1991

P.F.M. Nacken

A metric for line segments

Department of Operations Research, Statistics, and System Theory Report BS-R9126 December

CWI, nationaal instituut voor onderzoek op het gebied van wiskunde en informatica

CWI is the research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a non-profit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands organization for scientific research (NWO).

A Metric for Line Segments

P.F.M. Nacken

CWI

Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

This paper presents a way to construct a metric for comparing symbolic features or tokens in images. These are representations of geometrical structures in the image. The metric is used to measure how strongly two features are related or how much they are alike. We will describe the case of line segments generated by an edge detection algorithm. In the case of line segments, such a metric measures collinearity and separation. Different metrics can be defined with the technique described in this paper by choosing different neighbourhood functions. By choosing a proper neighbourhood function, a metric can be constructed that satisfies the needs of a specific application. The approach will be illustrated by constructing a metric on the set of all line segments. The usefulness of this metric is verified by using it in a clustering algorithm for detecting linear structures in images.

AMS 1980 Mathematics Subject Classification: 51K05, 60D05, 68T10, 68U10.

Keywords and Phrases: Metric, clustering, function spaces, image feature, image analysis, line segments, linear structure.

Note: This research was supported by the Foundation for Computer Science in the Netherlands (SION) with financial support from the Netherlands Organisation for Scientific Research (NWO). This research was part of a project in which the TNO Institute for Perception (Kampweg 5, 3769 DE Soesterberg, The Netherlands), the Centre for Mathematics and Computer Science (Kruislaan 413, 1098 SJ Amsterdam, The Netherlands) and the Faculty of Mathematics and Computer Science of the University of Amsterdam (Kruislaan 403, 1098 SJ Amsterdam, The Netherlands) cooperate.

1. Introduction.

An important task in pattern recognition is the detection of structure in a configuration of individual tokens or symbolic features. These are representations of a geometrical structure at some location in the image plane. The geometric structures can be very simple features such as points, but they can also have some internal structure, which is for example the case for line segments or circular blobs. These geometric structures are considered as single objects, rather than as a collection of pixels in a special configuration.

For the detection of structure in feature configurations it is necessary to have some measure indicating how well two features fit together or how much they are alike. The Euclidean metric is a distance measure often used when the features are simple points in the plane or when points in a Euclidean space represent the measurements of several quantities of objects in the image. Several existing algorithms use this metric. For detecting structure in configurations of more complicated features, a more complex measure for comparing them is required. This measure must take into account not only the location of the features in the image plane, but also their internal structure. The measure to be used will depend on the specific application.

Report BS-R9126
ISSN 0924-0659
CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Many measures for comparing features like line segments have been proposed (e.g.[3, 4]). Most of them are not metrics. It is conceptually attractive to use a metric, since it is the mathematical abstraction of the notion of distance. This paper describes a way to construct measures on feature spaces which are indeed metrics and which provide a more flexible way for defining the concept of nearness for a given type of symbolic features. Metric spaces have an interesting structure, which has been investigated extensively in mathematics [1].

The rest of this paper is organised as follows: In section 2 some general aspects of metric spaces are discussed and some examples are given. Section 3 describes how neighbourhood functions can be used to construct metrics on the set of all features of a given type. The construction is illustrated by two simple examples. The metric for lines segments is discussed in section 4. In section 5 it is discussed how the metric defined in section 4 can be used for the extraction of image structure using a hierarchical clustering procedure. The results of this procedure are presented in section 6, where the usefulness of the metric is discussed. Section 7 presents the conclusions of this paper.

2. Metrics.

Before the construction of metrics on sets of image features is discussed, some general remarks on metric spaces are made. $\mathbb{R}_{\geq 0}$ is defined as the set of non-negative reals.

1. Definition. A metric space is a pair (X, d) where X is some set and d is a function $X \times X \longrightarrow \mathbb{R}_{\geq 0}$ such that for all x, y and z in X:

(1) $d(x, y) = 0 \Leftrightarrow x = y$.
(2) $d(x, y) = d(y, x)$.
(3) $d(x, y) + d(y, z) \geq d(x, z)$.

The best known metric is the Euclidean distance, our daily-life metric. This is a metric on the Euclidean space $\mathbb{R}^n$. The third condition in the definition of a metric is called the triangle inequality. It is this inequality that is not satisfied in many of the measures used for comparing tokens.

In metric spaces, the notion of *betweenness* can be defined.

2. Definition. A point y in a metric space is said to lie *between* two other points x and z if $d(x, y) + d(y, z) = d(x, z)$.

Note that the left hand side in this definition can only be equal to or larger than the right hand side because of the triangle inequality. Therefore, the points y between two points x and z have the lowest possible value for the sum of their distances to x and z.

A point y that lies between x and z is said to lie *strictly* between x and z if it is not equal to either of them. If x and y are two points in Euclidean space, the points between them are exactly the ones on the line segment connecting x and y.

Let Y be a subset of X, where (X, d_X) is a metric space. Then Y can be provided with a metric in a straightforward way: for each pair $(y_1, y_2) \in Y \times Y$, the distance is defined as $d_Y(y_1, y_2) = d_X(y_1, y_2)$. This metric on Y is called the *induced* metric.

If U is a real vector space provided with an inner product, a metric on U can be constructed in a standard way.

3. Definition. *Let U be a real vector space. An inner product on U is a mapping $U \times U \longrightarrow \mathbb{R} : (x, y) \mapsto \langle x|y \rangle$ satisfying:*

- $\langle x|y \rangle = \langle y|x \rangle$ *for all* $x, y \in U$.
- $\langle x + y|z \rangle = \langle x|z \rangle + \langle y|x \rangle$ *for all* $x, y, z, \in U$.
- $\langle \lambda x|y \rangle = \lambda \langle x|y \rangle$ *for all* $x, y \in U, \lambda \in \mathbb{R}$.
- $\langle x|x \rangle > 0$ *for all* $x \in U, x \neq 0$.

4. Definition. *If U is a real vector space, a norm on U is a mapping $U \to \mathbb{R} : x \mapsto ||x||$ satisfying*

- $||x|| > 0$ *for all* $x \neq 0$.
- $||\lambda x|| = |\lambda| \, ||x||$ *for all* $\lambda \in \mathbb{R}, x \in U$.
- $||x + y|| \leq ||x|| + ||y||$ *for all* $x, y \in U$.

