Estimating Gaze from a Single View of a Face", *Proc. ICPR*, pp. 758-760, 1994.
- [3] Y. Gao, M. K. H. Leung, W. Wang, and S. C. Hui, "Fast face identification under varying pose from a single 2-D model view," *IEE Proc.-VISIP*, 148(4): 248-253, 2001.
- [4] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE PAMI*, 25(9): 1063-1074, 2003.
- [5] T.F. Cootes, G.V. Wheeler, K.N. Walker, and C.J. Taylor, "View-based Active Appearance Models", *Image and Vision Computing*, 20(9-10): 657-664, 2002.
- [6] M.B. Dias and B.F. Buxton, "Separating Shape and Pose Variations", *IVC*, 22(10): 851-861, 2004.
- [7] S. Ullman and R. Basri, "Recognition by Linear Combinations of Models", *IEEE PAMI*, 13: 992-1006, 1991.
- [8] D. Terzopoulos and K. Waters, "Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models," *IEEE PAMI*, 15: 569-579, 1993.
- [9] D. Reisfeld and Y. Yeshurun, "Preprocessing of face images: detection of features and pose normalization", *CVIU*, 71(3): 413-430, 1998.
- [10] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models - Their Training and Application", *CVIU*, 61(1): 38-59, 1995.
- [11] S. Zhao, W. Gao, S. Shan, and B. Yin, "Enhance the Alignment Accuracy of Active Shape Models Using Elastic Graph Matching", *Proc. ICBA*, pp. 52-58, 2004.
- [12] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms", *IEEE PAMI*, 22(10): 1090-1104, 2000.