

FACIAL GENDER RECOGNITION

FIELD OF INVENTION

5 The invention relates to a method and system for building a classifier for use in facial gender recognition, and to a method and system for facial image gender recognition.

BACKGROUND

10

Gender recognition is important as it can boost the performance of many applications, including person recognition and human-computer interfaces.

15

Face detection is the essential first step for almost all face information processing systems. Progress in face detection makes it possible for surveillance systems to take a human face as an input pattern and extract information from it.

20

Currently, the main methods used in gender classification are Neural Network and Support Vector Machine (SVM).

25

Face recognition from video sequences has recently become popular. Recent work has shown that good results can be obtained using spatiotemporal information for video-based face analysis. Conventional methods include the use of adaptive Hidden Markov Models (HMM) to perform video-based face recognition, and the use of Autoregressive Moving Average (ARMA) for video-based face recognition. However, research on gender recognition from video sequences is still in its infancy.

30

In facial gender recognition, face alignment, which transforms a detected face to harmonize the locations of facial features, can advantageously maximize recognition accuracy. Current existing gender recognition methods assume the faces are aligned manually or automatically. Automatic face alignment, e.g. using active shape model (ASM), involves searching a detected face image for basic facial features such as the eyes, nose, mouth, and chin. The face image is then transformed so that the detected facial features line up in all images. This leads to better gender classification performance. However, face alignment is a time-consuming process and a challenging problem especially for low resolution face images, where eye feature localization is relatively difficult.

Different regions of the face do not equally contribute to the identification of gender from a face image. The performance of facial gender recognition degrades significantly when some important regions, e.g. eye and eyebrow, are occluded. Most of the existing gender recognition studies have demonstrated that the eye, eyebrow, jaw and face outline are the most important parts for identifying gender. Unfortunately, the eye and eyebrow are often self-occluded, e.g. by eye-glasses or hair, and can significantly affect gender recognition. Developing a gender recognition method that is robust to self-occlusion is challenging because self-occlusion happens unpredictably and is hard to detect.

A need therefore exists to provide a method and system for real-time facial gender recognition that seeks to address at least one of the abovementioned problems.

SUMMARY

According to the first aspect of the present invention, there is provided a method of building a classifier for use in facial image gender recognition based on both LBP features and Gabor features, the method comprising extracting LBP features from respective images of a training set of facial images; extracting Gabor features from respective images of a training set of facial images; applying a weighted feature selection algorithm to the extracted LBP and Gabor features; and building the classifier based on selected features from the weighted feature selection algorithm.

The weighted feature selection algorithm may be applied to a concatenation of the extracted LBP and Gabor features.

The weighted feature selection algorithm may be applied separately to the extracted LBP and Gabor features respectively, and the building of the classifier comprises building an LBP classifier based on selected LBP features, and building a Gabor classifier based on selected Gabor features.

Extracting the Gabor features may comprise performing face component recognition on the respective images from the training set of facial images, and

extracting the Gabor features from two or more selected components of the respective images.

The method may further comprise performing an evaluation of the classifier on a sample set of facial images, and, if an accuracy of the classifier is below a first threshold value, applying the classifier to a new set of facial images, adding selected images to a training set of facial images, the selected images being incorrectly classified ones of the new set of facial images by the classifier, and re-training the classifier based on the training set of facial images including the selected images.

The method may further comprise processing respective images of a training set of facial images to occlude selected areas of the image, and training the classifier based on the processed training set of facial images.

According to a second aspect of the present invention, there is provided a method of facial image gender recognition based on both LBP features and Gabor features, wherein the LBP features and Gabor features may be determined from training of a classifier in accordance with the first aspect of the invention.

The method may comprise applying a first classifier based on a training set of facial images, and, if an absolute score of the classifier is below a first threshold value, applying a second classifier based on a modified training set of facial images, the modified training set of facial images comprising processed respective images of the training set of facial images to occlude selected areas of the image.

According to a third aspect of the present invention, there is provided a system for building a classifier for use in facial image gender recognition, the system configured such that the classifier is trained based on both LBP features and Gabor features, and the system comprising: means for extracting LBP features from respective images of a training set of facial images; means for extracting Gabor features from respective images of a training set of facial images; and means for applying a weighted feature selection algorithm to the extracted LBP and Gabor features and;

means for building the classifier based on selected features from the weighted feature selection algorithm.

5 The means for applying a weighted feature selection algorithm may be configured to apply the weighted feature selection algorithm to a concatenation of the extracted LBP and Gabor features.

10 The means for applying a weighted feature selection algorithm may be configured to apply the weighted feature selection algorithm separately to the extracted LBP and Gabor features respectively, and the means for building of the classifier may be configured to build an LBP classifier based on selected LBP features, and to build a Gabor classifier based on selected Gabor features.

15 The means for extracting the Gabor features may be configured to perform face component recognition on the respective images from the training set of facial images, and may extract the Gabor features from two or more selected components of the respective images.

20 The system may further comprise a means for performing an evaluation of the classifier on a sample set of facial images, and the system may be configured to, if an accuracy of the classifier is below a first threshold value, apply the classifier to a new set of facial images, to add selected images to a training set of facial images, the selected images being incorrectly classified ones of the new set of facial images by the classifier, and to re-train the classifier based on the training set of facial
25 images including the selected images.

The system may comprise means for processing respective images of a training set of facial images to occlude selected areas of the image, and for training the classifier based on the processed training set of facial images.

30

According to a fourth aspect of the present invention, there is provided a system for facial image gender recognition, the system configured such that the gender recognition is based on both LBP features and Gabor features, wherein the LBP features and Gabor features are provided by the system for building a classifier
35 in accordance with the second aspect of the invention.

