REAL-TIME GENDER RECOGNITION FOR UNCONTROLLED ENVIRONMENT OF REAL-LIFE IMAGES

Duan-Yu Chen and Kuan-Yi Lin
Department of Electrical Engineering, Yuan-Ze University, Taiwan

Keywords: Gender recognition, Uncontrolled environment, Real-life images.

Abstract: Gender recognition is a challenging task in real life images and surveillance videos due to their relatively low-resolution, under uncontrolled environment and variant viewing angles of human subject. Therefore, in this paper, a system of real-time gender recognition for real life images is proposed. The contribution of this work is fourfold. A skin-color filter is first developed to filter out non-face noises. In order to make the system robust, a mechanism of decision making based on the combination of surrounding face detection, context-regions enhancement and confidence-based weighting assignment is designed. Experimental results obtained by using extensive dataset show that our system is effective and efficient in recognizing genders for uncontrolled environment of real life images.

1 INTRODUCTION

Gender recognition is a challenging task in real life images and surveillance videos due to their relatively low-resolution, under uncontrolled environment and variant viewing angles of human subject. To recognize the gender of a human subject, the selection of a set of effective features on an appropriate part(s) of human body is necessary. The face of a human subject contains some information and could be a useful clue for recognizing emotion and facial expressions by (Wang et al. 2004). On the other hand, Andreu et al. 2009 apply a partial view of face for gender recognition. In (Andreu et al., 2009), they consider the eyes zone to recognize gender by using local feature vectors. In (Gallagher and Chen, 2009), they combine social context with appearance to recognize gender. In addition, the full body (Cao et al., 2008) of a human subject that provides the silhouette of a person was adopted for gender recognition. In the literature, some edge-based features were extracted from face zones, such as Haar-like features (Shen et al., 2009)(Lu and Lin, 2007), Gabor wavelets (Lin et al., 2006), LBP (Lian and Lu, 2007), LUT (Wu et al., 2003-2004) and quantized edge features (Lu et al., 2003). For color-based features, the PCA (Balci and Atalay, 2002)(Rodrigo et al., 2006)(Fang and Wang, 2008) and NNM (Nikolaus, 2007)(Lee and Seung, 1999) of relatively higher computation complexity are well known methods for analyzing the facial characteristics. However, for real-time applications, a feature set that is of light computational cost is unavoidable.

In the related works, most approaches focus on recognizing the gender of human subjects in the images obtained under well-controlled lighting condition and pure non-textured background. Besides, the face of human subjects captured is frontal and of high resolution. For the faces obtained from daily life images and surveillance videos, the resolution is relatively much lower than those from ID photos. Approaches of gender recognition proposed for ID photos could not work well for low-resolution ones. Under these circumstances, the face information could be insufficient. In order to tackle this problem, some researchers tried to extract features either from the internal or the external face zones, or both of them. (Lapedriza et al. 2005) recognized the gender by using the external face features. In (Lapedriza et al. 2006), features are computed from the external and internal face zones. The internal features composed by eyes, nose and mouth and the external features located in head, ears and chin. The fragment-based face features are thus extracted. It has proved that the external face zone can provide rich information for gender recognition. Therefore, in our proposed approach, both the internal and external face zones are our concern for feature extractions.

When a feature set is ready, a set of training dataset and an effective training approach are necessary. Mayo and Zhang aim to collect extensive

Chen D. and Lin K. (2010).
REAL-TIME GENDER RECOGNITION FOR UNCONTROLLED ENVIRONMENT OF REAL-LIFE IMAGES.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 357-362
DOI: 10.5220/0002823203570362

dataset and thus add deliberately misaligned faces (Mayo and Zhang, 2008) to improve the accuracy of gender recognition. Several well-known training approaches are like support vector machines (SVM) (Osuna and Freund, 1997)(Moghaddam and Yang, 2000), Adaboost (Freund and Schapire, 1996)(Schapire and Singer, 1999), and neural network, and the classifier obtained by Adaboost algorithm is proved to be the most efficient one among them (Shen et al., 2009). Since the real-time performance is the concern of our proposed approach, Adaboost approach is adopted in our work.

