*Research Article*

# Development of Biological Movement Recognition by Interaction between Active Basis Model and Fuzzy Optical Flow Division

## Bardia Yousefi and Chu Kiong Loo

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

Correspondence should be addressed to Chu Kiong Loo; ckloo.um@um.edu.my

Following the study on computational neuroscience through functional magnetic resonance imaging claimed that human action recognition in the brain of mammalian pursues two separated streams, that is, dorsal and ventral streams. It follows up by two pathways in the bioinspired model, which are specialized for motion and form information analysis (Giese and Poggio 2003). Active basis model is used to form information which is different from orientations and scales of Gabor wavelets to form a dictionary regarding object recognition (human). Also biologically movement optic-flow patterns utilized. As motion information guides share sketch algorithm in form pathway for adjustment plus it helps to prevent wrong recognition. A synergetic neural network is utilized to generate prototype templates, representing general characteristic form of every class. Having predefined templates, classifying performs based on multitemplate matching. As every human action has one action prototype, there are some overlapping and consistency among these templates. Using fuzzy optical flow division scoring can prevent motivation for misrecognition. We successfully apply proposed model on the human action video obtained from KTH human action database. Proposed approach follows the interaction between dorsal and ventral processing streams in the original model of the biological movement recognition. The attained results indicate promising outcome and improvement in robustness using proposed approach.

## 1. Introduction

Human action recognition in monocular video is one of the important subjects regarding video applications, for example, human computer interaction, video search, and so forth. It has been researched in different fields such as neurophysiology, psychophysics, and experimentations on imaging, and some areas in cortical engaged in it have been acknowledged. In general, human action recognition in the video stream using video processing and such methods in the proposed area is almost categorized by two techniques: one of them by using the global feature extraction form video streams tries to allocate a particular label to the whole video. The mentioned technique clearly needs unchanged observer within the video and considers the environments where actions occur [1].

Second technique considers local features regarding every frame and considers the label for distinct action. Afterward, by mechanisms of simple voting for global label regarding sequence can be attained. Temporal analysis for getting the features in every frame and classification is based on the observation in temporal window. It is important that both of these approaches have attained significant outcomes in such area (see [2]). One of the important factors of the complex action recognition and discriminating among unlike human motion styles and different individuals is learning [3] and also it is fundamental in recognition of 3D stationary human motion [4]. The human action recognition using frames of the video can be categorized as object recognition problem. It is supposed to handle the object variations (e.g., style, size, etc.). Meanwhile human brain is able to excellently categorize

human object in different classes of action; recent methods are inspired by biological outcomes of computational neuroscience [5, 6]. In the primary visual cortex (V1), procedure of images is more sensitive on barlike structures. Responses of V1 are combined together by extrastriate visual areas and passed to inferotemporal cortex (IT) for tasks of recognition [7].

## 2. Biologically Inspired Model

We follow the model of biological movement based on four assumptions, which are reliable by physiological and anatomical information [8]. The model splits into two corresponding preprocessing streams [9–12] parallel to dorsal and ventral streams which are specified for analysis of optic-flow and structure information, respectively. The model has used neural feature detector for extraction of optical flow and form features hierarchically considering size and style independency for both pathways; here we use synergetic neural network in both feedforward pathways for extraction of the structure and optical flow information. The corresponding results on the stationary human motion recognition reveal that discrimination can be accomplished through particularly small latencies, constructing an important role of top-down signals unlikely [9]. The motion is shown based on a set of patterns which are learned. The body shapes are determined by mentioned patterns like sequences of snapshots. The pathway regarding structure of body made by neurons and the complex patterns of motion pathway has been presented applying optical flow. This statement is a fundamental hypothesis of our model. The proposed model expands an earlier model used for the stationary objects [2, 7, 8, 12, 13] recognition by adding and combining the information over time in the dorsal and ventral pathway. Some visual physiologists have the regular belief regarding the proposed model [8]. It can be a good pertaining to quantity tool for organizing, summarizing, and interpreting existent information. The initial structure design is based on the data provided by neurophysiological and physiological evidences. This developed structure implements the quantitative estimation through computer simulations. Motion recognition and visual data have been involved in the model architecture. The proposed model has two separated pathways regarding form and motion information analysis. The information of two processing streams cooperates at few levels in the mammalian brains [14, 15]. Mentioned coupling is able to ease the model integration, for instance, in STS level [16] and it develops the performance of recognition without varying the fundamental results. Both pathways made up a hierarchy of neural feature detectors, which they use here for getting predefined templates regarding motion and form of the movement and make selective recognition.

*2.1. Form Pathway.* In the biological motion model, form pathway considers the system work as object recognition task and more involves in recognition of human body shape through sequential snapshots by camera. In general, few

models have been proposed which are plausible and neurophysiological about recognizing stationary form (e.g., [12]). Our proposed form pathway model follows an object recognition model [12] which is composed of form features detectors. The proposed approach has capability to be reliable like data obtained from neurophysiological information concerning scale, position, and sizes invariance which need further computational load along hierarchy. For modeling the cells in the primary visual cortex (V1) initial level of the structure pathway comprising detectors of local direction has been organized. Techniques having Gabor-like filters for modeling the detectors have good constancy by simple cells [17]. Two spatial scales by 2 factors for its differentiation and eight directions have been considered for the orientation detectors model. The neurons in monkey V1 range can influence the sizes of perceptive field in the receptive fields [18]. The scale and location detectors are located in the following level of this pathway that finds the information of local direction. Moreover, there is an approximated independency for scales and spatial location inside receptive fields. Perhaps, the complex-like cells in V1 area or in V2 and V4 are invariant regarding position varying responses (see [8]) and size independence is typical in area V4. These two areas (V2 and V4) are more selective for difficult form features, for example, junctions and corners whereas they are not suitable for recognition of the motion. To have an independent scale and position using mechanism of neurophysiologically through plausible model choosing the detectors responses by different directions and receptive field scales and locations. The pooling achieved through using maximum similarity operation as it mentioned in [19] some complex cells in cats visual cortex and areas V4 of macaques [20] reveal a maximum computing behavior. Afterward, the snapshots detectors are used for finding shapes of the human body model similar to area IT (inferotemporal cortex) of the monkey where the view-tuned neurons are located and model of complex shapes is tuned [21]. Snapshot neurons are similar to view-tuned neurons in area IT that give independent scale and position. Previous models used Gaussian Radial Basis functions for modeling and it adjusts during training. Then, the optimum set of training frames was analyzed and the learning rate was estimated (see [8, 12]). Neurons regarding motion pattern which are located in the form pathway highest level integrate the outcomes of snapshot neurons. Following presented biologically inspired methods regarding object recognition and its applications, we develop neurobiological model [2, 8, 22] of processing of the shape and motion in the dorsal stream in the visual cortex using active basis model as computational mechanisms into the feedforward aligned with motion pathway (optical flow).

*2.2. Motion Pathway.* In the motion pathway biological movements were recognized by using patterns of optical flow. The optical flow finds out the movement pattern which has consistency with neurophysiological information from hierarchy of neural detectors. In areas MT and V1 there are some neurons for motion and direction selection in first level of motion pathway, respectively. There are many

models for local motion estimation which are neurophysiologically plausible; we directly compute the response of motion-selective neurons and optical flow. Location form of the movement is analyzed in the second level of motion pathway. Local detector of optical flow is connected with motion patterns and the model comprises population of four directed neurons in area of MT. Also, motion edges selectors are sensitive in two opposite direction that find in areas of MT, MSTd, MSTl, and many parts of the dorsal steams and probably in the kinetic occipital area (KO) [8]. The proposed model and object specific motion position will be obtained by maximum pooling from motion position detector and considering motion selective edges which can be like MT [18] and MSTl [23] in macaque monkey. There is similarity between form pathway and third level motion pathway by snapshots neurons regarding model prediction. Motion pathway collects the optical flow patterns neurons outcome and smooths them to model them similarly with form pathway. The proposed approach active basis model [24] using spatiotemporal features intermediate complexity simulates the form pathway and optical flow [25] represents motion pathway for stimulating the proceeded areas MST and MT in dorsal stream where neurons have significant location and unchangeable scale. The obtained information will be processed by synergetic neural network [7, 13] and also predefined prototype which will be descried later. These patterns attain prototypes outcomes from active basis model and optical flow. Applying proposed approach is a simulation of both pathways in the primary visual cortex (V1) and projection of vertical stream in areas V2 and V4; see [8].

## 3. System Overview

The proposed system addresses a hierarchy of feature detectors which helps the system to be biologically inspired. Like [4, 8] and based on [8], we consider a basic consideration mechanism: our input is images obtained from video sequences and they have fixed size. The model is not sensitive regarding the background information.

