Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Network and Computer Applications 29 (2006) 62–80
Journal of NETWORK and COMPUTER APPLICATIONS
www.elsevier.com/locate/jnca

Facial feature tracking for cursor control

T. Morris*, V. Chauhan

Department of Computation, UMIST, P.O. Box 88, Manchester, M60 1QD, UK

Received 19 November 2003; received in revised form 12 July 2004; accepted 21 July 2004

Abstract

This work is motivated by the goal of providing a non-contact means of controlling the mouse pointer on a computer system for people with motor difficulties using low-cost, widely available hardware. The required information is derived from video data captured using a web camera mounted below the computer's monitor. A colour filter is used to identify skin coloured regions. False positives are eliminated by optionally removing background regions and by applying statistical rules that reliably identify the largest skin-coloured region, which is assumed to be the user's face. The nostrils are then found using heuristic rules. The instantaneous location of the nostrils is compared with their at-rest location; any significant displacement is used to control the mouse pointer's movement. The system is able to process 18 frames per second at a resolution of 320 by 240 pixels, or 30 fps at 160 by 120 pixels using moderately powerful hardware (a 500 MHz Pentium III desktop computer).

Keywords: Human–computer interaction; Perceptual interfaces; Face tracking; Enabling technologies

1. Introduction

For many people with physical disabilities, computers form an essential tool for communication, environmental control, education and entertainment. However, access to the computer may be made more difficult by a person's disability. A number of users employ head-operated mice or joysticks in order to interact with a computer and to type

* Corresponding author. Tel.: +44-161-200-3376; fax: +44-161-200-3324.
E-mail addresses: t.morris@co.umist.ac.uk (T. Morris), dtm@co.umist.ac.uk (T. Morris).

1084-8045/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2004.07.003

with the aid of an on-screen keyboard. Head-operated mice can be expensive. In the UK, devices that require the users to wear no equipment on their heads, other than an infrared reflective dot, for example Orin Instrument's HeadMouse (Orin Instruments, 2001) and Prentke Romich's HeadMaster Plus (Prentom, 2001), cost in excess of £1000 (€1400). Other devices are cheaper, notably Granada Learning's Head Mouse (Drew et al., 1998; Evans et al., 2000; Evans and Blenkhorn, 1999) (£200, €280), and Penny and Gilles' device (£400, €560). However, these systems require the user to wear a relatively complex piece of equipment on their head, an infrared transmitter and a set of mercury tilt switches, respectively.

The goal of this research is to develop and evaluate a non-contact interface that is both low cost and also does not require the user to wear any equipment. The most obvious way of doing this is to use a camera, interfaced to a PC. The PC/camera system will track the movement of the user; it will translate head movements into cursor movements, and could interpret changes in expression as button presses.

There have been many attempts at creating cursor control systems utilising head movements; these are reviewed below. The head monitoring system must also provide a mechanism for replacing the mouse buttons. We propose that this can be achieved by using the blinks of the eyes (or by dwelling the mouse or by using switches). Changes in expression could also be used as part of a separate device to provide a computer interface for more severely disabled persons.

A system such as the one proposed will confer significant benefits on its target audience. It will increase their independence by facilitating their communication with a computer and hence providing an avenue for communication and environmental control. The proposed interface may also have an application in the non-contact control of remotely controlled units for the able bodied; providing a perceptual control mechanism for such a unit will free the operator to concentrate on other demanding tasks.

The implementation of a system such as the proposed one presents several areas of difficulty:

1. Identifying and tracking the head location.
2. Identifying and tracking the location of a facial feature.
3. Being able to process the information in real-time using a moderately priced processor that will be running other applications in the foreground (for example, Microsoft Word).

Yang et al. (1999) presented a review of face detection methods, which fell into a small number of categories, the most important of which were methods depending of the colour and physiognomy of the face.

Störring et al. (1999) suggested that a face's apparent colour was due to two factors: the amount of melanin in the skin and the ambient illumination. Of the two, ambient illumination caused the greater variation to the perceived colour. They concluded that if normalised colour values were used (i.e. the effect of illumination were removed) then skin colours were consistently with a fixed and quite narrow set of limiting values, independent of the subject's natural skin colouration. This approach, and slight modifications, has become very popular due to its simplicity (Störring et al., 1999; Yang and Ahuja, 1998;

Toyama, 1998; Sobottka and Pittas, 1996a,b; Yang and Waibel, 1996; Qian et al., 1998; Menser and Brünig, 1999; Jie et al., 1998; Schiele and Waibel, 1995; Zitnick et al., 1999).

Template matching provides an attractive and simple approach to face detection that relies on the constant appearance of the facial features, but often presents difficulty in dealing with variations in scale, shape and pose. The simplistic approach would define a template that resembled a facial feature and cross correlate it with a face image. The location of the maximum response defines that feature's location. However, the response decreases as the resemblance between the template and the facial region diminishes, and in theory, a template is required for all different instances of the feature. To overcome this, flexible templates were originated, for example by Yuille et al. (1992), where templates of facial features are deformed to match against the features of the face. An energy function is used to link edges, peaks and valleys contained in the input image to corresponding values in the parameterised template. The best fit of the model is found by altering the parameter values—performing energy minimisation. The deformable template matching method goes some way to achieving scale and shape invariance, also proposed by the use of multiscale and multiresolution templates. Variations on this theme are widespread, (Orin Instruments, 2001; Yang et al., 1999; Yang and Ahuja, 1998; Sobottka and Pittas, 1996; Menser and Brünig, 1999; Brunelli and Poggio, 1993, Bakic and Stockman, 1999; Nikolaidis and Pitas, 2000; Morris et al., 2002).

To summarise, it is the aim of the research reported here to develop a cheap, non-contact computer interface that will be used primarily by people with severe motor difficulties. The system should require minimal initialisation/configuration.

