A dual fast and slow feature interaction in biologically inspired visual recognition of human action

Bardia Yousefi *, Chu Kiong Loo

Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Received 8 April 2015
Received in revised form 3 September 2017
Accepted 11 October 2017

Keywords:
Biologically inspired model
Interaction between ventral and dorsal stream
Human action recognition
Incremental slow feature analysis
Extreme learning machine
Dual processing pathways

ABSTRACT

Computational neuroscience studies have examined the human visual system through functional magnetic resonance imaging (fMRI) and identified a model where the mammalian brain pursues two independent pathways for recognizing biological movement tasks. On the one hand, the dorsal stream analyzes the motion information by applying optical flow, which considers the fast features. On the other hand, the ventral stream analyzes the form information with slow features. The proposed approach suggests that the motion perception of the human visual system comprises fast and slow feature interactions to identify biological movements. The form features in the visual system follow the application of the active basis model (ABM) with incremental slow feature analysis (IncSFA). Episodic observation is required to extract the slowest features, whereas the fast features update the processing of motion information in every frame. Applying IncSFA provides an opportunity to abstract human actions and use action prototypes. However, the fast features are obtained from the optical flow division, which gives an opportunity to interact with the system as the final recognition is performed through a combination of the optical flow and ABM-IncSFA information and through the application of kernel extreme learning machine. Applying IncSFA into the ventral stream and involving slow and fast features in the recognition mechanism are the major contributions of this research. The two human action datasets for benchmarking (KTH and Weizmann) and the results highlight the promising performance of this approach in model modification.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

A biologically inspired mechanism for human action recognition provides a new horizon in the fields of computer vision and video processing. Neurophysiological, physiological, and psychophysical evidence suggests two independent pathways in the system configuration where motion and form information are provided. The slowness principle provides a more stable form information than the current approaches [1–6] because of the independent behavior of temporal variations. In the motion pathway (dorsal stream), the motion information (dynamic pattern) is analyzed locally or globally [7–12] and is extracted by an optical flow that creates fast features. In the form pathway (ventral stream), the active basis model (ABM) [13] based incremental slow feature analysis (IncSFA) [14] generates slow features corresponding to the form of movements. The interaction between these two pathways defines the combination of information for making a recognition decision. Biologically inspired models have been analyzed in many research fields, such as artificial intelligence (e.g., early detection of the action [15] or fusion of homogeneous convolutional neural network classifiers [16]), soft computing (e.g., fuzzy shortest path problems [17]), computer vision (e.g., human action recognition system with projection-based metacognitive learning classifier [18]), and computational neuroscience, all of which follow the biological evidence. This research also follows the same evidence to develop a model for recognizing biological movements.

The number of temporal illusions, such as time perception distortions [19], bind synchronization of object features [20], and perception of time in object motion [21], has substantially increased. These illusions and the functional organization of visual systems are generally understood through temporal limits [22]. Multiple temporal resolution is caused by the temporal interval within video frames that can be considered for the perception of biological movements. Biological movements are recognized despite the variability in the conditions of lighting, human object locations, or even perspectives in time. Such feature represents an ability for the system [23] and a demand for devising a robust technique against the fast-varying parameters. In the primary sen-
sory system, even a minimal signal variation in the human object in time may generate different stimuli in the recognition results. Thus, considering the changes in time, the brain must produce different stimuli for representing the underlying cause, thereby creating an inner representation that does not change when irrelevant variations occur across time. The proposed approach in this work attempts to answer the question of how such unchanged representations can be established and influence the recognition of biological movements?

Invariant representations can be successfully learned via the input sensory adaptation to the extraction of slow-varying features. The temporal stability in the slowness principle is fundamental in the creation of slow features [23–25]. The applications of this approach have concentrated on visual system models that are predominantly focused on the self-organized configuration of complex cell receptive fields in the primary visual cortex [26,27]. Given that biological movement is an understandable task in the mechanism, the recognition of such actions is related to temporal movements and their visual understanding in the human brain. The proposed modification does not only analyze the slow or fast features but also strives to employ psychophysical and neurophysiological evidence for showing the independency of the pathways [1,5,28,6,29,30].

The proposed approach applies ABM followed by the slowness principle combined with motion information, which provides a different perspective toward the original model for recognizing biological movements [1]. The modification in the model occurs through the slow and fast features in the ventral and dorsal processing streams. The new development in the modeling of the ventral stream utilizes the slowness principle for the active basis calculated from the human object form. The two pathways interact in the decision-making part where significant results in human action recognition are obtained. However, each pathway separately facilitates the recognition of biological movements with a considerable disparity rate. One of a concrete evidence for this is the performances of two patients were analyzed, including DF who developed visual agnosia (i.e., damage to the ventrolateral occipital) and RV who developed optic ataxia (i.e., damage to the occipitoparietal cortex) [31]. The proposed approach contributes to the integration of the slowness principle into the model and the introduction of the slow and fast features concept (static and dynamic patterns [30]). This article is organized as follows. The next section presents the biological motivation. Sections 3 and 4 discuss our methodology and evaluates the simulation results and accuracy of the system, respectively. Section 5 presents the discussion, while Section 6 concludes the paper.

2. Biological motivation

The mechanism for recognizing biological movements can be split into two pre-processing streams [33–37] that are parallel to the dorsal and ventral streams, which are specified for analyzing motion and form information, respectively. Two streams have neural motion detectors, and form feature extraction hierarchically enables an in-dependency in size and style in both pathways and a classification of generated features from both feed-forward pathways to categorize biological movements. The results for stationary biological motion recognition indicate that discrimination can be accomplished through small latencies, thereby highlighting the important role of top-down unlikely signals [34]. The body shapes are determined through a set of patterns similar to sequences of “snapshot” [1], that maintain constant features throughout the entire action episode. The proposed approach expands upon an earlier model used for stationary object recognition [1–4,37,38] by adding and combining the temporal information in the pathways. This method can determine the appropriate quantity of existing information for organizing, summarizing, and interpreting by referring to data from neuro-physiological studies. The proposed approach quantitatively develops the original model for temporal analysis and computer simulations with respect to the previous model architecture. The slow and fast features are not used as criteria for measuring speed; rather, they are used to measure the rate of variation in the features. The motion pathway represents the function of the dorsal stream in the mechanism and involves information related to optical flow, which has a fast temporal variation. Fast-varying features (optical flow features) refer to the short changes between Frame(t) and Frame(t+1) rather than to the entire episode (such as the ABM-based IncSFA in the form pathway). The local detector of optical flow is connected to motion patterns, and the model is composed of a population of directed neurons in the area of MT. However, MT and V4 have a connection for motion and direction selection. The motion edge selectors in two opposite directions are present in the areas of MT, MSTd, and MSTI [39,40], the multiple regions of dorsal streams (most likely in the kinetic occipital area (KO) [1] and the motion selective edges that are similar to MT [39] and MSTI [40] in macaque monkeys. Few plausible models have been proposed for human body shape recognition that are considered appropriate for the neurophysiological recognition of the stationary form [37]. The proposed approach uses ABM [13], which inspired from V1-like Gabor filters in the luminance image (in V1) that has a contrasting normalization of filter outputs and summation of energy (by SUM-MAX in ABM) [41]. The addition of ABM, which is a supervised method that uses Gabor filters, into the mechanism [5,28,42,6,29] increases the constancy in simple cells [43]. The complex cells in V1, V2, or V4 are invariant in terms of the position of varying responses [1] whereas size independency is typically present in V4. V2 and V4 are more selective for difficult form features (e.g., junctions and corners) but are not suitable for motion recognition because of the temporal dependency in these two pathways. The snapshot detectors are used to identify the shape models that are similar to the area of the inferotemporal cortex (IT; Fig. 1) of monkeys where the view-tuned neurons are located and the model of complex shapes is tuned [39]. Snapshot neurons are similar to the view-tuned neurons in the area of IT and provide an independent scale and position. Previous models have used Gaussian radial basis functions for modeling and have been adjusted during training where they generate a key frame for the training sequences. This key frame can consider fast features (motion pathways) as it has been defined in time frames rather than episodic changes. Slowness features, which have Gabor-like features, are highly representative of the form information of biological movements. The proposed approach strongly suggests that motion contains spontaneous temporal variations and that form features are temporally slow-variation-ordered information throughout the movements. The novelty of this approach is more pronounced in computational intelligence and soft computing than in other relevant fields. Our development in the model follows the previous models and modifies the form pathway by applying ABM-based IncSFA as explained in the method Section. The computational simulations and testing method are presented in the results Section. The biological motion perception in the human visual system highlights an association between fast and slow features, which allows the recognition of biological movements.

