International Journal of Engineering & Technology, 3 (4) (2014) 535-539
©Science Publishing Corporation
www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v3i4.3633
Research Paper

Facial tracking in video transfer using watershed segmentation

P. Mayilvahanan [1]*, S. Purushothaman [2], R. Rajeswari [3]

[1] P. Mayilvahanan Professor Department of MCA VELS University, India-600117
[2] S. Purushothaman, Professor, PET Engineering College, India
[3] R. Rajeswari, Associate Professor, Institute of Technology, Haramaya University, Ethiopia
*Corresponding author E-mail: mayilkadir@yahoo.com

Abstract

This paper presents the implementation of watershed segmentation for facial tracking. It focuses on tracking face of human in a continuous video transfer. The database is obtained from internet resource. Tracking efficiency is 97.2%.

Keywords: *Facial Tracking, Watershed Segmentation, Motion Estimation.*

1. Introduction

Human face processing techniques are used for broadcast video, including face detection, tracking, and recognition has attracted a lot of research interest because of its value in various applications, such as video structuring, indexing, retrieval, and summarization. The main reason for this is that the human face provides rich information for spotting the appearance of certain people of interest, such as government leaders in news video, the pitcher in a baseball video, or a hero in a movie, and is the basis for interpreting facts. Face processing techniques have several applications.
In many face recognition systems, the input is a video sequence consisting of one or more faces. It is necessary to track each face over this video sequence so as to extract the information that will be processed by the recognition system. Tracking is also necessary for 3D model-based recognition systems where the 3D model is estimated from the input video. Face tracking can be divided along different lines depending upon the method used, e.g., head tracking, feature tracking, image-based tracking, and model-based tracking. The output of the face tracker can be the 2D position of the face in each image of the video (2D tracking), the 3D pose of the face (3D tracking), or the location of features on the face. Some trackers can output other parameters related to lighting or expression. The major challenges encountered by face tracking systems are robustness to pose changes, lighting variations, and facial deformations due to changes of expression, occlusions of the face to be tracked and clutter in the scene that makes it difficult to distinguish the face from the other objects.
Tracking, which is essentially motion estimation, is an integral part of most face processing systems. Face tracking is the process of locating a moving face or several of them over a period by using a camera. A given face is first initialized manually or by a face detector. The face tracker then analyzes the subsequent video frames and outputs the location of the initialized face within these frames by estimating the motion parameters of the moving face. The outcome is the position and scale of face detection in one single frame. Face tracking acquires information on multiple consecutive faces within consecutive video frames. More importantly, these faces have the same identity.

1.1. Benefits of face tracking

One of the main applications of face tracking is person retrieval from broadcast video, for example: “intelligent fast-forward,” where the video jumps to the next scene containing a certain person/actor; or retrieval of different TV segments, interviews, shows, etc., featuring a given person in a video or a large collection of videos. Face tracking is also used for face-name association, the objective of which is to label television or movie footage with the identity of the person present in each frame of the video.

1.2. Challenges in face tracking

The main challenges that face tracking methods have to overcome are (i) variations of pose and lighting, (ii) facial deformations, (iii) occlusion and clutter, and (iv) facial resolution. Robustness to pose and illumination variations often lead to loss of track. Low resolution will hamper performance of any tracking algorithm, face tracking being no exception.

1.3. Applications of face tracking

Video Surveillance: Since faces are often the most easily recognizable signature of identity and intent from a distance, video surveillance systems often focus on the face, Zhao [1]. In order to focus on the face, tracking the face over multiple frames is required.
Biometrics: Video-based face recognition systems require alignment of the faces before they can be compared. This alignment compensates for changes of pose. Face tracking, especially 3D pose estimation, is, therefore, an important component of such applications. Also, integration of identity over the entire video sequence requires tracking the face
Face Modeling: Reconstruction of the 3D model of a face from a video sequence using structure from motion requires tracking. The depth estimates are related non-linearly to the 3D motion of the object.
Video Communications and Multimedia Systems: Face tracking is also important for applications like video communications. Motion estimates remove the inter-frame redundancy in video compression schemes like MPEG and H.26x. In multimedia systems like sports videos, face tracking can be used in conjunction with recognition or reconstruction modules, or for focusing on the region of interest in the image.

