A RARITY-BASED VISUAL ATTENTION MAP - APPLICATION TO TEXTURE DESCRIPTION -

Matei Mancas[1], Céline Mancas-Thillou[1], Bernard Gosselin[1], Benoît Macq[2]

[1]Faculté Polytechnique de Mons, TCTS Lab, Belgium, e-mail : matei.mancas@fpms.ac.be
[2]Université Catholique de Louvain, TELE Lab, Belgium

ABSTRACT

This paper describes a simple and "pre-cortical" visual attention model, which does not take image directions into account. We compute rarity-based saliency maps and then we describe the relation between texture and visual attention. Finally we decompose the image into several textures with different regularities. Our purpose is to compress textures into images using small repeating patterns.

Index Terms— *Visual attention, texture, saliency*

1. INTRODUCTION

The human visual system (HVS) is a topic of increasing importance in computer vision research since Hubel's work [1] and the comprehension of the basics of biological vision. Mimicing some of the processes done by our visual system can help to improve the existing computer vision systems. Visual attention takes part to one of the most important tasks of the HVS, which is to extract salient features from the surrounding images in order to react in a relevant manner for our survival.

Grossberg's theory [2] is based on the comparison of current visual features with some previously learnt features. The novelty in a scene depends on how different the scene is from previously observed scenes.

Osberger and Maeder [3] first segment the images into homogeneous regions and then classify them according to some features. The first segmentation step is here extremely important and errors in this step could seriously affect the final result.

Itti and Koch [4] define a multi-resolution and multi-features based system which models the visual search in primates. Supervised learning is suggested to tune the feature weights for dedicated applications. A problem is that only local processing done by different kinds of cells is taken into account.

Walker at al. [5], Mudge at al. [6], Stentiford [7] and Boiman and Irani [8] base their saliency maps on the idea that important areas are unusual in the image. Even Itti [9] recently published a paper with a probabilistic approach which is very different from his initial ideas [4].

These techniques use comparisons between neighbourhoods of different shapes and at different scales to assign an attention score to a region. They provide satisfying results but they are sometimes too far from the biological system.

We propose a simple saliency measure based on the information theory which is log-inversely related to the pixel occurrence frequency.

We first define our visual attention model for both gray-level and color images. Then we apply our attention map to the textural description.

2. VISUAL ATTENTION (VA)

We deal here with the low-level visual attention which is pre-attentive (no eye movements are needed). Our definition will be based on the **rarity** concept. The eye is not attracted by particular features in an image but it is attracted by the features which are in minority in an image as it can be notticed by observing the following facts :

- Our vision can be attracted by homogeneous areas into a heterogeneous scene, but also by heterogeneous areas into a homogenous scene.
- Bright areas into a dark scene will be attractive for our eyes like dark areas into a bright scene.

Heterogeneous or homogeneous, dark or bright, symmetric or asymmetric, moving or static objects can all attract our visual attention. We also can note that the pairs of features we mentioned are opposite features describing the order and the disorder at several scales, in space and time. The HVS describes rare, so anomalous things in a scene as "**interesting**".

2.1. Rarity quantification

A pre-attentive analysis is achieved by humans in less than 200 milliseconds. How to modelize rarity in a simple and fast manner ?

The most basic operation is to count similar areas in the image. The histogram is an adequate statistical tool which counts equivalent pixels. Within the context of information theory, this approach based on the histogram is close to the so-called self-information. Let us note m a message containing an amount of information. A message self-information I is defined as:

$$I(m) = - log (p (m))$$

where $p(m)=Pr(M=m)$ is the probability that a message m is chosen from all possible choices in the message space M. We obtain an attention map by replacing each message m by its correspondig self-information $I(m)$.

The self-information is also known to describe the amount of surprise of a message inside its message space: rare messages are surprising, hence they attract our attention.

Nevertheless, comparing only isolated pixels is not efficient. In order to introduce a spatial relationship, areas surrounding each pixel should be considered. The mean and the variance can be used to describe the statistics of a pixel neighbourhood.

2.2. Adding spatial information

In order to take into account the spatial information, we compute the local mean and variance on a sliding window. We used a 3x3 window size as our experience showed that this parameter is not of primary importance.

Thus, we obtain two saliency maps: one for the image local mean and one for the image local variance that we average to obtain a unique saliency map.

Figure 1 shows the initial image on the left and its attention map on the right. Contours and statistically smaller areas get higher attention scores.

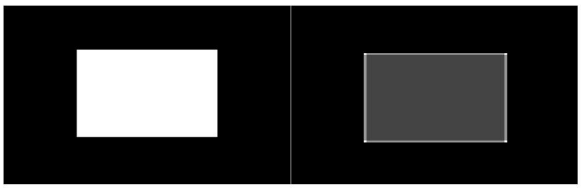


Fig. 1. *Initial image on the left, rarity map on the right (higher gray level = higher attention score)*

The window size is not very important but there is a second parameter which has more effect on the final result: the image quantification. After computing the mean and the variance on the initial image we can keep the original quantification or reduce it to 32 or 16 gray levels for exemple. The quantification value act like a sensitivity parameter in our method. If the quantification value is high (256 for exemple) pixels with little value difference will remain different. For lower quantification values (32 for exemple) pixels with little value difference are considered to have the same value. This fact obviousely has an impact on the pixel rarity computation. For natural images we use lower quantification values as 16 or 32 in order to group the "quite similar" pixels and to speed up the image reconstruction computation.

