



Review

# Depth Perception Based on the Interaction of Binocular Disparity and Motion Parallax Cues in Three-Dimensional Space

Shuai Li <sup>1,2</sup> , Shufang He <sup>1,\*</sup>, Yuanrui Dong <sup>1</sup>, Caihong Dai <sup>1</sup>, Jinyuan Liu <sup>1</sup>, Yanfei Wang <sup>1</sup>  
and Hiroaki Shigemasu <sup>3</sup> 

<sup>1</sup> Division of Optical Metrology, National Institute of Metrology, Beijing 100029, China; lees991119@163.com (S.L.); dongyr@nim.ac.cn (Y.D.); daicaihong@nim.ac.cn (C.D.); liujinyuan@nim.ac.cn (J.L.); wangyf@nim.ac.cn (Y.W.)

<sup>2</sup> Academy of Artificial Intelligence, Beijing Institute of Petrochemical Technology, Beijing 102627, China

<sup>3</sup> School of Information, Kochi University of Technology, Kami City 782-8502, Kochi, Japan; shigemasu.hiroaki@kochi-tech.ac.jp

\* Correspondence: hesf@nim.ac.cn

**Abstract:** Background and Objectives: Depth perception of the human visual system in three-dimensional (3D) space plays an important role in human–computer interaction and artificial intelligence (AI) areas. It mainly employs binocular disparity and motion parallax cues. This study aims to systemically summarize the related studies about depth perception specified by these two cues. Materials and Methods: We conducted a literature investigation on related studies and summarized them from aspects like motivations, research trends, mechanisms, and interaction models of depth perception specified by these two cues. Results: Development trends show that depth perception research has gradually evolved from early studies based on a single cue to quantitative studies based on the interaction between these two cues. Mechanisms of these two cues reveal that depth perception specified by the binocular disparity cue is mainly influenced by factors like spatial variation in disparity, viewing distance, the position of visual field (or retinal image) used, and interaction with other cues; whereas that specified by the motion parallax cue is affected by head movement and retinal image motion, interaction with other cues, and the observer’s age. By integrating these two cues, several types of models for depth perception are summarized: the weak fusion (WF) model, the modified weak fusion (MWF) model, the strong fusion (SF) model, and the intrinsic constraint (IC) model. The merits and limitations of each model are analyzed and compared. Conclusions: Based on this review, a clear picture of the study on depth perception specified by binocular disparity and motion parallax cues can be seen. Open research challenges and future directions are presented. In the future, it is necessary to explore methods for easier manipulating of depth cue signals in stereoscopic images and adopting deep learning-related methods to construct models and predict depths, to meet the increasing demand of human–computer interaction in complex 3D scenarios.

**Keywords:** human–computer interaction; virtual reality; human vision; depth perception; binocular disparity; motion parallax; fusion models; 3D space



Received: 18 February 2025

Revised: 28 April 2025

Accepted: 3 May 2025

Published: 17 May 2025

**Citation:** Li, S.; He, S.; Dong, Y.; Dai, C.; Liu, J.; Wang, Y.; Shigemasu, H. Depth Perception Based on the Interaction of Binocular Disparity and Motion Parallax Cues in

Three-Dimensional Space. *Sensors* **2025**, *25*, 3171. <https://doi.org/10.3390/s25103171>

**Copyright:** © 2025 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

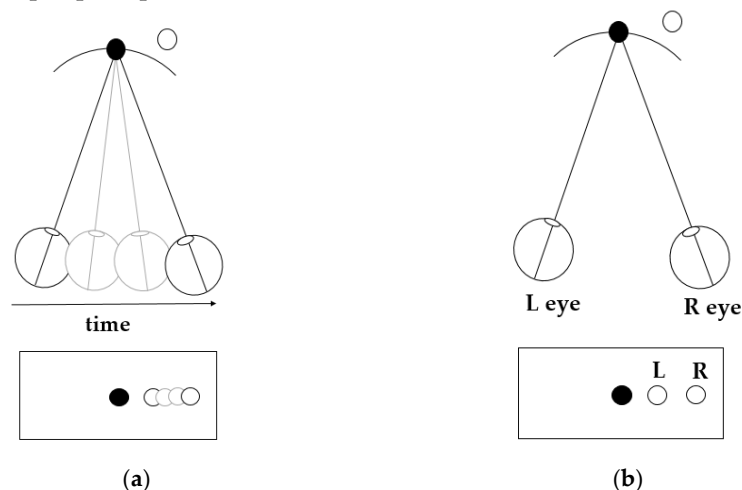
(<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the rapid development of artificial intelligence (AI) technology and the increasing demand for human–computer interaction, stereoscopic display technology represented by virtual reality (VR) and augmented reality (AR) devices is booming all over the world,

and has been widely used in medical treatment, education, industry, entertainment, and other fields. However, due to the complexity of AI and human–computer interaction demands, how to enhance the realism is one of the difficult problems for current technology development of stereoscopic displays.

In natural scenes, the human visual system mainly uses cues such as binocular disparity, motion parallax, shading, and texture to perceive depth in 3D space [1]. Among these, the first two cues are the most commonly used [2–5]. Binocular disparity is caused by the fact that the two eyes are separated in a certain distance in the horizontal direction. When viewing an object in 3D space, the retinal images in the left and right eyes have small differences, and these small differences will be further processed by the brain to form a 3D depth perception [2]. While motion parallax is caused by the ability that the brain can estimate the depth (or distance) of an object in 3D space based on the movement speed of its retinal image [6], it is a monocular cue, and can be generated by the movements of the objects or the observer themselves. In certain conditions, binocular disparity and motion parallax can produce equivalent depth perceptions [7]. For example, as shown in the top area of Figure 1a, when a monocular observer fixates on the black circle and moves an eye-distance to the right by self-motion, the retinal image shift of the observed object (the white circle in the top area of Figure 1a) generated by this motion parallax cue can be expressed as the distance between the leftmost and rightmost white circles in the bottom box of Figure 1a. While a binocular observer fixates on the black circle (in the top area of Figure 1b) and views the object (the white circle in the top area of Figure 1b) with two eyes simultaneously in a static condition, the difference between the left and right retinal images generated by this binocular disparity cue can be expressed as the distance between the two white circles in the bottom box of Figure 1b. The retinal image shift in Figure 1a is the same as the difference between the left and right retinal images in Figure 1b, thus causing equal depth perception [3].



**Figure 1.** (a) Schematic of depth perception generated by motion parallax; (b) schematic of depth perception generated by binocular disparity [3].

At present, most stereoscopic display devices are designed based on the principle of binocular disparity. By presenting image signals to the left and right eyes with a certain disparity, depth can be perceived after the observer's binocular fusion [8]. However, because of the accommodation–vergence (A–V) conflict and extensive binocular disparity caused by this kind of device, there is always visual fatigue which prevents the widespread use of stereoscopic display [9–11]. For example, Guo et al. have designed the Go/NoGo paradigm based on different disparity settings and clarified the neural mechanism related to depth perception and stereoscopic visual fatigue in VR [11]. To reduce visual fatigue,

some researchers tried to perform nonlinear disparity mapping to compress the binocular disparity in a certain range for stereoscopic display images [12]. However, there are always binocular disparity and motion parallax cues in these images. The nonlinear mapping might induce distortion, overestimation, or underestimation of the perceived depth. Moreover, the constraints of VR devices might also limit the realism of depth reproduction in many scenarios, like the low angular resolution of 3D display which might induce a small range of depth reproduction [13]. Since motion parallax is a depth cue that can be reproduced even in a 2D screen without any limits, some researchers have tried to manipulate both binocular disparity and motion parallax cues to improve the overall realism of depth reproduction in recent years [13].

Thus, it can be seen that to improve the visual comfort and realistic experience of stereoscopic display devices, the premise and key point is the study of depth perception based on the interaction of binocular disparity and motion parallax cues in 3D space. As a result, this review mainly analyzes the mechanisms and interactions of binocular disparity and motion parallax cues on depth perception. Firstly, we introduce the research trends of depth perception and the mechanisms of these two cues; then, we present several depth perception models based on the interaction of these two cues, and compare their respective merits and limitations; finally, we analyze the open challenges and look into the future directions about depth perception study.

