

Automatic Human Face Detection for Content-based Image Annotation

Richard M. Jiang, Abdul H. Sadka, Huiyu Zhou

Centre for Media Communications Research,
Brunel University, Uxbridge, London, UK

ABSTRACT

In this paper, an automatic human face detection approach using colour analysis is applied for content-based image annotation. In the face detection, the probable face region is detected by adaptive boosting algorithm, and then combined with a colour filtering classifier to enhance the accuracy in face detection. The initial experimental benchmark shows the proposed scheme can be efficiently applied for image annotation with higher fidelity.

1. INTRODUCTION

Content-based image annotation [1-2] has recently attracted a lot of interest from the research community, mainly due to the development of interactive multimedia technology over internet, wireless multimedia communication, and digital media broadcasting. In content-based image annotation, image analysis is the primary step to extract useful information from image databases. The difficulty in content-based image retrieval is how to summarize the low-level features into high-level or semantic descriptors to facilitate the retrieval procedure. Such a shift toward semantic visual data mining paradigm thus generates an urgent need to bridge the low level features with semantic understanding of the observed visual information.

To solve such a “semantic gap” problem [2], an efficient way is to develop a number of classifiers to identify the presence of semantic image components that can be connected to semantic descriptors. Among various semantic objects, human face is a very important semantic content [3], which is usually also the most concerned centric element in many images and photos. As shown in Fig.1, the presence of faces can usually be correlated to specific scenes with semantic inference according to a given ontology. Therefore, face detection can be an efficient tool to annotate images for semantic descriptors.

Face detection seems extremely easy for the human visual system. However, it is still a complex problem in computer-based systems. The difficulty resides in the fact that faces are non rigid objects. Face appearance may vary between two different persons but also between two photographs of the same person, depending on the light conditions, the emotional state of the subject and pose. Faces also vary apparently with added features, such as glasses, hat, moustache beards and hair style.

Colour based approach has been developed for face detection as an intuitive and efficient approach [4-12]. In comparison with geometric feature-based approach, detection algorithms using holistic representations have the advantage of finding small faces, pose-variant faces, or faces in poor-quality images [4-10]. A combination of holistic and feature-based approaches [6] is a promising approach to face detection. However, the colour-based approach face difficulties in robustly detecting skin colours in the presence of complex background

and different lighting conditions [8]. Jain [9] proposed a robust face detection algorithm by combining holistic features with geometry features and semantic facial component maps of eyes, mouth using a parametric ellipse. Aachen (RWTH) [10] combines face colour approach with principle component analysis (PCA) [12, 13] to achieve a high reliability in face detection. MIT [11] also exploit colour cues for face recognition while shape cues are degraded.

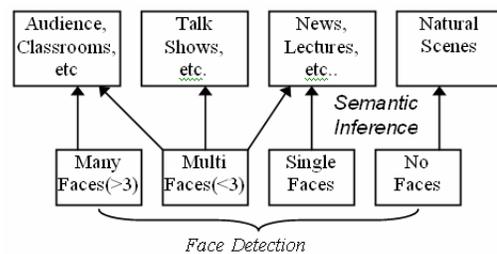


Fig.1 A Semantic Ontology for Image Annotation

Among geometric feature-based approach, adaptive boosting presented by Jones and Viola in 2001 [14] achieves a great success in face detection. AdaBoost is an aggressive learning algorithm which produces a strong classifier by choosing visual features in a family of simple classifiers and combining them in cascaded sequence. The family of simple classifiers contains simple rectangular wavelets which are reminiscent of the Haar basis. However, geometric feature based Haar wavelet classifier usually fails to discern false positive regions which have a complex textures in clothes or backgrounds, as shown in Fig.2.

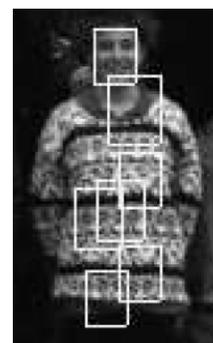


Fig.2 AdaBoost classifier fails in false positive test [15].

