Face Detection Based on Multiple Regression and Recognition Support Vector Machines

Jianzhong Fang and Guoping Qiu
School of Computer Science
University of Nottingham, Nottingham NG8 1BB, UK

jzf@cs.nott.ac.uk qiu@cs.nott.ac.uk

Abstract

This paper presents a novel approach to face detection. A potential face pattern is first filtered by a Gaussian derivative filter bank to generate a set of derivative images, which are then transformed by the Angular Radial Transform (ART) to form a compact set of representation feature vectors. Using these feature vectors for face detection is based on a two level multiple support vector machines (SVMs) strategy. At the first level, a separate SVM is trained for each derivative image to indicate the presence/absence of a face in the input based on the features from that derivative image alone. These SVMs are trained as binary classifiers but used for regression in the sense that they output continuous values. At the second level, a single SVM takes the outputs of the first level SVMs as input to make the overall and final decision to determine whether the current input is a face or nonface pattern. Experimental results are presented to demonstrate the effectiveness of the method.

1 Introduction

Face detection is a difficult problem, which confronts all the major challenges in computer vision and pattern recognition [1]. Over the years, many approaches to face detection have been proposed [2, 3]. Support Vector Machine (SVM) as a binary classifier has proved eminent in face detection literature achieving very high correct detection rates [4]. In [5], a series of support vector machines were trained based on Principal Component Analysis (PCA) coefficients to accomplish the task. In [6], 14 facial components were selected to model a face. Each facial component such as the nose, the eye and the mouth were each modelled with a specialised SVM classifier. The continuous outputs of all facial component classifiers were passed to the second level SVM classifier, which was referred to as geometrical configuration classifier, to make the final joint decision.

BMVC 2003 doi:10.5244/C.17.49

Our work employs the SVM technique in a way similar to that of [6]. What makes it different is the way in which the facial "components" were derived. We adopt a global approach rather than a local facial parts method to model each "component". We introduce a set of derivative images to represent face/nonface patterns. For each derivative image, we use the Angular Radial Transform (ART) [7] to derive a compact representation feature vector. At the first stage, a separate SVM is trained for each derivative image representation. Instead of making binary decisions, these first stage SVMs output continuous values in the way that SVMs were used for regression [14]. The output of each of these first stage SVM gives an indication of the likelihood that the component belongs to a face/nonface pattern. At the second stage, a SVM takes the outputs of the first stage SVMs as input to make a final decision whether the current input is a face or nonface pattern. Compared with the local facial component method [6], such an approach will not deliberately choose separable components, therefore, it is more suitable for generic object detection.

In the following section, we will give an overview of our method. In Section 3, we will discuss feature extraction by means of Gaussian derivative filters. In section 4, Angular Radial Transform (ART) is presented. In Sections 5 and 6, we give explanation of the training of the 2-stage SVMs and report experimental results.

2 Overview of the Method

The overall scheme of our proposed method is illustrated in Figure 1. In this section, we give a brief overview of each stage of the scheme.

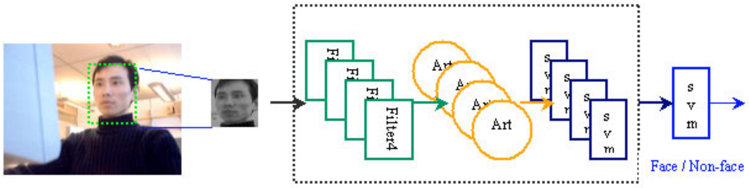

Figure 1: Flow chart of face detection. Normalised image patch goes through parallel filter bank, ART transform and SVM classifiers. The second level SVM classifier makes decision between face and non-face.

2.1 Representing Face using Image Derivatives

An image can be represented in different ways. Spatial derivatives are capable of extracting dynamic features hidden in the image. Our basic idea is to retrieve invariant face features not only from the original image alone, but from the spatial luminance change as well. By compiling derivative images together, a much more powerful face representation can be formed. Let *I*(*x*, *y*) denote a face image. The original face is represented as

$$FACE = I(x, y) \tag{1}$$

The kth-order derivatives of the image *I*(*x*, *y*) with respect to orientation $\vec{v} = (\cos\phi, \sin\phi)$ is

$$FACE_{k,\phi} = \frac{\partial^k I(x,y)}{\partial \vec{v}^k} \tag{2}$$

For a given choice order and direction of the derivative, the extended face description *FACE'* is written as

$$FACE^{'} = \{FACE_{k,\phi}\}, \;\; 0 \le \phi \le 2\pi, \;\; 0 \le k \le K_{\max} \tag{3}$$

where $K_{\max}$ is a predefined maximum derivative order.

