Lappeenrannan teknillinen korkeakoulu
Lappeenranta University of Technology

Ville Kyrki

LOCAL AND GLOBAL FEATURE EXTRACTION FOR INVARIANT OBJECT RECOGNITION

Thesis for the degree of Doctor of Science (Technology) to be presented with due permission for public examination and criticism in the Auditorium of the Student Union House at Lappeenranta University of Technology, Lappeenranta, Finland on the 5th of December, 2002, at noon.

Acta Universitatis
Lappeenrantaensis
133

Supervisor	Professor Heikki Kälviäinen Laboratory of Information Processing Department of Information Technology Lappeenranta University of Technology Finland
Reviewers	Professor John Illingworth Centre for Vision, Speech, and Signal Processing Department of Electronic and Electrical Engineering University of Surrey United Kingdom
	Professor Jouko Lampinen Laboratory of Computational Engineering Department of Electrical and Communications Engineering Helsinki University of Technology Finland
Opponent	Academy Professor Erkki Oja Laboratory of Computer and Information Science Helsinki University of Technology Finland

ISBN 951-764-698-4
ISSN 1456-4491

Lappeenrannan teknillinen korkeakoulu
Digipaino 2002

Preface

The work presented in this thesis has been carried out at the Laboratory of Information Processing in the Department of Information Technology of the Lappeenranta University of Technology, Finland, during the years 1999-2002.

I want to express my gratitude to my supervisor, Professor Heikki Kälviäinen, for his supportive guidance as well as providing me the facilities and financial support for my research.

I would like to warmly thank all other co-authors of the publications, Professor Pasi Fränti, Mr Joni Kämäräinen, Mr Alexey Mednonogov, and Mr Jani Peusaari. In addition, I would like to express my thanks to Professor Jari Porras and Dr Jouni Ikonen for their contribution in the earlier phases of the research on parallel Hough transform.

I thank the reviewers, Professor John Illingworth and Professor Jouko Lampinen, for critical reading of the manuscript and valuable comments.

I wish to thank also the colleagues at the Laboratory of Information Processing at Lappeenranta University of Technology for a stimulating atmosphere and CSC - Scientific Computing for providing parallel computing resources.

The financial support of the East Finland Graduate School on Computer Science and Engineering (ECSE), KAUTE foundation (Kaupallisten ja teknillisten tieteiden tukisäätiö), Jenny and Antti Wihuri fund, Foundation of the Lappeenranta University of technology, and Foundation of the Advance in Technology (Tekniikan edistämissäätiö) is gratefully acknowledged.

Finally, I wish to thank my dear wife Anna for her endless encouragements.

Lappeenranta, October 2002

Ville Kyrki

Abstract

Ville Kyrki
LOCAL AND GLOBAL FEATURE EXTRACTION FOR INVARIANT OBJECT RECOGNITION
Lappeenranta, 2002
115 p.

Acta Universitatis Lappeenrantaensis 133
Diss. Lappeenranta University of Technology

ISBN 951-764-698-4
ISSN 1456-4491

Perceiving the world visually is a basic act for humans, but for computers it is still an unsolved problem. The variability present in natural environments is an obstacle for effective computer vision. The goal of invariant object recognition is to recognise objects in a digital image despite variations in, for example, pose, lighting or occlusion.

In this study, invariant object recognition is considered from the viewpoint of feature extraction. The differences between local and global features are studied with emphasis on Hough transform and Gabor filtering based feature extraction. The methods are examined with respect to four capabilities: generality, invariance, stability, and efficiency. Invariant features are presented using both Hough transform and Gabor filtering. A modified Hough transform technique is also presented where the distortion tolerance is increased by incorporating local information. In addition, methods for decreasing the computational costs of the Hough transform employing parallel processing and local information are introduced.

Keywords: feature extraction, invariant object recognition, Hough transform, Gabor filtering, computer vision, image analysis

UDC 004.93'1 : 004.272.2 : 004.932.2

$a(k)$	Fourier descriptor
$g[x]$	transformation belonging to group $\mathcal{G}$ acting on signal x
$H(\rho, \theta)$	Hough transform accumulator
m_{pq}	geometric moment of order $(p+q)$
$r_\xi(t)$	response of 1-d Gabor filter
$r_\xi(x, y)$	response of 2-d Gabor filter
$T\{x\}$	measurement of signal x
$\mathfrak{F}\{f\}$	Fourier transform of f
η_{pq}	normalised central moment of order $(p+q)$
μ_{pq}	central moment of order $(p+q)$
ϕ	moment invariant
$\psi(t)$	1-d Gabor elementary function
$\psi(x, y)$	2-d Gabor elementary function
$\Psi(f)$	1-d Gabor elementary function in Fourier domain
$\Psi(u, v)$	2-d Gabor elementary function in Fourier domain

CRHT	Connective Randomized Hough transform
GEF	Gabor elementary function
GHT	Generalised Hough transform
HT	Hough transform
LFA	Local feature analysis
PCA	Principal component analysis
PE	processing element
RHT	Randomized Hough transform
SIMD	Single-instruction multiple-data
VLSI	Very large scale integration

I. Kyrki, V., Kälviäinen, H., "Combination of Local and Global Line Extraction", *Journal of Real-Time Imaging*, Vol. 6, No. 2, April 2000, pages 79–91.

II. Kyrki, V., Kälviäinen, H., "High Precision 2-D Geometrical Inspection", *Proceedings 15th International Conference on Pattern Recognition*, Barcelona, Spain, September 3–7, 2000, Vol. 4, pages 779–782.

III. Fränti, P., Mednonogov, A., Kyrki, V., Kälviäinen, H., "Content-Based Matching of Line-Drawing Images Using Hough Transform", *International Journal on Document Analysis and Recognition*, Vol. 3, No. 2, 2000, pages 117–124.

IV. Kyrki, V., Kämäräinen, J., Kälviäinen, H., "Invariant Shape Recognition Using Global Gabor Features", *Proceedings of the 12th Scandinavian Conference on Image Analysis SCIA 2001*, Bergen, Norway, June 11–14, 2001, pages 671–678.

V. Kyrki, V., Kämäräinen, J.-K., Kälviäinen, H., "Content-based Image Matching Using Gabor Filtering", *Advanced Concepts for Intelligent Vision Systems, ACIVS 2001*, Baden-Baden, Germany, July 30–August 3, 2001, pages 45–49.

VI. Kyrki, V., Peusaari J., Kälviäinen, H., "Intermediate level feature extraction in novel parallel environments", accepted for publication in *Machine Vision and Applications.*

In this thesis these publications are referred as *Publication I*, *Publication II*, *Publication III*, *Publication IV*, *Publication V*, and *Publication VI.*

Chapter I

Introduction

Efforts have been made over many years to equip computers with vision. Object recognition is a classical problem in computer vision that aims to categorise objects to known object classes using a digital image. It can be utilised in a vast number of applications including industrial, medical, and scientific ones.

Already in 1960, Selfridge and Neisser presented in the context of character recognition that the formulation of the recognition as a multi-stage process is beneficial [99]. Today, most recognition systems are based on a sequential model that has at least two main parts: feature extraction and classification. The goal of feature extraction is to extract useful information from an input image, which is then utilised to perform the classification. That is, the digital image is processed in order to find some structure, e.g. edges, which is then used as the criterion on which the categorisation is based. The feature extraction and classification are quite often also composed of several sub-stages.

The human visual system is nowadays thought to be a similar multi-stage model with several levels of preattentive feature analysis at the early stages of processing [107]. As in many other tasks, the skills of a computer and those of a human are largely complementary in object recognition. Humans can accurately determine the class of an object among hundreds of others almost regardless of the viewpoint, illumination, or partial occlusion. In contrast, in controlled environments computer vision can tirelessly find flaws in complex manufactured products that would not appear to a human observer.

The main difference between the capabilities seems to result from the fact that the machine is almost unerring in a controlled environment where it is known what visual features to seek. On the other hand, in a natural environment the viewing conditions vary, which is a remarkable problem for computer vision. For example, the object may be viewed from a different pose or the lighting may change. Thus, the problem of invariant object recognition is to recognise objects despite these variances.

Not much has changed since 1978 when Granlund [35] referred to feature extraction methods as "a bag of tricks" since a method is often chosen based on experience seemingly without explicit reasons. Nevertheless, the feature extraction method selected has a fundamental effect on the characteristics of a recognition system. Feature extraction can play different parts in invariant recognition, depending if the feature itself is invariant, such as moment invariants [39] or the Fourier-Mellin coefficients [18], or if the invariance is a characteristic of the system, as in labelled graph matching [65]. Regardless of the role of feature extraction, the following requirements and objectives can be stated: The system should be *general* enough to recognise a large variety of objects, *invariant* to natural variations, *stable* against distortions, and *computationally efficient.*

The aim of this thesis is to study some feature extraction methods that can be applied to invariant object recognition. While in general, invariant object recognition encompasses variations caused by the three-dimensional structure of the world, this thesis is limited to the study of two-dimensional variations. In addition, the methods presented assume that the object to be recognised has been segmented from the background, and that only a single object is considered at a time. Two well known feature extraction methods, the Hough transform (HT) and Gabor filtering, are examined and improved to be suitable for invariant recognition. Primarily the shape of the objects is inspected as it is supposed that the shape includes enough information for successful recognition. For that reason, the scope of the thesis is limited to two-dimensional gray level images and the methods presented are primarily applicable to man-made, geometric objects, where the object shape can be efficiently modelled.

The thesis is divided into five chapters. In Chapter 2, the role of invariance in object recognition is examined. Next, two feature extraction methods are presented and applied to object recognition, namely, the Hough transform in Chapter 3, and Gabor filtering in Chapter 4. Finally, the discussion and conclusions are presented in Chapter 5.

Summary of publications

In *Publication I* a new method for locating line segments was proposed. The method is an extension of the Randomized Hough transform (RHT) which exploits connectivity for the estimation of local structure of the image. The experiments indicated that the method was more tolerant to missing data when compared to its predecessor. The author of this thesis participated in the development of the method, performed the programming and the experiments, and wrote the major part of the publication.

In *Publication II* a system of inspecting two-dimensional planar objects was proposed. The inspection is based on a CAD model of the part. The system first estimates the position of the part followed by a local measurement of each feature. In the experiments, it was found that calibration was the most limiting factor for the accuracy of the system. The author developed the system, performed the experiments, and wrote the major part of the publication.

In *Publication III* two new methods for matching binary images were presented. The methods use the Hough transform to extract prominent line features which are then used to find matching images. The author participated in the method development, the experiments, and writing the publication.

In *Publication IV* the use of Gabor filters as direction sensitive edge detectors was studied. A translation, rotation, and scaling invariant distance measure was proposed for the Gabor features. Experiments were performed on images of digits to study the representation power of the features. The author participated in the development of the method, programming, experiments, and writing of the publication.

In *Publication V* the Gabor feature was applied to matching of binary line-drawing images. Rotation and scale invariance was demonstrated together with very good noise tolerance against salt-and-pepper noise. The author of the thesis participated in the development, programming, experiments, and writing of the publication.

In *Publication VI* the performance of the Hough transform was inspected in affordable parallel environments. A parallel HT algorithm was developed suitable both for shared memory multiprocessors and distributed memory workstation networks. Also, parallel implementations of the Randomized Hough transform were proposed. The approaches improved the performance notably compared to a sequential algorithm. However, the approaches did not seem to be suitable for large scale parallelisation. The author of the thesis developed the methods, participated in the programming and the experiments, and wrote the major part of the publication.

Chapter II

Invariance in Object Recognition

The fundamental problem of object recognition is to analyse an image and provide a high level symbolic representation of its contents. Each separate object in the image should be localised, i.e., its location should be resolved, and classified, assigned to a specific category. In addition to the three-dimensional structure of an object, there are several other factors that affect the image of the object, for example, lighting, perspective projection, and sensor noise. Invariant object recognition aims to solve the problem despite these variances. In this work, the focus will be on recognition of objects based on their shape because the shape is fairly independent of phenomena such as illumination, contrast, and colour.

In this chapter, invariance in object recognition is discussed. The first section deals with the role of invariance in general, concentrating on the definition, taxonomies, and advantages. Next, variations occurring in natural images are examined to define more precisely the types of invariance necessary for object recognition. In order to discuss invariance, the idea of a class is also considered. After that, two groups of methods for invariant object recognition are discussed. First, global invariants and invariant transforms are presented, forming the oldest approach to invariance in computer vision. Second, the use of local features and matching is considered. Finally, the role of invariance in human visual perception is briefly discussed.

2.1 The role of invariance

The term invariant refers to a representation of a signal that is constant under some transformation. For example, the measurement T of a signal x is invariant under transformation g when [68]

$$T\{x\} = T\{g[x]\}. \tag{2.1}$$

Using group theory this definition can be extended [29, 129]:

Assume that there is a group $\mathcal{G}$ which acts upon the set X of possible signals. Let T

denote the action of measurement. Then, if

$$T\{g[x]\} = T\{x\}h(g) \ \ \forall g \in \mathcal{G}, x \in X, \tag{2.2}$$

T is invariant under the action of the group. The group $\mathcal{G}$ can be, for example, the group of translations. $h(g)$ is a function of g alone. If $h(g) = 1$, T is a scalar invariant [29], as in (2.1). For example, the Fourier transform is invariant under the group of translations, while its amplitude is scalar invariant. Invariants can be categorised as strong and weak, according to whether the degree of the action g can be measured from the invariant representation. The representation is called strong if it contains a component that explicitly encodes the degree of transformation [16]. Thus, according to this definition, scalar invariants are weak invariants.

In computer vision and object recognition, invariants provide a method to measure images while keeping the measurement unaffected by known deformations or variations. For example, in image databases, geometric invariants can be used for indexing the database, allowing the retrieval of an image despite geometric transformations. It has been argued that object recognition is the search for invariants [127]. However, in this work, the emphasis is not on the construction of invariant features, but on methods that allow object recognition in the presence of variations. This distinction is important because invariant recognition is possible without explicit invariants.