The system may comprise means for applying a first classifier based on a training set of facial images, and means for applying a second classifier based on a modified training set of facial images if an absolute score of the classifier is below a first threshold value, the modified training set of facial images comprising processed
5 respective images of the training set of facial images to occlude selected areas of the image.

According to a fifth aspect of the present invention, there is provided a data storage medium having stored thereon computer program code means for instructing
10 a computer system to execute a method of building a classifier for use in facial image gender recognition, as described above.

According to a sixth aspect of the present invention, there is provided a data storage medium having stored thereon computer program code means for instructing
15 a computer system to execute a method of facial image gender recognition, as described above.

BRIEF DESCRIPTION OF THE DRAWINGS

20

Example embodiments of the invention will be better understood and readily apparent to one of ordinary skill in the art from the following written description, by way of example only, and in conjunction with the drawings, in which:

25 Figure 1 is a schematic illustrating a method and system implementing the fusion of Gabor and Local Binary Pattern (LBP) features for gender recognition, according to an embodiment of the present invention.

30 Figure 2 shows an example of a face image undergoing normalization, according to an embodiment of the present invention.

Figure 3 shows the differences in Gabor features with and without normalization, according to an embodiment of the present invention.

35 Figure 4 is a schematic illustrating an example implementation of the Gabor feature extraction module, the LBP feature extraction module, the feature selection

module and the classifier module to one image of a training data set, according to an embodiment of the present invention.

Figure 5 is a graph illustrating error rates versus the number of the boost
5 iterations, according to an embodiment of the present invention.

Figure 6 is a schematic illustrating a system for implementing real-time
gender recognition with unaligned face images, according an embodiment of the
present invention.

10

Figure 7 is an example of component images obtained from facial component
detection, according to an embodiment of the present invention.

Figure 8 is a graph illustrating error rates versus the number of the boost
15 iterations, according to an embodiment of the present invention.

Figure 9 is a flowchart illustrating a two-stage gender classification process to
reduce errors caused by a self-occlusion (e.g. eye-glasses or hair), according to an
embodiment of the present invention.

20

Figure 10 shows some training samples and their original images, according
to an embodiment of the present invention.

Figure 11 is a graph illustrating error rates versus the number of the boost
25 iterations, according to an embodiment of the present invention.

Figure 12 is a schematic of a computer system for implementing the
system and method for real-time facial gender recognition in example
embodiments.

30

Figure 13 is a flowchart illustrating a classifier learning method according to
an embodiment of the present invention.

Figure 14 is a flowchart illustrating an alternative classifier learning method
35 according to another embodiment of the present invention.

DETAILED DESCRIPTION

Embodiments of the present invention provided a method and system for gender recognition based on the fusion of two local feature representations in facial image analysis. The local feature representations are Gabor and Local Binary Pattern (LBP) and they are combined at the feature level to represent a face image. The process that the Gabor filters are convolved with the signal is closely related to processes in the primary visual cortex. Gabor and LBP are found to be complimentary in the sense that LBP captures small appearance details while Gabor features encode facial shape over a broader range of scales.

The fusion of Gabor and LBP features is, in one embodiment, incorporated with an iterative weighted feature selection scheme, such as an Adaptive Boosting (Adaboost) learning approach to advantageously provide more robust recognition and relatively higher positive recognition rates. The Adaboost learning approach selects weak learners at each iteration from fused Gabor and LBP features based on the lowest training error rate.

In order to provide a real-time gender recognition system that implements the fusion of Gabor and LBP features for gender recognition, an autoregressive moving average (ARMA) is adopted to describe the face dynamics and combines the decisions from an image sequence.

Some portions of the description which follows are explicitly or implicitly presented in terms of algorithms and functional or symbolic representations of operations on data within a computer memory. These algorithmic descriptions and functional or symbolic representations are the means used by those skilled in the data processing arts to convey most effectively the substance of their work to others skilled in the art. An algorithm is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. The steps are those requiring physical manipulations of physical quantities, such as electrical, magnetic or optical signals capable of being stored, transferred, combined, compared, and otherwise manipulated.

Unless specifically stated otherwise, and as apparent from the following, it will be appreciated that throughout the present specification, discussions utilizing terms such as "scanning", "calculating", "determining", "replacing", "generating",

"initializing", "outputting", or the like, refer to the action and processes of a computer system, or similar electronic device, that manipulates and transforms data represented as physical quantities within the computer system into other data similarly represented as physical quantities within the computer system or other
5 information storage, transmission or display devices.

The present specification also discloses apparatus for performing the operations of the methods. Such apparatus may be specially constructed for the required purposes, or may comprise a general purpose computer or other device
10 selectively activated or reconfigured by a computer program stored in the computer. The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general purpose machines may be used with programs in accordance with the teachings herein. Alternatively, the construction of more specialized apparatus to perform the required method steps
15 may be appropriate. The structure of a conventional general purpose computer will appear from the description below.

In addition, the present specification also implicitly discloses a computer program, in that it would be apparent to the person skilled in the art that the individual
20 steps of the method described herein may be put into effect by computer code. The computer program is not intended to be limited to any particular programming language and implementation thereof. It will be appreciated that a variety of programming languages and coding thereof may be used to implement the teachings of the disclosure contained herein. Moreover, the computer program is not intended
25 to be limited to any particular control flow. There are many other variants of the computer program, which can use different control flows without departing from the spirit or scope of the invention.

Furthermore, one or more of the steps of the computer program may be
30 performed in parallel rather than sequentially. Such a computer program may be stored on any computer readable medium. The computer readable medium may include storage devices such as magnetic or optical disks, memory chips, or other storage devices suitable for interfacing with a general purpose computer. The computer readable medium may also include a hard-wired medium such as
35 exemplified in the Internet system, or wireless medium such as exemplified in the GSM mobile telephone system. The computer program when loaded and executed

on such a general-purpose computer effectively results in an apparatus that implements the steps of the preferred method.