3. Method

The proposed approach modifies the recognition mechanism of a biologically inspired model [1] in the visual system and aggregates two types of information, namely, form information (slow features) and motion information (fast features). In the experiment, approximately 38,000 cuboid frames of human movements were analyzed
by using the proposed model. The response of each input was estimated throughout the time series data by applying IncSFA for slowness and combining these data with motion information. The proposed method contributes to the fields of computational vision and intelligence. However, the biological concept and some related evidence have vital roles in modifying the computational part of our methodology due to the link between the biological visual system and the proposed mechanism. After discussing ABM, IncSFA is described along with the method for merging ABM and IncSFA in the form pathway. The optical flow and its division approach for the interaction between pathways are explained following the kernel extreme learning machine (ELM) brief for the decision-making step in the recognition process.

3.1. Active basis model

The ABM [13] uses the Gabor wavelets (for the elements dictionary) consisting of a deformable biological template. A shared sketch algorithm (SSA) is followed through AdaBoost. In each iteration, SSA follows the matching pursuit chooses an element of the wavelet. It checks the object’s number in different orientations, locations and scales. By selecting a small number of elements from the dictionary for every image (Sparse coding), there can be a representation of the image using a linear combination of previously described elements by considering U as a minor residual.

\[ I = \sum_{i=1}^{n} c_i \beta_i + \epsilon \]  

Where \( \beta=(\beta_i, i=1, \ldots, n) \) is a set of Gabor Wavelet elements and components of sinusoid, \( c_i \) and \( \epsilon \) is unsolved image coefficient [13]. Using wavelet sparse coding a large number of pixels reduces to a small number of wavelet elements. Sparse coding can train natural patches of an image to a Gabor like wavelet element dictionary, which carries the simple cells in V1 properties [44]. The extraction of local shapes is separately performed for every frame [13] and filter computes the images considering the orientation and density. Also, ABM uses the Gabor filter bank but in a different form. A Gabor wavelets dictionary, comprising \( n \) directions and \( m \) scales is in the form of, \( GW(\theta, \omega), j=1, \ldots, m \times n \). Where, \( \theta \in \left[ \frac{\pi}{2}, k = 0, \ldots, n - 1 \right] \) and \( \omega = \left( \frac{2\pi}{k^2}, i = 1, \ldots, m \right) \). The object forms contain the small variance in size, location and posture. Though overall shape structure, it is considered to be maintained throughout the
process of recognition. Response (convolution) to each element offers form information with \( \theta \) and \( \omega \).

\[
B = \langle GW, I \rangle = \sum_{i=1}^{n} GW(x_0 - x, y_0 - y; \theta_i, \omega_i)l(x, y).
\]

\( GW \) is a \([x_g, y_g] \times [x_l, y_l] \) matrix, and the response of \( I \) to \( GW \) is \([x_l + x_g, y_l + y_g] \). Therefore, the previous convolution of both matrices must be padded through sufficient zeroes. Cropping the result can eliminate the consequences of convolution. An additional approach is to shift the center of the frequencies (zero frequency) back to the center of the image, although this process might result in lost data. For the obtained training image set \( \{I_m, m = 1, \ldots, M\} \), the joint sketch algorithm consecutively chooses \( B_l \) It fundamentally identifies \( B_l \) so that its edge segments obtained from \( I_m \) become maximal [45]. It is then necessary to compute \( I_m, \beta = |\psi | |I_m, \beta |^2 \) for different \( i \) where \( \beta \in \text{Dictionary} \) and \( \psi \) represents sigmoid, whitening and thresholding transformations. Then \( I_m, \beta \) will be maximized for all possible \( \beta \). Let \( \beta = (\beta_i, i = 1, \ldots, n) \) be the template; for every training image, \( I_m \) scoring will be based on:

\[
M(I_m, \theta) = \sum_{i=1}^{n} \delta_i | I_m, \beta_i | - \log \Phi(\lambda \delta_i).
\]

\( M \) is the match scoring function, \( \delta_i \) from \( \sum_{m=1}^{M} | I_m, \beta_i | \) addresses the steps selection and \( \Phi \) is a nonlinear function. The logarithmic likelihood relationship of the exponential model was attained from the template matching score. Vectors of the weight were calculated by the maximum likelihood technique and are revealed by \( \Delta = (\delta_i, i = 1, \ldots, n) \) [45].

\[
\max(x, y) = \max_{(x,y) \in D} M(I_m, \beta).
\]

\( \max(x, y) \) calculates the maximum matching score previously obtained and \( D \) represents the lattice of \( I \). Here, there is no summation because of updating the size based on the training system for frame \((t - 1) \). Moreover, the method tracks the object applying motion feature for obtaining displacement of the moving object.

### 3.2. ABM and Gabor kernel

ABM is a supervised learning Gabor wavelet model that has been successfully used for object recognition tasks [13]. Previously Gabor wavelet has been widely used as a kernel in various applications, e.g., face recognition [47] or even action recognition model [2–4].

