

# *Eye-tracking Applications in Multimedia Systems*

## **Introduction to Special Issue**

### **Introduction**

Monitoring eye movements offers insights into visual perception, attention mechanisms and cognitive processes. Interpretation of eye movement data can be based on the empirically validated assumption that when a person is performing a cognitive task, whilst watching a display, the location of his/her gaze corresponds to the symbol currently being processed in working memory and, moreover, that the eye naturally focuses on areas that are most likely to be informative. As such, eye-tracking systems can be used as a data-gathering device or can provide the user with interactive functionality.

Eye tracking techniques vary in their intrusiveness. At one end of the spectrum lie intrusive techniques, which require some form of equipment to be put in physical contact with the user, in the form of electrodes, contact lenses or head mounted devices. Techniques which rely on infrared light generated by a sensor and reflected by the eye retain a certain degree of intrusiveness (as the head should be in static positions). Such systems are accurate, however, they require special and expensive hardware, and are not robust to sunlight and glasses worn by users, because they are dependent on the reflective properties of infrared light. Non intrusive techniques, on the other hand, are primarily vision based, i.e. they use cameras to capture images of the eye.

Irrespective of the particular technique being used, recent improvements in accuracy and availability of eye tracking technologies have facilitated the diverse and interactive use of eye tracking in multimedia systems. Whilst traditionally eye tracking has been used for obtaining information about human perceptive and cognitive processes, its integration into multimedia systems remains a challenging endeavour that offers considerable potential to optimize and incorporate the user experience in the multimedia systems of the future.

Four inter-related themes permeate the contents of this special issue: appearance based models, analysis and coding, gaze contingent displays, and interactivity.

### **Special issue outline**

The first papers selected for the special issue deal with a particular category of non-intrusive eye tracking techniques, namely those based on appearance based models, also known as computer vision models. In these models, one considers active illumination, relying exclusively on natural light. Such methods typically use cameras (e.g. webcams), and monitor eye gaze shifts by performing iris localization and tracking.

Accordingly, Colombo et al. present an approach in which eye tracking is based on passive computer vision. The tracking approach is aimed at obtaining accurate and robust measurements of the iris/pupil position. To this end, a method for ellipse fitting is used, employing search constraints. Tracking also embeds an iris localization algorithm (working as a bootstrap multiple-hypotheses generation step), and a blink detector, which

can also serve to detect voluntary eye blinks in human-computer interaction applications. On-screen remapping incorporates a head tracking method capable of compensating for small user head movements. The approach operates in real time, under different light conditions, and in the presence of distractors.

The second approach considered in the special issue, by Wang et al., detects and tracks eyes using geometric surface features on topographic manifold of eye images. In the joint spatial intensity domain, a facial image is treated as a 3D terrain surface or image topographic manifold. In a particular, eye regions exhibit certain intrinsic geometric traits on this topographic manifold, i.e., the pit-labelled center and hillside-like surround regions. Applying a terrain classification procedure on the topographic manifold of facial images, each location of the manifold can be labelled to generate a terrain map. The approaches have been experimented under various imaging and with different facial appearances using both static images and video sequences background constraint.

Another important issue of eye-tracking is coding. Till now, video coding standards (e.g. MPEG2, MPEG4, H.264) have been context free i.e. they don't exploit the context, including the content of videos, that can influence the quality of the coded video as perceived by a specific target audience. For example the same coding rules will be applied to both a news and a handball video sequence, in spite of their very different content characteristics and viewer requirements. Considering contextual information in the coding process is not a new problem. Object based coding of MPEG4 is an important tentative. This relies on accurate frame by frame segmentation and object specification, both very complex tasks which do not necessarily lead to improved perceived quality unless the video objects are linked to the viewer's visual response. Most relevant literature concentrates on video conferencing applications where the context is very well defined and easy to recognize. As eye-tracking systems evolve, they will contribute to the emergence of specialized content providers. So, the viewers become more selective and the audience of specific content becomes also more specific. In such cases it could be argued that ignoring the content characteristics and possible viewing patterns that could be associated with well defined contexts results in a suboptimal codec.

Accordingly, the next paper of the special issue, by Agraftotis et al., presents a framework for model based, context dependent video coding, based on exploitation of characteristics of the human visual system. The system utilizes variable quality coding, based on priority maps which are created using mostly context dependent rules. The technique is demonstrated through two case studies of specific video context, namely open signed content and football sequences. Eye tracking analysis is employed for identifying the characteristics of each context, which are subsequently exploited for coding purposes either directly or through a gaze prediction model. The framework is shown to achieve a considerable improvement in coding efficiency.

