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

Corresponding author email: ckloo.um@um.edu.my

Abstract

Computational neuroscience studies that have examined human visual system through functional magnetic resonance imaging (fMRI) have identified a model where the mammalian brain pursues two distinct pathways (for recognition of biological movement tasks). In the brain, dorsal stream analyzes the information of motion (optical flow), which is the fast features, and ventral stream (form pathway) analyzes form information (through active basis model based incremental slow feature analysis ) as slow features. The proposed approach suggests the motion perception of the human visual system composes of fast and slow feature interactions that identifies biological movements. Form features in the visual system biologically follows the application of active basis model with incremental slow feature analysis for the extraction of the slowest form features of human objects movements in the ventral stream. Applying incremental slow feature analysis provides an opportunity to use the action prototypes. To extract the slowest features episodic observation is required but the fast features updates the processing of motion information in every frames. Experimental results have shown promising accuracy for the proposed model and good performance with two datasets (KTH and Weizmann).

Keywords: Biologically inspired model, ventral and dorsal streams interaction, human action recognition, active basis model, Incremental Slow Feature Analysis, Extreme Machine Learning.

1. Introduction

The visual system in the human brain has a spatial grain that provides small invisible details; for example, the surface that you are viewing is formed by many individual molecules (there are some valuable biological methods for Fuzzy Physarum Algorithm (FPA) for fuzzy shortest path problems [50] or human prediction related to sensor motor map [94]). Physiological and psychophysical studies have demonstrated that there are various processes involved in the analysis of biological motion. This analysis is initiated by the identification of energies of local motion to displacements (see [100], [11], [12], [13]). There is tuning in the spatial frequencies that consider the different variations and contrasts in luminance [11], [13]. In contrast, in different pathways, the qualitative aspects of motion analyses have been analyzed, i.e., the local or global motion or motion pattern. Tuning the temporal frequency has been identified as a mechanism of motion sensitivity in early vision. Many studies have suggested slow and fast motion detectors for the visual system. Long temporal integration properties have also been considered with respect to motion processes [13], [14]. When temporal processing is further considered, there has been substantial growth in temporal illusions, such as time perception distortions [15], object features that bind synchronization [16], object motion and time of perception [17]. A fundamental understanding of temporal illusions and the visual systems functional organization occurs via temporal limits [18]. Furthermore, multiple temporal resolutions due to the temporal interval within the video frames in the human visual system can be considered for the perception of biological movements. When recognizing a biological movement despite variability in the conditions of lighting, human object locations or even perspectives in time represent an ability that is undoubtedly an advantage in the system [19]. As previously discussed, it is amazingly easy for the brain to perform this task normally; thus, individuals may hardly be aware of difficulties in the recognition
problem. In the primary sensory system, even minimal signal variation in the human object in time may lead to vastly different stimuli in the recognition results. Thus, in some manner, the brain must produce relatively different stimuli considering the time changes for the representation of the underlying cause, creating an inner representation that is not changed when irrelevant variations occur across time. The approach presented here was derived from the question of how such unchanged representations could be established and influence the recognition of biological movements. Due to the limited quantity of genome information as well as the plasticity of neural expansion in divergent environments, it appears that the information required to form unchanged depictions is present at the start of individual development. Some information must be assembled from the sensory input experienced throughout interactions in an environment that must be learned. As previously discussed, the process of learning appears to be unsupervised to some extent. Thus, the brain requires heuristics as to the specific stimuli that should be similarly classified. There is also a possible stimulus indicator for the presentation of a similar object in temporal proximity. The visual scene is very unlikely to be completely different from one moment to the next. Recognizing a biological movement in spite of probable varies in terms of conditions of lighting, human object location or even perspective in time is an ability which is undoubtedly an advantage in the system [19]. The successful learning of invariant representations would occur via adaptation of the sensory input to the sensory system for the extraction of slowly varying features. The temporal stability or slowness principle is the fundamental basis of an entire learning class of algorithms [103], [19], [20]. The applications of this approach have concentrated on visual system models predominately focused on the self-organized configuration of complex cell receptive fields in the primary visual cortex [21],[22]. The influence of fast and slow features for the recognition of biological movement has been addressed. Considering that biological movements are understandable in the temporal mechanism, the recognition of such actions is substantially related through temporal movements and their visual understanding in the human brain. However, the recognition of biological movement is not limited to the analysis of more slow or fast features and follows the psychophysical and neurophysiological model regarding independent pathways that are a foundation of this field [23], [109], [95],[82], [96],[84]. In the following discussion, a different perspective of the original model along with an approach for utilizing the active basis model and incremental slow feature analysis are demonstrated. An application of the active bases through the slowness principle and its combination with the motion information provide a different perspective to the original model of the recognition of biological movements [23], and these results and discussion are subsequently presented.

Author Summary
The recognition of biological movements in the original model considered two parallel pathways in the mammalian visual system. The proposed perspective of the current model occurs through slowness and quickness in the ventral and dorsal processing streams. Modeling the ventral stream utilized the slowness principle for the active bases for the extraction of the human object form and motion information that attained the optical flow. The interaction between these two pathways occurs in the categorization, which obtains significant results in the decision and recognition of movements. However, each pathway can separately facilitate the recognition of biological movements by the considerable disparity rate. 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 occipitotoparietal cortex) (see additional details in [107]). The perspective of this approach considers the original model with a specified mathematical explanation of the system.

Methods and Materials
We considered the slowness features of a human object in the ventral stream that forms the pathway along with information about the motion pathways fast features, to modify the original biologically inspired model of the visual system. In the primary experiment, approximately 38,000 cuboid frames of human subject movements were entered into the proposed model. The information analyzed in the model included several factors, such as form, motion, slow and fast. The response of every input was estimated throughout the time series data by the application of IncSFA (see methods) for slowness and its combination with motion, i.e., fast information.
1.1. Model Motivation and Biologically Concept

The model presented here was for the recognition of biological movements and was motivated by the original biologically inspired models (see [23]) by development under the perspective of slowness and fastness in the context of the two parallel pathways of the mammalian visual system response for a sequence of operations (Figure 2A). It is mostly motivated through prediction using the Active Basis Model (ABM) [55], which is similar to 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) (see [52]). The novel component of this development was the utilization of the slowness principle over the ventral pathway. This model, similar to the original model, has two separate pathways for form and motion information.

To elaborate the concept of temporal features, it appears essential to consider the original model of biological movements, which followed four reliable assumptions based on physiological, anatomical and imaging experiments and several cortical areas [23], [24].

The motion pathway, which represents the dorsal stream in the mammalian brain and the biologically inspired model, involves information related to optical flow, which has fast temporal variations in nature. This pathway is consistent with neurophysiological data from neural detectors. Faster varying features are due to its achievements within short changes between Frame (t) and Frame (t+1) rather than for the entire episode, such as ABM-based incremental SFA in the form path. The local detector of optical flow is connected with motion patterns, and the model is composed of a population of four directed neurons in the area of MT; however, there is a connection between MT and V4 for motion and direction selection. Additionally, motion edge selectors in two opposite directions are present in areas of the MT, MSTd, MSTl [41], [42], many regions of the dorsal steams and most likely in the kinetic occipital area (KO) [23] as well as the motion selective edges that are similar to MT [41] and MSTl [42] in the macaque monkey.

