OULU 2009
C 330

ACTA UNIVERSITATIS OULUENSIS

Timo Ahonen

FACE AND TEXTURE IMAGE ANALYSIS WITH QUANTIZED FILTER RESPONSE STATISTICS

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

ACTA UNIVERSITATIS OULUENSIS
C Technica 330

TIMO AHONEN

FACE AND TEXTURE IMAGE ANALYSIS WITH QUANTIZED FILTER RESPONSE STATISTICS

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 28 August 2009, at 12 noon

OULUN YLIOPISTO, OULU 2009

Acta Univ. Oul. C 330, 2009

Supervised by
Professor Matti Pietikänen

Reviewed by
Doctor Sébastien Marcel
Professor Majid Mirmehdi

ISBN 978-951-42-9181-4 (Paperback)
ISBN 978-951-42-9182-1 (PDF)
http://herkules.oulu.fi/isbn9789514291821/
ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)
http://herkules.oulu.fi/issn03553213/

Cover design
Raimo Ahonen

OULU UNIVERSITY PRESS
OULU 2009

Ahonen, Timo, Face and texture image analysis with quantized filter response statistics
Faculty of Technology, Department of Electrical and Information Engineering, University of Oulu, P.O.Box 4500, FI-90014 University of Oulu, Finland; Infotech Oulu, University of Oulu, P.O.Box 4500, FI-90014 University of Oulu, Finland
Acta Univ. Oul. C 330, 2009
Oulu, Finland

Abstract

Image appearance descriptors are needed for different computer vision applications dealing with, for example, detection, recognition and classification of objects, textures, humans, etc. Typically, such descriptors should be discriminative to allow for making the distinction between different classes, yet still robust to intra-class variations due to imaging conditions, natural changes in appearance, noise, and other factors.

The purpose of this thesis is the development and analysis of photometric descriptors for the appearance of real life images. The two application areas included in this thesis are face recognition and texture classification.

To facilitate the development and analysis of descriptors, a general framework for image description using statistics of quantized filter bank responses modeling their joint distribution is introduced. Several texture and other image appearance descriptors, including the local binary pattern operator, can be presented using this model. This framework, within which the thesis is presented, enables experimental evaluation of the significance of each of the components of this three-part chain forming a descriptor from an input image.

The main contribution of this thesis is a face representation method using distributions of local binary patterns computed in local rectangular regions. An important factor of this contribution is to view feature extraction from a face image as a texture description problem. This representation is further developed into a more precise model by estimating local distributions based on kernel density estimation. Furthermore, a face recognition method tolerant to image blur using local phase quantization is presented.

The thesis presents three new approaches and extensions to texture analysis using quantized filter bank responses. The first two aim at increasing the robustness of the quantization process. The soft local binary pattern operator accomplishes this by making a soft quantization to several labels, whereas Bayesian local binary patterns make use of a prior distribution of labelings, and aim for the one maximizing the a posteriori probability. Third, a novel method for computing rotation invariant statistics from histograms of local binary pattern labels using the discrete Fourier transform is introduced.

All the presented methods have been experimentally validated using publicly available image datasets and the results of experiments are presented in the thesis. The face description approach proposed in this thesis has been validated in several external studies, and it has been utilized and further developed by several research groups working on face analysis.

Keywords: computer vision, face recognition, feature extraction, robust descriptors, texture analysis

Preface

The research work for this thesis was carried out in the Machine Vision Group of the Department of Electrical and Information Engineering at the University of Oulu. I want to thank Prof. Matti Pietikäinen for supervising this study and for all the support and advice. Dr. Abdenour Hadid and Dr. Topi Mäenpää provided me with much help and guidance, especially in the early stages of this work, and I gratefully acknowledge this. I would also like to thank Prof. Rama Chellappa and Dr. Gaurav Aggarwal for the advice and interesting collaboration during my research visit to University of Maryland.

I am indebted to the co-authors of the papers, Dr. Chu He, Dr. Jiří Matas, Dr. Esa Rahtu, Ville Ojansivu and Prof. Janne Heikkilä, for all the cooperation, ideas and discussions. I also want to thank Dr. Hadid and Dr. Rahtu for reading and commenting on the manuscript of this thesis.

The Machine Vision Group has been a wonderful place to work. For that, thanks go to the whole group. Especially, I would like to thank Marko Heikkilä, Vili Kellokumpu and Valtteri Takala for all the good discussions, not always related to computer vision.

I am grateful to Prof. Majid Mirmehdi and Dr. Sébastien Marcel for serving as official reviewers of the thesis and for the their constructive comments and feedback. I also would like acknowledge the help of Gordon Roberts for the language revision of the thesis.

The funding received from the Graduate School in Electronics, Telecommunications and Automation, the Academy of Finland, the EC project MOBIO, the Nokia Foundation, the Tauno Tönning Foundation, and the Kaute Foundation is gratefully acknowledged.

I am most grateful to my parents Seppo and Helena for the all the support and encouragement over the years. I also wish to thank my friends for the moments they took my mind off work.

I want to express my deepest gratitude to my wife Anna for all the love and for being there for me. Life, be it research or anything else, is just so much better shared with you. Finally, thanks go to our daughter Saana for bringing joy and sunshine to our days.

Oulu, May 2009 Timo Ahonen

Abbreviations

a	Discretized rotation angle
$b_p(i)$	p-th bit of binary representation of i.
β	Frequency parameter for local phase quantization
$\chi^2(\cdot,\cdot)$	Chi square distance of histograms
$\mathbf{C}$	Covariance matrix of pixels in the support area of local phase quantization filters
$\mathbf{D}$	Covariance matrix of local phase quantization filter responses
$\delta\{\cdot,\cdot\}$	Kronecker delta
$D(I_{lab},C)$	Decision function for face verification given labeled input image I_{lab} and claimed identity C
$F_1,\ldots,F_P$	A bank of P filter kernels
$F(u,v,x,y)$	Short term Fourier transform of an image at frequency (u,v) and spatial location (x,y)
$f_{\{0,1\},d}(\cdot)$	Fuzzy membership functions, d being the fuzzification parameter
$h(\cdot)$	Histogram
$h^{(x,y)}(\cdot)$	Kernel density estimate of local histogram at pixel location (x,y)
$H(n,\cdot)$	Discrete Fourier transform of n-th row of the LBP histogram ordered into rows corresponding to uniform patterns.
$I(x,y)$	An image
$I^{\alpha^\circ}(x,y)$	Image $I(x,y)$ rotated by α degrees.
$\mathbf{I}_f(x,y)$	A vector valued image resulting from applying a filter bank
$\mathbf{I}_w(x,y)$	A whitened vector valued image
$I_p(x,y)$	p-th element of vector valued image $\mathbf{I}_f(x,y)$ at location (x,y)
$I_{lab}(x,y)$	A label image
$g_1,\ldots,g_P$	Gray values of P sampling points in a circular neighborhood
g_c	Gray value of center pixel in a neighborhood
$\lambda_C(I_{lab}(x,y),x,y)$	Log likelihood of observing LBP label $I_{lab}(x,y)$ at location (x,y), given claimed identity C.
M	Number of distinct labels produced by a vector quantizer $q(\cdot)$
N_R	Number of local face regions

P	Number of filters in a filter bank, resp. number of sampling points in a neighborhood
$q(\cdot)$	A vector quantizer function
R	Radius of a circular sampling neighborhood
ρ	Correlation coefficient between neighboring pixels
$s_t(\cdot)$	Thresholding function with threshold t
$SLBP(x,y,i)$	Weight of soft LBP assignment of pixel (x,y) to bin i
$U(\cdot)$	Uniformity measure of a bit pattern
$U_P(n,r)$	Uniform bit pattern with P bits, n being the number of 1-bits and r the rotation of the pattern
$\mathbf{V}$	Whitening transformation matrix for local phase quantization filter responses
$w(j)$	Weight for region j in region based weighted information fusion for face recognition
$w(x,y)$	Weight for pixel location (x,y) in pixel based weighted information fusion for face verification

BANCA	A dataset and associated protocols for face verification experiments
BLBP	Bayesian local binary pattern
CMU PIE	CMU Pose, Illumination, and Expression (PIE) face dataset
EBGM	Elastic Bunch Graph Matching
FERET	A face image dataset and associated protocols for evaluating performance of face recognition algorithms
FLS	Filtering, labeling and statistic
FRGC	Face Recognition Grand Challenge
ICA	Independent Component Analysis
KDE	Kernel Density Estimation
LDA	Linear Discriminant Analysis
LBP	Local binary pattern
LBP^{u2}	Uniform local binary pattern operator
LBP^{riu2}	Uniform and rotation invariant local binary pattern operator
$\mathrm{LBP}_{P,R}$	Local binary pattern operator using P sampling points on a circle of radius of R
LBP-HF	Local binary pattern histogram Fourier features
LPQ	Local phase quantization

MAP	Maximum a posteriori
ML	Maximum likelihood
MRF	Markov Random Field
MR8	Maximum Response 8 filter bank
PCA	Principal Component Analysis
SIFT	Scale invariant feature transform
SLBP	Soft local binary pattern
SVM	Support vector machine
SQI	Self-Quotient Image

List of original articles

This thesis is based on the following articles, which are referred to in the text by their Roman numerals (I–VIII):

I Ahonen T & Pietikäinen M (2009) Image description using joint distribution of filter bank responses. Pattern Recognition Letters 30(4): 368–376.

II Ahonen T & Pietikäinen M (2007) Soft histograms for local binary patterns. In: Niskanen M & Heikkilä J (eds) Proc Finnish Signal Processing Symposium (FINSIG 2007), Oulu, Finland.

III He C, Ahonen T & Pietikäinen M (2008) A Bayesian local binary pattern texture descriptor. Proc. 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, FL.

IV Ahonen T, Matas J, He C & Pietikäinen M (2009) Rotation invariant image description with local binary pattern histogram Fourier features. Proc. 16th Scandinavian Conference on Image Analysis (SCIA 2009), Oslo, Norway. Lecture Notes in Computer Science 5575, Springer.

V Ahonen T, Hadid A & Pietikäinen M (2004) Face recognition with local binary patterns. Proc. European Conference on Computer Vision (ECCV 2004), Prague, Czech Republic. Lecture Notes in Computer Science 3021, Springer: 469–481.

VI Ahonen T, Hadid A & Pietikäinen M (2006) Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12): 2037–2041.

VII Ahonen T & Pietikäinen M (2009) Pixelwise local binary pattern models of faces using kernel density estimation. Proc. 3rd IAPR/IEEE International Conference on Biometrics (ICB 2009), Alghero, Italy. Lecture Notes in Computer Science 5558, Springer.

VIII Ahonen T, Rahtu E, Ojansivu V & Heikkilä J (2008) Recognition of blurred faces using local phase quantization. Proc. 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, FL.

The writing and the experiments for papers I, II and IV–VIII were the work of the present author, while the co-authors gave valuable suggestions and comments on the text, and participated by discussing the ideas and the results. For paper III, the experiments were conducted and writing was carried out for the most part by Dr. He, while the present author gave input by participating in the discussions leading to the ideas presented in the paper, and discussing the experiments and results.

Contents

1 Introduction

1.1 Background

Visual detection and classification is of the utmost importance in several applications. Is there a human face in this image and if so, who is it? What is the person in this video doing? Has this photograph been taken inside or outside? Is there some defect in the textile in this image, or is it of acceptable quality? Does this microscope sample represent cancerous or healthy tissue?

To facilitate automated detection and classification in these types of questions, both good quality descriptors and strong classifiers are likely to be needed. In the appearance based description of images, a long way has been traveled since the pioneering work of Bela Julesz in (Julesz 1962), and good results have been reported in difficult visual classification tasks, such as texture classification (Ojala *et al.* 2002b), face recognition (Phillips *et al.* 2007), and object categorization (Zhang *et al.* 2007b). Still, the capabilities of automated vision systems seem awfully limited.

The human visual system works as a source of inspiration, but also as a desperately challenging benchmark for researchers in computer vision. Humans are able to effortlessly extract information, even from severely degraded visual input. Tasks such as detecting faces or other objects, recognizing people from their appearance, or inferring the three-dimensional shape and mutual location of objects in a scene are solved subliminally by humans, but still pose a great challenge for automated vision systems.

What makes the problem of visual detection and classification challenging is the great variability in real life images. The two types of object categories this work focuses on, textures and human faces, both show major changes in their appearance as the viewpoint or lighting changes. Other sources of challenges are background clutter, possible occlusion, non-rigid deformations such as changes in facial expression, or folding of a cloth whose texture is analyzed, and change of appearance over time. Furthermore, image acquisition itself may present perturbations, like blur, due to the camera being out-of-focus, or noise.

Over the last few years, progress in the field of machine learning has manifested in learning based methods to cope with the variability in images. In practice, the system tries to learn the intra- and inter-class variability from, typically a very large set of, training examples. Despite the advances in machine learning, the maxim “garbage in, garbage out” still applies: if the features the machine learning algorithm is provided with do not convey the essential information for the application in question, good final results cannot be expected. In other words, good descriptors for image appearance are called for.

1.2 Motivation: face description with local binary patterns

The motivation for this work comes from the observation that face images can be effectively described with the Local Binary Pattern (LBP) texture operator. This model for face description was first presented in papers V and VI, and is discussed in more detail in Chapter 4. However, as face recognition work was the basis for this research, we briefly review it here.

Before the publication of Paper V, the LBP operator had become well known in texture description. Still, it had mostly been applied to traditional texture analysis problems. Paper V was the first paper to apply LBPs to face recognition, thus bringing a texture approach to face description.

The face descriptor of Paper V is outlined in Fig. 1. The method assumes a registered face image, and divides it into local rectangular regions. The texture or appearance of each region is then described using LBP histograms, and these histograms are concatenated to build a global description of a face. In the experiments, it was shown that this descriptor can give a good face recognition performance, and it is robust to changes in appearance due to facial expressions, lighting changes or aging of the subjects.

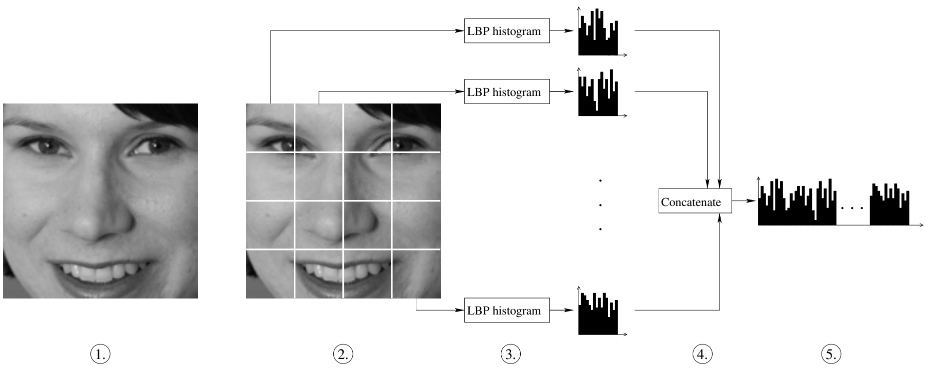

Fig 1. Face description with local binary patterns. 1. A registered face image is assumed. 2. The image is divided into local regions. 3. An LBP histogram is extracted independently from each region. 4. The histograms from regions are concatenated. 5. The concatenated histogram is used as descriptor for face recognition.

From the face recognition research community point of view, this paper has already had an impact. As discussed in Chapter 4, the approach of Papers V and VI has been adopted and further developed by several research groups.[1] More, extensions of this method have also been utilized in commercial applications, and such a system was used, for example, in the Beijing Olympics 2008 for identifying visitors (Ao *et al.* 2009).

These results also served as the motivation for the research leading to this thesis, as they showed that the local binary pattern methodology can be applied to the analysis of real world images, even problems that had not traditionally been considered as related to texture analysis.

1.3 Objectives of the research

Good descriptors for image appearance are called for. That being said, the objective of this work is to develop descriptors for the appearance of real life images. The goals are to have descriptors tolerant to the aforementioned challenges, discriminative enough for the application in question, and computationally feasible. Image description in general

[1]At the time of the writing, the Scopus database (http://www.scopus.com/) listed 70 citations to Paper VI, making it the most cited article published in IEEE Transactions on Pattern Analysis and Machine Intelligence in the year 2006.

is a very large field, and, in this thesis, the methods are presented in a rather general form, but the focus is on their application to *texture classification* and *face recognition*.

Looking back at the history of the development of image descriptors, even though some works were applying joint statistics already in 60's and 70's (e.g. Julesz 1962, Haralick *et al.* 1973) much of the emphasis was first on using marginal statistics for image description (e.g. Laws 1979, Manjunath & Ma 1996). Later, however, the research focused on joint statistics of either the pure pixel values or filter responses (Popat & Picard 1993, Valkealahti & Oja 1998, Malik *et al.* 1999, Portilla & Simoncelli 2000, Schiele & Crowley 2000, Ojala *et al.* 2001, Varma & Zisserman 2003). In this work, the aim is also to concentrate on joint statistics, more specifically, an estimate of a multi-dimensional histogram based on vector quantization.

Summarizing, the objective of this work is to study the use of statistics of quantized filter responses in image description, with its application to face recognition and texture analysis.

1.4 The contributions of the thesis

The contributions of this thesis are related to image description for face recognition and texture classification. With regard to all the contributions described in the following, due experimental validation to assess their applicability and performance is reported in Papers I–VIII.

The first contribution of this thesis, upon which the rest of the research leading to this thesis was based, is face representation using local binary pattern histograms computed in local rectangular regions. An important factor of this contribution is to view feature extraction from face images as a texture description problem. The second contribution for texture based face analysis is a revised model for local texture representation based on kernel density estimation.

Another contribution of this thesis is a general model for image description using vector quantization of filter bank responses to model their joint distribution. The most important aspect of this contribution is that the local binary pattern operator can also be presented as an operator producing vector quantized filter bank responses. This model allows for evaluating the significance of filter bank and vector quantizer selection, and it constitutes the general framework used in this thesis. The application of this model to texture and face analysis is examined.

Furthermore, the thesis presents two alternative approaches to increase the robustness of quantization based models. When there is uncertainty about quantization, soft local binary patterns offer soft quantization to several labels. Another proposed approach, the Bayesian local binary patterns, makes use of a prior distribution of labelings, and aims for the labeling maximizing the a posteriori probability.

Another contribution of this thesis to texture analysis is a novel statistic computed from LBP labels that is invariant to rotations of the input image, but retains more information about the structure of the image than the original rotation invariant LBP operator.

Finally, a face recognition method tolerant to image blur, using local phase quantization, is presented.

1.5 Organization of the thesis

This book is divided into two parts. The first part, consisting of five chapters, summarizes and presents the main results from the original papers. These eight papers form the second part of the book

Chapter 1 is an introduction to this work, presenting the background, motivation and objectives of the work as well as a brief summary of the original papers.

Chapter 2 first reviews the literature on general photometric image appearance description and its applications. Then, a general framework for image descriptors consisting of filtering, labeling and a statistic function is presented. The rest of the contributions are presented in this framework.

The next two chapters are concerned with the two application areas of this work: texture and face image description. First, in Chapter 3, we take a look at texture analysis problems and the literature. Then, three novel approaches to texture description are introduced, and their experimental evaluation is reported.

The theme of Chapter 4 is face description. Again, a literature review is first presented. Then a novel method of face description using local binary patterns is introduced, approaching the problem from a texture analysis viewpoint. Two extensions of this baseline are proposed, and experimental validations are reported. Finally, further work deriving from LBP based face representation is reviewed.

The last chapter of the first part provides a summary of the thesis.