The invention may also be implemented as hardware modules. More particular, in the hardware sense, a module is a functional hardware unit designed for use with other components or modules. For example, a module may be implemented using discrete electronic components, or it can form a portion of an entire electronic circuit such as an Application Specific Integrated Circuit (ASIC). Numerous other possibilities exist. Those skilled in the art will appreciate that the system can also be implemented as a combination of hardware and software modules.

Figure 1 is a schematic, designated generally as reference numeral 100, illustrating a method and system implementing the fusion of Gabor and LBP features for gender recognition, according to an embodiment of the present invention. The system comprises 2 feature extraction modules (a Gabor feature extraction module 102 and a LBP feature extraction module 104), a feature selection module 106 and a classification module 108. After a face (from a training data database 101) is detected and its features extracted by the two feature extraction modules 102/104, a sub-set of the extracted features are selected by the feature selection module 106. The selected features are then used to build the classifier module 108.

In the Gabor feature extraction module 102, Gabor features are extracted from the training data database 101. Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity. As a result, Gabor transformed face images produce salient local and discriminating features that are suitable for face recognition. Gabor wavelet representation of face images have been shown to be robust against variations in illumination and facial expression.

A Gabor wavelet $\varphi_{u,v}(z)$ is defined

$$\varphi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}} [e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}}] \quad (1)$$

where z represents a two-dimensional input point. The parameters u and v define the orientation and scale of the Gabor kernel. $\|\cdot\|$ indicates the norm operator, and σ refers to the standard deviation of the Gaussian window in the kernel.

5 The wave vector $k_{u,v}$ is defined as

$$k_{u,v} = k_v e^{i\phi_u} \quad - \quad (2)$$

where $k_v = \frac{k_{\max}}{f^v}$ and $\phi_u = \frac{\pi u}{8}$ if eight different orientations are chosen. $k_{\max}$ is the maximum frequency, and f' is the spatial frequency between kernels in the frequency domain.

10

In an example embodiment, five different scales and eight orientations of Gabor wavelets are used, e.g. $v \in \{0, \dots, 4\}$ and $u \in \{0, \dots, 7\}$. Gabor wavelets are chosen with the parameters $\sigma = 2\pi$, $k_{\max} = \pi/2$, and $f = \sqrt{2}$.

15

A Gabor wavelet representation of an image is the convolution of the image with the filter bank. The convolution of image I and a Gabor kernel $\phi_{u,v}(z)$ is defined as follows:

$$O_{u,v}(z) = I(z) * \phi_{u,v}(z) \quad - \quad (3)$$

and is called a Gabor feature.

20

As the response $O_{u,v}(z)$ to each Gabor kernel is a complex function with a real part $R\{O_{u,v}(z)\}$ and an imaginary part $I\{O_{u,v}(z)\}$, we use its magnitude

$$\|O_{u,v}(z)\| = \sqrt{R(O_{u,v}(z))^2 + I(O_{u,v}(z))^2} \quad - \quad (4)$$

to represent the Gabor features.

25

The complete set of Gabor wavelet representations of the image $I(z)$ is $G(I) = \{O_{u,v}(z) : u \in \{0, \dots, 7\}, v \in \{0, \dots, 4\}, z = (x, y)\}$. The resulting features for each orientation, scale and position are concatenated pixel by pixel to form a facial feature vector of the image.

30

In the LBP feature extraction module 104, LBP features are extracted from the training data database 101. Conventional LBP operators label the pixels of an

image by thresholding the 3×3 neighborhood of each pixel i_n , $n = 0, 1, \dots, 7$ with the center value i_c and considering the result as a binary number

$$LBP = \sum_{n=0}^7 S(i_n - i_c) 2^n \quad (5)$$

which characterizes the spatial structure of the local image texture. $S(x)$ is 1 if $x \geq 0$ and 0 otherwise.

A local binary pattern is considered uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular.

Local Binary Pattern Histogram Fourier features (LBP-HF) is a rotation invariant image descriptor based on uniform Local Binary Patterns. Unlike prior art local rotation invariant features, the LBP-HF descriptor is formed by first computing a non-invariant LBP histogram over the whole region and then constructing rotationally invariant features from the histogram. This means that rotation invariance is attained globally, and the features are thus invariant to rotations of the whole input signal but they still retain information about relative distribution of different orientations of uniform local binary patterns.

In an example embodiment, LBP-HF is used to extract features. Denoting a specific uniform LBP pattern by $U_P(n, r)$, the rotation invariant LBP is defined as:

$$LBP^{u2-HF}(n_1, n_2, u) = H(n_1, u) \bar{H}(n_2, u) \quad (6)$$

where $H(n, \cdot)$ is the DFT of n th row of the histogram $h_1(U_P(n, r))$, i.e.

$$H(n, u) = \sum_{r=0}^{P-1} h_1(U_P(n, r)) e^{-2\pi i u r / P} \quad (7)$$

and $\bar{H}(n_2, u)$ denotes the complex conjugate of $H(n_2, u)$.

25

A face image is divided into small regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram. The histogram provides an effective description of the face on two different levels of localisation: the labels for the histogram contain information about the patterns at the pixel level while the labels summed over a small region provide information at the regional level.

30

In the feature selection module 106 and classifier module 108, Gabor and LBP features are fused at the feature level. This can be done by concatenating the Gabor and LBP features to represent the face image. The dimension of the combined feature is typically very high and dimension reduction, e.g. principle component analysis (PCA), may be applied to the fused feature before classification. However, in example embodiments of the present invention, dimension reduction can be avoided because Adaboost is adopted for feature selection and classification. Furthermore, the computation of a detected face is fast because only the selected features are computed.