2.2 Developing Compact Face Descriptors via Angular Radial Transform

Angular Radial Transform (ART) is a moment-based image description method adopted in MPEG7 as a region-based shape descriptor [7]. It gives a compact and efficient way to express pixel distribution within a 2-D object region; it can describe both connected and disconnected region shapes. Here, we extend the application to our work for face detection. A set of orthogonal moment bases is defined along the angular and radial direction by the order of n and m across the unit disc. The transform renders an inner product between the base function $V_{nm}(\rho,\theta)$ and the image function $I(x,y)$ over a unit disc.

$$F_{nm} = \left\langle V_{nm}(\rho,\theta), I(x,y) \right\rangle = \left\langle V_{nm}(\rho,\theta), I(\rho\cos\theta, \rho\sin\theta) \right\rangle \tag{4}$$

where F_{nm} is an ART coefficient, $0 \le n < N$, $0 \le m < M$.

Parameters *N* and *M* indicate the number of moments employed in the transform determining the transform resolution. In this work, we set $N = 3$ and $M = 12$ and choose the real part of the ART bases to obtain a 36-dimensional feature vector for each derivative image. In Section 4, ART will be further discussed.

Let $ART_{NM}(I)$ be the full ART transform of $I(x, y)$, we have

$$ART_{NM}(I) = \{F_{nm}\}, \; 0 \le n < N, \; 0 \le m < M \; . \tag{5}$$

Accordingly, a face descriptor *FACE"* derived from the angular radial transform is represented as

$$FACE^{"} = ART_{NM}(FACE^{'}) = \{ART_{NM}(FACE_{k,\phi})\} \tag{6}$$

FACE" is a vector whose dimension is determined by the ART moment order *N* and *M*, and the number of derivative images chosen. For example, when we choose ART moment order $N = 3$ and $M = 12$, the dimensionality of the following face descriptor:

$$FACE^{"} = \{ART_{3,12}(FACE_{0,0}), ART_{3,12}(FACE_{1,0}), ART_{3,12}(FACE_{1,\pi/2})\}$$

will be calculated as 3 x 12 x 3 =108.

2.3 Two Level Multiple SVMs for Face Detection

Although we can train the support vector machine classifier directly based on the whole face descriptor *FACE"* rendered by the ART as described in previous sub sections, we decided to use a two stage strategy. At the first stage, we train one SVM for the ART coefficient vector of each of the derivative face, $ART_{MN}(FACE_{k,\phi})$. The rationale is as follows. Firstly, the aggregated feature *FACE"* may be of very high dimension which may cause the system to suffer from the "curse of dimensionality" phenomenon. Secondly, each of the derivative images may capture some unique characteristics of a face pattern, which by itself may be sufficient to determine the presence or absence of a face. Moreover, each derivative image may influence the decision making in a different way, and therefore we want to capture the role of and the way in which each derivative image plays in the final decision making. The SVMs in this stage are trained as binary classifiers but are used for regression after training. The continuous output values from each of the SVMs trained for each of the derivative image can be viewed as the likelihood of the component comes from a face or non face vector. Once these first stage SVMs are trained, we need to develop a method to use their outputs to make a final decision to determine whether the input image is a face pattern. SVM is once again an appropriate choice to model the second level decision. In the second stage, we train a single SVM, which takes the continuous output of the SVMs in the first stage to make an overall and final decision to determine whether the input image is a face or nonface.

Let *SVC*(*FACE"*) denote the SVM classifier output from the face descriptor *FACE"*, we will apply following rule to make the final decision

$$\begin{cases} SVC(FACE^{"}) > 0 & face \\ SVC(FACE^{"}) \le 0 & non-face \end{cases} \qquad (7)$$

SVC(*FACE"*) is a scalar output, the decision chain of our method can be expressed in terms of the first layer and the second layer SVM classifiers.

$$SVC(FACE^{"}) = SVC\{SVC(ART_{NM}(FACE_{k,\phi}))\} \qquad (8)$$

In our experiments, four derivative images were used, which were, the original face image, the first order derivative images at X and Y directions, and the Laplacian.

