5

Pixel-Level Decisions based Robust Face Image Recognition

Alex Pappachen James
Queensland Micro-nanotechnology center, Griffith University
Australia

1. Introduction

Face recognition is a special topic in visual information processing that has grown to be of tremendous interest to pattern recognition researchers for the past couple of decades (Delac & Grgic, 2007; Hallinan et al., 1999; Li & Jain, 2005; Wechsler, 2006; Zhao & Chellappa, 2005). However, the methods in general faces the problem of poor recognition rates under the conditions of: (1) large changes in natural variability, and (2) limitations in training data such as single gallery per person problem. Such conditions are undesirable for face recognition as the inter-class and intra-class variability between faces become high, and the room for discrimination between the features become less.

Major methods that are employed to reduce this problem can be classified into three groups: (1) methods whose so-called gallery set consists of multiple training images per person (e.g. Etemad & Chellappa (1997); Jenkins & Burton (2008)) (2) image preprocessing techniques that aim at feature restoration (e.g. Ahlberg & Dornaika (2004)), and (3) use of geometrical transforms to form face models (e.g. Ahlberg & Dornaika (2004)). Even though they show high performance under specific conditions they lack robust performance and in many cases have proved to be computationally expensive. Being distinct from these computational schemes, the human visual system, which is the best available natural model for face recognition, uses modular approach for classification of faces (Moeller et al., 2008).

This chapter presents a method (James, 2008; James & Dimitrijev, 2008) that implements the concept of local binary decisions to form a modular unit and a modular system for face recognition. This method is applied to formulate a simple algorithm and its robustness verified against various natural variabilities occurring in face images. Being distinct from a traditional approach of space reduction at feature level or automatic learning, we propose a method that can suppress unwanted features and make useful decisions on similarity irrespective of the complex nature of underlying data. The proposed method in the process do not require dimensionality reduction or use of complex feature extraction or classifier training to achieve robust recognition performance.

2. Proposed method

Understanding vision in humans at the level of forming a theoretical framework suitable for computational theory, has opened up various disagreements about the goals of cortical processing. The works of David Marr and James Gibson are perhaps the only two major attempts

to provide deeper insight. In majority of Marr's work (Marr, 1982), he assumed and believed vision in humans to be nothing more that a natural information processing mechanism, that can be modelled in a computer. The various levels for such a task would be: (1) computational model, (2) a specific algorithm for that model, and (3) a physical implementation. It is logical in this method to treat each of these level as independent components and is a way to mimic the biological vision in robots. Marr attempted to set out a computational theory for vision in a complete holistic approach. He applied the *principle of modularity* to argue visual processing stages, with every module having a function. Philosophically, this is one of the most elegant approach proposed in the last century that can suit both the paradigms of software and hardware implementations. Gibson on the other hand had an *"ecological"* approach to studying vision. His view was that vision should be understood as a tool that enables animals to achieve the basic tasks required for life: avoid obstacles, identify food or predators, approach a goal and so on. Although his explanations on brain perception were unclear and seemed very similar to what Marr explained as algorithmic level, there has been a continued interest in the *rule-based* modeling which advocates knowledge as a prime requirement for visual processing and perception.
Both these approaches have a significant impact in the way in which we understand the visual systems today. We use this understanding by applying the principles of modularity and hierarchy to focus on three major concepts: (1) spatial intensity changes in images, (2) similarity measures for comparison, and (3) decision making using thresholds. We use the following steps as essential for forming a baseline framework for the method presented in this chapter:

Step 1 Feature selection of the image data: In this step the faces are detected and localized. Spatial change detection is applied as a way to normalize the intensity features without reducing the image dimensionality.

Step 2 Local similarity calculation and Local binary decisions: The distance or similarity between the localized pixels from image to another image is determined. This results in a pixel-to-pixel similarity matrix having same size as that of the original image. Inspired from the binary nature of the neuron output we make local decisions at pixel level by using a threshold θ on the similarity matrix.

Step 3 Global similarity and decision: Aggregating all the local decisions, a global similarity score is obtained for the comparisons between a test image with different images. Based on the global similarity scores, they are ranked and the one with the highest similarity score selected as the best match.

These steps are summarised graphically in Fig. 1.

2.1 Feature Selection

The visual features mapped by the colour models used in the camera device are influenced by variations in illumination, spatial motions and spatial noise. Although noise and motion errors can be corrected at the camera itself, illumination correction or normalization is seldom done. The human eye on the other hand has inherent mechanical and functional mechanisms to form illumination invariant face images under a wide range of lighting conditions. Feature localization in humans is handled by feedback mechanisms linked to human eye and brain. However, in the case of automatic face image recognition, a perfect spatial localization of features is not possible using existing methods. Face detection methods are used to detect the face images and localize the feature with some degree of accuracy. Even after features are localised by any automatic detection methods, it is practically impossible to attain a perfect

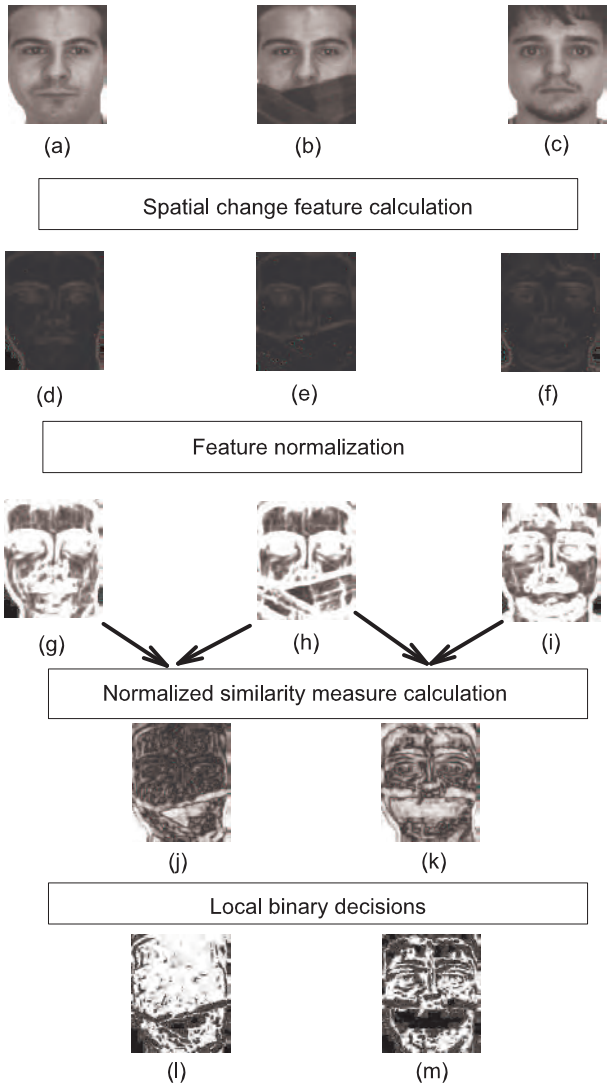

Fig. 1. An illustration of various steps in the baseline algorithm. The images labeled (a),(b), and (c) show the raw images, where (a) and (c) form the gallery images and (b) is a test image, all taken from the AR database (Martinez & Benavente, 1998). The images labeled (d), (e), and (f) show the output of a feature selection process, which corresponds to the raw images (a), (b), and (c), respectively. The normalized feature vectors are shown as the images labeled (g), (h), and (i), and are calculated from (d), (e), and (f), respectively. This is followed by comparisons of test image with gallery images. The normalized similarity measure when applied for comparing (h) with (g) and (h) with (i) results in images labeled (j) and (k), respectively. Finally, the local binary decisions when applied on (j) and (k) result in binary vectors labeled (l) and (m), respectively. Clearly, in this example, (b) is a best match to (a) due to more white areas (more similarity decisions) in (l) than in (m).