The remainder of the paper is organised as follows. In Section 2 the options for capturing the required data are reviewed. The characteristics of the selected hardware are reviewed, as they relate to the problem we are addressing. Section 3 presents an overview of the software architecture of the system we have developed. Sections 4 and 5 describe the two parts to the face detection algorithm: identifying possible face candidates and eliminating false positives, and identifying the extent of the face in an image. Section 2 describes how we identify the facial features whose instantaneous location is used to control the cursor's movements. The translation from feature location to cursor movement is described in Section 7. In Section 8 we present an evaluation of the system: how accurate is it and how rapidly can it process data. Finally, we draw conclusions in Section 9.

2. Data capture

As we are developing a non-contact interface that will not require the user to wear any kind of transducer, video imagery is the obvious data to process. There are two types of video capture interfaces to the PC: the standard video camera plus video capture hardware and digital cameras that interface directly to the system, that is Firewire or USB cameras. We discarded the video camera on the grounds of its cost, even though the image quality is likely to be better, and instead chose to investigate the use of a webcam interfaced via the USB, reasoning that this hardware is likely to be in the possession of a large proportion of computer users.

Fig. 1. Sample image captured using a Creative Labs Webcam Go.

The USB has an upper throughput rate of 12 Mb/s. Combinations of image spatial resolution, colour depth and frame rate must not exceed this. It has been shown that a minimum rate of 10 frames per second must be processed in order for a user to perceive real-time operation (Stroud, 1956). Therefore, each frame of data cannot exceed 1.2 Mb. We also consider that colour information is of primary importance and therefore require 24 bit colour information, which further reduces the frame size to be 0.4 million pixels or less. The largest 'standard' size image satisfying this criterion is the SIF resolution (320 by 240 pixels NTSC and 353 by 288 PAL). Fortunately it has been shown in many studies that satisfactory results may be obtained by processing images with this resolution, smaller images contain insufficient detail.

Fig. 1 shows a typical image captured during facial feature tracking, its spatial resolution is certainly sufficient to be able to detect the significant facial features, even if they cannot be outlined with any great degree of precision (if fact, this is more of an advantage than a problem, as will be shown below).

This single image fails to reveal the temporal nature of the data. We would expect the image of a uniform scene to be static over time. This was investigated by capturing sequences of images of a uniform scene (a Kodak midgrey test card) with the automatic gain control of the webcam turned on and off. For comparison, the equivalent data was captured using a Sony HyperHad camera interfaced to a Silicon Graphics workstation. The differences between the red, green and blue values of equivalent pixels in consecutive frames were computed. Table 1 presents the results,

Table 1
Image Quality—inter frame pixel differences

	Webcam		Sony HyperHad
	AGC on	AGC off	
Mean *R*	3	3	1
Mean *G*	2	2	1
Mean *B*	4	3	2
Average mean	2	1	1
Max *R*	37	34	22
Max *G*	29	22	18
Max *B*	48	41	37
Average max	24	20	16

showing the averages of these values (Mean *R*, Mean *G*, Mean *B*), averaged over all frames and the whole sequence; the average of these averages (Average mean); the maximum differences (Max *R*, Max *G*, Max *B*) and the average maximum difference (Average max).

As expected, the webcam's image quality is worse than a standard video camera's but not significantly worse.

The camera was further tested by capturing images of two people (Caucasian and Asian) under three different lighting conditions: afternoon sunlight, tungsten lighting and halogen lighting. The results are shown in Fig. 2.

According to Storring's results (Störring et al., 1999) we would expect quantitatively similar changes in the images as the lighting conditions are altered. This does not happen, and is especially obvious in the daylight images in which the Caucasian skin takes a bluish tinge. It is most likely that this deviation from Storring's result is due to the quality of the cameras used in these two studies. The practical consequence is that we would not expect to be able to specify universal skin colour ranges for the data captured by this camera. Rather, any system that will rely on colour to identify potential skin regions must be calibrated for each user and probably also for each use of the system.

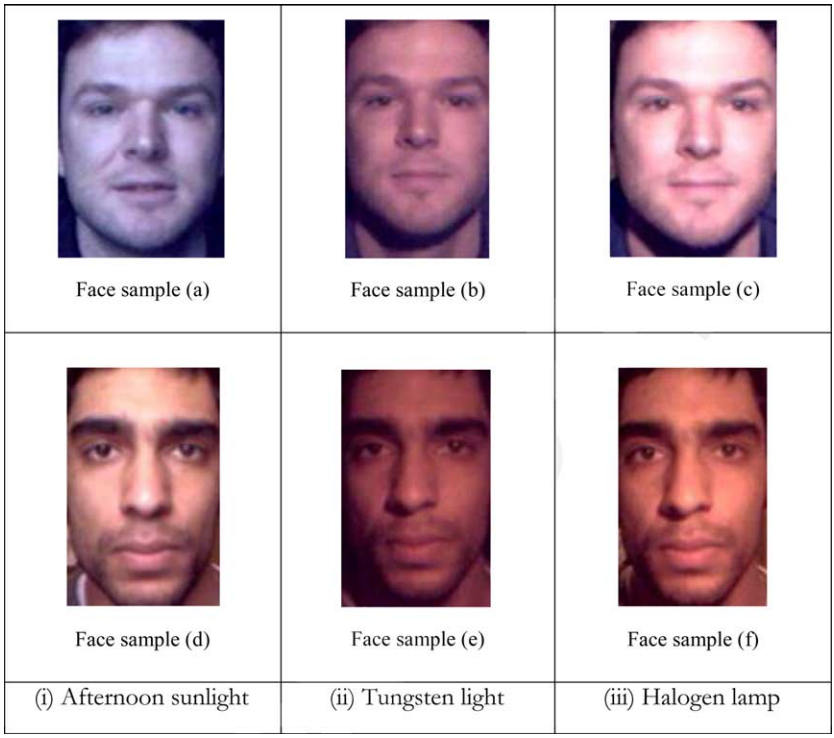

Fig. 2. Images of an Asian and a Caucasian captured under different illumination conditions using the Webcam Go.

