

4th Global Summit and Expo on

Multimedia & Artificial Intelligence

July 19-21, 2018 | Rome, Italy

Modeling human visual attention in multimedia indexing and retrieval

Jenny Benois Pineau

LABRI (UMR 5800) - University of Bordeaux, France

Since the beginning of the era of multimedia indexing and retrieval, the research communities in computer vision and in multimedia worked in parallel to some extent on the same problem: How can we incorporate our knowledge on human understanding of visual scenes into pattern recognition, visual information analysis, indexing and retrieval. It is quite natural to seek for driving visual content mining methods towards regions and areas which attract human attention. The first classical model to be combined with prediction of saliency was the bag-of-visual-words (BoVW), used as a signature for images and video retrieval, object and concept recognition in visual content. From simple weighting of BoVW to a non-uniform sampling and psychovisual filtering of the content, the incorporation of models of human visual system in content description, retrieval and classification has allowed obtaining very competitive results without heavy window-based scanning in training of models and in generalization. The most popular visual saliency models proposed for this purpose still remained those by L Itti and J Harel, but for specific kind of content such as wearable multimedia, specific models of prediction of visual attention had to be designed. With the advent of powerful classification models such as deep neural networks (Deep NN) prediction of visual saliency is being passing from ad-hoc a priori mastering of local features to fit Treisman's and Gelade's feature integration theory for visual attention, to a massive supervised learning approach, requiring understanding of inherent bottlenecks of these tools in terms of availability- and noise in training data, of initialization, of optimization algorithms and knowledge transfer. In the talk we will give a vast panorama of the topic including examples in various application domains of multimedia indexing and retrieval. We will present applications of human attention use in human assistance for wearers of neuro-prostheses.

benois-p@labri.fr