BULETINUL INSTITUTULUI POLITEHNIC DIN IAŞI
Publicat de
Universitatea Tehnică „Gheorghe Asachi” din Iaşi
Tomul LIX (LXIII), Fasc. 1, 2013
Secţia
ELECTROTEHNICĂ. ENERGETICĂ. ELECTRONICĂ

EXPLOITING QUATERNION RARE ORIENTATIONS FOR IMAGE SALIENCY DETECTION

BY

OANA-LOREDANA BUZATU*

"Gheorghe Asachi" Technical University of Iaşi
Faculty of Electronics, Telecommunications and Information Technology

Received: March 4, 2013
Accepted for publication: March 26, 2012

Abstract. Saliency detection has become lately a valuable tool for many image processing applications. Yet, methodology variations of existing algorithms still permit performances improvement. In this paper a state-of-the-art method, namely phase spectrum of quaternion Fourier transform (PQFT), is extended to a multi scale approach, by using steerable pyramids based decomposition. Considering different colour spaces, images are decomposed both in scale and four specific orientations, which are further captured in a quaternion vector representation. Based on each pyramid scale entropy, every level is weighted in a linear combination to form a saliency map. Qualitative and visual results proved that the proposed method has comparable performances with other state-of-the-art methods, and also can improve an image abstraction application.

Key words: visual attention; saliency; Steerable Pyramids; quaternion.

1. Introduction

The interest on selective visual attention mechanisms employed by the human brain in the visual tasks has gained lately increasing consideration in the computer vision applications. Thanks to these intelligent mechanisms, human brain can interact with and process almost instantly a large amount of data,

* *e-mail*: lbuzatu@etti.tuiasi.ro

quickly focusing on the most salient regions of the visual field. These regions often include significant or relevant semantic content for the current visual task. Such semantic content is trivially important for many image processing applications such as segmentation (Goferman *et al.*, 2012; Liu *et al.*, 2007, 2011), image editing (Fang *et al.*, 2012; Holtzman-Gazit *et al.*, 2010; Hou & Zhang, 2007b; Marchesotti *et al.*, 2009; Santella *et al.*, 2006), compression (Guo *et al.*, 2008; Guo & Zhang, 2010; Itti, 2004), object detection and recognition (Goferman *et al.*, 2012; Liu *et al.*, 2011; Wang & Li, 2008).

Current dedicated literature consists on a tremendous number of studies on human visual system understanding and simulating, which led to several computational models for saliency detection. Most of the models are built around the Feature Integration Theory (FIT) elaborated by Treisman & Gelade, (1980), which suggests that the visual information is analysed on different parallel channels. For every early visual feature, such as colour, luminance, motion, orientation, size, a topographical map is constructed, and then all these *conspicuity maps* are fused in a linear or nonlinear fashion into a final master map, usually called *saliency map*. Together with the Guided Search Model proposed by Wolfe *et al.*, (1989), and the first algorithmic model of attention conceived by Koch & Ullman, (1985), these theoretical studies provided the basic structure for the future computational models.

Most of the models follow the *bottom-up* attention mechanisms also called *automatic*, *reflexive* or *stimulus-driven* which are responsible for the automatic focus on the regions which are sufficient different with respect to their surroundings. But there are models which include also *top-down* attention mechanisms. These mechanisms are driven by cognitive factors such as expectations, knowledge or current tasks. Despite their significance, top-down cues are more difficult to model than bottom-up ones. Including other knowledge in a top-down matter is inspired by Guided Search Model proposed by Wolfe *et al.*, (1989), and is typically done by modulating the weights of the conspicuity maps based on previously known top-down information about the analysed scene or about the current task, including context information, face, object or text detectors, or based on different biases (Center Bias, Border Cut).

As for the performances evaluation, there are mainly two different possibilities to evaluate a computational attention system. First, the saliency map can be compared with the results from the psycho-physical experiments, and second, one can evaluate how well a system performed a certain task, compared with the standard algorithm results, or by comparing different systems results to each other. For the bottom-up approaches, many models compare their results with the human eye movements, namely by using certain image datasets, along with the corresponding fixation maps build with eye tracking devices. These evaluations are though not trivial, since there is a high variability between scanpaths of different human viewers. This is due to the fact that every scanpath is influenced by the individual unique cognitive factors for each subject and because it is not fully known how the bottom-up and top-down mechanisms interact nor, how to separate the two of them. Most of the models adopt the easiest evaluation method, by considering simple artificial scene containing *pop-outs* (Frintrop, 2006).

