Salient Point and Scale Detection by Minimum Likelihood

Kim Steenstrup Pedersen KIMSTP@DIKU.DK
Marco Loog LOOG@DIKU.DK
Department of Computer Science
University of Copenhagen
Universitetsparken 1
DK-2100 Copenhagen
Denmark

Pieter van Dorst
Department of Biomedical Engineering
Technische Universiteit Eindhoven
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Editor: Neil Lawrence, Anton Schwaighofer and Joaquin Quiñonero Candela

Abstract

We propose a novel approach for detection of salient image points and estimation of their intrinsic scales based on the fractional Brownian image model. Under this model images are realisations of a Gaussian random process on the plane. We define salient points as points that have a locally unique image structure. Such points are usually sparsely distributed in images and carry important information about the image content. Locality is defined in terms of the measurement scale of the filters used to describe the image structure. Here we use partial derivatives of the image function defined using linear scale space theory. We propose to detect salient points and their intrinsic scale by detecting points in scale-space that locally minimise the likelihood under the model.

Keywords: Image analysis, salient image point detection, scale selection, fractional Brownian images

1. Introduction

Following the widely recognised paradigm by Marr (1982), image feature detection is the basis which many vision and image analysis algorithms build upon. The definition of what constitutes an image feature is debatable, but there is a common agreement that it is either a point or curve at which the image intensities have a visual information carrying geometry. Examples of curve-like features include edges formed by abrupt contrast changes and ridges formed by bar- or valley-like intensity structures. Examples of point features include T-junctions formed by two edges crossing (usually caused by occluding objects), corners formed by two edges ending at a point, and blobs which are local extrema of the image intensity function. The term image feature (or just feature) should not be confused with the term feature common in pattern recognition. The former is represented by a position in the image domain.

An image feature has an intrinsic scale that describes its extent in space, e.g. a blob has a certain width and an edge has a certain sharpness. In this paper we use linear scale space theory (Koenderink, 1984; Witkin, 1983) for the analysis of local image geometry or structure at different scales. In this theory the scale of an image is changed by applying a Gaussian function filter to the image by convolution. The standard deviation of the Gaussian defines the scale of the processed image. There are several benefits of this theory, but the main point we are going to use is that by blurring the image (see Fig. 1 for an example) the problem of computing derivatives of the image becomes well-posed. The partial image derivatives describe the differential geometric properties of the image and are commonly used for defining image feature detectors. As an example, first order derivatives (the gradient) can be used to define an edge detector (see e.g. Lindeberg, 1998a).

In applications such as stereo matching, object recognition, and image retrieval the features used are often sparsely distributed point features and are usually denoted salient or interest points. General interest point detectors can be constructed using the probabilistic scale space framework presented by Majer (2000) in which pixel positions are considered particularly informative if the operator response is extremal. The actual type of detected interest point is dependent on the precise choice of the detection operator. As scale invariance is generally regarded as a desirable property, popular salient point detectors employ schemes that rely on some form of automatic scale selection (e.g. Lindeberg, 1998b), which makes the detection invariant to scale changes in the studied image. Classical point detectors, like the Harris corner detector (Harris and Stephens, 1988), have also been combined with scale selection techniques (see e.g. Dorkó and Schmid, 2006; Mikolajczyk and Schmid, 2004).

Loupias et al. (2000) and Sebe and Lew (2003) propose a different approach to the detection of salient points in a multi-resolution manner. A wavelet-based salient point detection scheme is presented that, similar to Lindeberg (1998b,a), but more heuristically and computationally inexpensive, does an over-scale comparison of image features. The point detection strategy is primarily designed to overcome some of the drawbacks of certain classical interest point detectors, i.e. corner detectors. It is demonstrated that a gain in performance can be achieved on, e.g. image retrieval tasks.

Another multi-scale algorithm for the selection of salient regions is introduced by Kadir and Brady (2001). It proposes an improved scale space-based version of an approach described by Gilles (1998), in which saliency is defined in terms of local signal complexity based on the Shannon entropy of local image descriptors. The applicability of this low-level approach is also demonstrated on several image processing problems.

The approach taken by Kadir and Brady (2001) and Gilles (1998) comes close to the probabilistic approach, we suggest here. However, the publications conceptually most closely related to our current work, are those of Loog et al. (2005) and Walker et al. (1997, 1998). In the latter two references certain features (in the pattern recognition sense) are extracted for all points in an image and a probability density is estimated for this collection of feature vectors. Points in the image that correspond to low-density areas are considered interesting and identified as salient points.