1.6 Summary of original papers

Paper I, an extension of (Ahonen & Pietikäinen 2008), presents a unified framework for image descriptors based on quantized joint distribution of filter bank responses, and it evaluates the significance of filter bank and vector quantizer selection. This paper constitutes the general framework within which this thesis is presented. Furthermore, it presents the local binary pattern operator as a linear filter and vector quantization operation.

The next two papers, II and III discuss the problems arising from uncertainty in the quantization process after linear filtering. To combat this challenge, Paper II proposes to replace the quantization operation with soft histograms. Paper III takes a different approach and proposes utilizing information from the local neighborhood of a pixel. This leads to a Markov Random Field (MRF) like model of the LBP labeling process based on a prior probability model of different labelings.

Paper IV considers the problem of rotation invariant texture classification. The paper proposes Local Binary Pattern Histogram Fourier features (LBP-HF), a novel rotation invariant image descriptor computed from discrete Fourier transforms of LBP histograms. Unlike most other histogram based invariant texture descriptors which normalize rotation locally, these invariants are constructed globally for the whole region to be described.

The remaining four papers in the thesis are related to face image representation. The first two describe a general model for face description, and the final two discuss two extensions to it.

Paper V presents a general model of face representation using local texture estimated in local rectangular blocks. Together with (Ahonen *et al.* 2004), it was later presented in a revised form in Paper VI. Paper VII extends this model by estimating the LBP distributions at each pixel with kernel density estimation. Furthermore, Paper VII proposes the use of weighted information fusion from individual pixels, based on a linear support vector machine (SVM).

In Paper VIII, recognition of blurred faces using the Local Phase Quantization (LPQ) operator is proposed. To the author's knowledge, this is the first paper explicitly constructing blur-invariant descriptors for face recognition.

2 Image appearance description

This chapter discusses photometric, i.e. appearance based, image description. A literature review is first presented and the framework for image description originally presented in papers I and III is then introduced.

The focus is on photometric description, i.e. methods that derive the descriptors from pixel values locally or globally. Other possible approaches, outside the scope of this thesis, include structural image description and pure geometric features.

2.1 Applications

Stable and reliable image appearance description has applications in a multitude of tasks related to image analysis and computer vision. The two applications that this thesis focuses on, texture analysis and face recognition, are discussed in Chapters 3 and 4, respectively. Some of the other applications are reviewed here.

Object detection, recognition and categorization are currently topics generating much activity in computer vision. The goal is to detect and recognize either specific instances of an object or broader object classes, such as cars or pedestrians in images. Over the last couple of years, the improvement shown in object classification and detection in the PASCAL visual challenges workshops[2] can be explained both by progress in appearance description and in the machine learning and classification methods used to analyze the images.

In its most simple form, content based image or multimedia retrieval (Lew *et al.* 2006, Datta *et al.* 2008) applies image appearance descriptors for searching a database for images that are visually similar to the query image. The user, however, is typically interested in higher level semantical concepts. This issue is referred to as the "semantic gap" (Smeulders *et al.* 2000).

Local appearance descriptors can be used for finding correspondences between images. This is needed in image stitching and mosaicing, which aim to combine several partly overlapping images into a single mosaic (Brown & Lowe 2007). Correspondences are also needed in stereo matching, and even though in that field a common approach is to search for correspondences using intensity values as such, some works applying local descriptors have been presented (Brown *et al.* 2003).

[2] http://pascallin.ecs.soton.ac.uk/challenges/VOC/

2.2 Image appearance descriptors

Attempts to categorize image appearance descriptors include, for example, dividing the description methods to global (holistic) and local (parts based) (e.g. Zhao *et al.* 2003), or making the distinction between sparse and dense descriptors (e.g. Lazebnik *et al.* 2005). Neither of these categorizations is complete: descriptors that are occasionally referred to as 'global' can be easily made local by computing them over a set of spatially restricted local regions. On the other hand, sparse descriptors typically take the dense description approach (i.e. account for every pixel) within the regions they are applied.

Image appearance descriptors that encode the appearance of the whole input image into a single descriptor or feature vector, have been studied extensively in texture analysis research. Furthermore, especially the early studies on content based image retrieval resorted to global image descriptors (Smeulders *et al.* 2000).

In this study, we concentrate on appearance description from monochromatic images, but it is well worth mentioning that color based and multi-band descriptors have also proven to be useful in a variety of image analysis tasks. Classical works in this area include the multi-band co-occurrence matrices and difference histograms discussed by Rosenfeld *et al.* (1982), color histograms by Swain & Ballard (1991) and color correlograms by Huang *et al.* (1999). For a recent comparison of color descriptors in image and video categorization, see (van de Sande *et al.* 2008).

2.2.1 Early works

In global image description for texture perception, pioneering work was done by Bela Julesz. In (Julesz 1962), it was postulated that textures which agree in their joint pixel densities up to N-th order and differ only in higher order joint statistics are preattentively indistinguishable to a human observer. Later a stronger statement was proposed, claiming that two textures could not be perceptually distinguished if they had identical second order statistics (Julesz *et al.* 1973). A counterexample to this was, however, presented, and it was found that some textures with identical second and third order statistics are visually distinct (Caelli & Julesz 1978). Finally, Julesz & Bergen (1981) formulated a theory of textons which presents texture as a composition of so called textons, which are visual events of the image, such as line terminations, intersections, etc.

An approach similar to Julesz's second order statistics, applied to gray scale images, is the gray-level co-occurrence matrix introduced by Haralick *et al.* (1973). The co-occurrence matrix records the number of occurrences of all different gray values pairs at pixels defined by a displacement vector, i.e. it is an estimate of second order joint probability of gray values occurring at pixel locations separated by the displacement vector. Yet another similar descriptor is the gray value difference histogram (Weszka *et al.* 1976) which records the difference of the gray values instead of the joint probability.

2.2.2 Statistics of linear filter outputs

The first methods applying linear filters to image description were developed in the late 1970's and early 80's (e.g. Laws 1979, Faugeras & Pratt 1980, Knutsson & Granlund 1983, Ade 1983, Coggins & Jain 1985). The early filter bank methods typically apply energy measures or simple statistics such as mean and variance of each of the filter responses to characterize the image appearance. A comparison of this type of descriptors in the context of segmentation was done by Randen & Husøy (1999).

Most of the newer descriptors aim to use joint statistics. Use of joint statistics has been found to be useful in texture synthesis by Popat & Picard (1993) and Bonet & Viola (1997). Portilla & Simoncelli (2000) showed in their texture synthesis work that joint statistics are needed in addition to marginal statistics of wavelet coefficients to get a good representation of the texture appearance.

It appears that there is no consensus among researchers on whether marginal statistics of filter responses are sufficient to build descriptors able to discriminate patterns that are visually distinct or whether joint statistics are needed. In contrast to the aforementioned findings in texture synthesis, some studies on texture classification have recently advocated the use of marginal statistics (Zhu *et al.* 1998, Do & Vetterli 2002, Liu & Wang 2003, Hadjidemetriou *et al.* 2004). On the other hand, Gluckman (2005) showed that for some commonly used filter banks, there exist visually distinct patterns that are similar in marginal and joint distribution of filter responses.

The term texton was reintroduced by Malik *et al.* (1999) to mean a cluster center in the space of filter bank responses. Then, a histogram of these textons would approximate the full joint probability density of responses. A similar clustering based approach was also presented by Rikert *et al.* (1999) for the task of object detection.

Quantizing filter responses and then computing a multi-dimensional histogram has also been studied. Pass & Zabih (1999) create a joint histogram of color and four other local features, and use it for image retrieval. Konishi & Yuille (2000) use six different filters and apply the representation to supervised image segmentation. Schiele & Crowley (2000), on the other hand, use joint statistics of Gaussian derivative filter responses over multiple scales for object recognition.

The clustering of filter responses for texture description was considered by Leung & Malik (2001), but that study explicitly considered three dimensional textures and required a stack of registered texture images from different viewpoints and lightings. Cula & Dana (2004) and Varma & Zisserman (2004, 2005) took this approach to more general image description by proposing improved texton models which use different filter banks and do not require a registered set of images. In those works, appearance changes due to viewpoint and illumination effects caused by the three-dimensional nature of the texture were handled using multiple models for each texture.

A critical viewpoint to filter bank based texture description was taken by Varma & Zisserman (2003). They showed that in texture classification, pixel values from small image patches can yield as representative codebooks as responses of filters with much larger spatial support. Earlier, Ojala *et al.* (2001) has proposed a very similar method for texture description with vector quantizing gray value differences in local neighborhoods. Yet another similar technique is the texem model by Xie & Mirmehdi (2007) which models the appearance of a texture with a Gaussian mixture model of pixel values of image patches at multiple scales.

2.2.3 Local region descriptors

An approach to image description that has been gaining interest since the 90's is local image descriptors. In contrast to global descriptors, where the features encode the appearance of the whole image, local descriptors consider smaller subregions of the image.

Depending on the application, the local regions to be described may be obtained around automatically detected interest points (Schmid *et al.* 2000) or affine covariant regions (Mikolajczyk *et al.* 2005). Random sampling was proposed by Nowak *et al.* (2006) and dense sampling, also typical in texture description, was applied in (Tuytelaars & Schmid 2007). A fixed grid was used for determining the local regions in Paper VI, and, for example, by Vogel & Schiele (2007).

As the local patch to be described has been determined and possibly transformed into some canonical orientation (c.f. Section 2.2.4), a variety of options for constructing photometric features exists. One can apply some of the global descriptors within the image patch, or use descriptors specifically designed for local regions. A comparison of local descriptors by Mikolajczyk & Schmid (2005) evaluated several local descriptors in a retrieval framework, and concluded that a gradient location-orientation histogram (GLOH) provided the best results, closely followed by a scale invariant feature transform (SIFT) descriptor by Lowe (2004). Both these descriptors are based on computing histograms of gradient orientation within subregions of the patch to be described. Speeded up robust features (SURF) is a recently introduced local descriptor by Bay *et al.* (2008). The SURF descriptor is based on Haar wavelets, and it is fast to compute. It is also claimed that it achieves better retrieval results than SIFT based description, despite its smaller descriptor dimension. (Bay *et al.* 2008).

The histogram of oriented gradients (HOG) descriptor (Dalal & Triggs 2005), based on a weighted histogram of gradient directions has shown good performance in human detection. Furthermore, also the local binary pattern operator has been extended to meet the needs of interest region description by Heikkilä *et al.* (2009).

2.2.4 Invariance of descriptors

Often it is desirable to have descriptors that are robust or totally invariant to certain kinds of perturbations of input image. Such perturbations include noise, blur, lighting changes, geometric transformations, or even non-rigid transformations of the objects.

Rotation invariance has been extensively studied in texture analysis, see (Zhang & Tan 2002) for a review.

A common approach to achieving invariant descriptors is to first extract some non-invariant features at each pixel location, and then apply a mapping which outputs invariant pixel level features. A global rotation invariant description is then obtained by computing translation invariant statistics such as the histogram of pixel level features over the image.

This general approach is used in the work by Arof & Deravi (1998) where neighbors of each pixel are first sampled on a circular neighborhood, and rotation invariant features are then computed from the samples and their differences to the central pixel taking the discrete Fourier transform magnitude. Siggelgov (2002) extracts rotation and translation invariant histograms by summing or computing the histogram over the group of translations and rotations. The rotation invariant LBP operator, presented in Section 2.2.5, also applies this general methodology, and Varma & Zisserman (2004) achieve rotational invariance by, at each pixel location, selecting the maximum response among different rotations of each filter type. In the approach of Schmid (2001), no mapping is needed, as it uses responses of isotropic filters, which are naturally rotation invariant.

Full affine invariance has been shown to be more difficult to achieve with descriptors still discriminative enough and robust to other possible challenges. Global affine invariant image descriptors include moment invariants (Hu 1962, Flusser & Suk 1993), the trace transform (Petrou & Kadyrov 2004) and multi-scale autoconvolution (Heikkilä 2002, Rahtu *et al.* 2005).

In image patch based representation, a common strategy is to represent the automatically detected local patch as an ellipse, which is then mapped to a circle of fixed size to remove the scale and shear ambiguities (Mikolajczyk *et al.* 2005). This circular region can be further normalized for rotation by using, for example, dominant local gradient direction (Lowe 2004), or a rotation invariant local descriptor can be applied. What makes a locally affine invariant description especially attractive is the notion that some stronger transformations, such as perspective transformation due to viewpoint changes or even non-rigid deformations of the objects, can be locally approximated with affine transformation. It has, however, been argued that full affine invariance may not be necessary in object and texture recognition, but using only scale and rotation invariance for local patches produces better recognition results (Zhang *et al.* 2007b).

Photometric normalization aims to compensate for photometric distortions due to lighting changes, image acquisition, etc, and makes the descriptors robust to such distortions. Arguably, domain-specific techniques produce most reliable normalization when applicable, as reported in surveys comparing generic and application specific illumination normalization methods for facial images (c.f. Section 4.1.1). Generic techniques for reducing the effect of illumination changes include local or global histogram equalization, normalizing local regions to zero mean and unit variance (Mikolajczyk & Schmid 2005), and normalizing filter response vectors using the L1 norm, as suggested by Malik *et al.* (2001).

2.2.5 Local binary patterns

The original local binary pattern

The basic local binary pattern operator, introduced by Ojala *et al.* (1994, 1996), was based on the assumption that texture has locally two complementary aspects, a pattern and its strength. In that work, an LBP was proposed as a two-level version of the texture unit (Wang & He 1990) to describe the local textural patterns. Another similar image operator, developed for image matching, is the census transform by Zabih & Woodfill (1994).

The original version of the local binary pattern operator works in a 3×3 pixel block of an image. The pixels in this block are thresholded by its center pixel value, multiplied by powers of two, and then summed to obtain a label for the center pixel. As the neighborhood consists of 8 pixels, a total of $2^8 = 256$ different labels can be obtained depending on the relative gray values of the center and the pixels in the neighborhood. See Figure 2 for an illustration of the basic LBP operator.

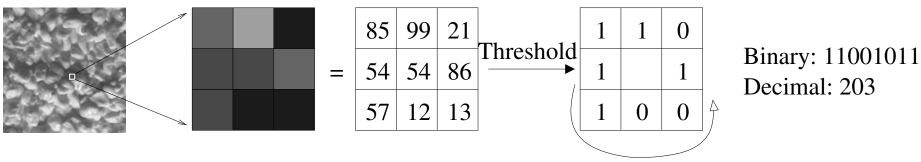

Fig 2. The basic LBP operator. [Paper VI] ©2006 IEEE.

Uniform and rotation invariant patterns

It can be argued, that despite the essential appearance of textures probably lies in the interpixel relations of pixels near each other, the information about relations of only neighboring pixels does not always capture the essence of the texture. Hence, the local binary pattern operator was later presented in a more generic revised form by Ojala *et al.* (2002b).

In the generic LBP operator, using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. In the following, the notation (P,R) will be used for pixel neighborhoods, which means P sampling points on a circle of radius R. See Figure 3 for an example of circular neighborhoods.

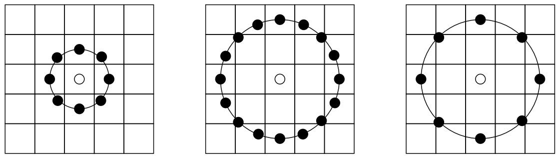

Fig 3. The circular (8,1), (16,2) and (8,2) neighborhoods. The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel. [Paper VI] ©2006 IEEE.

Let us denote the value of the center pixel (x,y) by $g_c = I(x,y)$ and the gray values of the P sampling points by $g_1 = I(x_1,y_1), g_2 = I(x_2,y_2), \ldots, g_P$, where the sampling points lie at coordinates $(x_p,y_p) = (x + R\cos(2\pi p/P), y - R\sin(2\pi p/P))$. Now the generic $\mathrm{LBP}_{P,R}$ operator is defined as

$$\mathrm{LBP}_{P,R} = \sum_{p=1}^{P} s(g_p - g_c)2^{p-1}, \tag{1}$$

where

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} . \tag{2}$$

Another extension to the original operator uses so called uniform patterns (Ojala *et al.* 2002b). For this, a uniformity measure of a pattern is used: U("pattern") is the number of bitwise transitions from 0 to 1, or vice versa, when the bit pattern is considered circular. A local binary pattern is called uniform if its uniformity measure is at most 2. For example, the patterns 00000000 ($U = 0$), 01110000 ($U = 2$) and 11001111 ($U = 2$) are uniform, whereas the patterns 11001001 ($U = 4$) and 01010011 ($U = 6$) are not. In uniform LBP mapping, there is a separate output label for each uniform pattern, and all the non-uniform patterns are assigned to a single label.

As the $\mathrm{LBP}_{P,R}$ patterns are obtained by circularly sampling around the center pixel, rotation of the input image has two effects: each local neighborhood is rotated into other pixel location, and within each neighborhood, the sampling points on the circle surrounding the center point are rotated into a different orientation (c.f. Fig. 10 on p. 46). Therefore, rotation invariant LBP labels can be obtained by applying mapping which circularly rotates each bit pattern to the minimum value. For instance, the bit sequences 1000011, 1110000 and 0011100 arise from different rotations of the same local pattern, and they all correspond to the normalized sequence 0000111.

We use the following notation for the LBP operator: $\mathrm{LBP}_{P,R}^{riu2}$. The subscript represents using the operator in a (P,R) neighborhood. Superscript *ri* denotes the use of rotation invariant mapping, and *u*2 stands for using only uniform patterns and labeling all remaining patterns with a single label.

2.3 The framework for filter bank and vector quantization based texture descriptors

This section presents the Filtering-Labeling-Statistic (FLS) framework for image descriptors. This framework was first presented in Paper I, and then in a slightly revised form in Paper III. The framework is illustrated in Fig. 4. Appendix 1 contains a summary presenting the methods proposed in this thesis, as well as some other works in the FLS framework.

Fig 4. The Filtering, Labeling and Statistic framework.

A widely used approach to texture analysis is to convolve an image with P different filters whose responses at a certain position (x,y) form a P-dimensional vector. At the learning stage, a set of such vectors is collected from training images, and the set is clustered using, for example, k-means to form a codebook. Then each pixel of a texture image is labeled with the label of the nearest cluster center, and the histogram of these labels over a texture image is used to describe the texture (e.g. Leung & Malik 2001, Cula & Dana 2004, Varma & Zisserman 2004). A different technique, not requiring codebook learning, is to apply thresholding to quantize each response into two or more values, and then compute their joint histogram (e.g. Schiele & Crowley 2000, Ojala *et al.* 2001).

More formally, let $I(x,y)$ be the image to be described by the texture operator. Now the vector valued image obtained by convolving the original image with filter kernels $F_1, F_2, \ldots, F_P$ is

$$\mathbf{I}_f(x,y) = \begin{bmatrix} I_1(x,y) = I(x,y) \star F_1 \\ I_2(x,y) = I(x,y) \star F_2 \\ \vdots \\ I_P(x,y) = I(x,y) \star F_P \end{bmatrix}. \tag{3}$$

The labeled image $I_{lab}(x,y)$ is obtained with a vector quantizer $q : \mathbb{R}^P \mapsto \{0,1,2,\cdots,M-1\}$, where M is the number of different labels produced by the quantizer. Thus, the labeled image is

$$I_{lab}(x,y) = q(\mathbf{I}_f(x,y)) \tag{4}$$

and the histogram of labels is

$$h(i) = \sum_{x,y} \delta\{i, I_{lab}(x,y)\}, i = 0,\ldots,M-1, \tag{5}$$

in which δ is the Kronecker delta

$$\delta\{i,j\} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}. \tag{6}$$

If the task is classification or categorization, as in this thesis, several possibilities exist for classifier selection. The most typical strategy is to use a nearest neighbor classifier with, for example, χ^2 distance measure (Leung & Malik 2001, Varma & Zisserman 2005). In (Varma & Zisserman 2004), the nearest neighbor classifier was compared to Bayesian classification, but no significant difference in the performance was found. In (Caputo *et al.* 2005), it was shown that the performance of a material categorization system can be enhanced by using a suitably trained support vector machine based classifier. In this thesis, the main interest is not in the classifier design, but in the local descriptors, and thus the nearest neighbor classifier with χ^2 distance was used in most experiments.

