Charles University in Prague

Faculty of Mathematics and Physics

Department of Software Engineering

DOCTORAL THESIS

Query by Pictorial Example

Mgr. Pavel Vácha

Supervisor: Prof. Ing. Michal Haindl, DrSc.

Prague, 2010

Abstract:

Ongoing expansion of digital images requires new methods for sorting, browsing, and searching through huge image databases. This is a domain of Content-Based Image Retrieval (CBIR) systems, which are database search engines for images. A user typically submit a query image or series of images and the CBIR system tries to find and to retrieve the most similar images from the database. Optimally, the retrieved images should not be sensitive to circumstances during their acquisition. Unfortunately, the appearance of natural objects and materials is highly illumination and viewpoint dependent.

This work focuses on representation and retrieval of homogeneous images, called textures, under the circumstances with variable illumination and texture rotation. We propose a novel illumination invariant textural features based on Markovian modelling of spatial texture relations. The texture is modelled by Causal Autoregressive Random field (CAR) or Gaussian Markov Random Field (GMRF) models, which allow a very efficient estimation of its parameters, without the demanding Monte Carlo minimisation. Subsequently, the estimated model parameters are transformed into the new illumination invariants, which represent the texture. We derived that our textural representation is invariant to changes of illumination intensity and colour/spectrum, and also approximately invariant to local intensity variation (e.g. cast shadows). On top of that, our experiments showed that the proposed features are robust to illumination direction variations and the image degradation with an additive Gaussian noise. The textural representation is extended to be simultaneously illumination and rotation invariant.

The proposed features were tested in experiments on five different textural databases (Outex, Bonn BTF, CUReT, ALOT, and KTH-TIPS2). The experiments, closely resembling real-life conditions, confirmed that the proposed features are able to recognise materials in variable illumination conditions and different viewpoint directions. The proposed representation outperformed other state of the art textural representations (among others opponent Gabor features, LBP, LBP-HF, and MR8-LINC) in the almost all experiments. Our methods do not require any knowledge of acquisition conditions and the recognition is possible even with a single training image per material, if substantial scale variation or perspective projection is not included. The psychophysical experiments also indicated that our methods for the evaluation of textural similarity are related to the human perception of textures.

Four applications of our invariant features are presented. We developed a CBIR system, which retrieves similar tiles. We integrated the invariants into a texture segmentation algorithm. And feasible applications were demonstrated in optimisation of texture compression parameters and recognition of glaucomatous tissue in retina images. We expect that the presented methods can improve the performance of existing CBIR systems or they can be utilised in specialised CBIR systems focused on e.g. textural medical images or tiles as in the presented system. Other applications include computer vision, since the analysis of real scenes often requires a description of textures under various light conditions.

Keywords: texture, color, illumination invariance, rotation invariance, Markov random field, content based image retrieval

Abstrakt:

Rostoucí množství digitálních fotografií vyžaduje nové metody třídění, organizace a vyhledávání. Toto je úkolem CBIR systémů, což jsou databázové systémy specializované na prohledávání rozsáhlých obrazových databází. Uživatel typicky zadá vstupní obrázek nebo sérii obrázků a úkolem CBIR systému je nalézt v databázi obrázky co nejvíce podobné. V ideálním případě by nalezené obrázky neměli záviset podmínkách, ve kterých byly pořízeny. Bohužel vzhled mnoha objektů a přírodních materiálů velmi závisí na světelných podmínkách a úhlu pohledu.

Tato práce se zaměřuje na reprezentaci a vyhledávání homogenních obrazů (textur) a odolnost této reprezentace vůči změnám osvětlení a otočení textury. Navrhujeme nové světelně invariantní texturní příznaky, která jsou založené na Markovovském modelování prostorových vztahů v textuře. Textura je modelována kauzálním autoregresním modelem (CAR) nebo Gaussovsko-Markovovským modelem náhodného pole (GMRF), které umožňují velmi efektivní odhad svých parametrů, bez použití časově náročné Monte Carlo minimalizace. Odhadnuté parametry jsou následně transformovány do světelných invariantů, které reprezentují texturu. Odvodili jsme, že tato texturní reprezentace je invariantní ke změně intensity a barvy/spektra osvětlení a je také téměř invariantní k lokálním změnám intensity (např. vržené stíny). Provedené experimenty navíc ukázaly, že navrhované texturní příznaky jsou robustní ke změnám směru osvětlení a degradaci obrázků Gaussovským šumem. Navrženou texturní reprezentaci jsme rozšířili, aby byla zároveň světelně i rotačně invariantní.

Navrhované texturní příznaky byly otestovány na pěti různých texturních databázích (Outex, Bonn BTF, CUReT, ALOT a KTH-TIPS2). Provedené experimenty, odpovídající reálným podmínkám, potvrdily, že představené texturní příznaky jsou schopné rozpoznat přírodní materiály za různých světelných podmínek a při různém směru pohledu. Výsledky navržené reprezentace překonaly nejlepší alternativní texturní reprezentace jako oponentní Gaborovy příznaky, LBP, LBP-HF a MR8-LINC v téměř všech experimentech. Naše metody pracují bez znalosti podmínek při pořízení snímku a rozpoznávání je možné i s jediným trénovacím obrázkem pro každý materiál, pokud není obsažena výrazná změna měřítka nebo perspektivní projekce. Psychovizuální experimenty také naznačují, že naše metody pro posuzování texturní podobnosti odpovídají lidskému vnímání textur.

Navržené příznaky byly využity při konstrukci systému pro vyhledávaní podobných obkladů a začleněny do algoritmu pro segmentaci textur. Také jsme ukázali možné aplikace pro optimalizaci parametrů při kompresi textur a rozpoznávání glaukomické tkáně na snímcích sítnice. Prezentované metody mohou být využity pro zlepšení funkčnosti stávajících CBIR systémů nebo pro konstrukci specializovaných systémů zaměřených např. na texturní medicínské snímky nebo na obklady jako v prezentovaném systému. Další možnosti aplikací se nachází v počítačovém vidění, protože analýza reálných scén často vyžaduje popis textur při měnících se světelných podmínkách.

Klíčová slova: textura, barva, světelná invariance, rotační invariance, Markovovo náhodné pole, prohledávání obrazových databází

I hereby declare that I have written the thesis on my own and using exclusively the cited sources. For any work in the thesis that has been co-published with other authors, I have the permission of all co-authors to include this work in my thesis. I authorise Charles University to lend this document to other institutions or individuals for academic and research purposes.

Pavel Vácha
Prague, October 8, 2010

Acknowledgements:

Contents

Contents

List of Figures

List of Tables

List of Acronyms

2D	2 Dimensional
3D	3 Dimensional
ALOT	Amsterdam Library of Textures
AP	Average Precision
BTF	Bidirectional Texture Function
BRDF	Bidirectional Reflectance Distribution Function
CAR	Causal Autoregressive Random field
CBIR	Content-Based Image Retrieval
CUReT	Columbia-Utrecht Reflectance and Texture database
DFT	Discrete Fourier Transform
GM	Gaussian Mixture
GMRF	Gaussian Markov Random Field
EM	Expectation Maximisation
FC	Fuzzy Contrast
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
fMRI	functional Magnetic Resonance Imaging
HGS	Hoang-Geusebroek-Smeulders segmenter
J2EE	Java 2 platform Enterprise Edition
JRE	Java Runtime Environment
JSP	Java Server Pages

List of Acronyms

JPEG	Joint Photographic Experts Group
K-L	Karhunen-Loève transformation
k-NN	k-Nearest Neighbours
LBP	Local Binary Patterns
LBP^{riu2}	rotation invariant uniform Local Binary Patterns
LBP^{u2}	uniform Local Binary Patterns
LBP-HF	Local Binary Patterns - Histogram Fourier features
LMS	Least Mean Squares
LPQ	Local Phase Quantization
LS	Least Squares
MAP	Mean Average Precision
MCMC	Markov Chain Monte Carlo
MFS	MultiFractal Spectrum
ML	Maximum Likelihood
MR8	Maximal Response 8
MR8-NC	Maximal Response 8 - Normalised Colours
MR8-INC	Maximal Response 8 - Intensity Normalised Colours
MR8-LINC	Maximal Response 8 - Locally Intensity Normalised Colours
MR8-SLINC	Maximal Response 8 - Shading and Locally Intensity Normalised Colours
MRF	Markov Random Field
MRMR	Maximum Relevance and Minimum Redundancy
MR-SAR	MultiResolution Simultaneous AutoRegressive model
MUSCLE	Multimedia Understanding through Semantics, Computation and Learning
ONH	Optic Nerve Head
PCA	Principal Component Analysis
RAR	Rotation Autoregressive Random model

RGB	Red, Green, Blue additive colour model
RNF	Retinal Nerve Fibres
RR	Recall Rate
SIFT	Scale Invariant Feature Transform
SSIM	Structure Similarity Index Metric
SVM	Support Vector Machine
TRF	Tactical Receptive Field
VDP	Visual Difference Predictor

List of Notations

$\tilde{\cdot}$	accent used for different illumination
$\hat{\cdot}$	accent used for estimate
∇G	gradient of image G
$\operatorname{tr} A$	matrix trace
A^T	matrix transpose
A^{-1}	matrix inverse
$|A|$	matrix determinant
$|I|$	set cardinality
$|a|$	absolute value
a^*	complex conjugate
$\operatorname{diag} A$	matrix diagonal
$\operatorname{diag} v$	matrix with vector v on the diagonal
$\operatorname{supp}(f)$	support of function f
$0_{n \times n}$	zero matrix with size $n \times n$
$1_{n \times n}$	identity matrix with size $n \times n$
α_ℓ	illumination invariants
$\alpha_{\ell,j}$	illumination invariants, j-th spectral plane
β_ℓ	illumination invariants
γ	model parameter vector
γ_j	model parameter vector, j-th spectral plane
$\hat{\gamma}$	estimate of γ

List of Notations

$\hat{\gamma}_t$	estimate of γ from history $Y^{(t)}$
$\hat{\gamma}_{t,j}$	estimate of γ from history $Y^{(t)}$, j-th spectral plane
$\Gamma(x)$	Gamma function of variable x
η	cardinality of contextual neighbourhood I_r
ϵ_r	noise at position r
$\epsilon_{r,j}$	noise at position r, j-th spectral plane
λ_t	statistic used for estimation of noise variance
$\mu(X)$	mean value of X
$\nu_{s,j}$	j-th eigenvalue of matrix A_s
ω	wavelength
$\psi(r)$	number of steps from the beginning to position r
$\sigma(X)$	standard deviation of X
σ_j^2	variance of noise $\epsilon_{r,j}$
$\hat{\sigma}_{t,j}^2$	estimate of σ_j^2 from history $Y^{(t)}$
Σ	covariance matrix of noise ϵ_r
$\hat{\Sigma}$	estimate of Σ
$\hat{\Sigma}_t$	estimate of Σ from history $Y^{(t)}$
A_s	model parameter matrix corresponding to relative position s
$a_{s,j}$	model parameter for relative position s, j-th spectral plane
B	illumination transformation matrix
c_{pq}	complex moment of order $p + q$
$\hat{c}_{pq}$	discrete complex moment of order $p + q$
$\mathbb{E}X$	expected value of random variable X
$E(\omega)$	illumination spectral power distribution
$f_\ell^{(T)}$	ℓ-th component of feature vector for texture T
I	image lattice

I_r	index shift set
$I_r^{\odot}$	circular index shift set
I_r^u	unilateral index shift set
K	number of levels in Gaussian down-sampled pyramid
L_p	Minkowski norm (p-norm)
M	model
r, t	pixel position multiindices (row, column index) $r = [r_1, r_2]$
s	relative pixel position multiindex
$R_j(\omega)$	j-th sensor response function
V_{yy}	data accumulation matrix of pixel vectors Y_r
V_{zy}	data accumulation matrix of vectors Z_r and Y_r
V_{zz}	data accumulation matrix of data vectors Z_r
$V_{zz,j}$	data accumulation matrix of data vectors $Z_{r,j}$
$V_{zz(t)}$	data accumulation matrix V_{zz} computed from history $Y^{(t)}$
$V_{zz(t),j}$	data accumulation matrix $V_{zz,j}$ computed from history $Y^{(t)}$
V_0	data accumulation matrix prior
Y_r	vector of values at pixel position r
$Y_{r,j}$	value at pixel position r, j-th spectral plane
$Y^{(t)}$	process history up to pixel t, including corresponding data vectors
Z_r	model data vector at pixel position r
$Z_{r,j}$	model data vector at pixel position r, j-th spectral plane

Chapter 1

Introduction

1.1 Motivation

Ongoing expansion of digital images requires improved methods for sorting, browsing, and searching through ever-growing image databases. Such databases are used by various professionals including doctors searching for similar clinical cases, editors looking for illustration images and almost everyone needs to organise their personal photos. Other applications comprise accessing video archives by means of similar keyframes, detection of unauthorised image use, or cultural heritage applications. Former approaches to the image indexation were based on text descriptions and suffered not only from laborious and expensive creation but also imprecise description. Textual descriptions are influenced by personal background and expected utilisation, which is difficult or even impossible to predict. Moreover, there are some properties that can be hardly described in text as the atmosphere of Edvard Munch's The Scream.