## 2. Research Trends of Depth Perception

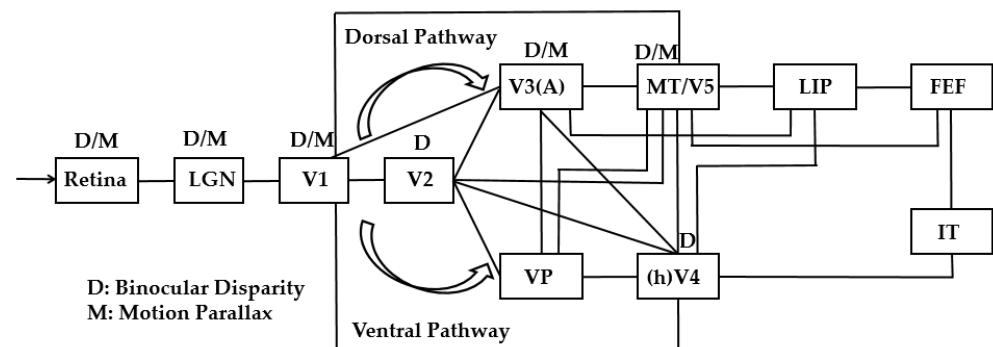
Regarding the studies on depth perception, in early days, research was mainly based on a single cue such as binocular disparity or motion parallax; later, as research deepened, scientists conducted qualitative studies on the interaction between two cues on depth perception; recently, some quantitative models were constructed for simple dynamic scenarios.

(1) Studies on a single cue of binocular disparity or motion parallax. Early studies on depth perception mainly focused on a single cue. Since Julesz proposed using random-dot stereograms to create binocular disparity images, scientists had conducted depth perception study in 3D space using psychophysical and neurobiological methods [14]. For example, Tyler studied 3D depth perception by using binocular disparity [15], and Rogers and Graham conducted related research using motion parallax cues [16]. Fang's team, through psychophysical and functional magnetic resonance imaging (fMRI) techniques, identified specific visual areas in the brain that respond to binocular disparity [17]. Other researchers have also explored depth perception study from other cues like color [18,19], but they mainly focused on a single cue, without studying the interaction between these two cues.

(2) Qualitative studies on the interaction of two cues on depth perception. Previous studies reported that there was interaction between these two cues on depth perception [20]. For example, through a series of psychophysical experiments, Bradshaw et al. found that the absolute values of depth perception thresholds obtained based on binocular disparity and motion parallax cues are different, but their depth perception thresholds are very similar when relative to the spatial frequency distributions of the sinusoidal stimuli, indicating that there is a close connection between the two cues on depth perception [21]. After adapting to the same or a different cue of binocular disparity or motion parallax cues, Bradshaw and Rogers found there was a within- and between-cue threshold elevation for 3D structure detection defined by either cue; for a compound stimulus containing both cues, the depth detection threshold was lower than the thresholds defined by either cue separately (namely, there was a sub-threshold summation) [22]. These experiments suggest that there might be a nonlinear interaction between binocular disparity and motion parallax cues [22].

Furthermore, researchers asked macaques or human observers to observe motion parallax- or binocular disparity-specified stimuli, and simultaneously detected their brain

activities by using a neurophysiological method or functional nuclear magnetic resonance (fMRI) approach. They found the following: (a) the primary visual cortex (V1, V2, V3, etc.), ventrolateral area (hV4, etc.), and dorsal visual areas (V3A, MT, etc.) show responses to binocular disparity signals [23]. (b) More than half of the neurons in the MT visual area respond strongly to both binocular disparity and motion parallax signals, as evidence of interaction between these two cues [22,24–28]. The schematic of visual processing in the visual cortex for binocular disparity and motion parallax cues can be summarized as shown in Figure 2 [22,24–29]. However, these studies mainly focused on qualitative analysis and did not explore the quantitative relationship between these two cues on depth perception.



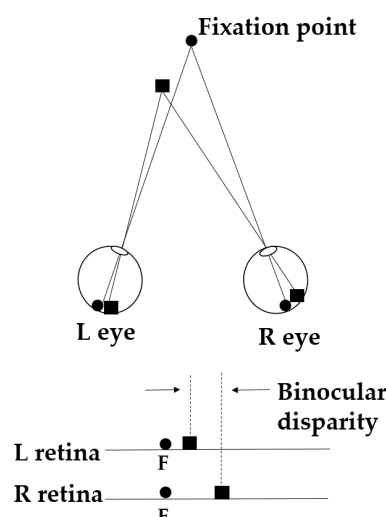
**Figure 2.** Schematic of visual processing in the visual cortex for binocular disparity and motion parallax cues.

(3) Quantitative studies with these two cues on depth perception. Qian explained the mechanism of binocular disparity-specified depth perception from the aspect of the neuronal receptive field, and constructed a binocular disparity calculation model based on complex cell response characteristics [30]. Nawrot and Stroyan proposed the “motion/pursuit ratio” law, linking binocular disparity and motion parallax cues to the ratio of depth to viewing distance [31]. In recent years, some researchers have constructed depth perception models, including the WF model, the MWF model, the SF model, and the IC model, to quantitatively analyze the contributions of binocular disparity and motion parallax cues on depth perception [22,32–34]. With these models, they revealed that the relationship between these two cues for depth perception might be linear, non-linear, and even more complicated based on different conditions. We will explain these models with more details in Section 5.

### 3. The Mechanism of Binocular Disparity Cue

The principle of binocular disparity can be illustrated as the schematic in Figure 3. The upper part of Figure 3 demonstrates how the fixation point (black circle) and the observed object (black square) are projected onto the retinas of the left and right eyes, respectively. Because the distance and viewing angle of the fixation point relative to both eyes are the same, its image is in the fovea of each retina (i.e., point F in the lower part of Figure 3). In contrast, the observed object is located at different distances and viewing angles relative to each eye, resulting in a positional offset between its retinal images in the left and right eyes (the distance between the two dashed lines in the lower part of Figure 3). This positional difference is known as binocular disparity [30].

Binocular disparity used as a cue for depth perception started as early as the 19th century. In 1838, Wheatstone invented the stereoscope and demonstrated the retinal differences between the two eyes could cause stereoscopic vision [35]. Nearly a century later, Julesz used random-dot stereogram stimuli, and found that the brain could only use binocular disparity information to perceive depth, even when other depth cues were absent [14].



**Figure 3.** Schematic of the binocular disparity principle [30].

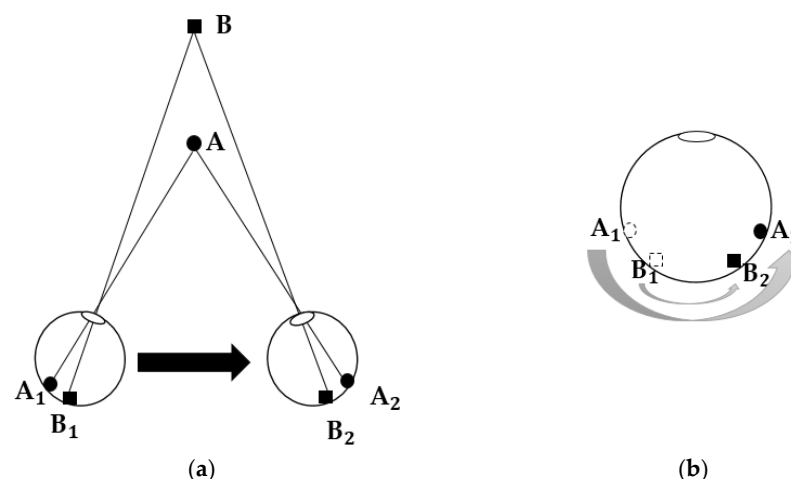
Recent research has revealed that the magnitude of binocular disparity can be influenced by spatial variation in disparity. For example, Hibbard explored how the spatial variation in disparity and its second-order luminance statistics had an impact on disparity tuning of the energy model. By modeling natural images using a binocular energy model, the author analyzed the neural responses of the model neurons in various disparity conditions. Results showed that model neurons tuned to small disparities responded most strongly, which was more obvious for vertical than for horizontal disparity, and also related to the eccentricity [36]. In addition, viewing distance is another factor that has impact on binocular disparity-specified depth perception. Studies have shown that when observers view a scene at a close distance, binocular disparity serves as an important depth cue for providing precise depth estimates [37–39].

Depth perception specified by binocular disparity is also related to the position of visual field (or retinal image) used. Hibbard and Bouzit found that when the stimuli were shown below fixation or the fixation distance was increased, the perceived depth tended to be closer than fixation [40]. This result was also supported by other physiological studies; neurons responding to the lower visual field tend to be more sensitive to crossed disparity, whereas those to the upper visual field are more sensitive to uncrossed disparity [41].