In this paper, in order to recognize the gender of human subjects under uncontrolled environment, a novel mechanism is proposed. First, we calculate the RGB ratio of a racial skin to eliminate the noises from the face detector. The context-regions enhancement, which enhances some regions related in spatial domain, is developed to make the evidence of these regions stronger than the original ones. To deal with the problem of variant viewing angles of human subjects, a voting strategy, in which gender information is collected from the surrounding faces of the original one, is weighted by the novel confidence ratio.

The remainder of this paper is organized as follows. In Section II, we describe the approaches of the face detection and skin-color-filter. Section III shows the proposed feature set and Section IV introduces the mechanism of gender recognition. We then detail our experimental results in Section V and present some concluding remarks in Section VI.

2 FACE DETECTION AND FACE FILTERING

2.1 Face Detection

Since face-based gender recognition is our concern, it is necessary to detect the frontal or near frontal faces. Therefore, face detection is the first important step to be accomplished. To satisfy the real-time requirement, we use the Viola and Jones face detector, which can provide fast detections of face regions. The face detector is trained by using Adaboost algorithm. Haar-like features including edge features and center surround features are extracted, in which integral images are employed for efficient computing. In order to detect the region rotated 45 degrees, we add the Haar-like features that rotate 45 degrees. The details of the face detector can be found in (Viola and Jones, 2001)

2.2 Skin-color Filter

The face detector can achieve success rate of 80%-90%. However, their detection rate of false positives is in the 10%-20% range. Most false positives are detected as faces due to their patterns of Haar-like features are highly similar to the real faces. To distinguish between real faces and non-faces, skin colors are important features for noise filtering. Considering the different skin colors of different races, one race from another should have distinct color characteristics. In this work, we focus on the race of Asians.

We compute the RGB ratio $P(r_a)$ from face images as follows:

$$\mu_{\mathrm{P(r)}} = \frac{1}{K}\sum_{a=0}^{K} P(r_a) \text{ ,} \tag{1}$$

$$P(r_a) = \frac{r_a}{r_a + g_a + b_a} \tag{2}$$

where the r_a , g_a , b_a are one pixel RGB values from face images. The mean $\mu_{\mathrm{P(r)}}$ of the RGB ratio is defined. K is the pixels from faces. The variance $\sigma^2_{P(r)}$ from the faces RGB ratio can thus be computed by

$$\sigma^2_{P(r)} = \frac{1}{K}\sum_{a=0}^{K}[P(r_a) - \mu_{\mathrm{P(r)}}]^2 \,. \tag{3}$$

According to the RGB color distribution, a face candidate is considered as a non-face if its color distribution was out of the range of 2 $\sigma_{\mathrm{p(r)}}$.

3 FEATURE EXTRACTION

In this section, we shall present the hybrid feature set and the training algorithm used in the proposed system. In the hybrid feature set, simple but effective block-based color and edge features are computed. Furthermore, the efficient algorithm of Adaboost is employed for training purpose.

3.1 A Hybrid Feature Set

To show the hybrid feature set, the color feature and the edge feature are demonstrated in Figs.1(a)-(b), respectively. In the first step, we transform the RGB color images into gray-scale images and then divide into 8×8 blocks for each image. The quantized gray-scale in one block region g , which is classed templates ℓ is defined as follows:

$$\omega^{\ell}_{(i,j)} = \mathrm{INT}(\tau^{\ell}_{(i,j)} / \rho), \tag{4}$$
$$\rho \in 2^{\phi}, 0 < \phi < 8; \phi \in N,$$

where the $\tau^{\ell}_{(i,j)}$ and $\omega^{\ell}_{(i,j)}$ are the original gray-scale value and the quantized gray-scale one in the coordinate (i, j), respectively.