The following three subsections discuss in more detail the parts that define an image descriptor in the proposed framework. These parts are the filter bank $F_1, F_2, \ldots, F_P$, the quantization function q, and the statistic function.

2.3.1 The filter bank

In Paper I, three different types of filter kernels that are commonly used in texture description were compared. The first filter bank is a set of oriented derivative filters whose thresholded output is shown to be equivalent to the local binary pattern operator. The other two filter banks included in that comparison are Gabor filters and the Maximum Response 8 filter set. In Paper VIII, a blur tolerant descriptor for face representation was proposed. That is based on Gabor like filters and is discussed in Section 4.2.3.

A novel way to look at the LBP operator arising from the FLS representation is to see it as a special filter-based texture operator. The filters for implementing LBP are approximations of image derivatives computed at different orientations. The filter coefficients are computed so that they are equal to the weights of bilinear interpolation of pixel values at sampling points of the LBP operator, and the coefficient at the filter center is obtained by subtracting 1 from the center value. For example, the kernels shown in Fig. 5 can be used for filter based implementation of the local binary pattern operator in the circular (8,1) neighborhood. The response of such a filter at location (x, y) gives the signed difference of the center pixel, and the sampling point corresponding to the filter. These filters, which will be called local derivative filters in the following, can be constructed for any radius and any number of sampling points.

0	0	0
0	−1	1
0	0	0

0	0	0
0	−0.914	0.207
0	0.207	0.5

0	0	0
0	−1	0
0	1	0

0	0	0
0.207	−0.914	0
0.5	0.207	0

0	0	0
1	−1	0
0	0	0

0.5	0.207	0
0.207	−0.914	0
0	0	0

0	1	0
0	−1	0
0	0	0

0	0.207	0.5
0	−0.914	0.207
0	0	0

Fig 5. Local derivative filters filters $F_1 \cdots F_8$ at (8,1) neighborhood. Revised from [Paper I] ©2009 Elsevier.

Applying the Maximum Response 8 descriptor in this framework is straightforward. In the filter bank design for Paper I, we followed the procedure of Varma & Zisserman (2005). This filter bank produces a 38 dimensional vector valued image. Selecting the maximum over orientations is handled in the vector quantizer, and it is described in more detail in the following section.

For Gabor filters, much work has been devoted to designing the filter bank and feature computation methods, see, e.g., (Manjunath & Ma 1996, Clausi & Jernigan 2000, Grigorescu *et al.* 2002). In Paper I, we applied the Gabor filters in the proposed image description framework, which is to say that the responses of the filter bank at a certain position are stacked into a vector which is used as an input for the vector quantizer. This resembles the Gabor filter based face description suggested by Zhang *et al.* (2007a) in which the Gabor filter responses are quantized and a histogram of them is then formed to encode a facial image.

In the design of the filter bank, i.e. selection of the scale and frequency parameters, the procedure of Manjunath & Ma (1996) was applied. Furthermore, at each chosen scale and frequency, the real and imaginary parts of the complex Gabor filter were treated as separate filters, so the total number of filters is $2n_sn_f$, in which n_s is the number of scales, and n_f is the number of orientations at each scale.

2.3.2 Vector quantization

The assumption upon which the proposed texture description framework is based is that the joint distribution of filter responses can be used to describe the image appearance. In this thesis, two alternatives for estimating the joint distribution are presented. First, we discuss the widely used approach of quantizing the response vectors into bins and then using a histogram as a density estimate. Second, in Section 3.3.1 we present a more robust approach, soft histograms, derived from fuzzy sets.

Alternative approaches would be kernel density estimation (Scott 1992), or parametric models such as modeling the joint distribution as a mixture of multivariate Gaussians. To the authors' knowledge, a full comparison of the performance of these alternatives with local derivative filters (LBPs) has not been made. When using distributions of Gaussian derivative filters for object recognition, Schiele & Crowley (2000) found only a marginal difference in the generalization capabilities of histogram and kernel function estimates.

As we first aim at estimating the joint distribution from the vectors in the image $\mathbf{I}_f(x,y)$ using quantization, the natural question is which quantization function should be selected, and how much does the selection affect the final recognition performance of the system. Here we present thresholding and codebook based quantizers, and as a more robust alternative, in Section 3.3.2 we introduce a Bayesian LBP that incorporates information from neighboring pixels into the quantization using a Markov random field approach.

A simple, non-adaptive way of quantizing the filter responses is to threshold them, and to compute the sum of thresholded values multiplied by powers of two:

$$I_{lab}(x,y) = \sum_{p=1}^{P} s_t\left(I_p(x,y)\right) 2^{p-1}, \tag{7}$$

where $s_t(z)$ is the thresholding function

$$s_t(z) = \begin{cases} 1, & z \geq t \\ 0, & z < t \end{cases}, \tag{8}$$

in which the parameter t is the threshold. Thresholding divides each dimension of the filter bank output into two bins. The total number of different labels produced by threshold quantization is 2^P, where P is the number of filters.

If the threshold $t = 0$ and the coefficients of each of the filters in the bank have zero mean (i.e. they sum up to zero), the value of $s\left(I_p(x,y)\right)$, and thus the value of $I_{lab}(x,y)$ is not affected by affine gray level changes $I'(x,y) = \alpha I(x,y) + \beta, \alpha > 0$. If the filter coefficients $F_p(x,y)$ sum up to zero, $I'(x,y) \star F_p = \alpha(I(x,y) \star F_p)$ and, assuming $\alpha > 0$, the sign is not changed. On the other hand, such global gray level changes can be easily normalized, but this holds also locally for such areas where changes in pixel values, for example, due to lighting changes, can be modeled as affine gray level change within the filter support area.

Now, let us consider the case where the filter bank used to obtain the image $\mathbf{I}_f(x,y)$ is the set of local derivative filters (e.g. the filters presented in Fig. 5), designed so that filter responses are equal to the signed differences, i.e. at each pixel location

$$I_p = g_p - g_c,$$

where I_p is the response of p-th filter at a given location, and g_c and g_p are the center pixel and p-th sampling point gray values (see Eq. (1)) at the same location. As the quantizer (7) with threshold $t = 0$ is applied to $\mathbf{I}_f$, it follows that the resulting label I_{lab} is equal to that resulting from local binary pattern operator $\text{LBP}_{P,R}$ using the same neighborhood. Therefore, the LBP operator can be represented in the proposed framework.

It should be noted that for LBPs, an even stronger invariance to gray level changes holds than that presented above. As discussed by Ojala *et al.* (2002b), the LBP labels are invariant to any monotonic mapping of gray values.

Due to these reasons, the choice of threshold $t = 0$ has been common, especially with the LBP operator. Still, under some circumstances a different choice may yield better results. For instance, when LBP features are used for background subtraction, choosing a non-zero threshold was observed to result in more stable codes in nearly flat gray areas, and, consequently, increased performance (Heikkilä & Pietikäinen 2006).

Another method for quantizing the filter responses is to construct a codebook of them at the learning stage, and then use the nearest codeword to represent the filter bank output at each location:

$$I_{lab}(x,y) = \arg\min_{m} \left|\left|\mathbf{I}_f(x,y) - \mathbf{c}_m\right|\right|, \tag{9}$$

in which $\mathbf{c}_m$ is the m-th vector (codeword) in the codebook. This approach is used in (Leung & Malik 2001) and (Varma & Zisserman 2005), who use k-means to construct the codebook whose elements are called textons. Codebook based quantization of signed differences of neighboring pixels (which correspond to local derivative filter outputs) was presented in (Ojala *et al.* 2001).

When comparing these two methods for quantizing the filter responses, one might expect that if the number of labels produced by the quantizers is kept roughly the same, the codebook based quantizer handles better the possible statistical dependencies between the filter responses. On the other hand, since codebook based quantization requires a search for the closest codeword at each pixel location, it is clearly slower than simple thresholding, even though a number of both exact and approximate techniques have been proposed for finding the nearest codeword without an exhaustive search through the codebook (Gray & Neuhoff 1998, Chávez *et al.* 2001).

2.3.3 The statistic

The last part of the FLS framework is the statistic, that is a function aiming to extract information about the distribution of labels obtained in the previous part.

The most common approach is to simply resort to the histogram of labels over the whole image (Eq. (5)), which then approximates the joint distribution of filter responses.

Also other statistics, for example co-occurrence matrices, could be used. Furthermore, a variety of methods have been presented to model directly the distribution of filter responses instead of the distribution of labels (Manjunath & Ma 1996, Rikert *et al.* 1999, Schmid 2001). In the following, we however concentrate on histogram based statistics.

The histogram of labels over the whole image is not a valid model if the input texture or input image cannot be assumed to be stationary, that is, if the distribution of filter responses is not independent of image coordinates. A typical situation for this to happen is a registered face image the analysis of which we discuss in Chapter 4. For this case, we present two approaches in Section 4.2 which estimate the distribution conditional on the image coordinates.

Finally, it can be observed that when suitable filter bank and quantization methods have been chosen, certain invariance properties can be achieved in the computation of statistics. The primary statistic of choice, the histogram, is naturally invariant to translations, but in Section 3.3.3 we introduce an LBP histogram Fourier statistic that, in addition to being translation invariant, is also invariant to rotations of the input image.

3 Texture analysis

Texture is a fundamental property of natural images, and thus it is of much interest in the fields of computer vision and computer graphics. There is, however, no generally accepted formal definition for that phenomenon. Mäenpää (2003) compares different attempts to define texture, but concludes: "Unfortunately, no one has so far been able to define digital texture in mathematical terms. It is doubtful whether anyone ever will.". One of the more recent definitions can be found in (Petrou & Sevilla 2006), where texture is defined to be "variation of data at scales smaller than the scales of interest"[3].

Chantler and Van Gool define texture analysis as image analysis producing measurements of the texture. These measurements may be low-level, such as statistics of local appearance, or a result of higher level processing, such as segmentation of an image into different regions or the class of the texture present in an image. (Chantler & Van Gool 2005). For general introductions to texture analysis, see (Tuceryan & Jain 1999, Petrou & Sevilla 2006, Mirmehdi *et al.* 2008).

3.1 Texture analysis problems

When applying texture analysis to a real life application, depending on the measurements required, different subproblems of texture analysis are encountered. Tuceryan & Jain (1999) list four texture analysis problems: texture classification, texture segmentation, texture synthesis and shape from texture. In (Petrou & Sevilla 2006), the list of texture analysis problems includes texture classification, texture segmentation and texture defect detection. In the following, these problems are discussed in more detail.

3.1.1 Texture classification

In texture classification, the aim is to assign an unseen texture sample into one of predefined classes. The assignment is done based on rules which are typically derived automatically from a training set consisting of texture samples with known classes.

Given a segmented texture image to be classified, the two critical components are

[3] In the preface of the book, they propose an alternative definition: "texture is what makes life beautiful; texture is what makes life interesting and texture is what makes life possible"

the feature extractor and the classification algorithm. A review of approaches to texture feature extraction is presented in Section 3.2. For a general introduction and review of statistical pattern classifiers, refer to, (Duda *et al.* 2001, Bishop 2006), for example, or for a critical view of the progress in classifier research, see (Hand 2006).

In texture classification, the nearest neighbor or k nearest neighbor classifier with different distance measures is a common choice (e.g. Ojala *et al.* 1996, Varma & Zisserman 2005)

Recently, the Support Vector Machine (Vapnik 1998) has gained interest, and it has been reported to outperform nearest neighbor classifier in texture classification by Hayman *et al.* (2004) and Zhang *et al.* (2007b), for example.

The problem of texture retrieval is to some extent related to texture classification. In essence, texture retrieval is content based image retrieval applied to texture images. Thus, the aim is to retrieve from a database as many samples of a requested texture as possible. However, texture is more often used as an additional feature for general image retrieval.

3.1.2 Texture segmentation

In texture segmentation, the goal is to divide an image into coherent regions using texture information. In supervised texture segmentation, the system has models of textures to be encountered in the images to be segmented. Unsupervised texture segmentation, on the other hand, aims at dividing an image into regions of similar texture without a priori information about the different textures. Image segmentation, even though an ill posed problem, has several practical applications and texture has proven to be a useful cue in segmentation.

Texture defect detection is a subproblem of texture segmentation, and it is commonly encountered in visual inspection. In that problem, we have a model of “acceptable” texture, and the goal is to analyze texture image to find defects, which are local events that derive from the model.

3.1.3 Shape from texture

The shape from texture problem deals with inferring the three dimensional (3D) shape of an object from its image. There is evidence that texture is an important cue in 3D shape perception by humans (e.g. Todd 2004). In computer vision research, one of the most recently proposed strategies for shape from texture is to model the deformation of individual texture elements as proposed by Lobay & Forsyth (2006). They do, however, point out in their paper that "applications for shape from texture have been largely absent".

3.1.4 Texture synthesis

Given a sample texture, the goal of texture synthesis is to synthesize more samples of perceptually similar texture. After several years of research aiming at texture synthesis by enforcing statistical constraints on output images (Popat & Picard 1993, Bonet & Viola 1997, Zhu *et al.* 1998, Portilla & Simoncelli 2000), currently the best results seem to be produced by image patch based approaches, first suggested in texture synthesis by Efros & Freeman (2001).

3.2 Texture description

Different ways of grouping texture models have been discussed in the texture analysis literature. Such taxonomies are seldom exhaustive, as some models may have properties from several model groups and some models seem to belong to none. These groupings still are useful in helping to understand the variety of different models.

In their survey, Tuceryan & Jain (1999) divide texture models into four groups: statistical methods such as co-occurrence matrices (Haralick *et al.* 1973) and autocorrelation features; geometrical methods based on analysis of geometric properties of texture primitives; model based methods aiming to provide a model that can be used to both describe and to synthesize a texture; and signal processing methods, typically using some energy measurements from filtered texture images as the texture descriptor.

The methods discussed in Section 2.2 can be considered as general image descriptors, but most of the global descriptors presented there actually originate from the texture analysis field. That is, they were originally proposed for texture classification or segmentation.

The use of Gabor filters in texture analysis dates back to the 1980's. Turner (1986) investigated the discriminability of different textures by Gabor filter responses. Later, Bovik *et al.* (1990) and Jain & Farrokhnia (1991) applied Gabor filters for unsupervised texture segmentation. From there, a wide variety of research using Gabor filters in different texture analysis problems has been spurred on. One of the best known Gabor based descriptors is that developed by Manjunath & Ma (1996) where texture features are obtained through computing means and standard deviations of Gabor filter responses over a texture image. For a brief survey and comparison of different approaches to Gabor filter based texture description, see (Grigorescu *et al.* 2002).

One of the more recently discovered issues in texture description is the problem arising from three dimensional surfaces. The perceived texture from these surfaces is not due to changes in surface albedo alone, but also self-shadowing and possible self-occlusions caused by small scale variations in the surface shape. This causes the variations due to illumination or viewing angle changes to be complex, and thus difficult to be modeled or handled by a texture descriptor.

Resorting to clustering of filter outputs, Leung & Malik (2001) proposed a 3D texture descriptor by precisely registering a set of images of a texture and convolving each registered image with a filter bank. Next, so called 3D textons could be formed by clustering the vectors consisting of responses of each filter at the same location of each registered image. The histogram of 3D textons was then used as a descriptor.

A recently proposed strategy for texture description is to take the local patch based representation (c.f. Section 2.2.3) and apply it to texture description. This approach was first proposed by Lazebnik *et al.* (2005), who presented a texture representation method using interest region detectors and clustering of the local descriptors computed within these regions.

3.3 Extensions of the Local Binary Pattern operator

This thesis presents three extensions related to the use of the local binary pattern operator in texture analysis. In the original papers, they were derived for the local binary pattern operator, but when viewed within the FLS framework, it becomes apparent that these extensions are more general and can be used also in combination with other filter banks.

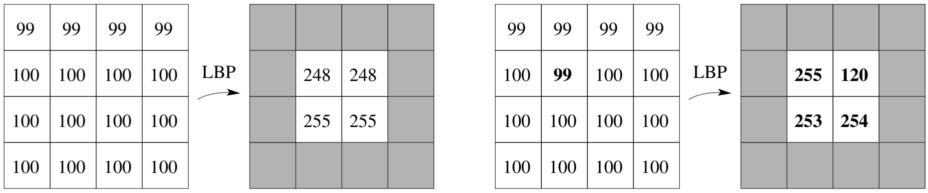

Fig 6. An example in which a small change of input gray values causes a large change in the resulting local binary pattern codes.

The first two of these extensions aim at increasing the robustness of the operator. Figure 6 presents a scenario where one gray value in a 4×4 region in an image is decreased by one, and this results in all the LBP labels in the region changing. The two proposed extensions, soft local binary patterns and Bayesian local binary patterns, make the operator more robust to changes of this type.

The third extension concerns the rotation invariance of LBP description.

3.3.1 Soft Local Binary Patterns

A drawback of the local binary pattern method, as well as all local descriptors that apply vector quantization, is that they are not robust in the sense that a small change in the input image would always cause only a small change in the output. To increase the robustness of the operator, Paper II proposes replacing the thresholding function $s_t(z)$ (Eq. (7)) with the following two fuzzy membership functions:

$$f_{1,d}(z) = \begin{cases} 0, & z < -d \\ 0.5 + 0.5\frac{z}{d}, & -d \leq z \leq d \\ 1, & z > d. \end{cases} \tag{10}$$

$$f_{0,d}(z) = 1 - f_{1,d}(z). \tag{11}$$

The parameter d controls the amount of fuzzification the function performs. These functions are plotted in Fig. 7.

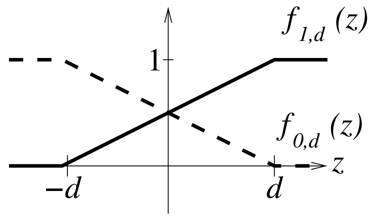

Fig 7. The functions $f_{0,d}(z)$ and $f_{1,d}(z)$.

As the number of filters applied to produce vector valued image $\mathbf{I}_f(x,y)$ (Eq. (3)) is P, the resulting histogram has bins numbered $0 \cdots 2^P - 1$. Now the contribution of a single pixel (x,y) to bin i of the histogram is

$$\begin{aligned} SLBP(x,y,i) &= \prod_{p=0}^{P-1} [b_p(i) f_{1,d}(I_p(x,y)) \\ &+ (1 - b_p(i)) f_{0,d}(I_p(x,y))], \end{aligned} \tag{12}$$

where $b_p(i) \in \{0,1\}$ denotes the numerical value of the p-th bit of binary representation of i. The complete soft histogram H_{SLBP} is computed by summing the contributions of all the pixels in the input image:

$$H_{SLBP}(i) = \sum_{x,y} SLBP(x,y,i), i = 0, \ldots, 2^P - 1. \tag{13}$$

When using the original LBP operator, one pixel always contributes to one bin of the histogram, but in the case of the soft histogram version, one pixel typically contributes to more than one bin. However, the sum of contributions of the pixel to all bins is always 1.

When d is set to 0, the resulting fuzzy membership functions $f_{0,0}(z)$ and $f_{1,0}(z)$ are almost equal to the thresholding function $s_t(z)$ with $t = 0$, the difference being that $f_{0,0}(0) = f_{1,0}(0) = 0.5$ whereas $s_0(0) = 1$. Because of this, the soft histogram representation differs from that of the original LBP, even when $d = 0$.