If U is provided with an inner product $\langle \cdot | \cdot \rangle$, a norm can be defined by $||x|| = \sqrt{\langle x | x \rangle}$. It can easily be checked that this is indeed a norm by using the Cauchy-Schwartz inequality

$$|\langle x | y \rangle| \leq ||x|| \, ||y||.$$

It can also be shown that equality in the Cauchy-Schwartz inequality can only occur if x and y are linearly dependent. Both this statement and the Cauchy-Schwartz inequality can be proven by observing that $\langle x - \lambda y | x - \lambda y \rangle \geq 0$ for all values of λ. The left hand expression is a quadratic form in λ and achieves its minimal value for $\lambda = \langle x | y \rangle ||y||^{-2}$. Inserting this in the above relation gives the Cauchy-Schwartz inequality; it can readily be seen that equality in the Cauchy-Schwartz inequality is only achieved if $||x - \lambda y|| = 0$, from which the second statement follows.

If $|| \cdot ||$ is a norm on U, a metric d on U can be defined by $d(x, y) = ||x - y||$. It can easily be checked that this is indeed a metric.

In the next sections, metrics on the set of functions on a given manifold will be used. If U is a manifold provided with a measure μ, $\mathcal{L}^2(U)$ is the set of all real valued function classes f for which $\int_U f(t)^2 \mathrm{d}\mu(t)$ exists and has a finite value. (Actually, instead of using functions, one should use equivalence classes of functions, defined by the equivalence relation $f \sim g \Leftrightarrow \int_U (f(x) - g(x))^2 d\mu(x) = 0$. In this paper only continuous functions are used, so this has no practical consequences.) $\mathcal{L}^2(U)$ is a vector space. An inner product on this vector space is defined by

$$\langle f | g \rangle = \int_U f(t) g(t) \mathrm{d}\mu(t).$$

In the sequel, $\mathcal{L}^2(U)$ is provided with a norm $|| \cdot ||$ and a metric in the way described above. The metric is then described by:

$$d(f, g) = \left[\int_U (f(t) - g(t))^2 \mathrm{d}\mu(t) \right]^{1/2}.$$

Let f and g be two functions in $\mathcal{L}^2(U)$. It can be shown that the only functions that lie metrically between f and g are functions β for which $\beta = \lambda f + (1 - \lambda) g$ for some $\lambda \in [0, 1]$. The subset $\mathcal{S}(U)$ defined by $\{f \in \mathcal{L}^2(U) | \, ||f||_2 = 1\}$ is the unit sphere in $\mathcal{L}^2(U)$. From now on $\mathcal{S}(U)$ will be provided with the induced metric.

If f and g are two functions on the unit sphere, there is no function h on the unit sphere between them. If such a function would exist, it would also lie between f and g in the space $\mathcal{L}^2(U)$ itself; hence it would be of the form $h = \lambda f + (1 - \lambda) g$ for some $\lambda \in (0, 1)$. But then

$$\begin{aligned} ||h||_2^2 &= \lambda^2 ||f||_2^2 + (1 - \lambda)^2 ||g||_2^2 + 2\lambda(1 - \lambda) \langle f | g \rangle \\ &< \lambda^2 + (1 - \lambda)^2 + 2\lambda(1 - \lambda) ||f|| \, ||g|| \\ &= 1. \end{aligned}$$

The inequality follows from the fact that $\langle f | g \rangle = ||f|| \, ||g||$ is only true if $g = 0$ or $f = \lambda g$ for some positive λ. If f and g are different points on the unit sphere, this is not possible. Thus h cannot lie on the unit sphere. Note that this is completely analogous to the case of the unit sphere in Euclidean space.

3. Metrics on Sets of Features.

Let us now return to our original problem, where X is the set of all features of a given type, for example line segments. Assume that the feature set can be parametrised by points in a manifold U.

Line segments can, for example, be parameterised by coordinates $(x, y, \theta, l) \in \mathbb{R}^2 \times (\mathbb{R}/\pi\mathbb{Z}) \times \mathbb{R}^+$. The point (x, y) is the center of the line segment, the coordinate θ indicates that the direction of the line segment is $(\cos\theta, \sin\theta)$ (The begin and endpoints of a line segment are indistinguishable. Therefore θ and $\theta + \pi$ denote the same direction.) and l is the length of the line segment.

If U is embedded in some $\mathbb{R}^n$, the Euclidean metric on $\mathbb{R}^n$ induces a metric on U: the distance between two points of U is simply the Euclidean distance between the corresponding points in $\mathbb{R}^n$. This will in general not be a metric that defines the sense of nearness which is useful in specific application. In the case of line segments discussed above, the distance between two line segments with fixed direction and length depends only on the Euclidean distance between the two center points. The distance does not depend on the direction of the line connecting the centers of the two segments. This is not desirable, since the metric should be sensitive to the relation between the directions of the line segments and the direction of the line joining them. In general, the metric induced by an embedding does not work well, because it does not take into account the geometrical interpretation of the parameters for the features they describe.

If tokens are parameterised by coordinate vectors $(x_1, \ldots, x_n)$, their nearness can also be defined using thresholds. If a set of thresholds θ_i is defined, two tokens x and y are said to be related if $|x_i - y_i| < \theta_i$ for all i. In relation to this approach, a metric d can be defined as $d(x, y) = \max_i |x_i - y_i|/\theta_i$. Then two tokens are related if their distance is less than 1. This metric suffers from the same problems as the metric induced by the Euclidean metric, since the geometric interpretation of the coordinates is not taken into account.

Now a more flexible type of metric will be discussed. This metric is constructed by embedding the feature space U in the function space $\mathcal{S}(U)$. The metric in $\mathcal{S}(U)$ induces a metric in U. (For normalisation purposes, the metric will be divided by a factor $\sqrt{2}$.) With each point u of U one associates some function $f_u \in \mathcal{S}(U)$ which satisfies $f_u(x) \geq 0$. This function will be called the *neighbourhood function* of u. The neighbourhood functions are chosen such that different points in U have different neighbourhood functions. As an illustration, consider the case where the feature space is the set $\mathbb{R}$ of real numbers. Let the neighbourhood function of u be the indicator function of an interval of length 1 centered at u: $f_u(x) = 1_{[u-1/2, u+1/2]}$.

