

# A multi-scale feature likelihood map for direct evaluation of object hypotheses<sup>\* \*\*</sup>

Ivan Laptev and Tony Lindeberg

Computational Vision and Active Perception Laboratory (CVAP)  
Dept. of Numerical Analysis and Computing Science  
KTH, S-100 44 Stockholm, Sweden

**Abstract.** This paper develops and investigates a new approach for evaluating feature based object hypotheses in a *direct* way. The idea is to compute a feature likelihood map (FLM), which is a function normalized to the interval  $[0, 1]$ , and which approximates the likelihood of image features at all points in scale-space. In our case, the FLM is defined from Gaussian derivative operators and in such a way that it assumes its strongest responses near the centers of symmetric blob-like or elongated ridge-like structures and at scales that reflect the size of these structures in the image domain. While the FLM inherits several advantages of feature based image representations, it also (i) avoids the need for explicit search when matching features in object models to image data, and (ii) eliminates the need for thresholds present in most traditional feature based approaches. In an application presented in this paper, the FLM is applied to simultaneous tracking and recognition of hand models based on particle filtering. The experiments demonstrate the feasibility of the approach, and that real time performance can be obtained by a pyramid implementation of the proposed concept.

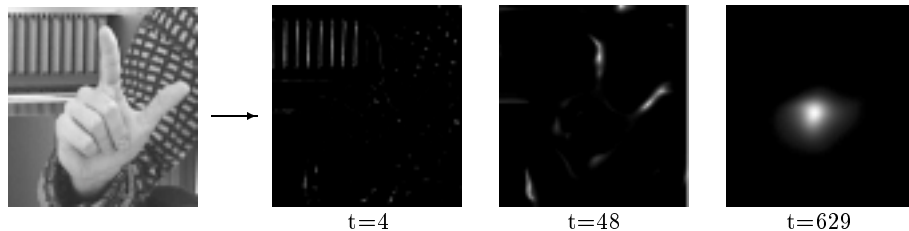
## 1 Introduction

When interpreting image data, the purpose of filtering is to emphasize and abstract relevant properties in the data while suppressing others. Common approaches for computing image descriptors involve either (i) the computation of sparse sets of image features (feature detection) or (ii) the computation of dense maps of filter responses (direct methods).

In this respect, a main strength of feature based approaches is that they provide an abstracted and compact description of the local image shape. Image features are usually invariant to absolute intensity values and can selectively represent characteristic visual properties of image patterns. In particular, using multi-scale feature detection it is possible to estimate the size of image structures

\* The support from the Swedish Research Council for Engineering Sciences, TFR, and from the Royal Swedish Academy of Sciences as well as the Knut and Alice Wallenberg Foundation is gratefully acknowledged.

\*\* Shortened version in IEEE Workshop on Scale-Space and Morphology, Vancouver, Canada, July 2001, M. Kerckhove (Ed.), Volume 2106 of Springer Verlag Lecture Notes in Computer Science, pages 98–110.



**Fig. 1.** The result of computing the proposed feature likelihood map on an image of a hand. At any point  $(x, y, t)$  in scale-space, this function approximates the likelihood of symmetric blob-like or ridge-like image structures. In this figure, the feature likelihood map is shown for three scale levels  $t = 4, 48, 629$ , which (here) are characteristic scales for the background, the fingers and the palm of a hand, respectively. Note that the response of the map is well localized in space and scale and that the response is invariant to the local amplitude of the image structures.

and to represent image patterns in a scale-invariant manner. Relations between features in terms of positions, scales, types and other attributes may then be effectively used for recognizing objects and image patterns.

The use of features for image representation, however, also has drawbacks. One is that image features may depend on thresholds used for separating relevant image structures from noise. This may make the results unstable for patterns with low contrast. Another disadvantage is that algorithms involving matching of sparse points in image space usually lead to combinatorial complexity.

The aim of this paper is to develop a *dense* image representation, which preserves the advantages of feature-based representation, while avoiding the problems of local thresholding and selection of sparse image features for matching. The idea is to compute a function on a multi-scale feature space, which is normalized to the interval  $[0, 1]$  and thus independent of the local contrast of the grey-level pattern. Moreover, the function will be defined in such a way that its response is localized in space and scale, with the strongest responses near the centers of blob-like and ridge-like structures. The proposed function, referred to as a *feature likelihood map*, can be used for approximating the likelihood of image features. Figure 1 illustrates this concept for an image of a hand.

A main reason behind this construction is to provide means for *direct* verification of feature-based object hypotheses. Given a hypothesis, the verification on such a map does not require explicit search and will therefore be highly efficient. In particular, this approach is convenient for object tracking and object recognition based on the recently developed approach of particle filtering. In this paper, the feature likelihood map will indeed be used for *simultaneous* hand tracking and hand recognition. The viability of the approach will be demonstrated with a pyramid implementation, which gives real time performance.

## 2 The feature likelihood map

The aim of the proposed likelihood map is to emphasize specific structures in the image domain, and to localize them in space and scale. To study this problem, we initially restrict ourselves to symmetric blob-like and elongated ridge-like image features. The general ideas behind this construction, however, are more general and apply to many other aspects of local image structures.

