Region segregation and saliency using colour information

J.A. Martins, J. Rodrigues and J.M.H. du Buf
Vision Laboratory, Inst. for Systems and Robotics (ISR) – University of the Algarve, Faro, Portugal
{jamartins, jrodrig, dubuf}@ualg.pt

Abstract

Saliency maps determine the likelihood that we focus on interesting areas of scenes or images. These maps can be built using several low-level image features, one of which having a particular relevance: colour. In this paper we present a new computational model, based only on colour features, which provides a sound basis for saliency maps for static images and video, plus region segregation and cues for local gist vision.

1. Introduction

Visual attention is one of the most important mechanisms in the perception of our environment. It guides our eye movements for scrutinising conspicuous areas. Our goal is to develop an artificial vision system which mimics the attention process of our visual system, for which we need to construct so-called *saliency maps*. Such maps can be built from many features, including keypoints, edges, colour, texture, orientation and motion, disparity, etc. [2]. Here we focus on creating saliency maps employing only colour features, as these provide the most important cues for attention [7], but in combination with low-level geometry features. We show that these maps can also be used for tasks such as region segregation [4] and motion tracking [5], and they provide cues to bootstrap a cortical architecture for invariant object recognition, as proposed by Rodrigues and du Buf (see [6]).

2 Methods

Entire processing of a scene or image, from original input image to saliency map, consists of the following steps: (a) colour normalisation, (b) adaptive smoothing, (c) contour detection, (d) colour conspicuity computation at contour points, and (e) low-level geometry processing.

The first step (a) consists of colour illuminant and geometry normalisation. We process the input image $I_{in}(x,y)$ using the method described by Finlayson et al. [3] in RGB colour space. Their method applies iteratively two steps: illuminant geometry independency (i.e., *chromaticity*), and global illuminant colour independency (i.e., *grey-world normalisation*), until colour convergence is achieved, usually after 4–5 iterations.

The colour-corrected RGB image is then converted to Lab colour space, as this provides an almost linear space which is more suitable for determining the conspicuity of borders between regions (see below). The new space consists of the a_{cc} and b_{cc} components (cc stands for colour-corrected), but combined with the *unmodified* L_{in} component from I_{in}, i.e., $I_{cc} = (L_{in}, a_{cc}, b_{cc})$. The reason for using L_{in} is that colour correction eliminates all spatial information in gray (uncoloured) image regions.

The top row in Figure 1 shows an original image (left) and two modified images, one with a blue tint (R –12%, G +4% and B +50%) (centre) and one with a warm white balance (right). As can be seen in the second row, colour correction yields very similar images despite the rather large differences in the input images.

The second step (b), adaptive filtering, aims at smoothing inhomogeneities in regions while maintaining or even improving the boundaries between regions. We apply the Context-Sensitive Weighting-Factor (CSWF) filter (see [1]) 5 times to each of the components of I_{cc}, which results in $I_{ci}(x,y) = \mathrm{CSWF}[I_{cc}(x,y)]$ where subscript ci stands for colour-improved. A result (surfer image) is shown in Fig. 1 (3rd row at left).

The third step (c) serves to detect boundaries. We apply a simple edge detector with thresholding to each colour component, after which the three results are OR-ed together in $I_{ed}(x,y)$ (3rd row, centre).

In the next step (d), colour conspicuity is calculated at each position where $I_{ed}(x,y) \neq 0$. We define conspicuity $\Psi(x,y)$ as the maximum difference between the colours in I_{ci} at four pairs of symmetric points at distance $d = 4$ from (x,y), i.e., on horizontal, vertical and two diagonal lines ($i = \{1,2,3,4\}$). Partial conspicuity is calculated independently for each of the colour components. If $\vec{x}_i$ and $\vec{x}_{i+4}$ are two symmetric points on line i relative to (x,y), total conspicuity is the sum of the three partial components:
$\Psi(x,y) = \sum_{L,a,b} \left(\max_i \left(|I_{ci}^{L,a,b}(\vec{x}_i) - I_{ci}^{L,a,b}(\vec{x}_{i+4})| \right) \right).$

Conspicuity $\Psi(x,y)$ is shown in Fig. 1 (3rd row at right).