Gabor wavelet (kernels filter) has been defined in previous works [47–50] as (Fig. 3):

\[
\psi_{\mu, \nu}(z) = \frac{|\kappa_{\mu, \nu}|^2}{\sigma^2} e^{i\kappa_{\mu, \nu}||z||^2/2\sigma^2} e^{i\kappa_{\mu, \nu}z - e^{i(\sigma^2/2)}}
\]

where \( \mu \) and \( \nu \) are the orientation and scale, respectively, of the Gabor kernels, \( z = (x, y), || \cdot || \) is the norm operator, and the wave vector \( \kappa_{\mu, \nu} \) is defined as follows:

\[
\kappa_{\mu, \nu} = \kappa_{\mu} \varepsilon ^{\mu/8}, \kappa_{\mu} = \pi \kappa_{\max} \text{is the frequency of the maximum, and } \varepsilon \text{ is the spacing factor between kernels in the frequency domain [51].}
\]

Unlike the Gabor wavelet kernel that was required to define the scales and orientation of the Gabor beams, these parameters in ABM attains by training. It represents the image obtained by the summation of the active basis families, which are obtained by dictionary and matching scoring function. Let \( l(x, y) \) be the gray level distribution of an image; the image convolution \( l \) and a Gabor kernel \( \psi_{\mu, \nu} \) are defined as follows:

\[
B_{\mu, \nu} = \psi_{\mu, \nu}(z) \ast l(z), i = 1, 2, \ldots, n
\]

where \( z = (x, y) \) denotes the convolution operator, and \( B_{\mu, \nu} \) is the active basis that corresponds to matching scoring at the proper orientation and scale. Consequently, the set \( S = \{B_{\mu, \nu}, \nu \in M, \nu \in O \} \) forms the Gabor wavelet representation of the image \( l(z) \) along with \( M \) and \( O \) which represents the orientations and scales of the Gabor wavelet dictionary. To include the various spatial localities, spatial frequencies (scales) and orientation selectivity, we concentrated on all depiction results and obtained a supplemented feature vector. \( X \) defines a set of active basis that have the highest matching scores based on the training sets that were used together to make the object form. This method prefers the integration of simple cells to make complex cells (Fig. 5).

### 3.3. Slowness principle for ventral processing streams

The perception of SFA relates the high-dimensional input data (2D signal sequence) and an abstraction of the entire stream to an image [52]. The input signal usually shows some variations in the environment with varying lighting conditions or noise levels, and the extraction of information from the sequence seems to be a challenging task. Slowness features can extract video attributes that slightly vary over time. These features have been recently entered in computer vision [52,32,53] and are typically connected to the visual cortex [54,55]. An incremental learning algorithm is used due to the application of SFA for each time step in an unknown video input. The incremental principal component analysis (IncPCA) is closely related to IncSFA [41,44] because the latter applies PCA and minor component analysis (MCA). Slowness features contain information regarding the active basis from multidimensional input, while fast features can recognize the biological movement task. However, the active basis that is extracted by ABM and IncSFA only summarizes the sequence. SFA provides instantaneous scalar input-output functions that generate a 2D signal output which carries important information that changes as slowly as possible. SFA is an unsupervised learning method that extracts slowly changing features. This method provides a characteristic representation of the entire set and eliminates the unrelated details selected by the sensors [32,56,54]. Given that a high-dimensional video is used as input, an otherwise stationary room may be searched and the data may be enhanced by combining the situation and direction with slow features [57]. SFA is typically concerned with the optimization of complexity; it is common that for the identification of \( x(t) \) as an input by the \( D \) dimension, \( x(t) = [x_1(t), \ldots, x_D(t)] \), there is a set of functions similar to \( f(x) \) that have \( L \) dimension, \( g(t) = [g_1(t), \ldots, g_L(t)] \), or that can produce the output for \( L \) dimension as \( y(t) \) so that \( y(t) = [y_1(t), \ldots, y_L(t)] \). Thus, the relationship between these sets is \( y(t) := g(t) \).

\[
\Delta_{i} = \Delta(y_i) = (y^2_i) \text{ is minimal}
\]

\[
(y_i) = 0, \text{(Zero mean)}.
\]

\[
(y_i^2) = 1, \text{(Unit variance)}.
\]

\[
\forall d < 1 : y_{d}y_{d} = 0, \text{(De – correlation and order)}.
\]

The general definitions, similar to (8) and (9), are the restrictions for having insignificant constants in the output, and (10) is for de-correlation restrictions for features that are the same. A representation of the evaluation for the derivative of \( y \) and the sequential average are considered, correspondingly. The problem will be defined by identifying the \( f(x) \) for generating the slow varying output. It is noticeable that for the solution of this problem, the optimization of variation calculus, similar to [19] is not applicable, but it is simple especially for the eigenvector method. Considering that \( \mu \) is constrained to be a linear function that consists of a com-
Features of Johnsson shown.

Then, we will have \( z(t) = f(x(t)) \). Based on the changes previously incorporated, the optimization problem will be introduced by minimizing (12), the \( w_1 \):

\[
\Delta(y_1) = (y_1^2) = w_1^T (zz^T) w_1.
\]

If the \( p \) functions are selected such that \( z \) has a unit of covariance matrix and a zero mean, then three restrictions will be satisfied, only if the weight vectors have an orthonormal difference. Whitening is a very common technique that is used for identifying \( p \). For whitening, the PCA of the input data is required to be calculated; considering the zero mean and the individuality covariance matrix, put the \( x \) to \( z \) and by this \( z \), SFA problem will be converted to the linear problem. Eq. (12) should be considered for minimizing the L-normed set of eigenvectors of \( (zz^T) \). The desired features will be obtained from the set of PCs of \( z \). The objective was to calculate the temporal slowness, \( \delta \)-value, features and \( (x) \) as instantaneous functions of the input 2D-signal. This eigenvector-based algorithm is guaranteed to obtain the global optimum and learn biologically plausible rules for the existing optimization problem [54,58,59]. The modified optimization problem for high-dimensional visual input uses the information of biological movements and the human

---

**Fig. 2.** A different approach to present the hierarchical model in terms of the theory and computation of the form (slow features) and motion (fast features) information are shown. ABM provides this property with human objects and the computation of slow features are performed with IncSFA, which is episodic approach and provides slow features of basis within the episode that were combined with optical flow division information. It creates an interaction between the pathways.

**Fig. 3.** The qualitative comparison of functional imaging experiments to ABM is revealed in the figure. The biological movements according to the research experiments of Johnson [46]: 10 light bulbs were located on the joints, and the actor was recorded performing complex movements. There is recognition of the action within the action episodes. In addition, the dots were spontaneously interpreted as a human. Similar to the point light technique, which presents static pictures, ABM has a good representation of biological movements, and adding it into IncSFA can be a strong tool for selection of best representative of basis in form pathway.
object through ABM. This pathway information is then combined with fast features with respect to the original model [1–4].

3.4. ABM-based IncSFA

SFA is a non-linear decomposition method that is solved by twofold PCA and is entirely based on second-order statistics. Bray and Martinez [60] presented the kernel based on the temporal slowness principle by using the objective function of Stone [61], which is, in some ways, dissimilar to SFA. The SFA solution is applied through PCA and MCA and is closely related to IncPCA [45,62,32]. Schölkopf et al. [63] proposed a kernel-PCA and proposed that the kernelized IncPCA must be considered in the case of IncSFA [64]. ABM can be applied as a subset of the Gabor filter for IncSFA to find the best representative basis of each action.

