OULU 2007
C 282

ACTA UNIVERSITATIS OULUENSIS

Esa Rahtu

A MULTISCALE FRAMEWORK FOR AFFINE INVARIANT PATTERN RECOGNITION AND REGISTRATION

FACULTY OF TECHNOLOGY,
DEPARTMENT OF ELECTRICAL AND INFORMATION ENGINEERING,
INFOTECH OULU,
UNIVERSITY OF OULU

ACTA UNIVERSITATIS OULUENSIS
C Technica 282

ESA RAHTU

A MULTISCALE FRAMEWORK FOR AFFINE INVARIANT PATTERN RECOGNITION AND REGISTRATION

Academic dissertation to be presented, with the assent of the Faculty of Technology of the University of Oulu, for public defence in Auditorium TS101, Linnanmaa, on November 2nd, 2007, at 12 noon

OULUN YLIOPISTO, OULU 2007

Acta Univ. Oul. C 282, 2007

Supervised by
Professor Janne Heikkilä

Reviewed by
Doctor Joni-Kristian Kämäräinen
Doctor Tomás Suk

ISBN 978-951-42-8600-1 (Paperback)
ISBN 978-951-42-8601-8 (PDF)
http://herkules.oulu.fi/isbn9789514286018/
ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)
http://herkules.oulu.fi/issn03553213/

Cover design
Raimo Ahonen

OULU UNIVERSITY PRESS
OULU 2007

Rahtu, Esa, A multiscale framework for affine invariant pattern recognition and registration
Faculty of Technology, University of Oulu, P.O.Box 4000, FI-90014 University of Oulu, Finland, Department of Electrical and Information Engineering, Infotech Oulu, University of Oulu, P.O.Box 4500, FI-90014 University of Oulu, Finland
Acta Univ. Oul. C 282, 2007
Oulu, Finland

Abstract

This thesis presents a multiscale framework for the construction of affine invariant pattern recognition and registration methods. The idea in the introduced approach is to extend the given pattern to a set of affine covariant versions, each carrying slightly different information, and then to apply known affine invariants to each of them separately. The key part of the framework is the construction of the affine covariant set, and this is done by combining several scaled representations of the original pattern. The advantages compared to previous approaches include the possibility of many variations and the inclusion of spatial information on the patterns in the features.

The application of the multiscale framework is demonstrated by constructing several new affine invariant methods using different preprocessing techniques, combination schemes, and final recognition and registration approaches. The techniques introduced are briefly described from the perspective of the multiscale framework, and further treatment and properties are presented in the corresponding original publications. The theoretical discussion is supported by several experiments where the new methods are compared to existing approaches.

In this thesis the patterns are assumed to be gray scale images, since this is the main application where affine relations arise. Nevertheless, multiscale methods can also be applied to other kinds of patterns where an affine relation is present.

An additional application of one multiscale based technique in convexity measurements is introduced. The method, called multiscale autoconvolution, can be used to build a convexity measure which is a descriptor of object shape. The proposed measure has two special features compared to existing approaches. It can be applied directly to gray scale images approximating binary objects, and it can be easily modified to produce a number of measures. The new measure is shown to be straightforward to evaluate for a given shape, and it performs well in the applications, as demonstrated by the experiments in the original paper.

Keywords: Affine invariant features, image alignment, image transforms, object recognition, pattern classification, shape analysis

Preface

The research for this thesis was carried out in the Machine Vision Group of the Department of Electrical and Information Engineering at the University of Oulu, Finland, and in the Albert-Ludwigs-Universität Freiburg, Germany during the years 2004-2007.

I am indebted to my supervisor Professor Janne Heikkilä for his guidance and support. He introduced me to the field of pattern recognition in general, and provided many interesting research topics throughout my studies. I also owe one to my friends and workmates Dr. Mikko Salo and Juho Kannala who had a great impact to the progress of the research and who made the work to be great fun. I am also grateful to Professor Matti Pietikäinen for offering me a possibility to work in the research group. Furthermore I would like to express my gratitude to Professor Hans Burkhardt, Alexandra Teynor, Lokesh Setia, and all other personnel in the Albert-Ludwigs-Universität Freiburg, who worked with me during my visit there. I would also like to thank Professor Jan Flusser for fruitful research collaboration.

I am grateful to Dr. Joni-Kristian Kämäräinen and Dr. Tomás Suk for reviewing the dissertation manuscript. I would like thank all my colleagues in the Machine Vision Group for building up a most pleasant and relaxed atmosphere in which to carry out the research work and discuss other (almost) related matters.

The financial support provided by the Infotech Oulu Graduate School, the Academy of Finland, the Nokia Foundation, the Foundation for the Promotion of Technology (TES), the Tauno Tönning Foundation, and the Seppo Säynäjäkangas Science Foundation, is gratefully acknowledged.

Finally special thanks go to my family and friends for their unconditional support and great times spent outside the scientific world.

Oulu, September 2007 Esa Rahtu

List of original articles

I Rahtu E, Salo M & Heikkilä J (2005) A new efficient method for producing global affine invariants. Proc 13th International Conference on Image Analysis and Processing (ICIAP 2005). Lecture Notes in Computer Science 3617: 407-414.

II Rahtu E, Salo M & Heikkilä J (2005) Affine invariant pattern recognition using multi-scale autoconvolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(6): 908-918.

III Kannala J, Rahtu E, Salo M & Heikkilä J (2005) A new method for affine registration of images and point sets. Proc 14th Scandinavian Conference on Image Analysis (SCIA 2005). Lecture Notes in Computer Science 3540: 224-234.

IV Kannala J, Rahtu E & Heikkilä J (2005) Affine registration with Multi-Scale Autoconvolution. Proc International Conference on Image Processing (ICIP 2005) 3: 1064-1067.

V Teynor A, Rahtu E, Setia L & Burkhardt H (2006) Properties of patch based approaches for the recognition of visual object classes. Proc 28th Jahrestagung der Deutschen Arbeitsgemeinschaft fr Mustererkennung (DAGM 2006). Lecture Notes in Computer Science 4174: 284-293.

VI Rahtu E, Salo M, Heikkilä J & Flusser J (2006) Generalized affine moment invariants for object recognition. Proc 18th International Conference on Pattern Recognition (ICPR 2006) 2: 634-637.

VII Rahtu E, Salo M & Heikkilä J (2006) A new affine invariant image transform based on ridgelets. Proc 16th British Machine Vision Conference (BMVC 2006) 3: 1059-1068.

VIII Rahtu E, Salo M & Heikkilä J (2006) J Multiscale autoconvolution histograms for affine invariant pattern recognition. Proc 16th British Machine Vision Conference (BMVC 2006) 3: 1039-1048.

IX Rahtu E, Salo M & Heikkilä J (2006) A new convexity measure based on a probabilistic interpretation of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9): 1501-1512.

X Rahtu E, Salo M & Heikkilä J (2007) Nonlinear functionals in the construction of multiscale affine invariants. Proc 15th Scandinavian Conference on Image Analysis (SCIA 2007). Lecture Notes in Computer Science 4522: 482-491.

The writing of Papers I, II, VI, VII, VIII, IX, and X, was mainly carried out by the author, who was also in charge of the experiments in these papers. The ideas were processed together with the co-authors, especially with Dr. Mikko Salo. With Papers III and IV, the author was closely involved in developing the ideas in the techniques introduced. With Paper V the author participated in the writing of the paper and performed some of the experiments, especially those involving multiscale autoconvolution.

Abbreviations

MSA	Multiscale autoconvolution
SMA	Spatial multiscale affine invariants
RANSAC	Random sample consensus
$\mathbb{R}$	Real numbers (Denoted also as **R** in the original publications)
$\mathbb{R}_+$	Positive real numbers
$\mathbb{R}^n$	n-dimensional space of real numbers
$\mathscr{A}$	Affine transformation, see Definition 1
$\mathscr{A}^{-1}$	Inverse of the affine transformation $\mathscr{A}$
$\mathscr{A}\{T,t\}$	Affine transformation with parameters T and t
$\mathscr{A}(x)$	Affine transformation of a coordinate vector x
$f:\mathbb{R}^n\to\mathbb{R}$	A function from $\mathbb{R}^n$ to $\mathbb{R}$
$L^p(\mathbb{R}^n)$	L^p space of functions $f:\mathbb{R}^n\to\mathbb{R}$
f	Function $f:\mathbb{R}^2\to\mathbb{R}$
$f\circ\mathscr{A}^{-1}(x)$	Affine transformation of the function f
f'	$f'(x)=f\circ\mathscr{A}^{-1}(x)$
χ_f	Characteristic function of f
$\|f\|_{L^p}$	L^p norm of function f
$\mu(f)$	Centroid of function f
$*$	Convolution operation
$\hat{f}$	Fourier transform of f, see Definition 5
$\tilde{f}$	Translation normalized version of f, see Definition 6
$\bar{g}$	Complex conjugate of g
$det(T)$	Determinant of the matrix T
$P(A)$	Probability of the event A
$\|K\|$	Area of a set K
v	Feature vector

Contents

Abstract

Preface 5

List of original articles 7

Abbreviations 9

Contents 9

1 Introduction 13

1.1 Background and motivation 13

1.2 The contribution of the thesis 16

1.3 Summary of original papers 17

2 Affine recognition and registration 21

2.1 The affine transformation of patterns 21

2.2 Recognition 24

2.2.1 Segmentation 24

2.2.2 Affine feature extraction 25

2.2.3 Classification 30

2.3 Registration 31

2.3.1 The area based approach 32

2.3.2 Feature based approach 32

3 Multiscale framework 35

3.1 Motivation 35

3.2 Definition 36

3.3 Related theorems in the spatial and Fourier domain 38

3.4 Multiscale affine invariants 39

3.4.1 Spatial multiscale affine invariants 39

3.4.2 Multiscale autoconvolution 40

3.4.3 Multiscale Fourier invariants 41

3.4.4 Multiscale ridgelet invariants 42

3.4.5 Generalized affine moment invariants 43

3.4.6 Multiscale autoconvolution histograms 43

3.4.7 Comparison based histograms 44

3.5 Multiscale affine registration 46
3.5.1 Spatial multiscale registration 46
3.5.2 Multiscale autoconvolution registration 47
3.6 Completeness 47
3.7 Experiments and implementational issues 48
3.7.1 Implementation 48
3.7.2 Experiments 50
4 Convexity measures **53**
4.1 Motivation and other convexity measures 53
4.2 Multiscale autoconvolution in convexity measurements 55
5 Conclusions **59**
References **60**
Original articles **64**

1 Introduction

1.1 Background and motivation

The word ”pattern” can refer to many different concepts. A pattern can be, for example, a spoken word, an image, written text, a human face, a motion sequence, a silhouette, the order of books on a shelf, or learned manners. In fact, it is quite difficult to define a general pattern in a rigorous and practical way. Watanabe (1985) defines a pattern as the opposite of a chaos; it is an entity, vaguely defined, that could be given a name. Even if this definition may not be very useful for applications, it nevertheless gives some idea what a pattern can be.

Humans deal with patterns in hundreds of situations every day. We are actually so used to recognizing, managing, and interpreting different patterns that most of the time we take it for granted. In fact, compared to machines, humans are by far the better pattern recognizers in most situations, and despite many years of research it is not known how humans do this (Jain *et al.* 2000). However, computer aided pattern processing is already mature enough to be applicable in limited settings. Table 1, taken from (Jain *et al.* 2000), gives some idea about where it has been used in different fields. The vast potential of general purpose pattern recognizers also encourages to continue the research in this field.

Constrained pattern processing problems usually arise in particular applications. Two applications, which are studied in this thesis, are recognition and registration. In recognition applications, we need to build systems that can associate a given pattern, say an image of an object, to one or more sample patterns. In registration applications, on the other hand, the requirement is to find the best possible alignment between two patterns. Tasks of this kind appear frequently in many applications, as already illustrated in Table 1. Registration is particularly important in medical imaging (Maintz & Viergever 1998) and in achieving the relative positions of images for 3D reconstruction (Hartley & Zisserman 2003).

Table 1. Examples of Pattern Recognition Applications.

Problem Domain	Application	Input Pattern	Pattern Classes
Bioinformatics	Sequence analysis	DNA/Protein sequence	Known types of genes/patterns
Data mining	Searching for meaningful patterns	Points in multidimensional space	Compact and well separated clusters
Document classification	Internet search	Text document	Semantic categories (e.g. business, sports, etc.)
Document image analysis	Reading machine for the blind	Document image	Alphanumeric characters, words
Industrial automation	Printed circuit board inspection	Intensity or range image	Defective/non-defective nature of product
Multimedia database retrieval	Internet search	Video clip	Video genres (e.g. action, dialogue, etc.)
Biometric recognition	Personal identification	Face, iris, fingerprint	Authorized users for access control
Remote sensing	Forecasting crop yield	Multispectral image	Land use categories, growth pattern of crops
Speech recognition	Telephone directory enquiry without operator assistance	Speech waveform	Spoken words

Pattern recognition problems can be approached from several different viewpoints. Four well-known paradigms are template matching, syntactic or structured matching, neural networks, and statistical classification (Jain *et al.* 2000). These models need not be independent, and there have also been attempts to design hybrid systems involving multiple models (Fu 1983).

In template matching, the idea is to test the pattern against a set of stored templates, while taking into account all possible pose changes. The quality of the different matches is evaluated with some measure, like correlation or mutual information. For example, let us have an image of a text and templates of all possible characters that can occur. Then to recognize the characters, we simply measure the correlation between the image and templates with, say, all possible translations and rotations. Those that have a correlation higher than some threshold are assumed to be correct matches. In general, the templates do not have to be the actual training patterns, as in this example, but can be learned from them. Especially if the number of possible poses is high, template matching becomes computationally inefficient. However, as computation power has increased, it has become feasible in some applications.