*3.1. Active Basis Model for Form Pathway.* Active basis model [2] applying Gabor wavelets (for elements dictionary) offers deformable biological template; however, Gabor filters and Histograms of Oriented Gradients are presented for finding the human object. Shared sketch algorithm (SSA) followed through Ada-boost (like [26]). In every iteration, SSA following matching pursuit chooses an element of wavelet. It checks the objects number in different orientation, location, and scale. Selecting the small number of elements from the dictionary for every image (Sparse coding), therefore there can be representation of image using linear combination of mentioned elements by considering U as a minor residual. Consider

$$I = \sum_{i=1}^{n} c_i \beta_i + \epsilon, \tag{1}$$

where $\beta = (\beta_i, i = 1, \ldots, n)$ 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 [24]. By using wavelet sparse coding the large number of pixels reduces to small number of wavelet elements. Sparse coding can train natural patches of image to a Gabor-like wavelet elements dictionary which carries the simple cells in V1 properties [6, 24]. The extraction of local shapes will be separately done for every frame like [2] which computes the responses of filter orientation and density for every pixels. Also, the active basis model [24] uses the Gabor filter bank but in different form (see Figure 2). A Gabor wavelets dictionary, comprising $n$ directions and $m$ scales, is in the form of $GW_j(\theta, \omega)$, $j = 1, \ldots, m \times n$ where $\theta \in \{k\pi/n, k = 0, \ldots, n-1\}$ and $\omega = \{\sqrt{2}/i, i = 1, \ldots, m\}$. Gabor wavelet features signify the object form as a small variance in size and location and posture. The shape structure considers being safe maintained throughout the procedure. Response (convolution) to each element offers form information with $\theta$ and $\omega$. Consider

$$B = \langle GW, I \rangle$$
$$= \sum \sum GW(x_0 - x, y_0 - y : \omega_0, \theta_0) I(x, y). \tag{2}$$

Let $GW_j$ be a $[x_g, y_g]$, let $I$ be $[x_i, y_i]$ matrices, and let response of $I$ to $GW$ be a $[x_i + x_g, y_i + y_g]$. Therefore, previous convolution of both matrices must be padded through sufficient zeros. Consequence of convolution can be eliminated via cropping the result. Additional technique would be to shift back the center of the frequencies (zero frequency) to center of the image though it might reason for loosing data. Obtaining training image set $\{I^m, m = 1, \ldots, M\}$, the joint sketch algorithm consecutively chooses $B_i$. The fundamental opinion is to find $B_i$ so that its edge segments obtained from $I_m$ become maximum [24]. Afterward, it is necessary to compute $[I^m \cdot \beta] = \psi|\langle I^m \cdot \beta \rangle|^2$ for different $i$ where $\beta \in Dictionary$ and $\psi$ represents sigmoid, whitening, and thresholding transformations and then maximizing $[I^m \cdot \beta]$ for all possible $\beta$ will be computed. Let $\beta = (\beta_i, i = 1, \ldots, n)$ be the template for every training image $I^m$; scoring will be based on

$$M(I^m, \theta) = \sum_{i=1}^{n} \delta_i |I^m, \beta| - \log \Phi(\lambda \delta_i). \tag{3}$$

$M$ is the match scoring function and $\delta_i$ is obtained from $\sum_{n=1}^{M}[I^m, \beta]$ regarding steps selection and $\Phi$ is nonlinear function. The logarithmic likelihood relation of exponential model attains from the score of template matching. Vectors of the weight are calculated by maximum likelihood technique and are revealed by $\Delta = (\delta_i, i = 1, \ldots, n)$ [24]. Consider

$$\text{Max}(x, y) = \max_{(x,y) \in D} M(I_m, \beta). \tag{4}$$

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

*3.2. Optical Flow for Motion Pathway.* Using optical flow (as it is aforementioned) is one of the effective methods in human action recognition (see [27]). For having the features regarding motion of subject, layer-wise optical flow estimation has been done. A mask which reveals each layer's visibility is the main difference between estimation of traditional and layer-wise optical flow. The mask shape is able to perform fractal and arbitrary considering match will occur inside mask in the pixel-wised form (see [25]). We use the layer-wise optical flow method in [25] which has baseline optical flow algorithm of [28–30]. $M_1$ and $M_2$ are visible masks for two frames $I_1(t)$ and $I_2(t - 1)$, the field of flow from $I_1$ to $I_2$ and $I_2$ to $I_1$ is represented by $(u_1, v_1)$ and $(u_2, v_2)$. The following terms will be considered for layer-wise optical flow estimation. Objective function consists of summing three parts; visible layer masks match to these two images using Gaussian filter which called data term matching $E_\gamma^{(i)}$, symmetric $E_\delta^{(i)}$, and smoothness $E_\mu^{(i)}$. Consider

$$E(u_1, v_1, u_2, v_2) = \sum_{i=1}^{2} E_\gamma^{(i)} + \rho E_\delta^{(i)} + \xi E_\mu^{(i)}. \qquad (5)$$

After optimization of objective function and using outer and inner fixed-point iterations, image warping, and coarse to fine search, we attain flow for both bidirections. Compressed optic flow for all the frames is calculated by straight matching of template to the earlier frame by applying the summation of absolute difference ($L1$ − norm). Though optic flow is particularly noisy, no smoothing techniques have been done on it as the field of flow will be blurred in gaps and specially the places where information of motion is significant [22]. To obtain the proper response of the optical flow regarding its application in the proposed model, optical flow will be applied for adjusting the active basis model and making it more efficient. To achieve a representation reliable through the form pathway, the optic flow estimates the velocity and flow direction. The response of the filter based on local matching of velocity and direction will be maximal as these two parameters are continuously changing.

*3.3. Synergetic Neural Network Classifier.* Analyzing the human brain cognitive processes [31, 32], particularly the visual analysis, we apprehend that the brain is persistently involved in a big amount of the perception reprocessing, subconscious mind, filtering, decomposition, and synthesis. The brain of human is a cooperative system; in some cases, cognitive processes can be supposed to depend on the self-organizing pattern formation. Based on this knowledge, Haken presents synergetic neural network as one pattern recognition process which performs in the brain of the human. A joint method for association of trained samples is the values of feature averaging (see [33]). He revealed a collaborative pattern recognition of a top-down thinking: pattern recognition process can be comprehended like a specific order parameter competition process for recognition mode $q$ can construct a dynamic process, so $q$ after middle state $q(t)$ into a prototype pattern $v_k$. Though it is not flexible enough for direction changing, therefore the boundaries of

these templates are not clear. Applying learning object in the same view is a technique for dealing with inflexibility which will limit the task of classification. Algorithm of melting is introduced by [3] for objects combination in diverse pose. Assume a trained object sample $\acute{I}_i$ contains $n$ pixel values. By reshaping $\acute{I}_i$ to $v_i$ which is a column vector matrix and normalization we will have

$$\sum_{j=1}^{n} v_i j = 0, \qquad \sum_{j=1}^{n} v_{ij}^2 = 1, \qquad (6)$$

where $q$ is the input mode and $q_0$ is the initial values of the state vector for attention parameters, which we will discuss later. Connected prototype matrix $V^+$ calculates $V^+ = (V^+ V)V(1)$. Let $V$ be all the learn samples set $v_i = 1, \ldots, m$. and every column satisfies condition of orthonormal $v_k^+ v_j = \delta_{ij}$, for all $j$ and $k$, where $\delta_{ij}$ is delta of Kronecker. For a sample examination $q$, parameters of order signify test sampling matching. Class parameter of order for $k$ derives as $\epsilon_k = v_k^+$, $k = 1, \ldots, m$. Due to pseudoinverse overfitting, sometimes melting fails to generalize the learning. A penalty function is presented as Most Probable Optimum Design (MPOD) to improve the generalization and classify face object pose application (see [34]). Following this modification, the melting combination of similar object patterns into a template is useful for classification. So synergetic template is

$$v_p^+ = E(V^T V + P_1 O + P_2 I)^{-1} V^T. \qquad (7)$$

$I$, $O$, $P_1$, and $P_2$ are identity matrix, unitary matrix, and coefficients of penalty. $E$ is an enhanced identity matrix; every element of $E$ is a row vector of size $j$ as the following:

$$E = \begin{bmatrix} e_1^{n(1)} & e_0^{n(2)} & \cdots & e_0^{n(M)} \\ e_0^{n(1)} & e_1^{n(2)} & \cdots & e_0^{n(M)} \\ \cdots & \vdots & \ddots & e_0^{n(M)} \\ e_0^{n(1)} & e_0^{n(2)} & \cdots & e_1^{n(M)} \end{bmatrix}, \qquad (8)$$

$$e_0^i = (0, \ldots, 0), \qquad e_1^i = (1, \ldots, 1).$$

It can be a relevant feedback and self self-attentive for similarity measurement. But here the proposed model uses synergetic neural network two times, once for making the templates in each pathway and in the second time in the final classification.