To compute the representative color for a block, the resulting color of a block is defined by

$$J^{\ell} = \sum_{j=1}^{c}\sum_{i=1}^{c} p(\omega^{\ell}_{(i,j)} \mid \ell) p(\ell), \quad (5)$$

$$G_d = \max(J^1, J^2, ..., J^{\ell}), \quad (6)$$

where the value J^{ℓ} is one template probability of the pixels in one block. The representative color G_d is determined by choosing the color with maximum probability.

To compute the edge features, Canny edge detector is employed to conduct the edge computing in one block. The feature vectors E are computed and considered as templates γ by:

$$E = \max(o^1, o^2, ..., o^{\gamma}), \quad (7)$$

where the sum of edge vectors o^{γ} in one block, which combines with $N \times N$ mask is defined by:

$$o^{\gamma}_{(i,j)} = \sum_{i=i-(N-1)/2}^{i+(N-1)/2} \sum_{j=j-(N-1)/2}^{j=j-(N-1)/2} \hbar^{\gamma}(i, j), \quad (8)$$

where γ is the type of template, and the binary function $\hbar^{\gamma}(i, j)$ is value '1' if this pixel value is not zero. Otherwise, the pixel is set to '0' if its value is zero.

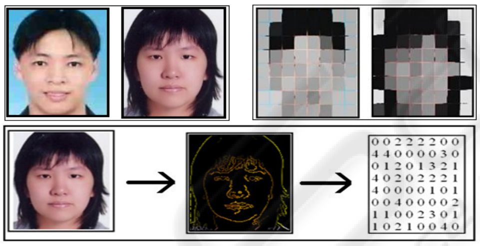

Figure 1: The process of (a) color and (b) edge features computation.

3.2 Training Mechanism

In the training mechanism, we collect 1948 face images to be the training dataset. To overcome the problem of the variant viewing angles of human subjects, we simulate the actual rotated faces by cutting a portion of the right or the left side face regions.

We then adopt the set of color and edge features, in which the total dimension is 128 and 64 for each feature in our experiment, for the Adaboost training algorithm. The algorithm of Adaboost is shown as follows.

Adaboost algorithm.

Input:

(1) n_+ female face images and n_- male face images. The face image label y_i is '1' for female or '0' for male as follows:

$\{(s_1, x_1), (s_2, x_2), ..., (s_i, x_i)\}$

(2) $X = \{s_1, s_2, ..., s_i\}$ are images with d dimension feature vectors

(3) **Initialization:**

The weight of training examples

$D_1(i) = 1/m, i = 1, ..., m.$

(4) For **weak classifiers** $t = 1, ..., T$.

1. Find the classifier φ_t that minimizes the $D_t(i)$ weighted error
2. $\varphi_t = \arg\min_{h_j \in H} \varepsilon_j$,
   where $\varepsilon_j = \sum_{i=1}^{m} D_t(i)$ (for $y_i \neq \varphi_j(x_i)$)
   as long as $\varepsilon_j < 0.5$;
   else quit
3. Set the φ_t voting weight $\alpha_t = 0.5 \times \log[\frac{1-\varepsilon_t}{\varepsilon_t}]$,
   where ε_t is the arg min error
   from step 2.
4. Update the weight:
   $D_{t+1}(i) = [D_t(i)\exp(-\alpha_t y_t \varphi_t(x_i))] / Z_t$,
   where Z_t normalizes the equation over all data point

Output: $H(x) = \sum_{t=1}^{T} \alpha_t \varphi_t(x)$

The value of the combination of weak classifiers in Adaboost algorithm is employed and considered as the confidence of the result of gender recognition. The confidence value is further used for the voting strategy among the faces which are detected in the surrounding regions near the original detected face. The novel voting strategy is detailed in Section IV.

4 GENDER RECOGNITION

In this section, we shall describe the gender recognition based on a novel decision mechanism. In order to make the system robust, a mechanism of decision making based on the combination of surrounding face detection, context-regions enhancement and confidence-based weighting assignment is designed.

