Biologically-Inspired Computational Neural Mechanism for Human Action/activity Recognition: A Review

Bardia Yousefi and Chu Kiong Loo*

1 Computer Vision and Systems Laboratory (CVSL), Department of Electrical and Computer Engineering, Laval University, Quebec City, QC G1V 0A6, Canada; Bardia.Yousefi.1@ulaval.ca
2 Advanced Robotic Laboratory, Department of Artificial Intelligence, University of Malaya, Kuala Lumpur 50603, Malaysia; ckloo.um@um.edu.my
* Correspondence: Bardia.Yousefi.1@ulaval.ca; Tel.: +1-267-669-9954

Received: 30 September 2019; Accepted: 8 October 2019; Published: 15 October 2019

Abstract: Theoretical neuroscience investigation shows valuable information on the mechanism for recognizing the biological movements in the mammalian visual system. This involves many different fields of research such as psychological, neurophysiology, neuro-psychological, computer vision, and artificial intelligence (AI). The research on these areas provided massive information and plausible computational models. Here, a review on this subject is presented. This paper describes different perspective to look at this task including action perception, computational and knowledge based modeling, psychological, and neuroscience approaches.

Keywords: human visual system; recognition of biological movement; human action recognition; biologically inspired model; ventral and dorsal streams interaction

1. Introduction

Analysis of biological motion recognition is categorized in different research fields, such as neurophysiology, neuropsychological, artificial intelligence (AI), and computer vision. The present article mainly investigates the AI perceptive of this task concerning neurophysiology, and neuropsychological evidence. This considers a subdivision of a computer science field dealing with machine intelligent behavior and learning. The primary experimental methods on this field have been classified into two main divisions:

1. **symbolic methods**, which follow classic approaches, are similar to expert systems and termed as connectionism approaches;
2. **scruffy methods** concentrate on the intelligence evolution or following artificial neural networks.

Both directions undergo rigorous restrictions. The untimely objectives, such as human behavior reproduction and simulation, are entirely overlooked. Machine and man have opposite abilities: man can estimate, infer, and recognize in parallel, whereas machine can sequentially perform quick computations. Biologically inspired models are similar to behavior-based AI, and have attracted attention of scientific communities. These models are more focused on the performance than on the internal processing of the machine. Many projects have intensive collection of the mentioned facts, but none can create a machine that takes direct advantage of this information. Modern machines can learn on the basis of statistics and follow few determined objectives. One objective is to deal with information extraction from large datasets, unsupervised learning, pattern recognition, and calculation based on statistical analysis statements. Such machines are practically used for actual problems, such as speech, image, and object recognition. Biological movement and its recognition comprise a multifield...
research that follows many biological principles and engineering approaches, and are based on the human (or mammalian) visual system. In the present study, computational model for recognition of biological movement is further reviewed in terms of mechanisms and models proposed for this task.

2. Motivation

Human action recognition in monocular video is an important subject regarding video applications, such as in human computer interaction, video search, and many other relevant tasks. Biologically-inspired recognition models provide more insight regarding the performance of such approaches compared to the mammalian visual system following neurophysiological, and psychophysical evidence. Understanding about the mammalian visual system provides an opportunity to improve the artificial intelligence (AI) models to be stronger with higher similar capability to human. This paper reviews different aspects of the biologically-inspired mechanisms for recognition of human action into the following overall categories:

- Perception of the motion
- Knowledge based modeling of the human action
- Psychology and neurophysiology of the motion

To better follow the different aspects of biological human action, these subjects are divided into several smaller subsections (there are other similar categorizations that exist i.e.,[1]).

Figure 1 represents a categorization of human action recognition based on neuro-physiological evidence and computer based modeling.

![Human action recognition diagram](image)

Figure 1. A categorization of human action recognition methods in different methodological perspectives.

3. Analysis of Biological Movement

3.1. Motion Patterns

Human action recognition is summarized by automatically determining the type of human action (human as a moving object in video sequences (image frames)). Motion recognition has been studied in neurophysiological, psychophysical, and imaging experiments. Marey and Muybridge carried out the initial studies on human movement in the 1850s; they photographed moving subjects and presented on its locomotion [2]. One of the earliest studies on visual perception and introduction of actual movement was conducted by Rubin (1927) and Duncker (1929) [3,4]. Johansson et al. (1973) [5] presented the initial part of a study on moving object characteristic and movement perception (similar to [6]). In general, the methods for motion patterns can be categorized into two different paths:
1. One technique uses global feature extraction from video streams to allocate a particular label to the whole video. This technique clearly needs an unchanged observer within the video, and the environments where actions are occurring should be considered [7].

2. The second technique considers local features in each frame and label for distinct action. Afterward, sequences can be attained through simple voting for global labeling. Temporal analysis for obtaining the features in each frame and classification is based on the observation in temporal window.

Both approaches showed significant outcomes [8–10]. Learning is also fundamental in recognizing 3D stationary human motion [11]. Human action recognition using video frames categorizes as an object recognition problem. Such recognition is supposed to handle object variations (e.g., style and size), and the human brain can excellently categorize human objects in different classes of action (recent bio-inspired methods in computational neuroscience [12,13]). In the primary visual cortex (V1), the image procedure is particularly sensitive on bar-like structures (Gabor-like methods and showed in Figure 2). V1 responses are combined by extrastriate visual areas and passed to the inferotemporal cortex (IT) for recognition [14].

![Neural activities for a comparison with functional imaging experiments](image)

**Figure 2.** A illustrative similarity comparison of MLD and ABM [15] from the perception of biological movements. These basis have considerable consistency to the point light technique (MLD), which presents static pictures. ABM also has a good representation of biological movements in form pathway. Picture is adapted from [16].

### 3.2. Kinetic–Geometric Model

Kinetic–geometric model presents the analysis of visual vector and basically expansion in mechanical form when biological motion perception and its patterns are combined. A classic moving light display (MLD) provides an excellent impetus for human motion perception in neuroscience analysis and study [5,17]. Recognition of human walker’s gender has been performed without knowledge cues by point light sources established on important joints of the human body, unlike versions of statistical experiments that have sufficient accuracy in this task. Changes in the speed of walking and degree of arm-swing especially in higher speeds are associated with females and upper body joints in lights analysis which are better in finding the accuracy for gender recognition [18]. Marr et al. (1978) showed the problem of computational process in human visual system and information obtained via retinal images; 3D shapes are considered for problem presentation by introducing some notes as follows: three criteria are introduced for shape recognition in judging, three aspects of design representation are considered (e.g., coordinate, primary shape unit information, and information organization), shape description (e.g., coordination of the object center, size variations, modular organization, view transferring mechanisms, and identification of natural axes), and constraints for conservation recognition applying further information from the image. Perrett et al. (1985) reported on the temporal cortex of macaque monkey; they found that most of the cells in the
brain region are sensitive to the type of movement and respond to specific body movements [19] (models showed similar behavior applying different receptive field showed in Figure 3). Two cell types introduced are sensitive to the rotation and view of body movements and the response of majority of cells in these areas of temporoparieto-occipital and PGA of temporal cortex has been considered for providing descriptions of view-centered and view-independent responses among the mentioned cells [20]. Goddard (1989) used connectionist techniques, along with spatial and temporal feature incorporation through diffused MLD data, and represented the walking recognition in 400 ms MLDs [21]. This integration occurs in the low-level features of shape and motion by this target to make high-level features. Low-level features include sequential trajectories points, and they are grouped in line segmentation with others to obtain proper lower and upper body limb forms (or even correlated features [22]). Figure 2 represents the similarity between MLD and active basis model (ABM) [15,16]. An unsupervised approach performs synthetic biological movement recognition [23] and shows great potential for use in the mechanism of biological movements and the importance on geometric model implying synthetic data.

Figure 3. (A) Schema of the model shown and symbols are shown following the brain areas and their functionality: MT: middle temporal area; V1: primary visual cortex; FFA: fusiform face area; STS: superior temporal sulcus; KO: kinetic occipital area. These areas and their functionalities are considered in their timing $t_1, t_2, \ldots, t_n$ for the input data frames and their encoded information gathered by radial base function and optical flow. In addition, (a) reveals the opponent motion detector; (b) shows the lateral coupling in complex optical flows; (c) response of the motion pattern detector. [24]); (B) The language production and perception performance review is shown in the figure [25]); (C) illustrate the model presented in [26] with concerns of receptive field as well.