In addition, the interaction between binocular disparity and other cues can further enhance the accuracy of depth perception. For instance, the combination of binocular disparity with texture direction, convexity, and/or color information can improve the accuracy of disparity estimation and depth perception [42–47]. There is a relationship between binocular disparity and luminance, meaning that objects that are lighter seem to be closer [48]. From the physiological aspect, this can be explained by the fact that neurons tuned to brighter stimuli are usually more sensitive to nearer distances, whereas neurons tuned to darker stimuli are more sensitive to farther distances [49].

#### 4. The Mechanism of Motion Parallax Cue

The principle of motion parallax can be illustrated by the schematic in Figure 4. When an observer moves his/her head from side to side and uses one eye (while closing the other) to view objects with different depths in 3D space, the object closer to the observer (object A in Figure 4a) seems to move more on the retina (from  $A_1$  to  $A_2$  on the retina) than that (object B, from  $B_1$  to  $B_2$ ) farther away from him/her (as Figure 4b). The human visual system uses the movement speeds of objects on the monocular retina to judge their depths; this is called motion parallax [50].



**Figure 4.** Schematic of the motion parallax principle.

Motion parallax used as a cue for depth perception also started in the 19th century. With the invention of the stereoscope, Wheatstone also pointed out that head movement could provide equivalent depth perception without the involvement of binocular disparity [35]. In 1925, Helmholtz defined this cue as motion parallax and pointed out that it could produce the same depth perception as binocular disparity [51]. Although some researchers questioned the effectiveness of motion parallax [52–54], Rogers and Graham used random-dot stereograms to construct an experimental paradigm of observer- or object-motion, and demonstrated that motion parallax could serve as an independent and effective depth cue [16].

Previous studies reported that head movement and retinal image motion have influence on motion parallax-specified depth perception [55]. Designing different kinds of motion parallax cues, Malla et al. proposed that even slight head movement (e.g., a few millimeters), it will have an influence on depth perception caused by these cues [55]. Fulvio et al. also confirmed that head jitter had an impact on motion-in-depth perception, which was not found in many experiments due to head fixation [56]. In addition, based on the motion parallax cue, Dokka et al. revealed that both observer's velocity and retinal speed had influence on depth perception [57].

In addition, studies have shown that the combination of motion parallax with other cues may also have an influence on depth perception. Buckthought et al. designed experiments to compare orthographic and perspective rendering, by using textures composed of random-dot and Gabor micro-pattern elements. In these experiments, observers were asked to perform depth sorting tasks in monocular viewing conditions. The results demonstrated that dynamic perspective cues (including small vertical displacement, lateral gradients of speed, and the speed differences between near and far surfaces) can enhance depth perception from motion parallax [58].

Moreover, researchers have also reported the impact of age on motion parallax-specified depth perception. Research by Norman et al. suggested that although older observers may perform less well in 3D shape perception than young people, they can still effectively perceive the magnitudes of depth specified by the motion parallax cue [59]. However, Holmin and Nawrot found that for older adults, their depth thresholds might increase, and the pursuit accuracy might decrease when perceiving depth based on the motion parallax cue. They suggested that these age-related results might be due to the changes of the pursuit signals for the older adults [60].



## 5. Models of Depth Perception Based on Interaction Between Binocular Disparity and Motion Parallax Cues

In practical applications, binocular disparity and motion parallax cues are always coexisting for depth perception. Researchers have revealed that there are interactions between these two cues and their respective contributions to depth perception may vary based on different conditions [38,39,61,62]. This section will introduce several integration models (the WF model, the MWF model, the SF model, and the IC model) with their concepts, advantages/disadvantages, and comparisons.

### 5.1. The WF Model

The WF model, also known as the Weak Observer, uses a method of weighted averaging to combine multiple depth cues [32,33]. As illustrated in Figure 5, the main idea of this model is as follows: (1) visual system computes depth maps independently based on each individual depth cue (marked as Cue<sub>A</sub> and Cue<sub>B</sub> in Figure 5); (2) then averages these depth maps to obtain the overall depth for the scene based on linear integration with weights Weight<sub>A</sub> and Weight<sub>B</sub> [34]. Here, take motion parallax and binocular disparity cues as an example; the integration can be expressed as Equation (1).

$$D = \alpha \cdot D_m + \beta \cdot D_d \quad (1)$$

where  $D_m$  and  $D_d$  represent the depth estimates based on motion parallax and binocular disparity cues, respectively. The weighting coefficients  $\alpha$  and  $\beta$  mean the weights of these two cues.

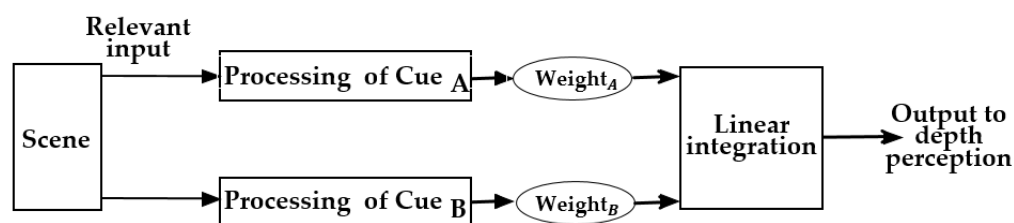


Figure 5. Principle of the WF model [63].

Previous studies used the ancillary measurement method with psychophysical observers to define the weights [64]. In the experiments, researchers used a staircase method to adjust the depth value caused by a single cue until the combined depth was perceived the same as that caused by two constant cues, which can be expressed as Equation (2) [64].

$$D = \alpha \cdot D_m + \beta \cdot D_d = \alpha \cdot D_m + \beta \cdot (D_m + \Delta_{cue}) \quad (2)$$

where  $\Delta_{cue}$  is the amount of the adjusted depth.

Since  $\alpha + \beta = 1$ , they further specialized Equation (2) into Equation (3).

$$\beta = \frac{D - D_m}{\Delta_{cue}} \quad (3)$$

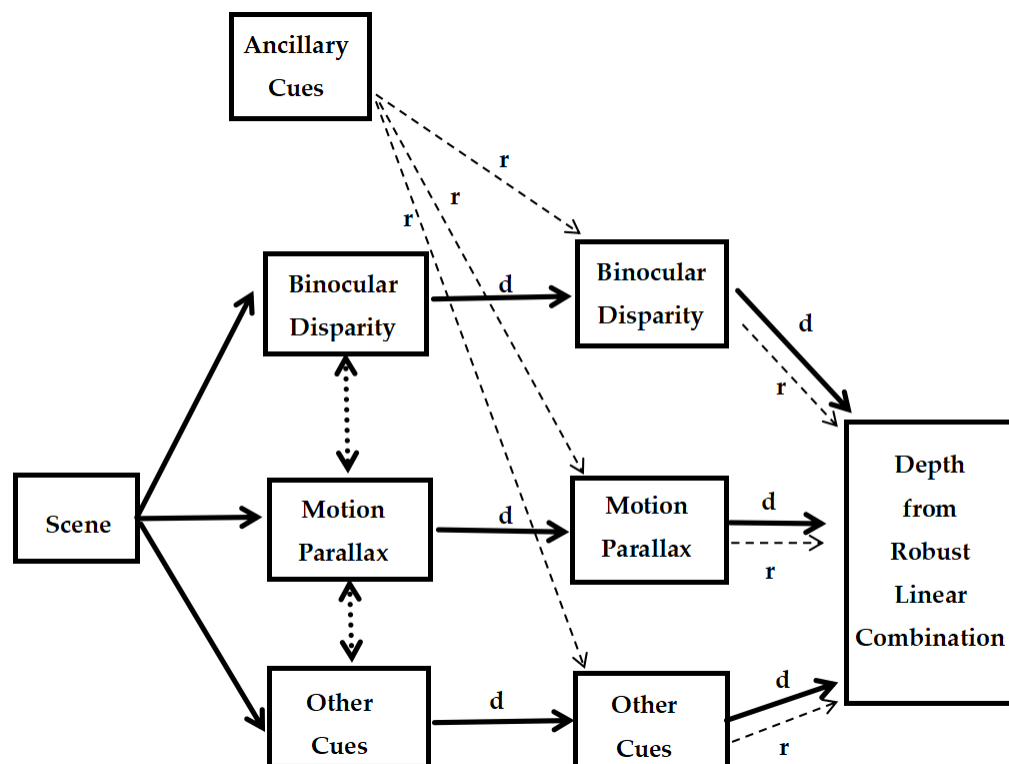
With this method, the weight  $\beta$  can be calculated as the ratio of the change in estimated depth to perturbed  $\Delta_{cue}$ , which can be obtained via psychophysical experimental data.