A possible problem of the soft histogram extension is the loss of invariance to monotonic gray-scale changes. The original LBP operator is invariant to monotonic gray-scale changes in the images. This is due to the thresholding operation and as thresholding is replaced by a fuzzy membership function, the resulting histogram is no longer completely invariant to such changes. However, as the operator only considers differences of gray values, adding a constant to the gray values does not change the soft histogram representation. Moreover, because of the robustness property induced by the fuzzy membership function, a small change of the gray scale causes only a small change in the operator output.

Another possible weakness of the soft histogram representation in comparison with the original LBP is the increased computational complexity. When computing the original LBP histogram, only one bin needs to be updated at each pixel location. In the case of the soft histogram, the contributions $SLBP(x,y,i)$ need to be computed for each of the 2^P bins if no optimizations are utilized.

Using soft histograms for image description has been previously introduced by Siggelgov (2002). The soft histograms for local binary patterns outlined here can be seen as a P-dimensional variant of Siggelkov's soft histograms, where each dimension, corresponding to one sampling point in the neighborhood, contains two bins ("greater than center pixel" and "less than center pixel"). Also, independently of this work, a similar extension of the texture spectrum operator was recently proposed by Barcelo *et al.* (2007). Soft local binary patterns have been used in medical image analysis in (Keramidas *et al.* 2008) and (Iakovidis *et al.* 2008).

3.3.2 Bayesian Local Binary Patterns

Paper III takes a different approach to considering the stochastic nature of image formation, aiming at reducing the sensitivity of the image descriptor to illumination changes and noise. As a result, a novel Bayesian LBP operator is proposed.

The Bayesian LBP models the label acquirement from filter responses as a stochastic process, and embeds a Markov Random Field in the label space. The labeling procedure is then considered as the maximum a posteriori estimation of the labels.

Let us denote the filtered vector valued image (Eq. (3)) with $\mathbf{I}_f$ and the label image with I_{lab}. Furthermore, let the labeling have a prior probability $P(I_{lab})$ and the filtered image have probability density $P(\mathbf{I}_f|I_{lab})$, given a certain labeling I_{lab}. Now the BLBP labels are obtained as the maximum a posteriori estimate for I_{lab}, which is

$$I_{BLBP} = I_{lab}^{MAP} = \arg\max_{I_{lab}} P(\mathbf{I}_f|I_{lab})P(I_{lab})$$

or equivalently

$$I_{BLBP} = \arg\min_{I_{lab}} -\log\left(P(\mathbf{I}_f|I_{lab})\right) - \log\left(P(I_{lab})\right). \tag{14}$$

Here, $P(I_{lab})$ is also called the smoothing term, and $P(\mathbf{I}_f|I_{lab})$ is the likelihood term.

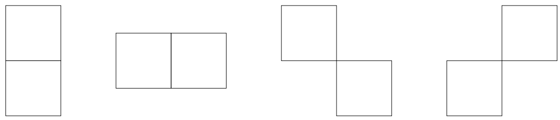

Fig 8. The clique set used with Bayesian local binary patterns. Revised from [Paper III] ©2008 IEEE.

One of the advantages of BLBP is its capability of adopting different likelihood and smoothing terms. In Paper III, the prior knowledge of the spatial consistency of the label space is described by an MRF and Potts model as with

$$P(I_{lab}) \propto \exp\left(\lambda \sum_{c\in C} V_c(I_{lab})\right). \tag{15}$$

The summation is over cliques C in the 8-neighbor clique set (See Fig. 8), λ is a smoothing parameter, and

$$\begin{aligned} V_c(I_{lab}) &= V_c(I_{lab}(x_{c1},y_{c1}), I_{lab}(x_{c2},y_{c2})) \\ &= \begin{cases} 1, & \text{when } I_{lab}(x_{c1},y_{c1}) = I_{lab}(x_{c2},y_{c2}) \\ 0, & \text{when } I_{lab}(x_{c1},y_{c1}) \neq I_{lab}(x_{c2},y_{c2}) \end{cases} \end{aligned} \tag{16}$$

in which $I_{lab}(x_{c1},y_{c1})$ and $I_{lab}(x_{c2},y_{c2})$ are the labels of sites (x_{c1},y_{c1}) and (x_{c2},y_{c2}) in clique c.

The likelihood term is modeled using the logsig function, which means that the distance of filter response from zero increases the likelihood to threshold 1 or 0. Assuming the elements $I_p(x,y)$ of the filter response vector $\mathbf{I}_f(x,y)$ are conditionally independent, the likelihood function can be written as

$$P(\mathbf{I}_f(x,y)|I_{lab}(x,y)) = \prod_{p=1}^{P} \frac{1}{1+\exp(\alpha I_p(x,y)(-1)^{b_p(I_{lab}(x,y))})}, \tag{17}$$

in which $b_p(I_{lab}(x,y))$ is the p-th bit of binary representation of the label at location (x,y), and α is a parameter of the likelihood function.

To obtain labeling I_{lab} which minimizes (14), the Graph Cut algorithm (Boykov & Kolmogorov 2004) is applied.

Considering the formulation of BLBP in (14), it can be noticed that the original LBP can be derived from BLBP by ignoring the smoothing term, i.e. taking the maximum likelihood (ML) estimate

$$I_{LBP} = I_{lab}^{ML} = \arg\min_{I_{lab}} -\log\left(P(\mathbf{I}_f|I_{lab})\right). \tag{18}$$

3.3.3 Local Binary Pattern histogram Fourier features

In Paper IV, a rotation invariant descriptor computed from uniform LBP pattern histograms was derived.

Let us denote a specific uniform LBP pattern with $U_P(n,r)$. The pair (n,r) specifies a uniform pattern so that n is the number of 1-bits in the pattern (corresponds to row number in Fig. 9) and r is the rotation of the pattern (column number in Fig. 9).

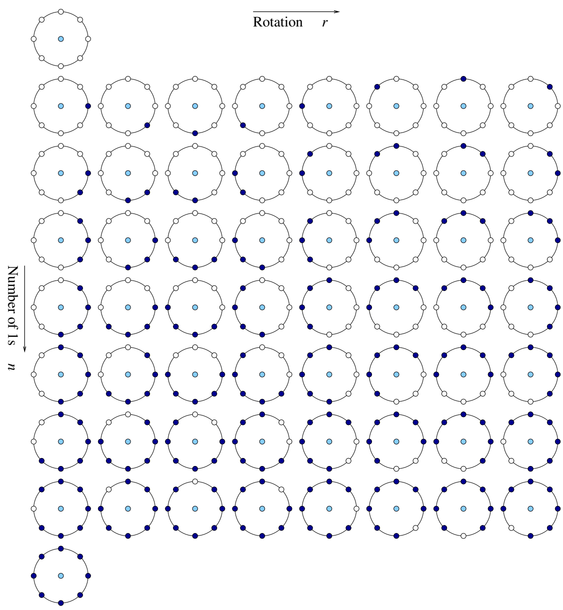

Fig 9. The 58 different uniform patterns in (8,R) neighborhood. [Paper IV] ©2009 Springer.

Now if the neighborhood has P sampling points, n gets values from 0 to $P+1$, where $n=P+1$ is the special label marking all the non-uniform patterns. Furthermore, when $1 \leq n \leq P-1$, the rotation of the pattern is in the range $0 \leq r \leq P-1$.

Let $I^{\alpha^\circ}(x,y)$ denote the rotation of image $I(x,y)$ by α degrees. Under this rotation, point (x,y) is rotated to location (x',y'). If we place a circular sampling neighborhood on points $I(x,y)$ and $I^{\alpha^\circ}(x',y')$, we observe that it also rotates by α°. See Fig. 10.

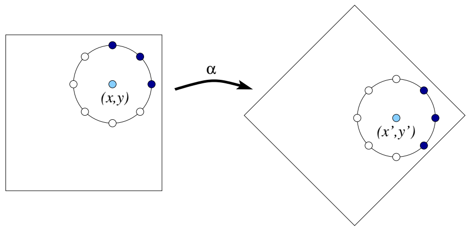

Fig 10. Effect of image rotation on points in circular neighborhoods. [Paper IV] ©2009 Springer.

If the rotations are limited to integer multiples of the angle between two sampling points, i.e. $\alpha = a\frac{360^\circ}{P}, a = 0,1,\ldots,P-1$, this rotates the sampling neighborhood by exactly a discrete steps. Therefore the uniform pattern $U_P(n,r)$ at point (x,y) is replaced by uniform pattern $U_P(n, r+a \mod P)$ at point (x',y') of the rotated image.

Now consider the uniform LBP histograms $h_I(U_P(n,r))$. The histogram value h_I at bin $U_P(n,r)$ is the number of occurrences of uniform pattern $U_P(n,r)$ in image I.

If the image I is rotated by $\alpha = a\frac{360^\circ}{P}$, based on the reasoning above, this rotation of the input image causes a cyclic shift in the histogram along each of the rows,

$$h_{I^{\alpha^\circ}}(U_P(n,r+a)) = h_I(U_P(n,r)) \tag{19}$$

For example, in the case of 8 neighbor LBP, when the input image is rotated by 45°, the value from histogram bin $U_8(1,0) = 000000001b$ moves to bin $U_8(1,1) = 00000010b$, the value from bin $U_8(1,1)$ to bin $U_8(1,2)$, etc.

Based on this observation, we propose a class of features that are invariant to rotation of the input image, namely such features, computed along the input histogram rows, that are invariant to cyclic shifts.

We use the Discrete Fourier Transform to construct these features. Let $H(n,\cdot)$ be the DFT of nth row of the histogram $h_I(U_P(n,r))$, i.e.

$$H(n,u) = \sum_{r=0}^{P-1} h_I(U_P(n,r))e^{-i2\pi ur/P}. \tag{20}$$

Now for DFT it holds that a cyclic shift of the input vector causes a phase shift in the DFT coefficients. If $h'(U_P(n,r)) = h(U_P(n,r-a))$, then

$$H'(n,u) = H(n,u)e^{-i2\pi ua/P}, \tag{21}$$

and therefore, with any $1 \leq n_1, n_2 \leq P-1$,

$$H'(n_1,u)\overline{H'(n_2,u)} = H(n_1,u)e^{-i2\pi ua/P}\overline{H(n_2,u)}e^{i2\pi ua/P} = H(n_1,u)\overline{H(n_2,u)}, \tag{22}$$

where $\overline{H(n_2,u)}$ denotes the complex conjugate of $H(n_2,u)$.

This shows that with any $1 \leq n_1, n_2 \leq P-1$ and $0 \leq u \leq P-1$, the features

$$\text{LBP}^{u2}\text{-HF}(n_1,n_2,u) = H(n_1,u)\overline{H(n_2,u)}, \tag{23}$$

are invariant to cyclic shifts of the rows of $h_I(U_P(n,r))$ and consequently, they are invariant also to rotations of the input image $I(x,y)$. The Fourier magnitude spectrum

$$|H(n,u)| = \sqrt{H(n,u)\overline{H(n,u)}} \tag{24}$$

can be considered a special case of these features. Furthermore it should be noted that the Fourier magnitude spectrum contains LBP^{riu2} features as a subset, since

$$|H(n,0)| = \sum_{r=0}^{P-1} h_I(U_P(n,r)) = h_{\text{LBP}^{riu2}}(n). \tag{25}$$

An illustration of these features is in Fig. 11

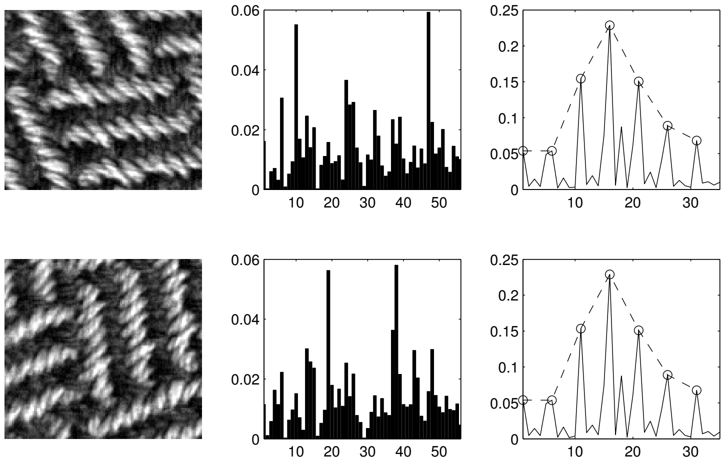

Fig 11. 1st column: Texture image at orientations $0°$ and $90°$. 2nd column: bins 1–56 of the corresponding LBP^{u2} histograms. 3rd column: Rotation invariant features $|H(n,u)|, 1 \leq n \leq 7, 0 \leq u \leq 5,$ (solid line) and LBP^{riu2} (circles, dashed line). Note that the LBP^{u2} histograms for the two images are markedly different, but the $|H(n,u)|$ features are nearly equal. [Paper IV] ©2009 Springer.

In preliminary tests for Paper IV, the Fourier magnitude spectrum was found to give the most consistent performance over the family of different possible features (Eq. (23)). Therefore, an LBP-HF feature vector consisting of three LBP histogram values (all zeros, all ones, non-uniform) and Fourier magnitude spectrum values was defined. The feature vectors have the following form:

$$
\begin{aligned}
fv_{\text{LBP-HF}} = \quad & [|H(1,0)|, \ldots, |H(1,P/2)|, \\
& \ldots, \\
& |H(P-1,0)|, \ldots, |H(P-1,P/2)|, \\
& h(U_P(0,0)), h(U_P(P,0)), h(U_P(P+1,0))]_{1\times((P-1)(P/2+1)+3)}.
\end{aligned}
$$

3.4 Texture analysis experiments

The results from experimental validation of the methods presented in texture analysis is in the following. For quick reference, the numerical results are also summarized in Table 1.

3.4.1 Choice of filter bank and quantization method

Paper I tested the filter bank and quantization framework in the material categorization task using the KTH-TIPS2 database[4]. The aim of the experiments was to validate the applicability of the framework, and systematically explore the relative descriptiveness of the different filter banks and vector quantization methods.

The filter banks that were included in the tests were local derivative filters, two different banks of Gabor filters and Maximum Response 8 (MR8) filters (Varma & Zisserman 2004). The local derivative filter bank was chosen to match the $\mathrm{LBP}_{8,1}$ operator which resulted in 8 filters (see Fig. 5). Two very different types of Gabor filter banks were tested, one with only 1 scale and 4 orientations and small spatial support (7×7), and another one with 4 scales and 6 orientations and larger spatial support (49×49).

In Paper I, the main interest was to examine the relative descriptiveness of different setups of the filter bank based texture descriptors. To facilitate this task, we chose a very challenging test setup that resembles the most difficult setup used in (Caputo *et al.* 2005). Using each of the descriptors to be tested, a nearest neighbor classifier using χ^2 distance was trained with one sample per material category. The remaining images were used for testing. This was repeated with 10000 random combinations as training and testing data, and the mean and standard deviations over the permutations were used to assess the performance.

The first experiment tested the effect of the codebook size when applying codebook based quantization. When testing codebooks of size of 16–256 vectors, it was noted that using a larger codebook enhances the texture categorization rate, but the selection of the filter bank is a clearly more dominant factor than the codebook size. For example, local derivative filters achieve a higher categorization rate with the smallest codebook size than the MR8 filters with any codebook size.

The second experiment involved thresholding quantization. It was noticed that non-zero threshold values provide the best results in texture categorization with Gabor(1,4) and local derivative filters, but in general the performance differences caused by changing the threshold are small.

Figure 12 shows the texture categorization rates using thresholding based quantization with $t = 0$, and codebook based quantization with a codebook size of 256. Codebook based quantization yields a slightly worse categorization rate than thresholding

[4]http://www.nada.kth.se/cvap/databases/kth-tips/

when using local derivative filters, but the difference is smaller than the standard deviation in the rates. With the Gabor(1,4) filter bank, thresholding performs worse than codebook based quantization, but interestingly with MR8 filters, thresholding yields a better rate. The local derivative filters give the best categorization rate over the tested filter sets with both quantization functions.

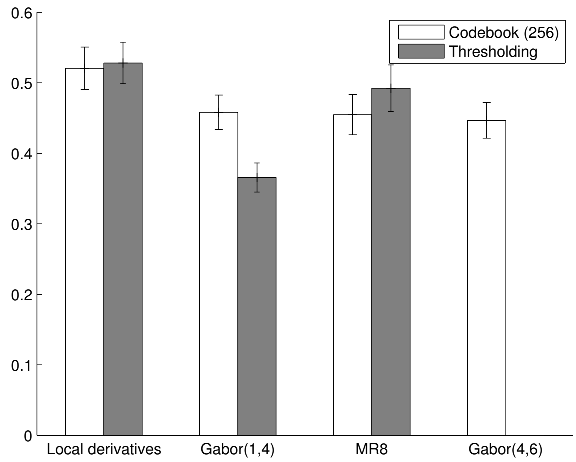

Fig 12. The KTH-TIPS2 categorization rates. [Paper I] ©2009 Elsevier.

The third experiment of Paper I tested whether it is possible to select a representative subset of filters from a large filter bank for thresholding based quantization. The number of labels produced by the quantizer is 2^P in which P is the number of filters, which means that the length of the label histograms grows exponentially with respect to the number of filters. Thus a small filter bank is desirable for thresholding quantization.

The Sequential Floating Forward Selection (SFFS) (Pudil *et al.* 1994) algorithm was used to select a maximum of 8 filters from a larger filter bank. The optimization criterion was the recognition rate over a training set. Two different initial filter banks were tested. First, 8 filters were selected from the 48 filters in the Gabor(4,6) filter bank. In the material categorization problem, this approach yielded a categorization rate of only 0.295.

Another tested initial filter bank for the SFFS, motivated by the findings that LBP and Gabor filter based information might be complementary (Zhang *et al.* 2005, Yan *et al.* 2007), was the union of local derivative filters and Gabor filters with small spatial support. In material categorization experiments, this resulted in a set of 6 local derivative and 2 Gabor filters, and the resulting filter bank reached a categorization rate of 0.544 which is significantly higher than the rate for the Gabor filters in the initial filter bank (0.366), and slightly higher than the rate of the local derivative filter bank (0.528).

3.4.2 Soft and Bayesian LBP

The performance of soft histograms was tested in Paper II using the supervised texture classification test images by Randen & Husøy (1999). In the first experiment, the 12 images in the test set were segmented using $\mathrm{LBP}_{8,1}$ with thresholding (Eq. (5)) and with soft histograms (Eq. (13)). The mean error rates were 0.123 and 0.108 for thresholding and soft histograms, respectively.

The second experiment tested the robustness of the methods to noise. Additive white Gaussian noise was added to the test images, and classification was performed again using the same operator as in the first experiment. Now, the mean error rates were 0.529 for thresholding, and 0.458 for soft histograms.

Examining the effect of fuzzification on the classification error, it can be noted that in both cases, using fuzzy histograms yields better performance than traditional hard histograms. The actual size of the decrease in the error rate depends on the test images.

In Paper III, Bayesian LBP was compared to the original LBP and its derivative Local Ternary Patterns (Tan & Triggs 2007a) and Multi-Block LBP (Liao *et al.* 2007). The comparison was made by applying unsupervised texture retrieval scheme using images from the Brodatz set. In these experiments, the Bayesian LBP gave constantly the best average recall rates, followed by the original LBP and then MB-LBP and LTP. However, the differences were not very large, as for example the average recall rate at 8 retrievals was 0.81 for BLBP, 0.79 for LBP, 0.78 for MB-LBP and 0.77 for LTP.

3.4.3 Rotation invariance

In Paper IV, the performance of the LBP-HF descriptor was tested in two texture related scenarios: texture classification and material categorization. Additionally, experiments in face recognition were also performed; see the original paper for these results.

