Robust Complex Salient Regions

[1]Sergio Escalera, [2]Oriol Pujol, and [1]Petia Radeva

[1]Computer Vision Center, Dept. Computer Science, UAB, 08193 Bellaterra, Spain
[2]Dept. Matemàtica Aplicada i Anàlisi, UB, Gran Via 585, 08007, Barcelona, Spain

Abstract. The challenge of interest point detectors is to find, in an unsupervised way, keypoints easy to extract and at the same time robust to image transformations. In this paper, we present a novel set of saliency features that takes into account the region inhomogeneity in terms of intensity and shape. The region complexity is estimated at real-time by means of the entropy of the grey-level information. On the other hand, shape information is obtained by measuring the entropy of normalized orientations. The normalization step is a key point in this process. We compare the novel complex salient regions with the state-of-the-art keypoint detectors. The new set of interest points shows robustness to a wide set of transformations and high repeatability. Besides, we show the temporal robustness of the novel salient regions in two real video sequences.

1 Introduction

Visual saliency [1] is a broad term that refers to the idea that certain parts of a scene are pre-attentively distinctive and create some form of immediate significant visual arousal within the early stages of the Human Vision System. The term 'salient feature' has previously been used by many other researchers [12][1]. Although definitions vary, intuitively, saliency corresponds to the 'rarity' of a feature [2]. In the framework of keypoint detectors, special attention has been paid to biologically inspired landmarks. One of the main models for early vision in humans, attributed to Neisser [6], is that it consists of pre-attentive and attentive stages. In the pre-attentive stage, 'pop-out' features are only detected. These are the salient local regions of the image which present some form of spatial discontinuity. In the attentive stages, relationships between these features are found, and grouping takes place in order to model object classes.

Region detectors have been used in several applications: baseline matching for stereo pairs, image retrieval from large databases, object retrieval in video, shot location, and object categorization [9][8], to mention just a few. One of the most well-known keypoint detector is the Harris detector [3]. The method is based on searching for edges at different scales to detect interest image points. Several variants and application based on the Harris point detector have been used in the literature, such as Harris-Laplacian [5], Affine variants [3], DoG [4], etc. In [11], the authors proposed a novel region detector based on the stability of the parts of the image. Nevertheless, the homogeneity of the detected regions makes the description of the parts ambiguous when considered in object recognition frameworks. Schmid and Mohr [3] proposed the use of corners as interest

points in image retrieval. They compared different corner detectors and showed that the best results were provided by the Harris corner detector [5]. Kadir et al [1] estimate the entropy of the grey levels of a region to measure its magnitude and scale of saliency. The detected regions are shown to be highly discriminable, avoiding the exponential temporal cost of analyzing dictionaries when used in object recognition models, as in [12]. Nevertheless, using the grey level information, one can obtain regions with different complexity and with the same entropy values. In [10], a method for introducing the cornerness of the Harris detector in the method of [1] is proposed. Nevertheless, the robustness of the method is directly dependent on the cornerness performance.

In this paper, we propose a model that allows to detect the most relevant image features based on their saliency complexity. We use the entropy measure based on the color or grey level information and shape complexity (defined by means of a novel normalized pseudo-histogram of orientations) to categorize the saliency levels. This new Complex Salient Regions can be related to the pre-attentive stage of the HVS. In this sense, they are biologically inspired since it is known that some neural circuits are specialized or sensitive to a restrictive set of visual shapes, as edge, contour and motion detectors as others related to color and spatial frequencies [7]. Although orientations have been previously used in the literature with very few success[1], our approach defines a normalized procedure that makes this measure very relevant and robust.

The paper is organized as follows: chapter 2 explains our Complex Salient Regions, section 3 shows experimental results, and section 4 concludes the paper.

2 Complex Salient Regions

In [1], Kadir et. al. introduce the grey-level saliency regions. The key principle is that salient image regions exhibit unpredictability in their local attributes and over spatial scale. This section is divided in two parts: firstly, we describe the background formulation, inspired in [1]. And, secondly, we introduce the new metrics to estimate the saliency complexity.