The identification of the components of eye movements (fixations and saccades) is an essential part in the analysis of visual behaviour, as these types of movements provide the basic elements used by further investigations of human vision.

A data mining method, presented by Urruty et al., examines eye fixations on static pictures and videos. The method analyses eye fixations and simplifies it by clustering eye positions using projections and projection aggregation. The method is robust to noise. Moreover, the method deals with videos, since it computes the dispersion in multi-user environment. One original aspect of the paper is that it integrates a data mining approach (projection clustering) and metrics (e.g. Gini) to simplify eye fixation representation.

Gaze contingent displays (GCDs) select the region of interest where the user is most likely to look, by actively tracking the viewer's eyes in real time and maintaining a high level of detail at the point of gaze. GCDs are especially useful in bandwidth-constrained environments or large-screen visual displays, and aim to minimise bandwidth use, whilst limiting any potential negative impact on user perception. Early GCDs suffered from limited display sizes, noticeable quality edges and limited control of resolution. However, increased screen and resolution sizes, as well as the falling cost of eye-tracking equipment, have all led to increased interest in attentive display research.

In our special issue, Duchowski and Coltekin present an elegant pixel shader algorithm for real-time GCD processing of both still and video content. Their paper also provides a comprehensive introduction to the area of GCDs, covering topics including medical applications of GCDs and perceptually lossless GCDs.

Loschky and Wolverton in their paper tackle the interesting issue of perceptual disruptions in GCDs – specifically examining perceptually acceptable update delays in multi-resolutional displays. They show that 60ms is perceptually acceptable, providing ample time for GCD update. Indeed, perceptual-based issues remain at the forefront of eye-tracking research and this paper marks a transition to those concentrating on interactivity issues in eye-tracking - the last theme of the special issue.

The use of eye-tracking in human-computer interaction is far from new. As such, eye-tracking is increasingly being employed in the design of user interfaces, where an efficient interface would ensure, for instance, that commonly-used controls are located in areas where gaze is most likely to rest, and that eye movement between these controls is minimal. Such eye-function-based interfaces would help users (especially disabled) to effectively execute specific interface actions, such as menu selection, eye-typing, and even mouse clicking, through the development of an 'eye-mouse'. Moreover, eye-tracking is increasingly being used for obtaining information about human perceptive and cognitive processes. For instance, knowing where a user's gaze is more likely to rest, when viewing a web page, has implications in usability. Therefore web design guidelines, based on results obtained using eye tracking technology, have been elaborated and are now being used by commercial web designers to develop more effective web pages. Eye tracking is also currently being used in virtual reality-based education and training, ranging from such diverse topics as aircraft inspection to driving safety.

The use of eye-tracking in virtual reality, and specifically in immersive environments, is the theme of the paper by Murray et al. By observing the interactions of users with avatars in such environments, as given by avatar gaze, they highlight the importance of

correct identification of gaze position in immersive scenarios, and that eye direction has to be combined with convergence in order for users to correctly identify the location of avatar gaze.

Inclusion of eye-tracking in interaction offers considerable benefits to the user; yet dependence on eye-tracking alone can lead to considerable problems. Rachovides et al. propose an alternative interaction method, the Conductor Interaction Method (CIM). CIM aims to provide a more natural and easily learned interaction technique. This novel interaction method employs a two-phased multi-modal interaction mechanism, using user gaze for selection and gesture for manipulation. When incorporated within a metaphor based environment, this provides a viable alternative for complete interaction with the computer.

### **Conclusion**

When selecting the contributions to be presented in this issue, we aimed at providing a good balance of research areas. The selected articles fall into broad categories that reflect the variety of research directions in eye-tracking and present some important contributions to the current state of art in the area. We believe that the special issue will certainly contribute to this exciting domain, by highlighting not only its challenges, but also some innovative approaches towards meeting them.

### **Acknowledgments**

We are grateful to the authors for their hard work during the special issue preparation. The special issue would not be possible without the great efforts of the authors and reviewers. Here is the list of the reviewers involved in the special issue reviewing process:

Gintautas Daunys (Siauliai University, Lithuania), Adel Lablack (LIFL, France), Paivi Majaranta (Tampere University, Finland), Sylvain Mongy (LIFL, France), Gabriel-Miro Muntean (Dublin City University, Ireland), John W Senders (University of Toronto, Canada), Laurent Sparrow (University of Lille3, France), Johnson P. Thomas (Oklahoma State University, USA).

Gheorghita Ghinea, Chabane Djeraba, Stephen Gulliver, Kara Pernice Coyne