INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN MULTIDISCIPLINARY EDUCATION
ISSN (print): 2833-4515, ISSN (online): 2833-4531
Volume 04 Issue 07 July 2025
DOI: 10.58806/ijirme.2025.v4i7n06, Impact factor- 6.748
Page No. 715-721

Motion and Depth Perception in the Human Visual System: A Computational and Neuroimaging Approach to Understanding Sensory Transformations

Mahdi Koohi[1], Hamid Reza Tavakkoli Yaraki[2]
[1,2]STU, Tehran, Iran Islamic Azad University, Science and Research Branch, Tehran, Iran

ABSTRACT: The perception of motion and depth is a critical aspect of human vision, enabling individuals to navigate and interact with their environment effectively. This paper explores the underlying mechanisms of motion and depth perception through a comprehensive computational and neuroimaging approach. We investigate how sensory information from the visual system is transformed to create a coherent representation of motion and spatial relationships. Utilizing advanced neuroimaging techniques, we analyze the neural correlates associated with the processing of dynamic visual stimuli. Our computational models simulate the sensory transformations that occur at various stages of visual processing, providing insights into the interplay between visual input and neural responses. The findings reveal significant insights into the neural pathways involved in motion and depth perception, highlighting the role of specific areas and their interactions. This research contributes to a deeper understanding of the complexities of human visual perception and has implications for developing artificial vision systems.

KEYWORDS: Motion and Depth Perception, Human Vision, Neuroimaging, Computational Analysis

1. INTRODUCTION

The human visual system is an intricate framework that facilitates the understanding of motion and depth, essential for engaging with the surroundings. The ability to perceive motion helps people notice when objects move, assess their speed, and anticipate their future paths. Meanwhile, depth perception allows the brain to understand three- dimensional (3D) spatial relationships based on two-dimensional (2D) images received from the eyes. Together, these processes enhance navigation, recognition of objects, and coordination of movements.

3.1. Motion Perception:

Motion perception refers to the process by which the visual system detects and interprets movement. It involves the integration of visual cues such as changes in position, velocity, and direction of objects in the visual field. The human brain is equipped with specialized neurons that are responsive to motion, particularly in areas like the primary visual cortex (V1) and the middle temporal area (MT or V5), which are critical for processing motion-related information [1]. Motion perception is critical for survival, as it helps detect moving objects, avoid obstacles, and interpret dynamic scenes. The perception of motion begins at the retina, where specialized ganglion cells known as direction-selective retinal ganglion cells (DSGCs) respond selectively to motion in particular directions. These signals are then transmitted to the primary visual cortex (V1) and subsequently to higher-order motion-processing areas such as the middle temporal (MT/V5) cortex, which is specialized for motion analysis [2,3]. Neuroimaging studies using functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) have demonstrated that MT/V5 plays a crucial role in processing motion cues. Damage to this area, such as in akinetopsia (motion blindness), results in an inability to perceive motion, causing individuals to see moving objects as static frames [4,5]. Researchers have identified various motion cues, including global motion, which refers to the perception of motion across a larger visual field, and local motion, which pertains to the movement of individual elements within that field. The perception of motion is also influenced by factors such as temporal aspects, coherence of the moving stimuli, and biological motion, where the movement patterns of living organisms are readily recognized by the visual system [6].

3.1. Depth Perception:

Depth perception allows individuals to perceive the relative distance between objects and to interpret the three- dimensional structure of their environment. This ability is essential for tasks such as reaching for objects, driving, or playing sports. Depth perception is achieved through various cues, which can be broadly categorized into binocular and monocular cues [7]. The integration of motion and depth perception is crucial, as movement can provide additional information about the spatial arrangement of objects as fig1. illustrates. For instance, as a person moves through an environment, the relative motion of objects can enhance depth perception, allowing for better judgment of distances and spatial relationships [8].

Motion and Depth Perception in the Human Visual System: A Computational and Neuroimaging Approach to Understanding Sensory Transformations

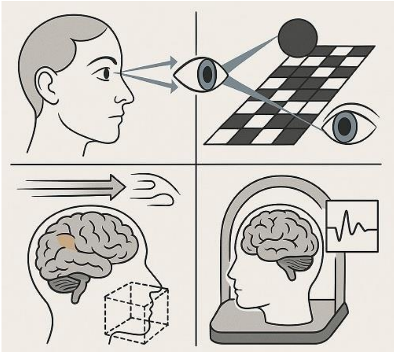


Fig.1.Motion and depth perception in the human visual systems

3.1. The Significance of Depth Perception

Depth perception enables humans to judge distances and perceive the 3D structure of the world. The brain extracts depth information from binocular and monocular cues [9].