*3.4. Fuzzy Optical Flow Division.* Fuzzy logic is a logic having multivalued that is originated from theory of fuzzy set found by Zadeh [35] and it deals with reasoning approximation [35]. It provides high level framework targeted at approximation reasoning which can appropriately deliver the imprecision and uncertainty together in linguistic semantics and model expert heuristics and handles requisite high level organizing principles. Fuzzy logic can be an important balancing method which is plausible and justifies combining approaches together for designing the classification, decision, and inference systems [36]. Various fuzzy inference systems have been proposed and have many applications through

Max-Min fuzzy operations. However, Leotamonphong and Fang mentioned that composition of max-min is suitable only when a system allows no compensation among the elements of a solution vector [37]. A time dependent fuzzy system also has been used many times regarding solution of control and classification and so forth. Chen and Liu [38] present a delay-dependent robust fuzzy control for a class of nonlinear delay systems via state feedback [38]. *Problem statement and preliminary*: after applying optical flow, the velocity of human object will be considered for both $x$ and $y$ directions. In general, $v_x, v_y \in \mathbb{R}^{m \times n}$ where $m$ and $n$ are sizes of image frame from input video stream.

(1) $\mu_{v_x}^{C_{1,2}}(x)$, $\mu_{v_x}^{C_{2,4}}(x)$, $\mu_{v_x}^{C_{1,2}}(y)$, and $\mu_{v_x}^{C_{2,4}}(y)$ are triangular membership functions for $v_x$ and it will be the same for $v_y$ velocity in $x$ and $y$ directions these represent the quaternion correlator for outputs of motion pathway. The fuzzification is done through triangular membership function as activation functions

$$\mu_{v_x}^{C_{1,2}}(x) = \begin{cases} \dfrac{x}{C_{1,3}} & 0 < x \le C_{1,3} \\ \dfrac{C_{2,4} - x}{C_{1,2} - C_{3,4}} & C_{1,3} < y \le C_{2,4}, \end{cases}$$

$$\mu_{v_x}^{C_{2,4}}(x) = \begin{cases} \dfrac{x - C_{1,3}}{C_{2,4} - C_{1,3}} & C_{1,3} < x \le C_{2,4} \\ \dfrac{C_{1,3} - x}{C_{1,3} - C_{2,4}} & C_{2,4} < x \le n, \end{cases}$$

$$\mu_{v_x}^{C_{2,4}}(y) = \begin{cases} \dfrac{y}{C_{3,4}} & 0 < y \le C_{3,4} \\ \dfrac{C_{1,2} - y}{C_{1,2} - C_{3,4}} & C_{3,4} < y \le C_{1,2}, \end{cases}$$

$$\mu_{v_x}^{C_{1,4}}(y) = \begin{cases} \dfrac{y - C_{3,4}}{C_{1,2} - C_{3,4}} & C_{3,4} < x \le C_{1,2} \\ \dfrac{m - y}{m - C_{1,2}} & C_{1,2} < y \le m. \end{cases} \tag{9}$$

The position of the highest velocity in $x, y$ estimated by evaluating the amount of membership functions and then membership function related to every cell will be based on aggregating $x, y$ for each velocity. It will evaluate both cases of velocities. $\mu_{v_x}^{C_i}$ and $\mu_{v_y}^{C_i}$ are showing the membership in each cell where $z$ is number of the cell ($i = 1, \ldots, 4$). Consider

$$\mu_{v_x}^{C_i} = \max \left\{ \mu_{v_x}^{C_i}, \mu_{v_y}^{C_i} \right\}. \tag{10}$$

(2) Determine the value of motion information in motion pathway in frame time **t**. As information of velocities can be unstable due to shaking the camera or different style in human object meanwhile he is acting in front of camera, the amount of velocity is dependent on time. The definition of time in this context is based on the frame time per second. Here, this dependency implements by considering the previous frame membership value. Proposed time dependent fuzzy optical flow division can be utilized for representing a class of optical flow divisions with fuzzy inference rules

concerning time regarding every frame of video stream as unit of the time is defined here, as follows:

$$\tilde{\mu}_v^{C_i}(t) = \tilde{\mu}_v^{C_i}(t - \tau) + \eta_v^{C_i}(t) \left( 1 - \tilde{\mu}_v^{C_i}(t - \tau) \right)$$
$$t \in [t_0, t_0 + k\tau], \quad k \in (0, 1, \ldots, N), \tag{11}$$

where $\tau$ is the frame time which is a parameter for camera and $k$ is numbers of frames pasted from the cell changing (it means $k$ will be reset after varying of the cell membership). $N$ is the maximum number of frame distance from present frame which does not unreasonably increase membership function value. We call $\eta_v^{C_i}(t)$ memory coefficient function and add to the membership function of the winner cell and define as follows:

$$\eta_v^{C_i}(t) = \begin{cases} \dfrac{1}{k + \beta} & \mu_v^{C_j}(t) \le \tilde{\mu}_v^{C_j}(t = \tau) \\ \dfrac{-1}{k + \beta'} & \mu_v^{C_j}(t) > \tilde{\mu}_v^{C_j}(t = \tau), \end{cases} \tag{12}$$

$$k \ge 0, \quad k \ne -\beta, \quad k \ne -\beta'$$

Let $\beta$ as adjustment parameter be tuned in the system. $C_j$ presents the cell which is different from $C_j$ and has maximum velocity among all cells in optical flow division. $t$ is the time of frame where one division of the optical flow has the highest membership amount as compared with other divisions and it will be restarted by changing the division.

(3) Gather values produced in previous memberships in every optical flow division in each frame by the following rules.

(a) Flow of upper limb is attained by association of optical flow fuzzy amounts for $C_1$ and $C_2$. Membership value reveals the flow for upper limb of human object. It is mentioned as follows:

$$\mu_{\text{Upper-Limb}}(t) = \tilde{\mu}_v^{C_1 \cup C_2}(t) = \max \left\{ \tilde{\mu}_v^{C_1}(t), \tilde{\mu}_v^{C_2}(t) \right\}. \tag{13}$$

(b) Flow related to lower limb calculates from union the amounts of optical flow in $C_1$ and $C_2$ with each other in time $t$:

$$\mu_{\text{Lower-Limb}}(t) = \tilde{\mu}_v^{C_3 \cup C_4}(t) = \max \left\{ \tilde{\mu}_v^{C_3}(t), \tilde{\mu}_v^{C_4}(t) \right\}. \tag{14}$$

*Optional.* Flow of left and right limb is calculated by considering the optical flow membership amount among $C_1, C_3$ and $C_2, C_4$, respectively:

$$\mu_{\text{Left-Limb}}(t) = \tilde{\mu}_v^{C_1 \cup C_3}(t) = \max \left\{ \tilde{\mu}_v^{C_1}(t), \tilde{\mu}_v^{C_3}(t) \right\},$$
$$\mu_{\text{Right-Limb}}(t) = \tilde{\mu}_v^{C_2 \cup C_4}(t) = \max \left\{ \tilde{\mu}_v^{C_2}(t), \tilde{\mu}_v^{C_4}(t) \right\}. \tag{15}$$

This part is optional suggested but we have not used it.

(4) For the following one fuzzy IF-THEN rule, perform defuzzification.

*R1s.* If every membership function from the subject has maximum degree in membership function as compared with

others, then the subject limits just some relevant action eligible for form pathway selection and selection possibility of other actions by form pathway is eliminated.

(5) Output of aforementioned membership values can be considered as belonging scores among the classes of actions which shows specific movements in human subject limbs. The biggest amount as the degree of belongs for each class will win among other amount. For example, running, jogging, and walking involve the lower limb activities whereas boxing, clapping, and waving make flow in the upper limb of human object.

*3.5. Selecting Video Frames.* Motion analysis, video processing, and action recognition are based on frame selection for temporal order. Choosing frames based on randomization methods of temporal order can destroy the biological perception of movement [8]. Frames selection through input movie follows proposed model of form and motion pathway connection from snapshot neurons. Snapshots follow temporal order regarding configuration motion patterns of object different activity in both pathways. Proposed model uses feedforward structure for form connection (active basis function) and motion pathway (optical flow). Three frames as minimum number of the frames for snapshots will be taken from video streams following temporal order and motion information induces active basis function through feedforward joining in share sketch algorithm and makes connection appropriate (see Figure 1).

*3.6. Relation to Existing Methods.* The proposed approaches like current techniques regarding human action recognition are basically very similar to each other. In this part we will signify differences and similarities. Similar with [2, 4, 8, 22], the proposed follows original model of biological movements. The approach is made based on object recognition following hierarchical feedforward designs like [22] and specially tries to develop a model that follows neurobiological motion processing in visual cortex and basically follows [8]. Object recognition task in form pathway has been changed within the researchers work from spatiotemporal features in [22, 40] and original Gabor filter [2] for proposed approach by using active basis model. However, active basis model has basic characteristic of previous features and basically uses Gabor wavelet but it decreases matching operation. It activates limited clutters and ensures the important amounts in points of interest which falls on the person subject. The motion feature which generated through layer-wised optical flow [25] has similarity with silhouette from moving object. In our work, it is used in helping active basic model to concentrate on the object and prevent wastage of Gabor beams. Additionally, as it is previously mentioned proposed approach follows biologically inspired model [8] through parallels to visual cortex.

