Visual saliency detection in colour images based on density estimation

Désiré Sidibé, F Mériaudeau

Visual saliency detection in colour images based on density estimation

D. Sidibé and F. Mériaudeau

A simple and effective method for visual saliency detection in colour images is presented. The method is based on the common observation that local salient regions exhibit distinct geometric and and texture patterns from neighbouring regions. We model the colour distribution of local image patches with a Gaussian density and measure the saliency of each patch as the statistical distance from that density. Experimental results with public datasets and comparison with other state-of-the-art methods show the effectiveness of our method.

Introduction: Visual saliency is an important research topic in the field of computer vision due to its numerous applications such as object detection and image segmentation [1], robots navigation and localization [2], video surveillance [3], object tracking [4] and image/video compression [5]. Visual saliency helps to focus on regions of interest instead of processing the whole image or video data, and it is a process which has been widely addressed in literature with several formulations. The main idea used in most of the methods is the so-called *pop-out* effect, i.e. an image region is considered as salient if it differs from its surrounding regions. Different image features such as colour or texture, can be used to compute a region's distinctness. Different methods proposed in literature exploit different features and adopt different functions to compute the unique characteristics of a region. A recent survey of state-of-the-art methods for visual saliency detection can be found in [6].

A common approach is to extract local image patches from the image and compute a dissimilarity measure between each local pacth and its k-nearest neighbours. Using only k neighbours reduces the computation time, but ignores the global distribution of the patches intensities within the image.

In this Letter, we propose a simple and effective saliency detection method based on modelling the statistics of local image patches with a Gaussian distribution. The saliency of a local patch is then computed as the distance from the distribution. Experimental results with publicly available datasets show the good performance of our approach in comparison with other methods in literature.

Local patches intensity distribution: The key idea of our method is to compare an image patch not only to its k-nearest neighbours, as done is many previous approaches, but to all other images patches. However, to avoid the computation burden of a comparison with all image patches, we first model the intensity distribution of local image patches with a Gaussian density. The density's parameters, mean and covariance, capture the internal statistics of the patches distribution within the image.

More specifically, given an image I, we extract all local patches $p_\mathbf{x}$ centred at pixel $\mathbf{x} = (x, y)$, in the image. The local patch $p_\mathbf{x}$ is represented as a vector of size $d = W \times H$, where $W \times H$ is the spatial size of the patch. For a colour image, with 3 channels, the vectors obtained from each channel are just concatenated. Thus, each local patch is represented as a point in a high dimensional vector space: $p_\mathbf{x} \in \mathbb{R}^d$, where $d = 3 \times W \times H$.

By putting together all local patches from I, we form a matrix $\mathbf{X}_I = [p_1, \ldots, p_M]$, M being the number of patches. We apply principal components analysis (PCA) to reduce the dimensionality of the data. In particular, we keep the first k dominant principal components corresponding to the k largest eigenvalues. We set the value of k such that at least 95% of the total variance of the data is retained.

Next, in the reduced k-dimensional space, we represent the distribution of local image patches with a multivariate Gaussian density described by its mean $\mu \in \mathbb{R}^k$ and covariance matrix $\Sigma \in \mathbb{R}^{k \times k}$. Note that it is also possible to use a Gaussian mixture model (GMM) rather than a single Gaussian. However, this requires more time to estimate the parameters, while not improving the results as we have observed in our experiments.

Saliency detection: The saliency value of each local patch $p_\mathbf{x}$ is finally given by the statistical distance of the patch from the Gaussian density, measured by the Mahalanobis distance:

$$S(p_\mathbf{x}) = (p_\mathbf{x} - \mu)^T \Sigma^{-1} (p_\mathbf{x} - \mu). \tag{1}$$

Using equation (1), we obtain a saliency value for each local patch and, therefore, a saliency map for a given image. In order to detect saliency regions of different sizes, we compute the saliency map for three different resolutions of the input image I: 100%, 50% and 25%, and the results are averaged.
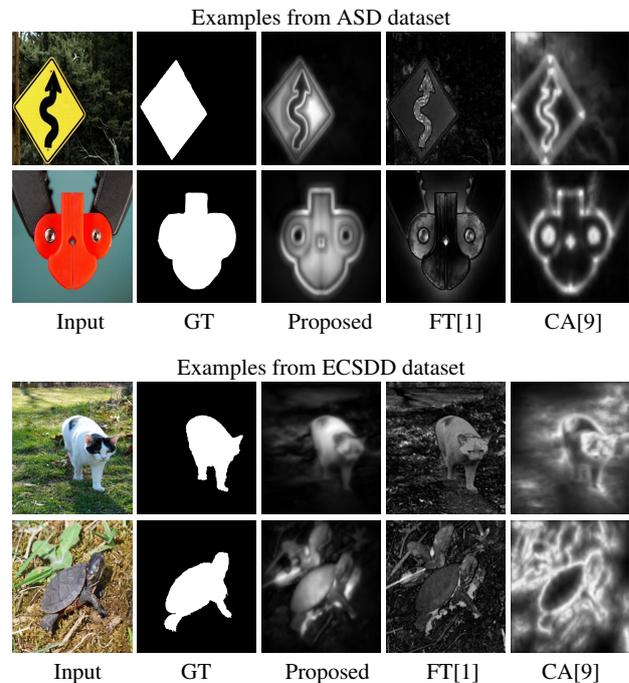
This saliency map is further refined using a centre prior, i.e. we filter the obtained saliency map with a Gaussian filter placed at the centre of the image and with high standard deviation. This takes into account the fact that people tend to place objects of interest near the centre of the image when taking a picture [7].

