FACE TRACKING SYSTEM USING CAMSHIFT ALGORITHM

Gîlcă Gheorghe, ***PhD., University of Craiova, Faculty of Automation, Computers and Electronics, Craiova, ROMANIA***
Bîzdoacă Nicu - George, ***Prof. PhD., University of Craiova, Faculty of Automation, Computers and Electronics, Craiova, ROMANIA***

Abstract: *This paper deals with an implementation of a face tracking system in a video sequence. The tracking algorithm consists of three subsystems: the primary system of face detection, the identification of the complexion features, and face tracking. In the first subsystem we have used the Viola-Jones algorithm and a classification model specially developed for detection, which is inserted in the Computer Vision System toolbox of Matlab , the second subsystem mainly selects the skin tone and pigment as a tracking feature – an unique facial feature which stays invariable even when the individual is in full motion and the third subsystem will apply the tracing method based on the histogram that has the CAMShift algorithm implemented.*

Key words: *face detection, face tracking, meanshift, camshift*

1. INTRODUCTION

Face detection is an important component in many applications such as: biometric, real-time recognition, human – computer interaction, automotive, robotics, video surveillance, industrial automation and video games. This paper is organized as follows: section 2 presents similar approaches of this topic. In section 3, we describe the main methods for face detection with the Viola-Jones algorithm, facial feature selection using a binary histogram to model skin tone and facial features tracking using the Camshift algorithm. In section 4 we show the experimental results and the performance of our face tracking system in a video sequence. Section 5 summarizes our work and presents the possibility of extending the system.

3. RELATED WORK

Face tracking is detailed in the following study [1]. The Viola and Jones object detection algorithm [2] and its implementation for face detection is one of the most popular techniques for object detection. It uses the Adaboost machine learning algorithm to train a set of weak classifiers in order to form a cascade of strong classifiers which are very fast and robust. They utilize a rectangular shaped weak classifier called the Haar-like features that are easy to compute using a pre computed image called integral image [2]. Michal Kawulok et al. [3] proposed an approach describing the skin probability maps using bayesian skin modeling. The method consists of analyzing the color histograms of the skin and non-skin pixels. Rachid Belaroussi and Maurice Milgram [4] present two approaches for tracking faces: the connected component segmentation and the coupled Camshift algorithm. Both approaches use a 32 x 32 bins histogram in the H–S plane as a parametric model of the skin color. Then, each pixel value of chrominance (H, S) is passed into a binary histogram, as number of occurrence of each color.

2. FACE TRACKING ALGORITHMS

Computer Vision System Toolbox™ (CVS) provides algorithms and tools for the design and simulation of computer vision and video processing systems. The toolbox includes algorithms for feature extraction, motion detection, object detection, object tracking, stereo vision, video processing, and video analysis [5].

2.1 Detection of faces using the Viola-Jones algorithm

Figure 1 illustrates the Viola and Jones object detection algorithm in two steps [6]: detection of the face in an image region and then applying on it a cascade of boosted classifiers.

The first step detection window is realised by scanning the same image many times, each time with a new size. During the second stage each window is passed through a cascade of classifiers which in its turn is divided in various steps, each step uses a set of weak learners. Each stage is trained to select only the wanted images using a technique called boosting. Boosting has the advantage of training a very accurate classifier to choose a weighted average of the decisions taken by the weak learners.

Each level of the labled region belonging to the classifier is defined as being the current location of the sliding window, being positive or negative. The detector reports a found face in the current location when the final level classifies the region as being positive.

If we have the k classifiers in a cascade, the result of detection rate, D, and false positive rate, F, is given by the product rates on each stage classifier [7]:

$$D = \prod_{i=1}^{K} d_i \qquad F = \prod_{i=1}^{K} f_i \qquad \textbf{(1)}$$

,where d_i is the detection rate of the *i* the classifier in the examples that get through to it and f_i is the false detection rate of the *i* the classifier in the examples that get through to it.

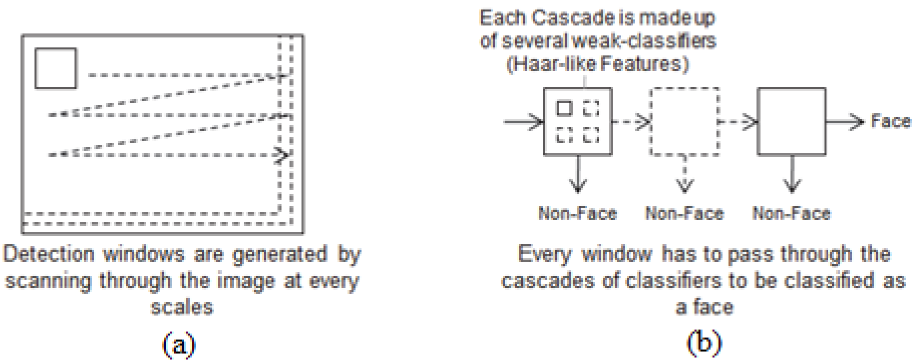


Figure 1. The Viola-Jones Object Detection Algorithm [6]: a) Detection windows ; b) Classified windows

