Biological and Active Vision: From Stereopsis to 3D Scene Understanding

Graduate course · 14 weeks · Mathematics, computational experiments, and research-oriented projects

The unifying scientific question would be:

How can an intelligent visual system control its sensing processes to construct useful representations of the three-dimensional world?

The course would deliberately cross four disciplines: visual neuroscience, projective geometry, machine learning, and active perception/control.

1. Curriculum

Biological foundations of visual perception Weeks 1–3

Retinal organization, receptive fields, foveation, binocular disparity, vergence, accommodation, saccades, eye–head coordination, motion parallax, and perceptual integration.

Learning objective: Identify biological mechanisms that suggest testable computational principles.

Geometry and computational stereopsis Weeks 4–6

Projective and multiview geometry, calibrated stereo, neural disparity models, correspondence, multiscale matching, depth uncertainty, learned stereo, and the geometry of eye and head motion.

Learning objective: Understand what can be measured geometrically, what requires inference, and what is unobservable.

Active and foveated perception Weeks 7–9

Nonuniform sampling, space-variant receptive fields, temporal sampling, gaze selection, information gain, Bayesian filtering, belief-state representations, and uncertainty-aware sensing.

Learning objective: Design a visual system that chooses what information to acquire.

3D reconstruction and scene understanding Weeks 10–12

Object-centric representations, surface exploration, segmentation, recognition, scene graphs, near/far decomposition, active reconstruction, and sequential integration.

Learning objective: Connect measured geometry to persistent, meaningful scene representations.

Research projects and scientific evaluation Weeks 13–14

Reproduction studies, prototype demonstrations, ablations, benchmark comparisons, failure analysis, and research presentations.

Learning objective: Develop and experimentally evaluate a falsifiable research hypothesis.

The mathematical prerequisites would include linear algebra, optimization, probability, and basic multiview geometry, plus familiarity with machine learning and programming.

2. The laboratory component is essential

I would organize the practical work around one progressively developed artificial visual system.

Students would begin with a simulated binocular pair observing a simple 3D scene.

First, they would implement conventional stereo and understand its limitations. Next, they would introduce nonuniform retinal sampling and compare it with uniform sensing. Finally, they would add controlled fixation, vergence, a simple gaze-selection policy, and sequential scene reconstruction.

The final project should not merely demonstrate that the system works.

It should answer a question such as:

Under a fixed observation budget, does active binocular foveation improve the recovery of object geometry and boundaries compared with passive acquisition?

That is a legitimate, measurable scientific question.

A useful secondary project would investigate how the results change when small head translations become available.

3. What should students learn beyond the technical content?

The most valuable outcome would be a change in scientific perspective.

Students should learn to distinguish:

- The information physically available in the light field.
- The information actually measured by the sensors.
- The information inferred using models and prior knowledge.
- The information required by a particular task.

These distinctions are frequently blurred in modern AI-based vision systems.

A PhD student who understands them should be better equipped to identify observability limits, distinguish predictions from measurements, and design experiments that reveal why a visual system succeeds or fails.

4. Bibliography:

Below is a consolidated bibliography of the literature cited in my previous analysis, organized into five thematic areas. I have verified the bibliographic details and included links to the publications.

I. Foundations of active and biologically inspired vision

1. Bajcsy, R. (1988). Active Perception. Proceedings of the IEEE, 76(8), 966–1005.

A foundational paper proposing perception as a process involving active control of sensing, rather than passive processing of images.

2. Ballard, D. H. (1991). Animate Vision. Artificial Intelligence, 48(1), 57–86.

Introduces a framework in which gaze control, visual attention, and behavior simplify visual computation. Particularly relevant to active fixation and foveated sensing.

3. Hung, Y.-C., Guo, Q., & Alexander, E. (2025). Bio-Inspired Computational Imaging: Components, Algorithms, and Systems. Annual Review of Vision Science, 11, 565–590.

A contemporary review of biological inspiration in optical design, sensors, computational imaging, and vision algorithms.

II. Human stereopsis and biological depth perception

4. Cumming, B. G., & DeAngelis, G. C. (2001). The Physiology of Stereopsis. Annual Review of Neuroscience, 24, 203–238.

A major review of neural mechanisms underlying binocular disparity processing and stereoscopic depth perception.

5. Gomes Tomaz, Â., Chopin, A., Alcalde, N., Levi, D. M., & Verghese, P. (2026). The Best Stereoacuity Is Rarely at the Fovea. Vision Research, 240, 108748.

An experimental study challenging the assumption that the fovea invariably provides the best stereoscopic acuity. Relevant to spatial resolution allocation in artificial binocular systems. Published online in December 2025, with a March 2026 journal issue date.

6. Wilcox, L. M., & Allison, R. S. (2026). Reexamining the Relationship Between Stereopsis and Motion Parallax. Annual Review of Vision Science, 12, 391–411.