In the syntactic approach the patterns are interpreted as being composed of simple subpatterns, which again can be built from even simpler subpatterns (Fu 1982). The simplest possible subpatterns are called primitives and a given complex pattern is represented in terms of these. The natural analogy to this approach is the syntax of a language (Jain *et al.* 2000). The patterns could be seen as sentences and the primitives as the alphabet. The sentences would be then constructed according to grammatical rules. The syntactic approach is appealing, but its implementations can run into problems, especially due to difficulties in the segmentation of patterns to primitives (Jain *et al.* 2000).

Neural networks can be seen as a massive parallel computing system, with a large number of simple processing units with many interconnections. The main characteristics of neural networks are that they can learn complex and nonlinear relationships between the input data and the outputs (Jain *et al.* 2000). These relations are learned using training samples, and they are used to recognize the upcoming patterns. Neural networks can also be seen as efficient models for statistical pattern recognition (Bishop 2006), and they are actually closely connected to this approach.

The statistical approach to pattern recognition is maybe the one which is most frequently applied. In this approach, the patterns are first mapped to feature vectors and the recognition is performed according to them (Jain *et al.* 2000). The sample patterns are used to adjust the recognition method to a specific problem. There are numerous ways of performing both the mapping of patterns to feature vectors, referred to as feature extraction, and the recognition, usually referred to as classification (Bishop 2006). The statistical approach is the one used with the methods in this thesis. In particular, the presented methods provide new approaches for the feature extraction step in this procedure.

There are also several approaches (Zitova & Flusser 2003) for registration. The main two paradigms here are somehow similar to template matching and the statistical approach. From the description of template matching, it is easy to see that it actually performs a registration between the pattern and the template. Hence, simply by using one of the registered patterns as a template, we have immediately a registration method. The other main approach to registration uses features similar to those in statistical recognition. There, we extract some salient characteristics of the patterns and use only them to perform the alignment, which can be computationally simpler than direct matching. As in the registration, the practical algorithms are not limited to follow only one of these paradigms, but can often apply a combination of them.

Recognition and registration, as described above, have thus far proved to be too complex problems to allow for a general solution which works efficiently in real applications. Thus it is necessary to even further restrict these problems. In this thesis, we study a case where the transitions within one pattern class can be depicted, or at least well approximated, by a particular geometric mapping, namely the affine transformation. In other words, in recognition we are associating the new pattern to the training pattern that can be mapped close to it by some affine transformation. In registration, we estimate the transformation parameters given a pattern and an affine transformed version of it.

Since a pattern can refer to many different things, we usually have to select the particular patterns under discussion. The selection can naturally rise from the application under study. For example, if we are recognizing objects using images, examples of possible patterns could be the 2D image intensity function, the contour of the object, or the corners of the object in the image. The new methods presented in this thesis apply to patterns that can be presented by image intensity functions.

1.2 The contribution of the thesis

A common difficulty in affine invariant recognition and registration methods is to produce sufficiently many discriminating descriptors from the patterns. Such problems are also present in approaches that in theory would give an unlimited number of features. This is because usually only part of these features are stable or the implementation of further features is not feasible. The low number of descriptors can result in situations where the methods are not discriminative enough to register or recognize complex patterns.

This thesis introduces a novel multiscale framework for constructing affine invariant recognition and registration methods. The idea in the proposed framework is to extend the given pattern to a set of affine covariant versions, which carry partially independent information from the original pattern. Then the known approaches are applied to each of these separately, increasing the total number of descriptors compared to the original method. The key part of the framework is the construction of the affine covariant set, and this is done by combining several scaled representations of the patterns. The advantages are the possibility of many variations and the extension of the existing methods to produce a number of features capturing more information on the patterns.

The application of the multiscale approach is demonstrated in nine constructions where different preprocessing steps, combination schemes, and feature extraction methods are applied. The resulting new techniques are comprehensively analyzed, and their performance is illustrated in the experiments, including comparisons to other similar affine recognition and registration approaches. Also implementational issues and computational complexities are considered, making it easier for application developers to use the proposed methods.

In addition to the recognition and registration applications one multiscale method, the multiscale autoconvolution, can be used to measure convexity. Convexity is a descriptor for the shape of an object and it is an important feature in many applications, especially in medical imaging. The introduced approach and its properties are thoroughly analyzed and assessed in experiments, comparing them to the other convexity measures. The new measure is also shown to be directly applicable to gray scale images approximating binary objects and it can be easily modified to produce a number of measures.

The multiscale methods are presented from the perspective of recognition and registration from image intensity functions. However, the ideas in the framework are more general and may provide new approaches for other applications which consider different types of patterns as well.

1.3 Summary of original papers

This thesis consists of ten publications. Since the multiscale framework, as presented in this thesis, was only developed later, the papers presented are originally not directly derived from the framework, but are more inspired by each other.

Paper I presents a new fast and simple affine invariant feature extraction method based on pointwise products of scaled image functions. In the experiments the proposed descriptors were shown to outperform the descriptors based on affine invariant moment polynomials.

Paper II presents the convolution based affine invariant feature extraction method, with a detailed discussion about its known properties. One of the strongest advantages compared to other approaches is that the method does not require any translation normalization. The presented method is also straightforward to implement and reasonably fast to compute. The experiments performed indicate that the new features perform well in classification.

Paper III introduces the affine registration technique using similar pointwise products of scaled image functions to those in Paper I. The method was also extended to cover registration of point sets. The new technique proved to be fast to compute, due to its easy operations.

Paper IV presents another affine registration method, based on the ideas of multiscale autoconvolution introduced in Paper II. This registration method has similar advantages to the features in Paper II. The two most important ones are the ability to deal with the translation component directly and that the same efficient evaluation through Fourier transform is applicable here. This registration method is also extended to the registration of point sets, as in Paper III.

Paper V examines the properties of patch based approaches in the classification of visual object classes. The main focus was to observe how different properties of the segmented image patches affect the classification performance. The multiscale autoconvolution descriptors, presented in Paper II, were one of the methods for which the examination was performed.

Paper VI introduces a new combination of the well known affine invariant moment polynomials and multiscale approach. Using this method it is possible to increase the number of descriptors achieved with low order moments. The additional features clearly increase the performance compared to the moment polynomials and some other methods.

Paper VII proposes a new way of using ridgelets to produce affine invariants. The approach was inspired by the Fourier form of multiscale autoconvolution, but brings in a new analyzing function with greater variability.

Paper VIII extends the convolution based multiscale approach presented in Paper II, to produce affine invariant histograms. The main advantage is that with the same computational load as for traditional multiscale autoconvolution, this new method is able to capture more information from the patterns. This is clearly illustrated in the experiments.

Paper IX presents a new convexity measure based on the ideas in multiscale autoconvolution, and gives a comprehensive analysis of its properties. The new measure introduces two special features compared to previous approaches. It can be applied directly to gray scale images approximating binary objects, and it is easily modified to give a number of different measures. The experiments performed indicate that the proposed technique is applicable in shape analysis.

Paper X introduces comparison operations to multiscale methods and demonstrates them in two examples similar to Papers I and II. The construction of the new features also takes advantage of the binary coding presented originally in local binary patterns. The new methods based on nonlinear operations demonstrated high discriminability in the experiments, achieving the same recognition rates with only a fraction of the computation time compared to the traditional methods.

2 Affine recognition and registration

This chapter presents a brief introduction to affine pattern recognition and registration. The discussion begins with a definition of the affine transformation and a short motivation for the study of recognition and registration in the affine case. The main focus of this thesis is in the feature extraction, and hence this point of view is clearly emphasized in the following introduction.

2.1 The affine transformation of patterns

The affine transformation of a coordinate vector $x \in \mathbb{R}^2$ is a linear mapping defined by a matrix T composed with translation by a vector t.

Definition 1 *Define the affine transformation $\mathscr{A} = \mathscr{A}\{T,t\}$ by*

$$x' = \mathscr{A}(x) = Tx + t, \tag{1}$$

where $t,x \in \mathbb{R}^2$ and T is a 2×2 nonsingular matrix whose elements belong to $\mathbb{R}$. Any such transformation is invertible with inverse $\mathscr{A}^{-1}(x) = T^{-1}x - T^{-1}t$.

Before applying this transformation to the actual patterns, they must be defined more specifically.

Definition 2 *The patterns examined in this thesis will be real valued positive functions defined in $\mathbb{R}^2$. We will further assume that the functions belong to $L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$, have compact support, and are not identically zero as L^1 functions.*

A function f belongs to $L^p(\mathbb{R}^2)$ if the integral $\int_{\mathbb{R}^2} |f|^p$ exists and has a finite value. f is identically zero as an L^1 function if $\int_{\mathbb{R}^2} |f| \equiv 0$. The additional requirements in Definition 2 have been imposed to ensure that the methods introduced later are well defined. This is explained in more detail in Paper II.

A particular pattern fulfilling Definition 2 is the intensity function of a gray scale image. In fact, since images are the main interest of this thesis, the words pattern and image are used interchangeably in the following, although the methods could be applied to any pattern fulfilling Definition 2. Most of the discussed methods, particularly the multiscale framework, may be extended in a straightforward manner to cover higher dimensional functions $f : \mathbb{R}^n \to \mathbb{R}$. The case $n = 3$ is particularly useful in applications

(Rigoutsos 1998, Ronneberg *et al.* 2002, Schael & Siggelkow 2002). Some discussion of such applications can be found in the original papers.

We define the action of affine transformations on image functions as follows:

Definition 3 *Suppose $f(x)$ is a pattern as defined in Definition 2. We may apply an affine transformation $\mathscr{A}$ to f, which gives a new function f' in $\mathbb{R}^2$, where*

$$f'(x) = f \circ \mathscr{A}^{-1}(x) = f(T^{-1}x - T^{-1}t). \tag{2}$$

We call f' the $\mathscr{A}$ transformed version of f.

The affine transformation arises in this thesis in two ways. In affine recognition, the goal is to find a method for extracting a descriptor vector v from the pattern f in such a way that the resulting v is the same for f and f' for any affine transformation $\mathscr{A}$. In affine registration, the objective is to recover the parameters T and t using the given pattern pair f and f'. The methods introduced apply to any application where the affine transformation between the patterns f and f' may arise. However, to motivate the practical importance of this transformation, the most common practical situation where this is encountered is presented in the following.

Let us have a pinhole camera acquiring a picture of a planar object from different viewing angles, as illustrated in Figure 1. Following from the camera geometry, the resulting images are connected by a projective transformation, also called homography (Hartley & Zisserman 2003). Using this knowledge in recognition, we could define that all such patterns that are the same up to some homography belong to the same object class or category. In registration, on the other hand, we are looking for a particular projective transformation, which aligns the two given patterns most accurately.

However, the unrestricted projective transformation gives a great deal of freedom to deform the images, and it is unfortunately very difficult to limit the examination only to the physically relevant cases. In recognition, this means that many patterns that we would like to distinguish can actually be mapped close to each other by some projective transformation. For example, a square may be mapped to be arbitrarily close to a triangle by a homography, although we would usually like to distinguish between these two shapes. Also any closed contour can be mapped very close to a circle with a series of homograhies (Åström 1995). Another drawback rises from the fact that the projective transformation of image functions is a nonlinear operation, which can be hard to manage in the algorithms.

Fig 1. Two images of a planar object, taken from different viewing angles with a pinhole camera, are connected by a projective transformation, also called homography.

A common solution is to relax the requirement of the accurate camera model and approximate the projective transformation by something that is easier to handle. On many occasions, a proper approximation has been provided by the first order Taylor expansion of the projective transformation, which is exactly the affine transformation. Due to the properties of Taylor representation, this is accurate at least locally, but if the field of view and the change of viewing angle are relatively small, the approximation is accurate for the entire image (Hartley & Zisserman 2003).

The image can also be segmented by looking at regions where the requirement of the affine transformation holds. This is exactly what is done in affine covariant area detectors (Mikolajczyk *et al.* 2005, Mikolajczyk & Schmid 2004, Matas *et al.* 2002, Tuytelaars & Gool 2004, Kadir *et al.* 2004), as illustrated in Figure 2.

The affine transformation is not the only way to approximate the projective transformation. Depending on the situation, it may be beneficial to use further simplified models, like translation, rotation, or scaling. A number of different approaches for recognition and registration exist for these simpler models and in those cases where this approximation is accurate, they are likely to perform better than affine techniques. Some of the presented affine methods can also be modified to operate only with these simpler transformations, as explained for example in Paper II, but generally such modification are not easy. However, in many real applications the translation, rotation, and scaling are not sufficent for modeling the distortions, and the affine model is required.

Fig 2. Affine covariant neighborhoods segmented from two images using a Hessian affine detector (Mikolajczyk & Schmid 2004).

2.2 Recognition

This section considers affine recognition. The approach studied here is statistical, and it is well described for example by Jain *et al.* (2000). Briefly stated, this means the following: the studied image, or pattern, is represented as a d-dimensional feature vector, which defines a point in d-dimensional feature space. Such points, created from the given sample set, are then given to a classifier that defines boundaries in the feature space separating the object classes. Finally, the examined new image, or pattern, is again mapped to a point in this feature space and classified according to the defined boundaries. The affine recognition is a special case, where we assume that the patterns in a particular class are the same up to affine transformation.

