

Biologically Inspired Dictionary Learning for Visual Pattern Recognition

A. Memariani and C. K. Loo
 University of Malaya, Kuala Lumpur, Malaysia
 E-mail: ali_memariani@siswa.um.edu.my, ckloo.um@um.edu.my.

Keywords: holonomic brain theory, dictionary learning, sparse coding, quantum particle swarm optimization, complex-valued synergetic neural network, body expression.

Received: March 4, 2013

Holonomic brain theory provides an understanding of neural system behaviour. It is argued that recognition of objects in mammalian brain follows a sparse representation of responses to bar-like structures. We considered different scales and orientations of Gabor wavelets to form a dictionary. While previous works in the literature used greedy pursuit based methods for sparse coding, this work takes advantage of a locally competitive algorithm (LCA) which calculates more regular sparse coefficients by combining the interactions of artificial neurons. Moreover the proposed learning algorithm can be implemented in parallel processing which makes it efficient for real-time applications. A complex-valued synergetic neural network is trained using a quantum particle swarm optimization to perform a classification test. Finally, we provide an experimental real application for biological implementation of sparse dictionary learning to recognize emotion using body expression. Classification results are promising and quite comparable to the recognition rate by human response.

Povzetek: Z zgledevanjem po bioloških sistemih je predstavljena je metoda učenja vizualnih vzorcev.

1 Introduction

Neural structure has been one of the inspirations of machine learning. However, the concept of axonal discharge is misunderstood. Pribram's holonomic brain theory, proposes the term neuromodulator rather than neurotransmitter to refer to the electrical gap in junctions (axodendritic and dendo-dendritic) caused by chemical synapses. Accordingly arrival patterns of a nerve impulse are described as sinusoidal fluctuating hyperpolarizations (-) and depolarizations (+) which are inadequately large to make a nerve impulse discharge instantly [1]. Maps of these hyper and polarizations are called receptive fields. These receptive fields of visual cortex contain multiple bands of excitatory and inhibitory areas which act as line detectors. Thus neurons are tuned to a limited bandwidth of frequencies to provide harmonic features; In other words neurons behave like active filters sensitive to oriented lines, movements and colours rather than Euclidean-based geometric features. A specific shape could be represented as a combination of filter responses (2-D convolution integrals). A set of filters is called a dictionary, since elements of a dictionary are not orthogonal to each other, there are many redundant features to represent an image (overcomplete approximation). A more sparse representation is obtained by selecting the best features among those with high correlation with each other. And remove others. Following an iterative strategy, the sparse coded

representation is generated in which selected features satisfy the orthogonality assumption.

This paper applies a locally competitive algorithm (LCA) [2] to extract the sparse coded definition of visual patterns. A synergetic neural network (SNN) is used to learn the visual features of a class of objects. SNN parameters are optimized with a quantum particle swarm approach.

1.1 Holonomic brain theory

The fact that for a harmonic oscillation we can either specify frequency or time (i.e. Heisenberg's principle of indeterminacy) has linked psychophysics and quantum mechanics. Gabor function is described as the modulation product of an oscillation with a given frequency (carrier) and an envelope in the form of normal distribution function. A biologically-plausible model for the visual pathway (retina, LGN, striate cortex) is described as a triple of convolutions. These triple convolutional preprocessing provides maximal coding of information. Biological Infomax visual cognition models such as independent component analysis (ICA) [3] and sparseness-maximization net [4] have better performance than classical Principal Component Analysis (PCA) or Hebbian models[5]. Relations between sparseness-maximization net and dendritic fields describes a dendritic implementation of sparseness-maximization net [6]; Though the dendritic

implementation is limited by infomax process which could be originated from top down lateral inhibition. Olshausen and Field formulate the reconstruction of stimuli in receptive fields of simple cell using sparse coding [4, 7]. Advantages of combining Gabor responses as in [4, 7] over ICA-like shapes are described by [6].

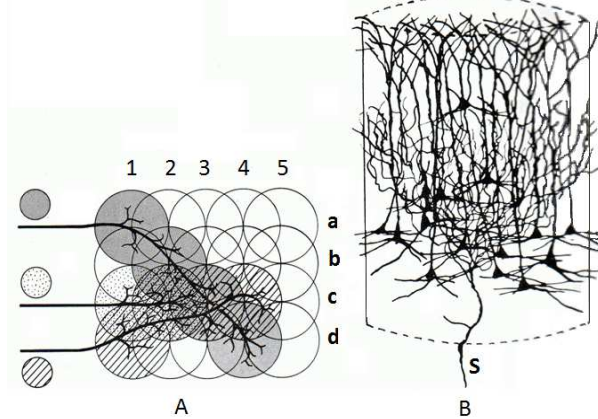


Figure 1: Microstructure of synaptic domains in cortex[1]. Overlapping line detectors (vertical and horizontal circles) combined to represent a stimulus (A) and interacting polarizations producing the dendritic fields (B).

Since then sparse coding is improved by many researchers. Though, most of them used greedy approaches to compute a sparse representation [8-10]. Accordingly, Biological realization of sparsification was unknown. However, in the recent work [2] a locally competitive algorithm (LCA) is proposed which is based on biological inhibition in neural circuits.

Table 1: Comparison of basic holonomic approaches[1].

