Real-Time Feature Based Face Detection and Tracking I-CURSOR

Sunita Barve, Dhaval Dholakiya, Shashank Gupta, Dhananjay Dhatrak

Abstract-This project aims to present an application that is able of replacing the traditional mouse with the human face as a new way to interact with the computer. Facial features (nose tip and eyes) are detected and tracked in real-time to use their actions as mouse events. In our work we were trying to compensate people who have hands disabilities that prevent them from using the mouse by designing an application that uses facial features (nose tip and eyes) to interact with the computer. It can be applied to a wide range of face scales. Our basic strategy for detection is fast extraction of face candidates with a Six-Segmented Rectangular (SSR) filter and face verification by a support vector machine[4][5]. A motion cue is used in a simple way to avoid picking up false candidates in the background. In face tracking, the patterns of between-the eyes are tracked with updating template matching.

1. Introduction

In the past few years high technology has become more progressed, and less expensive. With the availability of high speed processors and inexpensive webcams, more and more people have become interested in real-time applications that involve image processing. One of the promising fields in artificial intelligence is Human Computer Interface which aims to use human features (e.g. face, hands) to interact with the computer. One way to achieve that is to capture the desired feature with a webcam and monitor its action in order to translate it to some events that communicate with the computer. In our work we were trying to compensate people who have hands disabilities that prevent them from using the mouse by designing an application that uses facial features (nose tip and eyes) to interact with the computer.
The nose tip was selected as the pointing device; the reason behind that decision is the location and shape of the nose; as it is located in the middle of the face it is more comfortable to use it as the feature that moves the mouse pointer and defines its coordinates, not to mention that it is located on the axis that the face rotates about, so it basically does not change its distinctive convex shape which makes it easier to track as the face moves. Eyes were used to simulate mouse clicks, so the user can fire their events as he blinks. While

different devices were used in HCI (e.g. infrared cameras,

- *Mrs Sunita Barve is currently inMaharashtra Academy Of Engineering as an Assistant Professorof Dept. of Computer Engineering in University of Pune, India, PH-918149688922. E-mail: barve_sunita@rediffmail.com*
- *Shashank Gupta is currently pursuing Bachelor's degree program in computer engineering in Universityof Pune,India, PH-918149992748. E-mail:2shahji@gmail.com*

sensors, microphones) we used an off-the-shelf webcam that affords a moderate resolution and frame rate as the capturing device in order to make the ability of using the program affordable for all individuals. We will try to present an algorithm that distinguishes true eye blinks from involuntary ones, detects and tracks the desired facial features precisely, and fast enough to be applied in real-time.
We have aimed to design an application that uses facial features (nose tip and eyes) to interact with the computer. In this application, Facial features (nose tip and eyes) are detected and tracked in real-time to use their actions as mouse events. The coordinates and movement of the nose tip in the live video feed are translated to become the coordinates and movement of the mouse pointer on the user's screen. The left/right eye blinks fire left/right mouse click events. The only external device that the user needs is a webcam that feeds the program with the video stream.
In our work we are trying to compensate people who have hand disabilities that prevent them from using the mouse.

2. Related Work

Most previous approaches to facial feature tracking utilize skin tone based segmentation from single camera exclusively (Yang & Waibel, 1996; Wu et al., 1999; Hsu et al., 2002; Terrillon & Akamatsu, 1999; Chai & Ngan, 1999). However, color information is very sensitive to lighting conditions, and it is very difficult to adapt the skin tone model to a dynamically changing environment in real-time Kawato and Tetsutani (2004) proposed a mono camera based eye tracking technique based on six-segmented filter (SSR) which operates on integral images (Viola & Jones, 2001)[7].

3. Feature Based Face Localization Method

In contrast to the knowledge-based methods, research has been done to find invariant features of faces for detection.

The idea came from that human could easily detect faces and objects even if the illumination condition and poses changed. So there must exist properties or features that remain unchanged over different situations, or, could change in a predictable way. Feature invariant approaches search for face structural features that are invariant to changes in pose, viewpoint, illumination, and expression. An example of largely adopted feature is the skin color, several works [8], [9], [10] suggest modeling the skin color distribution with a Gaussian mixture model. Other facial features such as forehead, eyebrows, eyes, nose, cheeks and mouth also extracted as invariant features using edge detectors. The face image is usually divided into small regions that contain the extracted invariant features and a statistical model is built. As a result, the feature of each region, together with relationships between these regions suggests different facial expressions, illumination condition, viewpoints, etc. The drawback of this kind of method is that image features could be severely destroyed due to bad illumination condition, noise, and other occlusion, and the boundaries between features could be too weak to detect while the shadows could produce strong fake edges. Image invariants can be designed to fit the needs of specific systems. Some require only that it be non-discriminating to an object's geometric pose or orientation. Others may be only interested in it being insensitive to the change of illumination. More complex systems however demand that it be insensitive to a combination of several environmental changes. Clearly the latter case is more difficult to achieve.

