ELSEVIER

Image and Vision Computing 20 (2002) 77–84

image AND vision COMPUTING

www.elsevier.com/locate/imavis

AI-EigenSnake: an affine-invariant deformable contour model for object matching

Zhong Xue[a], Stan Z. Li[b], Eam Khwang Teoh[a,*]

[a]*School of Electrical and Electronic Engineering, Nanyang Technological University, Vice-Dean's Office, Block S2, Singapore, Singapore 639798*
[b]*Microsoft Research China, 3/F Beijing Sigma Center, No. 49 Zhichun Road, Beijing 100080, People's Republic of China*

Received 11 April 2000; received in revised form 3 March 2001; accepted 16 July 2001

Abstract

An affine-invariant (AI) deformable contour model for object matching, called AI-EigenSnake (AI-ES), is proposed in the Bayesian framework. In AI-ES, the prior distribution of object shapes is estimated from the sample data. This distribution is then used to constrain the prototype contour, which is dynamically adjustable in the matching process. In this way, large shape deformations due to the variations of samples can be tolerated. Moreover, an AI internal energy term is introduced to describe the shape deformations between the prototype contour in the shape domain and the deformable contour in the image domain. Experiments based on real object matching demonstrate that the proposed model is more robust and insensitive to the positions, viewpoints, and large deformations of object shapes, as compared to the Active Shape Model and the AI-Snake Model.

Keywords: Active shape model; Affine invariant; Deformable model; EigenSnake; Object matching

1. Introduction

Deformable models have been studied intensively during the last 10 years [1,3–6]. They have been demonstrated to be more effective in object matching and able to adapt themselves to fit objects more closely than the traditional rigid models. Generally, deformable models can be classified into two classes [6,17,24]: the free-form models and the parametric models. The free-form models, e.g. active contours or snakes, can be used to match an arbitrary shape that provided some general regularization constraints, such as continuity and smoothness, are satisfied [25,27]. Although the free-form models also act as interactive tools for object extraction and image segmentation, the parametric models are more constrained because some prior information of the geometrical shape is incorporated, and hence are more robust to irrelevant structures and occlusions when used to detect specific shapes of interest.

The parametric models, such as deformable templates [6–8,18,19,21,26], G-Snake [9] and Active Shape Model (ASM) [5,15,20], encode specific characteristics of a shape and its variations using a global shape model, which is formed by a set of feature parameters or well defined landmark/boundary points of that shape. A quite successful and versatile scheme in this class is that of Bayesian statistics-based shape models [6,9,16,23]. In these models, both prior knowledge and observation statistics are used to define the optimal Bayesian estimate. However, most existing parametric models in the Bayesian framework (e.g. G-Snake) encode the shape information in a 'hard' manner in that the prototype contour is fixed during the matching process. As a result, only a small amount of local deformation can be tolerated. To remedy this short-coming, a deformable model with the name of 'EigenSnake' is proposed in the Beyasian framework [10], where the prototype contour can be adaptively adjusted in the process of object matching. Comparative studies based on face extraction experiments verified that the EigenSnake performs better than its fixed counterpart.

In 3D object matching using 2D images, the object shapes are subject to projections such as affine transformations, therefore the algorithms developed should be able to deal with shearing of object shapes, as well as rotation, translation, and scaling. Unfortunately, most deformable models (including the G-Snake and EigenSnake models) are not affine invariant, and the performance deteriorates when the algorithms are applied to matching affine transformed object shapes. Therefore, different affine-invariant (AI) snake models are developed [2,11]. Recently Ip and Shen developed an AI snake model (AI-Snake) [11], in which an AI internal energy term of the deformable contour is

* Corresponding author. Tel.: +65-790-5393; fax: +65-792-0415.
E-mail address: eekteoh@ntu.edu.sg (E.K. Teoh).

0262-8856/02/$ - see front matter
PII: S0262-8856(01)00078-6

proposed to match the objects. Although AI-Snake has a virtue of being affine invariant, it is sensitive to global shape deformations, because its prototype contour remains unchanged in the matching process as that of the G-Snake model.