The advantages of this model are as follows: it is modular, and the rule of combination (weighted averaging) is simple. Based on this model, there is interaction among different cues only if they share common retinal input. However, since each depth cue might provide different depth information (in a different physical unit), it does not really make sense to make an average of this depth information, which might be a major problem of this model [64].

### 5.2. The MWF Model

Based on the WF model, researchers found that the interaction of binocular disparity and motion parallax is not just a linear combination, which cannot be simply defined as weak fusion [34]. Hence, they proposed an MWF model, which takes consideration of the interaction between cues, and involves two more stages as “cue promotion” and “dynamic weighting”, when comparing with the WF model [34].

As shown in Figure 6 (d means depth map, and r means reliability map), the main procedure of the MWF model is as follows:



**Figure 6.** Principle of the MWF model [34].

(1) Cues interact for the purpose of cue promotion, and a depth map is produced for each cue.

Since depth maps estimated from different cues might be in different physical units (e.g., in meters or ratios with dimension 1), there should be a process to change them into common units, which is shown as the interactions between cues (two-way arrows in the left part of Figure 6), namely “cue promotion” [34].

For example, depth from motion parallax might be an absolute measure ( $\text{depth}_p$ ) given by Equation (4) [34]:

$$\text{depth}_p = f_p(\text{velocity}) \quad (4)$$

where  $\text{depth}_p$  is the distance from the observer to the object, which is a function of velocity as  $f_p(\text{velocity})$ .

Depth from binocular disparity ( $\text{depth}_s$ ) can be expressed as Equation (5) [34]:

$$\text{depth}_s = d + d^2 f_s(\text{disparity}) \quad (5)$$

where  $d$  is the unknown viewing distance,  $f_s(\text{disparity})$  is the relative depth. After scaling by viewing distance as  $d^2 f_s(\text{disparity})$ , it is an absolute depth.



To promote the binocular cue, one can use Equation (6) to minimize the inconsistency between the disparity and motion parallax-specified depths and obtain a more stable-estimated—viewing distance  $d$  [34].

$$\min_d \sum_{x,y} \left[ d + d^2 f_s(\text{disparity}; x, y) - f_p(\text{velocity}; x, y) \right]^2 \quad (6)$$

(2) Using ancillary cues, each depth cue produces a depth map and a reliability map. Note that ancillary cues are cues (e.g., vestibular input) which concern the reliability of depth cues when conveying information.

(3) Depth from robust linear combination. Since the reliable depth information from each cue might vary across the scene, thus its weight might change accordingly, this is the process of “dynamic weighting” [34]. If the variance  $\sigma_i^2$  of each cue is known, the weight  $\hat{x}$  for each estimated value  $x_i$  can be defined by its inverse variance, as shown in Equation (7) [34].

$$\hat{x} = \frac{\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \quad (7)$$

The weights can also be measured by experiments. For example, researchers constructed two groups of stimuli specified by textural and motion cues for depth perception. The two groups of stimuli were displayed side by side, in which one side was in the consistent-cues condition, the other side in the mixed-cues condition. They asked subjects to indicate which side appeared to have larger depth, and the point of subjective equality (PSE) for each consistent-cues condition can be obtained. Finally, the slope of the PSE distribution suggested the weight of the textural cue was 0.46 [34]. It was found that the lower reliability of a cue, the lower weight it receives [34]. Moreover, the MWF model also can be approximated by using Bayesian methods.

This model has advantages like the involvements of cue interaction and cue modularity, which make better robustness and reliability for depth estimation, and can be used to predict depth across variable scenes [34,65].

### 5.3. The SF Model

The SF model, also known as the Strong Observer, is an alternative to the WF model [34]. Figure 7 shows how the depth perceived from each cue is integrated in the SF model: the processing of the two depth cues Cue<sub>A</sub> and Cue<sub>B</sub> (for example, binocular disparity and motion parallax cues, respectively) is not independent, but with mutual improvement through a priori constraints and recurrent constraints (namely, a feedback loop). After that, the outputs from these processings are integrated in a nonlinear way to produce output to depth perception [63].

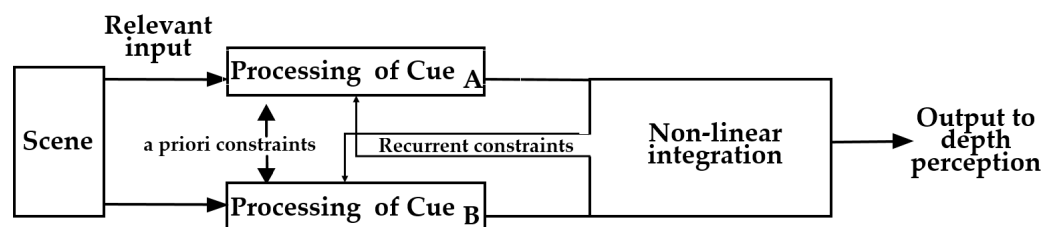


Figure 7. Principle of the SF model [63].

In a priori constraints process, fusion occurs by one cue changing a priori constraints on another cue. For example, in a stereo-vision algorithm, a smoothness constraint is altered to keep the extracted depth smooth. This process can be expressed as an energy minimization procedure as Equation (8) [66].

$$E(\vec{x}) = \int (I_l(\vec{x}) + I_r(\vec{x} + \vec{D}(\vec{x}))^2 + (e(\vec{x}) - 1) \|\nabla \vec{D}(\vec{x})\|^2 d\vec{x} \quad (8)$$

where  $\vec{D}(\vec{x})$  represents the disparity field,  $I_l$  and  $I_r$  respectively represent the intensity fields of the left and right images. The variable  $e(\vec{x})$  functions as the (binary) field of occluding edge locations ( $e(\vec{x}) = 1$  at an occluding edge, and is zero elsewhere). The occluding edge field is produced from a module that is independent of the disparity field estimation process [66].

In recurrent constraints process, a feedback loop is involved. Since this is a dynamic process, which might cause divergence or oscillation, the convergence and stability should be taken into consideration [66].

Finally, in the non-linear integration process, coupled Markov Random Field (MRF) method, Bayesian formulation, and other methods can be used for data fusion [66].

Unlike the WF model, the SF model does not estimate depth based on the information from different cues in modular, but more likely from the retinal data; thus, the procedure is not modular [34,67].

For example, Ichikawa et al. studied whether and/or how the integration of multiple cues for depth perception at near-threshold levels depended on cue types or the consistency of depth information from each cue [63]. In their experiments, they used sinusoidal stimuli with three depth cues (binocular disparity, motion parallax, and monocular configuration). The sinusoidal stimuli were manipulated in different spatial frequencies and different phases. Experimental results show, in binocular disparity and motion parallax cue condition, when cues specified the same spatial frequency and phase (in-phase condition), these two cues were integrated in the strong fusion process; while if cues specified different spatial frequencies or different phases (out-of-phase condition), the integration was a weak fusion [63]. In addition, in binocular disparity and monocular configuration cue condition, the integration was also a weak fusion process [63]. It can be seen that the visual system integrates depth cues differently depending on the cue types and their consistency. This suggests that cue integration is hierarchical and context-dependent, with early-stage interactions for disparity and parallax and later-stage integration for monocular cues. These findings highlight the complexity of depth perception [63].

#### 5.4. The IC Model

Since there is noise in measurement, the estimated depth from each separate cue might have error. Therefore, to reduce the measurement noise, Domini et al. proposed the IC model, which can obtain a composite depth by pooling all depth-related cues together [22].

As shown in Figure 8, when processing disparity and velocity signals, the model contains two stages [68].