Few plausible models have been proposed for the recognition of human body shape that are neurophysiological with regard to the recognition of the stationary form (e.g., [30]). The proposed approach follows an object recognition model [30], which is composed of form feature detectors and involves slowness through ABM-based IncSFA. This approach is reliable and follows the data obtained from neurophysiological information concerning the scale, position and size invariance in the case of adaptive ABM, which requires a further computational load with hierarchy. The methods, which have Gabor-like filters to model the detectors, have good constancy in simple cells [39]. The complex-like cells in the V1 area or in V2 and V4 are invariant in terms of position varying responses (See [23]) and the size independency typically present in the V4 area. V2 and V4 are more selective for difficult form features, e.g.,

Figure 1: Overall structure of the visual system analytical models. The approach aims to develop computational models for the recognition of biological movements and to characterize the responses for different actions. This model represents the perspective of the original model that consisted of particular computations of slow and fast features. The model can operate with a wide range of high-dimension input, and the outcome is a combination of the ventral and dorsal processing streams.
Figure 2: The hierarchical model follows the original model; the interpretation of the data reflects the perspective of a combination of slowness and fast features provided from the ventral and dorsal processing streams. An overview of the form and motion pathways is shown. The insert depicts the various types of neural detectors in diverse sections of the hierarchy. V1 and IT represent the primary visual cortex and the inferotemporal cortex, respectively; KO and STS represent the kinetic occipital cortex and the superior temporal sulcus, respectively. These and other abbreviations indicate the visual cortex in monkeys and humans (see [23]).

junctions and corners, but they are not appropriate 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 those of area IT (inferotemporal cortex) of the monkey, where the view-tuned neurons are located and the model of complex shapes is tuned [41]. Snapshot neurons are similar to view-tuned neurons in area IT and provide an independent scale and position. Previous models used Gaussian Radial Basis functions for modeling, and they were adjusted in training that performed a key frame with regard to training sequences. This paper elaborates the key frame, which is considered to be efficient and has a fast feature concept, whereas the shape of specific human movement is defined in the entire episode and is independent from fast temporal change. Slowness features, which have Gabor-like features, can serve as a better representative with regard to the form information of biological movements. We introduce a perspective of a model that follows the original models by utilizing ABM-based IncSFA, as explained in the methods section. This section might be skipped for the reader who is mathematically less inclined. The computational simulations and the testing method are presented in the results section. Finally, we conclude that biological motion perception in the human visual system comprises fast and slow feature associations, which induces the recognition of biological movements. For the examination of the proposed model in a broader range of high-dimensional video streams, we measured the responses in separate parallel pathways of the visual system. The results of an instance patterns model in the ventral path are shown in Figure 3. The proposed model has adequate performance for catching the constant pattern of responses of the ventral pathway to the human movements (Figure 3, upper processing stream). The model does not undervalue the responses of the dorsal pathway covering approximately half of the visual system. The slowness characteristic in the ventral stream has been hidden, and its response is underestimated in the recognition of the biological movement model. It can be clearly indicated through responses to different actions. Here, this pattern of slowness features is shown with respect to the visual system model regarding the application of Gabor-like stimuli for the object as an object recognition task throughout the ventral stream. ABM is a Gabor-based supervised method that can boost the responses of the stream directive and can demonstrate excellent interpretation of the human object. The proposed model attempted to increase the performance of the recognition of the biological movement model by incorporating the slowness features with fast features from the dorsal stream in the previous and original models (see [23]).
Figure 3: The schematic of the model is presented here for both pathways. In the ventral processing stream, i.e., the form pathway, a set of Gabor filters has been applied at different orientations, positions and phases; the outputs of the V1 are the outcomes of quadrature-phase pairs, summed, squared and square-rooted. Then, the outputs of the filter are normalized with regard to the local population. The filter outcomes are subsequently max pooled and summed across space. The MAX and SUM operations are based on the attained active bases of the object form. This initial part of the schematic occurs through the Active Basis Model [55] as the Gabor-based object recognition operation. Finally, the output of the ABM is utilized in the slowness principle method (incremental slow feature analysis) [7], [79] for the extraction of form slow features. The dorsal processing stream helps to obtain motion information throughout the high-dimensional input stream. The motion pathway is attained by Optical Flow. The average of these flows within the episode \((t_0, t_1, ..., t_n)\) plays the fast feature in this hierarchy, which the ventral stream requires temporally for utilizing IncSFA for the generation of slow features. However, each pathway can have its own decision with regard to categorization, and it justifies the two patients (DF and RV) performances (see [107]).

As a result of the biologically inspired model complication discussed briefly in the previous section, it appears difficult to provide an analytical statement about the model. In this section, the mathematical framework relevant to the actual prediction will be introduced. Individuals who are theoretically less inclined should feel free to skip this section. The model will consider a perspective of the recognition of the biological movement model, which concerns the task throughout two parallel pathways in slowness and fastness, outlined theoretically and conceptually.

### 1.2. Active Basis Model

The active basis model [55] that applies Gabor wavelets (for the elements dictionary) consists of a deformable biological template. A Shared Sketch Algorithm (SSA) is followed through AdaBoost. In each iteration, the SSA that follows the matching pursuit chooses an element of the wavelet. It checks the objects 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 \(\beta = (\beta_i, i = 1, ..., n)\) is set of Gabor Wavelet elements and components of sin and cosine, \(c_i = \langle I, \beta_i \rangle\) and \(\epsilon\) is unsolved image coefficient [55]. By using wavelet sparse coding large number of pixels reduces to small number of wavelet element. Sparse coding can train natural patches of image to a Gabor like wavelet elements dictionary which carries the simple cells in V1 properties [53]. The extraction of local shapes will be separately done for every frame and like [55] responses of filter orientation and density of each pixels computes. Also, the active basis model uses the Gabor filter bank but in different form. A Gabor wavelets dictionary, comprising \(n\) directions and \(m\) scales is in the form of, \(GW_j(\theta, \omega), j = 1, ..., m \times n\). Where, \(\theta \in \{\frac{k\pi}{n}, k = 0, ..., n-1\}\) and \(\omega = \{\frac{\sqrt{2}i}{m}, i = 1, ..., m\}\). Gabor wavelet
features signifies the object form as small variance in size and location and posture. Though overall shape structure, it considers to be maintained throughout the process of recognition. Response (convolution) to each element offers form information with $\theta$ and $\omega$.

$$B = (GW, I) = \sum \sum GW(x_0 - x_{y_0} - y : \omega_{0}, \theta_{0})I(x, y). \quad GW_{i}$$ is a $[x_{y}, y_{y}]$ matrix, and the response of $I$ to $GW$ is $[x_{i} + x_{y} + y_{y}]$. Therefore, the previous convolution of both matrices must be padded through sufficient zeroes. The consequences of convolution can be eliminated by cropping the result. An additional approach would be 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 $I^{m}, m = 1, ..., M$, the joint sketch algorithm consecutively chooses $B$. The fundamental opinion is to identify $B_{i}$ so that its edge segments obtained from $I_{m}$ become maximal [5]. It is then necessary to compute $[I^{m} \beta] = \psi | (I^{m} \beta) |$ 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 $\beta = (\beta_{i}, i = 1, ..., n)$ be the template; for every training image, $I^{m}$ scoring will be based on:

$$M(I^{m}, \theta) = \sum_{i=1}^{n} \delta_{i} | I^{m}, \beta | - \log \Phi(\lambda \delta_{i}). \quad (2)$$

$M$ is the match scoring function, $\delta_{i}$ from $\sum_{i=1}^{M}[I^{m}, \beta]$ addresses 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 $\Delta = (\delta_{i}, i = 1, ..., n)$ [5].

$$\text{MAX}(x, y) = \max_{x, y \in D} M(I_{m}, \beta). \quad (3)$$

$\text{MAX}(x, y)$ calculates the maximum matching score previously obtained. $D$ 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.

