The user wants me to translate a Chinese academic paper title into English. Let me analyze the requirements: 1. This is a title, not a full paper with paragraphs 2. There are no LaTeX commands, placeholders, or citation commands 3. There are no …tags 4. I need to maintain academic tone and for…

Authors: The user has provided "张孝伍" which is a Chinese name. Based on the context of being a professional scientific translator, this is likely a name that needs to be translated or transliterated.

In academic translation, Chinese names are typically transliterated using the Pinyin system, with the family name first followed by the given name. "张孝伍" would be:

- 张 (Zhāng) - family name, commonly written as "Zhang" in English
- 孝伍 (Xiàowǔ) - given name, typically written as "Xiaowu" or with a space as "Xiao Wu"

In academic contexts, the standard format is: 1. Family name first, given name second 2. No comma between them (unlike Western names) 3. First letter of each name capitalized 4. Given name parts can be combined or separated

So "张孝伍" should be translated as "Zhang Xiaowu" or "Zhang Xiao Wu". The more common and standard format in academic papers is "Zhang Xiaowu" (combined given name).

However, I need to check if there are any specific instructions about names in the prompt. The prompt says: - "Translate from Simplified Chinese to English with the highest accuracy and academic tone" - "Use precise technical terminology - obey the glossary strictly" - The glossary provided only includes "机器学习 => machine learning" and "深度学习 => deep learning"

Since this is just a name and not technical terminology, and there' s no specific name in the glossary, I should use standard academic transliteration.

The most appropriate translation would be "Zhang Xiaowu".

But wait - the user might be asking me to translate a document where this name appears. However, they only provided the name itself. Let me re-read the instructions.

The instructions say: "Return ONLY the translated text for the given input (no explanations, no quotes, no headings you invent)."

So if the input is just "张孝伍", the output should be just the translated name: "Zhang Xiaowu"

I should not add any explanations or additional text. Just the transliterated name.

Let me confirm the standard practice: - 张 is Zhang (family name) - 孝 is Xiao - 伍 is Wu

In academic contexts, given names are typically written together: Xiaowu

So the final answer should be: Zhang Xiaowu

I should not include any tags, Zhang Xiaowu

Date: 2023-10-12T00:00:00+00:00

Abstract

Stereoscopic vision is based on the depth perception principles of human binocular vision and insect compound eyes, employing multiple cameras to simultaneously acquire multiple digital images of the surrounding scenery from different viewpoints. Through stereo matching techniques, corresponding points across these images are identified, and three-dimensional object images are reconstructed from the disparity information of these correspondences. However, the biological mechanisms underlying existing multi-view visual three-dimensional reconstruction and stereo matching methods remain unknown, and it is unclear how human binocular vision and insect compound eyes perceive depth information and perform rapid, accurate stereo matching. By investigating the near-field high-order spatial coherence of thermal light, this study explores the neurobiological mechanisms of multi-view stereoscopic vision and derives three-dimensional reconstruction formulas for both binocular and multi-view stereoscopic vision. The obtained results are consistent with theories in visual neuroscience.

Full Text

Multi-view Stereo Imaging and Three-dimensional Reconstruction Method Based on Near-field Spatial Coherence of Thermal Light

Zhang Xiaowu
School of Science, Qingdao University of Technology, Qingdao, Shandong

266520, China

Abstract: Stereo vision is based on the depth perception principles of human binocular vision and insect compound eyes. It obtains multiple digital images of a scene from different viewpoints simultaneously using multiple cameras, identifies corresponding points across these images through stereo matching techniques, and reconstructs three-dimensional object images from the disparity information of these correspondences. The neurobiological mechanisms underlying existing multi-view visual three-dimensional reconstruction and stereo matching methods remain unknown, and it is unclear how human binocular vision and insect compound eyes perceive depth information and perform fast, accurate stereo matching. This paper investigates the neurobiological mechanisms of multi-view stereo vision through the near-field high-order spatial coherence of thermal light, and presents three-dimensional reconstruction formulas for both binocular and multi-view stereo vision. The results obtained are consistent with theories of visual neuroscience.