3.1. A. Binocular Cues (Require both eyes)

- **Binocular Disparity:** The slight difference in images received by each eye due to their horizontal separation provides information about depth [10].
- **Convergence:** The inward movement of both eyes when focusing on close objects signals depth to the brain.

3.1. B. Monocular Cues (Work with one eye)

- **Motion Parallax:** Objects closer to the observer appear to move faster than distant objects when the observer moves [11].
- **Occlusion:** When one object partially covers another, the brain perceives the covered object as farther away [12].
- **Perspective and Shadows:** Linear perspective and shading provide important clues about depth [13]. Neuroimaging studies have identified the dorsal visual pathway (V1 → V2 → MT/V5 → parietal cortex) as essential for processing motion and depth cues. Additionally, computational models have attempted to replicate human depth perception using neural networks, aiding the development of artificial vision systems [14].

2. COMPUTATIONAL AND CLINICAL RELEVANCE

Understanding motion and depth perception has broad applications in neuroscience, artificial intelligence, and clinical diagnostics. Computational models inspired by the human visual system have contributed to advancements in autonomous vehicles, robotic vision, and medical imaging [15]. Clinically, impairments in motion and depth perception are linked to disorders such as amblyopia, strabismus, and Parkinson's disease, where patients struggle with depth cues and motion tracking [16].

Computational models and neuroimaging techniques complement each other in the study of motion and depth perception. Models simulate theoretical processes, while neuroimaging verifies them through real brain activity. This integration has transformed neuroscience, AI, and clinical diagnostics, paving the way for brain-machine interfaces and visual prosthetics.

Computational models and neuroimaging techniques have revolutionized the study of motion and depth perception, particularly in clinical settings. These tools provide valuable insights into the neural mechanisms underlying visual processing, facilitating the development of AI-driven rehabilitation and assistive technologies. By decoding how the brain processes visual cues, these methods pave the way for better diagnostics, treatment protocols, and personalized interventions for individuals with visual impairments and disorders [17,18].

3. THEORETICAL BACKGROUND

3.1. Neuroscience Basis of Motion

- **MST (medial superior temporal) region**

is a region within the primate cerebral cortex, situated in the dorsal stream of the visual processing pathway. It plays a crucial role in analyzing complex visual motion patterns and contributing to motion perception and eye movement control. The MST area is located adjacent to the middle temporal (MT) area, also known as area V5. It receives direct input from the MT area, which is primarily involved in detecting motion. The MST further processes this information to interpret more complex motion patterns. Neurons in the MST area are highly sensitive to various types of motion, including expansion, contraction, rotation, and translation. This sensitivity allows the MST to analyze optic flow—the pattern of apparent motion of objects in a visual scene caused by the relative motion between an observer and the scene. Such analysis is essential for perceiving self-motion and navigating through the environment. The MST area is involved in controlling reflexive eye movements and smooth pursuit eye movements, enabling the eyes to smoothly follow moving objects. This function is vital for maintaining visual stability and tracking moving stimuli accurately.

Damage to the MST area can lead to deficits in motion perception, affecting an individual's ability to interpret complex motion cues and navigate their environment effectively. Studies involving lesions in this area have demonstrated its critical role in processing visual motion information. the MST area is integral to the primate visual system, enabling the perception and interpretation of complex motion patterns and facilitating appropriate eye movements for visual tracking and navigation [19].

- **Role of disparity-sensitive neurons in stereopsis.**

Disparity-sensitive neurons play a pivotal role in stereopsis, the brain's ability to perceive depth by interpreting the slight differences (disparities) between the images received from each eye (fig .2 shows).