## 4. Evaluation and Results

To estimate the ability of the proposed approach to human action recognition, a famous human action and the largest

databases, that is, the KTH human action dataset [39] and Weizmann human action recognition robustness set [31, 41], are implemented in the tests. KTH action dataset is the largest human action dataset including 598 action sequences that it comprises six types of single person actions as boxing, clapping, jogging, running, walking, and waving. These actions are performed by 25 people in different conditions: outdoors (s1), outdoors with scale variation (s2), outdoors with different clothes (s3), and indoors with lighting variation (s4). Here, using downsampling the sequences resolutions become 200 142 pixels. For our approach, we used 5 random cases (subjects) for training and making the form and motion predefined templates. As it is mentioned in the literature, KTH is a robust intrasubject variation with large set whereas the camera for taking the video during the preparation had some shacking and it makes the work with this database very difficult. Moreover, it has four scenarios which are independent, separately trained, and tested (i.e., four visually different databases, which share the same classes). Both alternatives have been run. For considering the symmetry problem of human actions, there is a mirror function for sequences along with vertical axis which can be available for testing and training sets. Here all possible overlapping of human actions within the training and testing sets has been considered (e.g., one video has 32 and 24 action frames).

*4.1. Contribution between Motion and Form Features.* Major strength compared with other human action recognition methods is utilizing fuzzy optical flow division for guidance of share sketch algorithm in active basis model. It combined the form and motion pathways with respect to original model. Regarding combination, a question may arise that is it necessary to combine these two pathways? and how do these two combine? The way that we have applied active basis model to form pathway and its adjustment using motion pathway information which makes the proposed method modified as compared with original model also using optical flow division guidance for active basis model is very much successful to prevention of Gabor beams wastage that it presents novelty as compared with common methods. We applied optical flow for updating active basis model point of application by evaluating the velocity of object with guidance of each optical flow division in form of fuzzy membership function (see Figure 3). It means that somehow information attained from motion pathway helps form pathway. However, combination of motion and form regularly overtakes both motion and form separately, in most of experiments conducted; combining information of these two pathways takes place in the final decision part (see [2, 8, 22]). Besides, relative feedforward structure from input data stream till final decision does not change and is similar across different datasets among two independent sets of features computed (see Figure 1 in [8] and Figure 2 in [2]). The proposed approach has been presented before [42] as the fuzzy optical flow divisions have not applied on it. Here, we have presented that with respect to the original model regarding both pathways, extracted features for each pathway can be relevant and feedforward structure has been
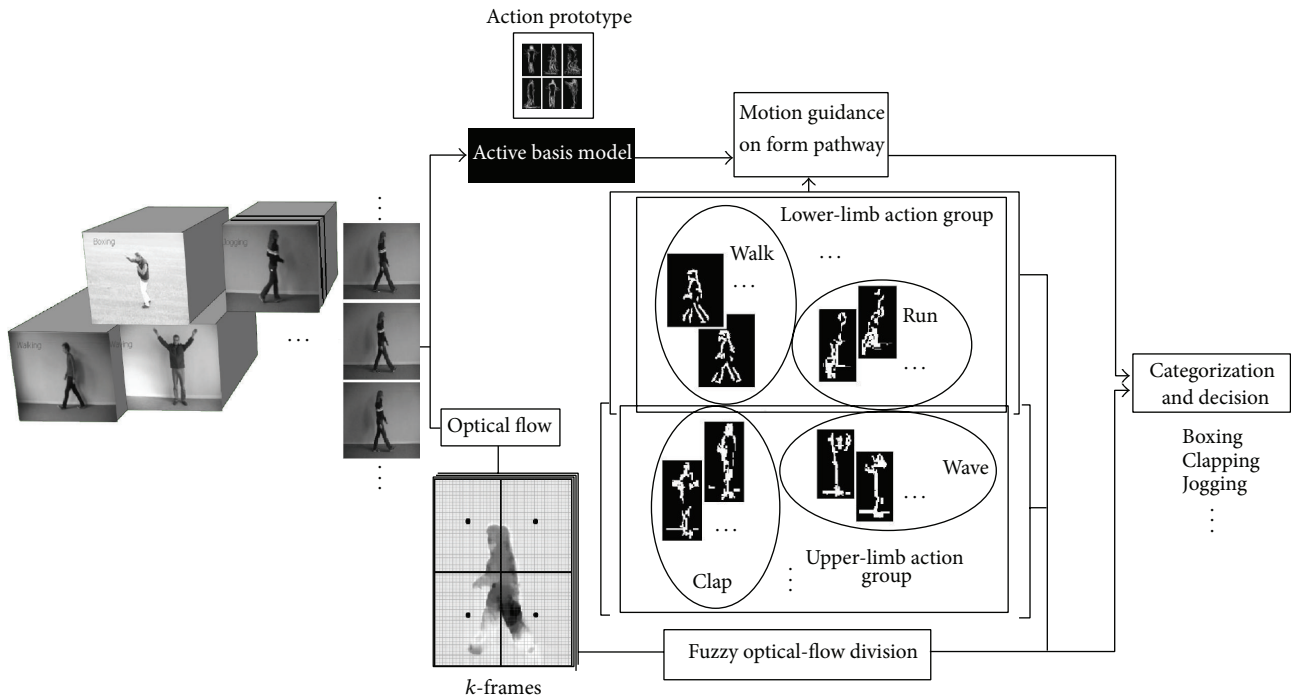
FIGURE 1: The figure reveals the flowchart of our algorithm regarding human action recognition. The flowchart presenting the hierarchical model in terms of theoretical and computation for combination of the form information and motion information is shown. A supervised Gabor based object recognition method, ABM, gives this property to have human object and computation of the form data and its combination with optical flow information, motion information, creates interaction between the pathways.
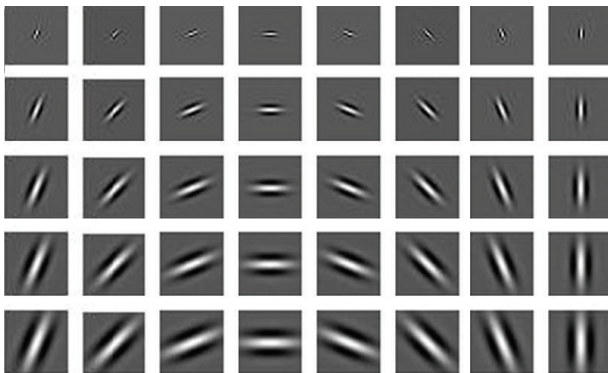


FIGURE 2: Gabor wavelets filter bank which has been used for the active basis model.

modified and extracted features for both pathways considered having dependent information.

*4.2. Results.* Here, the biologically inspired model for human action recognition has been studied. Principally, we have described the form features attained from active basis model that it represents features form from pathway. It is also mentioned that active basis model adjusts by motion pathway information and utilizing fuzzy optical flow division regarding adjustment for increasing the accuracy of recognition. Afterward, we proceeded for application of feature selection regarding the experiments. We also prepared action

prototype regarding every specific movement for human objects applying synergetic neural network. These templates are made by extracting two times prototypes from applying synergetic neural networks on train set of our human action dataset. Finally for examination of proposed approach, we applied the approach to very famous datasets which have been used for presentation of accuracy.