alignment due to random occlusions and natural variations that depend on environment. As a result, we need to integrate an error correction mechanism to reduce the impact of localization error by applying image perturbations. The perturbations can be applied with respect to an expected spatial coordinate such as eye coordinates. Ideally, any pixel shift from these expected coordinates results in rotation, scale or shift error. So to undo such errors, by the idea of reverse engineering, pixel shifts are applied to the expected coordinate to detect the face images. In this way any arbitrary N number of pixel shifts on an image results in N number of perturbed images, one of which will be localised the best.

After the raw features are localized, they are processed further to extract features through the detection of spatial change as an essential visual cue for recognition. Spatial change in images can be detected using spatial filtering and normalization mechanisms such as local range filtering, local standard deviation filtering, gradient filtering or gabor filtering.

The relative change of spatial intensity of a pixel in a raw image with respect to the corresponding pixels in its neighbourhood can be used to form features useful for recognition. In the baseline algorithm we can detect such features by calculating the local standard deviation on the image pixels encompassed by a window w of pixels of size $m \times n$ pixels. This type of spatial operation is known as a kernel based local spatial filtering. The local standard deviation filter is given by the following equation:

$$\sigma(i,j) = \sqrt{\frac{1}{mn}\sum_{z=-a}^{a}\sum_{t=-b}^{b}[I(i+z,j+t) - \overline{I(i,j)}]^2} \tag{1}$$

where $a = (m-1)/2$ and $b = (n-1)/2$. The local mean $\overline{I(i,j)}$ used in (1) is calculated by the following equation:

$$\overline{I(i,j)} = \frac{1}{mn}\sum_{s=-a}^{a}\sum_{t=-b}^{b} I(i+s,j+t) \tag{2}$$

In Fig. 1, the images labeled (a), (b), and (c) show the raw images, whereas the images labeled (d), (e), and (f) show the corresponding spatial change features [using Eq. (1)] respectively. The normalized spatial change features $\hat{x}$ are calculated using the following equation:

$$x(i,j) = \frac{\sigma(i,j)}{\overline{\hat{\sigma}}} \tag{3}$$

where the spatial change features σ are normalized using the global mean $\overline{\hat{\sigma}}$. The global mean is calculated by the following equation:

$$\overline{\hat{\sigma}} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\sigma(i,j) \tag{4}$$

In Fig. 1, the images labeled (g), (h), and (i) show the normalized spatial change features which is obtained by applying global-mean normalization on spatial change features images labeled (d), (e), and (f), respectively.

An extension to this class of filters is the V1-like features generated from Gabor filters that detect different types of spatial variations in the images. The advantage of Gabor filters for feature extraction in face recognition was evident through the works of (Zhang et al., 2005).

These suggest that like the Gradient filters Gabor filters can be used for preprocessing the images. Formally, Gabor filters are defined as:

$$\psi_{\mu,v}(z) = \frac{\| k_{\mu,v} \|^2}{\sigma^2} e^{(-\|k_{\mu,v}\|^2 \|z\|^2 / 2\sigma^2)} [e^{ik_{\mu,v}z} - e^{-\sigma^2/2}] \tag{5}$$

where μ defines the orientation, v defines the scale of the Gabor filters, $k_{\mu,v} = \frac{k_{max}}{\lambda^v} e^{i\frac{\pi\mu}{8}}$, λ is the spacing between the filters in frequency domain and $\| \, . \, \|$ denotes the norm operator. The phase information from these filters is not considered, and only its magnitude explored. For the experiments, we set the value of parameters as follows: $\lambda = \sqrt{2}$, $\sigma = 2\pi$ and $k_{max} = \pi/2$. Further by considering five scales $v \in \{0, \ldots 4\}$ and eight orientations $\mu \in \{0, \ldots, 7\}$ which on convolution result in 40 filters. Again, these class of filters work on the primary principle of local feature normalization through spatial change detection and provide a way to reduce natural variability present in intensity raw image. Following these filtering operations, the images are normalized using local mean filtering to readjust the signal strength locally.

2.2 Local similarity calculation and binary decisions

What is similarity? This question has eluded researchers from various fields for over a century. Although the idea of similarity seem simple, yet it is very different from the idea of difference. The difficulty lie in the idea of expressing similarity as a quantitative measure, for example, unlike a difference measure such as Euclidean distance there is no physical basis to similarity that can be explained. Although, perception favours similarity, the use of an exact mathematical equation dose not properly justify meaning of similarity.

Type	Equation
Min-max ratio	$min[x_g, x_t] / max[x_g, x_t]$
Difference	$\left\|x_g - x_t)\right\| / \gamma$
Exponential difference	$e^{-\left\|x_g - x_t\right\| / \gamma}$
	where γ is $max[x_g, x_t]$ or $[x_g + x_t]/2$ or $min[x_g, x_t]$

Table 1. Normalized similarity measures

The absolute difference between pixels is a well known distance measure used for the comparison of features and can be used to find the similarity. Further, element wise normalization of this similarity measure is done by taking the minimum of each feature within test image x_t and gallery image x_g under comparison. This feature by feature comparison results in a normalized similarity measure δ, which is given by:

$$\delta(i,j) = \frac{|x_g(i,j) - x_t(i,j)|}{min(x_g(i,j), x_t(i,j))} \tag{6}$$

Similarity measures based on this idea of measurement are shown in Table 1. However, they suffer from the inter-feature similarities being detected as true similarities from patterns involving natural variability. We find a way to get around this problem by reducing the inter-feature similarity and maintain only relevant differences through a combination of steps involving local similarity calculation and pixel-level binary decision. Inspired from the idea of ability of neurons to compare and make a binary decision at local level, we apply local similarity measures followed by a local binary decision (see Table 1). In the comparison of images this translates into pixel to pixel local similarity calculation followed by an application of a