1.3. **Slowness Principle for ventral processing streams**

The perception of slow feature analysis is connected to the hypothesis that the input information (e.g., actions or activities) included in a 2D signal sequence (e.g., a video) vary not rapidly but rather gradually over time [4]. Whereas the input signal normally has a high variability (e.g., due to variation in the environment and different lighting conditions or noise), the separation between informative changes is generally hidden in the rarely changing sequence features. The video attributes that vary least over time can be extracted by slowness features. Slowness features have recently been entered in the computer vision task [4], [7], [51] and are typically connected to the visual cortex [46], [49]. An incremental learning algorithm is used due to the application of the slow feature analysis for each time step in an unknown video input. The Incremental Principle Component Analysis (PCA) is closely related to the incremental SFA [52],[53] because the PCA and Minor Component Analysis (MCA) can solve the SFA. Slowness features, which have information regarding active bases from multidimensional input as well as involving fast features, can solve the recognition of the biological movement task. SFA provides instantaneous scalar input-output functions, which generate a signal output (2D signal) that carries the important information and changes as slowly as possible.

Slow Feature Analysis (SFA) is an unsupervised learning method. The functions, which include planning the input stream for the most slowly changing outcomes, are characteristic of a number of elementary representatives of world possessions, which summarizes unrelated details selected by the sensors [7], [44], [46]. Moreover, considering a mobile agent that has high-dimensional video input, it is possible to search an otherwise stationary room and encode the data by combining the situation and direction with slow features [43].

SFA is typically concerned with the optimization of complexity: it is common that for the identification of $x(t)$ as input by the $D$ dimension, $x(t) = \left[x_{1}(t), ..., x_{D}(t)\right]$, there is a set of functions similar to $f(x)$ that have $L$ dimension, $g(t) = \left[g_{1}(t), ..., g_{L}(t)\right]$, or that can produce the output for $L$ dimension as $y(t)$ so that $y(t) = \left[y_{1}(t), ..., y_{L}(t)\right]$. Thus, the relationship between these sets is $y(t) := g(t)$.

$$\Delta_{i} := \Delta_{y} := \langle y_{i} \rangle. \quad \text{isminimal} \quad (4)$$

$$\langle y_{i} \rangle = 0. \quad \text{(Zeromean),} \quad (5)$$
Figure 4: A different approach to present the hierarchical model in terms of the theory and computation of the form information, slow features and motion information; fast features are shown in figure 3. A supervised Gabor-based object recognition method, ABM, provides this property with human objects and the computation of the slowness features is performed with IncSFA, which is episodic and provides slow form features within the episode that were combined with optical flow information; fast features create an interaction between the pathways.

\[ \langle \dot{y}_l^2 \rangle = 1. \quad \text{(Unitvariance)}, \]  

(6)

\[ \forall d < 1 : \langle y_{dyl} \rangle = 0. \quad \text{(De - correlation and order)}, \]  

(7)

These general definitions, similar to 2 and 3, are the restrictions for having insignificant constants in the output, and 4 is for de-correlation restrictions for features that are the same but are not coded. 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 [15], is not applicable, but it is predominately straightforward especially for the eigenvector method. Considering that \( f_i \) is constrained to be a linear function that consists of a combination of a finite set of nonlinear functions \( p \), the output function will be:

\[ y_l(t) = f_l(x(t)) = w_{yl}^T p(x(t)). \]  

(8)

Then, we will have \( z(t) = p(x(t)) \). Based on the changes previously incorporated, the optimization problem will be introduced by minimizing (6), the \( w_l \).

\[ \Delta(y_l) = \langle \dot{y}_l^2 \rangle = w_{yl}^T \langle \dot{z} \dot{z}^T \rangle w_l. \]  

(9)

If the \( p \) functions are selected such that \( z \) has a unit of covariance matrix and a zero mean, then the three restrictions will be satisfied if, and only if, the weight vectors have an orthonormal difference. Whitening is a very common technique that is used for identifying \( p \). For whitening, the principle component of the input data are required; thus, considering the zero mean and the individuality covariance matrix, put the \( x \) to \( z \) and by this \( z \), the SFA problem will be converted to the linear problem. Equation (6) should be considered for minimizing the L-normed set of eigenvectors of \( \langle \dot{z} \dot{z}^T \rangle \). The desired features will be obtained from the set of principle components of \( \dot{z} \). The objective was to calculate the temporal slowness, \( \delta \)-value, features and \( (x) \) as instantaneous functions of the input 2D-signal.
Figure 5: Comparison of functional imaging experiments with the outcome of the active basis model regarding the features of active basis prior to generation of the slowness features in the form pathway. The biological movements according to the research experiments of Gunnar Johnsson [108]: ten light bulbs were located on the joints, and the actor was recorded performing complex movements. There is recognition of the action within the episode of the actions. 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 increasing the ventral pathway in the recognition task.

The eigenvector-based algorithm is guaranteed to obtain the global optimum and learn biologically plausible rules for the existing optimization problem [46], [47], [48]. The modified optimization problem for the high-dimensional visual input utilizes the information of biological movements and the human object through an active basis model as a Gabor-based kernel. Then, this pathway information is combined with fast features by optical flow in the motion pathway with respect to the original model [23], [33], [34], [35].

1.4. ABM based Incremental Slow Feature Analysis

The active basis model is a supervised learning Gabor wavelet model that has been successfully used for object recognition tasks. It is motivated to apply Olshausen and Fields representation [53] to modeling the particular image object category collections. Although the Olshausen and Fields model were proposed to provide an explanation of the role of simple cells in the primary visual cortex (V1), Riesenhuber and Poggio's theory [30] grasps that the local maximum pooling of simple cell responses has been performed in the V1 complex cells. Thus, the local perturbations for the orientations and locations of linear basis elements in the model of Olshausen and Field can be derived to a deformable template from the active basis and, prior to that, the linear basis [54]. Riesenhuber and Poggio's local maximum pooling represents the active basis deforming for the image data explanation. Multiple active bases are used for more articulate shape representations as it is the simplest example of the and-or graph in a compositional framework [55], [56]. Furthermore, the model of Gabor wavelets is very similar to the receptive field profiles of cortical simple cells [58]. Previously, kernel PCA conquered several restrictions of its linear characteristics by nonlinearly transferring to a space of high-dimensional features from the input space. Kernel PCA derives low-dimensional feature space and is nonlinear in the space of input [58]. It originated from Covers theorem regarding pattern separability and represents that in the input space, nonlinear separable patterns are linearly distinguishable with high possibility if the input space is nonlinearly converted to a high dimensional feature space. From the perspective of computation, kernel PCA receives the Mercer equivalence condition benefit as well as feasibility because the inner products in the high dimensional feature space are returned by those in the input space, whereas the complexity of computation is connected to the training sample numbers moderately compared with the feature space dimension (See [102]).

