Optimized multimodal biometric system based fusion technique for human identification

Muthana H. Hamd, Rabab A. Rasool
Department of Computer Engineering, University of Al Mustansiriyah, Iraq

ABSTRACT

This paper presents three novelty aspects in developing biometric system-based face recognition software for human identification applications. First, the computations cost is greatly reduced by eliminating the feature extraction phase and considering only the detected face features from the phase congruency. Secondly, a motivation towards applying a new technique, named mean-based training (MBT) is applied urgently to overcome the matching delay caused by the long feature vector. The last novelty aspect is utilizing the one-to-one mapping relationship for fusing the edge-to-angle unimodal classification results into a multimodal system using the logical-OR rule. Despite some dataset difficulties like Unconstrained Facial Images (UFI) which includes varying illuminations, expressions, occlusions, and poses, the multimodal system has highly improved the accuracy rate and achieved a promising recognition result, where the decision fusion is classified correctly (84, 92, and 72%) with only one training vector per MBT in contrast to (80, 62, and 68%) with five training vectors for Normal matching. These results are measured by Euclidean, Manhattan, and Cosine distance measure respectively.

1. INTRODUCTION

Phase congruency (PC) is a contrast and illumination invariant measure of valuable features. Unlike gradient-based feature detectors, which can only detect the first or second derivative features (also called step features), phase congruency correctly detects all kind of phase angle related features, and not just step features having 0 or 180 phase angle degrees. Although the complexity in calculating the PC features, it is considered an immune method against noise, contrast variations, and illumination. Also, the amount of redundant data is effectively reduced after extraction stage, which is an important process for building features that: speed-up learning process, phase generalization, and time saving. Generally, it is the step that classification process comes after. One of the biggest challenges in this work is the removing of the feature extraction stage for reducing the computations complexity and considering only the detected PC features. These features (edge-angle) are considered here for constructing separately two face recognition unimodal systems: edge-based modal and angle-based modal. Then the recognition accuracies for both modals are computed for different databases, training methods, and classifiers to determine how much they would be improved after applying one-to-one or edge-to-angle mapping relationship and decision fusion technique. Next sections demonstrate three different approaches in finding PC formulations in term of 1-D, Gabor Wavelets, and 2-D Laplass of Gaussian and Gabor.

Journal homepage: http://beei.org
2. RELATED WORKS

P. Kovesi was the most interested in the PC field since 1999, a calculation for 1-D signal was extended to 2-D images using wavelet high-pass filters to obtain image information at different scales where a universal threshold value over wide class of images was applied [1]. The advantages of using PC over Canny edge detectors was explained and approved using standard images [2]. PC-based face recognition technique for improving the recognition rates of the faces that are affected due to varying illuminations, partial occlusions, and varying expressions was proposed and implemented [3]. A descriptor based on the PC concept, called histogram of oriented phase (HOP) was applied to depict and represent the human objects more efficiently than the gradient based approach [4]. A modular kernel eigen spaces approach was implemented on the PC images to localize nonlinear feature selection procedure to help overcoming the bottlenecks of illumination variations, partial occlusions, and expression variations [5]. A 2-D multi-scale phase congruency (2D-MSPC) software for detecting and evaluation of image features could appropriately tune many parameters for optimal image features detection, these parameters are optimized for maximum and minimum moments [6]. A modified algorithm of PC to locate image features using Hilbert transform was implemented and the local energy was obtained by convoluting original image with two operators of removing direct current (DC) component over current window and 2-D Hilbert transform respectively. The local energy was divided with the sum of Fourier amplitude of current window to retrieve the value of PC [7]. A quality metric for evaluation of different video fusion methods using set of predefined 3D Gabor filters was employed to compute the spatial temporal PC for input and fused videos [8].

