

# Automatic Age Estimation System for Face Images

Regular Paper

Chin-Teng Lin<sup>1</sup>, Dong-Lin Li<sup>1,\*</sup>, Jian-Hao Lai<sup>2</sup>, Ming-Feng Han<sup>1</sup> and Jyh-Yeong Chang<sup>1</sup><sup>1</sup> Department of Electrical Engineering, National Chiao-Tung University, Taiwan, R.O.C.<sup>2</sup> Department of Computer Science, National Chiao-Tung University, Taiwan, R.O.C.

\* Corresponding author E-mail: donglin.eric.li@gmail.com

Received 15 Jun 2012; Accepted 29 Aug 2012

DOI: 10.5772/52862

© 2012 Lin et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract** Humans are the most important tracking objects in surveillance systems. However, human tracking is not enough to provide the required information for personalized recognition. In this paper, we present a novel and reliable framework for automatic age estimation based on computer vision. It exploits global face features based on the combination of Gabor wavelets and orthogonal locality preserving projections. In addition, the proposed system can extract face aging features automatically in real-time. This means that the proposed system has more potential in applications compared to other semi-automatic systems. The results obtained from this novel approach could provide clearer insight for operators in the field of age estimation to develop real-world applications.

**Keywords** Gabor Wavelet, Face Image, Age estimation, SVM

## 1. Introduction

A human face image contains abundant information about personal characteristics, including identity, emotional expression, gender, age, etc. Generally, a

human image can be considered as a complex signal composed of many facial attributes such as skin colour and geometric facial features. These attributes play a crucial role in real applications of facial image analysis. In such applications, various attributes estimated from a captured face image can infer further system reactions. Age, in particular, is more significant among these attributes. For example, users may require an age-specific human computer interaction system that can estimate age for secure system access control or intelligence gathering. Automatic human age estimation using facial image analysis has numerous potential real-world applications.

An automatic face image age estimation system is composed of two parts: face detection and age estimation. The purpose of face detection is to localize the faces in an image. It is quite challenging to detect the faces in images, because the detected results are highly dependent on many conditions, such as environment, movement, lighting, orientation and facial expressions. These variant factors may lead to changes in colour, luminance, shadows and contours of images. For this reason, Viola and Jones proposed the famous face detector system in 2004 [1]. The Viola-Jones classifier employs AdaBoost at each node in the cascade to learn a high detection rate at

the cost of a low rejection rate of a multi-tree classifier at each node of the cascade. The algorithm incorporates several innovative features: 1) the Haar-like input features - a threshold is applied to sums and differences of rectangular image regions; and 2) the integral image technique enables rapid computation of the value of rectangular regions or such regions rotated 45 degrees. This data structure is used to accelerate computation of the Haar-like input features. In addition, the AdaBoost algorithm 3) uses statistical boosting to create binary (face - non-face) classification nodes characterized by high detection and weak rejection; and 4) organizes the weak classifier nodes of a rejection cascade. In other words, the first group of classifiers is selected that most effectively detects image regions containing an object while allowing for many mistaken detections; the next classifier group is the second-best at detection with weak rejection. In test mode, an object is detected only if it makes it through the entire cascade.

Although the automatic face detection of an image is a mature technique involving many real-world applications, estimating human age from face images is still a challenging problem. Because the aging process is represented differently not only among races, but also within races, the process is almost personal. Moreover, this process is also determined by external factors, such as health, lifestyle, location and weather conditions. Therefore, "how to find a robust representation featuring" remains an open problem.

Overall, there are three categories of feature extraction for human facial age estimation in the proposed literature. The first category is statistical-based approaches. Xin Geng et al. [2][3] proposed the AGing pattErn Subspace (AGES) method for automatic age estimation. The idea of AGES is to model the aging pattern, which is defined as a sequence of personal aging face images, by learning a representative sub-space from EM-like (expectation-maximization) iterative learning Principle Component Analysis (PCA). In other major studies [4][5], Guodong Guo et al. compared three typical dimensionality reduction and manifold embedding methods, such as PCA, Locally Linear Embedding (LLE) and Orthogonal Locality Preserving Projections (OLPP). According to the data distribution in OLPP sub-space, they proposed the Locally Adjusted Robust Regression (LARR) method for learning and prediction of human ages. The LARR applies Support Vector Regression (SVR) to obtain a coarse prediction and determine a local adjustment within a limited range of ages centred on the predicted result using the Support Vector Machine (SVM).

The second category comprises appearance-based approaches. Using appearance information is the most intuitional method in all facial image analysis works.