3.5. Motion information in the dorsal stream and interaction between pathways

In the motion pathway, biological movements are recognized by the dynamic patterns of optical flow. The optical flow identifies the movement pattern, which is consistent with the neuropysiological information from the hierarchy of neural detectors [1]. Some neurons for motion and direction selection can be found in the MT and V1 areas, respectively.

This motion information extracted by applying layer-wise optical flow estimation. A mask that reveals the visibility of each layer mainly accounts for the difference in the estimations of traditional and layer-wise optical flows. The mask shape may be fractal and arbitrary, and matching only applies for those pixels that fall inside the mask [65]. We use the layer-wise optical flow method in [52] which has a previously described baseline optical flow algorithm [33–35]. M1 and M2 are visible masks for two frames t1(t) and t2(t−1), and the field of flow from t1 to t2 and t2 to t1 are represent by (u1, v1), (u2, v2). The following terms will be considered for layer-wise optical flow estimation. The objective function consists of summing three parts: visible layer masks then match to these two images using a Gaussian filter and are called data term matching E_{D}\text{sym} and smoothness E_{\mu}.

\[
E(u_1, v_1, u_2, v_2) = \sum_{i=1}^{2} (E_{D}^{(i)} + \rho E_{\mu}^{(i)}).
\]

We achieved a bidirectional flow after optimizing the objective function, using outer and inner fixed-point iterations, applying image warping, and performing a coarse-to-fine search process. The compressed optic flow for all frames was calculated by straight matching the template to the earlier frame through the summation of absolute difference (l1-norm). Although optic flow is particularly noisy, no smoothing techniques have been performed using this flow because the field of flow becomes blurred in gaps, especially in those locations where the motion information is significant [32]. To determine the proper response of the optical flow with regard to its application in the proposed model, the optical flow will be applied to adjust ABM and increase its efficiency. To achieve a reliable representation through the form pathway, the optic flow estimates the velocity and flow direction. The response of the filter based on the local matching of velocity and direction will be maximal as these two parameters are continuously changing. Optical flow division has been used for analyzing the interaction between the motion and form information [6]. The information of the optical flow division is also added to the representative of each action as calculated by ABM-based IncSFA. Such interaction confirms both pathways in each biological movement before this information is applied in the final decision making procedure using ELM.

3.6. Extreme learning machine (ELM)

Artificial neural networks (ANN) have been widely utilized in several research areas because of their capability to estimate difficult nonlinear mappings straight from the input sample and their ability to offer models for a large class of artificial and natural phenomena that are difficult to model using classical parametric techniques. Huang et al. (2004) proposed a novel algorithm for learning, namely, a single layer feed-forward neural network structure called ELM [66–68]. This algorithm solves those problems that are initiated by those algorithms that use gradient descent (e.g., the back propagation used in ANNs). ELM considerably diminishes the amount of time required for training in the neural network, shows a faster learning and generalization performance, requires fewer human interventions, and can run significantly faster compared with conventional techniques. ELM also routinely concludes the parameters of the entire network, thereby preventing unimportant external interventions from humans, and is highly effective in real-time applications. Several advantages of ELM include its simple usage, fast learning speed, excellent generalization performance, appropriateness for several nonlinear kernel functions, and activation function [69]. The single hidden layer feed-forward neural network (SLFN) function with hidden nodes [70,71] can be reflected on the mathematical explanation of SLFN, which integrates additive and Sigmoid hidden nodes altogether in a joint method.

\[
f_L(x) = \sum_{i=1}^{L} \beta_i G(s_i, b_i, x) x^T \in \mathbb{R}^d, a_i \in \mathbb{R}^d
\]

Let a and b represent the parameters of learning in hidden nodes and β represent the connecting weight of the ith output node of the hidden node. G(s, b, x) is the output of the ith hidden node with respect to the input x. For the additive hidden node with the activation function G(x) : R → R (e.g., sigmoid and threshold), G(s, b, x) is provided by G(a, b, x) = g(a + b^T x) b ∈ R. Let α represent the connecting weight vector of the input layer to the ith hidden node and b, represent the ith hidden node bias. For N, arbitrary different examples are indicated by (x_i, t_i) ∈ R^{d+m}. Now, x is a n vector of contribution, and t_i is in an m vector of target. If the SLFN by L hidden nodes can be estimated, then these N samples have zero error. This relationship infers a β, a, and b such that

\[
f_L(x) = \sum_{i=1}^{L} \beta_i G(s_i, b_i, x) x^T = 1, 2, \ldots, N.
\]

The equation above is described in a compacted form as follows:

\[
\begin{align*}
H\beta &= T
\end{align*}
\]

where H(α, b, x) with \( \alpha = a_1, \ldots, a_L \); \( \beta = b_1, \ldots, b_L \); \( \times = x_1, \ldots, x_N \); \( \beta = [\beta_1, \ldots, \beta_L]^T \); and \( T = [t_1, \ldots, t_N]^T \). Let H represent the hidden layer of the SLFN output matrix, with the ith column of H being the ith hidden node's output with respect to the inputs x_1, x_2, \ldots, x_N. β can be calculated by \( \beta = H^+ T \) where \( H^+ \) is the Moore-Penrose generalized inverse of matrix H [72].

4. Results

Through computer programming and simulations, the proposed approach applies the detailed theoretical framework to several movement patterns in different environments that resemble typical experimental settings for system benchmarking. In the datasets section, two human movement datasets are introduced as a diverse biological movement paradigm. The simulation results are obtained by testing two datasets for benchmarking. The results of the simulation for generating the slowest features in the ventral streams
and those of the combination technique for recognizing biological movements are also presented. The following sections present the results for recognition accuracy and confusion matrices. The experimental results are extensively discussed to reveal the effectiveness and to understand the perspective of the biological movements model. The performance of the model in recognizing different biological movements is evaluated through model simulations and by comparing this model with other state-of-the-art methods.

4.1. Datasets

KTH action dataset [73] is the largest human action dataset and includes 598 action sequences, which are composed of six types of single individual actions, including boxing, clapping, jogging, running, walking and waving. These actions were performed by 25 individuals in different conditions: outdoors (s1), outdoors with scale variations (s2), outdoors with different clothes (s3), and indoors with lighting variations (s4). Here, using down sampling, the sequence resolutions became 200 × 142 pixels. In our approach, we used 5 random cases (subjects) for training and designing the form and predefined templates. It means for every action, five different cases chosen to create action prototypes [5,29]. It provides the possibility to represent an action by only one image-template and ultimately template matching among the different actions. As discussed in the literature, KTH is a robust intra-subject variation with a large dataset, whereas the camera for video recording during the preparation had some shaking, making working with this database very difficult. In addition, it has four scenarios that are independent, separately trained and tested (i.e., four visually different databases that share the same classes) and both alternatives have been run. For considering the symmetry problem of human actions, a mirror function for sequences along the vertical axis was available for the testing and training sets. Here, all possible overlaps between human actions within the training and testing sets were considered (e.g., one video had 32 and 24 action frames.) Weizmann human action database [74] comprises nine types of single individual actions and had 83 video streams that revealed individuals performing nine different actions: running, galloping sideways, jumping in place on two legs, walking, performing jumping jacks, jumping forward on two legs, waving one hand, waving two hands, and bending. We tracked and stabilized the figures using the background subtraction masks that came with this dataset. A sample frame of this dataset is shown in Fig. 4. Aforementioned datasets have been extensively used to estimate the performance of the proposed techniques at recognizing biological movement examples. However they were concentrated on recognizing a single individual action (e.g., clapping or walking).