Content-Based Image Retrieval (CBIR) systems are search engines for image databases, which index images according to their content. A typical task solved by CBIR systems is that a user submits a query image or series of images and the system is required to retrieve images from the database as similar as possible. Another task is a support for browsing through large image databases, where the images are supposed to be grouped or organised in accordance with similar properties. Although the image retrieval has been an active research area for many years (see surveys Smeulders et al. (2000) and Datta et al. (2008)) this difficult problem is still far from being solved. There are two main reasons, the first is so called *semantic gap*, which is the difference between information that can be extracted from the visual data and the interpretation that the same data have for a user in a given situation. The other reason is called *sensory gap*, which is the difference between a real object and its computational representation derived from sensors, which measurements are significantly influenced by the acquisition conditions.

The *semantic gap* is usually approached by learning of concepts or ontologies and subsequent attempts to recognise them. A system can also learn from the interaction with a user or try to employ combination of multimedia information. However, these topics are beyond the scope of this work and we refer to reviews Smeulders et al. (2000)

Figure 1.1: Appearance of a real scene under natural changes of illumination conditions.

and Lew et al. (2006) for further information.

This work concerns with the second mentioned problem of finding a reliable image representation, which is not influenced by image acquisition conditions. For example, a scene or an object can be photographed from different positions and the illumination can vary significantly during a day or be artificial, which causes significant changes in appearance (see Fig. 1.1). More specifically, we focus on a reliable and robust representation of homogeneous images (textures), which do not comprise the *semantic gap*.

1.1.1 Existing CBIR systems

Early CBIR systems as QBIC (Flickner et al., 1995) and VisualSEEk (Smith and Chang, 1996) were based on image colours represented by a kind of colour histogram, which totally ignored structures of materials and object surfaces present in the scene. Visual appearances of such structured surfaces are commonly referred as textures and their characterisation is essential for understanding of real scene images.

Later systems attempted to include some textural description, e.g. based on wavelets as CULE (Chen et al., 2005), IBM Video Retrieval System (Amir et al., 2005) or Gabor features as MediaMill (Snoek et al., 2008). MUFIN (Batko et al., 2010), which is focused on efficiency and scalability, includes a simple texture representation by MPEG-7

descriptors. A CBIR system img(Anaktisi) (Chatzichristofis et al., 2010) is aimed at a compact representation, which was extracted by fuzzy techniques applied to colour features and wavelet based texture description. However, texture representations in these systems are more or less supplemental and the algorithms rely on colour features. Although retrieval results look promising, they are often provided by enormous image databases than exact image indexing. It is quite simple to fill the first result page with very similar images from a large database (e.g. sunsets, beaches, etc.), nevertheless, the lack of image understanding is revealed on further result pages.

In narrow image domains, CBIR systems are more successful e.g. trademark retrieval (Leung and Chen, 2002; Wei et al., 2009; Phan and Androutsos, 2010), drug pill retrieval (Lee et al., 2010) or face detection (Lew and Huijsmans, 1996) and similarity, which evolved in a separate field.

One of the reasons of disregarding textural features are that they are still immature for a reliable representation (Deselaers et al., 2008) and at least weak texture segmentation of images is required (Smeulders et al., 2000). If the segmentation is extracted, shape features and region relations can be employed (Datta et al., 2008), however, the reliable segmentation is a difficult problem on its own. Recent methods avoid the image segmentation by local descriptors as SIFT (Lowe, 2004), which were extended to colour images and used for image indexing (van de Sande et al., 2010; Burghouts and Geusebroek, 2009a; Bosch et al., 2008). However these keypoint based descriptors are more suitable for description of objects without large textured faces than homogeneous texture areas.

The other reason for marginalising textures is that a more precise description of textures also requires more attention to expected variations of acquisition conditions. Many existing systems do not care about such variations or they handle it in a very limited way. Recently, Shotton et al. (2009) demonstrated that textural features can be successfully used for image understanding, if the variation of acquisition circumstances is considered.

1.1.2 Invariance

A representation is referred as invariant to a given set of acquisition conditions if it does not change with a variation of these conditions. The invariance property allows recognition of objects or textures in the real world, where the conditions during an image acquisition are usually variable and unknown. It is necessary to keep in mind that an undesired invariance to a broad range of conditions inevitably reduces the discriminability and aggravates the recognition. (An absurd example is the representation by a constant; it is invariant to all possible circumstances, but it has no use.) Consequently, the optimal image representation should be invariant to all expected variations of acquisition conditions and still it is required to remain highly discriminative, which are often contrary requirements.

Alternative ways how to deal with changing acquisition conditions are normalisation or learning from all possible examples. The normalisation transforms representation or features to a canonical form, e.g. image rotation according to dominant edges. The draw-

Figure 1.2: Examples of materials from the Amsterdam Library of Textures (ALOT) and their appearance for different camera and light conditions. The two columns on the right are acquired from viewpoint with declination angle 60° from the surface macro-normal.

back is that this approach may suffer from instability or ambiguity in detection of the canonical form, which results in imprecise or totally wrong normalisation. On the other hand, the learning from all possible appearances offers a robust representation, but it is extremely time consuming. It is applicable mainly in cases where some approximative appearance can be artificially generated, e.g. in-plane rotation of flat surfaces. Unfortunately, very often the required measurements are neither available nor possible to collect; or the measurements would be too expensive to acquire.

The appearance of rough materials is highly illumination and view angle dependent, as demonstrated in Fig. 1.2. Unfortunately, the appearance under different conditions cannot be easily generated, unless strong additional requirements are adopted (e.g. three precisely registered images of each material acquired with different and known illumination direction (Targhi et al., 2008)). Therefore we focus on creating a reliable texture representation, which is invariant or at least robust to variation of view angle and illumination conditions. Additional examples of material appearance changes are presented in Figs. B.2, B.5, and B.6 in the Appendix.

1.2   Thesis contribution

This work is focused on a query by and retrieval of homogeneous images (textures) and on the robustness against image acquisition conditions, namely illumination variation and texture rotation. It is believed that this thesis contributes to the field of pattern recognition with the following original work:

1. The main contribution is a set of novel illumination invariant features, which are derived from an efficient Markovian textural representation based on modelling by either Causal Autoregressive Random models (2D CAR, 3D CAR) or a Gaussian Markov Random Field (GMRF) model. These new features are proved to be invariant to illumination intensity and spectrum changes and also approximately invariant to local intensity changes (e.g. cast shadows). The invariants are efficiently implemented using parameter estimates and other statistics of CAR and GMRF models.

2. The illumination invariants are extended to be simultaneously rotation invariant. The rotation invariance is achieved either by moment invariants or by combination with circularly symmetric texture model.

Although the proposed invariant features are derived with the assumption of fixed viewpoint and illumination positions, our features exhibit significant robustness to illumination direction variation. This is confirmed in thorough experiments with measurements of Bidirectional Texture Function (BTF) (Dana et al., 1999), which is currently the most advanced representation of realistic material appearance. Moreover, no knowledge of illumination conditions is required and our methods work even with a single training image per texture. The proposed methods are also robust to image degradation with an additive Gaussian noise.

The proposed invariant representation of textures is tested in the task of texture retrieval and recognition under variation of acquisition conditions, including illumination changes and texture rotation. The experiments are performed on five different textural databases and the results are favourably compared with other state of the art illumination invariant methods. The psychophysical tests with our textural representation indicate its relation to the human perception of textures.

We utilise our features in a construction of system for retrieval of similar tiles, which can be used in decoration industry and we show feasible application in optimisation of parameters in texture compression used in computer graphics. Finally, our illumination invariants are integrated into a texture segmentation algorithm and our textural features are applied in the recognition of glaucomatous tissue in retina images.

We expect that the presented results can be used to improve the performance of existing CBIR systems or they can be utilised on their own in specialised CBIR systems concerning narrow domain images as medical images or the presented tile retrieval system. Other possible applications include computer vision, since analysis of real scenes inevitably includes description of textures under various light conditions.

1.3   Thesis organisation

The thesis is organised as follows: state of the art textural representations and textural databases are reviewed in the next chapter. The proposed textural representation is described in Chapter 3. Chapter 4 concerns with illumination invariance and it contains derivation of novel illumination invariants based on the proposed textural representation. In Chapter 5 rotation invariance is incorporated into the textural representation. Experimental results of the proposed methods are presented in Chapter 6 and applications follow in Chapter 7. Finally, the thesis is concluded and further directions of development are outlined. Appendices include additional derivations, experiments and examples from texture databases.

Chapter 2

State of the Art

Informally, a texture can be described as an image that consists of primitives (micro structures) placed under some placement rules, which may be randomised somehow. This texture primitive may be considered to be an object, and vice versa many objects may form a texture, it all depends on the resolution scale. Crucial properties of all textures are homogeneity and translation invariance. The homogeneity is understood quite vaguely and it means that any subwindow of a single texture posses some common characteristics. The translation invariance implies that these texture characteristics do not depend on texture translation. To name a few examples, an appearance of many materials or regular patterns is perceived as a texture.

Although the notion of texture is tied to human perception, there is no mathematically rigorous definition that would be widely accepted. In our work we assume that texture is a kind of random field and the texture image is the realisation of random field.

The following review of textural representations begins with known findings of human perception, continues with representations used in computers, and then these representations are considered according to invariant properties they provide. Finally, existing texture databases and comparisons are listed.

2.1   Human perception of textures

Julesz (1962) published one of the first works on visual texture discrimination, and he devoted next thirty years (Julesz, 1991) to work on human perception of textures, which was highly influential for construction of texture discrimination algorithms.

In order to explain the psychophysical findings, some image statistics have to be clarified (Julesz, 1962),

> "The nth-order statistic (or nth-order joint probability distribution) of an image can be obtained by randomly throwing n-gons of all possible shapes on the image and observing the probabilities that their n vertices fall on certain colour combinations."

The n-gons are geometrical objects: points (1-gon), line segments (2-gons, or dipoles), triangles (3-gons), etc.

