SCALE SALIENCY : A NOVEL APPROACH TO SALIENT FEATURE AND SCALE SELECTION

Timor Kadir,[*] Michael Brady

Robotics Research Group, Oxford University, UK

ABSTRACT

This paper presents an overview of the Scale Saliency algorithm recently introduced in (10). Scale Saliency is a novel method for measuring the saliency of image regions and selecting optimal scales for their analysis. The model underlying the algorithm deems image regions salient if they are simultaneously unpredictable in some feature-space and over scale. The algorithm possesses a number of attractive properties: invariance to planar rotation, scaling, intensity shifts and translation; robustness to noise, changes in viewpoint, and intensity scalings. Moreover, the approach offers a more general model of feature saliency compared with conventional ones, such as those based on kernel convolution, for example wavelet analysis, since such techniques define saliency and scale only with respect to a particular set of basis morphologies. Finally, we present a generalised version of the original algorithm which is invariant to Affine transformations.

INTRODUCTION

Computer vision algorithms are, in general, information reduction processes. Brute-force approaches to image or image sequence analysis can quickly overwhelm most computing resources at our disposal. Fortunately, images are a redundant data source. The same set of inferences may be drawn from a variety of image characteristics. This becomes self-evident considering the array of different methodologies available for solving any particular vision task. Hence, the selection of a sufficient set of image regions and properties, or salient features, forms the first step in many computer vision algorithms.

Two key issues face the vision algorithm designer: the subset of image properties selected for subsequent analysis and the model used to represent those properties. For example, many image matching algorithms begin with a set of 'landmark' points which serve as a basis for estimating the image transformation that defines the match. In this case, well-localised and unique image regions are desirable to minimise the likelihood of false matches. For many tasks, geometric and photometric invariance properties are also beneficial. Finally, there is often an implicit, but difficult to quantify, requirement that the salient regions be relevant to the task of interest — in other words, the regions or descriptions subsequently extracted from them are somehow characteristic of the scene contents they are intended to signify.

[*]Timor Kadir is the recipient of a Motorola University Partners in Research grant.

Many definitions for saliency have been proposed. Perhaps the most popular have arisen out of the application of local surface differential geometry techniques to imaging. Such methods consider the image to be a discrete approximation to a surface and categorize it by application of differential operators. Closely related to these are basis projection and filtering methods. Common to both is the development of one or two dimensional features; one dimensional features include edges, lines, ridges (2, 3); two dimensional features are often referred to as Interest points or 'Corners' (4, 7). More recently, inspired by the pioneering of Lindeberg work (11), scale and affine adapted versions of Interest point detectors have been developed, based on Corner (12) and blob detectors (1).

In general, these methods share one assumption: that saliency can be defined directly with respect to some particular property of the geometry or morphology of the image surface. Consequently, such methods tend to respond well only to a relatively narrow set of image morphologies. Efforts to overcome this limitation and to generalise such methods to capture a broader range of salient image regions have had limited success. For example, one approach is to combine the output of a bank of 'complementary' feature detectors (13). However, such methodologies must address the cue integration problem, to which a satisfactory solution remains elusive to date.

In practice, such limitations become manifest in a number of ways, dependent on the particular application of interest. For example, in some applications such as image matching and registration, representations of images based on Corner or blob-based Interest point detectors can be useful since large numbers of points can be employed to estimate global transforms. In contrast however, recognition tasks benefit from compact descriptions to facilitate fast matching. In such applications, the relatively narrow set of feature morphologies to which such operators respond, results in a poor representation of the object class. To overcome this, large numbers of such features may be used, but at a penalty of increased computational load. Finally, as has been noted elsewhere (14), where image features do not correspond to the restricted saliency model, the scale selection of such detectors is poor.

We argue that a set of features derived from a more general definition of saliency can overcome such limitations and provide high quality compact representations of the image beneficial for tasks such as recognition, tracking and matching.

SCALE SALIENCY

Recently, we have proposed a novel approach to feature saliency (10), termed Scale Saliency. Motivated by the work of Gilles (6), our technique deems 'salient' those regions exhibiting unpredictable characteristics simultaneously in some feature-space and over scale.

In our formulation, Scale Saliency is a product of two terms, each a function of the PDF of local image attributes (e.g. intensity, colour) at multiple scales : $\mathcal{H}_D$ (Shannon Entropy) and $\mathcal{W}_D$ – which measure feature-space and inter-scale unpredictability, respectively. Extrema in $\mathcal{H}_D$ are used as the basis for scale selection. Under such a scheme the multi-scale representation is implicit in the estimation of the local PDF (e.g. using a histogram) at multiple scales. Critically, conventional blurring or wavelet decomposition methods are avoided since such methods alter the local image statistics.