A general requirement on the proposed feature likelihood map,  $\mathcal{M} : \mathbb{R}^2 \times \mathbb{R}_+ \mapsto \mathbb{R}$ , is that for a blob of size  $t_0$  located at a point  $(x_0, y_0)$  in space,  $\mathcal{M}$  should satisfy the following basic properties: (i)  $\mathcal{M}$  should assume its maximum value one at  $(x_0, y_0; t_0)$ , (ii)  $\mathcal{M}$  should assume high values in a small neighborhood of  $(x_0, y_0; t_0)$ , and (iii)  $\mathcal{M}$  should decrease monotonically towards zero elsewhere. Additionally,  $\mathcal{M}$  should not give preference to blobs of any particular size, position or amplitude, and should thus be invariant to scalings and translations in the image as well as local changes of the image contrast.

### 2.1 Scale-space representation

For any continuous signal  $f : \mathbb{R}^D \mapsto \mathbb{R}$ , the linear scale-space representation  $L : \mathbb{R}^D \times \mathbb{R}_+ \mapsto \mathbb{R}$  is defined as the convolution of  $f$  with Gaussian kernels  $g$

$$L(\cdot; t) = g(\cdot; t) * f(\cdot), \quad (1)$$

where  $g(x; t) = \exp(-(x_1^2 + \dots + x_D^2)/2t)/(2\pi t)^{D/2}$ , and  $x = (x_1, \dots, x_D)^T$ . One reason for considering such a representation is that the Gaussian derivatives

$$L_{x^m}(\cdot; t) = \partial_{x^m}(g * f) = (\partial_{x^m} g) * f = g * (\partial_{x^m} f) \quad (2)$$

(where  $m$  denotes the order of differentiation) constitute a canonical set of filter kernels given natural symmetry requirements on a visual front-end (Witkin 1983, Koenderink and van Doorn 1992, Lindeberg 1994, Florack 1997). Another reason is that the evolution over scales of a signal and its Gaussian derivatives provides important cues to local image structure. One such property, which we will make particular use of here, is based on the behavior over scales of  $\gamma$ -normalized Gaussian derivative operators (Lindeberg 1998)

$$L_{\xi^m, \gamma\text{-norm}}(\xi; t) = t^{m\gamma/2} L_{x^m}(x; t). \quad (3)$$

where  $\xi = x/t^{\gamma/2}$  denotes  $\gamma$ -normalized coordinates. It can be shown both theoretically and experimentally that the scales at which such normalized differential entities assume local maxima over scales reflect characteristic scales of local image patterns and can thus be used for, for example, local size estimation.

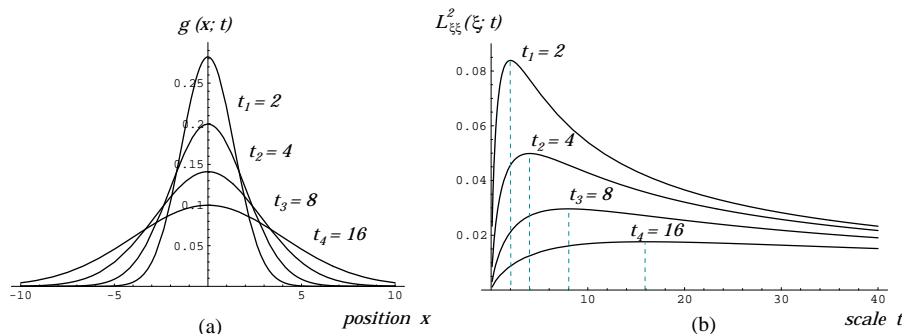
### 2.2 Likelihood map in the 1-D case

When we construct the feature likelihood map, let us first consider the one-dimensional case and take a Gaussian function  $f(x) = g(x; x_0, t_0)$  as a prototype

for a blob of size  $t_0$  centered at  $x_0$  (see Figure 2(a)). Using the semi-group property of the Gaussian kernel, it follows that the scale-space representation of  $f$  is  $L(x; t) = g(x; x_0, t + t_0)$ , and its  $\gamma$ -normalized second-order derivative:

$$L_{\xi\xi}(\xi; t) = t^{\gamma_2} L_{xx}(x; t) = -\frac{t^{\gamma_2}(t + t_0 + (x - x_0)^2)}{\sqrt{2\pi}(t + t_0)^5} e^{-\frac{(x - x_0)^2}{2(t + t_0)}}. \quad (4)$$

If we choose  $\gamma_2 = 3/4$ , then it can be shown (Lindeberg 1998) that  $L_{\xi\xi}$  assumes a local extremum over space and scale at the point  $(x_0, t_0)$  in scale-space that corresponds to the position  $x_0$  and the size  $t_0$  of the original blob  $f$ . Thus,  $L_{\xi\xi}^2$  satisfies some of the required properties of the desired likelihood map  $\mathcal{M}$ , however,  $L_{\xi\xi}^2$  is not invariant to the local amplitude of the signal (see Figure 2(b)).