In this paper, in order to enhance the confidence in face detection for image annotation, colour-based facial classifier is combined with cascaded Haar classifier to reduce false positive problems. Combined with colour-based parametric ellipse used in [8], the new scheme also adds detected profile-view faces complementary while AdaBoost misses due to its frontal-view training database.

In following passages, Section II introduces the Ada-Boost Haar classifier, Section III describes the colour-based face classifier and combined approach, Section IV gives the results, and Section V concludes this paper.

2. ADABOOST HAAR CLASSIFIER

In the machine learning community it is well known that more complex classification functions yield lower training errors yet run the risk of poor generalization. AdaBoost uses a cascaded scheme, as shown in Fig.3, to overcome this problem and achieve the best of both worlds: high detection rates and extremely fast classification [14]. The final classifier can be expressed as

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

Where, h_t is weak Haar classifiers in cascaded scheme, as shown in Fig.3.

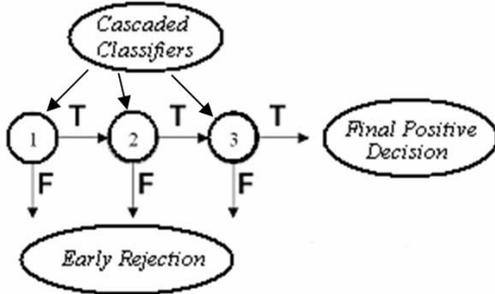


Fig.3 Cascaded Haar Classifiers in AdaBoost

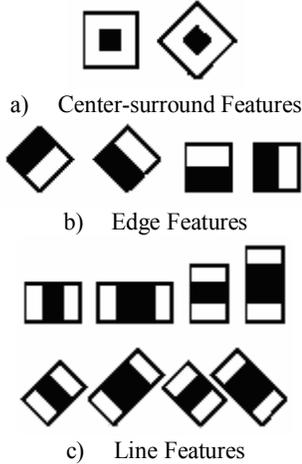


Fig.4 Haar Rectangular Features for Face Detection

In the cascaded classifier, using rectangular Haar features is a commonly adopted strategy for face detection. The wavelet function corresponding to Haar wavelet is:

$$\varphi(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq \frac{1}{2} \\ -1, & \text{if } \frac{1}{2} \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Haar features can be classified as three types, as shown in Fig. 4, including center-surround features, edge features, and line features. Integral image technique which yields a fast feature computation is applied to extract Haar-like features for the classifier training.

3. COLOUR BASED FACE DETECTION

3.1 Facial Colour Modelling

In the face detection using colour, the prior probability of a pixel belonging to a face can be estimated by its feature distance to the learned average value, usually described by Gaussian distribution, as,

$$P(f_{x,y}^i | \Lambda) = \sum_k \prod_i N(\xi_i^k, \sigma_i^k) \quad (2)$$

Λ is the observation, $f_{x,y}^i$ is the pixel features at $\{x, y\}$, k denotes the k -th facial colour model. ξ_i^k and σ_i^k are the model parameters learned in training procedure.

In the case of using skin colour as features of facial points, the RGB properties are usually converted into $YCrCb$ format. Because Y is closely related with surrounding illumination, only Cr and Cb are usually counted as particular features for the evaluation.

In Bayesian description of a real world, the event of a pixel belonging to a face is jointly connected with the probabilistic results of its neighbour pixels, which can be described as,

$$P(f_{x_0,y_0}^i | f_{x,y}^i, \Lambda) = \frac{\prod_{x,y \in \mathfrak{N}} P(f_{x,y}^i | f_{x_0,y_0}^i, \Lambda)}{\prod_{x,y \in \mathfrak{N}} P(f_{x,y}^i | \Lambda)} P(f_{x_0,y_0}^i | \Lambda) \quad (3)$$

\mathfrak{N} is the collection of neighbourhood pixels, e.g., a region of 3×3 pixels in this paper. With above joint inference, the single isolated face-like points will be precluded from the probabilistic inference. In total, the likelihood of a region \mathfrak{R}_k as a collection of face-like N points can be tentatively given as the total probability,

$$P(\mathfrak{R}_k | (f_i)_{x,y}) = \frac{1}{N} \sum_{x,y \in \mathfrak{R}} P(f_{x_0,y_0}^i | f_{x,y}^i, \Lambda) \quad (4)$$

which is the average probability of all detected points in this region.