The feature set has now been identified with a subset of $\mathcal{S}(U)$. We call this subset $\mathcal{F}(U)$. $\mathcal{F}(U)$ is provided with the induced metric, divided by $\sqrt{2}$ for normalisation purposes. This means

$$d(u, v) = \left[\frac{1}{2}\int_U ((f_u(t) - f_v(t))^2 \mathrm{d}\mu(t)\right]^{1/2}.$$

For the example mentioned above, it can easily be seen that $\int_{\mathbb{R}} f_u(x) f_v(x) \mathrm{d}x = \max(0, 1 - |u - v|)$. The metric thus defined on $\mathbb{R}$ is

$$d(u, v) = \sqrt{1 - \max(0, |u - v|)}.$$

This metric is illustrated in figure 1. There is translation invariance. $f_x(y)$ and $d(x, y)$ depend only on $x - y$. The upper curve shows $f_x(y)$ as a function of $x - y$. The lower curve shows $d(x, y)$ as a function of $x - y$.

Another metric on $\mathbb{R}$ can be obtained by taking the neighbourhood function

$$f_u(x) = \pi^{-1/2} e^{-(x-u)^2}.$$

The metric defined by this function is

$$d(u, v) = \sqrt{1 - e^{-(u-v)^2/2}}.$$

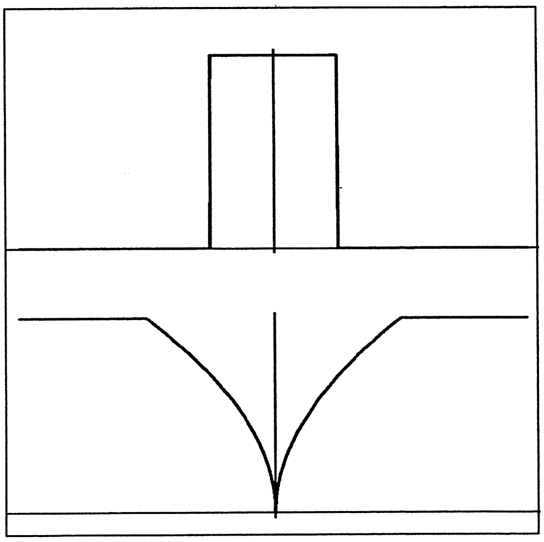


figure 1

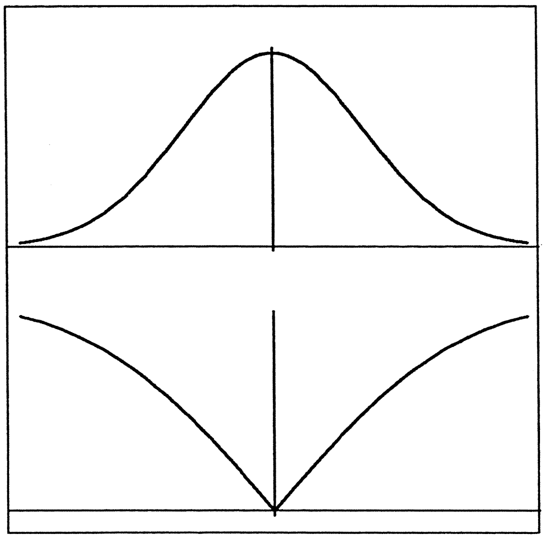


figure 2

This metric is illustrated in figure 2. There is again translation invariance, so $f_x(y)$ and $d(x,y)$ depend only on $x-y$. The upper and lower curves show $f_x(y)$ and $d(x,y)$ as a function of $x-y$, respectively. Note that the use of a Gaussian neighbourhood function has the practical advantages that the product of two Gaussians is again a Gaussian and that the integral of a Gaussian over $\mathbb{R}$ can be expressed in a closed form.

Let us now return to the case of line segments and try to develop an intuition for the rôle of the neighbourhood function in the construction of the metric. We have:

$$\begin{aligned} 2d(u,v)^2 &= \|f_u\|_2^2 + \|f_v\|_2^2 - 2\int_U f_u(t)f_v(t)\mathrm{d}\mu(t) \\ &= 2 - 2\int_U f_u(t)f_v(t)\mathrm{d}\mu(t). \end{aligned}$$

The neighbourhood functions are nonnegative. Hence the maximal value of $d(u,v)$ is 1. The distance decreases when $\int_U f_u(t)f_v(t)\mathrm{d}\mu(t)$ increases.

Now suppose that an image is created by measurement of some real linear structures in the scene. The measurement process is not perfect, so deformations occur. The neighbourhood function $f_u(t)$ is interpreted as the probability that a measured line segment t stems from measurement on a true structure t in the scene:

$$f_u(t) = P(X_{\mathrm{true}} = t | X_{\mathrm{meas}} = u).$$

In this interpretation, the neighbourhood function contains information on both the distribution of linear structures in the scene and on the characteristics of the measurement process. No knowledge of the scene is assumed. Therefore the distribution of linear structures in the scene is assumed to be described by a stochastic process with a uniform distribution. The neighbourhood function also depends on the corruptions which can be caused by the measurement process. This corruption can be the breaking of long structures in smaller segments and the displacement of position and orientation of these segments.

If it is assumed that different line segments in the image are generated by independent measurements of the scene, $f_u(t)f_v(t)$ is a measure for the probability that both u and v are generated by a measurement of the structure t. Integrating this product over t gives a measure for the probability that u and v have the same origin. This integral can be regarded as a measure of nearness of the two features.

This interpretation gives a hint on how to choose the neighbourhood functions: $f_u(t)$ should be a measure for the chance that a measured structure u is caused by a true structure t. Note that the interpretation discussed above is just an aid for choosing a neighbourhood functions. One can always put restrictions on the neighbourhood functions in order to provide the metric with some special property or in order to make computations more easy.