Firstly, Julesz (1962) experimented with a spontaneous visual discrimination of textural images, which were generated by the Markov process as a realisation of a random field. He posed a conjecture that textures cannot be spontaneously discriminated if they have the same first-order and second-order statistics and if they differ only in their third or higher order statistics. However, this conjecture was later disproved when several counterexamples were published (Julesz et al., 1978; Yellot, 1993). Consequently, such images cannot be discriminated by texture recognition algorithms that rely only on first or second order statistics (e.g. histograms or co-occurrence matrices). Our textural features (Section 3.1) use higher order statistics, although their interaction range is locally limited, so we expect their ability to recognise even textures with identical second-order statistics.

Yellot (1993) also proved that the third-order statistics of any monochromatic image of finite size uniquely determine this image up to translation. Although Julesz et al. (1978); Julesz (1991) presented examples of distinguishable textures with same second-order and third-order statistics, Yellot (1993) argued that the actual sample third-order statistics were not identical. It is worth to stress that the theorem of Yellot (1993) does not claim that images with close statistics up to the third order look similar.

In later work, Julesz (1991) tended to characterise textures by small texture elements (textons) instead of global statistics. Similar paradigm was adopted by micropattern and texton based texture representations (Sections 2.2.4, 2.2.5). Julesz (1991) also demonstrated that texture discrimination is not symmetric: a small piece of one texture can be distinguished from another texture background, but if the textures are swapped the discriminability is weaker. Finally, the human texture discriminability is not linear in the sense that if an image with two highly discriminable textures is added to a homogeneous texture, the textures in the resulting image may be nondiscriminable, because the texture elements became too complex (Julesz, 1991).

Rao and Lohse (1996) performed a psychophysical experiment with 56 textures, where the subjects were asked to group the textures and to describe the characteristics of created groups. Rao and Lohse (1996) concluded that texture can be described in three orthonormal dimensions:

repetitive/regular/non-random	*vs.*	non-repetitive/irregular/random
granular/coarse/low-complexity	*vs.*	non-granular/fine/high-complexity
low contrast/directional	*vs.*	high contrast/non-directional.

Rao and Lohse (1996) argued that the joint axis of contrast and directionality is a new complex texture dimension, similarly as is the perception of colour hue (which can be decomposed into red–green and yellow–blue opponent components). However, we doubt about that and we would decompose this axis into two different properties.

Natural materials are recognised not only from the texture, but also from their reflectance properties as lightness and gloss. Fleming et al. (2003) showed that humans are

usually able to estimate these properties irrespective of natural illumination conditions, however some artificial illuminations can confuse the human perception system (Fleming et al., 2003).

Recent technological advances allow exploration of human perception by more elaborate techniques. Drucker et al. (2009); Drucker and Aguirre (2009) used functional Magnetic Resonance Imaging (fMRI) to explore perception of colour and shape. Or Filip et al. (2009) exploited gaze tracking device to identify salient areas on textured surfaces.

2.2 Computational representation of textures

Let us assume that a texture is defined on a rectangular lattice I and it is composed of C spectral planes measured by the corresponding sensors (usually {Red, Green, Blue}). Consequently, the texture image is composed of multispectral pixels with C components $Y_r = [Y_{r,1}, \ldots, Y_{r,C}]^T$, where pixel location $r = [r_1, r_2]$ is a multiindex composed of r_1 row and r_2 column index, respectively.

We are concerned in statistical texture representations, where the texture is characterised by a set of features extracted from the texture image. The alternative approach is the structural texture representation (Haralick, 1979; Vilnrotter et al., 1986), which characterises the texture by a set of texture primitives and their placement rules. The statistical texture representations can be divided into the following groups according to techniques they use. The techniques can utilise histograms, filters or transformation, patterns, modelling, combination of these approaches or they may offer perceptual interpretation. We list these groups with representative methods and after that popular textural features are described more thoroughly.

The first group is based on statistics computed directly from images, usually histograms (Stricker and Orengo, 1995) or co-occurrence matrices (Haralick, 1979) (see Section 2.2.1).

The second group is composed of methods, which use various filters or transformations to extract information from texture in a more convenient form. Subsequently, the texture is characterised by statistics computed from the filtered images. Various filters were described by (Randen and Husøy, 1999; Rivero-Moreno and Bres, 2004) including Gabor filters (Manjunath and Ma, 1996; Jain and Healey, 1998) (see Section 2.2.2). The transformations comprise wavelets (Jafari-Khouzani and Soltanian-Zadeh, 2005; Pun and Lee, 2003), wavelet packets (Laine and Fan, 1993), ridgelets, and curvelets (Semler and Dettori, 2006).

Pattern based methods characterise texture by a histogram of micropatterns (Ojala et al., 2002b) or texture elements – textons (Varma and Zisserman, 2005) (see Sections 2.2.4, 2.2.5).

Model based methods try to model texture with a local model, whose parameters are estimated from the texture image and the texture is characterised by these model parameters (Mao and Jain, 1992; Kashyap and Khotanzad, 1986; Deng and Clausi, 2004). The textural representation we propose belongs to this group of textural representations.

Some methods employ a combination of approaches as Wold features (Liu and Picard, 1996; Liu, 1997), which measure how much is an image structured or unstructured and which express the image as the combination of periodic/structured and random/unstructured parts. The structured texture component is represented by the most important frequencies in Fourier spectrum whereas the unstructured texture component is characterised by an autoregressive model (Mao and Jain, 1992). The texture randomness is estimated from autocovariance function and it is used as the weighting factor of periodic and random components. Liapis and Tziritas (2004) combined separate representations of colours and texture, characterised by histograms in CIE Lab space and wavelet features, respectively.

The questions whether colour and texture should be represented jointly or separately is discussed by Mäenpää and Pietikäinen (2004). They argued that colour and texture should be treated individually, and that many published comparisons do not take into account the size of feature vectors. We oppose this statement from two reasons:

1. relations among pixels with same luminance are lost in grey-scale images

2. a separate colour representation is not feasible in conditions with illumination colour variation, which Mäenpää and Pietikäinen (2004) admitted. In this case the interspectral texture relations play the crucial role.

Finally, we mention methods which offer perceptual interpretation of their features as most of the other textural features are difficult to interpret. A Six-stimulus theory by Geusebroek and Smeulders (2005) describes statistics of pixel contrasts by Weibull-distribution and the authors showed the relation of Weibull-distribution parameters with perceived texture properties as regularity, coarseness, contrast, directionality. Padilla et al. (2008) proposed a descriptor of roughness of 3D surface, which is in accordance with the perceived roughness. Mojsilovic et al. (2000) built colour pattern retrieval system using separate representation of colours and textures, where the similarity is based on rules inferred from human similarity judgements. However, the similarity evaluation was performed only on 25 patterns, which we consider insufficient for the inference of general pattern similarity. Alvarez et al. (2010) decomposed texture into blobs in the shape of ellipse and characterised the texture by a histogram of these blobs. This method is not able capture blobs relations or their interactions as crossings.

2.2.1 Histogram based features

The simplest features used with textures are based on histograms of colours or intensity values. However, these features cannot be considered as proper textural features, because they are not able to describe spatial relations which are the key texture properties. The advantage of histogram based features is their robustness to various geometrical transformations, fast and easy implementation.

Stricker and Orengo (1995) proposed **cumulative histogram**, which is defined as the distribution function of the image histogram, the i-th bin H_i is computed as

$$H_i = \sum_{\ell \leq i} h_\ell \ , \tag{2.1}$$

where h_ℓ is the ℓ-th bin of ordinary histogram. The distance between two cumulative histograms is computed in L_1 metric defined in formula (2.2). The cumulative histogram is more robust than the ordinary histogram, because a small intensity change characterised by a one-bin shift in the ordinary histogram, have only negligible effect on the cumulative histogram. Descriptors based on colour histograms and dominant colours are also part of MPEG-7 features (Manjunath et al., 2001).

Alternatively, colour histogram can be represented by its moments (Stricker and Orengo, 1995). Paschos et al. (2003) used CIE XYZ colour space to gain robustness to intensity changes.

Hadjidemetriou et al. (2004) proposed multiresolution histograms computed on levels of Gaussian-downsampled pyramid, which partially incorporated some spatial relations in the texture. The spatial relations are also described by the well-known co-occurrence matrices Haralick (1979), which contain probabilities that two intensity values occur in the given distance. An extension of the co-occurrence matrices to colour textures was proposed by Huang et al. (1997), who also added rotation invariance.

2.2.2 Gabor features

The Gabor features are based on Gabor filters (Bovik, 1991; Randen and Husøy, 1999), which are considered to be orientation and scale tunable edge and line detectors. The statistics of Gabor filter responses in a given region are, subsequently, used to characterise the underlying texture information.

The Gabor function is a harmonic oscillator, composed of a sinusoidal wave of particular frequency and orientation, within a Gaussian envelope. A two dimensional Gabor function $g(r) : \mathbb{R}^2 \to \mathbb{C}$ can be specified as

$$g(r) = \frac{1}{2\pi \ddot{\sigma}_{r_1} \ddot{\sigma}_{r_2}} \exp\left[-\frac{1}{2} \left(\frac{r_1^2}{\ddot{\sigma}_{r_1}^2} + \frac{r_2^2}{\ddot{\sigma}_{r_2}^2} \right) + 2\pi i \ddot{V} r_1 \right] \ ,$$

where i is the complex unit, $\ddot{\sigma}_{r_1}$, $\ddot{\sigma}_{r_2}$, $\ddot{V}$ are the filter parameters. $\ddot{\sigma}_{r_1}$, $\ddot{\sigma}_{r_2}$, are standard deviations of the Gaussian envelope and $\ddot{V}$ is related to the detected frequency. The Fourier transform of Gabor function is a multivariate Gaussian function

$$G(u) = \exp\left\{ -\frac{1}{2} \left[\frac{(u_1 - \ddot{V})^2}{\ddot{\sigma}_{u_1}^2} + \frac{u_2^2}{\ddot{\sigma}_{u_2}^2} \right] \right\} \ ,$$

where $\ddot{\sigma}_{u_1} = \frac{1}{2\pi\ddot{\sigma}_{r_1}}$, $\ddot{\sigma}_{u_2} = \frac{1}{2\pi\ddot{\sigma}_{r_2}}$ are standard deviations of the transformed Gaussian function and the vector $u = [u_1, u_2]$ is composed of coordinates u_1 and u_2.

As it was mentioned, the convolution of the Gabor filter and a texture image extracts edges of a given frequency and orientation range. The texture image is analysed with a set of filters (Manjunath and Ma, 1996) obtained by four dilatations and six rotations of the function $G(u)$. The filter set was designed so that Fourier transform of the filters cover most of the image spectrum, see Manjunath and Ma (1996) for more details. Finally, given a single spectral image with values $Y_{r,j}$, $r \in I$, $j = 1$, its Gabor wavelet transform is defined as

$$W_{k\phi,j}(r_1, r_2) = \int_{u_1, u_2 \in \mathbb{R}} Y_{r,j} \, g^*_{k\phi}(r_1 - u_1, r_2 - u_2) \, du_1 \, du_2 \; ,$$

where $(\cdot)^*$ indicates the complex conjugate, ϕ and k are orientation and scale of the filter. The convolution is implemented by means of Fast Fourier Transform (FFT), which complexity $O(n \log n)$ is dominant in computational time of Gabor features. Moreover, the Gabor filters are supposed to model early visual receptive fields (V1 cells), see Jones and Palmer (1987) for details .

Monochromatic Gabor features

The Monochromatic Gabor features (Manjunath and Ma, 1996; Ma and Manjunath, 1996), usually referred just as Gabor features, are defined as the mean and the standard deviation of the magnitude of filter responses $|W_{k\phi,j}|$. The straightforward extension to colour textures is computed separately for each spectral plane and concatenated into the feature vector, which is denoted with "RGB" suffix in the experiments.