2. Literature survey

Ubiquitous application of eye tracking is precluded by the requirement of dedicated and expensive hardware, such as infrared high definition cameras, Valenti [2]. The systems are based solely on appearance are used in literature. These systems can successfully locate eyes; their accuracy is significantly lower than commercial eye tracking devices. Their aim is to perform very accurate eye center location and tracking, using a simple web cam. By means of a novel relevance mechanism, the method makes use of isopoda properties to gain invariance to linear lighting changes, to achieve rotational invariance and to keep low computational costs. They test their approach for accurate eye location and robustness to changes in illumination and pose, using the BioID and the Yale Face-B databases. They demonstrate that their system can achieve a considerable improvement in accuracy over state of the art techniques.
The use the state-of-the-art progress on visual tracking methods, classify them into different categories and identify future trends were implemented, Yang [3]. Visual tracking is a fundamental task in many computer vision applications and has been well studied in the last decades. Robust visual tracking remains a huge challenge. Difficulties in visual tracking can arise due to abrupt object motion, appearance pattern change, non-rigid object structures, occlusion and camera motion. They analyze the state-of-the-art feature descriptors which are used to represent the appearance of tracked objects. They categorize the tracking progresses into three groups; provide detailed descriptions of representative methods in each group examine their positive and negative aspects and the future trends for visual tracking research.
The human detection, tracking, and identification are presented in Mau-Tsuen [9]. Skin color modeling has been used for facial tacking, prashanth kumar [10].

3. Materials and methodology

Table 1: Sample Frames from a Video

	Start frame	End frame	Frame start	Frame end	Total frames
Video 3			34749	35125	376

3.1. Face tracking algorithm

A face appears in the successive frames.
The frames with face are available continuously in the next 2029 frames.
The tracking mode appears for person-1 after certain number of frames.

3.2. Schematic diagram of facial tracking

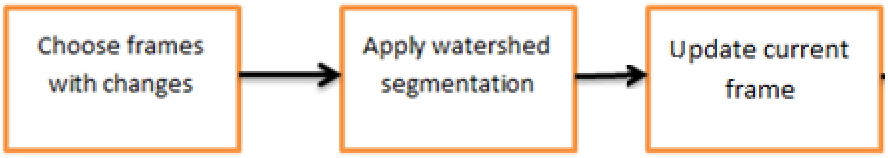

Fig. 1: Schematic Diagram for Frame Updation

In the video, many objects are present. We are mainly focusing of facial tracking to mainly achieve the following tasks. Reduce the consumption of bandwidth in transferring frames of video during a conversation in video conferencing. To transfer expression of the human face onto face of any other model.
The following procedure is used to implement watershed segmentation algorithm for facial tracking.
Step 1: Read frame.
Step 2: Take a portion of frame (eye/nose/lip etc.)
Step 3: Apply watershed segmentation.
Step 4: Find the mean of the segmented image.

3.3. Motion estimation using the m-estimator

The inter-frame motion is defined as follows:

$$F(x, t+1) = f(x-u(x;a),t), \tag{1}$$

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

$$u(x;a) = \begin{bmatrix} u(x,y) \\ v(x,y) \end{bmatrix} = \begin{bmatrix} a_0 + a_1 x + a_2 y \\ a_3 + a_4 x + a_5 y \end{bmatrix} \tag{2}$$

Where
$a = (a_0, a_1, a_2, a_3, a_4, a_5)^T$ are the parameters of the affine model. So, the dominant motion estimation of a given region 'R' is formulated as the following robust M-estimator,

$$\min_{(u,v)} E = \sum_{(x,y)} (uf_x + vf_y + f_t, \sigma) \tag{3}$$

Here, f_x, f_y and f_t are the partial derivatives of brightness function with respect to x, y and t, Nesi [4]; and σ is the scale parameter. To solve this problem, there are two different ways to find robustly the motion parameters: one is gradient-based, like the successive over relaxation method, Gauch [5]; another is least squares-based, such as Iterative Weights Least Square (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.
In static segmentation, the watershed algorithm of mathematical morphology is a powerful method, Gleischer [6]. 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. The former method is used. A severe drawback, to the computation of the watershed algorithm, is over-segmentation. Normally watershed merging is performed, along with the watershed generation. Over-segmentation is welcome; so, during tracking, the merging process is omitted, which saves some computational costs. Figure .2 shows procedure for watershed segmentation.

