Local Scale Selection for Gaussian Based Description Techniques

Olivier Chomat, Vincent Colin de Verdière, Daniela Hall, and James L. Crowley

Project PRIMA - Lab GRAVIR - IMAG
INRIA Rhône-Alpes, 655, avenue de l'Europe
38330 - Montbonnot - FRANCE
Olivier.Chomat@inrialpes.fr

Abstract. This paper addresses the problem of the local scale parameter selection for recognition techniques based on Gaussian derivatives. Patterns are described in a feature space of which each dimension is a scale and orientation normalized receptive field (a unit composed of normalized Gaussian-based filters).
Scale invariance is obtained by automatic selection of an appropriate local scale [Lin98b] and followed by normalisation of the receptive field to the appropriate scale. Orientation invariance is obtained by the determination of the dominant local orientation and by steering the receptive fields to this orientation.
Data is represented structurally in a feature space that is designed for the recognition of static object configurations. In this space an image is modeled by the vectorial representation of the receptive field responses at each pixel, forming a surface in the feature space. Recognition is achieved by measuring the distance between the vector of normalized receptive fields responses of an observed neighborhood and the surface point of the image model.
The power of a scale equivariant feature space is validated by experimental results for point correspondences in images of different scales and the recognition of objects under different view points.

1 Introduction

Object indexing is the problem of determining the identity of a physical object from an arbitrary viewpoint under arbitrary lighting conditions. Changes in the appearance of the object under variations in lighting and viewpoint make this a difficult problem in computer vision. The classic approach is based on the idea that the underlying 3D structure is invariant to viewpoint and lighting. Thus recovery of the 3D structure should permit the use of techniques such as back-projection and features matching [Fau93] to match the observed structure to a data base of objects.

An alternative to 3D reconstruction is to remain in the 2-D image space and to consider measurements of the object appearance. Sirovitch and Kirby [SK87] showed that in the case of face recognition, Principal Components Analysis (PCA) can be used to generate a low dimensional orthogonal sub space.

D. Vernon (Ed.): ECCV 2000, LNCS 1842, pp. 117–134, 2000.

The distance of two points in this space is determined by inner product, whose computation cost depends on the dimensionality of the space. Turk and Pentland [TP91] refined and popularized this approach, greatly enhancing the acceptance of principal components analysis as a vision technique. Murase and Nayar [MN95] extended this idea by expressing the set of appearances of objects as a trajectory in a PCA space. Black and Jepson [BJ96] demonstrated that the appearance of a hand making a gesture could also be expressed and matched as a trajectory in a PCA space.

All of the above techniques are sensitive to partial occlusions and scale normalization. The position of a projected image in the PCA space is coupled with the appearance of the image. Object translations within the image, variable background, differences in the image intensity or illumination color alter the position of the image in PCA space. Thus, PCA techniques require object detection, segmentation and precise normalisation in intensity, size and position. Such segmentation and normalization is very difficult, and there exists no approach that solves this problem in the general case.

Segmentation and normalization problems can be avoided by using local appearance based methods [Sch97,CC98a,RB95] that describe the appearance of neighborhoods by receptive fields. The effects of background and partial occlusion are minimized by considering small neighborhoods. The problem of object position within the image is solved by mapping the locally connected structures into surfaces or multi-dimensional histograms in a space of local appearances. Robustness to changes in illumination intensity are obtained by energy normalization during the projection from image to the appearance space. The resulting technique produces object hypotheses from a large data base of objects when presented with a very small number of local neighborhoods from a newly acquired image.

In this paper the local appearance technique proposed Colin de Verdière [CC98a] is extended to a local description technique which is scale and orientation invariant. A description of the local visual information is obtained using a set of Gaussian derivatives. The Gaussian derivatives responses to different images result in a family of surfaces. Such a surface is another representation of the model image. Recognition is achieved by projecting neighborhoods from newly acquired images into the local appearance space and associating them to nearby surfaces. This technique leads to the problem of the Gaussian derivatives parameterization in scale and orientation.

Lindeberg defines a set of Gaussian derivatives operators to select scale for edge, ridge, corner, and blob features, thus for feature detection [Lin98a]. We adopt this approach and apply it to all image points. Local appropriate scales are detected at every point allowing the parameterization and normalization of the Gaussian derivatives. This leads to a scale equivariant description. Also, detecting the dominant orientation of neighborhoods allows to normalize the receptive fields by orientation [FA91]. An orientation equivariant description is found. The scale parameter is important also for orientation normalization, because a scale that is not well adapted to the local structure makes orientation detection instable.

In this article we investigate local scale selection for appearance based recognition techniques. As demonstrated below the appropriate local scale is an important factor. We focus on the local scale selection according to the feature type. Experiments show that local scale selection with consideration of the feature type improves object recognition.

In the next section the pattern description and representation is explained. The proposed approach can be applied to patterns in spatial or frequency domain. Then the scale and orientation equivariance property is described according to the main publications in this area. As a result we explain our contribution to the local scale selection considering the feature type. Experiments in scale and orientation validate the proposed approach.

2 Pattern Description and Representation

The appearance of an object is the composition of all images of the object observed under different viewing conditions, illuminations, and object deformations. Adelson and Bergen [AB91] define the appearance space of images for a given scene as a 7 dimensional local function, $I\,(x, y, \lambda, t, V_x, V_y, V_z)$, whose dimensions are viewing position, (V_x, V_y, V_z), time instant, (t), position, (x, y), and wavelength, (λ). They call this function the "plenoptic function" from the Latin roots *plenus*, full, and *opticus*, to see. The use of description techniques and the use of representation models of descriptors responses allow the analysis of the plenoptic function for recognition problems.