3.3. What and Where Pathways

Shape and form pathways are hierarchically joined to detect the three levels of complexity, i.e., component, segment, and assembly levels that signify temporal series on procedures [21]. Goddard followed the biologically inspired human action recognition in determining the complex structured motion by using MLDs. He analyzed major computational problems, such as time-varying representation, visual stimuli integration, gestalt formation, contextual formation, and particular spatial location focusing on process and its representation. Moreover, he showed the process of “what” and “where” in the visual system tightly coupled in a synergistic manner [21,27] (we will discuss this in next sections). One of the famous biologically inspired model [26,28] has proposed two independent pathways, which model the dorsal and ventral processing streams in the mammalian visual system. Figure 2 shows a representation of active basis model (ABM) and its similarity with MLD. The form pathway in ventral stream, uses Gabor filter like function to obtain the shape and form information.
and as a good representation of simple and complex cells (Figure 3A,C). Motion pathway in the computational mechanism usually uses for optical flow to extract motion information.

4. Perception of The Motion

4.1. Perception And Actions

Action perception for the purpose of recognition is initially (in separate paths) presented by Goodale and Milner (1992), whom were proposed a separation of the perception and actions for recognition and identification on the ventral processing stream concerning the object recognition task. This separation allows the observer to move the hand for picking up of object and considers the projection perceptual information for object identification from striate and IT; furthermore, the posterior parietal region of the striate cortex has dorsal stream projection and needs visual sensorimotor transformations [29].

After that, Cédras and Shah (1995) reviewed motion development from the recognition aspect and emphasized on two main stages, which were presented for motion-based recognition by organizing it into motion models and matching unknown input with the constructed model. Several recognition, such as cyclic motion detection and recognition, hand gesture interpretation, tracking, and human motion recognition, were reported [30]. Perkins (1995) presented real-time animation, along with rhythmic and stochastic noise, for conveying only the texture of motion; this research also avoided the computational dynamic and constraint solvers [31]. He showed that each action has an internal rhythm and transition among movements, and realized a real-time animation. The detection of cyclic motion in frequency domain techniques of the magnitude information of Fourier transform and autocorrelation is represented as a curvature (1D signal) in function of time at 2D trajectories. Such detection is tested by synthetic and actual data of a walking person [32]. Such techniques brought this field in the motion capture direction.

4.2. Motion Patterns for Perception Of Action

4.2.1. Spatiotemporal Filter

Hubel and Wiesel (1968) suggested the occurrence of object recognition from simple to complex cells, which continued by Riesenhuber and Poggio (1999) [33,34]. But before that, Gallese et al. (1996) analyzed the electrical activity in the brain of macaque monkeys from 532 rostral parts of six inferior neuron areas and found out that the mirror neurons form a system aimed at matching observation, motor action execution in action recognition [35]. Quantitative modeling was conducted on biological feasibility for high-level recognition of the object (the model is based on MAX-like operation) [33] which showed more biological plausible than 3D object recognition in in man, monkey, and machine presented by Tarr & Bulthoff (1998) [36]. This method was expanded to a biologically inspired model approach for human arm movements and human action high-level abstraction by using the hierarchy of artificial neural networks and demonstrated the abstraction occurs in the visuo-motor control area of the brain and detected 37 degrees of freedom and biomechanical simulation with humanoids [37]. Having such approaches strengthened the idea of Gabor-like filter (or in general using spatiotemporal filters for form pathway). Figure 2 is a great representation of complex-cells implemented by ABM in form pathway [38]. The combination of spatiotemporal filter and hidden Markov model (HMM) technique was presented for MLD identification and provides decision based on the spatiotemporal sequence of the observed object features, and relatively little spatial information is caused by the segmentation of MLD image sequences, along with object identification; such information is highly temporal and is accessed by the HMM system, a major high classification rate [39] (was similar to [40]). An investigation on the spatiotemporal generalization of biological movement perception revealed the response of motion stimuli and interpolated this generalization among natural biological motion patterns [41]. A linear combination of spatiotemporal patterns is estimated using natural movement
patterns. The weight of prototypes in the morphs and the continuous and smooth variations in category probabilities are observed in this approach. A generalization exists within the motion patterns classes in the visual system [42], which also used for activity recognition [43]. However, these methods could not find more popularity than bio-inspired model-based spatiotemporal interesting points for human action recognition proposed by Cai et al. (2014) due to following the dual pathways and comparison between such points with the spatio-temporal interesting points (STIP) framework [44].

4.2.2. 3D Structural Method

The problem in most conventional methods might lie in the difficulty of creating computer-generated characters that display real-time, engaging interaction, and realistic motion, as well as the process of action perception in a common representational structure. Some methods tried to solve this issue by analyzing the action in 3D structure (such as [36]). The interpretation of human motion is divided into three following tasks detecting human body parts [45], tracking body using single/multiple cameras, and recognizing human activities through frame sequences, instead of simple translational model [46]. That involved low-level segmentation of human body parts involves joints and projection of the 3D structure of the human body in 2D representation [1,45,47]. Also this method could be more space-time dependent (like [28]) or geometry based movements approaches (such as [48]). A linear non-separable effects on 2D and 3D shape targets for visual search [49] and dynamic 3D recognition [50] were found to be decomposed under the law of perceptual organization. The understanding of action observation (in the observers) in another individual has a similar neural code used to produce the same actions. Evidence of this hypothesis includes brain image studies and examination of the functional segregation light of the perceptual mechanisms subtending visual recognition and the same mechanism used for action [51]. Same methodology with slight extension used for proto-object model based on the saliency map, which involved depth information using the 3D eye tracking datasets presented by Hu et al. (2016) [52], which was followed the evidence that supports the influence of eye movements on shape processing in the human visual system [6].

4.2.3. Motion Capture

**Optical capture with marker:** Applying MLD previously summarized and categorized a traditional optical capture technique using marker. Attaching markers to the actor’s body to have the 3D anatomic human body is used for skeleton visibility assessment (fitting), optical markers, motion capture, or mapping motion onto skeleton [53–57]. More advanced cameras helped on performing such analyses (i.e., Pan-Tilt Camera Tracking [54]). Some of the approaches here were not necessarily used for recognition of human action but applied for augmented reality or graphic applications [58,59], despite of employing deep network that can be justified as a biological-inspired model [60]. However, some of trajectory labelling applying permutation learning (using deep learning) could be more relevant [57].

**Motion capture markerless:** Current developments on the computer vision and imagery systems mitigate the necessity of using markers to assess the motion capture. Many of the summarized research works here involved markerless optical capture [61–64]. Giese and Poggio (2000) presented morphable models by linear combination of prototypical views to recognize biological movements and image synthesis for stationary 3D objects and its involvement in complex motion patterns. The linear combination of prototypical image sequences is used to recognize action patterns (even complex movements). The mentioned new approach can be used to analyze and synthesize the biological movement, which involves the actual and simulated video data and various patterns (which has local properties of the linear vector space) of locomotion [65]. Moeslund and Granum (2001) conducted a comprehensive survey on motion capture involving computer vision. This survey targeted an overview on the taxonomy of system functionalities and summarized it into four processes (initialization, tracking, estimation of agent pose, and recognition); each of these processes is divided into its subprocesses and various categories [66]. Moreover, Grèzes et al. (2001) analyzed human
perception in biological motion. Their study considers key role for action interpretation, identification, and predication. The main hypothesis of their approach lies under neural network specifications and its verifications through fMRI for 10 healthy volunteers. Seven types of visual motion displays are used: random dot cube, drifting random dots, random dot cube with masking elements, upright point-light walker display with masking elements, inverted point-light walker display with masking elements, upright point-light walker, and inverted point-light walker. In this approach, the hemodynamic responses of both rigid and non-rigid biological motions are connected (rigid motion responses are localized posteriorly to the rigid responses). The left intraparietal cortex is involved in non-rigid biological movement perception and associated with the posterior superior temporal sulcus (STS) and left anterior portion of Intraparietal sulcus-IPS responses. Regions, such as LOS/KO, MT/V5, and the posterior STS, are included in these activations [67]. An examination on the visual perception effects used point-light display for the arm movements of two actors for knocking and drinking movements. These actions were performed in 10 various effects. The point-light animation influenced by the phase-scrambled and upside-down versions of actions was shown to the actors for classification. The experimental results indicated that perception affects the corresponding action kinematics and movement of the phase related to the different limb segmentation [68]. In general, this part considered to be related to shape assessment pathway, whereas even with motion information can be tracked [69]. Kinect and improved Vicon cameras showed considerable contribution on markerless motion capture like [64,70]. Despite similarity of ABM (or similar approach for form pathway) and MLD, motion information in the biological model extracted from optical flow (which discussed in further sections).