We are developing a system that will be used as a non-contact computer interface. It is therefore obvious that the video camera will be situated such that it captures images of the computer user's face as he or she views the computer's monitor. We have chosen to place the camera below the monitor, pointing upwards at the user.

The software architecture of the system will now be described.

3. Software architecture

A diagram illustrating an overview of the system's operations is presented in Fig. 3. Note that this diagram shows the system's steady state operation, results from processing the previous frame inform the processing of the current frame, and the results of processing this frame will inform processing the following one.

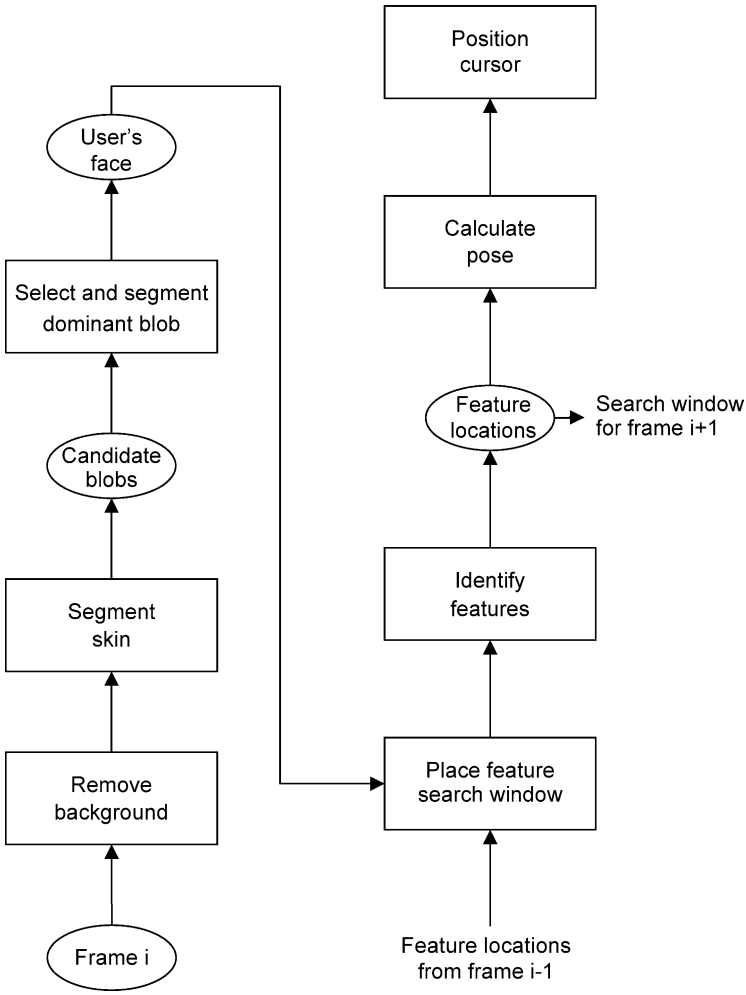

Fig. 3. Data flow diagram.

The system is divided into four components that are responsible for identifying blobs that might correspond to the face in an image, selecting and delineating the blob that corresponds to the face, identifying the facial features that are to be tracked and moving the cursor. The four components will be described in the following sections.

4. Face candidate identification

We have chosen to use colour information to identify the face. Our aim was to use the simplest reliable algorithm that will satisfy our requirements. We dismissed other methods of face detection as we deemed them to be overly complex for this application, and although they might yield correct results at an acceptable frame rate, we do not believe that they will leave sufficient processing resources to allow any useful software to be used without an unacceptable degradation in performance.

Storring's contention is that skin colours are compactly clustered, the size and shape of the cluster is unaffected by changes in illumination: only the location of the cluster changes. We investigated methods of removing the illumination dependence by investigating alternative colour spaces.

Facial images were downloaded from the University of Stirling (University of Stirling Face Database 2003). This database contains a demographically representative sample of full frontal face images captured under varying lighting conditions. Regions of each image containing skin only were manually extracted and the red, green and blue (*RGB*) components of each pixel recorded.

The *RGB* values were converted to normalised red, green and blue, *rgb*; log-opponent, and Y, C_r and C_b:

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B} \tag{1}$$

$$L(x) = 105 \log_{10}(x+1) \quad I = L(G) \quad R_g = L(R) - L(G)$$
$$B_y = L(B) - \frac{L(G)+L(R)}{2} \tag{2}$$

$$Y = 0.30R + 0.59G + 0.11B \quad C_r = 0.50R - 0.42G - 0.08B$$
$$C_b = -0.17R - 0.33G + 0.50B \tag{3}$$

Illumination independence was achieved by deleting any one of the *rgb* components, the I and the Y component.

Plots of the normalised skin values are shown in the scattergrams of Fig. 4a (plotting r and g), Fig. 4b (R_g and B_y) and Fig. 4c (C_r and C_b). To be considered suitable for this application, the points must be tightly clustered. It is also advantageous for the normalisation to be computationally simple, to enhance the data throughput.

Inspection of the scattergrams reveals that the normalised red, green and blue representation gives the tightest clustering, this representation was therefore chosen.