3.1 SSR Filter

For face candidate extraction, a rectangle is scanned on the input image. The rectangle is segmented into six parts as in Fig. 2. We denote an average pixel value within a segment S_i as S_i. Then, when one eye and eye brow are within S_3, we can expect

$$S_1 < S_2 \text{ and } S_1 < S_4 \qquad (1)$$
$$S_3 < S_2 \text{ and } S_3 < S_6 \qquad (2)$$

A point where (1) and (2) are satisfied can be a face candidate. We call this an SSR filter [4].

The proposed SSR filter, which is the rectangle divided into 6 segments as shown in Fig.2 (a), operates by using the concept of bright-dark relation around Between-the-Eyes area as explained by Fig.2 (b) and (c). We select Between-the-Eyes as face representative because it is common to most people and easy to find for wide range of face orientation [11].

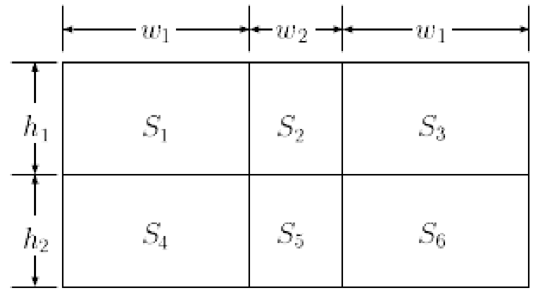

Figure 1: Six-Segmented Rectangular (SSR) Filter. Average pixel values in each segment are computed and compared with each other to find whether they satisfy certain conditions

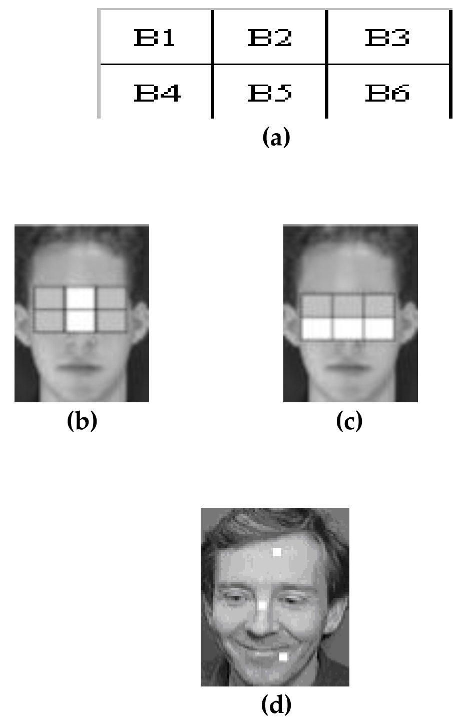

Fig 2: Concept of SSR Filter: (a) the proposed rectangular filter divided into six segments. (b) Nose area is brighter than right and left eye area. (c) Eye area is relatively darker than cheekbone area. (d) Example of the extracted Between-the-Eyes candidates.

4. Face ROI Localization

In general, face tracking approaches are either image based or direct feature search based methods [6]. Image based (top-down) approaches utilize statistical models of skin color pixels to find the face region first, accordingly pre-stored face templates or feature search algorithms are used to match the candidate face regions as in Chiang et al. (2003). Feature based approaches use specialized filters directly such as templates or Gabor filter of different frequencies and orientations to locate the facial features. Our work falls into the latter category. That is, first we find the eye candidate locations employing the integral image technique and the six segmented rectangular filter (SSR) method with SVM. Then, the similarities of all eye candidates are verified using the stereo system. The convex curvature shape of the nose and first and second derivatives around the nose tip are utilized for the verification. The nose tip is then utilized as a reference for the selection of the mouth ROI. At the current implementation, the system tracks the person closest to the camera only, but it can be easily extended to a multiple face tracking algorithm.

4.1 Eye Tracking

The pattern of the between the eyes are detected and tracked with updated pattern matching [1]. To cope with scales of faces, various scale down images are considered for the detection, and an appropriate scale is selected according to the distance between the eyes (Kawato and Tetsutani, 2004). The algorithm calculates the intermediate representation of the input image called "Integral image", described in Viola & Jones (2001). Then, a SSR filter is used for fast filtering of bright-dark relations of the eye region in the image. Resulting face candidates around the eyes are further verified by perpendicular relationship of nose curvature shape as well as the physical distance between the eyes, and eye level and nose tip [3].

4.2 Nose Bridge and Nose Tip Tracking

The human nose has a convex curvature shape and the ridge of the nose from the eye level to the tip of the nose lies on a line as depicted in Fig. 1. Our system utilizes the information in the integral intensity profile of convex curvature shape. The peak of the profile of a segment that satisfies (1) using the filter shown in Fig.2 is the convex hull point. A convolution filter with three segments traces the ridge with the center segment greater than the side segments, and the sum of the intensities in all three segments gives a maximum value on the convex hull point. Fig.2 shows an example filter with three segments that traces the convex hull pattern starting from the eye line. The criteria for finding the convex hull point on an integral intensity profile of a row segment is as follows,

$$S_1 < S_2 < S_3 \text{ and } \arg\{\max_j(S_1+3\,S_2+S_3)\}$$

Where S_i denotes the integral value of the intensity of a segment in the maximum filter, and j is the center location of the filter in the current integral intensity profile. The filter is convolved with the integral intensity profile of every row segment [1]. A row segment typically extends over 5 to 10 rows of the face ROI image, and a face ROI image typically contains 20 row segments. Integral intensity profiles of row segments are processed to find their hull points (see Fig.3 using (1) until either the end of the face ROI is reached or until (1) is no longer satisfied. For the refinement process, we found that the first derivative of the 3D surface data as well as the first derivative of the intensity at the nose tip is maximum, and the second derivative is zero at the nostril level [2].