The first stage is “dimensionality reduction” to obtain the best affine structure estimate. A Principal Component Analysis (PCA) was used to combine disparity signals  $\bar{d}_i$  and velocity signals  $\bar{v}_i$  to generate a composite score  $\rho_i$  which is highly correlated with the scaled depth  $z_i$  [22]. The main equations are as follows:

$$\bar{d}_i \approx \bar{\mu} z_i + \varepsilon \quad (9)$$

$$\bar{v}_i \approx \bar{\omega} z_i + \varepsilon \quad (10)$$

$$\rho_i = z_i \sqrt{\bar{\mu}^2 + \bar{\omega}^2} + \varepsilon \quad (11)$$

where  $\bar{\mu}$  and  $\bar{\omega}$  are the scaled convergence angle and angular velocity, respectively, and  $\varepsilon$  represents measurement noise. With this method, the best estimate of the affine structure can be obtained based on the first principal component  $PC_1$ .

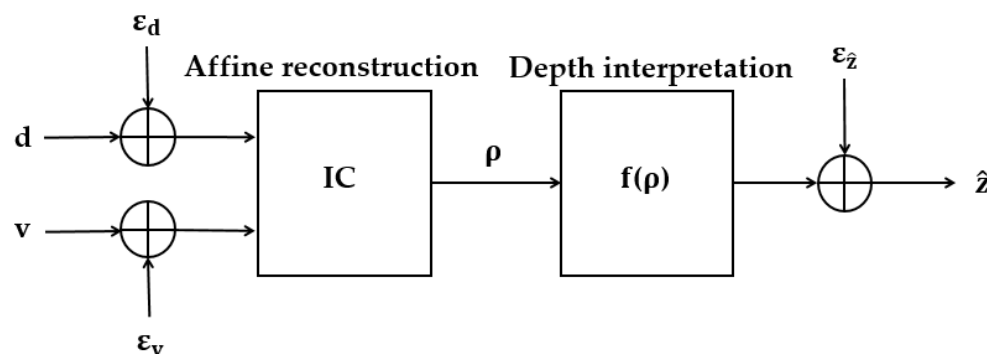


Figure 8. Principle of the IC model [68].

The second stage is “depth interpretation,” where the model recovers the true Euclidean depth  $z_i$  from the composite score  $\rho_i$  through a maximum likelihood estimation method [22]. The main equations are as follows:

$$\hat{z}_i = \operatorname{argmax}_{z_i} p(\rho_i | z_i) \quad (12)$$

$$p(\rho | z_i) \propto \int_{\mathbf{R}} p(\rho | z_i, \mathbf{R}) p(\mathbf{R}) d\mathbf{R} \quad (13)$$

Here,  $\mathbf{R}$  represents the global rotation,  $p(\rho_i | z_i, \mathbf{R})$  is the probability of the score given depth and rotation, and  $p(\mathbf{R})$  is the prior distribution of rotation. Since the specific prior distribution is usually unknown, the model assumes that the perceptual outcome primarily depends on the composite score  $\rho_i$ . With this model, the visual system can obtain the most stable interpretation to the perceived depth [22].

### 5.5. Comparisons and Applications of Above Models

#### 5.5.1. Comparisons of Above Models

To validate the WF, MWF, and SF models, Fine and Jacobs used ellipses with variable widths and depths as stimuli, and positioned them at different viewing distances [34,69]. The stimuli were set with three kinds of information: disparity, retinal motion, and vergence angle; and also in three noise conditions. The tasks were to perform depth judgement (the distance from the nearest point to the farthest point of the ellipse) or shape judgement (the ratio of the depth to the width of the ellipse). They compared the predictions of these three models with the performance of human observers [62]. Experimental results show the following: (i) In the no-noise condition, each model performed well on both tasks. While after adding noise, the shape judgement task seemed to be easier than the depth judgement task. The MWF model performed best in the depth judgement task, and was equivalent to the SF model in the shape judgement task, while the weak fusion model performs poorly. (ii) In the MWF model, the weights of motion and disparity cues were affected by factors like task, viewing distance, and noise model. In the shape judgement task, the weight of the motion cue was higher; while in the depth judgement task, the weight of the disparity cue was higher, and with the increase in viewing distance, the weight of the disparity cue increased. (iii) However, in some conditions or after adding noise, the prediction of the MWF model was inconsistent with the psychophysical data. For example, the weight of the disparity cue in the MWF model increased at a long distance, while in the psychophysical study, human observers relied more on motion cues [34,69].

Tassinari and Domini implemented an experiment to compare the IC/MWF predictions with human observers' performance [34,69]. In their experiment, they adopted a haploscope with two CRT monitors to present an elliptical hemicylinder specified by random-dot stereograms as stimuli. The elliptical hemicylinder was elongated in one of five different values, and presented in stereo-only, motion-only, and stereo-motion con-

ditions [34,69]. The observers' task was to judge whether the hemicylinder had larger or smaller depth than an apparently circular cylinder (ACC). Experimental results showed that in the MWF model, although the observer's point of subjective equality (PSE) in the stereo-only condition was closer to the veridical value than that in the motion-only condition, the PSE in the stereo-motion condition was closer to the value in the motion-only condition; the predicted depth of the MWF model in the stereo-motion condition was close to the observer's PSE in the stereo-only condition. These results suggest that the predictions of the MWF model could not totally match with observers' performance [34,69]. They discussed that this might be due to two reasons: it is still unclear how to guarantee a veridical Euclidean solution for each separate depth cue in the cue promotion stage; it is unspecified how to perform the dynamic cue re-weighting (for example, providing different weights to the same cue, which were influenced by the other cues in the scene) [34,69].

However, for the IC model, taking one observer ABB as an example (similar as most of the other observers in this experiment), the ACC was perceived for a simulated depth as in 18.21 mm in the stereo-motion condition, which was highly consistent with the observer's psychophysical performance (PSE = 18.02 mm), suggesting the effectiveness of the IC model [34,69]. Recently, Domini proposed that the IC model worked in a vector sum model, which used components of a multi-dimensional vector to represent individual cue estimates, and their norms determined the combined outputs. The IC model could explain not only the findings which were in support of the maximum-likelihood estimation (MLE) model, but also those which contradicted the MLE model [70–72].

As a result, comparisons of the WF, MWF, SF, and IC models can be summarized as shown in Table 1. In the no-noise condition, each model performs well. However, after adding noise, compared with the WF and SF models, the MWF model performs best in the depth judgement task; but in some conditions, the prediction of the MWF model is inconsistent with the psychophysical data. The IC model pools all depth cues to obtain the most stable interpretation of depth. It predicts systematic bias in performance, which is consistent with observers' psychophysical performance and also concerns the variability of their performance, but the underlying mechanisms still need to be clarified [22,68,71,72].

**Table 1.** Comparisons of the WF, MWF, SF, and IC models.

Model	Main Procedure	Characteristics	Validation Results
WF [32–34,63,64]	(1) Linear average	(1) Modular (2) Cue independent (no interaction)	(1) In the no-noise condition, each model performs well.
SF [34,63,66,67]	(1) A priori constraints (2) Recurrent constraints	(1) Non-modular (2) Unconstrained nonlinear interaction	(2) While after adding noise, the WMF model performs best in depth judgement task; but in some conditions, the prediction of the MWF model was inconsistent with the psychophysical data.
MWF [34,65]	(1) Cue promotion (2) Dynamic weighting	(1) Modular (2) Constrained nonlinear interaction	
IC [22,68,70,72]	(1) Affine reconstruction (2) Depth interpretation	(1) Non-modular for cues (2) Concern of reducing the measurement noise	The IC model could explain both findings which were in support of the MLE model, or contradicted the MLE model.

### 5.5.2. Applications of Above Models

The above models can be used in many fields, like VR/AR, 3D television (TV), and so on. For example, Temel et al. proposed a feedback-based modified weak fusion (MWF) model to reconstruct the depth maps at the receiver side of 3D TV [73]. In this model, from the source side, each monocular cue (like color, motion, texture, and so on) in the

depth-free streaming of 3D video was generated and sent; the combination of monocular cues at the receiver side was addressed as a nonlinear optimization problem with linear constraints. The initial weights were obtained by particle swarm optimization based on Peak Signal-to-Noise Ratio (PSNR), and then optimized by Active-Sets based on 3VQM (Video Quality Metric). The experimental results show that the quality of the combined depth maps is equivalent to that of the views based on ground truth depth maps, but the bandwidth savings were as high as 38.8% [73].

Some researchers reported that there was a superadditivity of depth cue combination (namely depth overestimation) for 3D shape estimation in VR. Campagnoli et al. used the IC model to predict the superadditivity, and found the prediction by this model was consistent with human observers' performance [74].