Before concatenating the Gabor and LBP features, it is preferable to normalize the features to zero mean and unit variance. A vector can be constructed by concatenating its rows (or columns). Normalization advantageously provides additional robustness against variation in illumination of a subject.

Figure 2 shows an example of a face image undergoing normalization. A face is detected and an example is shown in image 202. The image 202 preferably undergoes masking such that unwanted background and noise is removed. Here, a circle contour is applied to the detected face and an example of a masked face is shown in image 204. Accordingly, only the points of interest within the mask, i.e. the subject's face, is used for gender recognition. The masked image 204 undergoes feature extraction (e.g. Gabor feature extraction) and an example of the magnitude of the Gabor features obtained is shown in image 206. The features are normalized and an example output of the normalized Gabor features is shown in image 208.

Figure 3 shows the differences in Gabor features with and without normalization. A subject's face is detected and an example is shown in image 302. The same subject is detected again but now with less ambient lighting and an example is shown in image 304. The Gabor features of image 302 and 304 are normalized and example outputs of the normalized Gabor features are shown in images 306 and 308 respectively. With normalization, the difference in Gabor features is relatively small even with a significant difference in ambient illumination. Conversely, images 312 and 314 represent non-normalized Gabor features of images 302 and 304 respectively. Without normalization, the difference in Gabor features is relatively larger when there is a significant difference in ambient illumination.

Each detected face potentially requires the computation of 77440 ($44 \times 44 \times 40$) Gabor features and 7139 ($11 \times 11 \times 59$) LBP features before gender classification can be carried out. To advantageously reduce computation, the AdaBoost classification framework picks the T best features out of 77440+7139. The T best features are preferably in the range from $T = 100 - 1000$, and more preferably about $T = 150$. At each iteration, one (i.e. the best of the remaining features) is selected, for a total of T iterations, as shown in the boosting algorithm below.

The inventors have recognized that Gabor and Local Binary Pattern (LBP) are two local feature representations which separately have good performance in facial image analysis; and furthermore have recognized that Gabor and LBP features are complementary in the sense that LBP captures small appearance details while Gabor features encode facial shape over a wider range of scales.

By selectively computing only the best features for gender classification, the classification time required for each detected face is reduced from 70ms to 2ms (in an embodiment where $T = 150$). This speedup facilitates multiple face detection / classification within each frame (e.g. 20 persons simultaneously within a frame). It can also increase program stability and classification performance due to the ability to classify more frames and aggregate these classifications to obtain a more accurate final classification. The T best Gabor and LBP features are selected based on the training data, and are then "fixed" for a particular test sample.

An example boosting algorithm in terms of fused Gabor and LBP features is as follows:

Input: N training samples in two training sets: LBP features of the training samples $(x_1^L, y_1), (x_2^L, y_2), \dots, (x_N^L, y_N)$ and Gabor features of the training samples $(x_1^G, y_1), (x_2^G, y_2), \dots, (x_N^G, y_N)$; where $x_i^G \in X_G$ is the Gabor feature representation of the face images and $x_i^L \in X_L$ are the LBP representation of the face image; $y_i \in Y = \{-1, +1\}$ labels the male/female class.

Initialization $w(i) = \frac{1}{N}$ for all $i = 1, 2, \dots, N$.

Feature fusion by concatenating the Gabor and LBP features:

$$(x_i, y_i) = [x_i^G; x_i^L]; i = 1, \dots, N$$

*For $t = 1$ to T

Find the classifier $h_t : X \rightarrow \{-1, +1\}$ that minimizes the error with respect to the distribution w_t

$$h_t = \arg \min_{h \in H} \varepsilon_j$$

$$\text{where } \varepsilon_j = \sum_{i=1}^N w_t(i) [y_i \neq h_j(x_i)]$$

Prerequisite the weighted error rate of classifier $h_t < 0.5$, otherwise stop or start over.

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right)$$

where ε_t is the lowest weighted error rate of classifier h_t .

Update

$$w_{t+1}(i) = \frac{w_t(i) e^{-\alpha_t y_i h_t(x_i)}}{Z_t}$$

where Z_t is a normalization factor (chosen so that W_{t+1} will be a probability distribution, i.e. sum one over all x).

Loop back to * until T is reached.

Output the final classifier:

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad - (8)$$

Figure 4 is a schematic, designated generally as reference numeral 400, illustrating an example implementation of the Gabor feature extraction module 402, the LBP feature extraction module 404, the feature selection module 406 and the classifier module 410 to one image 411 of a training data set 412, according to the embodiment of the present invention as described above.

To advantageously improve robustness to take into account movement, pose and illumination, face tracking is preferably incorporated into the gender recognition system. Kernel-based mean-shift is adopted. Accordingly, in an example embodiment, face tracking is combined with face detection to minimize the loss of a tracking target. After face detection and gender recognition are performed, face tracking is performed for a pre-determined time, t (e.g.: 5 frames). After the pre-determined time, t , has lapsed, face detection is performed again to minimize the loss of a tracking target.

In another embodiment of the present invention, a classifier is provided that operates on video instead of still images. This advantageously allows video processing technologies to be applied to determine gender from a sequence of face images to achieve greater robustness.

Autoregressive moving average (ARMA) can be used to recognize gender from video sequences. ARMA is adopted to model the dynamical information of face videos of male and female classes and gender recognition is completed by estimating the distance of the probe and gallery model parameters. The ARMA model is built using the scores from the Adaboost algorithm. Raw face image data is used to estimate model parameters and the scores obtained by applying a strong classifier to a face image sequence are used to describe the face dynamic.

