International Journal of Neural Systems, Vol. 22, No. 6 (2012) 1250029 (17 pages)
© World Scientific Publishing Company
DOI: 10.1142/S0129065712500293

BIOLOGICALLY INSPIRED FACE RECOGNITION: TOWARD POSE-INVARIANCE

NOEL TAY NUO WI
Centre of Diploma Programmes, Multimedia University
JalanAyerKeroh Lama, 75450 Melaka, Malaysia
nwtay@mmu.edu.my

CHU KIONG LOO
Department of Artificial Intelligence
University of Malaya, 50603 Kuala Lumpur, Malaysia
ckloo.um@um.edu.my

LETCHUMANAN CHOCKALINGAM
Faculty of Information Science and Technology
Multimedia University, Ayer Keroh, 75450 Melaka, Malaysia
Chockalingam@mmu.edu.my

Accepted 11 October 2012
Published Online 13 November 2012

A small change in image will cause a dramatic change in signals. Visual system is required to be able to ignore these changes, yet specific enough to perform recognition. This work intends to provide biological-backed insights into 2D translation and scaling invariance and 3D pose-invariance without imposing strain on memory and with biological justification. The model can be divided into lower and higher visual stages. Lower visual stage models the visual pathway from retina to the striate cortex (V1), whereas the modeling of higher visual stage is mainly based on current psychophysical evidences.

Keywords: Biologically inspired vision; 3D pose-invariance; face recognition; hierarchical architecture.

1. Introduction

Applications of face recognition have been integrated in areas such as human–machine interaction.[1] and security applications. Face recognition maps the image of a person to the appropriate information corresponding to the person such that further processing can be achieved.[2–5] Technologies in face recognition are improving, but, they are still far inferior compared to biological visual system in virtually every way. Although capable of achieving near perfect recognition rate under controlled condition, the performance of face recognition is less than mediocre when applied to real-life uncontrolled situation. Variables found in real-life scenario such as facial expressions, viewing poses, external accessories and occlusions needed to be dealt with, in order to achieve successful recognition. However, there is also an issue of performance versus computational power.[6] Yet, biological visual system outperforms advance artificial visual systems, despite the neurons being a few tenths of magnitude slower than transistors, and has lower accuracy and resolution. Therefore, it is crucial to study how nature comes out with this efficient processing to perform feats of such magnitude. But this does not mean one needs to emulate birds to be able to fly, but rather, study the underlying principles that causes flight. In other words, we try to learn the "laws" exploited by nature in the course of evolution to come out with an efficient visual system.

Biological wise, there is a rapid progression in knowledge in neuroscience as well as developments in computational models to emulate and explain experimental observations of biological brain in the past 60 years. Human visual system is one of the most studied systems of the brain. But due to the complexity of the brain, scientists still do not have sufficient knowledge to fully understand the sophisticated processes that happen during recognition process. Currently, only deductions based on observations are made on the possible processing that took place for different parts of the brain.

Piecing together the puzzle is a monumental task. Many researchers have contributed in various ways to close the gap of understanding. Therefore, it is the intention of the researcher to also try to contribute a small piece for the puzzle, that is, to build a framework that can act as a guide to assist in piecing together the fragmented studies and information on visual invariance, as well as giving insight into a more efficient way of performing recognition such that it can be applied to current face recognition systems.

The framework is built based on well-established research and findings. There are also possible contradictions with other prominent models. As an example, the framework built assumes a feed-forward path for early visual stages due to the fact that spiking rate of neurons according to the conventional view of academic establishment is too slow for feedback to take place to obtain the speed observed in experiments during recognition tasks.[7,8] Yet, there are other justifications that feedback can happen that are biologically plausible such as through critical fusion frequency and neuron priming.[9] But this by no means indicates that feedback concept is refuted for early visual processes. A framework needs to be built in order to study the contradictions. As stated by Albright and Gross,[10] any systematic method will fail if they are without frameworks or prior models due to the magnitude of complexity of the nonlinear processes in the brain. Caution is taken to avoid compromise bias.

This paper is divided into five sections. Section 1 is on introduction and problem statement. Section 2 discusses the background literature for our framework. Since this work is about achieving invariance through lower and higher level feed-forward architecture which is biologically inspired, Sec. 2.1 discusses the concept of invariance, Sec. 2.2 discusses the hierarchical architecture which is used for lower level visual system, and Sec. 2.3 discusses some psychophysical experiments that help us model the higher level visual system. The framework is discussed in Sec. 3, which is divided into lower and higher level visual system. Results and discussions of the performance of the framework are given in Sec. 4. Finally, Sec. 5 gives the conclusion.

2. Background Literature

Extensive work on early biological visual systems has been performed. Functional mapping and information flow of the brain has been extensively worked on. One can refer to the paper by Ungerleider and Haxby[11] for the path ways and functional roles of parts in the brain. Feed-forward hierarchical model with strict biological justifications is developed for early visual processes.[12–14] Fazl *et al.* developed a model with different functional parts in the "where" and "what" stream contributing to attention selection through attentional shroud and saccadic eye movements.[15] However there is a lack of biological plausible models in neural processes of higher level computational tasks such as invariance and stable symbolic representation for recognition.[12,13,16–18] Probabilistic models like the top-down Bayesian inference approach[19] though can be used to interpret psychophysical experiments, lack correspondences between functional primitives and structural primitives. Thus, computational model for this research is built based on literatures of psychophysical observations yet provide room for neuron processing to take place. The problem intended to be solved is translation, scale and pose-invariance and low-level and high-level information linkage.

2.1. *Invariance*

Invariance means being able to identify an object when the object is being presented under different transformations, such as insensitivity to certain transformations like translation, scaling and rotation. More complicated invariance involves projection, illumination, depth transformation and facial emotions. For an ability that seems effortless for biological systems, ability to achieve invariance is an extremely subtle and elusive problem to be solved and emulated using current technologies at hand. Among the most basic invariance is two-dimensional shift[20] and scale invariance.[21] But there are limitations to how much such transformations can be tolerated by the visual system. For review, refer to the paper by Wiskott.[22]

The visual cortex consists of areas that have a hierarchical structure.[23] Along the visual stream of the visual cortex, complexity in the cells' preferred attributes and invariance increases. Starting from simple features, complexity increases to basic shapes, then to view-specific object and finally pose-invariant object. More precisely, early stage regions like the striate cortex V1 respond selectively to images of bars of certain orientation, V2 responds to angular stimuli which arrives from nonlinear combination of earlier response from much simpler cells, while V4 responds to more complicated shapes[24] like the letter "T" or a cross, and more so for inferior temporal cortex where neurons, selective to a wide range of attributes such as textures and colors, are tuned to more complicated shapes. Complexity increases further, where the neurons are tuned toward stimuli like faces and limbs. For example, there are neurons from the inferior temporal cortex that are observed to be face pose-invariant, regardless of the position of the face up to a certain threshold.[25] Therefore, it is believed that inferior temporal cortex is involved in visual information processing and identification of objects.[26,27] From electrophysiological observations, cells from inferior temporal cortex shows little attenuation under transformations.[28,29] Receptive field sizes (and thus, translation invariance), increases from bottom to top level of the visual processing hierarchy.[30]

To achieve invariance, proper metric needs to be defined that can be used as a comparison tool between the input and a reference subject to the design of the metric. Two hypotheses regarding shape constancy achieved by the reference are brought forward by Palmer,[31] that is, invariant features hypothesis and reference frame hypothesis. Invariant features hypothesis states that the reference is defined by features that do not change with transformations such as relative length, number of lines, angle size and number of angles. Reference frame hypothesis assumes an inner reference frame which neutralizes transformations that occurs to the external object. This neutralization by a counter transformation is called mental transformation. Experiment by Shepard shows that the more transformation that occurred to an object from its original position, the longer time it took for test subjects to identify it, thus, serves as an evidence for mental rotation.[32] This applies to scaling as well, where "mental adjustment" is applied.

2.2. *Simple/complex cells hierarchical architecture*

A rapid short presentation of visual signal, between 100 and 200 ms, is too fast for feedback modulation, eye movements and attention shifts to happen. Normalizing to a standardized frame through feedback modulation is considered impossible under the conventional concept of neuronal spike and process that it represents for such a short time. Yet recognition, like scene gist, can be easily achieved in less than 150 ms, as compared to fixation duration which is about 300 ms.[33] Human can easily detect objects in a natural scene and images or making accurate statistical judgments within 150 ms.[7,8,34–36] These works[12–14,37] provide a simple/complex cells hierarchical architecture of the visual cortex to explain the first 100–200 ms of visual processing, where attention modulation, feedbacks and eye shifts cannot be afforded due to such brief time. They also show that 2D translation and scaling invariance can be achieved through the simple and complex cell hierarchical structure. Serre *et al.* built a more sophisticated version which includes four layers of these processing units to perform classification.[37] Invariance is achieved due to pooling mechanism of complex cell which picks the maximum value from the simple cells it encompasses.