In this paper, an AI deformable contour model, called AI-EigenSnake (AI-ES), is proposed in the Bayesian framework. The AI-ES has the advantages of both the EigenSnake model and AI-Snake model: The deformable contour used to match an object is modeled as the affine-transformed and deformed version of the prototype contour, and this prototype contour is dynamically deformable to adapt the shape variations using the information gathered from the matching process. The prior distribution of the prototype contour obtained via PCA [5,12,22] which reflects the major shape deformations, is used to constrain the admissible shapes. In this way, large shape deformation due to the shape variations of samples can be tolerated. Moreover, to make the AI-ES model more robust to rotation, translation, scaling and shearing of object shapes, an AI internal energy term is defined and incorporated into the model to describe the shape deformations between the prototype contour in the shape domain and the deformable contour in the image domain.

Experiments comparing the AI-ES model with the well known ASM model and the AI-Snake model are presented in the context of matching hand objects. The aim is to evaluate how prior constraints of the prototype contour, object shape deformations and different viewpoints have affected the matching results. The results demonstrate that in terms of accuracy the AI-ES outperforms the other two models in matching object when there exists large shape deformations, and is less sensitive to the variations in viewpoints and positions of the object. The second set of experiment deals with face detection under in-plane rotation; the effectiveness of the AI-ES is illustrated by matching the frontal faces with different in-plane rotation angle.

The rest of the paper is organized as follows. Section 2 describes the AI-ES model in detail. In Section 3, a comparative study among the ASM, the AI-Snake and the proposed AI-ES is presented by matching a series of hand objects. The AI-ES model is also applied to extract the rotated frontal face in this section. Section 4 derives the conclusion from this study.

2. Affine-invariant deformable contour model: AI-ES

2.1. Bayesian framework and energy terms

The matching of a deformable contour to the object in a given image can be formulated as a maximizing a posteriori (MAP) estimation problem. Denote the mean of the sample contours in the shape domain as $\bar{f}_0$ (the mean contour), the deformed version of $\bar{f}_0$ and $\bar{f}$ (the prototype contour), and the deformable contour in the image domain as f, where $\bar{f}_0 \in R^{2\times N}$, $\bar{f} \in R^{2\times N}$ and $f \in R^{2\times N}$ are the matrices representing the corresponding contours formed by the coordinates of N landmark/boundary points. According to the Bayesian estimation, the joint posterior distribution of f and $\bar{f}$, $p(f,\bar{f}|d)$, is

$$p(f,\bar{f}|d) = \frac{p(d|f)p(f,\bar{f})}{p(d)} \tag{1}$$

where $p(d|f) = p(d|f,\bar{f})$ is the likelihood distribution of input image data d.

$$p(f,\bar{f}) = p(f|\bar{f})p(\bar{f}) \tag{2}$$

is the joint prior distribution of f and $\bar{f}$. For a given image d, the MAP estimates, f_{MAP} and $\bar{f}_{\mathrm{MAP}}$, can be defined as:

$$\begin{aligned}\{f_{\mathrm{MAP}},\bar{f}_{\mathrm{MAP}}\} &= \arg\max_{f,\bar{f}} \{p(f,\bar{f}|d)\} \\ &= \arg\max_{f,\bar{f}} \left\{ \frac{p(d|f)p(f|\bar{f})p(\bar{f})}{p(d)} \right\}\end{aligned} \tag{3}$$

When the densities can be modeled as Gibb's distribution, i.e.

$$\begin{aligned} p(\bar{f}) &= Z_1^{-1}\exp\{-E_{\mathrm{con}}(\bar{f})\} \\ p(f|\bar{f}) &= Z_2^{-1}\exp\{-E_{\mathrm{int}}(f|\bar{f})\} \\ p(d|f) &= Z_3^{-1}\exp\{-E_{\mathrm{ext}}(d|f)\} \end{aligned} \tag{4}$$

where Z_1, Z_2 and Z_3 are the partition functions, maximizing the posterior distribution is equivalent to minimizing the corresponding energy function of the contour:

$$\{f_{\mathrm{MAP}},\bar{f}_{\mathrm{MAP}}\} = \arg\min_{f,\bar{f}} \{E_{\mathrm{contour}}\} \tag{5}$$

where $E_{\mathrm{contour}} = E_{\mathrm{con}} + E_{\mathrm{int}} + E_{\mathrm{ext}}$. $E_{\mathrm{con}} = E_{\mathrm{con}}(\bar{f})$ is the constraint energy term of the adjustable prototype contour $\bar{f}$, which limits the variations of $\bar{f}$ and ensures that $\bar{f}$ is similar with $\bar{f}$ in shape. $E_{\mathrm{int}} = E_{\mathrm{int}}(f|\bar{f})$ is the internal energy term that describes the global and local shape deformation between f and $\bar{f}$. The external energy term $E_{\mathrm{ext}} = E_{\mathrm{ext}}(d|f)$ defines the degree of matching between f and the salient image features.