Adelson and Bergen propose to detect local changes along one or two plenoptic dimensions. The detector responses, that code the visual information, are represented by a table in which they are compared pairwise. Adelson and Bergen use low order derivative operators as 2-D receptive fields to analyse the plenoptic function. However, their technique is restricted to derivatives of order one and two. No analysis of three or more dimensions of the plenoptic function is investigated and little experimental work is published on this approach.

Nevertheless the plenoptic function provides a powerful basis for recognition systems. This paper deals with such framework where patterns are characterized by describing their local visual information and modeling the descriptor responses. The result is a recognition software based on local properties.

Consider the plenoptic function, $I\,(x, y, V_x, V_y, V_z)$, constrained to a single frame and a gray channel. $I\,()$ is analyzed by a set of receptive fields. An orthogonal basis of receptive field responses can be found that span the space of all receptive field responses. The goal is to decrease the dimensionality of this space by determining those receptive fields which allow an optimal description of the appearance. This basis can vary according to the nature of the recognition problem. The next section discusses the construction of receptive fields according to different signal decomposition techniques. Then two methods of pattern representation in the feature space are discussed. The first is a statistical representation where objects are characterized as the joint statistics of the receptive field responses and, the second is a structural approach where connected structures in the images are mapped as surfaces in the feature space.

2.1 Signal Decomposition

Classically the description of a signal is obtained by its projection onto a set of basis functions. Two widely used approaches for signal decomposition are the Taylor expansion (equation 1) and the Fourier transform (equation 2), corresponding respectively to the projection of the signal onto basis functions with modulated amplitude and to the projection of the signal onto a function base which is frequency modulated:

$$f(t) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(t_0) \cdot (t - t_0)^n \tag{1}$$

$$f(t) = \sum_{n=-\infty}^{\infty} \hat{f}(n) \cdot e^{i \cdot n \cdot t} \tag{2}$$

Note that there exist other local decomposition bases. The nature of the problem motivates the choice of the decomposition base. For example a frequency-based analysis is more suitable for texture analysis, and a fractal-based description for natural scene analysis. But independently from the basis choice, the receptive fields responses are estimated over a neighborhood which size is relative to the locality of the analysis.

The derivative operator of the Taylor expansion and the spectral operator of the Fourier transform can be formulated as generic operators. The concept of linear neighborhood operators was redefined by Koenderink and van Doorn [Kv92] as generic neighborhood operators. Typically operators are required at different scales corresponding to different sizes of estimation support. Koenderink and van Doorn have motivated their method by rewriting neighborhood operators as the product of an aperture function, $A(\boldsymbol{p}, \sigma)$, and a scale equivariant function, $\phi(\boldsymbol{p}/\sigma)$:

$$G(\boldsymbol{p}) = A(\boldsymbol{p}, \sigma)\, \phi(\boldsymbol{p}/\sigma) \tag{3}$$

The aperture function takes a local estimation at location $\boldsymbol{p}$ of the plenoptic function which is a weighted average over a support proportional to its scale parameter, σ. The Gaussian kernel satisfies the diffusion equation and can therefore serve as aperture function:

$$A(\boldsymbol{p}, \sigma) = \frac{e^{-\frac{1}{2}\frac{\boldsymbol{p} \cdot \boldsymbol{p}}{\sigma^2}}}{\left(\sqrt{2\pi}\sigma^D\right)} \tag{4}$$

The function $\phi(\boldsymbol{p}/\sigma)$ is a specific point operator relative to the decomposition basis. In the case of the Taylor expansion $\phi(\boldsymbol{p}/\sigma)$ is the n^{th} Hermite polynomials [MS65]:

$$\phi(\boldsymbol{p}/\sigma) = (-1)^n He_n(\boldsymbol{p}/\sigma) \tag{5}$$

In the case of the Fourier series $\phi(\boldsymbol{p}/\sigma)$ are the complex frequency modulation functions tuned to selected frequencies, $\boldsymbol{\nu}$:

$$\phi(\boldsymbol{p}/\sigma) = e^{2\pi j \boldsymbol{\nu} \cdot \boldsymbol{p}/\sigma} \tag{6}$$

Within the context of spatial, or spectral, signal decomposition the generic neighborhood operators are scale normalized Gaussian derivatives [Lin98b], and respectively scale normalized Gabor filters.

2.2 Pattern Representation in Feature Space

The computation of a vector of descriptors can be formally modeled as a projection from the image pixel space to a new space more suitable for indexing. This descriptor space is composed of N receptive fields corresponding to a set of Gaussian derivatives or Gabor filters for example. An image neighborhood which is a vector in a M dimensional space can be represented in the descriptor space by a N dimensional vector with $N \ll M$. The distance between two points in the descriptor space is a measure for the similarity of these neighborhoods and is used for recognition.

An object signature is obtained by representing or modelizing the receptive fields responses in the feature space, either statistically or structurally. Schiele [SC96] has shown that the local appearance of static objects can be represented statistically using multi-dimensional histograms of Gaussian derivative responses. Histograms of object classes are compared and the conditional probability is returned that the observed vector is part of the trained classes. Colin de Verdière [CC98a] has used a structural representation by sampling the local appearance of static objects. Such discrete sampling permits recognition from small neighborhoods by a process which is equivalent to table lookup.