In many cases, the relation between two features is invariant under the operation of some group G. In the case of line segments, for example, collinearity and nearness only depend on the relative positions of the features; they are invariant under translations and rotations of the whole feature configuration. It is natural to choose a measure on U which is invariant under the symmetry group and neighbourhood functions which are transformed into each other under the action of the group.

If G is transitive, only one neighbourhood function needs to be defined; the others can be constructed from this function. This reduces the number of choices to be made significantly. However, in many applications there will be not so many symmetries that the symmetry group becomes transitive. In the case of line segments, for example, there will be invariance under translation and rotation, but no scale invariance. Therefore, a different neighbourhood function must be defined for each l_0. The neighbourhood functions belonging to different x_0, y_0 and θ_0 are then fixed by these choices through the action of the group of Euclidean motions.

Note that the interpretation of the neighbourhood functions f actually requires that $\int_U f_u(t)\mathrm{d}t = 1$ or at least that the integral has the same value for all u. The construction however required that $\int_U {f_u}^2(t)\mathrm{d}t = 1$ for all u. If the symmetry group of the problem is transitive, the ratio between the integral of f_u and the integral of its square is fixed, as can be derived from the fact that all neighbourhood functions can be transformed into each other by a simple coordinate transformation. If the value of $\int_U f_u(t)\mathrm{d}t$ has a value not equal to one, this value can be regarded as a normalisation factor which forces the metric to have values in $[0, 1]$.

If the symmetry group is not transitive, it may still be possible to choose the neighbourhood functions in such a way that $\int_U f_u(t)\mathrm{d}t$ does not depend on u. In the example of line segments discussed in the next sections, the neighbourhood functions will be chosen in such a way that they satisfy this condition.

Note that the integral of f_u being constant is not needed for the construction of the metric itself. It allows an interpretation in terms of probabilities and thus is helpful in the choice of the neighbourhood functions.

It is also possible to interpret $f_u^2(t)$ in stead of $f_u(t)$ as the probability $P(X_{\mathrm{true}} = t | X_{\mathrm{meas}} = u)$. Then each measured segment u defines a probability distribution $f_u^2(x)$ on the parameter space U. The distance between u and v defined in this paper is then equal to the so-called Hellinger distance [4] between the probability distributions $f_u(x)$ and $f_v(x)$.

4. The Metric on Line Segments.

In this section, a metric on the set of all line segments is constructed. If an edge detector followed by a line fitting algorithm is used to calculate linear structures in images, distortions can occur. Lines can be interrupted or displaced from their original place and orientation. Additional processing is then necessary for the extraction of the complete linear structure in the scene. Therefore, a metric on the set of all line segments must be constructed which measures to what extent two line segments are likely to belong to the same linear structure.

Now let $u_0 = (x_0, y_0, \theta_0, l_0)$ be the coordinate representation of a line segment in the parametrisation discussed previously. In order to define a metric, the neighbourhood function $f_{u_0}(x, y, \theta, l)$ must be defined, which indicates how likely it is that a given measured segment u_0 stems from a measurement on a true linear structure with coordinates (x, y, θ, l).

Note that there is a desirable property of the metric which imposes an extra condition on the form of the neighbourhood function. If two line segments lie on the same support line while they are separated by a gap which is large compared with their own lengths, the distance between them should be large, even if they could be caused by a single, very long, linear structure in the scene. If such a long structure is really present, it is expected that the gap is filled with other line segments, making a detection of the linear structure in the scene possible. This condition on the metric is fulfilled if it is assumed that the centers of the true and measured segments are always close together (compared with the length of the measured segments), even if the true structure is very long. This can be interpreted as *local measurement*: the segment measurement process 'sees' just the central part of the (long) linear structures in the scene; the length distribution of the measured segments does not depend on the lengths of the linear structures in the scene.

We will take a neighbourhood function which has a maximum at the point u_0 itself and which decays towards 0 when moving away from this point. The neighbourhood function is constructed using Gaussians, which have a well-defined width. This choice implies that we assume that the orientation and the position of a measured line segment are normally distributed around the position and orientation of the underlying structure in the scene.

We can write:

$$P(X_{\mathrm{true}} = x | X_{\mathrm{meas}} = u) = \frac{P(X_{\mathrm{meas}} = u | X_{\mathrm{true}} = x)}{P(X_{\mathrm{meas}} = u)} \times P(X_{\mathrm{true}} = x).$$

The prior distribution of position and orientation of the linear structures in the scene is assumed to be uniform. Hence the second factor in the right hand side depends only on l_x. The local measurement property discussed above implies that the first factor does not depend on the length l_x of the linear structure in the scene. Hence we can take $f_u(x)$ of the form

$$f_u(x) = \psi_u(x_x, y_x, \theta_x)\phi(l_x).$$

There is no need to specify ϕ since

$$\int_U (f_{u_0} - f_{v_0})^2 \mathrm{d}\mu(t) = \int_{\mathbb{R}^+} \phi(l)^2 \mathrm{d}l \int (\psi_{u_0} - \psi_{v_0})^2 \mathrm{d}x\mathrm{d}y\mathrm{d}\theta$$

and $\int \phi(l)^2 \mathrm{d}l$ is a constant.

The notion of nearness one wants to quantify with the metric can be investigated to find some properties of the neighbourhood functions.

- The measured direction of the line segments can be corrupted. Hence, a line segment should be "near" to other line segments at the same position but having a somewhat different direction. This is expressed by the width of the neighbourhood function in the θ-direction (in U). The direction measured for a long line segment is a better indication for the direction of the underlying linear structure than the direction measured for a short segment. Therefore the width in the θ-direction of the neighbourhood function a short segment is larger than the width in the neighbourhood function of a long segment.
- If a linear structure is broken into pieces, several short segments can result. The midpoints of these segments are displaced from the center of the original structure, in the direction of the linear structure. Therefore the neighbourhood function of a line segment must have a relatively large width (proportional to the length of the measured line segment) in the direction in which the linear structure lies.
- A line segment can also be displaced from the linear structure in the direction orthogonal to the structure. The displacement in this direction will not be too large and does not depend on the length of the line segment.