	HNeT	Quantum Associative Net	ICA	field computing
Effectiveness	Very Effective	Effective	Very Effective	A general model with potentially very effective "sub-branches"
biological plausibility	Fundamental level only	Fundamental level only	bio-implausible but plausible output	fundamental level only
Possible quantum implementation	indirect similar core as QAN	direct	not yet known	partially direct
Main weakness	a mixture of natural and artificial features	limited to assoc. memory and pattern recognition	unknown bio implementability	Consciousness still missing

The striate cortex (V1) is the area of conscious visual perception in brain. Experimental results from functional magnetic resonance imaging (fMRI) supported that effect of the visual cortex in V1 in response to a stimuli can be estimated by a 2D Gabor function. A Gabor field I is the superposition of different Gabor functions' responses:

$$I = \sum_{j=1}^M a_j GW_j \tag{1}$$

where a_j and GW_j are Gabor coefficient and elementary Gabor function corresponding to the J_{th} element in the dictionary. The superposition of Gabor fields is in analogy to dictionary learning that represent the equation in similar form [4, 7, 9]. Therefore, the selection of Gabor coefficients can be performed by a sparse coding algorithm such as LCA so that an image is represented with minimum subset of Gabor elementary functions.

Output of V1 is projected to peri-striate cortex (V2) where probably retinal images are reconstructed. Triple-stage convolution in visual pathway has inspired convolutional neural networks acting as a course to fine process; though the research has focused mostly on magnitude data [11]. Some of the works included phase information to form an associative memory network [12]. Table 1 compares some of basic approaches of holonomic phase-magnitude encoding approaches.

Here we proposed a recognition algorithm based on the holonomic brain theory. Experimental results are compared to the state of the art algorithms. Furthermore, we applied the algorithm to recognize emotions based on body expression data which is inspired by the action based behavior in psychology. Classification results are compared to those of human recognitions.

2 Sparse coding

Representing an image with a few elementary functions is widely used in image processing and computer vision. Determining image component is useful to remove the noise. Also decomposition is used for compression by simplifying image representation.

In computer vision decomposition is a tool for feature extraction. An elementary function is called basis and set of bases functions is a dictionary. In early models choice of dictionary elements was subject to orthogonality condition. A complete representation of image is a linear combination of bases in the dictionary, derived by projection of image into bases. However, poor quality of representation in complete solutions resulted in relaxation of orthogonality condition and applying overcomplete dictionaries. Due to useful mathematical characteristics obtained by orthogonality (e.g. computing decomposition coefficients with projection), overcomplete dictionaries are still meant to be partially orthogonal. A common approach is to use an orthogonal subset of a large dictionary containing all possible elements.

Early works applied gradient descent to train the dictionary. Bayesian approaches also have been used to represent an image based on the MAP estimation of the dictionary components[13].

Textons are developed as a mathematical representation of basic image objects[14]. First images are coded by a dictionary of Gabor and Laplacian of Gaussian elements; Responses to the dictionary elements is Combined by transformed component analysis. Furthermore, sparse approximation helps to find a more general object models in terms of scale and posture[15].

Active basis model [16] provides a deformable template using Gabor wavelets as dictionary elements. They also proposed a shared sketch algorithm (SSA) inspired by AdaBoost.

2.1 Gabor wavelets

Biological models in object recognition are based on the findings of functional magnetic resonance imaging (fMRI) of mammalian brain. process of images in receptive fields (V1) is more sensitive on bar-like structures [17]. Responses of V1 are combined together by extrastriate visual areas and passed to inferotemporal cortex (IT) for recognition tasks. Research in computational neuroscience argued that recognition of objects in mammalian brain follows a sparse representation of responses to bar-like structures [4, 18]. Gabor wavelets are widely used as biologically inspired basis to model information encoding in receptive fields. 2D Gabor function centered at (x_0, y_0) is:

$$G(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left[\frac{(x-x_0)^2}{\sigma_x^2} + \frac{(y-y_0)^2}{\sigma_y^2}\right]} e^{i[\xi_0x + \nu_0y]} \quad (2)$$

where (ξ_0, ν_0) is optimal spatial frequency. Using wavelet transform a Gabor function can be rotated, dilated or translated. General form of Gabor wavelet function is:

$$GW(x, y, \omega, \theta) = \frac{\omega}{\sqrt{2\pi k}} e^{-\frac{\omega^2}{8k^2}(4(x\cos\theta+y\sin\theta)^2 + (-x\sin\theta+y\cos\theta)^2)} \left[e^{i\omega(x\cos\theta+y\sin\theta)} - e^{-\frac{k^2}{2}} \right] \quad (3)$$

where ω is the radial frequency and θ is the wavelet orientation. k is a constant representing bandwidth frequency [19]. Approximation of $k \approx \pi$ and $k \approx 2.5$ are common for 1 and 1.5 octave bandwidth (ϕ) respectively. Generally k is:

$$k = \sqrt{2 \ln 2} \left(\frac{2^\phi + 1}{2^\phi - 1} \right) \quad (4)$$

A dictionary of Gabor wavelets (as shown in Fig.2), including n orientations and m scales is in the form of: $GW_j(\theta, \omega)$, $j = 1, \dots, m \times n$, where

$$\theta = \left\{ \frac{k\pi}{n}, k = 1, \dots, n-1 \right\}, \quad (5)$$

and $\omega = \frac{\sqrt{2}}{i}, i = 1, \dots, m$.