4.1 Surrounding Faces Detection

In uncontrolled environment of daily life images, the viewing angle of human subjects would vary in different orientations. A slight rotation of the face would usually result in the disappearance of some important clues from internal and external face zones. The determination of the external face zones is based on the internal face detected by the face detector. Under these circumstances, the external face zones would be extracted including more background area than those obtained from frontal faces. Therefore, to recognize the gender without re-training and to solve the misalignment problem, we detect the faces in the surrounding regions near the original face detected by the face detector, which is so called surrounding faces detection. To emphasize some important regions of a detected face, an approach so called context-regions enhancement is proposed. Furthermore, we evaluate the confidence of the gender of the face with local region enhanced and compute the linear combination of these faces weighted by the confidence value. The details of context-regions enhancement are illustrated in the following section.

4.2 Context-regions Enhancement

In this section, we describe the method of context-regions enhancement. With applying this method, we can first enhance some important face zones and also reduce some disturbance of the gender recognition.

The symmetric structure in shape is important for recognizing genders. We observe that some effective features learned from Adaboost are from the symmetric regions. Insufficient information obtained from these regions would result in fewer evidence of their corresponding gender. Therefore, an approach named context-regions enhancement is proposed to overcome this problem. We enhance the symmetric regions in external face zones by

$$\varpi \times p(I_k \mid B = I_k) \geq 1, \tag{9}$$

where ϖ is the threshold, B is the maximum cumulative density of the contrast value, and I_k is a value of a context region. If I_k satisfies Eq.(9), then I_k is replaced by

$$I_k = I^*, \tag{10}$$

where the value I^* is the enhancement parameter determined empirically. We can examine the context regions based on Eqs.(9)-(10) to verify if they possess coherent features. In contrast, if a value of a context regions satisfies

$$\varpi \times p(I_k \mid B = I_k) \leq 1, \tag{11}$$

then

$$I_k = I^{\Delta}, \tag{12}$$

where the value I^{Δ} is the reduced disturbance parameter.

4.3 Confidence Evaluation and Voting Strategy

After detecting surrounding faces of the original one, we evaluate the confidence of the gender of the face with local region enhanced and compute the linear combination of these faces weighted by the confidence value. Obtaining from the Adaboost classifier, we compute two values that are the positive and negative values for the genders and then combine them together by measuring the distance between these two values. In this way, a normalized weighting for the confidence of the gender can be obtained. The distribution of our training dataset is shown in Fig. 2 and then a fitting line is approximated for the further weighting analysis.

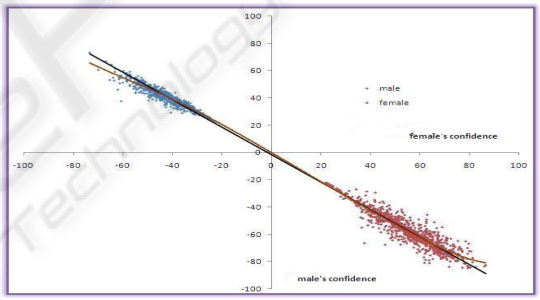

Figure 2: Distributions of the database generate from Adaboost algorithm.

From Fig. 2, we can get a fitting line and the fitting line is also called the confidence line computed by

$$ax + by + d = 0, \tag{13}$$

where a, b and c are constant. After the fitting line is obtained by Eq.(13), the extreme values of the fitting line should be estimated for normalizing confidence values. The input samples closer to the extremities of the fitting line are projected onto the line, which are $P_L(x_0, y_0)$, and the mean of the positive samples with high confidence is computed. The mean is then considered as the maximum confidence for the positive samples. The extreme value of the fitting line is formulated as:

$$B^f_{(x,y)} = \xi + \frac{1}{\eta}\sum_{L=1}^{\eta} P_L^{female} \times \bar{\lambda}^f, \tag{14}$$

where $B^{f}_{(x,y)}$ is the female's boundary position, ξ is the adjustable error, η is the numbers of female's data and P^{female}_{L} is the female's positions. When λ^{f} is '1', it means that P^{female}_{L} belongs to the high confidence values.