The suggested distance between feature vectors of textures T, S is $L_{1\sigma}(T, S)$, which is a normalised version of Minkowski norm L_p:

$$L_p(T, S) = \left(\sum_{\ell=0}^{m} \left| f_\ell^{(T)} - f_\ell^{(S)} \right|^p \right)^{\frac{1}{p}} \; , \tag{2.2}$$

$$L_{p\sigma}(T, S) = \left(\sum_{\ell=0}^{m} \left| \frac{f_\ell^{(T)} - f_\ell^{(S)}}{\sigma(f_\ell)} \right|^p \right)^{\frac{1}{p}} \; , \tag{2.3}$$

$$\tag{2.4}$$

where m is the feature vector size, $f_\ell^{(T)}$ and $f_\ell^{(S)}$ are the ℓ-th components of feature vectors of textures T and S, respectively. $\sigma(f_\ell)$ is standard deviation of the feature f_ℓ computed over all textures in the database.

Alternatively, a histogram of mean filter responses was used (Squire et al., 2000) in image retrieval.

Opponent Gabor features

The opponent Gabor features (Jain and Healey, 1998) are an extension to colour textures, which analyses also relations between spectral channels. The monochrome part of these features is:

$$\varrho_{k\phi,j} = \sqrt{\sum_{r \in I} W_{k\phi,j}^2(r)} \; ,$$

where $W_{k\phi,j}$ is the response of Gabor filter $g_{k\phi}$ on the j-th spectral plane of colour texture T. The opponent part of features is:

$$\xi_{kk'\phi,jj'} = \sqrt{\sum_{r \in I} \left(\frac{W_{k\phi,j}(r)}{\varrho_{k\phi,j}} - \frac{W_{k'\phi,j'}(r)}{\varrho_{k'\phi,j'}} \right)^2} \; ,$$

for all j, j' with $j \neq j'$ and $|k - k'| \leq 1$. The previous formula could be also expressed as the correlation between spectral plane responses. Jain and Healey (1998) suggested computation of the distance of feature vectors using $L_{2\sigma}(T, S)$ normalised Minkowski norm (2.4).

Although, the Gabor features are widely used in computer vision applications, some authors reported them as non-optimal: Randen and Husøy (1999) who compared many filter based recognition techniques and Pietikäinen et al. (2002) in comparison with LBP features.

Generally, the Gabor features are translation invariant, but not rotation invariant. The rotation invariant Gabor features can be computed as the average of Gabor filter responses for the same scale, but different orientations, see Haley and Manjunath (1995). However, this averaging aggravates recognition of isotropic vs. anisotropic textures with similar statistics. An invariant object recognition based on Gabor features was described by Kamarainen et al. (2006), who also gave insightful notes for practical implementation.

As an analogy to Gabor filter modelling of visual receptive field, Bai et al. (2008) built filters in accordance with touch perception – tactical receptive field (TRF). The TRF is composed of three Gabor subfilters which relative positions and orientations are not fixed, therefore the filter for detection of particular orientation of edges is not a simple rotation of the basic filter, but also the relative positions of subfilters changes.

2.2.3 Steerable pyramid features

The steerable pyramid (Portilla and Simoncelli, 2000) is an over complete wavelet decomposition similar to the Gabor decomposition. The pyramid is built up of responses to steerable filters, where each level of pyramid extracts certain frequency range. All pyramid levels, except the highest and the lowest one, are further decomposed into different orientations. The transformation is implemented using the set of oriented complex analytic filters B_ϕ that are polar separable in the Fourier domain (see details in

Simoncelli and Portilla (1998); Portilla and Simoncelli (2000)):

$$B_\phi(R,\theta) = H(R)G_\phi(\theta), \qquad\qquad \phi \in [0, \Phi-1],$$

$$H(R) = \begin{cases} \cos\left(\frac{\pi}{2}\log_2\left(\frac{2R}{\pi}\right)\right), & \frac{\pi}{4} < R < \frac{\pi}{2} \\ 1, & R \geq \frac{\pi}{2} \\ 0, & R \leq \frac{\pi}{4} \end{cases}$$

$$G_\phi(\theta) = \begin{cases} \alpha_\Phi\left[\cos\left(\theta - \frac{\pi\phi}{\Phi}\right)\right]^{\Phi-1}, & \left|\theta - \frac{\pi\phi}{\Phi}\right| < \frac{\pi}{2}, \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha_\Phi = 2^{\Phi-1}\frac{(\Phi-1)!}{\sqrt{\Phi[2(\Phi-1)!]}}$; R and θ are polar frequency coordinates, $\Phi = 4$ is the number of orientation bands, and $K = 4$ is the number of pyramid levels. Like Gabor filters, the used wavelet transformation localises different frequencies under different orientations. Unlike Gabor filters, the inverse transformation can be computed as convolution with conjugate filters and therefore the synthesis is much faster.

Despite the decorrelation properties of wavelet decomposition, the coefficients are not statistically independent (Simoncelli, 1997), for instance large magnitude coefficients tend to occur at the same spatial relative position in subbands at adjacent scales, and orientations. Moreover, the coefficients of image wavelet subbands have non-Gaussian densities with long tails and sharp peak at zero. This non-Gaussian density is probably caused by the fact that images consists of smooth areas with occasional edges (Simoncelli and Portilla, 1998). The textural representation suggested by Portilla and Simoncelli (2000) comprise following features:

- **marginal statistics:** Skewness and kurtosis at each scale, variance of the high-pass band; and mean, variance, skewness, kurtosis, minimum and maximum values of the image pixels.

- **raw coefficient correlation:** Central samples of auto-correlation at each scale before the decomposition into orientations. These features characterise the salient spatial frequencies and the regularity of the texture, as represented by periodic or globally oriented structures.

- **coefficient magnitude statistics:** Central samples of the auto-correlation of magnitude of each subband; cross-correlation of each subband magnitudes with other orientations at the same scale, and cross-correlation of subband magnitudes with all orientation at a coarser scale. These features represent structures in images (e.g. edges, bars, corners), and "the second order" textures.

- **cross-scale phase statistics:** Cross-correlation of the real part of coefficients with both the real and imaginary part of the up-sampled coefficients at all orientations at the next coarser scale. These features distinguish edges from lines, and help in representing gradients due to shading and lighting effects.

The experiments in Portilla and Simoncelli (2000), were focused on texture synthesis and they were performed with $\Phi = 4$ orientation bands, $K = 4$ pyramid levels. In our

experiments, we used the same parameters, but we omitted the phase statistics, because they specifically describe shading and lighting effects, which are not desired. We computed the features on all spectral planes and compared the feature vectors with the $L_{1\sigma}$ norm defined by formula (2.4).

2.2.4 Local binary patterns

Local Binary Patterns (LBP) (Ojala et al., 1996) is a histogram of texture micro patterns. For each pixel, a circular neighbourhood around the pixel is sampled, and then the sampled values are thresholded by the central pixel value. Given a single spectral image with values $Y_{r,j}$, $r \in I$, $j = 1$, the pattern number is formed as follows:

$$LBP_{P,R} = \sum_{s \in I_r^\odot} sg\left(Y_{r-s,j} - Y_{r,j}\right) 2^{o(s)}, \qquad sg\left(x\right) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, \qquad (2.5)$$

where $I_r^\odot$ is the circular neighbourhood, which contains P samples in the radius R, $o(s)$ is the order number of sample position (starting with 0), and $sg(x)$ is the thresholding function. Subsequently, the histogram of patterns is computed and normalised to have unit L_1 norm. Because of thresholding, the features are invariant to any monotonic change of pixel values. The multiresolution analysis is done by growing the circular neighbourhood size. The similarity between feature vectors of textures T, S is defined by means of Kullback-Leibler divergence.

$$L_G(T, S) = \sum_{\ell=1}^{m} f_\ell^{(T)} \log_2 \frac{f_\ell^{(T)}}{f_\ell^{(S)}},$$

$f_\ell^{(T)}$ and $f_\ell^{(S)}$ are the ℓ-th components of feature vectors of textures T and S, respectively.

Uniform LBP

A drawback of the original LBP features is that complex patterns usually do not have enough occurrences in a texture, which introduces a statistical error. Therefore Ojala et al. (2002b) proposed the uniform LBP features, denoted as LBP^{u2}, which distinguish only among patterns that include only 2 or less transitions between 0 and 1 at neighbouring bits in formula (2.5). The formalisation of the number of bit transitions for the particular pattern is:

$$U(LBP_{P,R}) = \sum_{\substack{s,t \in I_r^\odot \\ o(t)=0 \,\wedge\, o(s)=P-1}} \left|sg\left(Y_{r-s,j} - Y_{r,j}\right) - sg\left(Y_{r-t,j} - Y_{r,j}\right)\right|$$

$$+ \sum_{\substack{s,t \in I_r^\odot \\ o(t)-o(s)=1}} \left|sg\left(Y_{r-s,j} - Y_{r,j}\right) - sg\left(Y_{r-t,j} - Y_{r,j}\right)\right| \ .$$

Actually, the patterns distinguished by LBP^{u2} are single arcs, which differ only in their length or position in the circular neighbourhood $I_r^{\odot}$. See Ojala et al. (2002b) for implementation details.

The uniform LBP features can be also made rotation invariant (Ojala et al., 2002b). These features are denoted as $LBP_{P,R}^{riu2}$ and they consider uniform patterns regardless their orientations. The pattern number is, consequently, defined as

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{s \in I_r^{\odot}} sg\left(Y_{r-s,j} - Y_{r,j}\right) & \text{if } U\left(LBP_{P,R}\right) \leq 2 \\ P+1 & \text{otherwise.} \end{cases}$$

In fact, the pattern number of $LBP_{P,R}^{riu2}$ is the number of bits with value 1.

The LBP features were straightforwardly extended to colour textures by computation on each spectral plane separately, they are denoted by "LBP, RGB" (Mäenpää and Pietikäinen, 2004).

The best results were reported (Maenpaa et al., 2002; Pietikäinen et al., 2002) with "$LBP_{16,2}^{u2}$" and "$LBP_{8,1+8,3}$", which is combination of features "$LBP_{8,1}$" and "$LBP_{8,3}$". The comparison was performed on the test with illumination changes (test suite OUTEX_TC_00014), where they outperformed Gabor features. In the test with additional rotation invariance (test suite OUTEX_TC_00012), the best results were achieved with "$LBP_{16,2}^{riu2}$" and "$LBP_{8,1+24,3}^{riu2}$" features (Ojala et al., 2002b). However, they were outperformed by LBP-HF (Ahonen et al., 2009) described later.

LBP-HF

Local Binary Pattern Histogram Fourier features (LBP-HF), which were introduced by (Ahonen et al., 2009), are based on the rotation invariant $LBP_{P,R}^{riu2}$. Additionally, they analyse the mutual relations of orientations of each micropattern.

At first, a histogram of occurrences is computed for a single uniform pattern and all its rotations. Subsequently, Discrete Fourier Transformation (DFT) is computed from the histogram and the amplitudes of Fourier coefficients are the rotation invariant features. These features are computed for all uniform patterns.

The authors' implementation is provided in MATLAB at (implementation LBP). Ahonen et al. (2009) reported LBP-HF features to be superior to $LBP_{P,R}^{riu2}$ in rotation invariant texture recognition.

In general, the LBP features are very popular, because they are effective, easy to implement and fast to compute. However, if bilinear interpolation of samples is employed, it slows down computation significantly. The main drawback of the LBP features is their noise sensitivity (Vacha and Haindl, 2007a). This vulnerability was addressed by Liao et al. (2009), but used patterns are specifically selected according to the training set, which is not suitable for general purpose textural features. He et al. (2008) proposed Bayesian Local Binary Pattern (BLBP), which introduced smoothing of detected micropatterns before computation of their histogram. However, the used Potts model

and graph cut minimisation is very time demanding in comparison with other textural representations.