Statistical representation: The output from the set of receptive fields provides a measurement vector at each pixel. The joint statistics of these vectors allow the probabilistic recognition of objects. A multi-dimensional histogram is computed from the output of the filter bank. These histograms can be considered as object signature and provide an estimate of the probability density function that can be used with Bayes rule. Schiele [SC96] uses this methods for object recognition and Chomat [CC99a] extends the approach to the recognition of activity patterns.

Structural approach: At each pixel a measurement vector of the output from the set of receptive fields is stored in association with an identifier of the model image and its position within the model image. The storage of vectors associated to all model image points enables a simultaneous identity and pose recognition of object by matching of measurement vectors. This matching is performed efficiently by using a hierarchical data structure. Competitive evaluation of multiple vectors of an image provide a highly discriminant recognition of the learned object. As a result, this recognition scheme returns one (or multiple) object poses associated with a confidence factor based on the number of points detected on this pose [CC98a,CC99b]

In this paper a structural representation of data in a feature space is used for the recognition of static object configurations. In the structural approach the vectors with the shortest distance are searched. The class of the observed vector is the class of the vector with the shortest distance. The classification is based on searching all vectors that are within a sphere centered on the observed vector. For an efficient performance of the classification task, the storage structure is very important. Colin de Verdière [CC98a] proposes an indexation storage tree, in which each vectorial dimension is decomposed into 4 parts successively.

During the training phase new vectors can be added easily to the tree. The advantage of this structure is that all vectors within the search sphere centered on the observed vector can be computed efficiently.

Note that the addressed problem in this paper is not a critical study of indexing techniques, but the local scale parameter selection for Gaussian based descriptors in the context of object recognition. So, the choice of a structural full description of the appearance is very suitable to this study, but other recognition schemes can take profit of this automatic scale selection as for example : a statistical representation by histograms [CC99a,Sch97] or an interest points based approach like Schmidt [SM97] who selects a priori interesting points and modelizes only these points in her system.

2.3 Conclusion

The quality of recognition techniques depends on their ability to recognize objects in a scene under a minimum of assumptions. Generally the required properties are their robustness or invariance to illumination and view point variations.

The robustness to illumination variations and point of view changes is obtained by sampling the appearance of the object by including images with these changes into the training base. It is possible to view the set of different appearances of one object, that result in a trajectory in the appearance space [MN95, CC98b] parameterized by illumination and view point. For example, scale robustness is achieved by learning the object at several different scales and matching a new image of the object to each of the trained scales. Another approach to achieve robustness is to model the object at a fixed scale and then match the images of a pyramid of a new observed image of the object.

As shown in this section Gaussian based techniques are well suited for a scale equivariant description since the scale parameter of the aperture function and of the point operator is explicit. A scale invariant Gaussian based description can be obtained by an appropriate scale parameter selection. The problem of the detection of orientation is solved using the property of steerability of Gaussian derivatives. Freeman and Adelson [FA91] use Gaussian derivatives, to compute the n^{th} derivative under arbitrary orientation by a linear combination of a finite number of derivatives.

The approach we propose is to use the properties of scale and orientation parameterisation of Gaussian based local descriptors to design receptive fields that are equivariant to scale and orientation. The use of Gaussian derivatives as local descriptors provides an explicit specification of scale and orientation. Scale is specified by the σ parameter providing a scale invariant feature. Using steerable filters [FA91], it is possible to compute the n^{th} derivative under arbitrary orientation by a linear combination of a finite number of n^{th} order derivatives. In this paper a set of normalized Gaussian derivatives up to Order three is used to describe the plenoptic function. A scale detector and an orientation detector are used to normalize and steer Gaussian derivatives. The local description is equivariant to scale and orientation and allows recognition which is invariant to scale and orientation. Note that Gaussian derivatives can be efficiently computed by using a recursive implementation (see [YV95]).

3 Scale Invariance

Theoretically there exist specific features that are scale invariant such as corners and edges. Practically this is not the case because edges resemble more to a ramp over scales. However these features can be described using scale equivariant descriptors. The first paragraph of this section demonstrates the scale equivariance property of Gaussian derivatives by normalizing them according to the scale parameter. The next paragraph deals with two scale invariant representations. One is based on a multi-scale data representation (or pyramidal representation), and another one is based on local scale parameter selection.

Scale equivariant receptive fields A scale equivariant feature space is designed using normalized Gaussian derivatives taking into account their scaling property.

$$\partial_{x^n} g\left(s \cdot x, s \cdot \sigma\right) = \frac{1}{s^{n+1}} \partial_{x^n} G\left(x, \sigma\right) \tag{7}$$

Consider the Gaussian filter, $G\left(x, \sigma\right)$, and the one dimensional signal, $f\left(x\right)$. Let $L\left(x, \sigma\right)$ be the response of the Gaussian filter:

$$L\left(x, \sigma\right) = f\left(x\right) * G\left(x, \sigma\right) \tag{8}$$

The normalization of the Gaussian derivatives responses according to a selected scale parameter, σ, is:

$$\partial_{\xi^n} L\left(\xi, \sigma\right) = \frac{\partial^n L\left(x, \sigma\right)}{\partial_{x^n}} \cdot \sigma^n \quad with \quad \xi = \frac{x}{\sigma} \tag{9}$$

This scale normalization leads to a descriptor which is scale equivariant

$$\partial_{\xi'^n} L\left(\xi^{'}, \sigma^{'}\right) = \partial_{\xi^n} L\left(\xi, \sigma\right) \quad with \quad x^{'} = s \cdot x \quad and \quad \sigma^{'} = s \cdot \sigma \tag{10}$$

A local scale equivariant feature space can be built using such scale normalized descriptors but the a priori knowledge of the local feature scale is necessary.