There are several taxonomies for invariant recognition. According to Wood [129], there are two broad approaches: (1) Distinct feature extraction and classification; and (2) parameterised invariants. In the first approach, each feature is itself invariant and the classification can be performed based on the features without considering the symmetries in the features. With parameterised invariants, the parameters of the invariant are adopted to perform the classification. Invariants can also be classified as global and local. Global invariants use the knowledge of a shape as a whole, while local ones are based on local properties such as curvature. Most global invariants belong to the class of distinct feature extraction and classification. When considering invariant recognition of shape, the methods can also be classified as either region- or boundary-based [57]. Region-based methods use the whole region under consideration for determining the feature while boundary-based methods consider only the outer boundary. However, this taxonomy applies specifically to the global invariants of binary images and for that reason many methods cannot be classified using it.

2.2 Variation in natural images

In this thesis, it is assumed that all objects representing a class are rigid and visually similar to a very high degree. Thus, the intra-class geometric variations are very small compared to the inter-class variations. In practice this means that only rigid objects are considered. In addition, the classes are assumed to be disjoint, that is, each object can belong to a single class, and there is no class hierarchy similar to the human visual system. Based on these restrictions, the variations are now examined.

The process of imaging an object has a number of sources for variation in the result image. Wechsler [126] states that the variability is due to perspective, changing orientation and scale, sensor noise, occlusion, and illumination. Buhmann [15] gives the following list of variations: pattern translation, pattern distortion, perspective distortion, variation in background, partial occlusion, change in pattern size, change in orientation, and change in lighting. The variations can be categorised according to the cause. First, the three-dimensional position of the imaging equipment with respect to the object. Second, the change in environment, primarily the lighting, but also encompassing changes in background. Third, noise in the imager and in the environment. Fourth, the occlusion of the target object by other objects. Traditionally, the study of invariants has focused on the geometric variations, but it is important to note that while geometric variations are important when describing the perspective projection of image formation, they are only one class of image variation.

Geometric variations can be further categorised based on the allowed transforms. The simplest variation is the translation of the object in the image. Translation together with rotation forms the group of Euclidean transforms. Including isotropic scaling gives the group of similarity transforms. The group of similarity transforms can be used to describe the projection of a planar object which is orthogonal to the optical axis of the imager. Similarity transforms are extended by allowing non-isotropic scaling and skewing, forming the group of affine transforms. The most general group of geometric transforms is the group of projective transforms, which represents perspective projections.

Changes in environment and the effect of noise are sometimes neglected in the study of invariant recognition, although they definitely have an effect on applications. The issue of nonuniform illumination is usually considered separate from other invariances, but recently it has been proposed that geometric and illumination invariants should be treated together [1]. In the methods presented in this thesis, the objects are, however, assumed to be segmented from the background, as stated in the introduction.

In complex environments, invariant recognition in 3-D is also highly dependent on an efficient world model. Without a model that can explain the overlap of objects, the problem of occlusion is very likely to be impossible to solve. Also, the world model can give valuable cues for segmentation, especially in multi-object cases.

The robustness of an object recognition system is an important consideration. To be efficient in applications, invariant recognition has to be stable, general, and computationally efficient. Stability means that small changes in environment should not have a remarkable effect on the result because in natural scenes several types of variations are likely to occur. In addition to invariant recognition, the system should be able to characterise the variability because the variability itself carries important information [126].

2.3 Global features and invariant transforms

The two most famous invariant transforms in image processing are based on moments and the Fourier-Mellin transform. Moment invariants, presented by Hu in 1962 [39], are the first general invariants. The geometric moments of order $(p+q)$ of $f(x,y)$ can be defined as

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)\,dx\,dy, \quad p,q = 0,1,2,\cdots . \tag{2.3}$$

By shifting the image centroid to origin, central moments μ_{pq} can be defined as

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx\, dy \tag{2.4}$$

where the origin is $(\bar{x}, \bar{y}) = (m_{10}/m_{00}, m_{01}/m_{00})$. Central moments can be normalised to obtain scale invariance as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \quad \gamma = (p + q + 2)/2. \tag{2.5}$$

Hu presented that certain combinations of the normalised central moments are invariant under the class of similarity transforms. The Hu's moment invariants up to third order are

$$\begin{aligned}
\phi_1 =& \eta_{20} + \eta_{02} \\
\phi_2 =& (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \\
\phi_3 =& (\eta_{30} - 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2 \\
\phi_4 =& (\eta_{30} + \eta 12)^2 + (\eta_{03} + \eta 21)^2 \\
\phi_5 =& (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right] + \\
& (\eta_{03} - 3\eta_{21})(\eta_{03} + \eta_{21}) \left[(\eta_{03} + \eta_{21})^2 - 3(\eta_{12} + \eta_{30})^2 \right] \\
\phi_6 =& (\eta_{20} - \eta_{02}) \left[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2 \right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21}).
\end{aligned} \tag{2.6}$$

Geometric moment invariants do not decode the degree of transformation and are therefore scalar invariants. Affine invariant combinations of geometric moments were introduced later in [96]. A treatise on moment invariants can be found in [108] where several types of moments are analysed, in addition to the geometric moments by Hu. Zernike and pseudo-Zernike moments were found to have the best overall performance [108] in terms of noise sensitivity, information redundancy, and image representation capability. Surveys of moment based techniques for object recognition can be found in [95] and [5].

The Fourier-Mellin transform was presented in the context of invariant recognition by Casasent and Psaltis in [18]. The Fourier transform can be defined as

$$\mathfrak{F}\{f(x, y)\} = F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi(ux+vy)} dx\, dy \tag{2.7}$$

where j is the imaginary unit. It is translation invariant since

$$\begin{aligned}
\mathfrak{F}\{f(x + x_0, y + y_0)\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x + x_0, y + y_0) e^{-j2\pi(ux+vy)} dx\, dy \\
&= e^{j2\pi(ux_0+vy_0)} \mathfrak{F}\{f(x, y)\}.
\end{aligned} \tag{2.8}$$

Thus, the amplitude is invariant to translations while the phase shift codes the degree of translation. Mapping the amplitude into log-polar coordinates transforms rotation and scaling into translations [126]:

$$\theta = \tan^{-1} \frac{v}{u} \qquad\qquad t = \log\left(\sqrt{u^2 + v^2}\right). \tag{2.9}$$

Applying the Fourier transform again and rejecting the phase gives a signature that is invariant to translation, rotation, and isotropic scaling, that is the class of similarity transforms. This process is known as the Fourier-Mellin transform. Using a log-log mapping of the frequency space instead of the log-polar one,

$$s = \log u \qquad\qquad t = \log v \tag{2.10}$$

makes the result invariant to anisotropic scaling, but the rotation invariance is lost [123].

Moment invariants and Fourier-Mellin signatures are both region based methods of description. Several global contour based invariants have been proposed, the most notable being perhaps the Fourier descriptors [133]. The list of points representing a closed object boundary can be presented as a complex function of one variable as

$$u(n) = x(n) + jy(n), \;\; n = 0, 1, \ldots, N-1. \tag{2.11}$$

The discrete Fourier transform coefficients $a(k)$,

$$a(k) = \sum_{n=0}^{N-1} u(n) e^{-j2\pi \frac{kn}{N}}, \;\; k = 0, 1, \ldots, N-1, \tag{2.12}$$

are called the Fourier descriptors. They allow recognition invariant to similarity transforms, since rotation causes only phase shift, translation changes only the DC-component $a(0)$, and scaling scales the coefficients correspondingly. Fourier descriptors have been extended to full affine invariant recognition in [2]. Other boundary decompositions have also been used. For example, a wavelet decomposition of the boundary has been used to realise similarity transform invariant recognition [111] and full affine invariant recognition [57]. The main problem of the global boundary based methods is that they are usually restricted to objects which have a single, clear outer boundary.

Neural networks can be also used to obtain invariance in recognition. There seem to be two general types of neural network approach: (i) using invariant features as input to a neural network, and (ii) obtaining the invariance in the network by learning [90]. The former approach does not differ much from conventional invariant feature extraction, but the latter one is interesting because the invariance is inherent in the network structure and weights. The Neocognitron model by Fukushima is one of the earliest successful attempts to perform distortion tolerant recognition [30]. The invariances depend solely on the training inputs of the networks, and thus the use of learned invariances in general purpose tasks is not very strongly justified. However, the tangent propagation algorithm, introduced by Simard [103], trains a neural network to be invariant to a known transformation of input by minimising the gradient of the network in the direction of the transformation. Thus, the network can be made invariant to analytic transformations of the input.

Some problems are related particularly to global invariant features. First, global invariants do not tolerate occlusion, as they assume that the whole object or shape can be observed. Second, global invariants require segmentation if there are regions of the image that do not belong to the object to be recognised. This limitation, of course, can be overcome by a segmentation, equivalent to object detection [47]. However, it is not clear if object detection can be performed separately from recognition. Finally, the recognition of especially complex objects is often sensitive to noise and distortions as a small change in a part of an image can sometimes cause large changes in the invariant.

2.4 Local features and matching

The methods for invariant recognition that do not use global invariants will now be considered. The main body of the research is related to methods that try to achieve invariance through matching local features to a known model. The methods usually have two distinct parts, feature extraction and matching. These phases will be considered separately as they do not usually depend on each other.

The first approach that will be considered, template matching, does not have distinct feature extraction since the gray-level values are used directly in the matching. It is worth mentioning as it is one of the oldest techniques for object recognition [97]. In template matching, the cross-correlation between an image and a template is computed. The location of the template is then found as the location of the maximum correlation, and the goodness of the match is related to the normalised value of the correlation. Template matching does not itself produce invariance to geometric transforms other than translation. However, rotation invariant recognition has been realised by convolving an image with a radially symmetric filter to select possible candidates [20]. Also, if the position of the object is known, rotation and scale invariant recognition is possible in log-polar coordinates. In addition, template matching is sometimes used as a verification step after a hypothesis of the object pose has been established. While template matching can be made more efficient using Fourier and principal component transforms [116] or wavelet decomposition [114], it is nevertheless usually considered computationally too costly for invariant recognition because several templates are often needed.

Principal component analysis (PCA) has been applied in object recognition, for example in the eigenfaces approach [105], but PCA is fundamentally a global method, and thus does not offer robustness against changes in localised regions. Penev and Atick [89] have proposed Local Feature Analysis (LFA), which modifies principal component analysis by deriving local topographic receptive fields optimally matched to the second-order statistics of input data.

Two basic approaches can be used for the matching of local features between an input and a model. First, pairwise matching of each feature on the input to a feature in the model, and second, optimisation of transform parameters that brings the features in the best match. When the matching of local features is formulated as a graph matching problem, each feature in the input and the model are encoded as vertices of a graph [11] and the matching corresponds to finding the subgraph isomorphism where an error function has its minimum. Finding the subgraph isomorphism is an NP-complete problem, thus there are no efficient techniques for the general case. However, in image analysis the graphs are usually topological, that is, they represent the two-dimensional topology of the image, which reduces the complexity of the matching [15]. The search space can also be pruned using various heuristics, for example, relative geometric constraints such as distance and angle between graph vertices [11], matching of the attributes at graph vertices between the model and the input [42], matching of vertex attributes with a hash lookup [66, 67], or emphasising the topological coherence [101].

A limited matching procedure can be used to find likely candidates for the transform parameters as in [131] where only the graph vertices on a convex hull are used in the preliminary matching, as the points on the convex hull are invariant in affine transforms. Then, the search through the likely candidates is less exhausting than full search.

Another formulation for geometric invariants is an optimisation of the transform parameters. For the optimisation of parameters, least-squares [67] and evidence gathering [115] methods have been used. In *Publication II*, translation and rotation of an object are solved through the minimisation of the least-squares fitting error between the model and the input. In evidence gathering, local matching evidences are accumulated to make a global pose hypothesis. Evidence gathering methods for object recognition include the Generalised Hough transform (GHT) [4], pose clustering [106], and evidence accumulation [49, 77, 88]. Evidence gathering will be considered further in Chapter 3, where the Hough transform is presented.

Feature extraction is also necessary for an object recognition system that uses matching other than template matching. One of the simplest features to use are edge points. They are often used with Hough transform methods. Interest points, such as curvature maxima [48] and corners can also be used [131]. Lee et al. propose the use of line segments [72]. In *Publication II*, line segments are also used. More complex features like Gabor filter responses [15, 88] and local Fourier-Mellin descriptors [34] have also been used. Lowe [77] proposes to use maxima and minima of a difference-of-Gaussian function in a multi-scale grid to determine interest points, and then use the local gradients of the neighbourhood as the feature. For the local features, two distinct properties are desired: distinctiveness against false positives, and robustness against variations [88]. It is clear that by increasing distinctiveness, robustness is lost, and vice versa. More complex features often provide more distinctiveness but less robustness and large computational costs. The most important requirement for a feature is that it should remain the same regardless of the variations. In addition, it should provide invariant local information to allow the pruning of the search space.

The problem of partial occlusion can be solved to some degree by a system that achieves invariance through local features and matching. However, a reliable world model is still necessary for 3-D inference and missing data handling. The problems that arise in the matching of local features are often related to computational complexity. For example, graph matching and Hough transform are both computationally quite expensive because the search space is large. A second problem is the selection of the local features. An important question is, which kind of features are able to represent a wide variety of natural scenes.

In *Publication II*, a complete recognition system is presented that employs rotation invariant template matching for determination of an initial estimate of a two-dimensional object position. Then, local feature extraction is performed based on the initial estimate to extract straight edges and circular shapes based on the CAD model of the object. Finally, a least-squares fitting is performed for the local features and, based on the pose estimate, the difference between the object and the model is determined for each line segment and circular shape.

2.5 Object recognition in human visual system

Because invariant recognition is evidently an unsolved problem, it is natural to take a look at one known solution, the human visual system. It should be noted that the structure of the human visual system is not completely known and the following theories do not

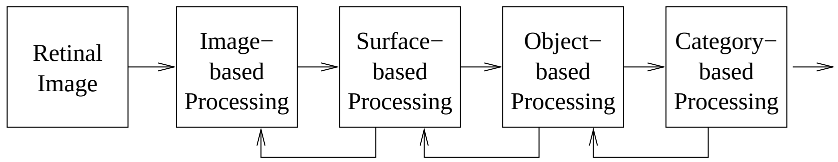


Figure 2.1: Four stages of visual processing [86].

capture the full power of shape recognition in human vision. The following theories are, however, based on observations in cognitive science and psychology.