Keywords: Stereo vision; Three-dimensional reconstruction; High-order coherence; Stereo imaging system; Real-time stereo matching

Classification: O431.2

1 Introduction

The fundamental principle of multi-view stereo vision is to observe the same scene from two or more different locations [1,2,3,4], acquiring multiple perceptual images from different viewpoints. Through three-dimensional reconstruction methods and stereo matching principles, the positional disparity of corresponding points between multi-view image pixels is calculated to reconstruct the three-dimensional shape of the scene. This process is analogous to the stereo perception mechanisms of human binocular vision and insect compound eyes. Three-dimensional reconstruction methods include projective reconstruction, affine reconstruction, and Euclidean geometric reconstruction. Binocular disparity includes retina-centered disparity, head-centered disparity, and relative disparity between two objects.

Projective reconstruction includes methods based on the fundamental matrix, double algebra, and trilinear constraints. O. Faugeras applied the fundamental matrix to projective reconstruction, and Hartley et al. subsequently improved and extended Faugeras' method. Scholars such as G. C. Rota refined 19th-century double algebra projective geometric reconstruction, Carlsson derived several invariant expressions in double algebra form, and A. Shashua utilized geometric invariant theory to obtain trilinear constraint relationships for corresponding point coordinates among three two-dimensional images.

Affine reconstruction includes methods based on modulus constraints and camera translation motion. Luong et al. employed the concept of modulus constraints for camera self-calibration, while Peter Sturm utilized vanishing points and modulus constraints in images to achieve affine reconstruction of objects. When a camera undergoes translational motion between two images, the affine transformation between these images enables three-dimensional reconstruction of the object.

Under Euclidean geometry, the general method for three-dimensional reconstruction involves calculating spatial lines to reassemble three-dimensional surfaces and quadratic surfaces when the intrinsic and extrinsic parameters of the left and right cameras are known, thereby reconstructing three-dimensional object images. When the coordinate systems of the left and right cameras are parallel, three-dimensional reconstruction formulas can be obtained through stereo imaging geometry. In general, a projective relationship exists between two-dimensional images and three-dimensional scenes, which is represented by the camera projection matrix. This projection matrix can be recovered from a small number of corresponding points in the left and right images. Using the projection matrices of both cameras, high-precision depth information for each image point can be obtained through least squares methods, enabling reconstruction of the object' s three-dimensional image.

Compound eye multi-view vision utilizes multi-view tensors from multi-view geometry [5] to generate multilinear relationships among measurement coordinates across multiple images. The camera matrix for each image is calculated from the tensor, and three-dimensional stereo images can be computed from the recovered cameras and the disparity of corresponding points across multiple images.

The human stereo vision system perceives three-dimensional information through binocular disparity. References [6,7,8] describe the fundamental principles and types of binocular disparity. As shown in Figure 1 [Figure 1: see original paper], retina-centered binocular disparity is···(text incomplete). Figure 1 shows retinal disparity and head-centered disparity.

Stereo matching in human binocular vision occurs naturally while viewing scene objects. However, binocular vision systems differ from human binocular vision, requiring specialized stereo matching methods to obtain corresponding points between left and right images. To date, no scholars have applied the near-field second-order coherence of quantum optical thermal light to investigate the perception mechanisms of binocular stereo vision.

References [9,10,11] provide detailed introductions to the first-order, second-order, and high-order coherence of thermal light near-field, as well as quantum imaging. The second-order coherence function of thermal light near-field is measured through direct coincidence detection by two photodetectors at spacetime points. As shown in Figure 2 [Figure 2: see original paper], the HBT interferometer [9] employs two independent photodetectors placed behind pinholes to perform joint observations of two distinct events occurring at spacetime points. The observation in Figure 2 is based on the joint current at an electrical linear multiplier:

The second-order coherence function is related to the first-order coherence functions: ⋯

Second-order coherence ⋯is the left camera coordinate system, ⋯is the right camera coordinate system, ⋯coordinate system, ⋯, ⋯, and ⋯are points in the three coordinate systems ⋯, ⋯, and ⋯respectively. Through stereo matching methods such as image feature point matching, epipolar geometry recovery, sparse and dense matching, local and global matching, corresponding points ⋯ and ⋯on the left and right images are found. To date, the biological mechanisms and theoretical models of binocular stereo vision and real-time stereo matching methods remain unknown.