A similar notion of saliency is used in the current work where it is integrated into the scale selection framework earlier presented by some of the authors (Loog et al., 2005). The latter presents a probabilistic localised scale selection principle based on maximum likelihood estimation under a Brownian image model. For each image point a generic scale is selected by maximising the likelihood of the image structure under the model. The hereby selected scale is considered as the intrinsic scale of the local image structure. Contrary to other scale selection approaches this method allows us to assign scales to all image positions irrespectively of whether they are image feature points or not.

In this paper, a similar approach is taken, however it is based on the idea that minimum likelihood (contrary to maximum likelihood) indicates that the measured local structure do not fit the underlying stochastic image model well and therefore indicates a point of interest. More specifically, we propose a definition of features and their scales as points in scale space occurring rarely locally under a stochastic model of generic images. The rationale is that image features or salient points are sparsely distributed rare events. This definition allows us to detect points in which the image structure differ from the structure found in the surrounding neighbourhood. As an example one may think of a texture consisting of a dotted pattern in which a box suddenly appears. In this case we would like to detect the centre of the box as the salient point of this pattern. We define locality in terms of the scale of the operators used to describe the point-wise image structure. Using this principle, the method simultaneously detects features and their intrinsic scales.

The proposed feature detection framework allows in general the use of various stochastic image models and descriptions of local image structure. In this paper we use the so-called scale space jet as descriptor of the local structure. The scale space jet is a vector of the partial derivatives of the image up to some order k and can be considered as a representation of a truncated (at order k) Taylor expansion of the image function. Using the jet representation allows us to study image structures at different differential orders. As an example of a stochastic image model we use the fractional Brownian image model as defined by Pedersen (2003) and Markussen et al. (2005). The motivation for using this model is based on the fact that the fractional Brownian image model has been proposed by several authors (Pentland, 1984; Pesquet-Popescu and Vehel, 2002; Pedersen, 2003) as a model of natural images which captures the self-similarity and other important properties of the covariance structure of such images. This image model is a zero mean Gaussian process model and is as such fully described by its covariance structure. The covariance structure for scale space jets of fractional Brownian images has been studied by Pedersen (2003) and Markussen et al. (2005) and an analytical expression for the covariance matrix of scale space jets can be derived. Based on this model we can calculate scale dependent likelihoods. Local minima in this likelihood 'scale space' can then be interpreted as image locations which, in the neighbourhood defined by the scale of the jet measurements, are unlikely and therefore interesting or, for that matter, salient. Contrary to other methods for image feature detection we do not rely on an explicit model of the image features we want to detect.

The apparent discrepancy between maximising (Loog et al., 2005) and minimising (the here presented method) the likelihood in order to select a scale is caused by the difference in goals in the two methods. In Loog et al. (2005) the purpose is to assign a scale independent of an image feature. Here we search for image features which are rare events in space and similarly we define the scale of the feature as a rare change of the image structure across scale.

2. Scale-Space and Scale-Space Jets

In order to describe local image geometry we use the scale space jet representation proposed by Koenderink and van Doorn (1987). The scale space $L : \mathbb{R}^2 \times \mathbb{R}_+ \rightarrow \mathbb{R}$ of an image $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by convolution with a Gaussian blurring kernel

$$L(\mathbf{x}; \sigma) = (G_\sigma * f)(\mathbf{x}) \tag{1}$$

where σ is the measurement scale, $\mathbf{x} = (x, y)^T$, and

$$G_\sigma(\mathbf{x}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma^2}\right) . \tag{2}$$

So-called scale normalised scale space derivatives may be computed by

$$L_{x^n y^m}(\mathbf{x}; \sigma) = \sigma^{n+m} \left(\frac{\partial^{n+m} G_\sigma}{\partial x^n \partial y^m} * f\right)(\mathbf{x}) . \tag{3}$$

The scale space k-jet at an image point is the vector of partial scale space derivatives up to order k of the image intensity function at that point,

$$j_\sigma(\mathbf{x}) = (L_x, L_y, \ldots, L_{x^n y^m})^T (\mathbf{x}; \sigma) \tag{4}$$

where $n+m = k$. The scale space k-jet describes the local differential geometry at position $\mathbf{x}$ measured at scale σ. We disregard the zeroth order term since it does not carry any relevant geometric information except a linear displacement of the intensity. Figure 1 displays examples of a scale space and corresponding derivatives.

