UABDIVULGA
BARCELONA RECERCA I INNOVACIÓ

04/2013

What attracts our visual attention?

A doctoral thesis conducted at the Department of Computer Science of the UAB has studied which parts of an image attract our attention the most and which parts seem aesthetically pleasing to us and which not. The study suggests that regions that attract our attention can be predicted by the color and contrast they have, depending on whether these are stressed or suppressed by the human visual system.

Humans are constantly perceiving and reacting to their perceptions of the visual world. One region of a visual scene may attract focused attention while large regions are completely ignored. The scene may elicit pleasant emotions or feelings of revulsion. It may make a lasting and memorable impression on the observer or may never again be recalled. It seems reasonable to hypothesize that some of these reactions, for example the attention we give to a visual stimulus and the way the stimulus makes us feel, may share similar or even common perceptual mechanisms.

This dissertation presents our attempt to evaluate this hypothesis, by adapting a *state-of-the-art* model of human visual perception and applying its modified version to different visual tasks. Specifically, we investigate two different aspects of how an observer experiences a natural image: (i) where we look, that is, where attention is guided; (ii) what we like, that is, whether or

not the image is aesthetically pleasing.

These two experiences are the subjects of increasing research efforts in computer vision. The ability to predict visual attention has wide applications, from object recognition to marketing. Aesthetic quality prediction is becoming increasingly important for organizing and navigating the ever-expanding volume of visual content available online and elsewhere.

Both visual attention and visual aesthetics can be modeled as a consequence of multiple interacting mechanisms, some driven by information reaching our eyes from the world (*bottom-up*), and others driven by our internal state of mind (*top-down*). In this work we focus on a *bottom-up* perspective as it is here that the links between aesthetics and attention may be more obvious and/or easily studied.

We first investigate bottom-up visual attention, which is often called saliency. We hypothesize that salient and non-salient image regions can be estimated to be the regions where color contrast is enhanced or suppressed by the human visual system. We prove this hypothesis by adapting a low-level model of color perception into a saliency estimation model. The proposed model outperforms the state-of-the-art at the task of predicting which image locations attract attention.

Next we investigate the problem of aesthetic visual analysis. We entertain the hypothesis that low-level visual information in our saliency model can also be used to predict visual aesthetics by capturing local image characteristics such as feature contrast, grouping and isolation, characteristics thought to be related to universal aesthetic laws. We demonstrate that visual features extracted from our saliency model achieve *state-of-the-art* performance on aesthetic quality classification.

As such, a promising contribution of this thesis is to show that several vision experiences -low-level color perception, visual saliency and visual aesthetics estimation- may be successfully modeled using a unified framework. This suggests a similar architecture in the low-level human visual system for both color perception and saliency and adds evidence to the hypothesis that visual aesthetics appreciation is driven in part by *bottom-up* mechanisms.

Naila Murray
naila.murray@gmail.com

References

"Predicting Saliency and Aesthetics in Images: A Bottom-up Perspective", Naila Murray's doctoral thesis supervised by Xavier Otazu Porter and Maria Vanrell Martorell.

View low-bandwidth version