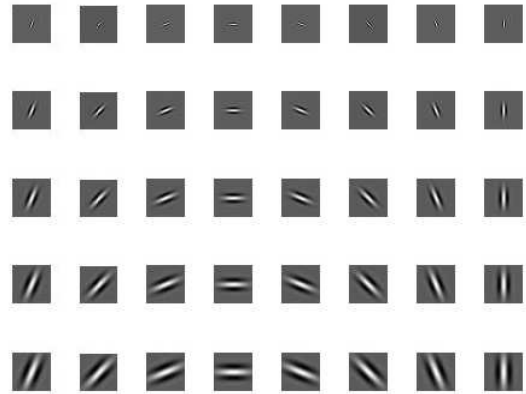


Figure 2: A dictionary of Gabor wavelets.

2.2 Sparse coding using locally competitive algorithm

Response to a dictionary of Gabor wavelets is an overcomplete representation. Sparse coding is the method of selecting a proper subset of responses to represent the image (signal). In addition to biological motivations, sparse coding is necessary to avoid redundant information. Having a fixed number of features, redundancy may cause loss of essential information which is going to be encoded in the lower levels (Fig.4).



Figure 3: Edge detection using Gabor wavelets, A. Original image[1], B. edge detected image with a large number of features without sparsity, C. edge detected image with a small number of features where sparsity is enforced.

Assuming an image (I_0) its sparse approximation I is derived according to (1). Optimal sparse coding tries to minimize the number of nonzero coefficients a_j , which is an NP-hard optimization problem.

We applied a locally competitive algorithm (LCA) [2] to enforce local sparsity. Unlike classical sparse coding algorithms, LCA uses a parallel neural structure inspired by biological model. LCA is applied to minimize the mean square error combined with a cost function in the local neighbourhood:

$$E(t) = \frac{1}{2} \|I(t) - I_0\|^2 + \lambda \sum_j C(a_j(t)). \quad (6)$$

Thresholds are useful to generate coefficients with exact zero value.

For a threshold function $T_{(\alpha, \gamma, \lambda)}(\cdot)$, cost function C is:

$$C_{(\alpha, \gamma, \lambda)}(a_j) = \frac{(1 - \alpha)^2 \lambda}{2} + \alpha |a_j|, \quad (7)$$

$$T_{(\alpha, \gamma, \lambda)}(u_j) = \frac{u_j - \alpha \lambda}{1 + e^{-\gamma(u_j - \lambda)}} \quad (8)$$

Limit of T as $\gamma \rightarrow \infty$ is called ideal thresholding function. $T_{(0,\infty,\lambda)}(\cdot)$ is hard thresholding function and $T_{(1,\infty,\lambda)}(\cdot)$ is soft thresholding function.

In previous works, there is no real application that has been applied using LCA, although some simulation results are shown. Here an empirical experiment based real application of body expression recognition, is proposed to provide an evidence for the practical utility of Holonomic Brain Model as dictionary learning method by LCA.

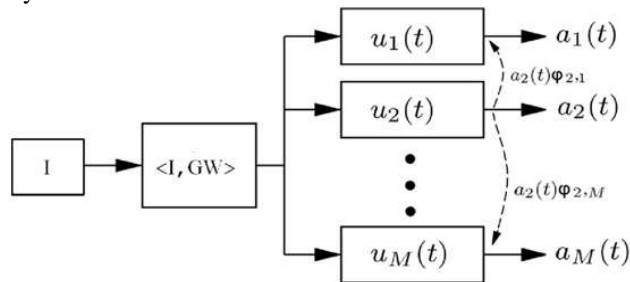


Figure 4: LCA structure [4].

LCA structure acts as a set of integrate and fire neurons. response to a dictionary of filter charges the internal state of the neurons and leads to the activity of the neuron. Neurons with higher charge (internal state) become active and fire signals to inhibit other neurons. A firing signal keeps other neurons that are highly correlated with the corresponding active neuron from being active by defusing their charge in an unidirectional inhibition.

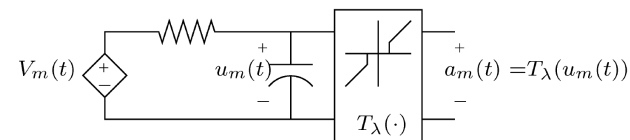


Figure 5: integration (charge up) and fire in a neural circuit [2].

3 Locally competitive active basis recognition

We applied a supervised algorithm to recognize two types of objects in images; First a pixel-wise approach for aligned objects which combines the learned samples of objects in each class to form a prototype and second a feature based approach for non-aligned objects in which Gabor wavelets are localized to represent a potential match between specific scale and orientation and edges of objects. Both approaches are fed into a synergetic neural network to perform a classification task.

Images are scaled to have the exact same size. Each image is convolved with all the elements in the dictionary. Then sparse coding is enforced to minimize the representing elements for each pixel. Finally, remaining parts are reconstructed to generate the sparse superposition of the image. For pixel values in the local area LCA has the following steps:

1. Compute the response (convolution) of I_i with all the elements in the dictionary.

$$C_j = \langle GW_j, I_j \rangle \tag{9}$$

(Set $t = 0$ and $u_j(0) = 0$, for $j = 1, \dots, n$).