In this work a bottom-up saliency detection method is proposed. Basic features (also called *primitive features* or *attributes*), which are early or pre-attentively processed in the human brain such as colour and orientations, are considered. The orientation features were considered based on the perceptual decomposition of the visual information in multiple processing channels of the human brain which is selectively sensitive to certain frequencies and orientations.

A majority of computational models of attention uses a colour space, which is mostly the default RGB colour space. As stated by Borji & Itti, (2012), choosing the right colour space has a major influence on the system performances. To ensure the best precision, the proposed method was applied in four different colour spaces. The method was tested on three different datasets against other state-of-the-art approaches. Resulted saliency maps were used furthermore for an image abstraction application.

2. Related Works on Saliency Modelling

Most of the existing saliency detection strategies are grounded in the FIT theory of preattentive vision, which extracts the saliency maps in a purely data-driven manner by considering contrast measurements. The earliest such computational approach is the biologically motivated model introduced by Itti *et al.*, (1998), which employs a multi-scale center-surround mechanism that imitates the retinal receptive fields functions to estimate salient image locations. In a similar manner, (Harel *et al.*, 2007), Graph Based Visual Saliency (GBVS) method extracts several feature maps at multiple spatial scales and then, represent them as fully connected graphs.

Gao & Vasconcelos, (2007), predict salient regions by maximizing the mutual information between several local features and class labels as center-surround operations. Bruce & Tsotsos, (2007), propose the Attention based on Information Maximization (AIM), method which employs bottom-up strategies to calculate salient regions as Shannon's self-information $-\log p(f)$, where f is a local visual feature.

Liu *et al.*, (2007), estimate saliency using multi-scale contrast, center-surround colour histograms and colour spatial distributions, which are combined into a Conditional Random Field (CRF). The resulted saliency maps are binary representations which separate salient proto-objects from background.

There are also studies that carry out visual attention computation in the frequency domain. These spectral methods are usually based on detecting rare magnitude picks in the frequency spectrum. Among these, Hou & Zhanh, (2007a), suggested a simple method to estimate salient locations, by computing the inverse Fourier transform of the spectral residual (SR) of the image, which is defined as the difference between the log amplitude spectrum of the image and its smoothed version. On the other hand, Guo *et al*., (2008), suggested that salient image regions are better predicted when using phase spectrum instead of amplitude spectrum of the Fourier transform. Their phase spectrum of the Quaternion Fourier Transform model leads to better results in the spatio-temporal domain salient regions prediction. Further, Guo & Zhang, (2010), extended the PQFT method to a multi scale approach.

Another category of models are those which integrate low-level cues following specific top-down knowledge by integrating face or objects detectors, or by considering global scene context information or pose saliency detection as a supervized learning problem (Liu *et al.*, 2007; Judd *et al.*, 2009). Judd *et al.*, (2009), have introduced a support vector machine (SVM) to learn saliency from a human eye fixations dataset. Biologically inspired low, mid and high level feature sets are used to estimate salient image locations.

Most of the visual attention computational models try to overcome the problem of feature integration. As suggested by FIT theory, most of the approaches combine multiple low-level features in a linear or nonlinear fashion by summation; others compute the optimal weight for each feature vector element (Liu *et al.*, 2007, 2011; Judd *et al.*, 2009), or use quaternion vector representations, making sure that feature interaction is also somehow accomplished (Guo *et al.*, 2008; Guo & Zhang, 2010; Holtzman-Gazit *et al.*, 2010). A similar approach is proposed in this paper. Using quaternion vector representations, orientation features are combined to form saliency maps in different colour spaces. More details about the method are depicted in the following sections, as well as the compressive evaluation results on several benchmark datasets and against other state-of-the-art methods, and brief concluding remarks.