Young H. Kwon et al. [6] used visual aging features to construct an anthropometric model. The primary features are the eyes, nose, mouth and chin. The ratios of those features are computed to distinguish different age ranges. In secondary feature analysis, a wrinkle geography map is used to guide the detection and measurement of wrinkles. Jun-Da Xia et al. [7] proposed an age estimation method using the Active Appearance Model (AAM) to extract the regions of age features. Each face requires 28 feature points and is divided into ten wrinkle feature regions. Shuicheng Yan et al. [8] presented a patch-based appearance model named Patch-Kernel. This method is designed to characterize the Kullback-Leibler divergence between the models which are derived from the global Gaussian Mixture Model (GMM) using Maximum a Posteriori (MAP) for any two images. The discriminating power is further enhanced using a weak learning process, called "inter-modality similarity synchronization". Kernel regression is employed for estimating age.

The third category comprises frequency-based approaches. In image processing and pattern recognition, frequency domain analysis is the most popular method for extracting image features. Guodong Guo et al. [9] investigated the biologically inspired features (BIF) for human age estimation from faces. Unlike the previous works in [4][5], Guo simulated the human visual process based on bio-inspired models [10] by applying Gabor filters. A Gabor filter is a linear filter used in image processing for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and have been found to be particularly appropriate for textural representation and discrimination. Furthermore, previous bio-inspired models are changed by proposing a novel "STD" operation.

Our proposed system used the cascaded Adaboost learning algorithm in face detection and achieved the age estimation mechanism using Gabor wavelets and OLPP. This paper is organized in the following sections. First, our presented face detection system includes histogram lighting normalization, feature selection, the cascaded Adaboost classifier and the region-based clustering algorithm. The age estimation process, including the feature extraction using Gabor wavelets, feature reduction and selection, and age classification, is then introduced. Finally, the experimental results and conclusions are provided and summarized.

This paper proposes a fully automatic age estimation system using Gabor wavelets to represent aging progress. The system we proposed has four main modules: 1) face detection; 2) Gabor wavelet analysis; 3) OLPP reduction; and 4) SVM classification. The input image comes from a

camera frame or image file. First, the face is captured from an image using a face detector which is achieved using the AdaBoost approach presented in [12], and the image is resized to 64 by 64 pixels. After face detection, using 40 Gabor wavelet kernels, features are extracted and reduced by OLPPs. Lastly, age estimating from features using the SVM classification is conducted.

The remainder of this paper is organized as follows. Section 2 describes the sub-system of face detection using AdaBoost. Section 3 shows a facial age estimation algorithm including textural analysis using Gabor wavelets, data reduction based on orthogonal locality preserving projections and classification. Section 4 shows experimental results and comparisons. Finally, the conclusions on this system are presented in Section 5.

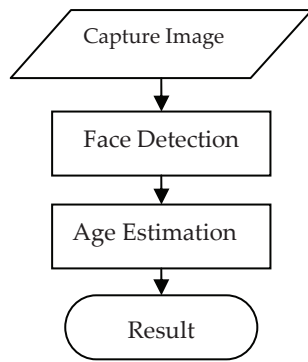


Figure 1. System overview.

## 2. Face Detection

Figure 1 displays the architecture of the automatic age estimation system in our work. The system consists of a face detection system localizing the facial regions in a captured image and an age estimator for the extracted face. Searching windows of various sizes are applied to an image to find multi-scale facial candidates as a result of object distance to camera during image capture. There are in total twelve block searching windows for multi-scale purposes and the window size is increased from the smallest (24x24) size with a scaling factor of 1.25. While a camera is acquiring an image, the camera may produce various illuminating intensities of image depending on the environment. The image can be more accurately recognized after its brightness was normalized.

### 2.1 Lighting normalization

The lighting normalization is based on the histogram fitting method. The primary task of histogram fitting is to transform the original histograms  $H(l)$  to the target histogram  $G(l)$ . The target histogram  $G(l)$  is chosen as the histogram of the image closest to the mean of the face database. Let the chosen target be image  $G(l)$  as shown in Figure 2(a), the images before and after normalization are shown in Figure 2(b)-(c). The input images that are too

dark or too light are normalized to the target image, by contrast, the histograms  $H(l)$  are fitted to  $G(l)$  by  $M_{H \rightarrow G}(l)$

$$M_{H \rightarrow G}(l) = M_{U \rightarrow G}(M_{H \rightarrow U}(l)) \quad (1)$$

where  $M_{H \rightarrow U}(l)$  and  $M_{U \rightarrow G}(l)$  are the histogram mapping and inverse mapping from  $H(l)$  and  $G(l)$  transforming into the histograms of uniform distributions, respectively.

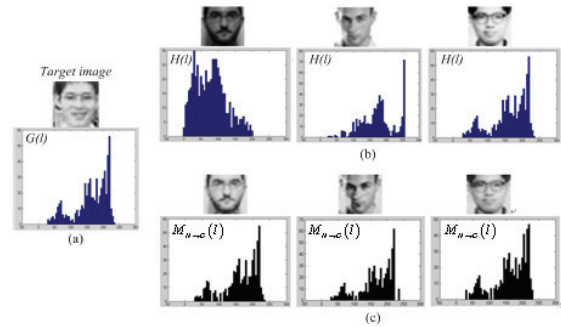


Figure 2. Lighting normalization. (a) Target image. (b) Input images. (c) Lighting normalization images.