2.2. Constraint energy of the prototype contour

The constraint energy term E_{con} of the prototype contour is caused by the prior distribution of the samples in the shape domain. The density of $\bar{f}$, $p(\bar{f})$, can be estimated by applying PCA to the sample contours. In cases where all the samples are aligned views of similar objects seen from a standard view, this distribution usually be approximated

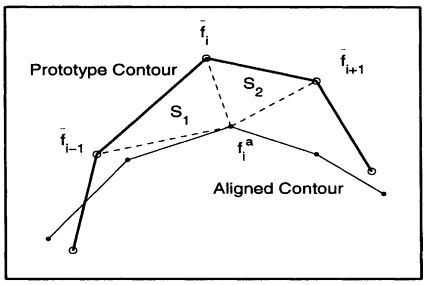

Fig. 1. A part of the prototype contour $\bar{f}$ and the aligned deformable contour f^a.

reasonably well by a single Gaussian distribution [12].

$$p(\bar{f}) = \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{w_i^2}{e_i}\right)}{(2\pi)^{M/2}\prod_{k=1}^{M}e_k^{1/2}} \tag{6}$$

where

$$w = \Phi_M^T(\bar{f} - \bar{f}_0) \tag{7}$$

is the vector of the shape parameters, and $\bar{f} - \bar{f}_0$ is the deformation from $\bar{f}_0$ to $\bar{f}$. Φ_M is the matrix composed of the eigenvectors corresponding to the largest M eigenvalues e_i, $(1 \le i \le M)$, which is computed from the covariance matrix of all the sample contours. Therefore, using the PCA, a prototype contour can be reconstructed from $\bar{f}_0$ and a given w,

$$\bar{f} = \bar{f}_0 + \Phi_M w \tag{8}$$

Note in Eq. (8), $\bar{f}$ and $\bar{f}_0$ have been expanded to $2N \times 1$ vectors. The PCA representation preserves the major linear correlations of the sample shapes and discards the minor ones, hence provides an optimized approximation of $\bar{f}$ in the sense of least squares error. This representation describes the most significant modes of the shape variations or the global shape deformations subject to the prior distribution of the prototype contour. From Eq. (6), the corresponding constraint energy is denoted as

$$E_{\text{con}} = \frac{1}{2}\sum_{i=1}^{M}\frac{w_i^2}{e_i} \tag{9}$$

The variations of the prototypes contour is limited by the plausible area of the corresponding shape parameters w, which is defined as:

$$\sum_{i=1}^{M}\frac{w_i^2}{e_i} \le M_{\text{t}} \tag{10}$$

The threshold, M_{t}, may be chosen using the χ^2 distribution [5]. The constraint energy term ensures that the dynamically adjustable prototype contour remains similar with the mean shape during the matching process, and at the same time, large shape variations and deformations subject to the prior distribution of the samples can be tolerated.

2.3. Affine-invariant internal energy

An AI internal energy term, $E_{\text{int}}(f|\bar{f})$, is defined and incorporated the AI-ES to deal with the affine transformations between the shape domain and the image domain, which describes the global and local shape deformations between $\bar{f}$ and f. Mathematically, $\bar{f}$ and f are related by $f_i = T(\mathbf{f}_i) + \boldsymbol{\epsilon} = A\bar{f}_i + \mathbf{t} + \boldsymbol{\epsilon}$, $(1 \le i \le N)$, where A is a 2×2 nonsingular matrix, $\mathbf{t}$ is a translation vector, and $\boldsymbol{\epsilon}$ represents the deformation. Define the least squares objective function as

$$E(A, \mathbf{t}) = \sum_{i=1}^{N}[(\bar{f}_i - A^{-1}(f_i - \mathbf{t}))^T \cdot (\bar{f}_i - A^{-1}(f_i - \mathbf{t}))] \tag{11}$$