Results: We evaluate the performance of the proposed method using two public datasets which both contain manually segmented ground truth regions. The first dataset, ASD [1], contains 1000 images with simple and smooth background structures, while the second one, ACSSD [8], contains 1000 images with more complex background structures and more than a single object in an image. Thus, detecting salient objects with ECSSD is more challenging than with ASD.

We also compare our method with one global method, the context-aware approach [9], and one local method, the frequency based approach [1]. These two methods have shown good results, especially on the ASD dataset. In particular, the local method (FT) was shown to perform better than five other saliency detection methods [1], and the global method (CA) was shown to outperform three state-of-the-art methods [9].

Some examples of results obtained with both datasets are shown in Fig. 1. As can be seen, the proposed method provides better saliency maps for both datasets. In particular, it captures both the outline and the internal pixels of salient objects, while the context-aware method (CA) detects the borders of the objects but not their interior. The proposed method is also less affected by background structures than other two methods. This is particularly true for images from the ECSSD dataset which contain complex backgrounds.

The main parameter of the proposed method is the spatial size of the local patches. In our experiments, we use square patches of size $W \times W$ and we have tried several values in the range $\{5, 7, 9, 11, 13, 15\}$. We observe that $W = 9$ provides the best results. For the other two methods, we used codes provided by authors with the default parameters.

Examples from ASD dataset

| Input | GT | Proposed | FT[1] | CA[9] |

Examples from ECSDD dataset

| Input | GT | Proposed | FT[1] | CA[9] |

Fig. 1 *Qualitative results*

Our proposed method provides better saliency map than the local method (FT) and the global method (CA). It captures both the outline and the internal pixels of salient objects

Table 1: Comparison of saliency detection methods: AUC values

Dataset	Proposed	Achanta [1]	Goferman [9]
ASD	0.9030	0.8617	0.8728
ECSSD	0.8070	0.650	0.7820

For a quantitative evaluation, we use Receiver Operating Charecteristics (ROC) curves and compute the Area Under Curve (AUC) which measures the similarity of the detected saliency map with the ground truth. A saliency map obtained from a method is first normalized to the range $[0, 1]$ and is binarized using varying thresholds $t \in [0, 1]$. Then, we compute the True Positive Rate (TPR) and False Positive Rate (FPR) using these binarized maps, and generate a ROC curve.

The quantitative results summarized in Table 1 show that the proposed density based approach performs better than other two methods. As expected, all accuracies are lower with the ECSSD dataset, which attests the difficulty of this datasets. However, our method significantly outperforms the local method [1] on this dataset. The accuracy is slightly better than the global method [9], but the proposed method is much faster since it is based on a simple distance calculation (equation (1)). On the ASD dataset, the average processing time per image is 0.7 second with the proposed method, and 27.4 seconds with the context aware method [9]. Note that these processing times are obtained using a Linux computer with 2.8GHz CPU and 8GB RAM.

Conclusion: We have proposed a simple and effective method for visual saliency detection in colour images. The method is based on representing the colour distribution of local image patches with a Gaussian density and measuring the saliency of each patch as the statistical distance from that density. Thus, we do not compare each patch only to its k-nearest neighbours, but to all other patches in the image in a very efficient way. Experimental results with two public datasets show the effectiveness of our method in comparison with two state-of-the-art methods.

We have used RGB colour colour space in this work, but the method can easily be applied with other colour or features representations.

D. Sidibé (*Université de Bourgogne-Franche Comté - Le2i UMR CNRS 6306, France*)

F. Mériaudeau (*CISIR, Univeristi Teknologi Petronas, Malaysia*)

E-mail: dro-desire.sidibe@u-bourgogne.fr

References

1 Achanta, R., Estrada, F., Susstrunk, S. and Hemami, S.: 'Frequency-tuned salient region detection', Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2009, pp. 1597-1604

2 Siagian, C. and Itti, I.: 'Biologically inspired mobile robot vision localization', *IEEE Trans. Robot.*, 2009, **25**, (4), pp. 861-873

3 Yubing, T., Cheikh, F. A., Guraya, F. F. E., Konik, H. and Trémeau, A.: 'A spatiotemporal saliency model for video surveillance', *Cognitive Computation*, 2011, **3**, (1), pp. 241-263.

4 Sidibé, D., Fofi, D. and Mériaudeau, F.: 'Using visual saliency for object tracking with particle filters', Proc. Euro. Conf. on Signal Processing, 2010, pp. 1776-1780

5 Guo, C. L. and Zhang, L. M.: 'A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression', *IEEE Trans. Image Process.*, 2010, **19**, (1), pp. 185 -198.

6 Borji, A. and Itti, L.: 'State-of-the-art in visual attention modeling', *IEEE Trans. on Pattern Anal. Mach. Intell.*, 2013, **35**, (1), pp. 185-207.

7 Judd, T., Ehinger, K., Durand, F. and Torralba, A.: 'Learning to predict where humans look', Proc. IEEE Int. Conf. on Computer Vision, 2009, pp. 2106-2113

8 Shi, J., Yan, Q., Xu, L. and Jia, J.: 'Hierarchical image saliency detection on extended CSSD', *IEEE Trans. on Pattern Anal. Mach. Intell.*, 2016, **38**, (4), pp. 717-729.

9 Goferman, S., Zelnik-Manor, L. and Tal, A.: 'Context-aware saliency detection', Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2010, pp. 459-468