Scale invariant modelization Traditionally, the scale parameter of the filters is defined intuitively, according to the size of the features to be recognized and a multi-scale strategy is adopted to overcome the problem of scale variations. Models of objects are built at several scales and matching is done by comparison within the different trained scales. Currently a similar strategy is adopted to be robust to changes in orientation by learning several orientations. The goal of such strategies is to become robust to scale changes. This robustness lays on the structure for data representation or modelization. The main problem still in the parameterization of receptive fields, and generally the scale parameter is fixed but not appropriate to the local information.

The approach we propose is to locally estimate the required scale at each point and to normalise the Gaussian derivative filters to the local scale. With such a method there exists no data representation at several scales (which is redundant), but only one single scale invariant data representation. Several maps of selected scales are used depending on the features to be analysed.

4 Detection of Orientation

Structured neighborhoods can have a dominant orientation. The dominant direction of a neighborhood can be found by determining the filter direction that gives the strongest response.

There are two ways to determine this filter. First, a set of filters can be generated that are rotated by a small angle. Then each filter of this set is applied to the neighborhood. If a precise orientation is required the number of generated filters is very high and also the computation cost of the operation.

A second possibility is to use only a small number of appropriate filters and interpolate between the responses. With an appropriate filter set and the correct interpolation rule the response of the neighborhood to a filter with an arbitrary orientation can be determined without explicitly applying this filter to the neighborhood. Freeman [FA91] uses the term steerable filter for such a filter class.

Steerable Filters Let G_n be the n^{th} order derivative of the Gaussian function. Let $()^\theta$ be the rotation operator so that a function $f(x,y)^\theta$ is the function $f(x,y)$ rotated by θ. The synthesized filter of direction, θ, can be obtained by a linear combination of $G_1^{0^\circ}$ and $G_1^{90^\circ}$ [FA91]

$$G_1^\theta = \cos(\theta) G_1^{0^\circ} + \sin(\theta) G_1^{90^\circ} \tag{11}$$

Equivariance of Orientation Let I be an image. In I let w be the neighborhood around p with the dominant orientation θ. Let I^ω be I rotated by ω. The to w corresponding neighborhood w' in I^ω has then the dominant direction $\theta + \omega$. $(G_1^{0^\circ})^\omega$ and $(G_1^{90^\circ})^\omega$ are the basis functions $G_1^{0^\circ}$ and $G_1^{90^\circ}$ rotated by ω and can be written as

$$(G_1^{0^\circ})^\omega = \cos(\omega) G_1^{0^\circ} + \sin(\omega) G_1^{90^\circ} = G_1^{0^\circ + \omega} \tag{12}$$

$$(G_1^{90^\circ})^\omega = -\sin(\omega) G_1^{0^\circ} + \cos(\omega) G_1^{90^\circ} = G_1^{90^\circ + \omega} \tag{13}$$

The equivariance results from

$$\begin{aligned} (G_1^\theta)^\omega &= \cos(\theta)(G_1^{0^\circ})^\omega + \sin(\theta)(G_1^{90^\circ})^\omega \\ &= \cos(\theta+\omega) G_1^{0^\circ} + \sin(\theta+\omega) G_1^{90^\circ} = G_1^{\theta+\omega} \end{aligned} \tag{14}$$

Orientation Invariance Taking into account the values of the gradient G_1 in x and y direction the dominant direction of a neighborhood can be determined by

$$\theta = \operatorname{atan2}\left(\frac{\delta}{\delta x} L(x,y;\sigma), \frac{\delta}{\delta y} L(x,y;\sigma)\right) \tag{15}$$

For each neighborhood in the image the dominant direction can be determined, which allows to normalise each neighborhood in orientation. Using the equivariance property two corresponding neighborhoods w^θ and w^ω will be normalized to the same neighborhood w^{normal}.

5 Local Scale Selection

Features in a scene appear in different ways depending upon the scale of observation. Traditionally, when scale is considered, image structures are represented in a multi-scale pyramid and processing is applied to a set of scales. Such techniques are sensitive to the fact that some features may disappear at too coarse scales or too fine scales. It is therefore necessary to determine an appropriate scale for each observed feature. Targeting this appropriate scale for the projection in the feature scale in association with scale invariant features computation enables a scale independent representation.

Lindeberg [Lin98b] proposes a framework for generating hypotheses about scale levels based on the assumption that local extrema over scales of normalized derivatives correspond to interesting structures. This approach gives rise to analytically derived results which correspond to intuition for scale selection for detecting image features.

This section provides experiments based on Lindeberg proposal for features scale selection. We are interested in receptive fields which are scale invariant and orientation invariant. Scale invariance is obtained by selecting a scale relative to the feature shape. Such scale selection is available using a Laplacian based operators. The detection of orientation is done at an appropriate scale where the gradient is stable.

5.1 Blob Features Scale Selection

The proposed general scale detector of equation (16) is expressed as a polynomial combination of normalized Gaussian derivatives, where the normalization controls the scale invariance.

$$Lap\left(x, y, \sigma_s\right) = \left(\sigma_s^2\right)\left(\partial_{xx} g\left(x, y, \sigma_s\right) + \partial_{yy} g\left(x, y, \sigma_s\right)\right) \tag{16}$$

The function $Lap\left(x, y, \sigma_s\right)$ is computed for a large set of scales, σ_s, corresponding to the scale-space feature signature. The maximum of the normalized derivatives, $Lap^{max}\left(x, y, \sigma_s = \sigma_0\right)$, along scales leads to the feature scale σ_0.