3 Computing Image Derivatives

Gaussian derivatives are widely used in the literature and well understood [8, 9]. In this work, we use Gaussian derivative filters to compute the derivative images. The 2-D Gaussian with the scale σ is defined by:

$$G^{\sigma}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (9)$$

The nth-order Gaussian derivative at direction $\vec{v} = (\cos\phi, \sin\phi)$ is defined by:

$$G^{\sigma}_{n,\phi}(x,y) = \frac{\partial^n}{\partial \vec{v}^n} G^{\sigma}(x,y) \qquad (10)$$

Based on such provision, the response of face derivative image going through a Gaussian filter is the convolution of the two functions: $FACE_{k,\phi} * G^{\sigma}(x,y)$ We can write:

$$FACE_{k,\phi} * G^{\sigma}(x,y) = \frac{\partial^k}{\partial \vec{v}^k} I(x,y) * G^{\sigma}(x,y) = I(x,y) * \frac{\partial^k}{\partial \vec{v}^k} G^{\sigma}(x,y)$$

Therefore, we have

$$FACE_{k,\phi} * G^{\sigma}(x,y) = I(x,y) * G^{\sigma}_{k,\phi}(x,y) \qquad (11)$$

In our work we use Gaussian derivatives up to the second order. We choose the scale $\sigma = 1$ and two directions with respect to X-axis and Y-axis. In total, five derivative images are obtained:

$$\{I * G^{\sigma}_{0,0}, I * G^{\sigma}_{1,0}, I * G^{\sigma}_{1,90}, I * G^{\sigma}_{2,0}, I * G^{\sigma}_{2,90}\}$$

Practically, we merge the second derivative images to a Laplacian of Gaussian:

$$L^{\sigma}(I) = (\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}) * G^{\sigma} = I * G^{\sigma}_{2,0} + I * G^{\sigma}_{2,90} \tag{12}$$

Eventually, four derivative images are selected for the representation of face image in this paper: $\{ I * G^{1}_{0,0} \quad I * G^{1}_{1,0} \quad I * G^{1}_{1,90} \quad L^{1}(I) \}$. Figure 2 shows an example of these derivative images.

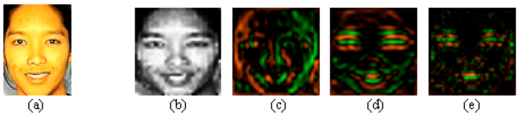

Figure 2 (a) original face; (b) 32x32-pixel normalised face after lighting correction and histogram equalisation; (c) derivative image with respect to X-axis; (d) derivative image with respect to Y-axis; (e) Laplacian image. Red colour (darker part) indicates positive value, green colour (lighter part) indicates negative value.

4 Angular Radial Transform

Several choices are available to derive face descriptors from face image. Principal component analysis is one of those very commonly used methods [10, 11, 12]. Here we introduce a new method to generate a compact face descriptor in face detection. Angular Radial Transform (ART) is an irreversible orthogonal transform, which is robust against noise. Independent experiments have shown that ART outperforms other moment-based transforms [13].

The base function forms the kernel of ART. A set of orthogonal bases with angular radial moment *n* and *m* is defined in polar coordinates. The basis function $V_{nm}(\rho,\theta)$ is separable along the angular and radial directions. It is defined by:

$$V_{nm}(\rho,\theta) = \frac{1}{2\pi} e^{jm\theta} R_n(\rho) \qquad R_n(\rho) = \begin{cases} 1, & n = 0 \\ 2\cos(n\pi\rho) & n \neq 0 \end{cases} \tag{13}$$

The ART transform of the function $f(\rho,\theta)$ is to perform inner product between the basis function and the input function over the unit disc. Suppose the transform coefficient corresponding to the basis $V_{nm}(\rho,\theta)$ is F_{nm}, then it is expressed as

$$F_{nm} = \langle V_{nm}(\rho,\theta), f(\rho,\theta) \rangle = \int_0^{2\pi} \int_0^1 V_{nm}(\rho,\theta) f(\rho,\theta) \rho d\rho d\theta \tag{14}$$

The coefficient F_{nm} is a complex number, we can either choose the real part or the imaginary part from the bases. Figure 3 is a set of real part ART bases, the angular and radial moment order is selected up to 3 and 12.