Here, we introduce the Active Basis Model as a subset of the Gabor wavelet kernel for incremental slow feature analysis. The inference behind this model was motivated by a schematic model of the visual cortex. The set of Gabor wavelet filters on the various phases, orientations, and positions is initially filtered by the input stimulus, the quadrature-phase output are squared, summed and square-rooted (energy of V1) and division normalization and summation occur across orientations (see [52]). The Active Basis Model has approximately similar operations and a
A mask that reveals each layer’s visibility is the main di-

level of the motion pathway. For the motion of the subject, the layer-wise optical flow estimation has been utilized.

detectors. In the MT and V1 areas, there are some neurons for motion and direction selection, respectively, in the first

movement pattern, which is consistent with the neurophysiological information from the hierarchy of neural

areas. Such kernels are similar to the role of the 2D receptive field in the mammalian cortical simple and complex cells; the orientation and selectivity exhibit desirable characteristics of spatial locality and are spatially localized in the optimal positions and domains of frequency. Previously, Gabor wavelet has been widely utilized as a kernel and for various applications, e.g., face recognition [102]. The Gabor wavelet (kernels filter) has been defined in previous works [102], [57], [61], [63]:

\[
\psi_{\mu, \nu}(z) = \|k_{\mu, \nu}\|^2 \frac{e^{-\frac{\|z\|^2}{\sigma^2}} e^{i 2 \pi \nu (z - \underline{y})}}{\sigma^2}
\]

(10)

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

\[
k_{\mu, \nu}, \nu \mathbf{z} = e^{12} = \kappa \psi_{\mu, \nu}
\]

(11)

Where \(\kappa = \kappa_{\text{max}} / f^2\) and \(\phi_{\mu} = \pi \mu / 8\). \(\kappa_{\text{max}}\) is the frequency of the maximum, and \(f\) is the spacing factor between kernels in the frequency domain [64]. The active basis model is a supervised learning Gabor wavelet, and it is considered a kernel that is related to the bases within it. Unlike the Gabor wavelet kernel that was required to define the scales, the orientations and pixels in the active basis model, these parameters and attained in a way that is dependent upon the training. The active basis model represents the image obtained by the summation of the active base families, which are obtained through the Gabor wavelet dictionary and match scoring function. Let \(I(x, y)\) be the gray level distribution of an image; the image convolution \(I\) and a Gabor kernel \(\psi_{\mu, \nu}\) are defined as follows:

\[
B_{z, \mu, \nu} = \psi_{\mu, \nu}(Z) * I(Z) \quad i = 1, 2, ..., n
\]

(12)

Where \(z = (x, y)\) denotes the convolution operator, and \(B_{z, \mu, \nu}\) is the active base that corresponds to match scoring at the proper orientation and scale. Consequently, the set \(S = \{B_{z, \mu, \nu}: \mu \in M, \nu \in O\}\) forms the Gabor wavelet representation of the image \(I(Z)\) along with \(M\) and \(O\) which represent 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 \(X\). \(X\) is defined as a set of active bases 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.

1.5. ABM based Slow Feature Analysis

A method for nonlinearization utilizes the fact that SFA is solved by twofold PCA and entirely based on second-

order statistics. Therefore, SFA is capable of being kernelled in line with the extension of PCA to Kernel-PCA by Schlkopf et al. 1998; thus, in the case of incremental SFA, kernelled incremental PCA must be considered [3]. The presentation and implementation of a kernel based in the principle of temporal slowness has been performed by Bray and Martinez 2002 [1] using the Stone 1996 [2] objective function, which was in some ways not similar to SFA. Incremental SFA is needed, while SFA is used for every time step. As the SFA solution can be reached through PCA and MCA (Minor Components Analysis), it is closely relevant to incremental PCA [5], [6], [7].

1.6. Motion information from dorsal pathway

In the motion pathway, biological movements are recognized by patterns of optical flow. The optical flow identifies the movement pattern, which is consistent with the neurophysiological information from the hierarchy of neural detectors. In the MT and V1 areas, there are some neurons for motion and direction selection, respectively, in the first level of the motion pathway. For the motion of the subject, the layer-wise optical flow estimation has been utilized. A mask that reveals each layers visibility is the main difference between the estimation of traditional and layer-wise
Figure 6: The figure depicts Weizmann and KTH human action datasets. To test the recognition of biological movements, two well-known human action recognition datasets were utilized. 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.

optical flow. The mask shape is able be fractal and arbitrary, and matching only applies for the pixels that fall inside the mask (see [71]). We use the layer-wise optical flow method in [4] which has a previously described baseline optical flow algorithm [26], [27], [28]. $M_1$ and $M_2$ are visible masks for the two frames $I_1(t)$ and $I_2(t-1)$, and the field of flow from $I_1$ to $I_2$ and $I_2$ to $I_1$ are represented by $(u_1, v_1), (u_2, v_2)$. 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 $E_{\gamma}^{(i)}$, symmetric $E_{\delta}^{(i)}$, and smoothness $E_{\mu}^{(i)}$.

$$E(u_1, v_1, u_2, v_2) = \sum_{i=1}^{2} E_{\gamma}^{(i)} + \rho E_{\delta}^{(i)} + \xi E_{\mu}^{(i)} .$$

After optimization of the objective function, the use of outer and inner fixed-point iterations, image warping and a coarse to fine search, we attained bidirectional flow. Compressed optic flow for all frames was calculated by straight matching the template to the earlier frame by applying the summation of the absolute difference (L1-norm). Although optic flow is particularly noisy, no smoothing techniques have been performed with it because the field of flow will be blurred in gaps and, in particular, in the locations where information regarding motion is significant [7]. To obtain the proper response of the optical flow with regard to its application in the proposed model, the optical flow will be applied to adjust the active basis model 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.

1.7. Extreme Learning Machine (ELM)

Neural Networks have been widely utilized in several research areas because of their capability to estimate difficult nonlinear mappings straight from the input sample as well as offering models for a large class of artificial and natural phenomena that are problematic to model via classical parametric techniques. Recently, Huang et al. [65], [66], [67] presented a novel algorithm for learning, i.e., a Single Layer Feed-forward Neural Network structural design named Extreme Learning Machine (ELM). ELM solves the problems initiated through algorithms that use gradient descent, e.g., the Back propagation used in ANNs. ELM considerably diminishes the time quantity required for training in the Neural Network and has greatly enhanced faster learning and generalization performance. It requires fewer human interventions and can run significantly faster than conventional techniques. It routinely concludes the parameters of the entire network, which evades unimportant external interventions by humans and is more effective in real-time and applications. Several advantages of Extreme Learning Machine include the simplicity of usage, quicker speed of learning, greater generalization performance, appropriateness for several nonlinear kernel functions and activation function [68]. The Single Hidden Layer Feed-forward Neural Network (SLFN) function with hidden nodes

Results

This approach applied the detailed theoretical framework through computer programming and simulations to several movement patterns in different environments that resemble typical experiments for benchmarking the system.
In the following section, Datasets, the two human movement datasets have been introduced as a diverse biological movement paradigm. The results of simulation are revealed by the mathematical analysis of the described datasets as case experiments for benchmarking. Consequently, the results of simulation for generating the slowest features in the ventral streams and the combination technique to recognize biological movements are shown. In the next sections, the results of recognition accuracy and confusion matrices are presented. The experimental results are extensively presented to reveal the effectiveness and understanding of the proposed perspective of the biological movements model. For model evaluation, the recognition of different biological movements is rated through simulation of the model and comparison with state-of-the-art methods.