The equivariance property of blob feature scale, that enables blob feature recognition at different scales, is demonstrated in the following experiment. A set of images representing a scene at several scales is taken (two of them are shown in figure 1). A target is tracked along scales and its scale signature is shown in the central figure. The over-lined feature has a signature which translates with scale. These curves present a maximum over scale. The σ parameter, which is characteristic to the local feature, is selected according to the observed maximum. Thus, the local scale can be used to parameterize the normalized Gaussian derivatives described in the previous section. Maps of the selected scale parameter σ of two images of the chocos object are shown in figure 2. These maps show that the scale parameter distribution is preserved over the scanned scale range. Figure 3 shows the first normalized derivative along the x axis computed on three "chocos" images. For all images the derivatives up to order three are computed. Each local feature is described by a nine dimensional receptive field, which corresponds to a point in a nine dimensional descriptor space. This vector

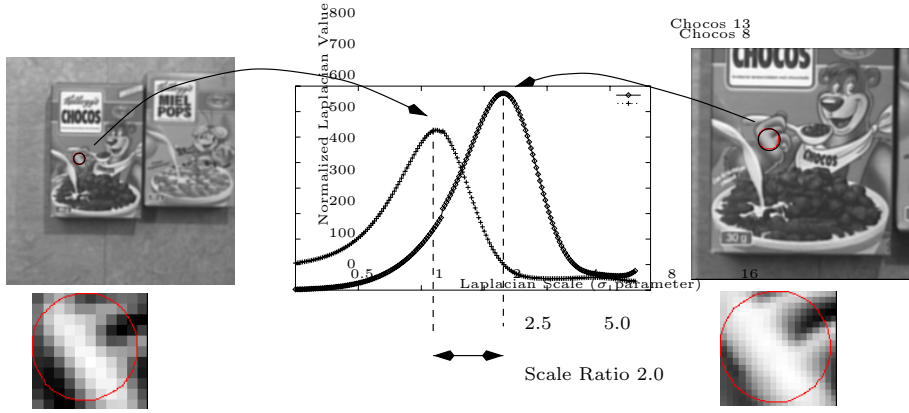

Fig. 1. *Automatic scale selection of a local feature for corresponde nce between two images. The curves present the evolution of normalized Laplacian with* σ*. Circles indicate a radius of* 2σ *(twice the selected scales parameter value). The ratio between the selected* σ *gives the scale ratio between the local features and therefore in this example, an approximate scale ratio of* 2 *between the images.*

Fig. 2. Images of the selected scale parameter σ of two images of the chocos object.

is scale invariant and provides a means to obtain point correspondences between images at different scales. In the first image, four points have been selected which corresponds to four features vectors. In the next images, their correspondents are successively detected by searching the most similar feature vectors.

5.2 Edge Features Scale Selection

In order to design a receptive field that is invariant to orientation the dominant direction of an image neighborhood needs to be determined. An important parameter for a reliable orientation normalization is the selected scale parameter.

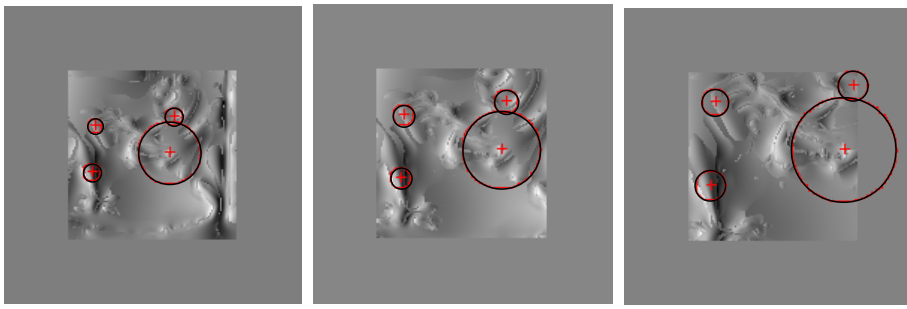

Fig. 3. Points correspondences obtained by projecting the points on a nine dimensional feature space and matching similar vectors between the images. The gray values in the images correspond to the intensity of the first derivative.

If this parameter is not appropriate the gradient information becomes instable. This results in an orientation error which makes classification difficult.

The neighborhoods size is chosen such that the gradient in at least one direction is stable. An appropriate measure is the gradient norm. It is isotropic and returns a maximum energy when a stable gradient is present. If none of the gradient filters are stable within a maximum filter size, the neighborhood contains very low energy edge data. An orientation normalisation is unstable, but because of the lack of edge data, this does not perturb the recognition.

In a previous experiment one single scale based on the Laplacian was selected for all derivatives. This scale is appropriate for blob features. Figure 4 compares the σ detected by the normalized gradient norm and the σ detected by the normalized Laplacian. The graphics show that different σ are detected by the gradient norm and the scale normalized Laplacian.

The graphic below the image shows an interesting case in figure 4. If the σ selected by the normalize Laplacian would b e applied for orientation detection, orientation errors can not be avoided. The scale where the normalize Laplacian has maximum energy, shows a normalized gradient norm with very weak energy. The σ detected by the normalized Laplacian is appropriate for the purple blob between the two white lines, which is an uniform region. For a stable orientation detection a much bigger size must be selected to obtain enough edge information. This is an extreme case. More often the normalize Laplacian selects a size where there is some gradient energy (see graphics left and right of the image in figure 4). However a higher gradient energy is found at a different scale.