2. Determine the active nodes by activity thresholding.
3. For each pixel calculate internal state u_j of element j .

$$u_j = \frac{1}{\tau} \left[C_j(t) - u_j(t) - \sum_{j \neq k} \Phi_{j,k} \cdot a_j(t) \right] \tag{10}$$

$$\Phi_{j,k} = \langle GW_j, GW_k \rangle \tag{11}$$

4. Compute sparse coefficients $a_j(t)$ for $u_j(t)$.

$$a_j(t+1) = T_\lambda(u_j(t)) \tag{12}$$

$$T_{(\alpha,\gamma,\lambda)}(u_j) = \frac{u_j - \alpha\lambda}{1 + e^{-\gamma(u_j - \lambda)}} \tag{13}$$

5. If $a_j(t-1) - a_j(t) > \delta$ then $t \leftarrow t+1$ and go to step 2, otherwise finish.

Original SNN used pixel-wised features to represent an object which is not robust in case objects are in a variable shapes (e.g. different body emotions of human). In this case, we construct a template model as a collection of Gabor wavelet features included in the dictionary which represents the general characteristics of all body posture classes. Test images are convolved with the components of the template model. Sparsity is then enforced to catch the best fit over the specific posture. LCA thresholding strategy enables us to remove redundancies effectively (producing sparse coefficients with exactly zero values). Number of output Gabor wavelets are fixed in order to make the comparison with trained prototype of each class. Features are selected based on their highest response to the training images; furthermore, each feature is allowed to perturb slightly in terms of location and orientation. In this aspect our template construction is a modification of shared sketch algorithm [16]. For each image i feature value v_{ij} corresponded to the selected Gabor wavelet j , is determined as the following:

$$v_{ij} = \gamma_i C_{ij} - \log(Z(\gamma_i)) \tag{14}$$

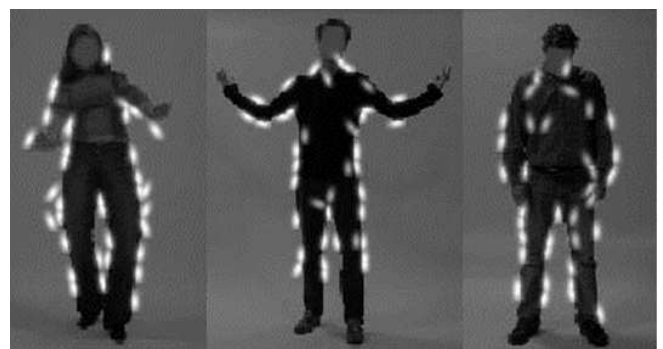


Figure 6: Gabor wavelet features detecting the edge pattern of different body postures.

where γ_i is derived by maximum likelihood estimation and Z is the partition function. Therefore, boundaries of

object are segmented out before the result is given to SNN.

3.1 Complex-valued synergetic neural network

Synergetic neural network (SNN) developed by Haken [20] describes the pattern recognition process in the human brain which tries to achieve the learned model with fast learning and no false state rather than traditional neural networks [21, 22] [20].

A common approach to combine learned samples is averaging the feature values. One way to deal with inflexibility is to use learning object in the same view which will restrict the classification task. A melting algorithm is proposed by [23] to combine objects in deferent poses. Suppose a learned sample object \hat{I} consists of n pixel values. \hat{I} is reshaped to a column vector v_i and normalized so that:

$$\sum_{j=1}^n v_{ij} = 0 \tag{15}$$

$$\sum_{j=1}^n |v_{ij}|^2 = 1 \tag{16}$$

A prototype V^\dagger is the Hermitian conjugate of V :

$$V^\dagger = (V^T V)^{-1} V = C(v) + iS(v) \tag{17}$$

A test samples q corresponding to a test image is normalized and compared to the prototype of each class, using the order parameters. For each prototype k order parameters ε_k is initialized as:

$$\varepsilon_k = v_k^\dagger \cdot q, \quad k = 1, \dots, m. \tag{18}$$

where v_k^\dagger is the k th row in the Hermitian conjugate V^\dagger . Order parameters are updated derived iteratively with the synergetic dynamics:

$$\dot{\varepsilon}_k = \frac{1}{D} (\lambda_k - D + B\varepsilon_k^2) \varepsilon_k + \varepsilon_k \tag{19}$$

$$D = (B + C) \sum_k \varepsilon_k^2 \tag{20}$$

where λ_k is the attention parameter for class k ; B, C are constants [24]. Attention parameters could be considered balance (equal and mostly unit) or unbalance. Attention parameters in the model are trained using a quantum particle swarm optimization in order to minimize the overall classification error in the test set.