4.3. Computational Models

Computational approaches for biological movement perception presented in the form of the computer–human interface algorithm [71], methods which involve kinetics of human body [72] or conducted visual system analysis on neural mechanisms, anatomical and functionality into two forms, and motion distinct pathways [73]. All of these approaches are considered to be biologically inspired approaches particularly last groups. Grossman and Blake (2002) conducted visual system analysis on neural mechanisms, anatomical and functionality into two forms, and motion distinct pathways. The analysis of point-light animation involves the mutual perception of form and motion within the act of biological movement. Their study referred to a previous work regarding the activation of posterior superior temporal sulcus (STSp) and presented a new finding for the activation of fusiform (FFA) and occipital within the biological movement and generation of neural signals, which can differentiate a biological motion from a non-biological one. LOC and EBA involved in human form perception were also presented. The neural in the form and motion pathways causes the biological motion perception [73]. Jastorff at el. (2002) proposed an approach to investigate the recognition process during neural mechanism and whether the brain can learn a completely new complex pattern of action. They generated a new artificial–biological motion using the linear combination of time–space prototypical trajectories recorded through motion capturing (similar to [1,28,45,47]). This method provides a significant improvement in discrimination for all the stimuli and showing that the human brain can learn entire novel action patterns [74]. Another salient foreground trajectory extraction method used saliency detection and low-rank matrix recovery for learning the discriminative features is proposed by Yi et al. (2017) [75].

4.3.1. Form Pathway

The ability to recognize moving human figure using moving point light is considered a biological motion perception and an evidence for processing form information on body shape and local motion signals to such a vivid perception. Most of the computational mechanisms considered an independent pathway for form information [1,24,28,45,47,73,76]. This section summarized approaches focused more on form information perception. Beintema and Lappe (2002) analyzed the perception of form
pattern of human action through moving light points. The biological motion stimulus follows limited
time perception of human motion without the local image motion, and the direction of the walker and
walking figure coherence [77] which also followed by Kilner et al. (2003) having hypothesized that
the overlap between observation and execution causes inconsistent performance [78]. The complex
shape orientation and local geometric attribute perceptual integration was also presented for global
representations through two-part shape adjustment and focused more on shape analysis of visual
system [79] and modified for only global form representation in [80]. A transfer perception for shape
information (object shape) is presented in [81] and argued that the shared components of objects
cause a high level of recognition in the objects, but the component transfer between the objects has
limitation. One of the impressive evidence on form perception is presented in a research on motion
blind patient (LM patient) who was suspected human homolog of V5/MT concerning the moving
stimuli which the patient did not report spatial disposition of the actor (using moving lights) and ability
for figural segregation on the movement basis cue and interpretation of the movements (moving parts)
independently [82]. An analysis of body postures in different viewpoints and human identification
using four experiments concluded that people who can identify the actions are basic-level objects
and that an abstraction occurs in the visual system [83] (like point lights [84]), whereas the low
presentation complexity and speed of pattern categorization and cognitive processing for peripheral
vision for low-level functions are also shown to be more relevant to form perception [85]. This also
has a relationship between biological motion and control unpredicted stimuli which involves shape
perception, motion neural subtractions, and motor imagery found in the lingual gyrus at the cuneus
border[86,87]. A study about single cells, neuroimaging data function, and field potential records
shows the visual mechanism in STS in primates and humans; it also simplifies biological motion
display by using point-light markers on the limbs of walkers [87–89].

4.3.2. Motion Pathway

The motion perception has been investigated in the aspect of psychology by Giese (2014) who
showed that body motion perception needs an integration of multiple visual processes involving
Gastalt-like pattern and aggregation of the bottom-up and up-down processes with recognition based
on learning [90]. Visual motion perception uses the integrated dynamic motion model to handle diverse
moving stimuli involving motion integration and detection to perceive correctly in decision [91] and
can be modelled like [92] or linked between the patterns and perception duration into three groups
according to their direction cues, namely, cardinal, diagonal, and toward diagonal [93].

The link between imagery and perception was investigated by putting the observers in the
dark, which rotates in the left or right. The velocity of chair rotation should be high, and the
direction of imagined rotation is different from physical rotation [94], this motivated suppression of
surrounding spatial information which seems losing information [95]. Matsumoto et al. (2015) analyzed
schizophrenia patients who have impairments in cognition, perception, and visual attention; they also
analyzed the biological motion perception in 17 patients and 18 healthy controls [96]. Ahveninen et al.
(2016) investigated the combination of spatial and non-spatial information in the auditory cortex (AC)
of two parallel streams, namely, “what” and “where” that are modulated for visual cortex subsystems,
as well as their integration regarding object perception. This approach uses animated video clips of two
audiovisual objects, namely, black and gray cats, and records the magneto- and electroencephalography
data. The events in sound are initially linked to object perception in posterior AC, with modulation
representations in anterior AC [97].

4.4. Summary

Action perception methods are dominant in computational models for recognition of actions and
different techniques used to computationalize this concept. Motion patterns methods (e.g., [10,17])
involved recognition of complex motions using global and local features, whereas more partially focus
on locomotion of human subject (forms or motions no discrimination). Kinetic-geometric models
Kinetic-geometric models also use modification of the methods involving “what” and “where” pathways and in general models with two separated pathways and improved form pathway (answer to “what”). Spatiotemporal filters often used to replace the MLD ([15,16,38,39,100–106]) and expanded into 3D structural methods ([1,51]), even new deep learning approaches involving 3D recognition of human action (i.e., [99]). Table 1 summarizes the pros and cons of action recognition approaches.