### 2.1 Feature selection

The intensity-based features employed in our work were based on Haar features. We selected four types of rectangular features, as illustrated in Figure 3: the vertical edge, horizontal edge, vertical line and diagonal edge, as proposed by Papageorgiou [13]. It is feasible to use a composition of multiple different brightness rectangles to present the light and dark regions in the image. The features are defined as:

$$value_{subtracted} = f(x, y, w, h, Type) \quad (2)$$

where  $(x, y)$  indicate the origin of the relative coordinate of rectangular features in the searching window. The significance of  $w$  and  $h$  denote the relative width and height of rectangular features, respectively.  $Type$  presents the type of rectangular features and  $value_{subtracted}$  is the sum of the pixels in the white rectangle subtracted from the dark rectangle.

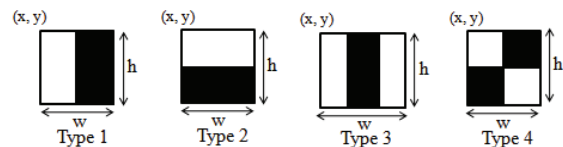
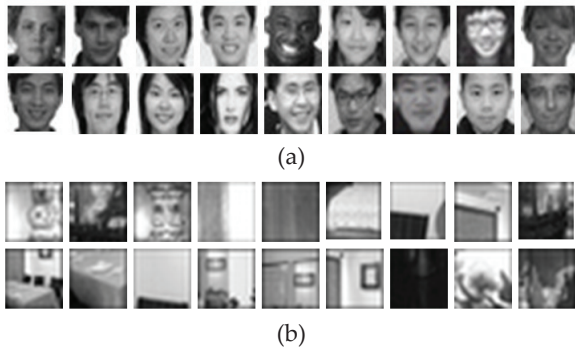


Figure 3. Four types of rectangle features.

A single rectangle feature which most effectively separates the face and non-face samples can be considered as a weak classifier  $h(x, f, p, \theta)$  as shown in the following equation:

$$h(x, f, p, \theta) = \begin{cases} 1, & \text{if } pf(x) < p\theta \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The weak classifier  $h(x, f, p, \theta)$  used to determine if the  $x$ -block image is a face or a non-face depends on a feature  $f(x, y, w, h, type)$ , a threshold  $q$  and a polarity  $p$ , indicating the direction of inequality. For each weak classifier, an optimal threshold is chosen to minimize misclassification. The selected threshold for each rectangle feature is trained by a face database, consisting of 4,000 face images and 59,000 non-face images. Figures 4(a)-(b) present examples from the face and non-face databases. In this procedure, we collect the distribution of each feature  $f(x, y, w, h, type)$  for each image in the database, and then choose a threshold that discriminates the two classes and obtains a detection rate higher than those of the others. Although each rectangular feature can be computed highly efficiently, computing the complete set is prohibitively expensive. For example, for the smallest (24x24) search window, the exhaustive set of rectangular features totals 160,000.



**Figure 4.** Database of face detection system. (a) Face images. (b) Non-face images.

The Adaboost method combines a collection of weak classifiers to form a stronger classifier. Although the stronger classifier is effective for face detection application, it is still time consuming. A structure of cascaded classifiers which improve the detection performance and reduce the computation time was proposed by Viola and Jones [14]. Based on this idea, our cascade Adaboost classifier will work stage by stage to classify a face and form a stronger classifier. In stage 1, if an image-block is classified as a face then it will allow entering stage 2, otherwise it is rejected. Likewise stage 3 can continue only if the object has been classified as a face at stage 2. The number of stages must be sufficient to achieve an excellent detection rate while minimizing computation. For example, if each stage has a detection rate of 0.99 (since  $0.9 \approx 0.99^{10}$ ), a detection rate of 0.9 can be achieved using a 10-stage classifier. While achieving this detection rate may sound like a daunting task, it is made significantly easier by the fact that each stage need only achieve a false positive rate of about 30%.

The procedure of the Adaboost process is described as follows: if  $m$  and  $l$  are the number of non-face and face, respectively, and  $j$  is the sum of non-face and face

samples, the initial weight  $w_{i,j}$  for  $i$ -stage can be defined as  $w_{i,j} = \frac{1}{2m}, \frac{1}{2l}$  for  $y_j = 0, 1$ . The normalized weighted error w. r. t the weak classifier can be expressed as below:

$$\varepsilon_i = \min_{f,p,\theta} \sum_j w_{i,j} |h(x_j, f, p, \theta) - y_j| \quad (4)$$

The weights are updating by Eq. (5) in each iteration, if the object is classified correctly then  $e_j = 0$ , otherwise  $e_j = 1$ .

$$w_{i,j} = w_{i,j} \beta_i^{1-e_j} \quad (5)$$

The final classifier for  $i$ -stage is defined below:

$$C(x_j) = \begin{cases} 1, & \alpha_i h(x_j, f, p, \theta) \geq \frac{1}{2} \alpha_i \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where  $\alpha_i = \log \frac{1}{\beta_i}$  and  $\beta_i = \frac{\varepsilon_i}{1 - \varepsilon_i}$