3.1 Face tracking using the Camshift algorithm

CamShift is mainly intended to make efficient head and face tracking in a perceptual user interface [8], it is based on the principles of the MeanShift algorithm. The CamShift algorithm requires the MeanShift one to compute the target centre in the probability distribution of pixels in the image, but also to find the orientation of the principal axis and the dimensions of the probability distribution. There is a difference between the MeanShift and CAMShift algorithms because the first algorithm relies only on a static distribution of pixels in the image, while CAMShift is based on a probability distribution by adjusting the pixel adaptive search window size for the next frame using the zero moment of the current pixel distribution [9]. For target tracking, in addition to calculating the center of mass with mean shift, the CAMShift algorithm uses histogram back projection to associate the pixel values of the image in correspondence with the bin histogram and weighted histogram to remove the effect of the pixels which are outside the selected region because they can change the probability distribution of the image with their frequency. The calculation of the MeanShift parameters is done as follows: the input in this algorithm is the window detected with the Viola-Jones algorithm. Consider the image I(x,y) in two dimensional space and (*x*, *y*) is the position of the face in the image. The moment of order zero (2) represents the area occupied by the shape of the frame.

$$M_{00} = \sum_x \sum_y I(x,y) \qquad \textbf{(2)}$$

The moments of order one (M_{10},M_{01}) are calculated as in (3)

$$M_{10} = \sum_x \sum_y x \times I(x,y)$$
$$M_{01} = \sum_x \sum_y y \times I(x,y) \qquad \textbf{(3)}$$

We find the face's centre of mass by means of the moments of zero order and one, (x_c, y_c) is expressed as in (4)

$$x_c = \frac{M_{10}}{M_{00}};\ y_c = \frac{M_{01}}{M_{00}} \qquad \textbf{(4)}$$

- The CamShift Algorithm

The CamShift algorithm can be written in the following steps [10]:

1. Set the region of interest (ROI) of the probability distribution image to the entire image.
2. Select an initial location of the Mean Shift search window. The selected location is the target distribution to be tracked.
3. Calculate a color probability distribution of the region centred at the Mean Shift search window.
4. Iterate Mean Shift algorithm to find the centroid of the probability image. Store the zeroth moment (distribution area) and centroid location.
5. For the following frame, center the search window at the mean location found in Step 4 and set the window size to a function of the zeroth moment. Go to Step 3.

4. EXPERIMENTS AND RESULTS

We have used a PC system with an Intel Core i3-2100 processor CPU and a 4GB memory. We have tested on two sequence videos on which Matlab 2012b with Computer Vision System Toolbox was installed. Fig. 2 shows the experimental results achieved by using the Viola-Jones algorithm which detects human face in a still image. The algorithm contains a classification model, Frontal Face (CART) which has the following features: the image size is used to train a two-element vector with height = 20 and width = 20, the scale factor is 1.1, and the merge threshold is 4. Fig. 3 shows how to get the skin tone information by extracting the Hue and Saturation from the image converted to the H-S color space. We have used a 16x16 bins histogram in the H-S plane as a model of the skin color.

The first expriment is on a sequence with 150 frames of spatial resolution 640x480. In this video, we will track the face while the person is in motion. Fig. 4a,4d shows face tracking in the frames 66 and 139 of the video sequence using the proposed CAMShift algorithm, figure 4b,4e shows face tracking for the same frames using the MeanShift algorithm and 4c,4f shows face tracking with the Viola-Jones algorithm. Table 1 shows the experimental results obtained with the tests sequences by the methods CAMShift, MeanShift and Viola-Jones algorithm for face tracking.

Table 1. Three methods for tracking face for a sequence with 150 frames of spatial resolution 640x480

Methods for tracking	***Total execution time***	***frame***	***Self time***	***rgb or hs tracking histogram***
MeanShift	115.275 (s)	150	42.497 (s)	67.015 (s)
CAMShift	71.935 (s)	150	15.082 (s)	49.759 (s)
Viola-Jones	14.575 (s)	150	1.099 (s)	7.237 (s)

From the resulting data in Table 1 we conclude that the Viola-Jones algorithm is the fastest because it does not lose much time for calculating the tracking histogram because it is a hs type while for the other two algorithms the calculation of the rgb histogram is much higher. The CAMShift algorithm is faster than Meanshift, but the second has little advantage in accuracy. Self time is the time spent effective in a function. We used a tracker based on a histogram which incorporates the continuously adaptive mean shift (CAMShift) algorithm for object tracking. It uses the histogram of pixel values to identify the tracked face. It begins by initializing the tracker through which to setup the tracked face by extracting it from the selected region from the 2-D input image I, or in this case the image is a hue channel of the H-S color space.