3. The Fractional Brownian Image Model

We will assume that images of naturally occurring scenes can be modelled as samples from the fractional Brownian image (fBi) model. This model is a non-stationary zero mean Gaussian random field with stationary increments and long range dependencies. The model was first considered by Mandelbrot and van Ness (1968) and there are several similar definitions of the fractional Brownian image model, see e.g. Pesquet-Popescu and Vehel (2002).

We will use the definition referred to as Levy Brownian motion (Pesquet-Popescu and Vehel, 2002). Following this definition a fractional Brownian image

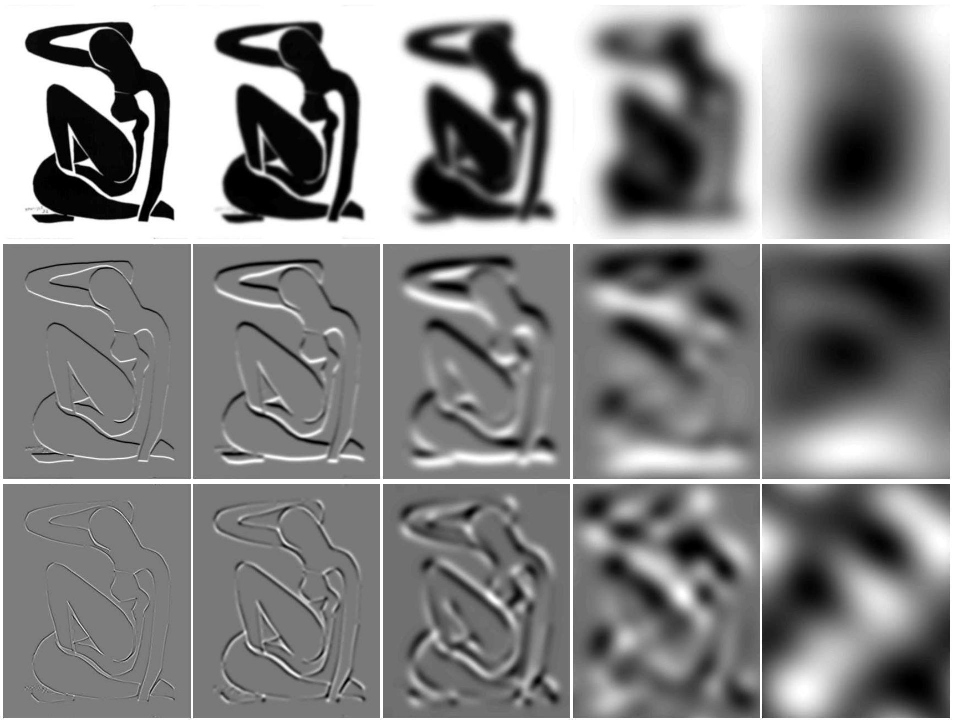

Figure 1: Fauvist scale space for a blue nude (#2) by Matisse (1952). The top row displays normal linear scale space at scales e^0, e^1,... e^4. Middle and last row show the corresponding derivatives L_y and L_{xy}, respectively, in which the same scales are used as for the first row. Note that for every image the intensities have been scaled linearly so to span the whole intensity range.

is a non-stationary zero mean Gaussian random field $B_H(\mathbf{x})$ with stationary increments $B_H(\mathbf{x}) - B_H(\mathbf{y})$. The correlation between two points $B_H(\mathbf{x})$ and $B_H(\mathbf{y})$ has the form

$$R_{B_H}(\mathbf{x}, \mathbf{y}) \propto \frac{\beta}{2} \left(\|\mathbf{x}\|^{2H} + \|\mathbf{y}\|^{2H} - \|\mathbf{x} - \mathbf{y}\|^{2H} \right) \ , \tag{5}$$

where H is known as the Hurst coefficient and can take the values $0 < H < 1$. This parameter dictates the spatial correlation structure of the model as well as its scaling behaviour ranging from self-similarity to true scale invariance. For $H = \frac{1}{2}$ the process $B_H(\mathbf{x})$ corresponds to Brownian motion. The β constant describes the variance of the intensities. The increment process $B_H(\mathbf{x}) - B_H(\mathbf{y})$ is a zero mean Gaussian process with variance proportional to $\|\mathbf{x} - \mathbf{y}\|^{2H}$. The increments are in general dependent, but independent only for $H = \frac{1}{2}$.