The proposed rotation invariant LBP-HF features were compared to non-invariant LBP^{u2} and the older rotation invariant version LBP^{riu2}. In the texture classification and material categorization experiments, the MR8 descriptor (Varma & Zisserman 2005) was used as an additional control method. The experiments were made following the setup of (Ojala *et al.* 2002b) for nonparametric texture classification.

In the first experiment, we used the *Outex_TC_0012* (Ojala *et al.* 2002a) test set, intended for testing rotation invariant texture classification methods. The experimental results showed that both rotation invariant features provide better classification rates than non-invariant features. The performance of LBP-HF features is clearly higher than that of MR8 and LBP^{riu2}. This can be observed on all tested scales, but the difference between LBP-HF and LBP^{riu2} is particularly large on the smallest scale $(8,1)$. Over all the tested scales, the best recognition rates are 0.595 for LBP^{u2}, 0.883 for LBP^{riu2}, and 0.925 for LBP-HF. The MR8 descriptor achieved a recognition rate of 0.761.

In the second experiment, material categorization tests using the KTH-TIPS2 database and setup of Paper I were performed. In these experiments, LBP-HF reaches, or with most scales even exceeds the performance of LBP^{u2}. The performance of LBP^{riu2} is consistently lower than that of the other two. In this experiment, the best correct categorization rates were 0.542 for LBP^{u2}, 0.514 for LBP^{riu2}, and 0.546 for LBP-HF. The MR8 descriptor gives the lowest recognition rate, 0.455. It should be noted that the categorization rate of LBP-HF features is higher than with any of the setups of Paper I.

The reason for LBP-HF not performing significantly better than non-invariant LBP in material categorization is most likely that different orientations are present in the training data, so rotational invariance does not benefit much here. Unlike with LBP^{riu2}, no information is lost either, but a slight improvement over the non-invariant descriptor is achieved instead.

Table 1. A summary of results from the original papers in the texture analysis experiments. The results are not comparable between papers due to the different datasets used.

Paper I. Texture categorization rates on the KTH-TIPS2 dataset using different filter banks and quantization methods.

	Codebook	*Thresholding*
Local derivative filters	0.521	0.528
Gabor(1,4)	0.458	0.366
Gabor(4,6)	0.447	-
MR8 filters	0.455	0.492
(SFFS) Local derivative + Gabor	-	0.544

Paper II. Correct classification rates in supervised texture segmentation using test images from (Randen & Husøy 1999) with and without additive Gaussian noise.

	No added noise	*With added noise*
Local binary patterns	0.877	0.471
Soft local binary patterns	0.892	0.542

Paper III. Average recall rate at 8 retrievals in texture retrieval using Brodatz texture images (111 classes, 9 samples per class).

Local binary patterns	0.792
Local ternary patterns	0.771
Multi-block local binary patterns	0.782
Bayesian binary patterns	0.806

Paper IV. Rotation invariant texture recognition rates on the *Outex_TC_0012* test set and texture categorization rates on the KTH-TIPS2 dataset.

	Outex_TC_0012	*KTH-TIPS2*
MR8 descriptor	0.761	0.455
LBP^{u2}	0.595	0.542
LBP^{riu2}	0.883	0.508
LBP-HF	0.925	0.546

3.5 Discussion and conclusion

In this chapter, we have studied the applicability of the FLS framework to texture description. The filter sets and vector quantization techniques for LBP, MR8 and Gabor filter based texture descriptors were compared.

The main result is the validity of the FLS framework. Experiments show that the effect of changing one of the components of the FLS chain while keeping the others can be measured. Furthermore, extensions and improvements of the texture description scheme were shown to be derived and analyzed within this framework.

In the material categorization task, the relative importance of the choice of filter set was found to be somewhat higher than the choice between thresholding or codebook quantization, or the number of codebook vectors. In the comparison, it was found that the local derivative filter responses are not only fastest to compute but also most descriptive in the texture categorization task. This result further attests to the previous findings that texture descriptors relying on small-scale pixel relations yield comparable or even superior results to those based on filters of larger spatial support (Ojala *et al.* 2002b, Varma & Zisserman 2003). The experiments on filter subset selection and Gabor filter responses showed that these filter sets may be complementary, and may yield better performance than either of the sets alone in texture description.

In the FLS framework, two alternative solutions to the labeling part were proposed for making the resulting descriptor more robust. The first approach, soft LBP, is based on softly assigning the vector of filter responses to several labels. Another method, Bayesian LBP, uses crisp assignment, but considers information from neighboring pixel labels to obtain more stable labelings.

The main finding in the experiments with Soft LBP is that when the images have a high amount of noise, lower error rates can be obtained by utilizing soft histograms. Furthermore, applying soft histograms seems to result in a slightly improved performance, even for good quality images, with only a low degree of camera noise.

The validity of the BLBP descriptor was experimentally tested in the texture retrieval setting using Brodatz images. It was shown that histograms of MAP labels in Bayesian LBP give better results in retrieval than the original LBP histograms, both with original and noisy images.

The main drawback of the two proposed labeling methods in comparison to the original LBP is the increased computational complexity. Additionally, the sensitivity to gray-scale changes may be increased.

The third contribution presented in this chapter relates to the computation of statistics, i.e. the last part of the FLS framework. It was shown that rotations of the input image cause cyclic shifts of the values in the uniform LBP histogram. Relying on this observation, we proposed discrete Fourier transform based features as a statistic to describe the texture. This method for constructing the LBP-HF statistics from uniform LBP histograms was shown to be invariant to rotations of input image. The main difference of the proposed invariant statistic from most other rotation invariant texture features is that the LBH-HF features are constructed to be invariant to global rotations, instead of local rotations at every pixel neighborhood.

In the experiments, it was shown that in addition to being rotation invariant, the proposed LBP-HF features retain the highly discriminative nature of LBP histograms. The LBP-HF descriptor was shown to outperform the MR8 descriptor and the non-invariant and earlier version of rotation invariant LBP in texture classification, material categorization and face recognition tests.

4 Face recognition

Face recognition, both by humans and computers is a widely studied issue, and the first reported attempts to automate face recognition date back to the 1970's. Having applications in content-based image retrieval, security, image and video compression, human computer interaction, etc, analysis of human faces is an intensively studied research area.

4.1 Literature review

Face recognition is a topic that generates much interest in computer vision research. Surveys of the research made in the field can be found in (Samal & Iyengar 1992, Chellappa *et al.* 1995, Zhao *et al.* 2003, Tan *et al.* 2006). The scope of this thesis is automatic face recognition from 2D visible light images. Closely related fields that however fall beyond the scope of this work include face recognition from 3D data (Bowyer *et al.* 2006) and from video (Tistarelli *et al.* 2009), as well as face recognition by humans (Sinha *et al.* 2006).

This section gives an overview of the research questions and proposed methods.

4.1.1 Face detection, alignment and preprocessing

In face image analysis systems, the first step after image acquisition, and possibly some preprocessing, is face detection in an image with the goal: "to determine whether or not there are any faces in the image and, if present, return the image location and extent of each face." (Yang *et al.* 2002). Much has happened in the field of face detection since the publication of the face detection literature review by Yang *et al.* (2002), but in terms of detection accuracy, some of the methods presented in this survey are still top-class.

One of the best performing detectors mentioned in the survey by Yang *et al.* (2002) is that developed by Schneiderman & Kanade (2000) which is based on wavelet transform and AdaBoost learning. The problem with this detector is its high computational cost.

The work by Viola & Jones (2001, 2004) can be considered somewhat of a breakthrough in face detection: the combination of an efficient computation of Haar-like features and a simple-to-complex cascade of AdaBoost based classifier result in a face detector capable of processing video in real time.

Li & Zhang (2004) developed the system of Viola and Jones further by replacing AdaBoost with FloatBoost, and developing the cascade of classifiers into a classifier pyramid which is able to detect faces at different poses in real time.

After face detection, a more precise alignment is often needed prior to extracting an actual face descriptor. One possible solution, taken by Li *et al.* (2007), for example, is to use an eye detector within the detected face area and then perform normalization of rotation and scaling by transforming the detected eye centers into fixed coordinates. A more sophisticated, but also computationally heavier, solution is, for example, the active appearance model introduced by Cootes *et al.* (2001).

In the research on illumination effects on the face appearance it has been concluded that illumination induces larger changes on the unprocessed gray-level image than differences between individuals (Adini *et al.* 1997). To compensate for these changes, several different image processing algorithms have been introduced. These typically aim to transform the face image into a canonical face image free of illumination artifacts prior to feature extraction. Experimental comparisons of preprocessing algorithms have been done by Short *et al.* (2004) and Tan & Triggs (2007a).

Many preprocessing algorithms proposed in the literature for illumination invariance try to extract the reflectance component free of effects caused by illumination. The Self-Quotient Image (SQI) by Wang *et al.* (2004) aims to solve the reflectance component of a face image by dividing the perceived image by an approximation of the lighting component. In the SQI model, the lighting component is approximated by a smoothed version of the input image.

In the local total variation model by Chen *et al.* (2006), the input image is decomposed into large and small-scale components, where the latter is assumed to contain mostly illumination invariant information. Anisotropic smoothing (Gross & Brajovic 2003), on the other hand, aims at finding a lighting component from the input image through a constrained optimization procedure.

In addition to making an experimental comparison of different illumination normalization methods, Tan & Triggs (2007a) propose a normalization procedure consisting of gamma correction, difference of Gaussian filtering and contrast equalization. This preprocessing was further developed by Holappa *et al.* (2008) by applying an optimization procedure to finding the filter coefficients.

4.1.2 Feature extraction from face images

The purpose of feature extraction is to represent the facial image in another, usually non-visual, way so that the new representation is more suitable for a classifier. In other words, the feature extraction process is a mapping from facial image space to feature space. Usually it is desirable that the dimensionality of the feature space is low, the features are robust to intra-class variation of facial images, and that the feature extractor is fast.

In 1995, Chellappa *et al.* divided face recognition into three classes: holistic methods, local feature based methods and hybrid methods. This division is still valid and often used, but currently the major trend in research seems to be the hybrid methods, in which the descriptor mostly extracts local appearance information from the face, but also encodes the global constitution in some manner.

Linear subspace methods

Linear subspace methods in face feature extraction assume that the facial image $\mathbf{i}$, with pixels reordered into a vector and mean having been subtracted, lies approximately on a linear subspace of the image space and therefore they can be approximated as a weighted sum of the basis vectors of the subspace. The face feature vector $\mathbf{f}$ is then extracted through a linear projection onto another space:

$$\mathbf{f} = \mathbf{W}\mathbf{i}. \tag{26}$$

Typically the dimension of $\mathbf{f}$ is much lower than of $\mathbf{i}$. The various linear subspace methods differ in how the basis vectors, i.e. rows of projection matrix $\mathbf{W}$ are obtained.

Linear subspace methods are well studied, and a very large number of papers discussing the issue has been published, even to the point that in an editorial of Pattern Recognition Letters, Duin *et al.* (2006) write: “Without doubt, a number of interesting studies and proposals have been published recently. However, many other papers have only a minor significance, or low citation value, as others have treated the same point.”

Principal Component Analysis (PCA) was the first one of the linear subspace methods to be applied to face recognition, and according to Zhao *et al.* (2003), it was “the first really successful demonstration of machine recognition of faces.” It was first applied to facial image representation by Sirovich & Kirby (1987), and later extended to face recognition by Kohonen (1988) and Turk & Pentland (1991). In face recognition

PCA is often referred to as the Eigenfaces method. In PCA, the basis of the subspace is obtained from the eigenvectors of the sample covariance matrix of the input (facial images), and as the eigenvectors corresponding to the largest eigenvalues are used as a basis, the resulting projection simultaneously maximizes the variance of projected data or minimizes the average projection cost (Bishop 2006).

Another classical linear projection method is Linear Discriminant Analysis or the Fisher Linear Discriminant, which aims at finding such projection directions that are useful for classifying. This is done by simultaneously maximising the between-class scatter and minimizing the within-class scatter (Duda *et al.* 2001). Belhumeur *et al.* (1997) applied LDA to facial images and experimentally showed that this leads to clearly better results than PCA. In later work it has been pointed out that LDA might not generalize well for persons who are not in the training set that was used for computing the projection matrix, but this can be helped by mapping the faces first into a lower-dimensional subspace with PCA and performing LDA in this subspace (Zhao *et al.* 1998). Furthermore, the training of LDA has been observed to be unstable when the number of training samples is low (Martínez & Kak 2001).

Another linear subspace method that has been applied to face recognition is Independent Component Analysis (ICA), which seeks basis vectors that are statistically independent under the assumption that the data is non-Gaussian (Hyvärinen *et al.* 2001). The application of ICA to face recognition has been studied by Bartlett *et al.* (2002).

The unified framework for subspace face recognition by Wang & Tang (2004) is a recent work which models the differences between facial images using three components, namely intrinsic difference (i.e. differences between people), transformation difference (differences between the images of one subject caused by illumination, expression, pose angle, etc) and noise. Based on this model, PCA, the Bayesian Intra/Extrapersonal model and LDA are combined to form a novel method for extracting the identity discriminating information in face images. The combination of these three subspace methods is shown to be more powerful in terms of recognition than any one of the methods alone.

Recently, interesting approaches that are similar in spirit to linear subspace methods but do not actually apply the linear projection of (26), have been proposed by Zhou *et al.* (2007) and Wright *et al.* (2009). Zhou *et al.* (2007) represent the face image as an illumination free linear combination of lighted albedo-shape matrices. The blending coefficients, corresponding to $\mathbf{f}$ of (26) and the lighting parameters are obtained through an interactive procedure. Wright *et al.* (2009), on the other hand, borrow the idea of compressed sensing from signal processing. They use the training samples as a possibly overcomplete basis for face images, and aim for such a sparse representation that only the coefficients corresponding to those basis vectors that represent the same individual as the probe image are non-zero.

Kernel based methods

Linear subspace methods, discussed above, assume that faces lie approximately on a low dimensional subspace of the image space, and use this assumption to extract features. Kernel methods take the opposite direction: they are based on an implicit non-linear projection to a higher dimensional space in which the classification or feature extraction problem is supposed to be easier to solve. Using the so called kernel trick, methods such as PCA or SVM can be applied in this implicit space without actually computing the projections.

Liu (2004) proposes combining Gabor filters and Kernel PCA for face recognition. The approach is similar to the Gabor-Fisher Classifier (Liu & Wechsler 2002) but Kernel PCA is used instead of a linear subspace projection. The Kernel PCA is a non-linear extension of the Principal Component Analysis that basically corresponds to mapping the data to a higher-dimensional space using nonlinear mapping and performing PCA in this space. In the paper (Liu 2004) it is shown that Gabor filtering and Kernel PCA together result in higher recognition rates than linear PCA on nonprocessed images or Gabor filtered images, or kernel PCA on nonprocessed images.

A similar, more recent work by Liu (2006) applies Kernel Fisher Analysis, a kernelized variant of the Fisher linear discriminant to Gabor filtered images. Tan & Triggs (2007b), on the other hand, use a similar kernel method named kernel discriminative common vectors (Cevikalp *et al.* 2006) to fuse Gabor and local binary pattern features, achieving clearly better results than with either representation alone.

Projection to higher-dimensional space to obtain better separability of classes is also deployed by support vector machines. Using support vector machines for face recognition has been studied by Guo *et al.* (2001) but in that work, only linear SVMs are used. The paper however shows that support vector machines can be used for the face recognition task.

In (Huang *et al.* 2003), support vector machines are used for component-based face recognition. The recognition system first uses linear SVMs to localize facial components, such as eyes and nose, and the histogram equalized gray values from these components are used as features for the polynomial kernel based support vector machine performing identification. This system uses the one vs. all approach to extend SVM to multi-class classification. The problem of obtaining enough training data is handled by using a 3D morphable model of faces to create artificial training images.

Local features

Local feature based methods represent the face as a compound of local features and possibly their spatial relations.

Modular eigenspaces (Pentland *et al.* 1994) and local feature analysis (Penev & Atick 1996) are similar to linear subspace methods discussed above, however extracting spatially localized information. The modular eigenspaces method brings the PCA towards feature-based face recognition by first finding facial components (eyes, nose and mouth) and then performing PCA on each of these components individually. Local Feature Analysis, on the other hand, uses projection basis vectors with local spatial support.

Elastic Bunch Graph Matching (EBGM), developed by Wiskott *et al.* (1997) from Dynamic Link Architecture representation (Lades *et al.* 1993) modelled by a facial graph consisting of edges and nodes positioned on facial landmarks such as pupils and corners of the mouth. Each edge is labeled with a distance vector and nodes are labeled with jets consisting of Gabor filter responses at the node locations.

3D face recognition

In the face recognition community, there is no consensus on whether 3D information from a face is more powerful in terms of recognition than a good quality 2D image. Acquiring real 3D data with a laser scanner or a similar device is not feasible in most scenarios and 3D face recognition from range data is outside the scope of this thesis.

However, the 3D shape and texture of a face can be estimated from single or multiple 2D images using, e.g. shape from shading. Morphable models of 3D faces (Blanz & Vetter 1999) have performed very well in estimating the 3D shape and texture of a face from a single image, and this methodology has received wide attention in face recognition.

When using morphable models for face recognition, one can take three different approaches. The morphable model can be fitted either on the gallery images, on the probe images, or on both. All three different approaches have been discussed in the literature. Blanz & Vetter (2003) fit a morphable model both on the probe and gallery images and use the model parameters as features for a classifier. Huang *et al.* (2003), take an image-based approach and fit a morphable model on gallery images and produce a set of artificial training images under varying poses and illumination conditions. They then use local parts of these training images to construct a Support Vector Machine based face recognition system. Blanz *et al.* (2005), on the other hand, fit a morphable model on the non-frontal probe images, and, based on the model, produce a frontal view of the face and use traditional appearance based face recognition methods for recognition.

4.1.3 Classification in face recognition

Depending on the underlying assumptions about the classes, different scenarios for face classification can be defined.

The problem of recognizing a person, given the assumption that the person is known by the system, is typically called face recognition, and it is probably the most common scenario in face recognition literature.

In real-life situations, however, it can be seldom assumed that all the persons to be encountered by the system have been enrolled in advance. This scenario is called full identification, open set face recognition or face recognition with a reject option. In this setting, the classifier must output either the identity of the person or, if the person is not known, a reject decision. Open set face recognition has been studied, e.g., by Li & Wechsler (2005), and in a more general setting, multi-class classifiers with a reject option by Tax & Duin (2008).

In the third type of classification problem, named face verification, the classifier needs to find if a person is the one they claim to be. In other words, this is a two-class classification or decision problem with the alternative decisions of accepting or rejecting the claimed identity.

4.2 Face description with distribution of filter responses

In this section, we present an approach to face feature extraction based on viewing this problem as one of texture description. This can be expressed conveniently in the FLS framework, though in the course of research, this framework was formulated only after the presentation of the first face description method.

Registered face images cannot be regarded as stationary texture, which inhibits the use of traditional texture descriptors as such in face description. The key observation in this study is that when this non-stationarity is taken into account by observing filter response distributions conditioned on spatial coordinates, very discriminative but still robust face descriptors can be achieved.

4.2.1 Joint distribution of filter responses for facial image representation

In papers V and VI, a local binary pattern based face descriptor was presented. In Paper I, this was generalized to different types of filters.