3. Proposed Method

The proposed method follows a multistage approach by recursively down sampling the original image at different sizes, and at each level four distinct conspicuity maps are extracted using band pass filters tuned at certain orientations, namely 0, 45, 90 and 135 degrees. The multi scale decomposition is accomplished by applying Steerable Pyramids filters, first proposed by Simoncelli & Freeman, (1995), which are able to linearly subdivide the image into multiple subbands localized in both scale and orientation. It has been shown that this biologically plausible decomposition correlates with human visual system behavior, which is more sensitive to particular orientations (mostly, vertical lines) and textures with specific frequencies. The Steerable Pyramids decomposition was also applied by Torralba *et al.*, (2006), where the 4 stages with 6 different orientation subbands were used to compute saliency as proposed by Rosenholtz, (1999). Judd *et al.*, (2009), employ Steerable Pyramids to calculate low-level feature-based saliency as a local energy.

It is worth mentioning that a multi scale strategy is mandatory in a visual attention computation, since an image may contain salient proto-objects of various structures and at different scales. This approach is also sustained by the idea that HVS acts as a zoom lens (Wolfe *et al.*, 1989) capable to capture all the relevant information from the visual field regardless of size. Also any resolution level must be equally treated since no level is more important than other and, in this way, the local and global saliency is implicitly incorporated (Holtzman-Gazit *et al.*, 2010).

For the proposed method the steerable decomposition was applied at three different scales. At each scale a three stage pyramid was built. This strategy was reasoned by the fact that the tests on the single pyramid decomposition (with maximum allowed number of scales) approach lead to decreasing accuracy, and also has been many times proved that choosing a certain number of scales for a linear combination is a difficult process, which shows great variability in results accuracy (Borji & Itti, 2012). The number of pyramids and of their subbands is due to the original resolutions of the used images (ranging between 405 × 1,024 and 1,024 × 1,024 pixels).

Saliency maps were calculated for four different colour spaces: the default red green blue (RGB), Lab (using basic Matlab conversions), Intensity Colour Opponent (ICOPP) space as depicted by Guo *et al.*, (2008), Guo & Zhang, (2010), and YUV consisting on luminance, *Y*, and two chromatic, *U* and *V*, channels (Bian & Zhang, 2008). Each input colour image was down sampled by a factor of two at three resolutions, starting from the original size. At each resulted resolution, for each colour channel, a three stage Steerable Pyramid was generated. Each pyramid level consisted on 4 subbands tuned to the previously mentioned orientations: 0, 45, 90 and 135 degrees, which were combined into a quaternion representation as specified in relation

$$q_s^c = S_0^c + S_{45}^c \mu_1 + S_{90}^c \mu_2 + S_{135}^c \mu_3 , \tag{1}$$

where μ_i, , $(\forall i = 1,2,3)$, satisfies relations $\mu_i^2 = -1$, $\mu_1 \perp \mu_2$, $\mu_3 \perp \mu_2$, $\mu_3 \perp \mu_1$, $\mu_3 = \mu_1 \mu_2$. In relation (1) *c* is one of the colour channels from Lab, RGB, ICOPP or YUV colour spaces, *s* – the current level of a pyramid, whereas S_0^c, S_{45}^c, S_{90}^c and S_{135}^c – the four oriented subbands from the scale, *s*.

A quaternion Fourier transform (QFT) (Ell & Sangwin, 2007) is applied to q_s^c, the magnitude is set to 1, and the inverse IQFT is calculated

$$Q_s^c = F^{-1}\left\{ \frac{F\left\{q_s^c\right\}}{\left\| F\left\{q_s^c\right\} \right\|} \right\}. \tag{2}$$

For example, if RGB is considered as the current working colour space, then saliency map corresponding to one of the three colour channels $c \in$ R, G, B, is computed based on the three pyramids, p , each having three scales, s, namely

$$S^c = \sum_{p=1}^{3} \alpha_p \sum_{s=1}^{3} \alpha_s Q_s^c \,. \tag{3}$$

In (3) α_p , $\alpha_s \in \{0.5;0.25\}$ are linear weights, where the best value 0.5 corresponds to the pyramid, respectively, scale with the minimum entropy, whereas the smallest value (0.25) corresponds to the rest of pyramids, respectively, scales. These are experimental values, which were chosen after several trials.