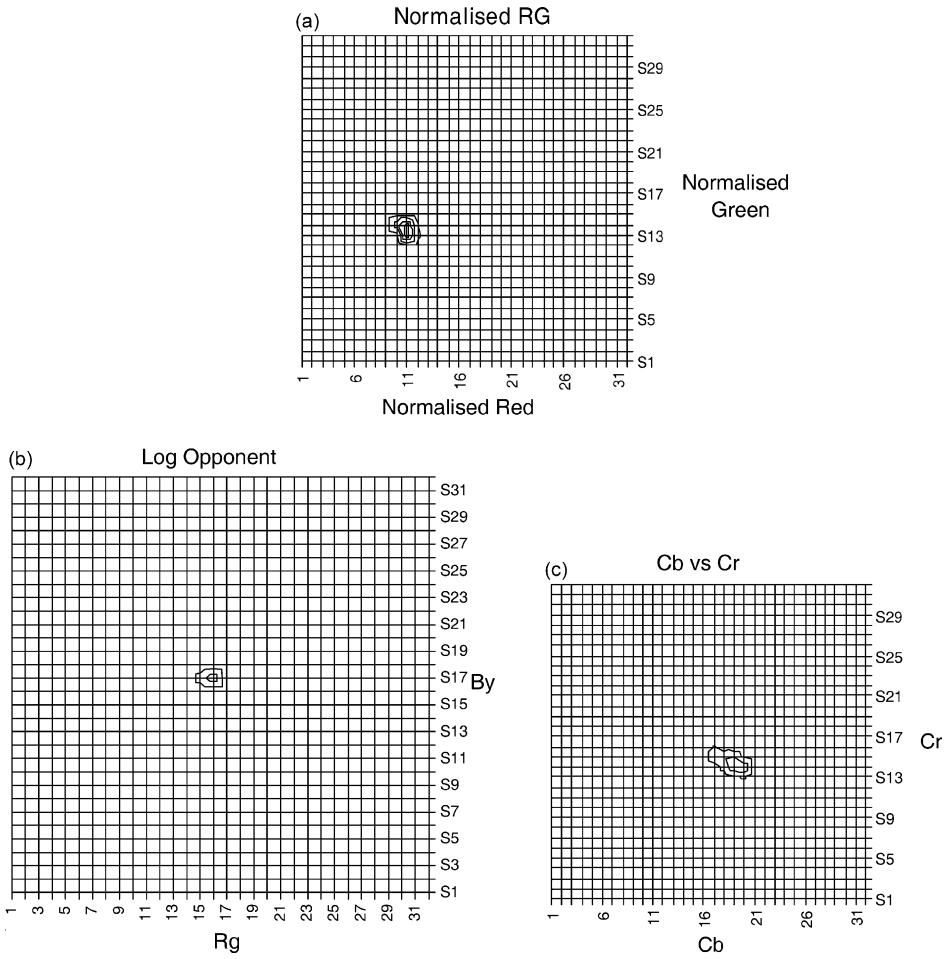

Fig. 4. Clustering of skin colours in various colour spaces, (a) normalised red green, (b) normalised log-opponent, (c) C_r and C_b.

We have demonstrated that it is possible to identify a range of normalised colours that include all skin colourations. Of course, these colours will not be unique to skin samples, the colours of other objects will also be found in this range. We have found that certain colours of paint and especially woodwork can give false positive results using this algorithm. These pixels are readily removed using background subtraction.

A background image may be acquired during a calibration phase, this image will be the view observed by the camera in its usual position, but without the user in place. During operation, each frame of data is compared to the background image. If pixels are sufficiently different, then the pixel is considered to be part of the foreground, that is the object being tracked and it is passed to the colour matching algorithm. Of course,

if the background to the scene is carefully controlled, it will not contain any objects of potentially confusing colours, and the background subtraction will not be required. Nevertheless, we have included this operation as an optional step as it can improve face detection in some environments.

The output of this stage of the processing cycle is an image map that indicates those pixels that are of a colour consistent with the skin colour model and are not part of the static background. Despite removing the background pixels, the map still contains a number of false positive results caused primarily by the poor quality data delivered by the camera. These are removed as a side effect of the following stage of the processing cycle, Face Region Growing, whose primary goal is to identify the region of the image that corresponds to the face.

5. Face region growing

Due to the location of the camera, the user's head and shoulders will be the dominant object in the images that are captured. Following the previous stage, the image map will contain a large region corresponding to the face (although this is occasionally fragmented due to shadowing effects) and a large number of much smaller regions due to image noise and imperfect colour matching. Fig. 5 represents a sample image map and the input that was used to generate it.

Connected component analysis can be used to identify the different contiguous groups of pixels (blobs), and determine their sizes. It is therefore a simple matter to select the largest blob and assume that it is the face. However, this algorithm takes no account of the blob's shape and can give erroneous results, Fig. 6.

Radial spanning (Toyama, 1998) has also been suggested as a means of detecting convex objects, by growing radial spokes from a seed point and requiring that adjacent spokes are of similar lengths.

Both of these methods require a seed point within the blob to be identified. Although this is a simple matter as any point can be used as a seed, it is a time consuming process to examine all blobs in this fashion.

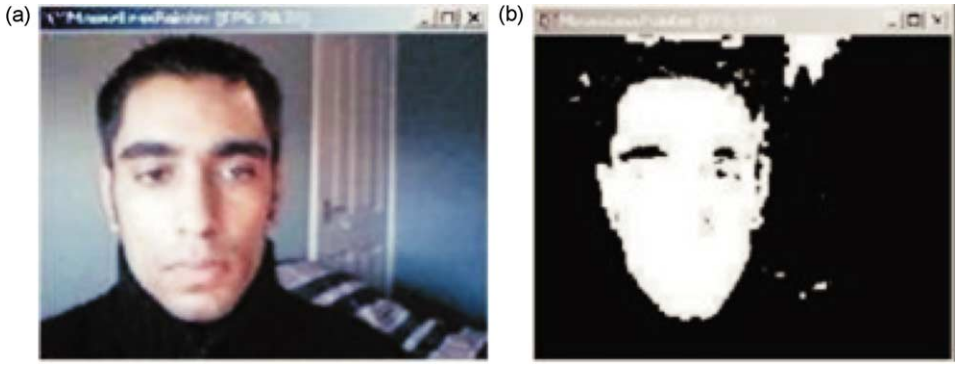

Fig. 5. (a) Typical input, (b) skin map generated by colour matching.