## 6. Open Research Challenges and Future Direction

### 6.1. Open Research Challenges

- (1) How to apply depth perception models to improve the realistic experience of depth perception in 3D space

As summarized in Section 5.5, the above models have different complexity and applications. However, based on our investigation, we did not find many applications of the above models. So how to apply these models to improve the realistic experience of depth perception in virtual environments is very important.

In one aspect, it is necessary to further improve the above models. For example, for the MWF model, we can further improve the noise model and consider the deviation of human observers in estimating viewing distance, so as to make it more consistent with the psychophysical data; for the IC model, we still need to clarify the underlying mechanisms.

In the other aspect, it is necessary to further improve the methods of manipulating the 3D sequence. Kellnhofer et al. conducted psycho-visual experiments to study the influence of motion parallax in stereo 3D display, and built up a joint disparity-parallax computational model [13]. To apply this model in autostereoscopic displays, at first, they extracted the optical flow and 3D transformations from the input stereo 3D sequence; then, they used their model to predict depth specified by motion parallax and binocular disparity cues; after that, they performed sampling and estimated necessary scaling for disparity gradients; finally, they integrated the scaled gradients to construct a new output stereo 3D sequence [13]. Furthermore, they evaluated whether the realistic experience of the stereoscopic content was improved after using their method, and results showed a statistically significant 60% preference of the 3D sequence with their method. However, how to extract the optical flow and 3D transformations, estimate the scaling for disparity gradients, and integrate the scaled gradients to improve the preference is still a challenge.

- (2) How to build up depth perception models for applications in complex 3D scenarios

As mentioned before, binocular disparity is a cue in a static scene and plays an important role especially in near-distance distances [61]. In contrast, motion parallax is a dynamic scene cue, important for far-distance depth perception, like object tracking and navigation [38,39]. Each cue has different contribution on depth perception at different viewing distances. Moreover, there are also other cues, like contour occlusions, texture, lightness, shading, and perspective, that might have influence on depth perception. This evidence indicates the challenge of depth perception study since it can be affected by multiple viewing conditions and cues [38,39,75].

Moreover, although Kellnhofer et al. obtained a statistically significant 60% preference of the 3D sequence with their method as described in Section 6.1 (1), they still found a problem: their model was built up based on motion parallax and binocular disparity cues,

but they applied this model for complex images, which might limit the performance of depth enhancement [13].

So how to build up depth perception models for applications in complex 3D scenarios is another challenge.

- (3) How to improve the performance of human–computer interaction via the study of depth perception in 3D space

As human–computer interaction is commonly applied in virtual environments (like VR, AR), how depth perception in 3D space has influence on human–computer interaction is also important. By asking participants to perform continuous Fitt’s pointing tasks with a computer mouse, Cheng and Lin investigated how depth perception in VR environments had an influence on hand–eye coordination, and found that factors like angle of declination (looking downward 20° or straight-ahead vision), visual depth (stereoscopic viewing or monoscopic viewing), space cue (no space, closed space, or closed space with shadow), and index of difficulty (different sizes and distances) had an influence on the performance of depth perception, thus had an impact on the hand-movement performance [76].

These results indicate that depth perception has an influence on human–computer interaction, and it is also a challenge to improve the performance of human–computer interaction since there are multiple factors that have influence on depth perception.

## 6.2. Future Directions

- (1) Exploring methods for easier manipulating depth cue signals in stereoscopic images

Although many researchers have been devoted to studying depth perception models based on interaction of different cues, how to apply these models to improve the quality of stereoscopic images is equally important. Up to now, there are some methods to apply models: some are mainly focusing on enhancing apparent depth, without expanding the overall disparity range; but they might not consider the contribution of the motion parallax cue [77,78]. Others account for both binocular disparity and motion parallax cues, and use disparity mapping method to reallocate disparities [13]. In their methods, they compressed disparity for regions which contain motion parallax information, and used the additional disparity budget in other static regions; but there are still some challenges to be solved as mentioned in Section 6.1 (1) [13].

In the future, it is necessary to explore methods (like optimizing hardware and software designs, improving algorithms) for easier manipulating of depth cue signals in stereoscopic images.

- (2) Adopting deep learning methods to construct models with multiple cues and/or factors and predict perceived depth for human visual system in complex 3D scenarios

As mentioned in Section 6.1, to solve the challenging problems of building up models in complex 3D scenarios and improving the performance of human–computer interaction, it is necessary to design experimental paradigms that are more in line with the actual complex 3D scenarios; conduct research on depth perception based on the interaction among multiple cues or factors; and construct depth perception models with high efficiency and robustness based on multiple cues’ (or factors’) interaction.

Similar to the human visual system, depth perception is also very important for many high-level computer vision applications, like object reconstruction, scene understanding, autonomous driving, and so on [79,80]. In the above fields, deep learning methods (including artificial neural network, ANN; convolution neural network, CNN) are widely used to predict 3D depth from monocular- and binocular cues of 3D images [79,80]. For example, by using the depth from focus/defocus (DfF/DfD) and stereo matching as the



monocular- and binocular information, Chen et al. emulated human perception and exploited unified learning-based hybrid techniques to combine the DfF/DfD and stereo matching information for depth perception [79]. In their experiments, at first, they constructed a comprehensive focal stack dataset, which contained stereo color images and ground truth disparity maps; then, they produced refocused images at each focal layer by using the algorithm from Virtual DSLR (digital single lens reflex); next, they proposed different network architectures for different inputs (including different combinations of the above monocular and binocular cues), integrated and optimized the separate networks to obtain high-fidelity disparity maps; and finally obtained the estimated depth with significantly improved accuracy and robustness [79]. In gesture recognition studies, Almasre and Al-Nuaim compared four support vector machine (SVM) classifiers with two depth sensors to recognize the hand gestures of Arabic Sign Language (ArSL) words [81].

As a next step, we may also use deep learning-relevant methods to construct models with multiple cues and/or factors and predict perceived depth for the human visual system in complex 3D scenarios.

## 7. Summary

Since AI and human–computer interaction are commonly applied in virtual environments (like VR, AR), how to improve the realistic experience of depth perception in virtual environments is very important. To this end, the premise and key point is the study on how the human visual system perceives depth in 3D space. Hence, this manuscript investigates the studies about depth perception based on the interaction of binocular disparity and motion parallax cues. At first, motivation and development trends of depth perception specified by these two cues are introduced. Then, the mechanisms of these two cues are presented. Binocular disparity-specified depth perception is mainly influenced by factors like spatial variation in disparity, viewing distance, position of the visual field (or retinal image) used, and interaction with other cues; while motion parallax-specified depth perception is mainly influenced by factors such as head movement and retinal image motion, interaction with other cues, and the observer’s age. Next, several models, including the WF model, the MWF model, the SF model, and the IC model for depth perception based on the interaction of these two cues are explained and compared. Based on these investigations and summaries, we provide perspectives of open research challenges regarding how to build up and/or apply depth models in 3D complex scenarios, and how to improve the performance of human–computer interaction via the study of depth perception in 3D space. Finally, the new insights for future directions as exploring methods for easier manipulating of depth cues and adopting deep learning methods to construct models and predict perceived depth are proposed, to meet the increasing demand of human–computer interaction in complex 3D scenarios.