1.8. Datasets

The KTH action dataset [77] 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. For our approach, we used 5 random cases (subjects) for training and designing the form and predefined motion templates. 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. Moreover, it has four scenarios that are independent, separately trained and tested (i.e., four visually different databases that share the same classes). 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.) The Weizmann human action database[78] comprised 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 Figure 6. The previously discussed datasets have been extensively utilized to estimate the performance of the proposed techniques at recognizing biological movement examples. However, they were concentrated on recognizing a single individuals action, e.g., clapping or walking. The other advantage of using these datasets is a comparison with the state-of-the-art. For our testing datasets, we illustrated the experimental results using a kernelled ELM algorithm to classify into the different kernel modes and performed a comparison with previous work that proposed biological human action models. We also performed a comparison of the classifications made between various kernels in the form pathway with their accuracy. The proposed methods are efficient, and the computational cost is due to feature extraction with regard to the two pathways form and motion and slow and fast features, applying an active basis model-based incremental feature analysis 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 [7], [55], [69], [70], [71], [103], [104], [105], [106].

1.9. Results of simulation for ventral stream slow features in biological movement paradigm.

It is instinctively obvious and has been revealed that the ventral stream results must be oriented around the concepts of shape and form features. This orientation follows the original biological movement model [23] and many approaches that are based on the visual system. Cooperation between information attained from two processing streams occurs at several levels in the mammalian brain [36], [37], and it simplifies the aggregation of the model (for instance in STS level [38]) and improves the performance. Holonomical features consider both pathways for predefined action templates. In the form pathway, the proposed approach followed Karl Pribram’s holonomic theory, which was based on evidence that dendritic receptive fields in sensory cortexes have been described mathematically by the Gabor functions [77] that are largely used by the active basis model [55]. As previously discussed, the primary stage includes local (in V1 cells) and model detectors (Gabor-like filters) in sixteen (including eight preferred) orientations and using a scale that depends upon the receptive field (see [32], [41]). The active basis model also played the role of a snapshot detector with regard to the human body shape model, which was consistent with area IT (inferotemporal cortex) of the monkey, where view-tuned neurons are located and the model of complex shape tunes [40] is implemented through ABM-based IncSFA. Specifically, the slow feature analysis technique facilitated the representation of
Figure 7: An explanatory diagram of the ventral processing of the active basis model [55], which represents the movement patterns and shape forms of biological objects within the movement episode. The active basis model is a Gabor-based, supervised object recognition method that can learn the object shape in the training stage and can be utilized with the object recognizer 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 walkers 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. Overall, ABM has two stages, SUM and MAX, which form the hierarchy from simple cells to complex cells, and, at the end, the entire human object shape by active bases.

the view-tuned neurons performance in area IT and the snapshot neurons with respect to providing in-dependency in scale and position. The proposed model followed the modeling and was adjusted through training with key frames. Utilizing the optical flow outcome and inferring it with information that was known as a form of the biological object presented an approach that covered a high level of integration of the snapshot neuron outcomes with motion pattern neuron information. Furthermore, the active basis model used a computational mechanism that recognized the human object form in addition to the slowness information throughout the entirety of the biological movements, which follows the neurobiological, neuro-computational and theoretical records [4],[23],[33]. The local direction has been organized in the initial level of the form pathway and Gabor-like modeling detector methods, i.e., the active basis model had good constancy by modeling the cells in the discussed section [39]. Using the mechanism of the proposed neurophysiologically plausible model, we will generate sixteen directions and two spatial scales using two differentiators to identify information about the local direction of the pathway and complex-like cells that have independent form features appropriate for the form pathway.

1.10. Slowness features in Ventral stream

The recognition of the biological movement patterns in the form pathway depend upon the slow features generated by IncSFA (see [7], [79]). The slowest features of the training set have been used as human action prototypes. First, we performed the multi-prototype predefined templates for each human action obtained by applying the IncSFA to the datasets. For the training map of every action, we divided every human movement sequence into training and testing sets. These action prototypes are considered preventative for different biological movements [109], [84]. The results
Figure 8: Simulation results for a simple biological movement paradigm based on ABM-based slow features 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 [78], and the second group of biological movements was obtained from the KTH dataset [77]. Simulation results for the active bases through incremental slow features 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 of the original model for the recognition of biological movements.
Figure 9: Simulation results with regard to the dorsal processing stream by applying optical flow [103] are shown. As episodic operation occurred within the ventral processing stream, the form information, motion information, and fast features must be considered during the time that the ventral stream was performing (t0,tn). Each row represents an action during its episode, and the average of the flow for the entire episode is shown at the end of each row. The images present the flow in color form, which can depict the biological movement flow within the episode. The average optical flow throughout the biological movement took into consideration the recognition of the biological movement by fast features and added a coefficient to form pathway results as the interaction between the ventral and dorsal streams. The different actions of these stimulations are presented from the KTH human action recognition dataset [77].

of the slowness features attained through the application of IncSFA are shown in Figure 8. For every dataset, there were two categories of biological movements, the use of different slowness prototypes was required and the actions were not directly comparable. The KTH and Weizmann human action databases have been used for benchmarking the approach performance; thus, consistency with the set of experiments used in [33], [51],[77] was required. We defined the set of our training map and test of the proposed technique for each dataset in which the mixture of four scenarios in the videos were used together (for KTH dataset). The datasets were split into a set of training maps with randomly selected subjects and a test portion with the residual subjects. IncSFA was then applied to the training sets and attained the slow feature prototypes that had a role in the form movement templates.

The consideration of the slowness features and the slowest feature in the ventral stream of the proposed biological movement model helped to avoid the use of a computer vision technique, such as bag-of-words. The regular concept of the previously discussed approaches comprised extracting the slowest features in the set of image frames for every action. It has been determined that the ABM-based IncSFA in the model performed considerably well on the discussed datasets. The whole model outperformed the reported state-of-the-art computational methods. It also provided better performance compared with other methods, i.e., bag-of-words and action key, which may not have been correct as these methods consider a set of patches that are locally selected and may ignore many structures; thus, they have been acknowledged as efficient object recognition methods. This method performed well (see Figure 10 and Figure 11) while the local distribution of the action sequence was very similar to the targeted action and very different from other sample sequences from different categories. Briefly, the intraclass variance was large, and the interclass variance was lower. In the case of a single human action recognition in particular, the intraclass variance was smaller [51]; consequently, this application of the model performed well for the recognition of biological movements.
1.11. Simulation Results for dorsal stream and information of Motion pathway in biological movement paradigm