3.2 Face Region Pre-selection

After a possible face region is detected by its colour features, a further decision diagram can be performed by matching the region to the average facial features, such as shape, lips, ears, nostril region, eyebrow, and eye regions.

In this work, we use the shape as one morphological criterion for further evaluation of the detected skin-similar region. Basically, a face can have a shape of ellipse. While the face region is detected, its contour should be fit into an elliptical

shape, with an acceptable axis ratio. The distance between the ellipse and the region shape is computed using the Hausdorff distance measure. In our work, we use a Gaussian model to evaluate the likelihood ρ_{shape} of the obtained Hausdorff distance as an acceptable facial component, rather than simply using a distance threshold to reject the detected regions. If the total likelihood is lower than a user-defined minimum accepted likelihood, the region is rejected. If the region has a face-similar shape, an artificial neural network (ANN) classifier[18] trained by 20 profile-view faces and 20 non-face samples is employed to testify if it is a profile-view face.

3.3 Face Detection Fusion Scheme

In a fusion scheme as shown in Fig.5, it is easy to combine geometric feature based AdaBoost classifier with colour-based facial region likelihood classifier, which are combined at image level, feature level and scoring level separately.

In the fusion scheme, the detected regions in the AdaBoost classifier are re-examined in correspondence to facial likelihood map at image level and feature level. Some face-like regions that are not detected by the AdaBoost classifier are forwarded to an ANN classifier to check if it is a profile face.

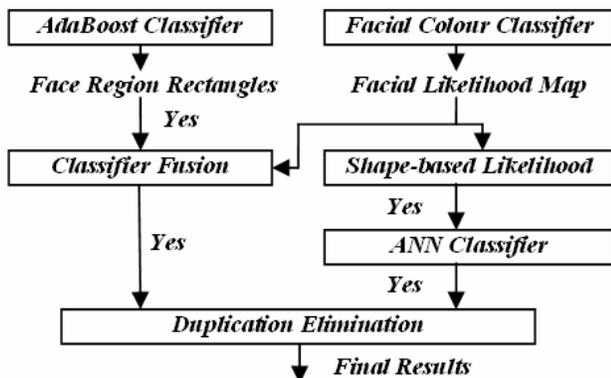


Fig.5 Face Classifier Fusion Scheme

4. EXPERIMENTAL RESULTS

The proposed algorithm in this paper was implemented by C++ language, and compiled into executable files. The software runs on Windows-XP with Intel 3.00GHz Pentium microprocessor. Fig.7 is the benchmark results.

We use AT & T database to train AdaBoost classifier, and selected 20 profile-view images to train the ANN classifier. Face databases designed for face recognition usually contains grey-scale images. In order to facilitate the program benchmark with colour face detection, a set of colour images taken from QMUL database [16] and other colour images downloaded from web internet were used as test bench in this paper.

As shown in Fig.7-a), after the facial colour filtering, a map of facial colour likelihood is generated, while a number of face-like regions can be easily clustered.

In Fig.7-b), while the AdaBoost Haar classifier detected two false region due to their similarity in their geometric components, as shown in the left column image, the wrong detection can be corrected by checking their facial likelihood map in these

rectangles according to eq.(4). Precluded by their total colour likelihood, the wrong detected regions were then removed from possible face-like candidates, as shown in right column image in Fig.7-b).