the affine transformational parameters can be estimated by:

$$\hat{A} = [(f - f_{\text{av}})(\bar{f} - \bar{f}_{\text{av}})^T] \cdot [(\bar{f} - \bar{f}_{\text{av}}) \cdot (\bar{f} - \bar{f}_{\text{av}})^T]^{-1} \tag{12}$$

$$\hat{T} = [\hat{\mathbf{t}} \cdot [1, 1, \ldots, 1]]_{2\times N} = f_{\text{av}} - A\bar{f}_{\text{av}} \tag{13}$$

where $\bar{f}_{\text{av}} = [(1/N\sum \bar{f}_i) \cdot [1, 1, \ldots 1]]_{2\times N}$ and $f_{\text{av}} = 1/N \times \sum f_i \cdot [1, 1, \ldots 1]_{2\times N}$ are the matrices formed by the average points of $\bar{f}$ and f, respectively. Matrix $[(\bar{f} - \bar{f}_{\text{av}}) \cdot (\bar{f} - \bar{f}_{\text{av}})^T]$ is always nonsingular provided there exist at least three points in $\bar{f}$, which are not located in the same line.

Let (A_1, t_1) be the estimated transformational parameters between $\bar{f}$ and f, if f is affine-transformed to f', i.e. $f_i' = A'f_i + \mathbf{t}'$, $1 \le i \le N$, according to Eqs. (12) and (13), the estimated transformational parameters between $\bar{f}$ and f' will be $(A'A_1, A'\mathbf{t}_1 + \mathbf{t}')$. Therefore, the values of the objective function remain unchanged under affine transformations, and hence the global internal energy term of the deformable contour is designed as,

$$E_{\text{gint}}(f|\bar{f}) = \frac{1}{N}\sum_{i=1}^{N}[(\bar{f}_i - \hat{A}^{-1}(f_i - \hat{\mathbf{t}}))^T \cdot (\bar{f}_i - \hat{A}^{-1}(f_i - \hat{t}))] \tag{14}$$

where the transformational parameters $(\hat{A}, \hat{\mathrm{t}})$ are calculated by Eqs. (12) and (13). This internal energy indicates the global matching degree between the deformable contour f and the prototype contour $\bar{f}$.

In addition, a local internal energy term is also defined by affine invariants: the proportion of area [11],

$$E_{\text{lint}}(f_i|\bar{f}) = \frac{(S_1' + S_2)AREA_{\text{proto}}}{AREA_{\text{aligned}}} \tag{15}$$

$S_1 = S(\bar{f}_{i-1}, f_i^a, \bar{f}_i)$ and $S_2 = S(\bar{f}_i, f_i^a, \bar{f}_{i+1})$, where $S(\,)$ is the area of the triangle formed by the three points inside the brackets. $AREA_{\text{proto}}$ and $AREA_{\text{aligned}}$ represent the interior areas formed by the hull of the prototype contour and the aligned deformable contour $(f_i^a = \hat{A}^{-1}(f_i - \hat{t}))$ in the shape domain, respectively. Fig. 1 shows the geometrical relationship between S_1 and S_2. It can be seen from the figure that,

when the areas of the triangles S_1 and S_2 are close to zero, the shape and position of $\bar{f}$ and f will also be close. When the contour represents an open shape, the local internal energy at the endpoint of the contour will not be calculated.

In summary, the AI internal energy is composed of both the global and local terms,

$$E_{\text{int}}(f|\bar{f}) = E_{\text{gint}}(f|\bar{f}) + \frac{1}{N}\sum_{i=1}^{N} E_{\text{lint}}(f_i|\bar{f}) \tag{16}$$

which reflects the degree of fitting between $\bar{f}$ and f.

2.4. External image constraints

The external energy term $E_{\text{ext}} = E(d|f)$ indicates the degree of matching between the deformable contour f and the salient image features. Minimizing E_{ext} adjusts f and moves it towards the object boundary in the image d. The external energy usually combines all the information of edge, texture, color and region, etc, so that it can provide an effective description of the matching. For example, the color information can be combined into the edge detection process, so that the edge maps will accurately stand for the boundaries of the interested objects. Among various matching rules and external energy terms used in the literature, the energy term including both the gradient and directional edge information is utilized because of its simplicity and efficiency [8].