A particularly relevant review of the relationships between binocular disparity, motion parallax, and eye–head movement.

III. Contemporary learning-based stereo and 3D reconstruction

7. Wen, B., Trepte, M., Aribido, J., Kautz, J., Gallo, O., & Birchfield, S. (2025). FoundationStereo: Zero-Shot Stereo Matching. Proceedings of CVPR 2025, 5249–5260.

A modern foundation-model approach to stereo depth estimation, illustrating the effectiveness of learned priors and generalization across domains.

8. Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C., & Novotny, D. (2025). VGGT: Visual Geometry Grounded Transformer. Proceedings of CVPR 2025, 5294–5306.

A unified neural architecture for recovering camera parameters, depth, point maps, and 3D tracks from image collections.

IV. Active perception and 3D reconstruction

9. Alsadik, B., Mahdi, H. A., & Abdulateef, N. A. (2025). A Structured Review and Taxonomy of Next-Best-View Strategies for 3D Reconstruction. ISPRS Open Journal of Photogrammetry and Remote Sensing, 17, 100098.

A systematic review of next-best-view techniques, including rule-based, uncertainty-based, sampling-based, learning-based, and prediction-based approaches.

10. Chen, L., Zhan, H., Chen, K., Xu, X., Yan, Q., Cai, C., & Xu, Y. (2025). ActiveGAMER: Active GAussian Mapping Through Efficient Rendering. Proceedings of CVPR 2025, 16486–16497.

Combines active exploration, information-driven viewpoint selection, and 3D Gaussian Splatting for scene mapping.

11. Li, Y., Guo, X., Zhang, H., Chen, J., Li, S., & Dai, X. (2026). Active Visual Perception: Opportunities and Challenges. Sensors, 26(20), 6544.

A recent review proposing a unified perspective on active observation acquisition, task-driven sensing, action selection, evaluation, and uncertainty.

V. Foveation, adaptive sensing, and neural rendering

12. Blauch, N., Alvarez, G. A., & Konkle, T. (2026). FOVI: A Biologically-Inspired Foveated Interface for Deep Vision Models. Proceedings of the 43rd International Conference on Machine Learning (ICML), PMLR 306, 8446–8469.

Introduces a retina-inspired variable-resolution sensing interface with a V1-like computational representation, demonstrating compatibility with contemporary neural vision models.

13. Ghosh, S., & Gallego, G. (2025). Event-Based Stereo Depth Estimation: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(10), 9130–9149.

Reviews asynchronous stereo depth estimation with event cameras, including geometric, learning-based, and temporal approaches.

14. Faramarzi, F., Linares-Barranco, B., & Serrano-Gotarredona, T. (2024). A 128×128 Electronically Multi-Foveated Dynamic Vision Sensor With Real-Time Resolution Reconfiguration. IEEE Access, 12, 192656–192671.

Demonstrates a physical sensor capable of dynamically configuring multiple regions of high spatial resolution. Particularly relevant to hardware realizations of adaptive foveation.

15. Shi, X., Wang, L., Liu, X., Wu, J., & Shao, Z. (2025). Scene-Aware Foveated Neural Radiance Fields. IEEE Transactions on Visualization and Computer Graphics, 31(9), 5039–5054.

Investigates foveated neural rendering for immersive environments, illustrating the exploitation of perceptual resolution requirements in 3D scene visualization.

16. Franke, L., Fink, L., & Stamminger, M. (2025). VR-Splatting: Foveated Radiance Field Rendering via 3D Gaussian Splatting and Neural Points. Proceedings of the ACM on Computer Graphics and Interactive Techniques, 8(1), Article 18.

A hybrid foveated rendering approach combining Gaussian representations for peripheral regions with more detailed neural representations around the fovea.

Suggested reading priorities

For a PhD course or a new research initiative, I would not assign equal importance to all papers.

I would prioritize these eight:

Priority	Reference	Principal role
1	Bajcsy (1988)	Foundations of active perception
2	Ballard (1991)	Gaze control and computational organization
3	Cumming & DeAngelis (2001)	Biological mechanisms of stereopsis
4	Hung et al. (2025)	Modern biological inspiration for artificial sensing
5	Wilcox & Allison (2026)	Binocular depth and observer motion
6	Blauch et al. (2026)	Practical variable-resolution visual representations
7	Alsadik et al. (2025)	Active 3D acquisition and next-best-view strategies
8	Li et al. (2026)	Integrated active-perception research challenges

The remaining publications provide important contemporary comparisons, enabling technologies, and application examples.

A particularly useful reading combination would be Bajcsy (1988), Ballard (1991), and Li et al. (2026). Together they offer a perspective spanning nearly four decades, from the original motivation for active perception to the problems that remain unresolved today.

For the specific research direction of active foveated binocular 3D scene understanding, I would additionally put Cumming & DeAngelis (2001), Wilcox & Allison (2026), and Blauch et al. (2026) at the center of the discussion.