The local binary pattern (LBP) on the ultra sound (US) medical images after computing their PCs was used to improve the reliable feature point localization [9]. The multi-focus image fusion applied new fusion rule and complex Gabor wavelet to obtain the benefits of PC sharpness in finding new focus measure [10]. Two different multi-spectral image fusion rules for nonsubsampled contourlet transform (NSCT) was introduced with: PC, principle component analysis (PCA), directive contrast, and entropy for developing integrity model [11]. A fusion of the bio-medical images and guided filter was implemented by decomposing the bio-medical images into two sub-bands using the non-subsampled contourlet transform. The guided filter was very effective in extracted the high frequency details for obtaining the final fused image in spatial domain [12]. The edges in the liver image was enhanced by improving the segmentation accuracy. The sum modified Laplacian (SML), contrast features measure, and non-subsampled shearlet transform (NSST) were integrated to reconstruct a model for better fusion of edge information [13]. Pearson correlation coefficient and Spearman rank order correlation coefficient [14] were used to evaluate the performance of the proposed metric. The proposal developed a novel image quality and metric as a set of ten extracted features from each distorted image, then relevance vector machine algorithm (RVM) was used to learn the mapping between the human opinion scores and combined features. A novel descriptor using multi-oriented PC and magnitude information was presented to show the advantages of the histograms of oriented magnitude (HOM) and Histograms of oriented phase congruency (HPC) combination, also it confirmed that the histograms of oriented magnitude and phase congruency (HOMPC) was so superior to the local feature in the state-of-the-art descriptors. Three datasets were composed of multi-sensor remote sensing images obtained from satellite platforms, unmanned aerial vehicle, and unmanned ground vehicle [15]. An immune model of PC against illumination and contrast variation was developed. An obtained result from variety of remote sensing images showed an excellent performance against the local invariant and robust of the model to both geometric and radiometric changes [16]. The negatively impact was avoided using the same fusion rule on different scales, also a Gaussian filler with multi-scale decomposition (MSD) of total variation (TV) and PC were proposed for designing the modal [17].

3. PHASE CONGRUENCY MEASURES

Three different methods are presented in the following sections to determine a new operator that employs the phase congruency principle moments for computing the edge and angle information which is invariant to the image contrast, noise, and illumination. [18-20]:

3.1. Fourier transform

The fourier series is expanded at some location point, \( x \) to define the one-dimension phase congruency function. Features are detected by founding the Fourier components that has maximum phase, as in (1) [1, 2].

\[
PC(x) = \max_{\theta \in [0, 2\pi]} \left( \sum \left| F_p \right| \cos(\theta(x) - \bar{\theta}(x)) \right)
\]

\[
\sum \left| F_p \right|
\]

(1)
Where: $\phi_u(x)$ and $\bar{\phi}_u(x)$ are the local and mean phase angles of the frequency component $F_{p_u}$ at $x$ respectively. The aim is to maximize Equation (1) by maximizing the amplitude weighted mean for local phase angle for all considered Fourier points of $\bar{\phi}_u(x)$. Hereby, phase congruency is a rather difficult quantity to be computed, as finding where phase congruency is a maximum is approximately equivalent finding where the weighted variance of local phase angles, relative to the weighted average local phase, is a minimum [1, 2].

3.2. Wavelet transform

The phase congruency in (1) is sensitive to the noise and it is not well localized because the measure changes with the difference in phase not in the small responses or magnitude, so for small angle value $(\converges to zero)$, the explains how phase difference effects the weighted magnitudes. So, an alternative approach is needed to find a maximum local energy as phase congruency is directly proportional to it [2]. The phase congruency performance is improved with presence of noise. The convolution of a signal, $I(x)$ with a set of cosine and sine Wavelets at scale $m$ will deliver the even and odd components at the $m^{th}$ scale. The amplitude of this transformation is the local energy as explained in (2, 3, 4, 5, 6 and 7) [21, 22]. The phase congruency measurement is defined in (8).

\[
\begin{align*}
[e_m(x), o_m(x)] &= [I(x) * M_{m}^e, I(x) * M_{m}^o] \\
A_m(x) &= \sqrt{e_m(x)^2 + o_m(x)^2}
\end{align*}
\]

where: $I(x)$: is the one-dimension signal, $M_{m}^e, M_{m}^o$: are the cosine and sine Wavelets at the $m^{th}$ scale respectively, $e_m(x), o_m(x)$: are the even and odd components, $A_m(x)$: is the local energy or the amplitude of transformed results, so the total energy $A$ is:

\[
\sum_m A_m(x) \approx \sum_m \sqrt{e_m(x)^2 + o_m(x)^2}
\]

Let, $F(x) \approx \sum_m e_m(x)$

and $H(x) \approx \sum_m o_m(x)$

Where $E(x) = \sqrt{F(x)^2 + H(x)^2}$

\[
PC(x) = \frac{E(x)}{\sum_m A_m(x) + \varepsilon}
\]