First, the image $d = \{d(x, y)\}$ is smoothed using Gaussian function, $G_\sigma(x, y)$, with the deviation σ.

$$d_\sigma(x, y) = G_\sigma(x, y)d(x, y) \tag{17}$$

Second, the normalized gradient of the smoothed image d_σ at each pixel location (x, y), denoted $d^g_\sigma(x, y) = (d^g_{\sigma x}(x, y), d^g_{\sigma y}(x, y))$, $(\|d^g_\sigma(x, y)\| \in [0, 1])$, is computed.

At last, constraints on the deformable contour f ensure that f moves towards the object boundaries: when the image pixels along the contour have high gradient magnitude, and the direction of the image gradient along the contour is perpendicular to the orientation of the contour, the external energy is small. Therefore, the external energy function is defined as:

$$E_{\text{ext}}(d|f) = \sum_{i-1}^{N} (1 - \|d^g_\sigma(x_i, y_i)\|)|\mathbf{n}(x_i, y_i)\cdot\mathbf{h}(x_i, y_i)| \tag{18}$$

where '·' is the dot product. $\mathbf{h}(x, y)$ is the direction of the gradient $d^g_\sigma(x, y)$, $\mathbf{h}(x, y) = d^g_\sigma(x, y)/\|d^g_\sigma(x, y)\|$ and $\|\mathbf{h}(x, y)\| = 1$. $\mathbf{n}(x_i, y_i)$ indicates the normal vector of the contour f at point $f_i = (x_i, y_i)$, with $\|\mathbf{n}(x_i, y_i)\| = 1$ and

$$\mathbf{n}(x_i, y_i) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{v}_i/\|\mathbf{v}_i\|,$$

where $\mathrm{v}_i = (f_{i+1} - f_i)/(\|f_{i+1} - f_i\|) + (f_i - f_{i-1})/(\|f_i - f_{i-1}\|)$ is the tangent vector of contour f at point f_i.

2.5. Computation of AI-ES solution

The computation of AI-ES is actually a solution finding procedure that minimizes the energy term of the deformable contour, E_{contour}. The strategies used in Refs. [6,5] are adopted: Gaussian pyramid for coarse-to-fine search and the rapid iterative approach to find the nearest plausible prototype shape $\bar{f}$ (estimate $\hat{A}$, $\hat{\mathbf{t}}$, and w) from f according to the edge profile information.

A two-stage procedure is used: coarse and fine. In the coarse matching, the coarse-to-fine search and the iterative schemes are utilized, and the deformation between f and $\bar{f}$ is not taken into account, so that a large initial range of the contour can be tolerated. Generally, in this stage, K is selected as three, and the convergence is declared when applying an iteration produces no significant change in the pose or shape parameters. In the fine matching stage, both of the deformations between f and $\bar{f}$, $\bar{f}$ and $\bar{f}_0$, are considered, which can fit the object closely. The main steps to compute the AI-ES solution are summarized as follows.

Stage 1 (coarse matching):
$k = K$, initialize contour f.

(a) Examine the image region around f, find out a new contour candidate f', which matches
image edge best (at resolution level k).
(b) Update the parameters of affine transform, $T(A, \mathbf{t})$, the shape parameters w, and the
prototype contour $\bar{f}'$, according to f' [5], where w is constrained in the plausible area.
(c) Update the current contour f to $T(\bar{f}')$, and the prototype contour $\bar{f}$ to $\bar{f}'$.
(d) Go to step (a) unless the newly updated f is close to the old one, or the maximum iterations have been supplied.
(e) if $k > 1$, then $k = k - 1$, goto (a), else exit.

Stage 2 (fine matching):
$k = 1$

(a) Calculate the internal and external energy terms, and update the contour f.
(b) Update the parameters of affine transform $T(A, \mathbf{t})$, the shape parameters w, and the prototype contour $\bar{f}$, according to f [5]. The constraints of w is applied to ensure the prototype contour is in a plausible area.
(c) Repeat (a)(b) until convergence.