In the ARMA framework, a moving face is represented by a linear dynamical system

$$z(t+1) = Az(t) + v(t), \quad x(t) \in R^n, \quad v(t) \sim N(0, R) \quad - \quad (9)$$

$$s(t) = Cz(t) + w(t), \quad s(t) \in R^m, \quad w(t) \sim N(0, Q) \quad - \quad (10)$$

where $s(t)$ is the noise version score at time instant t , $z(t)$ is a state {male, female} vector that characterizes the gender dynamic, A and C are matrices representing the

state and output transition, $v(t)$ and $w(t)$ are IID sequences driven from some unknown distribution.

- ARMA models are built for male and female classes respectively using collected training video sequences. The parameters A , C , Q and R are estimated for describing an ARMA model. The estimation of these parameters is in a closed-form, as shown below:

Parameters estimation of the ARMA model

- 10 Let $S^T = [s(1), \dots, s(t)] \in R^{m \times \tau}$ with $\tau > n$.

From (9), we arrive at $S^T = CX^T + W^T C \in R^{m \times n}$

If singular value decomposition (SVD) of S is

$$S^T = U \Sigma V^T \text{ then}$$

$$\hat{C}(T) = U$$

- 15 $\hat{X}(T) = \Sigma V^T$

$$\hat{A}(T) = \Sigma V^T D_1 V (V^T D_2 V)^{-1} \Sigma^{-1}$$

$$\text{where } D_1 = \begin{pmatrix} 0 & 0 \\ I_{T-1} & 0 \end{pmatrix} \text{ and } D_2 = \begin{pmatrix} I_{T-1} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\hat{Q}(T) = \frac{1}{T} \sum_{i=1}^T \hat{v}(i) \hat{v}^T(i)$$

$$\text{where } \hat{v}(t) = \hat{x}(t+1) - \hat{A}(T) \hat{x}(t)$$

20

The s in equations (8) and (9) is represented using the scores for successive frames of training video sequence:

$$\sum_{t=1}^T \alpha_t h_t(x) \quad - \quad (11)$$

- 25 where x is the Gabor + LBP representation of the current image, and h_t is the classified in t -th iteration; and $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$ where ε_t is the lowest weighted error rate of classifier h_t .

- 30 Figure 5 is a graph illustrating error rates versus the number of the boost iterations, according to an embodiment of the present invention. It can be seen that

embodiments of the present invention, incorporating fused Gabor and LBP features 502, exhibit higher recognition rates (about 91% accuracy) compared with single feature methods (about 81% for LBP alone 504 and 87% for Gabor alone 506).

5 Embodiments of the present invention were evaluated using a hybrid face database comprising of three databases: color FERRET, PEAL and a private database. The color FERRET database contains images of 994 people (591 male, 403 female) and only frontal images labeled "fa" and "fb" in the database with labeled eye coordinates were used, for a total of 2409 faces images (1495 male, 914
10 female). The CAS-PEAL face database contains images of 1040 individuals (595 males and 445 females) with varying pose, expression, accessories, and lighting (PEAL). Only the frontal images with neutral expression were used. The private database contains images of 51 people (38 male, 13 female). Half of the images in each database is used for Adaboost training and the remainder for testing.

15 Two-fold cross validation was executed and average accuracy was computed. Faces were detected using a OpenCV 1.0 face detector and faces were first normalized so that the centres of the two eyes were kept at fixed position.

20 The size of the normalized face image was 88×88. Gabor features at five scales and eight orientations were extracted at each pixel. Hence, the dimension of the Gabor representation was 309760 (88×88×40). In order to reduce the dimension of the feature vector, 16 times down sampling of the Gabor features was adopted, so the reduced dimension was 19360. Like the Gabor features, the LBP descriptor is
25 usually high dimensional. The 88×88 image was divided into 8×8 blocks and the histograms of these 121 blocks were concatenated into a vector to represent the image. Hence the dimension is 7139 (121 patches with 59 entries/patch).

Figure 6 is a schematic, designated generally as reference number 600,
30 illustrating a system for implementing real-time gender recognition with unaligned face images, according another embodiment of the present invention. The system comprises a face detection module 602, a face component detection module 604, a LBP feature extraction module 606, a Gabor feature extraction module 608, a holistic classifier module 610, a component classifier module 612, and a feature fusion
35 module 614. The feature fusion module 614 further comprises a probability model processing module 614a and a decision module 614b.

In the face detection module 602, face detection is performed. Face detection is the first step for almost all face information processing systems. It will be appreciated by a person skilled in the art that any appropriate face detector may be used. Embodiments of the present invention use the face detector available in
5 OpenCV. Detected faces are resized to a fixed size using the centre of the bounding box. The frontal view is considered and the eye and mouth can be extracted using a horizontal and vertical projection method. The location of the eyebrow and nose can be predicted from the eye and mouth.

10 In the face component detection module 604, face component detection is performed. Although holistic feature-based methods, where a face is represented as a whole and statistical techniques are used to extract feature from faces, has been successful, the importance of individual facial components to the gender recognition is equally important. For instance, eyebrows are found to be useful in aiding
15 automatic classification of the gender of a face image. An example of the component images obtained from face facial component detection, according to an embodiment of the invention, is shown in Figure 7. As component and holistic-based face representations have complementary strengths and weaknesses, a well designed visual perception system advantageously employs both types of representation for
20 gender recognition.

Feature extraction of the face and its components can be performed using the LBP feature extraction module 606 and Gabor feature extraction module 608, which has been described in detail above in relation to the previous embodiment. The LBP feature extraction module 604 extracts the holistic features of the face while
25 the Gabor feature extraction module 602 extracts the component features of the face.

Once the training samples of the holistic face and face components are obtained, the classifier for each of them can be trained, respectively. Unaligned faces can intentionally be included in the training set. AdaBoost learning can be adopted to
30 select features. It is to be understood that the AdaBoost algorithm as described above can be similarly applied here.