$\varepsilon$: is a small value between 0 and 1

3.2. Two-dimensional log-Gabor

The extension of frequency domain considerations to 2D image was made up by convolving set of frequency domain constructed filters with an image. So, a construction of Fourier domain filter like log-Gabor function is very suitable as it has complementary spreading functions and singularity at DC frequency. Starting from the low-pass Gaussian filter $g$ with $K$ different orientations has a transfer function defined in (9) [18, 19, 21, 23, 24].

\[
g(\theta_0, \theta) = \frac{1}{\sqrt{2\pi\sigma_s}} e^{-\frac{(\theta_0 - \theta)^2}{2\sigma_s^2}}
\]

where,

$\theta_0, \theta$: are the orientation and angle of the local orientation respectively

$\sigma_s$: controls the spreading around the orientation $\theta_0$

*Optimized multimodal biometric system based fusion technique for human identification (Muthana H. Hamd)*
The Laplace of Gaussian log and Gabor band-pass filter \( g \) have spreading function \( I_g(\omega, \omega_m) \) with \( M \) different scales as defined in (10).

\[
I_g(\omega, \omega_m) = \begin{cases} 
0 & \omega = 0 \\
-\frac{(\log(\omega/\omega_m))^2}{2(\log(\beta))^2} e^{-\frac{1}{2\pi\sigma_s^2} \omega^2} & \omega \neq 0
\end{cases}
\]  

(10)

where, \( \omega, \omega_m \) are the scale and the center frequency at this scale respectively, \( \beta \) controls bandwidth at this scale, \( \beta = k/\omega_m \), the functions combination will result a 2D filter \( I_{2D} \) which works at different scales and orientations as in (11).

\[
l_{2D}g(\omega, \omega_m, \theta_0, \theta) = g(\theta_0, \theta) \times I_g(\omega, \omega_m)
\]  

(11)

so, the convolution of this filter with image \( I(x, y) \) delivers the phase congruency measurement at \( M \) scales after inverting the fourier transformation,

\[
E(m)_{x,y} = z^{-1}(I_{2D}g(\omega, \omega_m, \theta_0, \theta))_{x,y} * I_{x,y}
\]  

(12)

\[
PC_{x,y} = \frac{\sum_{m=1}^{M}|E(m)_{x,y}|}{\sum_{m=1}^{M}|E(m)_{x,y}| + \varepsilon}
\]  

(13)

4. UNIMODAL SYSTEMS DESIGN

Two unimodal systems based face trait are designed, the first one is edge-based modal and the second one is angel-based modal. Figure 1 is an example of PC outputs applied on four dataset images with their edge-gradient and angle-gradient features, where the source code in [25] is applied with its experimental optimum parameters to compute the PC feature matrices. The tuned parameters include: no. of wavelet scales=4, no. of filter orientations=6, min Wavelength scale filter=3, ratio of the standard deviation of the Gaussian=0.55, filter orientations to the Gaussian standard deviation ratio =1.2, standard deviations of noise threshold point=2.0. The two modals experience serious challenge in the matching time causes by cancelling the feature extraction phase. Therefore, a new modification toward determining the mean vector for each class in the training datasets is turned on as explained in the next section.

![Figure 1. Phase congruency feature detection](image)
5. MEAN-BASED TRAINING

The MBT vector can be calculated directly during system running as explained in (14). For purpose of comparison and performance evaluation, matching stage is implemented in both modes: normal matching and MBT matching. Hereby, each unimodal system will output forty-eight classification results comes from: 4 databases × 2 unimodals (edge and angle) × 2 training-modes (MBT and normal) × 3 classifiers, as detailed in Figure 2. The recognition rate is calculated at minimum equal error rate (EER) value that resultant from intersection of false acceptance rate (FAR) and false rejection rate (FRR) curves. The four databases specifications are tabulated in Table 1 for farther explanation.

\[
MBT = \frac{\sum_{i=0}^{n} \vec{v}_i}{n}
\]  

(14)

where, \( \vec{v}_i \) : is a machine column vector, and \( n \) is the number of training vector in a class.