**Author Contributions:** Conceptualization, S.H. and H.S.; methodology, S.H., C.D. and J.L.; investigation, S.L., Y.D. and Y.W.; resources, S.H., C.D., J.L. and Y.W.; writing—original draft preparation, S.L. and S.H.; writing—review and editing, S.H., H.S. and S.L.; supervision, S.H., C.D., J.L. and Y.W.; project administration, S.H.; funding acquisition, S.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (NO. 62205323), the Fundamental Research Funds of National Institute of Metrology China (NO. AKYJJ2309, NO. AKYZZ2215), and the China Postdoctoral Science Foundation (NO. 2020M670412).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Brenner, E.; Smeets, J.B. Depth perception. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience: Sensation, Perception, and Attention*; Wiley: Hoboken, NJ, USA, 2018; pp. 385–414.
- Howard, I.P.; Rogers, B.J. *Perceiving in Depth, Volume 2: Stereoscopic Vision*; OUP: Oxford, UK, 2012.
- Kim, H.R.; Angelaki, D.E.; DeAngelis, G.C. The neural basis of depth perception from motion parallax. *Philos. Trans. R. Soc. B Biol. Sci.* **2016**, *371*, 20150256. [[CrossRef](#)]
- Howard, I.P. *Perceiving in depth, Vol. 3: Other Mechanisms of Depth Perception*; Oxford University Press: Oxford, UK, 2012.
- Welchman, A.E. The human brain in depth: How we see in 3D. *Annu. Rev. Vis. Sci.* **2016**, *2*, 345–376. [[CrossRef](#)]
- Howard, I.P.; Rogers, B.J. *Binocular Vision and Stereopsis*; Oxford University Press: Oxford, UK, 1995.
- Lillakas, L.; Ono, H.; Ujike, H.; Wade, N. On the definition of motion parallax. *Vision* **2004**, *16*, 83–92.
- Wang, S.; Ming, H.; Wang, A.; Xu, L.; Zhang, T. Three-Dimensional Display Based on Human Visual Perception. *Chin. J. Lasers* **2014**, *41*, 209007. [[CrossRef](#)]
- Lambooij, M.; IJsselsteijn, W.; Fortuin, M.; Heynderickx, I. Visual discomfort and visual fatigue of stereoscopic displays: A review. *J. Imaging Sci. Technol.* **2009**, *53*, 30201. [[CrossRef](#)]
- Koulieris, G.-A.; Bui, B.; Banks, M.S.; Drettakis, G. Accommodation and comfort in head-mounted displays. *ACM Trans. Graph. (TOG)* **2017**, *36*, 1–11. [[CrossRef](#)]
- Guo, M.; Yue, K.; Hu, H.; Lu, K.; Han, Y.; Chen, S.; Liu, Y. Neural research on depth perception and stereoscopic visual fatigue in virtual reality. *Brain Sci.* **2022**, *12*, 1231. [[CrossRef](#)] [[PubMed](#)]
- Lang, M.; Hornung, A.; Wang, O.; Poulakos, S.; Smolic, A.; Gross, M. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph. (TOG)* **2010**, *29*, 1–10. [[CrossRef](#)]
- Kellnhofer, P.; Didyk, P.; Ritschel, T.; Masia, B.; Myszkowski, K.; Seidel, H.-P. Motion parallax in stereo 3D: Model and applications. *ACM Trans. Graph. (TOG)* **2016**, *35*, 1–12. [[CrossRef](#)]
- Julesz, B. Binocular Depth Perception without Familiarity Cues: Random-dot stereo images with controlled spatial and temporal properties clarify problems in stereopsis. *Science* **1964**, *145*, 356–362. [[CrossRef](#)]
- Tyler, C.W. Depth perception in disparity gratings. *Nature* **1974**, *251*, 140–142. [[CrossRef](#)] [[PubMed](#)]
- Rogers, B.; Graham, M. Motion parallax as an independent cue for depth perception. *Perception* **1979**, *8*, 125–134. [[CrossRef](#)]
- Chen, N.; Chen, Z.; Fang, F. Functional specialization in human dorsal pathway for stereoscopic depth processing. *Exp. Brain Res.* **2020**, *238*, 2581–2588. [[CrossRef](#)] [[PubMed](#)]
- Li, Z.; Cui, Y.; Zhou, T.; Jiang, Y.; Wang, Y.; Yan, Y.; Nebeling, M.; Shi, Y. Color-to-depth mappings as depth cues in virtual reality. In Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology, 29 October–2 November 2022; pp. 1–14.
- Bailey, R.; Grimm, C.; Davoli, C.; Abrams, R. *The Effect of Object Color on Depth Ordering*; Technical Report WUCSE-2007-21; Department of Computer Science & Engineering, Washington University in St. Louis: St. Louis, MO, USA, 2007.
- Bradshaw, M.F.; Rogers, B.J. The interaction of binocular disparity and motion parallax in the computation of depth. *Vis. Res.* **1996**, *36*, 3457–3468. [[CrossRef](#)] [[PubMed](#)]
- Bradshaw, M.F.; Hibbard, P.B.; Parton, A.D.; Rose, D.; Langley, K. Surface orientation, modulation frequency and the detection and perception of depth defined by binocular disparity and motion parallax. *Vis. Res.* **2006**, *46*, 2636–2644. [[CrossRef](#)]
- Domini, F.; Caudek, C.; Tassinari, H. Stereo and motion information are not independently processed by the visual system. *Vis. Res.* **2006**, *46*, 1707–1723. [[CrossRef](#)]
- Minini, L.; Parker, A.J.; Bridge, H. Neural modulation by binocular disparity greatest in human dorsal visual stream. *J. Neurophysiol.* **2010**, *104*, 169–178. [[CrossRef](#)]
- DeAngelis, G.C.; Cumming, B.G.; Newsome, W.T. Cortical area MT and the perception of stereoscopic depth. *Nature* **1998**, *394*, 677–680. [[CrossRef](#)]
- Uka, T.; DeAngelis, G.C. Linking neural representation to function in stereoscopic depth perception: Roles of the middle temporal area in coarse versus fine disparity discrimination. *J. Neurosci.* **2006**, *26*, 6791–6802. [[CrossRef](#)]
- Nadler, J.W.; Angelaki, D.E.; DeAngelis, G.C. A neural representation of depth from motion parallax in macaque visual cortex. *Nature* **2008**, *452*, 642–645. [[CrossRef](#)]
- Kim, H.R.; Angelaki, D.E.; DeAngelis, G.C. A functional link between MT neurons and depth perception based on motion parallax. *J. Neurosci.* **2015**, *35*, 2766–2777. [[CrossRef](#)] [[PubMed](#)]
- Xu, Z.-X.; DeAngelis, G.C. Neural mechanism for coding depth from motion parallax in area MT: Gain modulation or tuning shifts? *J. Neurosci.* **2022**, *42*, 1235–1253. [[CrossRef](#)] [[PubMed](#)]