At least two coincident structures are supposed to appear in the human visual system, the sequential visual perception process and parallel pathways. Visual processing can be divided into four stages: image-based, surface-based, object-based, and category-based [86]. As illustrated in Fig. 2.1, the stages are sequential but feedback may exist to a preceding stage. The image-based processing is supposed to extract two-dimensional primitives of the retinal image. The surface-based processing attempts to recover the properties of surfaces such as local surface orientation and depth. This stage is not fully three-dimensional but employs local cues about the three-dimensionality, and for that reason it is often called 2.5-dimensional. The object-based processing is then concerned with the fact that an object is a three-dimensional entity which can have a totally different two-dimensional shape from different viewpoints. Finally, the category-based stage is mainly concerned with recovering the functional properties of an object by placing it into known categories.

In computer vision, the term classification can be related to either object- or category-based stages. The categories in computer vision are usually simple and disjoint, while in the human visual system there is essentially a hierarchy of classes. Thus, a single object can be categorised at several levels. It has been suggested that there are separate processes in visual system for different levels. It is also important to note that the focus of research in the categorisation behaviour of the human visual system has been on the entry level categories, that is, those categories that can be determined most quickly, while the computer vision has been traditionally related to subordinate level categorisation, that is, the categories which are more specific. For example, a typical computer vision application could be the classification of different types of bottles in a bottle recycling machine. However, for a person, the entry level class "bottle" would be the most natural one.

In addition to the sequential process framework, there is physiological evidence that there are parallel pathways in early visual processing. The proposal suggests that there are separate pathways for such features as form, colour, motion, and depth [75]. It has been shown that significant cross-talk exists among the pathways [119], but nevertheless, parallel processing seems to exist in the lower levels of visual processing.

The issue of object representation in the human visual system is under research but two main approaches can be seen, structural descriptions and view-based representations [107, 86]. Structural descriptions describe objects as compositions of three-dimensional volumetric primitives. The view-based representations describe the two-dimensional appearance of the object from a single view, while three-dimensional recognition is incorporated by using multiple views.

This thesis concentrates on recognition based on the two-dimensional shape and for that reason the discussion will be now limited to shape recognition. Palmer [86] approaches the question of shape recognition from two distinct points of view, shape equivalence and shape representation. Shape equivalence is the phenomenon that two shapes are perceived as having the same shape. This theory is mainly concerned with perceived shape and does not specify the processes that are used in determining the equivalence. An important finding is that while the perceived shape often does not depend on the orientation, in some cases it does [79]. Thus, the human visual system is not totally invariant to rotations even in the categorical level.

Shape representation considers how the shape of objects might be represented within the visual system and how such representations might be compared. Palmer presents the following as the four major theories of shape representation: a) templates, b) Fourier spectra, c) feature lists, and d) structural descriptions. It should be noted that none of these has been accepted generally as the definitive theory of shape representation in the human visual system, but they provide some insight to the issue of invariance.

Templates correspond closely to the template matching approach in computer vision. In biological vision, templates are seen as useful in providing simple low-level information about object parts but lacking the versatility required for the representation of complete shapes. This is in accord with the methods of local features and matching, where the local features can be extracted using a template matching type method.

The Fourier spectrum has been suggested as the shape representation [33, 50]. Early studies suggested the power spectrum as the representation. This corresponds closely to the use of Fourier transform in computer vision to obtain translation invariance. However, later studies have revealed that the phase spectrum dominates the perception [91]. This result may indicate the importance of phase information in Fourier domain feature extraction in computer vision. However, the existence of localised frequency sensitive cells has been shown at the lower levels of processing in mammalian visual systems [40, 41].

The feature lists approach states that an object's perceived shape can be described by a set of features, such as the number of lines and their relative lengths. The features need to be invariant to variations, relating this approach to the use of global invariant features. However, the basic approach of a set of features has a flaw because the interdependencies between the features cannot be taken into account. To solve this problem, feature maps have been proposed that include additional constraints to the mutual positions of features [76, 117]. However, pure feature-based approaches are often considered inadequate as describing the shape of an object often requires not only the components and global properties but also the spatial relations between the parts.

The need for relative information gives the motivation for structural descriptions. They can be compared to graphs in which vertices represent objects parts and labelled arcs represent the relations [85]. The representation is strong as the graph structure allows the relations to be separate from the parts. The structural descriptions are closely related to the graph based approaches in computer vision. As also in computer vision, the weakness of this approach is the complexity of representation, with also the problem that a set of sufficiently powerful primitives and relations needs to be identified. For example, it is very hard to find a set of simple primitives that could describe a human.

Hough Transform

In 1962, Paul Hough presented a method for detecting straight line tracks of charged particles in a bubble chamber [38]. In the 1970's the method gained popularity in the computer vision community and became known as the Hough transform (HT) [1] [27]. In 1981, Deans [25] showed that the mathematical roots of HT lie in the Radon transform, as the Hough and Radon transforms are essentially the same in certain circumstances [94]. There are numerous different variants of HT [44, 71, 52].

While Paul Hough presented the method as a means to detect straight lines, today the Hough transform is considered more generally as a tool for detecting parameterised shapes. HT processes binarised images. Usually, the binary image is produced using an edge detection method such as Prewitt [93] or Canny [17] edge detectors. The Hough transform is based on the principle that each edge point in the image gives positive evidence for curves through that point. This evidence is accumulated for all points. If there are many points giving evidence for a particular curve, then that curve is likely to exist in the image.

In this chapter, the Hough transform algorithm for finding line segments is first presented with its benefits and pitfalls. Then, the problem domain is extended to cover more complex shapes and an application to image matching. Finally, improvements to the computational complexity are overviewed with emphasis on probabilistic and parallel methods.

3.1 Hough transform for line detection

The Hough transform is essentially a transform from an image space to a parameter space. The parameters of curves that are present in the image appear as maxima in the parameter space. For example, a line can be represented in the slope-intercept form as

$$y = ax + b. \tag{3.1}$$

[1] The terms Hough transform and standard HT are used to refer to the same concept.

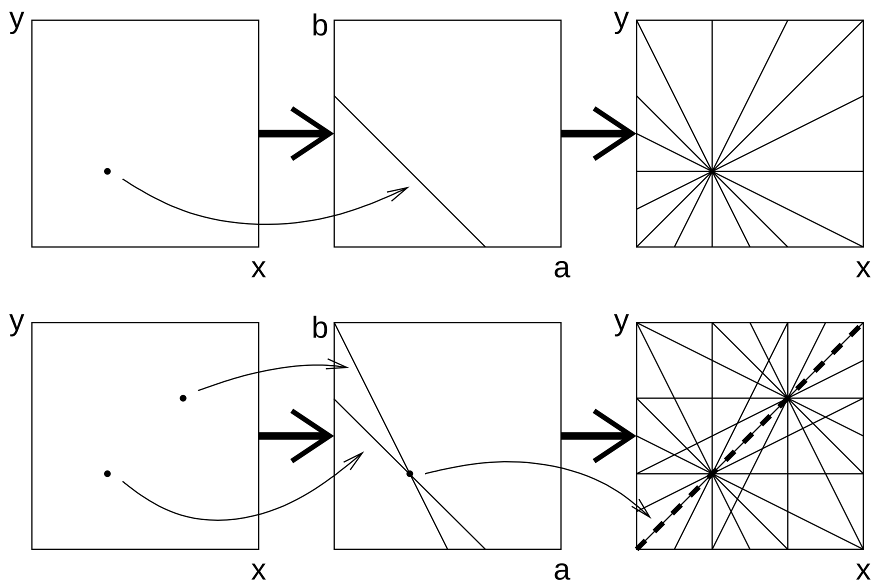


Figure 3.1: Projection of a point is a line in the parameter space. The line represents all the lines through the point in the image. The intersection of lines in the parameter space represents the line defined by the corresponding points.

In this equation, x and y denote the coordinates of a point, a is the slope of a line, and b is the intercept on the y-axis. By rearranging (3.1), it can be seen that each point (x, y) in the image defines a unique line in the parameter space (a, b) as

$$b = -xa + y. \tag{3.2}$$

When several points lie on the same line in the image, the corresponding parameter space lines intersect. As illustrated in Fig. 3.1, the line can be detected as the intersection in the parameter space called the accumulator. Each coordinate pair (a, b) denotes a specific line in the original image. To find the intersection in the parameter space, a procedure called accumulation is normally used. The parameter space is first discretised, that is, it is divided into a finite number of non-overlapping cells covering the whole parameter space. Then, accumulation is performed by transforming each image point into a parametric curve and incrementing counters associated with those accumulator cells for which the transformed locus intersects the cell. In this way, each point in the image gives partial evidence that a certain feature, with certain parameters, is present. Finally, local maxima are searched for in the accumulator. The locations of the local maxima specify the parameters of apparent lines. In addition, a verification may be performed where each local maxima is checked to assess whether it really represents the desired feature in the original image.

The accumulation can be specified more strictly as follows:

Algorithm 1 *Hough transform accumulation.*

Let D denote the set of points in the image. Let A denote the accumulator and S represent the indices used to index the accumulator. Let T be the function that projects a point in the image to a curve, that is a set of indices, in the parameter space.

1. *Initialise $A(s) = 0, \forall s \in S$.*
2. *For all d in D*
   - *2.1 For each s in $T(d)$*
     - *2.1.1 Increment $A(s) = A(s) + 1$.*

The slope-intercept parameterisation of (3.1) has a problem in that the parameter space is not bounded, since for a vertical line both parameters approach infinity. This problem can be overcome by using two accumulators, one for lines with slope $|a| \leq 1$ and the other for lines with $|a| > 1$. Another parameterisation, the normal parameterisation,

$$\rho = x \cos\theta + y \sin\theta \tag{3.3}$$

was introduced by Duda and Hart [27]. As illustrated in (3.2), θ represents the angle between the normal of the line and the x-axis and is bounded to interval $[0, \pi)$, while ρ is the distance from the line to the origin and its range depends on the image size. Other bounded parameterisations also exist for lines, for example the points of intersections with the edges of the image can be used [121]. However, the boundedness of the parameter space seems to be the most important restriction for the selection of the parametric representation.

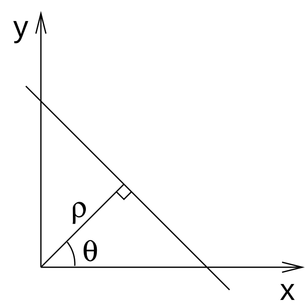


Figure 3.2: (ρ, θ)-parametrisation of a line.

Because only positive evidence is considered, HT is very robust against missing data, i.e. image points. This robustness is considered to be one of the main advantages of the Hough transform. The main drawbacks of the HT are related to computational complexity in terms of both time and space. The computation time is related to the number of edge points in the image times the larger dimension of the accumulator. Thus, the discretisation of the parameter space affects both the time and space complexity. In addition, the discretisation can present difficulties in the implementation of the algorithm that have to be taken into account [120].

3.2 Extending the problem domain

3.2.1 More complex shapes

Kimme et al. were some of the first researchers to extend the problem domain from the detection of lines to more complex shapes by presenting a HT based method for circle detection in 1975 [58]. In addition to using more complex analytic shapes, they discovered that the edge direction information could be used to decrease the computational complexity of the transform. In 1981, Deans suggested a framework in which the Hough transform may be generalised to detect any analytically defined curves. This is accomplished by first defining HT as an integral transform as

$$H(\rho,\theta) = \iint_{-\infty}^{\infty} I(x,y)\delta(\rho - x\cos\theta - y\sin\theta)\,dx\,dy. \tag{3.4}$$

Now, $I(x,y)$ is the image with edge points having a value 1 and all others value 0. $\delta(\cdot)$ is the Dirac delta function. The argument of the delta function evaluates to zero only on the line with parameters (ρ,θ). Deans proposed [25] that (3.4) can be generalised as

$$H(p,\xi) = \iint_{-\infty}^{\infty} I(x,y)\delta(p - C(x,y;\xi))\,dx\,dy \tag{3.5}$$

where $I(x,y)$ is a function defined on the (x,y)-plane representing the image. The argument of the delta function defines a family of curves parameterised by a scalar p and a vector $\xi = (\xi_1,\ldots,\xi_{n-1})$. With two-dimensional parameter space (as in lines), a point (x,y) is projected into a curve parameterised by scalar ξ. In more than 2 dimensions, the point is projected into an $n-1$-dimensional hypersurface in the n-dimensional parameter space. For each point (x,y), the accumulation is performed by calculating the values of p for each value of the vector ξ and incrementing the corresponding accumulator cells. Now, it is evident that the space complexity of the accumulator is $O(m^n)$ where m is the number of indices for each dimension of the accumulator. The time complexity is $O(m^{n-1})$ being also exponential. Finally, finding the maxima involves searching in the n-dimensional accumulator. Altogether, the computation becomes unrealistic for large n. As already presented, additional information such as edge direction can be used to constrain the computation. In terms of the methodology of Deans, this corresponds to using information besides the coordinates of a single pixel to lower the dimensionality of ξ. This can be formulated as

$$H(\mathbf{p},\xi) = \iint_{-\infty}^{\infty} I(x,y)\delta(\mathbf{p} - C(\mathbf{x};\xi))\,dx\,dy \tag{3.6}$$

where $\mathbf{p}$ is the vector of parameters that can be solved using $\mathbf{x} = (x,y,\ldots)$ and ξ, that is, an additional parameter can be solved for each additional element of $\mathbf{x}$. Thus, the dimensionality of ξ is reduced which reduces the computational complexity. In practice, the use of the edge direction is problematic because the estimation of the gradient is prone to errors due to noise.

The previous formulation of HT is valid for analytic curves. Nevertheless, already in 1975, Merlin and Farber presented a Hough transform type algorithm for detecting non-analytic curves in images [83]. However, their method is equivalent to convolution of the edge image and is impractical for real image data. In 1981, Ballard presented the Generalised Hough transform (GHT), which allows the detection of arbitrary shapes [4], giving also the theory for extending the method for rotation and scale invariance. In GHT, the edge direction is used to constrain the search. However, the parameter space is still four-dimensional if the rotation and scale invariances are allowed. Ballard also comments that pairs of edge points could also be used to constrain the search but the comment is left at that. GHT has been later improved by using local curvature and slope [78], higher level primitives [72], relative angles [110, 56], and edge point pairs [100]. A comparison of the variants can be found in [56].

