FACE DETECTION IN COLOUR IMAGES

J.D.Brand and Dr. J.S.D.Mason

Department of Electrical Engineering
University of Wales Swansea
SA2 8PP, UK
email: J.D.Brand@swansea.ac.uk

Mark Pawlewski

BTexaCT
Advanced Communication Technologies
Adastral Park
Martlesham Heath
Ipswich IP5 3RE

ABSTRACT

This paper describes the development and quantitative assessment of an approach to face detection (FD), with the application of image classification in mind. The approach entitled 'Crude Automatic Face Extraction' (CAFE), is based on skin colour segmentation, multiple image representations and multi-resolution scanning. It is a direct extension of an approach by Huang [Pattern Recognition 1994]. Assessment is in terms of false acceptance (FA) and false rejection (FR) scores using a large quantity of unconstrained typical Internet images, some of which contain faces and some of which do not.

Skin-based colour segmentation is introduced as a front-end, along with several additional low-level image representations, the benefits of which can be seen by a reduction in the FA score from 74% to 17%. It is also noted that a skin probability map can improve FD significantly.

1. INTRODUCTION

Within the field of computer vision, face detection (FD) has many potential applications. Image database searching is one of these. Here the task is to label images according to their content so that these images can be sorted or retrieved at some time in the future. A typical application might be searching the Internet for images containing certain items. One might ask for images containing sunsets or cathedrals. Here the item considered is the human face.

In the limits this task becomes a simple yes/no classifier: does a given image contain one or more faces? In a constrained context, for example in a face recognition system where persons are expected to present themselves in front of a camera, FD is a relatively simple task. However in contrast, when searching for faces in an almost unconstrained context, variations present significant difficulties. These variations are attributable to the quality of the image (camera and environmental conditions) and the natural variations of the face itself (particularly scale, pose, rotation and expression).

Detailed reviews concerning the various approaches to FD can be found in [1, 2, 3]. In particular, the recent work by Yow [1] summarises FD strategies into 6 general categories:

- shape-based [4],
- feature-based [1],
- pattern-based [5, 6, 7],
- colour-based [8, 9],
- motion-based [10],
- miscellaneous [11, 12].

The earlier work of Huang [7] uses a pattern based approach that applies knowledge-based rules to intensity images. One of the problems with this approach is the adverse effects of various lighting conditions, causing many incorrect faces to be found in images that do not contain faces. The Crude Automatic Face Extraction (CAFE) system presented here is founded on Huang's approach and addresses this problem. Huang's approach was chosen as a starting point primarily because it is an intensity only based approach with the potential to be extended using colour pre-processing, as demonstrated here. Also, the pyramid searching method used by Huang to cope with changes in scale can be efficiently computed using algorithms such as T-pyramids and Quad-trees [13].

A hybrid-based approach is developed by adding image pre-processors which significantly improve performance, as measured by false acceptance (FA) and false rejection (FR) rates. The paper is structured as follows. In Section 2 a brief review of Huang's approach to FD is presented. Section 3 then describes the hybrid CAFE approach, which is built on the fundamentals of Huang's approach. Section 4 gives details of how assessment was carried out, using a database of test images taken from the Internet. Finally Section 5 draws conclusions from the FD results.

2. HUANG'S APPROACH TO FACE DETECTION

In 1994 Huang presented a novel approach to FD, using grey-level mosaic images [7], later also investigated by Song [14] and Liu [15]. This hierarchical pattern-based approach is based on a set of intuitively determined rules, applied to a pyramid of mosaic intensity images. Pyramid systems are often used in computer vision to search images when scale is a factor.

Huang uses three levels of rules. Each level is applied to a different representation of the original image. As the detail of the image increases, so does the complexity of the rules. The increasing complexity of the rules is designed to progressively remove more and more false candidates from the image. All the rules are based on the relative intensities within a specified square image region, under the assumption that faces can be contained within such a region.