The idea of involving entropy into saliency calculation is inspired by Li *et al.*, (2013), where entropy is used to determine the optimal filtering scale for the amplitude spectrum of the quaternion Fourier transform. Due to the fact that the probabilistic approaches consider a saliency map as a probability map, Li *et al.*, (2013), explain this principle by the need of computing saliency maps as topographical representations with regions of interest assigned with high values, whereas the rest of the image is strongly suppressed. Such a map would probably have a histogram clustered around certain values, which will

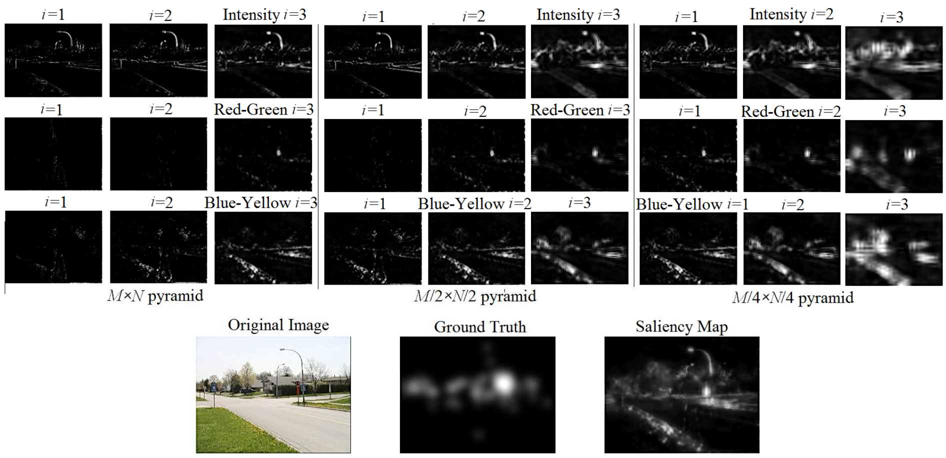

Fig. 1 – Resulted subbands for a given ICOPP image (on rows from left to right): Intensity, Red-Green, Blue-Yellow, respectively, three pyramids with three quaternion decomposed levels, and original image with fixation map and resulted saliency map. For every pyramid, scales with minimum entropy mark the pyramid colour channel.

correspond, according to entropy definition, to small entropy. In this approach all the detail levels are inspected, and thus, objects which appear as salient at different scales are captured. Using a single resolution, regions with high local contrast changes may appear salient; even they are in fact parts of a large textured uninteresting region. In Fig. 1 the decomposed subbands of an ICOPP image are given.

The final saliency map is computed by combining the three normalized maps corresponding to each colour channel of RGB, Lab, ICOPP or YUV space as described in relation (3), where the RGB colour space is considered.

In relation

$$S^{\mathrm{RGB}} = S^{\mathrm{R}} \circ S^{\mathrm{G}} \circ S^{\mathrm{B}} \tag{4}$$

$\circ$ suggests an integration scheme, which may correspond to {+, $*$, max, min}. Through experiments, was found that summation over the three elementary colour maps leads to slightly higher accuracy than the others. Thus, the final saliency map is given by relation

$$S = g * \sum_{c \in \mathrm{R,G,B}} N\left(S^{c}\right), \tag{5}$$

where $N(\ldots)$ is a normalization operator, and g is a 2D Gaussian filter. Since the result, after summation, is actually an image with high values on object edges, by applying the Gaussian filter, edges are smoothed, and objects are filled in the saliency map. Similar maps are computed for the other colour spaces. In Fig. 2 resulted saliency maps are depicted and, although similar, each colour space captures differently the salient regions. After running tests on three benchmark datasets, including *Toronto* (Bruce & Tsotsos, 2007), *MIT* (Judd *et al.*, 2009) and *ImgSal* (Li *et al.*, 2013), the saliency maps computed in ICOPP colour space proved to achieve slightly higher accuracy (see Table 1). Also, within each colour space and across all data bases the accuracy ranking is relatively stable (see Table 1): RGB colour space has the worst performances, while Lab and YUV change the rank between each of them from a data set to other, having close performances.