To implement this pathway, the proposed technique applied a common and noticeable tool, optical flow [71], to generate information regarding the motion pathway. Motion information for the recognition of biological movements was obtained by the analysis of optical flow patterns [23]. It contained the neural detector order for the features of optical flow for growing complications, which is reliable using neurophysiological data [86]. Here, the information about the motion processing stream was considered to be a fast feature from the perspective of the temporal changes 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 path has temporal features, and every feature represents the motion information in that specific temporal movement. Based on several proposed neurophysiologically plausible models for the approximation of local motion (e.g., [88], [89], [90], [91], [92], [23]), the first level of the motion pathway comprised the correspondence detectors for local motion, which included direction-selective neurons (see [87]) and motion-selective neurons of components in area MT [93]. In the simulation stage, the temporal optical flow patterns were directly calculated, and the motion-sensitive neuron responses were computed by realistic physiological parameters [93]. The size of the receptive field was in the range of the neuron for direction selection in V1 and (foveal neurons) MT [98], [23]. The second level of the motion pathway had larger receptive fields for the local flow structure, which induced the movement stimuli. The selective flow translation and neurons of motion pattern correspond in the area of MT [92], with bandpass or low tuning by considering the speed. Typically, in the original model, four direction neuron populations were preferred, and the local optical flow detectors were considered for the motion edges [23]. The output signals were calculated by the combination of two nearby subfields with contradictory preferred directions. The motion selective neurons and opponents have been identified in several areas of the dorsal processing streams that contain areas MT, MSTD and MSTl [41], [42], [23]. The optical flow pattern neurons in the third step of the motion path stream correspond to snapshot neurons in other pathways. Optical flow pattern neurons have been identified at different locations of the visual cortex (i.e., STS or fusiform and occipital face areas). The temporally optical flow pattern neurons were generated during action cycles and modeled like the form pattern neurons from the form processing stream by this difference in the feature space motion pattern features; they are considered fast features in front of the form features, which are the slowest features. Figure 9 reveals the motion pattern features throughout the action cycles with the consideration of integration into the processing stream.

![Figure 10: Confusion matrices of the proposed approach are presented and were obtained from the human action movements of the KTH dataset[77]. Three different kernels were for classification using the ELM algorithm [69], [70] in the decision making and categorization of the biological movement. From left to right, the RBF kernel-ELM, the wavelet kernel ELM and the Sigmoid-ELM confusion matrices have been depicted; the Sigmoid Kernel-ELM produced better results for the classification of biological movements.](image)

1.12. Evaluation of Interaction between Two Paths

To analytically assess the model, more than 38,000 frame cuboid forms of different human biological action movements were prepared. As the behavior of the recognition of the human movement model is episodic due to the
utilization of incremental slow features, the analysis training of this algorithm required many inputs frames, and the testing and evaluation of the model was very limited. The proposed models are valued via the level of the predictions matched with the data. Because the proposed model subsumes the original models and is concerned with slowness features and episodic processing, it conceivably has important results. However, there is no guarantee that the proposed model will reach a completely accurate level.

Subsequently, by computing features in both pathways with the previously described different settings, the system correspondingly trained and tested. For a specified test sequence, the action label was assigned to the action frames. The proposed model correctly classified the majority of the actions (see the confusion matrices revealed below). The most frequent mistakes were in distinguishing between running and jogging and between boxing, clapping and waving [33], [35], [51]. The intuitive reasoning for this difficulty is based on the resemblance between these groups of movements. However, the presented model dramatically diminished this issue due to the episodic learning and ABM-based slowness features in the ventral processing stream.

The result of each human movement scenario is presented in Tables 1 and 2, which represent the accuracy of the proposed approach in comparison with earlier methods that utilized the same datasets. However, this comparison is not precise because of differences in the experimental setups. The presented results are comparable with state-of-the-art systems, whereas the consideration of the various methods involves multiple differences in their setups, such as being supervised or unsupervised and with or without tracking, subtraction of the background or consideration of multiple action recognition.

The evaluation of the proposed approach through two human action datasets was performed, and the confusion matrices are shown (see Figures 10 and 11). Here, the performance of the proposed model was compared with previous approaches that utilized the same dataset in Tables 1 and 2. Additionally, we should note that the different methods listed in Table 1 have many variations in their experimental setups, e.g., different splits of the training/testing data, whether some pre-processing (e.g., tracking, background subtraction) was needed, whether there was supervision, whether per-frame classification was performed and whether the method handled multiple action classes in a video. The results of our model had more stability due to the slow features in the ventral stream and their combination with fast features, although we agree that comparisons with other methods are not strictly fair, despite other methods not completely covering the biological point of view (e.g., [77]).

Discussion

How should the proposed model with slow and fast processing streams be gauged for the task of recognizing biological movement? The results presented here suggest that the combination of form and motion, i.e., slow and fast information, was performed by ABM-based IncSFA. The temporal features in terms of the episodic or frame considerations of the features that were attained through pathways represent slow and fast features, which are often not far from
Figure 12: Confusion matrices for the two pathways separately represent the accuracy of each processing stream when considering no interaction between the pathways. This representation was obtained from the KTH human action dataset for benchmarking without interaction of the paths. It justifies the performances of two patients, DF who developed visual agnosia (i.e., damage to the ventrolateral occipital cortex) and RV who developed optic ataxia (i.e., damage to the occipitoparietal cortex) (see more details in [107]).

Table 1: The recognition results using the proposed method are compared with state-of-the-art methods that utilized the KTH action dataset.

| Methods        | Accuracy          | Years |
|----------------|-------------------|-------|
| Schuldt. [77]  | 71.72%            | 2004  |
| Niebles. [101] | 83.33%            | 2008  |
| Schindler. [33]| 92.7%             | 2008  |
| Wang. [85]     | 91.2%             | 2009  |
| Danafar [35]   | 93.1%             | 2010  |
| Zhang. [51]    | U-SFA:84.67%      |       |
|                | S-SFA:88.83%      |       |
|                | D-SFA:91.17%      | 2012  |
|                | SD-SFA:93.50%     |       |
| Proposed Model | 90.07%            | 2015  |

the original model. It is certainly a different perspective to overlook the original recognition of biological movements [23] and present a model viewpoint based on slow and fast features in the mammalian visual system and its motion analysis temporal response [12] along with sensory information gathered over diverse time scales [97]. Consequently, with respect to the original model (which achieved good performance in the targeted databases (see [33], [34], [35])), it is necessary to objectively consider the recognition task. From a biological viewpoint, the understanding of biological movement contains both pathways, and cooperation between the information attained from the two processing streams occurs at several levels in the mammalian brain [36], [37]. It can also simply be the aggregation of the model (e.g., at the STS level [38]) and improve recognition performance. Through current neuroscience and psychophysics research, the form signals effects on motion processing have become more widespread in comparison with earlier findings (more details in [26], the holonomical features that consider both pathways features for the recognition of biological movements). In the form pathway, the proposed approach followed Pribram’s holonomic theory, which was
Table 2: Comparison of the proposed approach and previous methods that utilized the Weizmann human action dataset.

| Methods      | Accuracy   | Years |
|--------------|------------|-------|
| Schuldt. [77]| 72.8%      | 2004  |
| Niebles. [101]| 72.8%    | 2008  |
| Schindler. [33]| 100%    | 2008  |
| Wang. [85]   | 100%       | 2009  |
| Zhang. [51]  | U-SFA:86.67%|       |
|              | S-SFA:86.40%|       |
|              | D-SFA:89.33%|       |
|              | SD-SFA:93.87%|      |
| Proposed Model| 97.5%      | 2015  |

based on evidence that the dendritic receptive fields in sensory cortices have been described mathematically by Gabor functions [76] that are vastly utilized by the active basis model. From this stream, it is known that visual information is treated incrementally in a series of cortical stages (e.g., motion and orientation as local features in neurons at early levels, such as V1 [98]). ABM-based IncSFA [55], [79], [7] (basically all slow feature analysis methods) can be a useful tool for the extraction of slow features from modeling the form processing streams in the ventral stream. As previously discussed, the primary stage includes local (in V1 cells) and model detectors (Gabor like filters) in the sixteen (including eight preferred) orientations, and using the proper scale depends upon the receptive field (see [33], [41]).