2.3. *Implications from change blindness experiments*

For pose-invariance, method employed in models from Fazl *et al.*, and Riensenhuber and Poggio involves storing all view images of the same object (represented by view-tuned cells), which in turn activates one view-invariant cell.[12,13,15]

Experimental work on monkeys showed such view-invariant cells reacting to novel objects (paper clips).[38–40] Multiple of these view-tuned cells activate a single cell (Grandfather cell), thus, it is invariant to different views of the same object.[13,41–44] This applies only when all views of different poses of the same object are stored, thus, enable recognition invariant to poses. This method is impractical for large number of objects since the system needs to be exposed to every view images, and to be able to store them all. A solution, though unknown implementation wise, is to generate virtual views for novel objects through interpolation between stored information when relevant templates are not available through previously learnt prototypes,[44–49] such that novel view-points can be aligned to the nearest view-point. Borders of columns in inferior temporal cortex contain overlaps which can enable continuous mapping of feature space as suggested by Tanaka.[50] This continuous mapping may contribute to the generation of object image in various views. Ideally, one should be able to predict transformed faces from only one view image (or test image) through past experiences. To deal with this problem, representation of different views needs to be generated from one particular view image, which requires substantial information. Yet, from observations and changed blindness experiments[51–53] internal representation of images in the brain is informationally impoverished — we do not have a full detailed image in our head when we visualize. For the current research work, dimension reduction is employed through principal component analysis (PCA).[54] PCA is used due to its simplicity and its ability to extract important components which is crucial for scarce storage. Besides, PCA is biologically supported by Oja learning rule that is a variation of the Hebbian learning rule.[55] Different view representation of faces is generated through geometrical transformation in this reduced space using only one canonical vector per person.

3. Method

Visual system requires fast feed-forward mechanism for rapid vision and feedbacks and modulators to extract relationships such as spatial links. The fast-feed-forward mechanism is to extract important information from the input image given a short presentation time within 100–200 ms. As had been explained in the literature, attention modulation, eye shifting and feedbacks cannot happen within this rapid time range. A simplified hierarchical architecture of simple and complex cells[13] will be employed for speed and to achieve 2D translation and scale invariance. For a more advanced model, refer to works by Serre *et al.*[37] There should also be an interface which connects the fragmented but information-rich signals from early visual stage with representations in higher level, which is impoverished but integrated. Besides, an appropriate dimension reduction is crucial such that 3D pose-invariance can be accounted for.

The work reported in this paper is on pose-invariant face recognition. The model used is based on the literature in Sec. 2. No feedback is performed for the work reported in this paper. The model can be divided into two sections, which is, the lower and higher level visual system. Lower level visual system assumes the processes that took place from LGN to the striate cortex (V1), while anything beyond V1 is considered higher level.

3.1. *Lower level visual system*

Early visual process is modeled according to the three-stage convolution system.[56] LGN performs convolution on the incoming signal with difference of Gaussian function.

Wavelets have been used for signal denoising effectively in various fields such as EEG-based diagnosis of neurological[57–64] and psychiatric disorders,[63,65–67] seismology,[68,69] intelligent transportation,[68,70–81,111] vibration control,[82–88] system identification,[89,90] health monitoring of structures[91] and image processing.[114] For V1, it is modeled using Gabor wavelet.[92] For the model, Gabor wavelet properties and its ensemble follows the over complete representation by Lee.[93] Gabor wavelet is used to model preliminary visual cortex due to biological and information theoretic reasons.[94] Gabor wavelet ψ used has the following properties:

- The aspect ratio of the elliptical Gaussian envelope is 2:1
- Wave's propagating direction along short axis of the envelope
- Bandwidth of 1.5 octaves. (Though for biological receptive fields, the bandwidth may span from 1 to 2 octaves)
- Zero mean

Where (x, y) is the center point of the wavelet, ω_0 the unit spatial frequency and θ the orientation.

$$\begin{aligned}&\psi(x, y, \omega_0, \theta)\\&\quad = \beta\left[\mathrm{e}^{i(\omega_0 x\cos\theta + \omega_0 y\sin\theta)} - \mathrm{e}^{-\frac{k^2}{2}}\right]\end{aligned}$$

$$\beta = \frac{\omega_0^2}{8k^2}\left[\mathrm{e}^{-\frac{\omega_0^2}{8k^2}(4(x\cos\theta+y\sin\theta)^2+(-x\sin\theta+y\cos\theta)^2)}\right]$$
$$k = \sqrt{2\ln 2}\left(\frac{2^{\phi}+1}{2^{\phi}-1}\right), \tag{1}$$

where ϕ is the bandwidth (1.5 octave for biological model).

Simple cell output is the Gabor coefficient value. A spatial area and a range of frequencies are connected to a single complex cell. The complex cell will then obtain the maximum value out of those simple cells connected to it. This is the MAX pooling method as described in the hierarchical architecture.[12,13] Biologically plausible suggestion of the MAX operation is provided by Yu *et al.*[95] Tests will be performed on different pooling methods, that is, using MAX method and the SUM method. SUM method means obtaining the linear summation value of the output of simple cells encompassed by the complex cell. MAX is a nonlinear approach, whereas SUM is a linear approach.[12] But for SUM, the output cannot be used to determine whether a certain feature is present within the receptive field. Feature specificity is lost due to contributions from all simple cells. For MAX, it can be determined since the maximum value of simple cells determines the output. There are cells in striate cortex[96] and inferior temporal[97] cortex that exhibits the MAX approach. For current work, MAX method is performed artificially just by choosing the maximum absolute value from the simple cells (since they are in complex value due to Gabor transformation). For a biological plausible approach, refer to Yu *et al.*[95]

3.2. *Higher level visual system*

Visible persistence only persists for a very short interval, after which it will decay or replaced by subsequent signals. But visible persistence is the only memory as discussed by Hollingworth[98] that supports visual phenomenology. Short and long term memory is too impoverished in terms of information to support this. Thus, there should be a block that establishes an interface between these two. The block stores the bases that reduce the dimension of the incoming information. The bases of choice are those that can reduce dimensions as much as possible, yet contain enough information to enable transformation for invariance recognition. The bases should extract important information from the information-rich yet fragmented signal from the primary visual area. The extracted information has very low dimension, which is crucial for higher level storage. They are informationally impoverished, but stable. They should also be capable of reconstructing back the original signal with high resemblance (visualization) such that an interface between the low and high visual area can be established. Visualization can be said a rough estimation of the actual face in high-dimensional space, but represented by information in low dimension. Thus, this provides a link between information-rich yet fragmented low level and impoverished but stable and integrated high level information. This acts as an object-centered mapping which enables cumulative learning through transformation of features that can contribute to methods as proposed in Ref. 99. At the moment, tests on visualization have yet to be performed, thus, not reported.

Principle components can be used to reconstruct images which are crucial in visualization. This is because unlike linear discriminant analysis (LDA) that extracts vectors for discrimination between classes, PCA extracts vectors that best describes the image. Scarcity of sample test images also pose a problem for LDA since different views need to be generated from only one test image. Low samples will reduce LDA's performance drastically.[100] Apart from that, LDA is unable to classify nonfrontal images given only frontal images.[101] Though it is not clear how these bases arise or, information processing wise, are the principle components or independent components actually been utilized, but for the current work, the purpose is to develop an invariance recognition framework that acts as a guide for further in-depth study. This is also supported by Oja learning rule,[55] which extracts principle components. Therefore, using PCA can provide insight into the interface between information-rich yet fragmented representation and integrated yet impoverished and abstract representation. High-dimensional space of the original 2D representation is used to interface with high-informational yet fragmented bottom-up signal. The low-dimensional space that is derived through PCA is used as abstraction and transformation. Therefore, this provides an interface between the fragmented yet information-rich representation and the impoverished yet integrated and stable representation.