In this paper, two independent joint second-order coherence measurements are transformed into joint coincidence measurements of a binocular imaging system, making the measurement results of the second-order coherence function equivalent to stereo imaging in binocular vision. The main results obtained in this paper are:

(1) Based on the triangular measurement principle, the three-dimensional reconstruction formula for binocular stereo vision imaging is obtained:

where ⋯is the magnification of the camera imaging system, ⋯and ⋯are the translation position vectors of the left and right cameras in the world coordinate system, ⋯is the retina-centered binocular disparity, and ⋯is the head-centered binocular disparity.

(2) Assuming ⋯cameras are located at the nodes of a connected graph ⋯, where ⋯and ⋯are the node set and edge set of ⋯respectively, and ⋯is the imaging point on the ⋯-th image, the three-dimensional reconstruction formula for multi-view stereo vision imaging is:

where ⋯is the translation position vector of the ⋯-th camera in the world coordinate system, and ⋯is the magnification of the ⋯-th camera.

(3) Based on the near-field second-order spatial coherence of thermal light, the three-dimensional reconstruction formula for binocular stereo imaging is obtained:

where ⋯is the first-order Bessel function, ⋯is the distance between the object and the imaging lens, ⋯is the distance between the imaging lens and the image plane, ⋯is the radius of the imaging lens, ⋯is the distribution function of the object light field, ⋯is the magnification of the imaging system, and ⋯is the binocular disparity.

(4) Assuming ⋯photodetectors are located at the nodes of a connected graph ⋯, where ⋯and ⋯are the node set and edge set of ⋯respectively, ⋯is a subgraph of ⋯with node set ⋯, and the subgraph ⋯contains no isolated nodes. If ⋯has isolated nodes, these nodes are added to ⋯to obtain a new subgraph without isolated nodes. Assuming all parameters of the ⋯ imaging systems are identical, for arbitrary nodes ⋯and ⋯, based on the

near-field ···-order spatial coherence of thermal light, the three-dimensional reconstruction formula for multi-view stereo imaging is obtained:

(5) The three-dimensional reconstruction formulas in equations (1) and (3), and equations (2) and (4) are similar respectively, indicating that stereo imaging perceived by human binocular vision and insect compound eyes results from ···-order coherence measurements of thermal light near-field performed by ocular dominance cells in the visual nervous system that perceive both binocular and compound eye inputs on corresponding points of multi-view images across multiple retinas. The interferometer measuring thermal light second-order coherence performs deep fusion perception of left and right images, and the process of measuring thermal light second-order spatial coherence automatically completes stereo matching of binocular corresponding points simultaneously with stereo imaging.

2 Three-dimensional Reconstruction Formulas for Binocular and Multi-view Vision Imaging

The triangular measurement principle is illustrated in Figure 3 [Figure 3: see original paper] [2]. The camera coordinate system is ···, the imaging plane is ···, ···is any point on the image plane, and ···is any point in space. Based on the principle of triangle similarity: ···, where ···is the imaging magnification, we have:

The geometric principle of binocular convergent optical axis stereo imaging is shown in Figure 4 [Figure 4: see original paper]. The coordinate transformations among the three coordinate systems ···are:

where ···and ···are the rotation matrices between coordinate systems. From equations (5) and (6), we obtain: ···and ···are both invertible matrices, thus we have ···When the magnifications of the left and right cameras are ···, from equation (7) we obtain:

Taking the vector length on both sides of these two equations yields: ···, thus obtaining equation (1), and we have: ···According to the definition of binocular disparity [7] and Figures 1 and 4, ···is the head-centered binocular disparity, and ···is the retina-centered binocular disparity.