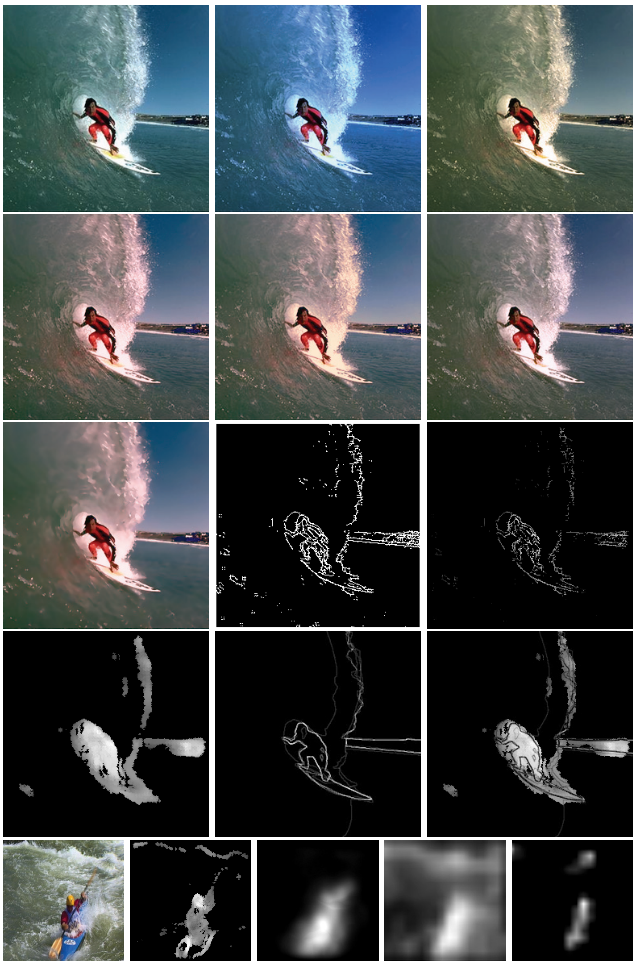

Figure 1. Results (see text).

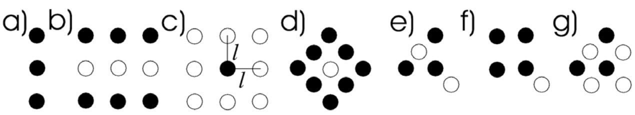

Figure 2. Geometry groupings (see text).

Most conspicuous are the surfer, the vertical wavefront and the coastline.

In the final step (e), the saliency map is built on top of the conspicuity map by using grouping cells which code local geometry. This process also serves to create regions *inside* boundaries; see Fig. 1 (4th row at left). Basically, grouping cells combine summation cells with a dendritic field of size $n \times n$ ($n = 5$), with their centres at a distance l ($l = 5$), and each grouping cell is devoted to a particular spatial configuration. Figures 2a to g show examples of configurations, respectively, a line (or an isolated contour), a bar (two parallel contours of a bar), two types of blobs and three types of corners. Configurations c and d are not rotated, but the five others are (horizontal, vertical and two diagonal orientations), so in total there are 22 configurations. Solid dots represent summation cells which must be excitated (at least one response must exist in their dendritic field) and open dots are cells which must not be excitated (zero response). The saliency map consists of the response of the grouping cell (of all 22) with maximum activity at each position (x, y).

3 Brief discussion

We presented a method to obtain a colour-based saliency map for attention, but it can also be used in tasks like region segregation and motion tracking. Figure 1 (4th row at left) shows our result which can be compared against ground-truth (centre), generated from 30 human observers [4] (see also http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/). The superposition of the two images, shown at right, indicates that the salient areas correspond to the ones chosen by observers in image segmentation and boundary detection tasks. The bottom row of Fig. 1 shows, from left, one frame of a kayak video from [5], our own saliency map, the saliency map from the authors, plus two maps using different algorithms from Itti et al. presented in [5]. Our result is much more precise and highlights regions with conspicuous colours. Finally, low-level geometry grouping maps in combination with the saliency map will provide the necessary cues for fast local gist (object) vision, necessary for bootstrapping a cortical architecture for invariant object recognition, as proposed by Rodrigues and du Buf [6].

Acknowledgements: This research is partly supported by the FCT (ISR/IST pluri-annual funding) through the POS_Conhecimento Program, which includes FEDER funds, and by the FCT project PTDC/EIA/73633/2006 – SmartVision: active vision for the blind.

References

[1] J. M. H. du Buf and T. G. Campbell. A quantitative comparison of edge-preserving smoothing techniques. *Signal Processing*, 21:289–301, 1990.

[2] L. Elazary and L. Itti. Interesting objects are visually salient. *Journal of Vision*, 8(3):3.1–15, 2008.

[3] G. Finlayson, B. Schiele, and J. Crowley. Comprehensive colour image normalization. *Proc. 5th Europ. Conf. Comp. Vision*, I:475–490, 1998.

[4] C. C. Fowlkes, D. R. Martin, and J. Malik. Local figure–ground cues are valid for natural images. *Journal of Vision*, 7(8):2:1–9, 2007.

[5] O. Le Meur, P. Le Callet, and D. Barba. Predicting visual fixations on video based on low-level visual features. *Vision Research*, 47:2483–2498, 2007.

[6] J. Rodrigues and J. du Buf. Invariant multi-scale object categorisation and recognition. *Proc. 3nd Iberian Conf. on Patt. Recogn. and Image Anal.*, LNCS 4477:459–466, 2007.

[7] J. van de Weijer, T. Gevers, and A. D. Bagdanov. Boosting color saliency in image feature detection. *IEEE Tr. PAMI*, 28(1):150–156c, 2006.