### 2.3 Region based clustering

The face detector usually finds more than one face candidate even though only a single face appears in an image, as illustrated in Figure 5. Therefore, a region-based clustering method is used to solve this kind of problem. The proposed region-based clustering method consists of two levels of clustering, local and global-scale clustering. The local-scale clustering is used to cluster the blocks in the same scale and design a simple filter to determine the number of blocks within clusters. While the number of blocks in some clusters is more than one, that cluster will be reserved as a possible face candidate, otherwise it will be discarded. The local-scale clustering judges if the blocks meet the decision rule in:

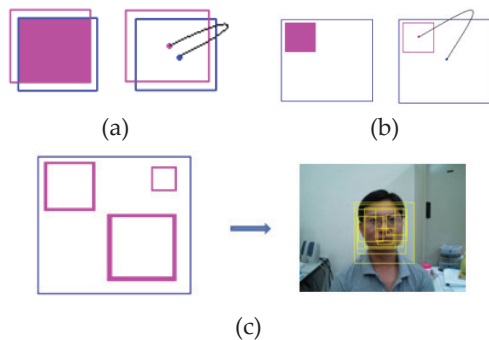
$$cluster(x, y) = \begin{cases} 1, & \text{if } overlap\_rate(x, y) \geq TH_{overlap\_rate} \\ & \text{and } distance(x, y) \leq TH_{distance} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

In Eq. (7), the overlap rate  $(x, y)$  is the percentage overlapped between two detected regions,  $x$  and  $y$ , and distance  $(x, y)$  is the distance of centres in these two regions. The equality,  $cluster(x, y) = 1$  means the block  $x$  and  $y$  are in the same cluster and the regions are completely overlapped

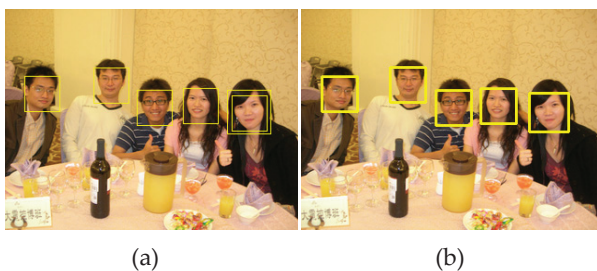


**Figure 5.** Face detector result.

Figure 6 shows several cases of the clustering process. In Figure 6(a), the two blocks are processed as the same cluster, and in Figure 6(b) the two blocks are processed as different clusters because the distance of the centres does not satisfy  $distance(x,y) \leq TH_{distance}$ . For special cases, as shown in Figure 6(c), they are all considered as facial candidates but most of them are false accept blocks. Therefore in this paper for practical applications we only choose one block that satisfies Eq. (7) rather than select multiple blocks. At the end, the global-scale clustering will use the blocks obtained from local-scale clustering, and label the facial regions according to the average size of all available blocks. Some results in the entire region-based clustering process for both local-scale and global-scale levels will be shown in Figure 7. From the right image in Figure 5, in fact, only one block will be precisely clustered as a facial region after applying our local and global clustering processes, even though more than five facial candidates are obtained for an image with only five faces.



**Figure 6.** The chart of overlapped regions and distances of the centres of two blocks. (a) Case 1. (b) Case 2. (c) Special case in the cluster: more than two blocks overlapping.



**Figure 7.** Region-based clustering result. (a) The results of clustering in local-scale and (b) in global-scale.

### 3. Age Estimation

There are three major parts to our age estimation system in this work, age feature extraction, feature reduction and feature classification. The age feature extractor is constructed using Gabor wavelets that were used for image analysis because of their biological relevance and computational properties. Gabor wavelet kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibiting strong characteristics of spatial locality and orientation selectivity, and are

optimally localized in space and frequency domains. The Gabor wavelet transform is *generally acknowledged* to be particularly suitable for image decomposition and representation when the goal is the derivation of local and discriminating features. Moreover, Donato et al. [15] have experimentally shown that the Gabor wavelet representation resulted in higher performance for classifying facial actions. This section introduces the basics of the Gabor wavelet feature representation of images, describes the feature reduction and selection, and derives a Gabor feature vector for age estimation.