Even though the fBi model is non-stationary it is possible to define a power spectrum of the fBi in terms of the spectral properties of the stationary increments. Reed et al. (1995) does this by constructing a generalised power spectrum and Flandrin (1989) and Hoefer et al. (1993) uses the Wigner-Ville spectrum. By using such techniques it is possible to derive a meaningful power spectrum of the fBi and it is proportional to $|\omega|^{-\alpha}$, where $\alpha = 2H + 1$ and $1 < \alpha < 3$. Using the form of the power spectrum the covariance matrix of scale normalised jets of fractional Brownian images can be computed analytically (see derivation in Pedersen, 2003). In the following $n_1 m_1$ and $n_2 m_2$ indicates the derivation order for both filters. The covariance is

$$\Sigma_{nm}^{\{\sigma,\alpha\}} = (-1)^{\frac{n+m}{2}+n_2+m_2} \frac{\beta(n-1)!!(m-1)!!}{4\pi\sigma^{2-\alpha}(n+m)!!} \Gamma\left(\frac{n+m-\alpha}{2}+1\right) \tag{6}$$

whenever both $n = n_1 + n_2$ and $m = m_1 + m_2$ are even integers, otherwise $\Sigma_{nm}^{\{\sigma,\alpha\}} = 0$. Double factorial is defined as $n!! = n(n-2)(n-4)\cdots$. Choosing $H = \frac{1}{2}$ (implying $\alpha = 2$), leads to the scale invariant Brownian image model which can be seen by the fact that the scale dependence of the covariance matrix $\Sigma^{\{\sigma,\alpha\}}$ given in (6) vanishes.

Since the fractional Brownian image is a zero mean Gaussian process and due to the linearity of the scale-space derivative filters, we have that the probability density on jets must be zero mean Gaussian

$$p_\alpha(j_\sigma(\mathbf{x})) = \frac{1}{Z} \exp\left(-\frac{1}{2} j_\sigma^T(\mathbf{x})(\Sigma^{\{\sigma,\alpha\}})^{-1} j_\sigma(\mathbf{x})\right) \tag{7}$$

where Z is a normalisation constant and the covariance $\Sigma^{\{\sigma,\alpha\}}$ is given by (6).

4. Detecting Salient Points and Scales

As stated in the introduction, we define salient points and their intrinsic scale as points in scale space of locally minimal probability of occurrence under the fractional Brownian image model, i.e. a probabilistic feature detector based on local minimum likelihood.

Our method can be summarised as follows: For a particular choice of α, we find points in scale space $(\hat{x}, \hat{y}, \hat{\sigma})$ which locally minimise the likelihood $p_\alpha(j_\sigma(\mathbf{x}))$ under the model given by (7). In practise, this is done by comparison of the probability at the current position (x, y, σ) with its six nearest neighbours in the sampled scale-space.

The probabilities at the detected points $p_\alpha(j_{\hat{\sigma}}(\hat{\mathbf{x}}))$ give us a measure of saliency by which we can order the detected points.

We find it interesting to note that minimising the likelihood bears some resemblance to the well-known feature and scale detection method proposed by Lindeberg (1998b). Lindeberg maximises so-called measures of feature strength which are polynomials of image derivatives. In our setting, $-\frac{1}{2} j_\sigma^T (\Sigma^{\{\sigma,\alpha\}})^{-1} j_\sigma$ correspond to such a feature strength, and maximising this measure is equivalent to minimising the likelihood.

5. Illustrative Results

As a demonstration of our method, we include examples of detecting features on a photo of a painting by Matisse (1952) and four images (see Figures 1 and

Figure 2: Four images used in the following demonstration. These are reference images taken from the Affine Covariant Features image database (ACF, 2006).

2) taken from the Affine Covariant Features image database (ACF, 2006) which has been used in various evaluation studies of salient point and region detectors (e.g. Mikolajczyk and Schmid, 2004, 2005). The free parameters of the method are α and the jet order k. The β constant is estimated for each image as the variance of the intensities in the image. We applied our method to these images using jet orders $k = 1, 2, 3, 4$, model $\alpha = 1.5, 2, 2.5$, and logarithmically sampled scales σ in the range 0.6 to 54. We also include examples of ranking the detected points after the saliency measure $p_\alpha(j_{\hat{\sigma}}(\hat{\mathbf{x}}))$.