The other advantage of using these datasets is to compare with the state-of-the-art approaches. For our testing datasets, we illustrated the experimental results using a kernelized ELM algorithm to classify into the different kernel modes and performed a comparison with previous work that proposed biological human action models. We also compared the classifications between various kernels in the form pathway in term of their accuracy. The proposed method is not efficient in term of the computational cost. The computational load is usually due to feature extraction in two pathways form and motion features, applying ABM-based IncSFA and optical flow. The system infers that a new video requires time in our un-optimized MATLAB implementation, in which it is combined with existing codes for the motion and form pathway in MATLAB/C [32,13,70,71,65,24,75–77]. For providing an estimation of the computational time with an i7, 2.80 GHz CPU, 64–bit operating system machine with 12 GB memory (RAM) and our un-optimized codes under MATLAB R2016a, optical flow took around 0.7 s and ABM and IncSFA took around 1 and 0.02 second for 200 × 142 and 64 × 64 pixels input image resolution, respectively. It is noticeable that the aforementioned times are estimated for only one cycle of the approaches for the paths (Fig. 6).

4.2. Results of the simulation for the slowness features in the form pathway

The results for the ventral stream must be oriented around the concepts of shape and form features. This orientation follows the original biological movement model [1] and many approaches that are based on the visual system.

The recognition of biological movement patterns in the form pathway depends on the slow features generated by IncSFA. The slowest features of the training set have been used as human action prototypes [5]. We generated multi-prototype predefined templates for each obtained human action by applying IncSFA to the datasets. We divided every human movement sequence into training (55% of data) and testing (45% of data) sets. We used five random subjects from the training set for each action to obtain the action prototypes. These action prototypes are considered preventative for different biological movements [5,29].

The slowness features that are obtained through the application of IncSFA are shown in Fig. 8. For each dataset, two categories of biological movements are identified, using different slowness prototypes is required, and the actions are not directly comparable. KTH and Weizmann human action databases were used for benchmarking the performance of the proposed approach; thus, our experiment process should be consistent with that presented in [2,5,73]. We defined the set of our training map and tested the performance of our proposed technique for each dataset in which four scenarios in videos were used together (for the KTH dataset). The datasets were split into a set of training maps (from 55% of the dataset) with randomly selected subjects (five subjects) and a test portion with residual subjects (whatever remained). These two sets did not overlap. IncSFA was then applied to the training sets and generated slow feature prototypes that played a role in the form movement templates. We did not mainly depend on the training set as we applied slow feature prototypes. The ELM was trained and tested based on the splitting of datasets into 55% and 45% non-overlapping sets. The prototype selection was performed on the training set. The SFA in the ventral stream prevents us from using complex computer vision techniques, such as “bag of words.” The proposed approach involved the extraction of slowest features from a set of image frames for every action. The ABM-based IncSFA in the model performed more successfully compared with other state-of-the-art computational methods. This approach also greatly outperformed the other methods, such as “bag of word” and action key, both of which might generate inaccurate results. This method considers a set of patches that are locally selected and may ignore many structures; therefore, this technique has been acknowledged as a well-performing efficient object recognition method (Figs. 8 and 9). The local distribution of the action sequence was very similar to the targeted action and differed from other sample sequences taken from different categories. The intraclass variance was generally large except when recognizing a single human action [53]. Consequently, the model showed a fair performance in the recognition of biological movements.

4.3. Simulation results for motion pathway

To implement the motion pathway, the proposed technique applied a common and noticeable tool, optical flow [65], to generate motion pathway information that is estimated by analyzing optical flow patterns [1]. This pathway contains the neural detector order for motion information and optical flow features are providing such information in our computational model, which is followed by neurophysiological data [78]. From a temporal variation perspective,
Fig. 4. The figure depicts Weizmann and KTH human action datasets. To test the recognition of biological movements, two well-known human action recognition datasets were used. Here, the left set of image samples demonstrate actions from the Weizmann dataset; the second set, right-hand side, shows the KTH human action dataset. It is noticeable that the KTH dataset is one of the largest human action datasets, including six various human actions in four different scenarios.

Fig. 5. An explanatory diagram of ABM in the ventral processing [13] which represents the moving patterns and basis within the movement is shown in the figure. It shows the application of Gabor filter into a supervised object recognition method. It learns the object shape in the training stage within the action episode. (a) Representation of the Gabor bank filter in different scales and orientations. (b) Simulation results of training the ABM system for biological walking movements using the KTH human action recognition dataset. At the end, the walker’s shape is presented at the top of the figure. (c) The processing diagram of the ABM process for identifying the human object presented. The similarities between the method and biological findings at different levels have been discussed for different stages. In overall, ABM has SUM and MAX operands which constructs the hierarchy from simple cells to complex cells, and at the end, the entire human object shape by active basis.

the motion information here is considered a fast feature through which biological movements occur. These features were not generated as constant representative features throughout the whole episode but were focused at the temporal order within the current frame. In contrast to the form pathway, the motion pathway has temporal features, with each feature representing the motion information in a specific temporal movement. The optical flow pattern neurons were generated during action cycles and modeled similarly to the form pattern neurons from the form pathway. These two types of neurons primarily differed in terms of their temporal variations. The average optical flow throughout a biological movement is taken as a fast feature and added in the form of a coefficient to the form pathway results that are generated by the interaction between the ventral and dorsal streams (combining the fast and slow features). Fig. 7 shows the motion pattern features throughout the action cycles that are integrated into the processing stream.

4.4. Evaluation of the interaction between two pathways

To analytically assess the model, more than 38,000 frame cuboid forms of different human biological action movements were prepared. Given that the human movement recognition model shows an episodic behavior due to its utilization of slowness features (InSFA provides a better condition), using this algorithm for training analysis required many input frames and the testing and evaluation of the model were very limited. The proposed models were evaluated based on the level of predictions that were matched with the data. The proposed model assumes that the original models focus on slowness features and episodic processing. However, one cannot guarantee that the proposed model will reach a completely accurate level. The system was then trained and tested by computing the features in both pathways based on the previously described settings. For a specified test sequence, the action label
Fig. 6. Simulation results for a simple biological movement paradigm based on ABM-based IncSFA in the ventral processing stream are shown. Each row within the panel reveals the response of ABM during the episode as well as the slowness features generated for each different action. The first set of biological movements was obtained from the Weizmann human action recognition dataset [74], and the second group of biological movements are the results for KTH dataset [73]. Simulation results for the active basis through IncSFA follow the theoretical prediction with regard to the simplification of recognition using ABM-based IncSFA and its application in the ventral processing stream for opening a new perspective for the mechanism for recognition of biological movements.