5.3 Scale and Orientation Invariant Description

Lindeberg uses normalized derivatives for adaptively choosing the scales for blob, corners, edges or ridges scale detection. The selected scale is also a good cue for tuning the parameters of Gaussian derivatives for appearance description. Tuning Gaussian derivatives with the map of selected scales leads to a scale

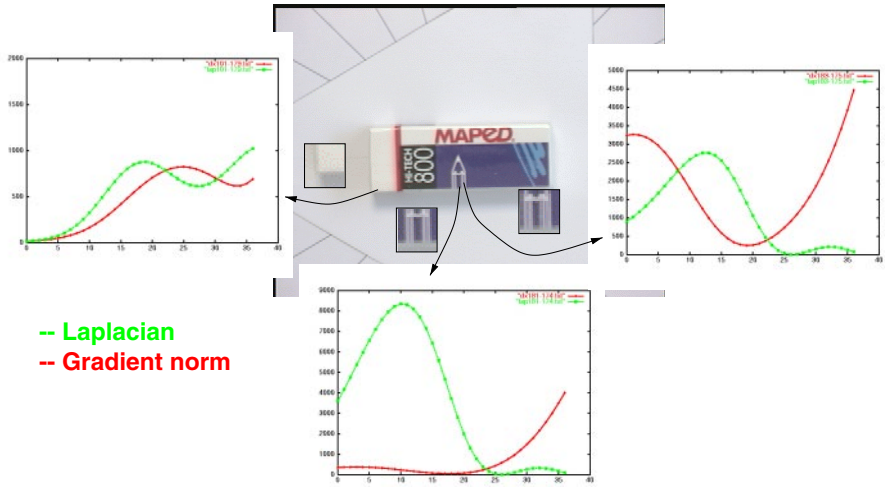

Fig. 4. *Comparison of σ detection between using normalized gradient norm or normalize Laplacian.*

invariant description. And steering the filters to the dominant orientation leads to an orientation invariant description.

The choice of the appropriate scale is an important parameter, and therefore must be chosen very carefully. The scales detected by the Laplacian adapt very well to blob features, where as the scale detected by the gradient norm designes a neighborhood such that a stable gradient can be found within this neighborhood. Both features are important for a reliable recognition and essential for scale and orientation invariant description.

Figure 4 displays the energy differences of gradient norm energy and Laplacian energy over different scales. It can be observed that the two filter methods detect in all cases different scales. This is due to the fact that the filter methods adapt to either blob features or edge features. The presence of a blob feature and an edge feature of the same scale in the same neighborhood is a contradiction. As a consequence we investigate in the following experiments the impact of using both scales selected by the two methods in order to improve the stability in scale and orientation normalisation.

Figure 5 displays this enhancement on an example. The orientation displayed in the angle image in figure 5(d) obtained from the gradient norm scales is much more stable than the angle image in figure 5(c) obtained from the scales selected by the Laplacian. Many discontinuities can be observed in figure 5(c) that are not present in figure 5(d). This improvement in the orientation detection is due to the fact that the scale is based on derivatives of order one.

To obtain a description which is both scale and orientation invariant two scales maps are used. The map of selected scales obtained with the gradient norm is used to parameterize odd derivatives and the map of selected scales

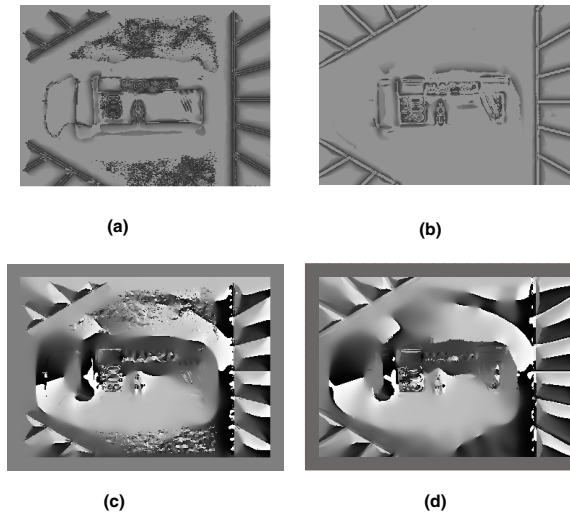

Fig. 5. *(a) sigma image obtained from Laplacian (b) sigma image obtained from gradient norm (c) angle image resulting from sigma selected by Laplacian (d) angle image resulting from sigma selected by gradient norm.*

obtained with the Laplacian is used to parameterize even derivatives. The next section provides recognition results using such a scale and orientation invariant description.

6 Application to Indexation

6.1 Experiments on Scale and Orientation Detection

For validation of the presented approach, automatic scale selection is applied to indexation. Two experiments are compared to show the stabilization of the orientation normalisation using gradient norm for scale selection. In the first experiment the scale is selected by the Laplacian. In the second experiment two scales are selected, one based on the gradient norm for the first derivatives, a second based on the Laplacian for the 2nd derivatives.

A set of 13 images are taken of one single object. The object is rotated by 15 degrees in-between two frames. One image is used for training. The other 12 images are used for testing.