**Fig. 2.** (a): Gaussian kernels of various widths; (b): Evolution over scales of the second order normalized derivative operator  $L_{\xi\xi}^2$  in the case when  $\gamma_2 = 3/4$ .

*Quasi-quadrature.* A standard approach for amplitude estimation in signal processing is in terms of quadrature filter pairs  $(h_+, h_-)$ , from which the amplitude can be estimated as  $Q = (h_+ * f)^2 + (h_- * f)^2$ . Strictly, a quadrature filter pair is defined from a Hilbert transform, in such a way that  $Q$  is phase-independent. Within the framework of scale-space derivatives, the quadrature entity  $Q$  for first- and second-order derivatives can be approximated by a pair of normalized first- and second-order Gaussian derivative operators (Lindeberg 1998):

$$Q_1 L = A L_{\xi}^2 + L_{\xi\xi}^2 = A t^{\gamma_1} L_x^2 + t^{\gamma_2} L_{xx}^2. \quad (5)$$

where  $A$  is a constant and  $L_{\xi} = t^{\gamma_1/2} L_x(x; t)$  is the normalized first-order derivative operator, where we choose  $\gamma_1 = 1/2$  to match  $\gamma_2 = 3/4$ . Moreover, the value of  $A$  can be chosen to  $A \approx 4/e$ , such that the response of the  $Q_1 L$  is approximately constant over space in the neighborhood of  $(x_0; t_0)$  (Lindeberg 1998). This quadrature entity is, however, not phase-independent over scales.

*Including stability over scales.* To include the stability of image structures over scales (corresponding to low values of derivatives with respect to scale), and to also increase approximate phase invariance with respect to space and scale

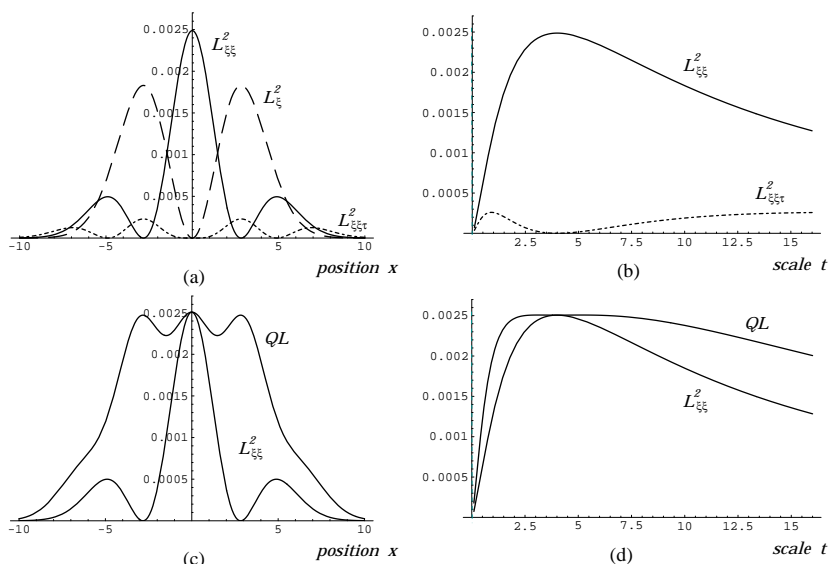
simultaneously, we propose to include the derivative of  $L_{\xi\xi}$  with respect to effective scale  $\tau = \log t$ . Using  $\partial_\tau = t \partial_t$  and the fact that all Gaussian derivatives satisfy the diffusion equation  $\partial_t(L_{x^\alpha}) = 1/2 \partial_{xx}(L_{x^\alpha})$ , it follows that:

$$L_{\xi\xi\xi\xi}(\xi; t) = t \partial_t L_{\xi\xi\xi\xi}(\xi; t) = \gamma_3 t^{\gamma_3} L_{xx} + t^{\gamma_3} L_{xxxt} = \gamma_3 t^{\gamma_3} L_{xx} + \frac{t^{\gamma_3+1}}{2} L_{xxxx}. \quad (6)$$

By adding this expression to (5), we thus propose to extend  $Q_1 L$  into

$$Q_2 L = A L_\xi^2 + B L_{\xi\xi\xi\xi}^2 + L_{\xi\xi}^2. \quad (7)$$

Figures 3(a) and (b) illustrate the evolution of the components in this expression, i.e.  $L_\xi^2$ ,  $L_{\xi\xi\xi\xi}^2$  and  $L_{\xi\xi}^2$ , over space and scale. As can be seen, the responses of  $L_\xi^2$  and  $L_{\xi\xi\xi\xi}^2$  complement the response of  $L_{\xi\xi}^2$  by assuming high values where  $L_{\xi\xi}^2$  is low and vice versa. Thus, one can expect that by an appropriate choice of the weights  $A$  and  $B$ ,  $Q_2 L$  will approximately be constant in a neighborhood of  $(x_0, t_0)$ . Such a behavior is illustrated in Figures 3(c) and (d).