The algorithms applying the statistical recognition scheme usually contain three main steps, namely segmentation, mapping to a feature vector, and classification. Since the main subject of this thesis is to construct mappings from patterns to feature vectors, the other two steps are only briefly introduced in the following.

2.2.1 Segmentation

A typical problem in pattern recognition is that one has large amounts of data, from which only part is relevant for the recognition at hand. A common way to approach this problem is to introduce a preprocessing step, where the essential information is separated from nonrelevant parts. This kind of action is referred to as segmentation. In the literature, one also uses the forms grouping, perceptual organization, or fitting (Forsyth & Ponce 2003). Segmentation is a wide ranging and very challenging research area, and Duda *et al.* (2001) even referred to it as one of the deepest problems in the whole pattern recognition.

Despite the difficulties, the segmentation is very important part of many applications, particularly in image recognition. For example, often one needs to recognize if particular sample objects are present in a given image. It would be difficult, if not impossible, to generate the feature vectors directly so that the recognition could be done only using them. Instead, one could first segment the image by detaching those parts from the scene where the objects of interest might be. Then the feature extraction would be applied only to the segmented parts.

All the affine feature extraction methods discussed in this thesis require segmentation to be done in one way or another. However, even in the case of images, this can involve many different operations and the choice again depends heavily on the application. Some recent and generally interesting approaches for image segmentation would include texture based methods (Chen & Kundu 1995, Panjwani & Healey 1995), mean shift (Comaniciu & Meer 2002), normalized cuts (Shi & Malik 2000), and affine covariant regions (Mikolajczyk *et al.* 2005). The last-mentioned approach is not usually referred to as segmentation, but it can be interpreted in this way also, since it segments the interest points and the affine neighborhoods around them. Plenty of discussion about image segmentation can be also found from (Forsyth & Ponce 2003, Umbaugh 2005, Petrou & Bosdogianni 1999).

2.2.2 Affine feature extraction

In feature extraction, the idea is to convert the studied pattern to a d-dimensional feature vector describing a point in d-dimensional feature space, which is usually $\mathbb{R}^d$. The affine feature extraction which is studied here is a special case of general feature extraction. There the idea is to construct a mapping from patterns f to vectors v, which would result in the same v for f and f' if, and only if, the two images are connected by an affine transformation. Such features are called complete. In reality, if v has finite length, no known feature extraction method satisfies the ”only if” part of this condition. In the applications, this is not necessarily a problem, if the functions which result in the same feature vector are reasonably close to being the same up to affine transformation.

What is reasonable here often depends on the application.

There are two main approaches to affine feature extraction, namely normalization and invariants. Those feature extraction methods applying the first approach have a distinguishable normalization part where the pattern is rectified to a canonical form before the actual mapping to the feature vector. The methods in the latter approach map the pattern directly to a feature vector without any normalization phase. Figure 3 illustrates the difference between these two approaches. Another possibility is to use a combination of normalization and invariants. In this kind of approach, one only partially normalizes the image and then applies a mapping that is insensitive to the deformations left unnormalized. It must be emphasized that in the following discussion, as well as in Chapter 3, it is assumed that the patterns f are segmented and the background is set to be zero.

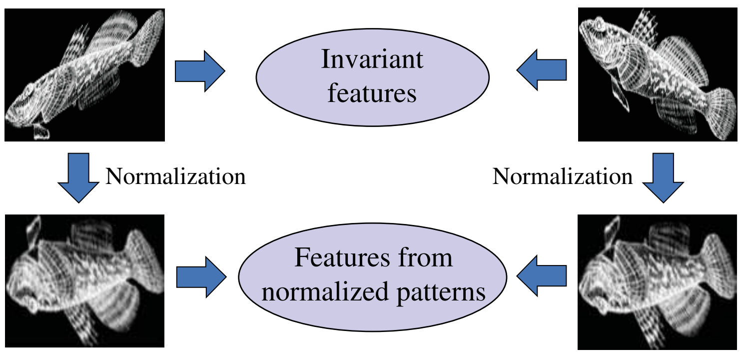


Fig 3. Illustration of the differences between the invariant and normalizing approaches to the affine recognition.

The normalization approach

The idea in the normalization approach is to rectify the image to a canonical representative before extracting the actual features. The advantage here is that the method producing the features after normalization need not have any invariance. Usually, it can be even better if they do not have much invariance in order to achieve strong discrimination. Some examples of such feature extraction methods, appearing frequently in applications, could be gradient orientation histograms (SIFT) (Lowe 2004) and the image moments (Gonzalez *et al.* 2002). Promising results have also been produced by the recently introduced CS-LBP (Heikkilä *et al.* 2006).

The drawback of this approach is, however, that robust affine normalization is not an easy task, and the amount of possible approaches is limited. Maybe the most important one is based on normalizing the image sequentially starting from the translation t and continuing to the linear part T. A common example of this procedure is the following three step algorithm:

1. Given a pattern f, normalize the translation by moving its centroid to the origin.
2. Normalize the linear part of the affine transformation, excluding final rotation, by transforming f so that its second moment matrix will become the identity matrix.
3. Normalize the final rotation by turning the pattern in such a way that it has some fixed dominant gradient orientation.

If the above procedure has an unique solution for f, then it should result in the same normalized form also for any $f \circ \mathscr{A}^{-1}$, with arbitrary $\mathscr{A}$. Problems are encountered, however, if no unique solution exist, which often happens for instance if f has rotational symmetries.

There are many variations of this procedure, where, for example, the translation normalization is based on some point other than the image centroid, the second moment matrix is computed from the image gradient magnitudes or substituted with the Hessian matrix, or the final rotation is normalized based on moments as presented by Suk & Flusser (2005). Suk & Flusser (2005) also deal with symmetric patterns.

Another possibility for applying normalization is to use it only partially, and then to apply invariant feature extraction to the result. One example, which is used in some moment based features and in some examples of the multiscale framework, is to normalize only the translation part t and leave out the rest of the transformation. Another example could be that we exclude the final step of the presented normalization procedure and apply rotation invariant features. For such features, and other types of invariants, there is a vast number of different approaches. For some examples one can refer to (Burkhardt & Siggelkow 2001, Rodrigues 2000, Reiss 1993).

The invariant approach

Feature extraction methods using the invariant approach do not have a distinguishable normalization part, but the features are created directly from the given pattern. The only exception in this thesis is that also methods that may normalize the translation are categorized as affine invariant feature extraction methods. The advantage compared to the normalizing approach is that one possible error prone step is completely omitted in the process.

Several affine invariant feature extraction methods are known so far. Maybe the most trivial ones are the normalized integral

$$If = \frac{1}{\|\chi_f\|_{L^1}} \int_{\mathbb{R}^2} f(x)dx \tag{3}$$

and the normalized histogram

$$Hf = \Big[\frac{1}{\|\chi_f\|_{L^1}} \int_{(f)^{-1}(B_j)} dx\Big]_{j=1}^{m}, \tag{4}$$

where $B_1,\ldots,B_m$ are disjoint subsets of $\mathbb{R}_+$ (bins), and $\chi_f(x) = 1$ if $f(x) \neq 0$, and $\chi_f(x) = 0$ otherwise. In the case of images, these two descriptors are more commonly known as the mean gray value, and the normalized gray value histogram. While they are very easy to evaluate, they have serious problems as descriptors. The first one of these, which applies to both (3) and (4), is that we completely ignore the spatial distribution of the gray values in the image. For example, if we discretize the integrals and apply them to digital images, we can permute the pixels into arbitrary order without changing the features. This results in a situation where the set of patterns having the same feature vector is very large. Another problem is the limited number of features. This is especially seen in the case of a normalized integral (3), which gives only one feature. It is easy to imagine that this would not be enough to classify complex patterns. Hadjidemetriou *et al.* (2004) present a modified form of image histograms which are able to include spatial information of the image. These constructions are, however, invariant only to rigid motion of f.

A less trivial solution is provided by the affine invariant moment polynomials, introduced in 1960's by Hu (1961, 1962), though later corrected by Flusser & Suk (1993) and Reiss (1991). These descriptors make use of the two-dimensional cross products allowing us to incorporate also spatial information to the features. A general formulation, from which all the affine invariant moment polynomials can be constructed, is given by Suk & Flusser (2004) as

$$MPf = \frac{1}{\|f\|_{L^1}^{w+N}} \int_{\mathbb{R}^{2N}} \prod_{1\leq k<l\leq N} C((x_k-\mu(f)),(x_l-\mu(f)))^{n_{kl}} \prod_{i=1}^{N} f(x_i)\,dx_i, \tag{5}$$

where $x_i, x_k, x_l \in \mathbb{R}^2$, $C(x_k, x_l) = \det[x_k x_l]$, $N \geq 2$ and $n_{kl} \geq 0$ are integers, $w = \sum_{k,l} n_{kl}$, and $\mu(f) \in \mathbb{R}^2$ is the centroid of f. Integration is over vectors $x_1, \ldots, x_N$. This formulation can be further simplified to the more familiar form as a polynomial of the image central moments

$$m_{pq}(f) = \int_{\mathbb{R}^2} (x_1 - \mu(f)_1)^p (x_2 - \mu(f)_2)^q f(x)\, dx, \tag{6}$$

where $x = (x_1, x_2)$, $\mu(f) = (\mu(f)_1, \mu(f)_2)$, and the value $p+q$ is called the order of the moment. Only some of these polynomials are independent, and thus worth using as descriptors (Suk & Flusser 2004). The resulting features are computationally very efficient and there are number of them. However, they do have a serious deficiency, since in order to produce more features one needs to use higher order moments, which soon results in poor performance in noisy conditions. The amount of useful features is hence limited.

Another moment based approach was introduced by Van Gool *et al.* (1996). There, in addition to the moments computed from image f, they use moments that are computed from the support χ_f of f. In this way, it is possible to construct features that are also insensitive to certain changes in the values of f, which can occur when photographing under varying illuminations. The drawback of these invariants seems to be their very strong sensitivity to the changes in the image. Also the number of features is rather small.

One new solution was presented by Yang & Cohen (1999). The key part of this approach was the introduction of the cross-weighted moment

$$\mu_{klpq}(f) = \int (x_1 - \mu(f)_1)^k (x_2 - \mu(f)_2)^l (y_1 - \mu(f)_1)^p (y_2 - \mu(f)_2)^q w(x,y) f(x) f(y)\, dx dy, \tag{7}$$

where x and y are as above, and $w(x,y)$ is a weighting function defined by Yang & Cohen (1999). Compared to the invariants constructed using traditional moments, this new technique seems to provide clearly more robustness. The unfortunate trade off is the increased computational cost, which very easily becomes impractically high.

Yet another approach to the affine invariants is proposed by Petrou & Kadyrov (2004). The method is called the trace transform and it was based on applying a particular combination of functionals to f. The features seem to perform well, but the implementation of the method is far from trivial.

In the methods described above, it was assumed that the patterns f have nonzero support size. In the case of images, this means that there are segments which have a nonzero area, and the images do not consist solely of points, line segments, conics, or curves. Another approach, however, to constructing affine invariants is to use exactly these components. In fact, line segments and conics can be even used to construct projective invariants (Mundy & Zisserman 1992). A number of methods using curves and points to produce affine invariants have been introduced in the literature. We will describe some of these in the following.

One of the first approaches presented by Burkhardt *et al.* (1991) and Arbter *et al.* (1990) is based on closed contours. These curves can be interpreted as periodic functions, and one may compute the corresponding Fourier coefficients. These can then be used to produce the affine invariants as described by Arbter *et al.* (1990). The main idea here is to take advantage of the fact that the affine mapping scales all areas with the same factor. The method they presented is, however, not translation invariant, and this must be normalized. The Fourier coefficients of the boundary curve were also used by Jin & Yan (1992) to construct affine invariants.

Other techniques for producing affine invariants using contours were introduced by Khalil & Bayoumi (2002, 2001). The construction in these methods was based on wavelet instead of Fourier transform of the boundary curve. The experiments performed in these papers indicate that wavelet transform provides some advantages, in particular increased robustness compared to the Fourier approaches.

Affine invariants can also be constructed using points. One possible construction is presented by Forsyth & Ponce (2003). In their method, the idea is to use three non-collinear points to form an affine coordinate frame and use the coordinates of the other points as features. The only problem here is that we would need to use the corresponding three points in all patterns to construct the affine coordinate frame. One solution to this is provided by geometric hashing (Forsyth & Ponce 2003), which is basically based on trying different combinations and voting for the best one.

2.2.3 Classification

The task of the classifier is to assign a category to the pattern according to the corresponding feature vector. Before this can be done, the classifier has to be trained to the specific problem using a given set of training patterns. These patterns can either have preassigned categories or not, in which case they must be also learned from the data. In the training phase, the classifier defines boundaries in the feature space separating the object classes. The simplest way to do this is based on the nearest neighbor method, where the boundaries are defined so that each point in feature space will be assigned to the same class as the nearest training sample. This is done according to some distance measure, which could be selected according to the application. Other commonly used methods to find the decision boundaries are based on support vector machines (Vapnik 1998) or different discriminant functions (Bishop 2006).

The classifier can hardly ever achieve perfect performance. Therefore sometimes the classifier can be also asked to determine only the probabilities for each possible category. The classification is considered in the context of the examples in the original papers, but for a comprehensive discussion one can refer to (Duda *et al.* 2001, Bishop 2006, Vapnik 1998).