<table>
<thead>
<tr>
<th>Approaches in Perception</th>
<th>Topic of Each Approach</th>
<th>Connection to Other Researches</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. J. Marey and E. Muybridge (1880s)</td>
<td>moving photographs presenting locomotion</td>
<td></td>
</tr>
<tr>
<td>Rubin (1927)</td>
<td>visual perception of real movement</td>
<td></td>
</tr>
<tr>
<td>Johansson et al. (1973)</td>
<td>motion patterns for humans as biological motion (MLD)</td>
<td></td>
</tr>
<tr>
<td>Turaga et al. (2008)</td>
<td>locomotion analysis</td>
<td></td>
</tr>
<tr>
<td>Johansson (1973)</td>
<td>perception of human motion in neuroscience analysis</td>
<td>perception</td>
</tr>
<tr>
<td>Kozlowski &amp; Cutting (1977)</td>
<td>with females and upper body MLD, gender recognition</td>
<td>perception</td>
</tr>
<tr>
<td>Marr et al. (1978)</td>
<td>computational process in the human visual system, 3D shapes</td>
<td>perception-computer</td>
</tr>
<tr>
<td>Perrett et al. (1985)</td>
<td>the temporal cortex of macaque monkey analysis found two cells in the brain sensitive for rotation and view of the body movements</td>
<td>perception</td>
</tr>
<tr>
<td>Perrett et al. (1989)</td>
<td>view centered, view independent responses among the brain cells</td>
<td>perception-computer</td>
</tr>
<tr>
<td>Goddard (1989)</td>
<td>spatial and temporal feature incorporation through diffuse MLD data</td>
<td>perception</td>
</tr>
<tr>
<td>Goddard (1992)</td>
<td>synergistic manner of the process of “what” and “where” in visual system</td>
<td>perception</td>
</tr>
<tr>
<td>Goodale &amp; Milner (1992)</td>
<td>projection perceptual information from striate and inferotemporal cortex</td>
<td>perception</td>
</tr>
<tr>
<td>Cédras and Shah (1995)</td>
<td>motion-based recognition into motion models</td>
<td>modelling</td>
</tr>
<tr>
<td>Perkins (1995)</td>
<td>animated real-time, texture of motion, avoiding computational</td>
<td>modelling</td>
</tr>
<tr>
<td>Tsai et al. (1993)</td>
<td>detection of cyclic motion, applying Fourier transform</td>
<td>highly related to</td>
</tr>
<tr>
<td>Fielding &amp; Ruck (1995)</td>
<td>Hidden Markov Model (HMM) technique for classification</td>
<td>computer vision</td>
</tr>
<tr>
<td>Gallese et al. (1996)</td>
<td>analysis the electrical activity in macaque monkey’s brain</td>
<td>computer vision</td>
</tr>
<tr>
<td>Aggarwal et al. (1999)</td>
<td>human motion analysis review and computer vision approaches</td>
<td>computer vision</td>
</tr>
<tr>
<td>Rangarajan et al. (1992)</td>
<td>matching the biological motion trajectories (object recognition)</td>
<td>computer vision</td>
</tr>
<tr>
<td>Mccleod (1996)</td>
<td>motion blind patient, the homologue of V5/MT concerning the moving stimuli</td>
<td>computer vision</td>
</tr>
<tr>
<td>Wiley &amp; Hahn (1997)</td>
<td>virtual reality approach regarding the computer-generated characters</td>
<td>computer vision</td>
</tr>
<tr>
<td>Neri et al. (1998)</td>
<td>visual system ability to integrate the motion information of walkers</td>
<td>computer vision</td>
</tr>
<tr>
<td>Hill &amp; Pollick (2000)</td>
<td>temporal differences in MLD, recognition of the exaggerated motions</td>
<td></td>
</tr>
<tr>
<td>Giese &amp; Poggio (2000)</td>
<td>linear combination of prototypical views, 3D stationary object recognition</td>
<td></td>
</tr>
<tr>
<td>Moeslund &amp; Granum (2001)</td>
<td>a comprehensive survey on the motion capture</td>
<td></td>
</tr>
<tr>
<td>Pollick et al. (2001)</td>
<td>visual perception effects used point-light display (MLD)</td>
<td></td>
</tr>
<tr>
<td>Song et al. (2001)</td>
<td>Computer-human interface using joint probability density function (PDF)</td>
<td></td>
</tr>
<tr>
<td>Servos et al. (2002)</td>
<td>relationship between biological motion and control unpredicted stimuli</td>
<td></td>
</tr>
<tr>
<td>Jastorff et al. (2002)</td>
<td>investigating of recognition process in the neural mechanism</td>
<td></td>
</tr>
<tr>
<td>Giese &amp; Lapte (2002)</td>
<td>spatio-temporal generalization of the biological movement perception</td>
<td></td>
</tr>
<tr>
<td>Beintema &amp; Lapte (2002)</td>
<td>analysis of the perception of form pattern of human action by MLD</td>
<td></td>
</tr>
<tr>
<td>Fuce &amp; Perrett (2003)</td>
<td>analysis of single cells, neuroimaging data and records of field potential</td>
<td></td>
</tr>
<tr>
<td>Kilner et al. (2003)</td>
<td>analysis of the action in motor programs</td>
<td></td>
</tr>
<tr>
<td>Cohen &amp; Singh (2006)</td>
<td>perceiving the complex shape orientation and local geometric attributes</td>
<td></td>
</tr>
<tr>
<td>Lange et al. (2006)</td>
<td>moving human figure using MLD</td>
<td>computer vision</td>
</tr>
<tr>
<td>Trejo &amp; Westhoff (2006)</td>
<td>data retrieval of direction from scrambled MLD in humans and animals review in perception</td>
<td>computer vision</td>
</tr>
<tr>
<td>Blake &amp; Shilliyar (2007)</td>
<td>comparative research on human MLD</td>
<td></td>
</tr>
<tr>
<td>Pyles et al. (2007)</td>
<td>perception of object recognition analyzing the human features</td>
<td></td>
</tr>
<tr>
<td>Olivi &amp; Gilbert (2009)</td>
<td>analysis of body postures is different viewpoints and human identification</td>
<td></td>
</tr>
<tr>
<td>Daems &amp; Verfaillie (2010)</td>
<td>analyzing STSp region and its functionality underlying the BLOD response</td>
<td>modelling</td>
</tr>
<tr>
<td>Strasburger et al. (2011)</td>
<td>complex pattern recognition mechanism and motion perception</td>
<td></td>
</tr>
<tr>
<td>Giese (2014)</td>
<td>integrated dynamic motion model (IDM) to handle diverse moving stimuli</td>
<td>modelling</td>
</tr>
<tr>
<td>Tapia et al. (2015)</td>
<td>perception and modeling in the visual motion, characterized the patterns and perception duration,</td>
<td></td>
</tr>
<tr>
<td>Meso &amp; Masson (2015)</td>
<td>link between the imagery and perception</td>
<td></td>
</tr>
<tr>
<td>Nigmatulina (2015)</td>
<td>spatially suppress the surrounding by perception information</td>
<td></td>
</tr>
<tr>
<td>Tadie (2015)</td>
<td>analyzed the biological motion perception in Schizophrenia patients</td>
<td></td>
</tr>
<tr>
<td>Matsumoto et al. (2015)</td>
<td>combination of spatial and non-spatial information in auditory cortex (AC)</td>
<td></td>
</tr>
</tbody>
</table>
| iwenin et al. (2016) | | }

5. Knowledge Based Modeling Approaches

Modeling of biological movements into systematical and mathematical models follows the neuro-physiological, physiological, and neuro-scientific evidence. This modeling is increasingly

...
developed and considered one branch of this research field; many computer vision approaches also underlie this model. Sometimes engineering approaches have mixed into such modelling methods such as using HMM and features based bottom-up approaches in time sequence images \([11,107,108]\), however, some of them can be interpreted in the perspective of biologically inspired approaches (e.g., \([75,98,104,109–113]\)). We divide the methodology in these approaches into several directions based on their popularity, which followed neurophysiological evidence discussed in the previous sections. Figures 3 and 4 represent computational model proposed based on biological evidence.

**Figure 4.** An overall trend of biologically inspired model for human action recognition overview diagram. The two parallel processing streams are considered for the form and motion information.

### 5.1. Gabor Filter in Form Pathway

Furthermore, Webb et al. (2008) presented the mechanism in intermediate levels of visual processing and investigation to detect circular and radial forms. This mechanism analyzes the detection of the global structure in spiral form using the array consisting of 100 Gabor that is randomly positioned within the window. The Gabor filter randomly rotates, and the structure can be detected when the mask and test have the same spiral pitch (Figures 3–5). The Gabor filter is extensively used in the form pathway, and the approach is significant for elucidating the mechanism of visual processing streams \([114]\). The Gabor filter (assumed as zero-centered) is the product of a sinusoid and a Gaussian and can be presented by following formula:

\[
GW(x, y; \lambda, \theta, \phi, \gamma) = \exp \left( - \frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \right) \cos \left( 2\pi \frac{x'}{\lambda} + \phi \right),
\]

where \(x' = x \cos \theta + y \sin \theta, y' = -x \sin \theta + y \cos \theta\). Let wavelength is the number of cycles/pixel is \(\lambda\). Orientation is the angle of the normal to the sinusoid is \(\theta\) and phase is the offset of the sinusoid is \(\phi\). Aspect Ratio is showed by an ellipticity is produced with \(\gamma < 1\). Some studies investigate bio-inspired models of human action recognition that focuse more on the influence of spiking neural networks in the visual cortex \([15,38,101,102,115]\). Shu et al. (2014) used a 3D Gabor filter was tailored for V1 cells based on a hierarchical architecture and considered two visual cortical areas, namely, middle temporal area (MT) and primary visual cortex (V1) for motion processing \([115]\). This model is
modified to be a supervised Gabor-based object recognition approach called ABM [15] in the ventral processing stream) [38,102] (Figure 5a,c). A fuzzy-based optical flow proposed for dorsal streams was used to improve the model [100,101]. Furthermore, an approach to involve slow features was presented for ventral processing stream [16,103]. The active basis model [15] that applies Gabor wavelets (for the elements dictionary) consists of a deformable biological template. A Shared Sketch Algorithm (SSA) is followed through AdaBoost.