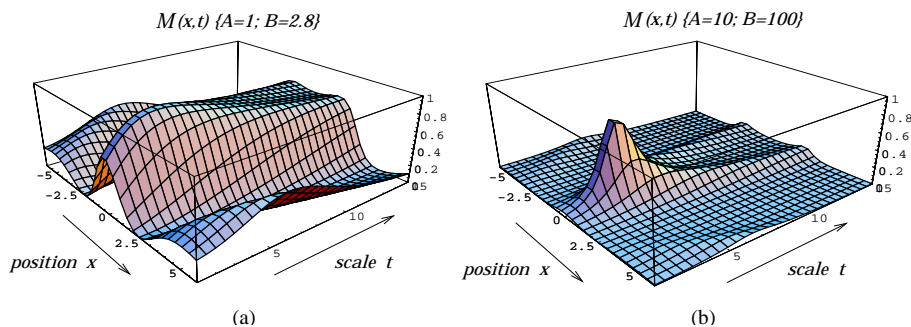
**Fig. 3.** (a)-(b): Evolution of  $L_\xi^2$ ,  $L_{\xi\xi\xi\xi}^2$  and  $L_{\xi\xi}^2$  over space and scale when applied to a Gaussian blob centered at  $x_0 = 0$  and with variance  $t_0 = 4$ ; (c)-(d): Evolution of  $L_{\xi\xi}^2$  and  $Q_2 L$  when using the parameter values  $A = 1$  and  $B = 2.8$ . Note that  $Q_2 L$  is approximately constant over space and scale in the neighborhood of  $(x_0, t_0)$ .

*Invariance properties.* If we consider the ratio  $L_{\xi\xi}^2/Q_2 L$ , it is apparent that the amplitude cancels between the numerator and the denominator. Thus, we achieve local contrast invariance. Moreover, since  $Q_2 L \geq L_{\xi\xi}^2 \geq 0$ , it follows that the ratio  $L_{\xi\xi}^2/Q_2 L$  will always be in the range  $[0, 1]$ . Scale invariance of  $Q_2$  holds if we for  $\gamma_2 = 3/4$  take  $\gamma_1 = 1/2$ . Moreover, it can be shown that for a Gaussian blob, the scale-space maximum of  $Q_2 L$  is assumed at  $t_0$  if and only if  $\gamma_3 = 1$ . The relative magnitudes of  $L_{\xi\xi}^2$  and  $Q_2 L$  are illustrated in Figures 3(c) and (d).

To conclude, the ratio  $L_{\xi\xi}^2/Q_2L$  satisfies all the stated requirements on the feature likelihood map, and we define

$$\mathcal{M} = \frac{L_{\xi\xi}^2}{Q_2L} = \frac{L_{\xi\xi}^2}{AL_{\xi}^2 + BL_{\xi\xi\tau}^2 + L_{\xi\xi}^2}. \quad (8)$$

*Determination of the free parameters A and B.* Concerning the choice of  $A$  and  $B$ , it can be verified that  $A \approx 1$  and  $B \approx 3$  give an approximately constant behavior of the denominator of  $\mathcal{M}$  around  $(x_0; t_0)$ . This was the original design criterion when the quasi quadrature entity (5) was proposed. Figure 4(a) shows the behavior of  $\mathcal{M}$  in this case. Notably, the peak around  $(x_0; t_0)$  is rather wide in the scale direction, and there are two quite strong side lobes in the spatial direction. For the purpose of dense scale selection with application to recognition, it is desirable to have a more narrow and localized response with respect to scale and space. For this reason, we increase the parameters to  $A = 10$  and  $B = 100$  and obtain a desired behavior of  $\mathcal{M}$  as illustrated in Figure 4(b).



**Fig. 4.** Evolution of the likelihood map  $\mathcal{M}$  over space and scale for the different values of parameters  $A$  and  $B$  when applied to a Gaussian blob ( $x_0 = 0, t_0 = 4$ ).

### 2.3 Likelihood map in the 2-D case

The likelihood map defined in (8) can be easily extended to two dimensions. Consider again a Gaussian kernel  $f = g(x, y; x_0, y_0, t_0)$  as a prototype image blob of size  $t_0$  centered at  $(x_0, y_0)$ . The scale-space representation of this signal is given by  $L(x, y; t) = g(x, y; x_0, y_0, t+t_0)$  and the normalized Laplacian operator

$$\nabla_{norm}^2 L = L_{\xi\xi} + L_{\eta\eta} = t^{\gamma_2} L_{xx}(x, y; t) + t^{\gamma_2} L_{yy}(x, y; t) \quad (9)$$

assumes a local extremum at  $(x_0, y_0, t_0)$  if  $\gamma_2 = 1$ . To construct a quadrature entity  $Q$ , we consider the gradient magnitude (with  $\gamma_1 = 1$ )

$$|\nabla_{norm} L| = \sqrt{L_{\xi}^2 + L_{\eta}^2} = t^{\gamma_1/2} \sqrt{L_x^2 + L_y^2}, \quad (10)$$