Given the tensor $\bar{F}_{p\theta}(x)$ which contains all training facial images, where the subscript $p =$ individual identification index, $\theta =$ pose index, $N =$ spatial dimension (high dimension) and $M =$ number of individuals:

$$\bar{F}_{p\theta} = \begin{bmatrix} \bar{F}_{1\ \theta}(1) & \bar{F}_{2\ \theta}(1) & \dots & \bar{F}_{M\ \theta}(1) \\ \bar{F}_{1\ \theta}(2) & \bar{F}_{2\ \theta}(2) & & \\ \vdots & & \ddots & \\ \bar{F}_{1\ \theta}(N) & & & \bar{F}_{M\ \theta}(N) \end{bmatrix}. \tag{2}$$

For clarity, $\bar{F}_{1\ \theta}(2)$ is the gray scale scalar value at spatial location 2 of the person 1 with a facial pose of θ, Eq. (3) shows a reduced dimension tensor through PCA, $\bar{P}_{p\theta}$, where $n =$ number of reduced dimension:

$$\bar{P}_{p\theta} = \begin{bmatrix} \bar{P}_{1\ \theta}(1) & \bar{P}_{2\ \theta}(1) & \dots & \bar{P}_{M\ \theta}(1) \\ \bar{P}_{1\ \theta}(2) & \bar{P}_{2\ \theta}(2) & & \\ \vdots & & \ddots & \\ \bar{P}_{1\ \theta}(n) & & & \bar{P}_{M\ \theta}(n) \end{bmatrix}. \tag{3}$$

Transformation between poses is performed on the reduced dimension space $\bar{P}_{p\theta}$. For the current work, setting $n = M$, transformation $\bar{T}_{\theta1\ \theta2}$ between two poses (from θ_1 to θ_2) is directly obtained through:

$$T_{\theta1\ \theta2} = \begin{bmatrix} \bar{P}_{1\ \theta2}(1) & \bar{P}_{2\ \theta2}(1) & \dots & \bar{P}_{M\ \theta2}(1) \\ \bar{P}_{1\ \theta2}(2) & \bar{P}_{2\ \theta2}(2) & & \\ \vdots & & \ddots & \\ \bar{P}_{1\ \theta2}(n) & & & \bar{P}_{M\ \theta2}(n) \end{bmatrix} \times \begin{bmatrix} \bar{P}_{1\ \theta1}(1) & \bar{P}_{2\ \theta1}(1) & \dots & \bar{P}_{M\ \theta1}(1) \\ \bar{P}_{1\ \theta1}(2) & \bar{P}_{2\ \theta1}(2) & & \\ \vdots & & \ddots & \\ \bar{P}_{1\ \theta1}(n) & & & \bar{P}_{M\ \theta1}(n) \end{bmatrix}. \tag{4}$$

The obtained transformation matrix is overfitting. This means effect of noise is also equally emphasized as other more important features. A nonoverfitting transformation can be obtained as[102]:

$$T_{\theta1\ \theta2} = \bar{P}_{p\ \theta2}\bar{P}_{p\ \ \theta1}^{\ \mathrm{T}}(\bar{P}_{p\ \theta1}\bar{P}_{p\ \ \theta1}^{\ \mathrm{T}} + \alpha I)^{-1}, \tag{5}$$

where superscript $T =$ transpose, $\alpha =$ constant (determined through experiments) and $I =$ identity matrix.

But it is by no means being claimed that transformation is being performed as such by the brain. A more appropriate dynamical method needs to be found with biological support and implementable by cognitive wet-ware. This can be likened to developing a training method that trains perception toward the input image proposed in Ref. 103. Transformation provides invariance by generating a template that can properly interface with the input image according to its transformed state. Geometrically, the different state vectors of different transformations of an object meant the same thing, just that they are represented in different manifold. Manifolds are projections from the ambient space. Transformation matrix linked these manifolds together. Representation will be transformed to a canonical state first before the transmission. This fits the finding on mental rotation from Shepard[104] and Jolicoeur.[32] Reaction times depend on how much transformation is applied to achieve the original position. Palmer[31] found that test subjects generally have a canonical representation of an object they want to recognize.

For current work, the manifold is determined through inner product with the adjoint face vector of different poses and pick the winner. Adjoint vectors are vectors which are being "melted" with different face vectors of the same pose as proposed in Ref. 105.

Given a face represented by vector $v_\theta^I = [v_\theta^I(1)\ \ v_\theta^{\mathrm{T}}(2)\ \ \cdots\ \ v_\theta^I(n)]^{\mathrm{T}}$, where $\theta =$ pose and $I =$ person identifier index. All these vectors are combined to obtain the matrix:

$$V = \begin{bmatrix} v_1^1(1) & v_1^2(1) & \dots & v_1^M(1) & v_2^1(1) & v_2^2(1) & \cdots & v_{\bar{\theta}}^M(1) \\ v_1^1(2) & v_1^2(2) & \dots & v_1^M(2) & v_2^1(2) & v_2^2(2) & & v_{\bar{\theta}}^M(2) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ v_1^1(N) & v_1^2(N) & & v_1^M(N) & v_2^1(N) & v_2^2(N) & & v_{\bar{\theta}}^M(N) \end{bmatrix},$$

with $N =$ dimension of image representational space, $M =$ total persons and $\bar{\theta} =$ total pose. To obtain the pose of an image vector is to apply an operator to the image vector to produce the result of the pose. If the operation is linear, it means inner product with a set of vectors to obtain the order parameter which signifies the pose. Thus, the solution is to find a vector or a linear classifier that, when acted on the input vector, will produce a finite value with the magnitude depending on the represented component in the input vector. Directly using prototype image vectors as operators will cause cross-talk since covariant components are extracted, which is interdependent under nonorthogonal condition.

To prevent cross-talk, contra-variant component is extracted instead of covariant. This can be achieved by applying an appropriate metric tensor to the nonorthogonal operators. Contra-variant component allows reconstruction of the original vector through parallelogram summation given that the input vector can be fully represented by it. Given the matrix above, the contra-variant component can be obtained as follows:

$$S^{\alpha} = \widehat{I}(V^{\mathrm{T}}V)^{-1}V^{\mathrm{T}}X,$$

where S^{α} is the order parameter. More details can be obtained from Ref. 106. For distinct identification for every image residing in vector V, $\hat{I}$ is an identity matrix. But for pose detection,

$$\hat{I} = \begin{bmatrix} \hat{i} & 0 & & 0 \\ 0 & \hat{i} & & 0 \\ & & \ddots & \\ 0 & 0 & & \widehat{i} \end{bmatrix} \quad \text{with } \widehat{i} = \begin{bmatrix} 1 & 1 & & 1 \\ 1 & 1 & & 1 \\ & & \ddots & \\ 1 & 1 & & 1 \end{bmatrix},$$

with the number of rows for the square matrix $\hat{I}$ equals the number of poses p and the rows for the square matrix $\widehat{i}$ equals the number of persons m. $\hat{I}$ can be considered a correlation matrix, which determines how much any two vectors reside in Vcorrelate with each other. Here, any persons' faces with the same poses will be given maximum correlation. Therefore, a vector, which is a result of the combination of these faces (which is also called "melting"), is produced. These vectors act as linear classifier to determine the pose of a given face image.

After the pose is determined, transformation is performed to transform the input to the predetermined canonical pose.

Comparison between the input and the stored representations is performed through inner product, which produces an order parameter vector. Order parameter vector (containing M elements indicating M number of people) holds the similarity magnitude between the input and the people in the database regardless of pose. The winning face will be chosen through winner-take-all (meaning choosing the person that has the highest order parameter magnitude). To include temporal effects, evidence accumulation mechanism is employed. It is being considered that a population of cells provides evidence in terms of neuronal activity, which happens through time and space.[107] This is performed using nonlinear shunting network equation[108] with faster-than-linear signal function. The dynamic equation is as follows:

$$\begin{aligned}\frac{\partial x_i}{\partial t} &= -A_i x_i + (B_i - x_i)(I_i + S(x_i)) \\ &\quad - (x_i + C_i)\left(\sum_{j\neq i} I_j + \sum_{j\neq i} w_{ij} S(x_j)\right) + \tilde{N}(t),\end{aligned} \tag{6}$$

where

$I_i =$ order parameter input of the ith face which perturbs the system.
$A_i =$ decay factor of the internal state variable x_i.

B_i and C_i defines the upper and lower limit of the state variable x_i.

$w_{ij} =$ the feedback weights
$\widetilde{N}(t) =$ noisy fluctuation with 0 mean
$S(x_i) = ax^2$, which is a faster-than-linear signal function.

For this model, decay factor and upper/lower limits are assumed the same for all state variables x_i. A state variable is a symbolic representation of a person. w_{ij} is assumed to be identity such that all state variables have equal contribution.

The units handled by the shunting network represents one particular person (not a view-point of a person), which is pose-invariant. It can be likened to the view-invariant cells of inferior temporal cortex.

To handle 10 persons, $A = A_i = 1$, $B = B_i = 2$, $C = C_i = 0$ and $a = 10$. $\widetilde{N}(t)$ is a Gaussian noise with 0 mean and standard deviation of 0.5. Parameters are set through calibration. The motivation is for the dynamics to be able to perform winner-take-all. The attraction of the attractor should be strong enough to ignore unwanted signals such as noise and distortions from transformations, yet weak enough such that the dynamics will approach a new attractor given a sudden change of perturbation. More analysis is required on the equation since the size of storage requires different parameter settings for stable and optimum dynamics. Figure 1(a) shows the recognition performance over time (given a 7 unit translated and 10% noisy image). As shown, over time (iterations), information is constantly being collected which slowly drives the recognition magnitude of the correct person. Figure 1(b) shows a change in attractor when perturbation changes. In the case of face recognition, this means, a sudden change of view to another person.