However, widely invariant behavior is referred from the neuron response of the central nervous system [19] (e.g., early vision complex cells phase invariance [98] in hippocampal place cells of head direction invariance [99]). Human action cycles are unlikely to have invariant poses that are independent from the environment, different lighting conditions or the pose of the actions. These invariant forms of the actions can be an important criterion that represents the form processing stream information. The slowness principle applies the perception and inferences of neurons trained by these invariances by favoring slowly changing outputs in 2D (more details at [99]). A good implementation of this principle is SFA, which represents the mean square from the temporal derivative of the output and serves as a good model for the physiological properties of complex cells in the visual cortex [21] and, in particular, other invariances in the visual system [49]. This combination also played a role in the snapshot detectors with regard to human body shape model finding, such as in area IT (inferotemporal cortex) of the monkey, where the view-tuned neurons are located and the model of complex shape tuning [40] is implemented in the synergetic neural network. The IncSFA can be considered to be view-tuned neurons in area IT and snapshot neurons with regard to the independency of the scale and form. The proposed model followed the modeling through the slowest feature of the ventral stream and adjusted it through unsupervised learning methods. Fast features that used optical flow and inferred it with the slowness information of the other pathway can represent high-level integration of the snapshot neuron outcomes with the motion pattern neuron information.

CONCLUSION

In conclusion, the presented perspective of the recognition of the biological movement model followed the original model [23] in that the human visual system has two distinct pathways. These form and motion pathways are represented in the ventral and dorsal processing streams. The slowness principle using ABM-based IncSFA has been utilized for the extraction of form information (denoted as slow features). Optical flow generates the motion information considered to be fast features. The model analyzed the original recognition of biological movements by the perspective of a combination of fast and slow features. Furthermore, the integration of these features, i.e., slow and fast features, had good performance in terms of the recognition of human actions, which were evaluated through the
KTH and Weizmann human action datasets for benchmarking.

Acknowledgments

The authors would like to thank Ce Liu in Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology for layer-wised optical flow codes and Ying Nian Wu in UCLA (University of California, Los Angeles) for active basis model code. Also we are thankful to Matthew Luciw in the Swiss AI Lab IDSIA (Istituto Dalle Molle di Studi sull’Intelligenza Artificiale) for MATLAB codes of incremental slow feature analysis. We are very much grateful to Guang-Bin Huang and Zhou Hongming in Nanyang Technological University for their guidance in using ELM. We are thankful to Arthur Gretton in Gatsby Computational Neuroscience Unit and University College London for his short but very useful guidance in kernel methods and Diana Sima at the ESAT-SCD laboratory, Katholieke Universiteit Leuven for her help and comments. This research was sponsored by grants from: contract No. UM.C/HIR/MOHE/FRGS/10, High Impact Research (HIR) foundation in University Malaya (UM) Malaysia.

2. Reference

[1] A. Bray, D. Martinez, Kernel-based extraction of slow features: Complex cells learn disparity and translation invariance from natural images. Adv. Neural. Info. Process Syst. (NIPS) (2003) 269-276.

[2] J. V. Stone, Learning perceptually salient visual parameters using spatiotemporal smoothness constraints. Neural Comput. 8 (1996) 1463-1492.

[3] H. Nickisch, Extraction of visual features from natural video data using Slow Feature Analysis. Diploma thesis, (2006) Technische Universit"at Berlin.

[4] S. Liwicki, S. Zafeiriou, M. Pantic, Incremental slow feature analysis with indefinite kernel for online temporal video segmentation. Computer Vision ACCV (2013) 162-176. Springer Berlin Heidelberg.

[5] A. Levy, M. Lindenbaum, Sequential Karhunen-Loeve Basis Extraction and its Application to Images, IEEE Trans. Image Proc. 9(2000) 1371 1374.

[6] D. Ross, J. Lim, R. Lin, M. Yang, Incremental Learning for Robust Visual Tracking. Int. J. Comp. Vision 77 (2008) 125 141.

[7] V. Kompella, M. Luciw, J. Schmidhuber, Incremental Slow Feature Analysis. In Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume 2 (2012) (1354-1359) AAAI Press.

[8] W. Bhmer, S. Grnewlder, H. Nickisch, K. Obermayer, Regularized sparse kernel slow feature analysis. In Machine Learning and Knowledge Discovery in Databases (2011) 235-248 Springer Berlin Heidelberg.

[9] L. Wiskott, Slow feature analysis: A theoretical analysis of optimal free responses. Neural Comput 15 (2003) 2147-2177.

[10] T.J. Chin, D. Suter, Incremental kernel principal component analysis. IEEE Trans Image Proc 16(2007) 1662-1674.

[11] E. H. Adelson, J. R. Bergen, Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A. 2 (1985) 284-299.

[12] S. Shioiri, P. Cavanagh, ISI produces reverse apparent motion. Vision Res. 30 (1990) 757-768.

[13] S. Shioiri, K. Matsumiya, Motion mechanisms with different spatiotemporal characteristics identified by an MAE technique with superimposed gratings. J. Vision 9 (2009).

[14] D. Regan, Orientation discrimination for objects defined by relative motion and objects defined by luminance contrast. Vision Res. 29 (1989) 1389-1400.

[15] A. Johnston, D. H. Arnold,S. Nishida, Spatially localized distortions of event time. Curr. Biol. 16 (2006) 472479.

[16] K. Moutoussis, S. Zeki, A direct demonstration of perceptual asynchrony in vision. Proc. Biol. Sci. 264(1997) 393399.

[17] D. Whitney, I. Murakami, Latency difference, not spatial extrapolation. Nat. Neurosci. 1 (1998) 656667.

[18] A. O. Holcombe, Seeing slow and seeing fast: two limits on perception. Trends. Cogn. Sci. 13 (2009) 216-221.

[19] H. Sprekeler, C. Michaelis, L. Wiskott, Slowness: An Objective for Spike-Timing Dependent Plasticity?. PLoS Comput. Biol. 3(6) (2007) e112.
M. Franzius, H. Sprekeler, L. Wiskott, Slowness and sparseness lead to place, head-direction, and spatial-view cells. PLoS. Comput. Biol. 3 (2007) e166.

W. Hashimoto, Quadratic forms in natural images. Network-Comp. Neural 14 (2003) 765-788.

H. Sprekeler, C. Michaelis, L. Wiskott, Slowness: An objective for spiketiming-plasticity? PLoS Comput. Biol. 3 (2007) 112.

L. Wiskott, T.J. Sejnowski, Slow feature analysis: Unsupervised learning of invariances. Neural comput. 14 (2002) 715-770.

Y. Zhang, Z. Zhang, Y. Deng, S. Mahadevan, A biologically inspired solution for fuzzy shortest path problems. Appl. Soft Comput. 13(5) (2013) 2356-2363.

Z. Zhang, D. Tao, Slow feature analysis for human action recognition. Intelligence, IEEE Trans. Pat. Anal. Mach. 34 (2012) 436-450.

K.N. Kay, J. Winawer, A. Rokem, A. Mezer, B.A. Wandell, A Two-Stage Cascade Model of BOLD Responses in Human Visual Cortex. PLoS Comput. Biol. 9(5) (2013) e1003079. doi:10.1371/journal.pcbi.1003079.

B.A. Olshausen, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381 (1996) 607-609.