Figure 2 shows the attention map computation on the "cameraman" image. The mean and variance attention maps are on top and their final map is on bottom-right. We used a quantification value of 32.

Fig. 2. Top: mean (left) and variance (right) attention maps. Bottom: original image (left) and final attention map (right). Quantification=32.

2.3. From gray-scale to color images

The retina cells and then the lateral geniculate nucleus (LGN) have three kinds of ganglion cells. The so-called mango cells (M) which deal mostly with the luminance, the parvo cells (P) with a red-green opposition and the K cells with a blue-yellow opposition.

This color system is known as the opponent color system and it is defined from the retina receptors: O_1 = Y (luminance), O_2 = Y – B (blue), O_3 = R(red) – G(green).

Following that, we apply a visual attention map computation at each of these three color channels.

Finally, in order to fuse visual attention data coming from each color channel, we use the maximum operator: if a feature is likely to attract human eyes in at least one of the three channels, it will attract the eyes in the final visual attention map.

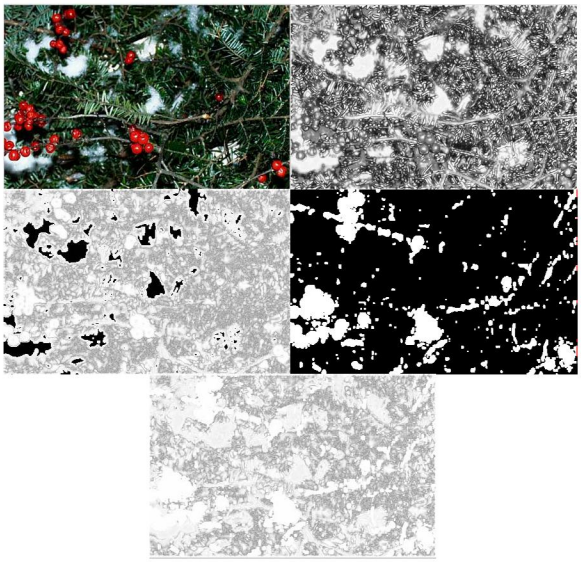


Fig 3. Top: initial image (left), luminance VA map (right). Middle: Red/Green component VA map (left), Blue/Yellow VA map (right), Bottom: final VA map

Figure 3 shows on top-left the initial color image. We can see that the luminance VA map is quite high for the snowy white parts which are in minority, but the red fruits are not very well highlighted because from a luminance point of view, they are not so different from the background.

On the middle line of Figure 3, we can see that the red fruits are very well highlighted because from a color opposition point of view, the fruits are very rare so very discriminant for the human eye. On Figure 3 (bottom image) the rare structures are well highlighted and they correspond quite well with the areas we believe we look at first.

3. TEXTURE CHARACTERISATION

3.1. From visual attention to texture

There is no precise definition of texture but the most commonly accepted one is that texture is a combination of repeated patterns with a regular frequency. This definition is only used for artificial or human made textures, but we also use the word "texture" when we talk about grass, wood or other natural images also made from patterns but which do not repeat exactly in the same way.

As we cannot define precisely a texture, we will assume a "degree" of texture, of pattern repetition. Some regions in an image could be more regular, so closer to the academic definition of repeating patterns, but others could be more complex but we still perceive them as "textures".

Let us now introduce the relation between texture and visual attention. As showed in section 2.1, our VA map computation is done by counting similar pixels. Rare pixels will get a very high attention score, but repeating pixels will get lower scores. So, more a texture will be regular, more it will repeat itself and less salient it will be.

A homogenous area could be considered as the most regular texture. In our case, it will have an attention score very close to 0. For more irregular textures, we will obtain higher attention scores. In this way, we can directly link our visual attention map to the texture regularity: more a region is salient, less regular is its texture.

3.2. Texture information from the VA map

We characterize in this subsection textural regions from perfectly regular texture to highly irregular texture. We first find an automatic VA map threshold by using the Otsu [10] method which chooses the threshold to minimize the intraclass variance of the thresholded black and white pixels. This automatic threshold will roughly divide the image into a foreground (more salient than the threshold) and a background part (less salient than the threshold).

Then, we arbitrary divide the foreground and the background into five equally important parts. We obtain in this way ten levels of visual attention (five of them mainly dealing with the foreground and other five with the background).

