International Journal of Soft Computing 3 (4): 271-276, 2008
ISSN: 1816-9503

Facial Tracking in Video Using Artificial Neural Networks

[1]E. Mohan, [2]C. Puttamadappa, [3]B. Jayaraman, [4]E. Anbalagan and [5]Srinivasarao Madane
[1]Vinayaka Missions University, Salem, Tamil Nadu, India
[2]New Horizon College of Engineering, Bangalore-560087, India
[3]Department of Computer Science and Engineering, [4]Department of Information Technology,
Pallavan College of Engineering, Thimmasamudram, Kanchipuram-631 502, Tamilnadu, India
[5]Adhiparasakthi College of Engineering, Kalavai, Tamilnadu, India

Abstract: A region-based method for facial tracking is proposed in this study. In this method, the facial information of temporal motion and spatial luminance are fully utilized. The dominant motion of the tracked facial object is computed. Using this result, the object template is warped to generate a prediction template. A method is proposed which incorporates an Artificial Neural Network (ANN) with Back-Propagation Algorithm (BPA) to modify the prediction. A decision approach with a threshold is used to detect if there is any change in the object of the successive frames. The accuracy of the result depends upon the number of nodes in the hidden layer and learning factor. The number of nodes in the hidden layer is 10 and the learning factor is 1. The performance of the algorithm in reconstructing the tracked object is about 96.5% in terms of reduced time and quality of reconstruction.

Key words: Back-Propagation Algorithm (BPA), watershed algorithm, motion estimation, video frames

INTRODUCTION

Tracking objects in image sequences is an important task for vision-based control, Human Computer Interaction (HCI), content-based video indexing and structure from motion. A great variety of visual tracking algorithms have been proposed: they can be classified roughly into 2 categories (Hagger and Belhumeur, 1998). The first is the feature-based method. A typical instance in this category estimates the 3D pose of a target object to fit into the image features such as contours given a 3D geometric model of the object. The second is the region-based method. Compared to the feature-based methods, the region-based methods are more robust, insensitive to small partial occlusions. The region-based methods can be subdivided into 2 groups: the view-based method and the parametric method. The proposed method belongs to the latter group.

Some related research in the literature of region-based visual tracking is discussed below. Shi and Tomasi (1994) put forward the criterion of "good features" by its texture and used it in affine feature tracking. Parry *et al.* (1996) introduced a region-based (formed by segmentation) tracking method, mainly updating the template by projecting it around the detected positions of the target and considering its overlap with the segmented image. The tracking results shows good performance when the camera moves towards the object. Hager and Belhumeur (1998) developed a general framework for region tracking which includes models for image changes due to motion, illumination and partial occlusion. They used a cascaded parametric motion model and a small set of basis images to account for shading changes, which will be solved in a robust estimation framework in order to handle small partial occlusion. Gleicher (1997) introduced difference decomposition to solve the registration problem in tracking, where the difference would be linear combination of a set of basis vectors. Sclaroff and Isidoro (1990) used this idea for template registration in region-based non-rigid tracking, where the non-rigid deformation was represented in terms of eigenvectors of a finite element method. Photometric variation is considered and a modified Delaunay refinement algorithm is used to construct a consistent triangular mesh for the region of the tracked object.

Corresponding Author: E. Mohan, Department of Computer Science and Engineering, Pallavan College of Engineering, Thimmasamudram, Kanchipuram-631 502, Tamilnadu, India

Nguyen and Worring (2000) made their contribution by introducing a contour tracking method incorporating static segmentation by the watershed algorithm. Their method utilized kinds of edge maps from motion (optic flow), intensity (watershed) and prediction (contour warping) to update the object contour. It was claimed that this method yielded accurate and robust results.

This study proposes a region-based method for motion estimation undergoing object tracking. Tracking is performed by means of motion segmentation. The proposed method fully utilizes information of temporal motion and spatial luminance. Computation of dominant motion of the tracked object is done by a robust Iterative Weights Least Square (IWLS) method. Static segmentation is incorporated to modify this prediction, where the warping error of each watershed segment and its rate of overlapping with warped template are utilized to help classification of some possible watershed segments near the object border.

The concept of ANN with supervised algorithm is proposed in addition to the above described method for computing the affine transformation-taking place in the current frame with respect to previous frame. This is achieved, when there is a significant change in the output of the neural network which indicates the change in position of the object in the current frame. To detect the change in position of the object, the network has to be trained in advance under supervised mode.