The nonlinear shunting dynamic equation is currently too slow and requires more extensive analysis for higher number of faces. A more artificial but faster and more efficient method is employed to deal with high number of faces, which is:

$$\frac{\partial x_i}{\partial t} = a\left(\frac{V_i^2}{\|V^2\|} - \frac{V_i}{\|V\|}\right), \tag{7}$$

where

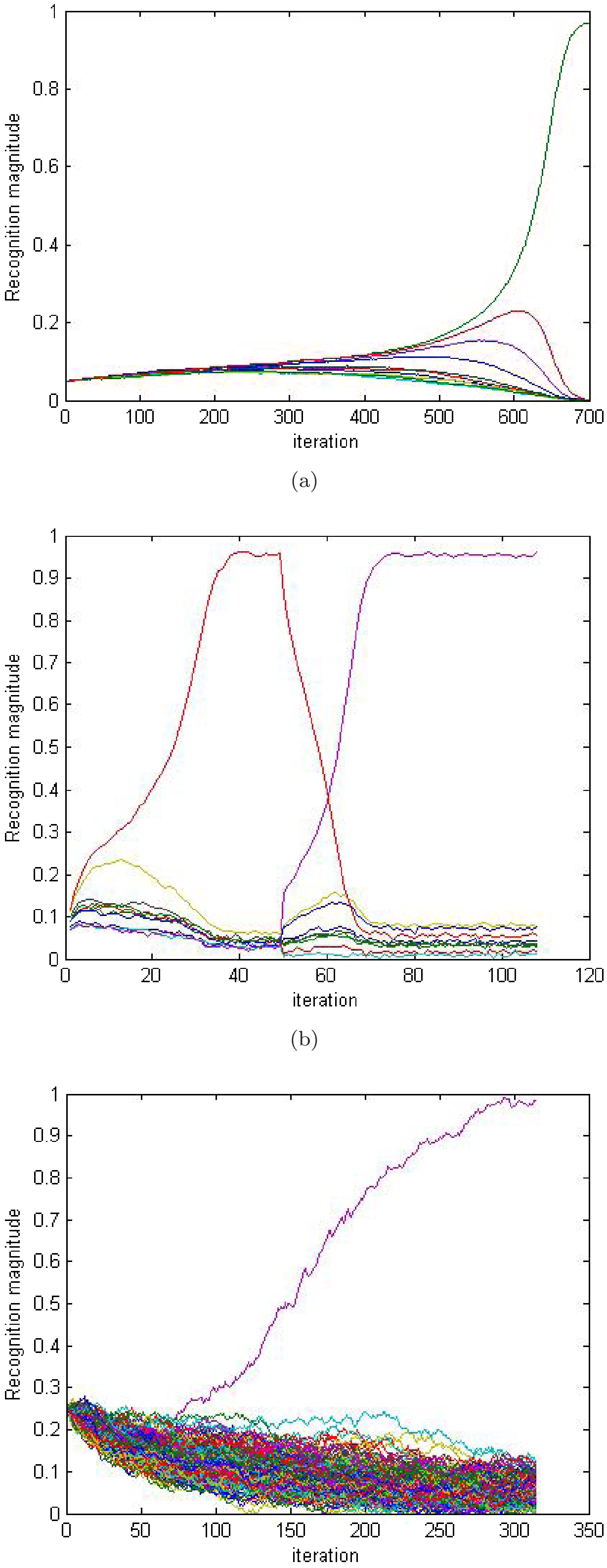
$$V_i = x_i I_i \quad \text{and} \quad \|V\| = \sqrt{\sum_i V_i^2}.$$

Figure 1(c) shows the dynamic of recognition process for 200 persons with 7 unit translation, 70% noise and random pose. The notable jaggedness is due to noise and high a value to hasten recognition process. Figure 2 shows the working diagram of the high level visual system of the model.

4. Results and Discussion

Tests are performed on the 2D translation and scaling invariance capabilities of the simple to complex cell structure, determination of pose through melting method, 3D pose-invariance capability of geometrical transformation of the principle components reduced space and information accumulation of the shunting network.

For the experiments, face database used is the facial recognition technology (FERET) face database. FERET database is used to obtain insight into real world face recognition problems, and also for comparison purposes with existing face recognition system which supports pose-invariance. It contains 200 subjects with different pose image each, which is a large number compared to other databases. This is important for principle components extraction, since low number of subjects can lead to lower generalization which is insufficient for representation of images in lower dimensional space. Images are cropped 62 units at the top and 66 units at the bottom of the image, making the image 256 × 256. It is then resized



(a)

(b)

(c)

Fig. 1. Dynamics of recognition process under shunting network and artificial network.

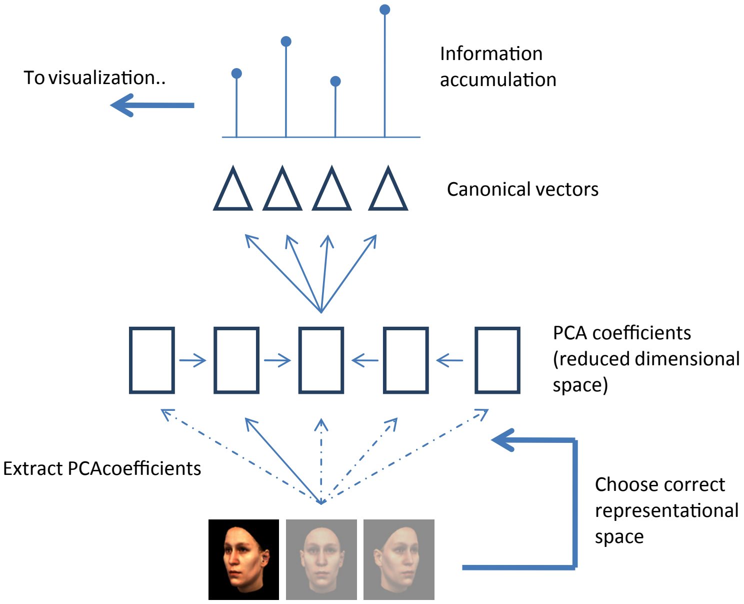


Fig. 2. Diagram of the high level visual system.

to 100 × 100. No face-centered adjustments are performed in order to test the effectiveness of the simple to complex cell architecture in solving translation and scaling invariance. Any other form of normalization (except converting the image vector to one unit length) is not performed. Figure 3 shows example of pre-processed (cropped and resized) faces of one pose to be used for experiment.

150 subjects will be randomly chosen for every test to be trained to obtain their principle components and transformation matrix. The remaining 50 will be used as testing subjects. For the experiments, every subject has seven images each representing a particular pose (left to right). Up to down pose is not performed, but the basics are the same. Emotion and illumination are assumed to be null. For every pose of every subject, only 1 image will be used for that particular pose per person for training. After training, there are seven representation spaces with six transformation matrices. For simple cells, the Gabor wavelets used are wavelets with 4 unit frequency (0.1 per pixel) with orientations 0, 0.4, 0.8, 1.2, 1.6, 2, 2.4 and 2.8 radians. The 0.1 per pixel spatial frequency is chosen empirically. Experiment on

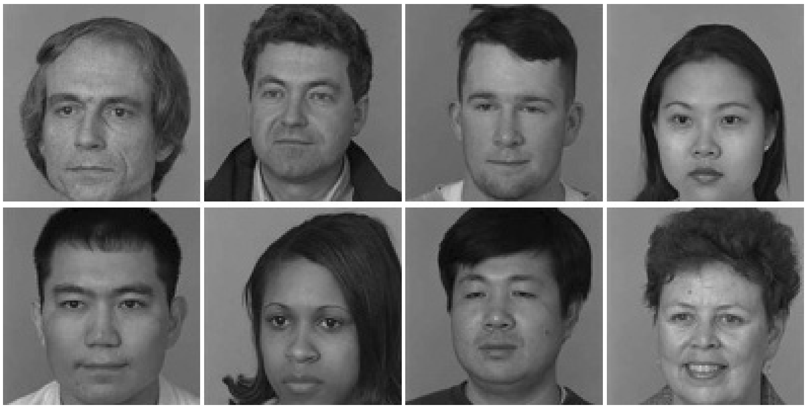


Fig. 3. Example of pre-processed faces for one particular pose.

recognition is performed (randomly pose-wise) using Gabor wavelet of spatial frequency from 1 (0.05 per pixel) to 9 units (0.3175 per pixel) frequency.

Tests are performed to determine the size of the receptive field L for the complex cell and the distance between two center points D that produces the best recognition result. Values for L are $3, 5, 7, 9, 11$ and 13, whereas values for D are 4, 5, 6, 7 and 8. For the test, 150 persons are randomly chosen for training. The remaining 50 are testing sets. The testing images are randomly translated using normal distribution with mean equal to 0 translation and standard deviation of 8. Poses for test subjects are randomly chosen.

