

# 단안 깊이 지각의 보완 가능성: 동물의 시야와 비교하여

이채헌, 허재필\*

성균관대학교

bethelight@skku.edu, \*jaepilheo@skku.edu

## Improving Monocular Depth Estimation: Biological Inspirations

Lee Chae Heon, Heo Jae-Pil\*

Sungkyunkwan Univ.

### 요 약

컴퓨터 비전에서 깊이 지각 및 예측은 장면 분할과 더불어 중요한 문제로 꼽힌다. 최근에는 monocular depth estimation을 통해 한 장의 이미지에서 깊이를 예측하는 방법이 사용되고 있다. 허나 동시에 취약점이 존재함은 여러 성공적인 real-world attack에 의해 증명되었다. 깊이 지각은 또한 동물에서도 생존에 중요한 문제이다. 따라서 본 연구에서는 동물의 양안시 및 단안시적 깊이 지각 방법에 대한 비교 분석을 통해 monocular depth estimation의 보완에 대한 아이디어를 제시하고자 한다.

### I. 서 론

컴퓨터 비전은 영상, 비디오 등 다양한 매체를 통해 받아들이는 시각적 자료에서 유의미한 정보를 추출할 수 있으며, 또한 유의미한 행동이나 결정을 내리는 데에 해당 정보를 사용할 수도 있다[22]. 컴퓨터 비전은 CT와 X-ray 결과의 분석, 생산 설비에서의 불량 확인 등 여러 분야에서의 활용이 연구되고 있다.

컴퓨터 비전을 비롯하여 인공지능이 효율적으로 작동해야 하는 환경 중 많은 경우는 사람에게 맞추어 설계되어 있다[16]. 안구 각각의 시야가 별개로 존재하는 경우 단안시(monocular view)를 가진다고 할 수 있다. 사람은 상대적으로 두개골 전면에 두 개의 안구가 위치하므로 양안시(binocular view)를 가지고 있다.

### II. 본론

#### 1. 동물에서의 깊이 지각

##### 1) 양안 깊이 지각

양안시의 경우 단안시에 비해 몇 가지 유리한 점을 지닌다. 양쪽의 시야가 겹치는 부분이 커질수록 빛을 효율적으로 활용할 수 있다. 즉, 빛에 대한 민감성이 증가하는 것이다. 이는 다양한 환경에 시각이 적응할 수 있음을 의미한다. 이는 한쪽 눈이 기능하지 못하더라도 시각을 완전히 잃어버리지 않는다는 점에 의해서도 뒷받침된다. 또한, 같은 장면에 대한 양쪽 시야의 정보 차이 역시 여러 방면으로 활용된다.

#### 2) 단안 깊이 지각

사람의 경우 깊이 지각에 대한 단안단서(monocular cue)로는 선명도, 중첩, 결의 밀도, 직선적 조망(linear perspective), 결의 밀도, 명암이 있다[2]. 이 중 직선적 조망과 결의 밀도는 각기 멀어질수록 사이의 간격이 가까워 보인다는 면에서 유사하다. 다만, 이러한 단안 단서는 양안 불일치와 같은 양안단서와 결합하여 사용된다.

#### 2. 컴퓨터 비전에서의 깊이 예측

##### 1) Stereo Depth Estimation

컴퓨터 비전에서의 깊이 예측은 자율주행과 로봇틱스, 3D 복원, 가상현실 등 여러 분야에서 중요한 역할을 한다. 정확하고 밀도가 높은 깊이 지도(depth map)를 그리는 것은 이러한 과제를 해결하는 데에 필수적이며, 따라서 여러 방법이 고안되었으며 사용 중이다.

딥러닝 상용화 이전에는 양안시에 대한 연구를 통해 운동 시차(motion parallax)와 불일치를 기반으로 계산과 등극선 기하(epipolar geometry)를 통해 깊이 지도를 그렸으며 카메라의 위치 조정에 따른 초점 흐림과 텍스처를 통해서도 예측이 가능했다[37].

이러한 방법은 calibration, 즉 카메라의 초점거리와 방향, 각도와 같이 외내부의 지표에 대한 의존도가 높다. 또한, 여러 장의 이미지와 정밀한 직접적 측정을 요구하는 만큼 필요한 장비 및 계산 자원, 복잡한 캘리브레이션 조건 등으로 인해 활용에 한계가 존재한다.