Figures 3 and 4 show detected points on the Matisse painting including respectively all detected points and the 10% lowest probability points. Figure 5 shows results on an image of a wall with graffiti including the 30% lowest probability points. Finally to show the variation for different images Figure 6 show results on the four images from Figure 2. Here the model is $\alpha = 2$ and the jet order $k = 2$ and we show the 20% lowest probability points.

As one would expect the $k = 1$ jet seem to detect edges for $\alpha \leq 2$. For the $k = 2$ jet the method seems to detect corners, blobs and ridges for any α. The $k = 3$ and $k = 4$ jets seem to detect combinations of edges, ridges, corners, and blobs and other higher order structures.

The scales being selected changes with the choice of α. There is a general trend that for $\alpha > 2$ our method selects mostly small scales whereas for $\alpha < 2$ the method prefers large scales. This corresponds well with the fact that for $\alpha > 2$ samples from the fractional Brownian image model will consist mostly of large scale structures making small scale structures unlikely events. Similarly, for $\alpha < 2$ samples from the model will in general include mostly small scale structures, hence large scale structures will become unlikely and will therefore be preferred by our method.

As mentioned in the introduction the most popular salient point detectors are based on some form of corner detectors. Our method clearly finds other types of points. Looking at the results for $k = 2$ and $\alpha = 2$ we see that, besides corners, we also detect blobs at scales comparable to the underlying image structure. These points do indeed represent salient structure if we look at higher scales. However, a thorough evaluation and comparison of our method with other detectors is needed in order to judge the quality of the detected points. Such an evaluation could follow the method used by both Schmid et al. (2000) and Mikolajczyk and Schmid (2005). This, however, is not a trivial task and we consider it outside the scope of this workshop publication and leave it for future work.

6. Conclusion

We proposed a novel method for detecting salient points in images and their intrinsic scales. This was done by finding local minima in scale space of the likelihood of the local image structure under the fractional Brownian image model. In principle our method performs a variable threshold outlier detection, where the threshold is determined based on the neighbouring image structure. We presented promising illustrative results demonstrating the method. The next step is a thorough evaluation on a large set of natural images. Such an evaluation could follow the method used by both Schmid et al. (2000) and Mikolajczyk and Schmid (2005).