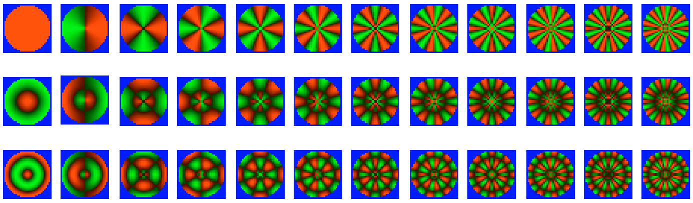

Figure 3: An example of real part ART bases with moment order up to 3 and 12. Red colour (darker part) indicates positive value, green colour (lighter part) indicates negative value.

5 Training Support Vector Classifiers

5.1 Sample Preparation

We collected face samples from the undergraduate student database in our department. We also collected face samples from Corel database with various scales and appearance. We selected some 600 frontal view upright positioned faces, and subsequently rotated them up to 14 degrees in both clockwise and counter-clockwise directions. All the face samples were clipped in a rectangular shape keeping the aspect ratio roughly to height : width = 1.25 : 1. A total of 2083 face samples were then re-scaled to a uniform scale 60 x 80 pixels in size.

Non-face samples included initial non-face samples and false positive candidates generated from the bootstrapping process [11]. The initial set of non-face samples was collected from Corel database over a broad range of categories. The categories we used include scenery, texture, surface, fur, desert, underwater shimmer, art and skin. More than 3000 non-face samples were chosen to form an initial negative training set. The aspect ratio and scale of each non-face was kept consistent to that of the face samples.

5.2 Bootstrapping Process

We use the bootstrapping process [11] to make up for the deficiency of non-face samples. Such method is important in SVM training. The false positive candidates were collected and used as non-face training samples. False positive samples are those non-faces but mistakenly detected as faces by the classifier. A total of 2543 non-face samples are generated as a result of the bootstrapping process. Table 1 shows the training kernel types and training results in each model.

SVM Classifier	**Kernel Type**	**Bound Support Vectors**	**Non-Bound Support Vectors**	**Mis-Classified Samples**	**Mis-Classification Rate**
$I * G^1_{0,0}$	**RBF**	**1368**	**964**	**347**	**4.53%**
$I * G^1_{1,0}$	**RBF**	**1348**	**1128**	**338**	**4.41%**
$I * G^1_{0,90}$	**RBF**	**1832**	**1472**	**394**	**5.14%**
$L^1(I)$	**RBF**	**1590**	**1322**	**468**	**6.10%**
2nd Level Classifier	**Linear**	**194**	**6**	**66**	**0.86%**

Table 1: Kernel type and support vectors in each classifier. A total of 7666 samples were collected consisting of 2083 face and 5583 non-face samples.

6 Experimental Results

Our system accepts both colour and grey scale test images (chromatic information was not used in the results presented here). We employed exhaustive search by moving evaluation windows of various sizes over the whole image. 18 different face sizes were used, the smallest was 25 x 31 pixels the biggest was 149 x 186 pixels. Each consecutive scale was about 10% larger or smaller in size. Before being operated on by the Gaussian derivative filters, the input image patch was normalised to 32 x 32 pixels. Luminance correction and histogram equalisation techniques were in turn used to eliminate uneven lighting conditions and to remove variations in image brightness and contrast. Eventually, the ART transform of 4 images (the original, two first order derivatives and the Laplacian) were then evaluated by the 2-level SMV classifiers.

We first tested the method on our own database. Our database consisted of two types of images. The first type was synthetic. We used various images from the Corel photo CD as background and embedded the students head and shoulder faces (outside the training set) in them. Before used to form testing images, these faces have been subjected to various distortion processing. The second type of images was photographs contained people and various backgrounds. Examples can be seen in Figure 5. There were 81 images with 308 faces in them (we will make this testing database available to researchers). The result on this database was, 295 faces were correctly detected, achieving a detection rate of 95.8%. There were 55 false positives with respect to 14,406,020 evaluated patterns. The false positive rate was as low as only 1 in every 261,927 patches evaluated.