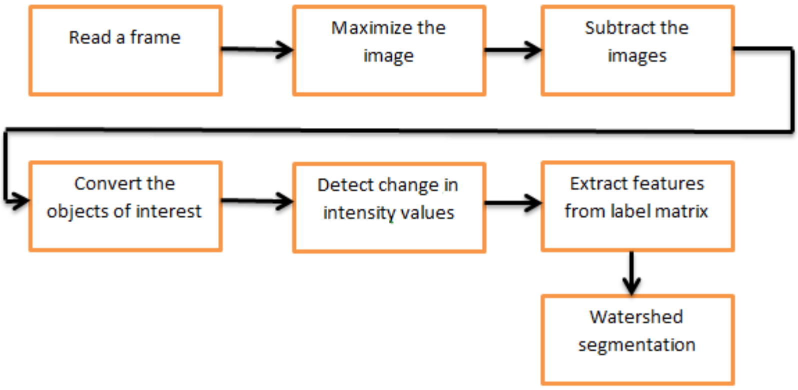

Fig. 2: Flow Chart for Watershed Algorithm

3.4. 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, Vincent [7], 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, as given in eq(4):

$$r_i = C_{Pi} / C_i \quad (4)$$

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

1) When $r_i > r_0$, then, classify r_i as part of the final object template;

2) When $r_0 \geq r_i \geq r_1$ (here $r_1 = 0.4$), another measure as MAE (Mean Absolute Error) of the difference between the warped frame and the current frame is taken into account, equation (5)

$$M_i = \sum |f(x,t+1) - f^W(x,t)| / C_i \quad (5)$$

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, 10), R_i is still regarded as part of the updated template; otherwise, exclude R_i out of the object region.

3) When $r_i < r_1$, r_i will not be included in the updated template.

When people make facial expression movements, especially behaving emotionally, (mainly, six universal facial expressions are to be discussed, i.e. disgust, sadness, happiness, fear, anger and surprise), in most of the cases, head motion is accompanied. The procedure is divided into two steps:

1) Head tracking is realized first, and then the estimated motion is used to stabilize the face region;
2) The local motion of each facial feature is estimated relatively to the stabilized face.

Human face motion is complex with rigid and non-rigid movements; hence the idea in, Yan [8], 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.

4. Results and discussion

Figure 3 shows various facial expressions of a standard 'Ruth' video database.

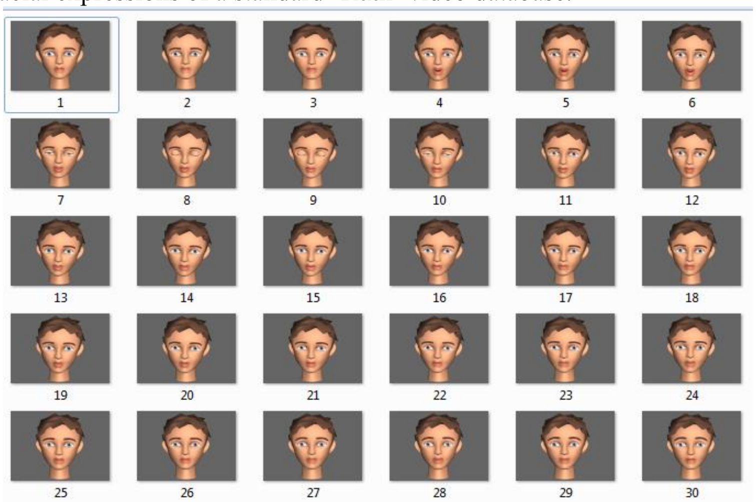

Fig. 3: Frames 1-30 of 'Ruth' Video

There are about 3000 frames in video. Frames 968 till 2997 contain the person. Figure.4 shows frame numbers in x-axis and each frame's summed intensity value along y-axis. There is a lot of variations in the y-axis for many frame numbers. A threshold has to be fixed such that only frame whose summed threshold is above a certain value should be considered for extracting information