2.3 Registration

In the affine recognition in Section 2.2, we were looking for something in the patterns that is independent of any possible affine transformation. In affine registration, we are especially interested in the transformation instead of the particular patterns. In a basic registration assignment, we are given two patterns f and f', which are known to be connected by some affine transformation $\mathscr{A}$, i.e $f' = f \circ \mathscr{A}^{-1}$. We are then asked to solve the parameters T and t of this transformation. In practice, the affine relation $f' = f \circ \mathscr{A}^{-1}$ does not necessarily hold exactly, but it is a good approximation of the mapping between the patterns.

One way of categorizing the affine registration approaches is to divide them to the area and feature based approaches (Zitova & Flusser 2003). In the area based methods, we make some initial guess for the transformation parameters and then apply a standard optimization method with some selected error criteria to refine them. In the feature based techniques, we first extract salient features, from which we can later solve the transformation parameters. The multiscale framework, presented in this thesis, is basically applicable to both of these approaches, but in the area based approach it is not likely to provide significant additional advantage. Hence this approach is only briefly introduced and the discussion concentrates more on the feature based methods.

2.3.1 The area based approach

The idea in area based registration is to obtain the transformation parameters directly, using standard optimization methods with a proper quality measure. This measure is usually computed by first mapping one pattern on the other, according to the current estimate, and then evaluating some function giving the alignment error. The choice of this function is maybe the most crucial step in the process and many different options have been proposed (Zitova & Flusser 2003). A simple example for such an error function is the sum of absolute pointwise differences between the aligned patterns. Another possible example is a measure based on the mutual information between the patterns (Viola & Wells 1997).

The actual optimization phase is done by applying some of the standard gradient decent methods, for example the Gauss-Newton or Levenberg-Marquardt algorithms. The best selection, of course, depends on the selected error measure and the type of the aligned patterns. However, regardless of this choice, the quality of the initial guess for transformation parameters has a strong effect on the performance and a poor initial guess can cause the algorithm to converge very far from the optimal solution. Thus, it can be even beneficial to use some feature based registration method to find the initial guess, and then to refine the result using the area based approach. One can find a more comprehensive introduction, with a number of references from (Zitova & Flusser 2003), which also includes approaches for nonaffine situations.

2.3.2 Feature based approach

In the feature based approach, the transformation parameters are not estimated directly, but use salient features extracted from the patterns. The extraction of these features is the most important step here, and many different ways of performing it have been introduced. The main difference in these methods is whether they find the features using some local information or the entire pattern at once. These two approaches are referred to as local and global feature extraction, respectively.

Many examples of local feature extraction methods are listed by Zitova & Flusser (2003), where the possible features range from interest regions to different kinds of lines and points. The common characteristic to all of these features is that they are extracted based on small local regions of the pattern, which means that they remain unaltered no matter what changes are made outside these neighborhoods. Another way of interpreting the local feature extraction is to see it as a segmentation of distinguishable local characteristics.

In the context of affine registration for images, those methods using the affine covariant neighborhoods (Mikolajczyk *et al.* 2005) or interest points (Mikolajczyk & Schmid 2004) are maybe the most frequently used. The registration using these techniques usually follows the three step algorithm:

1. Extract affine covariant neighborhoods or interest points from both images.
2. Find the correspondences between the extracted features.
3. Estimate the transformation parameters using the matched features.

The correspondence search in step 2 can be performed by applying the recognition techniques described in Section 2.2, where the input patterns are now the extracted affine covariant areas or the neighborhoods around the interest points. The estimation of the transformation in the third step can then be done, for example, by minimizing the reprojection error between corresponding features (Hartley & Zisserman 2003). This estimate is however not without problems, since there are usually false matches referred to as outliers in step 2. This may happen for example if the detected regions are small, or if the interest point matching is done using nonaffine covariant neighborhoods. The outliers can result in significant errors in the final estimate of the transformation parameters. One way to approach this problem is to apply the RANSAC algorithm (Fischler & Bolles 1981), which has proved to be efficient in applications. More discussion about RANSAC and its application to registration can be found for example in (Hartley & Zisserman 2003).

In general, the local approaches perform well, if the patterns have distinctive local characteristics which can be accurately extracted and reasonably well matched. These are normally fulfilled with standard photographs, but for example medical images usually lack such strong details (Zitova & Flusser 2003).

Another option is to extract the features using the entire pattern. These so called global features have the characteristic that they do change if the pattern is altered at any place. Another common character is that using these techniques the correspondence matching, step 2 in the registration algorithm, is achieved automatically. This happens since each feature corresponds to a certain parameter configuration in the feature extraction method, and hence we know that the features achieved with similar configurations in two patterns must be the corresponding ones. This gives the additional advantage that no outlier detection is required.

The simplest example of the global approach is to use the image centroid as a feature. This would only produce one point correspondence, which alone would be enough to determine the translation t. However, as will be described in Section 3.5, we can use the multiscale framework to produce a set of images from which to compute enough corresponding centroids to solve the entire affine transformation. Other methods in this category include the cross weighted moments (Yang & Cohen 1999), the affine moment descriptors (Heikkilä 2004), the trace transform (Kadyrov & Petrou 2006), and the affine invariant spectral signatures (Ben-Arie & Wang 1998). The affine invariant spectral signatures are not actually a purely feature based approach, but partially apply optimization to find some of the transformation parameters.

The global methods presented encode the affine transformation parameters in different ways, and the final estimate for the transformation parameters must be formulated accordingly. Often some type of least squares approach is used in these constructions.

3 Multiscale framework

This chapter defines the multiscale approach and discusses some related issues. Several invariants and registration methods based on the framework are also given. The methods are related to the original publications, and hence some mathematical formalities have been left out in order to make the main ideas easier to understand. The details can be found in the corresponding publications. The discussion begins with a short motivation in the following section.

3.1 Motivation

As already mentioned in Section 2.2.2, previously proposed affine invariants have some important disadvantages. First of all, the amount of useful descriptors is limited. An extreme example is the normalized integration, which produces only one feature. Another example could be the affine invariant moment polynomials (Flusser & Suk 1993), where in theory we can construct an infinite amount of features, but in reality only the few invariants which use low order moments are applicable in practice.

The second disadvantage was related to normalized integration and image histogram descriptors, which without any modifications do not include any information about the spatial distribution of gray values. The third problem is related to the computational and implementational difficulties. An example of these could be the trace transform (Petrou & Kadyrov 2004), which is able to produce a great number of features with spatial information, but the implementation and evaluation of these features is complicated and time consuming.

Similar problems are also associated with the affine registration methods. There, the low amount of descriptors can even result in a situation where the affine transformation cannot be fully recovered.

To overcome these problems we would ideally like to find a simple way to construct arbitrarily many descriptors which capture both intensity and the spatial characteristics of the patterns. While this can be difficult, it is possible to modify the previous methods so that they overcome these deficiencies.

3.2 Definition

The basic idea in the multiscale approach can be presented as a three step algorithm:

1. Given an image f, represent it in n different scales $f(\alpha_1 x), \ldots, f(\alpha_n x)$.
2. Combine the scaled images to a new image $Gf(x)$. The combination is required to be affine covariant, which means that for any affine transformation $\mathscr{A}$, one has

$$G(f \circ \mathscr{A}^{-1})(x) = (Gf)(\mathscr{A}^{-1}(x)). \tag{8}$$

3. Extract the affine invariant features, or perform affine registration, using f, Gf, and some of the methods presented in Chapter 2.

Figure 4 illustrates the overall procedure. The advantage of the approach is that by varying the scales α_i and combinations G, we are able to generate a great variety of different descriptors. This is possible regardless of the inherent number of features offered by the descriptor used in step 3. It is also significant that the values of Gf can already carry the spatial information so that this is not necessarily required from the operations in the last step.

The first step, scaling of images, is straightforward. The third step can also be quite simple, depending on which methods are selected, but if we take the mean gray values, gray value histograms, or image centroids, the implementation is not too difficult. The second step is often the most complicated one and is the key part of the whole approach. There, one needs to take the scaled images $f(\alpha_i x)$ and combine them to a new image $Gf(x)$ so that Gf and $G(f \circ \mathscr{A}^{-1})$ are related with an affine transformation $\mathscr{A}$. The reason for this requirement is quite obvious, since we need to maintain the affine relationship of the new images $G_i f$ in order to get the methods in step three to work.

The requirement for the affine covariance also provides the motivation for us using scaling instead of other transformations in the construction of Gf. This is because in general only scaling commutes with matrix products, and this makes it easier to find combinations for which this requirement is fulfilled. The translation component of $\mathscr{A}$ can, however, cause some problems. It may either be normalized away by computing the image centroid, or one may choose G more carefully, so that translation invariance is obtained without finding the centroid.

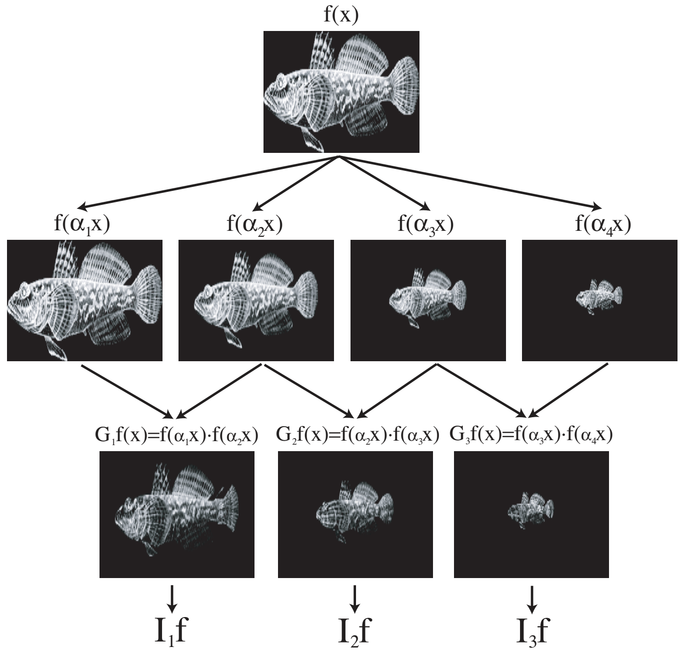


Fig 4. Example of a multiscale scheme where two scaled representations are combined together with a pointwise product. The final invariants are computed from the resulting images $G_i f$.

We consider also one possible extension to the basic multiscale framework. We preprocess the image by taking some transform $\mathscr{T}$ of it. The resulting new image $\mathscr{T}(f)$ is then considered as an input to the basic multiscale algorithm. Some extra caution must be paid here, since the relation between these new images is not likely to be the original affine transformation. If the new relation is again affine, then the invariant features can be computed in a straightforward manner. If we can find inverse mapping from the resulting relation to $\mathscr{A}$, also the registration can be performed. However, problems may arise because the resulting relation is not affine. In some cases, we also have to make exceptions to the affine covariance requirement in the second step. Examples of this extension are discussed in Sections 3.4.3 and 3.4.4.

3.3 Related theorems in the spatial and Fourier domain

This section lists a number of spatial and Fourier domain theorems and definitions that are used in the discussion of multiscale methods in this chapter, as well as in the original publications. All of them are well known and can be found in textbooks (Kreyszig 1998, Bracewell 2000), but they are included here for convenience. Throughout the section, the functions f and g are assumed to fulfill Definition 2.

Definition 4 *The L^1 norm of f is given by*

$$\|f\|_{L^1} = \int_{\mathbb{R}^2} |f(x)| dx. \tag{9}$$

The L^2 norm of f is given by

$$\|f\|_{L^2} = \left(\int_{\mathbb{R}^2} |f(x)|^2 dx \right)^{1/2}. \tag{10}$$

Lemma 1 *The L^1 norms of f and the affine transformed version $f \circ \mathscr{A}^{-1}$ are related by*

$$\|f \circ \mathscr{A}^{-1}\|_{L^1} = |\det(T)| \|f\|_{L^1}, \tag{11}$$

where $\det(T)$ denotes the determinant of the matrix T.

Definition 5 *The Fourier transform $\mathscr{F}$ of function f is given by*

$$\mathscr{F}(f) = \int_{\mathbb{R}^2} f(x) e^{-j2\pi x \cdot \xi} dx = \hat{f}(\xi) \tag{12}$$

with inverse $\mathscr{F}^{-1}$

$$\mathscr{F}^{-1}(\hat{f}) = \int_{\mathbb{R}^2} \hat{f}(\xi) e^{j2\pi x \cdot \xi} d\xi = f(x). \tag{13}$$

Remark 1 *If $f(x) \geq 0$ for all $x \in \mathbb{R}^2$, then $\|f\|_{L^1} = \hat{f}(0)$. This is particularly the case with the image intensity functions.*

Lemma 2 *The Fourier transforms of function f and the affine transformed version $f \circ \mathscr{A}^{-1}$ are related as*

$$\mathscr{F}(f(T^{-1}x - T^{-1}t)) = |det(T)| e^{-j2\pi t \cdot \xi} \hat{f}(T^t \xi). \tag{14}$$

Particularly if $\mathscr{A}$ is nonzero scaling, we have

$$\mathscr{F}(f(ax)) = \frac{1}{a^2} \hat{f}(\frac{\xi}{a}), \tag{15}$$

where the scaling factor $a \in \mathbb{R} \setminus \{0\}$.

Lemma 3 *The Fourier transform of the convolution* $(f * g)(x) = \int_{\mathbb{R}^2} f(y)g(x-y)dy$ *is related to the Fourier transforms of* f *and* g *as*

$$\mathscr{F}(f * g) = \hat{f}(\xi)\hat{g}(\xi). \tag{16}$$

Lemma 4 *The inner product of two functions* f *and* g *is related to the inner product of their Fourier transforms* $\hat{f}$ *and* $\hat{g}$ *according to Plancherel formula (see (Rudin 1987)) as*

$$\int_{\mathbb{R}^2} f(x)\bar{g}(x)dx = \int_{\mathbb{R}^2} \hat{f}(\xi)\bar{\hat{g}}(\xi)d\xi, \tag{17}$$

where $\bar{g}$ *denotes the complex conjugate of* g.