The trend of this research is comparable to the work of Nguyen and Worring (2000). The idea of "active blob"discusses the non-rigid deformation. The Delaunay triangulation of computer graphics is used to generate some mesh of the object region (Sclaroff and Isidoro, 1998).

MULTILAYER ARCHITECTURE

An Artificial Neural Network (ANN) is a mathematical way of simulating the capability of human brain. The category of supervised method requires inputs and target outputs.

The Back-Propagation Algorithm (BPA) is a supervised method that uses steepest-descent method to reach global minima. The flowchart for the BPA is given in Fig. 1. The number of layers and number of nodes in each layer is decided. The connections between nodes are initialized with random weights. A pattern from the training set is presented in the input layer of the network and the error is calculated in the output layer. The error is

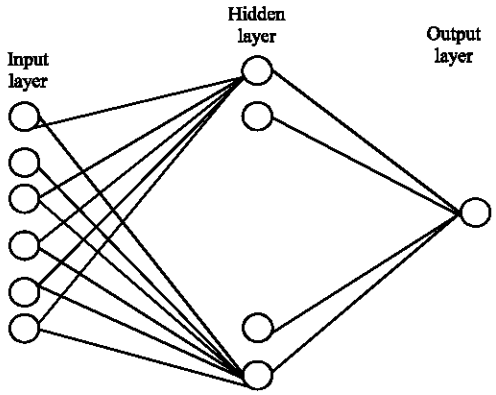

Fig. 1: Back propagation network

propagated backwards towards the input layer and the weights are updated. This procedure is repeated for all the training patterns.

At the end of iteration, test patterns are presented to ANN and the classification performance of ANN is evaluated. Further training of ANN is continued till the desired classification performance is reached.

Steps involved

Forward propagation: The output of each node in the successive layers is calculated.

$$O(\text{output of a node}) = 1/(1+\exp(-\Sigma w_{ij}\, x_i)) \quad (1)$$

The error E(p) of a pattern number p is calculated

$$E(p)=(1/2)\,\Sigma(d(p) - o(p))^2 \quad (2)$$

Reverse propagation: The error δ for the nodes in the output layer is calculated

$$\delta(\text{output layer}) = o(1-o)(d-o) \quad (3)$$

The new weights between output layer and hidden layer are updated

$$W(n+1) = W(n) + \eta\, \delta(\text{output layer})\, o(\text{hidden layer}) \quad (4)$$

The error δ for the nodes in the hidden layer is calculated

$$\delta(\text{hidden layer}) = o(1-o)\, \Sigma\delta(\text{output layer})\, W \text{ (updated weights between hidden and output layer)} \quad (5)$$

The weights between hidden and input layer are updated.

$$W(n+1) = W(n) + \eta\delta(\text{hidden layer})\, o(\text{input layer}) \quad (6)$$

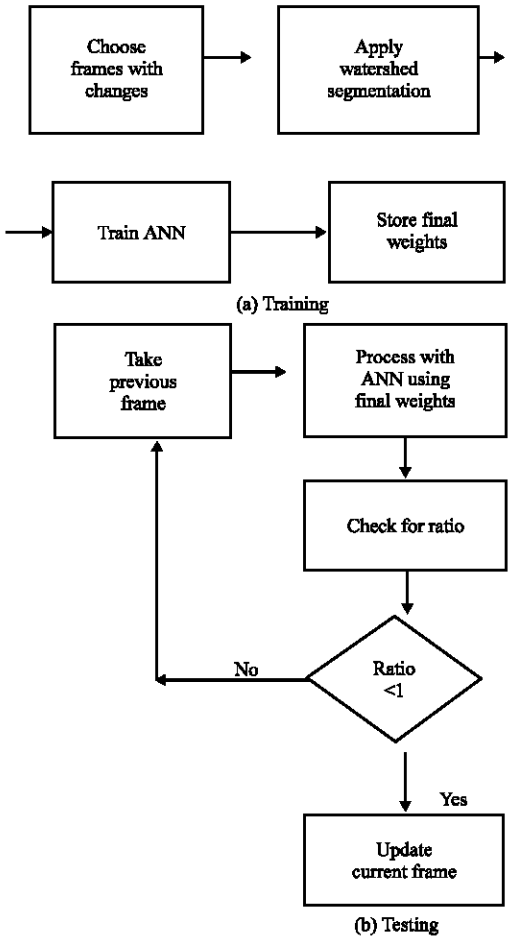

Fig. 2: Procedure for object tracking

The above steps complete one weight updation. The remaining training patterns are presented and Eq. 2.1-2.6 are followed which form one iteration. The training of the network is stopped once the desired Mean Squared Error (MSE) is reached as given:

$$E(MSE) = \Sigma E(p) \quad (7)$$

The final updated weights are saved for testing the video transfer.