Table 1
Results on Three Datasets for the Proposed Method in the Four Considered Colour Spaces

Dataset	RGB	Lab	ICOPP	YUV
Toronto	0.8006	0.8183	**0.8205**	0.8180
ImgSal	0.7938	0.8197	**0.8245**	0.8212
MIT	0.8063	0.8225	**0.8233**	0.8220

Since ICOPP is the best colour model across all three datasets, further it will be used as default colour space for the proposed method. In the next section performances measurements are discussed and comparative results are depicted on the three benchmark datasets.

4. Evaluation and Results

Three benchmark datasets were employed for the performances assessment due to the fact that each dataset has some specific characteristics regarding the image statistics, stimulus variety, biases (*e.g.* center bias) and parameters setup for the eye tracking devices. To validate the saliency model the most widely used score is adopted, namely the aria under the receiver operator characteristic curve (ROC AUC). Furthermore, each saliency map is treated as a binary classifier which separates the positive samples (set of all humans' fixation points on an image) from the negatives ones (the union of all eye fixation points across all images of the dataset excluding the positive samples set). By varying the threshold of the classifier the ROC curve is drawn and the area under the curve suggests how well a saliency map can predict human eye fixations. A value of 0.5 leads to chance, while smaller/bigger values indicate a negative/positive correlation. Perfect prediction is achieved for AUC = 1.

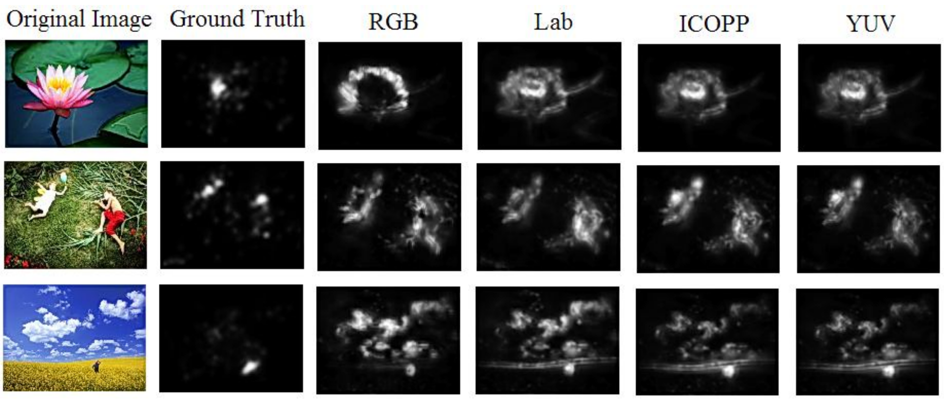

Fig. 2 – Example images from *ImgSal* database, on columns: original images, fixation maps and resulted saliency maps for each colour space with the proposed method.

As for the benchmarking databases, Toronto dataset contains 120 images with fixed resolution 511 × 681 pixels posing indoor and outdoor environments presented to 20 subjects. MIT dataset is the largest existing benchmark for saliency evaluation with 1,003 images (resolutions ranging between 405 × 1,024 and 1,024 × 1,024 pixels) and consists of 779 landscapes and 228 portraits which were visualized by 15 subjects. Whereas ImgSal dataset contains 235 images, most of them are repeating, (resolution 480 × 640 pixels) labelled by 21 naive subjects. Different from Toronto dataset, is that ImgSal is split into 6 image classes, with highly salient regions.

Along with the proposed method other 5 state-of-the-art bottom-up methods (with publically available software) are evaluated: SR, PQFT, Itti's method, AIM and GBVS. Saliency maps, with original images sizes, for the three datasets were calculated using the original setups (resolutions, biases, post-processing operations), as proposed by their authors. Quantitative and qualitative results are further depicted.

