Denman, Simon, Butler, Darren, Sridharan, Sridha, & Chandran, Vinod (2004) ***Infra-red pupil detection for use in a face recognition system***. In: Proceedings of The 3rd Workshop on the Internet, Telecommunications and Signal Processing (WITSP '04), 20-22 December 2004, Adelaide, SA.

Infra-Red Pupil Detection for use in a Face Recognition System

Simon Denman, Darren Butler, Sridha Sridharan, Vinod Chandran
Image and Video Research Laboratory
Queensland University of Technology
GPO Box 2434, Brisbane 4001, Australia
{s.denman, de.butler, s.sridharan, v.chandran}@qut.edu.au

Abstract

This paper presents a new method of eye localisation and face segmentation for use in a face recognition system. By using two near infrared light sources, we have shown that the face can be coarsely segmented, and the eyes can be accurately located, increasing the accuracy of the face localisation and improving the overall speed of the system. The system is able to locate both eyes within 25% of the eye-to-eye distance in over 96% of test cases.

1 Overview

With the increasing security requirements of both the private and public sector, the need for reliable face recognitions systems is increasing.

Typically, face recognition systems need to process an image to locate and extract a face before any recognition can be done. This often requires a substantial amount of complex image processing and can be very demanding on computational resources. A wide range of techniques in the areas of skin segmentation of feature localisation are used in these situations. These methods are often dependent on models that are trained based on specific environments.

Features within the face must also be accurately located to provide reference points for extraction and verification. If the face itself, or its features cannot be located, then entire the process is likely to fail. As soon as imaging conditions are altered, the likelihood of localisation failure increases, until the conditions are so different such that the system is no longer viable.

This paper looks at two possible improvements to an existing face recognition system that aim to reduce the systems dependence on trained models

1. The addition of motion detection
2. The use of face and pupil detection through two switched IR light sources

and assesses the improvements made to the system though the addition of these features .

2 Literature Review

Many face detection/recognition systems rely on locating skin regions or facial features to aid in face localisation and segmentation. Features are located using a variety of techniques, with methods using colour information being quite popular. Rein-Lien *et al* [1] use chrominance images and morphological operations to locate the eyes and mouth, while Seo *et al* [2] uses contour detection, to look for a contour that has a specific colour on the inside and outside. Zhi-fang *et al* [3] uses skin detection followed by edge analysis and PCA, with He *et al* [4] applying gray scale projection to skin regions segmented using colour to obtain feature locations. Eigenfeatures and PCA are also popular, used in methods such as those by Sun *et al* [5] and Zhi-fang *et al* [3]. Neural networks have also been used [6, 7] for face localisation, and allow systems to be more invariant to in-plane rotations. The QUT system [8] uses a fusion of five techniques; chrominance difference image, shifted chrominance difference image, distance from features space, edge detection and discriminatory space map; to obtain eye locations.

All of these methods rely on the use of models to locate features, and thus their performance is dependent on how the models can adapt to any change in lighting or camera setup. In addition to this they are quite complex and are unable to track a face in real time.

An alternative method to locate the eyes involves the use of two near infrared (0.7u - 1.1u) light sources. Pupils can be located using two switched sources; one closely surrounding the camera lens (on-axis) and one source further away from the lens (off-axis). By illuminating the on-axis source, light is reflected of the back of the eye and the pupil appears as 'bright spot'. When the off-axis source is illuminated the pupil will appear as its usual black colour.

Recently several efforts have been made to use infrared techniques to locate eyes for reasons such as person detection and tracking [9, 10, 11], person counting [12], face pose estimation [13] and gaze detection [14].

Systems for pupil detection and tracking such as those developed by Haro, Morimoto and Kopoor [9, 10, 11] use two concentric rings of LEDs as the light sources. Bright eye and dark eye images are subtracted from one another to obtain a difference image, to which a threshold is applied to locate potential pupils. Haro used a simple adaptive threshold, where a histogram for the difference image was created and then back integrated, from which a threshold that would keep around 0.1% of the pixels was determined and applied.

Difference image pixels must be grouped into candidates, and from these pupils can be located and tracked. Haro chose to track all candidates, allowing candidate behavior to be observed over several frames to get a better idea of which candidates where behaving like pupils. Morimoto used the last known pupil position as a guide to search for the pupils in the next frame while Kapoor stored the average distance between the pupils, their last known x and y co-ordinates and restricted the search space to a bounding box centered on the last known pupil position.

To reduce the number of false positives when detecting pupils, Morimoto uses more than two images (consecutive bright and dark eye pair) for pupil location, allowing the pupils to be detected as high contrast regions, and help eliminate any motion artifacts. Haro used probabilistic PCA to help eliminate spurious candidates picked up within the difference image.