2.2.5 Textons

Texton representation proposed by Leung and Malik (2001); Varma and Zisserman (2005) characterizes textures by histogram of texture micro-primitives called textons. The textons are acquired during learning stage, when all available images are convolved with the chosen filter set to generate filter responses. The filter responses are subsequently clustered and the cluster representatives are the textons.

During the classification stage, the filter responses for the given pixel are computed and the pixel is assigned to the texton number with the most similar filter responses. The texture is characterised by the texton histogram, which is normalised to have unit L_1 norm, and the similarity of histograms is evaluated with χ^2 statistic.

MR8-*

The previous texton representation was modified to be rotation invariant by Varma and Zisserman (2005) who recorded only the maximal response of different orientations of the same filter, the method is denoted as VZ MR8. Recording of maximal responses is advantageous compared to the averaging over filter orientations, because it enables to distinguish between isotropic and anisotropic textures. The co-occurrence statistics of relative orientations of maximal response filters can be studied as well, but it may be unstable and noise sensitive (Varma and Zisserman, 2005).

Partial illumination invariance is achieved by an image normalisation to zero mean and unit standard variation. Of course, each filter is L_1 normalised so that the responses of each filter lie roughly in the same range.

Later on, Varma and Zisserman (2009) demonstrated that filters are not necessary. They took VZ MR8 algorithm and replaced the filter responses by image patches, consequently, the textons were learned from these image patches. Quite surprisingly, the recognition accuracy remained the same or even improved, however, this modification is no more rotation invariant.

The VZ MR8 algorithm was extended by Burghouts and Geusebroek (2009b) to incorporate colour information and to be colour and illumination invariant. The extension is based on the Gaussian opponent colour model (Geusebroek et al., 2001), which separates colour information into intensity, yellow–blue, and red–green channels when applied to RGB images. Four modifications were proposed differing in range of illumination invariance:

MR8-NC applies VZ algorithm to the Gaussian opponent colour model (Geusebroek et al., 2001), which is computed directly from RGB pixel values. Since the VZ algorithm normalizes each channel separately, the method normalises colours, however, it also discards chromaticity in the image.

MR8-INC normalises all channels by variance of intensity channel and therefore preserves chromaticity.

MR8-LINC is similar to MR8-INC, but the normalisation is performed locally instead of global one.

MR8-SLINC additionally to MR8-LINC, it offers shadow and shading invariance.

In the experiments of Burghouts and Geusebroek (2009b), MR8-NC and MR8-LINC features were reported with the best performance. Shotton et al. (2009) applied MR8-NC features in a texture based representation of images.

2.3 Invariance

As it was mentioned, textures are important clues to specify objects present in a visual scene. Unfortunately, the visual appearance of natural materials is highly dependent on circumstances during image acquisition, which complicates their recognition. We review methods that are able to recognise textures under variations of these conditions and group them according to provided invariance.

2.3.1 Illumination invariance

Appearance of natural materials, especially with rough surfaces, significantly varies with illumination direction and viewpoint position (Chantler, 1995; Dana et al., 1999). Without illumination invariance, recognition methods require multiple training images captured under a full variety of possible illumination and viewing conditions for each class (Varma and Zisserman, 2005; Suen and Healey, 2000). Such learning is obviously clumsy and very often even impossible if required measurements are not available.

Intensity invariance

Invariance to global change of illumination intensity can be achieved by simple normalisation of intensity channel, analogically to formula (2.6). LBP (Ojala et al., 2002b) are invariant to any monotonic changes of illumination intensity. This includes changes caused by histogram equalization or the gamma correction, which is usually applied to image data during the coding process. Because of local nature of the LBP features, they successfully cope with local intensity changes. Masotti and Campanini (2008) took advantage of ranklet transformation to introduce textural features, which are also invariant to any monotonic changes of illumination intensity. Since the ranklet transformation exploits ordering of values, unlike thresholding in LBP, the feature can be expected to be less noise sensitive than LBP.

Spectrum invariance

Colour constancy algorithms, represented by Finlayson (1995), attempt to recover the image illuminated by some standard illumination, which is an unnecessarily complex task and it induces additional assumptions on a recognised scene.

Channel normalisation is a simple method, which achieve invariance to global change of illumination spectrum. The image spectral planes are normalised as follows

$$Y'_{r,j} = \frac{Y_{r,j}}{\sum_{t \in I} Y_{t,j}} \qquad \forall j = 1, \ldots, C \ . \tag{2.6}$$

In our experiments, we tested this normalisation prior to computation of Gabor features, opponent Gabor features or steerable pyramid features in order to make them comparable to another illumination invariant methods.

Healey and Wang (1995); Yang and Al-Rawi (2005) employed properties of correlation functions between different spectral channels to achieve invariance to illumination spectrum changes. Geusebroek et al. (2001) introduced colour invariants, some of which are even invariant to local changes or shadow and shading effects. They were also employed in colour invariant modifications of MR8 textons (Burghouts and Geusebroek, 2009b), which were previously described in Section 2.2.5. Hoang et al. (2005) combined these invariants with the Gabor filters in texture segmentation algorithm.

Illumination direction invariance

Drbohlav and Chantler (2005) solved similarity of images under different illumination direction, however, they require surfaces of uniform albedo, smooth and shallow relief and most importantly the knowledge of illumination direction for all involved (trained as well as tested) textures. This method bypasses the theoretical limitation proved by Jacobs et al. (1998); Chen et al. (2000),

> "Discounting interreflections, all illumination invariants of objects with Lambertian reflectance under point light sources at infinity are nondiscriminative."

However, probability based discrimination is still possible (Chen et al., 2000). Chantler et al. (2005) proposed texture classification with simultaneous estimation of texture direction.

Fortunately, multispectral images additionally offer relations of spectral planes. From the theoretical point of view, the invariance to illumination direction could be achieved with pixel-wise intensity normalisation. However, many images include very dark areas, where the normalisation of RGB triplets would produce not only a huge amount of noise, but it is also undefined for black pixels. Alternatively, invariants based on pixel hues are ambiguous on the black-white axis and they are not suitable for rough surfaces with uniform colours. The normalisation proposed by Finlyason and Xu (2002) is able to cancel changes of illumination colour, lighting changes caused by the surface shape and even the gamma correction. However, since the method normalises lighting changes

caused by the surface shape it completely wipes out the structure of rough textures and therefore it destroys the possibility to distinguish such textures. An interesting approach of quasi-invariants (van de Weijer et al., 2005) relieves the condition of full invariance and therefore it is less sensitive to noise.

Another approach (Targhi et al., 2008) generated unseen training images using the photometric stereo approach. Although it improves classification accuracy, this algorithm has strong requirements of three mutually registered images with different illumination direction for each material.

2.3.2 Rotation invariance

Rotation invariant textural features can be divided into two main groups. The first group contains filter based features such as mentioned Gabor features (Haley and Manjunath, 1999); Circular-Mellin features (Ravichandran and Trivedi, 1995) that are additionally scale insensitive. The second group composes of model based features, which were in the invariant form introduced by Kashyap and Khotanzad (1986), who used an autoregressive model of pixel value and averages on concentric circles around it. The disadvantage of this model and its multiresolution extension (Mao and Jain, 1992) is the insensitivity to anisotropic texture properties. This weakness was removed in Anisotropic Circular Gaussian Markov Random Field (ACGMRF) model by Deng and Clausi (2004), which computes the Fourier descriptors of estimated model parameters. Alternatively, the texture anisotropy can be estimated together with symmetry and texture regularity (Chetverikov and Haralick, 1995).

Another possibility to deal with rotated textures is a rotation normalisation (Jafari-Khouzani and Soltanian-Zadeh, 2005), where the principal texture direction is detected by the Radon transformation. Still, a detection of the principal direction may be ambiguous and it can be influenced by the illumination direction.

Rotation invariance and illumination invariance

A combined rotation and illumination invariance is the property of the following, previously mentioned, features: LBP^{riu2}, LBP-HF, which are invariant to any monotonous changes of image values. Variants of MR8-* texton method provide various degrees of illumination invariance – from colour invariance to cast shadows and shading. Also multiresolution histograms (Hadjidemetriou et al., 2004) are rotation and intensity invariant.

Local SIFT descriptor (Lowe, 2004), which is rotation normalised and scale robust, was extended to incorporate colour invariants Burghouts and Geusebroek (2009a).

It is worth to note that the rotation of materials with rough surfaces cannot be modelled as a simple image rotation, because their appearance also depends on light position (Chantler, 1995). Consequently, additional assumptions or measurements are required: McGunnigle and Chantler (1999) require multiple mutually registered images to apply Photometric stereo and to reconstruct surface height variation. Or rotation invariants can be combined with the mentioned approach of Drbohlav and Chantler (2005), which require uniform albedo surface and known light direction.

2.3.3 Other invariances

Invariance to viewpoint position is provided by Schaffalitzky and Zisserman (2001), where the perspective projection is locally modelled as an affine transformation of texture patches. The patches are normalised based on the covariance matrix of gradient image and, subsequently, any rotation invariant texture descriptor can be used. Invariance to linear intensity changes of grey-scale images is incorporated as well. The method proposed by Lazebnik et al. (2005) describes a texture by histograms of textons computed over local regions that are transformed to be affine invariant. Diplaros et al. (2006) proposed combination of colour and shape for object recognition with illumination-viewpoint invariance.

Rotation and scale invariant features were introduced by Pun and Lee (2003) who computed log-polar transformation of input image to achieve the invariance and, subsequently, texture is represented by statistics of wavelet subbands. Xu et al. (2006) introduced multifractal spectrum (MFS) features, which are invariant to linear intensity changes and smooth texture transformations including perspective projection and texture warping.

Ojansivu et al. (2008) introduced Local Phase Quantization (LPQ) features, which are based on locally computed Fourier transform and which are rotation invariant and blur insensitive. The authors showed significant improvement over LBP and Gabor features in the recognition of blurred textures.

Finally, it is possible to recognize textures whose appearance changes in time (Fazekas et al., 2009; Zhao and Pietikainen, 2007), examples of these dynamic textures are leaves in wind, flowing river, burning fire, etc.

2.4 Texture databases

The following databases or datasets are freely available for comparison of texture analysis algorithms. The databases are categorised according to variation of conditions, used in acquisition of material samples.

No variation

At first, we list traditional texture databases, which do not include any significant variation of acquisition conditions or these conditions are not explored systematically.

The oldest database is Brodatz texture database (Brodatz, 1966), which contain grey-scale images and it is freely available for research purposes at (database Brodatz). VisTex (database VisTex) contains colour textures, however, it is no longer maintained. MeasTex (Ohanian and Dubes, 1992) is a framework for evaluation of texture classification algorithms, which includes grey-scale subimages of Brodatz and Vistex textures, it is available at (database MeasTex). We consider all previous databases as obsolete and it is quite surprising how often experiments are performed only on the Brodatz database, despite the fact that much better texture databases are available.

Illumination variation

The following databases include variations of illumination spectrum, illumination direction or both of them. Some databases also include rotation or viewpoint variation.

The best variations of illumination spectra is comprised in Outex database (Ojala et al., 2002a), which consists of colour texture images acquired under three different illumination spectra and nine in-plane rotations. Outex also define several classification tests, which differ in recognition conditions.