was assigned to the action frames. The proposed model correctly classified the majority of the actions (mentioned in the confusion matrices). The most frequent mistakes were found in distinguishing “running” from “jogging” and in distinguishing “boxing”, “clapping,” and “waving” [2,4,53]. Such difficulty can be attributed to the similarities among these movements, but the presented model dramatically diminishes such similarities via episodic learning. The results of each human movement scenario are presented in Tables 1
and 2. Unlike that of previous methods that utilized the same datasets, the accuracy of the proposed approach was reflected on the results of each human movement scenario. However, this comparison is not precise because of differences in the experimental setups. These results are comparable with those of other state-of-the-art approaches despite the differences in their methodologies, such as being supervised or unsupervised, the presence or absence of tracking, the subtraction of the background, or the consideration of multiple action recognition. The proposed approach was then evaluated using two human action datasets. Figs. 8 and 9 present the confusion matrices. The performance of the proposed model was compared with that of previous approaches that utilized the same dataset presented in Tables 1 and 2. The different methods listed in Table 1 show many variations in the experimental setups, such as in the partitions of training and testing data, the pre-processing procedure (e.g., tracking and background subtraction), the presence or absence of supervision, and the per-frame classification performance. The results of our model were more stable than those of other models due to its adoption of the slowness principle and its combination with fast features. However, the comparisons of our model with other methods are not strictly fair even though the other methods do not completely cover the biological perspective [73] (Fig. 10).
the of agnosia representation art

5.1. Fig.

9. Fig.

10. Table

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schult et al. [73]</td>
<td>71.72%</td>
<td>2004</td>
</tr>
<tr>
<td>Niebles et al. [79]</td>
<td>83.33%</td>
<td>2008</td>
</tr>
<tr>
<td>Wang and Mori [80]</td>
<td>91.2%</td>
<td>2009</td>
</tr>
<tr>
<td>Danafar et al. [4]</td>
<td>93.1%</td>
<td>2010</td>
</tr>
<tr>
<td>Zhang and Tao [53]</td>
<td>U-SFA:84.67%</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>S-SFA:88.83%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D-SFA:91.17%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD-SFA:93.50%</td>
<td></td>
</tr>
<tr>
<td>Babu et al. [18]</td>
<td>90.44%</td>
<td>2015</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>90.07%</td>
<td>2017</td>
</tr>
</tbody>
</table>

Table 1

The recognition results using the proposed method are compared with state-of-the-art methods that utilized the KTH action dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schult et al. [73]</td>
<td>72.8%</td>
<td>2004</td>
</tr>
<tr>
<td>Niebles et al. [79]</td>
<td>72.8%</td>
<td>2008</td>
</tr>
<tr>
<td>Schindler and Van Kool [2]</td>
<td>100%</td>
<td>2008</td>
</tr>
<tr>
<td>Wang and Mori [80]</td>
<td>100%</td>
<td>2009</td>
</tr>
<tr>
<td>Zhang and Tao [53]</td>
<td>U-SFA:86.67%</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>S-SFA:86.40%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D-SFA:89.33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD-SFA:93.87%</td>
<td></td>
</tr>
<tr>
<td>Babu et al. [18]</td>
<td>85.56%</td>
<td>2015</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>97.5%</td>
<td>2017</td>
</tr>
</tbody>
</table>

Table 2

Comparison of the proposed approach and previous methods that utilized the Weizmann human action dataset.

5. Discussion

5.1. ABM and Gabor kernel in form pathway

ABM is motivated to apply the Olshausen and Fields representation [44] to model image object category collections. Although the Olshausen and Fields model was proposed to explain the role of simple cells in the primary visual cortex (V1) (Fig. 5), Riesenhuber and Poggio theory [37] suggests that the local maximum pooling of simple cell responses is performed in V1 complex cells. Thus, the local perturbations for the orientations and locations of linear basis elements in the Olshausen and Field model can be derived in the form of a deformable template from the active basis [81]. The local maximum pooling by Riesenhuber and Poggio serves as an active basis for deforming image data. Multiple active bases are used for
highly articulate shape representations, such as an and-or graph in a compositional framework [13,82].

The Gabor wavelets model is very similar to the receptive field profiles of cortical simple cells [63]. Kernel PCA previously overcame some restrictions in its linear characteristics by nonlinearly transferring to a space of high-dimensional features from the input space. Kernel PCA derives a low-dimensional feature space and is considered nonlinear in the input space [63]. This technique originates from the theorem of Covers regarding pattern separation in the input space, which posits that nonlinear separable patterns are linearly distinguishable if the input space is nonlinearly converted into a high-dimensional feature space. From the computational perspective, kernel PCA follows the Mercer equivalence condition because in the input space return the inner products in the high-dimensional feature space; the complexity of computation is related more to the number of training samples than to the feature space dimension [47]. We treat the application of ABM [5,28,6,29] as a subset of the Gabor wavelet kernel for IncSFA. A schematic model of the visual cortex motivated by the inference behind this model is shown in Fig. 2. A set of Gabor wavelet filters with various phases, orientations, and positions is initially filtered by the input stimulus, the quadrature-phase output is squared, summed, square-rooted (energy of V1), and divided, and normalization and summation are performed across orientations [41]. The operations of ABM are almost similar to and significantly consistent with this procedure. Given the recognition property of the supervised learning object, the ABM can identify and focus on human objects more precisely and robustly than the Gabor wavelet filter. ABM was proposed to analyze the images more biologically significant than Gabor filter itself [47,43,49,83–85]. The role of ABM is similar to that of the 2D receptive field in mammalian cortical simple and complex cells. The orientation and selectivity of these kernels exhibit desirable characteristics of spatial locality. These kernels are also spatially localized in the optimal positions and domains of frequency.

5.2. Static and dynamic patterns

How does the proposed model with slow and fast processing streams recognize biological movement? The results suggest that ABM-based IncSFA combines form with motion (i.e., slow and fast information). The temporal features were calculated by two independent pathways that represent slow and fast features or static and dynamic patterns. Therefore, the proposed model offers a different perspective of biological movements [8] by gathering sensory information over diverse time scales [86]. Consequently, the recognition task follows the original model (which shows a fair performance in the targeted databases [2–4]). The proposed approach is significantly motivated by the holonomic theory (particularly form pathway) of Pribram that is based on the fact that the dendritic receptive fields in sensory cortices are described mathematically by Gabor functions [87] that are widely used by ABM. The visual information is incrementally treated in a series of cortical stages (e.g., motion and orientation as local features in neurons at each level, such as V1 [88]). ABM-based IncSFA [13,14,32] (multiple SFA methods) extracts slow features for the incremental modeling of the form processing streams in the ventral stream. As previously discussed, the primary stage involves the use of local (in V1 cells) and model detectors (Gabor-like filters) in 16 (including 8 preferred) orientations, and using the proper scale depends on the receptive field [2,39]. However, an invariant behavior is referred from the neuron response of the central nervous system [23] (e.g., early vision complex cells phase invariance [88] in hippocampal place head direction invariance cells [89]). Human action cycles are unlikely to have invariant poses that are independent from the environment, different lighting conditions, or action poses. These invariant forms of actions can serve as important criteria that represent the form processing stream information. The slowness principle applies the perception and inferences of neurons that are trained by these invariant actions by favoring the slowly changing outputs in 2D [89]. A good implementation of this principle is SFA, which represents the mean square from the temporal derivative of the output and serves as a fine method for identifying the physiological properties of complex cells in the visual cortex [26,55]. IncSFA considers the similarities between the view-tuned neurons in area IT and the snapshot neurons with regard to the independency of their scale and form. The proposed approach is modeled through the slowest feature of the ventral stream, and such feature is adjusted by using unsupervised learning methods. Fast features use and infer the optical flow by using the slowness information of the other pathway to achieve a high-level integration of neuron motion pattern information into the snapshot neuron outcomes.