#### 3.1 Feature extraction using Gabor wavelets

A Gabor wavelet  $\psi_{\mu,\nu}$  can be defined as follows [16]:

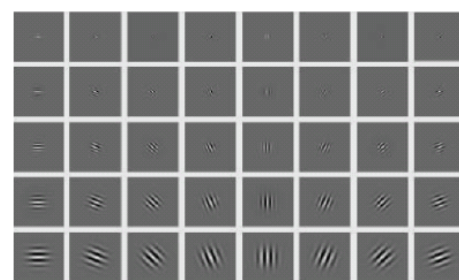
$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}} \left[ e^{ik_{\mu,\nu}z} - e^{-\frac{\sigma^2}{2}} \right] \quad (8)$$

where  $\mu$  and  $\nu$  define the orientation and scale of the Gabor kernels,  $z = (x,y)$ ,  $\|\cdot\|$  demotes norm operator, and the wave vector  $k_{\mu,\nu}$  is defined as follows:

$$k_{\mu,\nu} = k_\nu e^{i\varphi_\mu} \quad (9)$$

where  $k_\nu = k_{\max} / f^\nu$  and  $\varphi_\nu = \pi\nu / 8$ .  $k_{\max}$  is the maximal frequency, and  $f$  is the spacing factor between kernels in the frequency domain. Generally, the Gabor kernels in Eq. (8) are all self-similar because they can be generated from one filter, the mother wavelet, by scaling and rotation via the wave vector  $k_{\mu,\nu}$ . Each kernel is a product of a Gaussian envelope and a complex plane wave, while the first term in the square brackets in Eq. (9) determines the oscillatory part of the kernel, and the second term compensates for the DC value. The parameter  $\sigma$  is related to the standard derivation of the Gaussian envelope's width to the wavelength.

In most cases, researchers use Gabor wavelets at five different scales,  $\nu \in \{0, \dots, 4\}$ , and eight orientations,  $\mu \in \{0, \dots, 7\}$  [16-18]. Figure 8 shows the real part of the Gabor kernels at five scales and eight orientations, and their magnitudes, with the following parameters:  $\sigma = 2\pi$ ,  $k_{\max} = \pi / 2$ , and  $f = \sqrt{2}$ .



**Figure 8.** Region-based clustering result. (a) The results of clustering in local-scale and (b) in global-scale.

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined using Eq. (8). Let  $I(z)$  be the grey level distribution of an image. The convolution output of image  $I$  and  $\psi_{\mu,\nu}(z)$  is defined as follows:

$$O_{\mu,\nu}(z) = I(z) * \psi_{\mu,\nu}(z) \quad (10)$$

where  $z=(x, y)$  and  $*$  denotes the convolution operator.

To apply the convolution theorem, the Fast Fourier Transform (FFT) is used to derive the convolution output. Eq. (11) and Eq. (12) are the definition of convolution via FFT.

$$\mathfrak{F}\{O_{\mu,\nu}(z)\} = \mathfrak{F}\{I(z)\}\mathfrak{F}\{\psi_{\mu,\nu}(z)\} \quad (11)$$

$$O_{\mu,\nu}(z) = \mathfrak{F}^{-1}\{\mathfrak{F}\{I(z)\}\mathfrak{F}\{\psi_{\mu,\nu}(z)\}\} \quad (12)$$

where  $\mathfrak{F}$  and  $\mathfrak{F}^{-1}$  denote the Fourier and inverse Fourier transform, respectively.



Figure 9. Sample image and magnitude of 40 convolution outputs.

Figure 10 shows the magnitude of convolution outputs of a sample image. The outputs exhibit strong characteristics of spatial locality, as well as scale and orientation selectivity corresponding to those displayed in Fig. 9. Such characteristics produce salient local features that are suitable for visual event recognition. Hereafter, we indicate with  $O_{\mu,\nu}(z)$  the magnitude of the convolution outputs.

### 3.2 Feature reduction by scheme

Generally, Principle Component Analysis (PCA) or other algorithms will follow with Gabor wavelet feature extraction to reduce dimensionality of the transformed data [19][20]. The convolution results corresponding to all Gabor wavelets are put together as a whole to enhance

the computational efficiency when Principle Component Analysis (PCA) is applied to dimensional reduction. Three different schemes have been proposed: (a) Parallel Dimension Reduction Scheme (PDRS): Gabor wavelet features are extracted from each sample as shown in Figure 10. Training each PCA projection matrix in every channel and combining these features using a voting method. (b) Ensemble Dimension Reduction Scheme (EDRS): the EDRS is the most common scheme used for Gabor wavelet feature. As shown in Figure 11, the difference between PDRS and EDRS is that the EDRS concatenates Gabor wavelet features instead of using them in parallel. (c) Multi-channel Dimension Reduction Scheme (MDRS). Xiaodong Li et al. [21] proposed MDRS in 2009 - as shown in Figure 13, the main idea of MDRS is training a PCA projection matrix for the same channel between different samples. In [21], Xiaodong Li et al. have already proved that MDRS has higher performance than EDRS in facial feature extraction using a Gabor wavelet transform.

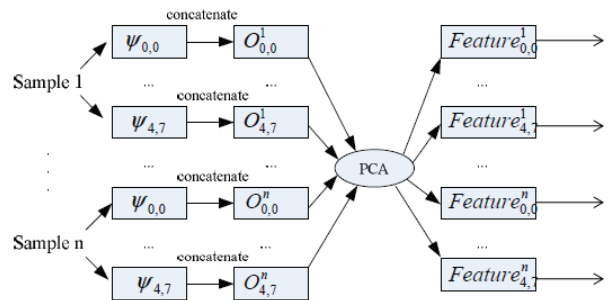


Figure 10. Parallel Dimension Reduction Scheme.