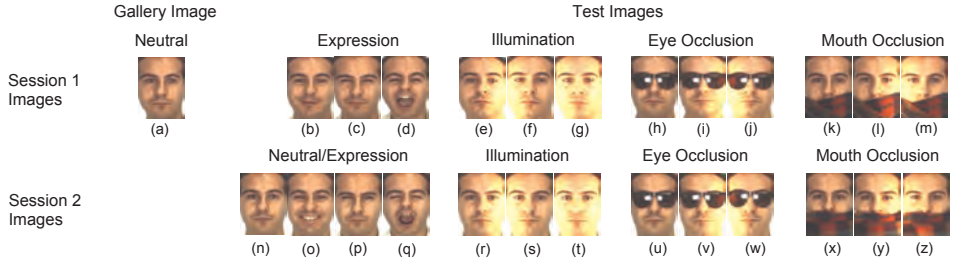

Fig. 2. The illustration shows the images of a person in the AR database (Martinez & Benavente, 1998; 2000) and its organization for the single training sample per person problem depicted in this article. The session 1 image having a neutral facial expression is selected as the gallery image. The remaining 25 images from session 1 and session 2 are used as test images.

binary decision using a threshold. The inherent ability of neurons to exhibit a logic high or logic low state based on the ionic changes occurring due to the presumably threshold limited variations in input connections inspires this idea of local decision. The resulting output for the local similarity measure $S_l(i,j)$ that is defined as to represent $S_l(i,j) = 0$ as least similar and $S_l(i,j) = 1$ as most similar, when applied on a threshold θ to form the binary decision $B(i,j)$ takes the form $B(i,j) = 1$ if $S_l(i,j) <= \theta$ and $B(i,j) = 0$ if $S_l(i,j) > \theta$. The values generated by B represents the local decision space of the image comparison.

2.3 Global similarity and decision

Local decisions on similarity give the similarity match at pixel level, this however is only useful if it can be used at a higher level of decision level abstraction. A reduction of the decision space is necessary to obtain a global value of the image comparison between a test and the gallery image. The simplest possible way to achieve this is by aggregating the local decisions to form a global score which we refer to as global similarity score S_g. The comparison of a test image with any arbitrary M number of gallery images results in M global similarity score S_g. Including the N perturbations done on the test image, this number increases to $M \times N$. These generated similarity scores are then ranked and the top rank is selected to represent the best match. This idea of ranking top rank is no different from threshold logic based decisions at global level (wherein threshold can be thought of being applied between the top rank and second most top rank). Overall, this process represents the global decision making process through a simple approach of global similarity calculation and selection.

3. Experimental Analysis

Unless specified otherwise, all the experiments presented in this section are conducted using the AR face database (See Fig. 2) with the following numerical values: 0.25 for θ, 160×120 pixels for the image size, and 7×5 pixels for the kernel window size of the standard deviation filter.

3.1 Effect of Spatial Intensity Change Used as Features

An analysis using spatial change features and raw features suggest that inter-pixel spatial change within an image is the essential photometric or geometric visual cue that contributes to the recognition of the objects in it. This can be observed from the results presented in Table 2.

The performance analysis using various features with and without mean normalization is shown in Table 2. The importance of spatial change as features for face recognition is analysed by comparing its performance with raw and edge features. For this comparison the standard nearest neighbour (NN) classifier (Cover, 1968; Cover & Hart, 1967; Gates, 1972; Hart, 1968) and the proposed classifier are used.
A raw face image in itself contains all the identity information required for face recognition. However, occurrence of external occlusions, expressions, and illumination in face images can result in loss of such identity information. Further, raw image intensities are highly sensitive to variations in illumination, which make recognition on raw images a difficult task. The comparison shown in Table 2 between spatial change features and raw image features clearly shows that spatial change features outperform the raw features significantly. This superior performance of spatial change features over raw features can be attributed to the facts that spatial change features (1) show lower local variability in the face images under various conditions such as expression, illumination, and occlusion, and (2) preserve the identity information of a face.
Most edge detection techniques are inaccurate approximations of image gradients. Spatial change detection techniques are different from standard edge detection techniques. Majority of the edge detection techniques result in the removal of medium to small texture variations and are distinct from spatial change detection techniques that preserve most of the texture details. Such variations however contain useful information for identification and show increased recognition performance. These observations are shown in Table 2. They further confirm the usefulness of spatial change features in face recognition and show the relative difference of spatial change features as opposed to the edge features.
Figure 3 is a graphical illustration of the overall impact of using spatial change features. The plot shows a normalized histogram of similarity scores S_g resulting from inter-class and intra-class comparisons. The 100 gallery images from the AR database described in the Section 3 form the 100 classes and are compared against 2500 test images in the AR database. The inter-class plots are obtained by comparing each of these test images with the gallery images belonging to a different class, whereas intra-class plots are obtained by the comparison of each test image against a gallery image belonging to its own class. Further, a comparison is done between spatial change features (Fig. 3a) and raw image features (Fig. 3b). The overlapping region of the two distributions indicates the maximum overall probability of error when using the proposed classifier. This region also shows the maximum overall false acceptance and false rejection that can occur in the system. A smaller area of overlap implies better recognition performance. Clearly, it can be seen that the use of feature vectors in Fig. 3a as opposed to the raw-image features in Fig. 3b results in a smaller region of overlap and hence better recognition performance.
An analysis is done to study the effect of using a spatial change filter window w of various sizes [w is described in Section (2.1)]. It can be observed from Fig. 4 that with an increase in resolution of the spatial change features (or the raw image) the recognition performance shows increased stability against variation in spatial change filter window size. Further, it can also be seen that higher resolution images show better recognition accuracies.

3.2 Normalization

The baseline algorithm contains two different types of normalization. They are: (1) global mean normalization of the feature vectors and (2) similarity measure normalization employed in the classifier. The relative importance of using these normalization methods is presented

Index	Feature Type	Recognition accuracy (%)[a]	
		NN Classifier	proposed Classifier
	With global mean normalization[a]		
	Raw features		
r1	Raw	46.0	63.8
	Spatial change features		
s1	Local Standard Deviation	67.6	84.9[b]
s2	Local Range	68.6	84.4
	Edge		
e1	Sobel edges	69.0	80.3
e2	Prewitt edges	69.2	80.4
	Without global mean normalization		
	Raw features		
r2	Raw	38.5	50.8
	Spatial change features		
s1	Local Standard Deviation	59.3	84.7
s2	Local Range	63.0	83.4
	Edge		
e1	Sobel edges	50.4	80.8
e2	Prewitt edges	49.4	80.8

[a] Global mean normalization is achieved using Eq. (3) and Eq. (4). While for raw features normalization is done by replacing $\sigma(i, j)$ with $I(i, j)$ in Eq. (3) and Eq. (4).
[b] proposed baseline algorithm with global mean normalization.

Table 2. Effect of global mean normalization and feature type

in Table 3. It is observed that normalization of the distance measures results in higher recognition accuracies. It can also be observed that global mean normalization shows improved recognition accuracy only when similarity measure normalization is used, which also shows that global mean normalization in isolation does not improve the recognition performance. In the following sections the effect of these two normalization is further studied and alternative methods are attempted. This is done to provide a better technical insight into the normalization methods. This also helps in understanding the unique features that contribute to the overall recognition performance.