*4.3. How Are Action Prototypes Created?* As it is previously mentioned, predefined templates for each human action are obtained by applying the synergetic neural network on the human action image. For making the training map of every action, we divide every human action sequence to five primitive basic movements. One can create the whole action sequence using these five basic actions. Besides, considering the style invariance difficulties regarding diverse object in the same action, the proposed training map attains using five different subjects from targeted human action databases. For easing the explanation, we consider five snippets in different actions $A_1 - A_5$ and each subject from targeted database $D_1 - D_5$. First, synergetic neural network applies to $A_1$ in $D_1 - D_5$ and outcome shows by $P_1$ as first prototype obtains from first action snippet. The number of prototypes will be completed by applying the synergetic neural network and calculating the residual prototypes that they have called $P_1 - P_5$. Calculated prototype images considering style invariance represent the one action within five snapshots. Afterward, these prototypes melt together using second time synergetic neural network for attaining the final prototypes where each
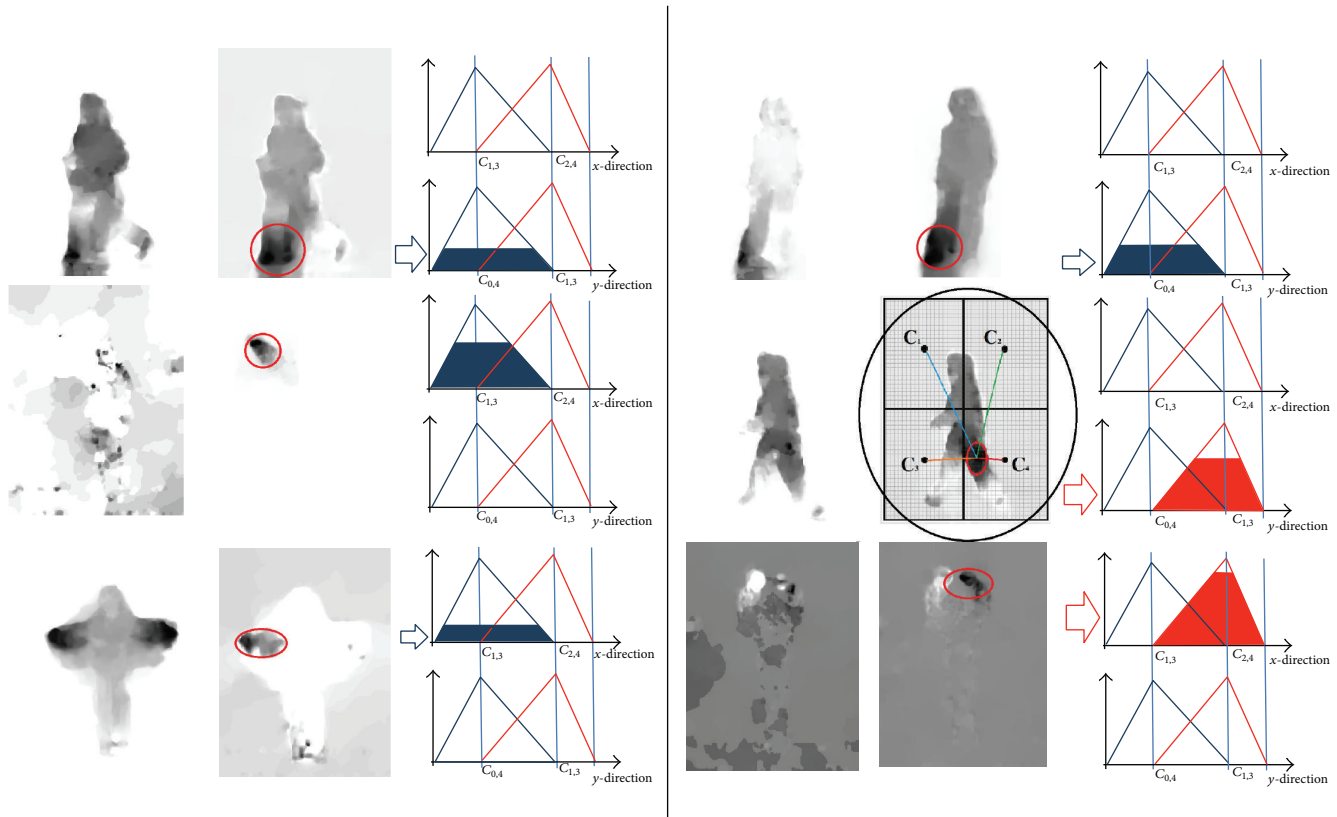
FIGURE 3: The results of dorsal processing stream obtains through optical flow [25] along with division procedures have been depicted. The resolution of divisions is designed for categorization of actions group to have additional interfere of dorsal and ventral processing streams. It can be a good representative of the interaction on MT, middle temporal of dorsal stream, and V4, ventral stream (for shape and orientation), or the MST area with inferior temporal (IT) (see more details in [32, 33]). The membership function of the action will be estimated from the position of maximum flow in the flow image. Membership values are aggregated through the proposed technique to increase the robustness. The input image of action mentioned in the figures is obtained from KTH human action recognition dataset [39].

of them represents the specific action within different action snippets and considering style invariance property. Let $F_t$ represents outcome of melting $P_1 - P_5$ in specified action. The final prototypes along with the method which they calculated are depicted in Figure 4. The confusion metric reveal the significant of using fuzzy inference system inside the model of the biological movement. The difference among these two confusion matrixes is very big and it can prove the advantage of using fuzzy optical flow division for this context. Furthermore, Table 1 reveals a comparison of our method with other methods in terms of recognition of accuracy in which the accuracy result indicates that the accuracy of proposed technique is relatively comparable with state-of-art by considering that there are somehow two categories using not very similar paradigms, which cannot be straightly compared. Here, the experimental result of proposed approach is presented. As KTH human action database [39] has been used for benchmarking the accuracy of consistency with set of experiments used in [2, 22, 27, 43, 44], we made a set of our training map and test set for proposed technique on the entire dataset, in which the mixture of four scenarios videos was together. The dataset split into a set of training maps with five randomly selected

subjects and a test part by residual subjects. Afterward, we measured the average performance over five random splits. The training map dataset was very small and comprised five videos frames snippets randomly obtained from the mixture dataset. Figure 5 presents classification confusion matrices for KTH dataset. Rows of confusion matrix represent the corresponding classification results, where each column signifies the instances to be classified. In proposed approach, the highest confusion happens among walking, jogging, and running. Discriminating these actions is difficult as the performance of actions by some subjects has similarity (see Figure 6). Also, another misclassification happens principally between similar classes, like previous confusion or hand clapping, hand, and waving (see confusion matrices in Figure 5).

### 4.4. Does the Fuzzy Optical Flow Division Help to Have Better Accuracy?

As it is mentioned in the previous parts of the paper, regarding obtaining action prototypes through synergetic neural network during one whole action frames. This method can give a good abstract from the action video but it has a problem [42] which decreases our accuracy due to

FIGURE 4: The figure depicts KTH human action dataset. To test the recognition of biological movements one of the well-known human action recognition datasets has been utilized in its performance. Here, the set represents KTH human action dataset. It is noticeable to mention that KTH dataset is one of the largest human action datasets having six various human actions in four different scenarios.
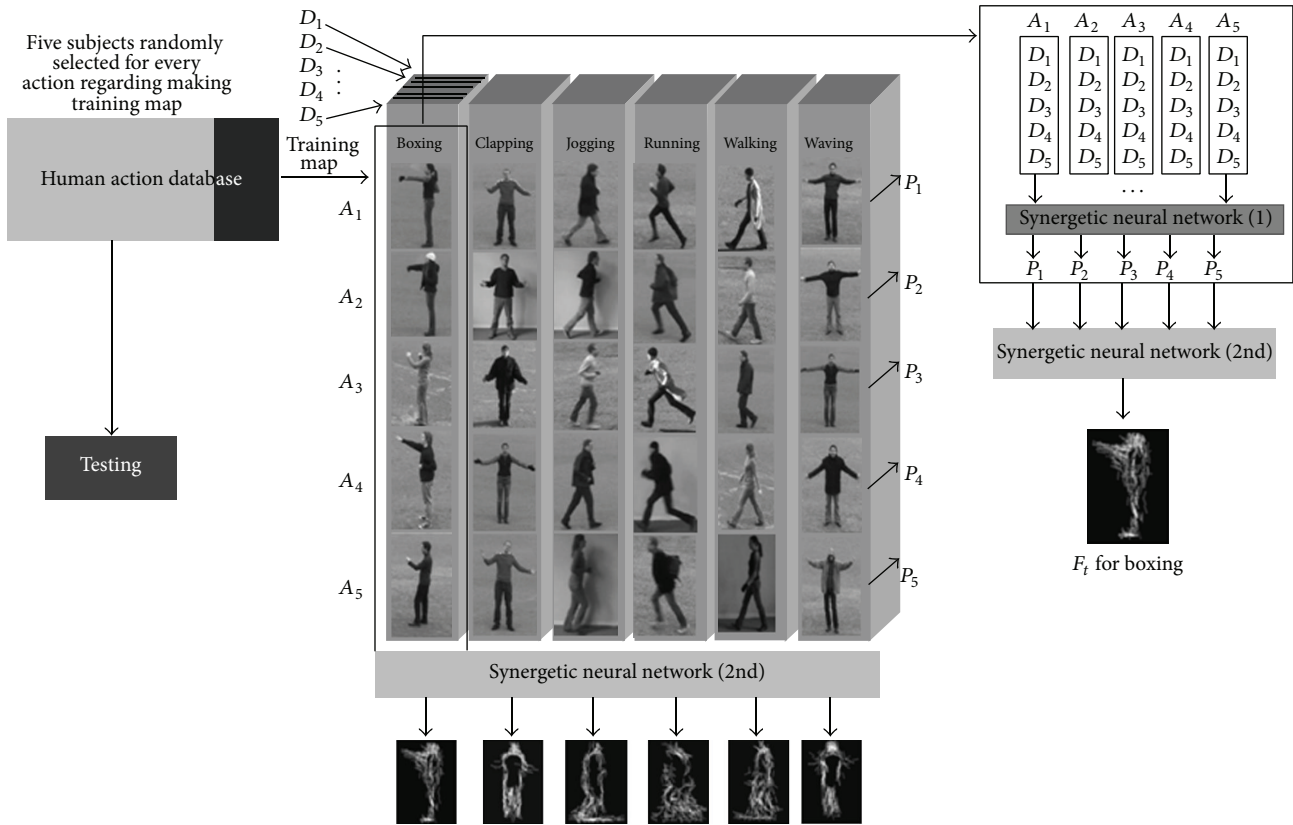