Intuitively, the method works as follows. $\mathcal{H}_D$, the entropy of local attributes measures the predictability of a region with respect to an assumed model of simplicity. In the case of entropy of local intensities, the model of simplicity is a region with a single intensity value. In such a scheme, the simplest and hence least salient region is a piecewise constant region. For example, consider the image in Figure 1. At a particular scale, the PDF of intensities in the eye region is flatter and hence has a higher entropy than that of the cheek region. Other attributes may be used to define alternative benchmarks against which a regions predictability is measured. For example, edge orientations, colour, optical flow, or texture may be used (9, 10). $\mathcal{W}_D$, the measure of inter-scale saliency applies a second constraint to salient features — that they are unpredictable over scale.

$\mathcal{W}_D$ is simply a measure of the magnitude change in the local PDF as a function of scale. This places a geometric constraint on the spatial configurations of pixels which maximise Scale Saliency. For example, the circular window used in (10) biases the method towards isotropic features. In this paper, we present a generalisation of this restriction that allows the method to become invariant to the full set of Affine transformations. Further details and analyses of the Scale Saliency method may be found in (9, 10).

In the discrete case, Scale Saliency is defined as:

$$\mathcal{Y}_D(\mathbf{s_p}, \mathbf{x}) \triangleq \mathcal{H}_D(\mathbf{s_p}, \mathbf{x}) \ \mathcal{W}_D(\mathbf{s_p}, \mathbf{x}) \qquad (1)$$

where entropy $\mathcal{H}_D$ is defined by:

$$\mathcal{H}_D(s, \mathbf{x}) \triangleq - \sum_{d \in D} p_{d,s,\mathbf{x}} \log_2 p_{d,s,\mathbf{x}} \qquad (2)$$

and where $p_{d,s,\mathbf{x}}$ is the probability as a function of scale s, position $\mathbf{x}$ and descriptor value d which takes on values in D, the set of all descriptor values. The inter-scale saliency measure, $\mathcal{W}_D(s, \mathbf{x})$, is defined by:

$$\mathcal{W}_D(s, \mathbf{x}) \triangleq \frac{s^2}{2s-1} \sum_{d \in D} |p_{d,s,\mathbf{x}} - p_{d,s-1,\mathbf{x}}| \qquad (3)$$

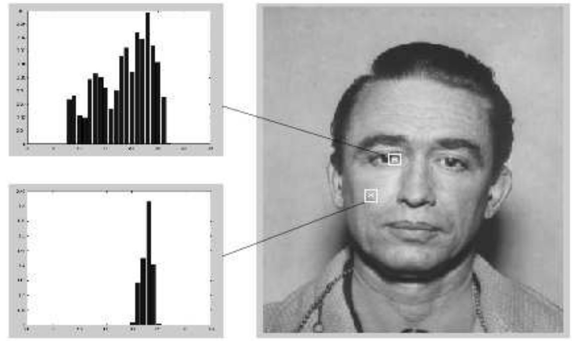

Figure 1: High saliency regions, such as the eye, exhibit unpredictable local intensity hence high entropy. Image from NIST Special Database 18, Mugshot Identification Database.

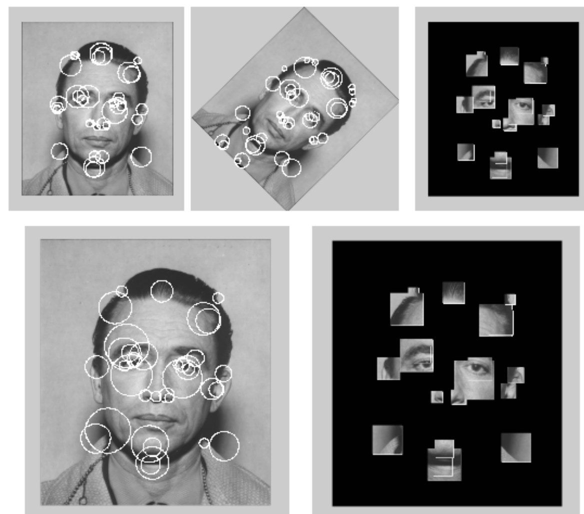

Figure 2: Salient scale region selection and salient icons (10% most salient shown) are robust to rotation (45° clockwise) and scaling (60% of original size).