3.3 Effect of Mean Normalization and Study of Alternative Normalization

From the experimental results obtained in Table 3, it is found that the normalization of spatial change features by a global mean is not robust against the recognition performance. Clearly, the feature normalization performed by Eq. (3) does not improve the performance considerably, which leads us to investigate alternative local mean normalization techniques. Equation (4) is now replaced by the following equation to calculate the local mean of spatial change

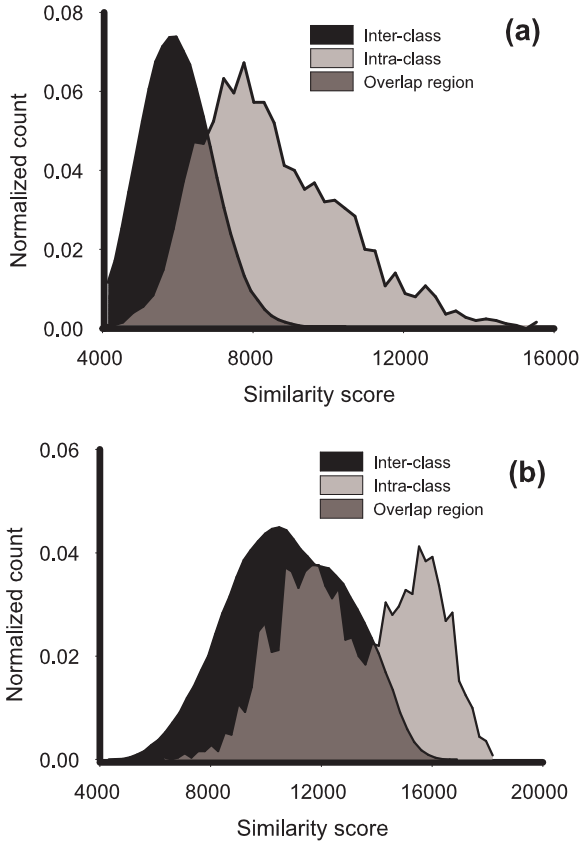

Fig. 3. Graphical illustrations showing the overall influence of using spatial change features. The graphs show a normalized frequency distribution of similarity scores S_g when using (a) spatial intensity change features (b) raw image features.

features:

$$\overline{\sigma(i,j)} = \frac{1}{kl} \sum_{s=-a1}^{a1} \sum_{t=-b1}^{b1} \sigma(i+s, j+t) \tag{7}$$

where the moving window of pixels is of size $k \times l$ pixels, $a1 = (k-1)/2$ and $b1 = (l-1)/2$. Local mean normalization is applied on spatial change features by using Eq. (7) followed by Eq. (3).

An investigation on the performance of using local mean normalization with local mean windows of different sizes is done. Figure 5 shows the effect of variation in local mean window on the recognition performance when using spatial change features and raw features. Further, the same graph shows a comparison of its performance with global mean normalization. It is observed that recognition performance increases when features are normalized using the local mean normalization described by Eq. (7) and Eq. (3). The improvement in recognition

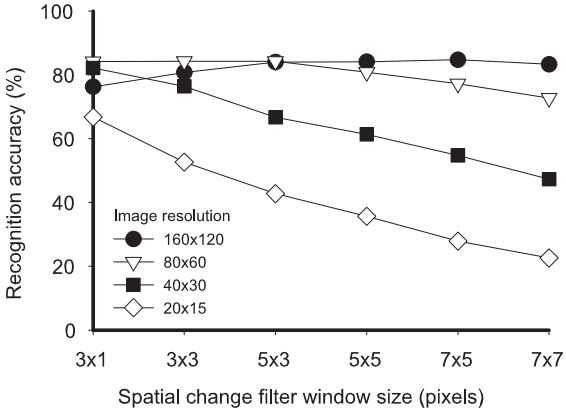

Fig. 4. A graphical illustration showing the recognition performance of the proposed algorithm under the variation of spatial change features filter window size at various image resolutions.

accuracy while using local mean normalization compared to global mean normalization is relatively large in the case of the raw features while having very little impact on spatial change features. Further, in comparison with the raw features, the spatial change features is stable for a broader range of local mean normalization filter window size. The algorithm using spatial change features provides robust performance within the local mean normalization filter window range of 80×60 pixels to 40×30 pixels as shown in Fig. 4.
Table 4 shows the effect of using local mean normalization on spatial change features. Clearly, in comparison with Table 3, the local mean normalization on spatial change features shows an increase in recognition performance when using the proposed classifier. However, the recognition performance shows no improvement when using an NN classifier. Further, Fig. 5 shows that local mean normalization improves the overall recognition performance and provides a wider stable range of threshold than when using global mean normalization [see Fig. 6 and Fig. 7]. It can be observed that in comparison with global mean normalization on similarity measure, the local mean normalization on similarity measure shows increased stability in recognition accuracy with respect to a varying threshold. All these effects make local mean normalization the preferred choice for use in a feature normalization process.

3.3.1 Effect of Similarity Measure Normalization and Study of Alternative Normalization

Normalization of the similarity measures also helps in increasing the recognition accuracy of the proposed algorithm and enables a stable threshold. This is evident from: (1) Table 3 and Table 4, showing the superiority of similarity measure normalization over mean normalization techniques and (2) Fig. 6 and Fig. 7 showing the relative importance of similarity measure normalization in stabilizing the threshold range and increasing the recognition performance. Further, the improvement of recognition performance provided by normalizing the similarity measure can be observed from Table 5. It can be observed that all of the normalized similarity measures outperform the corresponding direct similarity measures in the recogni-

	Normalization[a]		Recognition accuracy (%)	
Condition	Features	Similarity measure	NN Classifier	proposed Classifier[b]
(a)	Yes	Yes	67.6	84.9
(b)	Yes	No	67.6	76.0
(c)	No	Yes	59.3	84.7
(d)	No	No	59.3	78.4

[a] Feature extraction filter window used in Eq. (2) has a size of 7×5 pixels for a raw image I with a size of 160×120 pixels. Normalized similarity measure described using Eq. (6) is used for these simulations.

[b] The results are shown for the best accuracies by optimizing the threshold θ. The optimized values of the threshold for the condition indexes (a), (b), (c) and (d) are 0.5, 0.25, 0.35 and 0.85 respectively.

Table 3. Effect of Global Mean Normalization of Features and Similarity Measure Normalization

	Normalization[a]		Recognition accuracy (%)	
Condition	Features	Similarity measure	NN Classifier	proposed Classifier[b]
(a)	Yes	Yes	62.0	86.2
(b)	Yes	No	62.0	81.9
(c)	No	Yes	59.3	84.7
(d)	No	No	59.3	78.4

[a] Feature extraction filter window used in Eq. (2) has a size of 7×5 pixels for a raw image I with a size of 160×120 pixels. The size of local mean normalization window $w1$ used in Eq. (7) is set to 80×60 pixels. Normalized similarity measure described using Eq. (6) is used for these simulations.

[b] The results are shown for the best accuracies by optimizing the threshold θ. The optimized values of the threshold for the normalization conditions (a),(b),(c) and (d) are 0.5, 0.25, 0.35 and 0.85 respectively.