3.2.2 Invariant image matching

The accumulator of the straight line Hough transform has also been used as a description of an image content. In *Publication III*, a translation, rotation, and scaling invariant image matching method is presented which uses HT as a feature extraction method. An invariant feature vector can be constructed as follows: Let the (ρ, θ) accumulator be presented as a matrix A, where each row corresponds to a particular value of ρ, and each column to a value of θ. First, the matrix is thresholded as

$$A'_{ij} = \begin{cases} A_{ij}, & \text{if } A_{ij} > t \\ 0, & \text{if } A_{ij} \leq t \end{cases} \quad \forall i = 1 \ldots m, j = 1 \ldots n \tag{3.7}$$

where t is the threshold value. Next, the thresholded matrix is shrunk to a one-dimensional θ vector by summing the matrix across columns as

$$F_j^0 = \sum_{i=1}^{m} A'_{ij} \;\; \forall j = 1 \ldots n. \tag{3.8}$$

This vector is then normalised with respect to the mean as

$$F_j = F_j^0 \frac{1}{n} \sum_{j=1}^{n} F_j^0 \;\; \forall j = 1 \ldots n. \tag{3.9}$$

Intuitively, the vector can be thought of as an orientation histogram of lines exceeding t in length. The vector is translation invariant because only angular information is used to build it. The normalisation makes the vector invariant also to scaling. Rotational invariance can be obtained by using a rotation invariant distance to measure the similarity between two feature vectors. If R and S are the feature vectors, their rotation invariant distance can be calculated as

$$d(R, S) = \min_{k=1 \ldots n} \sum_{j=1}^{n} \left(R_j - S_j^{(k)} \right)^2 \tag{3.10}$$

where $S^{(k)}$ is a cyclic rotation of the vector s by k steps to the right. The problem of this approach is that it cannot be used for large and complex images because the orientations of lines do not carry a sufficient amount of information for the matching. The method can also be modified to include spatial information by incorporating the ρ-dimension of the accumulator.

3.3 Computational improvements

As stated in the previous section, one of the main problems of HT is its computational complexity. Several lines of research have emerged to address this problem [71, 52]. The improvements may be based on deterministic versions of the original algorithm, probabilistic methods, or using parallel hardware. Deterministic methods aim to decrease the computation time, for example, by recursive partitioning of the accumulator as in the Fast Hough transform [74] and the Adaptive Hough transform [43]. Other deterministic methods include using constraints on parameter calculation, as presented in the previous section, and multi-stage methods. However, the most active line of algorithmic research seems to be the probabilistic methods, which use random sampling for selecting a subset of the input data.

3.3.1 Probabilistic methods

Probabilistic Hough transforms are a family of methods that apply random selection in order to reduce the computational complexity. Already Ballard [4] commented that several points can be used to restrict the accumulation. The Randomized Hough transform [130] (RHT) is a method that is based on the selection of n points at random to map them into a single point in the accumulator (see Fig. 3.3). However, the use of all possible point sets would lead to a combinatorial explosion. For this reason, the random sampling and accumulation is continued only until an accumulator cell exceeds a predefined threshold value. Then, that cell presents a candidate feature that is verified in the image. If a feature is found, the points corresponding to it can be removed and the accumulation can be continued. This method corresponds to a Monte-Carlo approximation of the integral in (3.6). Thus, when a larger threshold is selected, it is more probable that the maximum found corresponds to a maximum of the standard Hough transform. The RHT algorithm can be presented as follows:

Algorithm 2 *Kernel of Randomized Hough transform.*
Let t denote the detection threshold.

1. *Create the set D of all edge points in a binary image.*
2. *Select two points d_i and d_j from D at random.*
3. *Calculate the line parameters (a, b) corresponding to d_i and d_j.*
4. *Accumulate $A(a, b) = A(a, b) + 1$.*
5. *If a threshold is not reached, $A(a, b) < t$, go to Step 2. Otherwise, the parameters a and b describe the parameters of the detected curve.*

In the Probabilistic Hough transform by Kiryati et al. [59], a subset of points is sampled and a standard Hough transform is then performed. In the framework of (3.6), this corresponds to computing the integral only over a random subset of the xy-plane. Because the Hough transform tolerates missing data, this method performs well as long as enough data is retained. The Dynamic Combinatorial Hough transform [8] uses a pairing of a

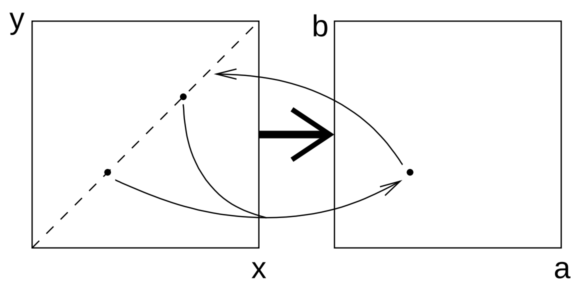


Figure 3.3: In the Randomized Hough transform for a line, a pair of image points is mapped to a single point in the accumulator.

random seed point with all other points to perform the accumulation in a 1-dimensional accumulator. Thus, if the seed point is part of a line, it is found at this stage. Because the line must be through the seed point, only a 1-dimensional accumulator is needed. The Dynamic Generalised Hough transform [70] is an extension, where a connected seed point is first determined and then a procedure similar to RHT is followed. In the Connective Hough transform [132], seed points are used with a row-wise accumulation of a 1-dimensional accumulator.

Probabilistic methods can also be used in conjunction with local data, such as the gradient direction. The Randomized Generalised Hough transform [31] combines the efficiency of RHT with the ability to detect general shapes of GHT. As in GHT, the gradient angles are locally estimated from a gray level image. In the Window Randomized Hough transform [52, 51], a window is placed over a randomly selected point and a line is fitted to the points inside the window. If the fitting error is small, the accumulator cell corresponding to the line parameters is accumulated. The Connective Randomized Hough transform [51] (CRHT) is a similar method that also restricts the fitted points to be connected to the centre point. In addition, the search for connected points is performed as a sectored scan, which means limiting the search directions to the three adjacent compass directions each time to reject paths with loops and sharp bends.

In *Publication I*, an extension to CRHT is presented that loosens the requirement of a connected line by allowing gaps in the line. The algorithm for the Extended Connective Randomized Hough transform (ECRHT) can be presented as:

Algorithm 3 *Kernel of Extended Connective Randomized Hough transform.*

1. *Create the set D of all edge points in a binary edge picture.*
2. *Select a point d_i randomly from the set D.*
3. *Create a $w \times w$ window centred at d_i, and locate points connected to d_i using the maximum gap criterion and the sectored scan.*
4. *If enough connected points are found, fit a curve to those points using the least square sum method and calculate the line parameters (a, b); Otherwise go to Step 2.*

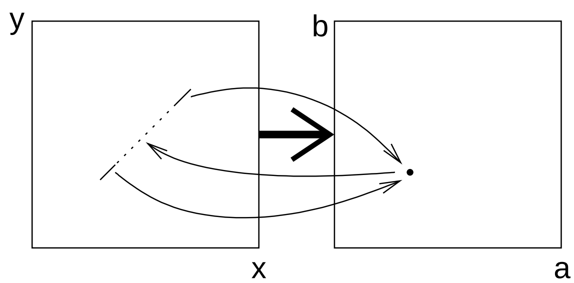


Figure 3.4: In the Extended Connective Randomized Hough transform, local fitting around a single image point is used to present evidence about a particular parameter combination.

5. *If the fitting error is within a tolerance, accumulate the cell $A(a, b)$ in the accumulator space; Otherwise go to Step 2.*
6. *If the $A(a, b)$ is equal to the threshold t, the parameters a and b describe the parameters of the detected curve; Otherwise continue to Step 2.*

The parameters a and b are calculated from the least-squares approximation

$$a = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}, \tag{3.11}$$

$$b = \frac{\sum y_i - a\sum x_i}{n} \tag{3.12}$$

where x_i and y_i denote the coordinates of pixel i and n denotes the number of pixels. Using the fitting error e,

$$e = \frac{a^2\sum x_i^2 + \sum y_i^2 - 2a\sum x_i y_i + 2ab\sum x_i - 2b\sum y + nb^2}{n(a^2+1)}, \tag{3.13}$$

only reliable estimates of the local gradient are accumulated. The idea is that accumulating only good local estimates decreases the total number of accumulations compared for example to the randomised Hough transform. This improves the performance especially with images with many lines. The local line segments do not need to be very long as they do not yet qualify as lines but just present evidence about a line (see Fig. 3.4). Compared to CRHT, the main advantage of ECRHT is its tolerance to missing data. Compared to RHT, the advantage is the decreased amount of computation and storage because local information is used to verify accumulations. The main disadvantage is that if a large amount of data is missing, the algorithm will fail because sufficient local information will not be available.

3.3.2 Parallel processing

The parallel Hough transform algorithms have been studied widely for both dedicated hardware architectures and general purpose parallel environments. The implementations

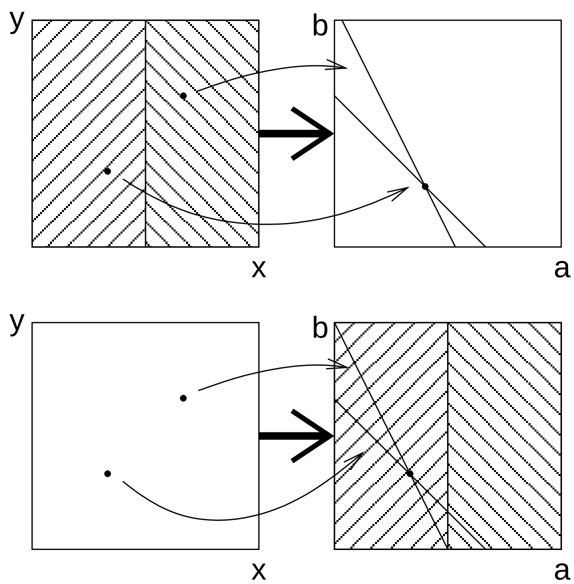


Figure 3.5: Division of the data and computation in parallel Hough transforms: Either source image or the accumulator can be divided between processing elements.

can be categorised into two classes depending on the division of the data and computation: either (a) the source image, or (b) the accumulator is split between the processors. This is illustrated in Fig. 3.5. The differences in these approaches have been evaluated by Thazhuthaveetil et al. [109]. An approach to split both the image and the accumulator has also been proposed [9].

Special hardware architectures include mesh-connected SIMD processors [98], discrete logic [23], transputers in a hypercube [22] reconfigurable processor array [55], reconfigurable multi-ring network [10], and VLSI content addressable memory [84, 82]. These specialised hardware solutions can offer the highest performance at the time they are built but the rapid advance of computer technology makes them grow old almost overnight. For this reason, more general parallel architectures are of great interest.

General purpose hardware architectures for the Hough transform include shared memory multiprocessors [21], distributed memory multiprocessors like Cray T3D [22], IBM SP2 [60], and Fujitsu AP1000 [118], and general purpose digital signal processors [9, 23, 63]. An implementation of the Hough transform in a distributed workstation network is reported in *Publication VI*, where the slow communication environment is taken into account by minimising the communication by using division of the accumulator instead of the image. In addition to decreasing the computation time, this approach decreases the storage requirement as the accumulator is divided between the processing elements. In addition to the workstation environment, the implementation is examined in a multiprocessor PC, demonstrating that inexpensive general purpose hardware can be used to speed up the computation remarkably.

Improved performance has been obtained by using modified Hough Transform techniques. The Multiresolution Hough Transform has been successfully implemented using a pyramid architecture [3]. Also the Fast Hough Transform [74] has been improved for use in a multiprocessor architecture [36]. Parallel implementation on a distributed memory system reported in [9] uses digital signal processors and overlapping shared memory.

In *Publication VI*, algorithms for parallel RHT are presented for a shared-memory multiprocessor. In RHT frequent communication between processing elements (PE) is necessary. Therefore, the algorithm uses shared memory which allows relatively easy and fast interaction between PEs compared to message passing parallel systems. The data structure for the list of points is an important determination, since it should allow parallel creation and a parallel update mechanism when lines are found. The image is divided into n parts and each PE creates the list of points corresponding to that part. The list update is performed by each PE updating the corresponding list, regardless of the line found. Parallel algorithm for RHT can be presented as:

Algorithm 4 *Parallel Randomized Hough Transform.*
Let n denote the number of processing elements (PE). The identifier of a PE is denoted by $id = \{0, 1, \ldots, n-1\}$. The same algorithm is executed in each PE.

1. *Make a list of the points D_{id} in the image region where $id \times x_{max}/n \leq x < (id + 1) \times x_{max}/n$. x_{max} is the maximum value of the x-coordinate.*
2. *Synchronise.*
3. *If $id = 0$ then calculate the total number of points in the lists $D_i, 0 \leq i < n$.*
4. *Synchronise.*
5. *Select two points d_i and d_j at random.*
6. *Accumulate with the line parameters corresponding to the points.*
7. *If threshold t is reached, set the flag F_{found}.*
8. *If the flag F_{found} is not set, return to Step 5.*
9. *Synchronise.*
10. *Update the point list D_{id} by removing points corresponding to the line found.*
11. *If all lines are not found, return to Step 2.*

There are two main parts in the algorithm, accumulation (Steps 5 to 8) and point list update (Step 9). Synchronisation points are needed to separate these. At each synchronisation point, no PE can continue before all designated PEs have reached the synchronisation point.

Because the overhead of parallelisation varies depending on the variable number of points in the image and the number of processes, a dynamic control of the number of processes is necessary for good performance. The overhead of the parallelisation is measured for each line sought and the number of processes is controlled to minimise the execution time. The dynamic parallel algorithm for RHT can be presented as follows:

Algorithm 5 *Dynamic Parallel Randomized Hough transform.*
Let n denote the total number of processing elements (PE) and p denote the number of active PEs. Initially $p = n$.