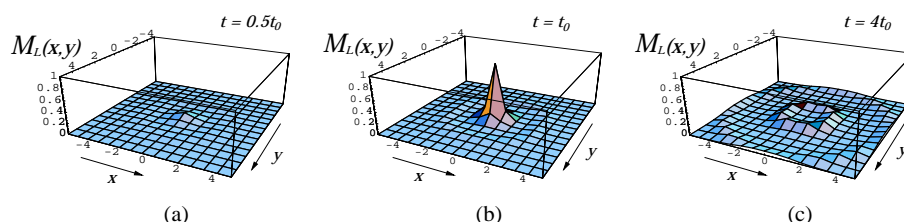
as the analogue to  $L_{\xi}$  in the one-dimensional case, and take

$$\partial_{\tau}(\nabla_{norm}^2 L) = L_{\xi\xi\tau} + L_{\eta\eta\tau} = \gamma_3 t^{\gamma_3} (L_{xx} + L_{yy}) + \frac{t^{\gamma_3+1}}{2} (L_{xxxx} + L_{yyyy} + 2L_{xxyy})$$

as the analogue to  $L_{\xi\xi\tau}$ . Then, we define the feature likelihood map as

$$\mathcal{M}_L = \frac{(L_{\xi\xi} + L_{\eta\eta})^2}{A(L_{\xi}^2 + L_{\eta}^2) + B(L_{\xi\xi\tau} + L_{\eta\eta\tau})^2 + (L_{\xi\xi} + L_{\eta\eta})^2}. \quad (11)$$

Clearly,  $\mathcal{M}_L$  is rotationally invariant, and invariant with respect to scale and local contrast; it assumes values in the range  $[0, 1]$  and for a Gaussian blob the maximum value 1 is assumed at  $(x_0, y_0; t_0)$ . Hence,  $\mathcal{M}_L$  has essentially similar properties as the likelihood map (8) in the one-dimensional case. Figures 5(a)-(c) illustrate how, with  $A = 10$  and  $B = 100$ ,  $\mathcal{M}_L$  assumes a rather sharp maximum at  $(x_0, y_0, t_0)$  and rapidly decreases with deviations from this point.



**Fig. 5.** Evolution of the likelihood map  $\mathcal{M}_L$  over space and scale for a two-dimensional Gaussian blob defined by  $(x_0 = 0, y_0 = 0, t_0 = 1)$ . Plots in (a),(b) and (c) illustrate  $\mathcal{M}_L$  for scale values  $t = 0.5, 1$  and  $4$ .

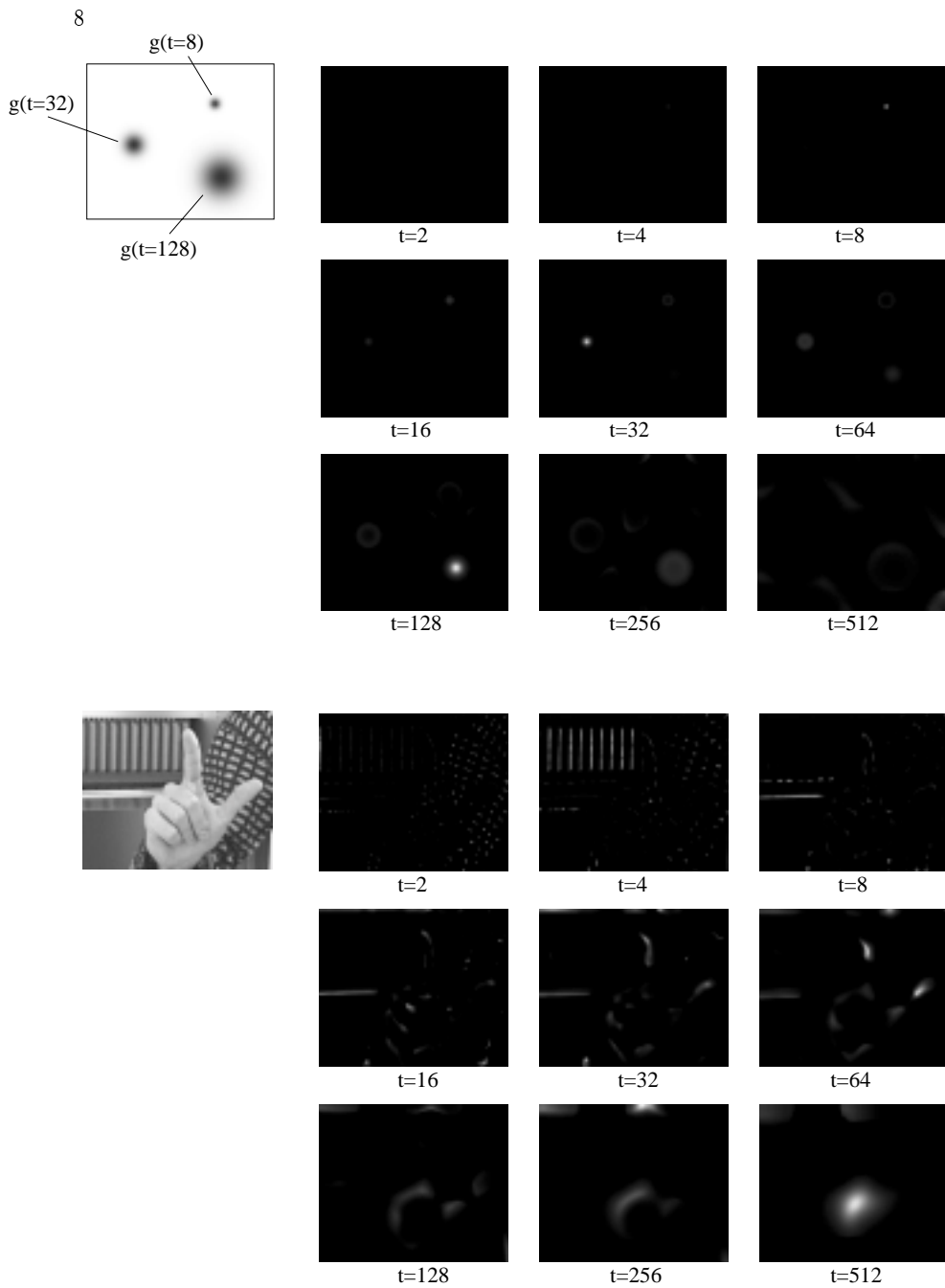
*Suppression of saddle regions and noise.* Besides blobs and ridges, however,  $\mathcal{M}_L$  will also respond to certain saddle points. This occurs when  $\nabla_{norm} L = 0$  and  $\partial_\tau(\nabla_{norm}^2 L) = 0$ . To suppress such points, introduce a saddle suppression factor