Table 4. Effect of Local Mean Normalization and Distance Normalization

tion accuracy. Fig. 8 shows the influence of variable threshold on the normalized and direct similarity measures. Clearly, for every threshold the normalized similarity measures show better recognition performance than those without similarity measure normalization. These results suggest that normalization of similarity measures is an important factor that helps in improving the recognition performance of the proposed algorithm.

3.3.2 Effect of Local Binary Decisions and Threshold

Binary decisions are made by transforming the normalized similarity measure to a binary decision vector by using a predefined global threshold. A threshold θ is used to set similar features to a value of one, whereas dissimilar features are set to a value of zero. The proposed classifier applies the binary decisions to individual pixels, which means that it can utilize the maximum available spatial change features in the image.

The importance of local binary decisions in the proposed classifier is shown in Fig. 9. The comparison of recognition performance with thresholding and without thresholding shows

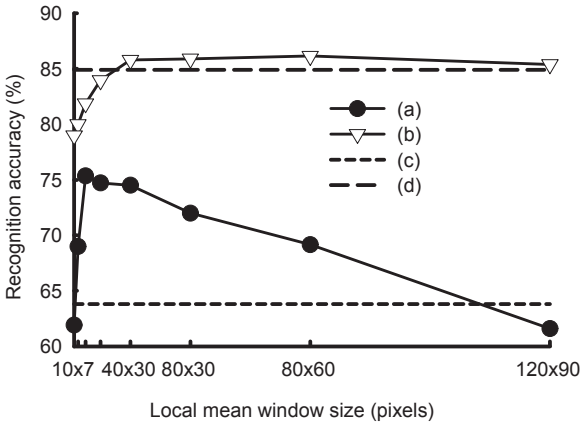

Fig. 5. Graphical illustration showing improved performance of local mean normalization compared to global mean normalization. The graph show the following conditions: **(a)** local mean normalization applied to raw features, **(b)** local mean normalization applied to spatial change features, **(c)** global mean normalization applied to raw features, and **(d)** global mean normalization applied to spatial change features. The image size is 160×120 pixels; w is of size 7×5 pixels; the local mean filter window size is varied from 10×7 pixels to 160×120 pixels; for each local mean filter window size the best recognition accuracy is selected by optimizing the threshold. Normalized similarity measure given by Eq. (6) is used for these simulations.

a very large change from 86.2% to 13.8% respectively. This shows the relative importance of local binary decisions, confirming it as the essential component of the algorithm. The local binary decisions result in the removal of noisy information associated with the natural variability. Although, it can be argued that such thresholding results in loss of information, but we find that for natural recognition problems it is the relative number of pixel information in intra-class and inter-class features that would effect the overall performance, and not the individual loss of information due threshold. For example, occlusions and facial expressions remove identity information from the face and can also add information that may seem to be relevant (false similarity) to a non-binary classifier such as the NN classifier. Without the binary decisions, the noisy information gets accumulated when forming a global similarity score (note that similarity scores are formed by adding the values of the elements in the similarity measure vector). Since the global similarity score has significant contribution of such noisy information (or false similarity), the result is a reduced recognition performance. As opposed to this, every feature is used for making local decisions in the case of the proposed classifier. In this case, the global similarity score does not accumulate the effect of less similar features, resulting in a better recognition performance.

Figure 10 shows the performance of the proposed algorithm with a change in threshold when using various normalized similarity measures. We can observe that the recognition accuracy is stable over a broad range of threshold values irrespective of the normalized similarity measures employed. The stability of the threshold and increased recognition performance can be attributed to the use of normalized similarity measures [see Fig. 8]. Further, the stability of

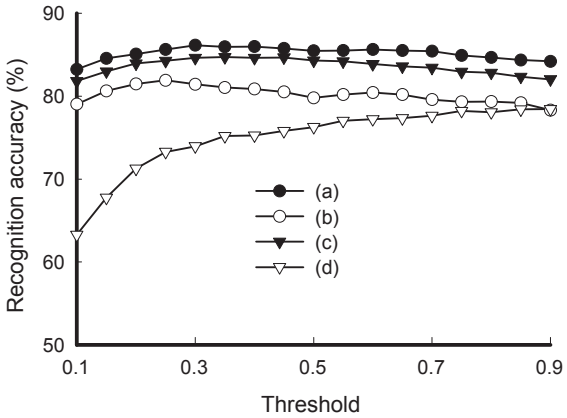

Fig. 6. Graphical illustration showing the effect of local mean normalization and similarity measure normalization on the performance of the proposed algorithm. The graph show the following conditions: **(a)** local mean normalization applied to spatial change features and with normalization similarity measure for comparison, **(b)** local mean normalization applied to spatial change features and with similarity measure without normalization for comparison, **(c)** spatial change features without normalization and with normalized similarity measure comparison, and **(d)** spatial change features without normalization and with similarity measure without normalization for comparison. Normalization of features is performed using global mean normalization of spatial change features using Eq. (4) and Eq. (3). This feature normalization is tried in combination with normalized similarity measure and the performances are compared.

the threshold enables the use of any of the possible similarity measures to form the proposed classifier. A stable threshold in turn implies that the recognition performance of the algorithm is least sensitive to threshold variation. Further, this allows for the use of a single global threshold across different databases containing images of various types of natural variability.

3.3.3 Effect of Resolution

The recognition performance with respect to variation in resolution can be studied by (1) varying the raw image resolution and (2) increasing the decision block size. In the first case, reducing the image resolution from a higher resolution will result in a smaller number of normalized spatial change features. The reduction of a higher resolution image to a lower resolution image can be achieved by averaging a block of pixels to form a single pixel. This averaging results in a loss of features and hence it is natural to expect that recognition performance will drop with lower resolution images which tends to have fewer features. We can observe from Fig. 11 that with lower resolution images the recognition performance drops considerably (this situation is labeled as *average before*).

In the second case, the resolution of spatial change features are kept to a maximum of 160×120 pixels, followed by the calculation of δ. The reduction in resolution is achieved by averaging on a block of elements in δ. Block by block reduction across the entire δ results in a lower resolution of δ. This situation is labeled as *average after* in Fig. 11. We can observe

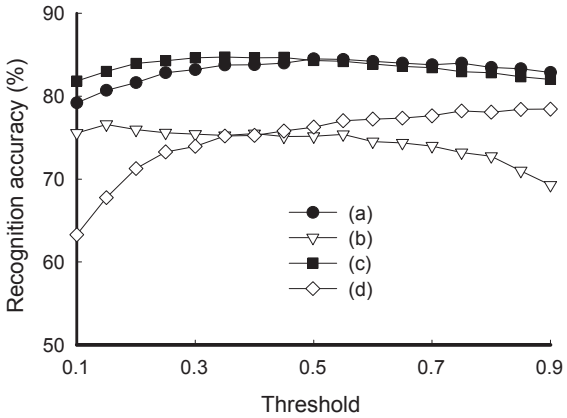

Fig. 7. Graphical illustration showing the effect of global mean normalization and similarity measure normalization on the performance of the proposed algorithm. The graph show the following conditions: **(a)** global mean normalization applied to spatial change features and with normalization similarity measure for comparison, **(b)** global mean normalization applied to spatial change features and with similarity measure without normalization for comparison, **(c)** spatial change features without normalization and with normalized similarity measure comparison, and **(d)** spatial change features without normalization and with similarity measure without normalization for comparison.Normalization of features is performed using global mean normalization of spatial change features using Eq. (4) and Eq. (3). This feature normalization is tried in combination with normalized similarity measure and the performances are compared.

from Fig. 11 that in the case of *average after*, the reduction in resolution results in a slight reduction of the recognition performance, which however, again shows that a larger number of features helps to increase the recognition performance. Further to this, Figure 11 also shows the importance of having a larger number of features irrespective of the decision block size. A larger number of features and a smaller decision block size results in increased recognition performance. Further, as observed from Fig. 4, an increased resolution of features extends the stable range of spatial change filter window size.