Grids of different sizes are exhaustively passed over the appropriate image representation, and are subject to the corresponding set of rules. Only image regions that pass the first set of rules can be passed to the second set of rules, and so forth. Only grids that pass all 3 sets of rules are deemed to contain faces.

Level 1 rules look for general face shapes in very low-resolution images. Level 2 rules search for eyes, nose and mouth and are applied to an image twice the resolution of the previous image. Finally the Level 3 rules are applied to a binary edge representation of the original grey level image, and confirm or otherwise, the location of the facial features. The details of the rules can be found in [7].

3. THE CAFE SYSTEM

The Crude Automatic Face Extraction (CAFE) system developed here extends the fundamental ideas of Huang's approach. Huang's approach applied knowledge-based rules to intensity images, which are known for their sensitivity to changes in lighting conditions. This leads to false faces being picked up in complex backgrounds that have similar intensity distributions to human faces.

The CAFE system directly extends Huang's approach by integrating colour and intensity information. Several pre-processing steps are added to Huang's approach, the most crucial of which is a colour skin probability map (SPM) front-end. Rather than classifying pixels as skin and non-skin, each pixel is assigned a weight, showing the likelihood of the pixel belonging to skin [16]. This idea is also presented in [17]. To create such a map requires a large labeled database [18] [1].

The colour map has the potential to reduce the image search space and improve FD accuracy. This is done by highlighting face regions while reducing the adverse effects of various lighting conditions that tend to plague Huang's original approach.

In total five pre-processing image maps are added to Huang's FD approach. Some of the maps are designed to highlight face boundaries or features (positive face maps), while others are designed to diminish the effect of background information (negative face maps). A block diagram illustrating CAFE is shown in Figure 1.

The left hand side of Figure 1 shows Huang's original approach to FD, resulting in a single FD image classification decision. In this diagram H1 represents the Level 1 rules, while H2 is an umbrella term that incorporates both Level 2 and Level 3 rules. The right hand side of Figure 1 shows how Huang's approach has been extended by using modules as preprocessors to Huang's approach. Here seven classifier outputs are given: **C1** to **C5** represent the incremental performance gains as each module of CAFE is incorporated. **C6** and **C7** are decisions based on Huang's set of rules applied to intensity images. **C6** is directly equivalent to **H1** and **C7** to **H2**.

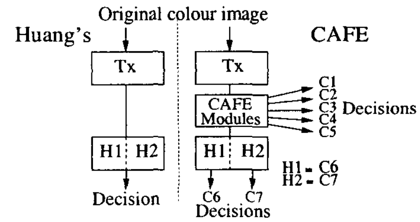

Fig. 1. A block diagram of the CAFE system with respect to Huang's approach

The pre-processing maps and their appropriate abbreviations with respect to Figure 1 can be summarised thus:

- **C1** - morphological closing of the SPM image (CSPM) to form closed, featureless faces.
- **C2** - a morphologically opened Hue Saturation Intensity image (MO-HSI). The hue and saturation components can have low variances for some examples of skin colour. After morphological filtering to remove holes in the image, the two components are separately thresholded and utilised in the face detection process.
- **C3** - this has two distinct components based on Hough line detection (HLD). First, Hough Line Detection of the CSPM (HLD-CSPM) gives face boundaries from the CSPM output. Second, Hough line detection of the BE image (HLD-BE) seeks strong, long lines *not* associated with faces. The second map is therefore classified as a negative face image unlike the previous two cases.
- **C4** - a simple Binary Edge (BE) image. The number of post-thresholding features with a closed range gives an indication of faces.
- **C5** - a mosaic-based horizontal feature detection image (M-HFD). This component seeks horizontally biased features associated with eyes and eyebrows, nostrils and mouth.

Thresholds for the above maps are empirically determined on a sub-set of 60 training images; 40 contain faces, the other 20 are considered potential problem non-face images. The seven components are cascaded in the final CAFE implementation. However, each can be added in turn, on a step-by-step basis, in order to assess its contribution.