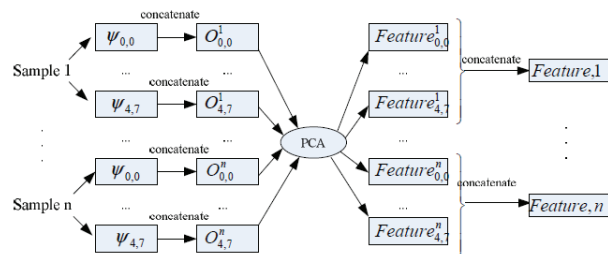


Figure 11. Ensemble Dimension Reduction Scheme.

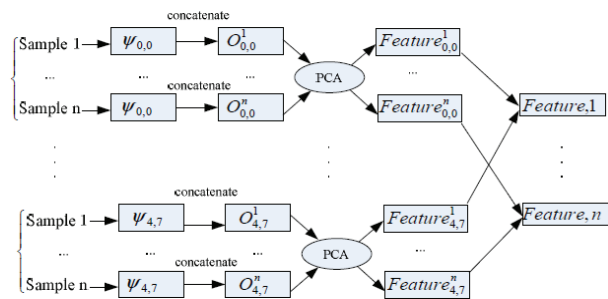


Figure 12. Multi-channel Dimension Reduction Scheme.

To compare the performance of PDRS and MDRS, the K-Nearest Neighbour (KNN) classifier is used for experimentation. For PDRS, we used a voting method called "Gaussian voting" to combine 40 channels. The

concept of Gaussian voting is described as using a KNN classifier for each channel to predict 40 ages. Each predicted age is treated as the mean value of a Gaussian distribution and is counted as a histogram. The highest peak is the final predicted answer. For MDRS, we use the concatenated feature directly.

The FG-NET Aging Database [22] is adopted for experiments. The database contains 1,002 high-resolution colour and grey-scale face images with large variations in lighting, pose and expression. There are a total of 82 subjects (multiple races) ranged in age from 0 to 69 years. We used the mean absolute error (MAE) criterion to evaluate the performance of each age estimation. The MAE denotes the average of the absolute errors between the estimated ages and ground truth ages. The mathematical function is defined as:

$$MAE = \frac{\sum_{k=1}^N |\hat{l}_k - l_k|}{N} \quad (13)$$

where  $\hat{l}_k$  is the ground truth age for the test image  $k$ ,  $l_k$  is the estimated age and  $N$  is the total number of test images. Table 1 shows the experimental results of two schemes. The results of experiments demonstrate that the MDRS is a better scheme than the PDRS.

Scheme	PDRS	MDRS
MAE	13.39	11.89

Table 1. MAE of PDRS and MDRS.

### 3.3 Feature selection

The dimensionality of the Gabor wavelet feature space is overwhelmingly high, even though the dimension reduction scheme has already been applied. Therefore it is important to select the more significant features and to further reduce the dimension to a low-dimensional space. Three typical dimensionality reduction methods have been proposed in past research besides PCA. (a) Linear Discriminant Analysis (LDA) is similar to the PCA method [23]; the difference is that LDA uses class information to improve itself. (b) Locality Preserving Projections (LPP) search the sub-space that preserves essential manifold structure by measuring the local neighbourhood distance information [24]. (c) Orthogonal Locality Preserving Projections (OLPP) produce orthogonal basis functions based on LPP and preserve the metric structure [25]. To determine which reduction method from the above is most suitable for use in age features from Gabor wavelets, we used the KNN classifier for experimentation and use the MAE criterion to evaluate performance. In the experiment, we changed the affinity weight of LPP and OLPP to obtain more detail. Table 2 shows the MAE of each reduction method. The OLPP with cosine distance affinity weight has the best performance in age estimation

Method	Best dimension	MAE
LDA	4096	11.15
LPP	61	10.52
LPP_Heat	254	10.99
LPP_Cosine	45	10.28
OLPP	49	9.24
OLPP_Heat	353	10.99
OLPP_Cosine	43	8.89

Table 2. MAEs of different reduction methods.

### 3.4 Age classification

The Gabor wavelet features are used in the SVM classifier to identify how old the face is. Support Vector Machines (SVMs) have considerable potential as classifiers of sparse training data as they were developed to solve classification and regression problems. SVMs have similar roots with neural networks, and they demonstrate the well-known ability of being universal approximates of any multivariable function to any desired degree of accuracy. This approach was produced by Vapnik et al. using a statistical learning theory [25-27]. Table 1 and Figure 11 show the comparisons of results using our conditional entropy-based feature selection approach with those by others for feature selection and classification. All the comparisons in this paper used the same training and testing database. The database was composed of 1002 high-resolution colour or grey-scale face images with large variations in lighting, pose and expression. There are 82 subjects (multiple races) in total with ages ranging from 0 to 69 years. Our used input dimensions of SVM in the comparison process were 43 which are shown in Table 2. In addition, we also compared the accuracy rate with the same Gabor wavelet features and KNN in classification.