To extend our test coverage, we performed tests on the CMU database as well. The ROC curve of the detector is shown in Figure 4. The face detection rate was 66.6% when the false positive rate reduced to zero; the face detection rate reached 90% when the false positive rate was only 3.83E-06 (against 19,845,419 evaluated patterns). Figure 5 shows examples of face detection results (Please note that the detection square may not clearly visible in printout hardcopies, in that case please refer to the electronic copy). Inspecting the actual images of our detection results, and comparing them with those published by other authors using the same database, reveals that our results are at least comparable and sometimes better than the results available in the published papers in terms of correct detection, the accuracy of the detected face sizes and their exact locations.

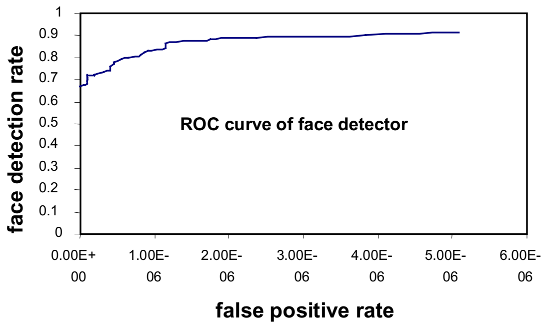

Figure 4: ROC curve of the face detector.

7 Concluding Remarks

A new method using multiple support vector machines for human face detection has been developed. We have presented very encouraging results of the method. Our method is very general. It is not only suitable for face detection but also general object detection as well. The features used by the first level SVMs are not restricted to derivative images, other suitable features, such as Gabor or Wavelet filter outputs can also be used. More importantly, the first level SVMs can be used as feature selectors to determine the usefulness of various features in a particular application, which will in turn help reduce the complexity of the method and improve its performance. We are currently working along this direction and will present results to the conference.

References

[1] Ming-Hsuan Yang, David Kriegman and Narendra Ahuja. Detecting Face in Images: A Survey. IEEE Transaction on PAMI, vol. 24, no. 1, pp. 34-58, January 2002.

[2] Henry Schneiderman and Takeo Kanade. A Statistical Method for 3D Object Detection Applied to Faces and Cars. IEEE CVPR, Hilton Head, USA, 2000.

[3] Henry A. Rowley, Shumeet Baluja and Takeo Kanade. Neural Network-Based Face Detection. IEEE Transaction on PAMI, vol. 20, no. 1, pp. 23-38, January 1998.

[4] E. Osuna, R. Freund, and F. Girosi. Training Support Vector Machines: An Application to Face Dectection. Proceedings of IEEE CVPR, Pages 130-136, 1997.

[5] Yongmin Li, Shaogang Gong and Heather Liddell. Support Vector Regression and Classification Based Multi-view Face Detection and Recognition. IEEE International Conference on Face Gesture Recognition, Grenoble France, March 2000.

[6] Bernd Heisele, Purdy Ho and Tomaso Poggio. Face Recognition with Support Vector Machines: Global versus Component-based Approach. IEEE International Conference on Computer Vision, Vancouver Canada, July 2001.

[7] Miroslaw Bober. MPEG-7 Visual Shape Descriptors. IEEE Transaction on Circuits and Systems for Video Technology, Vol. 1, No. 6, June 2001.

[8] W. Freeman and E. Adelson. The Design and use of Steerable Filters. IEEE Transactions on PAMI. Vol. 13, No. 9, pp. 891-906, 1991.

[9] Bernt Schiele and James Crowley. Recognition without Correspondence using Multidimensional Receptive Field Histograms. International Journal of Computer Vision, 36(1), 31-52, 2000.

[10] Matthew Turk and Alex Pentland. Eigenfaces for Recognition. Journal of Cognitive Neuroscience, 3(1):71-86,1991.

[11] Kah-Kay Sung and Tomaso Poggio. Example-Based Learning for View-Based Human Face Detection. IEEE Transactions on PAMI, vol. 20, no. 1, January 1998.

[12] H. Murase and S. Nayar. Visual Learning and Recognition of 3-D Objects from Appearance. International Journal of Computer Vision, 14(1), 5-24, 1995.