3.3.4 Effect of Color

Color images are formed of three channels, namely, red, green, and blue. Table 6 shows that the use of color images also helps to improve the recognition performance. Similarity scores for a comparison between a color test image and a color gallery image can be obtained by one-to-one comparison of red, green, and blue channels of one image to the other. To obtain an overall similarity score, an additive combination of the independent similarity scores observed across the red, green, and blue channels are taken. Table 6 lists some of the combinations that are used in our analysis. Table 6 further illustrates that the use of independent channels alone are not sufficient for robust performance. It can be also observed that utilizing the additive combination of similarity scores obtained from the channels of color images provides a higher recognition accuracy than when using gray images. This can be seen from the

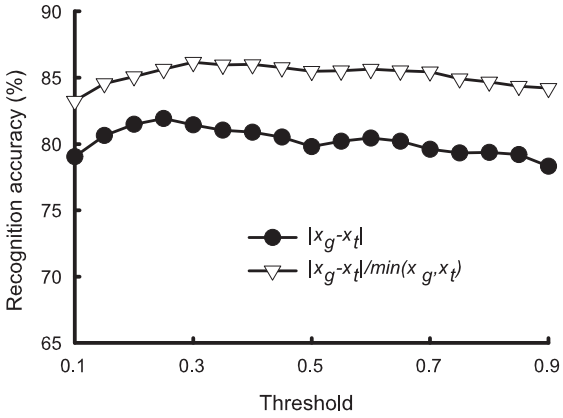

Fig. 8. Graphical illustration showing a comparison of normalized similarity measure with a direct similarity measure. The image size is 160×120 pixels; the size of w is 7×5 pixels; the size of local mean filter window $w1$ is set to 80×60 pixels.

recognition performance of the proposed algorithm when using the combination of the color channels (see c8 listed in Table 6). Although several other combinations can also be tried, analysis is limited to the extend to form a simple model for color, which is achieved through c8 listed in Table 6.

3.3.5 Effect of Localization

Automatic face detection and alignment is a difficult problem when natural variability in images is high. In any method that is based on pixel-by-pixel comparisons, it is essential that the features of the compared images are well aligned. Irrespective of the face detection method employed, natural variability can cause pixel-level misalignments. To compensate for the localization errors that occur after an automatic or manual alignment, we apply either test or gallery image shifts with respect to a set of registration points in the feature vectors. For example, the localization of face images can be achieved by detecting the location of eye coordinates. An error in localization means the eye coordinates are shifted. A scale error means that the eye coordinates are shifted towards each other or away from each other. A rotation error causes shifts of the two eye coordinates in opposite vertical directions. We pertubate the reference eye coordinates by applying such shifts and re-localize the face images using the shifted eye coordinates.

Using the above mentioned idea, two techniques that can be employed to reduce localization errors in the proposed algorithm are (a) application of modifications such as shift, rotation, and scaling on the test image, followed by comparison with gallery, and (b) perturbation of the eye-coordinates of the gallery images to form several sets of synthetic gallery images. In both cases, each comparison of a test image with a gallery image, results in a similarity score S_g^* for the baseline algorithm. The final similarity score S_g for the test image with a compared gallery image is found by selecting the maximum S_g^*. Table 7 shows the recognition performance using both techniques using color and gray scale images. For these simulations the values of number of perturbations is set to 15, composed of 5 horizontal, 5 vertical and 5

Index	Similarity measure[a]	Recognition accuracy (%)[b]
	Normalized	
n1	$\frac{min[x_g,x_t]}{max[x_g,x_t]}$	85.9
n2	$\frac{\left\|x_g-x_t\right\|}{max(x_g,x_t)}$	86.1
n3	$\frac{\left\|x_g-x_t\right\|}{min(x_g,x_t)}$	86.2
n4	$\frac{\left\|x_g-x_t\right\|}{mean(x_g,x_t)}$	86.1
n5	$e^{\frac{-\left\|x_g-x_t\right\|}{max(x_g,x_t)}}$	86.0
n6	$e^{\frac{-\left\|x_g-x_t\right\|}{min(x_g,x_t)}}$	86.1
n7	$e^{\frac{-\left\|x_g-x_t\right\|}{mean(x_g,x_t)}}$	86.1
	Direct	
d1	$\left\|x_g - x_t\right\|$	81.9
d2	$e^{-\left\|x_g-x_t\right\|}$	81.6

[a] Feature extraction filter window used in Eq. (2) has a size of 7×5 pixels for a raw image I with a size of 160×120 pixels. The size of local mean normalization window $w1$ used in Eq. (7) is set to 80×60 pixels.

[b] θ is optimised for highest accuracies on each similarity measure under consideration.

Table 5. Direct and Normalized Similarity Measures

diagonal perturbations. This performance difference is due to the fact that modification of test images is performed after cropping and results in loss of useful spatial information during comparison. This is different from the perturbation of the gallery images that preserves all the information from the original image.

4. Experimental Details

The algorithm is applied to AR (Martinez & Benavente, 1998), ORL(Samaria, 1994), YALE (Belhumeur et al., 1997), CALTECH (Lab, 1999), and FERET (Phillips et al., 2000) standard face image databases. At any specific time, illumination, occlusions, face expressions, and time gap between the gallery and test images form variabilities that make the face recognition difficult. A difficult and practically important face-recognition task is created by limiting the gallery to a single image per person. Unless otherwise specified, the results presented in this chapter are obtained by this kind of open-set testing.

For each image in the AR, YALE, and CALTECH databases, the eye coordinates of the face images are registered manually. For FERET database, the eye coordinates provided in the FERET distribution DVD is used for face alignment. The face alignment is done by rotating, shifting, and scaling the faces so that for all the faces the distance between the eyes remains constant and in fixed spatial coordinates. All the images were aligned and cropped to image size of

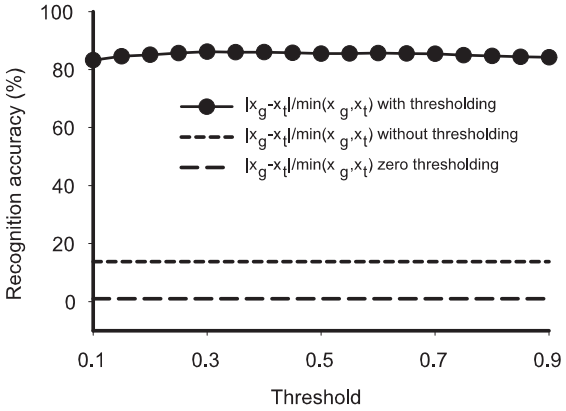

Fig. 9. Graphical illustration showing the effect of local binary decisions. “Without thresholding" is the situation when no threshold is used, which means that no local binary decisions being made. “Zero thresholding" is the situation when the threshold value is set to zero.