3. Experimental results

Two sets of experiments are presented to illustrate the effectiveness of the AI-ES. The first compares the AI-ES with the ASM and AI-Snake models for matching hand

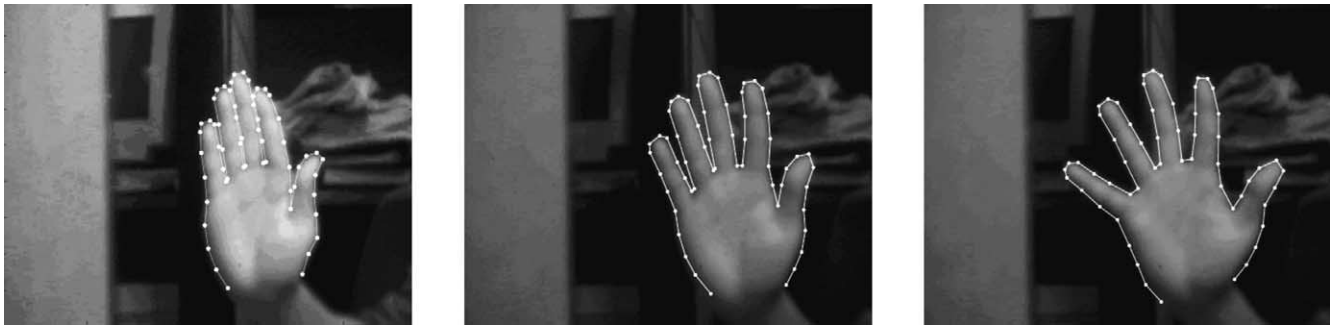

Fig. 2. Three training samples.

objects with different viewpoints and positions. The Eigen-Snake model has not been compared because it is by virtual unable to deal with affine transformation. The second is carried out to demonstrate the good performance of the AI-ES in matching the frontal faces with different rotation angles (rotation in the image plane).

3.1. Comparative study

The AI-ES is compared with the ASM and AI-Snake models in matching objects of hands in different viewpoints and positions. The results of ASM are obtained by using the ASM Toolkit (Version 1.0) of the Visual Automation Ltd (http://www.wiau.man.ac.uk/VAL). The experiment data is a series of hand images of size 240×179. 57 landmark/boundary points are used to describe the boundary of a hand in the image, and they are manually placed on the hand's outline of each training image. Fig. 2 plots three examples of the sample contours. To characterize the main shape of hand and its variations, the training set is acquired using the images of hand with fingers in different stretching angles. All the seven sample contours are aligned so that they are normalized into a standard view (see Fig. 3). These sample contours are then utilized to estimate the prior distribution or energy constraints of the prototype contours by applying PCA. The training set is the same for both ASM and AI-ES; whereas in AI-Snake, the prototype contour is fixed as the mean of the sample contours.

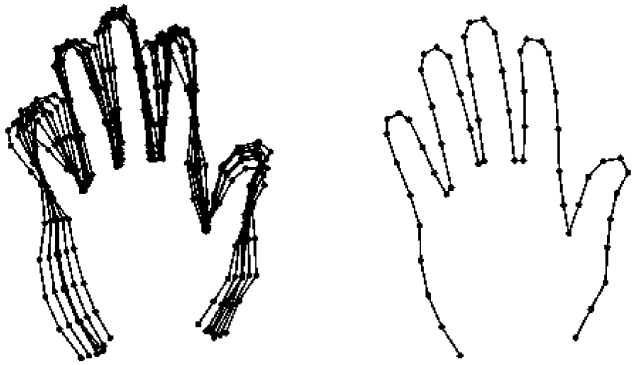

Fig. 3. All the seven aligned and normalized sample contours and the mean contour of hand.

Fig. 4 shows several examples of the results obtained by using the three models, in which the viewpoints of the hands are quite different from those of the samples, or/and the fingers are in different stretching angles. The initial contour for each test image is the same for all the models. It can be seen from the figures that with the fixed prototype contour used in the AI-Snake model, the matching results are poor when the stretching angle of the fingers changes greatly (it is unable to deform globally). For the ASM, good matching performance can be obtained when the position and viewpoint is similar to the sample data (refer to row 3, column 1). However, because the ASM model relies greatly on the training data (including the object and the background of the sample images), when the viewpoints and/or the shape change, it may not match the hand object closely (row 3, column 2–4). In comparison to that, the AI-ES is able to match objects more closely with different viewpoints and stretching angle of fingers (last row). The experiment indicates that by using the specific prior distribution of shapes and the AI internal energy constraint, the AI-ES is more robust and insensitive to large deformation and viewpoint of objects, resulting a much better performance that the other models.