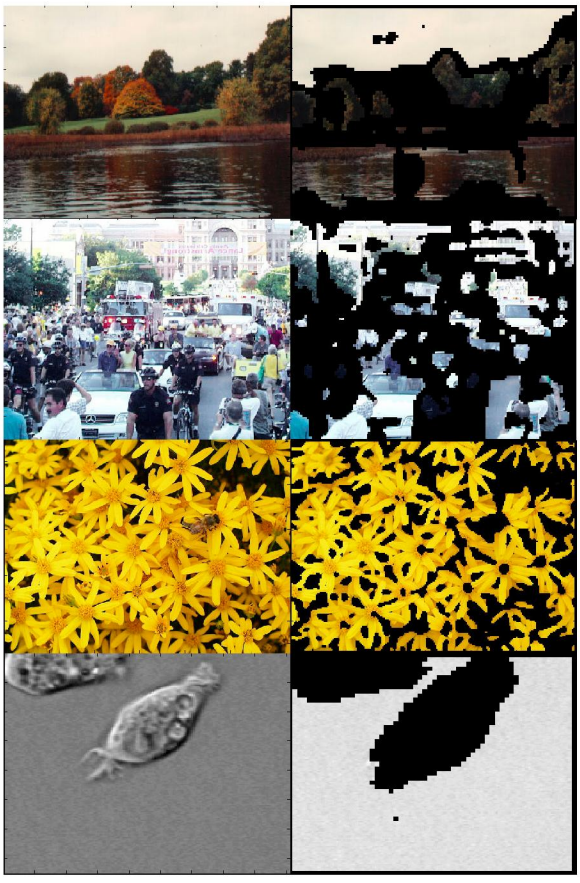
Fig. 4. Top: initial image (left), using 9 VA levels (right), Bottom: texture using 7 VA levels (left), and using 5 VA levels (right)

There are two parameters which are important for texture: the VA level (from 1 to 10) and the spatial density of these levels. A high density means a texture with a simple pattern as all the gray-levels of the pattern are quite close. If the density is not very high and the regions are spread over several VA levels, it means that the texture pattern is quite complex and it is composed from very different gray levels.

Figure 4 shows well how taking fewer VA levels induces more and more regular texture. Figure 5 also shows some results on different color or gray-level natural or medical images.

4. CONCLUSION AND FUTURE WORK

We presented in this article a simple and efficient pre-attentive rarity-based visual attention map. This map is then used in texture characterization and we show that there is a direct relation between visual attention and the texture regularity. In this way, we can select in an image several textures from regular to very irregular. A very challenging application of this technique is in efficient image coding: once we detect a texture, we can only code and transmit its basic pattern.



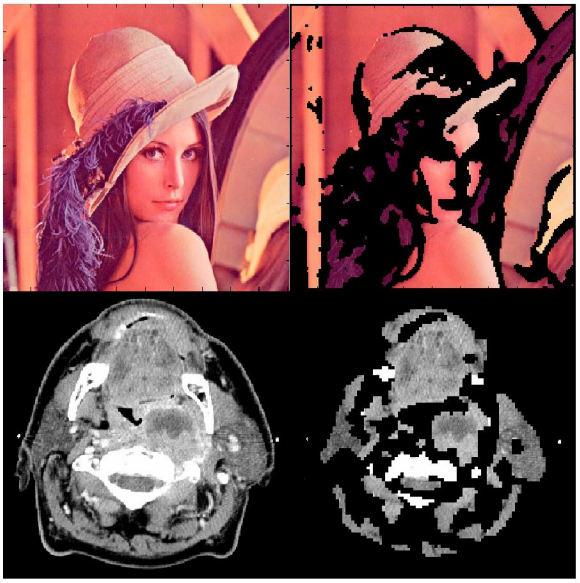


Fig. 5. First column: original images, second column texture detection after, from top to bottom, 8 VA levels, 6 VA levels, 8 VA levels, 7 VA levels, 8 VA levels and 7 VA levels

5. REFERENCES

[1] D. H. Hubel, "Eye, Brain and Vision", New York: Scientific American Library, 1989

[2] S. Grossberg, "The link between brain, learning, attention, and consciousness", Consciousness & Cognition, 8, 1-44, 1999

[3] W. Osberger and A.J. Maeder, "Automatic identification of perceptually important regions in an image", 14th IEEE Conf. on Pattern Recognition, 1998

[4] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention", Vision Research, 2000

[5] K.N. Walker, T.F. Cootes and C. J. Taylor, "Locating salient object features", Proceedings of BMVC, 1998

[6] T.N. Mudge, J.L. Turney and Volz, "Automatic generation of salient features for the recognition of partially occluded parts", Robotica, 1987

[7] F.W.M. Stentiford, "An estimator for visual attention through competitive novelty with application to image compression", Picture Coding Symposium, Seoul, 2001

[8] O. Boiman and M. Irani, "Detecting irregularities in images and in video", Proceedings of ICCV, 2005

[9] L. Itti, P. Baldi, "A Principled Approach to Detecting Surprising Events in Video", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 631-637, 2005.

[10] N. Otsu, "A threshold selection method from gray-level histograms", IEEE Trans. Syst. Man Cybernet., 9(1):62–66, 1979