Definition 6 *The translation normalized version* $\tilde{f}$ *of function* f *is defined by*

$$\tilde{f}(x) = f(x + \mu(f)), \tag{18}$$

where the centroid $\mu(f) = \int_{\mathbb{R}^2} x f(x)dx$.

Lemma 5 *Let* $g(x) = f \circ \mathscr{A}^{-1} = f(T^{-1}x - T^{-1}t)$. *Now* $\tilde{f}$ *and* $\tilde{g}$ *are related as* $\tilde{g}(x) = \tilde{f}(T^{-1}x)$.

3.4 Multiscale affine invariants

In this section, we illustrate the multiscale framework by applying it to produce a set of affine invariant feature extraction methods. These techniques have originally been introduced in Papers I,II,VI,VII, and VIII. They are presented here very briefly and in a slightly different form, representing the viewpoint of the multiscale framework. Many details and further properties can be found in the original publications. We will begin with the simplest construction, and continue to more complex ones, presenting also several possible variations.

3.4.1 Spatial multiscale affine invariants

The first example is the Spatial multiscale affine invariants (SMA), originally presented in Paper I. Here, we choose Gf to be a product of the original f and two scaled representations of it, $f(\alpha x)$ and $f(\beta x)$. In this formulation, the translation component must, however, be normalized, and it is done by replacing f with the normalized version $\tilde{f}$ given by Definition 6. The operator G is given by

$$Gf(x) = \tilde{f}(x)\tilde{f}(\alpha x)\tilde{f}(\beta x), \tag{19}$$

where $\alpha, \beta \in \mathbb{R}$. For the construction of the final invariant, we choose the normalized integration. The resulting final SMA invariants are then given by

$$Sf(\alpha,\beta) = \frac{1}{\|f\|_{L^1}} \int_{\mathbb{R}^2} \tilde{f}(x)\tilde{f}(\alpha x)\tilde{f}(\beta x)\,dx. \tag{20}$$

Due to its simplicity, $Sf(\alpha,\beta)$ is very fast to evaluate, and the possibility of varying the scales results in an infinite number of different descriptors.

The choice to use exactly two scaled versions to form the invariant is somewhat arbitrary. However, the computational cost of one scaling is $O(n)$, where n is the number of points in the scaled image. Therefore by taking more scales the descriptors would get less efficient to evaluate. On the other hand, using only one scaling we would probably compromise the discriminability. Hence the choice of two scales can be considered as a trade-off between the discriminability and the efficiency.

3.4.2 Multiscale autoconvolution

Another example is multiscale autoconvolution (MSA), originally presented by Heikkilä (2002) and later more comprehensively discussed in Paper II. In this case, we use a combination of convolutions and products to form Gf. One advantage of this approach is the fact that the translation component does not have to be considered separately. Define

$$Gf(x) = \frac{1}{\|f\|_{L^1}^2} f(x)(f_\alpha * f_\beta * f_\gamma)(x), \tag{21}$$

where $\alpha, \beta \in \mathbb{R}$, $\gamma = 1-\alpha-\beta$, $f_a(x) = a^{-2}f(x/a)$, and $*$ denotes convolution. The third scaling factor γ is set to $1-\alpha-\beta$ in order to eliminate the translation component. The actual invariant features are again constructed by normalized integration, which gives the MSA transform

$$Mf(\alpha,\beta) = \frac{1}{\|f\|_{L^1}^3} \int_{\mathbb{R}^2} f(x)(f_\alpha * f_\beta * f_\gamma)(x)\,dx. \tag{22}$$

This formulation is not computationally appealing, but fortunately by applying the results in Section 3.3 Mf can be expressed using the Fourier transform $\hat{f}$ as

$$Mf(\alpha,\beta) = \frac{1}{\hat{f}(0)^3}\int_{\mathbb{R}^2} \hat{f}(-\xi)\hat{f}(\alpha\xi)\hat{f}(\beta\xi)\hat{f}(\gamma\xi)\,d\xi. \tag{23}$$

3.4.3 Multiscale Fourier invariants

As mentioned in Section 3.2, it is possible to extend the basic multiscale approach by making a transform $\mathscr{T}$ to the image function f before application of the multiscale framework. We now give one example of this, where we select $\mathscr{T}$ to be the Fourier transform $\mathscr{F}$. Recall from Section 3.3 that the Fourier transformed versions of f and $f \circ \mathscr{A}^{-1}$ are connected as

$$\mathscr{F}(f\circ\mathscr{A}^{-1}) = \mathscr{F}(f(T^{-1}x - T^{-1}t)) = |det(T)|e^{-j2\pi t\cdot\xi}\hat{f}(T^t\xi). \tag{24}$$

From (24) one can see that if $t \neq 0$ and $det(T) \neq 1$, the resulting relationship is not affine and the application of the standard multiscale framework would fail.

In order to make this approach work, we need to do something about these problems. The issue with the $det(T)$ is easily fixed by defining a normalized Fourier transform as

$$\hat{f}_n(\xi) = \frac{1}{\hat{f}(0)}\hat{f}(\xi). \tag{25}$$

The problem with the translation is, however, slightly more complicated. We could of course normalize the translation by computing the image centroid, but a better choice is to choose the combination G so that this effect is eliminated. This is done by defining,

$$Gf(\xi) = \hat{f}_n(\alpha_1\xi)\hat{f}_n(\alpha_2\xi)\ldots\hat{f}_n(\alpha_N\xi), \tag{26}$$

where we require that $\alpha_1 + \alpha_2 + \ldots + \alpha_N = 0$. This additional requirement makes the translation parts cancel, and we see that the overall relation becomes $Gf'(\xi) = Gf(T^t\xi)$, which is affine. In particular, if we select a normalized integration to produce the final invariant and $\alpha_1 = -1$, $\alpha_2 = \alpha$, $\alpha_3 = \beta$, and $\alpha_4 = 1-\alpha-\beta$, we have

$$Mf(\alpha,\beta) = \frac{1}{\hat{f}(0)^3}\int_{\mathbb{R}^2} \hat{f}(-\xi)\hat{f}(\alpha\xi)\hat{f}(\beta\xi)\hat{f}(\gamma\xi)\,d\xi, \tag{27}$$

which is exactly the Fourier form of the MSA transform. Note that here the integral is normalized by multiplying with $\|f\|_{L^1}$. The affine invariance of (27) is easy to show using the results in Section 3.3.

3.4.4 Multiscale ridgelet invariants

In Section 3.4.3 we considered the utilization of the Fourier transform before applying the multiscale framework. This is not the only possibility, and another prominent choice would the ridgelet transform considered in Paper VII. We define the ridgelet transform of a function f as

$$Rf(\xi) = \int_{\mathbb{R}^2} f(x)\psi(x \cdot \xi)dx, \tag{28}$$

where $\xi \in \mathbb{R}^2$ and ψ is a wavelet. The function $\psi(x \cdot \xi)$ looks like a ridge, and such functions are called ridgelets in the literature (Candés 1998, Candés & Donoho 1999, Do & Vetterli 2003). An example of $\psi(x \cdot \xi)$ is illustrated in Figure 5. $Rf(\xi)$ behaves well under the affine transformation except for the translation part. This must be normalized and we replace f with the normalized version $\tilde{f}$ given by Definition 6. Further by denoting $f' = f \circ \mathscr{A}^{-1}$ we have the relation

$$R\tilde{f}'(\xi) = |\det(T)|R\tilde{f}(T^t\xi). \tag{29}$$

Now define,

$$Gf(\xi) = \frac{1}{\|f\|_{L^1}^3} R\tilde{f}(\xi)R\tilde{f}(\alpha\xi)R\tilde{f}(\beta\xi), \tag{30}$$

where $\alpha, \beta \in \mathbb{R}$. Then, by applying normalized integration to produce the final invariant, we have

$$RIf(\alpha,\beta) = \frac{1}{\|f\|_{L^1}^2} \int_{\mathbb{R}^2} R\tilde{f}(\xi)R\tilde{f}(\alpha\xi)R\tilde{f}(\beta\xi)d\xi. \tag{31}$$

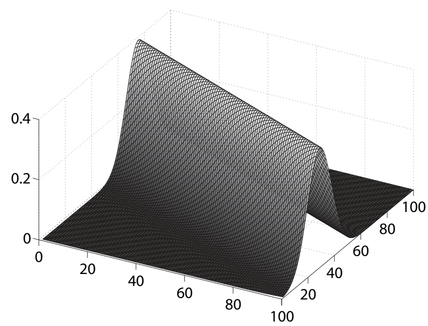


Fig 5. An example of a ridge function, using Gaussian wavelet.

3.4.5 Generalized affine moment invariants

In the previous examples, we used normalized integration, which is perhaps the simplest possible invariant at the third step of the multiscale approach. Instead, we could use something different, such as affine invariant moment polynomials. This is exactly what is done in the generalized affine moment invariants, presented in Paper VI. Now let G be one of the operators given in previous sections and define

$$If = \frac{1}{\|f\|_{L^1}^{w+N}} \int_{\mathbb{R}^{2N}} \prod_{1 \le k < l \le N} C((x_k - \mu(f)), (x_l - \mu(f)))^{n_{kl}} \prod_{i=1}^{N} Gf(x_i)\, dx_i, \tag{32}$$

where $x_i, x_k, x_l \in \mathbb{R}^2$, $C(x_k, x_l) = \det[x_k x_l]$, $N \ge 2$ and $n_{kl} \ge 0$ are integers, $w = \sum_{k,l} n_{kl}$, and $\mu(f) \in \mathbb{R}^2$ is the centroid of f. Integration is over vectors $x_1, \ldots, x_N$. With particular choices of N and n_{kl}, If will give all the possible affine invariant moment polynomials of Gf (Suk & Flusser 2004). These can then be expressed as polynomials of central moments of Gf, given by

$$m_{pq}(G, f) = \int_{\mathbb{R}^2} (x_1 - \mu(f)_1)^p (x_2 - \mu(f)_2)^q Gf(x)\, dx, \tag{33}$$

where $x = (x_1, x_2)$, and $\mu(f) = (\mu(f)_1, \mu(f)_2)$.

As an example, the first two affine invariants would be

$$\begin{aligned} I_1 &= (m_{20}m_{02} - m_{11}^2)/\|f\|_{L^1}^4, \\ I_2 &= (-m_{30}^2 m_{03}^2 + 6m_{30}m_{21}m_{12}m_{03} - 4m_{30}m_{12}^3 \\ &\quad -4m_{21}^3 m_{03} + 3m_{21}^2 m_{12}^2)/\|f\|_{L^1}^{10}. \end{aligned} \tag{34}$$

3.4.6 Multiscale autoconvolution histograms

In addition to normalized integration and moment polynomials, one may apply a histogramming operation to Gf in order to produce affine invariants. Formally, this is given by

$$If = \left[\frac{1}{\|\chi_{Gf}\|_{L^1}} \int_{(Gf)^{-1}(B_j)} dx \right]_{j=1}^{m}, \tag{35}$$

where $B_1, \ldots, B_m$ are disjoint subsets of $\mathbb{R}_+$ (bins), and $\chi_{Gf}(x) = 1$ if $Gf(x) \neq 0$ and $\chi_{Gf}(x) = 0$ otherwise. These operations have natural discrete analogues. A particular example of this histogramming approach is presented in Paper VIII, where

$$Gf(x) = \frac{1}{\|f\|_{L^1}^2}(f_\alpha * f_\beta * f_\gamma)(x). \tag{36}$$

In that paper, the histogram bins are, however, chosen according to the original function f rather than Gf as in (35).

3.4.7 Comparison based histograms

So far, all the examples of multiscale invariants presented construct the combination G, using convolutions or pointwise products. These linear operations can behave robustly in many settings, but they can also compromise the discriminability of the methods to some extent. For this reason, it might be better to utilize nonlinear functionals in the combination of the scaled images.

One example of such an approach, presented in Paper X, is to use pointwise comparison operations. The motivation for using this method comes from the local binary patterns (LBP) (Ojala *et al.* 2002), where similar operations were used to construct highly discriminative texture descriptors. Paper X illustrates this idea in two different cases, based on a formulation similar to SMA and MSA.

In the first of these constructions, define

$$Gf(x) = G_\alpha f(x) = X(\tilde{f}(x), \tilde{f}(\alpha x)), \tag{37}$$

where $\alpha \in \mathbb{R}$, and

$$X(a,b) = \begin{cases} 1 & \text{if } a > b, \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

The G in (37) corresponds to the one defined for SMA in Section 3.4.1, with some modifications. In (37) the pointwise product is replaced with the comparison operation X, and the affine covariant combination is computed using only one scaled version of f.

The second formulation, presented in Paper X, for G is based on convolution

$$C_\alpha f(x) = \frac{1}{\|f\|_{L^1}}(f_\alpha * f_{1-\alpha})(x), \tag{39}$$

where $\alpha \in \mathbb{R} \setminus \{0\}$ and $f_\alpha(x) = \alpha^{-2} f(x/\alpha)$. Using (39) define

$$Gf(x) = G_\alpha f(x) = X(f(x), C_\alpha f(x)). \tag{40}$$

This formulation is clearly related to MSA, with similar modifications as in the case of SMA and G defined in (37).