The second experiment is on a sequence with 200 frames of spatial resolution 640x480. In this video, we will track the face, occluded with other objects or human parts with the same skin tone (i.e. hand) . Fig. 5a,5d, shows face tracking in the frames 86 and 163 of the video sequence using the proposed CAMShift algorithm, figure 5b,5e, shows face tracking for the same frames using the MeanShift algorithm and fig. 5c,5f shows face tracking with Viola-Jones algorithm. Table II shows the experimental results obtained with the tests sequences by the proposed method (CAMShift), MeanShift and Viola-Jones algorithm for face tracking while the face is occluded with other objects.

Table 2. Three methods for tracking face in occlusion with other objects for a sequence with 200 frames of spatial resolution 640x480

Methods for tracking	***Total execution time***	***frame***	***Self time***	***rgb or hs tracking histogram***
MeanShift	82.241 (s)	200	3.468 (s)	68.923 (s)
CAMShift	72.285 (s)	200	2.895 (s)	61.772 (s)
Viola-Jones	17.733 (s)	200	1.453 (s)	9.170 (s)

From the resulting data in Table 2 we conclude that the Viola-Jones algorithm is the fastest, but it has very bad results, the proposed method based on the weighted-histogram of the CAMShift algorithm is faster and more acurrate than MeanShift.

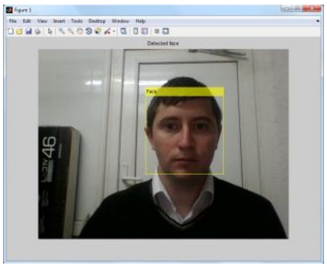


Fig. 2. Face detection with the Viola-Jones algorithm

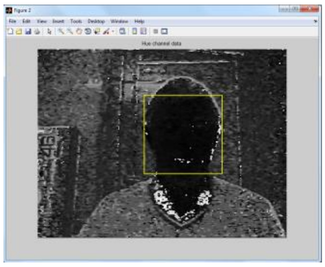
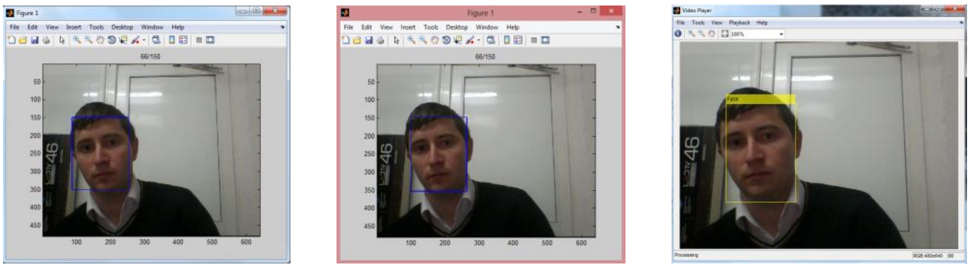
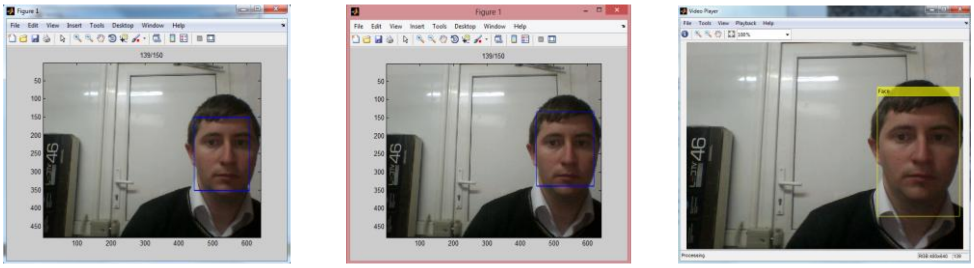