FIGURE 5: The figure shows procedure of making the action active basis templates by applying two times synergetic neural networks melting on the training map which calculates from randomly selected video frames from KTH human action database [39].

cluttered areas in the action prototypes. Following this problem, there was similarity among the matched image frames which is the cause of disparity in accuracy and it is very clearly revealed in confusion matrix (see Figure 5 upper confusion matrix). Presented approach has considered 0.1 second for, 2 for, $k = 3$ means three frame time considered as dependency of prevent membership function value and attained from training on our training set. During our experiment, we just have applied upper and lower limb membership functions and left and right limb functions can be suggested for more complex actions. After applying fuzzy optical flow division, disparity dramatically diminished. Confusion matrix after applying this method has been shown in Figure 5 (second confusion matrix).

*4.5. Related Work.* Human action recognition tasks are generally categorized as two separated classes. First class prefers

to track the part of image which is object (human) exists [45]. Mentioned groups of techniques might not be useful in less articulated objects. However, they are considered as successful approaches. The other popular class is addressed on low resolution videos or high locally resolution images [46] or by using spatiotemporal features [4, 40]. As it has previously been discussed regarding neurobiologically inspired model for analysis of movement in dorsal stream visual cortex and psychological and physiological information, our proposed approach is categorized as second group of methods. Previous method [8] has constant translation lack and a limited handcrafted features dictionary in intermediate periods [47], Jhuang et al. (2007) [22] and Schindler et al. [2].

*4.6. Discussion.* In the previous sections, we revealed the biologically inspired model regarding human action recognition using active basis model and fuzzy optical flow division.

**(a)**

| | Boxing | Clapping | Jogging | Running | Walking | Waving |
|---|---|---|---|---|---|---|
| Boxing | 0.94 | 0.001 | 0.02 | 0.01 | 0.001 | 0.02 |
| Clapping | 0.26 | 0.52 | 0.02 | 0.02 | 0.02 | 0.15 |
| Jogging | 0.02 | 0.01 | 0.70 | 0.06 | 0.15 | 0.04 |
| Running | 0.015 | 0 | 0.08 | 0.88 | 0.01 | 0.01 |
| Walking | 0.03 | 0 | 0.18 | 0.01 | 0.77 | 0.01 |
| Waving | 0.21 | 0.03 | 0.05 | 0.01 | 0.01 | 0.69 |

**(b)**

| | Boxing | Clapping | Jogging | Running | Walking | Waving |
|---|---|---|---|---|---|---|
| Boxing | 0.86 | 0 | 0 | 0 | 0 | 0.13 |
| Clapping | 0.07 | 0.81 | 0 | 0 | 0 | 0.11 |
| Jogging | 0 | 0 | 0.77 | 0.09 | 0.11 | 0 |
| Running | 0 | 0 | 0.09 | 0.91 | 0 | 0 |
| Walking | 0 | 0 | 0.17 | 0.02 | 0.80 | 0 |
| Waving | 0.13 | 0.16 | 0 | 0 | 0 | 0.71 |

FIGURE 6: Confusion matrices SNN classifying KTH dataset obtained by adapted active basis model as combination of form and motion pathways. Confusion matrices of the proposed approach has been presented for the case of without fuzzy interference system, left matrix, and after it, right matrix which are achieved from human action movements of KTH dataset [39]. The robustness of the method after adding the fuzzy interference stabilizer is considerably increased. The wrong recognitions in the left confusion matrix have been decreased especially in case of some actions, that is, clapping. Moreover, soar of robustness helps to increase the overall accuracy and better results in classification of biological movement. The accuracy of categorizations using unbalanced SNN reached 86.46%.

TABLE 1: The recognition results of proposed method has presented along with comparison between previous human action recognition method (Bio or non bio-inspired) on KTH human action dataset.

| Methods | Accuracy (%) | Years |
|---|---|---|
| Wang et al. [40] | 71.72 | 2004 |
| Niebles et al. [43] | 83.33 | 2006 |
| Jhuang et al. [22] | 91.79 | 2008 |
| Schindler and Van Gool [2] | 92.79 | 2009 |
| Wang and Mori [27] | 91.29 | 2009 |
| Zhang and Tao [44] | U-SFA: 86.67 S-SFA: 86.40 D-SFA: 89.33 SD-SFA: 93.87 | 2012 |
| Proposed Method | 86.46 | 2013 |

Here, we discuss advantages and limitations of our models, in addition to comparative explanation plus relationship of the proposed approach with existing models (see [2, 8, 48, 49]). We demonstrated how to apply a supervised learning Gabor based method which successfully has been utilized before for the task of object recognition previously [24] for the form pathway. As the form pathway is considered for ventral stream representation, it has a task of object recognition biologically. Active basis model can learn the human object considering the prototypes and is able to find it within frames. Such a property is very much desirable for visual system representation in the model; however, this part has been done by Gabor wavelet in the previous models [8, 48, 49] and similar works [2]. It could follow the involving encode object shape [32]. The object shape concern in the form pathway and ventral stream has been properly considered based on training stage and human prototypes. Using active basis model is somehow considered as the Gabor action stimulus for pin down form processing at two levels local information about limb angle from Gabor orientations and global body structure signaled by the spatial arrangement of Gabor paths. On the other hand, using optical flow for extraction of motion information has followed the second attribute and involves filtering by direction selection sensors and its integration for solving the famous aperture problem. Motion information presents both types of motion signals local velocity and joint motion trajectories will be signals to form path by guiding SSA in active basis model [50] as a good representation of cross-connection between V4 and MT. It follows the predominant view of form and motion processing in the human visual system which assumes that these two attributes are handled by independent and separate modules ([2, 8, 48, 49]). It has been recognized that form signal information can influence processing of motion more extensively than previously thought (see [32]) and the proposed approach considers direct effect on motion information on the form processing. The connectivity within the visual system is characterized by cross-connections in respect of parallel feedforward connection ([33, 51, 52]). Using optical flow division technique provides connection and interaction of bottom-up and top-down processing among brain regions along the dual computational streams. Also dorsal stream

is assumed to preform complementary spatial computation (where) and ventral stream for performing object recognition (what) in the cortical areas V1, V2, V4, and IT (inferotemporal cortex) along with current evidence in opposition to a complete segregation of where and what information in the brain of macaque (see [52, 53]) representing that information about position and size of objects is also represented in inferotemporal cortex of macaques as top layer of ventral stream. However, in the proposed approach an early isolation of spatial configuration and identity into divided processing pathways need heavy computation in hardware. But having low resolution optical flow divisions (four divided parts) could be a good parameter for diminishing this computational load. The correctly classified sequences are reported as highest results in literature. To place proposed technique in this context, we have presented it with state-of-the-art. Our method similar with other methods which is frame-based runs for all frames of action sequences. Then the individual labels obtained from training map are simply compared to a sequence label through majority voting (it is like a bag-of-frames model and like [2, 40, 48]). The comparison with state-of-the-art has been done and it is revealed in Table 1. It accuracy considering comparing with other methods indicates relative compatibility for proposed approach. In terms of contribution among motion and form features, we can mention applying active basis model which modified the form pathway itself considering it as a Gabor based model and its ability for learning the object increases the robustness of the system which is tested using Weizmann robustness dataset. Moreover, optical flow guidance for SSA as cross-connection among the dual computational streams plus prevention of application of Gabor beams for nontargeted objects is depicted in Figure 7. Considering fuzzy optical flow division keeps system robust, the proposed method can be categorized as an improvement in this field. However, the natural question (see [2, 48]) regarding whether this combination is necessary or how to improve it is still there and researchers are still trying to expand the model and make it more accurate. We have performed experiments following presented method, in which we have modified form pathway and made it combined with motion path and made a relation for these two independent feature sets and its connection which revealed promising results.

