Pattern Recognition ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Recognizing faces with normalized local Gabor features and Spiking Neuron Patterns

Fadhlan Kamaruzaman [a,1], Amir Akramin Shafie [b,*]

[a] Department of Computer Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[b] Department of Mechatronics Engineering, International Islamic University Malaysia, P.O. Box 10, 50728 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 14 February 2015
Received in revised form
15 August 2015
Accepted 25 November 2015

Keywords:
Gabor Wavelets
Feature representation
Face recognition
Spiking neurons
Dimensionality reduction

ABSTRACT

Gabor Wavelets (GW) have been extensively used for facial feature representation due to its inherent multi-resolution and multi-orientation characteristics. In this work we extend the work on Local Gabor Feature Vector (LGFV) and propose a new face recognition method called LGFV//LN//SNP, which employs local normalization filter in pre-processing stage. We propose a novel Spiking Neuron Patterns (SNP) as a dimensionality reduction method to reduce the dimensions of local Gabor features. SNP is acquired from projection of LGFV//LN features using Spike Response Model (SRM), a neuron model describing the spike behavior of a biological neuron. Results on AR, FERET, Yale B and FRGC 2.0 face datasets showed that SNP implementation delivered significant improvement in accuracy. Comparisons with several previously published results also suggested that LGFV//LN//SNP achieved better results in some tests. Additionally, LGFV//LN//SNP requires relatively smaller number of GW than LGFV//LN to produce optimal results.

1. Introduction

There are several ways to recognize a person from another person. Face, fingerprint, DNA, gait and iris are among biometrics properties that are widely used for person recognition. However, face recognition is the leading approach due the non-expensive implementation and non-obtrusive nature of the image acquisition which is possible without active subject participation [1,2]. Furthermore, the accuracy of face recognition in ideally controlled settings is equivalent to fingerprints and iris recognition [3]. However, in unconstrained environment, several factors such as illuminations, noise, variation in poses, facial expressions, occlusions and disguises were identified previously as contributing factors to degradation of face recognition performance [4,5].

In order to improve the robustness of face recognition against the aforementioned factors, special attention need to be given towards one of the most critical components of face recognition that is feature representations. These representations of internal structure of the face image are the key to a successful face recognition system [6], and are vital to ensure a computationally feasible and robust face processing. A good face representation according to [7] should be easy to compute, possess good separation of intraclass and interclass variations while maintaining robustness against illumination and other noises and factors.

There are mainly two types of approaches of representing facial features, namely the geometric-based and appearance-based methods. Geometric-based method usually defines the feature representation by the geometric facial features representing the spatial and configural information of facial components. Appearance-based method on the other hand relies on a set of feature vectors representing the face such as output from image filters or based on simply image intensities. One of the earliest appearance-based method, the Eigenface method, uses image of known individuals to find the Eigenface through Principal Component Analysis (PCA) [8]. Tan et al. [9] proposed a local SOM approach called SOM-face method while Wright et al. [10] used Sparse representation-based methods (SRC) to reconstruct the test images by linear combination of training images. Later, Qiao et al. [11] introduced Sparsity Preserving Discriminant Analysis (SPDA) which is based on graph-based semi-supervised dimensionality reduction approach. Jiwen et al. [12] proposed Discriminative Multimanifold Analysis (DMMA) that treats the face recognition problem as manifold matching and maximizes the manifold margins of dissimilar persons by learning multiple DMMA feature spaces. Recently, Yin et al. [13] proposed a method called Double Linear Regression (DLR) whose objectives are to find best discriminating subspace and preserve the sparse representation structure. Some other popular appearance-based methods are Local Binary Patterns (LBP) [7], Uniform Pursuit (UP) [14], and Partial Distance Measure (PDM) [15].

* Corresponding author. Tel.: +60 361966320; fax: +60 361964049.
E-mail addresses: fadhlan@salam.uitm.edu.my (F. Kamaruzaman), aashafie@iium.edu.my (A.A. Shafie).
[1] Tel.: +60 355435090.

http://dx.doi.org/10.1016/j.patcog.2015.11.020
0031-3203/© 2015 Published by Elsevier Ltd.

Nevertheless, Gabor Wavelet (GW) [16–18] has been one of the prominent and successful appearance-based feature representation whose effectiveness is attributed to its biological relevance [6,16–19]. GW uses kernels similar to receptive field on cortical cells with inherent spatial locality and is orientation selective, thus optimally localized in both space and frequency domains. Some popular implementations based on GW are Elastic Bunch Graph Matching (EBGM) [20], Gabor Fisher Classifier (GFC) [21], Local Gabor Binary Pattern Histogram Sequence (LGBPHS) [22], and Histogram of Gabor Phase Patterns (HGPP) [23]. Su et al. [24] proposed weighted fusion of Local Gabor Feature Vector (LGFV) and global Fourier transform called Hierarchical Ensemble Classifier (HEC). Besides, Jie et al. [25] proposed Local Matching Gabor (LMG) where ensembles of Borda count classifier were used to classify the Gabor features independently. Later, several improvements to LMG have been proposed over the years including several works by Perez et al. [19,26,27]. Moreover, recent improvements to LMG namely LMGEW//LN has been reported in [28] where the authors improved LMG using entropy-like weighting (EW) strategy and Local Normalization (LN) approach [28]. Besides the EW strategy, various further improvements to LMGEW//LN have also been made in [28]. This includes combination with other previously published methods, such as the use of Borda count Threshold (BTH) called LMGEW//LN-BTH, fusion at score level with variants of Gabor features called LGXP [29] and LGBP [29], denoted as LMGEW//LN+LGXP, and LMGEW//LN+LGBP respectively. These combined methods managed to produce state-of-the-art results surpassing most previously published methods.