The set of scales $\mathbf{s_p}$, at which entropy peaks, is defined by:

$$\mathbf{s_p} \triangleq \{s : \mathcal{H}_D(s-1, \mathbf{x}) < \mathcal{H}_D(s, \mathbf{x}) > \mathcal{H}_D(s+1, \mathbf{x})\} \qquad (4)$$

Scale Saliency possesses a number of attractive properties. It is invariant to similarity transforms and intensity shifts, and robust to small changes in viewpoint and intensity scalings. It offers a more general model of feature saliency and scale compared to conventional feature detection techniques, such as those employing basis projection, differential geometry or Scale-Space approaches. It also incorporates an intrinsic notion of inter-scale saliency and thereby provides a technique for scale-selection. Compared to visual search and attention techniques such as (8), Scale Saliency offers a coherent methodology; its relation to tasks such as cue selection and image description are well understood and closed-

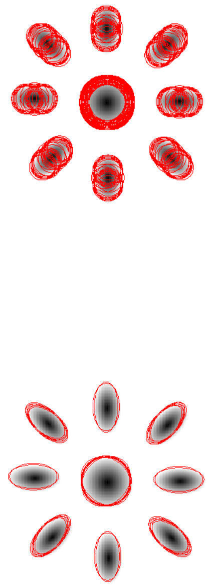

Figure 3: Comparing the original isotropic (top) and modified anisotropic (bottom) Scale Saliency algorithms.

form expressions for its maximisation exist.

Figure 2 illustrates the Scale Saliency algorithm applied to three versions of a face image: the original, a 60% scaled, and a 60% scaled and $45°$ rotated. The white circles show the 10% most salient regions at their respective scales. The example demonstrates that the algorithm can select relevant features that are stable across similarity transformations. The images on the right are the so-called Sparse Iconic Representation proposed by Gilles (6) as a minimal representation of an image.

ANISOTROPIC SCALE SALIENCY

In this section, we generalise the circular sampling function to the anisotropic case using an ellipse parameterised by a scale parameter, a rotation and aspect ratio. This modification enables the method to become invariant to anisotropic scaling and shear; that is, the full affine set of transformations. Another benefit is that orientation information can be captured. Similar approaches have been used to generalise Corner based interest point detectors to the affine invariant case (1, 12).

The modification is quite straight-forward and requires replacing the single parameter sampling function with a three-parameter version. The entropy is calculated at each step of each of the three parameters. There is no need to modify, $\mathcal{W}_D$, the inter-scale saliency because as shown in (9) the shape that causes the largest $\mathcal{W}_D$ is the one that matches the feature shape. Furthermore, including the rotation angle in $\mathcal{W}_D$ would cause a bias against

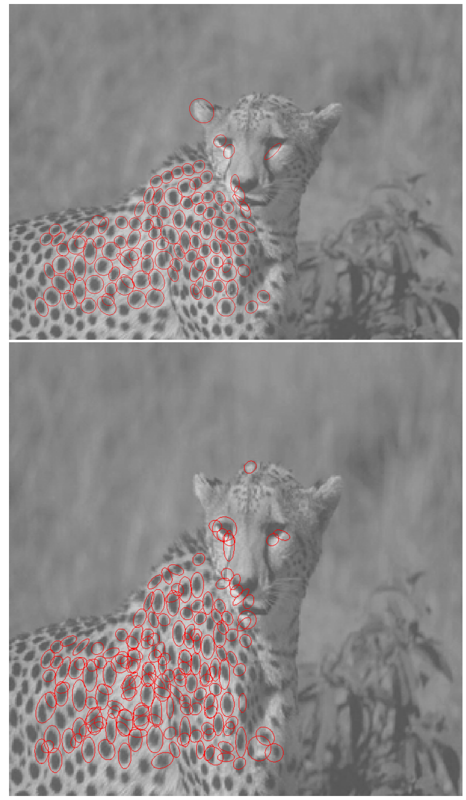

Figure 4: The anisotropic Scale Saliency applied to a Cheetah image and its stretched version. The 0.5% (by number) clustered most salient features are shown. In each case, many of the Cheetah's spots are correctly identified at their appropriate local (anisotropic) scale.

isotropic features. Similarly, the peak detection is not modified. As in the isotropic case, scales are selected at peaks of entropy over scale.