##### 2) Monocular Depth Estimation

기존의 컴퓨터 비전에서는 기하학적인 방법을 바탕으로 깊이 예측이 이루어졌다면 딥러닝의 등장으로 여러 정보를 종합함으로써 더 정확하고 다양한 방법으로 깊이 예측이 가능하게 되었다. 이 중에서도 특히 영향을 받은 MDE 는 한 장의 컬러, 즉 RGB 이미지를 이용하여 깊이 예측이 이루어지는 방법으로, 요구되는 자료가 적은 만큼 활용 범위가 더 넓다.

### III. 결론

#### 1. 보완 방법 및 가능성

인간을 비롯하여 동물들도 사람의 시지각에 사용되는 것과 유사한 단안단서를 이용하여 깊이를 예측한다[8]. 사람의 경우 한쪽 눈으로 보더라도 해당 안구에서 상의 위치 뿐만 아니라 양안시의 경험적 근거에 기반하여 황반중심좌에 대한 상대적 위치를 예측하여 깊이 측정을 진행한다[25, 43]. 즉, 단안적 단서를 기반으로 한 정확한 깊이 예측을 위해서도 이러한 경험의 적용이 필요하다고 할 수 있다. 현재 MDE 에 대한 real-world adversarial attack 에 대한 대응으로는 주로 위와 같은 적대적인 자료 자체를 훈련 자료에 포함되는 방식이 사용된다[17]. 이러한 방식의 영향 중 하나는 MDE 자체의 일반화 가능성(generalization capability)을 높인다는 점이다. 또한, 라벨링되지 않은 데이터를 이용하더라도 자기지도학습을 통해 훈련 데이터의 여러 무작위적인 변형을 적대적 공격에 유사하게 생성하고 학습하는 방식으로 대응하는 방법도 사용된다[26]. 이와 같이, 주어진 자료에서는 상대적인 지표를 기준으로 예측하더라도 양안시적 경험에 해당하는 넓은 범위의 학습을 통해 상대성 근간의 범위의 확장으로 예측의 정확도를 높일 수 있을 것으로 예상된다..

### ACKNOWLEDGMENT

#### 참 고 문 헌

[1] 고희빈. (2012). 스테레오스코피 3D 공간에 대한 고찰. 현대영화연구, 8(1), 154-193.

[2] 감기택. & 한광희. (1999). 양안 부동 처리에 있어서 구별되는 두 시야 영역. 한국심리학회지: 인지 및 생물, 11(1), 1-16.

[3] 김삼섭. (2010). 특수교육의 심리학적 기초(개정판). 시그마프레스.

[4] Basgöze, Z., White, D. N., Burge, J., & Cooper, E. A. (2020). Natural statistics of depth edges modulate perceptual stability. Journal of Vision, 20(8), 10. <https://doi.org/10.1167/jov.20.8.10>

[5] Bedinghaus, T., OD. (2022, November 25). Depth perception: What is it and how does it work? Verywell Health. <https://www.verywellhealth.com/depth-perception-3421547>

[6] Brunetti, A., Buongiorno, D., Trotta, G. F., & Bevilacqua, V. (2018). Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing, 300, 17- 33. <https://doi.org/10.1016/j.neucom.2018.01.092>

[7] Carlini, N., & Wagner, D. (2016). Towards evaluating the robustness of neural networks. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.1608.04644>

[8] Cavoto, B. R., & Cook, R. G. (2006). The contribution of monocular depth cues to scene perception by pigeons. Psychological Science, 17(7), 628- 634. <https://doi.org/10.1111/j.1467-9280.2006.01755.x>

[10] Eigen, D., Puhrsch, C., & Fergus, R. (2014). Depth Map Prediction from a Single Image using a Multi-Scale Deep Network. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.1406.2283>

[11] Feather, J., Leclerc, G., Ma dry, A., & McDermott, J. H. (2023). Model metamers reveal divergent invariances between biological and artificial neural networks. Nature Neuroscience, 26(11), 2017- 2034. <https://doi.org/10.1038/s41593-023-01442-0>