In Table 2 the average value of the saliency maps across each dataset are presented. AUC values are calculated for saliency maps at original image size. It can be noticed that performances depend substantially on the dataset images characteristics. If the two spectral algorithms, SR and PQFT, achieved similar results across the all databases, AIM outperformed both of them with similar performances on MIT and ImgSal data sets. FIT based algorithm, proposed by Itti *et al.*, (1998), achieves the worst results, while the graph-based version of this model, GBVS, has similar results with the proposed algorithm. It is well known that GBVS predicts saliency with high accuracy; mostly because it has a global centre bias incorporated. As stated by Judd *et al.*, (2009), blurrier and centred models predict better human eye fixations. The proposed multi scale model, which incorporates a smoothing operation, achieves slightly higher results than GBVS state-of-the-art method on Toronto and MIT data sets. In Fig. 3 visual results are presented.

Table 2
Area under the Receiver Operator Characteristic Curve (Mean Value)

Data set	SR	PQFT	Itti's	AIM	GBVS	**Proposed**
ImgSal	0.7381	0.7337	0.6913	0.7618	**0.8350**	0.8245
Toronto	0.7191	0.7169	0.6546	0.7477	0.8160	**0.8205**
MIT	0.7043	0.6994	0.6548	0.7631	0.8160	**0.8233**

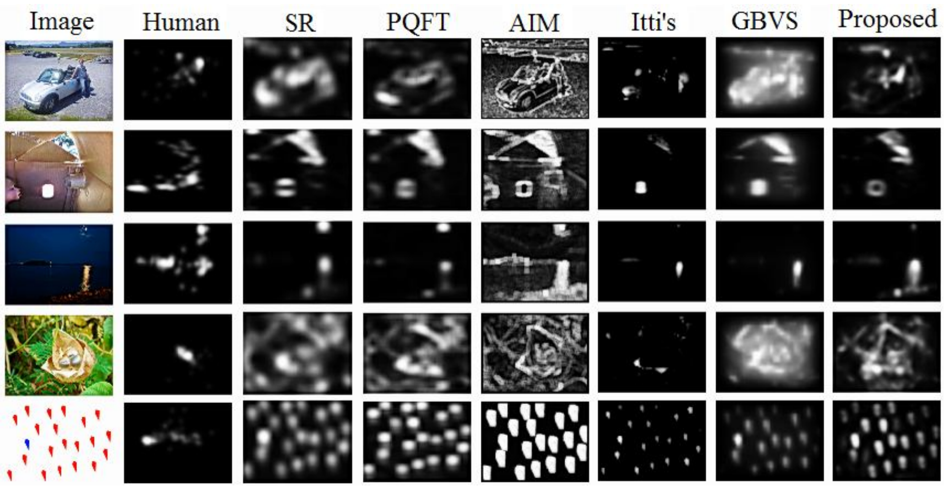

Fig. 3 – Visual comparison between proposed algorithm and other five state-of-the-art methods over resized sample images from MIT data set.

Since a saliency algorithm is important due to the fact that it can be used in an image application enhancement, the proposed multistage method is furthermore used for an image editing application. Using the steerable decomposition, which enables edge detection (Simonceli & Freeman, 1995), the resulted saliency maps highlight the edges of salient objects (Fig. 4 – saliency maps before smoothing), hence an application that uses these edges it will be suitable. Such an example is the image cartoonization that abstracts the image into a cartoon like image. Images obtained by Winnemoller *et al.*, (2006), are abstracted using bilateral filtering and colour quantization. Thus most of textures and edges are strongly highlighted, resulting in a clatter of black lines (see Fig. 4, fourth column). However, by simply taking the product between the saliency maps calculated as in relation (5), without Gaussian smoothing and the edge map obtained by Winnemoller *et al.*, (2006), only salient edges will be emphasized in the final abstracted image.