The neighbourhood function is now chosen as follows: For fixed θ, the function $\psi_{u_0}(x, y, \theta)$ is a two-dimensional Gaussian function centered at (x_0, y_0) with its long axis oriented in the θ-direction of the $x - y$-plane. The width σ_{long} in the θ-direction is proportional to the length l_0 of the line segment. The width σ_{short} is fixed to some small value. The amplitude of the Gaussians varies with direction: it is large for θ close to the direction θ_0, while it is small for values θ further away from the line direction θ_0.

Thus, the neighbourhood function is described by:

$$f_{u_0}(x,y,\theta,l) = G_{\sigma_{\text{angle}}}(\theta-\theta_0)G_{\sigma_{\text{long}}\sigma_{\text{short}}}(R_{\theta_0}^{-1}(x-x_0,y-y_0))\phi(l)$$

Here $G_\sigma(x)$ is a Gaussian of width σ:

$$G_\sigma(x) = \pi^{-1/4}\sigma^{-1/2}e^{-(x/\sigma)^2}.$$

G is a two-dimensional Gaussian function:

$$G_{\sigma_x\sigma_y}(x,y) = \frac{1}{(\pi\sigma_x\sigma_y)^{-1/2}}e^{-(x/\sigma_x)^2-(y/\sigma_y)^2}.$$

R_θ is a rotation over θ.

The spread in the θ-direction is defined by a Gaussian function with a width σ_{angle}. In the actual implementation, the integration in the θ-direction is approximated by discretising the θ-values, while the integration in the x- and y-directions is performed using the analytical expressions for the integral of a Gaussian function. As mentioned before, the factor ϕ cancels in later calculations.

The symmetry group underlying this situation is the group of Euclidean motions. There is no scale invariance, since the uncertainty in the orientation of a segment depends on its length. Note that, once the choice for Gaussian neighbourhoods is made, there are only three degrees of freedom left: σ_{long}, σ_{short} and σ_{angle}. Note that the values of these parameters depend only on the length l_0 of the line segment. The value of $\int_U f_{u_0}(t)\mathrm{d}t$ is

$$(2\pi)^{3/4}(\sigma_{\text{long}}\sigma_{\text{short}}\sigma_{\text{angle}})^{1/2}\int_0^\infty \phi(x)\mathrm{d}x.$$

This value does not depend on u_0 if the product $\sigma_{\text{long}}\sigma_{\text{short}}\sigma_{\text{angle}}$ does not depend on l_0. In the next section, the values of σ_{long}, σ_{short} and σ_{angle} will be chosen such that this condition is satisfied.

5. Segment Clustering Using the Metric.

To verify the usefulness of the metric defined in the previous section, a hierarchical clustering method based on this metric is examined. The clustering method used is a standard hierarchical clustering method [2]. It acts on a configuration of a number of features. The algorithms proceeds step by step. In each step, the two nearest features are replaced by a single one, such that a hierarchy of configurations with smaller and smaller numbers of features is created.

Let $\{s_1,\ldots,s_n\}$ be the set of all features in the original configuration. Let $L^{(k)} = \{s_1^{(k)},\ldots,s_{n-k}^{(k)}\}$ be the k-th configuration in the hierarchy. $L^{(0)}$ is equal to $\{s_1,\ldots,s_n\}$. Successive levels are built by replacing the feature pair having minimal distance by a single feature:

$$L^{(k+1)} = L^{(k)}\backslash\{s_p^{(k)},s_q^{(k)}\}\cup\{c(s_p^{(k)},s_q^{(k)})\}.$$

Here $s_p^{(k)}$ and $s_q^{(k)}$ are chosen such that $d(s_p^{(k)},s_q^{(k)}) = \min_{i\neq j} d(s_i^{(k)},s_j^{(k)})$ and $c(s_p^{(k)},s_q^{(k)})$ is the *join* of $s_p^{(k)}$ and $s_q^{(k)}$, a line segment which replaces the original two segments.

When clustering in a convex feature space like $\mathbb{R}^n$, a very simple choice for the calculation of joins can be made, such as $c(\vec{x},\vec{y}) = \frac{1}{2}(\vec{x}+\vec{y})$. When clustering complex features such as line segments, the construction of joins is more involved.

In the general case, it would be nice if two points u and v could be replaced by a point that lies strictly between them. In the present case, this is not possible. Since all functions belonging to features lie in the unit sphere, there are no features strictly between two different features.

We have chosen the following approach. For each pair of line segments to be joined, a line is calculated such that the sum of the squares of the distances of the four endpoints of the segments to the line is minimal. The join of the two segments is a segment of this line. The end points of the join are calculated by looking at the orthogonal projections of the two segments on the line. The join is the smallest segment of the line which contains both projections.

6. Experimental Results

The clustering method is applied to a number of line segment configurations, derived from natural images. A sobel gradient filter is applied to the original image. The edges in the image are detected by a peak filter. The edges detected by the peak filter are one pixel thick. Line segments are then fitted to the curves by connecting the first and the last point. If a curve is longer than a given number of pixels (in this case 6), the curve is cut into shorter pieces before fitting the line segments. In this way, line segments are fitted with reasonable accuracy to curves which contain bends or angles. Moreover, this procedure simulates corruption of the line segment configuration by breaking longer straight curves. In a practical application one would use a more advanced line segment fitting algorithm which can produce longer segments, such that the task of clustering algorithm is made easier.