In contrast, the negative samples further to the origin of the fitting line are projected onto the line and the mean of these points are computed and considered as the minimum value of negative samples $B^{m}_{(x,y)}$. Confidence of the recognized gender defined by this approach can reveal its normalized value. For an unknown face, the sample is first projected onto the fitting line for measuring its confidence. The projection $\| P_L - B^{m}_{(x,y)} \| \times \cos(\theta)$ and confidence values $\mathrm{C_L}$ can be obtained by

$$\mathrm{C_L} = \frac{\| P_L - B^{m}_{(x,y)} \| \times \cos(\theta)}{\| B^{f}_{(x,y)} - B^{m}_{(x,y)} \|}, \tag{15}$$

where θ is the angle between the fitting line and a vector building $P_L(x_0, y_0)$ and $B^{m}_{(x,y)}$. After the confidence of each surrounding face is evaluated, we make use of voting strategy method as follows:

$$\Omega = \frac{1}{Q}\sum_{L=1}^{Q} C_L. \tag{16}$$

where Ω is the final result, and Q are the number of surrounding face.

5 EXPERIMENTAL RESULTS

The performance of the skin-color filter is shown in Table 1. We can observe that the error rate is reduced and the accuracy of face detection is improved to 96%. The performance of gender recognition using our proposed approach is shown in Table 2 and a SVM-based approach is compared. We recognize 469 faces which are real life images downloaded from on Google search. From Table 2, the proposed approach with novel decision mechanism achieves the best performance.

In our proposed approach, nine surrounding regions are searched if any face can be detected. The voting strategy with supported by these nine surrounding regions could improve almost 6.5% and up to 88% of recognition rate. This proves that the voting strategy from surrounding regions is effective. For the real-time applications, the execution time of the system is also critical. Thus, in Table 3, it can be observed that combining the surrounding faces for recognizing genders would cost more time. However, it shows that our proposed approach can still conduct gender recognition in the real-time manner.

Table 1: The accuracy of face detection.

	Face Numbers	Errors	Total	Accuracy
w/o SKF	469	42	511	91.78%
w/ SKF	469	**22**	**491**	**95.55%**

Table 2: The accuracy of gender recognition, where Ada. is Adaboost algorihm and CRE is context-regions enhancement.

	Male (210) Error	Female (259) Error	Error	Accuracy
SVMs	45	48	93	80.17%
Ada.+ CRE	58	30	88	81.24%
Ada.+CRE +Voting	**36**	**22**	**58**	**87.63%**

Table 3: Cost time of method in gender recognition system.

Method	SVMs	Ada. With CRE	Ada. With CRE + Voting strategy (9effects)
Time(ms)	123.51	**59.88**	540.66

Some examples of the result of gender recognition in real life images are illustrated in Fig. 3. These images are of different lighting conditions, different size of human subjects, different groups of people, etc. It can be observed that most human subjects can be recognized by their gender successfully. A few human subjects are not detected because the face detector employed does focus on frontal or near frontal faces. However, in this paper, this face detector satisfied our requirement since recognizing genders in frontal or near frontal faces is our primary concern.

6 CONCLUSIONS

In this work, a system of real-time gender recognition for real life images has been proposed.
The contribution of this work is four-fold. A skin-color filter has been developed to filter out non-face noises. In order to make the system robust, a mechanism of decision making based on the combination of surrounding face detection, context-regions enhancement and confidence-based weighting assignment has been designed. Experimental results obtained by using extensive

dataset have shown that our system is effective and efficient in recognizing genders in uncontrolled real life images.

Figure 3: Demonstration of the recognizing results.