In the feature fusion module 614, the individual classified holistic LBP and Gabor component features are fused by a probabilistic model. A face is defined as one of the components. Assume a probabilistic face model, where the holistic face and each component have some uncertainty. With the inclusion of uncertainties, the face model is flexible to describe a variety of possible male and female faces. Assuming Gaussian distributions in the face model, a set of confidences with means m_i , and deviation D_i , $i = 1, 2, \dots, N$, where N is the number of components. The face model quantifies how much contribution each of the components contributes to the gender classification. AdaBoost is used for selecting features to train a classifier. The face models for male and female are trained from known male and female face examples, respectively. Hence, the contribution of each component can be measured by the output of the AdaBoost.

After the component gender classifiers are trained, the confidence of the component in the input image is

$$A_i = \sum_{t=1}^T \alpha_t h_t(x), i = 1, 2, \dots, N \quad - \quad (12)$$

where N is the number of components;

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right) \quad - \quad (13)$$

where ε_t is the lowest weighted error rate of t -th weak classifier h_t in AdaBoost.

Confidence A_i is assumed normalized across all the components. As the shape of A_i is smooth and Gaussian-like, a Gaussian shape can be used to approximate it.

With the face model $\{m_i, D_i\}$, $i=1,2,\dots, N$, the overall gender likelihood function is:

$$L = \prod_{i=1}^N \left[\frac{1}{[2\pi D_i]^2} \exp \left(-\frac{1}{2} (A_i - m_i)^T D_i^{-1} (A_i - m_i) \right) \right] \quad - \quad (14)$$

The goal is to find the gender with maximal L . This can be done by comparing the outputs of the male and female likelihoods.

Figure 8 is a graph illustrating error rates versus the number of the boost iterations, according to an embodiment of the present invention. The accuracy of gender recognition method based on manually aligned face images 802/804 are slightly better than the one based on unaligned faces 806. The accuracy is best when the faces are normalized using the two outer eye corners 804, while the system without alignment can obtain about 91% accuracy.

Embodiments of the present invention were evaluated using a hybrid face database comprising of three databases: color FERRET, PEAL and a private database. The color FERRET database contains images of 994 people (591 male, 403 female) and only frontal images labeled "fa" and "fb" in the database with labeled eye coordinates were used, for a total of 2409 faces images (1495 male, 914 female). The CAS-PEAL face database contains images of 1040 individuals (595 males and 445 females) with varying pose, expression, accessories, and lighting (PEAL). Only the frontal images with neutral expression were used. The private database contains images of 51 people (38 male, 13 female). Half of the images in each database is used for Adaboost training and the remainder for testing.

Performance was evaluated using three data sets which are obtained by: (1) faces detected using OpenCV 1.0 face detector and the face rectangles are resized to a fixed size; (2) faces manually normalized using outer corners of the two eyes; and (3) faces normalized manually using the centre of the eyes. The size of the normalized face image is 88×88. Each 88×88 image is divided into 8×8 blocks and the histograms of these 121 blocks are concatenated into a vector to represent the image, for a total feature dimension of 7139. (121 patches with 59 entries/patch).

For evaluation of embodiments of the present invention, a webcam was used to capture face images. The frame rate was about 6 frames/second. The face detector of the OpenCV library and kernel-based mean shift are adopted to detect and track faces.

Figure 9 is a flowchart, designated generally as reference numeral 900, illustrating a two-stage gender classification process to reduce errors caused by a self-occlusion (e.g. eye-glasses or hair), according to an embodiment of the present invention.

At step 902, training data that comprises face images with and without self-occlusion is obtained. At step 904, Adaboost is performed on the training data and a sequence of features is fed to a first classifier module. It is to be understood that LBP and Gabor feature extraction and fusion is performed, in accordance with the
5 embodiments as described above, prior to being fed to the classifier. At step 906, face detection is performed on a set of samples. At step 908, if the absolute gender score is less than a pre-determined threshold (δ), a second classification stage is invoked. An absolute gender score less than the pre-determined threshold (δ) may imply that self-occlusion is detected. Otherwise, a decision on the gender can be
10 made at step 910.

Accordingly, once the score of a face is found to be near the threshold (δ) of a first strong classifier, it can be passed to a second strong classifier which is obtained by training AdaBoost on artificially occluded samples (e.g. the eye and eyebrow are
15 occluded). The principle behind this "two-stage" approach is that the feature selection of the AdaBoost can be controlled by carefully choosing the training samples. An independent processor module can be used for artificially occluding selected areas of the sample and for training the classifier based on the processed training set of facial
20 images.

In the second classification stage, training data that comprises face images with self-occlusion is obtained at step 914 and is subsequently subjected to Adaboost at step 916. A sequence of features is fed to the second strong classifier to facilitate gender recognition at step 918. Again, it is to be understood that LBP and Gabor
25 feature extraction and fusion is performed, in accordance with the embodiments as described above, prior to being fed to the classifier. A second strong classifier trained by self-occluded faces can be applied for this purpose. The eye or eyebrow is not selected during the boosting rounds of this second AdaBoost classifier because they are assigned a constant intensity in both male and female training samples and
30 consequently they have no discriminant information for the gender. Some training samples for the second stage (bottom row) and their original images (top row) are shown in Figure 10.

Visual information from different face areas are believed not to contribute
35 equally to a human's ability to recognize gender. However, most of the existing studies on gender recognition treat all areas equally. Accordingly, in this

embodiment, regions that carry the most information about the gender of a subject are identified and the loss of information caused by the occlusion of these informative regions are compensated.

At each round of boosting, a feature is selected. A first strong classifier is formed by a linear combination (see equation (7)) where α_i are coefficients found in the boosting process (see equation (6)). The order of the boosting round represents the importance of the corresponding feature. In other words, the more discriminative a feature is, the earlier the feature is selected.