In terms of implementation strategy, GW can be applied on either whole face or specific local region on face to extract the desired features. It is vital to remark that local approach is well-known to be more effective against localized variations such as expression variations whilst holistic approach is more suitable against variations that act globally on the image such as pose variations. For the local approach, GW is used to extract Gabor features at specific position or local patches (LP) of the face image. Then, a group of feature vectors called LGFV are formed by combining the spatially grouped Gabor features [24]. Similar to implementation in [28], LGFV representation in this work is enhanced into LGFV//LN features by applying LN filter on the images to compensate the varying illuminations that exist. However, one major problem faced from using Gabor features is the curse of dimensionality due to the high dimensionality of the resulting feature representations. Several ways to overcome this problem include applying feature selection to select most prominent Gabor features or Gabor jets and dimensionality reduction by projecting Gabor features onto smaller subspaces. Since feature selection method such as LDA requires multiple training samples to work, dimensionality reduction such as PCA, ICA, SOM or LLE seems to be a much practical solution to the problem. However, the dependency of these methods on the statistical distribution of the data or the manifold embedding of the neighborhoods requires recomputation of base vectors each time new face samples are added to the gallery. To overcome this, the dimensionality reduction should be non-statistical and relies only on the appearance of original features.

Thus, in this work, we propose novel Spiking Neuron Pattern (SNP) as a non-statistical dimensionality reduction method that relies only on the appearance of the original data to reduce the dimension of local Gabor features. SNP is computed based on Spike Response Model (SRM) [30] used to model the spike timing in biological neurons. We then used the ensembles of kNN classifier [9] that exploits the spatially grouped projections for face identification. Using AR, FERET and Extended Yale B databases, we validate the result and compare the proposed method with previously published results.

2. Related works

2.1. Gabor Wavelets

The Gabor kernel $\psi_{u,v}$ used to compute the wavelets in this work can be found using (1):

$$\psi_{u,v} = \frac{|k_{u,v}|^2}{\sigma^2} \exp\left(-\frac{|k_{u,v}|^2 |z|^2}{2\sigma^2}\right)\left[\exp(ik_{u,v}\cdot z) - \exp\left(-\frac{\sigma^2}{2}\right)\right]$$
$$z=(x,y)^T, \quad k_{u,v} = \frac{k_{max}}{f^v}\left[\cos\left(\pi\frac{u}{8}\right),\ \sin\left(\pi\frac{u}{8}\right)\right]^T \tag{1}$$

where z is the pixel, u is the orientation, $v$is the scale, $f$is the step in frequency, and k_{max} is the maximum frequency. As in [24,25,28] we use 8 orientations ($0 \le u \le 7$) and 5 scales ($0 \le v \le 4$) resulting into 40 GW of different scales and orientations. A Gabor feature $G_{u,v}$ is product of convolution between image $I(z)$ and GW kernel $\psi_{u,v}$ such that $G_{u,v} = I(z) * \psi_{u,v}$. $G_{u,v}$ can have both real and imaginary part and in this feature representation, only the Gabor magnitudes are used since small displacements can linearly affect Gabor phases [25]. The magnitude $M_{u,v}$ can be calculated from (2):

$$M_{u,v} = \sqrt{Im(G_{u,v}) + Re(G_{u,v})} \tag{2}$$

From the definition of Gabor kernel in (1), GW consists of a planar sinusoid multiplied by a 2D Gaussian. Due to the use of Gaussian, the region close to the center of the image would dominate the convolution process. In other words, the frequency information near the center of the Gaussian has more effect on the convolution than the frequency information far from the center of the Gaussian. Thus GW can effectively extract information within some local areas in the face.

2.2. Review on LGFV

In the original implementation of LGFV [24], the GW are applied before partitioning the face into LPs, while in LMG implementation [25], GW called are applied on specific locations on the face thus producing Gabor jets. LGFV implementation proposed that Gabor features are extracted by convolving the Gabor kernel $\psi_{u,v}$ with the sub-window sliding single image $I(x,y)$ pixels by pixels. Since 40 GWs are used, 40 Gabor Images (GI) are acquired from the convolutions on $I(x,y)$. Then, each GI is partitioned into m LPs. In [24], the LPs used are constructed from local facial components such as eyes, noses, mouth, etc., which are learned beforehand by selecting the patches with the highest discrimination and lowest correlation. The Gabor features are then grouped together according to their spatial location to obtain LGFV where each LGFV corresponds to a local facial component of the face.

2.3. Spike Response Model (SRM)

One of the most popular neuron model is the integrate-and-fire (I&F) model. The formulation of I&F model is further simplified and can be represented by SRM [30]. One advantage of SRM over I&F model is the use of arbitrarily chosen kernels instead of differential equations which enables SRM to be more universal than I&F model. For example, with appropriate choice of kernels SRM can reproduce 90% of the firing times in Hodgkin–Huxley model with ± 2 ms precision [31]. Let w_{ij} be the weight between post-synaptic neuron i and presynaptic neuron j, τ_s is the synaptic time constant, τ_{rec} is recovery time constant, I^{ext} is the external current, t_j^f is the time of presynaptic spikes, $\hat{t}_i$ is time of output spike, η, ϵ_0 and κ_0 are kernels, $\delta(t)$ is Dirac delta function, and $\eta_0 > 0$, according to [30] the membrane potential $u_m(t)$ can be computed

Please cite this article as: F. Kamaruzaman, A.A. Shafie, Recognizing faces with normalized local Gabor features and Spiking Neuron Patterns, Pattern Recognition (2015), http://dx.doi.org/10.1016/j.patcog.2015.11.020