1. *Make a list of points D_{id} in the image region where $id \times x_{max}/n \leq x < (id + 1) \times x_{max}/n$. x_{max} is the maximum value of the x-coordinate.*
2. *Synchronise.*
3. *If $id = 0$ calculate the total number of points in lists $D_i, 0 \leq i < n$.*
4. *Synchronise.*
5. *Select two points d_i and d_j at random.*
6. *Accumulate with the line parameters corresponding to the points.*
7. *If threshold t is reached, set the flag F_{found}.*
8. *If the flag F_{found} is not set, return to Step 5.*
9. *Update point lists, each process updates n/p lists.*
10. *If synchronisation time is greater than the effective accumulation time, set flag F_{reduce}.*
11. *If F_{reduce} is set, terminate processes $p/2 + 1 \ldots p$ and set $p = p/2$.*
12. *Synchronise. If flag F_{reduce} is set, change the number of processes arriving to the next barrier.*
13. *If all lines are not found, return to Step 3.*

The parallel HT presents very good speedups compared to the sequential algorithm. Because little communication is needed, the parallel algorithm is scalable and can be applied efficiently also in widely parallel environments. However, the efficiency of RHT is remarkable compared to HT, exceeding also that of the parallel HT in multiprocessor workstations. However, the Parallel RHT requires a substantial amount of communication. Therefore, it cannot be applied in a standard networked workstation environment. The parallelisation overhead is thus larger with RHT than with HT, and the scalability of RHT is not as good. By incorporating the dynamic allocation of work, the communication overhead can be constrained and the Parallel RHT outperforms the sequential algorithms as well as the parallel HT.

Gabor Filtering

Gabor analysis [32] is a method for combining time and frequency domain analyses. In computer vision, Gabor filtering has been used for a number of tasks, including edge and line detection [81, 19], texture classification [12], and compression [92, 73].

In this chapter, the use of Gabor filters in the context of feature extraction and object recognition is examined. First, the formulation of the Gabor elementary functions is covered in the one-dimensional case. Then, the formulation is extended for two-dimensional images with the focus on filter normalisation, parameter selection, and the advantages of Gabor filtering. Next, the applications of Gabor filtering in object recognition are examined noting their invariance properties. Finally, a scale, rotation, and translation invariant Gabor feature is presented for recognising binary images.

4.1 Gabor elementary functions

4.1.1 One-dimensional time-frequency space

In 1946, Dennis Gabor presented a new method of analysing signals in which time and frequency analyses are combined [32]. The effective widths of a signal in time and frequency domains, Δt and Δf, are defined based on the variance of the signal in the corresponding domains, which leads to the definition of minimal uncertainty as the product of the effective widths,

$$\Delta t \, \Delta f \geq \frac{1}{2}. \tag{4.1}$$

Gabor then shows that the signal for which the product $\Delta t \Delta f$ is minimal and turns the inequality into an equality, is

$$\psi(t) = \underbrace{e^{-\alpha^2 (t-t_0)^2}}_{Gaussian} \underbrace{e^{j(2\pi f_0 t + \phi)}}_{sinusoid} \tag{4.2}$$

where j is the imaginary unit. The function represents a complex sinusoidal wave modulated by a Gaussian probability function. α, t_0, f_0, and ϕ are constants which denote

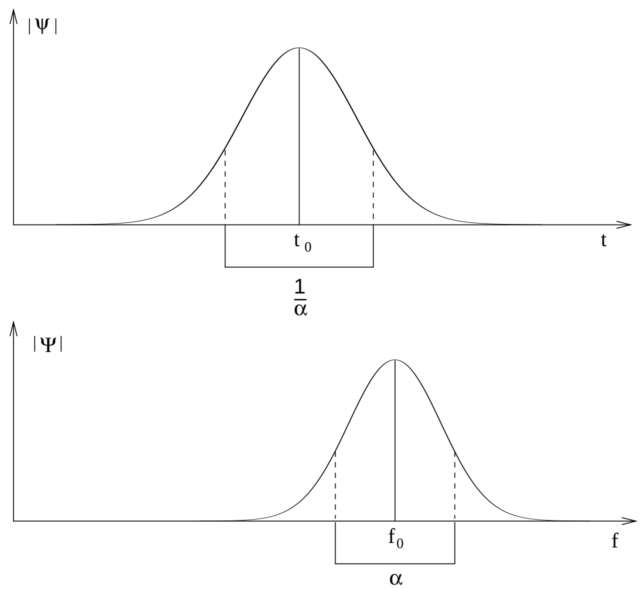


Figure 4.1: One-dimensional Gabor elementary function in time and frequency domains.

the sharpness of the Gaussian, centre of the Gaussian in time domain, and the frequency of the sinusoid and its phase shift. Applying Fourier transform to (4.2) gives

$$\Psi(f) = \frac{\sqrt{\pi}}{\alpha} \underbrace{e^{-\left(\frac{\pi}{\alpha}\right)^2 (f-f_0)^2}}_{Gaussian} \underbrace{e^{j(-2\pi t_0 (f-f_0)+\phi)}}_{sinusoid} \tag{4.3}$$

which also is a complex sinusoid modulated by a Gaussian. The effect of the parameters is illustrated in Fig. 4.1 which shows the Gaussian envelope of the function in both the time and frequency domains. The real and imaginary parts are separately shown in Fig. 4.2. The function in (4.2) became later known as the Gabor elementary function (GEF) and can be used to analyse signals by decomposing them into a number of GEFs. Gabor motivated this decomposition by the minimal joint uncertainty of the GEF in time-frequency space. It can also be noticed that the decomposition includes the Fourier analysis and the time-domain analysis as special cases, namely,

$$\begin{aligned} &\alpha \rightarrow 0 \Rightarrow \psi(t) \rightarrow \text{sinusoid} \Rightarrow \text{Fourier analysis} \\ &\alpha \rightarrow \infty \Rightarrow \psi(t) \rightarrow \text{Dirac-delta} \Rightarrow \text{time description.} \end{aligned} \tag{4.4}$$

The Gabor decomposition is a predecessor of multiresolution and wavelet analyses. The problem of the decomposition is that the GEFs are not orthogonal, which causes computational problems when they are used as a wavelet basis.

Instead of decomposition, signals can be analysed by convolving them with GEFs. The filter response to input $\xi(t)$, that is the convolution, can be defined as:

$$r_\xi(t) = \int_{-\infty}^{\infty} \psi(t-\tau)\xi(\tau)\, d\tau. \tag{4.5}$$

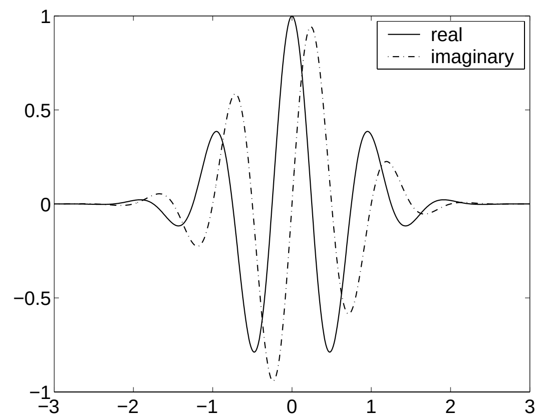


Figure 4.2: The real and imaginary parts of a Gabor elementary function.

In this thesis, the convolution is also used in the two-dimensional analysis of images because it is more straightforward than decomposition.

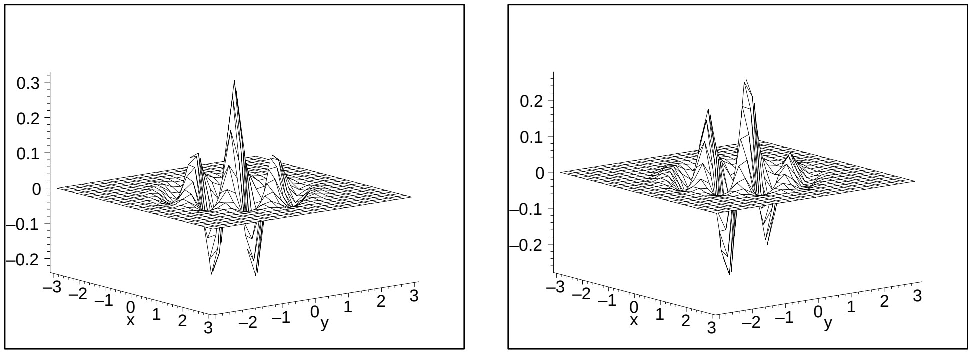
4.1.2 Two-dimensional spatial-frequency space

In 1978, Granlund presented a general picture processing operator which is a two-dimensional counterpart of the GEF [35], arguing that a single image processing operator which could detect and describe structure at different levels was of interest. However, the event that increased the popularity of Gabor analysis in the computer vision community was the result that the two-dimensional GEF resembles the receptive fields of simple cells in a mammalian visual system [80, 24]. When the GEFs are used as filters, the filter can be centred at origin and the phase and time shift parameters, ϕ and t_0, dropped. The filtering will be considered in continuous domain for simplicity.

The two-dimensional Gabor filter is a complex sinusoidal plane wave modulated by an elliptical Gaussian probability density function (see Fig. 4.3). Several forms of Gabor filters have been proposed for the two-dimensional case [35, 24]. Following Gabor's formulation for the one-dimensional GEF, a two-dimensional Gabor filter $\psi(x, y)$ can be defined as

$$\begin{aligned} \psi(x,y) =& \underbrace{e^{-(\alpha^2 x'^2 + \beta^2 y'^2)}}_{Gaussian} \underbrace{e^{j2\pi f_0 x'}}_{planewave}, \\ x' &= x\cos\theta + y\sin\theta, \\ y' &= -x\sin\theta + y\cos\theta, \end{aligned} \tag{4.6}$$

where f_0 is the frequency of the sinusoidal plane wave, θ is the anti-clockwise rotation of the Gaussian and the plane wave, α the sharpness of the Gaussian along the axis parallel to the wave, and β is the sharpness along the axis perpendicular to the wave. It should be noted that the 2-D Gabor filter in (4.6) is not in the most general form of the 2-D



(a) (b)

Figure 4.3: 2-D Gabor filter. a) real part; b) imaginary part.

Gabor elementary function but the orientation of the elliptical Gaussian is the same as the orientation of the plane wave. In addition, the shifts in space and phase have been dropped because the function is used as a convolution filter. The response to input image $\xi(x, y)$ is then

$$r_\xi(x, y) = \iint_{-\infty}^{\infty} \psi(x - x', y - y')\xi(x', y')\, dx'\, dy'. \tag{4.7}$$

The filter in (4.6) can be normalised by fixing the ratio of the frequency of the wave and the sharpness values of the Gaussian, i.e., $\gamma = \frac{f_0}{\alpha}$, $\eta = \frac{f_0}{\beta}$. Thus, the spatial filter includes a constant number of waves. This formulation fixes the behaviour of the response regardless of the frequency and makes the DC-response identical for all frequencies. It is desired that the DC-response is small, otherwise the average image intensity affects the response. This can be controlled by setting parameter γ large enough. Another approach is to use a modified Gabor filter where a Gaussian with the same DC-response is subtracted from the filter [128]. To make the area under the Gaussian unity, a normalisation factor $\frac{\alpha\beta}{\pi}$ has to be used. Thus, a normalised filter can be presented as

$$\psi(x, y) = \frac{f_0^2}{\pi\gamma\eta} e^{-\left(\frac{f_0^2}{\gamma^2}x'^2 + \frac{f_0^2}{\eta^2}y'^2\right)} e^{j2\pi f_0 x'}. \tag{4.8}$$

Using Fourier transform, (4.8) can be presented in the frequency domain as

$$\begin{aligned} \Psi(u, v) &= e^{-\frac{\pi^2}{f_0^2}\left(\gamma^2\left(u' - f_0\right)^2 + \eta^2 v'^2\right)}, \\ u' &= u\cos\theta + v\sin\theta, \\ v' &= -u\sin\theta + v\cos\theta. \end{aligned} \tag{4.9}$$

Thus, in the frequency domain the filter is a real Gaussian with centroid at frequency f_0 at orientation θ. The parameters of a frequency domain filter are illustrated in Fig. 4.4.

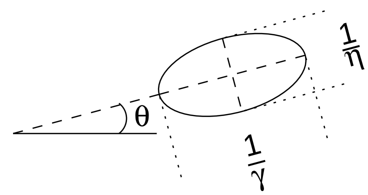


Figure 4.4: Gabor filter parameters in frequency domain.

Several Gabor filters are usually combined to form a filter bank. The filter bank is usually composed of filters in several orientations and frequencies, with equal orientation spacing and octave frequency spacing, as presented in Fig. 4.5, while the relative widths γ and η stay constant. That is,

$$\theta_k = \frac{k\pi}{n_\theta} \quad k = \{0, 1, \ldots, n_\theta - 1\} \qquad f_l = s^{-l} f_{max} \quad l = \{0, 1, \ldots n_f - 1\}, \tag{4.10}$$

where θ_k is the kth orientation, n_θ the number of orientations, f_l the lth frequency, f_{max} the maximal frequency, n_f the number of frequencies, and s ($>$1) the ratio between two consecutive frequencies. As shown in the figure, only a half of the frequency plane needs to be covered, because the input to the filters is assumed to be real and thus its frequency representation is symmetric and Hermitian [13].

As Bovik et al. note, parameter selection is a nontrivial problem having no simple solution [12]. The most common criteria proposed in the literature include manual selection (e.g., [102]), biological considerations (e.g., [73]) and optimisation (e.g., [45, 28]). If the selection is based on the knowledge about primate visual system, there is a drawback that it is unclear, what has been the goal of evolution in the visual system, that is, to which purpose the visual system has evolved. On the other hand, optimisation based selection has the drawback of being dependent on the training set.

If the feature to be detected can be formulated analytically, the selection can also be based on analysis of the filter response. Mehrotra et al. optimise the filter parameters for edge detection based on the criteria proposed by Canny [81]. Chen et al. select some parameters based on experimental results while others are selected based on mathematical analysis [19]. Analytic parameter selection has also been used in image representation using GEFs [73]. In *Publication IV*, a method for selecting Gabor filter parameters is presented for direction sensitive edge detection in binary images. The method is based on the analysis of the filter spatial size and its effect on the angular accuracy.