Nonetheless, to some extent, feature selection can be controlled. For instance, a moustache can be an important feature which can set apart male from female. However, there is a possibility that the moustache is not selected at the earlier stage because most of the male face images in a training database belong to Asian men, who typically have no moustache. Hence, the system could make an error classification for a people with moustache. One way to overcome the shortcoming is that to add more face images with a moustache so that the moustache is selected as one features at the earlier stage.

Gender recognition is sensitive to self-occlusion as the eye and eyebrow (two occluded parts) are the first top two features selected by the AdaBoost which contribute much to the final decision of the strong classifier. It is to be understood that the AdaBoost algorithms as described above in relation to the previous embodiments can be similarly applied here. $s = \sum_{t=1}^T \alpha_t h_t(x)$ in equation (8) is defined

as the score of a face. A negative score implies a male is detected and a positive score implies that a female is detected. The scores of a self-occluded face are usually around zero. In other words, the gender of self-occluded faces can be wrongly determined even if a person keeps still in front of the camera because the sign of the score can change occasionally.

Figure 11 is a graph illustrating the testing error rates versus the number of the boost iterations between one-stage and "two-stage" approach of embodiments of the invention. It can be seen that the two-stage approach 1102 yields relatively better classification accuracy than a one-stage approach 1104.

Embodiments of the present invention were evaluated using a hybrid face database comprising of four databases: color FERRET, PEAL, AR and a private database. The color FERRET database contains images of 994 people (591 male, 403 female) and only frontal images labeled "fa" and "fb" in the database with
5 labeled eye coordinates were used, for a total of 2409 faces images (1495 male, 914 female). The CAS-PEAL face database contains images of 1040 individuals (595 males and 445 females) with varying pose, expression, accessories, and lighting (PEAL). Only the frontal images with neutral expression were used. The AR database contains many occluded images and only face images with Neutral
10 expression are used. The private database contains 5000 images of 1210 people (710 male, 500 female). About ¼ of these 5000 images are self-occluded. Half of the images in each database are randomly selected for Adaboost training and the remainder for testing. The faces are detected using Viola's face detector and are then normalized to 44×44 images.

15 The method and system of the example embodiment can be implemented on a computer system 1200, schematically shown in Figure 12. It may be implemented as software, such as a computer program being executed within the computer system 1200, and instructing the computer system 1200 to conduct the
20 method of the example embodiment.

The computer system 1200 comprises a computer module 1202, input modules such as a keyboard 1204 and mouse 1206 and a plurality of output devices such as a display 1208, and printer 1210.

25 The computer module 1202 is connected to a computer network 1212 via a suitable transceiver device 1214, to enable access to e.g. the Internet or other network systems such as Local Area Network (LAN) or Wide Area Network (WAN).

30 The computer module 1202 in the example includes a processor 1218, a Random Access Memory (RAM) 1220 and a Read Only Memory (ROM) 1222. The computer module 1202 also includes a number of Input/Output (I/O) interfaces, for example I/O interface 1224 to the display 1208, and I/O interface
35 1226 to the keyboard 1204.

The components of the computer module 1202 typically communicate via an interconnected bus 1228 and in a manner known to the person skilled in the relevant art.

5 The application program is typically supplied to the user of the computer system 1200 encoded on a data storage medium such as a CD-ROM or flash memory carrier and read utilising a corresponding data storage medium drive of a data storage device 1230. The application program is read and controlled in its execution by the processor 1218. Intermediate storage of program data maybe
10 accomplished using RAM 1220.

It will be appreciated by a person skilled in the art that numerous variations and/or modifications may be made to the present invention as shown in the embodiments without departing from a spirit or scope of the invention as broadly
15 described. The embodiments are, therefore, to be considered in all respects to be illustrative and not restrictive.

For example, for the detection of faces, besides using OpenCV, two other current popular face detectors are LBP and Harr. The advantage of LBP is that it is
20 relatively fast, but true faces may be lost. The advantage of Harr is that all true faces are kept, but false alarms may happen. Accordingly, fast Harr-like feature based detection can be performed during a first stage and support vector machine (SVM) face/non-face detection is performed during a second stage. False faces detected in the first stage are advantageously rejected in the second stage.

25 Figure 13 is a flowchart, designated generally as reference numeral 1300, illustrating a classifier learning method that can be used in the embodiments described above. At step 1308, seed images 1306 are used for Adaboost training of the strong classifier 1310. The strong classifier 1310 performs gender recognition on
30 an evaluation set 1312 and a test set 1314. If the performances of the strong classifier on the evaluation set data 1312 and test set 1314 are above two pre-determined accuracy levels (T_e and T_t) respectively, the learning process is considered satisfactory and complete. However, if for example the performance is below the pre-determined accuracy level T_t , new data 1302, i.e. a new set of images,
35 is provided, and all new images are added at step 1304 to the training. On the other hand, if the performance is below the pre-determined accuracy level T_e , an alternative training set (e.g. subset of the original training set or a full different set) is

used, or the number of the boost stages (e.g. continue training more stages) is adjusted, as shown at step 1316. Step 1308 is performed again.

The evaluation set and the test set are two independent sets. The purpose of the evaluation set is so that an alternative training set (e.g. subset of the original training set or a full different set) can be used, or the number of the boost stages (e.g. continue training more stages) is adjusted, until the evaluation accuracy is above T_e . The test set (which may comprise unseen data) is used for predicting the performance of embodiments of the present invention.

Gender recognition performance is dependent on the training data. For example, low quality training data (e.g. blurred and lower resolution faces) affects the performance of a classifier. Further, theoretically, a classifier need not use all the data to fully learn the data. The challenging samples, i.e. those that are potentially hard to classify, are generally more informative.