In the new formulations introduced, the resulting G is a binary image, from which we would have to compute the invariant. This could be done following the same ideas as in the previous case, but since we have such a specific form we can do something better. Paper X proposes to compute $Gf(x)$ with a few different α values, say $\alpha_1, \alpha_2, \ldots, \alpha_n$, and then to use a binary code construction as in LBP to combine all the information to a new function

$$Bf(x) = G_{\alpha_1} f(x) + G_{\alpha_2} f(x) \cdot 2 + \ldots + G_{\alpha_n} f(x) \cdot 2^{n-1}. \tag{41}$$

The function Bf has integer values from 0 to $2^n - 1$, and it completely encodes the information in the functions Gf. To compute the final affine invariant, one could just evaluate the normalized integral of Bf. This approach, however, does not make sense, since the different combinations would have an uneven impact on the resulting invariant value. Instead, we construct a histogram $HSf(k)$, with $2^n - 2$ bins, from nonzero values of Bf, and we normalize the histogram so that the sum over all bins is equal to one. Compared to the direct integration of $G_\alpha f$, this makes it possible to preserve the relative spatial arrangements in the functions $G_{\alpha_1} f, \ldots, G_{\alpha_n} f$. The construction of the invariant is illustrated in Figure 6.

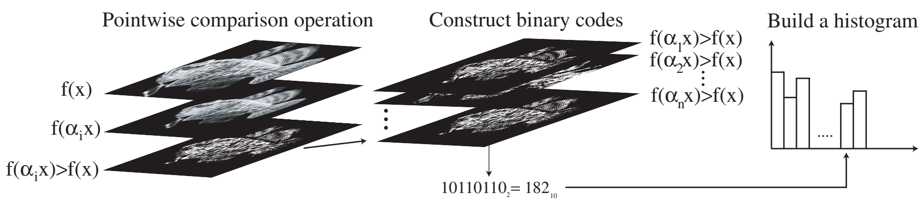


Fig 6. Illustration of the process for generating the invariant histograms.

These two examples illustrate the application of comparison operations in the construction of affine invariants. The approach is however not restricted to these examples, and may be applied to other constructions as well.

3.5 Multiscale affine registration

In the previous section, we considered the construction of affine invariant descriptors using a multiscale framework. Here we concentrate on using similar ideas in the affine registration between two images f and $f' = f \circ \mathscr{A}^{-1}$. We approach this through two examples originally given in Papers III and IV. These methods provide easily implementable and computable solutions for the registration task. According to the experiments, the accuracy of these methods is reasonable, and if more precision is required, these solutions serve as a good initial estimate for the complex iterative registration approaches. Note that as in feature extraction, in the following discussion it is assumed that the registered patterns are segmented and the background is set to zero.

3.5.1 Spatial multiscale registration

In general, one would need to fix six constraining equations in order to be able to solve all six degrees of freedom in the affine transformation. This can be done for example by locating three noncollinear corresponding coordinate points. In the global approach, such points should be produced using the entire image function instead of some local properties. The image centroid $\mu(f)$ could easily work as one of these, since it transforms exactly according to $\mathscr{A}$. However, at least two other similar points have to be located for the complete affine registration. Applying the multiscale approach, we can produce multiple images Gf and Gf' that are connected with the same $\mathscr{A}$ as the original images f and f'. Now by computing the image centroids $\mu(Gf)$ from all these, we can construct enough corresponding points to solve the transformation $\mathscr{A}$.

Paper III presents the kind of method that is based on using Gf defined for SMA in Section 3.4.1. Writing this out explicitly gives

$$\mu(Gf) = \mu(G_{\alpha,\beta}f) = \frac{1}{\|f\|_{L^1}} \int_{\mathbb{R}^2} x\tilde{f}(x)\tilde{f}(\alpha x)\tilde{f}(\beta x)\,dx, \tag{42}$$

where $\tilde{f}$ is the translation normalized version of f, as defined in Section 3.3. Clearly, for any $\alpha, \beta \in \mathbb{R}$ it holds that $\mu(Gf') = T\mu(Gf)$, where T is the linear part of $\mathscr{A}$. Now, by computing $\mu(Gf)$ and $\mu(Gf')$ for at least two different α, β pairs, we have enough information to solve T. In the case where we have more than two α, β pairs, the estimate for T can be obtained as a least-squares solution

$$\min_T \sum_i^n \|\mu(G_{\alpha_i,\beta_i}f') - T\mu(G_{\alpha_i,\beta_i}f)\|^2 \tag{43}$$

(Hartley & Zisserman 2003). The translation part t is then estimated as $t = \mu(f') - T\mu(f)$.

3.5.2 Multiscale autoconvolution registration

In the previous section, we formulated the affine registration based on the Gf defined for SMA in Section 3.4.1. The drawback there was that the translation component had to be considered separately using only one computed correspondence, namely $(\mu(f), \mu(f'))$. Now, instead we can select Gf to be the same convolution based combination as for MSA in Section 3.4.2. This approach is considered in Paper IV, where it is also shown that for this selection of Gf we have the relation

$$\mu(Gf') = T\mu(Gf) + tMf, \tag{44}$$

where T and t are linear and translation parts of $\mathscr{A}$ respectively, and Mf denotes MSA as defined in (22). One can observe that in (44) the translation is now incorporated in all pairs $(\mu(Gf), \mu(Gf'))$. This enables us to make a least squares estimate for both T and t directly from four or more corresponding $(\mu(Gf), \mu(Gf'))$ pairs. This, however, involves also computation of Mf with the same parameters α, β, but, as shown in Paper IV, this is achieved with only a small increase in the computation load. In addition, as in MSA the computation of the $\mu(Gf)$ can be performed efficiently in Fourier space.

3.6 Completeness

Recall from Section 2.2.2 that a system of features is called complete if the feature vectors extracted from patterns f and f' are the same exactly when the two images are connected by an affine transformation. Ideally one would like to use feature sets that are complete to ensure that they would distinguish between any patterns up to affine transformation. One example of a complete system is the geometric moments computed from an affinely normalized image (Shen & Ip 1997). Also affine invariant moments (Suk & Flusser 2004) are likely to form a complete system. However, as already mentioned, if the feature vector v has a finite length, no known feature extraction method is complete.

In applications, incompleteness is not necessarily a problem if the functions which result in the same feature vector are reasonably close to being the same up to affine transformation. It must also be emphasized that even if we have a complete feature set, it does not guarantee good performance. This is because the completeness does require features to be stable or tolerant to nonaffine distortions. Neither does it require that two feature vectors that are close to each other in the feature space are resulting from patterns similar to each other.

The multiscale framework by itself does not cause completeness problems, but the methods derived from it may be incomplete. SMA presented in Section 3.4.1, and the SMA type comparison invariants presented in Section 3.4.7 have known incompleteness related to starlike objects and reflections of image sectors with respect to the centroid. We refer to Section 4 in Paper I for more details. Similar issues arise with MSA if one performs the aforementioned operations in Fourier space. This however almost always results in noncompactly supported images, which are not interesting in practice. The generalized moment based invariants, presented in Section 3.4.5, are likely to have similar completeness to the affine moment invariants. The completeness of the other presented multiscale methods is not known in detail. It can be difficult to analyze what kind of incompleteness is present in a particular method, and what kind of effect this has in applications. This knowledge would, however, be valuable for implementations.

3.7 Experiments and implementational issues

This section discusses some general issues in the implementation of the multiscale methods, and the principles of the experiments performed in the original papers. A more detailed description of the experiments can be found in the corresponding papers, as well as detailed discussions of implementational issues of particular method. The idea here is only to give some guidelines.

3.7.1 Implementation

In practice, the patterns examined are not available as continuous functions, and we only have some amount of discrete samples. This is particularly the case with the gray-scale digital image functions, which can be interpreted as $N_1 \times N_2$ matrices consisting of samples from the real pattern. In such circumstances, the operations presented in Sections 3.4 and 3.5 must be discretized for the implementation. In most methods, this involves the evaluation of the integrals and performing the scaling.

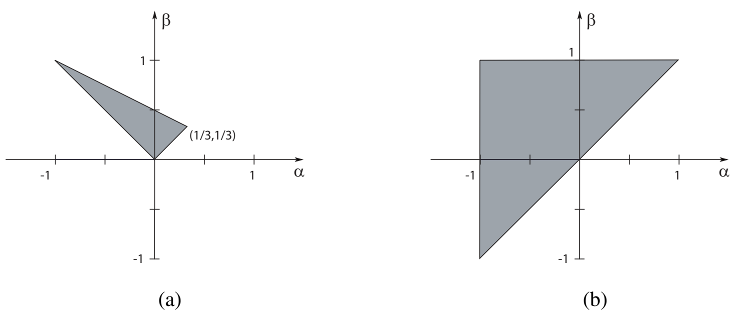
There are several approaches for carrying out these approximations, and it may be difficult to name the optimal ones. However, in the experiments we did not find much difference between the various approaches. Hence, we chose to use simple methods for both of these operations. The integrals were approximated as the sum

$$\int_{\mathbb{R}^2} f(x)dx \approx \frac{1}{N_1N_2}\sum_{i=0}^{N_1N_2-1} f(w_i), \tag{45}$$

where w_i are $N_1 \times N_2$ points in a rectangular grid in $\mathbb{R}^2$. The scaling of images was performed by using standard linear interpolation. With MSA, however, the interpolation was done by adding zeros between the known samples and the decimation by dividing the image into equal sized regions, and then summing the elements in them to form one element in the decimated image. This different method was chosen due to the probabilistic interpretation of MSA.

In addition to integration and scaling, we needed to evaluate also Fourier and Ridgelet transforms. The first one of these was easily achieved using the standard discrete Fourier transform. The second one, the Ridgelet transform, was more complicated and the standard techniques were found unusable, since they evaluate the transform in an unsuitable grid. The implementation that we created is introduced in Paper VII.

When the working implementation has been achieved we face another problem, which is the parameter selection. In the methods presented in Sections 3.4 and 3.5, the scaling parameters were allowed to be any real numbers. Consequently, there is an infinite amount of different combinations to select from. With all of these methods, however, there exist some symmetries which restrict the useful parameter space. For example, in the case of MSA and SMA, such areas would be the triangles illustrated in Figure 7.



(a) (b)

Fig 7. Regions in the (α, β) plane that will give all the (a) MSA or (b) SMA invariant values.

In spite of having these limited areas from which to choose the parameters, we still have infinitely many possibilities. Ideally, we would like to take those samples that would give us the best possible discrimination. However, we do not yet have a way of performing this. Hence, in the experiments the parameter values were picked either randomly, or from a uniform grid. For a particular application, it is also possible to take a large number of candidate parameters, and then select the best ones using sample data.

3.7.2 Experiments

Several types of experiments involving the new multiscale methods have been performed in the original papers. In the case of affine invariants, most of these can be divided into two categories.

The experiments in the first category examined the effect of certain nonaffine distortions on the performance. The setting was such that we had a group of training images which were then modified to form the test set by randomly affine transforming them and adding particular distortion. Features were computed from all images and then a classifier, typically SVM or nearest neighbor, was used to determine the correspondences. The performances with different distortions and distortion strengths were evaluated and compared to other multiscale or affine invariant approaches.

In the second class of experiments, the training and test sets were both taken in real photographing conditions and separately from each other. The feature extraction and classification was then performed in a similar manner to above. In these settings, it was possible to examine the combined effect from several real distortion types, originating from the scene, lenses, sensors, etc.

The experiments demonstrated that the multiscale approach clearly increases the performance compared to standard approaches. The MSA and the Ridgelet descriptors seemed to be the most accurate, but also computationally the most demanding. The histogram methods provided high discriminability with low computational costs, and they are recommended for situations with low resources. The challenge here is, however, to incorporate many histograms with different parameters in an efficient way.

Paper V illustrates yet another type of experiment. There the invariant features were computed from small local regions in the images. This basically differs from the previous experiments by the fact that the areas were much smaller. The results of these tests illustrated that MSA is applicable to these situations, but compared to performances in the other experiments we believe that MSA is likely to perform better when applied to larger segments. This is our impression also in the case of other multiscale methods.

The experiments in Papers III and IV, involving multiscale affine registration, are very similar to the first type of those involving invariants. There, instead of recognition, we performed registration of the images and evaluated the accuracy using some error measure, like pointwise absolute differences. The registration experiments also included a special case where the registered patterns were point sets.

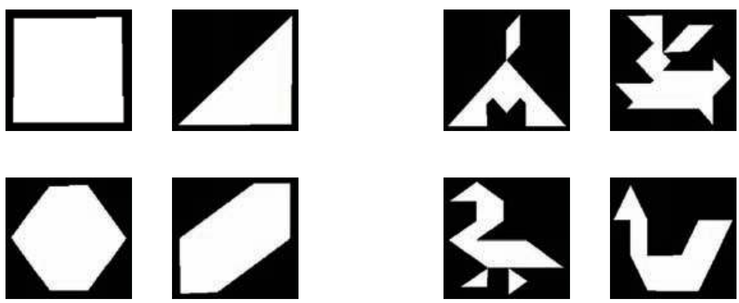
The registration methods provide a simple way of getting an estimate for the transformation parameters. In the cases where the accuracy of the methods is not sufficient, they give a good initial guess for the iterative algorithms.

4 Convexity measures

Besides the presented recognition and registration assignments, the multiscale approach may also be used in other applications. One specific example of these, described in Paper IX, is to use the multiscale autoconvolution for measuring object convexity. This chapter presents a short introduction to this method and to convexity measures in general. The presentation is slightly different from that in Paper IX in order to emphasize the connection to the MSA transform. Some mathematical formalities have also been left out. They can be found in Paper IX, where also a collection of different comparison experiments using the new measures are presented.