Assuming ···cameras are located at the nodes of a connected graph ···, where ···and ···are the node set and edge set of ···respectively. For multi-view stereo vision, the geometric model of the ···-th camera is ···, and by combining camera models on adjacent nodes ···the three-dimensional reconstruction formula (2) is obtained. The three-dimensional reconstruction formula for multi-view stereo vision is obtained by averaging.

2 Three-dimensional Reconstruction Formulas for Binocular and Multi-view Vision Imaging

The triangular measurement principle is illustrated in Figure 3 [Figure 3: see original paper] [2]. The camera coordinate system is ···, the imaging plane is ···, ···is any point on the image plane, and ···is any point in space. Based on the principle of triangle similarity: ···, where ···is the imaging magnification, we have:

The geometric principle of binocular convergent optical axis stereo imaging is shown in Figure 4 [Figure 4: see original paper]. The coordinate transformations among the three coordinate systems ···are:

where ···and ···are the rotation matrices between coordinate systems. From equations (5) and (6), we obtain: ···and ···are both invertible matrices, thus we have ···When the magnifications of the left and right cameras are ···, from equation (7) we obtain:

Taking the vector length on both sides of these two equations yields: ···, thus obtaining equation (1), and we have: ···According to the definition of binocular disparity [7] and Figures 1 and 4, ···is the head-centered binocular disparity, and ···is the retina-centered binocular disparity.

Assuming ···cameras are located at the nodes of a connected graph ···, where ···and ···are the node set and edge set of ···respectively. For multi-view stereo vision, the geometric model of the ···-th camera is ···, and by combining camera models on adjacent nodes ···the three-dimensional reconstruction formula (2) is obtained. The three-dimensional reconstruction formula for multi-view stereo vision is obtained by averaging.

3 Three-dimensional Reconstruction Formulas Based on Near-field High-order Spatial Coherence of Thermal Light

The classical imaging system for thermal light near-field is shown in Figure 5 [Figure 5: see original paper] [9]. The light field on the image plane is:

The Green' s function of the imaging system is:

where ···, ···, and ···are the two-dimensional coordinate vectors on the object plane, lens plane, and image plane respectively, ···is the Fresnel phase factor (Gaussian function), ···is the distance between the object and the imaging lens, ···is the distance between the imaging lens and the image plane, and ···is the distribution function of the light field on the object.

Utilizing the properties of the Gaussian function, the above equation simplifies to the following form: ···For a finite-size lens with radius R, the second double integral in the above equation simplifies to yield the point spread function of this imaging system: ···where ···is the first-order Bessel function and ···is the magnification of the imaging system.

The near-field second-order spatial coherence function of thermal light [9] is: ··· Substituting the free-propagation Green' s function ···into the above equation yields the first-order spatial coherence function ···, thus obtaining:

As shown in Figure 6 [Figure 6: see original paper], transforming the joint measurement of the ···interferometer at two independent observation points into a joint measurement of two imaging systems converts the ···interferometer into a stereo imaging measurement for binocular vision.

The Gaussian lens formula for left and right camera imaging is: ···Substituting the imaging system' s Green' s function (10) into equation (12) yields: ···Assuming the light field distributions on the object are identical for the left and right imaging systems, the above equation can be further simplified. When the Gaussian lens formula holds, we have:

Thus, the three-dimensional reconstruction formula (3) for the near-field first-order coherence function of thermal light is obtained. In the first-order coherence function ···, the expansion function ···contains vectors in the world coordinate system, and the expansion function ···includes disparity information of corresponding points ···and ···on the two images.

Therefore, based on equation (3), joint coincidence measurement of the two images in binocular vision reconstructs the three-dimensional shape of the object, and the expansion function ···simultaneously performs real-time stereo matching of corresponding points in the left and right images.