Equations 2 to 4 can be modified for the anisotropic case by replacing the scalar s parameter with a vector, $\mathbf{s} = (s, r, \theta)$ corresponding to the scale, ratio and orientation. The vector of scales at which the entropy peaks, $\mathbf{s_p}$, becomes a matrix, $\mathbf{S_p}$ with three rows, one for each of the scale variables and as many columns as peaks at that position. For completeness, the modified equations are as follows:

$$\mathcal{Y}_D(\mathbf{S_p}, \mathbf{x}) \triangleq \mathcal{H}_D(\mathbf{S_p}, \mathbf{x})\ \mathcal{W}_D(\mathbf{S_p}, \mathbf{x}) \tag{5}$$

$$\mathcal{H}_D(\mathbf{s}, \mathbf{x}) \triangleq -\sum_{d \in D} p_{d,\mathbf{s},\mathbf{x}} \log_2 p_{d,\mathbf{s},\mathbf{x}} \tag{6}$$

$$\mathcal{W}_D(\mathbf{s}, \mathbf{x}) \triangleq \frac{s^2}{2s-1} \sum_{d \in D} |p_{d,s,\mathbf{x}} - p_{d,s-1,\mathbf{x}}| \tag{7}$$

$$\mathbf{S_p} \triangleq \{\mathbf{s} : \mathcal{H}_D(s-1, \mathbf{x}) < \mathcal{H}_D(s, \mathbf{x}) > \mathcal{H}_D(s+1, \mathbf{x})\} \tag{8}$$

The original isotropic and modified anisotropic Scale Saliency algorithms are compared in Figure 3 where they have been applied to a synthetic image. The anisotropic version correctly identifies the scales of the ellipses and the circle, whereas the isotropic version correctly detects only the circle and finds numerous features along the ellipses. In Figure 4 the anisotropic Scale Saliency is applied to an original and stretched version of an image of a Cheetah. The image sizes for the original and stretched versions were 262x340 and 340x340 respectively. The features have been thresholded and clustered using the algorithm described in (10) (modified to work in this new space). The parameters of the clustering were set such that the images shown are fairly clear. It can be observed that in both images many of the spots (in this case the most salient features) have been identified at a scale, ellipse ratio and orientation that is appropriate to the local feature. The extra information brought through the use of the three parameter scale-space provides a more accurate representation of the image and a richer descriptor set.

However, it should be noted that we have found the modified algorithm to be quite sensitive to noise in the image and further developments are necessary before this approach can be applied. For example, alternative parameterisations of the ellipse might prove more stable, for example using the scales of the two axes and a rotation. Ultimately this becomes an issue of statistical geometry. Another problem is that the increased parameterisation results in large increase in computational load; processing time for Figure 4 was 160 seconds. However, it is expected that there may be many opportunities for optimisation, for example by using local adaptation in the manner of (5, 6). Such investigations are ongoing.

CONCLUSION

In this paper, we have presented an overview of the Scale Saliency algorithm. Scale Saliency offers are more general model of feature saliency compared to conventional methods and also possesses some attractive properties such as photometric robustness and similarity invariance. The output of the algorithm is a set of locations and scales ranked in order of saliency. These may be used as a basis for compact representations on images which facilitate matching and recognition tasks. We have also presented a generalisation of the algorithm which extends the method to full affine invariance.

ACKNOWLEDGEMENTS

This research was sponsored by a Motorola University Partners in Research grant. We would like to thank Dr. Paola Hobson of the Motorola UK Research Laboratory for the many very useful discussions we have had during the course of this work.

REFERENCES

[1] A. Baumberg. Reliable feature matching across widely separated views. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 774–781, 2000.

[2] F. Bergholm. Edge focusing. In *Proc. International Conference on Pattern Recognition*, pages 597–600, 1986.

[3] J. Canny. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[4] R. Deriche and T. Giraudon. A computational approach for corner and vertex detection. *International Journal of Computer Vision.*, 10(2):101–124, 1993.

[5] J. Garding and T. Lindeberg. Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision*, 17(2):163–191, 2001.

[6] S. Gilles. *Robust Description and Matching of Images*. PhD thesis, University of Oxford, 1998.

[7] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conference*, pages 189–192, 1988. Manchester.

[8] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[9] Timor Kadir. *Scale, Saliency and Scene Description*. PhD thesis, University of Oxford, 2002.

[10] Timor Kadir and Michael Brady. Scale, saliency and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[11] T. Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318, 1993.

[12] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *International Conference on Computer Vision*, pages 525–531, 2001.

[13] R. Milanese. *Detecting Salient Regions in an Image: From Biological Evidence to Computer Implementation*. PhD thesis, University of Geneva, 1993.

[14] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". In *Proc. European Conference on Computer Vision*, pages 414–431, 2002.