## 4. Experimental Results

The database that we adopt for age estimation experiments is the FG-NET Aging Database [20]. This database is a publically available age database containing 1002 high-resolution colour or grey-scale face images with large variations in lighting, pose and expression. There are 82 subjects (multiple races) in total with ages ranging from 0 to 69 years. Figure 13 shows a serial of sample images from the same subject at different ages.

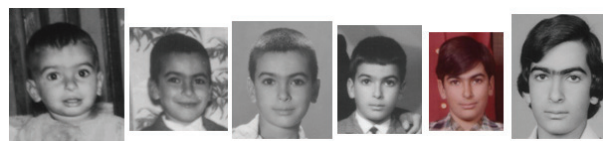


Figure 13. Sample images of the same subject at different ages.

To evaluate the age estimation performance, the facial area in each image was located by the face detector described in Section 2. A leave-one-person-out (LOPO) test scheme was used in the experiments. Each face image was cropped and resized to 64 x 64 pixels and the colour information was transformed to 256 grey level. We used the classifier of SVMs with parameters of the RBF kernel, where cost  $c$  is 0.5 and gamma  $g$  is 0.0078125 for the FG-NET database. Our main focus is the new feature based on Gabor wavelets.

The performance of age estimation can be measured by two different measures: the mean absolute error (MAE) and the cumulative score (CS). The MAE is defined as the average of the absolute errors between the estimated ages and the ground truth ages. The MAE measure has been used previously in [2-10]. The cumulative score is defined as

$$CS(j) = N_{e \leq j} / N \times 100\% \quad (14)$$

where  $N_{e \leq j}$  is the number of test images on which the age estimation makes an absolute error no higher than  $j$ .

Table 3 shows the experimental results. We compare our results with all previous methods reported on the FG-NET age database. The Gabor-OLSS method of this study has the MAEs of 8.43 and 5.71 years for using KNN and SVM respectively, which are explicitly smaller than most previous results under the same experimental protocol. Our method offers approximately 16% deductions of MAEs over the result of AGES [2]. In Table 3, we can see that the LARR [4] method and BIF [9] method have more favourable MAEs of 5.07 and 4.77 than do ours.

Method	MAE
WAS [2]	8.06
AGES [2]	6.77
MLPs [28]	6.98
LARR [4]	5.07
BIF	10.32
Ours (KNN)	8.43
Ours (SVM)	5.71

Table 3. MAEs of different methods.

As mentioned previously, our purpose is to build a “fully-automatic” age estimation system. The LARR method uses the AAM features FG-NET provided directly, meaning it usually needs human involvement in aligning the feature points. In our survey, there is still no efficient method that can automatically align feature points quickly and correctly. For applications, the LARR method may require considerable effort in aligning feature points. The MAE of the BIF is explicitly more

favourable than the method we propose. In order to verify their results, we tried to implement the BIF method. The result of the implemented BIF is quite poor, having an MAE of 10.32. Furthermore, the BIF method requires a large amount of time when extracting the aging features. Compared to our system, the BIF method requires an extraction time more than twice that of ours. Our method increases the performance of feature extraction to approximately 12 to 15 images per second.

The comparisons of cumulative scores are shown in Figure 14. Our Gabor-OLPP method performs much better than WAS and MLPs methods. The method of AGES is close to our Gabor-OLPP method in low age error levels, but lower than those of Gabor-OLPP when error level is larger than five.

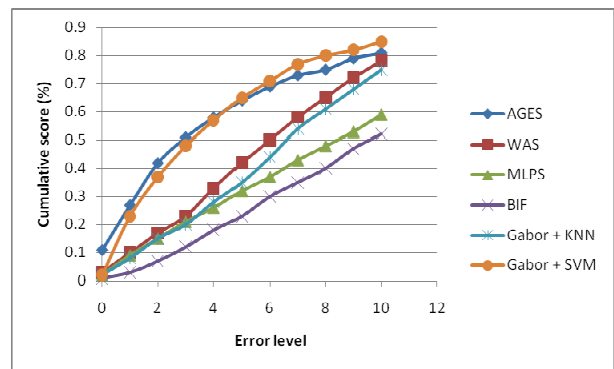


Figure 14. CS of each method.

## 5. Conclusions

In this paper, we propose a new framework for automatic age estimation of face images. A Gabor wavelet transform is first introduced for age estimation to achieve real-time and fully-automatic aging feature extraction; SVMs have considerable potential as classifiers of sparse training data and provide robust generalization ability.