[13] Dengsheng Zhang and Guojun Lu. A Comparative Study of Three Region Shape Descriptors. Digital Image Computing Techniques and Applications, 21-22, 2002.

[14] Nello Cristianini and John Shawe-Tayor. An introduction to Support Vector Machines: and Other Kernel-based Learning Methods. Cambridge Uni. Press, 2000.

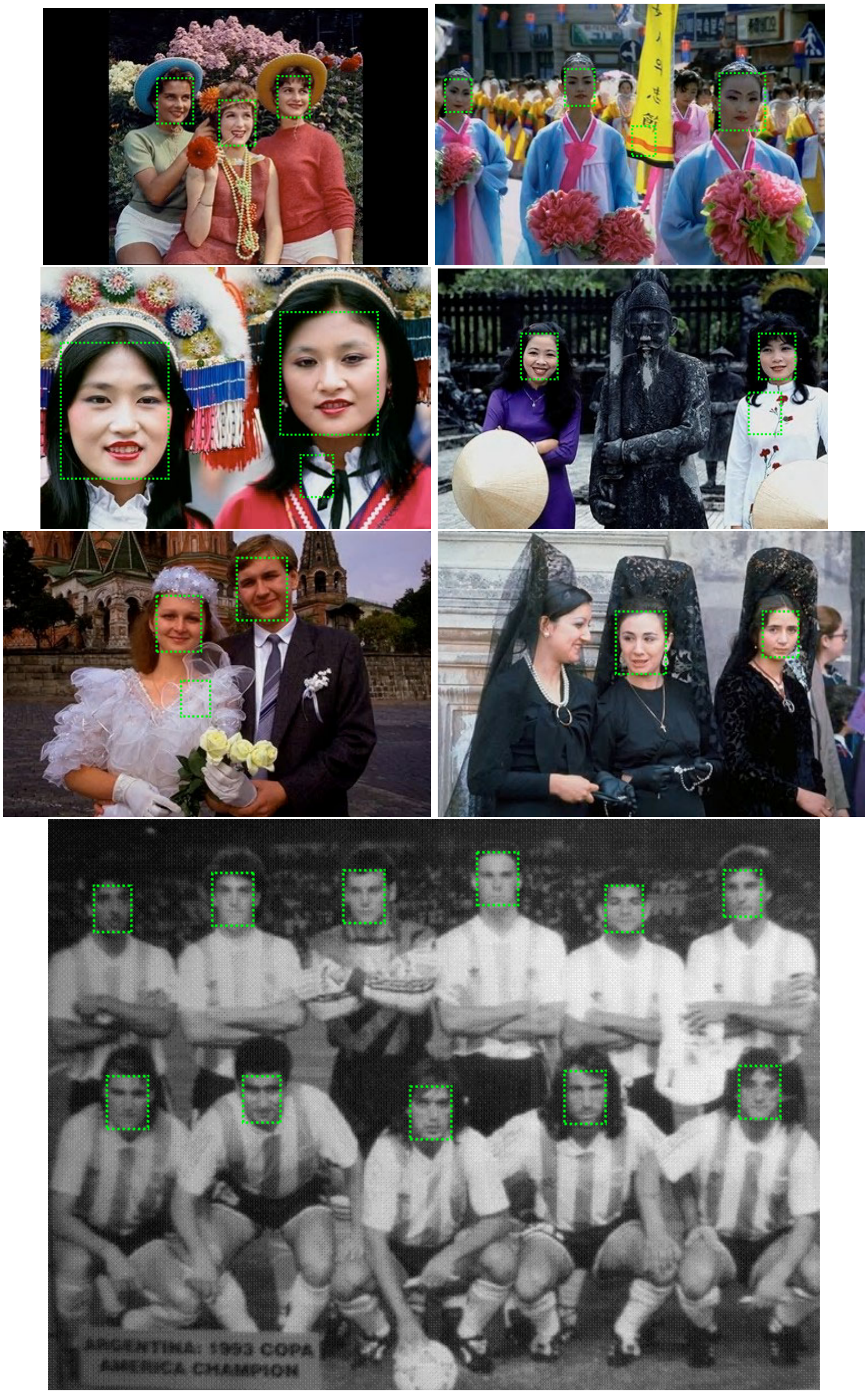

Figure 5: Examples of detection results (contd.)