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

(2)

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 [15,103]. A Gabor wavelets dictionary, comprising \( n \) directions and \( m \) scales is in the form of, \( GW_i(\theta, \omega), j = 1,...,m \times n \). Where, \( \theta \in \{ \frac{k\pi}{m}, k = 0,...,n - 1 \} \) and \( \omega = \{ \sqrt{\frac{k^2}{m}}, i = 1,...,m \} \). To find the object’s shape, GW features used with small size, location, and posture variance, which concludes the overall shape structure and stays throughout the recognition process. Response (convolution) to every element offers form information with \( \theta \) and \( \omega \).

\[ B = \langle GW, I \rangle = \sum GW(x_0 - x, y_0 - y : \omega, \theta) I(x,y). \]  

(3)

\( GW \) is a \([x_0, y_0]\), \( I \) is the \([x, y]\) matrix, and the response of \( I \) to \( GW \) is \([x_0 + x, y_0 + y]\) (with zero-padding). \([I^m, m = 1,...,M] \) denotes the obtained training set, and \( B_i \) is chosen by the joint sketch algorithm. The objective is to identify \( B_i \) so that its edge segments obtained from \( I_m \) become maximal [15]. It is then necessary to compute \([I^m, \beta] = \psi \mid (I^m, \beta) \mid^2\) for different \( i \) where \( \beta \in \text{Dictionary} \) and \( \psi \) represents sigmoid, whitening and thresholding transformations (\( \beta \) maximizes \([I^m, \beta]) \). Let the template; for every training image be \( \beta = (\beta_i, i = 1,...,n) \), and \( I^m \) scores as below:

\[ M(I^m, \theta) = \sum_{i=1}^{n} \delta_i \mid I^m, \beta \mid - \log \Phi(\lambda \delta_i). \]  

(4)

\( M \) shows the match scoring function, \( \Phi \) is a nonlinear function, and \( \delta_i \) from \( \sum_{i=1}^{M} |I^m, \beta| \) performs steps selection. Weight vectors are calculated by the maximum likelihood technique (by \( \Delta = (\delta_i, i = 1,...,n) \)).

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

(5)

The maximum matching score is calculated by \( Max(x,y) \), where \( D \) represents the lattice of \( I \), and used for recognition of object recognition in form pathway [15,16,102].

5.2. Deep Learning

In addition, a fully automatic system for human action recognition is presented using convolutional neural networks (CNNs) in the uncontrolled environment. CNN is a deep learning approach, which is bio-inspired and develops a 3D CNN for the task. This approach extracts the features from spatial and temporal dimensions via a convolutional network, captures the motion information encoded from adjacent frames, and generates and combines multiple channel information. The presented approach has successfully implied a bio-inspired method through CNN and motion information combination for actual environments [99] (Another similar approaches are [116–125], long-short term memory (LSTM) [98,104,108,126]).
Figure 5. The results of simulation for a biological movement paradigm using ABM-based IncSFA modelled for form pathway (a,c) and motion pathway (b) by applying optical flow [127], are shown [16]. In (a,c) every row shows the response of ABM as well as the slowness features generated for actions. In (b) rows represent actions and a representation of motion information in false color form.

5.3. Sparse Representation

Due to the nature of sparsity, it has been extensively used for modelling ventral stream [128–131]. Lehky et al. (2011) investigated the characteristic of sparseness selection in the anterior inferotemporal cortex on a large dataset which involved the information on 674 monkey inferotemporal cells for 806 object photographs and the two-way analysis of the responses of the entire neurons in single image (population sparseness) and column-wise (response of single neurons to all images). This approach also represents inconsistent structural-based object recognition tasks, and the objects are decomposed into small standard features [128]. Complex visual understanding Lobula giant movement detectors (LGMD) and directional selective neurons, in visual pathways of locusts involved a model that tunes these two networks for collision tasks, compares them separately, and analyzes them co-operationally [129]. A sparse representation is shown in a set of overcompleted basis (dictionary) on different actions includes vector quantization (VQ) and clustering [132]. A non-negative sparse coding is used to learn the basic patterns of motion [133]. Nayak & Roy-Chowdhury (2014) presented an approach using spatiotemporal features and their unsupervised relationship to dictionary learning in the model of activity recognition. This approach provides an unsupervised sparse decomposition
framework for the relationship between the spatiotemporal features and the local information from descriptors, which create classifiers through multiple kernel learning [134,135].

5.4. Dynamic Representation Of Action

The dynamic representation of action recognition was analyzed by different approaches to track the dynamics representation of movements rather than shape of the approaches [109,112,113,136–139]. Some of these approaches involved the human silhouette such as [136], which used a pose descriptor called histogram of oriented-rectangles to represent the human action recognition in the video streams. Such approaches require more consistent connection between the frames of the stream, such as bag-of-word (BOW) representation and time dependent techniques [109,110,137,138,140]. Sometimes BOW mixed up with the basis of vector quantization (VQ) [140], maxima of the sparse interest point operators [130], or Sparse coding [131] to improve the performance of model.

Optical flow is widely used in the computational model to provide motion information (i.e., the layer-wise optical flow) [15,16,38,100–103,141,142]. $M_1$ and $M_2$ are visible masks for 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 represent 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 $E_{\mu}^{(i)}$.

$$E(u_1,v_1,u_2,v_2) = \sum_{i=1}^{2} E_{\gamma}^{(i)} + \rho E_{\delta}^{(i)} + \theta E_{\mu}^{(i)}.$$  (6)

In the next section, a summary of knowledge based modeling approaches and their approaches to combine the motion and form information is briefly summarized (Figure 5b).

The approach was considered by Marc Jeanerod as the basic method of action (semantics and pragmatics) and movement. The ordinary representational resources of pragmatics and semantic types of actions following the evidence of simulation and language understanding were investigated. Three theoretical frameworks were mentioned by Prinz (2014):

1. Semantics is based on pragmatics;
2. Pragmatics is anchored on semantics; and
3. Pragmatics is a part and parcel of semantics (taken from [143]).

This approach analyzes adaptive local space time features that are captured in the local events located in the video [144]. A computational model follows the neural plausibility assumptions for the interaction of the form and motion signals in biological motion perception from figural form cues; the receptive fields in the images of a static human body were also analyzed [145].

5.5. Interaction Between Pathways

Having two pathways involves the system with two types of data and for making decisions, the system usually combined these information and played an important role [146]. There are several methods proposed for interaction between pathways based on biological evidence (e.g., quantitative models and neurophysiologically plausible tools for model establishment [26]) (Figures 3c and 5–7). The dominant form and motion (optic flow) feature extraction in the mid-level of the moving subject is conducted by using principal component analysis (PCA) from spatial localization are considerably effective to recognize biological movement [24] (has similarity with [147]). Another skeletal representation of human body as the articulated interconnections incorporated a dual square-root function (DSRF) descriptor, where it decomposed the skeleton into five parts [111] and a relevant hierarchy that processes them and provides position-invariant feature detection [106]. An interactive approach was used that estimates motion using optical flow through a dynamic Bayesian network and involves interaction with motion information from two-filter inference in online and offline parameter optimization [142].
A bio-inspired feed-forward of spiking network model was performed for the influences of the motion system (V1 and MT) on human action recognition. Two characteristics of neural code, namely, neuron synchronization and their firing rate, were considered. Spiking networks can be a potential alternative in actual visual applications [148]. Guthier et al. (2014) studied the interaction and combination of the pathways in the visual system. The investigation focused on the recognition of complex biological articulated movements. Figure 7 represents a general overview on computational model, which potentially a representation of advancements in the pathways. They introduced a model that utilizes gradient and optical flow. The patterns are used by an unsupervised learning algorithm, translation-invariant nonnegative sparse coding called VNMF, and shaped prototypical optical flow patterns. In the learning processes, a lateral reserve term that eliminates competing pattern activations provide small sparse activations [149]. An interaction of motion and spatial information was presented using Motion Binary Pattern (MBP) and Volume Local Binary Pattern (VLBP) with Optical Flow for recognizing human actions and space-time volume binary patterns [150]. A fuzzy multiplication of form and motion information is proposed by combining active basis model with optical flow [38,100,101].