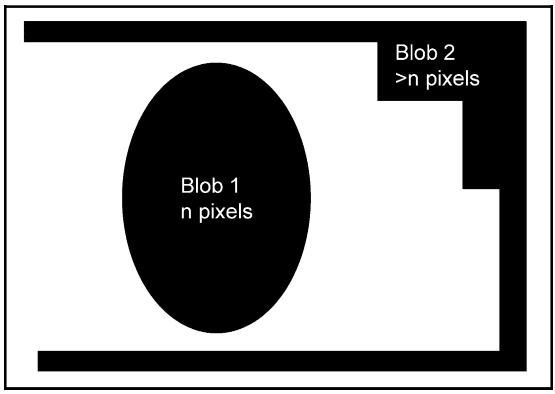

Fig. 6. An example of connected component analysis giving incorrect results.

We have used robust statistical analysis (Qian et al., 1998) to identify the dominant blob in the image. This is an iterative process that consistently identifies that centre of the largest blob in the image.

Two one dimensional histograms, $h_x(i)$ $0 \leq i <$ NumRows, and $h_y(j)$ $j \leq j <$ NumCols, were created by summing the image's columns and rows. The means and standard deviations of these were computed in the usual ways. The ranges of the summations can be restricted so as to exclude blobs neighbouring the image boundaries, thus it is possible to prevent the error shown in the example image. The two means give an initial estimate of the centre of the dominant blob in the image which is refined by repeating the calculation of the means and standard deviation but using only those pixels lying within one standard deviation of the mean in each distribution. This step was repeated until the change in the values of the two means was a negligible. The effect of this process is to locate the centre of the most dominant blob in the image, irrespective of any other blobs that might be present, Fig. 7.

The values of the standard deviations can be used to compute an approximate bounding box for the face region. We may also perform a connected component analysis to identify the region more accurately. Since the purpose of locating the face is to define a region to search for the facial features we are tracking, we do not require that the face is accurately delimited. Instead, we are satisfied with the more rapid, but less accurate process of defining the search region using the two standard deviations. Fig. 8 illustrates the regions identified using radial spanning and robust statistical analysis.

The region defined by this process is passed to the feature location module.

6. Feature location

To determine face pose would normally require that three non-collinear facial features were identified, such as the eyes and nose. Recognising that facial features are usually darker than the surrounding skin, some authors have used local minima in the brightness function to define feature locations, regardless of whether these locations correspond to

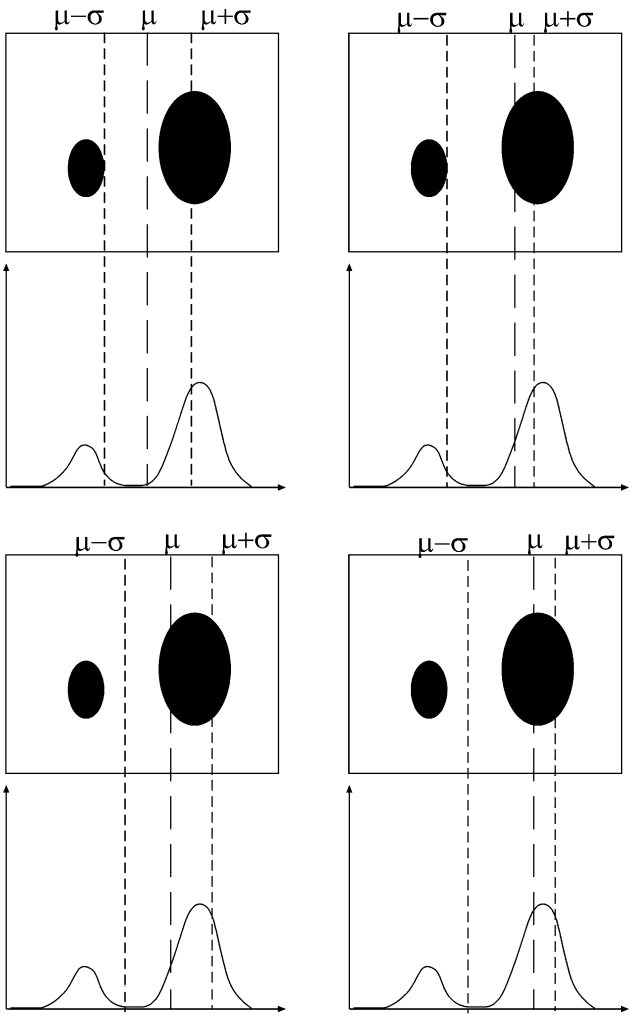

Fig. 7. Dominant blob selection. The rows and columns of the image are summed. The arrays of row sums and column sums are treated in the same manner. Firstly the mean and standard deviation is computed. Secondly the mean and standard deviation of values lying within one standard deviation of the original mean is computed. This calculation is iterated until the mean converges, at which time the means derived from the row and column sums define the co-ordinates of a point within the dominant blob.

a physical feature. It is also difficult to identify features consistently as the head pose is altered, as the head is bowed, the shadows under the eyebrows become darker, which can confuse some feature trackers.