$$\mu = \frac{\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2}{\lambda_1^2 + \lambda_2^2 + 2|\lambda_1\lambda_2|} = \frac{L_{\xi\xi}^2 + L_{\eta\eta}^2 + 2L_{\xi\xi}L_{\eta\eta}}{L_{\xi\xi}^2 + L_{\eta\eta}^2 - 2L_{\xi\eta}^2 + 2|L_{\xi\xi}L_{\eta\eta} - L_{\xi\eta}^2|}, \quad (12)$$

where  $\lambda_1$  and  $\lambda_2$  denote the eigenvalues of the Hessian matrix. Then, it can be seen that  $\mu$  is equal to one when  $\lambda_1$  and  $\lambda_2$  have the same sign (i.e., for emphasized blob and ridge structures), while  $\mu$  decreases towards zero if  $\lambda_1$  and  $\lambda_2$  have equal magnitude and opposite sign. Moreover, to suppress the influence of spurious noise structures of amplitude lower than  $\varepsilon_N$ , we introduce a small normalising parameter  $\varepsilon_N$  in the denominator of the expression for the FLM. Thus, we define a saddle- and noise-suppressed feature likelihood map as

$$\tilde{\mathcal{M}}_L = \mu^k \mathcal{M}_L = \frac{\mu^k (L_{\xi\xi} + L_{\eta\eta})^2}{A(L_{\xi}^2 + L_{\eta}^2) + B(L_{\xi\xi\tau} + L_{\eta\eta\tau})^2 + (L_{\xi\xi} + L_{\eta\eta})^2 + \varepsilon_N^2}. \quad (13)$$

Examples of other feature likelihood maps, with exclusive emphasis on specific types of image structures are presented in (Lindeberg 2001).



**Fig. 6.** The result of computing feature likelihood map on a synthetic image and a real image, where in the second case, the response of the FLM has been set to zero for points with  $\nabla^2 L > 0$ , in order to enhance the response to bright image structures. From the first image, it can be verified that the FLM gives a correct localization of blobs in space and scale. In the second image, it can be seen that the FLM clearly separates different image structures, such as the fingers and the palm of a hand, according to their size.



## 2.4 Experiments on synthetic and real data

Figure 6 shows the result of computing this feature likelihood map for a synthetic image with three Gaussian blobs. As can be observed, the high values of  $\mathcal{M}$  are well localized in space and scale, and the peaks over space and scale correspond to the positions and the sizes of the original blobs. Figure 6 shows the result of computing  $\mathcal{M}$  for an image of a hand. Here, it can be seen that  $\mathcal{M}$  responds not only to circular structures but also to elongated ridge-like structures, such as fingers. The reason for this is that the Laplacian operator, besides responding to circular blob-like structures, also gives a reasonably high response to elongated structures. From these results it can be clearly seen how  $\mathcal{M}$  separates between small structures in the background, the fingers and the palm of a hand. Moreover, despite the varying contrast of the image structures,  $\mathcal{M}$  gives equally high response to weak ridges in the background and to the fingers of higher contrast. In many cases, this is a desirable property of a recognition system aimed at classifying local image structures irrespective of illumination variations.

## 3 Hand tracking and recognition

To experimentally investigate the proposed direct approach for evaluation of feature hypotheses, we will in this section present an application of the feature likelihood map in combination with particle filtering for simultaneous tracking and recognition of hands in image sequences. By necessity the presentation is heavily condensed; more details can be found in (Laptev and Lindeberg 2001).