4.1 Motivation and other convexity measures

Shape is an effective descriptor in many applications (Lee *et al.* 2003, Costa & Cesar Junior 2001, Hyde 1997). One approach to measuring shape is to study the convexity of the object. Recall that an object is said to be convex if the line segment between two points in the object also belongs to the same object (Valentine 1964). Figure 8 illustrates some examples of both convex and nonconvex objects.



(a) (b)

Fig 8. (a) Samples of convex objects. (Note the slight deviation from convexity due to discrete sampling) (b) Samples of nonconvex objects.

In shape description, we are not so interested in classifying the objects as convex and nonconvex, but more in measuring how close to convex a particular object is. In other words, we consider a convex object to be somehow regular, and we are interested in how irregular the inspected objects are in this sense. One instance, where this idea is particularly suitable is in medical image analysis, where such irregularity can be used in classification of skin lesions (Lee *et al.* 2003, Rosin & Mumford 2006).

Before considering any convexity measures, it is reasonable to fix some properties that are required from a valid measure in order to make it behave well and produce sensible results. A presentation of such properties is given by Zunic & Rosin (2004), who list the following four.

Definition 7 *A convexity measure should have the following basic properties:*

1. *The convexity measure is a number from* $(0,1]$.
2. *The convexity measure of a given shape is equal to* 1 *if and only if the shape is convex.*
3. *There are shapes whose convexity measure is arbitrarily close to* 0*, implying that there is no gap between* 0 *and the minimal possible value of the measure.*
4. *The convexity measure is invariant under appropriate geometric transformations of the shape.*

The first approach to constructing a measure arises directly from the definition of convexity.

Definition 8 *Let $K \subseteq \mathbb{R}^2$ be a compact set representing a planar shape and let X and Y be independent random variables drawn uniformly from the set K. Define a convexity measure*

$$C_{ls}(K) = P([X,Y] \subseteq K), \tag{46}$$

where $P(A)$ denotes the probability of the event A and $[x,y]$ is a line segment between x and y.

This measure provides a direct way of estimating the convexity. However, the difficulties in computing the value of C_{ls} basically forbid the application of this measure in practice (Zunic & Rosin 2004).

Another approach is based on the idea of approximating an object by larger or smaller convex sets. This leads to two new measures.

Definition 9 *Let K be as in Definition 8. Define two convexity measures*

$$C_{ch}(K) = \frac{|K|}{|\mathrm{ch}(K)|}, \tag{47}$$

$$C_{mcs}(K) = \frac{|\mathrm{mcs}(K)|}{|K|}, \tag{48}$$

where $\mathrm{ch}(K)$ *is the convex hull of K,* $|K|$ *is the area of a set K, and* $|\mathrm{mcs}(K)|$ *is the supremum of the areas of convex subsets of K.*

Of these two measures, C_{ch} is the one most frequently appearing in the literature and in applications. It has proved to be robust and easy to compute. C_{mcs}, introduced by Rosin & Mumford (2006), has good performance in applications, but the challenges in the computation of $|\mathrm{mcs}(K)|$ may restrict the use of this measure in practice.

All of the measures C_{ls}, C_{ch}, and C_{mcs} can be thought of as being area based, since their values depend on the areas of different parts of the object. Consequently, such measures are usually tolerant with respect to small defects in the objects, which can be caused, for example, by noise or insufficient segmentation. In some applications, such robustness is not a desired attribute, and in such cases it is better to use so called boundary based measures. Two such measures, for polygons, can be defined as follows (Zunic & Rosin 2004).

Definition 10 *For a given planar shape K, where* $K \subseteq \mathbb{R}^2$ *is a compact and connected polygon, define*

$$C_{chp}(K) = \frac{\mathrm{Per}_2(\mathrm{ch(K)})}{\mathrm{Per}_2(\mathrm{K})}, \tag{49}$$

where $\mathrm{Per}_2(\mathrm{K})$ *is the* L_2 *perimeter of K and* $\mathrm{ch}(K)$ *is the convex hull of K. Further, let* $R(K,\alpha)$ *denote the minimal rectangle with edges parallel to the coordinate axes which contains the polygon K rotated by angle* α*. Define*

$$C_{poly}(K) = \min_{\alpha \in [0,2\pi]} \frac{\mathrm{Per}_2(\mathrm{R(K,\alpha)})}{\mathrm{Per}_1(\mathrm{K})}, \tag{50}$$

where $\mathrm{Per}_1(\mathrm{K})$ *is the* L_1 *perimeter, for definition see (Zunic & Rosin 2004), of K.*

4.2 Multiscale autoconvolution in convexity measurements

The idea that the multiscale autoconvolution can be used to measure convexity was first introduced by Rahtu *et al.* (2004). The approach there started directly from MSA, showing that with certain parameters the transformation acts as a convexity measure. Later, Paper IX presented a more comprehensive study and generalization of this type of convexity measure. The construction there was different, starting from the viewpoint of convex combinations, and showed the connection to MSA through the explicit expressions of the measure. The introduction here follows a similar idea to that in Paper IX, but presents it in a slightly modified form in order to illustrate the interconnectedness with multiscale autoconvolution more directly. Also for possible generalizations of the measure, one can refer to Paper IX.

The idea in the new measure follows directly from the alternative definition to convexity approaching it though convex combinations.

Definition 11 *A set $A \subseteq \mathbb{R}^2$ is called convex if and only if for any points $x_1, \ldots, x_n \in A$, any convex combination $a_1x_1 + \ldots + a_nx_n$, where $a_j \geq 0$ and $a_1 + \ldots + a_n = 1$, is in A.*

Definition 12 *Let $\alpha, \beta, \gamma > 0$, $\alpha + \beta + \gamma = 1$, and let K be the set representing the object under study. Define a convexity measure*

$$C_{\alpha\beta}(K) = P(\alpha X + \beta Y + \gamma Z \in K), \tag{51}$$

where X, Y, and Z are independent random variables drawn uniformly from K.

It can be shown, see Paper IX, that $C_{\alpha\beta}$ is a convexity measure fulfilling all the requirements set in Definition 7. Note that γ is determined by α and β because of the requirement $\alpha + \beta + \gamma = 1$. Figure 9 illustrates possible α, β values resulting in $\alpha, \beta, \gamma > 0$.

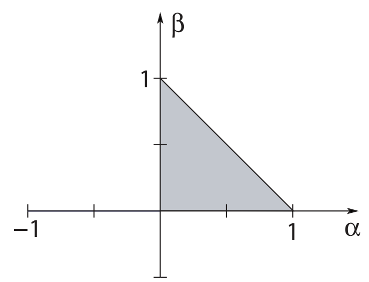


Fig 9. Region of valid α, β values for $C_{\alpha\beta}$.

One can observe that the formulation of $C_{\alpha\beta}$ is similar to C_{ls}, which originates from another definition of convexity. The fundamental difference, however, is that $C_{\alpha\beta}$ deals with points rather than line segments. This gives two significant advantages, the first one being that by varying the parameters α and β one can construct infinitely many different convexity measures. The second advantage is that there is a computationally efficient way of evaluating $C_{\alpha\beta}$, unlike in the case of C_{ls} which must be estimated by sampling. The connection of $C_{\alpha\beta}$ to MSA, or how to compute the $C_{\alpha\beta}$ is not obvious from the definition, but this will be clarified though the explicit expressions derived next.

Let X, Y, and Z be drawn independently from the set K. Thus, they all have the same probability density function

$$p(x) = \frac{1}{|K|}\chi_K(x), \tag{52}$$

where $\chi_K(x)$ is a characteristic function of K defined as

$$\chi_K(x) = \begin{cases} 1 & \text{if } x \in K, \\ 0 & \text{otherwise.} \end{cases} \tag{53}$$

Now let α, β, and γ be as in Definition 12 and define a new random variable

$$U_{\alpha\beta} = \alpha X + \beta Y + \gamma Z. \tag{54}$$

Using Lemma 2.1 in Paper II, it can be shown by a straightforward derivation that the probability density function of $U_{\alpha\beta}$ can be written in terms of p as

$$p_{U_{\alpha\beta}}(u) = \frac{1}{\alpha^2\beta^2\gamma^2}\left(p(\frac{x}{\alpha}) * p(\frac{x}{\beta}) * p(\frac{x}{\gamma})\right)(u), \tag{55}$$

if all $\alpha, \beta, \gamma > 0$, and with straightforward modifications if some of these are equal to zero.

From the definitions of $U_{\alpha\beta}$ and $C_{\alpha\beta}$, it is obvious that $C_{\alpha\beta}(K) = P(U_{\alpha\beta} \in K)$. Writing this out in terms of a probability density function gives

$$\begin{aligned} C_{\alpha\beta}(K) &= \int_K p_{U_{\alpha\beta}}(u)du \\ &= \int_{\mathbb{R}^2} \chi_K(u) p_{U_{\alpha\beta}}(u)du \\ &= \frac{1}{\alpha^2\beta^2\gamma^2}\int_{\mathbb{R}^2} \chi_K(u)\left(p(\frac{x}{\alpha}) * p(\frac{x}{\beta}) * p(\frac{x}{\gamma})\right)(u)du. \end{aligned} \tag{56}$$

Substituting here expression (52) for p, this becomes

$$C_{\alpha\beta}(K) = \frac{1}{|K|^3\alpha^2\beta^2\gamma^2}\int_{\mathbb{R}^2} \chi_K(u)\left(\chi_K(\frac{x}{\alpha}) * \chi_K(\frac{x}{\beta}) * \chi_K(\frac{x}{\gamma})\right)(u)du. \tag{57}$$

Applying the Plancherel formula, $\int_{\mathbb{R}^2} f\bar{g} = \int_{\mathbb{R}^2} \hat{f}\bar{\hat{g}}$ (see Section 3.3) and noting that the Fourier transform takes convolutions to products, one can reformulate (57) as

$$C_{\alpha\beta}(K) = \frac{1}{|K|^3}\int_{\mathbb{R}^2} \hat{\chi}_K(-\xi)\hat{\chi}_K(\alpha\xi)\hat{\chi}_K(\beta\xi)\hat{\chi}_K(\gamma\xi)\,d\xi, \tag{58}$$

where $\hat{\chi}_K$ denotes the Fourier transform of χ_K. Comparing (58) to the Fourier expression of MSA (23) it can be easily observed that

$$C_{\alpha\beta}(K) = M\chi_K(\alpha,\beta), \tag{59}$$

identifying the direct connection of the convexity measure to the multiscale autoconvolution. Formula (58) also gives a simple and efficient way of evaluating the measure $C_{\alpha\beta}$ from the given shape K using the Fourier transform.

Compared to the previously proposed measures C_{ls}, C_{ch}, C_{mcs}, C_{chp}, and C_{poly}, $C_{\alpha\beta}$ has two new properties. First of all, it offers a basis for constructing a whole set of new convexity measures, by altering the parameters α and β. The new measures, achieved in this way, can be then used to improve the discrimination as demonstrated in the Pollen experiment in Paper IX. Secondly, as explained in detail in Paper IX, $C_{\alpha\beta}$ can also be applied to gray scale images approximating a shape K. In this case, the gray values may be interpreted as the probabilities that particular points belong to K. Being able to apply the convexity measure directly to these situations removes the need for limiting the image to a strictly binary one.

Each convexity measure captures different, sometimes partially overlapping, characteristics of the object shape. For this reason, they are not generally easy to compare, since what is important in the shape can vary so much from application to application. However, being very straightforward to evaluate, and offering numerous possibilities for variations, the new MSA based convexity measure brings a valuable addition to the set of convexity measures. Figure 10 illustrates a set of objects classified according to their convexity measured with $C_{0.5,0.5}$.

Fig 10. Six objects ranked according to their convexity using the $C_{0.5,0.5}$ measure.

5 Conclusions

Pattern recognition and registration are two important tasks in many applications. Since a general solution to these problems has not yet been found, the questions have to be studied in restricted settings. In image processing, one particularly important situation arises when objects are photographed from various viewing angles. In this case, one may use a geometric transformation to describe the relations between patterns in one class, and use it as a basis for the recognition or registration. In many instances, affine mapping is a reasonable choice for this transformation, providing sufficient accuracy, and being still essentially linear. Hence, affine recognition and registration often appear in vision applications.

Both affine recognition and registration can be studied from several different perspectives. However, feature based approaches have been the ones which are most frequently applied. This thesis presented a new framework for constructing invariant feature extraction methods applied in these settings. The idea in the new approach is to extend the given pattern to a set of affine covariant versions, each carrying slightly different information, and then to apply known affine invariants to each of them separately. The key part of the framework is the construction of the affine covariant set, and this is done by combining several scaled representations of the original pattern.

The usage of the novel framework was illustrated in nine constructions, where different preprocessing methods, combination schemes, and final feature extraction techniques were applied. A short presentation of each of the new methods was offered, and a more thorough discussion is given in the corresponding original publications. There the methods are also assessed in several experiments, comparing them to previous approaches. The constructions presented are not in any way a complete set of possible ways to apply the multiscale approach, and further research may bring new interesting possibilities.

An additional application of one multiscale based technique in convexity measurements was introduced. Convexity is a descriptor of object shape. The proposed new measure proved to be simple to evaluate and it offered the possibility of many variations. Variability together with the ability to use the measure directly to gray-scale approximations of the binary shapes are properties previously not present in convexity measures.

The idea of applying scaled versions of the image in the multiscale framework was introduced particularly due to the properties of the affine transformation. In situations where the interconnections of the patterns are described with some other mappings, similar ideas may be applicable when the scaling is replaced by a suitable operation. For example, in order to construct rotation invariants, one could use combinations of rotated versions of the original pattern.