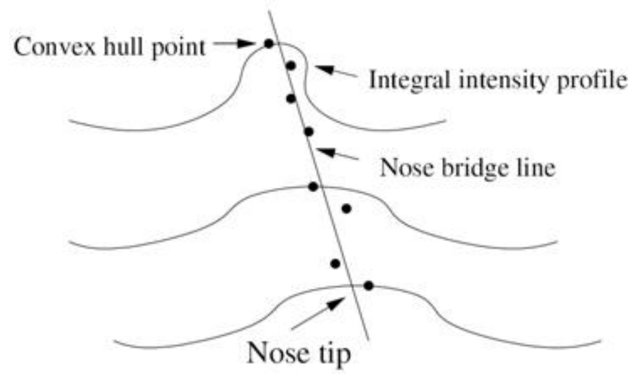

Fig. 3: Nose bridge line using its convex hull points from integral intensity projections.

5. Future Works

Feature works may include improving the tracking robustness against lighting conditions; perhaps by using more sophisticated and expensive capturing devices such as infrared cameras that can operate in absence of light and give more accurate tracking results.Adding the double left click (detecting the double left eye blink) and the drag mode (enabling/disabling with the right double eye blink) functionalities.

Adding voice commands to launch the program, start the detection process, and to enable/disable controlling the mouse with the face.

6. Conclusion

We are aimed to implement scale-adaptive face detection and tracking system using JAVA (J2ME) for face candidate detection, a six-segmented rectangle. (SSR) filter is scanned over the entire input image. This approach is similar to the window-scanning technique often used in the image-based approach. However, once the bright-dark relations between the six segments indicate a face candidate, eye candidate and nose tip regions are searched in the manner of the feature-based approach. Then, based on the locations of a pair of eye candidates and nose tip, the scale, orientation and gray levels are normalized.

The use of Java Millennium Edition helps it to be compatible even with pocket devices so, hopefully every new gadget or electrical appliance could be used by handicaps in future.

References

[1] Chiang, C. C., Tai, W. K., Yang, M. T., Huang, Y. T. & Huang, C. J., (2003). *A novel method for detecting lips, eyes and faces in real-time. Real-Time Imaging* 9, 277-287.

[2] Gurbuz, S., Kinoshita, K., & Kawato, S., (2004a).Real-time human nose bridge tracking in presence of geometry and illumination changes. *Second International Workshop on Man-Machine Symbiotic Systems, Kyoto, Japan.*

[3] S. Kawato and J. Ohya. Two-step approach for real-time eye tracking with a new filtering technique. *Proc. Int. Conf. on System , Man & Cybernetics*, pages 1366–1371, 2000.

[4] O. Sawettanusorn, Y. Senda, S. Kawato, N. Tetsutani, and H. Yamauchi. Real-time face detection using six-segmented rectangular filter. accepted in *ISPAC 2003*.

[5] Real-Time Face Detection Using Six-Segmented Rectangular Filter (SSR Filter) http://www.design-reuse.com/articles/6899/real-time-face-detection-using-six-segmented-rectangular-filter-ssr-filter-real-time-face detection-using-six-segmented-rectangular-filter-ssr-filter.html

[6] Real-Time Face Detection System http://www.mint.se.ritsumei.ac.jp/project/COE/ FirstSymposium/proc/Yamauchi- FaceDetect.pdf

[7] P.Viola and M.Jones, "*Rapid object Detection using a Boosted Cascade of Simple Features*," Conf.CVRP,1, pp.511-518,2001.

[8] H. Wu, Q. Chen, and M. Yachida. Face detection from color images using a fuzzy pattern matching method. *"IEEE Transactions Pattern Analysis and Machine Intelligence*", 21:557–563, 1999.

[9] S. McKenna, S. Gong, Y. Raja. "*Face recognition in dynamic scenes"*. British Machine Vision Conference, Essex, 1997.

[10] S.Kawato and N. Tetsutani, *"Real-time detection of Between-the-Eyes with a Circle Frequency Filter,"* ACCV2002: The 5th Asian Conference on Computer Vision, 23-25 January 2002, Melbourne, Australia.

[11] J.C. Terrillon, M.N. Shirazi, H. Fukamachi. "*Comparative performance of different skin chrominance models and chrominance for the automatic detection of human faces in color images"*. Proceedings of the IEEE International conference of Face and Gesture Recognition, pages 54–61, 2000.