The training is performed according to [CC98a] using local appearance description by Gaussian derivatives, recursive filtering, automatic scale selection and orientation normalisation by steerable filters. At each point of the training image the most appropriate scale is detected using the two different strategies in the two experiments. Then the neighborhoods are normalized by orientation and a 8 dimensional filter response to the steered Gaussian derivatives is computed, which is stored in an indexation storage tree. The filter response serves to identify the neighborhood and its similar samples in the test images.

Overlapping neighborhoods are sampled of the test images with a step size of 3 pixels. At each point the orientation normalized filter response in the appropriate scale is computed. To evaluate the experiment only the training vector with the smallest distance is considered. In general the approach returns all vectors from the indexation tree that are within a sphere centered on the newly observed vector. It is a restriction to look only at the closest vector. The recognition rates are naturally lower than in a system which takes into account the entire list of hypotheses. Two values are computed. Firstly, the percentage that the closest vector is correct, is measured. This means, the answer obtained from the indexation tree indicates the correct neighborhood. Secondly, the average error in the orientation normalisation is measured. This value indicates the precision of the orientation normalisation.

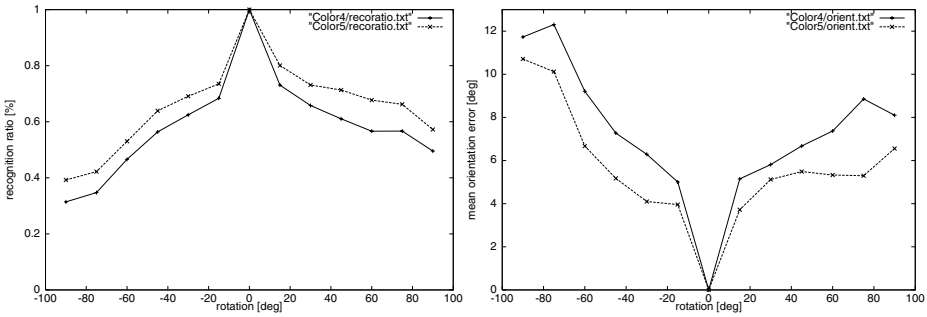

Fig. 6. *left: Percentage of correct first answers. The image at 0 degrees is the training image. right: average orientation error observed during indexation process.*

Figure 6 shows the results of the two experiments. The solid curve corresponds to the experiment in which only one scale was selected. The dashed curve corresponds to the experiment in which different scales were selected for first and second derivatives. The graphics show that the second technique produces a higher percentage in correct first answers and a higher precision in the orientation normalisation. These two values illustrate the gain that is obtained by using the gradient norm for the scale selection of the first derivative.

6.2 Object Recognition under Scale Variations

This object recognition experiment is evaluated on a basis of 28 objects (figure 7). One single image is learned. The results are shown on examples for objects "Chocos" and "Robot", that are highlighted in figure 7. Figures 8 and 9 show the recognition rates based on the receptive field responses. The first column displays the results for object "Chocos" and the second column shows the results for object "Robot". For each object two graphs are presented. The first one shows recognition rates with a fixed scale parameter, and the second one represents the recognition rate with automatic scale selection. The algorithm returns a list of hypothesis ordered with increasing distance to the observed receptive field

Fig. 7. Set of images of objects seen at different scales.

response. The three curves in the graphs are computed taking into account this hypothesis list and corresponds to three recognition cases:

- [*a*] The object corresponding to the answer with the smallest distance is correct.
- [*b*] The correct object is among the list of hypothesis, but other objects have a smaller distance.
- [*c*] Percentage of accepted neighborhoods. The description vector is rejected due to missing discrimination. The list of hypothesis is either empty or too large to be processed.

The percentage of accepted neighborhoods is very low for recognition in the case of a fixed scale parameter. Some neighborhoods have a quasi constant grey level, which leads to a very long list of hypothesis. These neighborhoods are ambiguous and not suitable for recognition. The automatic scale selection increases the percentage of accepted neighborhoods, because the scale is adapted to the feature. The figure 8 shows that recognition rate is unsatisfactory for scale variations above 20% whereas in figure 8 the recognition rate remains above 50%. Recognition is possible using a voting or a prediction-verification algorithm.

7 Conclusion

The appearance of features depends upon the scale of observation. In order to capture a maximum number of features, the scale of observation needs to be variable. This paper has shown very promising results of recognition under variable scale. Lindeberg proposes automatic scale selection for the determination of the appropriate feature scale. He assumes that maxima over scales of normalized derivatives reflect the scale of patterns. The selected scale corresponds to the Gaussian scale parameter at which the inner product of the derivative operator and the local image signal gives the strongest response. The application of this approach to all image neighborhoods allows the recognition of objects at

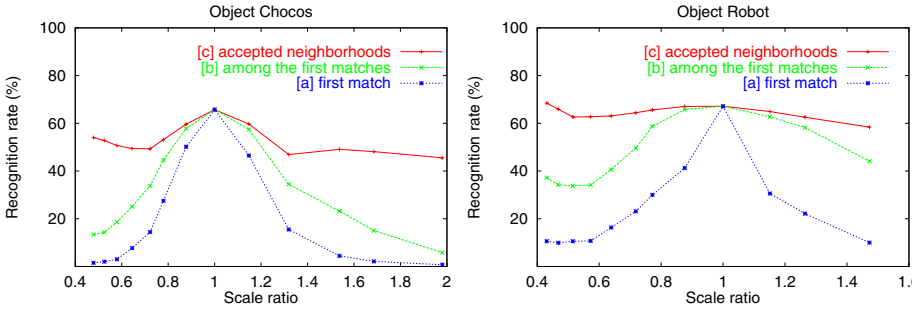

Fig. 8. *Recognition rate of object seen at different scales. The graphs abscissa is the scale ratio between the analysed image and the model image. Left graphs deals with the object "Chocos" and the right ones deals with the "Robot" object. Recognition is done with a fixed scale parameter.*