3 New System

The face recognition system resulting from this work, creates a system that is no longer dependent of models to locate the face and eyes. Motion detection is combined with IR face detection to locate the face, and IR pupil detection to locate the eyes. This simplifies the system, improves it's overall speed, and decreases its dependence on GMM's and eigenspaces compared to the system in [8], making the system more easily adaptable to other environments.

3.1 Pupil Detection Hardware

A black and white CCD camera, capable of seeing both near IR and visible light is used. The pupil detection hardware consists of two sets of eight infrared LEDs of variable brightness, and two transistor switches to control them. The LEDs are positioned in two concentric rings about the camera, with the inner ring having a diameter approximately equal to that of the camera, and the outer ring being sufficiently large such it can generate dark eye images. This configuration was chosen as similar configurations [9, 10, 11] have produced good results.

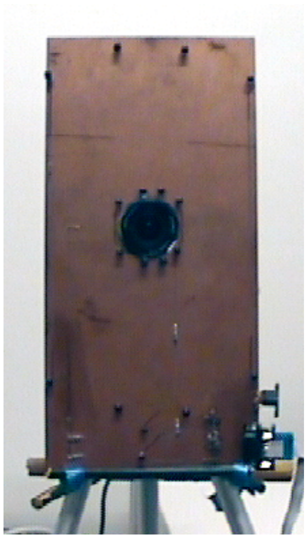


Figure 1: IR Pupil Detector Setup

The pupil detection system connects to the computer via a serial port, allowing the system to be controlled from within the face recognition system. The on and off axis LEDs are activated on alternate frames, generating a sequence of bright eye-dark eye images.

3.2 Motion Detection

Adaptive background detection methods [15, 16] allow changes in the scene to be incorporated into the background model, or multi-modal background situations to be correctly handled. Stauffer and Grimson [16] create a Gaussian Mixture Model for each pixel. Incoming pixels are matched against the GMM, and classified as foreground or background. The GMM is adapted after each frame, allowing the system to adapt to lighting changes and repetitive movement of background objects. However as a GMM is used for each pixel, it is very demanding on computational resources and cannot be implemented in real-time.

Butler *et al.* [15] propose a similar algorithm where each pixel is modeled by a group of clusters (a cluster consists of a weight, and a centroid) rather than a GMM. Pixels are matched and clusters are updated in a similar manner to Stauffer and Grimson's method. This method is substantially faster and capable of running in real time, and is used in our system.

The use of motion in this system has two main advantages:

1. It allows spurious candidates that may be inadvertently generated by an object in the background (i.e. a reflective object such as a glass or computer monitor) to be removed automatically
2. It allows the search space to be automatically restricted to the area of the person before any further information about the eye position is known

Motion detection is only performed on frames that are captured when the off-axis LEDs are illuminated. By using only the off-axis frames, no additional motion is detected due to the changing illumination of objects. This also allows for a more computationally efficient system, as pupil detection is performed on the on-axis frames. Motion for the on-axis frame is determined by applying the motion mask from the previous off-axis frame to the incoming image.

Once the motion detection is completed, the non-motion parts of the image can be filtered out before the image is passed onto the pupil detector and face verifier.

3.3 Face Detection through IR Illumination

Although the two sets of infrared LEDs are positioned such that the illumination from each set should be the same, there is still a slight difference in the illumination of objects positioned near the camera (such as a face when using a face verification system). As a result of this, it is possible to coarsely locate the face, and use this to restrict search space for pupil detection.

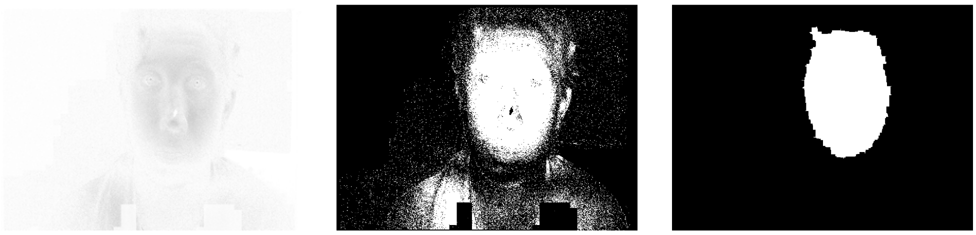
By calculating the absolute difference between dark eye and bright eye images, applying a threshold and some simple morphological operations, the face can be roughly segmented from the rest of the persons body, or any other objects that are close to the camera.

This method of face detection allows for further limiting the search space, enabling candidates appearing in a persons hair, or on their shirt to automatically removed. However, it will not work if the person is more than a couple of meters from the camera, as at this depth the illumination is identical.

3.4 Pupil Detection

For each bright-eye dark-eye pair that is captured, a difference image is calculated. This is done by subtracting the dark eye image from the bright eye image and applying a threshold. Only a segment of the difference image is calculated based on the current search space, which helps to improve speed and reduce errors.