Schematic diagram of object tracking is shown in Fig. 2.

Motion Estimation using the M-estimator: The inter-frame motion is defined as

$$f(x, t+1) = f(x-u(x; a), t) \quad (8)$$

where f (x, t) is the brightness function at the time instant t, x = (x, y) as the coordinate of the image pixel and u (x; a) is the motion vector. Without loss of generality, simple affine transform model is selected as the motion model:

$$u(x:a) = \begin{bmatrix} u(x,y) \\ v(x,y) \end{bmatrix} = \begin{bmatrix} a_o + a_1 x + a_2 y \\ a_3 + a_4 x + a_5 y \end{bmatrix} \quad (9)$$

where, $a = (a_0, a_1, a_2, a_3, a_4, a_5)^T$ are the parameters of the affine model. The dominant motion estimation of the given region R is formulated as the following robust M-estimator,

$$\min E_p = \sum \rho(uf_x + vf_y + f_t, \sigma \quad (10)$$

Here f_x, f_y, f_t are the partial derivatives of brightness function with respect to x, y and t, the ρ function is chosen as the German-McClure function (Black and Yaqoob, 1995) and σ is the scale parameter. To solve the problem, there are two different ways to find robustly the motion parameters: One is gradient-based, like the SOR method in (Black and Yaqoob, 1995), another is least squares-based, such (IWLS) method. The algorithm begins by constructing the Gaussian pyramid (three levels are set up). When the estimated parameters are interpolated into the next level, they are used to warp (realized by bilinear interpolation) the last frame to the current frame. In the current level only the changes are estimated in the iterative update scheme.

Static segmentation by watershed: In static segmentation, the watershed algorithm of mathematical morphology is a powerful method (Vincant and Soilie, 1991). Early watershed algorithms are developed to process digital elevation models and are based on local neighborhood operations on square grids. Some approaches use "immersion simulations"to identify watershed segments by flooding the image with water starting at intensity minima. Improved gradient methods are devised to overcome plateaus and square pixel grids (Gauch, 1999). Here the former method is used. A severe drawback to the computation of watershed algorithm is over-segmentation. Normally watershed merging is performed along with the watershed generation. Here over-segmentation is welcome, so during tracking the merging process is omitted, which saves some computational costs. Figure 3 shows procedure for watershed segmentation.

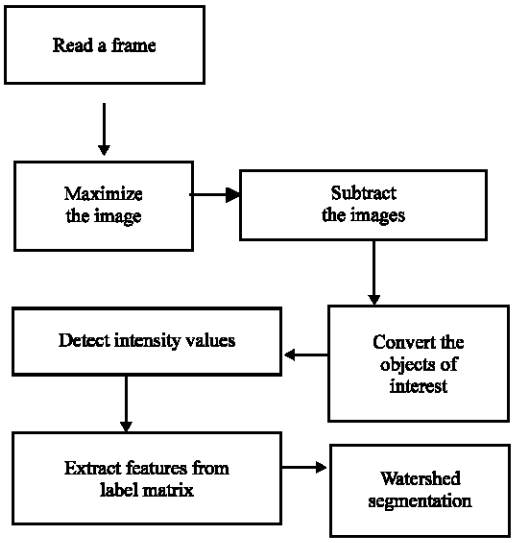

Fig. 3: Flow chart for watershed algorithm

TEMPLATE WARPING AND REGION ANALYSIS

Once the motion parameters have been computed, warp the object template from the last frame to the current frame. Then the warped template is used to determine which watershed segments enter the template according to the following measure: Given that the number of pixels belonging to the warped template in the number of all pixels in R_i is C_i, a ratio r_i is computed,

$$r_i = C_{pi}/C_i \quad (11)$$

Based on this measure the classification problem of each sub-region is discussed in the following cases:

- When $r_i > r_0$, then classify R_i as part of the final object template.
- When $r_0 \geq r_i \geq r_1$ (here $r_1 = 0.4$), another measure as MAE (Mean Absolute Error) of difference between the warped frame and the current frame is taken into account.

$$M_i = \sum |f(x, t+1) - f^w(x, t)| / C_i \quad (12)$$

where, $f^w(x, t)$ the warped image of $f(x, t)$ using the estimated dominant motion parameters; If the warped error M_i of R_i is smaller enough (less than a of $f(x, t)$ using given threshold, for instance, R_i is still regarded as part of the updated template; otherwise, exclude R_i out of the object region.