29. Li, Z. *Understanding Vision: Theory, Models, and Data*; Oxford University Press: Oxford, UK, 2014.
30. Qian, N. Binocular disparity and the perception of depth. *Neuron* **1997**, *18*, 359–368. [[CrossRef](#)] [[PubMed](#)]
31. Nawrot, M.; Stroyan, K. The motion/pursuit law for visual depth perception from motion parallax. *Vis. Res.* **2009**, *49*, 1969–1978. [[CrossRef](#)]
32. Doshier, B.A.; Sperling, G.; Wurst, S.A. Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vis. Res.* **1986**, *26*, 973–990. [[CrossRef](#)]
33. Rogers, B.J.; Collett, T.S. The appearance of surfaces specified by motion parallax and binocular disparity. *Q. J. Exp. Psychol. Sect. A-Hum. Exp. Psychol.* **1989**, *41*, 697–717. [[CrossRef](#)]
34. Landy, M.S.; Maloney, L.T.; Johnston, E.B.; Young, M. Measurement and modeling of depth cue combination: In defense of weak fusion. *Vis. Res.* **1995**, *35*, 389–412. [[CrossRef](#)]
35. Wheatstone, C., XVIII. Contributions to the physiology of vision.—Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philos. Trans. R. Soc. Lond.* **1838**, *128*, 371–394.
36. Hibbard, P.B. Binocular energy responses to natural images. *Vis. Res.* **2008**, *48*, 1427–1439. [[CrossRef](#)]
37. Kytö, M.; Nuutinen, M.; Oittinen, P. Method for measuring stereo camera depth accuracy based on stereoscopic vision. In Proceedings of the Three-Dimensional Imaging, Interaction, and Measurement, San Francisco, CA, USA, 24–27 January 2011; pp. 168–176.
38. Gillam, B.; Palmisano, S.A.; Govan, D.G. Depth interval estimates from motion parallax and binocular disparity beyond interaction space. *Perception* **2011**, *40*, 39–49. [[CrossRef](#)]
39. He, S.; Shigemasa, H.; Ishikawa, Y.; Dai, C. Effects of Different Cues on Object Depth Perception in Three-Dimensional Space. *Acta Opt. Sin.* **2019**, *39*, 1033002.
40. Hibbard, P.B.; Bouzit, S. Stereoscopic correspondence for ambiguous targets is affected by elevation and fixation distance. *Spat. Vis.* **2005**, *18*, 399–411.
41. Sprague, W.W.; Cooper, E.A.; Tošić, I.; Banks, M.S. Stereopsis is adaptive for the natural environment. *Sci. Adv.* **2015**, *1*, e1400254. [[CrossRef](#)] [[PubMed](#)]
42. Jordan III, J.R.; Geisler, W.S.; Bovik, A.C. Color as a source of information in the stereo correspondence process. *Vis. Res.* **1990**, *30*, 1955–1970. [[CrossRef](#)]
43. Simmons, D.R.; Kingdom, F.A. Contrast thresholds for stereoscopic depth identification with isoluminant and isochromatic stimuli. *Vis. Res.* **1994**, *34*, 2971–2982. [[CrossRef](#)] [[PubMed](#)]
44. Burge, J.; McCann, B.C.; Geisler, W.S. Estimating 3D tilt from local image cues in natural scenes. *J. Vis.* **2016**, *16*, 2. [[CrossRef](#)]
45. Burge, J.; Fowlkes, C.C.; Banks, M.S. Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception. *J. Neurosci.* **2010**, *30*, 7269–7280. [[CrossRef](#)]
46. Qiu, F.T.; Von Der Heydt, R. Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. *Neuron* **2005**, *47*, 155–166. [[CrossRef](#)]
47. Cottareau, B.R.; McKee, S.P.; Ales, J.M.; Norcia, A.M. Disparity-tuned population responses from human visual cortex. *J. Neurosci.* **2011**, *31*, 954–965. [[CrossRef](#)]
48. Potetz, B.; Lee, T.S. Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *JOSA A* **2003**, *20*, 1292–1303. [[CrossRef](#)]
49. Samonds, J.M.; Potetz, B.R.; Lee, T.S. Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6313–6318. [[CrossRef](#)]
50. Snowden, R.J.; Snowden, R.; Thompson, P.; Troscianko, T. *Basic Vision: An Introduction to Visual Perception*; Oxford University Press: Oxford, UK, 2012; pp. 220–222.
51. Von Helmholtz, H. *Helmholtz's Treatise on Physiological Optics*; Optical Society of America: Rochester, NY, USA, 1925; Volume 3.
52. Gibson, E.J.; Gibson, J.J.; Smith, O.W.; Flock, H. Motion parallax as a determinant of perceived depth. *J. Exp. Psychol.* **1959**, *58*, 40. [[CrossRef](#)] [[PubMed](#)]
53. Epstein, W.; Park, J. Examination of Gibson's psychophysical hypothesis. *Psychol. Bull.* **1964**, *62*, 180. [[CrossRef](#)]
54. Gogel, W.C.; Tietz, J.D. Eye fixation and attention as modifiers of perceived distance. *Percept. Mot. Ski.* **1977**, *45*, 343–362. [[CrossRef](#)]
55. de la Malla, C.; Buiteman, S.; Otters, W.; Smeets, J.B.J.; Brenner, E. How various aspects of motion parallax influence distance judgments, even when we think we are standing still. *J. Vis.* **2016**, *16*, 8. [[CrossRef](#)] [[PubMed](#)]
56. Fulvio, J.M.; Miao, H.; Rokers, B. Head jitter enhances three-dimensional motion perception. *J. Vis.* **2021**, *21*, 12. [[CrossRef](#)] [[PubMed](#)]
57. Dokka, K.; MacNeilage, P.R.; DeAngelis, G.C.; Angelaki, D.E. Estimating distance during self-motion: A role for visual-vestibular interactions. *J. Vis.* **2011**, *11*, 1–16. [[CrossRef](#)]
58. Buckthought, A.; Yoonessi, A.; Baker, C.L. Dynamic perspective cues enhance depth perception from motion parallax. *J. Vis.* **2017**, *17*, 10. [[CrossRef](#)]

59. Norman, J.F.; Clayton, A.M.; Shular, C.F.; Thompson, S.R. Aging and the perception of depth and 3-D shape from motion parallax. *Psychol. Aging* **2004**, *19*, 506. [\[CrossRef\]](#)
60. Ono, H.; Wade, N.J. Depth and motion perceptions produced by motion parallax. *Teach. Psychol.* **2006**, *33*, 199–202.
61. Mansour, M.; Davidson, P.; Stepanov, O.; Piche, R. Relative Importance of Binocular Disparity and Motion Parallax for Depth Estimation: A Computer Vision Approach. *Remote Sens.* **2019**, *11*, 1990. [\[CrossRef\]](#)
62. Johnston, E.B.; Cumming, B.G.; Landy, M.S. Integration of stereopsis and motion shape cues. *Vis. Res.* **1994**, *34*, 2259–2275. [\[CrossRef\]](#)
63. Ichikawa, M.; Saida, S.; Osa, A.; Munechika, K. Integration of binocular disparity and monocular cues at near threshold level. *Vis. Res.* **2003**, *43*, 2439–2449. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Maloney, L.T.; Landy, M.S. A statistical framework for robust fusion of depth information. In Proceedings of the Visual Communications and Image Processing IV, Tokyo, Japan, 8–11 December 2024; pp. 1154–1163.
65. Fine, I.; Jacobs, R.A. Modeling the combination of motion, stereo, and vergence angle cues to visual depth. *Neural Comput.* **1999**, *11*, 1297–1330. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Clark, J.; Yuille, A. *Data Fusion for Sensory Information Processing Systems*; Springer: Boston, MA, USA, 1990; pp. 71–104.
67. Nakayama, K.; Shimojo, S. Experiencing and perceiving visual surfaces. *Science* **1992**, *257*, 1357–1363. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Domini, F.; Caudek, C. The intrinsic constraint model and Fechnerian sensory scaling. *J. Vis.* **2009**, *9*, 25. [\[CrossRef\]](#)
69. Tassinari, H.; Domini, F.; Caudek, C. The intrinsic constraint model for stereo-motion integration. *Perception* **2008**, *37*, 79–95. [\[CrossRef\]](#)
70. Domini, F. The case against probabilistic inference: A new deterministic theory of 3D visual processing. *Philos. Trans. R. Soc. B* **2023**, *378*, 20210458. [\[CrossRef\]](#)
71. MacKenzie, K.J.; Murray, R.F.; Wilcox, L.M. The intrinsic constraint approach to cue combination: An empirical and theoretical evaluation. *J. Vis.* **2008**, *8*, 5. [\[CrossRef\]](#)
72. Kemp, J.T.; Cesanek, E.; Domini, F. Perceiving depth from texture and disparity cues: Evidence for a non-probabilistic account of cue integration. *J. Vis.* **2023**, *23*, 13. [\[CrossRef\]](#)
73. Temel, D.; Lin, Q.; Zhang, G.; AlRegib, G. Modified weak fusion model for depthless streaming of 3D videos. In Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), San Jose, CA, USA, 15–19 July 2023; pp. 1–6.
74. Campagnoli, C.; Hung, B.; Domini, F. Explicit and implicit depth-cue integration: Evidence of systematic biases with real objects. *Vis. Res.* **2022**, *190*, 107961. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Peters, T.M.; Linte, C.A.; Yaniv, Z.; Williams, J. *Mixed and Augmented Reality in Medicine*; CRC Press: Boca Raton, FL, USA, 2018.
76. Cheng, L.Y.; Lin, C.J. The effects of depth perception viewing on hand–eye coordination in virtual reality environments. *J. Soc. Inf. Disp.* **2021**, *29*, 801–817. [\[CrossRef\]](#)
77. Chapiro, A.; O’Sullivan, C.; Jarosz, W.; Gross, M.H.; Smolic, A. Stereo from Shading. In Proceedings of the EGSR (EI&I), Zaragoza, Spain, 19–21 June 2013; pp. 119–125.
78. Didyk, P.; Ritschel, T.; Eisemann, E.; Myszkowski, K.; Seidel, H.-P. Apparent stereo: The cornsweet illusion can enhance perceived depth. In Proceedings of the Human Vision and Electronic Imaging XVII, Burlingame, CA, USA, 22–26 January 2012; pp. 180–191.
79. Chen, Z.; Guo, X.; Li, S.; Yang, Y.; Yu, J. Deep eyes: Joint depth inference using monocular and binocular cues. *Neurocomputing* **2021**, *453*, 812–824. [\[CrossRef\]](#)
80. Ming, Y.; Meng, X.; Fan, C.; Yu, H. Deep learning for monocular depth estimation: A review. *Neurocomputing* **2021**, *438*, 14–33. [\[CrossRef\]](#)
81. Almasre, M.A.; Al-Nuaim, H. Comparison of four SVM classifiers used with depth sensors to recognize Arabic sign language words. *Computers* **2017**, *6*, 20. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.