4. ASSESSMENT

Classification score are determined using 1244 'unseen' test images (530 with faces, 714 without faces). Two assessments of classification performance are presented: CAFE versus Huang's original FD and a step-by-step analysis of the CAFE approach.

[1]The authors gratefully acknowledge the use of the Compaq skin database

4.1. Classification Error Rates

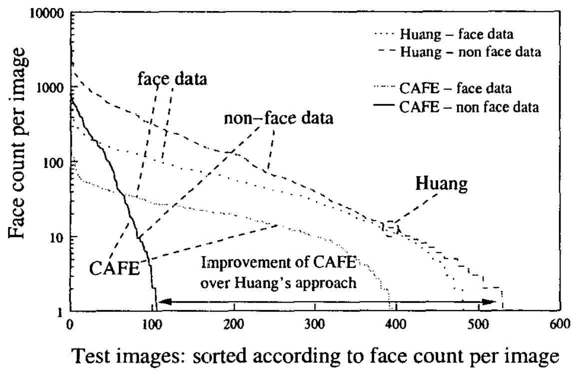

Fig. 2. A comparison of two approaches to face detection: Huang's FD and CAFE. Face count per image plotted against number of images classified as containing a face

Figure 2 shows four profiles relating to Huang and CAFE, face and non-face data. Referring back to Figure 1, it is seen that the difference between the two algorithms is the inclusion of 5 extra components **C1** to **C5** in CAFE.

The vertical axis for the plots shows the number of faces found in any one image and the horizontal axis shows the number of images deemed to contain one or more faces, rank ordered according to the face hit count. Where the profiles cross the x-axis indicates false acceptance (FA) rates for non-face data and false rejection (FR) rates for face data. Considering first the 530 images with faces, it is seen that Huang's approach finds 490 (8% FR) whereas CAFE finds only 400 (25% FR). Thus Huang's approach correctly accepts 17% more faces. However, there is a much bigger difference in FA - images wrongly declared as containing faces. Huang's approach falsely accepts 520 non-face images as having faces from a total of 714, giving an FA rate of 74%, compared with CAFE: 102 from 714 (17% FA).

Figures 3 shows two sets of profiles. Each graph illustrates the incremental improvements for face (top of Figure 3) and non-face data (bottom of Figure 3), for the CAFE approach as each module (**C1** to **C7**) is incorporated.

The general trend of both graphs is the same: as each new stage is added the curves shift from right to left. Working from right to left, the first profile represents the CAFE approach using only the CSPM image (**C1**) i.e. each image is exhaustively scanned and a face is said to exist if a given percentage of pixels in that region are skin coloured. Empirical experiments suggest that face regions contain approximately 70% skin.

The next profile incorporates the rules of the HLD-BE image and the HLD-CSPM image on top of the CSPM image (**C2**). Both of these images use Hough line detection to highlight face boundaries (HLD-CSPM) or background objects (HLD-BE). This is a robust method for finding curves that may contain gaps or be partially occluded.

After this the MO-HSI image and its corresponding set of colour statistics rules are applied (**C3**). Next, some simple 'feature counting' rules are applied to the BE image (**C4**). After this, potential face regions are searched for horizontally biased facial features, highlighted by the M-HFD map (**C5**).

The final two profiles are applied to mosaic representations of the intensity image. In particular, the penultimate profile is equivalent to Huang's stage 1 after pre-processing with five CAFE modules (**C1 - C5**). The final profile is equivalent to Huang's stage 2 and adds both Level 2 and Level 3 rules to the system.