We have chosen to track the user's nostrils and use their position with respect to the edges of the search region input to this module to define the head pose. The nostrils confer two major advantages to a simple tracking system. Firstly, they are clearly separated from any other features that could be confused with them. Secondly they are relatively small

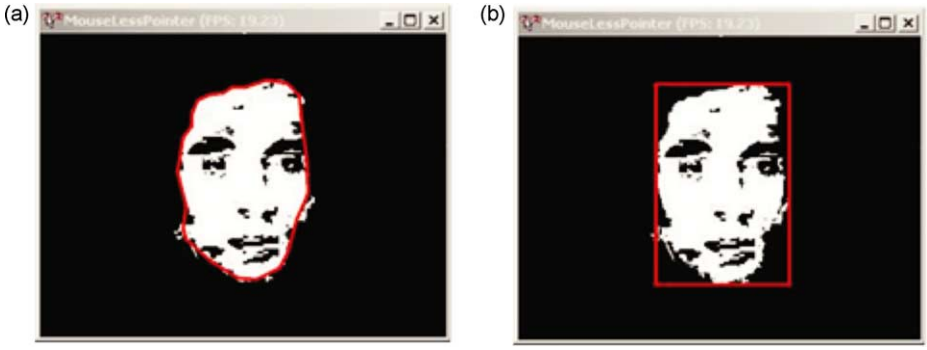

Fig. 8. Face regions detected by radial spokes and robust statistical analysis.

and situated away from the face boundary, this means that they remain visible even under extreme facial poses.

The window passed to this module defines the face region. The central third of this is used as a search region for the nostrils, this is large enough that we can be sure it contains the nostrils and small enough to allow rapid processing. The nostrils are located by a thresholding process; the data within the search region is thresholded with a gradually increasing threshold until two regions are found that match the nostril heuristics, that is they are of suitable and similar sizes and separation. The 'suitable size' is defined by the capture resolution and the identified face size. The centroid of the two nostril regions is taken as the active point that is the single location that is required by the tracking algorithm. Having located the nostrils, their location is used to update the centre of the search region for the following frame. If the nostril search fails and we are unable to update the search region using this mechanism, the search for the nostrils is reinitiated.

Samples of the search region and the identified nostrils are illustrated in Fig. 9. The translation of this information into cursor movement is the subject of the following section.

7. Cursor movement

Given the co-ordinates of the nostrils' active point and the co-ordinates of the face search area, we may derive signals for driving the cursor's movement.

Jitter, in this context, is defined as randomised apparent movement of the nostrils due to small amplitude, random head movements (tremor) and errors in the estimation of nostril location. Ideally, the cursor would move smoothly following decisive and well-controlled head movements. However, many of the intended users of this system have poor control of their movement. The system must therefore be capable of recognising jitter and eliminating it.

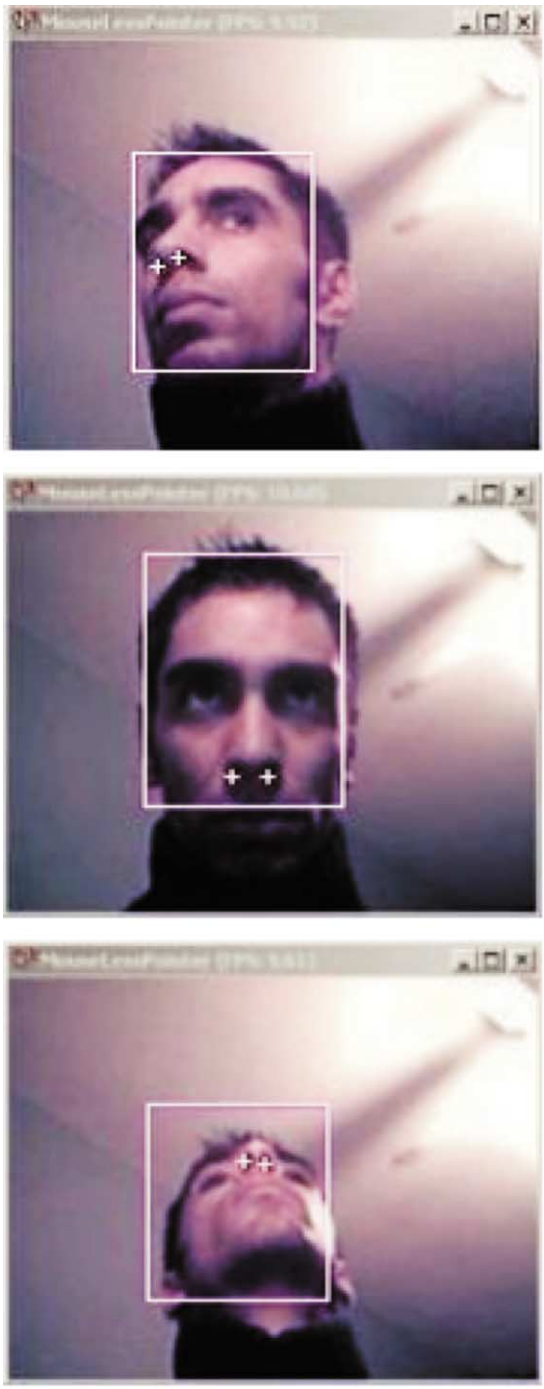

Fig. 9. Sample results of the nostril tracking stage.

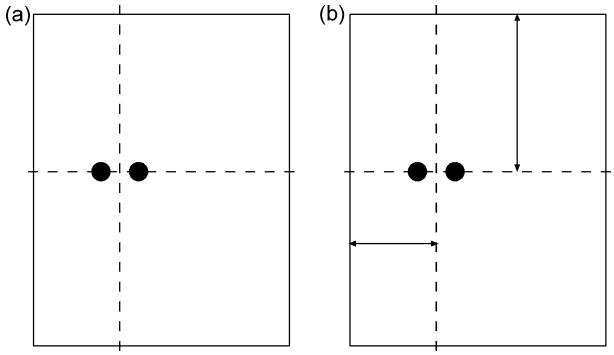

Fig. 10. Translation of nostril and search window locations to cursor control signals.

This is accomplished by monitoring the changes in the nostril location and the face search area co-ordinates. If these values change by more than a predefined threshold, the new values are used to update the cursor position, otherwise the position and these values remain unchanged.