The face description procedure consists of using the texture descriptor to build several local descriptions of the face, and combining them into a global description. Instead of striving for an holistic description, this approach was motivated by two reasons. First, the local feature based or hybrid approaches to face recognition have been gaining interest (Penev & Atick 1996, Heisele *et al.* 2003), which is understandable given the limitations of the holistic representations. These local feature based and hybrid methods seem to be more robust against variations in pose or illumination than holistic methods. The second reason for selecting the local feature based approach is that trying to build a holistic description of a face using texture methods is not reasonable since texture descriptors tend to average over the image area. This is a desirable property for ordinary textures, because texture description should usually be invariant to translation, or even rotation of the texture and, especially for small repetitive textures, the small-scale relationships determine the appearance of the texture, and thus the large-scale relations do not contain useful information. For faces however, the situation is different; retaining the information about spatial relations is important.

This reasoning leads to the basic texture based face descriptor. The facial image is divided into local regions and texture descriptors are extracted from each region independently. The descriptors are then concatenated to form a global description of the face. See Figure 13 for an example of a facial image divided into rectangular regions.

Fig 13. A facial image divided into 7×7, 5×5 and 3×3 rectangular regions. [Paper VI] ©2006 IEEE.

The basic histogram can be extended into a spatially enhanced histogram which encodes both the appearance and the spatial relations of the facial regions. As the N_R facial regions have been determined, a histogram is computed independently within each of the N_R regions. The resulting N_R histograms are combined, yielding the spatially enhanced histogram. The spatially enhanced histogram has a size $M \times N_R$ where M is the length of a single LBP histogram. In the spatially enhanced histogram, we effectively have a description of the face on three different levels of locality: the LBP labels for the histogram contain information about the patterns on a pixel-level, the labels are summed over a small region to produce information on a regional level and the regional histograms are concatenated to build a global description of the face.

It should be noted that when using the histogram based methods, despite the examples in Figure 13, the local regions do not need to be rectangular. Neither do they need to be of the same size or shape, and they do not necessarily have to cover the whole image. For example, they could be circular regions located at the fiducial points, as in the EBGM method. It is also possible to have partially overlapping regions. If recognition of faces rotated in depth is considered, it may be useful to follow the procedure of Heisele *et al.* (2003) and automatically detect each region in the image instead of first detecting the face, and then using a fixed division into regions.

The idea of a spatially enhanced histogram can be exploited further when defining the distance measure. An indigenous property of the proposed face description method is that each element in the enhanced histogram corresponds to a certain small area of the face. Based on psychophysical findings, which indicate that some facial regions (such as eyes) play more important roles in human face recognition than other features (Zhao *et al.* 2003), it can be expected that in this method some of the facial regions contribute more than others in terms of extrapersonal variance. Utilizing this assumption, the regions can be weighted based on the importance of the information they contain. For example, the weighted χ^2 distance can be defined as

$$\chi_w^2(x,\xi) = \sum_{j,i} w(j) \frac{(x_{i,j} - \xi_{i,j})^2}{x_{i,j} + \xi_{i,j}}, \tag{27}$$

in which $\mathbf{x}$ and ξ are the normalized enhanced histograms to be compared, indices i and j refer to i-th bin in the histogram corresponding to the j-th local region and $w(j)$ is the weight for region j.

4.2.2 Kernel density estimation of local distributions

Paper VII proposes a different model for estimating local LBP distributions. This method is based on kernel density estimation in the *xy*-space. The paper also proposes the use of a support vector machine for combining the information coming from different parts of the face because such a learning method can be expected to result in higher performance than a simple summation.

Researchers applying LBP histograms for face description have used the block based model presented in Paper VI, either using a manually fixed grid (e.g. Hadid *et al.* 2004, Shan *et al.* 2009) or by finding the division and weights denoting the relative importance of each region through some optimization procedure such as AdaBoost (e.g. Yang & Ai 2007, Zhang *et al.* 2004). In some other works, the histogram based model has been omitted and LBP is used as preprocessing prior to classification (Heusch *et al.* 2006) or the query and gallery label images are compared directly using a Hausdorff like distance measure (Tan & Triggs 2007a).

This block based representation, however, suffers from two problems. Firstly, finding the optimal division of face image area into N_R regions is a non-trivial task. Secondly, if the values of the histogram are used as estimates of probability distribution of LBP labels for pixels within the corresponding region, it is questionable whether the estimate is reliable near the borders of the region.

To address these two problems of the block based approach, we propose replacing it with Kernel Density Estimation (KDE) (Scott 1992) in the *xy*-space. KDE allows computing an estimate of probability distribution of LBP labels at any image coordinates (x,y) as a normalized sum of kernels positioned at occurrences of each LBP label. Here we use the symmetric Gaussian kernel, so the kernel density estimate of LBP distribution at location (x,y) becomes

$$h^{(x,y)}(i) = n^{(x,y)} \sum_{s,t} \delta\left\{i, I_{lab}(s,t)\right\} \exp\left\{-\frac{(x-s)^2 + (y-t)^2}{2\sigma^2}\right\}, \tag{28}$$

in which δ is the Kronecker delta, and normalizing constants $n^{(x,y)}$ are set so that $\sum_i h^{(x,y)}(i) = 1$. The difference between block based and kernel density estimation of LBP local occurrence probability is illustrated in Fig. 14.

If there is more than one label image available, this means simply having more samples in the xy-space. In practice the resulting model can be computed by applying (28) to each of the images independently, and then averaging over the obtained values.

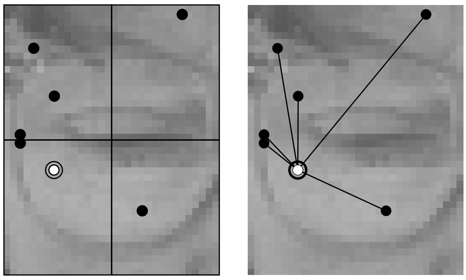

Fig 14. Comparison of block based and kernel density estimation of LBP histograms in xy-space. Left: In block based estimation the probability of occurrence of a specific LBP label in arbitrary coordinates (x,y) is the number of this label in the region to which (x,y) belongs divided by the total number of pixels in the region. Right: In kernel density estimation, the probability estimate is the sum of kernels positioned at occurrences of the label. [Paper VII] ©2009 Springer.

Paper VII follows the idea of Rodriguez & Marcel (2006) of a generative model for LBP based face verification. For that, both client and non-client models need to be estimated for each person known to the system. A client model is an estimated probability distribution of observing certain features given that they originate from this client, whereas a non-client model is the estimated distribution of features given that they do not originate from the client. Paper VII uses an LBP labeled face image as features, and the models are estimated using Eq. (28). A world model, i.e. a generic face model estimated from a large amount of training face images, is used as a common non-client model for all clients.

Using the estimated client and world models, Paper VII derives a log-likelihood ratio

$$\lambda_C(I_{lab}(x,y),x,y) = \log(\hat{h}_C^{(x,y)}(I_{lab}(x,y))) - \log(\hat{h}_W^{(x,y)}(I_{lab}(x,y))) \tag{29}$$

for observing the LBP label $I_{lab}(x,y)$ at location (x,y) given the estimated model $\hat{h}_C^{(x,y)}(\cdot)$ for client C and the world model $\hat{h}_W^{(x,y)}(\cdot)$.

Assuming independence of the labels at different locations, the decision rule for the face verification given LBP labeled input image I_{lab} and the claimed identity C then becomes

$$D(I_{lab},C) = \text{sign}\left\{\sum_{x,y}\lambda_C(I_{lab}(x,y),x,y) - \tau\right\}, \tag{30}$$

in which the decision threshold τ is learnt by minimizing the expected error over a development set.

In essence, the decision function in (30) is a sum of log likelihood ratios corresponding to individual pixels. As demonstrated in Paper VI, and some other studies applying LBP to face recognition (e.g. Zhang *et al.* 2004), different facial areas contribute unequally in terms of recognition. Therefore, Paper VII proposes replacing the sum in the decision function in (30) by a weighted sum, i.e. a general linear discriminant function, obtaining

$$D(I_{lab},C) = \text{sign}\left\{\sum_{x,y}w(x,y)\lambda_C(I_{lab}(x,y),x,y) - \tau\right\}. \tag{31}$$

A convenient way of computing the weights $w(x,y)$ and the threshold τ is to resort to the linear Support Vector Machine (Vapnik 1998). Previously SVMs have been proposed for information fusion for person verification by Ben-Yacoub *et al.* (1999), for example, who used SVM to combine information from face and voice modalities in multi-modal person verification. In face analysis, Heisele *et al.* (2007) used a support vector machine for fusing information from component-wise SVMs in face detection and recognition.

The linear SVM is obtained by finding such weights $w(x,y)$ that the minimum distance between the decision boundary and the training samples is maximized, i.e. it is a maximum margin classifier. The SVM can be seen as a structural risk minimization solution to the linear classification problem, and thus it is less prone to overtraining than some other learning methods. This is especially important in our case, since the input dimension is rather high with respect to the number of training samples. Despite this, in our experiments it was observed that the SVM did not show significant overfitting, but it performed well also on unseen data.

4.2.3 Blur insensitive face description

Paper VIII discusses the problem of face recognition from blurred images. Blur may be present in face images due to motion of the subject or the camera during the exposure, the camera not being in focus, or the low quality of the imaging device, such as an analog web camera.

Blur insensitive features are extracted using Local Phase Quantization (Ojansivu & Heikkilä 2008). This is based on the observation that when the point spread function of a linear, shift-invariant blur operator is centrally symmetric and its Fourier transform is positive at low frequencies, the Fourier phase at low frequencies is a blur invariant property.

In LPQ, the phase is examined in local $(2R+1)$-by-$(2R+1)$ neighborhoods $\mathscr{N}_{(x,y)}$ at each pixel position (x,y) of the image $I(x,y)$. These local spectra are computed using a short-term Fourier transform defined by

$$F(u,v,x,y) = \sum_{s=-R}^{R} \sum_{t=-R}^{R} I(x-s,y-t)e^{-j2\pi(us+vt)}. \tag{32}$$

Assume a set of $P/2$ frequencies $\{(u_n,v_n)\}_{n=1}^{P/2}$ at which the image is analyzed being given. Computing transform (32) at all pixel locations can be conveniently expressed in the FLS framework as convolutions of the input image with kernels $F_1,\ldots,F_P$ of the form

$$\begin{aligned} F_{2n-1}(x,y) &= \mathrm{real}\left(e^{-j2\pi(u_n x+v_n y)}\right), \\ F_{2n}(x,y) &= \mathrm{imag}\left(e^{-j2\pi(u_n x+v_n y)}\right), \\ &\quad \text{where } -R \leq x,y \leq R, 1 \leq n \leq P/2. \end{aligned} \tag{33}$$

All these kernels are separable, so the computation can be made even faster by doing simply 1-D convolutions for the rows and columns successively.

Following the procedure by Ojansivu & Heikkilä (2008), local Fourier coefficients are computed at four frequency points $(\beta,0)$; $(0,\beta)$; (β,β); $(\beta,-\beta)$, where β is a sufficiently small scalar to satisfy $H(u,v) > 0$ for all the four points, where H is the transfer function of the linear shift invariant blur operator. Hence we get at total of 8 real valued filters composed of the real and imaginary part of the complex short term Fourier transform.

The phase information in the Fourier coefficients is recorded by observing the signs of the real and imaginary parts of each component in $F(u,v,x,y)$. In the FLS framework, this is done by using a thresholding quantizer (Eq. (7)).

An interesting property of thresholding quantization, presented in the context of LPQ by Ojansivu & Heikkilä (2008), is that in quantization the information is maximally preserved if the samples to be quantized are statistically independent. Since the neighboring pixels of natural images are highly correlated, there is dependency between the elements in the vector valued image $\mathbf{I}_f(x,y)$ (Eq. (3)), which are then quantized. The situation can however be improved by adding a simple decorrelation to the labeling process prior to quantization.

Assume that the image function $I(x,y)$ is a result of a first-order Markov process, where the correlation coefficient between adjacent pixel values is ρ, and the variance of each sample is 1. As a result the covariance between positions (x_i,y_i) and (x_j,y_j) becomes $\sigma_{ij} = \rho^{\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}}$. With this information we can construct the covariance matrix $\mathbf{C}$ of the $(2R+1)^2$ samples in the neighborhood $\mathscr{N}_{(x,y)}$.

Rearrange the coefficients of filter kernels $F_1, F_2, \ldots, F_N$ into vectors so that vector $\mathbf{w}_{F_n}$ contains the coefficients of F_n and stack these into a matrix $\mathbf{W}$:

$$\mathbf{W} = [\mathbf{w}_{F_1}, \ldots \mathbf{w}_{F_N}]^T. \tag{34}$$

Now the covariance matrix $\mathbf{D}$ of elements of $\mathbf{I}_f(x,y)$ can be obtained from $\mathbf{D} = \mathbf{W}\mathbf{C}\mathbf{W}^T$. The covariance matrix $\mathbf{D}$ can then be used to derive a whitening transform $\mathbf{I}_w(x,y) = \mathbf{V}^T\mathbf{I}_f(x,y)$, where $\mathbf{V}$ is an orthogonal matrix derived from a singular value decomposition as $\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^T$.

The elements of $\mathbf{I}_w(x,y)$ are uncorrelated, which, assuming Gaussian distributions, is equal to independence. Using $\mathbf{I}_w(x,y)$ in the quantization instead of $\mathbf{I}_f(x,y)$ results in more information to be stored in most cases, even though the correlation model or the Gaussian assumption would not hold exactly. If the transfer function of the blur operator gets the same value at the chosen frequency points, it can be shown that the decorrelation does not affect the blur invariance. Note also that $\mathbf{V}$ can be solved in advance for a fixed ρ.

4.3 Face recognition experiments

In the following, experimental face recognition results from Papers I and V–VIII are reported. The numerical results are also summarized in Table 2.

4.3.1 Face recognition using texture descriptors

In Papers V and VI, the proposed face descriptor was tested on the FERET face dataset (Phillips *et al.* 2000) using the publicly available CSU face identification evaluation system by Beveridge *et al.* (2005). In the following, the mean recognition rates for different face recognition methods are reported. These were obtained by randomly permutating the gallery and probe image sets, and computing the mean correct recognition rate over permutations. A more detailed description of the experimental protocol as well as more experimental results can be found in Papers V and VI.

Paper V reports the extensive experiments that were done with regard to different possible parameters of the LBP based face description. Then LBP was compared to several publicly available benchmark algorithms, the PCA, LDA, Bayesian Intra/Extrapersonal Classifier (BIC) (Moghaddam & Pentland 1997) and EBGM. In this experiment it was shown that LBP yields clearly higher recognition rates than the control algorithms in all the FERET test sets. The obtained mean recognition rates were 0.81 for weighted LBP, 0.72 for PCA, 0.72 for BIC and 0.66 for EBGM.

In Paper VI, the LBP description was compared to a gray-level difference histogram (Weszka *et al.* 1976), a homogeneous texture descriptor (Manjunath *et al.* 2001), and a texton histogram, which is equivalent to the use of local derivative filters and codebook quantization in Paper I. No weighting of local regions was applied. The obtained mean recognition rates were 0.63 for the difference histogram, 0.62 for the homogeneous texture descriptor, 0.76 for the texton histogram and 0.76 for the LBP.

Finally, Paper VI presents an experiment evaluating the robustness of the proposed description to face localization errors. LBP, with two different window sizes, was compared to PCA. The experiments showed that when no localization error, or only a small error, is present, LBP with small local regions works well; but as the localization error increases, using larger local regions produces a better recognition rate. Most interestingly, the decrease of the recognition rate of the local region based methods was shown to be significantly slower than that of the PCA.

In Paper I, we performed a more thorough comparison of different filter banks and quantization methods, using the general framework of face description from Paper VI, and performing the tests using the CMU Pose, Illumination, and Expression (PIE) face dataset (Sim *et al.* 2003). The same filter banks and quantization methods as with texture experiments (Section 3.4.1) were tested.

First, the effect of codebook size was examined in codebook based quantization. The results were similar to those with texture. Increasing the codebook size from 16 to 256 did result in improvement in recognition rates, albeit very small, and the choice of filter set had a much more important role. Gabor filters with a large support area achieve better recognition rates than Gabor filters with only small spatial support or local derivatives. Each of these however performs better than the MR8 descriptor. A probable cause for this is the rotational invariance built into the MR8 descriptor that might actually lose some useful information.

The second experiment involved testing the effect of a threshold value in thresholding quantization. With all the filter sets, the best results are obtained with threshold $t = 0$. This effect is likely to be due to lighting effects, since as discussed in Section 2.3.2, using threshold $t = 0$ with filters having zero mean yields in a descriptor that is invariant to affine changes of gray values.

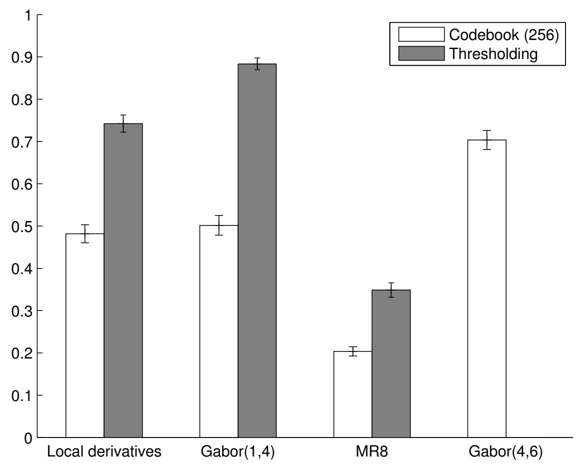

Fig 15. The CMU PIE recognition rates. [Paper I] ©2009 Elsevier.

Figure 15 shows the face recognition rates using thresholding based quantization with $t = 0$ and codebook based quantization with a codebook size of 256.

When testing the filter bank subset selection in the face recognition problem, very good results were obtained when selecting 8 filters from the 48 filters in the Gabor filter bank using the SFFS algorithm. The selected filter subset performed superbly on the test image set, achieving a recognition rate of 1.000. In experiments aimed at combining the local derivative filters and Gabor filters with small spatial support, performance gains in face recognition were not achieved.

4.3.2 Kernel density estimation

Paper VII tested the kernel density based approach in the face verification problem using the English section of the BANCA database (Bailly-Bailliére *et al.* 2003).

In the original paper, results on all BANCA protocols are reported. Probably the most representative of those is the pooled (P) protocol, in which data taken under controlled conditions is used for training, and all data in the database is used for testing; the results from the P protocol are therefore reported here. The proposed LBP kernel density estimation and the SVM weighting based verification method were compared to block based estimation of LBP distributions (Rodriguez & Marcel 2006).

In the BANCA P protocol, the block based estimation of LBP histograms by Rodriguez & Marcel (2006) yielded an error rate of 0.192, and by comparison, the best benchmark algorithm of that paper gave an error rate of 0.148. Our approach using kernel density estimation resulted in an error rate of 0.176 when uniform weights over the whole face were applied, and an error rate of 0.116 when the weights were optimized with a linear SVM. Finally, when applying the face preprocessing proposed by Tan & Triggs (2007a) prior to computing the LBP labels, the weighted KDE method obtained an error rate of 0.091.

4.3.3 Blur tolerant face recognition

The efficiency of the LPQ based blur tolerant face descriptor in the face recognition scenario was tested using the CMU PIE dataset with artificial blur, and the Face Recognition Grand Challenge (FRGC) experiment 1.0.4, in which many of the probe images contain significant out-of-focus blur.

In the experiments with the CMU PIE images, an LPQ based description was shown to produce a better recognition result than LBP, even with no blur (0.992 vs. 0.927). When adding artificial Gaussian blur with varying σ to the probe images, it was noticed that the LBP descriptor tolerates slight blur very well, but as blur increases from $\sigma = 0.5$, the recognition rate drops rapidly. At $\sigma = 2.0$, the achieved recognition rate is 0.701. Local phase quantization, on the other hand, tolerates increasing blur significantly better. At $\sigma = 1.0$, the recognition rate is 0.986, after which the rate starts to decrease slightly faster, but even at $\sigma = 2.0$ the recognition rate still is 0.935, higher than using LBP on images with no blur.