160×120.[1] However, as ORL images are approximately localized images, manual alignment are not done on it and are resized to 40×32 pixels.

Since the eye coordinates of the faces in AR, Yale, and Caltech databases are detected manually they show shift errors after processing. The eye coordinates of the faces in the gray FERET database are provided within the FERET distribution DVD, and when used, show rotation and scaling errors. Perturbation to eye coordinates are done to compensate for these localization errors. These modifications are in the range of 1 to 6 pixels.

Unless otherwise specified, the following global settings are used for the set of proposed parameters. To calculate spatial intensity change, the local standard deviation filter [see Eq. (1)] is used with optimal window size of 7×5 and 3×3 pixels when image size is 160×120 and 40×30 pixels respectively. The min-max similarity ratio shown in Table 1 is used. Finally, the value of the global threshold θ is set to 0.7 which is selected empirically. The number of perturbation used for compensating localization errors in every case is set to a value of 15.

5. Results and Discussion

The overall recognition accuracy for the 2500 gray scale test images and the gallery size of 100 in the AR database is 91%. This very high accuracy level is possible due to the consistent performance over the large number of variable conditions that are individually listed in Table 8. Similar accuracy levels are obtained for YALE, ORL and CALTECH databases as shown in Table 9. As expected, increased variations correspond to decreased recognition accuracies in all databases. The demonstrated robustness of the algorithm is consistent with the fact that the baseline algorithm does not require any prior knowledge of the specific condition that causes the dominant variations. To substantiate the claim of robustness, it is important to report the performance for a large gallery set. In practice, an increased gallery size decreases the overall

[1] This is done using the Unix script provided for face normalization in the CSU Face Identification Evaluation System, Version 5.0 (Beveridge et al. (2003)).

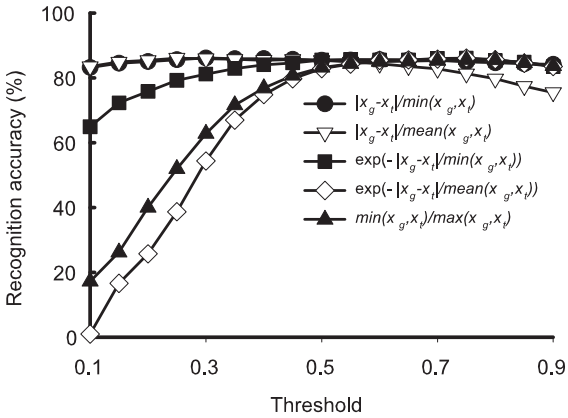

Fig. 10. Graphical illustration showing the stability of the threshold against various normalized similarity measures. The image size is 160×120 pixels, the size of the standard deviation filter is 7×5 pixels, and the value of the global threshold θ is varied from 0.1 to 0.9.

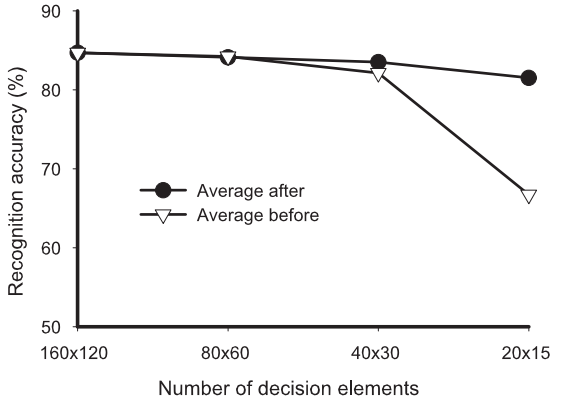

Fig. 11. Graphical illustration showing the recognition performance of the proposed algorithm with variation in resolution of the normalized similarity measure δ under comparison. Averaging is performed to reduce the resolution of δ. *Average before* shows the case when raw images at various resolutions are used, whereas *average after* shows the case when spatial change features at various resolutions are formed from a 160×120 pixels raw image.

Index[a]	Color combination	Recognition accuracy (%)		
		AR (b)-(z)	AR (b)-(m)	AR (n)-(z)
c1	Gray	86.16	94.75	78.23
c2	Red	68.86	76.29	62.00
c3	Green	86.00	95.00	77.69
c4	Blue	87.64	96.33	79.61
c5	Red+Green	81.55	90.16	73.61
c6	Blue+Green	88.96	97.00	81.54
c7	Red+Blue	85.84	95.00	77.38
c8	max(c5,c6,c7)	89.60	97.00	82.76

[a] Similarity score calculated from (c1) gray images, (c2) red channel alone, (c3) green channel alone, (c4) blue channel alone, (c5) combination of scores from red and green channels, (c6) combination of scores from blue and green channels, (c7) combination of scores from red and blue channels, and (c8) the maximum of scores obtained as a result of operations c5 to c7

Table 6. Effect of color on single training samples per person scheme

Color image	Recognition Accuracy (%)		
	Perturbation		
	No	Yes	
		Test image	Gallery image
Yes	89.6	94.0	94.8
No	86.2	91.0	92.0

Table 7. Effect of Localization Error Compensation

recognition accuracy of any face recognition system. The results of testing with the FERET database, also shown in Table 9, demonstrate that the robustness is maintained under this condition.

Using the AR database, the effects of block size used to make the local binary decisions is analyzed and the results are shown in Fig. 12. The maximum recognition accuracy is achieved when the local binary decisions are made at the level of individual pixels (block size of one pixel) with a steep drop in the recognition accuracy as the block size is increased. This directly implies that larger image resolutions could further improve the recognition accuracy.

The impact of different implementations of the similarity measure is also analyzed. Using the implementations listed in Table 1, the change observed in the recognition accuracy is within 1%. Furthermore, the global threshold θ for making the local decisions is not a sensitive parameter. It is found that the recognition accuracy remains within 1% across various databases for a range of threshold values from 0.6 to 0.8. This confirms the general applicability of localised decisions on similarity as a concept.

The impact of the spatial change as features in the baseline algorithm are studied by using raw images as the feature vectors instead of spatial change feature vectors. The recognition accu-

Test conditions	Recognition accuracy on AR database (%)	
	Localization error compensation	
	Yes[a]	No
Session 1 images		
Expression	99	98
Illumination	97	94
Eye occlusion	100	100
Eye occlusion, Illumination	95	80
Mouth occlusion	97	93
Mouth occlusion, Illumination	93	86
Neutral	99	96
Expression	86	80
Illumination	85	80
Eye occlusion	90	83
Eye occlusion, Illumination	77	62
Mouth occlusion	89	74
Mouth occlusion, Illumination	78	60
Overall accuracy	91	84

[a] Proposed algorithm depicted here uses test image perturbations of ± 5 pixels.
[b] Results not available from the literature.