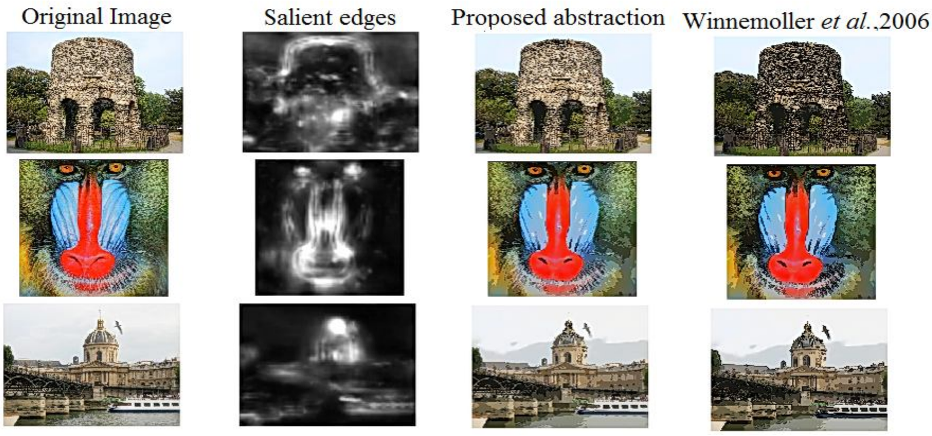

Fig. 4 – Abstraction with or without salient edges: incorporating the proposed salient edges, cartoon improved images and black highlighted irrelevant edges vanish.

5. Conclusions

A multi scale spectral saliency detection bottom-up algorithm is presented which integrates orientation information by using quaternion representation. Steerable decomposed levels are used hierarchically, based on the minimum 2D entropy of each scale in order to compute a saliency map. Also the importance of choosing the right colour space is discussed. Visual results show that the proposed method highlights salient edges, which smoothed by Gaussian kernels, can cover larger salient regions and achieve comparable performances with state-of-the-art methods. Also the method proved its efficiency in improving an image abstraction application.

Acknowledgments. This work was realized with the support of Posdru Cuantumdoc “Doctoral Studies for European Performances in Research and Innovation” ID79407 project funded by the European Social Fund and Romanian Government.

REFERENCES

Bian P., Zhang L., *Biological Plausibility of Spectral Domain Approach for Spatiotemporal Visual Saliency*. ICONIP, **1**, 251-258 (2008).

Borji A., Itti L., *Exploiting Local and Global Patch Rarities for Saliency Detection.* 'CVPR'. IEEE, 2012, 478-485.

Bruce N., Tsotsos J., *Attention Based on Information Maximization.* J. of Vision, **7**, *9*, 950-950 (2007).

Ell T., Sangwin S., *Hypercomplex Fourier Transform of Color Images*. IEEE Trans. on Im. Proc., 16, 22-35 (2007).

Fang Y., Lin W., Lee B.-S., Lau C.T., Chen Z., Lin, C.-W., *Bottom-Up Saliency Detection Model Based on Human Visual Sensitivity and Amplitude Spectrum.* IEEE Trans. on Multimedia **14**, *1*, 187-198 (2012).

Frintrop S., *VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search.* Springer, Berlin-Heidelberg, 2006.

Gao V., Vasconcelo N., *Bottom-Up Saliency as a Discriminat Process*. IEEE Conf. on Comp. Vision, Rio de Janeiro, Brasil, 2007.

Goferman S., Zelnik-Manor L., Tal A., *Context-Aware Saliency Detection*. IEEE Trans. Pattern Anal. Mach. Intell., **34**, *10*, 1915-1926 (2012).

Guo C., Ma Q., Zhang L., *Spatio-Temporal Saliency Detection Using Phase Spectrum of the Quaternion Fourier Transform*. Proc. Internat. Conf. Vis. Pat. Rec., 2008, 1-8.

Guo C., Zhang L., *A Novel Multiresolution Spatiotemporal Saliency Detection Model and its Applications in Image and Video Compression*. IEEE Trans. on Im. Proc., **19**, *1*, 185-198 (2010).

Harel J., Koch C., Perona P., *Graph-Based Visual Saliency*. NIPS, 2007, 545-552.

Holtzman-Gazit M., Zelnik-Manor L., Yavne I., *Salient Edges: A Multi Scale Approach*. ECCV Workshop on Vis. for Cogn. Tasks, 2010.

Hou X., Zhang L., *Saliency Detection: A Spectral Residual Approach.* IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR), 2007a, 1-8.

Hou X., Zhang L., *Thumbnail Generation Based on Global Saliency*. Internat. Conf. on Cognitive Neurodynamics, 2007b, 999-1003.

Itti L., *Automatic Foveation for Video Compression Using a Neurobiological Model of Visual Attention.* IEEE Trans. on Im. Proc., **13**, *10*, 1304-1318 (2004).