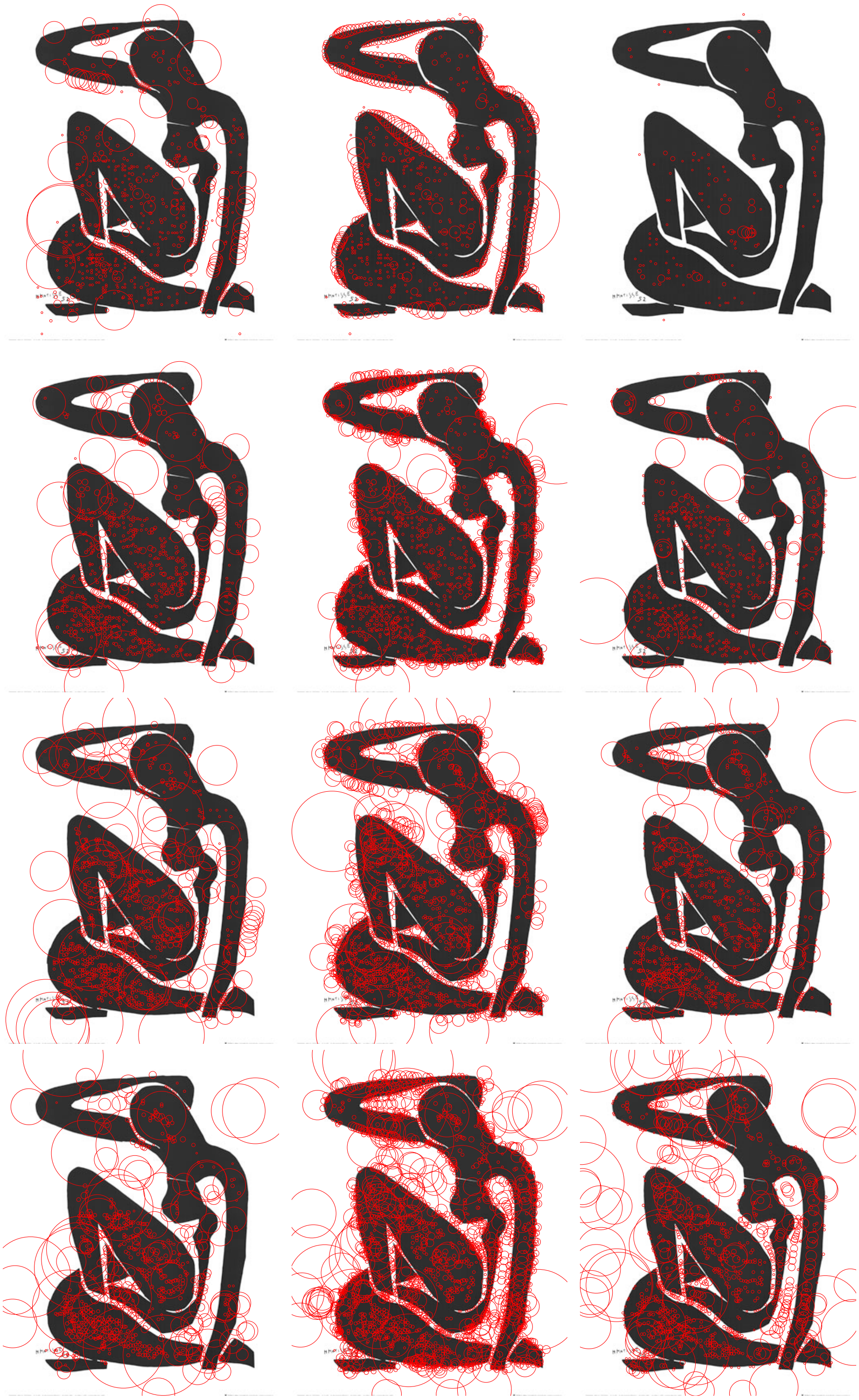

Figure 3: First row shows results on a painting by Matisse (1952) for jet order $k = 1$ and $\alpha = 1.5, 2, 2.5$. The other rows show similar results for jet order $k = 2, 3, 4$. The circles are centred on the detected points and the radii correspond to the selected scales.

Figure 4: First row shows the 10% lowest probability points for jet order $k = 1$ and $\alpha = 1.5, 2, 2.5$. The other rows show similar results for jet order $k = 2, 3, 4$. The circles are centred on the detected points and the radii correspond to the selected scales.

