Chapter 23
Artificial Visual Attention Using Combinatorial Pyramids

E. Antúnez
Universidad de Málaga, Spain

R. Marfil
Universidad de Málaga, Spain

Y. Haxhimusa
Vienna University of Technology, Austria

W. G. Kropatsch
Vienna University of Technology, Austria

A. Bandera
Universidad de Málaga, Spain

ABSTRACT

Computer vision systems have to deal with thousands, sometimes millions of pixel values from each frame, and the computational complexity of many problems related to the interpretation of image data is very high. The task becomes especially difficult if a system has to operate in real-time. Within the Combinatorial Pyramid framework, the proposed computational model of attention integrates bottom-up and top-down factors for attention. Neurophysiologic studies have shown that, in humans, these two factors are the main responsible ones to drive attention. Bottom-up factors emanate from the scene and focus attention on regions whose features are sufficiently discriminative with respect to the features of their surroundings. On the other hand, top-down factors are derived from cognitive issues, such as knowledge about the current task. Specifically, the authors only consider in this model the knowledge of a given target to drive attention to specific regions of the image. With respect to previous approaches, their model takes into consideration not only geometrical properties and appearance information, but also internal topological layout. Once the focus of attention has been fixed to a region of the scene, the model evaluates if the focus is correctly located over the desired target. This recognition algorithm considers topological features provided by the pre-attentive stage. Thus, attention and recognition are tied together, sharing the same image descriptors.

INTRODUCTION

Attention in humans defines the cognitive ability to select stimuli, responses, memories or thoughts that are behaviorally relevant among the many others that are irrelevant. Thus, attention has been often compared to a virtual spotlight through which our brain perceives the world. Based on concepts that emanate from the human perception system, computational attention models aim to develop this ability in artificial systems. Humans and animals are able to delineate, detect and recognize objects in complex scenes 'at a blink of an eye'. One of the most valuable and critical resources in human visual processing is time (Evolution conditioned the usage of this resource sparsely, because of survival necessity), therefore a highly parallel model is the biological answer dealing satisfactorily with this resource, since 'all complex behaviors are carried in less than 100 steps' (Feldman et al, 1982) (called the 100 step rule). That is, since neurons have a computational speed of a few milliseconds and each perceptual phenomenon occurs in a few hundreds of milliseconds yield that biologically motivated algorithms must be carried out in less than 100 steps. Tsotsos (1988, 1990, 1992) performed complexity analysis to show that hierarchical internal representation and hierarchical processing are the credible approach to deal with space and performance constrains, observed in human visual systems.

In the last years mobile robots have begun to address complex tasks that require them to obtain a detailed description of the environment. Human-robot interaction and object recognition are two examples of tasks that could be hardly achieved using range sensors and that usually need the use of vision. In these cases, the broad amount of information provided by vision systems makes its use more computationally expensive, a problem that can be solved by dealing only with a set of image entities (regions, points or edges). Following this feature-based strategy, it is now easier to find proposals that solve the simultaneous localiza-

tion and mapping problem or the human motion capture problem using vision, without employing external beacons or markers. If a mobile robot needs to solve several different tasks, we must consider that each task will need the detection of a specific set of features (local points of interest, human body parts...), so the perception system should be also changed according to the task. In this way, not only the generality of use is lost but also the robot will need to simultaneously manage different perception modules, as it will need to correctly attend to a very diverse set of situations. In biological vision systems, the attention mechanism is responsible for preselecting possible relevant information from the sensed field of view so that the complete scene can be analyzed using a sequence of rapid eye saccades. In recent years, efforts have been made to imitate such attention behavior in artificial vision systems, because it allows optimizing the computational resources as they can be focused on the processing of a set of selected regions only. Moreover, although these models can be influenced by the task to reach, they also include a bottom-up component to choose the more relevant item of the scene independently of the task. This allows to link perception and action, with perception influenced by the task to reach and the action by the perceived items.

The aim of this proposal is to present a new object-based framework of visual attention. With respect to previous approaches, our main contribution will be the representation of objects not only using appearance information, but also its internal topological configuration. This system will integrate bottom-up (data-driven) and top-down (model-driven) processing. The bottom-up component will determine salient 'pre-attentive objects' by integrating different features into the same hierarchical structure. Specifically, we propose to achieve this perception-based grouping process using a Combinatorial Pyramid (Brun and Kropatsch, 2001). Using this framework, the image topology will be preserved at upper levels; allowing correctly encoding relationships among

Related Content

An Image Forgery Detection Approach Based on Camera's Intrinsic Noise Properties
Shikha Gautamand Anand Singh Jalal (2018). *International Journal of Computer Vision and Image Processing (pp. 92-101).*
www.irma-international.org/article/an-image-forgery-detection-approach-based-on-cameras-intrinsic-noise-properties/201465

Measures of Image and Video Segmentation
Pushpajit A. Khaireand Roshan R. Kotkondawar (2017). *Applied Video Processing in Surveillance and Monitoring Systems (pp. 28-53).*
www.irma-international.org/chapter/measures-of-image-and-video-segmentation/168899

SVM-Based Switching Filter Hardware Design for Mixed Noise Reduction in Digital Images Using High-Level Synthesis Tools
Abdulhadi Mohammad din Dawraynand Muhammad Bilal (2022). *International Journal of Computer Vision and Image Processing (pp. 1-16).*
www.irma-international.org/article/svm-based-switching-filter-hardware-design-for-mixed-noise-reduction-in-digital-images-using-high-level-synthesis-tools/283966

Intelligent Control of Vehicles' Number Plates on Toll Gates in Developing Nations
Tawanda Mushiri, Charles Mbohwaand Simbarashe Sarupinda (2018). *Computer Vision: Concepts, Methodologies, Tools, and Applications (pp. 1023-1071).*
www.irma-international.org/chapter/intelligent-control-of-vehicles-number-plates-on-toll-gates-in-developing-nations/196992

An Iterative Method for 3D Body Registration Using a Single RGB-D Sensor
Victor Villena-Martinez, Andres Fuster-Guillo, Marcelo Saval-Calvoand Jorge Azorin-Lopez (2017). *International Journal of Computer Vision and Image Processing (pp. 26-39).*
www.irma-international.org/article/an-iterative-method-for-3d-body-registration-using-a-single-rgb-d-sensor/188759