Recognition rate is obtained through the average performance of all the tests. The optimum receptive field size is 7×7 with a distance of 5 units between two center points, which has a recognition rate of 0.72.

Experiment on recognition rate after transformation to various angles given a particular pose is performed. The experiment begins by defining a canonical representation space. Canonical representation space is where comparison between input and database image are performed. For example, if face representation at 0° pose is treated as the canonical space, then every other face with different pose needs to be transformed to this 0° space before recognition takes place. The input image is transformed to the canonical representation space such that evaluation is performed. (Example, 15 to 0 means 0° pose representation is the canonical representation, whereas the 15° pose is the input image). Results for the same absolute angles are averaged due to the assumption that face is symmetrical. Therefore, as an example, there's no transformation from -30 to 60°. For FERET, the angles are $+/-$ 0, 15, 25 and 40°. For each test, two methods of pooling of complex cells are carried out, which is the MAX method (extracting the maximum value of the simple cells encompassed by the complex cell) and SUM method (extracting the linear summation values of the simple cells encompassed by the complex cell).

Altogether, there are three tests being carried out, which are, recognition without geometrical transformation, recognition with gradual transformation (overfitting and nonoverfitting) and with direct transformation (overfitting and nonoverfitting). The parameter α from nonoverfitting transformation is obtained by performing a series of tests on different values of α, where the optimum value is chosen.

Recognition without transformation means no geometrical transformation is being carried out to transform the representational space to the canonical space. This is to test the contribution of input image to recognition rate without applying any algorithms. The motivation is to find out whether it is the transformation alone that contributes to higher recognition rate or due to other low level information.

Recognition with direct transformation and with gradual transformation both involves geometrical transformation. But the difference is that for gradual transformation, transformation is performed gradually from manifold to manifold. Geometrically, it means the path strictly occurs on the surface of the whole atlas of manifold. For direct transformation, direct transform from one manifold to the other can occur regardless of whether they are neighbors or not.

Gradual transformation is the desirable mode of transformation since information flow should be gradual and their relevance depends on how near their metric distance is to each other. In terms of pose generation in reduced space for face recognition, gradual transformation supports the concept that the greater the angle from the canonical pose, the lesser the relevant information is present for recognition. But this can be enhanced by gradual information accumulation, where accumulation occurs in terms of the current manifold the system is working on. Experiment on direct transformation is performed for comparison purposes. In ideal case, the matrix of direct transformation should be able to be obtained through operations of a chain of gradual transformation given that both have the same starting and ending points.

As can be seen from Table 1, when no transformation is applied, recognition rate is very low with a range of around 2% to 3%. This shows that no low level information is contributing to pose-invariant recognition. Except for a few anomalies which can be ignored, no preference in pose toward recognition is observed from the data. Relationship between recognition performance and distance between poses are not observed. This means information extraction prior to comparison is crucial since it is required for recognition tasks, as opposed to mere inner product. This also shows the effectiveness of the framework

Table 1. Recognition rate under no transformation.

Degree (°)	Max (%)	SUM (%)
15 to 0	2.0	2.0
25 to 0	2.4	2.0
40 to 0	4.0	2.4
0 to 15	2.2	1.8
25 to 15	2.2	1.6
40 to 15	1.4	2.0
0 to 25	6.0	6.0
15 to 25	1.8	2.2
40 to 25	2.4	2.4
0 to 40	3.6	4.0
15 to 40	4.0	2.6
25 to 40	2.0	2.2

(results shown subsequently), which does not rely on low level information such as global statistical information that is invariant to transformation.

Comparing gradual and direct overfitting transformation from Tables 2 and 3, both gives approximately the same performance. Gradual transformation works equally as well as the direct mode. Although more transformation is required for gradual approach, which is more likely to introduce distortion, from the test results, the distortion is negligible or nonexistent. This shows that local transformation is general and sufficient enough to embed all the manifolds together in the reduced dimensional space to achieve left–right pose-invariance. But in terms of application, gradual transformation is slower since every transformation takes up time regardless how small the transformation is. This problem can be significant if the number of poses within certain range of angle (pose resolution) increases.

Table 2. Recognition rate under gradual transformation with 150 training subjects and 50 testing images.

Degree (°)	Max (%)		SUM (%)	
	Nonoverfit	Overfit	Nonoverfit	Overfit
15 to 0	98.8	98.0	99.3	93.5
25 to 0	12.6	90.5	22.0	83.3
40 to 0	3.6	56.2	2.0	51.3
0 to 15	97.6	96.3	96.2	94.4
25 to 15	97.3	94.0	95.5	90.1
40 to 15	13.2	74.5	9.0	67.8
0 to 25	12.4	86.0	19.6	80.6
15 to 25	98.2	93.1	98.0	93.0
40 to 25	93.4	89.6	87.3	84.2
0 to 40	4.2	55.0	2.6	59.1
15 to 40	6.6	73.4	8.0	74.2
25 to 40	91.1	84.2	90.0	82.0

Table 3. Recognition rate under direct transformation with 150 training subjects and 50 testing images.

Degree (°)	Max (%)		SUM (%)	
	Nonoverfit	Overfit	Nonoverfit	Overfit
15 to 0	98.8	96.5	99.3	96.0
25 to 0	94.3	90.5	87.6	75.3
40 to 0	57.0	56.3	53.2	61.8
0 to 15	97.6	95.4	96.2	94.3
25 to 15	97.3	92.1	95.5	88.2
40 to 15	78.0	70.8	65.3	64.6
0 to 25	91.2	85.6	86.2	82.0
15 to 25	98.2	90.3	98.0	86.7
40 to 25	93.4	89.6	87.3	84.1
0 to 40	52.8	55.0	58.0	56.7
15 to 40	73.7	74.2	77.3	62.5
25 to 40	91.1	84.0	90.0	76.4

For nonoverfitting transformation, direct transformation gives an even better result compared to the former. But for gradual transformation, recognition is performed very badly. This may be due to some features that are crucial for continual transformation being regarded as noise during the extraction of nonoverfitting transformation matrix. More information theoretic work needs to be performed on this area to provide a clearer picture of the problem.

From Tables 2 and 3, when two poses are nearer to each other, recognition rate is high because it only requires minimal transformation. As for larger variation of poses, multiple transformations need to be performed to obtain the correct representation space, and therefore, can lead to distortion if the input image is not fully representable by the principle components. Another interpretation is that for larger pose difference, there is less correlated information as their representation manifolds are further from each other. Transformations that involves 40° pose have lower relative recognition rate. This may be due to inappropriate combination of Gabor wavelet or frequency, which might not extract sufficient information for transformation to other poses or vice versa. As stated by Keil,[109] spatial frequency and combination of Gabor wavelet depends on the task performed.

In terms of pooling method (MAX or SUM), MAX method gives a better result for FERET database as can be seen from the table. Therefore, MAX pooling method is applied to complex cells.

Gradual overfit, direct overfit and direct nonoverfitting transformation is then used with adjoint vector for pose-invariant test. As discussed before, adjoint vector is used to determine the pose, and thus, determining the starting manifold. It will then be transformed to the canonical representation. Adjoint vector is also obtained using the prior 150 training persons. Results are shown in Table 4. From the result, it can be seen that although pose detection accuracy with respect to the claimed angle associated with the face image is low, recognition rate is still high. Therefore, it is more appropriate to regard the adjoint vector as determiner of the appropriate transformation matrix. Results can be improved given more accurate pose detection.

Comparison of result at Table 5 is performed on Ref. 110, where classification is performed using virtual faces generated by one image. Coarse alignment of the eyes for the training and test images are performed. Input face image is first projected to a lower dimensional subspace. The transformed vector is then used for classification. Various classifiers are used for recognition. The best result is 100% for 0° pose, 83.5% for +/− 15° pose and 79.3% for +/− 25° pose. These are compared to results of the current work with 0° pose as the canonical representation.

Table 4. Pose-invariant recognition result.

Degree (°)	Gradual overfitting (%)	Direct overfitting (%)	Direct nonoverfitting (%)	Pose detection accuracy (%)
0 to 0	99.8	100	100	66.1
15 to 0	90.8	89.8	94.6	51.3
25 to 0	86.4	83.4	87.0	68.3
40 to 0	53.2	59.0	51.4	78.0
0 to 15	94.6	94.2	97.2	65.7
15 to 15	99.2	100	100	54.4
25 to 15	92.2	94.0	96.2	67.5
40 to 15	68.0	75.6	76.4	76.8
0 to 25	86.2	84.6	88.0	68.2
15 to 25	93.4	92.8	94.6	50.6
25 to 25	100	100	99.4	67.9
40 to 25	87.0	84.2	86.4	77.0
0 to 40	59.0	53.0	61.0	66.8
15 to 40	66.8	66.4	72.2	52.0
25 to 40	81.2	84.2	85.8	65.6
40 to 40	96.0	98.6	99.4	78.8

Table 5. Comparisons with other research works.