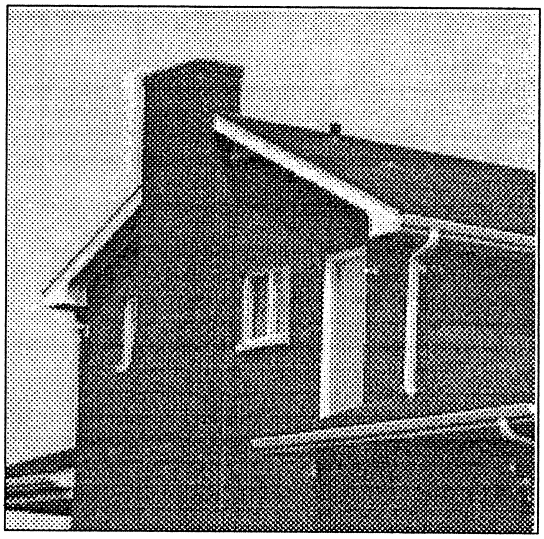


figure 3: the house image

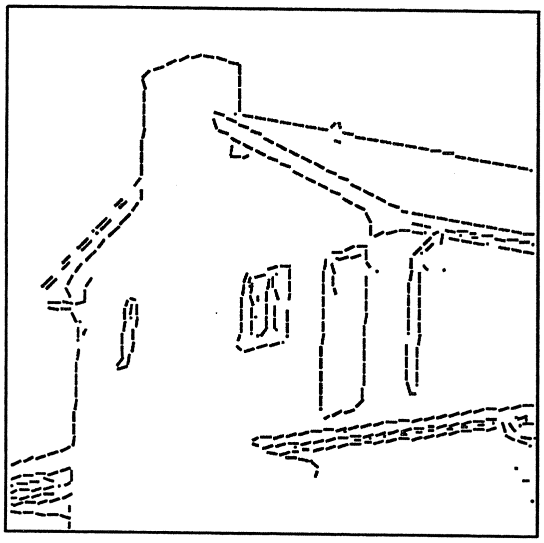


figure 4: extracted segments

The size of the original image is 256 × 256 pixels. The parameters for the neighbourhood functions were defined by:

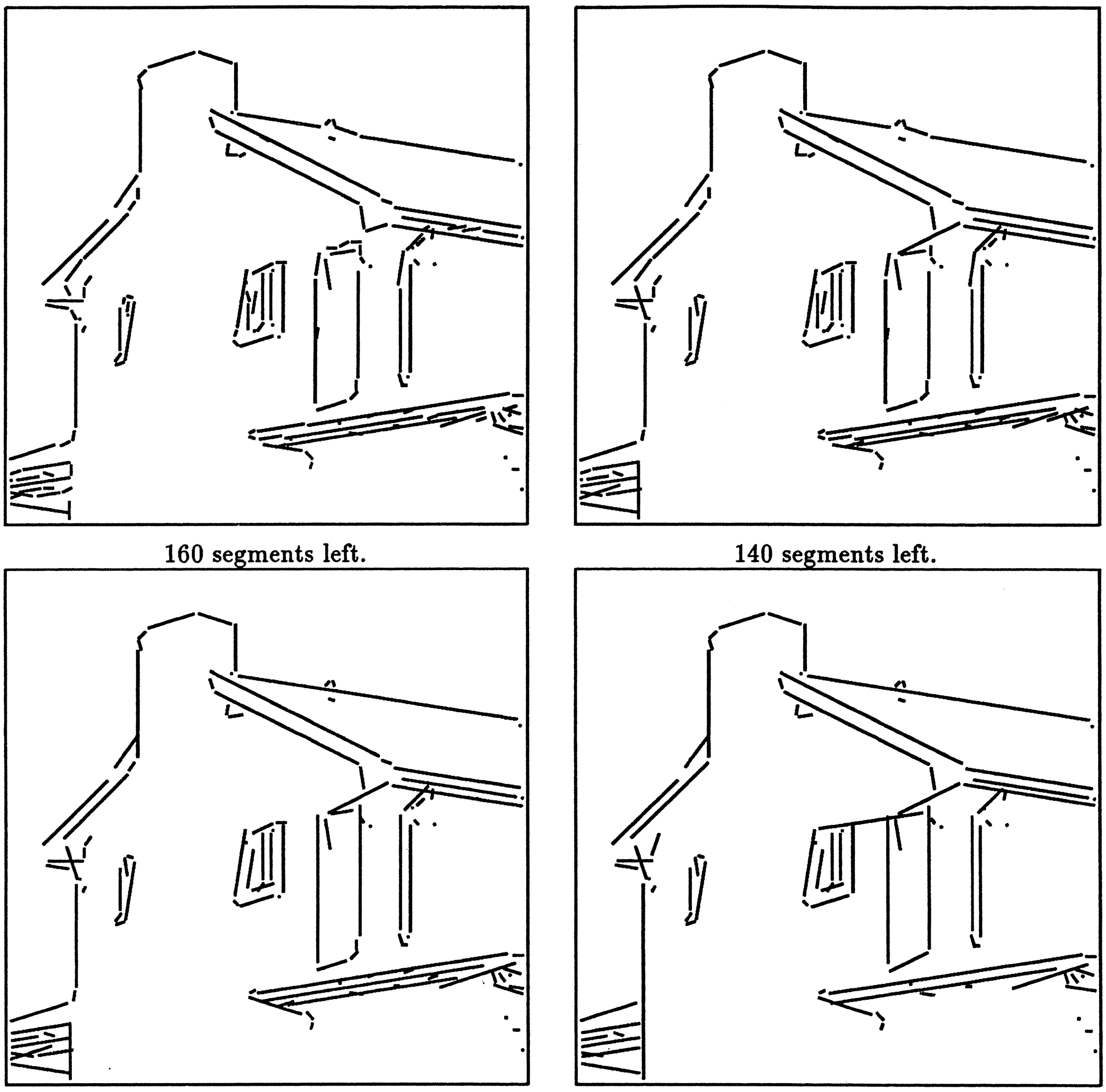
$$\sigma_{\mathbf{length}} = p \times \sqrt{L * l_0}$$
$$\sigma_{\mathbf{width}} = q$$
$$\sigma_{\mathbf{angle}} = r/\sqrt{L * l_0}$$

The choice of these parameters reflects the assumptions on the behaviour of the edge detection algorithm. The parameter $\sigma_{\mathbf{length}}$ is related to the length of the gaps that can exist in a linear structure. $\sigma_{\mathbf{length}}$ grows less than proportionally with the length of the segment. This implies that the gap between short segments can be relatively long with respect to the segment length. The parameter L determines the scale at which relatively long gaps are no longer allowed. In all examples shown in this paper, L is equal to 30.

The parameter $\sigma_{\mathbf{angle}}$ is inversely proportional to the root of the segment length. This implies that the direction of a short segment is assumed to be less exact than the direction of a longer segment.

The choice of parameters described above satisfies the conditions previously derived: The value of $\sigma_{\mathbf{length}}\sigma_{\mathbf{width}}\sigma_{\mathbf{angle}}$ is fixed. The parameters which still have to be fixed are (p, q, r). Our first example is the house image (figure 3). The initial configuration (figure 4) contains 415 line segments. The parameters are $(p, q, r) = (2, 2, 30)$. Figure 5 shows the stages of the clustering process in which 160, 140, 120 and 100 segments are left.