Despite the problems of parameter selection, Gabor filters have a number of attractive qualities. First, they can be used to extract various kinds of visual features, including texture [12], edges [81], lines [19], and shapes [53]. This is illustrated in Fig. 4.6 where the absolute responses of a coin image to filters in several orientations and frequencies are shown. Also the similarity with the simple cells of the mammalian visual cortex supports the representation power of the filters. Second, Gabor filter responses are robust. It is known that the amplitudes of complex Gabor coefficients are invariant under small translation, rotation, and scaling. An explanation for this is the shiftability

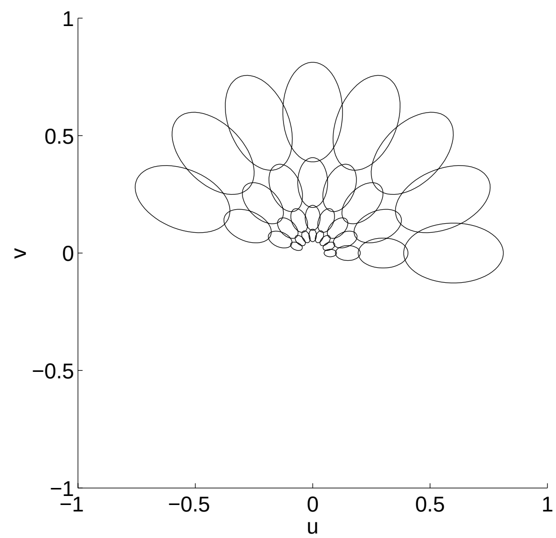


Figure 4.5: A bank of 2-D Gabor filters in frequency domain.

of the Gabor filters in spatial and frequency domains. While the shiftability requires a nonorthogonal, overcomplete representation [104], it provides the required robustness. In addition, shiftability makes it possible to interpolate responses both in spatial and frequency coordinates. Furthermore, the Gaussian nature of the filters makes them tolerant to noise [54]. However, the stability of the responses concerns primarily the amplitudes of the complex filter responses, while quite often only the real or imaginary part of the filter is used. In some applications, such as edge detection, this is justified, but generally the robustness and smoothness of the filter response is lost.

4.2 Applications in object recognition

Gabor filters have been applied to several tasks in object recognition, including segmentation, representing local features, and extracting size information. Some applications depend on a separate segmentation step before recognition can be performed. For that reason, the use of Gabor filters in segmentation will be considered first before moving to their use in the representation of objects, which will form the most significant part of this section.

The power of inspecting locally prominent frequencies has made Gabor filters popular in texture analysis. This has been utilised by Jain et al. for the segmentation of objects in complex backgrounds [47]. A multi-channel decomposition of an image is first performed, followed by a selection of channels based on minimising the reconstruction error. The responses are subjected to a sigmoid nonlinearity, averaged, and then clustered to form the segments. A bottom-up merging has also been used to segment images based on Gabor filter responses and colour information [102].

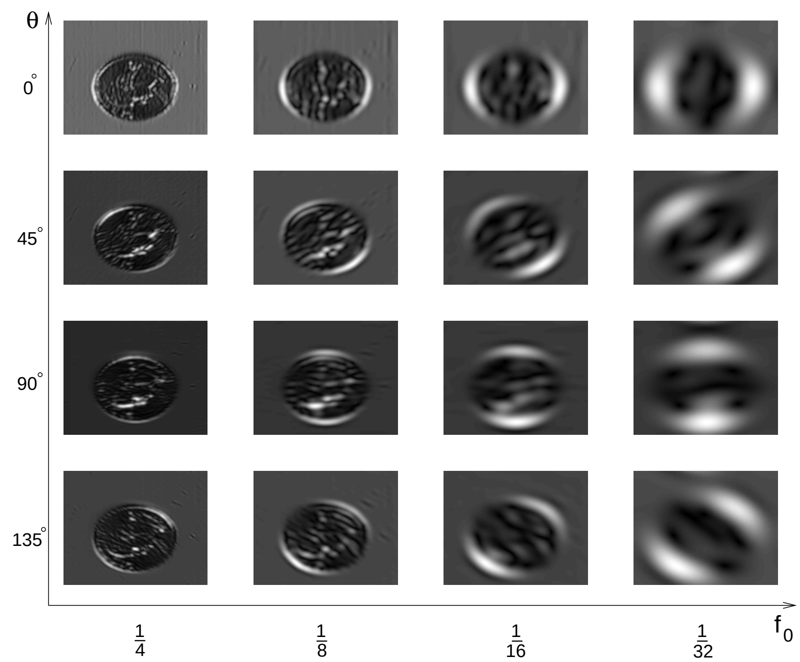


Figure 4.6: Absolute filter responses in different frequencies and orientations.

In addition to segmentation, Gabor filters have been applied as features to represent objects. These approaches can be divided into four categories depending on the type of the information used: a) response at a single point is used; b) responses on a predefined grid of points are used; c) responses on a deformable grid are used; d) global features of responses over an area are used.

Gabor filter response on a single point has been used to realise affine invariant recognition of image patches [6, 123, 7]. Full affine invariance was established by a log-log mapping of the frequency domain together with filters in several orientations. However, the image needs to be segmented in order to find the centres of the image patches. Weber and Casasent performed distortion invariant recognition by an entirely different concept using a linear combination of several separately optimised filters [124, 125]. The relative filter positions were also optimised in addition to other filter parameters. While the distortion tolerance of the system is good, there was no invariance to geometric transformations. The author has used Gabor filters to perform distortion tolerant rotation invariant recognition of electronic components [54, 53]. The recognition is based on a robust estimation of the dimensions of an object in different orientations around a single point. For a single sampling point, it is possible to realise invariances using analysis of the filter response.

The Gabor responses on a rigid predefined grid have been used to recognise handwritten numerals [37]. Filter parameters were optimised by hand with a 1-NN classification of the response vectors. Walter and Arnrich propose to use a grid of nine real (even) Gabor filters with a backpropagating multi-layer perceptron to position a robot manipulator [122]. Jain et al. use even symmetric (real) Gabor filters to extract the local ridge structures of fingerprints [46]. The fingerprint is inspected using a circular grid around a detected reference point. For each cell of the grid, the deviation of the filter responses for each filter is used to construct the feature vector. Altogether, none of the systems employing rigid sampling of responses achieve rotation or scale invariances.

Invariance to small scale geometric deformations has been achieved through the use of a deformable graph to represent the Gabor filter responses [15, 14, 65]. A single node of a graph represents the filter responses of a bank of filters, often called a Gabor jet [15]. The model graph of an object can be constructed either using sampling on a regular grid [15, 101] or sampling in interest locations [112]. The model graph is then compared with an image to find the locations for each node in the image that minimise both the geometric deformation of the graph and the difference of filter responses at these locations. This approach is invariant to small changes in the locations of the graph nodes which may result, for example, from geometric transformations. However, full invariance to rotation or scale is not achieved as the filter responses are directly compared yet the responses are invariant only to *small* changes in scale and rotation. This restriction is alleviated by storing several model graphs for objects of different sizes. The approach has been applied to face recognition [15, 14, 65, 128] and classification of hand postures [112, 113]. Krüger and others present a variant that utilises Gabor features to extract local line segments which are then fed to a learning process that outputs the model graph [61, 62]. Park and Yang combine the use of local Gabor features with Randomized Hough transform type evidence accumulation to estimate the parameters of a similarity transform (translation, rotation, and scaling) [88]. Local Gabor features similar to Gabor jets are used to find candidates of local structural elements which are then used as partial evidences. The matching of local elements is performed in a rotation invariant manner which should guarantee full rotational invariance while the invariance in scale remains small.

Gabor filter responses can be combined over an area to achieve a global descriptor for that area. Lampinen and Oja [69] employ clustering of the responses of a Gabor jet to assign each pixel to a single cluster. Then, the histogram of clusters over a Gaussian window is used as an input to a subspace classifier. Douville [26] uses the responses summed over the image. Shioyama et al. [102] compute a histogram of responses over a segmented image region as a description of that segment. They demonstrate the performance by detecting cars in real traffic scenes. The global systems presented are only scale and translation tolerant and do not achieve full invariance.

In *Publication IV*, invariant recognition properties of Gabor filters are examined in the context of directional edge detection in binary images. A method to select filter parameters is presented which assures that the circular Gaussian is able to capture the required number of angles. Translation invariance is achieved by summing the filter responses over the image. That is, a feature vector $\mathbf{G}$ is constructed from the individual responses $r(x, y; \theta)$ as

$$\begin{aligned}\mathbf{G} &= \begin{pmatrix} G_1 & G_2 & \cdots & G_n \end{pmatrix} \\ &= \begin{pmatrix} \sum_{x,y} |r(x,y;\theta_1)| & \sum_{x,y} |r(x,y;\theta_2)| & \cdots & \sum_{x,y} |r(x,y;\theta_n)| \end{pmatrix}\end{aligned} \tag{4.11}$$

where n is the number of filters in different orientations. Because the filter responses represent the edges in different orientations, scale invariance can be realised by normalising the feature magnitudes. Let $\mathbf{G}'$ denote the normalised feature,

$$\mathbf{G}' = \frac{1}{\sum_{i=1}^{n} G_i}\mathbf{G}. \tag{4.12}$$

Rotation invariant distance measure for the features can be presented by first defining the rotation of the feature vector as

$$\mathbf{G}^{(\theta+k)} = \begin{pmatrix} G_{n-k+1} & \cdots & G_n & G_1 & \cdots & G_{n-k} \end{pmatrix} \tag{4.13}$$

where $k = 0 \ldots n-1$ is the rotation index. Then the rotation invariant squared Euclidean distance of two feature vectors can be defined as the minimum over all rotations, that is,

$$d(\mathbf{G}, \mathbf{H}) = \min_k \left\{ \sum_{i=1}^{n} \left(G_i' - H_i^{\theta+k} \right)^2 \right\} \tag{4.14}$$

where $\mathbf{G}$ and $\mathbf{H}$ are the feature vectors. In *Publication IV* the invariances are demonstrated with an experiment with digits. A preliminary version of *Publication IV* was published as [64].

In *Publication V*, the recognition method is applied to matching of binary symbols. The method is inspected using the larger data set presented in more detail in *Publication III*. Both the Gabor filtering and Hough transform methods are based on the construction of a histogram of edges in different orientations. The noise tolerance of the Gabor filtering based image matching is found out to be superior to the Hough transform generated features.

Chapter V

Discussion

Invariant object recognition has been recognised as one of the most central problems in computer vision. Feature extraction is an important part of a vision system which converts digital image data into higher level features. In this thesis, new feature extraction methods have been presented and analysed. Methods based on Hough transform and Gabor filtering have been improved. The publications will now be reviewed based on the objectives of an object recognition system, and then discussed one by one concentrating on the contributions and restrictions.

In Chapter 1, the objectives of an object recognition system were defined as: "The system should be *general* enough to recognise a large variety of objects, *invariant* to natural variations, *stable* against distortions, and *computationally efficient*." These can be used to survey the contents of this thesis. Both the line segments detected by the Hough transform and the Gabor features can be thought of as general shape primitives because they are not limited to a single application and are able to represent many objects, particularly those which are man made and have geometric shapes. However, it seems likely that these low level features cannot be used directly but must be used in the generation of a higher level description.

It is doubtful if global shape type features can be used for the recognition of complex objects. The shapes of such objects can be complex and often cannot be described by a single contour. Also, the global features can often be spatially instable, that is, they are not equally sensitive over the whole shape, and the objects need to be segmented from the background in order to use global features. While local features can overcome these problems, at least partially, the question is how local and global information should be combined. The idea of combining local and global information can be seen both in the Hough and Gabor approaches. The Hough transform is inherently an evidence gathering process that accumulates local evidence to examine the global configuration. In *Publication I* a robust local estimation technique is presented to estimate the parameters of a line segment inside a local window to be used in the evidence gathering. The inspection system presented in *Publication II* utilises information on local line segments and centroids of circles fitting and compares it to a known global model. Likewise, Gabor filtering provides a way to robustly extract local directional edges and lines, which can then be combined to a global description as in *Publication IV* and *Publication V*.

Translation, rotation, and scale invariant recognition using global Gabor features is presented in *Publication IV* and *Publication V*. In *Publication II* the rotation invariant recognition was realised using local features with known spatial relationships. The Hough transform based translation, rotation, and scale invariant features presented in *Publication III* were based on global and local line orientations.

The existence of truly invariant features is questionable even in the human visual system, as it has been found that humans recognise certain views of an object more quickly than others [87]. In addition, the human visual system can be primed to certain poses of an object. For example, alphabetic characters are recognised more quickly in their normal position than upside down. An important note is that a human being can still recognise these characters regardless of the pose even though the recognition time changes. While this evidence is contradictory to the existence of global invariants, research in invariant recognition is certainly warranted and the evidence may, in addition, provide ideas for further approaches to computer vision. It is not very likely that the global features will provide enough information for discriminating images in large data sets. It seems that responses of Gabor filters are powerful as descriptors of local parts but the local parts have to be combined in a global description.

The stability against distortions is one of the most important qualities of a computer vision system which operates in natural surroundings. In *Publication I* the noise tolerance of the connective randomised Hough transform was improved with a new algorithm for the local estimation. The experiments in *Publication V* verify the noise tolerance of the Gabor filtering based features. Comparing the Hough and Gabor approaches it seems that the local features extracted by Gabor filtering are more stable against such distortions as noise and small displacements of the target. However, the evidence gathering nature of the Hough transform makes it more tolerant against missing data. Thus, it can only be concluded that while the stability of features is desirable, the optimally stable feature extraction method depends on the application.

The Hough transform is known to be computationally demanding. In *Publication I* a variant was proposed that reduces the computational burden by employing local information. While the method for local estimation is more complex than in the preceding method, CRHT, the total computation time is decreased in some cases due to the better estimates of the local structure. While separate analyses of parts of the algorithm can reveal important information, they cannot be combined directly to form a truth about the whole.