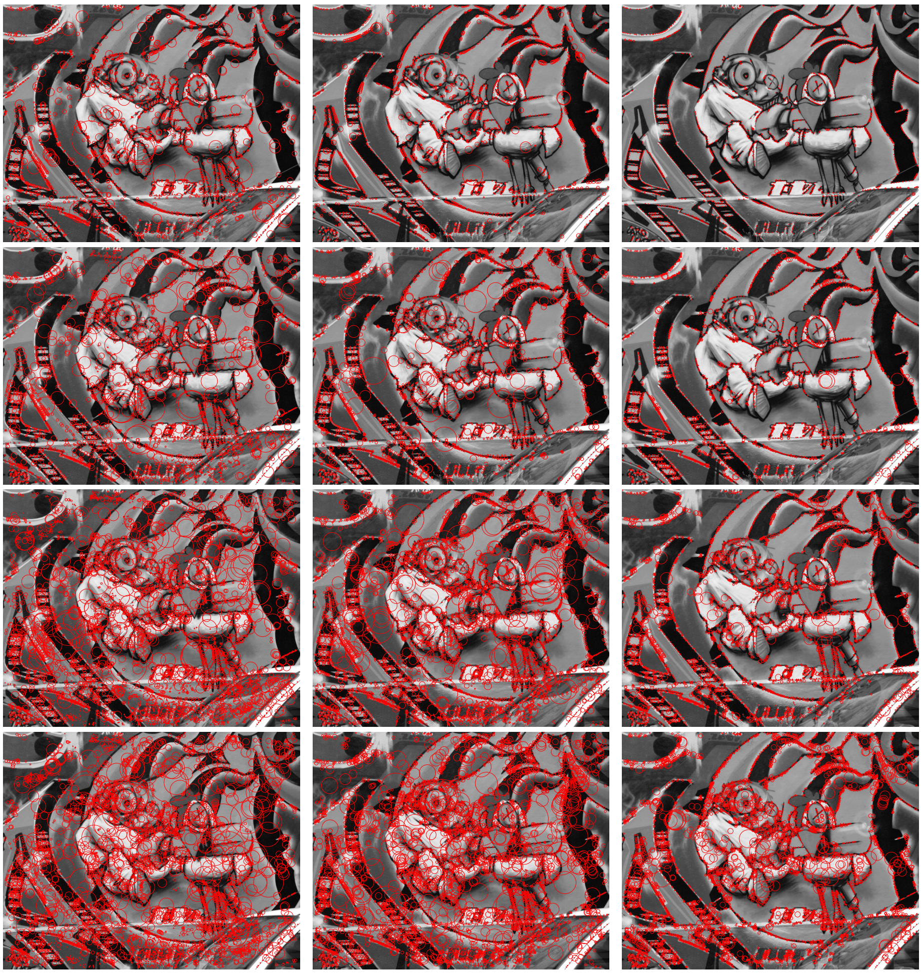

Figure 5: First row shows the 30% lowest probability points for jet order $k = 1$ and $\alpha = 1.5, 2, 2.5$. The other rows show similar results for jet order $k = 2, 3, 4$. The circles are centred on the detected points and the radii correspond to the selected scales.

Figure 6: Each image shows the 20% lowest probability points detected using the model $\alpha = 2$ and the jet order $k = 2$. The circles are centred on the detected points and the radii correspond to the selected scales.

References

ACF. Affine covariant features image database. Web: http://www.robots.ox.ac.uk/~vgg/research/affine/, 2006.

G. Dorkó and C. Schmid. Maximally stable local description for scale selection. In *Proceedings of the European Conference on Computer Vision*, pages 57–69. Springer, 2006.

P. Flandrin. On the spectrum of fractional brownian motions. *IEEE Transactions on Information Theory*, 35(1):197–199, January 1989.

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

C. Harris and M. Stephens. A combined corner and edge detector. In *Fourth Alvey Vision Conference*, pages 147–151, 1988.

S. Hoefer, F. Heil, M. Pandit, and R. Kumaresan. Segmentation of textures with different roughness using the model of isotropic two-dimensional fractional brownian motion. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1993. ICASSP-93*, volume 5, pages 53–56, 1993.

T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

J. J. Koenderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–154, November 1998a.

T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, November 1998b.

M. Loog, K. S. Pedersen, and B. Markussen. Maximum likely scale estimation. In *Proceedings of International Workshop on Deep Structure, Singularities and Computer Vision*, June 2005.

E. Loupias, N. Sebe, S. Bres, and J.M. Jolion. Wavelet-based salient points for image retrieval. In *Proceedings of the International Conference on Image Processing*, volume 2, pages 518–521, 2000.

P. Majer. *A Statistical Approach to Feature Detection and Scale Selection in Images*. PhD thesis, University of Göttingen, 2000.

B. B. Mandelbrot and J. W. van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, October 1968.

B. Markussen, K. S. Pedersen, and M. Loog. A scale invariant covariance structure on jet space. In *Proceedings of International Workshop on Deep Structure, Singularities and Computer Vision*, June 2005.

D. Marr. *Vision*. W. H. Freeman, New York, 1982.

H. Matisse. Blue nude (#2). Musee National d'Art Moderne, 1952. Gouache on paper, cut and pasted on white paper, 116×82cm.

K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10): 1615–1630, 2005.

K. S. Pedersen. Properties of brownian image models in scale-space. In *Proceedings of the 4th Scale-Space conference*, LNCS 2695, pages 281–296. Springer, 2003.

A. P. Pentland. Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):661–674, November 1984.

B. Pesquet-Popescu and J. L. Vehel. Stochastic fractal models for image processing. *IEEE Signal Processing Magazine*, 19(5):48–62, September 2002.

I. S. Reed, P. C. Lee, and T. K. Truong. Spectral representation of fractional brownian motion in n dimensions and its properties. *IEEE Transactions on Information Theory*, 41(5):1439–1451, September 1995.

C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(4):151–172, 2000.

N. Sebe and M.S. Lew. Comparing salient point detectors. *Pattern Recognition Letters*, 24:89–96, 2003.

K.N. Walker, T.F. Cootes, and C.J. Taylor. Correspondence using distinct points based on image invariants. *Proceedings of the 8th British Machine Vision Conference*, pages 540–549, 1997.

K.N. Walker, T.F. Cootes, and C.J. Taylor. Locating salient object features. *Proceedings of the 9th British Machine Vision Conference*, 2:557–566, 1998.

A. P. Witkin. Scale space filtering. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, volume 2, pages 1019–1023, Karlsruhe, Germany, 1983.