When 160 segments are left, the linear structures begin to show quite clearly. When 120 segments are left, the detection of linear structure in the image is completed.



160 segments left. 140 segments left.

120 segments left. 100 segments left.

figure 5: results from the clustering algorithm

There are some short stray segments left. When these segments participate in the clustering process in later stages, unnatural results are found, as can be seen from the situation where 100 segments are left. It is therefore best to stop the clustering before undesirable clusterings occur. If it is undesirable to cluster two line segments into a single one, their distance will be large. Practical examples also indicate that undesirable clusterings often occur if one of the two segments is a short noise segment which is not the result of previous clusterings, but has survived for a long time on its own without participating in the process.

If the clustering process is stopped before unnatural clusterings occur and the remaining short segments are removed from the resulting configuration, the quality of the resulting description is good enough for passing it to other high-level algorithms.

In order to examine the range of images for which the method gives good result, the clustering is applied to two other images from the grandfield benchmark set, which contain curved structures. The results are shown in figure 6. The pictures show, from top to bottom, the original image, the initial line segment configuration and the line configuration which is reached after a number of clustering operations, just before unwanted clusterings start to occur.

In all cases the choice of parameters was $(p, q, r) = (2, 2, 30)$. The linear structures in all images are

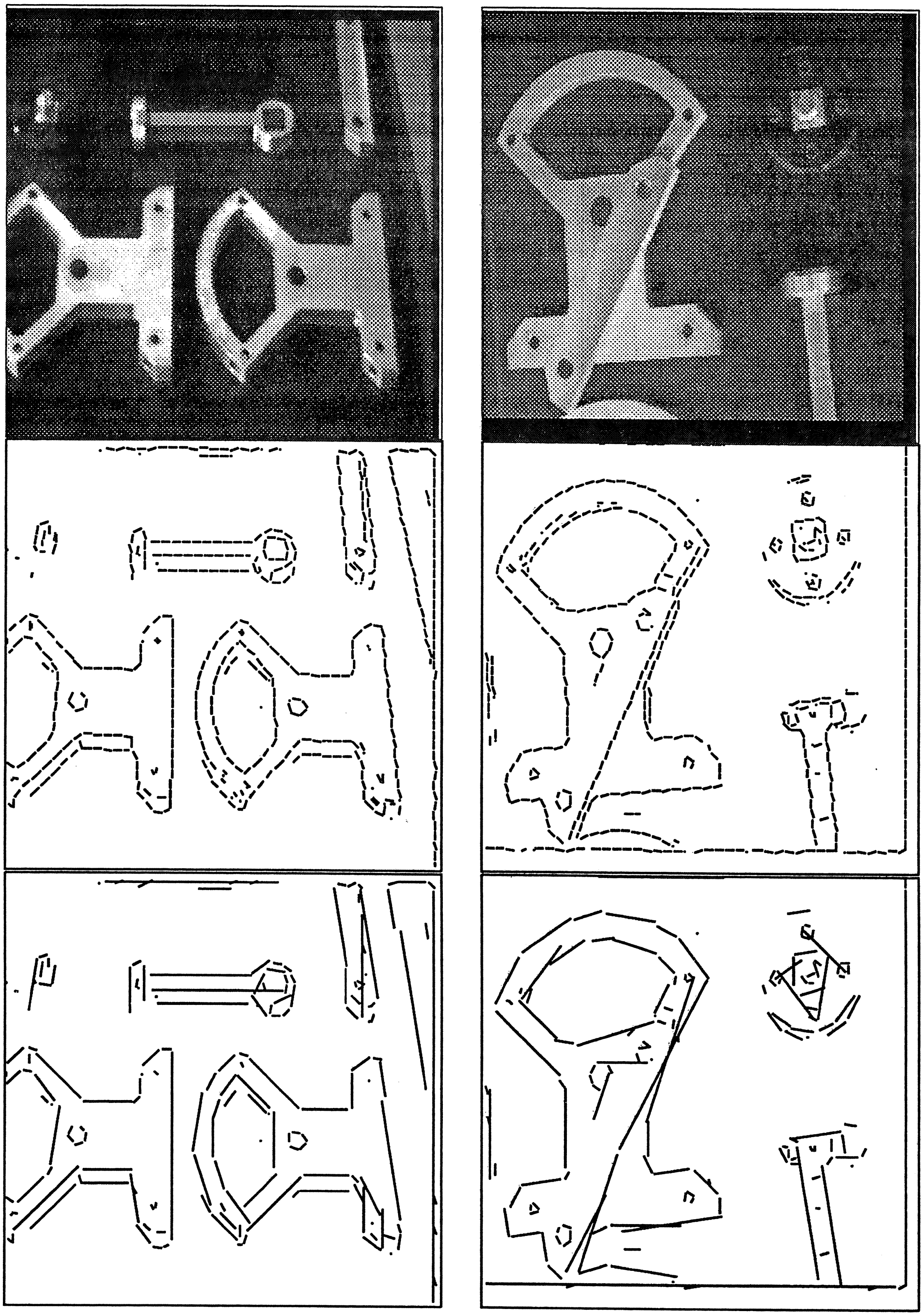
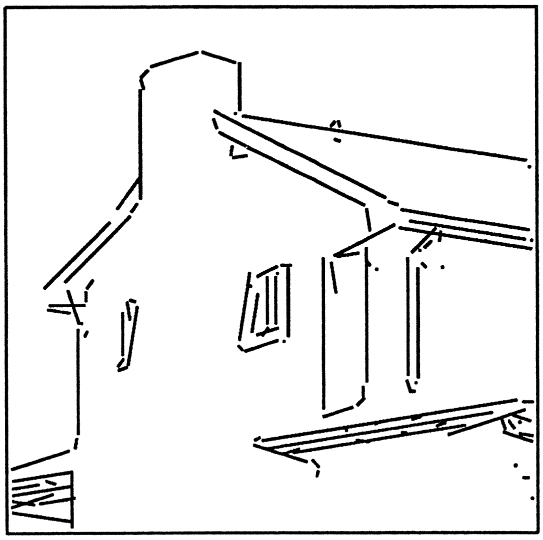