2.1 Detection of salient regions

The framework to detect the position and scale of the saliency regions uses a saliency estimation (defined by the Shannon entropy) at different scales of a given point. In this way, we obtain a function of the entropy in the space of scales. We consider significant saliency regions those that correspond to maxima of that function, where the maxim entropy value is used to estimate the complex salient magnitude. Now we define the notation and description of the stages of the process.

Let H_D be the entropy of a given descriptor D, S_p the space of significant scales, and W_D the relevance factor (weight). In the continuous case, the saliency measure γ_D, a function of scale s and position x, are defined as:

$$\gamma_D(S_p, x) = W_D(S_p, x)H_D(S_p, x) \tag{1}$$

for each point x and the set of scales S_p at which entropy peaks are obtained. Then, the saliency is determined by weighting the entropy at those scales by W_D. The entropy H_D is defined as $H_D(s,x) = -\int p(I,s,x)\log_2 p(I,s,x)dI$, where $p(I,s,x)$ is the probability density of the intensity I as a function of scale s and position x. In the discrete case, for a region R_x of n pixels, the Shannon entropy is defined as

$$H_D(R_x) = -\sum_{i=1}^{n} P_{D,R_x}(i) log_2 P_{D,R_x}(i) \tag{2}$$

where $P_{D,R_x}(i)$ is the probability of descriptor D taking the value i in the local region R_x, for n grey levels. The set of scales S_p is defined by the maxima of the function H_D in the space of scales $S_p = \left\{s : \frac{\partial H_D(s,x)}{\partial s} = 0, \quad \frac{\partial^2 H_D(s,x)}{\partial s^2} < 0\right\}$

These equations are illustrated by the detected local maxima in fig. 1. In the figure, a point x is evaluated in the space of scales, obtaining two local maxima. These peaks of the entropy estimation correspond to the representative scales for the analyzed image point.

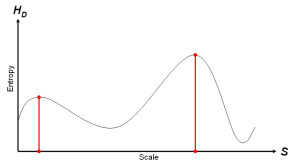


Fig. 1. Local maxima of function H_D in the scale space S

The relevance of each position of the saliency at its representative scales is defined by the inter-scale saliency measure $W_D(s,x) = s\frac{\partial}{\partial s}H_D(s,x)$.

Considering each scale s of S_p and the pixel x, we estimate W_D in the discrete case as,

$$W_D(s,x) = s\frac{|H_D(s-1,x) - H_D(s,x)| + |H_D(s+1,x) - H_D(s,x)|}{2} \tag{3}$$

where $s \in [1,...,S]$, for S the total number of scales. Using the previous weighting factor, we assume that the significant salient regions correspond to that locations with high distortion in terms of the Shannon entropy and its peak magnitude.

2.2 Traditional grey-level and orientation saliency

Kadir et. al. [1] used the grey-level entropy to define the saliency complexity of a given region. However, this approach falls short in front of clear cases of different complexities. In fig. 2 one can observe different regions with the same amount of pixels for each grey level and different visual complexity. Note that the approach proposed by [1] gives the same entropy value for all of them.

A natural and well founded measure to solve this pathology is the use of complementary orientation information. In the same work [1], Kadir et. al. considered the use of orientations with very limited and inconclusive results. The use

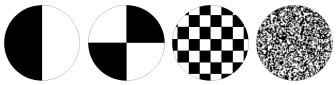
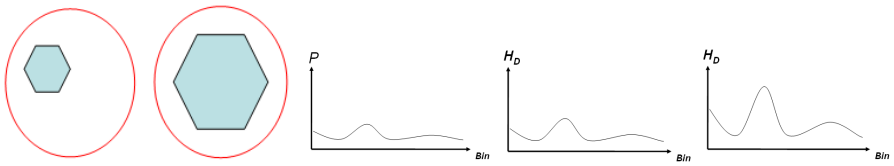


Fig. 2. Regions of different complexity with the same grey level entropy.

of orientations as a measure of complexity involves several problems. In order to exemplify those problems, suppose that we have the regions (a) and (b) of fig. 3. Both regions have the same pdf (fig. 3(c)), although contain different number of significant orientations with the same proportion (histograms of fig. 3(d) and (e)).