Illumination direction variations are included in the following databases. Photex texture database (database PhoTex) is focused on rough textures illuminated from different directions and it contains 12-bit monochrome images. Most of the materials have uniform albedo and they differ only in surface structure, which makes the recognition challenging. Columbia-Utrecht Reflectance and Texture Database (CUReT) by Dana et al. (1999) comprise materials acquired under combinations of viewing and illumination directions. Based on CUReT, Varma and Zisserman (2005) selected a dataset used for recognition tests, the dataset can be downloaded from (dataset CUReT). University of Bonn material measurements (Meseth et al., 2003) contains colour images of materials acquired in a controlled setup of 81 viewpoints times 81 illumination directions. Some of the measurements are now publicly available (database Bonn BTF).

Combined spectrum and direction variations are properties the following two databases. KTH-TIPS2 database (Caputo et al., 2005) includes colour samples with different scales and rotations, three illumination directions and one additional illumination spectrum. KTH-TIPS2 database can be freely downloaded from (database KTH-TIPS2). The newest and the most extensive texture database is Amsterdam Library of Textures (ALOT) (Burghouts and Geusebroek, 2009b), which comprise of 250 natural materials, each acquired with varying viewpoint and illumination positions, and one additional illumination spectrum. The ALOT database can be freely downloaded from (database ALOT).

3D shape and perspective projection

Finally, we list three databases with samples of textured non-planar surfaces. Microsoft textile database (database Microsoft Textiles) contains colour images of folded textile materials, which include nonlinear deformations and also some materials exhibit significant non-labertian reflection effects. The experimental results on this database were published by Varma and Zisserman (2009). UIUC texture database (Lazebnik et al., 2005) contains grey-scale images of materials acquired under significant viewpoint and scale variations. The acquisition conditions were uncontrolled and some materials were captured on non-planar surfaces or under substantial perspective projection. The database is freely available at (database UIUC). University of Maryland texture database (Xu et al., 2006) comprise grey-scale images of materials with significant scale variations and perspective projection, it is available online at (database Univeristy of Maryland).

2.5 Comparison

Randen and Husøy (1999) compared many classical textural features in recognition of grey-scales images from Brodatz, Vistex and Meastex texture databases. They concluded that the best performance was achieved by their own Quadrate Minor Filters (QMF) (Randen and Husoy, 1994) and they also stated that Gabor filters and co-occurrence are clearly not superior. Of course, newer features as LPB and MR8-* textons were not included.

A comparison of local descriptors was performed by Mikolajczyk and Schmid (2005) with conclusion that SIFT features performed the best. Burghouts and Geusebroek (2009a); van de Sande et al. (2010) compared colour extensions of SIFT features in object recognition tests with variable illumination conditions. Although local descriptors are mainly used for object recognition, they were also employed to retrieve textured surfaces (Sivic and Zisserman, 2004).

Deselaers et al. (2008) compared various features in image retrieval task performed on five different image databases. They concluded,

> "One clear finding is that colour histogram, often cited as a baseline in CBIR, clearly are a reasonable good baseline for general photographs. However, approaches using local image descriptors outperform colour histogram in various tasks but usually at the cost of much higher computational costs. ... Furthermore, it has been shown that, despite more than 30 years in research on texture descriptors, still none of the texture features presented can convey a complete description of the texture properties of an image. "

However, tested textural features comprised: Tamura features, co-occurrence matrices, Gabor features, and computationally inexpensive MPEG-7 features (Manjunath et al., 2001). Newer textural features as LBP and MR8-* textons were not included. Moreover, most of the textural features were computed globally from whole images, without any prior segmentation. This is not suitable setup, because textural features were averaged over areas with different textures.

Chapter 3

Textural Features

Our textural representation takes advantage of a fast estimate of Markovian statistics. A texture is locally modelled by an Markov Random Field (MRF) model and the model parameters become the texture characterisation (Haindl and Vacha, 2006; Vacha and Haindl, 2007a). A special wide sense Markov model is used, because it enables a fast analytical estimate of its parameters and thus to avoid time-consuming Monte Carlo minimisation prevailing in most of alternative MRF models (Gimel'Farb, 1999; Li, 2009).

The proposed model resembles MultiResolution Simultaneous AutoRegressive model (MR-SAR) described by Mao and Jain (1992). The difference is that we use restricted shape of neighbourhood, which allows efficient parameter estimation and we model interaction of colours.

3.1 Markov random field textural representation

Let us recall that a texture is defined on a rectangular lattice I and it is composed of C spectral planes measured by the corresponding sensors (usually {Red, Green, Blue}). The multispectral pixels are $Y_r = [Y_{r,1}, \ldots, Y_{r,C}]^T$, where pixel location $r = [r_1, r_2]$ is a multiindex composed of r_1 row and r_2 column index, respectively.

The texture analysis starts with a spatial factorisation of the texture into K levels of the Gaussian down-sampled pyramid. All spectral planes are factorised using the same pyramid. Each pyramid level is either modelled by a 3-dimensional MRF model or a set of C 2-dimensional MRF models. In case of 2D models the image spectral planes are mutually decorrelated by Karhunen-Loève transformation (Principal Component Analysis – PCA) prior to the spatial factorisation by the pyramid. The MRF model parameters are estimated, optionally illumination/colour or rotation invariants are computed, and textural features are formed from them. Finally, the textural features from all the pyramid levels are concatenated into a common feature vector. The overview of the texture analysis algorithm with a set of 2D models is displayed in Fig. 3.1.

The rest of the section introduces textural features based on three different models: 3-dimensional Causal Autoregressive Random field (3D CAR), 2-dimensional Causal Autoregressive Random field (2D CAR), and 2-dimensional Gaussian Markov Random
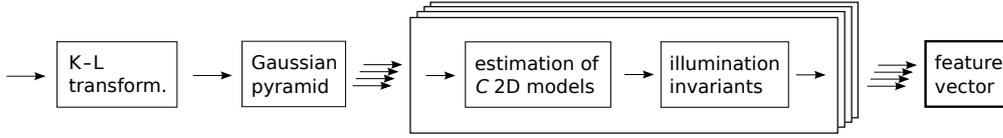
Figure 3.1: Texture analysis algorithm by means of a set of 2D models and with computation of illumination invariants.

Field (GMRF). The construction of illumination/colour and rotation invariant textural features is presented in the consecutive chapters.

3.1.1   Karhunen-Loève transformation

Karhunen-Loève transformation (K-L transformation) is a projection of image values, which decorrelates image spectral planes. K-L transformation is used prior to modelling by 2-dimensional (2D) models, because they are not able to model interspectral relations.

The vectors Y_r are mean centred $\dot{Y}_r$ and projected into a new coordinate axes $\bar{Y}_r$. These new basis vectors are eigenvectors of the second-order statistical moment matrix

$$\Xi = \mathbb{E}\left\{\dot{Y}_r \dot{Y}_r^{T}\right\} \ .$$

The projection of centred vector $\dot{Y}_r$ onto the K-L coordinate system uses the transformation matrix

$$T = [\dot{u}_1, \dot{u}_2, \ldots, \dot{u}_C]^{T} \ , \tag{3.1}$$

where columns vectors $\dot{u}_j$ are eigenvectors of the matrix Ξ:

$$\bar{Y}_r = T \dot{Y}_r \ . \tag{3.2}$$

Components of the transformed vector $\bar{Y}_r$ are mutually decorrelated (covariance matrix $\mathbb{E}\{\bar{Y}_r \bar{Y}_r^{T}\}$ is diagonal). If we further assume that random vectors $\bar{Y}_r$ are Gaussian, the components are also independent and they can be independently modelled by monospectral random fields.

3.1.2   Gaussian down-sampled pyramid

The Gaussian pyramid is a sequence of images in which each one is a low-pass down-sampled version of its predecessor. The employed Gaussian filter is approximated by the weighting function (Finite Impulse Response – FIR generating kernel) w which is chosen to comply with (Burt, 1983):

separability	$w_s = \dot{w}_{s_1} \dot{w}_{s_2}$
normalization	$\sum_{\ell=-\dot{m}}^{\dot{m}} \dot{w}_\ell = 1$
symmetry	$\dot{w}_\ell = \dot{w}_{-\ell}$
equal contribution	$\dot{w}_0 = 2\dot{w}_1 \qquad (\dot{m} = 1)$

where $\dot{m}$ bounds support of the kernel function and multiindex $s = [s1, s2]$ is composed of s_1 row and s_2 column index. The equal contribution constraint requires that all nodes at the given level contribute the same total weight to the nodes at the next higher level. The solution of above constraints for the kernel size 3×3 ($\dot{m} = 1$) is $\dot{w}_0 = 0.5$, $\dot{w}_1 = 0.25$.

The Gaussian pyramid for reduction factor n (for $n = 2$ the $N \times N$ image is down-sampled to $\frac{N}{2} \times \frac{N}{2}$) is defined as

$$\ddot{Y}_{\bullet,j}^{(k)} = Y_{\bullet,j} \ , \qquad\qquad\qquad\qquad k = 1 \ ,$$
$$\ddot{Y}_{r,j}^{(k)} = \downarrow^n (\ddot{Y}_{\bullet,j}^{(k-1)} \otimes w) \ , \qquad\qquad k = 2, \ldots, K \ ,$$

where $\ddot{Y}_{r,j}^{(k)}$ is the j-th spectral plane at the pixel position r of k-th pyramid level, the operator $\downarrow^n$ denotes down-sampling with the reduction factor n and $\otimes$ is the convolution operation. Convolution can be substituted using

$$\ddot{Y}_{r,j}^{(k)} = \sum_{s_1,s_2=-\dot{m}}^{\dot{m}} \dot{w}_{s_1} \dot{w}_{s_2} \ddot{Y}_{nr+(s_1,s_2),j}^{(k-1)} \ .$$

This multiscale pyramid approach is employed, because it allows us to incorporate larger spatial relations with smaller models, which have more concise and robust parameter sets than larger models.

3.1.3   3D causal autoregressive random field

The each level of Gaussian pyramid level is modelled separately and in the same way. Therefore we omit the level index k and we work generally with multispectral texture pixels Y_r.

The 3D CAR representation assumes that the multispectral texture pixel Y_r can be locally modelled by a 3D CAR model (Haindl and Šimberová, 1992) as a linear combination of neighbouring pixels. The shape of contextual neighbourhood is restricted to causal or unilateral neighbourhood, which allows efficient parameter estimation (see examples Fig. 3.2).

We denote I_r a selected contextual causal or unilateral neighbour index shift set and its cardinality $\eta = |I_r|$. Let Z_r is a $C\eta \times 1$ data vector, which consists of neighbour pixel values for a given pixel position r:

$$Z_r = [Y_{r-s}^T : \forall s \in I_r]^T \tag{3.3}$$

where r, s are multiindices. The matrix form of the 3D CAR model is:

$$Y_r = \gamma Z_r + \epsilon_r \ , \tag{3.4}$$

where $\gamma = [A_s : s \in I_r]$ is the $C \times C\eta$ unknown parameter matrix with square submatrices A_s. The white noise vector ϵ_r has zero mean and constant but unknown covariance matrix Σ. Moreover, we assume the probability density of ϵ_r to have the normal distribution independent of previous data and being the same for every position r.
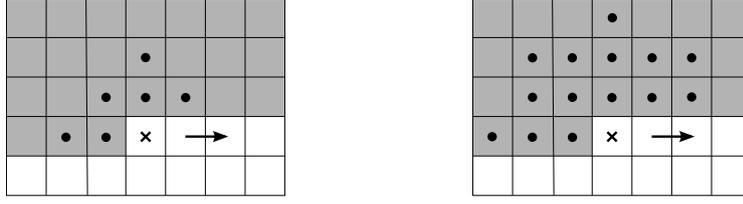
Figure 3.2: Examples of contextual neighbourhood I_r. From the left, it is the unilateral hierarchical neighbourhood of third and sixth order. X marks the current pixel, the bullets are pixels in the neighbourhood, the arrow shows movement direction, and the grey area indicate permitted pixels. The causal neighbourhood is a subset of unilateral neighbourhood which includes only pixels in the upper left quadrant from X.