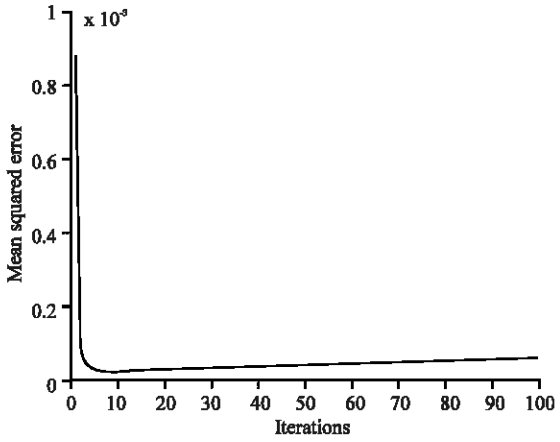

Fig. 4: Error curve

- When $r_i < r_1$, R_i will not be included in the updated template.

Multiple objects tracking: When people make facial expression movements, especially behaving emotionally (mainly 6 universal facial expressions are to be discussed, i.e., Disgust, Sadness, Happiness, Fear, Anger and Surprise), in most of cases head motion is accompanied. The procedure is divided into 2 steps: Head tracking is realized first, then the estimated motion is used to stabilize the face region. The local motion of each facial feature is estimated relative to the stabilized face.

Human face motion is complex with rigid and non-rigid movements; hence the idea in Black and Yacoob (1995) adopted using a modified affine model to describe the local motion of facial features (mouth, eyes and eyebrows) and a planar projective transform to model the head motion. The IWLS method is used to estimate these motion parameters.

RESULTS

The project is implemented using Matlab 7. The time taken for processing each frame is at an average of 1.4 sec. This includes segmentation and processing with neural network. The topology of the ANN is 100×50×1. The mean squared error (Fig. 4) at which the training stopped was 0.01. The total number of frames in the video are 91 (Fig. 5). For the experiment only 8 frames (Fig. 6) are considered which show significant changes in the lip movements.