3.2 Centroidal Voronoi Tessellation (CVT)

As mentioned in section 3.1 unbalance attention parameters should be tuned. We applied a CVT in order to cover the whole feasible space in the initial state of the random search. A set of generators are considered as a group of points in the space forming a Tessellation. Generators are associated with subsets and points are nearer to its corresponding generators rather than any of other generators according to the distance function (e.g., the l_p norm). Note that the generators are not quite evenly distributed throughout the space. Dividing the feasible

space into the partitions, several generators set at almost precisely the same point in the space. CVT overcomes the poor and non-uniform distribution of some Voronoi cells by choosing the generators at centre [25-27]. Assuming λ_{max} as the maximum potential attention parameter search space is defined as:

$$0 < \lambda_i < \lambda_{max}, i = 1, \dots, m. \tag{21}$$

Given a set of Voronoi regions $T_\xi (\xi = 1, \dots, \Xi)$ in the space $\Omega \subset R^m$, each initial position p_ξ is the Centroid of its region.

$$T_\xi = \left\{ x \in \Omega, |x - p_\xi| < |x - p_{\hat{\xi}}| \text{ for } \hat{\xi} = 1, \dots, \Xi, \hat{\xi} \neq \xi \right\} \tag{22}$$

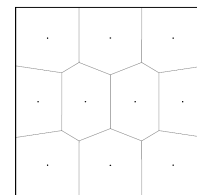


Figure 7: Centroidal Voronoi tessellation dividing a square into 10 regions[28].

3.3 Quantum particle swarm optimization (QPSO)

Initial attention parameters are tuned using a QPSO in order to minimize the overall classification error in the test set. Each particle position X , is updated based on the movement framework in the quantum mechanics.[29] State of the particle is described by a wave function,

$$\psi(Y) = \frac{1}{\sqrt{L}} e^{-\frac{|Y|}{L}} \tag{23}$$

$$Y = X - p \tag{24}$$

$$L = \frac{h^2}{my} \tag{25}$$

where y is called intensity of the potential well at point p , m is the particle mass and h is a constant. Finally, for particle i , j th element of the position $X_{i,n}^j$ can be updated as:

$$X_{i,n+1}^j = p_{i,n}^j \pm \frac{L_{i,n}^j}{2} \ln \left(\frac{1}{u_{i,n}^j} \right), \tag{26}$$

$$u_{i,n+1}^j \approx U(0,1), \tag{27}$$

$$L_{i,n}^j = 2\alpha |X_{i,n}^j - C_n^j|, \tag{28}$$

$$C_n^j = \frac{1}{M} \prod_{i=1}^M p_{i,n}^j, \tag{29}$$

where C_n^j is the average of all particle positions. α is a positive real number which could be constant or change dynamically in total N iteration as:

$$\alpha = \frac{0.5 * (N - n)}{N} + 0.5 \quad (30)$$

To improve the accuracy an adaptive penalty function [30] is added to the overall error:

$$Z = f(\Lambda) + \sum_{i=1}^m k_j v_j \quad (31)$$

$$k_j = |f(\Lambda)| + \frac{v_j(\Lambda)}{\sum_{i=1}^m v_j(\Lambda)^2} \quad (32)$$

$$\Lambda = (\lambda_1, \dots, \lambda_k) \quad (33)$$

Figure 8 shows an overview of the recognition method. QPSO is used to iteratively tune the attention parameters $\lambda_i, i = 1, \dots, k$ where k is the number of classes.

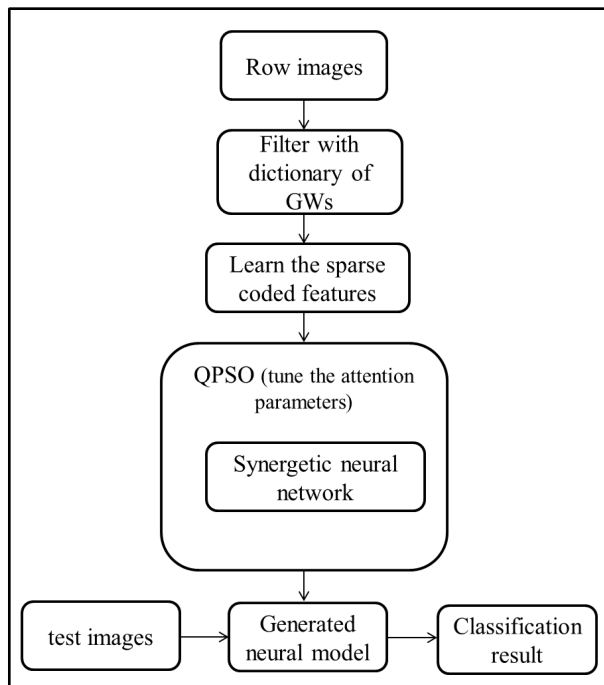


Figure 8: Scheme of proposed visual recognition model.

4 Emotion recognition using body expression and results

Even though most of the works in the area of emotion recognition has been focused on facial expressions, some of psychological theories considered emotional appraisals that are not facially expressive [31–33]. In that sense, emotions are described based on the state of action readiness that they cause in the whole body (either impulsive or intentional)[31]. Intentional actions might differ person to person though impulsive actions only depend of the nature of their action readiness.

Accordingly, impulsive actions can be used to recognize emotions considering the body expressions.

Facial expression has been combined with upper body gestures to recognize emotions [34]. Movements of hands are detected using color segmentation and represented by centroid of the area; face components is also detected using skin detection techniques. Facial features (eyebrows, mouth, chin, etc.) are then combined with hand movements to set up the features. similar works has consider body feature along with facial features for fear detection [35] and anger detection [36, 37].

Body gestures are also merged with speech based features derived by acoustic analysis. Together with facial expressions [38] developed a framework in which face and body data was recorded with different resolutions and synchronized with subjects' speech interaction. They applied a Bayesian classifier to recognize the emotions.