[12] Firestone, C. (2020). Performance vs. competence in human-machine comparisons. Proceedings of the National Academy of Sciences, 117(43), 26562- 26571. <https://doi.org/10.1073/pnas.1905334117>

[13] Forte, J., Peirce, J. W., & Lennie, P. (2002). Binocular integration of partially occluded surfaces. Vision Research, 42(10), 1225- 1235. [https://doi.org/10.1016/s0042-6989\(02\)00053-6](https://doi.org/10.1016/s0042-6989(02)00053-6)

[14] Fox, R., Lehmkuhle, S. W., & Bush, R. C. (1977). Stereopsis in the Falcon. Science, 197(4298), 79- 81. <https://doi.org/10.1126/science.867054>

[15] Garg, R., BG, V. K., Carneiro, G., & Reid, I. (2016). Unsupervised CNN for single view Depth estimation: Geometry to the rescue. In Lecture notes in computer science (pp. 740- 756). [https://doi.org/10.1007/978-3-319-46484-8\\_45](https://doi.org/10.1007/978-3-319-46484-8_45)

[16] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and Harnessing Adversarial Examples. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.1412.6572>

[17] Goutcher, R., Barrington, C., Hibbard, P. B., & Graham, B. (2021). Binocular vision supports the development of scene segmentation capabilities: Evidence from a deep learning model. Journal of Vision, 21(7), 13. <https://doi.org/10.1167/jov.21.7.13>

[18] Groenendijk, R., Karaoglu, S., Gevers, T., & Mensink, T. (2019). On the benefit of adversarial training for monocular depth estimation. Computer Vision and Image Understanding, 190, 102848. <https://doi.org/10.1016/j.cviu.2019.102848>

[19] Guesmi, A., Hanif, M. A., Alouani, I., Ouni, B., & Shafique, M. (2024). SSAP: a Shape-Sensitive Adversarial patch for comprehensive disruption of monocular depth estimation in autonomous navigation applications. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2786- 2793. <https://doi.org/10.1109/iros58592.2024.10802252>

[20] Haji-Esmaili, M. M., & Montazer, G. (2023). Large-scale monocular depth estimation in the wild. Engineering Applications of Artificial Intelligence, 127, 107189. <https://doi.org/10.1016/j.engappai.2023.107189>

[21] Heesy, C. P. (2009). Seeing in stereo: The ecology and evolution of primate binocular vision and stereopsis. Evolutionary Anthropology Issues News and Reviews, 18(1), 21- 35. <https://doi.org/10.1002/evan.20195>