Figure 14 is a flowchart, designated generally as reference numeral 1400, illustrating an alternative classifier learning method that can be used in the embodiments described above. At step 1412, seed images 1410 are used for Adaboost training of the strong classifier 1414. The strong classifier 1414 performs gender recognition on an evaluation set 1416 and a test set 1418. If the performances of the strong classifier on the evaluation set data 1416 and test set 1418 are above the two pre-determined accuracy levels (T_e and T_t) respectively, the learning process is considered satisfactory and complete.

However, if the performance is below the pre-determined accuracy level T_t , the classifier is applied at step 1404 to a new set of facial images 1402. At step 1406, selected images are added to a training set of facial images, the selected images being incorrectly classified ones of the new set of facial images by the classifier. At step 1412, the classifier is re-trained based on the training set of facial images including the selected images.

If the performance is below the pre-determined accuracy level T_e , an alternative training set (e.g. subset of the original training set or a full different set) is used, or the number of the boost stages (e.g. continue training more stages) is adjusted, as shown at step 1420. At step 1412, the classifier is re-trained based on

the alternative training set of facial images or the previous training set with the boost stages adjusted.

5

CLAIMS

1. A method of building a classifier for use in facial image gender recognition based on both LBP features and Gabor features, the method comprising:
5 extracting LBP features from respective images of a training set of facial images;
extracting Gabor features from respective images of a training set of facial images;
applying a weighted feature selection algorithm to the extracted LBP and
10 Gabor features; and
building the classifier based on selected features from the weighted feature selection algorithm.
2. The method as claimed in claim 1, wherein the weighted feature
15 selection algorithm is applied to a concatenation of the extracted LBP and Gabor features.
3. The method as claimed in claim 1, wherein the weighted feature
20 selection algorithm is applied separately to the extracted LBP and Gabor features respectively, and the building of the classifier comprises building an LBP classifier based on selected LBP features, and building a Gabor classifier based on selected Gabor features.
4. The method as claimed in claim 3, wherein extracting the Gabor
25 features comprises performing face component recognition on the respective images from the training set of facial images, and extracting the Gabor features from two or more selected components of the respective images.

5. The method as claimed in any one of the preceding claims, further comprising performing an evaluation of the classifier on a sample set of facial images, and, if an accuracy of the classifier is below a first threshold value, applying
5 the classifier to a new set of facial images, adding selected images to a training set of facial images, the selected images being incorrectly classified ones of the new set of facial images by the classifier, and re-training the classifier based on the training set of facial images including the selected images.

10 6. The method as claimed in any one of the preceding claims, comprising processing respective images of a training set of facial images to occlude selected areas of the respective images, and training the classifier based on the processed training set of facial images.

15 7. A method of facial image gender recognition based on both LBP features and Gabor features, wherein the LBP features and Gabor features are determined from training of a classifier according to any one of claims 1 to 6.

20 8. The method as claimed in claim 7, comprising applying a first classifier based on a training set of facial images, and, if an absolute score of the classifier is below a first threshold value, applying a second classifier based on a modified training set of facial images, the modified training set of facial images comprising processed respective images of the training set of facial images to occlude selected areas of the respective images.

25 9. A system for building a classifier for use in facial image gender recognition, the system configured such that the classifier is trained based on both LBP features and Gabor features, and the system comprising:

30 means for extracting LBP features from respective images of a training set of facial images;

means for extracting Gabor features from respective images of a training set of facial images; and

means for applying a weighted feature selection algorithm to the extracted LBP and Gabor features and;

5 means for building the classifier based on selected features from the weighted feature selection algorithm.

10 10. The system as claimed in claim 9, wherein the means for applying a weighted feature selection algorithm is configured to apply the weighted feature selection algorithm to a concatenation of the extracted LBP and Gabor features.

11. The system as claimed in claim 9, wherein the means for applying a weighted feature selection algorithm is configured to apply the weighted feature selection algorithm separately to the extracted LBP and Gabor features respectively, and the means for building of the classifier is configured to build an LBP classifier
15 based on selected LBP features, and to build a Gabor classifier based on selected Gabor features.

12. The system as claimed in claim 11, wherein means for extracting the Gabor features is configured to perform face component recognition on the
20 respective images from the training set of facial images, and to extract the Gabor features from two or more selected components of the respective images.

13. The system as claimed in any one of claims 9 to 12, further comprising a means for performing an evaluation of the classifier on a sample set of
25 facial images, and the system is configured to, if an accuracy of the classifier is below a first threshold value, apply the classifier to a new set of facial images, to add selected images to a training set of facial images, the selected images being incorrectly classified ones of the new set of facial images by the classifier, and to re-

train the classifier based on the training set of facial images including the selected images.

14. The system as claimed in any one of claims 9 to 13, comprising
5 means for processing respective images of a training set of facial images to occlude selected areas of the respective images, and for training the classifier based on the processed training set of facial images.

15. A system for facial image gender recognition, the system configured
10 such that the gender recognition is based on both LBP features and Gabor features, wherein the LBP features and Gabor features are provided by the system for building a classifier according to any one of claims 9 to 14.

16. The system as claimed in claim 15, comprising means for applying a
15 first classifier based on a training set of facial images, and means for applying a second classifier based on a modified training set of facial images if an absolute score of the classifier is below a first threshold value, the modified training set of facial images comprising processed respective images of the training set of facial images to occlude selected areas of the respective images.

20

17. A data storage medium having stored thereon computer program code means for instructing a computer system to execute a method of building a classifier for use in facial image gender recognition, as claimed in any one of claims 1 to 6.

25

18. A data storage medium having stored thereon computer program code means for instructing a computer system to execute a method of facial image gender recognition, as claimed in claim 7 or 8.