Moreover, a combination of slowly varying features and fast varying features proposed by Yousefi et al. (2016) [103]. Haghigh et al. (2016) studied human-like movements processed in the human brain and motor control. The study involved the concept of artificial intelligence and robotics, as well as learning the latent simple motions for imitation in more complex movements. It proposed MOSAIC structure in motor control modeling [151]. An interaction of form and motion pathways using representations of both paths for each selected key pose frame \( k_{\text{key}} \in K \), a form representation of spatial derivatives combination and concatenating orientation selective maps with following formula:

\[
I_{\text{form}} = \frac{\log(1 + 5|I_{\text{con}}|)}{\max(\log(1 + 5|I_{\text{con}}|))}
\]  

Replacing \( I_{\text{con}} \) with optical flow \( I_{\text{flo}} \) (using vertical and horizontal components instead of spatial resolution) and concatenating them and making similar formula for form information as follow:

\[
I_{\text{motion}} = \frac{\log(1 + 5|I_{\text{flo}}|)}{\max(\log(1 + 5|I_{\text{flo}}|))}
\]

Then, form and motion information are combined as an input to train of the deep convolutional neural network (DCNN) [119,152].

Ward et al. (2010) investigated the reference frames applied in terms of visual information by using fMRI. The analysis considers the receptive field scene processing areas, such as transverse occipital sulcus (TOS), retrosplenial complex (RSC), and parahippocampal place area (PPA). PPA and TOS show the position response curves on the fixation points to the screen (or the pattern), whereas RSC area does not [153].

Bio-inspired features in action recognition are presented by involving the motion in the models of cortical areas V1 and MT (shape and characteristics of their receptive field). A model with different surrounding geometries for MT cell receptive field is presented, which leads to bio-inspired features regarding the average activity of MT cells and how these features are used as a standard in the classification of activity recognition [154].

5.6. Summary

Computational mechanism for action recognition follows a dual pathways model, which involves form and motion information in two independent pathways. Form pathway corresponds to shape information and usually involves object recognition task having Gabor like (spatiotemporal) features motivated by simple and complex cells to grasp the object shape [15,38,100,102,155] (showed in Figure 6). Current developments in deep learning have considerable impact on improvement of this
pathway and researchers still trying to modified this pathway modifying deep learning configuration in a bid of increasing the accuracy or robustness of current approaches. For this pathway, deep learning showed a significant performance using different configurations [116–121,156]. On the other pathway, motion information extracted by optical flow, which was considered as fast varying features [103]. Interaction between dual pathways considered to be a challenge between recognition and space-time dependent patterns, which usually followed the neuro-physiological, physiological evidence rather than heuristic techniques. Table 2 summarizes knowledge based approaches.

Table 2. The Knowledge based modeling approaches presented with their contribution in the field.

<table>
<thead>
<tr>
<th>Psychological and Neuroscience Approaches</th>
<th>Topic of the Approach</th>
<th>Connection to Other Researches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jellema et al. (2000)</td>
<td>analysis of the cellular population located in the temporal lobe of the macaque monkey</td>
<td>Psychology</td>
</tr>
<tr>
<td>Billard et al. (2008, June)</td>
<td>action imitation considered the actions high-level abstractions</td>
<td></td>
</tr>
<tr>
<td>Vaina et al. (2003)</td>
<td>investigation regarding the neural network, MRI in MLD</td>
<td></td>
</tr>
<tr>
<td>Goodale &amp; Westwood (2004)</td>
<td>evaluating the labour division at visual pathways</td>
<td></td>
</tr>
<tr>
<td>Banquert et al. (2005)</td>
<td>associative learning for object location level, in CA3-CA1 region</td>
<td></td>
</tr>
<tr>
<td>Milner &amp; Goodale (2006)</td>
<td>involvement of dorsal stream in movement to target following ventral stream</td>
<td></td>
</tr>
<tr>
<td>Cook et al. (2009)</td>
<td>ASCs for comparing detection of non-biological and biological motion</td>
<td></td>
</tr>
<tr>
<td>Wyk et al. (2012)</td>
<td>action representation at STS</td>
<td>Psychology</td>
</tr>
<tr>
<td>Schenk (2012a)</td>
<td>DF patient analyzes the ability to get the object</td>
<td></td>
</tr>
<tr>
<td>Schenk (2012b)</td>
<td>Using MRI the functionality of DF patient</td>
<td></td>
</tr>
<tr>
<td>Fleischer et al. (2013)</td>
<td>visual recognition from motion</td>
<td></td>
</tr>
<tr>
<td>Hesse &amp; Schenk (2014)</td>
<td>visuomotor performance of a DF patient tested for letter-posting task</td>
<td></td>
</tr>
<tr>
<td>Theusner et al. (2014)</td>
<td>motion energy based on the luminance of objects motion detectors</td>
<td></td>
</tr>
<tr>
<td>Whitwell et al. (2014)</td>
<td>test by different width of the objects and DF, distinguish the shape perceptually</td>
<td>perception</td>
</tr>
<tr>
<td>Whitwell et al. (2015)</td>
<td>ability grip scaling is may rely on online visual or haptic feedback (for DF patient)</td>
<td></td>
</tr>
<tr>
<td>Krügel et al. (2013)</td>
<td>review of the behavior using EEG</td>
<td>Psychology</td>
</tr>
<tr>
<td>Cavina-Pratesi et al. (2015)</td>
<td>brain circulation using MRI regarding the word recognition ability</td>
<td></td>
</tr>
<tr>
<td>Ganos et al. (2015)</td>
<td>voluntary movement considering GTS area</td>
<td></td>
</tr>
<tr>
<td>Yamamoto &amp; Miura (2016)</td>
<td>visual object motions on time perception</td>
<td></td>
</tr>
<tr>
<td>Schindler &amp; Bartels (2016)</td>
<td>on 3 dimensional visual cue involving the motion parallax analyzing</td>
<td></td>
</tr>
<tr>
<td>Venezia et al. (2016)</td>
<td>the sensorimotor integration of visual speech through the perception</td>
<td></td>
</tr>
<tr>
<td>Harvey &amp; Dumoulin (2016)</td>
<td>visual motion effects on neural receptive field and MRI response</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. (a) The scheme representation of a stream of visual processing located in human cerebral cortex and the ventral stream information of early visual areas (V1p) and its projection in the occipito-temporal cortex and dorsal stream (blue) projection of the information to the posterior parietal cortex are shown. The indicated routes represented by the arrows and involvement of complex interconnections [155]; (b) also follows the scheme representation of human (or any objects) in ventral stream from simple cells to complex composites; (c) a presentation of receptive field from early vision till recognition in both pathways.
6. Psychological and Neuroscience Point Of View