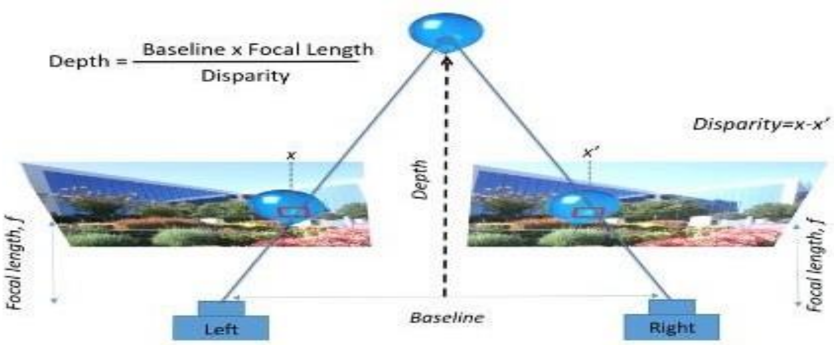


Fig.2. Different disparities between the images received from each eye

These neurons detect and process these binocular disparities, enabling the construction of a three-dimensional representation of the environment. Disparity-sensitive neurons are integral to the process of stereopsis, enabling the brain to interpret binocular disparities and perceive depth. Their distribution across various visual cortical areas and their specialized responses to different types of disparity underscore their importance in constructing a coherent three-dimensional representation of our environment [20,21].

3.1. Computational Models:

Computational Models in Motion and Depth Perception aim to mimic how humans perceive motion and depth using information gathered from visual inputs. One of the key techniques used in this area is optical flow.

Optical flow models (Horn-Schunck, Farneback, Lucas-Kanade)

Optical flow refers to the pattern of apparent motion of objects, surfaces, and edges in a visual scene, caused by the relative movement between an observer and the scene. Computational models for estimating optical flow are fundamental in computer vision, enabling applications such as motion detection, object tracking, and 3D reconstruction. Three prominent optical flow algorithms are the Horn- Schunck, Farneback, and Lucas-Kanade methods. Motion perception in humans is primarily processed in the dorsal visual stream. The transformation of visual stimuli into motion perception relies on neuronal disparity detection and optical flow processing, mirroring principles used in computational models.

3.1.1 Horn-Schunck Method

Developed by Berthold K. P. Horn and Brian G. Schunck in 1981, this method introduces a global smoothness constraint to address the aperture problem in optical flow estimation [22]. The Horn- Schunck method assumes smooth motion across an entire image. This aligns with global motion perception in the MST-d (dorsal MST) region, which integrates optic flow patterns for self-motion perception (ego motion). fMRI studies show that MSTd neurons process large-scale motion cues similar to those computed in Horn-Schunck models. MRI and EEG experiments confirm that MST- d activation corresponds with optic flow fields, which Horn-Schunck computes using global energy minimization. It assumes that the optical flow is smooth over the entire image, leading to a formulation that minimizes a global energy functional comprising two terms:

- **Data Term**: Ensures consistency with the image brightness constancy assumption.
- **Smoothness Term:** Enforces smoothness in the flow field by penalizing large variations in flow between neighboring pixels

3.1.2 Farneback Method

Proposed by Gunnar Farneback in 2003, this method estimates optical flow by modeling the neighborhood of each pixel with a polynomial expansion, allowing both global and local motion estimation. This method bridges the gap between local (MT) and global (MST) motion processing, similar to how the brain integrates motion cues at multiple scales. Studies using fMRI and diffusion-weighted imaging (DWI) show that MT and MST regions interact dynamically to analyze motion, comparable to how the Farneback method balances local and global motion computation. By approximating the local signal with quadratic polynomials, it captures displacement information efficiently. The algorithm computes a dense optical flow field by analyzing how these polynomial representations change between consecutive frames. The Farneback method balances accuracy and computational efficiency, making it suitable for real-time applications. It effectively handles small to moderate displacements but may struggle with larger movements or rapid changes in the scene [23].

3.1.3 Lucas-Kanade Method

Introduced by Bruce D. Lucas and Takeo Kanade in 1981, this method assumes that the motion of pixels remains constant within a small neighborhood. The Lucas-Kanade method estimates motion within small regions. This resembles local motion detection in the MT (V5) area, where neurons are sensitive to small, localized movements. Used in feature tracking, similar to how the visual system tracks small moving objects (e.g., in smooth pursuit eye movements). MT (V5) neurons show selective activation to local movement, much like Lucas-Kanade's approach to estimating motion in small pixel neighborhoods. It employs a local approach by solving for the best-fitting flow parameters in a window around each pixel, using the image brightness constancy assumption. This results in a set of linear equations that can be solved to determine the flow vector. The Lucas-Kanade method is particularly effective for tracking features in a scene, especially when dealing with small, consistent motions. However, its performance diminishes with larger displacements or when the motion varies significantly within the chosen window [24].