<table>
<thead>
<tr>
<th>Database</th>
<th>Subjects</th>
<th>Test image</th>
<th>Training images</th>
<th>Total samples</th>
<th>Image dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>50</td>
<td>1</td>
<td>5</td>
<td>250</td>
<td>120 x 165</td>
</tr>
<tr>
<td>ORL</td>
<td>40</td>
<td>1</td>
<td>8</td>
<td>320</td>
<td>112 x 92</td>
</tr>
<tr>
<td>VIS-CH</td>
<td>40</td>
<td>1</td>
<td>10</td>
<td>400</td>
<td>225 x 321</td>
</tr>
<tr>
<td>UFI</td>
<td>50</td>
<td>1</td>
<td>5</td>
<td>250</td>
<td>128 x 128</td>
</tr>
</tbody>
</table>

6. DECISION FUSION

The edge-angle classification results of 180 subjects are fused together using the logical OR-gate as explained in Figure 2. This step is greatly improved the overall system performance as the experimental results approved that in Table 2 to Table 7.

![Figure 2. Edge-angle decision fusion construction using two training modes and three classifiers](image)

**Table 1. Face databases descriptions**

**Table 2. MBT accuracy using edge, angle, and fusion methods with Euclidean classifier**

<table>
<thead>
<tr>
<th>Database</th>
<th>Maximum Accuracy%</th>
<th>Euclidean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Edge</td>
<td>Angle</td>
</tr>
<tr>
<td>AR</td>
<td>79.6</td>
<td>96.2</td>
</tr>
<tr>
<td>ORL</td>
<td>80.9</td>
<td>81.1</td>
</tr>
<tr>
<td>VIS-TH</td>
<td>84.6</td>
<td>83.3</td>
</tr>
<tr>
<td>UFI</td>
<td>57.6</td>
<td>53.8</td>
</tr>
</tbody>
</table>

*Optimized multimodal biometric system based fusion technique for human identification (Muthana H. Hamd)*
Table 3. MBT accuracy using edge, angle, and fusion methods with Manhattan classifier

<table>
<thead>
<tr>
<th>Database</th>
<th>Maximum Accuracy%</th>
<th>Manhattan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Edge</td>
<td>Angle</td>
</tr>
<tr>
<td>AR</td>
<td>82.7</td>
<td>92.8</td>
</tr>
<tr>
<td>ORL</td>
<td>81.5</td>
<td>82.6</td>
</tr>
<tr>
<td>VIS-TH</td>
<td>85</td>
<td>80.9</td>
</tr>
<tr>
<td>UFI</td>
<td>57.7</td>
<td>55.8</td>
</tr>
</tbody>
</table>

Table 4. MBT accuracy using edge, angle, and fusion methods with Cosine classifier

<table>
<thead>
<tr>
<th>Database</th>
<th>Maximum Accuracy%</th>
<th>Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Edge</td>
<td>Angle</td>
</tr>
<tr>
<td>AR</td>
<td>94.6</td>
<td>91.5</td>
</tr>
<tr>
<td>ORL</td>
<td>89.9</td>
<td>82.1</td>
</tr>
<tr>
<td>VIS-TH</td>
<td>97</td>
<td>76.2</td>
</tr>
<tr>
<td>UFI</td>
<td>59.4</td>
<td>57.4</td>
</tr>
</tbody>
</table>

Table 5. Normal accuracy using edge, angle, and fusion methods with Euclidean classifier

<table>
<thead>
<tr>
<th>Database</th>
<th>Maximum Accuracy%</th>
<th>Euclidean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Edge</td>
<td>Angle</td>
</tr>
<tr>
<td>AR</td>
<td>83.8</td>
<td>88.9</td>
</tr>
<tr>
<td>ORL</td>
<td>85.5</td>
<td>78.4</td>
</tr>
<tr>
<td>VIS-TH</td>
<td>84.3</td>
<td>86.2</td>
</tr>
<tr>
<td>UFI</td>
<td>57.2</td>
<td>58.7</td>
</tr>
</tbody>
</table>

Table 6. Normal accuracy using edge, angle, and fusion methods with Manhattan classifier

<table>
<thead>
<tr>
<th>Database</th>
<th>Maximum Accuracy%</th>
<th>Manhattan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Edge</td>
<td>Angle</td>
</tr>
<tr>
<td>AR</td>
<td>88.3</td>
<td>96</td>
</tr>
<tr>
<td>ORL</td>
<td>85.8</td>
<td>91.8</td>
</tr>
<tr>
<td>VIS-TH</td>
<td>87.6</td>
<td>93.6</td>
</tr>
<tr>
<td>UFI</td>
<td>59.9</td>
<td>59.4</td>
</tr>
</tbody>
</table>