6.1. Biological Evidence Using Fmri

The cellular population located in the temporal lobe of macaque monkeys’ inner superior temporal sulcus (partially called STPa or STSa) was analyzed. The responses of these cells were associated with the agent’s action performance as it reached the targeted position [157]. Some other types of actions (oscillatory movements, repetitive form of the arms and legs, and exact movements of reaching and grasping) investigated in [158], which was expanded by fMRI and point light research involving rigid and non-rigid motion responses of biological motion with gender concerns [90,159–161] (Reference [90] presented body posture for spatiotemporal receptive fields, Reference [160] studied hand actions, and Reference [161] presented an engineering approach but inspired by biological evidence). An investigation using fMRI for 37 children showed both groups of gender bilaterally showed the activation at the posterior STS; in response to incongruence, children showed response changes in the STS regions, also an incongruency effect was observed in the older children and adolescents in the experiment [162] (also similar to [163]). Goodale and Westwood (2004) presented another approach that evaluates the labor division at visual pathways and completed their hypothesis on the primate cerebral cortex between ventral streams dedicated to visual perception and dorsal stream for visual control in action. The study analyzed the psychological evidence and the response to visual motor control; in particular, the neurobiological challenge in mapping these behavioral findings onto the brain was analyzed and compared with known information about ventral and dorsal streams (in primate neurophysiology and human neurophysiology) [155]. An associative learning type analysis was presented for object location level, spatiotemporal level in the CA3–CA1 region, and movement-related information in the entorhinal cortex. This letter also analyzed the behavioral implementation and multi-modal integration, which suggested the functional interpretation in hippocampo-cortical systems [164]. Yamamoto and Miura (2016) analyzed visual object motions on time perception. They investigated the line segments in front or behind the occluders in different speeds and followed the association of time perception with global motion processing [165]. A study on 3D visual cue involving motion parallax analyzed the link between the visual motion and scene processing by using fMRI. Parallax-selective responses were found in parietal regions IPS4 and IPS3, and in the region of occipital place area. Some regions such as the RSC and PPA do not respond to the...
parallax [166]. Venezia et al. (2016) analyzed the sensorimotor integration of visual speech through perception. The study used fMRI on healthy individuals to identify new visiomotor circuit speech production [167]. A research on how visual motion affects neural receptive fields and fMRI response amplitudes was carried out to examine visual motion neural position preferences in the hierarchy of visual field maps using high-field fMRI and population receptive field. The results showed that visual motion induces the transformation of visuo-spatial representations through the visual hierarchy [168].

6.2. Biological Model And Imitation

Another approach on action imitation explains the natural action through visual analysis of actions and motor representation of the nervous system. The evidence of the existing system is mentioned by mirror system for mapping in primates and humans [169]. The human imitation of machines has been investigated for the purpose of flexibility, usefulness, and development of user-friendly machines. The approach concentrates on understanding how robots determine what to imitate, as well as the process of mapping perception onto the action it is imitating [170]. A mathematical approach that tackle parts of the imitation problem and the motor side of the imitation were investigated. The results argued that the perceptual system for movement identification and the spatial information correspond to these actions [171]. The cognitive development agent in the imitation and its architecture in the recognition of action was presented and implemented in the robots. The understanding and generation of actions, as well as the ability to learn new composite actions during the mentioned architecture, were also investigated [172] (similar to [173] another bio-inspired robotic imitation/recognition approach).

6.3. Visual System Impairment and Pathways

A study on autism spectrum conditions (ASCs) compared the detection of non-biological and biological motion in human adults through psychological evidence obtained from participants who watched biological (hand movements) and non-biological (falling tennis balls) stimuli. The ASC group did not show proper responses to perturbation from biological motions based on velocity profile [174]. The dorsal stream was suggested to be involved in movement to the target following ventral stream visual representation processing delay [175]. The visuomotor performance of a DF patient was tested through a letter-posting task. The absence of environmental cues was observed in the DF patient, causing them to be unaffected by delay (aforementioned). The findings suggest that ventral stream damage does not consistently influence delayed movements but affects the visual feedback and environmental landmarks [176]. Another investigation on DF patients analyzed their ability to acquire an object by distinguishing its geometry. Using fMRI, the functionality of a DF patient uses the intact visuomotor system housed within the posterior parietal lobe in the dorsal stream. Moreover, Schenk (2012a, 2012b) described the non-functioning of visuomotor networks in the dorsal stream, which was caused by a haptic feedback of the targeted object’s edges [177,178]. A test was conducted using different object widths, and the DF patients could grasp them within the healthy range (unlike the hypothesis that they should not); moreover, haptic feedback did not improve the ability of the DF patient to distinguish the shape perceptually [179]. Another research mentioned that ability grip scaling may rely on online visual or haptic feedback. The grip scaling of the DF patient did not activate while her vision was suppressed in a grasp movement, showing that the haptic feedback after perception impairs the DF patient’s performance. The research showed that DF patient’s spared grasping task relies on dorsal stream functioning at the normal mode [180]. Krigolson et al. (2015) presented a review of the behavior on three areas, namely, feedback processing, feed-forward control, and target perturbation; electroencephalography (EEG) was utilized to determine the temporal nature application in the goal-directed action. The cognitive potential and neural processing timing to motor control were further analyzed [181]. A research used fMRI to investigate brain circulation during word recognition in the left fusiform and left inferior frontal areas of the gyrus, as well as the left middle temporal cortex of the DF patient. The left fusiform activations called visual word form area appeared from the FFA and hypothesized that this area lies outside the LO [182]. A research on voluntary
movement in the Gilles de la Tourette syndrome comprised 25 patients. The results suggested that the brain learns voluntary control by perceptually discriminating signals from noise [183].

6.4. Summary

Neurophysiological evidence shows the connection between different parts of the brain during visual recognition of motion. Despite there being many studies in this area, this field is still considered to be an action research direction in a bid to develop computational neural mechanism [184–186]. The impairments in visual system reported in the evidence proved the concept of dual pathways (i.e., form and motion). Moreover, evidences gathered by fMRI and EEG continues to modify current developments in the computational model. Table 3 and Figure 7 summarize the psychological and neuroscience approaches along with targeted areas of the brain with their computational tool in the model.

Table 3. The psychological and neuroscience approaches presented with their contribution in the field.