Methods	0° (%)	+/− 15° (%)	+/− 25° (%)
Sharma *et al.*[110]			
LDA (Belhumeur *et al.* 1997)	99	79.8	76.5
2D LDA (Kong *et al.* 2005)	100	81.3	76.8
PCA (Turk and Pentland 1991)	96.5	70.8	66.3
CLAFIC (Cevikalp *et al.* 2009)	100	71.0	65.5
CLAFIC-μ (Cevikalp *et al.* 2009)	100	73.3	67.8
RADON (Jadhav *et al.* 2009)	98	69.0	66.0
WRADON (Jadhav *et al.* 2009)	100	75.5	67.3
NFL (Pang *et al.* 2007)	100	80.0	76.3
ONFL (Pang *et al.* 2009)	100	83.5	79.3
Overfitting gradual transform	99.8	90.8	86.4
Overfitting direct transform	100	89.8	83.4
Nonoverfitting direct transform	100	94.6	87.0

In order to have a more in depth observation of the translation and scaling invariance property, test is performed to compare the performance between using MAX pooling and without it. Without using MAX pooling means using Gabor coefficients straight from simple cells, omitting pooling by complex cells. Procedures are same as previously discussed. Gradual overfitting transformation is used. Recognition rate average between transformation 15 and 0° and 25° and 0° is used for comparison. 40 to 0° test is not used due to bad recognition rate which might interfere with the comparisons. Figure 4 shows the result for translation and Fig. 5 shows result for scaling. Both results show improvements in translation and scaling invariance using MAX pooling. There is a slight drop in recognition rate at 0 translation and scaling for the test without using MAX pooling. This is due to noncentered faces which introduces slight translation and scaling. MAX pooling contributes by eliminating the effects of these slight changes.

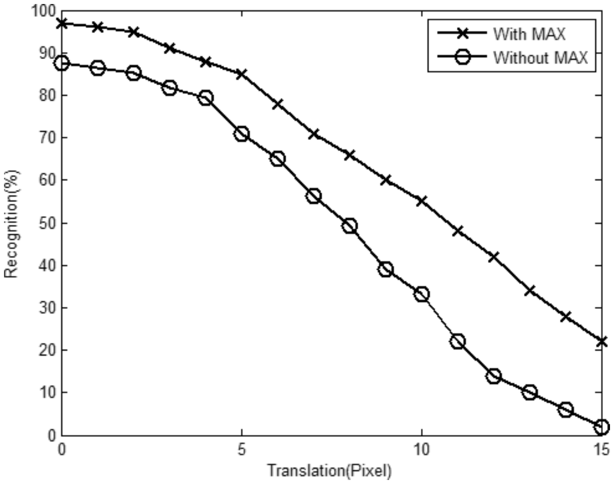


Fig. 4. Translation invariance with and without MAX pooling.

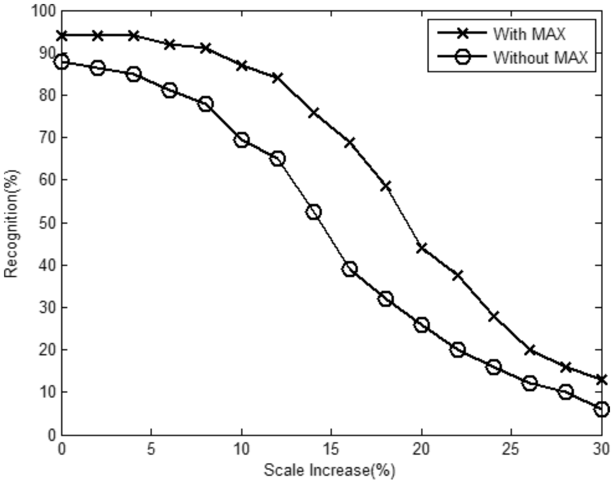


Fig. 5. Scaling invariance with and without MAX pooling.

5. Conclusion

Invariance is a subtle problem for face recognition. Yet, for biological visual system, invariance is easily achieved. For this research work, a translation, scaling and 3D pose-invariant (left–right) face recognition system has been built based on biological visual system with promising results. This can also act as a framework for further research in this field.

For early visual processes, suitable Gabor wavelets based on the work of Lee[93] are used to model simple cell receptive fields on striate cortex. Simple/complex hierarchical architecture[12,13] is employed to achieve translation and scaling invariance and also provide an explanation for rapid vision. MAX pooling of complex cells on simple cell will eliminate effects of translation and scaling up to a certain threshold. These early architectures are based on observations on biological visual system. From the results, MAX method performs better compared to SUM method.

For higher level visual system, our model is based on psychophysical evidences. From these observations, it is found that brain performs geometrical transformation to achieve pose-invariance (though unsure whether it is mental rotation or perspective transformation). Apart from this, information content from early visual stages is high but fragmented. But information content is low yet integrated for higher level visual stages (from V1 onwards). To account for these, for this model, transformation is performed on a reduced dimensional space where the components are the coefficients for principle components (which is supported by Oja learning rule). Therefore, a face is represented and processed in this low dimensional space.

Comparison and information accumulation are performed after the representation in reduced dimensional space is transformed to its equivalent canonical state. Information accumulation is performed using nonlinear shunting network with higher than linear transfer function to simulate winner take all. The representation element at this stage exists in an even lower dimension, as it only represents the identity of a face (regardless of the pose).

Recognition is performed without face detection, salient point detection and alignment. Calibration is performed to obtain optimum results. Based on the result obtained for gradual overfit, direct overfit and direct nonoverfit transformation using PCA, recognition performance (with translation, different scaling and poses) is very promising. Besides, MAX pooling provides more translation and scaling invariance.

But the tests are limited to 0, 15, 25 and 40° pose as provided by FERET database. What about the angles in between such as 10 and 22.6°? Given the current work on gradual transformation, we would like to suggest the construction of Lie group encompassing allowable transformations. But collecting the dataset for the construction of the group and methods of determining the parameters or coordinates of the Lie group are subjects of further research.

Acknowledgments

This work was supported by UMRG grant (RG059-11ICT) and HIR-MOHE grant (H-22001-00-B000010) of University of Malaya.

References

1. R. Riego, J. Otero and J. Ranilla, A low cost 3D human interface device using GPU-based optical flow algorithms, *Integr. Comput.-Aided Eng.* **18**(4) (2011) 391–400.
2. S. L. Hung and H. Adeli, Parallel backpropagation learning algorithms on Cray Y-MP8/864 supercomputer, *Neurocomputing* **5**(6) (1993) 287–302.
3. H. Adeli and S. L. Hung, An adaptive conjugate gradient learning algorithm for effective training of multilayer neural networks, *Appl. Math. Comput.* **62**(1) (1994) 81–102.
4. H. Adeli and S. L. Hung, *Machine Learning — Neural Networks, Genetic Algorithms, and Fuzzy Systems* (John Wiley and Sons, New York, 1995).
5. Q. D. Tran and P. Liatsis, Improving fusion with optimal weight selection in face recognition, *Integr. Comput.-Aided Eng.* **19**(3) (2012) 229–237.
6. Y. Tsai and Y. Huang, Automatic detection of deficient video log images using a histogram equity index and an adaptive gaussian mixture model, *Comput.-Aided Civ. Infrastruct. Eng.* **25**(7) (2010) 479–493.
7. S. J. Thorpe, Spike arrival times: A highly efficient coding scheme for neural networks, in *Parallel Processing in Neural Systems and Computers* (Elsevier, North-Holland, 1990), pp. 91–94.
8. S. J. Thorpe, D. Fize and C. Marlot, Speed of processing in the human visual system, *Nature* **381** (1996) 520–522.
9. C. Bundesen and A. Larsen, Visual transformation of size, *J. Exp. Psychol. Hum. Percep. Perform.* **1** (1975) 214–220.
10. T. D. Albright and C. G. Gross, Do inferior temporal cortex neurons encode shape by acting as fourier descriptor filters? in *Proc. Int. Conf. Fuzzy Logic and Neural Networks* (1990), pp. 375–378.
11. L. G. Ungerleider and J. V. Haxby, 'What' and 'where' in the human brain, *Curr. Opin. Neurobiol.* **4** (1994) 157–165.
12. M. Riesenhuber and T. Poggio, Hierarchical models of object recognition in cortex, *Nat. Neurosci.* **2**(11) (1999) 1019–1025.
13. M. Riesenhuber and T. Poggio, Models of object recognition, *Nat. Neurosci.* **3** (2000) 1199–1204.
14. T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman and T. Poggio, A theory of object recognition: Computations and circuits in the feedforward path of the ventral stream in primate visual cortex, AI memo 2005-036/CBCL memo 259, Artificial intelligence laboratory MIT, Cambridge, MA (2005).
15. A. Fazl, S. Grossberg and E. Mingolla, View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds, *Cogn. Psychol.* **58** (2009) 1–48.
16. D. I. Perrett and M. Oram, Neurophysiology of shape processing, *Image Vis. Comput.* **11** (1993) 317–333.
17. S. J. Thorpe and J. Gautrais, Rapid visual processing using spike asynchrony, *Adv. Neural Inf. Process. Syst.* (1997) 901–907.
18. G. Wallis and E. T. Rolls, A model of invariant object recognition in the visual system, *Progress Neurobiol.* **51** (1997) 167–194.
19. T. S. Lee and D. Mumford, Hierarchical Bayesian inference in the visual cortex, *J. Opt. Soc. Am.* **20**(7) (2003) 1434–1448.
20. I. Biederman and E. E. Cooper, Evidence for complete translational and reflectional invariance in visual object priming, *Perception* **20** (1991) 585–593.
21. C. S. Furmanski and S. A. Engel, Perceptual learning in object recognition: Object specificity and size invariance, *Vision Res.* **40**(5) (2000) 473–484.
22. L. Wiskott, How does our visual system achieve shift and size invariance? in *Problems in System Neuroscience* (Oxford University Press, 2003), pp. 322–341.
23. D. J. Felleman and D. C. vanEssen, Distributed hierarchical processing in the primate cerebral cortex, *Cereb. Cortex* **1** (1991) 1–47.
24. E. Kobatake, G. Wang and K. Tanaka, Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys, *J. Neurophysiol.* **80** (1998) 324–330.
25. M. J. Tov'ee, E. T. Rolls and P. Azzopardi, Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque, *J. Neurophysiol.* **72**(3) (1994) 1049–1060.
26. C. G. Gross, Representation of visual stimuli in inferior temporal cortex, *Philos. Trans. R. Soc. Lond.* **335** (1992) 3–10.
27. S. J. Thorpe and M. Fabre-Thorpe, Seeking categories in the brain, *Science* **291** (2001) 260–263.
28. M. C. Booth and E. T. Rolls, View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex, *Cereb. Cortex* **8** (1998) 510–523.
29. R. Desimone and C. G. Gross, Visual areas in the temporal cortex of the macaque, *Brain Res.* **178** (1979) 363–380.
30. M. W. Oram and D. I. Perrett, Modeling visual recognition from neurobiological constraints, *Neural Netw.* **7**(6) (1994) 945–972.