Kleinsmith et al [39] argued that emotions can be recognized by humans from body postures when their face is removed. They also developed a recognition model to recognize the affection of faceless avatars in computer games.

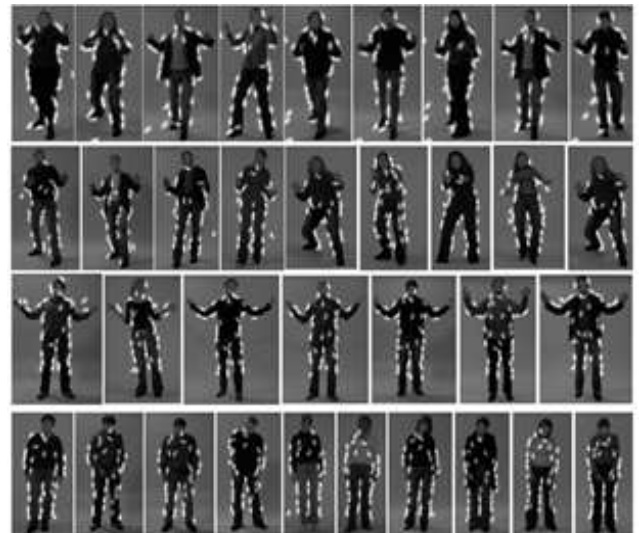


Figure 9: Extracted features for four classes of emotions top down as, anger, fear, happiness, sadness.

Human actions caused by emotions could be detected using point-light animations [40]. Ross et al perform a test to compare recognition ability of students in primary and secondary schools and adults [41]. Faces of the test subjects were covered and recognition task performed on both full-light display and a point-light display where only main parts of the body postures are shown in a black and white format. Their result shows that adults have a better ability of bodily emotion recognition and display full-light is more expressive than in point-light for the task.

Table 2: Classification Accuracies for different QPSOs.

	Anger (%)	Fear (%)	Happiness (%)	Sadness (%)	Overall Error (%)
QPSO1	92.31	68.97	72.0	93.10	18.35
QPSO2	36.54	93.10	62.0	93.10	27.52
QPSO3	92.31	72.41	74.0	93.10	16.97
QPSO4	36.54	93.1	64.0	93.10	27.06
QPSO5	92.31	86.21	60.0	93.10	16.51
QPSO6	92.31	68.97	72.0	93.10	18.35
QPSO7	36.54	94.83	64.0	93.10	26.61
QPSO8	82.69	89.66	66.0	93.10	16.51
BEAST (Human Recognition)	93.6	93.9	85.4	97.8	

In order to validate the perception of body expression tests have been developed and validated by human recognition. Atkinson et al developed a dataset for both static and dynamic body expressions; The dataset contains 10 subjects (5 female) and covers five emotions (anger, disgust, fear, happiness and sadness)[42]. The bodily expressive action stimulus test (BEAST) [43] provides a dataset for recognizing four types of emotions (anger, fear, happiness, sadness) which is constructed using non-professional actors (15 male, 31female). Body expressions are validated with a human recognition test.

We applied a supervised approach to recognize two types of objects in images; First a pixel-wise approach for aligned objects which combines the learned samples of objects in each class to form a prototype and second a feature based approach for non-aligned objects in which Gabor wavelets are localized to represent a potential match between specific scale and orientation and edges of objects (figure 9). Both approaches are fed into a synergetic neural network to perform a classification task.

We applied the BEAST data set¹ to classify four classes of basic emotions. Gabor wavelets are generated in a (20, 20) matrix and images are resized to have 500 pixels in row and relatively scaled pixels in column. Images are divided into train and test sets for each class 10 images are selected randomly to form the train data and the rest are included for test. Different scenarios are considered to train the model:

1. Static QPSO with $\alpha=0.75$ and randomly initialized.
2. Static QPSO with synergetic melting prototype [44].
3. Dynamic QPSO where α changes according to (29) and randomly initialized.
4. Static QPSO with $\alpha=0.75$ and initialized with CVT.
5. Dynamic QPSO as (29) and initialized with CVT.
6. Dynamic QPSO as (29), initialized with CVT and a synergetic melt prototype.
7. Static QPSO with $\alpha=0.75$, initialized with CVT and penalized with (30).

8. Dynamic QPSO as (29), initialized with CVT and penalized with (30).

Classification accuracies of different trained SNNs are compared with results of human recognition (table2). In some cases happiness and anger are misclassified as fear, this happened more frequently in static learning. However regardless of the learning scenario, happiness turns to be the most difficult one to detect and the reason is not clear for the authors.

Figures 10 and 11 show the learning rate for each scenarios during the learning iterations. CVT has improved the accuracy with Dynamic learning scenario.