In the experiment with FRGC 1.0.4 images with no illumination normalizing preprocessing, LBP and LPQ gave recognition rates of 0.326 and 0.459, respectively. When illumination normalizing preprocessing (Tan & Triggs 2007a) was applied prior to computing the features, the resulting recognition rate was 0.643 for LBP and 0.745 for LPQ.

Table 2. A summary of results from the original papers in the face recognition experiments. The results are not comparable between papers due to the different datasets used.

Papers V, VI. Mean face recognition rates on the FERET database using different descriptors.

PCA, MahCosine	0.72
Bayesian, MAP	0.72
EBGM_Optimal	0.66
Difference histogram	0.63
Homogeneous texture	0.62
Texton histogram	0.76
LBP (nonweighted)	0.76
LBP (weighted)	0.81

Paper I. Face recognition rate on the CMU PIE dataset using different filter banks and quantization methods.

	Codebook	*Thresholding*
Local derivative filters	0.482	0.742
Gabor(1,4)	0.502	0.883
Gabor(4,6)	0.704	-
MR8 filters	0.204	0.348
(SFFS) Gabor(4,6)	-	1.000

Paper VII. Correct face verification rates using the P and G protocols of the BANCA dataset.

	P	*G*
Block based LBP (Rodriguez & Marcel 2006)	.808	.950
Kernel density estimation + LBP	.824	.960
Weighted KDE + LBP	.884	.970
Illumination preprocessing KDE + LBP	.899	.981
Ill. preproc, Weighted KDE + LBP	.909	.985

Paper VIII. Face recognition rates on CMU PIE images with artificial Gaussian blur and on FRGC Experiment 1.0.4 with local binary patterns and local phase quantization.

	LBP	*LPQ*
CMU PIE images, no blur	0.927	0.992
CMU PIE images, blur $\sigma = 2.0$	0.701	0.935
FRGC Experiment 1.0.4	0.326	0.459
FRGC Exp 1.0.4 with illumination preprocessing	0.643	0.745

4.4 Further work using local binary pattern based face description

Since the publication of the first results on the LBP based face description in Paper V, our methodology has attained an established position in face analysis research. This is attested by the increasing number of studies which adopted a similar approach. At the time of the writing, Google Scholar[5] lists 132 citations of Paper V, and the local binary pattern bibliography[6] includes over 80 publications using local binary patterns for face image analysis. Several of these were cited in the previous sections, and in the following we review some further works applying LBP to face image analysis.

Local binary patterns computed in local regions for face detection was used by Hadid *et al.* (2004). In that work, LBP features from local regions combined with a histogram representing the whole face area yielded an excellent face detection rate when used as features for a support vector machine classifier. Later, LBP features for face detection on mobile phones has been investigated by Hannuksela *et al.* (2008).

Using LBP histograms for facial expression recognition from still images has been studied by Feng *et al.* (2005) and Shan *et al.* (2009). Zhao & Pietikäinen (2007) introduced a variant of LBP for analyzing dynamic textures and applied it to facial expression recognition from videos. Use of LBP features for demographic classification has been studied in (e.g. Yang & Ai 2007, Hadid & Pietikäinen in press). Li *et al.* (2007) built a highly accurate, illumination-invariant face recognition system by combining near-infrared imaging with an LBP-based face description and AdaBoost learning.

Computing LBP features from images obtained by filtering a facial image with 40 Gabor filters of different scale and orientation are shown to yield an excellent recognition rate on all the FERET sets in (Zhang *et al.* 2005). A downside of the method proposed in that paper is the high dimensionality of the feature vectors. Other approaches to combine LBP and Gabor features have been investigated by Yan *et al.* (2007) and Tan & Triggs (2007b).

Other proposed enhancements include multi-scale block LBP which considers mean gray values from larger pixel blocks than original LBP (Liao *et al.* 2007) and using patterns at multiple scales for representation (Chan *et al.* 2007).

[5] http://scholar.google.com/

[6] http://www.ee.oulu.fi/mvg/page/lbp_bibliography

4.5 Discussion and conclusion

The main contribution of this chapter is the efficient facial representation proposed in papers V and VI. It is based on dividing a facial image into small regions and computing a description of each region using local binary patterns. These descriptors are then combined into a spatially enhanced histogram or feature vector. The texture description of a single region describes the appearance of the region and the combination of all region descriptions encodes the global geometry of the face.

Furthermore, we proposed a different, revised method for estimating the local distributions of LBP labels for face description. This revised method is based on kernel density estimation in *xy*-space, and it provides much more spatial accuracy than the earlier, block-based methods.

The kernel density estimation based method alleviates two problems related to computing histograms within local blocks. First, finding an optimal way to divide the facial area into local blocks has proven to be a difficult problem. Second, the block histogram estimates of LBP label probabilities are likely to be unreliable near the block boundaries. Both these problems are avoided when using the proposed method. We also proposed and evaluated the use of a support vector machine in information fusion from individual pixels for the binary classification task for identity verification.

In this thesis, the proposed methodology is assessed with the face recognition and verification task. However, the block based method has already yielded in outstanding performance in face detection (Hadid *et al.* 2004) and facial expression recognition (Feng *et al.* 2005, Shan *et al.* 2009). We believe that pixel based density estimation will prove similarly useful. Moreover, the developed approaches are not limited to these few example applications as they can be easily generalized to other types of object detection and recognition tasks.

To combat a challenge often present in real life face images, blur, we analyzed the applicability of the blur tolerant local phase quantization operator. It was shown that LPQ can reach higher recognition rates than the LBP operator. Prior to the work in this thesis, the applicability of LPQ was only considered in the context of texture images with artificial blur (Ojansivu & Heikkilä 2008). Our work experimentally showed its effectiveness also with face images, and real out-of-focus blur.

In this thesis, it was shown that the LPQ operator also can be explained in the FLS framework. Presenting LPQ in this framework shows that whitening transformation of the LPQ filters enhances their descriptiveness. This observation begs the question whether a similar transformation of the other filters considered in the framework would be beneficial.

5 Summary

In this thesis, we aimed at studying and developing robust image descriptors especially for texture and face recognition tasks.

This work was based on our findings in face description using local binary patterns. This texture operator was applied to feature extraction from face images, yielding excellent recognition performance. These results highlight the connection between texture and face image analysis, and show that discriminative and robust face description can be achieved from a texture analysis standpoint.

To facilitate the task of descriptor development and analysis, we presented a novel unified filtering-labeling-statistic framework under which histogram based image description methods such as the local binary pattern and MR8 descriptors can be explained and analyzed. Even though this is still far from a complete unified theory of statistical image description, the framework makes the differences and similarities between the methods apparent.

Moreover, the presented framework allows for systematic comparison of different descriptors, and the parts that they are built of. Such an analytic approach can be useful in analyzing descriptors, as they are usually presented in the literature as a sequence of steps whose relation to other description methods is unclear. The framework presented in this study allows for explicitly illustrating the connection between the parts of the LBP and several other descriptors, and allows experimenting with the performance of each part.

The framework was utilized in face image analysis by considering statistics that retain information about the spatial structure of the labels. The first proposed approach, computing histograms in local rectangular regions and concatenating the histograms has already been widely accepted in face image analysis. A different, more precise approach was also presented in which the label distributions are estimated for each pixel by using kernel density estimation.

In the FLS framework, two alternative solutions to the labeling part were proposed to increase the robustness of the resulting descriptors. The first approach, soft LBP, is based on soft quantization of the vector of filter responses to several labels. The other proposed method, Bayesian LBP, uses a prior distribution of labelings, and aims for the labeling maximizing the a posteriori probability. Furthermore, a statistic computed from uniform LBP labels, and invariant to rotations of the input image was proposed and evaluated.

In face recognition, a filter set for a blur tolerant description of faces using local phase quantization was presented, and it was shown to be capable of achieving higher rates than the local derivative filter set, even with face images with no blur.

In analyzing the significance of the results presented in this thesis and assessing whether the objectives of the research were met, one can say that the proposed FLS framework has shown its usefulness, though not yet widely and explicitly applied by other researchers. Our experiments showed experimental the success of the methods, but often at the cost of increased computational load or memory requirements. It remains to be seen if these will be adopted by the research community. However, the proposed face representation has been proven with certainty to be useful, as it has fared well in independent comparisons, and it has been adopted both in research and commercial applications.

References

Ade F (1983) Characterization of textures by 'eigenfilters'. Signal Processing 5(5): 451–457.

Adini Y, Moses Y & Ullman S (1997) Face recognition: The problem of compensating for changes in illumination direction. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7): 721–732.

Ahonen T & Pietikäinen M (2008) A framework for analyzing texture descriptors. Proc. 3rd International Conference on Computer Vision Theory and Applications (VISAPP 2008), 2: 507–512.

Ahonen T, Pietikäinen M, Hadid A & Mäenpää T (2004) Face recognition based on appearance of local regions. Proc. 17th International Conference on Pattern Recognition (ICPR 2004), III: 153–156.

Ao M, Yi D, Lei Z & Li SZ (2009) Face recognition at a distance: System issues. In: Tistarelli M, Li SZ & Chellappa R (eds) Handbook of Remote Biometrics for Surveillance and Security. Springer.

Arof H & Deravi F (1998) Circular neighbourhood and 1-d dft features for texture classification and segmentation. Vision, Image and Signal Processing, IEE Proceedings 145(3): 167–172.

Bailly-Bailliére E, Bengio S, Bimbot F, Hamouz M, Kittler J, Mariéthoz J, Matas J, Messer K, Popovici V, Porée F, Ruíz B & Thiran JP (2003) The BANCA database and evaluation protocol. Proc. Audio- and Video-Based Biometric Person Authentication (AVBPA 2003), Springer, 2688: 625–638.

Barcelo A, Montseny E & Sobrevilla P (2007) Fuzzy texture unit and fuzzy texture spectrum for texture characterization. Fuzzy Sets and Systems 158(3): 239–252.

Bartlett MS, Movellan JR & Sejnowski TJ (2002) Face recognition by independent component analysis. IEEE Transactions on Neural Networks 13(6): 1450–1464.

Bay H, Ess A, Tuytelaars T & Van Gool LJ (2008) Speeded-up robust features (surf). Computer Vision and Image Understanding 110(3): 346–359.

Belhumeur PN, Hespanha JP & Kriegman DJ (1997) Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7): 711–720.

Ben-Yacoub S, Abdeljaoued Y & Mayoraz E (1999) Fusion of face and speech data for person identity verification. IEEE Transactions on Neural Networks 10(5): 1065–1074.

Beveridge JR, Bolme D, Draper BA & Teixeira M (2005) The CSU face identification evaluation system: Its purpose, features, and structure. Machine Vision and Applications 16(2): 128–138.

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

Blanz V, Grother P, Phillips PJ & Vetter T (2005) Face recognition based on frontal views generated from non-frontal images. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), IEEE Computer Society, Washington, DC, USA, 454–461.

Blanz V & Vetter T (1999) A morphable model for the synthesis of 3d faces. Proc. 26th annual conference on Computer graphics (SIGGRAPH 1999), ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 187–194.

Blanz V & Vetter T (2003) Face recognition based on fitting a 3D morphable model. IEEE

Transactions on Pattern Analysis and Machine Intelligence 25(9): 1063–1074.
Bonet JSD & Viola P (1997) A non-parametric multi-scale statistical model for natural images. Proc. Advances in Neural Information Processing, MIT Press, 773–779.
Bovik A, Clark M & Geisler W (1990) Multichannel texture analysis using localized spatial filters. Pattern Analysis and Machine Intelligence, IEEE Transactions on 12(1): 55–73.
Bowyer KW, Chang KI & Flynn PJ (2006) A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. Computer Vision and Image Understanding 101(1): 1–15.
Boykov Y & Kolmogorov V (2004) An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(9): 1124–1137.
Brown M & Lowe DG (2007) Automatic panoramic image stitching using invariant features. International Journal of Computer Vision 74(1): 59–73.
Brown MZ, Burschka D & Hager GD (2003) Advances in computational stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(8): 993–1008.
Caelli T & Julesz B (1978) On perceptual analyzers underlying visual texture discrimination: Part I. Biological Cybernetics 28: 167–176.
Caputo B, Hayman E & Mallikarjuna P (2005) Class-specific material categorisation. Proc. Tenth IEEE International Conference on Computer Vision (ICCV 2005), 1597–1604.
Cevikalp H, Neamtu M & Wilkes M (2006) Discriminative common vector method with kernels. IEEE Transactions on Neural Networks 17(6): 1550–1565.
Chan CH, Kittler J & Messer K (2007) Multi-scale local binary pattern histograms for face recognition. Proc. International Conference on Advances in Biometrics (ICB 2007), 4642: 809–818.
Chantler MJ & Van Gool LJ (2005) Editorial: Special issue on "texture analysis and synthesis". International Journal of Computer Vision 62(1-2): 5.
Chávez E, Navarro G, Baeza-Yates RA & Marroquín JL (2001) Searching in metric spaces. ACM Computing Surveys 33(3): 273–321.
Chellappa R, Wilson CL & Sirohey S (1995) Human and machine recognition of faces: a survey. Proceedings of the IEEE 83(5): 705–741.
Chen T, Yin W, Zhou XS, Comaniciu D & Huang TS (2006) Total variation models for variable lighting face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9): 1519–1524.
Clausi DA & Jernigan ME (2000) Designing Gabor filters for optimal texture separability. Pattern Recognition 33(11): 1835–1849.
Coggins J & Jain A (1985) A spatial filtering approach to texture analysis. Pattern Recognition Letters 3: 195–203.
Cootes TF, Edwards GJ & Taylor CJ (2001) Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6): 681–685.
Cula OG & Dana KJ (2004) 3d texture recognition using bidirectional feature histograms. International Journal of Computer Vision 59(1): 33–60.
Dalal N & Triggs B (2005) Histograms of oriented gradients for human detection. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 886–893.
Datta R, Joshi D, Li J & Wang JZ (2008) Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys 40(2): 5:1–5:60.

Do MN & Vetterli M (2002) Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance. IEEE Transactions on Image Processing 11(2): 146–158.

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

Duin RP, Loog M & Ho TK (2006) Recent submissions in linear dimensionality reduction and face recognition. Pattern Recognition Letters 27(7): 707 – 708.

Efros AA & Freeman WT (2001) Image quilting for texture synthesis and transfer. Proc. 28th Annual Conference on Computer Graphics (SIGGRAPH 2001), 341–346.

Faugeras O & Pratt W (1980) Decorrelation methods of texture feature extraction. IEEE Transactions on Pattern Analysis and Machine Intelligence 2(4): 323–332.

Feng X, Pietikäinen M & Hadid A (2005) Facial expression recognition with local binary patterns and linear programming. Pattern Recognition and Image Analysis 15(2): 546–548.

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

Gluckman J (2005) Visually distinct patterns with matching subband statistics. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(2): 252–264.

Gray RM & Neuhoff DL (1998) Quantization. IEEE Transactions on Information Theory 44(6): 2325–2383.

Grigorescu SE, Petkov N & Kruizinga P (2002) Comparison of texture features based on Gabor filters. IEEE Transactions on Image Processing 11(10): 1160–1167.

Gross R & Brajovic V (2003) An image preprocessing algorithm for illumination invariant face recognition. Proc. Audio- and Video-Based Biometric Person Authentication (AVBPA 2003), 10–18.

Guo G, Li SZ & Chan KL (2001) Support vector machines for face recognition. Image and Vision Computing 19(9–10): 631–638.

Hadid A & Pietikäinen M (in press) Combining appearance and motion for face and gender recognition from videos. Pattern Recognition .

Hadid A, Pietikäinen M & Ahonen T (2004) A discriminative feature space for detecting and recognizing faces. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), 2: 797–804.

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

Hand DJ (2006) Classifier technology and the illusion of progress. Statistical Science 21: 1–34.

Hannuksela J, Sangi P, Turtinen M & Heikkilä J (2008) Face tracking for spatially aware mobile user interfaces. Proc. 3rd International Conference on Image and Signal Processing (ICISP 2008), 405–412.

Haralick RM, Shanmugam K & Dinstein I (1973) Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics 3(6): 610–621.

Hayman E, Caputo B, Fritz M & Eklundh JO (2004) On the significance of real-world conditions for material classification. Proc. 8th European Conference on Computer Vision (ECCV 2004), Springer, 3024: 253–266.

Heikkilä J (2002) Multi-scale autoconvolution for affine invariant pattern recognition. Proc. 16th International Conference on Pattern Recognition (ICPR 2002), 119–122.

Heikkilä M & Pietikäinen M (2006) A texture-based method for modeling the background and detecting moving objects. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(4): 657–662.

Heikkilä M, Pietikäinen M & Schmid C (2009) Description of interest regions with local binary patterns. Pattern Recognition 42(3): 425–436.

Heisele B, Ho P, Wu J & Poggio T (2003) Face recognition: component-based versus global approaches. Compter Vision and Image Understanding 91(1–2): 6–21.

Heisele B, Serre T & Poggio T (2007) A component-based framework for face detection and identification. International Journal of Computer Vision 74(2): 167–181.

Heusch G, Rodriguez Y & Marcel S (2006) Local binary patterns as an image preprocessing for face authentication. Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FG 2006), IEEE Computer Society, 9–14.

Holappa J, Ahonen T & Pietikäinen M (2008) An optimized illumination normalization method for face recognition. Proc. IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2008), 1–6.

Hu MK (1962) Visual pattern recognition by moment invariants. Information Theory, IEEE Transactions on 8(2): 179–187.

Huang J, Heisele B & Blanz V (2003) Component-based face recognition with 3d morphable models. Proc. 4th International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 2003), Springer, 27–34.

Huang J, Ravi Kumar S, Mitra M, Zhu WJ & Zabih R (1999) Spatial color indexing and applications. International Journal of Computer Vision 35(3): 245–268.

Hyvärinen A, Karhunen J & Oja E (2001) Independent Component Analysis. John Wiley & Sons, New York.

Iakovidis DK, Keramidas EG & Maroulis D (2008) Fuzzy local binary patterns for ultrasound texture characterization. Proc. International Conference on Image Analysis and Recognition (ICIAR 2008), 5112: 750–759.

Jain AK & Farrokhnia F (1991) Unsupervised texture segmentation using gabor filters. Pattern Recognition 24(12): 1167 – 1186.

Julesz B (1962) Visual pattern discrimination. IRE Transactions on Information Theory 8(2): 84–92.

Julesz B & Bergen R (1981) Textons, the elements of texture perception, and their interactions. Nature 290: 91–97.

Julesz B, Gilbert EN, Shepp LA & Frisch HL (1973) Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited. Perception 2(4): 391–405.

Keramidas EG, Iakovidis DK, Maroulis D & Dimitropoulos N (2008) Thyroid texture representation via noise resistant image features. Proc. IEEE International Symposium on Computer-Based Medical Systems, 560–565.

Knutsson H & Granlund GH (1983) Texture analysis using two-dimensional quadrature filters. Proc. IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management, 206–213.

Kohonen T (1988) Self-Organization and Associative Memory. Springer, Berlin.

Konishi S & Yuille A (2000) Statistical cues for domain specific image segmentation with performance analysis. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000), I: 125–132.

Lades M, Vorbruggen JC, Buhmann J, Lange J, von der Malsburg C, Wurtz RP & Konen W (1993) Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers 42(3): 300–311.

Laws K (1979) Texture energy measures. Proc. DARPA Image Understanding Workshop, 47–51.