Based on several proposed neurophysiologically plausible models for approximating local motion [100–103,92,1], the first level of the motion pathway comprises the correspondence detectors for local motion, which includes direction-selective neurons [91] and motion-selective neurons [104]. In the simulation stage, the temporal optical flow patterns were directly calculated and the motion-sensitive neuron responses were computed by realistic physiological parameters [104] with several size of the receptive field in different levels. The second level of the motion pathway had larger receptive fields for the local flow structure, which induces the movement of stimuli (more neurophysiological evident mentioned in Table 2). The selective flow translation and neurons of motion pattern correspond to bandpass or low tuning by considering their speed [92]. In the original model, four direction neuron populations are typically preferred, and local optical flow detectors are considered for the motion edges [11]. The output signals were calculated by combining two nearby sub-fields with contradictory preferred directions. The optical flow pattern neurons in the third step of the motion path stream correspond to the snapshot neurons in the other pathways. Optical flow pattern neurons were identified at different locations of the visual cortex (i.e., STS or fusiform and occipital face areas).

5.3. Interaction between pathways

The mechanism for recognizing biological movements involves two independent pathways. The information of these two pathways interact at several levels of the mammalian brain [93,94]. The proposed approach presented an interaction between pathways by combining the results of IncSFA with the flow information (similar to [6]) and follows supporting evidence to generate the final recognition results. Such interaction simplifies the aggregation of the model (at the STS level for instance [98]) and improves its performance. Holonomic features consider both pathways for predefined action templates. In the form pathway, the proposed approach follows the holonomic theory of brain [87,90]. The dendritic receptive fields in sensory cortices are mathematically described by using Gabor functions [73] and are largely used by ABM [13]. As previously discussed, the primary stage involves the use of local (in V1 cells) and model detectors (Gabor-like filters) in 16 (including 8 preferred) orientations and a scale that depends on the receptive field [38,39]. ABM also acts as a snapshot detector in the human body shape model that is consistent with the IT area (inferotemporal cortex) of monkeys where view-tuned neurons are located. The model for complex-shaped tunes [99] is implemented through ABM-based IncSFA. Specifically, SFA represents the performance of view-tuned neurons in the IT area and that of snapshot neurons to achieve an in-dependency in scale and position. The proposed model follows the modeling and is adjusted through training with key frames. The results of the optical flow
are integrated with the form information that covers a high level of interaction between the snapshot neurons (static patterns) and the motion pattern neurons (dynamic patterns). The ABM and IncSFA for form information summarization are based on neurobiological, neuro-computational, and theoretical evidence [52,1,2]. The local direction is organized at the initial level of the form pathway, and Gabor-like modeling detector methods (i.e., ABM) achieve fair constancy by modeling the cells as described in the aforementioned section [43]. We can generate 16 directions and 2 spatial scales by using the mechanism of the proposed neurophysiologically plausible model. We also use two differentiators to gather information about the local direction of the pathway and complex-like cells with independent form features that are appropriate for the form pathway.

5.4. Comparison with related works

The proposed approach shows a considerably higher accuracy compared with extant approaches, most of which show similar accuracy ranges. Table 1 compares the accuracy of our approach with that of some state-of-the-art methods that are applied to the KTH human action database. The accuracy of the proposed approach is relatively comparable with that of the fairly accurate methods of Schindler and Van Kool [2], Wang and Mori [80], Danafar et al. [4], and Zhang and Tao [53]. The methods of Danafar [4] and Zhang [53] show great similarity with the proposed method in terms of biological motivation. Table 2 compares the accuracy of the proposed approach with that of some state-of-the-art methods that are applied to the Weizmann human action database. Schindler and Van Kool [2] and Wang and Mori [80] proposed extremely accurate approaches. The former shows a considerable accuracy for both human action databases but fails to consider the temporal features in the configuration. The method of Zhang and Tao [53] is the first SFA to be used for recognizing human actions yet is limited by its poor biological configuration according to biological evidence.

In terms of mechanism configuration, the proposed developments in the model are based on the models of Giese and Poggio [1], Schindler and Van Kool [2], and Danafar et al. [4], all of which have Gabor filters and optical flow in their ventral and dorsal streams. Our approach greatly improves the model by using an object-based ventral stream via ABM and by achieving an interaction between optical flow and object recognition via the optical flow division technique. However, the model of Danafar et al. [2] shows an increased computational time for each frame. Moreover, the high-
dimensional reproducing kernel Hilbert space data transformation of this model shows a favorable degree of interaction. The proposed approach strives to achieve this data interaction advantage by combining the motion and form information for ELM classification. The interaction between the slow and fast features, which is not present in the approaches of Zhang and Tao [53] (which used slow features) or Wang and Mori [80] (which used optical flow), provides a new perspective for developing the computational model.

5.5. Our contribution

The major contribution of this work approach lies in its application of computer vision as a biologically inspired approach and as a principal constituent of soft computing (similar to [105,106]) by integrating IncSFA into the mechanism for recognizing biological movements. Gabor wavelet-like filters have been previously applied in the ventral pathway [1–3] by integrating ABM (a supervised Gabor-based method for object recognition [75]) [5,42,107] into the ventral stream for recognizing the shape of movements. The proposed approach modifies the recognition mechanism by applying IncSFA, which eliminates the non-robust behavior of ABM to find form information in the ventral stream, and strives to overcome the limitation in the approach of Zhang and Tao [53] via modified configuration. Moreover, the slow and fast features concept definition for the recognition mechanism increases the novelty of this approach.

6. Conclusion

This paper develops a mechanism for recognizing biological movements in the visual system. The form and motion pathways are represented in the ventral and dorsal processing streams. Following the slowness principle, we use ABM-based IncSFA to extract form information from the human movement shape and use IncSFA to extract the slowest features in this pathway. The optical flow generates the motion information and extracts the fast features. The form information is abstracted by using IncSFA and is compared with the action prototypes afterward. However, the motion information involves an optical flow division that achieves an interaction between two pathways by aggregating information. For the decision-making process, a kernel ELM recognizes the actions using the information of both pathways as its attributes. Integrating IncSFA into the ventral stream and involving slow and fast features in the recognition mechanism are the major contributions of this research. The model analyzes and modifies the previous mechanism for recognizing biological movements by combining fast and slow features. The integration of these features shows a favorable performance in recognizing human actions, and such performance is evaluated by using the KTH and Weizmann human action datasets for benchmarking. Future studies on recognition capability may modify dual pathways by using dictionary learning approaches.

