A Biologically-inspired Robotic Vision System for Tracking Fast Moving Objects

Emerson J. Olaya and L. Abril Torres-Méndez
Robotics and Advanced Manufacturing Group, Cinvestav Saltillo
Ramos Arizpe, Coah, 25900, Mexico
Phone: +52 844 4389600 Fax: +52 844 4389610
e-mail: abril.torres@cinvestav.edu.mx

Abstract

In this work we present a complete framework of a robotic vision system. The system involves the design, construction of the active vision system and it considers some biological strategies from human visual systems in order to improve its capabilities in terms of speed and accuracy. Trying to imitate the complexity of biological processes to any level constitutes a big challenge in many research areas. The area of robotics is not the exception, as the human perceptual system comprises a set of mechanisms difficult to match by current technology. However, it also comprises a set of strategies which, once implemented over modern technology, improve its performance in a meaningful way. One of the main objectives of this research work is to build an active robotic vision system of open architecture, that exceeds the performance of previous designs, at least when compared to other commercial devices of the same kind, as well as setting a technological precedent useful for future investigations, giving the best possible use to the resources available. Our mission along the development of this project has been to make of this visual system more than an image storage system in a computer. According to this, we have also considered that the system be able to achieve complex tasks, such as the tracking of moving objects and the depth recovery for tridimensional reconstruction of scenes. We have also explored new horizons in the area of artificial intelligence such as functions for mapping from percepts to actions in order to propose some useful algorithms based on visual attention to scan the scene efficiently, and govern de system behavior.
In this context, our motivation is based on the field of cognitive neurology, where one of the main goals is to study the eyes movement in order to obtain information about the human visual attention.

The approach we adopted, rather than focusing on the eyes (cameras), it focus first on deciding which visual characteristics may be more relevant within a dynamic scene, and then direct the cameras to them. Our visual attention algorithm is of special interest when there is no a desired specific task for the system. Our motivation is to explore the scene in a more natural way compared to simply sweep out, in some order, the scene point by point or to wait patiently that an object of interest appear. We assure with this that all relevant visual information in the scene is taken into account according to the priorities and objectives of the system. We combine three of the existing methods in the literature. One is the von Helmholtz's approach [1], who observed the relation between the eyes and an involuntary segmentation process of the visual field, where the eyes are attracted to the new. Another approach was presented by James [2], in which the eyes are fixed voluntarily over the selected region, with the goal of identifying it, exploring it or just not losing it from the sight. These two approaches were reinforced by Nakayama and Mackeben [3], who gave evidence of this dichotomy in the attention. The third approach is due to Klein [4], who presented evidence of a new component in the attention called *inhibition of return* which consists in a type of selective attenuation of regions on a saliency map, avoiding that the focus of attention is directed to regions already visited. This new component has certain similarity with the Helmoltz's approach from the point of view that when eyes are attracted to new regions is somewhat equivalent to be repelled from already explored regions. We have considered the above methods in our visual attention algorithm. The algorithm allows us to select regions of interest over which we forced the

fixations of a vision mechanism. We create saliency maps with characteristics that highlight within the scene. The amount of maps that can be extracted in an image is huge, so we just use some of them to avoid high latencies that can harm the performance of our system. We integrate visual characteristics such as color, geometry, optical flow and intensity. These characteristics are weighted depending on the task to solve and after integrating them to a saliency map we use the Winner Take All algorithm (WTA) to select the most salient region. The difficulty in the estimation of this kind of maps increases when the vision system is active —when we move a camera, a visual flow is generated in the whole image. For that reason, our motion maps will be estimated at the end of a saccadic movement, when the camera motion is practically none. We have evaluated the performance of our system on real environments.

REFERENCES

[1] von Helmholtz, H. V., Treatise on Physicological Optics, Vol. 3. Rochester, NY.: Optical Society of America, 1925.
[2] James, W. The Principles of Psychology. Vol. 1,2. Harvard University Press Cambridge, MA, 1981 (original 1890).
[3] Nakayama, K., Mackeben, M. Sustained and Transient Components of Visual Attention, en Journal of Experimental Psychology: Human Perception and Performance, Vol. 14, pp. 453-471, 1989.
[4] Klein, R. M. Inhibition of Return, en Trends Cognitive Science Vol. 4, pp. 138147, 2000.