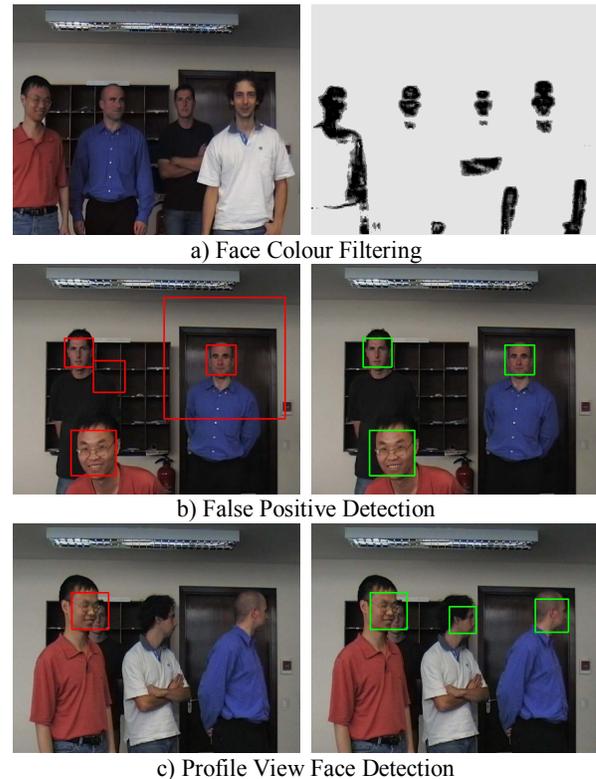


Fig.6 The experimental results. Left column in b) and c) is the initial results detected by AdaBoost algorithm, and right column is the adjusted results after classifier fusion.

Fig.7-c) demonstrates the detection of the profile-view faces. While lacking of profile-view database, the AdaBoost classifier failed to detect those profile-view faces. After the segmented face-like regions are evaluated by its shape likelihood, the possible rectangular regions were forwarded to profile-view face classifier trained by a small database set. Corrected with this complementary classifier, as shown in Fig.7-c), two undetected faces in left column image were annotated in left column image.

In image annotation, if a classifier detects 75 faces in 100 faces, and detect 30 false positive non-face regions as face in 100 non-face images, its total correct annotation rate is then only 60%, making the classifier as a useless cue for image retrieval because it's almost identical to 50%-50% dicing. Therefore, high accuracy is critically demanded by efficient image annotation.

From the experimental results, it is verified the fusion approach can efficiently improve the face detection rate and reduce the false positive detection rate. The initial experiment shows the above fusion scheme can be useful to supply a reliable face classifier for image annotation.

5. CONCLUSIONS

In conclusion, a robust face detections scheme combining the cascaded Haar classifier with face colour based classifier is

presented in this paper, which successfully enhanced the total performance in face detection and reduced the rate in non-face region false positive detection. The initial experimental benchmark shows the proposed scheme can supply higher classification fidelity for content-based image annotation, which requires accurate relevance feedback in image retrieval.