Itti L., Koch C., Niebur E., *A Model of Saliency-Based Visual Attention for Rapid Scene Analysis*. IEEE Trans. on Pat. An. and Mac. Int., **20**, *11*, 1254-1259 (1998).

Judd T., Ehinger K.A., Durand F., Torralba A., *Learning to Predict where Humans Look*. 'ICCV' , IEEE, 2009, 2106-2113.

Koch C., Ullman S., *Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry*. Human Neurobiol., **4**, *4*, 219-227 (1985).

Li J., Levine M.D., An X., Xu X., He H., *Visual Saliency Based on Scale-Space Analysis in the Frequency Domain.* IEEE Trans. Pattern Anal. Mach. Intell., **35**, *4*, 996-1010 (2013).

Liu T., Sun J., Zheng N.N., Tang X., Shum H.Y., *Learning to Detect a Salient Object.* Proc. of IEEE Comp. Soc. Conf. on CVPR, 2007, 1-8.

Liu T., Yuan Z.J.S., Wang J., Zheng N., Tang X., Shum H.-Y., *Learning to Detect a Salient Object*. IEEE Trans. Pat. Anal. Mach. Intell. **33**, *2* (2011).

Marchesotti L., Cifarelli C., Csurka G., *A Framework for Visual Saliency Detection with Applications to Image Thumbnailing*. ICCV, IEEE, 2009, 2232-2239.

Rosenholtz R., *A Simple Saliency Model Predicts a Number of Motion Popout Phenomena.* Vis. Res., **39**, 3157-3163 (1999).

Santella A., Agrawala M., DeCarlo D., Salesin D., Cohen M.F., *Gaze-Based Interaction for Semi-Automatic Photo Cropping*. ACM, 2006, 771-780.

Simoncelli E.P., Freeman W.T., *The Steerable Pyramid: a Flexible Architecture for Multi Scale Derivative Computation*. 2nd IEEE Internat. Conf. on Im. Proc., 1995, 444-447.

Torralba A., Oliva A., Castelhano M.S., Henderson J. M., *Contextual Guidance of Eye Movements and Attention in Real-World Scenes: the Role of Global Features on Object Search*. Psychological Review, **113**, *4*, 766-786 (2006).

Treisman A.M., Gelade G., *A Feature-Integration Theory of Attention*. Cogn. Psychol., **12**, 97-136 (1980).

Wang Z., Li B., *A Two-Stage Approach to Saliency Detection in Images*. ICASSP, IEEE, 2008, 965-968.

Winnemoller H., Olsen S.C., Gooch B., *Real-Time Video Abstraction*. ACM Trans. Graph., **25**, 1221-1226 (2006).

Wolfe J.M., Cave K.R., Franzel S.L., *Guided Search: An Alternative to the Feature Integration Model for Visual Search*. J. of Exper. Psychol.: Hum. Percep. a. Perform., **15**, *3*, 419-433 (1989).

EXPLORAREA ORIENTĂRILOR CUATERNIENE RARE PENTRU DETECȚIA DE „SALIENCY” ÎN IMAGINI

(Rezumat)

Detecția de „saliency” a devenit în ultimul timp un instrument valoros pentru multe aplicații de procesare a imaginilor. Cu toate acestea, variațiile metodologice ale algoritmilor existenți permit încă îmbunătățirea performanțelor. În această lucrare, o metodă „state-of-the-art”, şi anume „Phase Spectrum of Quaternion Fourier Transform” (PQFT), este extinsă într-o abordare multiscală, în care se utilizează descompunerea bazată pe piramide orientabile. Considerând diferite spații de culoare, imaginile sunt descompuse, la diverse scale şi pe patru orientări, care sunt ulterior reprezentate vectorial folosind reprezentări cuaterniene. Bazat pe entropia fiecărei scale a piramidelor orientabile, toate nivelele sunt ponderate într-o combinație liniară care generează o hartă „saliency”. Rezultatele calitative şi vizuale demonstrează că metoda propusă are performanțe comparabile cu cele ale altor metode „state-of-the-art”, şi, de asemenea, poate îmbunătăți o aplicație de abstractizare a imaginilor.