The difference image calculation uses a simple adaptive threshold. The threshold changes based on the number of candidate eye pairs that it detects. If no pupils are detected in an image pair, the threshold is lowered by a small amount, to allow more differences through and potentially more candidates to be detected. If too many pupils are detected, the threshold is increased, to try and remove spurious detections. Pupil colour should experience the greatest (or close to) change in the image, so lowering the threshold when nothing is detected



(a) Absolute Difference Image (Inverse) (b) Threshold Image (c) Detected Face Region

Figure 2: Face Detection

(a) Bright Eye Image (b) Dark Eye Image (c) Difference Image

Figure 3: Pupil Detection

should ensure that the first differences that appear within the image will be the pupils, and raising it when too many are detected should ensure that the erroneous changes are ruled out first.

To prevent the threshold from continually lowering when the system is not in use, the threshold will not be altered when only a small amount of motion (or no motion) is detected in the scene (i.e. as would occur when there is no one using the system). The threshold is also changed according to the pixel value of the located pupils in the difference image, as the pixel value for the pupils in the difference image should be fairly consistent in consecutive frames.

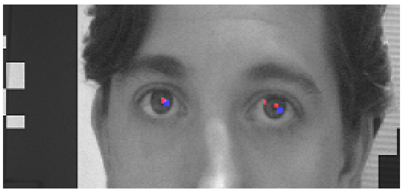
To help prevent incorrect detections that may be caused by strong reflections off hair or other objects, difference pixels must also have their corresponding pixels in the dark eye image checked for intensity. For a pupil, the colour should be close to black, and so any pixel that is too bright in the dark eye image, no matter how large the difference between bright and dark eye images, is likely to be something other than a pupil.

Once all different pixels have been located, they can be built into candidates (group of connected pixels that may or may not represent a pupil). Candidates are constructed by determining connected blocks within the difference image, and are then sorted and either accepted as possible pupils, or rejected as errors, based on the size and shape of the object.

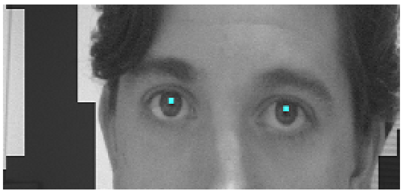
The remaining candidates are compared with one another to find the most likely pair of eyes. Each candidate is matched against with every other candidate, checking positions relative to one another and determining if they are likely to form a pair of eyes. Depending on the situation, this may result in multiple possible eye pairs.

Each eye candidate set is compared against the previous eye position, with the closest pair deemed to be the correct pair. For most cases however, only one valid pair of eyes is located.

If no pairs of eyes are found, then a single matching pupil is searched for. It is possible that only one pupil will be visible in the difference image, if the subject has quickly glanced away or had one eye

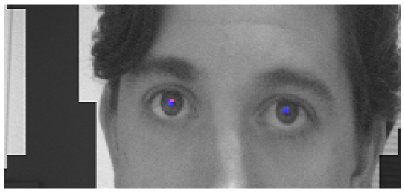


(a) Previous Eye Location and Current Candidates

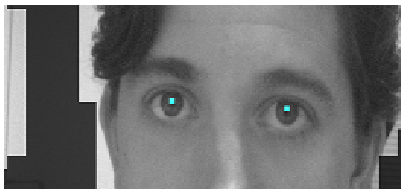


(b) New Eye Position

Figure 4: Updating Eye Position from Detected Candidate



(a) Previous Eye Location and Current Candidates



(b) New Eye Position

Figure 5: Updating Eye Position from a Single Candidate

obscured. In this case, provided the position of the eyes has been previously determined, it is still possible to accurately approximate the eye position.

The list of candidates is processed once more, with each candidate checked to see how close it is to the previous two eye locations. If a candidate that is suitably close to one of the eyes is found, it is assumed that this is the new position of the appropriate eye. The new position of the second eye can then be determined by offsetting its previous position by the same amount that the located eye has moved.

4 Results

A database was acquired in-house for testing, as no existing database contains suitable images, and by acquiring our own database it also allowed testing of the hardware. The database contained sets of 10 sequential bright eye-dark eye image pairs, with 4 or 5 sets of images for 10 subjects. In all, the database contained 830 bright eye-dark eye image pairs. Eyes for each pair of images were located by hand and these values were compared with positions determined by our algorithm. Eyes positions were correctly located if they satisfied equation 1 [17] with an error of less than 0.25.

$$e_{eye} = \frac{max(d_l, d_r)}{d_{eye}} \quad (1)$$

The system was tested with and without the face detection, and results are given for both. Results are also given for the accuracy of the last six pairs of images only. When implemented into the face recognition system, images are only passed to the verifier when a stable location of the eyes has been determined. This typically takes 3 to 4 frames, but can be faster of slower depending on the amount of motion from the user and in the background. Motion detection was not tested, as its performance has been previously documented in [8, 15].