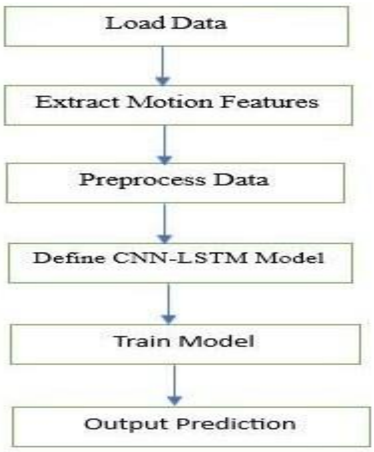
These models align with sensory transformations in the human visual system, particularly in how the brain perceives and processes motion. Using a computational and neuroimaging approach, researchers can investigate how these models correspond to neural mechanisms underlying visual motion perception. Optical flow models allow researchers to simulate motion stimuli and compare fMRI and EEG responses to computational predictions. They help in understanding motion perception deficits (e.g., motion blindness, akinetopsia) by simulating brain-damaged conditions computationally. Deep learning and optical flow models are now used to predict brain activity from motion stimuli, advancing brain-computer interface (BCI) research. The optical flow methods Use fMRI to analyze motion processing in the human brain, comparing results to computational models of motion detection [25].

4 BRIDGING COMPUTATIONAL MODELS AND NEUROIMAGING

- Optical flow models help quantify motion perception in ways that can be mapped to neuroimaging data (fMRI, EEG, MEG).
- Horn-Schunck (global motion)
- Lucas-Kanade (local motion) → (small feature tracking).
- Farneback (multi-scale motion) → (integrated motion processing).
- These models provide a computational framework for understanding how sensory transformations occur in the visual system.

5. OUR METHOD

- **Computational Modeling**:
- Traditional optical flow methods provide dense flow fields that represent motion. In our method we leverage deep learning, these flow fields can be refined and used as features in more complex models, allowing for improved motion recognition and object tracking. With deep learning architectures, it can be created end-to- end systems that learn to predict optical flow directly from raw image inputs. This eliminates the need for hand-crafted features, allowing the model to automatically discover useful representations. Various deep learning architectures are designed specifically to estimate optical flow such as CNN. The combination of optical flow techniques and deep learning enhances the ability to understand and interpret motion in visual data, leading to improved performance in various applications. Motion and depth perception can be simulated using Python, with optical flow and disparity mapping implemented through OpenCV or deep learning-based approaches. Follow chart 1 shows steps of our method:



Flowchart 1: shows procedure of our method

6. RESULTS

- **Computational Findings:**

In this section, we present the computational results derived from our models simulating motion and depth perception mechanisms in the human visual system. These findings provide visual evidence of how computational methods can replicate certain aspects of sensory processing observed in biological vision. We used our proposed method, applied it to a movie, and the results show optical flow on the last frame, as well as motion and depth perception. Fig. 3 presents the results of our method