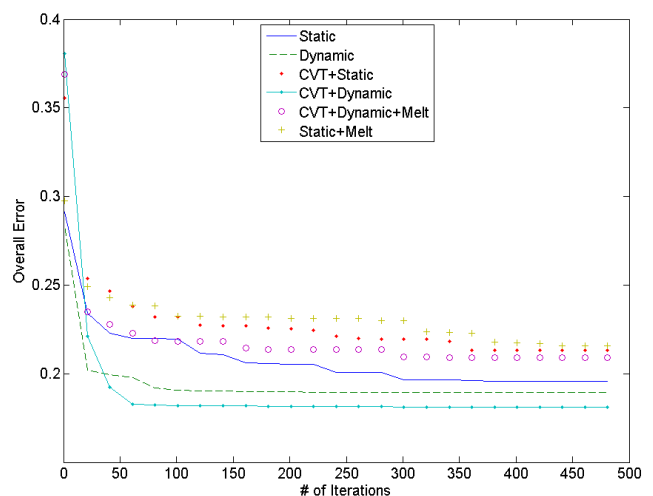


Figure 10: Average learning rates for different QPSOs.

¹ <http://www.beatricedegelder.com/beast.html>

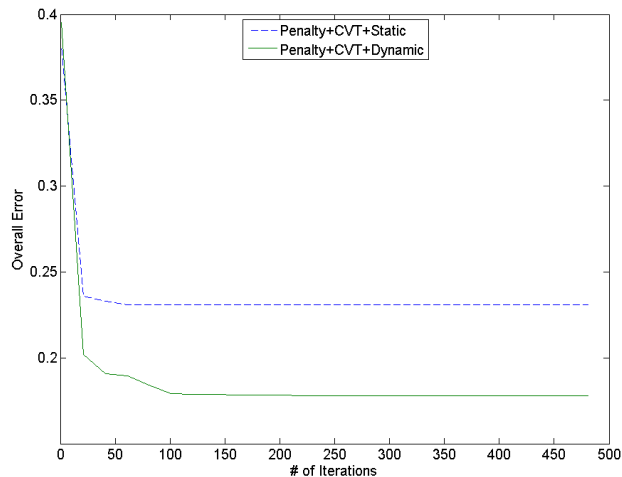


Figure 11: Average learning rates for penalized objective functions.

5 Conclusion

We proposed a biologically-plausible approach for recognition of aligned and non-aligned objects. Our dictionary learning algorithm is inspired by the holonomic brain theory. LCA is applied to enforce sparsity on a dictionary of Gabor wavelets. Regarding the parallel structure of the learning method, implementation could be optimized via parallel processing which is essential for real-time applications.

Furthermore, synergetic neural network is combined with Gabor wavelet features which make it applicable for recognition of non-aligned objects. Gabor features also enhance the SNN to use images with different size for both construction of the Hermitian conjugate and test images. Effect of background is also removed because of recognition is based on the pattern of edges; Though sparse coding is robust in presence of classical noise since dot noise does not follow any meaningful shape pattern intrinsically.

Experimental results supported the real application of Holonomic Brain Model as dictionary learning method using a biological implementation.

Acknowledgment

This work is supported by Flagship research grant of University of Malaya (FL006-2011) “PRODUCTIVE AGING THRU ICT” and HIR-MOHE research grant (H-22001-00-B000010) of University of Malaya.

References

- [1] Pribram, K.H., *Brain and perception : holonomy and structure in figural processing*1991, Hillsdale, N.J.: Lawrence Erlbaum Associates. xxix, 388 p.
- [2] Rozell, C.J., et al., *Sparse coding via thresholding and local competition in neural circuits*. Neural Computation, 2008. **20**(10): p. 2526-2563.
- [3] Makeig, S., et al., *Independent component analysis of electroencephalographic data*. Advances in neural information processing systems, 1996: p. 145-151.
- [4] Olshausen, B.A. and D.J. Field, *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*. Nature, 1996. **381**(6583): p. 607-609.
- [5] Peruš, M. and C.K. Loo, *Biological and Quantum Computing for Human Vision: Holonomic Models and Applications*2010: Medical Information Science Reference.
- [6] Peruš, M., *Image processing and becoming conscious of its result*. Informatica, 2001. **25**: p. 575-592.
- [7] Olshausen, B.A. and D.J. Field, *Sparse coding with an overcomplete basis set: A strategy employed by V1?* Vision Research, 1997. **37**(23): p. 3311-3325.
- [8] Chen, S.S.B., D.L. Donoho, and M.A. Saunders, *Atomic decomposition by basis pursuit*. Siam Journal on Scientific Computing, 1998. **20**(1): p. 33-61.
- [9] Olshausen, B.A. and D.J. Field, *Sparse coding of sensory inputs*. Current Opinion in Neurobiology, 2004. **14**(4): p. 481-487.
- [10] Mallat, S.G. and Z.F. Zhang, *MATCHING PURSUITS WITH TIME-FREQUENCY DICTIONARIES*. Ieee Transactions on Signal Processing, 1993. **41**(12): p. 3397-3415.
- [11] Haykin, S.S., *Neural networks: a comprehensive foundation*1994: Macmillan.
- [12] Hopfield, J.J., *Neural networks and physical systems with emergent collective computational abilities*. Proceedings of the National Academy of Sciences, 1982. **79**(8): p. 2554-2558.
- [13] Kreutz-Delgado, K., et al., *Dictionary learning algorithms for sparse representation*. Neural Comput., 2003. **15**(2): p. 349-396.
- [14] Zhu, S.C., et al., *What are textons?* International Journal of Computer Vision, 2005. **62**(1-2): p. 121-143.
- [15] Figueiredo, M.A.T., *Adaptive sparseness for supervised learning*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2003. **25**(9): p. 1150-1159.
- [16] Wu, Y.N., et al., *Learning Active Basis Model for Object Detection and Recognition*. International Journal of Computer Vision, 2010. **90**(2): p. 198-235.
- [17] Riesenhuber, M. and T. Poggio, *Neural mechanisms of object recognition*. Current Opinion in Neurobiology, 2002. **12**(2): p. 162-168.
- [18] Daugman, J.G., *TWO-DIMENSIONAL SPECTRAL-ANALYSIS OF CORTICAL RECEPTIVE-FIELD PROFILES*. Vision Research, 1980. **20**(10): p. 847-856.
- [19] Tai Sing, L., *Image representation using 2D Gabor wavelets*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1996. **18**(10): p. 959-971.
- [20] Haken, H., *Synergetic Computers and Cognition: A Top-Down Approach to Neural Nets*2004: Springer.