## 5. Conclusion

In this paper, a human action recognition method has been proposed; this method is based on interrelevant calculated motion and form information followed the biologically inspired system. The active basis model applied for generating the form information and optical flow guides the share sketch algorithm regarding better concentration on human object in the video frames and it can represent cross-connection of V4 and MT in brain [49]. Synergetic neural network is used twice on training set finding action prototypes for each action. The approach has been tested for KTH and robustness-Weizmann human action dataset and experimental assessment of the proposed technique has shown

promising results which was relatively comparable with state-of-the-art methods and benefit of proposed cross connection into the feed-forward method on the biological movement. Also it had good performance on different datasets and its training is done by less computational load regarding final action prototypes learning and reasonable computational cost. As a limitation of the proposed approach, it presently does not have mechanisms for invariance against rotation and viewpoint changes whereas it is capable to put mechanism regarding multiscale. Also active basis model is very sensitive algorithm which needs attention on while it is training. As open questions, motion sequences consistently represent recognition of video stream is done in which frame. How much these two pathways clearly follow the biologically inspired movement of mammalian brain. Future work will extend the proposed approach better integration of form and motion information in pathways. Another extension is to find better techniques to not using action prototypes regarding human action recognition.

## Appendix

In this part, we are presenting an example of proposed approach in fuzzy optical flow division part which is shown and proving the performance of time dependency in suggested fuzzy inference system for proposed application. Suppose that system in steady-state situation at the time of $t_0$, $\beta = \beta'$ is equal to 2 and $\tau = 0.1\,s$. Consider

$$\mu_v^{C_1}(t) = 0.6, \quad \mu_v^{C_2}(t) = 0.2,$$
$$\mu_v^{C_3}(t) = 0.4, \quad \mu_v^{C_4}(t) = 0.5. \tag{A.1}$$

Cell $C_1$ is winner and represents the highest optical flow. So, we will have timing influence by

$$k = 0, \qquad t = t_0, \qquad \widetilde{\mu}_v^{C_z}(t - \tau) = 0, \tag{A.2}$$
$$\widetilde{\mu}_v^{C_z}(t_0) = 0.6, \tag{A.3}$$

For $k = 1$, $t' = t_0 + \tau$,

$$\mu_v^{C_1}(t') = 0.59, \qquad \mu_v^{C_2}(t') = 0.3,$$
$$\mu_v^{C_3}(t') = 0.1, \qquad \mu_v^{C_4}(t') = 0.5. \tag{A.4}$$

As the highest cell did not change (condition of $\mu_v^{C_j}(t') \leq \widetilde{\mu}_v^{C_i}(t_0)$), having memory coefficient function ($\eta_v^{C_1}(t)$) chosen $1/(k + \beta)$ and considering $\widetilde{\mu}_v^{C_1}(t - \tau) = 0.6$, we will have

$$\widetilde{\mu}_v^{C_1}(t') = 0.6 + \frac{1}{1 + 2}(1 - 0.6) = 0.73. \tag{A.5}$$

Experimental outcomes of active basis model and outputs of motion pathway



(a)

Result of active bases for Wiezmann Robustness dataset
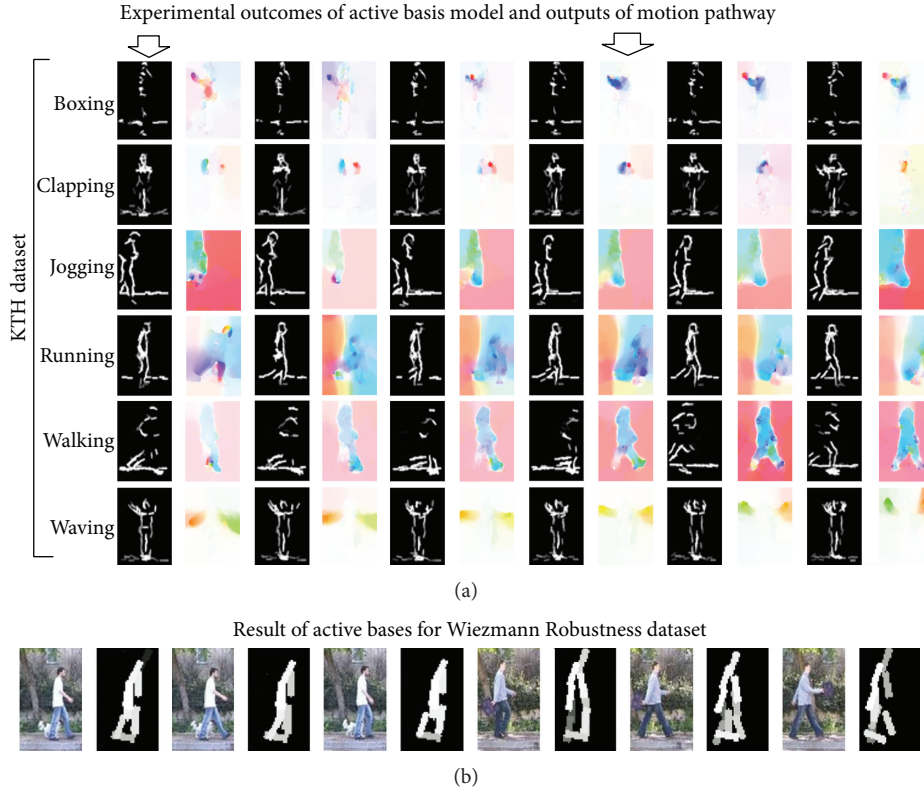


(b)

FIGURE 7: Simulation results for simple biological movement paradigm based on ABM [24] in the ventral processing stream and optical flow [25] in dorsal stream are shown. Each row within the panel reveals the response of ABM during the episode as well as flow generated for every different action. The set of biological movements that belongs to the biological movements is from KTH dataset [39]. (a) The simulation results of the different actions of KTH dataset along with results of optical flow simulation; (b) the figure depicts some results of Weizmann robustness dataset. It reveals an increase in the robustness of the proposed approach due to utilization of ABM [24] in the ventral stream.

Also having $N = 3$, we will have

$$k = 2, \qquad t'' = t_0 + 2\tau;$$

$$\mu_v^{C_1}\left(t''\right) = 0.7, \qquad \mu_v^{C_2}\left(t''\right) = 0.5,$$

$$\mu_v^{C_3}\left(t''\right) = 0.25, \qquad \mu_v^{C_4}\left(t''\right) = 0.45,$$

$$k = 3, \qquad t''' = t_0 + 3\tau;$$

$$\mu_v^{C_1}\left(t'''\right) = 0.8, \qquad \mu_v^{C_2}\left(t'''\right) = 0.1,$$

$$\mu_v^{C_3}\left(t'''\right) = 0.4, \qquad \mu_v^{C_4}\left(t'''\right) = 0.6, \qquad \widetilde{\mu}_v^{C_1}\left(t'\right) = 0.73,$$

$$\widetilde{\mu}_v^{C_1}\left(t'''\right) = 0.73 + \frac{1}{2+2}0.26 = 0.796.$$

$$(A.6)$$

After this by having maximum optical flow division for first cell, $k$ will be reset to zero and this cycle will be repeated by choosing the maximum membership function for next

step. But if suddenly after $k = 0$, some other cell gets highest optical flow, we will have

$$k = 3; \qquad t' = t_0 + \tau,$$

$$\mu_v^{C_1}\left(t'\right) = 0.4, \qquad \mu_v^{C_2}\left(t'\right) = 0.3, \qquad (A.7)$$

$$\mu_v^{C_3}\left(t'\right) = 0.6, \qquad \mu_v^{C_4}\left(t'\right) = 0.5.$$

$C_3$ won having the highest optical flow in its division, so cellular changing occurred but the result of the membership function is changing gradually and $1/(k + \beta')$ will be chosen for memory coefficient function and we will have

$$\widetilde{\mu}_v^{C_1}\left(t - \tau\right) = 0.73,$$

$$\widetilde{\mu}_v^{C_1}\left(t'\right) = 0.73 - \frac{1}{1+2}0.26 = 0.64.$$

As $\widetilde{\mu}_v^{C_1}(t') > \widetilde{\mu}_v^{C_3}(t')$ still first cell wins

$$k = 2; \qquad t'' = t_0 + 2\tau,$$

$$\mu_v^{C_1}\left(t''\right) = 0.2, \qquad \mu_v^{C_2}\left(t''\right) = 0.5,$$

$$\mu_v^{C_3}\left(t''\right) = 0.68, \qquad \mu_v^{C_4}\left(t''\right) = 0.6,$$

$$\widetilde{\mu}_v^{C_1}\left(t'\right) = 0.64, \qquad \widetilde{\mu}_v^{C_1}\left(t''\right) = 0.64 - \frac{1}{2+2}0.36 = 0.55.$$

$$(A.9)$$

And by considering the time when first cell won this amount will be increased but if the change continues third cell will win and we will have reset in $k$ and final choosing of the cell. Also, if the amount of membership function changes in $k = 2$ and back to the previous cell, we will have

$$k = 2; \quad t'' = t_0 + 2\tau,$$

$$\mu_v^{C_1}\left(t''\right) = 0.7, \qquad \mu_v^{C_2}\left(t''\right) = 0.5,$$

$$\mu_v^{C_3}\left(t''\right) = 0.4, \qquad \mu_v^{C_4}\left(t''\right) = 0.6,$$

$$\widetilde{\mu}_v^{C_1}\left(t'\right) = 0.64, \qquad \widetilde{\mu}_v^{C_1}\left(t''\right) = 0.64 - \frac{1}{2+2}0.36 = 0.73.$$

$$(A.10)$$

And there will be no sudden changes on final decision. It means first cell will be constantly selected. Also by selecting the suitable $\beta'$ the system will have minor sensitivity regarding cellular varying. But it is abundantly related to the application and video condition.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

## References

[1] M. J. Santofimia, J. Martinez-del-Rincon, and J. C. Nebel, "Episodic reasoning for vision-based human action recognition," *The Scientific World Journal*. In press.