Most previous studies have used PCA only to reduce the dimensionality of the Gabor wavelet features; but PCA exhibits inadequate efficiency when we use general Gabor wavelet features directly. By exchanging efficiency for accuracy of classification, previous researchers have usually attempted to select only the features they require, rather than using all the features. Therefore, data reduction methods are more convenient for selecting the target features. We compare four different typical data reduction methods; OLPP provides the lowest dimensionality of feature vectors and the most favourable discrimination from further extraction.

## 6. Acknowledgments

This work was supported in part by the Department of Industrial Technology under grant: 100-EC-17-A-02-S1-032, and supported in part by the Taiwan National Science Council under grant: NSC-100-2218-E-009-023.



## 7. References

- [1] Paul V, Jones M.J. (2004) Robust Real-Time Face Detection. *International Journal of Computer Vision* 57(2), 137-154
- [2] Geng X, Zhou Z-H, Zhang Y, Li G, Dai H. (2006) Learning from facial aging patterns for automatic age estimation, In *ACM Conf. on Multimedia*, pages 307–316
- [3] Geng X, Zhou Z-H, Smith-Miles K. (2007) Automatic age estimation based on facial aging patterns. *IEEE Trans. on PAMI*, 29(12): 2234–2240
- [4] Guo G, Fu Y, Dyer, C.R., Huang, T.S. (2008) Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression. *IEEE Trans. on Image Processing*, 17(7): 1178-1188
- [5] Guo G, Fu Y, Huang T.S. and Dyer, C.R. (2008) Locally Adjusted Robust Regression for Human Age Estimation. *IEEE Workshop on Applications of Computer Vision*, pages 1-6.
- [6] Kwon Y, Lobo N. (1999) Age classification from facial images. *Computer Vision and Image Understanding*, 74(1): 1–21
- [7] Txia J-D and Huang C-L. (2009) Age Estimation Using AAM and Local Facial Features. *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 885-888
- [8] Yan S-C, Zhou X and Liu M. Hasegawa-Johnson, M., Huang, T.S. (2008) Regression from patch-kernel. *IEEE Conference on CVPR*, pages 1-8
- [9] Guo G, Mu G, Fu Y and Huang T.S. (2009) Human age estimation using bio-inspired features. *IEEE Conference on CVPR*, pages 112-119.
- [10] Serre T, Wolf L, Bileschi S, Riesenhuber M and Poggio T. (2007) "Robust Object Recognition with Cortex-Like Mechanisms. *IEEE Trans. on PAMI*, 29(3): 411–426
- [11] Lin C-T, Siana L, Shou Y-W, Yang C-T (2010) Multi-client Identification System using Adaptive Probabilistic Model. *EURASIP Journal on Advances in Signal Processing*. Vol. 2010
- [12] Paul V and Jones M.J. (2004) Robust Real-Time Face Detection. *International Journal of Computer Vision* 57(2), 137-154
- [13] Papageorgiou C. P, Oren M and Poggio T. (1998) A general framework for object detection. in *Proceedings of the 6th IEEE International Conference on Computer Vision*, pp. 555–562
- [14] Viola P and Jones M.J. (2004) Robust real-time face detection. *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154
- [15] Donato G, Bartlett MS, Hager J.C., Ekman P and Sejnowski T.J. (1999) Classifying facial actions. *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 974–989
- [16] Wiskott L, Fellous J, Kruger N and Malsburg C. (1997) Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775–779
- [17] Liu C and Wechsler H. (2002) Gabor feature based classification using enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, vol. 11, pp. 467–476
- [18] Liu C. (2004) Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 572–581.
- [19] Belhumeur P.N., Hespanha J.P. and Kriegman D.J. (1997). "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7): 711-720.
- [20] Duda R.O., Hart P.E., and Stork D.G. (2000) *Pattern Classification*, 2<sup>nd</sup> ed. New York: Wiley Interscience
- [21] Li X, Fei S and Zhang T. (2009) Novel Dimension Reduction Method of Gabor Feature and Its Application to Face Recognition. *International Congress on Image and Signal Processing, 2009. CISP '09*. 2<sup>nd</sup>, Page(s): 1-5
- [22] The FG-NET Aging Database [Online]. Available: <http://www.fgnet.rsunit.com/>
- [23] He X-F, Yan S-C, Hu Y-X, Niyogi P and Zhang H-J. (2005) Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(3): 328-340.
- [24] Cai D, He X-F, Han J-W and Zhang H-J. (2006) Orthogonal Laplacianfaces for Face Recognition. *IEEE Transactions on Image Processing* 15(11): 3608-3614.
- [25] Mercier G and Lennon M. (2003) Support vector machines for hyperspectral image classification with spectral-based kernels. in *Proc. IGARSS, Toulouse, France, July 21–25*.
- [26] Abe S. (2005) *Support Vector Machines for Pattern Classification*. London: Springer-Verlag London Limited.
- [27] Wang L. (2005) *Support Vector Machines: Theory and Applications*. New York: Springer, Berlin Heidelberg
- [28] Lanitis A, Draganova C and Christodoulou C. (2004) Comparing different classifiers for automatic age estimation. *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 621–628