31. S. Palmer, The psychology of perceptual organization: A transformational approach in *Human and Machine Vision* (Academic Press, New York, 1983), pp. 269–339.
32. R. Shepard and J. Metzler, Mental rotation of three dimensional objects, *Science* **171**(3972) (1971) 701–703.
33. M. C. Potter, Meaning in visual search, *Science* **187** (1975) 565–566.
34. D. Ariely, Seeing sets: Representation by statistical properties, *Psychol. Sci.* **12** (2001) 157–162.
35. J. B. Debruille, F. Guillem and B. Renault, ERP and chronometry of face recognition: Following-up Seeck *et al.* and George *et al.*, *Neuroreport* **9** (1998) 3349–3353.
36. A. Oliva, Gist of the scene in *The Neurobiology of Attention* (Elsevier, San Diego, 2005), pp. 251–256.
37. T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio, Robust object recognition with cortex-like mechanisms, *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(3) (2007) 426–441.
38. N. K. Logothetis, J. Pauls, H. H. Bulthoff and T. Poggio, View-dependent object recognition by monkeys, *Curr. Biol.* **4**(5) (1994) 401–414.
39. N. V. Manyakov and M. M. Van Hulle, Decoding grating orientation from microelectrode array recordings in monkey cortical area V4, *Int. J. Neural Syst.* **20**(2) (2010) 95–108.
40. L. Ronan, R. Pienaar, G. Williams, E. Bullmore, T. J. Crow, N. Roberts, P. B. Jones, J. Suckling and P. C. Fletcher, Intrinsic curvature: A marker of millimeter-scale cortico-cortical connectivity? *Int. J. Neural Syst.* **21**(5) (2011) 351–366.
41. H. H. Bulthoff and S. Edelman, Psychophysical support for a two-dimensional view interpolation theory of object recognition, *Proc. Nat. Acad. Sci. USA* **89**(1) (1992) 60–64.
42. H. H. Bulthoff, S. Y. Edelman and M. J. Tarr, How are three-dimensional objects represented in the brain? *Cere. Cortex* **5**(3) (1995) 247–260.
43. M. Seibert and A. M. Waxman, Adaptive 3-D object recognition from multiple views, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(2) (1992) 107–124.
44. T. Vetter, A. Hurlbert and T. Poggio, View-based models of 3D object recognition: Invariance to imaging transformations, *Cereb. Cortex* **5**(3) (1995) 261–268.
45. D. J. Beymer, Face recognition under varying pose, A.I. Memo, 1461 Artificial intelligence lab, Cambridge, MA: MIT (1993).
46. T. Poggio and T. Vetter, Recognition and structure from one 2D model view: Observations on prototypes, object classes and symmetries, A.I. Memo 1347, Artificial intelligence laboratory, Cambridge, MA, MIT, 1992).
47. M. J. Tarr and S. Pinker, When does human object vision use a viewer-centred reference, *Psychol. Sci.* **2** (1990) 207–209.
48. M. J. Tarr and S. Pinker, Orientation-dependent mechanisms in shape-recognition-further issues, *Psychol. Sci.* **2** (1991) 207–209.
49. S. Ullman, Aligning pictorial descriptions: An approach to object recognition, *Cognition* **32** (1989) 193–254.
50. K. Tanaka, Inferotemporal cortex and object vision, *Annu. Rev. Neurosci.* **19** (1996) 109–139.
51. R. A. Rensink, The dynamic representation of scenes, *Visual Cogn.* **7** (2000) 17–42.
52. R. A. Rensink, The modeling and control of visual perception, in *Integrated Models of Cognitive Systems* (2007), pp. 132–148.
53. R. A. Rensink, Seeing seeing, *Psyche* **16** (2010) 68–78.
54. S. Ghosh-Dastidar, H. Adeli and N. Dadmehr, Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection, *IEEE Trans. Biomed. Eng.* **55**(2) (2008) 512–518.
55. E. Oja, A simplified neuron model as a principal component analyzer, *J. Math. Biol.* **15** (1982) 267–273.
56. K. H. Pribram, *Brain and Perception Holonomy and Structure in Figural Processing* (Lawrence Erlbaum Associates, Hillside, NJ, 1991).
57. H. Adeli, Z. Zhou and N. Dadmehr, Analysis of EEG records in an epileptic patient using wavelet transform, *J. Neurosci. Methods* **123**(1) (2003) 69–87.
58. H. Adeli, S. Ghosh-Dastidar and N. Dadmehr, A wavelet-chaos methodology for analysis of EEGs and EEG sub-bands to detect seizure and epilepsy, *IEEE Trans. Biomed. Eng.* **54**(2) (2007) 205–211.
59. S. Ghosh-Dastidar, H. Adeli and N. Dadmehr, Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection, *IEEE Trans. Biomed. Eng.* **54**(9) (2007) 1545–1551.
60. H. Adeli, S. Ghosh-Dastidar and N. Dadmehr, A spatio-temporal wavelet-chaos methodology for EEG-based diagnosis of Alzheimer's disease, *Neurosci. Lett.* **444**(2) (2008) 190–194.
61. M. Ahmadlou and H. Adeli, Wavelet-synchronization methodology: A new approach for EEG-based diagnosis of adhd, *Clin. EEG Neurosci.* **41**(1) (2010) 1–10.
62. A. Ahmadlou, H. Adeli and A. Adeli, Fractality and a wavelet-chao methodology for EEG-based diagnosis of Alzheimer's disease, *Alzheimer Dis. Assoc. Disord.* **25**(1) (2011) 85–92.
63. M. Ahmadlou and H. Adeli, Graph theoretical analysis of organization of functional brain networks in ADHD, *Clin. EEG Neurosci.* **43**(1) (2012) 5–13.