- [21] Koruga, D., et al. *Synergy of classical and quantum communications channels in brain: neuron-astrocyte network*. in *Neural Network Applications in Electrical Engineering, 2004. NEUREL 2004. 2004 7th Seminar on*. 2004.
- [22] Bin, L. and T. Yuru. *The research of learning algorithm of synergetic neural network*. in *Computer Science and Information Processing (CSIP), 2012 International Conference on*. 2012.
- [23] Hogg, T., D. Rees, and H. Talhami. *Three-dimensional pose from two-dimensional images: a novel approach using synergetic networks*. in *Neural Networks, 1995. Proceedings., IEEE International Conference on*. 1995.
- [24] Gao, J., et al., *Optical-electronic shape recognition system based on synergetic associative memory*. 2001: p. 138-148.
- [25] Richards, M. and D. Ventura. *Choosing a starting configuration for particle swarm optimization*. in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*. 2004.
- [26] Du, Q., V. Faber, and M. Gunzburger, *Centroidal Voronoi Tessellations: Applications and Algorithms*. SIAM Rev., 1999. **41**(4): p. 637-676.
- [27] Nguyen, H., et al., *Constrained CVT meshes and a comparison of triangular mesh generators*. *Comput. Geom. Theory Appl.*, 2009. **42**(1): p. 1-19.
- [28] Burkardt, J., et al., *User manual and supporting information for library of codes for centroidal Voronoi point placement and associated zeroth, first, and second moment determination*. SAND Report SAND2002-0099, Sandia National Laboratories, Albuquerque, 2002.
- [29] Sun, J., et al., *Quantum-Behaved Particle Swarm Optimization: Analysis of Individual Particle Behavior and Parameter Selection*. *Evolutionary Computation*, 2012. **20**(3): p. 349-393.
- [30] Barbosa, H.J.C. and A.C.C. Lemonge, *A new adaptive penalty scheme for genetic algorithms*. *Inf. Sci.*, 2003. **156**(3-4): p. 215-251.
- [31] Frijda, N.H., *THE LAWS OF EMOTION*. *American Psychologist*, 1988. **43**(5): p. 349-358.
- [32] Frijda, N.H., *Not Passion's Slave*. *Emotion Review*, 2010. **2**(1): p. 68-75.
- [33] Frijda, N.H., *Impulsive action and motivation*. *Biological Psychology*, 2010. **84**(3): p. 570-579.
- [34] Gunes, H. and M. Piccardi, *Bi-modal emotion recognition from expressive face and body gestures*. *Journal of Network and Computer Applications*, 2007. **30**(4): p. 1334-1345.
- [35] van Heijnsbergen, C.C.R.J., et al., *Rapid detection of fear in body expressions, an ERP study*. *Brain Research*, 2007. **1186**(0): p. 233-241.
- [36] Pollick, F.E., H. Paterson, and P. Mamassian, *Combining faces and movements to recognize affect*. *Journal of Vision*, 2004. **4**(8): p. 232.
- [37] Paterson, H.M., F.E. Pollick, and E. Jackson., *Movement and faces in the perception of emotion from motion*, in *Perception, ECVP Glasgow Suppl2002*. p. 232.
- [38] Kessous, L., G. Castellano, and G. Caridakis, *Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis*. *Journal on Multimodal User Interfaces*, 2010. **3**(1): p. 33-48.
- [39] Kleinsmith, A., N. Bianchi-Berthouze, and A. Steed, *Automatic Recognition of Non-Acted Affective Postures*. *Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 2011. **41**(4): p. 1027-1038.
- [40] Blake, R. and M. Shiffrar, *Perception of human motion*, in *Annual Review of Psychology 2007*, *Annual Reviews: Palo Alto*. p. 47-73.
- [41] Ross, P.D., L. Polson, and M.-H. Grosbras, *Developmental Changes in Emotion Recognition from Full-Light and Point-Light Displays of Body Movement*. *PLoS ONE*, 2012. **7**(9): p. e44815.
- [42] Atkinson, A.P., et al., *Emotion perception from dynamic and static body expressions in point-light and full-light displays*. *Perception*, 2004. **33**(6): p. 717-746.
- [43] de Gelder, B.V.d., Stock J., *The Bodily Expressive Action Stimulus Test (BEAST). Construction and Validation of a Stimulus Basis for Measuring Perception of Whole Body Expression of Emotions*. *Frontiers in Psychology*, 2011. **2**:181. doi:10.3389/fpsyg.