Acknowledgments

The authors would like to thank anonymous reviewers and editorial board of Applied Soft Computing journal (specially Bas Van Vlijmen) for their constructive comments which helped us to improve the quality of the paper also Ce Liu in Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology for layer-wise optical flow codes, Ying Nian Wu in UCLA (University of California, Los Angeles) for active basis model code, Matthew Luciw in the Swiss AI Lab IDSIA (Istituto Dalle Molle di Studi sull'Intelligenza Artificiale) for MATLAB codes of incremental slow feature analysis, and Guang-Bin Huang and Zhou Hongming in Nanyang Technological University for their guidance in using ELM. We acknowledge Paria Yousefi from school of computer science, Bangor University for her help and comments. This research was done under the High Impact Research (HIR) foundation grant: contract No. UMC/HIR/MOHE/FCST/10, in University Malaya (UM), Malaysia.

References

temporal
Slow
Racicot,
cortex,
572–581.
Wurtz,
recognition,
model
profiles,
H.
D.A.
H.
Z.
B.
J.P.
S.
Thorpe,
Eifuku,
Zito,
Bray,
Memariani,
Nickisch,
Franzius,
Riesenhuber,
Riesenhuber,
Oram,
Hashimoto,
Nickisch,
T.
Kana,
Inferotemporal
cortex
and
pathways
for
879–892.
Kanade,
Generalized
Inverse
of
Matrices
and
Its
Applications,
C.
Schuldt,
I.
Laptev,
B.
Caputo,
Recognizing
human
actions:
a
local
SVM
approach,
in:
Proceedings
of
the
17th
International
Conference
on
Pattern
Recognition,
2004,
ICPR
2004,
vol.
3,
IEEE,
pp.
32–36.
L.
Corelick,
M.
Blank,
E.
Shechtman,
M.
Irani,
R.
Basri,
Actions
as
space-time
shapes,
IEEE
Trans.
Pattern
Anal.
Mach.
Intell.
29
(12)
(2007)
2247–2253.
Y.
Wu,
Z.
S.
H.
Gong,
S.
Zhu,
Active
Baseline
Model,
Shared
Sketch
Algorithm,
and
Sum-Max
Maps
for
Representing,
Learning
and
Recognizing
Deformable
Templates,
Tech.
Rep.,
2009.
V.R.
Kompella,
M.
Luciw,
J.
Schmidhuber,
Incremental
slow
feature
analysis:
adapting
low-computational
power
feature
updating
from
high-dimensional
input
streams,
Neural
Comput.
24
(11)
(2012)
2994–3024.
Y.
Yi,
W.
Z.
S.
H.
Gong,
S.
Zhu,
Active
Baseline
Model,
Shared
Sketch
Algorithm,
and
Sum-Max
Maps
for
Representing,
Learning
and
Recognizing
Deformable
Templates,
Tech.
Rep.,
2009.
V.R.
Kompella,
M.
Luciw,
J.
Schmidhuber,
Incremental
slow
feature
analysis:
adapting
low-computational
power
feature
updating
from
high-dimensional
input
streams,
Neural
Comput.
24
(11)
(2012)
2994–3024.
Y.
Yi,
W.
Z.
S.
H.
Gong,
S.
Zhu,
Active
Baseline
Model,
Shared
Sketch
Algorithm,
and
Sum-Max
Maps
for
Representing,
Learning
and
Recognizing
Deformable
Templates,
Tech.
Rep.,
2009.
V.R.
Kompella,
M.
Luciw,
J.
Schmidhuber,
Incremental
slow
feature
analysis:
adapting
low-computational
power
feature
updating
from
high-dimensional
input
streams,
Neural
Comput.
24
(11)
(2012)
2994–3024.
Y.
Yi,
W.
Z.
S.
H.
Gong,
S.
Zhu,
Active
Baseline
Model,
Shared
Sketch
Algorithm,
and
Sum-Max
Maps
for
Representing,
Learning
and
Recognizing
Deformable
Templates,
Tech.
Rep.,
2009.
V.R.
Kompella,
M.
Luciw,
J.
Schmidhuber,
Incremental
slow
feature
analysis:
adapting
low-computational
power
feature
updating
from
high-dimensional
input
streams,
Neural
Comput.
24
(11)
(2012)
2994–3024.
Y.
Yi,
W.
Z.
S.
H.
Gong,
S.
Zhu,
Active
Baseline
Model,
Shared
Sketch
Algorithm,
and
Sum-Max
Maps
for
Representing,
Learning
and
Recognizing
Deformable
Templates,
Tech.
Rep.,
2009.
V.R.
Kompella,
M.
Luciw,
J.
Schmidhuber,
Incremental
slow
feature
analysis:
adapting
low-computational
power
feature
updating
from
high-dimensional
input
streams,
Neural
Comput.
24
(11)
(2012)
2994–3024.
Y.
Yi,
W.
Z.
S.
H.
Gong,
S.
Zhu,
Active
Baseline
Model,
Shared
Sketch
Algorithm,
and
Sum-Max
Maps
for
Representing,
Learning
and
Recognizing
Deformable
Templates,
Tech.
Rep.,
2009.
V.R.
Kompella,
M.
Luciw,
J.
Schmidhuber,
Incremental
slow
feature
analysis:
adapting
low-computational
power
feature
updating
from
high-dimensional
input
streams,
Neural
Comput.
24
(11)
(2012)
2994–3024.
Y.
Yi,
W.
Z.
S.
H.
Gong,
S.
Zhu,
Active
Baseline
Model,
Shared
Sketch
Algorithm,
and
Sum-Max
Maps
for
Representing,
Learning
and
Recognizing
Deformable
Templates,
Tech.
Rep.,
2009.
V.R.
Kompella,
M.
Luciw,
J.
Schmidhuber,
Incremental
slow
feature
analysis:
adapting
low-computational
power
feature
updating
from
high-dimensional
input
streams,
Neural
Comput.
24
(11)
(2012)
2994–3024.
Y.
Yi,
W.
Z.
S.
H.
Gong,
S.
Zhu,
Active
Baseline
Model,
Shared
Sketch
Algorithm,
and
Sum-Max
Maps
for
Representing,
Learning
and
Recognizing
Deformable
Templates,
Tech.
Rep.,
2009.
V.R.
Kompella,
M.
Luciw,
J.
Schmidhuber,
Incremental
slow
feature
analysis:
adapting
low-computational
power
feature
updating
from
high-dimensional
input
streams,
Neural
Comput.
24
(11)
(2012)
2994–3024.
Y.
Yi,
W.
Z.
S.
H.
Gong,
S.
Zhu,
Active
Baseline
Model,
Shared
Sketch
Algorithm,
and
Sum-Max
Maps
for
Representing,
Learning
and
Recognizing
Deformable
Templates,
Tech.
Rep.,
2009.
V.R.
Kompella,
M.
Luciw,
J.
Schmidhuber,
Incremental
slow
feature
analysis:
adapting
low-computational
power
feature
updating
from
high-dimensional
input
streams,
Neural
Comput.
24
(11)
(2012)
2994–3024.