Parameter estimation

The texture is analysed in a chosen direction, where multiindex t changes according to the movement on the image lattice e.g. $t - 1 = (t_1, t_2 - 1)$, $t - 2 = (t_1, t_2 - 2), \ldots$. The task consists in finding the parameter conditional density $p(\gamma \,|\, Y^{(t-1)})$ given the known process history $Y^{(t-1)} = \{Y_{t-1}, Y_{t-2}, \ldots, Y_1, Z_t, Z_{t-1}, \ldots, Z_1\}$ and taking its conditional mean as the textural feature representation. Assuming normality of the white noise component ϵ_t, conditional independence between pixels and the normal-Wishart parameter prior, it was shown (Haindl and Šimberová, 1992) that the conditional mean value is:

$$\mathbb{E}[\gamma \,|\, Y^{(t-1)}] = \hat{\gamma}_{t-1} \ ,$$

where the following notation is used:

$$\hat{\gamma}_{t-1}^T = V_{zz(t-1)}^{-1} \, V_{zy(t-1)} \ , \tag{3.5}$$

$$V_{t-1} = \begin{pmatrix} \sum_{r=1}^{t-1} Y_r Y_r^T & \sum_{r=1}^{t-1} Y_r Z_r^T \\ \sum_{r=1}^{t-1} Z_r Y_r^T & \sum_{r=1}^{t-1} Z_r Z_r^T \end{pmatrix} + V_0$$

$$= \begin{pmatrix} V_{yy(t-1)} & V_{zy(t-1)}^T \\ V_{zy(t-1)} & V_{zz(t-1)} \end{pmatrix} \ , \tag{3.6}$$

and V_0 is a positive definite matrix representing prior knowledge, e.g. identity matrix $V_0 = 1_{C\eta+C}$ for uniform prior. Noise covariance matrix Σ is estimated as

$$\hat{\Sigma}_{t-1} = \frac{\lambda_{t-1}}{\psi(t)} \ ,$$

$$\lambda_{t-1} = V_{yy(t-1)} - V_{zy(t-1)}^T \, V_{zz(t-1)}^{-1} \, V_{zy(t-1)} \ , \tag{3.7}$$

$$\psi(t) = \psi(t-1) + 1 \ , \ \psi(0) > 1 \ .$$

The parameter estimation $\hat{\gamma}_t$ can be accomplished using fast, numerically robust and recursive statistics (Haindl and Šimberová, 1992):

$$\hat{\gamma}_t^T = \hat{\gamma}_{t-1}^T + \frac{V_{zz(t-1)}^{-1} Z_t (Y_t - \hat{\gamma}_{t-1} Z_t)^T}{1 + Z_t^T V_{zz(t-1)}^{-1} Z_t} \quad , \tag{3.8}$$

and λ_t can be evaluated recursively too. The numerical realisation of the model statistics (3.5) – (3.8) is discussed in Haindl and Šimberová (1992). In principle, the parameter estimation process is very efficient, because matrix $V_{zz(t-1)}^{-1}$ is kept and updated in the form of Cholesky decomposition, which avoids computation of full matrix inverse. The computational complexity of parameter estimation process is linear with respect to the number of analysed pixels and quadratic in the size of contextual neighbourhood (data vector).

Alternatively, the model parameters can be estimated by means of Least Squares (LS) estimation, which minimise sum of prediction error squares:

$$\hat{\gamma}_t = \arg \min_{\gamma} \left\{ \sum_{r=1}^{t} (Y_r - \gamma Z_r)^T (Y_r - \gamma Z_r) \right\} \quad . \tag{3.9}$$

The estimation leads to the formally same equations as (3.5) – (3.7) with zero matrix $V_0 = 0_{C\eta+C}$.

Both methods for the parameter estimation (Bayesian and LS) have to deal with boundary conditions. Either the texture is periodically duplicated, which is related to a toroidal image lattice. Or the estimate is performed on a subset $J \subset I$ of the image lattice so that

$$\forall r \in J \ \wedge \ \forall s \in I_r \Rightarrow r + s \in I \ , \tag{3.10}$$

all data vectors lie in the image lattice. Moreover, it is advantageous to estimate the model parameters on the mean centred values, which simplifies the modelling. The original data can be whenever reconstructed with mean addition.

After the estimation of model parameters, the pixel prediction probability can be computed:

$$p\left(Y_t | Y^{(t-1)}\right) = \frac{\Gamma\left(\frac{\psi(t)-C\eta+C+2}{2}\right)}{\Gamma\left(\frac{\psi(t)-C\eta+2}{2}\right) \pi^{\frac{C}{2}} \left(1 + Z_t^T V_{zz(t-1)}^{-1} Z_t\right)^{\frac{C}{2}} |\lambda_{t-1}|^{\frac{1}{2}}}$$

$$\left(1 + \frac{(Y_t - \hat{\gamma}_{t-1} Z_t)^T \lambda_{t-1}^{-1} (Y_t - \hat{\gamma}_{t-1} Z_t)}{1 + Z_t^T V_{zz(t-1)}^{-1} Z_t}\right)^{\frac{\psi(t)-C\eta+C+2}{2}} \quad , \tag{3.11}$$

where $\Gamma(x)$ is the Gamma function.

Optimal support set estimation

The optimal contextual neighbourhood I_r can be found analytically by maximising the corresponding posterior probability $p(M_\ell | Y^{(t-1)})$, where model M_ℓ uses contextual

neighbourhood I_r^ℓ. Using the Bayesian formula, the most probable model can be selected without computing of normalisation constant. Therefore, the maximum of $p(M_\ell|Y^{(t-1)})$ can be found by maximising of $p(Y^{(t-1)}|M_\ell)$ or its logarithm. If we assume uniform model priors (Haindl and Šimberová, 1992), the optimal model can be found by maximising:

$$\ln p(Y^{(t-1)}|M_\ell) = K_1(\psi(t-1)) \; - \; \frac{C}{2} \ln \left| V_{zz(t-1)} \right| \; - \; \frac{\psi(t) - C\eta + C + 1}{2} \ln |\lambda_{t-1}|$$

$$+ \frac{C^2\eta}{2} \ln \pi$$

$$+ \sum_{m=1}^{C} \left[\ln \Gamma \left(\frac{\psi(t) - C\eta + C + 2 - m}{2} \right) - \ln \Gamma \left(\frac{\psi(0) - C\eta + C + 2 - m}{2} \right) \right] \;, \quad (3.12)$$

$K_1(\psi(t-1))$ is a constant dependent only on the number of analysed data, and it is omitted during the maximisation of (3.12). All used statistics (3.5) – (3.7) are related to the model M_ℓ and they are computed with its the contextual neighbourhood I_r^ℓ. The determinants $|V_{zz(t)}|, |\lambda_t|$ can be evaluated recursively.

Textural features

Textural features are composed of parameter matrices $\hat{\gamma}_t = [A_s : \forall s \in I_r]$ and $\sqrt{\hat{\Sigma}_t}$ estimated from all possible pixels (t is the last pixel position in the chosen direction of texture analysis), the square root of sigma denotes square root of the matrix elements. The textural features are (Vacha and Haindl, 2007a):

1. A_s , $\forall s \in I_r$,

2. $\sqrt{\hat{\Sigma}_t}$.

As it is required, the proposed textural features are not dependent on a texture sample size. However, the sufficient sample size is necessary for the reliable parameter estimation.

Because the CAR models analyse a texture in some fixed movement direction, we have experimented with additional directions to capture supplementary texture properties. In that case, the texture is optionally analysed in four orthogonal directions: row-wise top-down and bottom-up, column-wise leftwards and rightwards. Subsequently, the estimated features for all the directions are concatenated into a common feature vector.

3.1.4 2D causal autoregressive random field

The 2D CAR textural representation is very similar to 3D CAR representation. The texture pixels at the k-th Gaussian pyramid level are locally modelled by an 2D CAR model (Haindl and Šimberová, 1992), which additionally assumes that each spectral plane can be modelled separately. Therefore texture spectral planes are decorrelated by means of K-L transformation prior to modelling. The decorrelation is not mandatory, but any interspectral relations would be discarded by the 2D CAR model.

We again omit the level index k and work with the multispectral texture pixels $Y_r = [Y_{r,1}, \ldots, Y_{r,C}]^T$. These multispectral pixels are modelled by a set of C models and j-th spectral plane is described by

$$Y_{r,j} = \gamma_j \, Z_{r,j} + \epsilon_{r,j} \ , \qquad\qquad Z_{r,j} = [Y_{r-s,j} : \forall s \in I_r]^T \ , \qquad\qquad (3.13)$$

where $Z_{r,j}$ is the $\eta \times 1$ data vector, $\gamma_j = [a_{s,j} : \forall s \in I_r]$ is the $1 \times \eta$ unknown parameter vector. Some selected contextual causal or unilateral neighbour index shift set is denoted I_r and its cardinality $\eta = |I_r|$.

The set of 2D models can be stacked into the 3D model equation (3.4), where the parameter matrices A_s become diagonal $A_s = \mathrm{diag}[a_{s,1}, \ldots, a_{s,C}]$. Additionally, uncorrelated noise vector components are assumed, i.e.,

$$\mathbb{E}\{\epsilon_{r,l}\epsilon_{r,j}\} = 0 \qquad \forall r, l, j, \, l \neq j \ .$$

Parameter estimation

The model parameter estimation follows equations (3.5) – (3.8) for 3D case, so as the estimation of optimal contextual neighbourhood (3.12). The difference is that the estimation is performed for each spectral plane separately, $j = 1, \ldots, C$:

$$\hat{\gamma}_{t-1,j}^T = V_{zz(t-1),j}^{-1} \, V_{zy(t-1),j} \ , \qquad\qquad (3.14)$$

$$V_{t-1,j} = \begin{pmatrix} \sum_{r=1}^{t-1} Y_{r,j} Y_r^T & \sum_{r=1}^{t-1} Y_{r,j} Z_{r,j}^T \\ \sum_{r=1}^{t-1} Z_{r,j} Y_{r,j}^T & \sum_{r=1}^{t-1} Z_{r,j} Z_{r,j}^T \end{pmatrix} + V_{0,j}$$

$$= \begin{pmatrix} V_{yy(t-1),j} & V_{zy(t-1),j}^T \\ V_{zy(t-1),j} & V_{zz(t-1),j} \end{pmatrix} \ , \qquad\qquad (3.15)$$

and noise variance σ_j^2 is estimated as

$$\hat{\sigma}_{t-1,j}^2 = \frac{\lambda_{t-1,j}}{\psi(t)} \ ,$$

$$\lambda_{t-1,j} = V_{yy(t-1),j} - V_{zy(t-1),j}^T \, V_{zz(t-1),j}^{-1} \, V_{zy(t-1),j} \ . \qquad\qquad (3.16)$$

The superscript or subscript $(-_{.,j})$ denotes parameters or statistics related to the j-th spectral plane, e.g. $Y^{(t-1),j}$ is the history of $t-1$ pixels $Y_{r,j}$ and $Z_{r,j}$, $\hat{\gamma}_{t-1,j}$ is the estimate of parameter γ_j from this history, and $\hat{\sigma}_{t-1,j}^2$ is the estimate of noise variance for j-th spectral plane from the same pixel history.

Alternatively, the LS estimation leads to the formally same equations as (3.14) – (3.16) with zero matrices $V_{0,j} = 0_{\eta+1}$.