The standard Hough transform is normally impractically burdensome. In *Publication VI* it was shown that with modern parallel environments computation can be sped up considerably. The focus was on the Randomized Hough transform, which has been found out to outperform the standard Hough transform. A proposed parallel algorithm was discovered to be superior both to sequential randomised and parallel standard Hough transforms in the chosen environments. From the research the conclusion can be drawn that a low-scale parallelisation of known algorithms is often possible in an efficient way. On the other hand, in highly parallel environments the algorithm design should be strongly related to the characteristics of the environment, otherwise the performance will suffer considerably.

Publication I presents a new method to extract local information to improve the robustness of evidence accumulation in CRHT. The underlying idea of using the local information has been presented by other authors, but the novel contribution of the paper is the robust local estimation method. A restriction of the technique is that it only helps in the detection problem, that is, it is helpful when a random subset of evidences, i.e., edge points, is missing causing short gaps in local line segments. In addition, if only short gaps are present, the method can lower execution time because of the better estimates of line parameters. However, the method is not applicable for long gaps in the lines because the execution time grows intolerable.

Publication II presents a system for inspecting two-dimensional sheet metal parts. The contribution of the publication is in combining existing methods into a real-world application, and in the analysis of the measurement accuracy. A drawback of the system is that it is based on two-dimensional models of the parts. Thus, the calibration is estimated using two-dimensional similarity transform, the part and the image plane of the camera must lie parallel, and the thickness of the parts cannot be taken into account.

Publication III presents new global features based on Hough transform for matching images. Originally, it was desired that these features could be used for discriminating between different types of documents such as electrical and architectural drawings and maps because in these technical drawings, the orientations of lines often depend on the type of the drawing. However, it was discovered that the feature vector did not have strong correlation to the drawing type. Nevertheless, the features were found capable of discriminating line drawing symbols. A major restriction of the method as a global feature is that successful segmentation must be performed before recognition. For recognising the drawing type, it seems that a more effective scheme could be constructed by identifying typical primitives for each type, and performing the recognition based on these.

Publication IV presents a global, Gabor filtering based feature for recognition. The publication has two primary contributions, the selection of filter parameters and the global feature. The method of parameter selection is based on a formulation of a Gabor filter that forces the filter envelopes to be circular Gaussians. In addition, the relative frequency bandwidth is selected. Therefore, the spatial size of a filter could be controlled only using the frequency. However, as stated in Chapter 4, the frequency and orientation bandwidths can be controlled independently using parameters γ and η, which would seem to be a more reasonable way to solve the problem of parameter selection. The histogram like scaling and rotation invariant global feature presented is to author's knowledge original, but it is closely related to the energy on a certain band, which has previously been used and is also computationally lighter to compute. The necessity of segmentation applies also to this feature as the result of its global nature, and presents one of the major limitations for the application of the feature.

Publication V presents the application of the global Gabor features to symbol recognition and compares the results to those in *Publication III*. The noise tolerance of the Gabor feature is found superior to Hough transform based features. This is an important attribute of using Gabor type features, as the noise tolerance is often very good because the use of a limited range of the frequency spectrum suppresses noise on other frequencies. It would be beneficial to compare the presented features to other global features such as moment invariants to better demontrate their strengths and weaknesses in applications.

Publication VI presents parallel Hough transform algorithms. The contributions of the publication are the parallel algorithm for the Randomized Hough transform and the assessment of suitability of the parallel environment to the Hough transform. While the results show that the use of networked PCs can decrease the computation time of HT, it is not very likely that the environment would be useful in a real application because in the case of standard network hardware, the network latencies are always considerable and thus real time performance is hard to obtain if even possible. However, the use of multiprocessor PCs seems to provide a cost efficient platform for image processing and computer vision. A problem with the parallel RHT algorithm is that while good performance is obtained, the algorithm does not scale up very well to a large number of processors because of the random unpredictable nature of the algorithm. For optimal performance in multiprocessor environments it could be beneficial to study other efficient HT variants, such as the Adaptive HT, that do not have random behaviour and would be more suitable to a parallel environment.

Some directions for future research can already be seen. First, incorporating the higher level structure to the local features seems to be necessary, since global invariant features do not seem to be effective in discriminating large and complex data sets. The spatial relationships of the local features should be included in the representation. The current graph based approaches suffer from the fact that while the features or the model matching can be invariant, the non-invariant knowledge about local features, such as the orientation, is not included in the matching. Another area of research is the application of learning. Research on artificial neural networks has produced the means to incorporate knowledge through training by examples. However, the downside of neural networks is that their acquired knowledge is often very hard to interpret after the learning. In contrast, traditional artificial intelligence techniques, such as decision trees, can be interpreted, but they often do not tolerate uncertainty and also cannot describe complex relationships. Therefore, future research is clearly necessary in the use of learning, both inside the feature extraction and classification stages and in the interaction between the stages. In 1960, Selfridge and Neisser [99] wrote: “No current program can generate test features of its own.” While this is no longer true as general low-level features are known, knowledge about the interaction between the stages of a vision system is still mostly nonexistent, and it is not clear, what kind of general higher level descriptions should be used.

Bibliography

[1] Alferez, R., and Wang, Y.-F. Geometric and illumination invariants for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 6 (June 1999), 505–536.

[2] Arbter, K., Snyder, W. E., Burkhardt, H., and Hirzinger, G. Application of affine-invariant Fourier descriptors to recognition of 3-D objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence 12*, 7 (July 1990), 640–647.

[3] Atiquzzaman, M. Pipelined implementation of the multiresolution Hough transform in a pyramid multiprocessor. *Pattern Recognition Letters 15*, 9 (1994), 841–851.

[4] Ballard, D. H. Generalizing the Hough transform to detect arbitrary shapes. *Patter Recognition 13*, 2 (1981), 111–122.

[5] Belkasim, S. O., Shridhar, M., and Ahmadi, M. Pattern recognition with moment invariants: A comparative study and new results. *Pattern Recognition 24*, 12 (1991), 1117–1138.

[6] Ben-Arie, J., and Wang, Z. Pictorial recognition using affine-invariant spectral signatures. In *Proceedings. 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1997), pp. 34–39.

[7] Ben-Arie, J., and Wang, Z. Pictorial recognition of objects employing affine invariance in the frequency domain. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20*, 6 (June 1998), 604–618.

[8] Ben-Tzvi, D., Leavers, V. F., and Sandler, M. B. A dynamic combinatorial Hough transform. In *Proceedings of the 5th International Conference on Image Analysis and Processing* (Positano, Italy, Sept. 1989), pp. 152–159.

[9] Ben-Tzvi, D., Naqvi, A. A., and Sandler, M. Efficient parallel implementation of the Hough transform on a distributed memory system. *Image and Vision Computing 7*, 3 (Aug. 1989), 167–172.

[10] Bhandarkar, S. M., Arabnia, H. R., and Smith, J. W. A reconfigurable architecture for image processing and computer vision. *International Journal of Pattern Recognition and Artificial Intelligence 9*, 2 (Apr. 1995), 201–229.

[11] Bolles, R. C., and Cain, R. A. Recognizing and locating partially visible objects: The local feature focus method. *International Journal of Robotics Research 1* (1982), 57–82.

[12] Bovik, A. C., Clark, M., and Geisler, W. S. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence 12*, 1 (Jan. 1990), 55–73.

[13] Bracewell, R. N. *The Fourier Transform and Its Applications*, 2 ed. McGraw-Hill, Inc., 1978.

[14] Buhmann, J., Lades, M., and von der Malsburg, C. Size and distortion invariant object recognition by hierarchical graph matching. In *IJCNN International Joint Conference on Neural Networks* (1990), vol. 2, pp. 300–311.

[15] Buhmann, J., Lange, J., and von der Malsburg, C. Distortion invariant object recognition by matching hierarchically labeled graphs. In *IJCNN International Joint Conference on Neural Networks* (1989), vol. 1, pp. 155–159.

[16] Caelli, T., Ferraro, M., and Barth, E. Aspects of invariant pattern and object recognition. In *Neural Networks for Perception*, H. Wechsler, Ed., vol. 1. Academic Press, 1991, pp. 234–247.

[17] Canny, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8*, 6 (Nov. 1986), 679–698.

[18] Casasent, D., and Psaltis, D. New optical transforms for pattern recognition. *Proceedings of the IEEE 65*, 1 (Jan. 1977), 77–78.

[19] Chen, J., Sato, Y., and Tamura, S. Orientation space filtering for multiple orientation line segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 5 (May 2000), 417–429.

[20] Choi, M.-S., and Kim, W.-Y. A novel two stage template matching method for rotation and illumination invariance. *Pattern Recognition 35* (2002), 119–129.

[21] Choudhary, A., and Ponnusamy, R. Implementation and evaluation of Hough transform algorithms on a shared-memory multiprocessor. *Journal of Parallel and Distributed Computing 12*, 2 (1991), 178–188.

[22] Comino, N., and Narasimhan, V. L. Analysis of two image processing algorithms on parallel computer systems. In *1st IWPP. Proceedings of the First International Workshop on Parallel Processing* (Bangalore, India, Dec. 1994), pp. 287–291.

[23] Da Fontoura Costa, L. Assessing concurrent hardware alternatives for the implementation of the binary Hough transform. *Applied Signal Processing 1*, 4 (1994), 195–203.

[24] Daugman, J. G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A 2*, 7 (July 1985), 1160–1169.

[25] Deans, S. R. Hough transform from the Radon transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-3*, 2 (Mar. 1981), 185–188.

[26] Douville, P. Real-time classification of traffic signs. *Real-Time Imaging 6* (2000), 185–193.

[27] Duda, R., and Hart, P. Use of Hough transformations to detect lines and curves in pictures. *Comm. ACM 15*, 1 (1972), 11–15.

[28] Dunn, D. F., and Higgins, W. E. Optimal Gabor-filter design for texture segmentation. In *ICASSP-93. 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Minneapolis, MN, USA, Apr. 1993), vol. 5, pp. 37–40.

[29] Forsyth, D., Mundy, J. L., and Zisserman, A. Transformational invariance — a primer. *Image and Vision Computing 10*, 1 (Jan. 1992), 39–45.

[30] Fukushima, K., Miyake, S., and Ito, T. Neocognitron: a neural model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics SMC-13* (1983), 826–834.

[31] Fung, P.-F., Lee, W. S., and King, I. Randomized generalized Hough transform for 2-D grayscale object detection. In *Proceedings of the 13th International Conference on Pattern Recognition* (Vienna, Austria, Aug. 1996), vol. 2, pp. 511–515.

[32] Gabor, D. Theory of communications. *Journal of International Electrical Engineers 93* (1946), 427–457.

[33] Ginsburg, A. P. Spatial filtering and visual form perception. In *Handbook of perception and human performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds. Wiley, New York, 1986, pp. 1–41.

[34] Götze, N., Drüe, S., and Hartmann, G. Invariant object recognition with discriminant features based on local Fast-Fourier Mellin transform. In *Proceedings of 15th International Conference on Pattern Recognition* (Barcelona, Spain, Sept. 2000), vol. 1, pp. 948–951.

[35] Granlund, G. H. In search of a general picture processing operator. *Computer Graphics and Image Processing 8* (1978), 155–173.

[36] Guil, N., and Zapata, E. L. Fast Hough transform on multiprocessors: A branch and bound approach. *Journal of Parallel and Distributed Computing 45*, 1 (1997), 82–89.

[37] Hamamoto, Y., Uchimura, S., Watanabe, M., Yasuda, T., and Tomita, S. Recognition of handwritten numerals using Gabor features. In *Proceedings of the 13th International Conference on Pattern Recognition* (Vienna, Austria, Aug. 1996), vol. 3, pp. 250–253.

[38] Hough, P. Method and means for recognizing complex patterns. U.S. Patent 3069654, 1962.

[39] Hu, M.-K. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory IT-8* (Feb. 1962), 179–187.

[40] Hubel, D. H., and Wiesel, T. N. Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology 148* (1959), 574–591.

[41] Hubel, D. H., and Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology 195* (1968), 215–243.

[42] Huttenlocher, D. P., and Ullman, S. Object recognition using alignment. In *Proceedings 1st International Conference on Computer Vision* (London, UK, 1987), pp. 102–111.

[43] Illingworth, J., and Kittler, J. The adaptive Hough transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9*, 5 (Sept. 1987), 690–698.

[44] Illingworth, J., and Kittler, J. A survey of the Hough transform. *Computer Vision, Graphics, and Image Processing 44*, 1 (1988), 87–116.

[45] Jain, A. K., and Farrokhnia, F. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition 24*, 12 (1991), 1167–1186.

[46] Jain, A. K., Prabhakar, S., Hong, L., and Pankati, S. Filterbank-based fingerprint matching. *IEEE Transactions on Image Processing 9*, 5 (May 2000), 846–859.

[47] Jain, A. K., Ratha, N. K., and Lakshmanan, S. Object detection using Gabor filters. *Pattern Recognition 30*, 2 (1997), 295–309.

[48] Jia, L., and Kitchen, L. Object-based image similarity computation using inductive learning of contour-segment relations. *IEEE Transactions on Image Processing 9*, 1 (Jan. 2000), 80–87.

[49] Jones, III, G., and Bhanu, B. Recognition of articulated and occluded objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 7 (July 1999), 603–613.

[50] Kabrisky, M. *A proposed model for visual information processing in the human brain.* University of Illinois Press, Urbana, IL, 1966.

[51] Kälviäinen, H., and Hirvonen, P. An extension to the randomized Hough transform exploiting connectivity. *Pattern Recognition Letters 18*, 1 (1997), 77–85.

[52] Kälviäinen, H., Hirvonen, P., Xu, L., and Oja, E. Probabilistic and non-probabilistic Hough transforms: Overview and comparisons. *Image and Vision Computing 13*, 4 (1995), 239–252.

[53] Kamarainen, J.-K., Kyrki, V., and Kälviäinen, H. Fundamental frequency Gabor filters for object recognition. In *16th International Conference on Pattern Recognition (to be published).*

[54] Kamarainen, J.-K., Kyrki, V., and Kälviäinen, H. Noise tolerant object recognition using Gabor filtering. In *14th International Conference on Digital Signal Processing (to be published)* (July 2002).