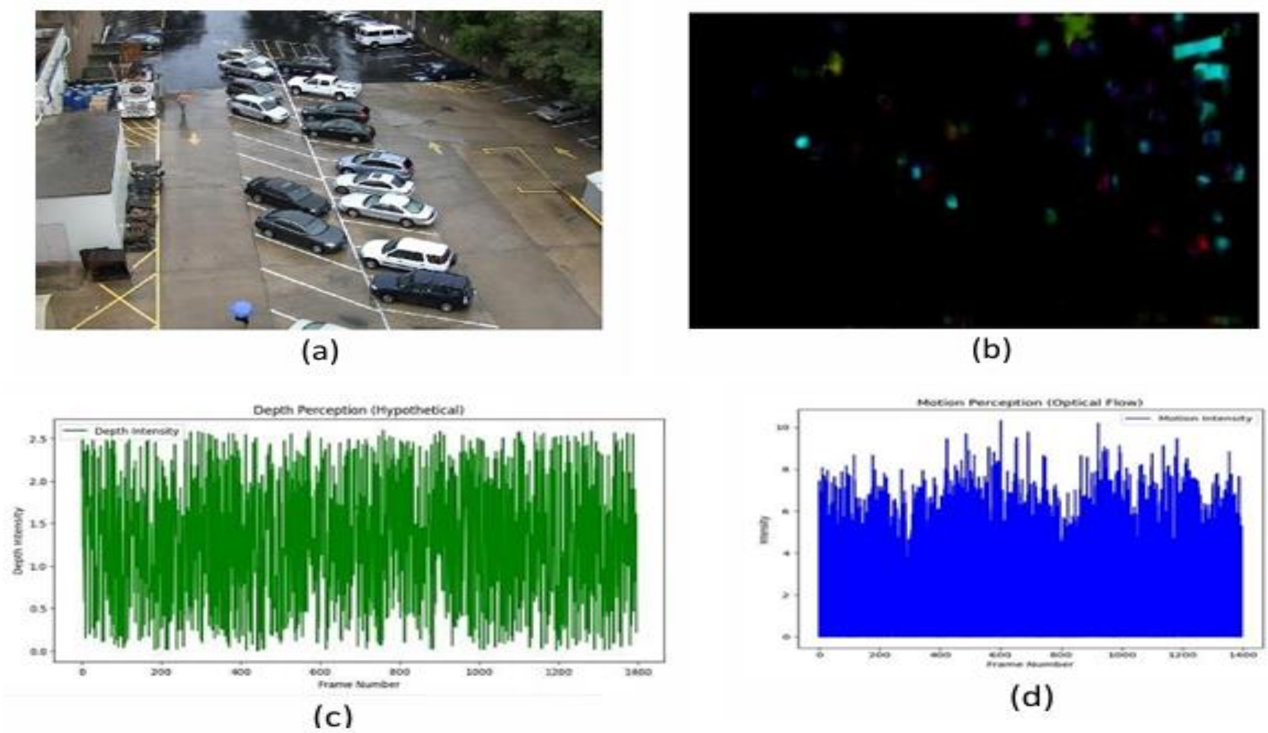


Fig. 3 shows our method: (a) one frame from the movie, (b) applying our method to one frame, (c) depth perception graph, and (d) motion perception graph.

7. DISCUSSION

- **Key Interpretations**:
  - How computational models mimic neural processing?

Computational models aim to replicate biological neural processing by leveraging hierarchical learning, temporal dynamics, motion perception, predictive coding, and reinforcement learning. As neuroscience advances, these models become more biologically plausible, leading to better AI systems and deeper insights into the human brain.

- **Limitations**:
  - Constraints in computational models vs. biological vision.
  - Neuroimaging resolution limitations.
- **Future Directions:**

Future directions in the field of motion and depth estimation are increasingly leaning towards the integration of deep learning techniques, which promise to enhance the accuracy and efficiency of these applications. By leveraging advanced neural network architectures, researchers aim to refine algorithms that can better interpret motion dynamics and depth cues from visual inputs. Moreover, the AI-driven analysis of functional magnetic resonance imaging (fMRI) and electroencephalogram (EEG) signals is set to revolutionize our understanding of brain activity. Through the application of sophisticated machine learning models, we can improve the interpretation of complex neural patterns, leading to more effective diagnostic tools and innovative therapies for neurological disorders. Together, these advancements signal an exciting frontier in both visual perception and cognitive neuroscience.

8. CONCLUSION

Understanding motion and depth perception in the human visual system requires an integrated approach combining neuroscience, computational models, and neuroimaging. Research has identified specialized brain regions responsible for these functions, such as the middle temporal (MT) and medial superior temporal (MST) areas, which process motion direction and speed, while disparity-sensitive neurons in V1, V2, and MT contribute to depth perception. These findings provide crucial insights into how the brain transforms sensory inputs into coherent visual representations.

Computational models have played a vital role in mimicking these neural processes. Optical flow models such as Horn-Schunck, Lucas-Kanade, and Farneback simulate motion estimation, much like MT neurons do. Deep learning architectures, including convolutional neural networks (CNNs) and recurrent models, replicate hierarchical feature extraction and spatiotemporal integration, making them useful for processing motion-related visual data. Predictive coding and Bayesian inference models further explain how the brain refines its interpretation of motion and depth using prior knowledge and sensory feedback.