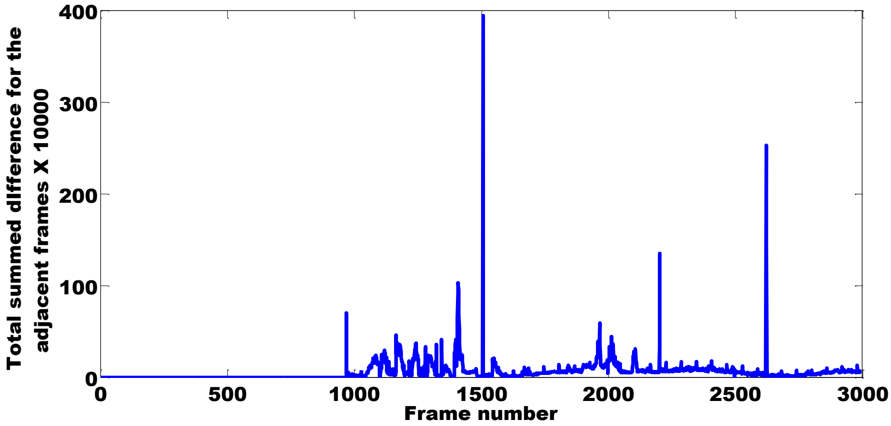

Fig. 4: Frame Differences for the Person

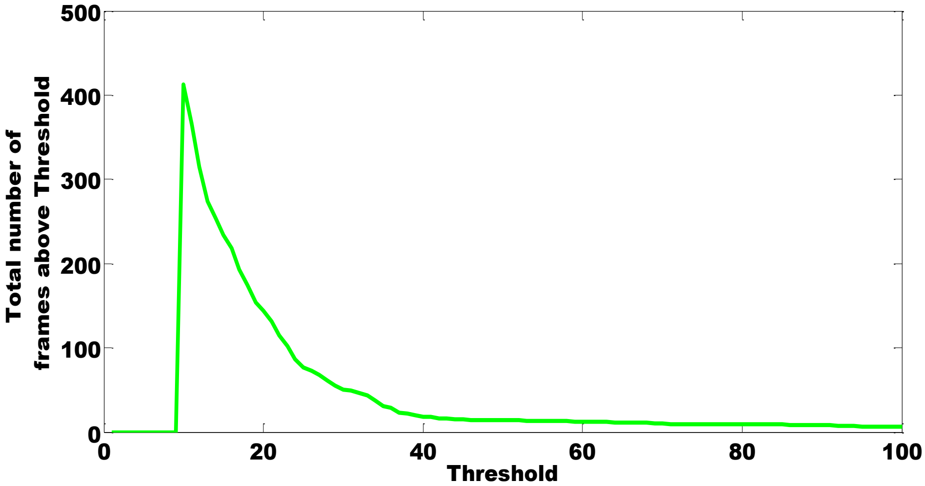

Fig. 5: Number of Frames above a Threshold for Figure 4, the Total Number of Frames that are Considered is 80.

5. Conclusion

The local features of the face are extracted and compared among successive frames. The full face of the sample videos shown is tracked. This concept can be extended for many faces that are present in video.

References

[1] W.Zhao, R.Chellappa, P.Phillips, A.Rosenfeld, Face Recognition: A Literature Survey. ACM Transactions, (2003), 399–458.

[2] R.Valenti and T.Gevers, Accurate eye center location and tracking using isopoda curvature, Intelligent Systems Lab Amsterdam University of Amsterdam, The Netherlands., (2008).

[3] H.Yang, L.Shao, F.Zheng, L.Wang, and Z.Song, Recent advances and trends in visual tracking: a review, Neuro computing, 74, 18, (2011), 3823–3831.

[4] P.Nesi, and R.Magnolfi, 1996, Tracking and Synthesizing Facial Motions with Dynamic Contours, Real-Time Imaging, 2, 2, 67–79. http://dx.doi.org/10.1006/rtim.1996.0008.

[5] J.Gauch, Image segmentation and analysis via multi-scale gradient watershed hierarchies, IEEE T-IP, 8, 1, (1999), 69-79.

[6] M.Gleicher, Projective registration with difference decomposition, IEEE CVPR'97, (1997), 331-337.

[7] L.Vincent Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, IEEE T-PAMI, 13, 6, (1991), 583-589.

[8] Yan Tong, Yang Wang, Zhiwei Zhu, Qiang Ji, 2007, Robust facial feature tracking under varying face pose and facial expression, Pattern Recognition, 40, (2007), 3195–3208. http://dx.doi.org/10.1016/j.patcog.2007.02.021.

[9] Mau-Tsuen Yang and Shen-Yen Huang, Appearance-Based Multimodal Human Tracking and Identification for Healthcare in the Digital Home, Sensors 2014, 14, 14253-14277; doi:10.3390/s140814253. http://dx.doi.org/10.3390/s140814253.

[10] G.Prashanth Kumar and M.Shashidhara, Real Time Detection and Tracking of Human Face Using Skin Color Segmentation and Region Properties. International Journal of Signal Processing Systems 2, 2, (2014), 102-107, doi: 10.12720/ijsps. http://dx.doi.org/10.12720/ijsps.