REFERENCES

- [1] Wei Jiang, Guihua Er, Qionghai Dai, Jinwei Gu, "Similarity-based online feature selection in content-based image retrieval", *IEEE Trans. Image Processing*, Volume 15, Issue 3, March 2006 Page(s):702 – 712.
- [2] Krishnapuram, R.; Medasani, S.; Sung-Hwan Jung; Young-Sik Choi; Balasubramaniam, R.; "Content-based image retrieval based on a fuzzy approach", *IEEE Trans. Knowledge and Data Engineering*, Vol. 16, Issue 10, Oct. 2004, pp.1185.
- [3] Y. Sheng, A. H. Sadka, A. M. Kondoz, "Automatic 3D face synthesis using single 2D video frame", *Electronics Letters*, Vol. 40, Issue 19, 16 Sept. 2004, pp.1173;
- [4] O. Schreer, S. Askar, Y. Kondratyuk, K. Elazouzi, P. Kauff, "Vision-based skin-colour segmentation of moving hands for real-time applications", 1st European Conference on Visual Media Production (ECVMP), 15-16 March 2004, p.79.
- [5] E. Saber, A.M. Tekalp, "Frontal-view face detection and facial feature extraction using colour, shape and symmetry based cost functions", *Pattern Recognition Letters*, 19(8), pages 669-680, 1998.
- [6] Chia-Feng Juang, Shih-Hsuan Chiu, Shen-Jie Shiu, "Fuzzy System Learned Through Fuzzy Clustering and Support Vector Machine for Human Skin Color Segmentation", *IEEE Tran. Systems, Man, & Cyber.*, Vol. 37, No. 6, Nov. 2007, p.1077.
- [7] K. Lam, H. Yan, "An Analytic-to-Holistic Approach for Face Recognition Based on Single Frontal View", *IEEE Tran. PAMI*, Vol.29, No.5, 1996, p.771.
- [8] R. Feraud, O. J. Bernier, "A Fast and Accurate Face Detection based on Neural Network", *IEEE Tran. PAMI*, Vol.23, No.1, Jan. 2001, p.42.
- [9] Rein-Lien Hsu, M. Abdel-Mottaleb, A.K. Jain, "Face detection in color images", Vol.24, Issue 5, 2002, p.696.
- [10] B. Menser, F. Muller, "Face detection in colour images using principal components analysis", *Seventh International Conference on Image Processing and Its Applications*, Vol.2, 13-15 July, 1999, p.620.
- [11] Andrew Yip, Pawan Sinha, "Role of color in face recognition," MIT tech report (ai.mit.com) AIM-2001-035 CBCL-212, 2001.
- [12] P. Belhumeur, J. Hespanha, D. Kriegman, "Eigenfaces versus fisherfaces: Recognition using class specific linear projection", *IEEE Trans. Pattern Analysis and machine Intelligence*, 19 (7), 1997, p.711.
- [13] K-K Sung, "Learning and Example Selection for Object and Pattern Detection", PhD Thesis, Massachusetts Institute of Technology, 1996.
- [14] Paul Viola, Micheal Jones, "Robust real-time object detection", *Second International Workshop on Statistical Learning and Computational Theories of Vision Modeling, Learning, Computing and Sampling*, July 2001.
- [15] Julien Meynet, "Information theoretic combination of classifiers with application to face detection", PhD Thesis, EPFL, no 3951 (2007).
- [16] <http://www.elec.qmul.ac.uk/mmv/>, QMUL Face Detection Database.
- [17] F. Samaria, "Face Recognition Using Hidden Markov Models", PhD Thesis, University of Cambridge, 1994.
- [18] Y. Patrice, Y. Simard, L. Bottou, P. Haffner, L. Yann, "Boxlets: a fast convolution algorithm for signal processing and neural networks", *Advances in Neural Information Processing Systems*, Vol.11, p.571, 1999.
- [19] E. Osuna, E. Freund, F. Girosi, "Training support vector machines: an application to face detection", *Proc. Computer Vision and Pattern Recognition*, 1998, p.45.
- [20] V. Pavlovic, A. Garg, "Efficient Detection of Objects and Attributes using Boosting", *IEEE Conf. Computer Vision and Pattern Recognition*, 2001.
- [21] K. C. Yow, R. Cipolla, "Scale and orientation invariance in human face detection", *Proc. British Machine Vision Conference*, p.745, 1996.
- [22] V. Govindaraju, S. N. Srihari, D. B. Sher, "A computational model for face location", *Proc. of the Third International Conference on Computer Vision*, p.718, 1990.
- [23] Y. Moses, Y. Adini, S Ullman, "Face Recognition: the problem of the compensating for changing in illumination direction", *European Conference on Computer Vision*, p.286, 1994.
- [24] T. Leung, M Burl, P. Perona, "Finding Faces in cluttered scenes using labelled random graph matching", *Proc. 5th Int. Conf. on Computer Vision*, p.637, MIT, Boston 1995.
- [25] M. Indovina, U. Uludag, R. Snelick, A. Mink, A. Jain. "Multimodal Biometric Authentication Methods: A COTS Approach", *Proc. MMUA 2003, Workshop on Multimodal User Authentication*, p.99, Santa Barbara, CA, December 11-12, 2003.
- [26] A. Ross, A.K. Jain, "Information Fusion in Biometrics", *Proc. of AVBPA, Halmstad. Sweden*. p.354, June 2001.
- [27] B. Achermann, H. Bunke, "Combination of classifiers on the decision level for face recognition", *Technical Report IAM-96-002, Insitut fur Informatik und angewandte Mathematik, Universitat Berlin, Berlin*, 1996.
- [28] B. Brunelli and D. Flavigna. *Personal identification using multiple cues*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10), p.955, 1995.
- [29] L. Hong and A. Jain. *Multimodal biometrics*. In A. Jain, R. Bolle, and S. Pankanti, editors, *Biometrics: Personal Identification in Networked Society*. Kluwer, 1999.
- [30] J. Kittler, M. Hatef, R. Duin, and J. Matas. *On Combining Classifiers*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), p.226, March 1998.