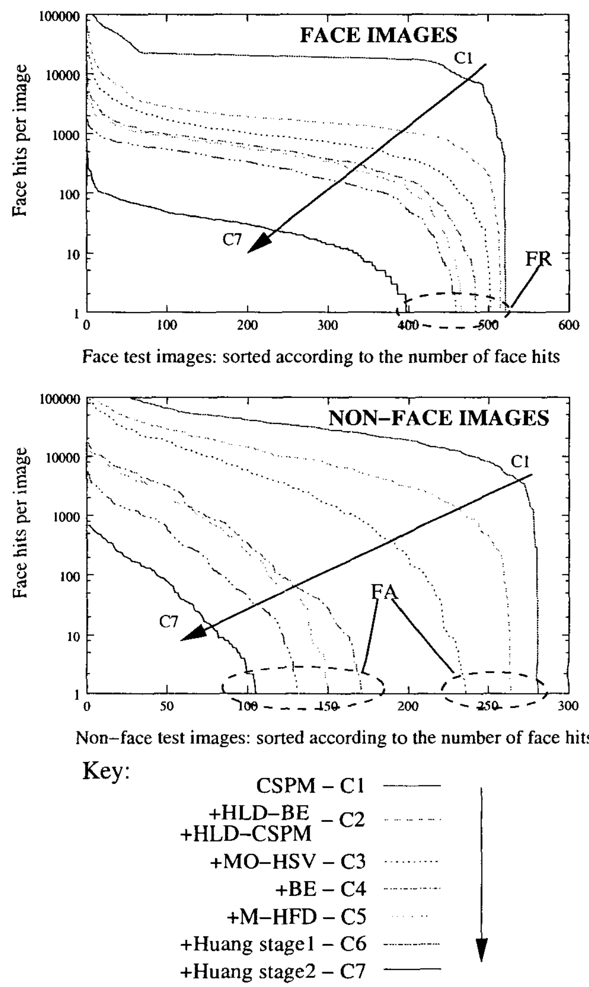

Fig. 3. Incremental improvements of CAFE modules for face (top) and non-face (bottom) test images. Axes as per Figure 2

For the face data (top of Figure 3) the first 6 profiles are closely bunched falling from 520 to just 460 images. The seventh profile, shows a large step and reduces the number to 400, thereby increasing the FR score from 14% to 25%. This rejection of true face images comes about as a result of the strictness of the rules for **C7**, which searches a high-resolution mosaic representation of the intensity image. Here, geometrical constraints are imposed on the candidate face regions in order to find eye, nose and mouth locations. These constraints (unwisely) assume the faces to be almost full frontal with minimum rotation. Faces that do not obey these criteria are rejected as the rules are too specific to encompass general face variations.

The profiles in the bottom half of Figure 3 relate to non-face data. As each CAFE module is added there is a progressive reduction in the FA rate as well as a reduction in the number of face hits per image. Interestingly, when more modules and rules are added to the *face* image set, (top of Figure 3), the drop in face hits seems very gradual until **C7** where Level 2 and Level 3 rules are applied. However, when the same rules are applied to non-face images, there is a noticeable reduction of both face hits and accepted images after each and every stage. This can be seen by the equal spacing of the curves. This is a positive indication that the rules imposed by each image are having the desired effect. Ultimately, when all images and rules have been applied, only 109 out of 714 non-face images remain. This gives an FA score of only 15%. The bulk of face hits per image is between 10 and 100.

Figure 4 shows FR and FA scores for the step-by-step addition of the CAFE modules. Here both FA and FR error rates have been plotted with respect to CAFE modules (x-axes). The CSPM is seen to reject over 60% of non-face images (top left corner of Figure 4). The scores suggest that the BE image (**C4**) has a considerable impact on the CAFE approach as the reduction in FR score is 9% (from 33% to 24%), with a trade-off of only 3% (from 5% to 8%) FA. As each CAFE module is added to the system, the FA score (top curve) can be seen to fall, while the FR score (bottom curve) is seen to rise. Where the two curves intersect, at approximately 17% is the conceptual equal error rate (EER).