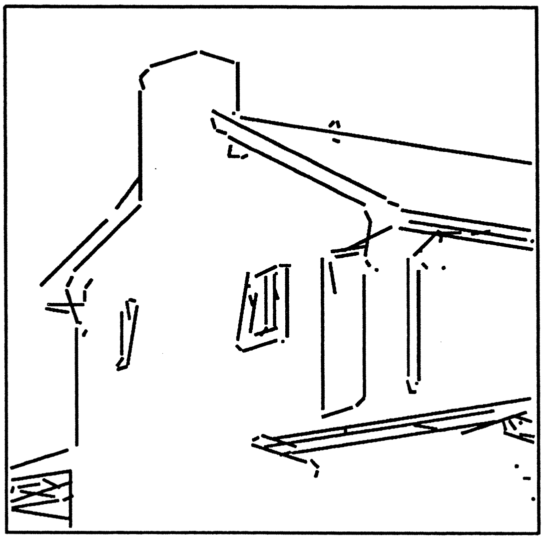
figure 6: result for different images.

detected quite well. If curves are present, a number of short segments is fitted along these curves. The algorithm can get confused in the present of much detail, as can be seen, for example, in the upper right hand corner of the third image.

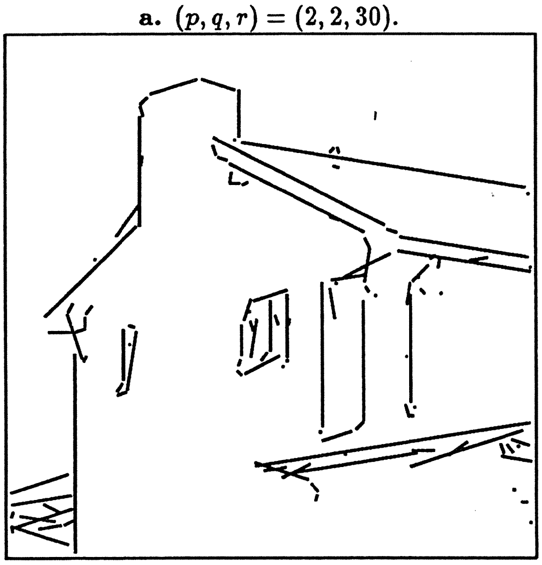
In order to test the sensitivity to parameter changes, the clustering method is applied to the house image with the parameter sets (2, 2, 30), (3, 2, 20), (2, 4, 30) and (2, 2, 40). The results are shown in figure 7. For all parameter settings, the linear structures are detected reasonably well. The are no large differences between the results. The most striking effect can be seen in figure 7c, in which the spread of the neighbourhood functions is larger in the direction orthogonal to the line segments. The result is that parallel lines have a stronger tendency to cluster compared with the other parameter settings.



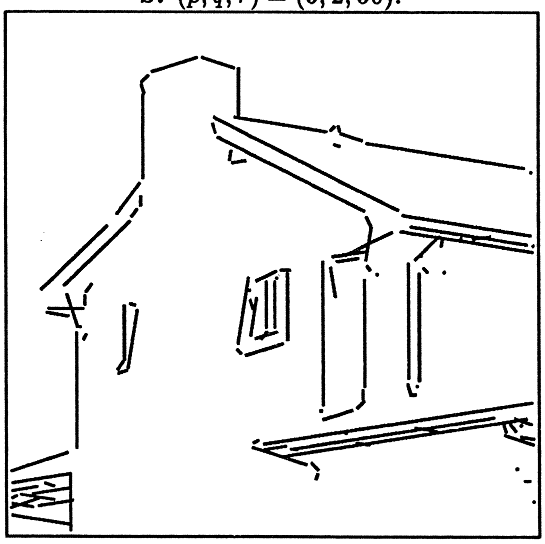
a. $(p, q, r) = (2, 2, 30)$.



b. $(p, q, r) = (3, 2, 30)$.



c. $(p, q, r) = (2, 4, 30)$.



d. $(p, q, r) = (2, 2, 40)$.

figure 7: results for different parameter settings.

7. Conclusions

This paper has presented a technique for constructing a metric for comparing symbolic features in images. The technique for constructing metrics on sets of features is founded on well known mathematical techniques. The construction of the metric indicates a way in which the notion of nearness that is desired for a specific application can be implemented in the metric by choosing appropriate neighbourhood functions.

The technique allows a limitation of the number of degrees of freedom. The number of degrees of freedom that are left decreases when the symmetry of the problem increases. It is our hope that by using the metrics we propose, the arbitrariness and the number of "free parameters" (especially thresholds) in pattern recognition algorithms can be reduced. In the case of line segments, for example, choices have to be made only for the different lengths occurring in the feature set.

Experiments with a metric on the set of all line segments have shown that the type of metrics discussed in this paper can be useful in a practical situation. It is possible to detect linear structures in images by a clustering method using the metric. The performance of this algorithm can be called satisfactory, especially if it is taken into account that the algorithm is conceptually very simple. In the last stages of the clustering, however, over-clustering occurs. Therefore a decision criterion is needed that determines when the clustering should stop because all linear structures have been detected.

Acknowledgements.

The author would like to thank F. Groen, H. Heijmans and L. Toet for many useful discussions and the thorough reading of the manuscript.

References.

[1] L.M. Blumenthal, *Theory and Applications of Distance geometry*, Oxford University Press, Oxford, U.K (1953)

[2] Anil K. Jain, Richard C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, Englewood Cliffs, N.J., U.S.A. (1988)

[3] Eric Saund, *Symbolic Construction of a 2-D Scale-Space Image*, Trans. PAMI **12** no. 8 (August 1990)

[4] Rakesh Mohan, Ramakant Nevatia, *Using Perceptual Organisation to Extract 3-D Structures*, IEEE Trans. PAMI **11** no. 11 (November 1998)

[4] L. Le Cam, *Notes on Asymptotic Methods in Statistical Decision Theory*, Centre de Recherches Mathématiques, Université de Montréal, Canada (1974).