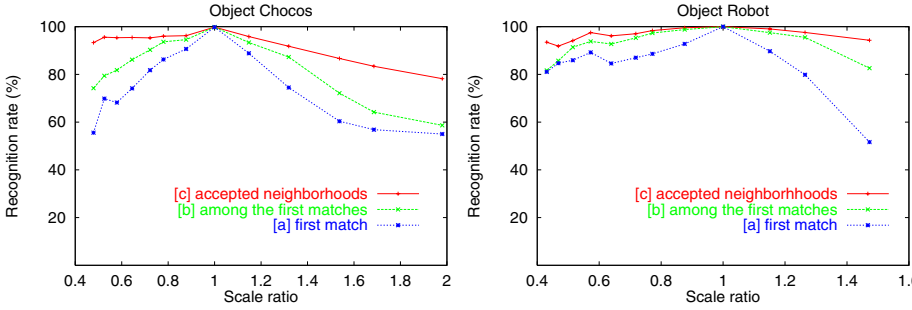

Fig. 9. *Recognition rate of object seen at different scales. The graphs abscissa is the scale ratio between the analysed image and the model image. Left graphs deals with the object "Chocos" and the right ones deals with the "Robot" object. Recognition is done with automatic scale selection.*

different scales. A map of appropriate scales is obtained that can be used to normalize the receptive fields. With steerable filters the dominant orientation of a neighborhood can be detected, which results in orientation invariance. Scale and orientation invariance are achieved by normalizing local descriptors. A remarkable gain in recognition and in the precision of the orientation normalization is achieved compared to the approach in which the feature type is ignored.

These results demonstrate that the approach is promising for real-world applications. The precise performances of the approach still requires a theoretical and quantitative evaluation, taking into account that the scale selection fails in some cases. There rest problems with point correspondences between images, in which the objects undergo important scale variations (factor 5 and more). Recognition requires that local patterns remain in a valid scale range between images. This range has to be evaluated and then the scale invariant recognition scheme must be applied on a set of images featuring high scale variations (larger

than three). The proposed approach has been tested with good results for image pairs with scale factor up to three.

Another interesting case is the detection of several characteristic scales for one feature. In this case several local extrema in the normalized Laplacian curves are present in function of the parameter σ. Considering all detected scales of a feature leads to a description that preserves a higher amount of information, which can result in a superior recognition system.

References

[AB91] E.H. Adelson and J.R. Bergen. *Computational Models of Visual Processing*, chapter The Plenoptic function and the elements of early vision. M.Landy and J.A.Movshons, Cambridge, 1991. MIT Press.

[BJ96] M. J. Black and A. D. Jepson. EigenTracking: Robust Matching and Tracking of Articulated Objects using a View-Based Representation". pages 329–342, 1996.

[CC98a] V. Colin de Verdière and J.L. Crowley. Visual recognition using local appearance. In *European Conference on Computer Vision*, pages 640–654, 1998.

[CC98b] V. Colin de Verdière and J.L. Crowley. Visual recognition using local appearance. In *Fifth European Conference on Computer Vision, ECCV '98*, Freiburg, June 1998.

[CC99a] O. Chomat and J.L. Crowley. Probabilistic recognition of activity using local appearance. In *Conference Computer Vision and Pattern Recognition*, 1999.

[CC99b] V. Colin de Verdière and J.L. Crowley. A prediction-verification strategy for object recognition using local appearance. Technical report, PRIMA Group, GRAVIR Lab, Grenoble, France, 1999. available at ftp://ftp.inrialpes.fr/pub/prima/.

[FA91] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *Pattern Analysis and Machine Intelligence*, 13(9), 1991.

[Fau93] O. Faugeras. *Three-Dimensional Computer Vision : A Geometric Viewpoint*. The MIT Press, 1993.

[Kv92] J.J. Koenderink and A.J. van Doorn. Generic neighborhood operators. *Pattern Analysis and Machine Intelligence*, 14(6):597–605, june 1992.

[Lin98a] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–154, 1998.

[Lin98b] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[MN95] H. Murase and S.K. Nayar. Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.

[MS65] M.bramowitz and I.A. Stegun. *Handbook of Mathematical Functions*, chapter 22. m.landy and j.a.movshons (eds) edition, Dover, New York, 1965.

[RB95] R.P.N. Rao and D.H. Ballard. An active vision architecture based on iconic representations. *Artificial Intelligence Journal*, 78, 1995.

[SC96] B. Schiele and J.L. Crowley. Probabilistic object recognition using multidimensional receptive field histograms. In *International Conference Pattern Recognition*, Vienna, 1996.

[Sch97] B. Schiele. *Object Recognition Using Multidimentional Receptive Field Histograms*. PhD thesis, Institut National Polytechnique de Grenoble, july 1997.

[SK87] I. Sirovich and M. Kirby. Low-Dimensional Procedure for the Caracterization of Human Faces. *J. Opt. Soc Am. A*, 4(3):519–524, March 1987.

[SM97] C. Schmid and R. Mohr. Local Grayvalue Invariants for Image Retrieval. 19(5):530–534, 1997.

[TP91] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[YV95] I.T. Young and L.J. Vliet. Recursive implementation of the gaussian filter. *Signal processing*, 44(2), 1995.