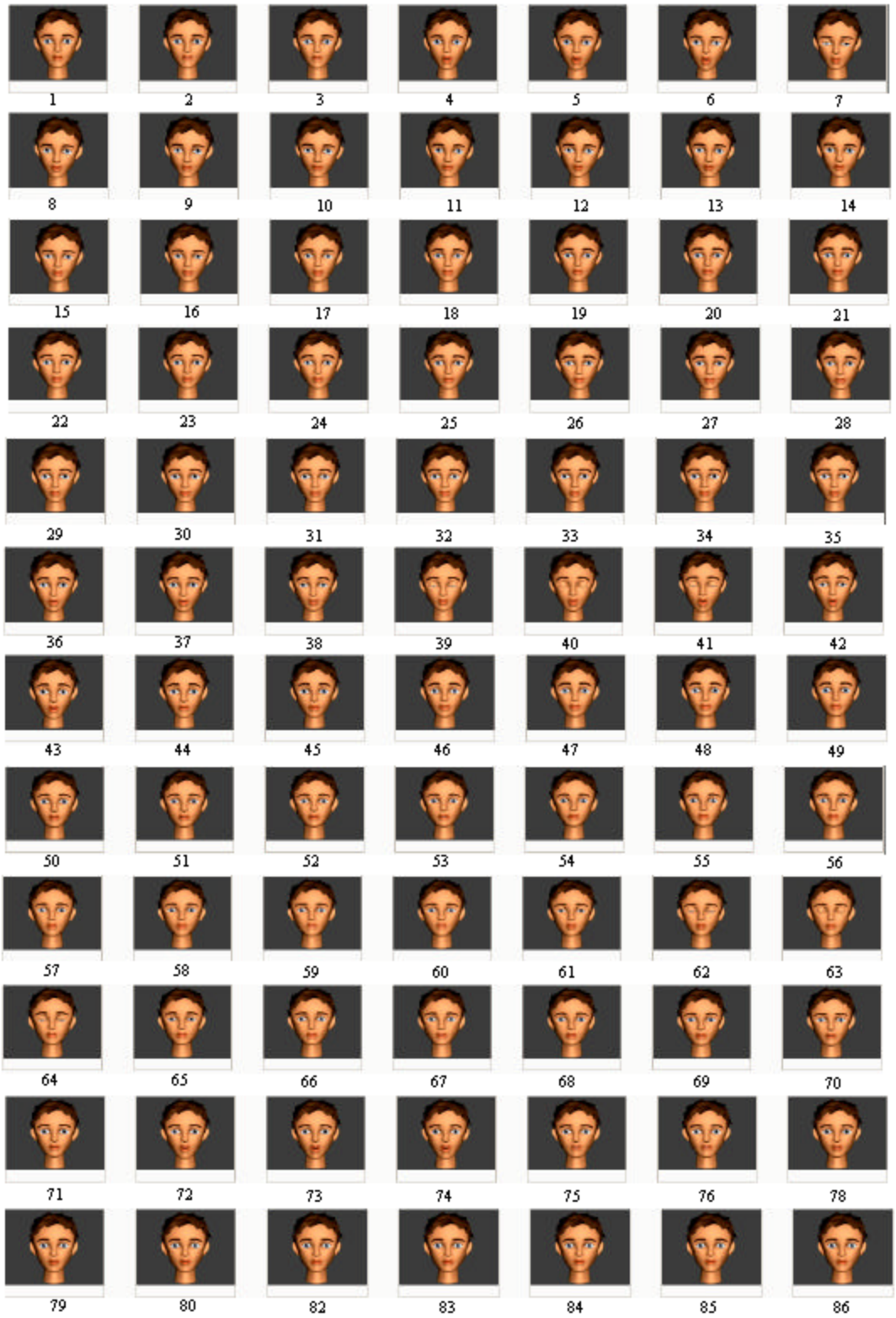

Fig. 5: Video frames

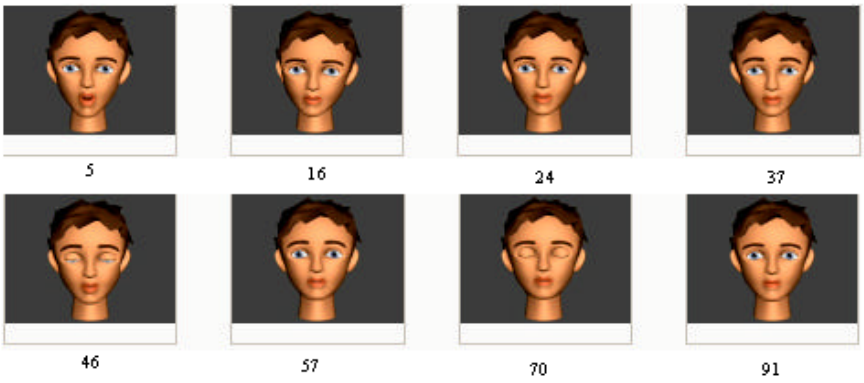

Fig. 6: Experimental results relating to lip movements

CONCLUSION

In this study an ANN based approach is proposed for motion estimation undergoing object tracking. The lip movements are mainly focused in this research. The template warping by watershed segmentation and ANN for quick decision of frame updation is implemented. The MSE used was 0.01 and the learning parameter as 1 for training the ANN. Applications of this method in facial expression tracking can be expressed for other parts of face.

REFERENCES

Black, M. and Y. Yacoob, 1995. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion, ICCV.

Gleicher, M., 1997. Projective registration with difference decomposition. IEEE. CVPR., pp: 331-337.

Gauch, J., 1999. Image segmentation and analysis via multi-scale gradient watershed hierarchies. IEEE T-IP, 8: 69-79.

Hager, G. and P. Belhumeur, 1998. Efficient region tracking with parametric models of geometry and illumination. IEEE T-PAMI, 20: 1025-1039.

Nguyen and Worring M., 2000. Multi-feature object tracking using a model-free approach. IEEE. CVPR., pp: 145-150.

Parry *et al*., 1996. Region Template Correlation for FLIR Target Tracking, British Machine Vision Conference.

Purushothaman, S. and Y.G. Srinivasa, 1998. A procedure for training an artificial neural network with the application of tool wear monitoring. Int. J. Prod. Res., 36: 635-651.

Shi, J. and C. Tomasi, 1994. Good features to track. In: Proc. Computer Vision and Pattern Recognition.

Sclaroff, S. and J. Isidoro, 1998. Active blobs, ICCV.

Vincent L. Soille, 1991. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE T-PAMI., 13: 583-589.