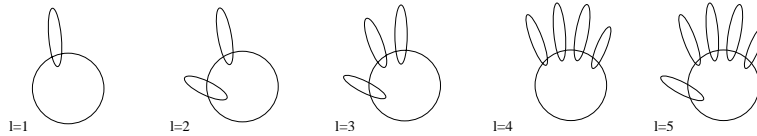
### 3.1 Hand model

An image of a hand can be expected to give rise to blob and ridge features corresponding the fingers and the palm of a hand. These image structures together with information about their relative orientation, position and scale can be used for defining a simple but discriminative, view-based model of a hand. Thus, we represent a hand by a set of blob and ridge features as illustrated in Figure 7, and define different hand states, depending on the number of open fingers.

To model translations, rotations and scalings of hands, we define a parameter vector  $X = (x, y, s, \alpha, l)$  which describes the global position  $(x, y)$ , the size  $s$  and the orientation  $\alpha$  of a hand in the image, together with its discrete state  $l = 1 \dots 5$ . The vector  $X$  uniquely identifies the hand configuration in the image and estimation of  $X$  from image sequences corresponds to simultaneous hand tracking and recognition.

### 3.2 Model evaluation

Given a feature-based object model, the feature likelihood map provides a *direct* way to evaluate the model on image data. To obtain the likelihood that a model configuration  $X$  gives rise to an image  $I$ , one can simply multiply the likelihood



**Fig. 7.** Feature-based hand models in different states. The circles and ellipses correspond to blob and ridge features. When aligning models to images, the features are translated, scaled and rotated according to the parameter vector  $X$ .

values for model features which are directly available from  $\mathcal{M}_L$ . Hence, we define the likelihood  $p$  for a model hypothesis  $X$  and an image  $I$  as

$$p(I|X) = (1 - \varepsilon)^{N-n} \prod_{i=1}^n \mathcal{M}_L(x_i, y_i, t_i), \quad (14)$$

where  $\mathcal{M}_L$  is computed on the image  $I$ ,  $x_i$ ,  $y_i$  and  $t_i$  denote the position and the size of the  $i^{th}$  feature in the model, while  $\varepsilon \in (1, 0)$  accounts for a maximal admissible matching error and enables for comparison of models with the different number of features  $n$  ( $N = \max_j(n_j)$ ). In addition, this likelihood is multiplied by a prior on skin colour computed from colour histograms of human hands.

Notably, the described evaluation does not involve any search and is simple and efficient to compute. Therefore it is highly useful for real-time applications.

### 3.3 Tracking and recognition

To detect, recognize and track hands in image sequences, we search for a hand configuration defined by a parameter vector  $X_k$  that maximizes the posterior distribution  $p(X_k|I_k)$  on a given image  $I_k$  at a time moment  $k$ . Using Bayes rule, the posterior can be estimated by

$$p(X_k|I_k) = h p(I_k|X_k) p(X_k|I_{k-1}) \quad (15)$$

where  $p(I_k|X_k)$  is the likelihood of  $X_k$  given  $I_k$ ,  $p(X_k|I_{k-1})$  is the prior distribution of  $X_k$  derived from a previous time step and  $h$  is a normalization constant. Since the likelihood distribution above has no closed-form expression, the desired posterior must be approximated. For this reason, we apply particle filtering to estimate and approximate the posterior by a set of  $N$  samples (here  $N \approx 1000$ ) distributed in a parameter space (see (Isard and Blake 1996) for an introduction). Given the posterior  $p(X_k|I_k)$ , we compute its mean  $X_{k,mean}$  and consider it as the estimate of a hand pose at time moment  $k$ .

Particle filters spend most of their time on evaluating the likelihood of model hypotheses (samples). As described in the previous section, the proposed feature likelihood map is highly efficient for this purpose and we use it for evaluating the likelihood of samples within the framework of particle filtering. The efficient evaluation enables recognition and tracking to be done in real-time (currently at the frame rate 5–10 Hz). Figure 8 illustrates the result of combined tracking and recognition using the described framework.



**Fig. 8.** Results of combined hand tracking and pose recognition using particle filtering and evaluation of feature-based hand models on feature likelihood maps.

### 3.4 Implementation details

In practice, the abovementioned scheme has been implemented in a pyramid framework using a fixed set of scale levels. The resolution at scale level  $t_i$  was obtained by sub-sampling the original image with a factor  $\kappa_i = \sqrt{t_i/t_f}$ , and the derivatives have been computed using filter kernels of fixed scale  $t_f$ . In the experiments, we found  $t_f \approx 2.0$  to be sufficiently large for obtaining a satisfactory quality of  $\mathcal{M}$  on one hand, while on the other hand being sufficiently small to enable fast computations. On a modest 550 MHz Pentium III processor our current implementation (without extensive optimization) requires about 0.1 s to compute the feature likelihood map on a  $100 \times 100$  image and about 0.04 s to perform the particle filtering using 1000 hypotheses.