Assuming ···photodetectors are located at the nodes of a connected graph ··· , transforming the joint measurement of the ···interferometer at ···independent observation points into a joint measurement of ···imaging systems converts the ···interferometer into a stereo imaging measurement for multi-view vision. The near-field ···-order spatial coherence function of thermal light [9] is:

Reference [9] measured the nontrivial third-order coherence function of thermal light, which exhibits a series of unusual and interesting properties. Experimental results and theoretical simulations agree well within statistical errors, and similar methods can be used to experimentally verify the correctness of high-order coherence for ···imaging systems. Assuming the Green' s function of the ··· -th imaging system is: ···where ···, ···, and ···are the two-dimensional coordinate vectors on the object plane, lens plane, and image plane respectively, and ···is the magnification of the imaging system.

For a connected graph ···, the near-field ···-order spatial coherence function of thermal light is obtained from the third-order spatial coherence function in reference [9]: ···For arbitrary nodes ···and ···, with the photodetectors of the ··· imaging systems located on a sphere of radius ···, and identical object light field distributions, the three-dimensional reconstruction formula (4) for multi-view stereo imaging is obtained.

4 Mechanism of Binocular Stereo Vision Based on Near-field Second-order Coherence of Thermal Light

Binocular stereo vision arises from the horizontal disparity of objects imaged in both eyes. Relative to the head, the perception of scene objects by the two eyes follows the same head-centered visual direction, equivalent to viewing external objects from a "cyclopean eye" located between the two eyes. The visual center directions of the two eyes form a common visual direction, which is the average direction of the two eyes [8]. The three-dimensional reconstruction formulas (1) and (3) derived in this paper also happen to include the average center coordinates ···on the imaging planes of the two eyes, consistent with visual neuroscience theories.

The human visual system includes visual cortices such as V1 and V2 [20,21]. In different regions of layer V1, layer 4C cells receive monocular information input from the lateral geniculate nucleus of both the ipsilateral and contralateral sides. In subsequent layers, binocular information converges to varying degrees, with over 80% of cells outside layer 4C being binocularly driven. The secondary visual cortex V2 area is adjacent to V1 and receives its orderly projections. Most disparity cells in the V2 area, which are driven by both eyes, participate in stereo visual information processing. Binocular-driven cells have corresponding receptive fields in the left and right eyes, with both receptive fields located at corresponding spatial positions in the two visual fields, possessing similar spatiotemporal characteristics. Most cells are highly sensitive to spatial disparity between the two receptive fields. Therefore, binocular-driven cells are equivalent to the joint coincidence measurement circuit for near-field second-order coherence measurements of thermal light.

Cyclopean stereo perception is illustrated in Figure 7 [Figure 7: see original paper] [22,23] using random dot stereograms (Wandell, 1995; Julesz, 1964). One eye views a random dot pattern, while the stimulus for the other eye is generated by copying the first image, horizontally shifting a specific region, and then filling the gaps with random dot samples. When both eyes view the left and right images simultaneously but independently (by using a stereoscope or converging/diverging the eyes through stereo matching to fuse the images), the plane of the shifted region (square) appears to protrude upward, and the brain perceives depth information. Beyond the two eyes, a third "cyclopean eye" perceives the stereo image. The random dot pattern is meaningless when viewed monocularly, yet the binocular system can perceive depth information solely through disparity information. Similarly, near-field second-order coherence measurement of light only requires corresponding information from stereo matching of left and right images to form a stereo image, and the jointly measured HBT interferometer serves as the "cyclopean eye."

Figure 7 Binocular fusion perception diagram of Julesz random dot stereogram. Therefore, binocular stereo perception in the brain' s visual system is a second-order coherence measurement of the object light field performed by two photosensitive observation cells in the left and right eyes. The theoretical knowledge about visual neuroscience presented in this section provides a neuroscientific foundation for equations (1) and (3) in this paper.

References

[1] Gao Hongwei. Computer Binocular Stereo Vision[M]. Beijing: Electronic Industry Press, 2012: 129-149.

[2] Zhang Guangjun. Visual Measurement[M]. Beijing: Science Press, 2008: 134-174.