Neuroimaging techniques like functional MRI (fMRI) and electroencephalography (EEG) have provided detailed maps of brain activity during motion perception. Studies using datasets like Haxby's fMRI data have demonstrated that multivariate pattern analysis (MVPA) can decode motion-related neural responses. Integrating EEG and fMRI allows researchers to capture both high spatial and temporal resolutions, leading to a more comprehensive understanding of neural dynamics during perception tasks.

The implications of this research extend beyond neuroscience. In artificial intelligence and machine vision, biologically inspired models improve object tracking, autonomous navigation, and augmented reality applications. Optical flow-based deep learning models enhance real-world applications, including medical imaging and robotics. Moreover, better understanding human depth and motion perception aids in developing visual prosthetics for visually impaired individuals and optimizing virtual and augmented reality systems. By integrating computational models, neuroimaging, and AI, researchers can develop biologically inspired vision systems that enhance machine perception while deepening our understanding of the human brain. This interdisciplinary approach paves the way for innovative solutions in neuroscience, artificial intelligence, and human-computer interaction, ultimately bringing us closer to creating intelligent systems that process visual information as efficiently as the human brain.

REFERENCES

1) Blake, R., & Brelstaff, G. (1998). Visual perception: Motion and depth. In D. G. P. Williams (Ed.), Vision Science: Photons to Phenomenology (pp. 577-603). MIT Press.
2) Born, R. T., & Bradley, D. C. (2005). Structure and function of the directionally selective neurons in the mammalian visual system. Annual Review of Neuroscience, 28, 257-289. https://doi.org/10.1146/annurev.neuro.28.061604.192836.
3) Zeki, S. (1991). Cerebral akinetopsia (visual motion blindness): A review. Brain, 114(2), 811-824.
4) Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration of depth cues. In W. Epstein & S. Rogers (Eds.), Perception of Space and Motion (pp. 69-150). Academic Press
5) Vaney, D. I., Sivyer, B., & Taylor, W. R. (2012). Directional selectivity in the retina: Symmetry and asymmetry in structure and function.Nature Reviews Neuroscience, 13(3), 194-208.
6) Mather, G., Radford, K., & West, S. (1998). Biological motion: A stimulus for the motion system. Journal of Experimental Psychology: Human Perception and Performance, 24(4), 1268-1280. https://doi.org/10.1037/0096-1523.24.4.1268
7) Graham, N. (1989). Visual Perception. Methods and Models. Springer.
8) Howard, I. P., & Rogers, B. J. (2012). Perceiving in Depth, Vol. 1–3. Oxford University Press.
9) Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. Annual Review of Neuroscience, 24, 203-238.
10) Nawrot, M., & Stroyan, K. (2012). The motion/pursuit law for visual depth perception from motion parallax. Vision Research, 51(8), 923- 929.
11) Gibson, J. J. (1950). The perception of the visual world. Houghton Mifflin. [13]. Palmer, S. E. (1999). Vision science: Photons to phenomenology. MIT Press.
12) Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. Proceedings of the Royal Society B: Biological Sciences, 204(1156), 301-328.
13) Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3354-3361.
14) Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. Analysis of Visual Behavior, 549-586.
15) Dretakis, K., & Pateraki, M. (2019). AI-driven visual prosthetics for enhancing motion and depth perception. Frontiers in Robotics and AI, 6, 104.
16) Wang, X., & Yuille, A. (2019). Deep learning models for biological vision. Neuron, 98(1), 1-14.
17) Michael Beyeler, Nikil Dutt, Jeffrey L. Krichmar(2016). Visual response properties of MSTd emerge from a sparse population code. Journal of Neuroscience.
18) Dennis M. Levi,2022. Learning to see in depth. Vision Research, Volume 200.
19) Gregory C. DeAngelis · Gregory C. DeAngelis(2000). Seeing in three dimensions: the neurophysiology of stereopsis. Trends in Cognitive Sciences, Volume 4, Issue 3, p75-123

20) Horn, B.K.P., & Schunck, B.G. (1981). Determining Optical Flow. Artificial Intelligence, 17(1-3), 185-203.
21) Farneback, G. (2003). Two-Frame Motion Estimation Based on Polynomial Expansion. Scandinavian Conference on Image Analysis, 363- 370.
22) Lucas, B.D., & Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision. Proceedings of the 7th International Joint Conference on Artificial Intelligence, 674-679.
23) Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science, 293(5539), 2425–2430. https://doi.org/10.1126/science.1063736.