- [22] Hoffman, D. M., & Banks, M. S. (2010). Focus information is used to interpret binocular images. *Journal of Vision*, 10(5), 13. <https://doi.org/10.1167/10.5.13>
- [23] IBM. (2024, August 13). What is Computer Vision? Retrieved September 20, 2024, from <https://www.ibm.com/topics/computer-vision>
- [24] Ingle, D. (1972). Depth vision in monocular frogs. *Psychonomic Science*, 29(1), 37– 38. <https://doi.org/10.3758/bf03336560>
- [25] Junayed, M. S., Sadeghzadeh, A., Islam, M. B., Wong, L., & Aydin, T. (2022). HIMODE: a hybrid monocular omnidirectional depth estimation model. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). <https://doi.org/10.1109/cvprw56347.2022.00569>
- [26] Kim, M., Tack, J., & Hwang, S. J. (2020). Adversarial Self-Supervised contrastive learning. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2006.07589>
- [27] Kral, K. (2003). Behavioural- analytical studies of the role of head movements in depth perception in insects, birds and mammals. *Behavioural Processes*, 64(1), 1– 12. [https://doi.org/10.1016/s0376-6357\(03\)00054-8](https://doi.org/10.1016/s0376-6357(03)00054-8)
- [28] Karsch, K., Liu, C., & Kang, S. B. (2014). Depth Transfer: Depth Extraction from Video Using Non-Parametric Sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11), 2144– 2158. <https://doi.org/10.1109/tpami.2014.2316835>
- [29] Lamb, T. D., Collin, S. P., & Pugh, E. N. (2007). Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. *Nature Reviews. Neuroscience*, 8(12), 960– 976. <https://doi.org/10.1038/nrn2283>
- [30] Menegaz, R. A., & Kirk, E. C. (2009). Septa and processes: convergent evolution of the orbit in haplorhine primates and strigiform birds. *Journal of Human Evolution*, 57(6), 672– 687. <https://doi.org/10.1016/j.jhevol.2009.04.010>
- [31] Nagata, T., Koyanagi, M., Tsukamoto, H., Saeki, S., Isono, K., Shichida, Y., Tokunaga, F., Kinoshita, M., Arikawa, K., & Terakita, A. (2012). Depth Perception from Image Defocus in a Jumping Spider. *Science*, 335(6067), 469– 471. <https://doi.org/10.1126/science.1211667>
- [32] Nityananda, V., & Read, J. C. A. (2017). Stereopsis in animals: evolution, function and mechanisms. *Journal of Experimental Biology*, 220(14), 2502– 2512. <https://doi.org/10.1242/jeb.143883>
- [33] Poggio, G. F., & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology*, 40(6), 1392– 1405. <https://doi.org/10.1152/jn.1977.40.6.1392>
- [34] Smolyanskiy, N., Kamenev, A., & Birchfield, S. (2018). On the Importance of Stereo for Accurate Depth Estimation: An Efficient Semi-Supervised Deep Neural Network Approach. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1803.09719>
- [35] Tao, T., Chen, Q., Feng, S., Hu, Y., & Zuo, C. (2018). Active depth estimation from defocus using a camera array. *Applied Optics*, 57(18), 4960. <https://doi.org/10.1364/ao.57.004960>
- [36] Timney, B., & Keil, K. (1996). Horses are Sensitive to Pictorial Depth Cues. *Perception*, 25(9), 1121– 1128. <https://doi.org/10.1068/p251121>
- [37] Timney, B., & Keil, K. (1999). Local and global stereopsis in the horse. *Vision Research*, 39(10), 1861– 1867. [https://doi.org/10.1016/s0042-6989\(98\)00276-4](https://doi.org/10.1016/s0042-6989(98)00276-4)
- [38] Trychin, S., & Walk, R. D. (1964). A study of the depth perception of monocular hooded rats on the visual cliff. *Psychonomic Science*, 1(1– 12), 53– 54. <https://doi.org/10.3758/bf03342786>
- [39] Tyrrell, L. P., & Fernández-Juricic, E. (2017). Avian binocular vision: It's not just about what birds can see, it's also about what they can't. *PLoS ONE*, 12(3), e0173235. <https://doi.org/10.1371/journal.pone.0173235>
- [40] Veerabadran, V., Goldman, J., Shankar, S., Cheung, B., Papernot, N., Kurakin, A., Goodfellow, I., Shlens, J., Sohl-Dickstein, J., Mozer, M. C., & Elsayed, G. F. (2023). Subtle adversarial image manipulations influence both human and machine perception. *Nature Communications*, 14(1). <https://doi.org/10.1038/s41467-023-40499-0>
- [41] Wheatstone, C. (1838). XVIII. Contributions to the physiology of vision. —Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 128, 371– 394. <https://doi.org/10.1098/rstl.1838.0019>
- [42] Wong, A., Mundhra, M., & Soatto, S. (2020). Stereopagnosia: Fooling Stereo Networks with Adversarial Perturbations. *AAAI Conference on Artificial Intelligence*.
- [43] Xie, Cihang & Zhang, Zhishuai & Wang, Jianyu & Zhou, Yuyin & Ren, Zhou & Yuille, Alan. (2018). Improving Transferability of Adversarial Examples with Input Diversity. 10.48550/arXiv.1803.06978.
- [44] Yamanaka, K., Matsumoto, R., Takahashi, K., & Fujii, T. (2020). Adversarial patch attacks on monocular depth estimation networks. *IEEE Access*, 8, 179094– 179104. <https://doi.org/10.1109/access.2020.3027372>
- [45] Yildiz, G. Y., Sperandio, I., Kettle, C., & Chouinard, P. A. (2021). A review on various explanations of Ponzo-like illusions. *Psychonomic Bulletin & Review*, 29(2), 293– 320. <https://doi.org/10.3758/s13423-021-02007-7>