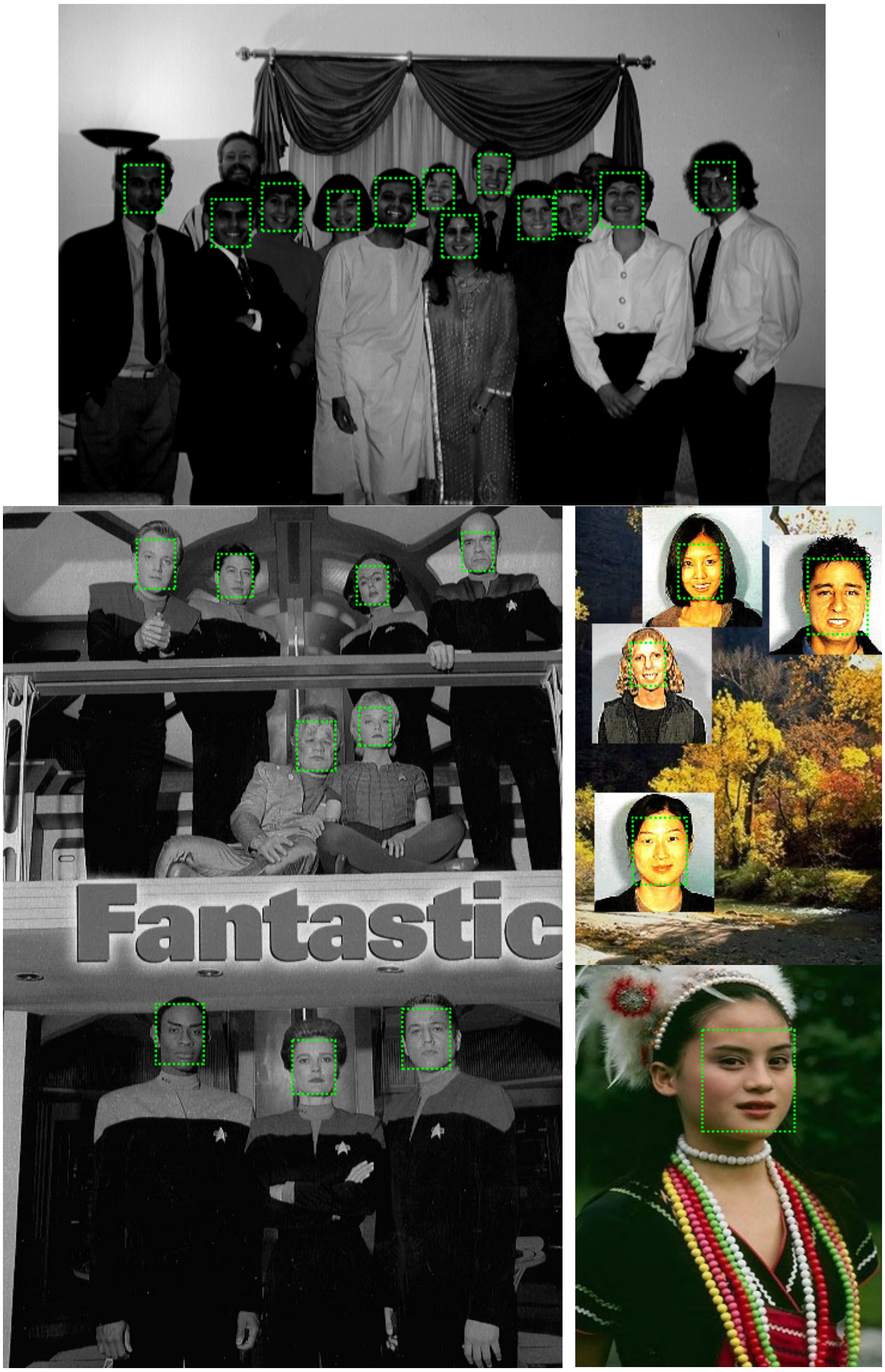

Figure 5: Examples of experimental results. The grey scale images were from the CMU database and the colour ones were our own testing data (notice that chromatic information was not used in the actual detection). Please note that the detection square may not clearly visible in printout hardcopies, in that case please refer to the electronic copy.