(a) (b) (c) (d) (e)

Fig. 3. (a)(b) Two circular regions with the same content at different resolutions. (c) Coincident pdf for the regions (a) and (b). (d) Orientations histogram for (a), and (e) orientations histogram for (b).

To solve the commented problems, we propose a design of the normalized orientation.

2.3 Normalized orientation entropy measure

The normalized orientation entropy measure is based on computing the entropy using a pseudo-histogram of orientations. The usual way to estimate the histogram of orientations of a region is to use a range from 0 to 2π radians. However, a very important information related to the orientation is omitted, the lack of orientation, referred from now on as '*non-orientation*'. Our proposed orientation metric consists of computing the saliency including this *non-orientations* in the modified orientation pdf.

Considering the $k \leq K$ most significant orientations using an experimental threshold, where K is the total orientation magnitudes from a given region, we compute the histogram h_O. The normalization bin is then added as $h_O(n+1) = K-k$. In this way, the modified orientation pdf for the histogram h_O is obtained by means of:

$$PDF_O(i) = \frac{h_O(i)}{\sum_{j=1}^{n+1} h_O(j)}, \forall i \in [1, .., n+1] \quad (4)$$

In order to obtain the orientation entropy value, we consider the first n values of the normalized histogram. Note that the $n+1$ position is not included in the entropy evaluation since its goal is to normalize the first n positions, as shown in eq. (4).

2.4 Combining the saliency

In our particular case, the grey-level histogram is combined with the pseudo-histogram of orientations. In this way, once estimated the two corresponding pdf, we apply equations (1), (2), and (3) to each one, and the final measure combination is obtained by means of the simple addition[1] $\gamma = \gamma_G + \gamma_O$, where γ_G and γ_O are estimated by equation (1) for the grey and orientation saliency, and γ is the result, where the final significant saliency positions, magnitudes (level of complexity), and scales are defined. This new saliency measure gives a high complexity value when the region contains different grey levels information (non-homogeneous region), and the shape complexity is high (high number of gradient magnitudes at multiple orientations). The complexity order to detect the salient regions is $O(dl)$, where d is the number of image pixels, and l is the number of scales searched for each pixel.

3 Results

We compare the presented CSR with the Harris-Laplacian, Hessian-Laplacian, and the grey-level saliency in terms of repeatability and false alarm rate. The parameters used for the region detectors are the default parameters given by the authors [11][1][3]. The number of regions obtained by each method strongly depends on the image type since each one responds to different type of features. Nevertheless, we use the 20% maximum responses of each detector to analyze the robustness of the most significant salient regions.

In order to validate our results, we selected the samples of fig. 4 from the public Caltech repository database. In this set of samples, we applied a set of transformations: rotation (10 degrees per step up to 100), white noise addition (0.1 of the variance per step up to 1.0), scale changes (15% per step up to 150), and affine distortions (5 pixels x-axis distortion per step up to 50). The mean results for the repeatability and false alarm ratios are shown in fig. 5. We consider the repeatability defined as the percentage of the initial detected regions that is maintained in the space of transformations, and the false alarm rate as the percentage of detected regions that do not have a correspondence in the initial image. Observing the figures, one can see that the CSR regions obtain better performance in terms of repeatability, obtaining the highest percentage of intersected regions for all types of image distortions. For the case of false alarm rate, the CSR and the Hessian Laplace methods are the best, obtaining similar results.

The next experiment is to apply the CSR regions to video sequences to show its temporal robustness. We have used the video images from the Ladybug2 spherical digital camera from Point Grey Research group [13]. The car system has six cameras that enable the system to collect video from more than 75% of the full sphere [13]. Besides, we have tested road video sequences from the

[1] We have experimentally observed that this simple combination obtains the most relevant results in comparison with other kinds of combinations.