<table>
<thead>
<tr>
<th>Knowledge Based Modelling Approaches</th>
<th>Topic of the Approach</th>
<th>Connection to Other Researches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yamato (1992, June)</td>
<td>HMM and feature based bottom up approaches in time sequence images</td>
<td>modelling</td>
</tr>
<tr>
<td>Giiese &amp; Poggio (1999)</td>
<td>Linear combination of motion sequence prototypical views, 3D object recognition</td>
<td>Computer vision</td>
</tr>
<tr>
<td>Gavriella (1999)</td>
<td>survey article in visual analysis regarding the human movement</td>
<td>perception</td>
</tr>
<tr>
<td>Wachtler &amp; Nagel (1997, June)</td>
<td>the quantitative description of the geometry of the human object</td>
<td>perception</td>
</tr>
<tr>
<td>Gises &amp; Poggio (2003)</td>
<td>dual processing pathways in the visual system</td>
<td>perception, neuroscience</td>
</tr>
<tr>
<td>SchlaF et al. (2004)</td>
<td>adaptive local space-time features</td>
<td>perception, neuroscience</td>
</tr>
<tr>
<td>Castle &amp; Giese (2005)</td>
<td>multilevel generalization using simple mid-Level optic flow features</td>
<td>perception</td>
</tr>
<tr>
<td>Arbib (2005)</td>
<td>analysis of neural and its functionality grounding for the Language skills</td>
<td></td>
</tr>
<tr>
<td>Valato &amp; Panic (2006)</td>
<td>comparison of Logical and biological inspired methods for facial expression</td>
<td></td>
</tr>
<tr>
<td>Demirisi &amp; Kadhouri (2006)</td>
<td>computational architecture and HAMMER for motor control systems</td>
<td></td>
</tr>
<tr>
<td>Lange &amp; Lappe (2006)</td>
<td>Neural plausibility assumptions for interaction of the form and motion signals</td>
<td></td>
</tr>
<tr>
<td>Willert et al. (2007)</td>
<td>estimating the motion using optical flow by dynamic Bayesian network</td>
<td></td>
</tr>
<tr>
<td>Bjuang et al. (2007)</td>
<td>hierarchical feed forward architecture on the object recognition</td>
<td></td>
</tr>
<tr>
<td>Minler and Goodale (2007)</td>
<td>analysis of two cortical systems regarding the vision in action</td>
<td></td>
</tr>
<tr>
<td>Fathi &amp; Mori (2008)</td>
<td>mid-Level Learning for the motion features</td>
<td></td>
</tr>
<tr>
<td>Schindler &amp; Goof (2006, June)</td>
<td>recognition of simple actions instantaneously by short sequences (snippets) 1-10 frames</td>
<td></td>
</tr>
<tr>
<td>Schindler &amp; Goof (2008)</td>
<td>recognition of form (Local shape) and motion (Local flow) features</td>
<td></td>
</tr>
<tr>
<td>Webb et al. (2008)</td>
<td>intermediate Levels of visual processing, detection circular, and radial form</td>
<td></td>
</tr>
<tr>
<td>Willert &amp; Eggert (2009)</td>
<td>estimation of motion to analyze the small number of temporal consecutive frames</td>
<td></td>
</tr>
<tr>
<td>Yue et al. (2009)</td>
<td>interaction of vision and touch, PCA for patterns shape features identification</td>
<td></td>
</tr>
<tr>
<td>Escobar et al. (2009)</td>
<td>bi-inspired feed-forward of spiking network model</td>
<td>neuroscience</td>
</tr>
<tr>
<td>Ekiiler &amp; Duygulu (2009)</td>
<td>analyzing the dynamic representation of action recognition using HOR</td>
<td></td>
</tr>
<tr>
<td>Wang &amp; Mori (2009)</td>
<td>visual features-as visual word and semi-Latent topic models</td>
<td></td>
</tr>
<tr>
<td>Dean et al. (2009, December)</td>
<td>Learning sparse spatiotemporal codes from the basis vectors</td>
<td></td>
</tr>
<tr>
<td>Shabani et al. (2010)</td>
<td>multiscale salient features from motion energy</td>
<td>modelling</td>
</tr>
<tr>
<td>Puppe (2010)</td>
<td>Visual based human action recognition</td>
<td>computer vision</td>
</tr>
<tr>
<td>Sun et al. (2010)</td>
<td>median filtering of the intermediate flow fields</td>
<td></td>
</tr>
<tr>
<td>Ward et al. (2010)</td>
<td>references frames applied for visual information using fMRI</td>
<td></td>
</tr>
<tr>
<td>WeinLand et al. (2011)</td>
<td>review paper for human action/activity recognition</td>
<td></td>
</tr>
<tr>
<td>Lehky et al. (2011)</td>
<td>characteristic of selection of sparseness in anterior inferotemporal cortex</td>
<td></td>
</tr>
<tr>
<td>Willert &amp; Eggert (2011)</td>
<td>representation of visual motion processing</td>
<td></td>
</tr>
<tr>
<td>Mathe &amp; Smichescu (2012)</td>
<td>BOW in maxima of sparse interest operators</td>
<td></td>
</tr>
<tr>
<td>Escobar &amp; Kornprobol (2012)</td>
<td>analysis motion in the models of cortical areas V1 and MT</td>
<td>neuroscience-perspective</td>
</tr>
<tr>
<td>Cadieu &amp; Olshausen (2012)</td>
<td>intermediate-level visual presentation</td>
<td>neuroscience-perspective</td>
</tr>
<tr>
<td>Guha &amp; Ward (2012)</td>
<td>human action in the sparse representation in overcompleted basis (dictionary) set</td>
<td></td>
</tr>
<tr>
<td>Guthrie et al. (2012)</td>
<td>non-negative sparse coding on biological motion</td>
<td></td>
</tr>
<tr>
<td>Yousef et al. (2013)</td>
<td>introducing Active Basis Model for ventral stream</td>
<td></td>
</tr>
<tr>
<td>Ji et al. (2013)</td>
<td>fully automatic system for human action recognition by CNN</td>
<td></td>
</tr>
<tr>
<td>Fitzhix (2013)</td>
<td>motion analysis approach considers the movements in all directions</td>
<td></td>
</tr>
<tr>
<td>Yue &amp; Kin (2013)</td>
<td>detection of collisions, analysis of two types neurons: LGM and DSNs</td>
<td></td>
</tr>
<tr>
<td>Cai et al. (2014)</td>
<td>spatiotemporal feature in the bio-inspired model, BIM-STIP</td>
<td></td>
</tr>
<tr>
<td>Guthrie et al. (2014)</td>
<td>survey, modelling using nonnegative spare coding, VNMF</td>
<td></td>
</tr>
<tr>
<td>Yousef et al. (2014)</td>
<td>bio-inspired modeling human action recognition, spiking neural network</td>
<td></td>
</tr>
<tr>
<td>Nayak &amp; Roy-Chowdhury (2014)</td>
<td>spatiotemporal features, unsupervised way into a dictionary Learning</td>
<td></td>
</tr>
<tr>
<td>Yousef &amp; Loo (2014)</td>
<td>fuzzy optical flow division in Dorsal stream</td>
<td>Computer vision</td>
</tr>
<tr>
<td>Esmali et al. (2014)</td>
<td>robust recognition of face using C2 features in HMAX</td>
<td></td>
</tr>
<tr>
<td>Yousef &amp; Loo (2014)</td>
<td>Interaction between dorsal and ventral streams</td>
<td>Computer vision</td>
</tr>
<tr>
<td>Princ (2014)</td>
<td>analysis of action semantics and pragmatics</td>
<td>perception</td>
</tr>
<tr>
<td>Yousef &amp; Loo (2015)</td>
<td>basic shape extraction of action group sparse coding employed BOW</td>
<td></td>
</tr>
<tr>
<td>Moayedi (2015)</td>
<td>Hybrid Max-Product Neuro-Fuzzy Classifier and Quantum-Behaved PSO in the model</td>
<td></td>
</tr>
<tr>
<td>Tian (2015)</td>
<td>BOW method, VQ, CLLC, GSRC in the human action recognition</td>
<td></td>
</tr>
<tr>
<td>Yousef &amp; Loo (2016)</td>
<td>Slowness prototypes in the ventral stream</td>
<td></td>
</tr>
<tr>
<td>Hu et al. (2016)</td>
<td>proto-object based on the saliency map</td>
<td></td>
</tr>
<tr>
<td>Haghigh et al. (2016)</td>
<td>human-Like movements</td>
<td></td>
</tr>
</tbody>
</table>
7. Future Directions

Several open problems should be solved to allow us to develop the methodology further. Here, we discuss some future directions in biologically inspired action recognition that might be interesting to explore.

**Neurophysiological evidences:** One direction involves further investigation of neuroscience, and neurophysiological evidences for a better perception of the biology of the brain. Evidently, the presented approaches carefully follow the existing evidence in the field, and further framework requires explicit details in biological studies of visual system, which can initiate further developments in the mechanism. (i.e., [187–189])

**Development in two pathways methodology:** The current frameworks include different methods to cover the requirements for the model with two pathways. The computational load of these combinations should also be considered. This step would allow adapting or extending into more complex analyses, which provide advancements in the current mechanism, which involves structural configurations or methodological developments (i.e., [190]).

**Deep learning and the biologically inspired mechanism:** Finally, in terms of machine learning, another possibility would be to create another machine learning framework (with respect to biological evidences) and modify the system from episodic recognition or the frame recognition with overall understanding of the movements. In this area, with the recent developments in deep learning approaches, this concept is implemented and would be a good methodology in term of involving shape features (from form pathway) and motion information (from motion pathway) (some examples are deep learning applications, which can be justified with biological connection [98,99,104,108,116–126]). One particular trend is on applying the framework to learn more complexity in biological movements depending on deep learning based machine vision applications.

8. Conclusions

The presence of neuroscience and neurophysiological evidence as motivating facts for modifying models changes the research focus about human action recognition from a computer vision method to computational neural mechanism. We have presented a relatively complete survey of state-of-the-art methods for biologically-inspired action recognition. The reviewed techniques merged several emerging fields focusing on different perspectives of biological movements. We gathered several aspects of such perspective involving action perception, computational and knowledge based modeling, psychological, and neuroscience approaches in the field along with providing future research directions.

**Funding:** This research was partially funded under the High Impact Research (HIR) foundation grant: contract No.UM.C/HIR/MOHE/FCSIT/10, at University of Malaya (UM), Malaysia.

**Acknowledgments:** The authors tried to consider the state-of-the-art approaches and a large set of scientific papers where, for the sake of accessibility for the reader, have been collected and we have largely restricted ourselves to the relevant papers. There was a selection inevitably and hence some papers have been omitted as this is the nature of any review. We apologize to those authors whom their papers have not been included. Also, we would like thank anonymous reviewers for their insightful and constructive comments.

**Conflicts of Interest:** Authors declare no conflict of interest.

**Abbreviations**

The following abbreviations are used in this manuscript:

- AI: artificial intelligence
- MLD: moving light display
- PGA: Parliamentarians for Global Action
- HMM: hidden Markov model
- 2D: two dimension
- 3D: three dimension
Electronics 2019, 8, 1169

References


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).