The multiscale methods and the corresponding experiments were presented using image intensity functions as patterns, since this is the main application where affine relations arise. Nevertheless, the new approach can be applied to other types of patterns where the affine transformation is present, as long as the patterns can be represented by functions. An example of such patterns could be volume data or probability densities. We are looking forward to such new applications in future.

References

Arbter K, Snyder WE, Burkhardt H & Hirzinger G (1990) Application of affine-invariant fourier descriptors to recognition of 3-d objects. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(7): 640–647.

Åström K (1995) Fundamental limitations on projective invariants of planar curves. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(1): 77–81.

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

Bishop CM (2006) Pattern classification and machine learning. Springer, New York.

Bracewell R (2000) The Fourier transform and its applications. MGraw-Hill.

Burkhardt H & Siggelkow S (2001) Invariant features in pattern recognition - fundamentals and applications. In: Pitas I & Kotropoulis C (eds) Nonlinear Model-Based Image/Video Processing and Analysis, 269–307. John Wiley & Sons, New York.

Burkhardt H, Simon JC & Neuvo Y (eds) (1991) From Pixels to Features: Parallelism in Image Processing: Proceedings of a Workshop, Bonas, France, 27 Aug.-1 Sept., 1990. Elsevier Science Inc., New York.

Candés EJ (1998) Ridgelets: Theory and applications, Ph.D. dissertation. Dept. Statistics, Stanford Univ., Stanford, USA.

Candés EJ & Donoho DL (1999) Ridgelets: A key to higher-dimensional intermittency? Phil. Trans. R. Soc. Lond. A. 2495–2509.

Chen JL & Kundu A (1995) Unsupervised texture segmentation using multichannel decompositions and hidden markov models. IEEE Transactions on Image Processing 4(5): 603–619.

Comaniciu D & Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5): 603–619.

Costa L & Cesar Junior RM (2001) Shape Analysis and Classification. CRC Press.

Do MN & Vetterli M (2003) The finite ridgelet transform for image representation. IEEE Transactions on Image Processing 12(1): 16–28.

Duda RO, Hart PE & Stork DG (2001) Pattern classification. John Wiley & Sons, New York.

Fischler MA & Bolles RC (1981) Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of ACM 24(6): 381–395.

Flusser J & Suk T (1993) Pattern recognition by affine moment invariants. Pattern Recognition 26(1): 167–174.

Forsyth DA & Ponce J (2003) Computer Vision a Modern Approach. Prentice Hall, New Jersey.

Fu KS (1982) Syntactic pattern recognition and applications. Prentice Hall.

Fu KS (1983) A step towards unification of syntactic and statistical pattern recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 5(2): 200–205.

Gonzalez R, Woods R & Eddins S (2002) Digital Image Processing. Prentice Hall.

Hadjidemetriou E, Grossberg M & Nayar S (2004) Multiresolution histograms and their use for recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(7): 831–847.

Hartley R & Zisserman A (2003) Multiple view geometry in computer vision, second edition.

Cambridge university press, United Kindom.
Heikkilä J (2002) Multi-scale autoconvolution for affine invariant pattern recognition. Proc. Proc. International Conference of Pattern Recognition, Québec, Canada, 119–122.
Heikkilä J (2004) Pattern matching with affine moment descriptors. Pattern recognition 37(9): 1825–1834.
Heikkilä M, Pietikäinen M & Schmid C (2006) Description of interest regions with center-symmetric local binary patterns. Proc. Proc. Computer Vision, Graphics and Image Processing, Madurai, India, 58–69.
Hu MK (1961) Pattern recognition by moment invariants. Proceedings of the IRE 49: 1428.
Hu MK (1962) Visual pattern recognition by moment invariants. IEEE Transactions on Information Theory 8: 179–187.
Hyde S (1997) The language of shape. Elsevier.
Jain AK, Duin RP & Mao J (2000) Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1): 4–37.
Jin Q & Yan P (1992) A new method of extracting invariants under affine transform. Proc. Proc. International Conference of Pattern Recognition, Hague, Netherlands, 1: 742–745.
Kadir T, Zisserman A & Brady M (2004) An affine invariant salient region detector. Proc. Proc. European Conference on Computer Vision, Prague, Czech Republic, 404–416.
Kadyrov A & Petrou M (2006) Affine parameter estimation from the trace transform. IEEE Transactions on Pattern Analysis and Machine Intelligence 28: 1631 – 1645.
Khalil MI & Bayoumi MM (2001) A dyadic wavelet affine invariant function for 2d shape recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(10): 1152–1164.
Khalil MI & Bayoumi MM (2002) Affine invariants for object recognition using the wavelet transform. Pattern Recognition Letters 23: 57–72.
Kreyszig E (1998) Advanced Engineering Mathematics (8th edition). John Wiley & Sons.
Lee T, McLean D & Atkins M (2003) A new irregularity measure for cutaneous lesions. Medical Image Analysis 7(1): 47–64.
Lowe D (2004) Distinctive image features from scale invariant keypoints. International Journal of Computer Vision 2(60): 91–110.
Maintz J & Viergever M (1998) A survey of medical image registration. Medical Image Analysis 2(1): 1–36.
Matas J, Chum O, Urban M & Pajdla T (2002) Robust wide baseline stereo from maximally stable extremal regions. Proc. Proc. British Machine Vision Conference, Cardif, UK, 384 – 393.
Mikolajczyk K & Schmid C (2004) Scale & affine invariant interest point detectors. International Journal of Computer Vision 60(1): 63–86.
Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T & Gool LV (2005) A comparison of affine region detectors. International Journal of Computer Vision 65(1): 43–72.
Mundy JL & Zisserman A (1992) Geometric Invariance in Computer Vision.
Ojala T, Pietikäinen M & Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7): 971–987.
Panjwani D & Healey G (1995) Markov random field models for unsupervised segmentation of textured color images. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(10): 939–954.

Petrou M & Bosdogianni P (1999) Image Processing. John Wiley & Sons, New York.

Petrou M & Kadyrov A (2004) Affine invariant features from the trace transform. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(1): 30–44.

Rahtu E, Salo M & Heikkilä J (2004) Convexity recognition using multi-scale autoconvolution. Proc. Proc. International Conference of Pattern Recognition, Cambridge, United Kingdom, 1: 692–695.

Reiss TH (1991) The revised fundamental theorem of moment invariants. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(8): 830–834.

Reiss TH (1993) Recognizing planar objects using invariant image features. Springer-Verlag, NJ.

Rigoutsos I (1998) Well-behaved, tunable 3d-affine invariants. Proc. Proc. Computer Vision and Pattern Recognition, Santa Barbara, USA, 455 – 460.

Rodrigues MA (2000) Invariants for pattern recognition and classification. World Scientific, Singapore.

Ronneberg O, Burkhardt H & Schultz E (2002) General-purpose object recognition in 3d volume data sets using gray-scale invariants – classification of airborne pollen-grains recorded with confocal laser scanning microscope. Proc. Proc. International Conference of Pattern Recognition, Québec, Canada, 290 – 295.

Rosin PL & Mumford CL (2006) A symmetric convexity measure. Computer vision and image understanding 103: 101–111.

Rudin W (1987) Real and Complex Analysis. McGraw-Hill.

Schael M & Siggelkow S (2002) Invariant grey-scale features for 3d sensor-data. Proc. Proc. International Conference of Pattern Recognition, Québec, Canada, 531 – 535.

Shen D & Ip H (1997) Generalized affine invariant image normalization. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(5): 431–440.

Shi J & Malik J (2000) Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8): 888–905.

Suk T & Flusser J (2004) Graph method for generating affine moment invariants. Proc. Proc. International Conference on Pattern Recognition, Cambridge, United Kingdom, 2: 192–195.

Suk T & Flusser J (2005) Affine normalization of symmetric objects. Proc. Proc. International Conference on Advanced Concepts for Intelligent Vision Systems, Antwerp, Belgium, 3708: 100–107.

Tuytelaars T & Gool LV (2004) Matching widely separated views based on affine invariant regions. International Journal of Computer Vision 59(1): 61–85.

Umbaugh S (2005) Computer imaging. CRC press, Florida.

Valentine F (1964) Convex sets. McGraw-Hill, San Fransisco.

Van Gool L, Moons T & Ungureanu D (1996) Affine/ photometric invariants for planar intensity patterns. Proc. Proc. European Conference on Computer Vision, Cambridge, United Kingdom, 1: 642 – 651.

Vapnik VN (1998) Statistical Learning Theory. John Wiley & Sons, New York.

Viola P & Wells WM (1997) Alignment by maximization of mutual information. International Journal of Computer Vision 24: 137–154.

Watanabe A (1985) Pattern Recognition: Human and Mechanical. John Wiley & Sons, New York.

Yang Z & Cohen FS (1999) Cross-weighted moments and affine invariants for image registration and matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(8): 804–814.

Zitova B & Flusser J (2003) Image registration methods: a survey. Image and Vision Computing 21: 977–1000.
Zunic J & Rosin PL (2004) A new convexity measure for polygons. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(7): 923–934.

Original articles

I Rahtu E, Salo M & Heikkilä J (2005) A new efficient method for producing global affine invariants. Proc 13th International Conference on Image Analysis and Processing (ICIAP 2005). Lecture Notes in Computer Science 3617: 407-414.

II Rahtu E, Salo M & Heikkilä J (2005) Affine invariant pattern recognition using multi-scale autoconvolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(6): 908-918.

III Kannala J, Rahtu E, Salo M & Heikkilä J (2005) A new method for affine registration of images and point sets. Proc 14th Scandinavian Conference on Image Analysis (SCIA 2005). Lecture Notes in Computer Science 3540: 224-234.

IV Kannala J, Rahtu E & Heikkilä J (2005) Affine registration with Multi-Scale Autoconvolution. Proc International Conference on Image Processing (ICIP 2005) 3: 1064-1067.

V Teynor A, Rahtu E, Setia L & Burkhardt H (2006) Properties of patch based approaches for the recognition of visual object classes. Proc 28th Jahrestagung der Deutschen Arbeitsgemeinschaft fr Mustererkennung (DAGM 2006). Lecture Notes in Computer Science 4174: 284-293.

VI Rahtu E, Salo M, Heikkilä J & Flusser J (2006) Generalized affine moment invariants for object recognition. Proc 18th International Conference on Pattern Recognition (ICPR 2006) 2: 634-637.

VII Rahtu E, Salo M & Heikkilä J (2006) A new affine invariant image transform based on ridgelets. Proc 16th British Machine Vision Conference (BMVC 2006) 3: 1059-1068.

VIII Rahtu E, Salo M & Heikkilä J (2006) J Multiscale autoconvolution histograms for affine invariant pattern recognition. Proc 16th British Machine Vision Conference (BMVC 2006) 3: 1039-1048.

IX Rahtu E, Salo M & Heikkilä J (2006) A new convexity measure based on a probabilistic interpretation of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9): 1501-1512.

X Rahtu E, Salo M & Heikkilä J (2007) Nonlinear functionals in the construction of multiscale affine invariants. Proc 15th Scandinavian Conference on Image Analysis (SCIA 2007). Lecture Notes in Computer Science 4522: 482-491.

Original publications are not included in the electronic version of the dissertation.

265. Suikki, Raija (2007) Changing business environment—effects of continuous innovations and disruptive technologies

266. Harri Haapasalo & Päivi Iskanius (Eds.) (2007) The 1st Nordic Innovation Research Conference—Finnkampen

267. Pikka, Vesa (2007) A Business Enabling Network. A case study of a high-tech network; its concepts, elements and actors

268. Noora Ervasti, Elina Pietikäinen, Eva Wiklund, Maarit Hiltunen ja Esa Vesmanen (2007) Yksilölliset toiveet asuntotuotannossa

269. Fabritius, Tapio (2007) Optical method for liquid sorption measurements in paper

270. Miao, Honglei (2007) Channel estimation and positioning for multiple antenna systems

271. Wang, Lingyun (2007) The key activities of partnership development in China—a study of Sino-Finnish partnerships

272. Aikio, Janne P. (2007) Frequency domain model fitting and Volterra analysis implemented on top of harmonic balance simulation

273. Oiva, Annukka (2007) Strategiakeskeinen kyvykkyyden johtaminen ja organisaation strateginen valmius. Kahden johtamismallin testaus

274. Jokinen, Hanna (2007) Screening and cleaning of pulp—a study to the parameters affecting separation

275. Sarja, Tiina (2007) Measurement, nature and removal of stickies in deinked pulp

276. Tóth, Géza (2007) Computer modeling supported fabrication processes for electronics applications

277. Näsi, Jari (2007) Intensified use of process measurements in hydrometallurgical zinc production processes

278. Turtinen, Markus (2007) Learning and recognizing texture characteristics using local binary patterns

279. Sarpola, Arja (2007) The hydrolysis of aluminium, a mass spectrometric study

280. Keski-Säntti, Jarmo (2007) Neural networks in the production optimization of a kraft pulp bleach plant

281. Hamada, Atef Saad (2007) Manufacturing, mechanical properties and corrosion behaviour of high-Mn TWIP steels

Book orders:
OULU UNIVERSITY PRESS
P.O. Box 8200, FI-90014
University of Oulu, Finland

Distributed by
OULU UNIVERSITY LIBRARY
P.O. Box 7500, FI-90014
University of Oulu, Finland

UNIVERSITY OF OULU P.O. Box 7500 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

ISBN 978-951-42-8600-1 (Paperback)
ISBN 978-951-42-8601-8 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)