Fig. 3. The corresponding skin map

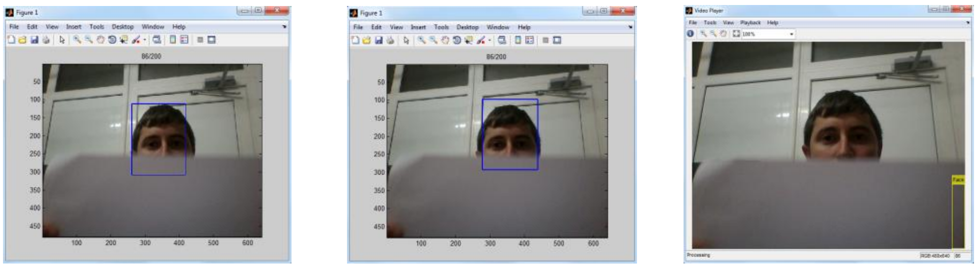


a) b) c)

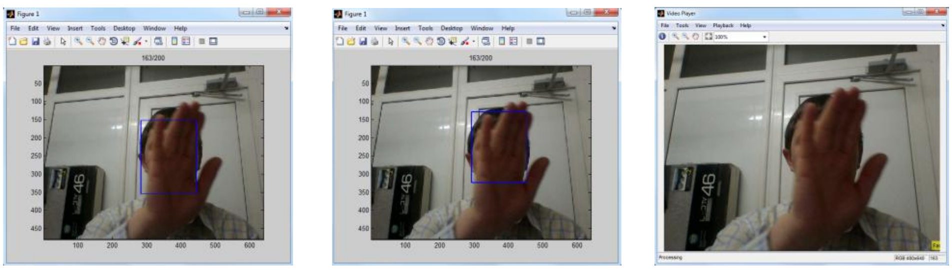


d) e) f)

Fig. 4.Tracking results in frame 66 and 139 from video sequence by the methods:CAMShift (a,d), MeanShift (b,e), Viola-Jones algorithm (c,f)



a) b) c)



d) e) f)

Fig. 5.Tracking results in frame 86 and 163 from video sequence by the methods:CAMShift (a,d), MeanShift (b,e), Viola-Jones algorithm (c,f)

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an efficient approach for face detection and tracking. The 20x20 cascade of boosted classifier has been used for target initialization, and the faces were tracked based on skin color which was modelled using a non-parametric approach. A selection of the skin tone is made with a histogram of 16x16 bins which extracts information from 2D images to the H-S color space. The Viola-Jones algorithm has proven to be the fastest, allow a detection that runs in less than 97 ms/frame on a PC with an Intel Core i3-2100 processor CPU, but concerning face tracking results, they are good only if the face is not occluded by other objects or by a human hand. MeanShift algorithm detects and tracks the human face even if it is occluded, its disadvantage is the slow execution time. The proposed method based on weighted-histogram of the CAMShift algorithm is faster than MeanShift, follows the face even if it is occluded by objects or by the human hand and it is also more accurate than MeanShift.

As future work the same algorithm will be extended to recognize human facial expressions in a real-time system which will be implemented on a mobile robot and on a car.

REFERENCES

[1] Cha Zhang and Zhengyou Zhang. A survey of recent advances in face detection. Technical report, 2010.

[2] P. Viola, and M. Jones, "Robust Real-Time Face Detection", International Journal of Computer Vision, Vol. 57, pp. 137-154, (2004).

[3] Michal Kawulok, Jolanta Kawulok, Jakub Nalepa, Spatial-based skin detection using discriminative skin- presence features, Pattern Recognition Letters, Volume 41, 1 May 2014, Pages 3-13, ISSN 0167-8655.

[4] Rachid Belaroussi, Maurice Milgram, A comparative study on face detection and tracking algorithms, Expert Systems with Applications, Volume 39, Issue 8, 15 June 2012, Pages 7158-7164, ISSN 0957-4174.

[5] Information on http://www.mathworks.com/help/vision/gs/product-description.html

[6] Yasir Mohd Mustafah, Amelia Wong Azman, Out-of-Plane Rotated Object Detection using Patch Feature based Classifier, Procedia Engineering, Volume 41, 2012, Pages 170-174, ISSN 1877-7058.

[7] Modesto Castrillón, Oscar Déniz, Daniel Hernández, Javier Lorenzo, A comparison of face and facial feature detectors based on the Viola–Jones general object detection framework, Volume 22, Issue 3, May 2011, Pages 481-494, ISSN 0932-8092.

[8] Bradski, G. (1998). Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal, Q2(15).

[9] John G. Allen, Richard Y. D. Xu, and Jesse S. Jin, Object tracking using CamShift algorithm and multiple quantized feature spaces, In Proceedings of the Pan-Sydney area workshop on Visual information processing (VIP '05), ISBN 1-920682-18-X, pp. 3-7, 2004.

[10] Intel Corporation (2001): Open Source Computer Vision Library Reference Manual, 123456-001.