64. Z. Sankari, H. Adeli and A. Adeli, Wavelet coherence model for diagnosis of Alzheimer's disease, *Clin. EEG Neurosci.* **43**(4) (2012) 268–278.
65. M. Ahmadlou, H. Adeli and A. Adeli, New diagnostic EEG markers of the alzheimer's disease using visibility graph, *J. Neural Transm.* **117**(9) (2010) 1099–1109.
66. M. Ahmadlou, H. Adeli and A. Adeli, Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of Autistic spectrum disorder, *J. Clin. Neurophysiol.* **27**(5) (2010) 328–333.
67. M. Ahmadlou and H. Adeli, Fuzzy synchronization likelihood with application to attention-deficit/hyperactivity disorder, *Clin. EEG Neurosci.* **42**(1) (2011) 6–13.
68. Z. Zhou and H. Adeli, Time-frequency signal analysis of earthquake records using Mexican hat wavelets, *Comput.-Aided Civ. Infrastruct. Eng.* **18**(5) (2003) 379–389.
69. Z. Zhou and H. Adeli, Wavelet energy spectrum for time-frequency localization of earthquake energy, *Int. J. Imaging Syst. Technol.* **13**(2) (2003) 133–140.
70. H. Adeli and A. Samant, An adaptive conjugate gradient neural network — wavelet model for traffic incident detection, *Comput.-Aided Civ. Infrastruct. Eng.* **13**(4) (2000) 251–260.
71. A. Samant and H. Adeli, Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis, *Comput.-Aided Civ. Infrastruct. Eng.* **13**(4) (2000) 241–250.
72. A. Samant and H. Adeli, Enhancing neural network incident detection algorithms using wavelets, *Comput.-Aided Civ. Infrastruct. Eng.* **16**(4) (2001) 239–245.
73. A. Karim and H. Adeli, Incident detection algorithm using wavelet energy representation of traffic patterns, *J. Transp. Eng. ASCE* **128**(3) (2002) 232–242.
74. A. Karim and H. Adeli, Comparison of the fuzzy — wavelet RBFNN freeway incident detection model with the california algorithm, *J. Transp. Eng. ASCE* **128**(1) (2002) 21–30.
75. A. Karim and H. Adeli, Fast automatic incident detection on urban and rural freeways using the wavelet energy algorithm, *J. Transp. Eng. ASCE* **129**(1) (2003) 57–68.
76. H. Adeli and S. Ghosh-Dastidar, Mesoscopic-wavelet freeway work zone flow and congestion feature extraction model, *J. Transp. Eng. ASCE* **130**(1) (2004) 94–103.
77. X. Jiang and H. Adeli, Wavelet packet-autocorrelation function method for traffic flow pattern analysis, *Comput.-Aided Civ. Infrastruct. Eng.* **19**(5) (2004) 324–337.
78. X. Jiang and H. Adeli, Dynamic wavelet neural network model for traffic flow forecasting, *J. Transp. Eng. ASCE* **131**(10) (2005) 771–779.
79. S. Ghosh-Dastidar and H. Adeli, Neural network-wavelet micro-simulation model for delay and queue length estimation at freeway work zones, *J. Transp. Eng. ASCE* **132**(4) (2006) 331–341.
80. D. Boto-Giralda, F. J. Díaz-Pernas, D. González-Ortega, J. F. Díez-Higuera, M. Antón-Rodríguez and M. Martínez-Zarzuela, Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks, *Comput.-Aided Civ. Infrastruct. Eng.* **25**(7) (2010) 530–545.
81. B. Ghosh, B. Basu and M. O'Mahony, Random process model for traffic flow using a wavelet — Bayesian hierarchical technique, *Comput.-Aided Civ. Infrastruct. Eng.* **25**(8) (2010) 613–624.
82. H. Adeli and H. Kim, Wavelet-hybrid feedback least mean square algorithm for robust control of structures, *J. Struct. Eng. ASCE* **130**(1) (2004) 128–137.
83. H. Kim and H. Adeli, Hybrid control of smart structures using a novel wavelet-based algorithm, *Comput.-Aided Civ. Infrastruct. Eng.* **20**(1) (2005) 7–22.
84. H. Kim and H. Adeli, Wavelet hybrid feedback-LMS algorithm for robust control of cable-stayed bridges, *J. Bridge Eng. ASCE* **10**(2) (2005) 116–123.
85. H. Kim and H. Adeli, Hybrid control of irregular steel highrise building structures under seismic excitations, *Int. J. Numer. Methods Eng.* **63**(12) (2005) 1757–1774.
86. H. Kim and H. Adeli, Wind-induced motion control of 76-story benchmark building using the hybrid damper-tuned liquid column damper system, *J. Struct. Eng. ASCE* **131**(12) (2005) 1794–1802.
87. X. Jiang and H. Adeli, Dynamic fuzzy wavelet neuroemulator for nonlinear control of irregular highrise building structures, *Int. J. Numer. Methods Eng.* **74**(7) (2008) 1045–1066.
88. X. Jiang and H. Adeli, Neuro-genetic algorithm for nonlinear active control of highrise buildings, *Int. J. Numer. Methods Eng.* **75**(8) (2008) 770–786.
89. X. Jiang and H. Adeli, Dynamic wavelet neural network for nonlinear identification of highrise buildings, *Comput.-Aided Civ. Infrastruct. Eng.* **20**(5) (2005) 316–330.
90. H. Adeli and X. Jiang, Dynamic fuzzy wavelet neural network model for structural system identification, *J. Struct. Eng. ASCE* **132**(1) (2006) 102–111.
91. X. Jiang, S. Mahadevan and H. Adeli, Bayesian wavelet packet denoising for structural system identification, *Struct. Control Health Monit.* **14**(2) (2007) 333–356.

92. Z. He, X. You, L. Zhou, Y. Cheung and Y. Y. Tang, Writer identification using fractal dimension of wavelet subbands in gabor domain, *Integr. Comput.-Aided Eng.* **17**(2) (2010) 157–165.
93. T. S. Lee, Image representation using 2D Gabor wavelets, *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(10) (1996) 1–13.
94. N. W. Tay, C. K. Loo and M. Peruš, Application of Gabor wavelet in quantum holography for image recognition, *Int. J. Nanotechnol. Mol. Comput.* **2** (2010) 44–61.
95. A. J. Yu, M. A. Giese and T. Poggio, Biophysiologically plausible implementations of the maximum operation, *Neural Comput.* **14** (2002) 2857–2881.
96. K. Sakai and S. Tanaka, Spatial pooling in the second-order spatial structure of cortical complex cells, *Vision Res.* **40**(7) (2000) 855–871.
97. T. Sato, Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake monkeys, *Exp. Brain Res.* **77** (1989) 23–30.
98. A. Hollingworth, Visual memory for natural scenes: Evidence from change detection and visual search, *Vis. Cogn.* **14** (2006) 781–807.
99. H. Wersing, S. Kirstein, M. Götting, H. Brandl, M. Dunn, I. Mikhailova, C. Goerick, J. Steil, H. Ritter and E. Körner, Online learning of objects in a biologically motivated visual architecture, *Int. J. Neural Syst.* **17**(4) (2007) 219–230.
100. A. M. Martinez and A. C. Kak, PCA versus LDA, *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(2) (2001) 228–233.
101. D. L. Swets and J. Weng, Using discriminant eigenfeatures for image retrieval, *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(8) (1996) 831–836.
102. S. Lucey and T. Chen, A view-point invariant, sparsely registered, patch based, face verifier, *Int. J. Comput. Vis.* **80**(1) (2007) 58–71.
103. J. Menke and T. Martinez, Improving supervised learning by adapting the problem to the learner, *Int. J. Neural Syst.* **19**(1) (2009) 1–9.
104. P. Jolicoeur, Identification of disoriented objects: A dual system theory, *Mind Lang.* **5** (1990) 387–410.
105. R. W. Frischholz, F. G. Boebel and K. P. Spinnler, Face recognition with the synergetic computer, *International Conference on Applied Synergetics and Synergetic Engineering*, Erlagen, Germany (1994), pp. 100–106.
106. G. Resconi, C. K. Loo and N. W. Tay, Quantum morphogenetic system in image recognition, *International Joint Conference on Neural Networks* (2009), pp. 2642–2648.
107. D. Perret, M. Oram and E. Ashbridge, Evidence accumulation in cell populations responsive to faces: An account of generalization of recognition without mental transformations, *Cognition* **67** (1998) 111–145.
108. S. Grossberg, Non-linear neural networks: Principles, mechanisms and architectures, *Neural Netw.* **1**(1) (1988) 17–61.
109. M. S. Keil, “I look in your eyes, honey”: Internal face features induce spatial frequency preference for human face processing, *PLoS Comput. Biol.* **5**(3) (2009) 1–13.
110. A. Sharma, A. Dubey, P. Tripathi and V. Kumar, Pose invariant virtual classifiers from single training image using novel hybrid-eigenfaces, *Neurocomputing* **73** (2010) 1868–1880.
111. S. Ghosh-Dastidar and H. Adeli, Wavelet-clustering-neural network model for freeway incident detection, *Comput.-Aided Civ. Infrastruct. Eng.* **18**(5) (2003) 325–338.
112. X. Jiang and H. Adeli, Pseudospectra, MUSIC and-Dynamic wavelet neural network for damage detection of highrise buildings, *Int. J. Numer. Methods Eng.* **71**(5) (2007) 606–629.
113. M. Ahmadlou, H. Adeli and A. Adeli, Fuzzy synchronization likelihood-wavelet methodology for diagnosis of Autism spectrum disorder, *J. Neurosci. Methods*, in press (2012).
114. L. Ying and E. Salari, Beamlet transform based technique for pavement image processing and classification, *Comput.-Aided Civ. Infrastruct. Eng.* **25**(8) (2010) 572–580.