Lazebnik S, Schmid C & Ponce J (2005) A sparse texture representation using local affine regions. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8): 1265–1278.
Leung T & Malik J (2001) Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision 43(1): 29–44.
Lew MS, Sebe N, Djeraba C & Jain R (2006) Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications and Applications 2(1): 1–19.
Li F & Wechsler H (2005) Open set face recognition using transduction. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(11): 1686–1697.
Li S, Chu R, Liao S & Zhang L (2007) Illumination invariant face recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4): 627 – 639.
Li SZ & Zhang Z (2004) Floatboost learning and statistical face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(9): 1112–1123.
Liao S, Zhu X, Lei Z, Zhang L & Li SZ (2007) Learning multi-scale block local binary patterns for face recognition. Proc. International Conference on Advances in Biometrics (ICB 2007), 4642: 828–837.
Liu C (2004) Gabor-based kernel pca with fractional power polynomial models for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5): 572–581.
Liu C (2006) Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5): 725–737.
Liu C & Wechsler H (2002) Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. IEEE Transactions on Image Processing 11(4): 467–476.
Liu X & Wang DL (2003) Texture classification using spectral histograms. IEEE Transactions on Image Processing 12(6): 661–670.
Lobay A & Forsyth DA (2006) Shape from texture without boundaries. International Journal of Computer Vision 67(1): 71–91.
Lowe DG (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60: 91–110.
Mäenpää T (2003) The local binary pattern approach to texture analysis – extensions and applications. Ph.D. thesis, University of Oulu, Acta Univ Oul C 187.
Malik J, Belongie S, Leung TK & Shi J (2001) Contour and texture analysis for image segmentation. International Journal of Computer Vision 43(1): 7–27.
Malik J, Belongie S, Shi J & Leung T (1999) Textons, contours and regions: Cue integration in image segmentation. Proc. IEEE International Conference on Computer Vision (ICCV 1999), 918–925.
Manjunath BS & Ma WY (1996) Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(8): 837–842.
Manjunath BS, Ohm JR, Vasudevan VV & Yamada A (2001) Color and texture descriptors. IEEE Transactions on Circuits and Systems for Video Technology 11(6): 703–715.
Martínez AM & Kak AC (2001) PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2): 228–233.
Mikolajczyk K & Schmid C (2005) A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10): 1615–1630.
Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T & Van

Gool LJ (2005) A comparison of affine region detectors. International Journal of Computer Vision 65(1-2): 43–72.

Mirmehdi M, Xie X & Suri J (eds) (2008) Handbook of Texture Analysis. Imperial College Press, London, UK.

Moghaddam B & Pentland A (1997) Probabilistic visual learning for object recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7): 696–710.

Nowak E, Jurie F & Triggs B (2006) Sampling strategies for bag-of-features image classification. Proc. Computer Vision - ECCV 2006, Springer, 3954: IV: 490–503.

Ojala T, Mäenpää T, Pietikäinen M, Viertola J, Kyllönen J & Huovinen S (2002a) Outex - new framework for empirical evaluation of texture analysis algorithms. Proc. 16th International Conference on Pattern Recognition (ICPR 2002), 1: 701 – 706.

Ojala T, Pietikäinen M & Harwood D (1994) Performance evaluation of texture measures with classification based on kullback discrimination of distributions. Proc. 12th IAPR International Conference on Pattern Recognition (ICPR 1994), 1: 582–585 vol.1.

Ojala T, Pietikäinen M & Harwood D (1996) A comparative study of texture measures with classification based on featured distribution. Pattern Recognition 29(1): 51–59.

Ojala T, Pietikäinen M & Mäenpää T (2002b) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7): 971–987.

Ojala T, Valkealahti K, Oja E & Pietikäinen M (2001) Texture discrimination with multidimensional distributions of signed gray-level differences. Pattern Recognition 34(3): 727–739.

Ojansivu V & Heikkilä J (2008) Blur insensitive texture classification using local phase quantization. Proc. International Conference on Image and Signal Processing (ICISP 2008), 236–243.

Pass G & Zabih R (1999) Comparing images using joint histograms. Multimedia Systems 7(3): 234–240.

Penev PS & Atick JJ (1996) Local feature analysis: A general statistical theory for object representation. Network – Computation in Neural Systems 7(3): 477–500.

Pentland A, Moghaddam B & Starner T (1994) View-based and modular eigenspaces for face recognition. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 84–91.

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

Petrou M & Sevilla PG (2006) Image Processing: Dealing with Texture. John Wiley & Sons, Ltd.

Phillips JP, Scruggs TW, O'Toole AJ, Flynn PJ, Bowyer KW, Schott CL & Sharpe M (2007) FRVT 2006 and ICE 2006 large-scale results. Technical report, National Institute of Standards and Technology.

Phillips PJ, Moon H, Rizvi SA & Rauss PJ (2000) The FERET evaluation methodology for face recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10): 1090–1104.

Popat K & Picard R (1993) Novel cluster-based probability model for texture synthesis, classification, and compression. Proc. Visual Communications and Image Processing, 756–768.

Portilla J & Simoncelli EP (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. International Journal of Computer Vision 40(1): 49–70.

Pudil P, Novovicová J & Kittler J (1994) Floating search methods in feature selection. Pattern Recognition Letters 15(10): 1119–1125.

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

Randen T & Husøy JH (1999) Filtering for texture classification: A comparative study. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(4): 291–310.

Rikert T, Jones M & Viola P (1999) A cluster-based statistical model for object detection. Proc. IEEE International Conference on Computer Vision (ICCV 1999), 1046–1053.

Rodriguez Y & Marcel S (2006) Face authentication using adapted local binary pattern histograms. Proc. 9th European Conference on Computer Vision (ECCV 2006), 4: 321–332.

Rosenfeld A, Wang CY & Wu AY (1982) Multispectral texture. IEEE Transactions on Systems, Man and Cybernetics 12(1): 79–84.

Samal A & Iyengar PA (1992) Automatic recognition and analysis of human faces and facial expressions: a survey. Pattern Recognition 25(1): 65 – 77.

van de Sande KEA, Gevers T & Snoek CGM (2008) Evaluation of color descriptors for object and scene recognition. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008).

Schiele B & Crowley JL (2000) Recognition without correspondence using multidimensional receptive field histograms. International Journal of Computer Vision 36(1): 31–50.

Schmid C (2001) Constructing models for content-based image retrieval. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001), 39–45.

Schmid C, Mohr R & Bauckhage C (2000) Evaluation of interest point detectors. International Journal of Computer Vision 37(2): 151–172.

Schneiderman H & Kanade T (2000) A statistical method for 3d object detection applied to faces and cars. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000), IEEE Computer Society, Washington, DC, USA, 1: 746–751.

Scott DW (1992) Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley-Interscience.

Shan C, Gong S & McOwan PW (2009) Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing 27(6): 803–816.

Short J, Kittler J & Messer K (2004) A comparison of photometric normalisation algorithms for face verification. Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2004), 254–259.

Siggelgov S (2002) Feature historgrams for content-based image retrieval. Ph.D. thesis, Albert-Ludwigs-Universität Freiburg.

Sim T, Baker S & Bsat M (2003) The cmu pose, illumination, and expression database. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(12): 1615–1618.

Sinha P, Balas B, Ostrovsky Y & Russell R (2006) Face recognition by humans: Nineteen results all computer vision researchers should know about. Proceedings of the IEEE 94(11): 1948–1962.

Sirovich L & Kirby M (1987) Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A 4(3): 519–524.

Smeulders AWM, Worring M, Santini S, Gupta A & Jain R (2000) Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12): 1349–1380.

Swain MJ & Ballard DH (1991) Color indexing. International Journal of Computer Vision 7(1): 11–32.

Tan X, Chen S, Zhou ZH & Zhang F (2006) Face recognition from a single image per person: A survey. Pattern Recognition 39(9): 1725–1745.

Tan X & Triggs B (2007a) Enhanced local texture feature sets for face recognition under difficult lighting conditions. Proc. Analysis and Modeling of Faces and Gestures (AMFG 2007).

Tan X & Triggs B (2007b) Fusing Gabor and LBP feature sets for kernel-based face recognition. Proc. Analysis and Modeling of Faces and Gestures (AMFG 2007), 235–249.

Tax D & Duin R (2008) Growing a multi-class classifier with a reject option. Pattern Recognition Letters 29(10): 1565 – 1570.

Tistarelli M, Bicego M & Grosso E (2009) Dynamic face recognition: From human to machine vision. Image and Vision Computing 27(3): 222 – 232.

Todd JT (2004) The visual perception of 3d shape. Trends in Cognitive Sciences 8(3): 115 – 121.

Tuceryan M & Jain AK (1999) Texture analysis. In: Chen CH, Pau LF & Wang PSP (eds) The Handbook of Pattern Recognition and Computer Vision (2nd Edition), 207–248. World Scientific Publishing Co.

Turk M & Pentland A (1991) Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1): 71–86.

Turner MR (1986) Texture discrimination by gabor functions. Biological Cybernetics 55(2-3): 71–82.

Tuytelaars T & Schmid C (2007) Vector quantizing feature space with a regular lattice. Proc. IEEE 11th International Conference on Computer Vision (ICCV 2007), IEEE, 1–8.

Valkealahti K & Oja E (1998) Reduced multidimensional co-occurrence histograms in texture classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(1): 90–94.

Vapnik VN (1998) Statistical Learning Theory. Wiley-Interscience.

Varma M & Zisserman A (2003) Texture classification: Are filter banks necessary? Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2003), 2: 691–698.

Varma M & Zisserman A (2004) Unifying statistical texture classification frameworks. Image and Vision Computing 22(14): 1175–1183.

Varma M & Zisserman A (2005) A statistical approach to texture classification from single images. International Journal of Computer Vision 62(1–2): 61–81.

Viola P & Jones M (2001) Rapid object detection using a boosted cascade of simple features. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), 1: 511–518.

Viola P & Jones MJ (2004) Robust real-time face detection. International Journal of Computer Vision 57(2): 137–154.

Vogel J & Schiele B (2007) Semantic modeling of natural scenes for content-based image retrieval. International Journal of Computer Vision 72(2): 133–157.

Wang H, Li SZ & Wang Y (2004) Face recognition under varying lighting conditions using self quotient image. Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2004), 819–824.

Wang L & He DC (1990) Texture classification using texture spectrum. Pattern Recognition 23(8): 905–910.

Wang X & Tang X (2004) A unified framework for subspace face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(9): 1222–1228.

Weszka J, Dyer C & Rosenfeld A (1976) A comparative study of texture measures for terrain classification. IEEE Transactions on Systems, Man and Cybernetics 6(4): 269–286.

Wiskott L, Fellous JM, Kuiger N & von der Malsburg C (1997) Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 19: 775–779.

Wright J, Yang A, Ganesh A, Sastry S & Ma Y (2009) Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2): 210 – 227.

Xie X & Mirmehdi M (2007) Texems: Texture exemplars for defect detection on random textured surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(8): 1454–1464.

Yan S, Wang H, Tang X & Huang TS (2007) Exploring feature descriptors for face recognition. Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007), IEEE Signal Processing Society.

Yang MH, Kriegman DJ & Ahuja N (2002) Detecting faces in images: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(1): 34–58.

Yang Z & Ai H (2007) Demographic classification with local binary patterns. Proc. International Conference on Advances in Biometrics (ICB 2007), 4642: 464–473.

Zabih R & Woodfill J (1994) Non-parametric local transforms for computing visual correspondence. Proc. 3rd European Conference on Computer Vision (ECCV 1994), B:151–158.

Zhang B, Shan S, Chen X & Gao W (2007a) Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition. IEEE Transactions on Image Processing 16(1): 57–68.

Zhang G, Huang X, Li SZ, Wang Y & Wu X (2004) Boosting local binary pattern (LBP)-based face recognition. Proc. Advances in Biometric Person Authentication, 3338: 179–186.

Zhang J, Marszalek M, Lazebnik S & Schmid C (2007b) Local features and kernels for classification of texture and object categories: A comprehensive study. International Journal of Computer Vision 73(2): 213–238.

Zhang J & Tan T (2002) Brief review of invariant texture analysis methods. Pattern Recognition 35(3): 735–747.

Zhang W, Shan S, Gao W, Chen X & Zhang H (2005) Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. Proc. Tenth IEEE International Conference on Computer Vision (ICCV 2005), 1:786–791.

Zhao G & Pietikäinen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6): 915–928.

Zhao W, Chellappa R & Krishnaswamy A (1998) Discriminant analysis of principal components for face recognition. Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition, 336–341.

Zhao W, Chellappa R, Phillips PJ & Rosenfeld A (2003) Face recognition: A literature survey. ACM Computing Surveys 35(4): 399–458.

Zhou SK, Aggarwal G, Chellappa R & Jacobs DW (2007) Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(2): 230–245.

Zhu SC, Wu YN & Mumford D (1998) Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. International Journal of Computer Vision 27(2): 107–126.

Appendix 1 The Filtering, Labeling and Statistic framework

The following two tables summarize how the methods introduced in this thesis, and some other image descriptors are presented within the Filtering, Labeling and Statistic framework.

Table 3. Methods presented in this thesis in the FLS framework.

Paper	Filtering	Labeling	Statistic	Applications
I: FLS framework	Derivative, Gabor, MR8	Thresholding (2 levels), codebook	Histogram	Material categorization, face recognition
II: Soft LBP	Derivative	Soft thresholding	Soft histogram	Supervised texture segmentation in noisy conditions
III: Bayesian LBP	Derivative	Maximum a posteriori estimation	Histogram	Texture retrieval
IV: LBP histogram Fourier features	Derivative	Thresholding, uniform patterns	Fourier features from histogram	Rotation invariant texture classification, material categorization, face recognition
V: Face recognition with LBP	Derivative	Thresholding, uniform patterns	Histograms from local regions concatenated	Face recognition
VI: Face recognition with LBP	Derivative	Thresholding, codebook	Histograms from local regions concatenated	Face recognition
VII: Kernel density estimation of LBP histograms	Derivative	Thresholding	Histograms for each pixel with kernel density estimation	Face verification
VIII: Local phase quantization for face description	Quadrature filters	Whitening + thresholding	Histogram	Face recognition from blurred images

Table 4. Other appearance descriptors presented within the FLS framework.

Method	Filtering	Labeling	Statistic	Applications
(Ojala *et al.* 2001)	Gray level differences of neighboring pixels (2, 4 or 8 neighbors)	Codebook, thresholding	Histogram	Texture classification
LM (Leung & Malik 2001)	48 filters (Gaussian, First and second order derivatives of Gaussian, Laplacian of Gaussian)	Codebook	Histogram	Three-dimensional texture classification
CD (Cula & Dana 2004)	45 filters (Gaussians, Derivatives of Gaussians)	Codebook	Manifold of histograms projected to lower dimension with PCA	Texture classification
VZ-MR8 (Varma & Zisserman 2004)	Bar and edge filters (derivatives of Gaussian), Gaussian, Laplacian of Gaussian	Select maximum over rotations, then codebook	Histogram	Texture classification
VZ-MRF (Varma & Zisserman 2003)	None, use raw pixel values from image patch	Exclude central pixel, then codebook	2-dimensional histogram of central pixel value and codebook label	Texture classification
Graylevel texems (Xie & Mirmehdi 2007)	None	-	Multivariate Gaussian mixture model	Texture defect detection
(Konishi & Yuille 2000)	Gaussian, Nitzberg, Gradient, Laplacian of Gaussian	Quantize each dimension into 6 bins	Joint histogram	Scene image segmentation

Method	Filtering	Labeling	Statistic	Applications
(Rikert *et al.* 1999)	Oriented derivatives at multiple scales	-	Gaussian mixture model	Object detection
(Schiele & Crowley 2000)	Gaussian derivatives at three scales	Quantize each dimension into 24 bins	Joint histogram	Object recognition
SIFT (Lowe 2004)	Difference of Gaussians	Phase angle of the response vector quantized	Weighted histograms from small spatial subregions	Several; originally proposed for wide baseline matching
HoG (Dalal & Triggs 2005)	Horizontal and vertical derivative filters	Phase angle of the response vector quantized	Weighted histograms from small spatial subregions	Several; originally proposed for human detection

Original articles

I Ahonen T & Pietikäinen M (2009) Image description using joint distribution of filter bank responses. Pattern Recognition Letters 30(4): 368–376.

II Ahonen T & Pietikäinen M (2007) Soft histograms for local binary patterns. In: Niskanen M & Heikkilä J (eds) Proc. Finnish Signal Processing Symposium (FINSIG 2007), Oulu, Finland.

III He C, Ahonen T & Pietikäinen M (2008) A Bayesian local binary pattern texture descriptor. Proc. 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, FL.

IV Ahonen T, Matas J, He C & Pietikäinen M (2009) Rotation invariant image description with local binary pattern histogram Fourier features. Proc. 16th Scandinavian Conference on Image Analysis (SCIA 2009), Oslo, Norway. Lecture Notes in Computer Science 5575, Springer.

V Ahonen T, Hadid A & Pietikäinen M (2004) Face recognition with local binary patterns. Proc. European Conference on Computer Vision (ECCV 2004), Prague, Czech Republic. Lecture Notes in Computer Science 3021, Springer: 469–481.

VI Ahonen T, Hadid A & Pietikäinen M (2006) Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12): 2037–2041.

VII Ahonen T & Pietikäinen M (2009) Pixelwise local binary pattern models of faces using kernel density estimation. Proc. 3rd IAPR/IEEE International Conference on Biometrics (ICB 2009), Alghero, Italy. Lecture Notes in Computer Science 5558, Springer.

VIII Ahonen T, Rahtu E, Ojansivu V & Heikkilä J (2008) Recognition of blurred faces using local phase quantization. Proc. 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, FL.

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

313. Hannuksela, Jari (2008) Camera based motion estimation and recognition for human-computer interaction

314. Nieminen, Timo (2009) Detection of harmful microbes and their metabolites with novel methods in the agri-food production chain

315. Marjala, Pauliina (2009) Työhyvinvoinnin kokemukset kertomuksellisina prosesseina–narratiivinen arviointitutkimus

316. Ahola, Juha (2009) Reaction kinetics and reactor modelling in the design of catalytic reactors for automotive exhaust gas abatement

317. Koskimäki, Heli (2009) Utilizing similarity information in industrial applications

318. Puska, Henri (2009) Code acquisition in direct sequence spread spectrum systems using smart antennas

319. Saari, Seppo (2009) Knowledge transfer to product development processes. A multiple case study in two small technology parks

320. Komulainen, Mikko (2009) Bandwidth enhanced antennas for mobile terminals and multilayer ceramic packages

321. Ronkanen, Anna-Kaisa (2009) Hydrologic and hydraulic processes in northern treatment peatlands and the significance for phosphorus and nitrogen removal

322. Liedes, Toni (2009) Improving the performance of the semi-active tuned mass damper

323. Marina Tyunina & Orest Vendik (Eds.) (2009) Proceedings of the 16th International Student Seminar "Microwave and optical applications of novel phenomena and technologies", June 8–9, Oulu, Finland

324. Belt, Pekka (2009) Improving verification and validation activities in ICT companies—product development management approach

325. Harri Haapasalo & Hanna Kropsu-Vehkaperä (Eds.) (2009) The 3rd Nordic Innovation Research Conference - NIR 2008—IEM Arctic Workshop

326. Selek, István (2009) Novel evolutionary methods in engineering optimization—towards robustness and efficiency

327. Härkönen, Janne (2009) Improving product development process through verification and validation

328. Peiponen, Kai-Erik (2009) Optical spectra analysis of turbid liquids

329. Kettunen, Juha (2009) Essays on strategic management and quality assurance

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.B. 7500 FI-90014 UNIVERSITY OF OULU FINLAND

A C T A U N I V E R S I T A T I S O U L U E N S I S

ISBN 978-951-42-9181-4 (Paperback)
ISBN 978-951-42-9182-1 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)

UNIVERSITY of OULU
OULUN YLIOPISTO