Fig. 4. Caltech database samples used to test the keypoint detectors.

Geovan Mobile Mapping process from the Institut Cartogràfic de Catalunya [14]. For both experiments we have analyzed 100 frames, using the SIFT descriptor [4] to describe the regions. The matching is done by similar regions descriptors in a neighborhood of the detected CSRs. The smoothed oriented maps from CSR matchings are shown in fig. 6 and fig. 7. Fig. 6(a) shows the oriented map in the first analyzed frame of [13]. Fig. 6(b) focuses on the right region of (a). One can see that the matched complex regions correspond to singularities in the video sequence and approximates roughly the video movement. From the road experiment of fig. 6, where appear cars and traffic signs (fig. 6(a) and (b)), the oriented map is shown in fig. 6(c), where the amplified right region shown in fig. 6(d) shows the correct temporal behavior of the road video sequences.

4 Conclusions

We have presented a novel set of salient features, the Complex Salient Regions (CSR). These features are based on complex image regions estimated at real-time using an entropy measure. The presented CSR analyzes the complexity of the regions using the grey-level, and orientations information. We introduced a novel procedure to consider the anisotropic features of image pixels that makes the image orientations useful and highly discriminable in object recognition frameworks. One can use the complexity criteria to adjust the detector requirements in a compromise between robustness and computational time. The novel set of features is highly invariant to a great variety of image transformations, and leads to a better repeatability and lower false alarm rate than the state-of-the-art keypoint detectors. These novel salient regions show robust temporal behavior on real video sequences, and can be potentially applied to real-time matching and image retrieval problems (less than 1 second in 800×640 medium resolution images), avoiding the exponential number of features and time complexity of the exhaustive methods.

5 Acknowledgements

This work was supported in part by the projects, FIS-G03/1085, FIS-PI031488, MI-1509/2005, and TIN2006-15308-C02-01.

References

1. T. Kadir, and M. Brady, "Saliency, Scale and Image Description", *Intl. J. of Computer Vision*, vol. 45, issue 2, pp- 83-105, 2001.