[3] Xie Jianbin, Chen Zhangyong, Liu Tong, et al. Principles and Applications of Visual Bionics[M]. Beijing: Science Press, 2013: 95-128.

[4] Zhang Xiaowu. 3D Reconstruction Method Constrained by Homogeneous Linear Equations[J]. Journal of Ningxia University: Natural Science Edition, 2021(4): 383-390.

[5] Richard Hartley, Andrew Zisserman. Multiple View Geometry in Computer Vision (2nd Edition)[M]. Wei Sui, Zhang Quanbing trans. Beijing: China Machine Press, 2020: 203-354.

[6] Kong Qingqun, Ming Yansheng, Hu Zhanyi. Disparity Computation in Visual Cortex[J]. Acta Automatica Sinica, 2011(6): 645-657.

[7] Kong Qingqun, Ming Yansheng, Hu Zhanyi. Weighted Disparity Energy Model[J]. Acta Automatica Sinica, 2014(2): 227-235.

[8] John S. Werner, Leo M. Chalupa. The New Visual Neurosciences[M]. The Massachusetts Institute of Technology Press, 2014: 809-824.

[9] Shi Shuohua. Introduction to Quantum Optics—Single-Photon and Two-Photon Physics[M]. Xu Ping trans. Beijing: Higher Education Press, 2016: 44-104, 254-289.

[10] Wu Jian. Classical Second-Order Coherence of Light[M]. Beijing: Science Press, 2017: 50-99.

[11] Wang Xiaoxi, Gao Chao, Yao Zhihai. Computational Ghost Imaging Technology from the Perspective of Classical Theory[M]. Beijing: Science Press, 2022: 8-24.

[12] Li Ronghua. Theory and Applications of Binocular Stereo Vision[M]. Beijing: Science Press, 2020: 14-135.

[13] Li Minglei. 3D Measurement and Modeling in Computer Vision[M]. Beijing: Electronic Industry Press, 2022: 1-29, 116-151.

[14] Peng Yali. Multi-View 3D Reconstruction Technology[M]. Beijing: Science Press, 2021: 1-50.

[15] Wu Fuchao. Mathematical Methods in Computer Vision[M]. Beijing: Science Press, 2008: 46-78, 227-244.

[16] Zhang Zonghua, Liu Wei, Liu Guodong, et al. Advances in 3D Vision Measurement Technology and Applications[J]. Journal of Image and Graphics, 2021(6): 1483-1498.

[17] Long Xiaoxiao, Cheng Xinjing, Zhu Hao, et al. Frontiers in 3D Vision[J]. Journal of Image and Graphics, 2021(6): 1389-1428.

[18] Zheng Yajing, Yu Zhaofei, Huang Tiejun. A Survey of Neural Network Coding Models of Biological Visual Systems[J]. Journal of Image and Graphics, 2023(2): 335-357.

[19] Zheng Taixiong, Huang Shuai, Li Yongfu, et al. A Survey of Key Technologies for Vision-Based 3D Reconstruction[J]. Acta Automatica Sinica, 2020(4): 631-652.

[20] Eric R. Kandel, John D. Koester, Sarah H. Mack, et al. Principles of Neural Science (Sixth Edition)[M]. Copyright 2021 by McGraw Hill, all rights reserved: 496-597.

[21] A. Robert Martin, David A. Brown, Mathew E. Diamond. From Neuron to Brain (Sixth Edition)[M]. Oxford University Press, 2020: 23-59.

[22] Dale Purves, George J. Augustine, David Fitzpatrick, et al. Neuroscience (Sixth Edition)[M]. Oxford University Press, 2018: 233-280.

[23] Harvey Richard Schiffman. Sensation and Perception (5th Edition)[M]. Li Leshan trans. Xi' an: Xi' an Jiaotong University Press, 2014: 229-258.

Author Contributions: Zhang Xiaowu: Conceived the research idea, designed the research plan; drafted the manuscript; revised the final version.

Note: Figure translations are in progress. See original paper for figures.