

Research Statement

Jan Neumann

jneumann@cs.umd.edu • www.videogeometry.com

The central question that I ask in my research is “How should a computer see the world?” When we think about vision, we usually think of interpreting the images taken by (two) eyes such as our own - that is, perspective images acquired by cameras based on the pinhole principle. Fundamentally though, there is no need to confine a machine to see the world with human eyes as long as we do not equip the machine with the capabilities of a human brain at the same time. Therefore, in my research I investigate how I can tailor the “eyes” of the machine to its computational capabilities and the task we are trying to solve. I believe that only with a unified approach to sensor and algorithm design we will be able to develop truly effective autonomous cars and robots, 3D shape and motion capture solutions, pervasive human-computer interfaces and reliable surveillance and recognition systems.

This concept is illustrated in Figure 1. Current computer vision research focuses on processing images (components enclosed by the solid red box). If instead we optimize all aspects of the vision process, including the image acquisition stage (components enclosed by the dashed blue box), we can develop solutions that are specifically optimized with respect to the capabilities of the machine and the task at hand. To de-

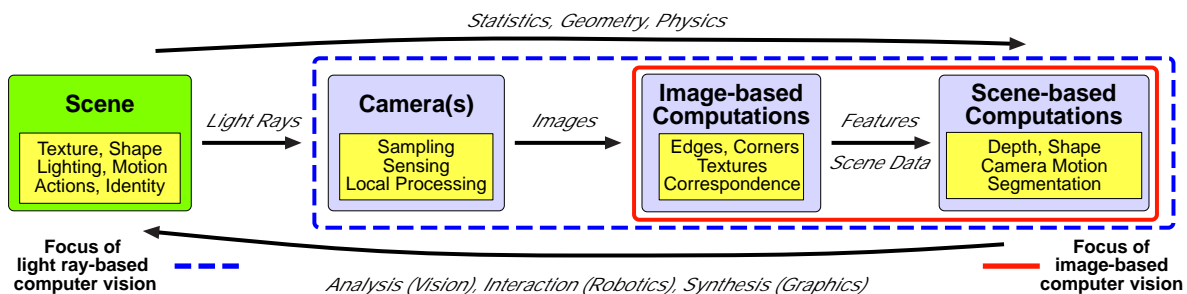


Figure 1: The processing pipeline in computer vision.

termine the optimal camera design and processing algorithms for a given task and machine, I study the properties of the most complete representation of visual information, the time-varying space of light rays. Using tools from computer science, mathematics, and engineering I simultaneously investigate how to process and how to sense the visual information:

Geometry of Light Rays: How is the information of interest encoded in the space of light rays?

To understand how the geometric structure of the space of light rays can be related to the properties of the scene, I combine signal processing and harmonic analysis with differential and multi-view geometry. The space of light rays naturally integrates signal processing and geometrical computations, enabling the definition of statistically optimal objective functions that account for both image-based intensity and scene-based geometric constraints. The integration of multi-view geometry and signal processing is essential to develop efficient and robust algorithms that can accurately recover spatio-temporal descriptions of an environment. Up to now I utilized this insight to develop a scene-independent linear 3D motion estimation algorithm for multi-perspective cameras [2, 4] and I applied these concepts to compute global 3D shape and motion models using multi-resolution subdivision surfaces [1] and level sets.

My light ray based approach to computer vision allows me to study many useful vision systems that cannot be adequately analyzed in terms of individual images. In the future I want to investigate how we can best recover shape and action descriptions using non-standard “cameras” such as small “insect-like” compound eyes, dynamic camera networks consisting of portable vision systems carried by soldiers and emergency personnel, or in the near future, the so-called “brilliant rocks”, a collection of distributed, low-resolution imaging sensors.

Sampling of Light Rays: How can we design the best camera for our computations?

For an optimal solution the design of the camera must be addressed jointly with the algorithm design. For example, efficient mathematics will demand a particular sampling of the light rays, but the sensors

may not be able to deliver a sufficient signal-to-noise ratio for a given sampling strategy. This in turn would make the computations too noise sensitive for real world applications. In my research I interpret cameras as spatio-temporal filters in the space of light rays which are defined by the camera optics, the sensor electronics and the geometry and distribution of the imaging surfaces. By extending techniques from signal processing and approximation theory to the space of light rays, I can accurately describe the imaging process for any assembly of cameras in mathematical terms. This description makes it possible to establish a quantitative relation between the design parameters of a camera or camera network and the accuracy of the computations we perform based on the camera output. We can then evaluate this quantitative relationship using statistical models of the environment in which the camera is supposed to operate to define a fitness function on the space of cameras. By optimizing over the camera design parameters we can determine the best camera design with respect to the task at hand. So far I have defined fitness functions on the space of cameras for 3D motion estimation [3] and 3D photography [5]. Currently, I am validating the accuracy of the fitness functions by implementing different camera prototypes.

Applications of camera design and multi-perspective image processing

I am very excited about the technological promises of optical MEMS-technology and distributed sensor networks. Soon it will be possible for anyone to use assemblies consisting of mini-cameras, micro-mirrors or light guides as input devices for the tasks they try to solve. Since these camera assemblies can be reconfigured easily, the design of new eyes will be possible with little effort, enabling us to customize the design of cameras and their distribution, as we already customize the algorithms. In the future I plan to extend the framework of my thesis to study these camera assemblies. This would allow me to make the adaptive sensor an integral part of the computational solution.

The quantitative analysis of the sampling properties of camera systems in the space of light rays makes it also possible to develop optimal algorithms to collaboratively analyze multiple video streams that observe the same scene (light fields). I think that the integrated development of sensors and algorithms for multiple-view processing will be a fundamental component in the design of the next generation of immersive environments, smart rooms, 3D imaging sensors and camera networks.

Besides the sensing and algorithmic components of visual information processing, I am also interested in developing multi-scale and probabilistic representations for the objects of interests and efficient algorithms to visualize and interact with them. Examples are multi-resolution data structures that allow for efficient computations of the 3D shape and motion of human body parts such as hands or faces while incorporating the uncertainty present in the sensing process [1].

Conclusion

I believe that to make full use of the opportunities offered by camera networks and custom-designed vision sensors, we need to jointly optimize all components of the sensing process within a unified mathematical framework. By integrating signal processing with geometry and sensor design with algorithm development, my research will deliver the necessary tools to build the vision systems of the future. More information about my research including publications and some illustrative videos can be accessed at www.videogeometry.com.

References

- [1] J. Neumann and Y. Aloimonos. Spatio-temporal stereo using multi-resolution subdivision surfaces. *International Journal of Computer Vision*, 47(1/2/3):181–193, 2002.
- [2] J. Neumann and C. Fermüller. Plenoptic video geometry. *Visual Computer*, 19(6):395–404, 2003.
- [3] J. Neumann, C. Fermüller, and Y. Aloimonos. Eye design in the plenoptic space of light rays. In *Proc. International Conference on Computer Vision*, volume 2, pages 1160–1167, 2003.
- [4] J. Neumann, C. Fermüller, and Y. Aloimonos. Polydioptric camera design and 3d motion estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume II, pages 294–301, 2003.
- [5] J. Neumann, C. Fermüller, and Y. Aloimonos. A hierarchy of cameras for 3d photography. *Computer Vision and Image Understanding*, 2004. in press.