[2] K. Schindler and L. Van Gool, "Action Snippets: how many frames does human action recognition require?" in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 3025–3032, Anchorage, Alaska, USA, June 2008.

[3] T. Hogg, D. Rees, and H. Talhami, "Three-dimensional pose from two-dimensional images: a novel approach using synergetic networks," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1140–1144, Perth, WA, Australia, December 1995.

[4] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," in *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV '03)*, pp. 726–733, Nice, France, October 2003.

[5] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, vol. 20, no. 10, pp. 847–856, 1980.

[6] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[7] M. Riesenhuber and T. Poggio, "Neural mechanisms of object recognition," *Current Opinion in Neurobiology*, vol. 12, no. 2, pp. 162–168, 2002.

[8] M. A. Giese and T. Poggio, "Neural mechanisms for the recognition of biological movements," *Nature Reviews Neuroscience*, vol. 4, no. 3, pp. 179–192, 2003.

[9] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, no. 6582, pp. 520–522, 1996.

[10] M. W. Oram and D. I. Perrett, "Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey," *Journal of Neurophysiology*, vol. 76, no. 1, pp. 109–129, 1996.

[11] G. Johansson, "Spatio-temporal differentiation and integration in visual motion perception—an experimental and theoretical analysis of calculus-like functions in visual data processing," *Psychological Research*, vol. 38, no. 4, pp. 379–393, 1976.

[12] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999.

[13] M. Riesenhuber and T. Poggio, "Models of object recognition," *Nature Neuroscience*, vol. 3, pp. 1199–1204, 2000.

[14] Z. Kourtzi and N. Kanwisher, "Activation in human MT/MST by static images with implied motion," *Journal of Cognitive Neuroscience*, vol. 12, no. 1, pp. 48–55, 2000.

[15] K. S. Saleem, W. Suzuki, K. Tanaka, and T. Hashikawa, "Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey," *Journal of Neuroscience*, vol. 20, no. 13, pp. 5083–5101, 2000.

[16] M. A. Giese and L. M. Vaina, "Pathways in the analysis of biological motion: computational model and fMRI results," *Perception*, vol. 30, pp. 119–119, 2001.

[17] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.

[18] B. M. Dow, A. Z. Snyder, R. G. Vautin, and R. Bauer, "Magnification factor and receptive field size in foveal striate cortex of the monkey," *Experimental Brain Research*, vol. 44, no. 2, pp. 213–228, 1981.

[19] I. Lampl, D. Ferster, T. Poggio, and M. Riesenhuber, "Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex," *Journal of Neurophysiology*, vol. 92, no. 5, pp. 2704–2713, 2004.

[20] T. J. Gawne and J. M. Martin, "Responses of primate visual cortical V4 neurons to simultaneously presented stimuli," *Journal of Neurophysiology*, vol. 88, no. 3, pp. 1128–1135, 2002.

[21] N. K. Logothetis, J. Pauls, and T. Poggio, "Shape representation in the inferior temporal cortex of monkeys," *Current Biology*, vol. 5, no. 5, pp. 552–563, 1995.

[22] H. Jhuang, T. Serre, L. Wolf, and T. Poggio, "A biologically inspired system for action recognition," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1253–1260, Rio de Janeiro, Brazil, October 2007.

[23] S. Eifuku and R. H. Wurtz, "Response to motion in extrastriate area MST1: center-surround interactions," *Journal of Neurophysiology*, vol. 80, no. 1, pp. 282–296, 1998.

[24] Y. N. Wu, Z. Si, H. Gong, and S.-C. Zhu, "Learning active basis model for object detection and recognition," *International Journal of Computer Vision*, vol. 90, no. 2, pp. 198–235, 2010.

[25] C. Liu, *Beyond pixels: exploring new representations and applications for motion analysis [Ph.D. thesis]*, Massachusetts Institute of Technology, 2009.

[26] J. Tanha, M. Van Someren, and H. Afsarmanesh, "An AdaBoost algorithm for multiclass semi-supervised learning," in *Proceedings of the IEEE 12th International Conference on Data Mining (ICDM '12)*, pp. 1116–1121.

[27] Y. Wang and G. Mori, "Human action recognition by semilatent topic models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1762–1774, 2009.

[28] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sánchez, "Symmetrical dense optical flow estimation with occlusions detection," *International Journal of Computer Vision*, vol. 75, no. 3, pp. 371–385, 2007.

[29] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proceedings of the European Conference on Computer Vision*, pp. 25–36, Prague, Czech Republic, 2004.

[30] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods," *International Journal of Computer Vision*, vol. 61, no. 3, pp. 211–231, 2005.

[31] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 1395–1402, October 2005.

[32] G. Mather, A. Pavan, R. B. Marotti, G. Campana, and C. Casco, "Interactions between motion and form processing in the human visual system," *Frontiers in Computational Neuroscience*, 2013.

[33] L. L. Cloutman, "Interaction between dorsal and ventral processing streams: where, when and how?" *Brain and Language*, vol. 127, no. 2, pp. 251–263, 2013.

[34] G. C. Lee and C. K. Loo, "Facial pose estimation using modified synergetic computer," in *Proceedings of the 2nd World Congress on Nature and Biologically Inspired Computing (NaBIC '10)*, pp. 322–327, Fukuoka, Japan, December 2010.

[35] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[36] S. Kumar, *Neural Networks: A Classroom Approach*, McGraw-Hill International Editions, 2004.

[37] J. Leotamonphong and S. Fang, "An efficient solution procedure for fuzzy relation equations with max product composition," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 4, pp. 441–445, 1999.

[38] B. Chen and X. Liu, "Delay-dependent robust H∞ control for T-S fuzzy systems with time delay," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 544–556, 2005.

[39] C. Schüldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local SVM approach," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, pp. 32–36, Cambridge, UK, August 2004.

[40] B. Wang, Y. Liu, W. Wang, W. Xu, and M. Zhang, "Multi-scale locality-constrained spatiotemporal coding for local feature based human action recognition," *The Scientific World Journal*, vol. 2013, Article ID 405645, 11 pages, 2013.

[41] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2247–2253, 2007.

[42] B. Yousefi, C. K. Loo, and A. Memariani, "Biological inspired human action recognition," in *Proceedings of the IEEE Workshop on Robotic Intelligence in Informationally Structured Space (RiiSS '13)*, vol. 4, pp. 63–70, Singapore, 2013.

[43] J. C. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," *International Journal of Computer Vision*, vol. 79, no. 3, pp. 299–318, 2008.

[44] Z. Zhang and D. Tao, "Slow feature analysis for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 436–450, 2012.

[45] C. Bregler, J. Malik, and K. Pullen, "Twist based acquisition and tracking of animal and human kinematics," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 179–194, 2004.

[46] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Proceedings of the 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS '05)*, pp. 65–72, Beijing, China, October 2005.

[47] C. Fanti, L. Zelnik-Manor, and P. Perona, "Hybrid models for human motion recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 1166–1173, San Diego, Cailf, USA, June 2005.

[48] S. Danafar, A. Gretton, and J. Schmidhuber, "Characteristic kernels on structured domains excel in robotics and human action recognition," in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 264–279, Springer, 2010.

[49] N. Kruger, P. Janssen, S. Kalkan et al., "Deep hierarchies in the primate visual cortex: what can we learn for computer vision?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1847–1871, 2013.

[50] S. M. Thurman, "Complex interactions between spatial, orientation, and motion cues for biological motion perception across visual space," *Journal of Vision*, vol. 13, no. 2, p. 118, 2013.

[51] C. Distler, D. Boussaoud, R. Desimone, and L. G. Ungerleider, "Cortical connections of inferior temporal area TEO in macaque monkeys," *Journal of Comparative Neurology*, vol. 334, no. 1, pp. 125–150, 1993.

[52] C. P. Hung, G. Kreiman, T. Poggio, and J. J. DiCarlo, "Fast readout of object identity from macaque inferior temporal cortex," *Science*, vol. 310, no. 5749, pp. 863–866, 2005.

[53] S. R. Lehky, X. Peng, C. J. McAdams, and A. B. Sereno, "Spatial modulation of primate inferotemporal responses by eye position," *PLoS ONE*, vol. 3, no. 10, Article ID e3492, 2008.