Table 7. Normal accuracy using edge, angle, and fusion methods with Cosine classifier

<table>
<thead>
<tr>
<th>Database</th>
<th>Maximum Accuracy%</th>
<th>Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Edge</td>
<td>Angle</td>
</tr>
<tr>
<td>AR</td>
<td>95.5</td>
<td>94.8</td>
</tr>
<tr>
<td>ORL</td>
<td>89.3</td>
<td>90.1</td>
</tr>
<tr>
<td>VIS-TH</td>
<td>94.7</td>
<td>90</td>
</tr>
<tr>
<td>UFI</td>
<td>56.4</td>
<td>59.3</td>
</tr>
</tbody>
</table>

7. RESULTS AND DISCUSSION

The 180 individuals are classified into genuine or imposter according to many considerations such as: edge or angle feature, MBT or Normal training, and type of classifier. Generally, the maximum performance results in Table 2 to Table 7 can be described and categorized into unimodal-based MBT performance, unimodal-based Normal training performance, and decision fusion performance as follows:

7.1. MBT training performance

The Euclidean classifier based unimodal system satisfied 96.2% maximum accuracy while the decision fusion improved this rate to 98%. The increment percentage is 1.8. The Manhattan classifier based unimodal system satisfied 92.8% maximum accuracy while the decision fusion improved this rate to 96%. The increment percentage is 1.2. The Cosine classifier based unimodal system satisfied 97% maximum accuracy while the decision fusion improved this rate to 98%. The increment percentage is 1.

7.2. Normal training performance

The Euclidean classifier based unimodal system satisfied 88.9% maximum accuracy while the decision fusion improved this rate to 98%. The increment percentage is 9.1. The Manhattan classifier based unimodal system satisfied 96% maximum accuracy while the decision fusion improved this rate to 98%. The increment percentage is 2. The Cosine classifier based unimodal system satisfied 95.5% maximum accuracy while the decision fusion improved this rate to 100%. The increment percentage is 4.5.
Hereby, the MBT training mode has achieved better performance than Normal training mode when the Euclidean and Cosine measurements is applied. Furthermore, the decision fusion technique could improve the unimodal accuracies for all cases and from Table 7, it achieved 100% accuracy rate with increment percentage of 4.5 using: AR dataset, Normal training mode, and Cosine classification. In contrast to the high accuracy rates of AR dataset as explained in the Figures 3 and 4, the UFI dataset contains worst facial images that have different face representation and poses. From Table 3, the decision fusion technique achieved 92% accuracy rate using Manhattan classifier in contrast to 57.7% and 55.8% for edge and angle unimodal systems respectively. This means that, the decision fusion has improved the increment percentage by 34.3 according to the difference between the two maximum accuracies 92% and 57.7%.

Figure 3. Maximum accuracy of multimodal system using MBT training

Figure 4 Maximum accuracy of multimodal system using normal training

8. CONCLUSION

This paper presented three new concepts in development of human identification system, based on the face recognition approaches. First, the regular feature extraction methods had not been employed within the design, therefore, the phase congruency was the only used approach as a facial feature’s detection. This step had an advantage in reducing the biometric system complexity and computation costs. Secondly, the large feature vector size that resulted from the phase congruency detection had now been appropriately resolved with the use of mean-based training technique. The recognition rates for this technique had advanced the normal training in most cases. At last, this work was the first to fuse the decision results for the two separable face features with edge-to-angle relationship. The decision fusion outputs highly improved the overall system performance, even for a low quality dataset such as the UFI. The multimodal system was able to accurately recognize 92% in contrast to 57.7% for a unimodal system, using mean-based training and Manhattan classifier. The new vision of those approved concepts can be a guidance for new developers in their future biometric system plans.

ACKNOWLEDGEMENTS

Our sincere thanks and gratitude to Al Mustansiriyah University for its support and encouragements.

REFERENCES


BIOGRAPHIES OF AUTHORS

Muthana H. Hamd is post-doctorate from Wollongong University, Australia 2007. He took Ph.D from Baghdad University in 2004. His research interest area is digital image processing, machine learning, deep learning, computer vision.

Rabab A. Rasool is lecturer in Computer Engineering Department, Al-Mustansiriya University, Baghdad, Iraq. M.Sc from Al Nahrain University 2012. Her research interest area is network engineering and internet, computer vision, digital image processing, feature extraction, deep learning.