## 4 Related work

The subject of this paper relates to multi-scale approaches for image representation, computation of differential invariants, detection of image features as well as tracking and recognition of view-based object models. Because of the scope of these areas, it is not possible to give an extensive review, and only a few closely related works will be mentioned. Crowley and Sanderson (1987) considered a graph-like image representation containing links between blobs at different scales. Pizer et al. (1994) proposed the use of multi-scale medial-axis representations computed directly from image patterns distributions. Multi-scale image differential invariants (Koenderink and van Doorn 1992, Lindeberg 1994, Florack 1997) have been computed by several authors, including Schmid and Mohr (1997) who apply such descriptors at interest points for image indexing and retrieval. Explicit scale selection for extraction of multi-scale image features has been investigated by Lindeberg (1998). A similar approach by Shokoufandeh et al. (1999) extracts extrema in a wavelet transform. Lindeberg (1998), Chomat et al. (2000) and Almansa and Lindeberg (2000) have computed dense descriptors for estimating the characteristic scale at any image point. With respect to object tracking, Isard and Blake (1996) developed a particle filtering approach for tracking contour-based models. Black and Jepson (1998) used eigenspace models of gray-value patterns for tracking deformable models. The approach by Bretzner and Lindeberg (1999) is closer to ours and applies a hierarchy of multi-scale features for representing and tracking hands.

## 5 Summary and future work

In this paper, we have presented a new approach for probabilistic and dense image representation by feature likelihood maps. Such maps are invariant to the amplitude of patterns and emphasize local structures in images by assuming high values at certain points in feature space. We derived the feature likelihood map for symmetric blob-like image structures and analyzed its behavior on synthetic and real images. Using the dense structure of the feature likelihood map, we have

shown how it can be applied for direct and efficient evaluation of feature-based object hypotheses. Based on this evaluation procedure, we developed a particle filtering approach for recognizing and tracking hands in image sequences.

By analogy with the developed likelihood map for symmetric blob-like structures, similar maps can be constructed for other types of local image structures. For this purpose, the expression in (11) must be redefined by substituting the normalized Laplacian operator and its quadrature by other differential entities emphasizing the desired image properties. Examples of other feature likelihood maps constructed in this way are presented in (Laptev and Lindeberg 2001), as well as ways to incorporate colour information into this framework.

Another interesting direction of future research concerns the extension of feature likelihood maps to spatio-temporal domain. Here, the general ideas of this presentation could be combined with the concept of normalized derivatives in spatio-temporal scale-space. The resulting maps could then be used in order to analyze, capture and recognize temporal events in image sequences.

## References

- Almansa, A. and Lindeberg, T. (2000). Fingerprint enhancement by shape adaptation of scale-space operators with automatic scale-selection, *IEEE-IP*, **9**(12): 2027–2042.
- Black, M. and Jepson, A. (1998). Eigen tracking: Robust matching and tracking of articulated objects using view-based representation, *IJCV* **26**(1): 63–84.
- Bretzner, L. and Lindeberg, T. (1999). Qualitative multi-scale feature hierarchies for object tracking, *JVCIR* **11**: 115–129.
- Chomat, O., de Verdiere, V., Hall, D. and Crowley, J. (2000). Local scale selection for gaussian based description techniques, *ECCV'00*, Dublin, Ireland, 117–133.
- Crowley, J. and Sanderson, A. (1987). Multiple resolution representation and probabilistic matching of 2-D gray-scale shape, *IEEE-PAMI* **9**(1): 113–121.
- Fleck, M., Forsyth, D. and Bregler, C. (1996). Finding naked people, *ECCV'96*, Cambridge, UK, II:593–602.
- Florack, L. M. J. (1997). *Image Structure*, Kluwer.
- Isard, M. and Blake, A. (1996). Contour tracking by stochastic propagation of conditional density, *ECCV'96*, Cambridge, UK, I:343–356.
- Koenderink, J. J. and van Doorn, A. J. (1992). Generic neighborhood operators, *IEEE-PAMI*, **14**(6): 597–605.
- Laptev, I and Lindeberg, T. (2001) “A multi-scale feature likelihood map for direct evaluation of object hypotheses”, Technical report ISRN KTH/NA/P-01/03-SE. <http://www.nada.kth.se/cvap/abstracts/cvap249.html>
- Lindeberg, T. (1994). *Scale-Space Theory in Computer Vision*, Kluwer.
- Lindeberg, T. (1998). Feature detection with automatic scale selection, *IJCV*, **30**(2): 77–116.
- Pizer, S. M., Burbeck, C. A., Coggins, J. M., Fritsch, D. S. and Morse, B. S. (1994). Object shape before boundary shape: Scale-space medial axis, *JMIV* **4**: 303–313.
- Schmid, C. and Mohr, R. (1997). Local gray value invariants for image retrieval, *IEEE-PAMI* **19**(5): 530–535.
- Shokoufandeh, A., Marsic, I. and Dickinson, S. (1999). View-based object recognition using saliency maps, *IVC* **17**(5/6): 445–460.
- Witkin, A. P. (1983). Scale-space filtering, *Proc. IJCAI'83*, Karlsruhe, West Germany, 1019–1022.