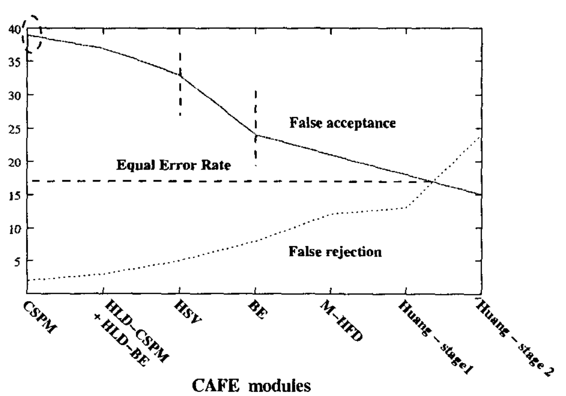

Fig. 4. FA and FR profiles for the CAFE approach with respect to the incremental addition of CAFE modules

5. CONCLUSION

A face detector has been designed and implemented. The incremental performance on the classification of 1244 Internet images has been assessed. The final stage of CAFE (**C7**), using Huang's Level 2 and Level 3 rules on the intensity image, seems less desirable than other components. The rules are deemed to be too strict for the wide variation in the images under test.

In contrast the skin map improves classification performance significantly. This and the other four components of CAFE contribute positively giving an overall FR of 13% and an FA of 18% for the given setup. These values can be traded against each other by changing thresholds.

6. REFERENCES

[1] K.C. Yow, "Automatic Human Face Detection and Localization," *PhD thesis, Downing College, Cambridge University*, 1998.

[2] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," *Pattern Recognition*, vol. 25, pp. 65–77, 1992.

[3] M.H. Yang, N. Ahuja, and D. Kriegman, "Detecting Faces in Images: A Survey," *Submitted to IEEE Trans. PAMI*, 2000.

[4] J. Wang and T. Tan, "A new face detection method based on shape information," *PRL*, vol. 21, pp. 463–471, 00.

[5] H. Rowley, S. Baluja, and T. Kanade, "Human Face Detection in Visual Scenes," *Technical Report*, 1995.

[6] M.H. Yang, N. Ahuja, and D. Kriegman, "Mixtures of linear subspaces for face detection," *ICAFGR*, 2000.

[7] T.S.Huang and G.Z.Yang, "Human Face Detection in a Complex Background," *Pattern Recognition*, vol. 27, pp. 53, 1994.

[8] J. Cai and A. Goshtasby, "Detecting human faces in colour images," *IVC*, vol. 18, pp. 63–75, 99.

[9] R.J. Qian, M.I. Sezan, and K.E. Matthews, "A Robust Real-Time Face Tracking Algorithm," in *Proc. ICIP*, 1998, pp. 131–135.

[10] H.P. Graf, T. Chen, E. Petajan, and E. Cosatto, "Locating Faces and Facial Parts," in *Int. Workshop on Automatic Face and Gesture Recognition*, 1995, pp. 41–46.

[11] M. Kapfer and J. Benois-Pineau, "Detection of human faces in colour sequences with arbitrary motions fro very low bit-rate videophone coding," *PRL*, vol. 18, pp. 1503–1518, 97.

[12] C.Wang and M.Brandstein, "Multi-Source Face Tracking with Audio and Visual Data," *IEEE MMSP*, p. 168, 1999.

[13] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, PWS Publishing, 1999.

[14] Q. Song, *Robust Face Detection and Pose Estimation*, PhD, Memorial University of Newfoundland, 1998.

[15] A. Liu and I. Wu, "Summarization of Face Objects in a Video Sequence," *http://www.ctr.columbia.edu/~ifongwu/visrep.html*, 1996.

[16] J.D. Brand, J.S.D. Mason, and M. Roach, "A Comparative Assessment of Three Approaches to Pixel-level Human Skin-Detection," *ICPR*, vol. 1, pp. 1056–1059, 2000.

[17] M.Jones and J.Rehg, "Statistical Color Models with Application to Skin Detection," *CVPR*, p. 274, 1999.

[18] M.Jones and J.Rehg, "Compaq skin database," *http://www.crl.research.digital.com/publications/tech-reports/abstracts/98_11.html*.