3.2. Matching rotated frontal faces using AI-ES

Face detection, extraction, and recognition play an important role in automatic face recognition systems. In practice, the users may be expected to detect, match, or recognize faces at any angles. For example, in Ref. [13], a template matching is used, and the rotations are dealt with by enumerations of various rotated versions of the template. In the following experiments, the AI-ES is used to match the frontal faces with different rotation angles.

A contour with 89 landmark/boundary points, which is composed of six facial features: the face outline, brows, eyes, and a mouth, is recruited to describe a full face. The face images used for training are selected from the AR frontal face database [14] (http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html), which consists of various expression, male and female frontal face images. Fig. 5 shows two examples of the marked faces. All the sample contours are aligned and normalized using least squares

Fig. 4. Comparison of the three methods. Top: initial. Second row: Tracking results of AI-Snake. Third row: Results of ASM. Last row: Results of the proposed AI-ES.

error method to constitute the shape space. The set of all sample contours used in this experiment and the mean contour are also plotted in Fig. 6. The face images used for testing are chosen from the database of the Vision and Autonomous System Center (VASC) of CMU (http://www.cs.cmu.edu/afs/cs.cmu.edu/user/har/Web/faces.html), including images with expressions, different rotation angles, and complex image background.

In facial feature matching procedure, first the whole face contour model, as well as the separate shape models of its various components (i.e. outline, eyebrow, eye, and mouth), are built from the training samples. Then the whole face contour model is utilized for matching in the coarse searching stage. Because at this time the deforming face contour is relatively faraway from the target contour, the outline points are recruited to update the whole face contour, while the

Fig. 5. Samples of the marked frontal faces.

Fig. 6. All the sample contours and the mean contour.

Fig. 7. Examples of AI-ES results for matching frontal faces(I).

other points are passively updated according to this matching results. Therefore the initial component contours can be quite far away from the target contours. In the fine matching stage, all the 89 points are used to match the full face accurately. In this procedure, two steps are carried out iteratively: first each component contour is used to match its corresponding facial feature separately under the framework of AI-ES, then a shape constraint is calculated based on the whole face model to guarantee the correctness of the matching result, and the corresponding affine transformational parameters are updated. In this way, both global and local facial features can be matched accurately.

Figs. 7 and 8 shows some examples, including the initial contours and the final matching results, where the initial contours are properly placed manually as a common practice for snakes. It can be seen from the figures that AI-ES matches the faces closely even when they are in different positions and in-plane rotation angles. Table 1 gives some quantitative results of the matching. The accuracy of matching is defined as the average square error among the aligned points of the matching result and those of the target object marked manually. From the table it can be seen that the average errors of the matching results are very small, which indicates the target objects are matched accurately. In addition, the computational time for extracting the facial features (89 landmark points) is recorded, which is around 0.12 s per iteration under the environment of Matlab 5.2 using SUN Ultra 10 workstation, and it generally takes less than 20 iterations for the AI-ES to converge. The experimental results for facial feature extraction again verify the effectiveness of the proposed AI-ES model: flexible to large shape deformations subject to the sample data, and at the same time, robust to the positions and viewpoints of objects.

Table 1
Accuracy of matching of AI-ES (Unit in pixel)

Figure	Average error of initial contours	Average error of matching results
Fig. 7(a)	32.9	2.4
Fig. 7(b)	35.3	2.9
Fig. 7(c)	39.4	3.5
Fig. 7(d)	41.3	4.1
Fig. 8	38.4	4.7

Fig. 8. Examples of AI-ES results for matching frontal faces(II).

4. Conclusions

An effective AI deformable contour model, AI-ES, is developed in the Bayesian framework. The AI-ES allows large deformations for object matching under affine transformations. The global shape variation is modeled using the prior shape knowledge by PCA, and an AI internal energy is utilized to describe the shape deformations between the prototype contour in the shape domain. As such the contour is deformed to fit the object in a way which not only reflects shape variations in the training set, but also matches object closely under different translations, rotations, scaling and shearing of object shapes. Comparative studies show that the AI-ES produces more accurate results than the ASM and AI-Snake models. The effectiveness is further demonstrated through face extraction experiments. Although the AI-ES is used in this paper for face/hand extraction, the model may be applied to other applications such as object segmentation and tracking, as long as set of training samples of object shapes is available.