The translation of image co-ordinates to cursor control signals is illustrated in Fig. 10. The distances from the nostril point to the boundaries of the face region in the vertical and horizontal directions is computed. This is translated directly into cursor position co-ordinates by linear scaling. The scaling coefficients are defined by a user specific calibration stage, we require that the user's maximum nostril movements in the horizontal and vertical directions will map into movements of the cursor that completely cover the monitor's screen. In a future version of the software we will replace the absolute cursor positioning with a joystick-like control method, whereby the cursor's velocity is controlled by the nostrils' positions as this offers improved cursor positioning (Drew et al., 1998; Evans et al., 1999; Evansand Blenkhorn, 2000.

The cursor is constrained to remain on the screen; cursor co-ordinates are therefore clipped to the screen.

8. Implementation and performance evaluation

The system was implemented and tested on various platforms running Windows. It executes in two phases: there is an initial calibration phase that is followed by the real-time tracking phase.

The initialisation phase performs two tasks. The first is the acquisition of the background image, if this is warranted. Recall from the discussion above that the blob finding module gave many false positive results when there were regions in the background with similar colours to the skin. Bare or varnished wood surfaces were especially problematical. Secondly, the initialisation phase captures the skin colour values for this particular user. Storring's conclusion that normalised skin colour values of all people fall within tightly constrained limits was found to be partly true, our experience has

been that normalised skin colours vary by a small amount but it is not possible to set global threshold values due to the variability in colour response of the cameras that are being used. Whilst it is possible to automate the process of acquiring a specific skin colour model, at present we use manual initialisation, even though this violates our requirement of minimal or zero initialisation.

Having performed the initialisation, the system moves into the real-time tracking phase that executes as previously described.

The system was evaluated using several metrics:

- The maximum throughput rate at various frame sizes,
- The accuracy with which the nostrils were located,
- Its robustness with respect to extreme head positions,
- Its robustness with respect to varying lighting conditions and
- The accuracy of the cursor positioning.

Using an entry level personal computer (at the time of development, a PIII processor with a clock speed of 500 MHz) we have achieved throughput rates of 30 frames per second at a resolution of 160 by 120 pixels, and 18 frames per second at 320 by 240 pixels. These results are intended to be illustrative of the speed that this system can achieve, but they do not necessarily reflect the speeds that would be achieved in practice, for three reasons. Firstly, an 'entry level personal computer' is more powerful now than when the development of this system commenced. Secondly, the tests were performed with the system providing graphical feedback to the developer, which would not be given in the same way in a release version. Thirdly, the tests were performed without the system running any other applications, such as a word processor. Given these factors, we expect that the system will be viable.

The system is able to detect the required facial features reliably, provided that the scene is adequately illuminated and the illumination remains unchanged. Adequate illumination is a reasonable requirement as the location in which this system will be used is expected to be well lit. As the illumination is reduced, darker skin tones result in tracking failure sooner than lighter tones due to the lower contrast between the skin and the nostril areas.

The system can be forced into tracking errors by introducing any large skin-coloured object into the field of view. If this object merges with the face, then the face region will change catastrophically.

The system is able to maintain a fix on the nostrils, even under movements that are more extreme than would usually occur when the user is controlling the cursor. This is a desirable feature given the physical characteristics of the intended users.

Several naïve users tested the system. They all found the system simple to use and were able to place the cursor within a 1 cm target on a $19''$ monitor with no difficulty. Although the system does not remove jitter completely, users compensate for slight remaining inaccuracies via the visual feedback that is provided by the on-screen cursor.

Although some of the users wore glasses, and this did not pose any problems, none had any facial hair. Moderate amounts of facial hair should not affect the performance of the system, especially if the hair is similar in colouration to the face. In fact, facial hair ought not to influence the system until it actually obscures the nostrils themselves (Fig. 11).