A.L. Yuille, P.W. Hallinan, D.S. Cohen, Feature extraction from faces using deformable emplates. Int. J. Comput. Vis. 8 (1992) 99-111.

Y.N. Wu, Z. Si, H. Gong, S.C. Zhu, Learning active basis model for object detection and recognition. Int. J. Comput. Vis. 90 (2010) 198-235.

S.C. Zhu, D.B. Mumford, A stochastic grammar of images. Found Trends Comput. Graph. Vis 2 (2007) 259-362.

M. Lades, J.C. Vorbrggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wrtz, W. Koen, Distortion invariant object recognition in the dynamic link architecture. IEEE Trans Comput 42(1993) 300-311.

B. Schlkopf, A. Smola, K.R. Miller, Nonlinear component analysis as a kernel eigenvalue problem. Neural comput. 10(1998) 1299-1319.

J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional cortical lters, J. Opt. Soc. Amer. 2 (1985) 1160-1170.

J.G. Daugman, Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression, IEEE Trans. Pat. Anal. Mach. Intel. 36 (1988) 1169-1179.

J.P. Jones, L.A. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J. Neurophysiol. 58 (1987) 1233-1258.

S. Marelja, Mathematical description of the responses of simple cortical cells. J. Opt. Soc. Amer. 70 (1980) 1297-1300.

J.G. Daugman, Two-dimensional spectral analysis of cortical receptive field profiles. Vis. Res. 20 (1980) 847-856.

B. Moghaddam, Principal manifolds and probabilistic subspaces for visual recognition. IEEE Trans. Pat. Anal. Mach. Intel. 24 (2002) 780-788.

G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks. IEEE Int. J. Conf. Neu. Net. (IJCNN) 2 (2004) 985-990.

G.B. Huang, D.H. Wang, Y. Lan, Extreme Learning Machines: A Survey. Int. J. Mach. Learning Cyb. 2(2011) 107-122.

G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications. Neurocomputing 70 (2006) 489-501.

R. Rajesh, J.S. Prakash, Extreme learning machines-a review and state-of-the-art. Int. J. Wisodom. Comput. 1 (2011) 35-49.

N.Y. Liang, G.B. Huang, P. Saratchandran, N. Sundararajan, A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans. Neural Net. 17 (2006) 1411-1423.

G.B. Huang, L. Chen, C.K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans. Neural Net. 17 (2006) 879-892.

C. Liu, Beyond pixels: exploring new representations and applications for motion analysis (Doctoral dissertation, Massachusetts Institute of Technology) (2009).

M. Aslan, A. Sengur, Y. Xiao, H. Wang, M.C. Ince, X. Ma, Shape Feature Encoding via Fisher Vector for Efficient Fall Detection in Depth-Videos. Applied Soft Computing (2015).
[73] T. Brox, A. Bruhn, N. Papenberg, J. Weickert, High accuracy optical flow estimation based on a theory for warping. Comput. Vision-ECCV (2004) 25-36, Springer Berlin Heidelberg.

[74] A. Bruhn, J. Weickert, C. Schnr, Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. Int. J. Comput. Vis. 61 (2005) 211-231.

[75] L. Alvarez, R. Deriche, T. Papadopoulo, J. Sneath, Symmetrical dense optical flow estimation with occlusions detection. Comput Vision-ECCV (2002) 721-735. Springer Berlin Heidelberg.

[76] K.H. Pribram, Brain and perception: Holonomy and structure in figural processing. Psychology Press (1991).

[77] C. Schuldt, I. Laptev, B. Caputo, Recognizing Human Actions: A Local SVM Approach. Proc. IEEE Intl Conf. Pat. Rec. 3(2004) 32-36.

[78] Y. Wang, G. Mori, Human action recognition by semilatent topic models. IEEE Trans. Pat. Anal. Mach. 31 (2009) 1762-1774.

[79] Y. M. Marghi, F. Towhidkhah, S. Gharibzadeh, A two level real-time path planning method inspired by cognitive map and predictive optimization in human brain. Appl. Soft Comput. 21 (2014) 352-364.

[80] Y. M. Marghi, F. Towhidkhah, S. Gharibzadeh, A two level real-time path planning method inspired by cognitive map and predictive optimization in human brain. Appl. Soft Comput. 21 (2014) 352-364.

[81] Rastegar, S., Babaiean, A., Bandarabadi, M., & Toopchi, Y. (2009, March). Airplane detection and tracking using wavelet features and SVM classifier. In System Theory, 2009. SSST 2009. 41st Southeastern Symposium on (pp. 64-67). IEEE.

[82] Tashk, A. R. B., & Faez, K. (2007, August). Boosted Bayesian kernel classifier method for face detection. In Natural Computation, 2007. ICNC 2007. Third International Conference on (Vol. 1, pp. 533-537). IEEE.

[83] U. Hasson, E. Yang, J. Vallines, D.J. Heeger, N. Rubin, A hierarchy of temporal receptive windows in human cortex. J. Neurosc. 28 (2008) 2539-2550.

[84] D. Hubel, T. Wiesel, Receptive fields and functional architecture of monkey striate cortex. J. Physiol. 195 (1968) 215-243.

[85] R. Muller, E. Bostock, J.S. Taube, J.L. Kubie, On the directional firing properties of hippocampal place cells. J. Neurosci. 14 (1994) 72357251.
[100] H. Sprekeler, C. Michaelis, L. Wiskott, Slowness: An Objective for Spike-Timing Dependent Plasticity?. PLoS Comput. Biol. 3 (2007) e112.

[101] J.C. Niebles, H. Wang, L. Fei-Fei, Unsupervised learning of human action categories using spatial-temporal words. Int. J. Compu. Vis. 79 (2008) 299-318.

[102] C. Liu, Gabor-based kernel PCA with fractional power polynomial models for face recognition. IEEE Trans. Pat. Anal. Mach. 26 (2004) 572-581.

[103] C. Liu, Optical Flow Matlab/C++ Code (2009), http://people.csail.mit.edu/celiu/OpticalFlow.

[104] Y.N. Wu, Z. Si, H. Gong, S.C. Zhu, Active Basis Model, Shared Sketch Algorithm, Sum-Max Maps for Representing, Learning, and Recognizing Deformable Templates (2010), http://www.stat.ucla.edu/~ywu/AB/ActiveBasisMarkII.html

[105] V.R. Kompella, M. Luciw, J. Schmidhuber, Incremental Slow Feature Analysis: Adaptive Low-Complexity Slow Feature Updating from High-Dimensional Input Streams (2012). http://www.idsia.ch/~luciw/incsfa.html

[106] H.M. Zhou, G.B. Huang, Kernelled ELM Algorithms, (2012) http://www.ntu.edu.sg/home/egbhuang/elmcodes.html.

[107] M.A. Goodale, J.P. Meenan, H.H. Blthoff, D.A. Nicolle, K.J. Murphy, C.I. Racicot, Separate neural pathways for the visual analysis of object shape in perception and prehension. Cur. Biol. 4 (1994)604-610.

[108] G. Johansson, Visual perception of biological motion and a model for its analysis. Percept. Psychophys. 14(1973) 201211.

[109] Rastegar, S., Babaeean, A., Bandarabadi, M., & Bahmani, G. (2009, March). Metric distance transform for kernel based object tracking. In System Theory, 2009. SSST 2009. 41st Southeastern Symposium on (pp. 54-58). IEEE.