References

[1] A. Blake, M. Isard, Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion, Springer, Berlin, 1998.
[2] A. Blake, R. Curwen, A. Zisserman, Affine-Invariant Contour Tracking with Automatic Control of Spatiotemporal Scale, International Conference on Computer Vision (ICCV93), 1993, pp. 66–75.
[3] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, International Journal of Computer Vision 1 (4) (1988) 321–331.
[4] D. Terzopoulos, A. Witkin, M. Kass, Constraints on deformable models: recovering 3D shape and nonrigid motion, Artificial Intelligence 36 (1) (1988) 91–123.
[5] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application, Computer Vision and Image Understanding 61 (1) (1995) 38–59.
[6] A.K. Jain, Y. Zhong, S. Lakshmanan, Object matching using deformable templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (3) (1996) 267–278.
[7] S. Lakshmanan, D. Grimmer, A deformable template approach to detecting straight edges in radar images, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (4) (1996) 438–443.
[8] A.L. Yuille, P.W. Hallinan, D.S. Cohen, Feature extraction from faces using deformable templates, International Journal of Computer Vision 8 (2) (1992) 99–111.
[9] K.F. Lai, R.T. Chin, Deformable contours: modeling and extraction, IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (11) (1995) 1084–1090.
[10] Stan Z. Li, J. Lu, Modeling Bayesian estimation for deformable contours, Proceedings of Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece 1999, pp. 991–996.
[11] H.H.S. Ip, D. Shen, An affine-invariant active contour model (AI-Snake) for model-based segmentation, Image and Vision Computing 16 (1998) 125–146.
[12] B. Moghaddam, A. Pentland, Probablistic visual learning for object representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 696–710.
[13] H.A. Rowley, A. Baluja, T. Kanade, Rotation invariant neural network-based face detection, International Conference on Computer Vision and Pattern Recognition (CVPR'98), 1998, pp. 38–44.
[14] A. Martinez, R. Benavente, The AR face database, CVC Technical Report #24, June, 1998.
[15] T. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, Proceeding of Fifth European Conference on Computer Vision, vol. 2, 1998, pp. 484–498.
[16] U. Grenander, K.M. Keenan, Y. Chow, A Pattern Theoretic Study of Biological Shapes, Springer, New York, 1991.
[17] A.K. Jain, Y. Zhong, M-P. Dubuisson-Jolly, Deformable template models: a review, Signal Processing 71 (1998) 109–129.
[18] A.K. Jain, D. Zongker, Representation and recognition of handwritten digits using deformable templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (12) (1997) 1386–1391.
[19] M-P.D. Jolly, S. Lakshmanan, A.K Jain, Vehicle segmentation and classification using deformable templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (3) (1996) 293–308.
[20] A. Lanitis, C.J. Taylor, T.F. Cootes, Automatic interpretation and coding of face images using flexible models, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 743–756.
[21] K.V. Mardia, W. Qian, D. Shah, K.M.A. de Souza, Deformable template recognition of multiple occluded objects, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (9) (1997) 1035–1042.
[22] A.T. Matthew, A.P. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
[23] D. Mumford, Pattern theory: a unified perspective, in: D. Knill, W. Richard (Eds.), Perception as Bayesian Inference, Cambridge University Press, Cambridge, 1996.
[24] L.H. Staib, J.S. Duncan, Boundary finding with parametrically deformable models, Transactions on Pattern Analysis and Machine Intelligence 14 (11) (1992) 1061–1075.
[25] A. Singh, D.B. Goldgof, D. Terzopoulos, Deformable Models in Medical Image Analysis, IEEE Computer Society, Los Alamitos, CA, 1998.
[26] Y.S. Akgul, C. Kambhamettu, M. Stone, Automatic extraction and tracking of the tongue contours, IEEE Transactions on Medical Imaging 18 (10) (1999) 1035–1045.
[27] H. Li, B.S. Manjunath, S.K. Mitra, A contour-based approach to multisensor image registration, IEEE Transactions on Image Processing (1995) 320–334.