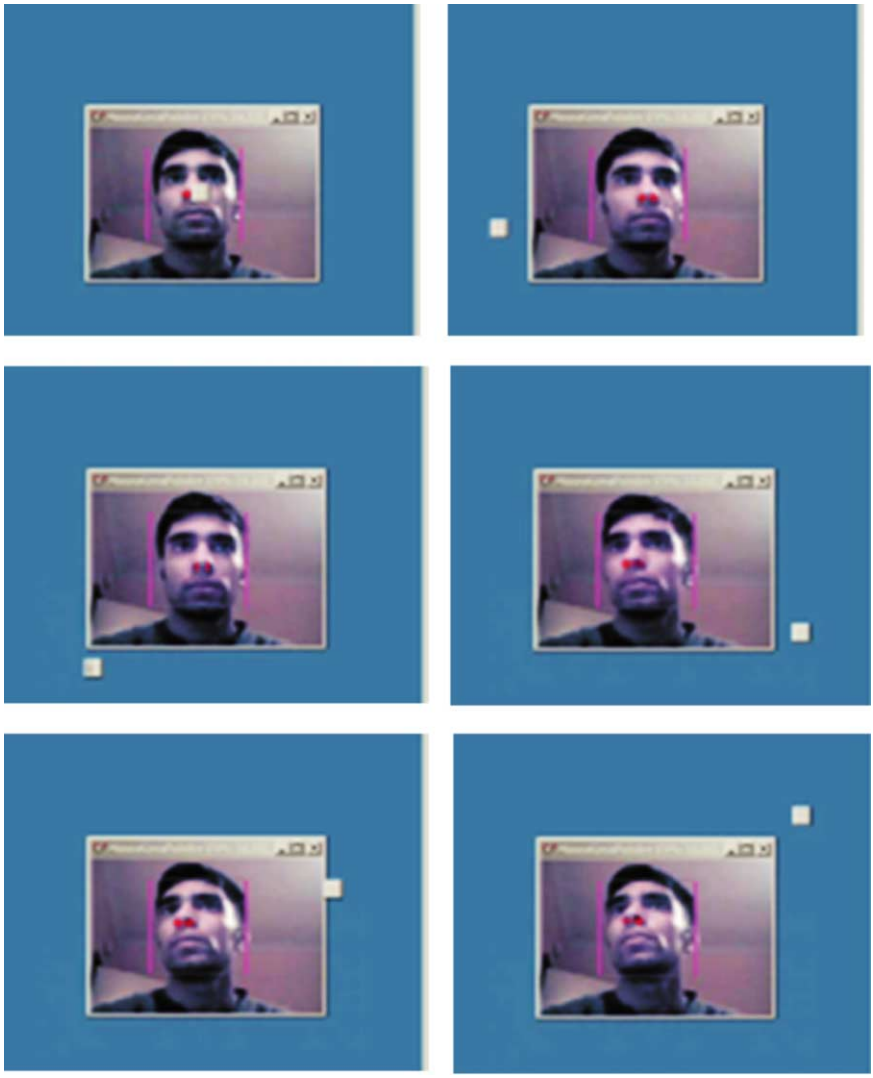

Fig. 11. Selected frames from a 30 s period of tracking. The images are screenshots of the system's output showing a video frame with tracking results superimposed: the rectangular facial region and two blobs indicating, where the system believes the nostrils to be located. Each image also contains a small square illustrating, where the system would place the cursor, note that the cursor's movement is a mirror image of the nostrils' movements.

Fig. 11 (*continued*)

9. Conclusion

A method of controlling the position of a cursor using video data has been presented. The system is intended to be used primarily by people with motor difficulties, although it could be used by the able bodied to enhance the mouse. We have aimed to develop a technically and computationally simple system. Input is captured using a webcam, simple processing methods have been employed.

The system tracks the user's nostrils, which are located by first finding a large skin-coloured region that is assumed to be the user's face. The position of the nostrils relative to the face region is used to control the position of the cursor. The system has been shown to be accurate and reliable under normal conditions of illumination and subject movement. The system is also able to track the desired features in real-time, lending support to our claim that this can be a valid and useful method of human computer interface for a certain class of computer users.

References

Bakic V, Stockman G. Menu Selection by Facial Aspect Proceedings on vision interface '99, Quebec, Canada 18–21 May 1999.

Brunelli R, Poggio T. Face recognition; features versus templates. IEEE Trans Pattern Anal Mach Intell 1993; 15(10):1042–3.

Drew R, Pettitt S, Blenkhorn P, Evans DG. A head operated 'joystick' using infrared. In: Edwards ADN, Arato A, Zagler WL, editors. Computers and Assistive Technology ICCHP '98, Proc XV IFIP World Computer Congress. Österreichische Computer Gesellschaft.

Evans DG, Blenkhorn P. A head operated joystick—experience with use Fourteenth Annual International Conference on Technology and Persons with Disabilities, Los Angeles, March 1999.

Evans DG, Drew R, Blenkhorn P. Controlling mouse pointer position using an infrared head operated joystick. IEEE Transactions on Rehabilitation Engineering 2000;8:107–17.

Jie Y, Lu W, Waibel A. Skin-color modeling and adaption Third asian conference on computer vision. Hong Kong: Hong Kong University of Science and Technology; Jan 1998. p. 687–94.

Menser B, Brünig M. Segmentation of human faces in color images using connected operators Proceedings of the IEEE international conference on image processing ICIP99, Kobe, Japan, 3 October 1999. p. 632–36.

Morris T, Zaidi F, Blenkhorn P. Blink detection for real-time eye tracking. J Netw Comput Appl 2002;25(2): 129–43.

Nikolaidis A, Pitas I. Facial feature extraction and pose determination. Pattern Recognit 2000;33(11):1783–91.

Orin Instruments, 2001 http://www.orin.com/access/.

Prentrom 2001, http://store.prentrom.com/.

Qian RJ, Sezan MI, Metthews KE. Face tracking using robust statistical estimation Proceedings of the workshop on perceptual user interfaces, San Francisco, California, November 1998.

Schiele B, Waibel A. Gaze tracking based on face-color Proceedings of the international workshop on auto. Face and gesture recognition, Zurich 1995. p. 344–49.

Sobottka J, Pittas I. Segmentation and tracking of faces in color images. In Proceedings of the Second International Conference on Automatic Face and Gesture Recognition 1996. p. 236–41.

Sobottka K, Pitas I. Extraction of facial regions and features using color and shape information International Conference on Pattern Recognition (ICIP), Vienna, Austria, August 1996.

Störring M, Andersen HJ, Granum E. Skin colour detection under changing lighting conditions. In: Araujo H, Dias J, editors. Seventh International Symposium on Intelligent Robotic Systems, Coimbra, Portugal, 20-23 July 1999. p. 187–95.

Stroud JM. The fine structure of psychological time. In: Quastlar H, editor. Information Theory in Psychology, Freepress, Glencoe, Ill.

Toyama K. Look, Ma—No Hands! hands free cursor control with real-time 3D face tracking Proceedings of Workshop on Perceptual User Interface (PUI'98) 1998.

University of Stirling Face Database 2003, http://pics.psych.stir.ac.uk/.

Yang M-H, Ahuja N. Detecting human faces in color images Proceedings of IEEE international conference on image processing, Chicago, IL October 4–7 1998. p. 127–30.

Yang J, Waibel A. A real-time face tracker. Proceedings of third IEEE workshop on application of computer vision 1996;142–7.

Yang M-H, Ahuja N, Kriegman D, survey A. A survey on face detection methods IEEE Transactions on Pattern Analysis and Machine Intelligence 1999.

Yuille A, Hallinan P, Cohen D. Feature extraction from faces using deformable templates. In International Journal of Computer Vision 1992;8:99–111.

Zitnick CL, Gemmell J, Toyama K. Manipulation of video eye gaze and head orientation for video teleconferencing. In: Technical Report MSR-TR-99-46. Redmond, WA: Microsoft Research; 1999. p. 12.