ACKNOWLEDGEMENTS

This work is supported by the National Science Council under Contract No. NSC98-2218-E-155-001.

REFERENCES

Wang, Y., Ai, H., and Wu, B., Huang, C., 2004. Real Time Facial Expression Recognition with Adaboost. *ICPR'04, International Conference on Pattern Recognition*.

Andreu, Y., Mollineda, R.A., and Garc´ıa-Sevilla, P., 2009. Gender Recognition from a Partial View of the Face Using Local Feature Vectors. *Lecture Notes in Computer Science*.

Gallagher, A.C. and Chen, T., 2009. Understanding Images of Groups of People. *CVPR'09, IEEE Conference on Computer Vision and Pattern Recognition*.

Cao, L., Dikmen, M., Fu, Y., and Huang, T.S., 2008. Gender Recognition from Body. *ACM international conference on Multimedia*.

Shen, B.C., Chen, C.S., and Hsu, H.H., 2009. Fast Gender Recognition by Using A shared-Integral-Image Approach. ICASSP'09, *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Lu, H. and Lin, H., 2007. Gender Recognition using Adaboosted Feature. *International Conference on Natural Computation*.

Lin, H., Lu, H., and Zhang, L., 2006. A New Automatic Recognition System of Gender, Age and Ethnicity. *The Sixth World Congress on Intelligent Control and Automation*.

Balci, K. and Atalay, V., 2002. PCA for gender estimation: which eigenvectors contribute? *ICPR'02, International Conference on Pattern Recognition*.

Rodrigo, V., Javier, R. D. S., and Mauricio, C., 2006. Gender Classification of Faces Using Adaboost. *CIARP'06*.

Fang, Y. and Wang, Z., 2008. Improving LBP Features for Gender Classification. *International Conference on Wavelet Analysis and Pattern Recognition*.

Lian, H.C. and Lu, B.L., 2007. Multi-View Gender Classification Using Multi-Resolution Local Binary Patterns and Support Vector Machines. *International Journal of Neural Systems*.

Wu, B., Ai, H. and Huang, C., 2003. LUT-Based Adaboost for Gender Classification. *Audio- and Video-Based Biometric Person Authentication*.

Wu, B., Ai, H. and Huang, C., 2004. Facial image retrieval based on demographic classification. *ICPR'04, International Conference on Pattern Recognition*.

Lu, H., Huang, Y., Chen, Y. and Yang, D., 2003. Automatic gender recognition based on pixel-pattern-based texture feature. *Journal of Real-Time Image Processing*.

Nikolaus, R., 2007. Learning the Parts of Objects using Non-negative Matrix Factorization. *Term Paper*, Feb 2007.

Lee, D.D. and Seung, H.S., 1999. Learning the parts of objects with nonnegative matrix factorization. *Nature*.

Lapedriza, A., Masip, D. and Vitria, J., 2005. Are External Face Features Useful for Automatic Face Classification? *CVPR'05, Computer Vision and Pattern Recognition*.

Lapedriza, A., Marin-Jimenez, M.J. and Vitria, J., 2006. Gender Recognition in Non Controlled Environments. *International Conference on Pattern Recognition*.

Mayo, M. and Zhang, E., 2008. Improving Face Gender Classification By Adding Deliberately Misaligned Faces to The Training Data. *Image and Vision Computing New Zealand*.

Osuna, E. and Freund, R., 1997. An Improved training algorithm for Support Vector Machine. *IEEE Workshop on Neural Networks for Signal Processing*.

Moghaddam, B. and Yang, M.H., 2000. Gender Classification with Support Vector Machines. *Int'l Conf. on Automatic Face and Gesture Recognition.*

Freund, Y. and Schapire, R. E., 1996. Experiments with a New Boosting Algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference*.

Schapire, R. E. and Singer, Y., 1999. Improved boosting algorithms using confidence-rated predictions. *Machine. Learning*.

Viola, P. and Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. *IEEE Conference on Computer Vision and Pattern Recognition*.