[55] Kao, T. W., Horng, S. J., and Wang, Y. L. An O(1) time algorithms for computing histogram and Hough transform on a cross-bridge reconfigurable array of processors. *IEEE Transactions on Systems, Man and Cybernetics 25*, 4 (Apr. 1995), 681–687.

[56] Kassim, A. A., Tan, T., and Tan, K. H. A comparative study of efficient generalised Hough transform techniques. *Image and Vision Computing 17* (1999), 737–748.

[57] Khalil, M. I., and Bayoumi, M. M. A dyadic wavelet affine invariant function for 2D shape recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 10 (Oct. 2001), 1152–1164.

[58] Kimme, C., Ballard, D. H., and Sklansky, J. Finding circles by an array of accumulators. *Communications of the ACM 18* (1975).

[59] Kiryati, N., Eldar, Y., and Bruckstein, A. M. A probabilistic Hough transform. *Pattern Recognition 24*, 4 (1991), 303–316.

[60] Krishnaswamy, D., and Banerjee, P. Exploiting task and data parallellism in parallel Hough and Radon transforms. In *Proc. Int. Conference on Parallel Processing* (1997), pp. 441–444.

[61] Krüger, N., and Lüdtke, N. Orassyll: object recognition with autonomously learned and sparse symbolic representations based on local line detectors. In *BMVC98 Ninth British Machine Vision Conference* (1998), vol. 2, pp. 488–497.

[62] Krüger, N., and Peters, G. ORASSYLL: Object recognition with autonomously learned and sparse symbolic representations based on metrically organized local line detectors. *Computer Vision and Image Understanding 77* (2000), 48–77.

[63] Kumarasamy, S., and Shakaff, A. Y. M. Comparative study of implementation performance for intermediate-level image-processing functions on transputer- and DSP-based multiprocessor systems. In *Transputer Research and Application 7. Proceedings of the Seventh Conference of the North American Transputer Users Group* (Atlanta, GA, USA, 1995), pp. 79–90.

[64] Kyrki, V., Kamarainen, J.-K., and Kälviäinen, H. Invariant shape recognition using global Gabor features. Research Report 72, Lappeenranta University of Technology, Department of Information Technology, Finland, 2000.

[65] Lades, M., Vorbrüggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R. P., and Konen, W. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers 42*, 3 (Mar. 1993), 300–311.

[66] Lamdan, Y., Schwartz, J. T., and Wolfson, H. J. Object recognition by affine invariant matching. In *Proceedings CVPR '88* (Ann Arbor, MI, USA, June 1988), pp. 335–344.

[67] Lamdan, Y., Schwartz, J. T., and Wolfson, H. J. Affine invariant model-based object recognition. *IEEE Transactions on Robotics and Automation 6*, 5 (Oct. 1990), 578–589.

[68] Lampinen, J. *Neural Pattern Recognition: Distortion Tolerance by Self-Organizing Maps.* PhD thesis, Lappeenranta University of Technology, 1992.

[69] Lampinen, J., and Oja, E. Distortion tolerant pattern recognition based on self-organizing feature extraction. *IEEE Transactions on Neural Networks 6*, 3 (May 1995), 539–547.

[70] Leavers, V. F. The dynamic generalized Hough transform: Its relationship to the probabilistic Hough transforms and an application to concurrent detection of circles and ellipses. *CVGIP: Image Understanding 56*, 3 (1992), 381–398.

[71] Leavers, V. F. Survey — Which Hough transform? *CVGIP: Image Understanding 58* (1993), 250–264.

[72] Lee, H. M., Kittler, J., and Wong, K. C. Generalised Hough transform in object recognition. In *Proceedings. 11th IAPR International Conference on Pattern Recognition* (Hague, Netherlands, Aug. 1992), vol. 3, pp. 285–289.

[73] Lee, T. S. Image representation using 2D Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*, 10 (Oct. 1996), 959–971.

[74] Li, H., Lavin, M. A., and Le Master, R. J. Fast Hough transform: A hierarchical approach. *Computer Vision, Graphics, and Image Processing 36*, 2–3 (1986), 139–161.

[75] Livingstone, M. S., and Hubel, D. Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science 240* (1988), 740–749.

[76] Lowe, D. G. *Perceptual Organization and Visual Recognition.* Kluwer Academic, Boston, MA, 1985.

[77] Lowe, D. G. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision* (Corfu, 1999), pp. 1150–1157.

[78] Ma, D., and Chen, X. Hough transform using slope and curvature as local properties to detect arbitrary 2D shapes. In *9th International Conference on Pattern Recognition* (Rome, Italy, Nov. 1988), pp. 511–513.

[79] Mach, E. *The analysis of sensations.* Open Court, Chicago, 1914/1959.

[80] Marcelja, S. Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am. 70* (1980), 1297–1300.

[81] Mehrotra, R., Namuduri, K., and Ranganathan, N. Gabor filter–based edge detection. *Pattern Recognition 25*, 12 (1992), 1479–1494.

[82] Meribout, M., Nakanishi, M., and Ogura, T. Accurate and real-time image processing on a new PC-compatible board. *Real-Time Imaging 8*, 1 (Feb. 2002), 35–51.

[83] Merlin, P. M., and Farber, D. J. A parallel mechanism for detecting curves in pictures. *IEEE Transactions on Computers C24* (1975), 96–98.

[84] Nakanishi, M., and Ogura, T. A real-time CAM-based Hough transform algorithm and its performance evaluation. In *Proceedings of the 13th International Conference on Pattern Recognition* (Vienna, Austria, Aug. 1996), vol. 2, pp. 516–521.

[85] Palmer, S. E. Visual perception and world knowldege: Notes on a model of sensory-cognitive interaction. In *Explorations in cognition*, D. A. Norman and D. E. Rumelhart, Eds. W. H. Freeman, San Francisco, 1975.

[86] Palmer, S. E. *Vision Science — Photons to Phenomenology.* MIT Press, 1999.

[87] Palmer, S. E., Rosch, E., and Chase, P. Canonical perspective and the perception of objects. In *Attention and Performance*, J. Long and A. Baddeley, Eds., vol. 9. Erlbaum, Hillsdale, NJ, 1981, pp. 135–151.

[88] Park, H. J., and Yang, H. S. Invariant object detection based on evidence accumulation and Gabor features. *Pattern Recognition Letters 22* (2001), 869–882.

[89] Penev, P., and Atick, J. Local feature analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems 7*, 3 (1996), 477–500.

[90] Perantonis, S. J., and Lisboa, P. J. G. Translation, rotation, and scale invariant pattern recognition by high-order neural networks and moment classifiers. *IEEE Transactions on Neural Networks 3*, 2 (Mar. 1992), 241–251.

[91] Piotrowski, L. N., and Campbell, F. W. A demontration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception 11*, 3 (1982), 337–346.

[92] Porat, M., and Zeevi, Y. Y. The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence 10*, 4 (July 1988), 452–468.

[93] Prewitt, J. M. S. Object enhancement and extraction. In *Picture Processing and Psychopictorics*, B. S. Lipkin and A. Rosenfeld, Eds. Academic Press, New York, 1970.

[94] Princen, J., Illingworth, J., and Kittler, J. A formal definition of the hough transform: properties and relationships. *Journal of Mathematical Imaging and Vision 1* (1992), 153–168.

[95] Prokop, R. J., and Reeves, A. P. A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP Graphical Models and Image Processing 54*, 5 (1992), 438–460.

[96] Reiss, T. H. The revised fundamental theorem of moment invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence 13*, 8 (Aug. 1991), 830–834.

[97] Rosenfeld, A., and Kak, A. C. *Digital Picture Processing*. Academic Press, Inc., 1982.

[98] Rosenfeld, A., Ornelas, Jr., J., and Hung, Y. Hough transform algorithms for mesh-connected SIMD parallel processors. *Computer Vision, Graphics, and Image Processing 41*, 2 (1988), 293–305.

[99] Selfridge, O. G., and Neisser, U. Pattern recognition by machine. *Scientific American 203*, 2 (1960), 60–68.

[100] Ser, P. K., and Siu, W. C. Non-analytic object recognition using the Hough transform with the matching technique. *IEE Proceedings Computers and Digital Techniques 141*, 1 (Jan. 1994), 11–16.

[101] Shams, L. B., Brady, M. J., and Schaal, S. Graph matching vs mutual information maximization for object detection. *Neural Networks 14* (2001), 345–354.

[102] Shioyama, T., Wu, H. Y., and Mitani, S. Object detection with Gabor filters and cumulative histograms. In *Proceedings 15th International Conference on Pattern Recognition* (Barcelona, Spain, Sept. 2000), vol. 1, pp. 704–707.

[103] Simard, P. Y., LeCun, Y. A., Denker, J. S., and Victorri, B. Transformation invariance in pattern recognition—tangent distance and tangent propagation. *International Journal of Imaging System and Technology 11*, 3 (2001), 181–194.

[104] Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J. Shiftable multiscale transforms. *IEEE Transactions on Information Theory 38*, 2 (Mar. 1992), 587–607.

[105] Sirovich, L., and Kirby, M. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America 4* (1987), 519–524.

[106] Stockman, G. Object recognition and localization via pose clustering. *Computer Vision, Graphics, and Image Processing 40*, 3 (Dec. 1987), 361–387.

[107] Tarr, M. J. Visual representation: From features to objects. In *Encyclopedia of Human Behavior*, V. S. Ramachandran, Ed., vol. 4. Academic Press, San Diego, CA, 1994, pp. 503–512.

[108] Teh, C.-H., and Chin, R. T. On image analysis by the methods of moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence 10*, 4 (July 1988), 496–513.

[109] Thazhuthaveetil, M. J., and Shah, A. V. Parallel Hough transform algorithm performance. *Image and Vision Computing 9*, 2 (1991), 88–92.

[110] Thomas, A. D. H. Compressing the parameter space of the generalised Hough transform. *Pattern Recognition Letters 13*, 2 (Feb. 1992), 107–112.

[111] Tieng, Q. M., and Boles, W. W. Recognition of 2D object contours using the wavelet transform zero-crossing representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 8 (Aug. 1997), 910–916.

[112] Triesch, J., and von der Malsburg, C. Robust classification of hand postures against complex backgrounds. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition* (Killington, VT, USA, Oct. 1996), pp. 170–175.

[113] Triesch, J., and von der Malsburg, C. A system for person-independent hand posture recognition against complex backgrounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 12 (Dec. 2001), 1449–1453.

[114] Tsai, D.-M., and Chiang, C.-H. Rotation-invariant pattern matching using wavelet decomposition. *Pattern Recognition Letters 23* (2002), 191–201.

[115] Turney, J. L., Mudge, T. N., and Volz, R. A. Recognizing partially occluded parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-7* (July 1985), 410–421.

[116] Uenohara, M., and Kanade, T. Use of Fourier and Karhunen-Loeve decomposition for fast pattern matching with a large set of templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 8 (Aug. 1997), 891–898.

[117] Ullman, S. Aligning pictorial descriptions: An approach to object recognition. *Cognition* (1989), 193–254.

[118] Underhill, A., Atiquzzaman, M., and Ophel, J. Performance of the Hough transform on a distributed memory multiprocessor. *Microprocessors and Microsystems 22*, 7 (Jan. 1999), 355–362.

[119] Van Essen, D. C., and DeYoe, E. A. Concurrent processing in the primate visual cortex. In *The cognitive neurosciences*, M. S. Gazzaniga, Ed. MIT Press, Cambridge, MA, 1995, pp. 383–400.

[120] Van Veen, T. M., and Groen, F. C. A. Discretization errors in the Hough transform. *Pattern Recognition 14* (1981), 137–145.

[121] Wallace, R. S. A modified Hough transform for lines. In *Proceedings CVPR '85: IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Francisco, CA, USA, June 1985), pp. 665–667.

[122] Walter, J., and Arnrich, B. Gabor filters for object localization and robot grasping. In *Proceedings. 15th International Conference on Pattern Recognition* (Barcelona, Spain, Sept. 2000), vol. 4, pp. 124–127.

[123] Wang, Z., and Ben-Arie, J. SVD and log-log frequency sampling with Gabor kernels for invariant pictorial recognition. In *Proceedings of International Conference on Image Processing* (Santa Barbara, CA, USA, 1997), vol. 3, pp. 162–165.

[124] Weber, D., and Casasent, D. Quadratic Gabor filters for object detection. In *Proceedings of the 1997 South African Symposium on Communications and Signal Processing* (Grahamstown, South Africa, Sept. 1997), pp. 69–74.

[125] Weber, D. M., and Casasent, D. P. Quadratic Gabor filters for object detection. *IEEE Transactions on Image Processing 10*, 2 (Feb. 2001), 218–230.

[126] Wechsler, H., and Zimmerman, G. L. 2-D invariant object recognition using distributed associative memory. *IEEE Transactions on Pattern Analysis and Machine Intelligence 10*, 6 (Nov. 1988), 811–821.

[127] Weiss, I. Geometric invariants and object recognition. *International Journal of Computer Vision 10*, 3 (1993), 207–231.

[128] Wiskott, L., Fellous, J.-M., Krüger, N., and von der Malsburg, C. Face recognition by elastic bunch graph matching. *IEEE Transaction on Pattern Analysis and Machine Intelligence 19*, 7 (July 1997), 775–779.

[129] Wood, J. Invariant pattern recognition: A review. *Pattern Recognition 29*, 1 (1996), 1–17.

[130] Xu, L., Oja, E., and Kultanen, P. A new curve detection method: Randomized Hough transform (RHT). *Pattern Recognition Letters 11*, 5 (1990), 331–338.

[131] Yang, Z., and Cohen, F. S. Image registration and object recognition using affine invariants and convex hulls. *IEEE Transactions on Image Processing 8*, 7 (July 1999), 934–946.

[132] Yuen, S. Y. K., Lam, T. S. L., and Leung, N. K. D. Connective Hough transform. *Image and Vision Computing 11*, 5 (1993), 295–301.

[133] Zahn, X. T., and Roskies, R. S. Fourier descriptors for plane closed curves. *IEEE Transactions on Computers C-21* (Mar. 1972), 269–281.