Table 8. Recognition performance of the proposed algorithm (Single training sample per person problem) on gray scale images

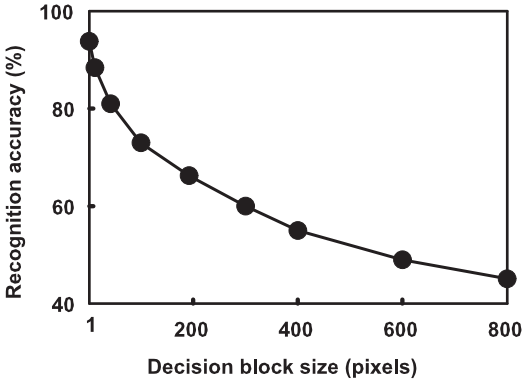

Fig. 12. The dependence of the overall recognition accuracy on the block size used to make the local binary decisions. The resolution of the images is 160×120 pixels. The window size of the standard-deviation filter is 7×5 pixels and the size of the normalization window is 80×60 pixels.

Condition index [a]	Database [b]	Top rank recognition accuracy (%)	
		Localization error compensation	
		No	Yes
(a)	CALTECH	89	95
(a)	YALE	93	95
(b)	ORL	72	84
(c)	FERET:Fb	85	96
(d)	FERET:Fc	71	90
(e)	FERET:Dup I	50	68
(f)	FERET:Dup II	40	65

[a] (a) Expression and illumination with a small gallery; (b) Small pose variation on small gallery (c) Expression on large gallery (Fb); (d) Illumination on large gallery (Fc); (e) Large gallery with mean time gap of 251 days (Dup I); (f) Large gallery with mean time gap of 627 days (Dup II).

[b] Single training image per person is used to form the gallery set. The sizes of the gallery sets are 28 in CALTECH, 15 in YALE, 40 in ORL and 1196 in FERET databases; the sizes of the test sets are 150 in the YALE database, 406 in the CALTECH database, 360 in the ORL database, 1194 in set Fb, 194 in set Fc, 722 in Dup I, and 234 in Dup II of the FERET database.

Table 9. Summary of the results on different databases

racy for the AR database dropped from 91% to 63%. Furthermore, investigation on different filters for calculating the spatial intensity changes shows that the variation of the recognition accuracy with the standard local spatial filters: standard deviation, range and gradient, is within 1%. Based on this and the clear performance difference between the use of raw images and the spatial intensity changes as the feature vectors, it is concluded that the spatial intensity change is the visual cue for face recognition.

Increased number of filters to form feature vectors can further improve the recognition accuracy. As an example, using 40 Gabor filters, the recognition performance on color images in AR database reaches around 97% from a baseline value of 91% on gray images in AR database.

6. Conclusions

In this chapter, the local binary decisions is identified an important concept that is required for recognition of faces under difficult conditions. In addition, spatial intensity changes is identified as the visual cue for face recognition. A baseline algorithm, formed by implementing the local binary decisions based classifier and the spatial intensity changes based feature extractor, shows a robust performance under difficult testing conditions. To increase the recognition performance, a baseline system is formed by including perturbation scheme for localization error compensation. Using this baseline system the effect of localization errors is analysed. Further, the analysis shows that the application of the principles of local binary decisions and modularity results in a highly accurate face recognition system. The presented algorithm does not use any known configurational information from the face images, which makes it applicable to any visual pattern classification and recognition problem. Furthermore, classifiers based on the local binary decisions on similarity can be used in other pattern recognition applications.

7. References

Ahlberg, J. & Dornaika, F. (2004). *Handbook of Face Recognition*, Springer Berlin / Heidelberg.

Belhumeur, P. N., Hespanha, J. P. & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Machine Intell.* **19**(7): 711–720. Special Issue on Face Recognition.

Beveridge, R., Bolme, D., Teixeira, M. & Draper, B. (2003). The csu face identification evaluation system users guide:version 5.0. Available from http://www.cs.colostate.edu/evalfacerec/algorithms5.html.

Cover, T. M. (1968). Estimation by the nearest-neighbor rule, *IEEE Transactions on Information Theory* **14**(1): 50–55.

Cover, T. M. & Hart, P. E. (1967). Nearest neighbor pattern classification, *IEEE Transactions on Information Theory* **13**(1): 21–27.

Delac, K. & Grgic, M. (2007). *Face Recognition*, I-Tech Education and Publishing, Vienna, Austria.

Etemad, K. & Chellappa, R. (1997). Discriminant analysis for recognition of human face images, *J. Opt. Soc. Am. A* **14**: 1724–1733.

Gates, G. W. (1972). The reduced nearest neighbor rule, *IEEE Transactions on Information Theory* **18**(5): 431–433.

Hallinan, P. W., Gordon, G., Yuille, A. L., Giblin, P. & Mumford, D. (1999). *Two- and Three-Dimensional Patterns of the Face*, AK Peters,Ltd.

Hart, P. E. (1968). The condensed nearest neighbor rule, *IEEE Transactions on Information Theory* **14**(5): 515–516.

James, A. P. (2008). *A memory based face recognition method*, PhD thesis, Griffith University.

James, A. P. & Dimitrijev, S. (2008). Face recognition using local binary decisions, *IEEE Signal Processing Letters* **15**: 821–824.

Jenkins, R. & Burton, A. M. (2008). 100% accuracy in automatic face recognition, *Science* **319**: 435.

Lab, C. V. (1999). Caltech face database. Available from http://www.vision.caltech.edu/html-files/archive.html.

Li, S. Z. & Jain, A. K. (2005). *Handbook of Face Recognition*, Springer Berlin / Heidelberg.

Marr, D. (1982). *Vision*, New York: W.H. Freeman and Company.

Martinez, A. M. & Benavente, R. (1998). The ar face database, CVC Technical Report 24.

Martinez, A. M. & Benavente, R. (2000). Ar face database. Available from http://rvl.www.ecn.purdue.edu/RVL/database.htm.

Moeller, S., Freiwald, W. A. & Tsao, D. Y. (2008). Patches with links: A unified system for processing faces in the macaque temporal lobe, *Science* **320**: 1355–1359.

Phillips, P. J., Moon, H., Rauss, P. J. & Rizvi, S. (2000). The feret evaluation methodology for face recognition algorithms, *IEEE Trans. Pattern Anal. Machine Intell.* **22**: 1090–1104.

Samaria, F. S. (1994). *Face Recognition Using Hidden Markov Models*, University of Cambridge.

Wechsler, H. (2006). *Reliable Face Recognition Methods*, Springer Berlin / Heidelberg.

Zhang, W., Shan, S., Gao, W., Chen, X. & Zhang, H. (2005). Local gabor binary pattern histogram sequence (lgbphs):a novel non-statistical model for face representation and recognition, *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCVŠ05)*.

Zhao, W. & Chellappa, R. (2005). *Face Processing : Advanced Modeling and Methods*, ACADEMIC PRESS.