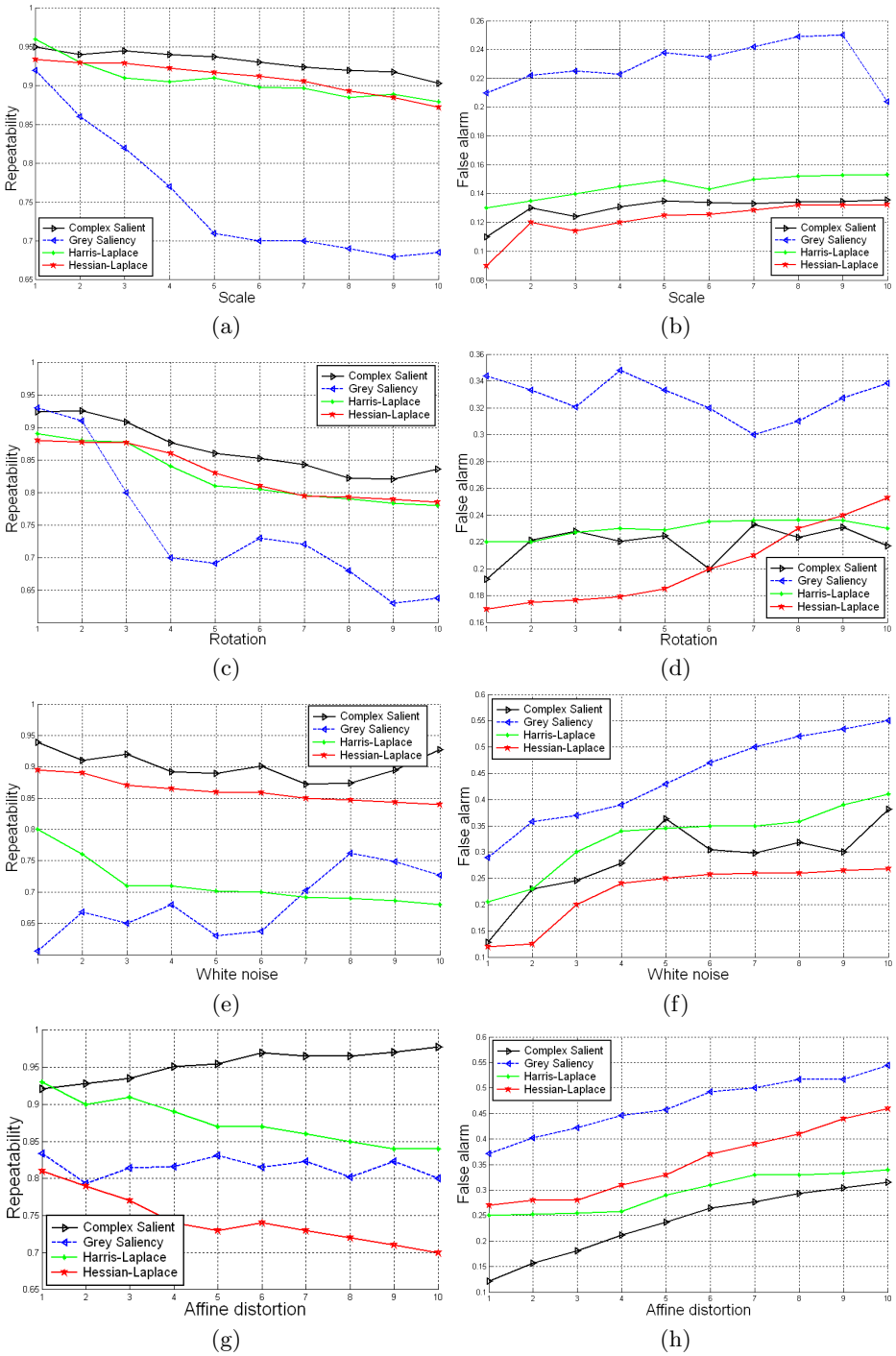


Fig. 5. (a)(b)Hit rate (H) and false alarm rate (FA) for scale, (c)(d) rotation, (e)(f) white noise, and (g)(h) affine invariants in the space of transformations.

(a) (b)

Fig. 6. (a) Smoothed oriented CSR matches, (b) Zoomed right region.

(a) (b) (c) (d)

Fig. 7. (a)(b) Samples, (c) Smoothed oriented CSR matches, (d) Zoomed right region.

2. D. Hall, B. Leibe, B. Schiele, "Saliency of Interest Points under Scale Changes", *proc. of the British Machine Vision Conference*, 2002.
3. K. Mikolajczyk and C. Schmid, "Scale & Affine Invariant Interest Point Detectors", *International Journal of Computer Vision*, vol. 60, pp. 63-86, 2004.
4. D. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 20, pp. 91-110, 2003.
5. C. Harris and M. Stephens, "A combined corner and edge detector", *Alvey Vision Conference*, pp. 147-151, 1999.
6. U. Neisser, "Visual Search", Scientific American, vol. 210, issue 6, pp. 94-102, June 1964.
7. W.E.L. Grimson, "From Images To Surfaces: A Computational Study of the Early Human Visual System", *MIT Press*, 1981.
8. C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, issue 5, pp. 530-535, 1997.
9. R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning", *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, USA, 2003.
10. F. Fraundorfer and H. Bischof, "Detecting Distinguished Regions by Saliency", *Image Analysis*, Springer, vol. 2749, pp. 208-215, 2003.
11. J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide baseline Stereo from Maximally Stable Extremal Regions", *Proc. of the British Machine Vision Conference*, vol. 1, pp. 384-393, 2002.
12. T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio, "A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex", *AIM*, vol. 36, 2005.
13. http://ptgrey.com/products/ladybug2/samples.asp
14. http://www.icc.es