Method	Left	Right	Both
IR	91.0%	90.5%	89.5%
IR with Face	95.1%	96.3%	95.0%
IR with Face (last 6 pairs)	96.3%	97.2%	96.3%

Table 1: Pupil Detection Accuracy

This accuracy is a substantial improvement over the previous system [8], which achieved an accuracy of 93.75% on image from the XM2VTS database.

Preliminary testing has shown that the use of the infrared light makes the system more robust to changes in the lighting environment. The off-axis set of LEDs effectively acts as a lighting booth, creating a more constant lighting environment. In addition, the pupil detection is very computationally efficient, and is able to process 25 frames per second at 320x240 resolution at approximately 50% processor utilisation on a 2.4GHz workstation. This is a significant improvement upon the old system which could take several seconds to locate the eyes in a single image.

5 Conclusion

We have shown that the use of an infrared pupil detector can improve the performance of eye localisation within a face recognition system, an thus improve the systems overall performance. The use of the infrared light sources is also able to adequately segment to the face, to restrict the search space and further improve the accuracy.

Further testing will be conducted to determine the effectiveness of the infrared light sources as a lighting booth.

Acknowledgments

This research was supported by the Office of Naval Research (ONR), USA, under Grant Award No: N000140310663 and by the Australian Research Council(ARC) through Discovery Grant Scheme, Project ID DP452676, 2004-6

References

[1] Hsu Rein-Lien, M Abdel-Mottaleb, and A.K Jain, "Face detection in color images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696 – 706, 2002.

[2] Kap-Ho Seo, Won Kim, Changmok Oh, and Ju-Jang Lee, "Face detection and facial feature extraction using color snake", in *IEEE International Symposium on Industrial Electronics*, 2002, vol. 2, pp. 457 – 462, Face Detection.

[3] Liu Zhi-fang, You Zhi-sheng, A.K. Jain, and Wang Yun-qiong, "Face detection and facial feature extraction in color image", in *Computational Intelligence and Multimedia Applications, 2003. ICCIMA 2003. Proceedings. Fifth International Conference on*, 2003, pp. 126–130.

[4] Xin He, Zhi-Ming Liu, and Ji-Liu Zhou, "Real-time human face detection in color image", in *Machine Learning and Cybernetics, 2003 International Conference on*, 2003, vol. 5, pp. 2915–2920 Vol.5.

[5] Darui Sun and Lenan Wu, "Face boundary extraction by statistical constraint active contour model", in *Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference on*, 2003, vol. 2, pp. 1274–1276 Vol.2.

[6] You-Wei Yuan, La-Mei Yan, and Han-Hui Zhan, "A robust face detection system using evolving neural networks", in *Machine Learning and Cybernetics, 2003 International Conference on*, 2003, vol. 5, pp. 2960–2964 Vol.5.

[7] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, "Rotation invariant neural network based face detection", 1998.

[8] Darren Butler, Chris McCool, Matthew McKay, Scott Lowther, Vinod Chandran, and Sridha Sridharan, "Robust face localisation using motion, colour fusion", in *Seventh International Conference on Digital Image Computing: Techniques and Applications*, 2003, pp. 899–908.

[9] A Haro, M Flickner, and I Essa, "Detecting and tracking eyes by using their physiological properties, dynamics, and appearance", in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2000, vol. 1, pp. 163–168 vol.1.

[10] C.H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources", *Image and Vision Computing*, vol. 18, pp. 331–335, 2000.

[11] A Kapoor, Y Qi, and R.W Picard, "Fully automatic upper facial action recognition", in *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, 2003, pp. 195–202.

[12] I Haritaoglu and M Flickner, "Attentive billboards", in *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on*, 2001, pp. 162–167.

[13] Qiang Ji and Xiaojie Yang, "Real time 3d face pose discrimination based on active ir illumination", in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, 2002, vol. 4, pp. 310–313 vol.4.

[14] M Ohtani and Y Ebisawa, "Eye-gaze detection based on the pupil detection technique using two light sources and the image difference method", in *Engineering in Medicine and Biology Society, 1995. IEEE 17th Annual Conference*, 1995, vol. 2, pp. 1623–1624 vol.2.

[15] D Butler, S Sridharan, and V. M Bove Jr, "Real-time adaptive background segmentation", in *ICASSP '03*, 2003.

[16] C Stauffer and W.E.L Grimson, "Adaptive background mixture models for real-time tracking", in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, 1999, vol. 2, p. 252 Vol. 2.

[17] O. Jesorsky, K. Kirchberg, and R. Frischholz, "Robust face detection using the hausdorff distance", in *Thrid International Conference on Audio and Video Based Biometric Person Authentication*, 2001, pp. 90–95.