The prediction probability $p\left(Y_{t,j}|Y^{(t-1),j}\right)$ and formula $\ln p(Y^{(t-1),j}|M_\ell)$ used in optimal model selection are computed according to equations (3.11), (3.12), which are used for each spectral plane separately (with parameter $C = 1$).

Textural features

The textural features are defined in the same form as features for the 3D CAR model. The set of 2D CAR models is stacked into the form (3.4) with diagonal matrices A_s and the noise covariance matrix is composed as

$$\hat{\Sigma}_t = \mathrm{diag}[\hat{\sigma}_{t,1}^2, \ldots, \hat{\sigma}_{t,C}^2] \ .$$

Textural features are again (Haindl and Vacha, 2006):

1. A_s , $\forall s \in I_r$,

2. $\sqrt{\hat{\Sigma}_t}$,

where the square root of sigma denotes square root of the matrix elements.

3.1.5   2D Gaussian Markov random field

The last textural representation assumes that spectral planes of each pyramid level are locally modelled using a 2-dimensional GMRF model Haindl (1991). This model is obtained if the local conditional density of the MRF model is Gaussian:

$$p(Y_{r,j}|Y_{s,j} \ \ \forall s \in I_r) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left\{ -\frac{(Y_{r,j} - \gamma_j Z_{r,j})^2}{2\sigma_j^2} \right\} \ ,$$

where $Y_{r,j}$ are mean centred values and j is the spectral plane index $j = 1, \ldots, C$. The data vector and parameter vector are again defined as

$$Z_{r,j} = [Y_{r-s,j} : \forall s \in I_r]^T \ , \qquad \gamma_j = [a_{s,j} : \forall s \in I_r] \ . \tag{3.17}$$

The contextual neighbourhood I_r is non-causal and symmetrical. Similarly as 2D CAR model, also GMRF model is not able to model interspectral relations. Therefore spectral planes are decorrelated by means of K-L transformation before the estimation of model parameters.

The GMRF model for centred values $Y_{r,j}$ can be expressed also in the matrix form of the 3D CAR model (3.4), but the driving noise ϵ_r and its correlation structure is now more complex:

$$\mathbb{E}\{\epsilon_{r,l}\epsilon_{r-s,j}\} = \begin{cases} \sigma_j^2 & \text{if } (s) = (0,0) \text{ and } l = j, \\ -\sigma_j^2\, a_{s,j} & \text{if } (s) \in I_r \text{ and } l = j, \\ 0 & \text{otherwise}, \end{cases} \tag{3.18}$$

where $\sigma_j, a_{s,j} \ \forall s \in I_r$ are unknown parameters. Also topology of the contextual neighbourhood I_r is different, because GMRF model requires a symmetrical neighbourhood.

Parameter estimation

The parameter estimation of the GMRF model is complicated because either Bayesian or Maximum Likelihood (ML) estimate requires an iterative minimisation of a nonlinear function. Therefore we use an approximation by the pseudo-likelihood estimator, which is computationally simple although not efficient. The pseudo-likelihood estimate for parameters $\hat{\gamma}_j, \hat{\sigma}_j^2$ has the form

$$\hat{\gamma}_j^T = \left[\sum_{\forall r \in I} Z_{r,j} \, Z_{r,j}^T \right]^{-1} \sum_{\forall r \in I} Z_{r,j} \, Y_{r,j} \ , \tag{3.19}$$

$$\hat{\sigma}_j^2 = \frac{1}{|I|} \sum_{\forall r \in I} (Y_{r,j} - \hat{\gamma}_j \, Z_{r,j})^2 \ , \tag{3.20}$$

where $j = 1 \dots C$, and I is the image lattice.

Let as additionally define $V_{zz,j}, V_{zy,j}$ analogically to the 2D CAR model

$$V_{zz,j} = \sum_{\forall r \in I} Z_{r,j} \, Z_{r,j}^T \ , \qquad\qquad V_{zy,j} = \sum_{\forall r \in I} Z_{r,j} \, Y_{r,j} \ . \tag{3.21}$$

Consequently, parameter estimate $\hat{\gamma}_j$ can be expressed as

$$\hat{\gamma}_j^T = V_{zz,j}^{-1} \, V_{zy,j} \ , \tag{3.22}$$

which is formally same as equation (3.14) with zero matrix $V_{0,j} = 0_{\eta+1}$.

The boundary conditions are again handled by either a toroidal lattice or the estimate is computed on an appropriate subset of the image lattice.

The optimal neighbourhood can be detected using the correlation method (Haindl and Havlíček, 1997) favouring locations of neighbours corresponding to large correlations over those with small correlations.

Textural Features

The estimated parameters for separate spectral planes are stacked together to produce multispectral representation (Haindl and Vacha, 2006; Vácha, 2005):

$$A_s = \text{diag}[a_{s,1}, \dots, a_{s,C}] \ , \qquad\qquad \hat{\Sigma} = \text{diag}[\hat{\sigma}_1^2, \dots, \hat{\sigma}_C^2] \ , \tag{3.23}$$

and the resulting textural features are in the same form as for CAR models (again the square root of sigma denotes square root of the matrix elements):

1. A_s, $\forall s \in I_r$,

2. $\sqrt{\hat{\Sigma}}$.

3.2 Feature comparison

All three previously described textural representations characterise texture with a feature vector, which is an element of a vector space. Feature vectors are used either directly, i.e. in combination with a classifier to build a class representation, or distance of feature vectors is computed to evaluate similarity of respective textures.

The distance between feature vectors of two textures T, S is computed using Minkowski norms (p-norms) L_1, $L_{0.2}$, defined in (2.2), or fuzzy contrast FC_3 proposed by Santini and Jain (1999). The norms L_1, $L_{0.2}$ are preferred to usual L_2, because they are more robust (Jacobs et al., 2000), while L_2 is sensitive to variation of even a single feature f_ℓ.

Fuzzy contrast FC_{ab} models features as predicates in fuzzy logic using sigmoid truth function τ. Subsequently, the dissimilarity of two feature vectors is defined as

$$
FC_{ab}(T,S) = m - \left\{ \sum_{\ell=1}^{m} \min\left\{ \tau(f_\ell^{(T)}), \tau(f_\ell^{(S)}) \right\} - a \sum_{\ell=1}^{m} \max\left\{ \tau(f_\ell^{(T)}) - \tau(f_\ell^{(S)}), 0 \right\} \right.
$$
$$
\left. - b \sum_{\ell=1}^{m} \max\left\{ \tau(f_\ell^{(S)}) - \tau(f_\ell^{(T)}), 0 \right\} \right\} , \qquad (3.24)
$$
$$
\tau(f_\ell) = \left(1 + \exp\left(-\frac{f_\ell - \mu(f_\ell)}{\sigma(f_\ell)} \right) \right)^{-1} ,
$$

where m is the feature vector size, $f_\ell^{(T)}$ and $f_\ell^{(S)}$ are the ℓ-th components of feature vectors of textures T and S, respectively. $\mu(f_\ell)$ and $\sigma(f_\ell)$ are average and standard deviation of the feature f_ℓ computed over all textures in the database. It is worth to note that FC_{ab} is not a metric, because it does not hold $FC_{ab}(T,T) = 0$ and it is not necessary symmetrical. However, we use only its symmetrical form FC_3, where $a = b = 3$.

In experiments with texture classification (Chapter 6), we use the simple k-Nearest Neighbours (k-NN) classifier, which classifies a texture according to majority of k-nearest training samples. The distance to training samples is computed with L_1, $L_{0.2}$, or FC_3 dissimilarity measures. However, a more elaborate classifier can be used, either trained directly from feature vectors as Support Vector Machine (SVM) (Cortes and Vapnik, 1995; Cristianini and Taylor, 2000) or from the dissimilarities (Pekalska and Duin, 2002; Pekalska et al., 2006). In image retrieval applications, we retrieve a given number of images that are nearest according to one of the previous dissimilarities.

3.3 Discussion

All the previously described MRF models are estimated on the levels of Gaussian downsampled pyramid, because it enables to capture larger spatial relations in a texture. An alternative method is a texture analysis by means of models with larger contextual neighbourhoods. Unfortunately, parameters of such larger models tend to be more sensitive and they can fluctuate with insignificant changes in a texture. On contrary,

models on the Gaussian pyramid are more robust, because Gaussian smoothing and down-sampling suppress insignificant details. Moreover, models on the Gaussian pyramid are more efficient, since the computational complexity is polynomial with respect to radius of contextual neighbourhood.

In our texture recognition or retrieval applications, we use models with a fixed contextual neighbourhood I_r for all processed textures. Although different optimal neighbourhood could be found for each texture (3.12), it would be difficult to compare the model parameters for different neighbourhoods. It is also possible to estimate the model parameters for different neighbourhood sizes and combine them in a feature vector. The parameters of a model with smaller neighbourhood I'_r can be efficiently estimated during the estimation of model with neighbourhood I_r, if $I'_r \subset I_r$. See Haindl and Šimberová (1995) for more details.

The final remark concerns monospectral (grey-scale) textures. They can be either modelled as a single spectral textures or the models can be estimated on the gradient image $\nabla Y_r = [\frac{\partial Y_r}{\partial r_1}, \frac{\partial Y_r}{\partial r_2}]^T$, which enlarge feature vector and simplifies the modelling. Additionally, the gradient image is more robust to illumination changes. Moreover, it will be derived (Section 4.3) that the part of feature vector which include only features A_s, $\forall s \in I_r$ is invariant to simple brightness changes. However, full invariants to illumination colour, brightness and other conditions will be derived in the following chapter.

Chapter 4

Illumination Invariance

Illumination conditions of an image acquisition can change due to various reasons. In our approach, we allow changes of brightness and spectrum of illumination sources, and we derive illumination invariants based on the textural features from the previous chapter. It enables us to create textural representation, which is invariant to illumination colour brightness (Vacha and Haindl, 2007a, 2010a).

We assume that a textured surface is illuminated with several illumination sources and that positions of viewpoint and illumination sources remain unchanged. We start with the assumption of a single illumination, which is far enough to produce uniform illumination, and planar Lambertian surfaces with varying albedo and surface texture normal. However, these restrictive assumptions will be further relieved to incorporate more illumination sources, nonuniform illumination, and surfaces with a natural reflectance model. Still, the assumption of fixed illumination positions might sound limiting. Nevertheless, our experiments with natural and artificial surface materials (Sections 6.1.3 and 6.1.4) show that the derived features are very robust even if the illumination positions changes dramatically.

4.1 Illumination models

Let us assume that a textured Lambertian (ideally diffuse) surface is illuminated with one uniform illumination. The value acquired by the j-th sensor at the pixel location r can be expressed as

$$Y_{r,j} = \int_{\Omega} E(\omega)\, S(r,\omega)\, R_j(\omega)\, d\omega \ , \tag{4.1}$$

where ω is wavelength, $E(\omega)$ is the spectral power distribution of a single illumination, $S(r,\omega)$ is a Lambertian reflectance coefficient at the position r, $R_j(\omega)$ is the j-th sensor response function, and the integral is taken over the visible spectrum Ω. The Lambertian reflectance term $S(r,\omega)$ depends on surface normal, illumination direction, and surface albedo.

Following the works of Finlayson (1995); Healey and Wang (1995) we approximate the surface reflectance $S(r, \omega)$ by a linear combination of a fixed basis of functions $s_c(\omega)$:

$$S(r, \omega) = \sum_{c=1}^{C} d_{r,c} \, s_c(\omega) \ . \tag{4.2}$$

The functions $s_c(\omega)$ are optimal basis functions that represent the data. The method for finding suitable basis was introduced by Marimont and Wandell (1992). They also concluded that, given the human receptive cones, a 3-dimensional basis set is sufficient to model colour observations. However, finding such basis set is not needed in our method, because the key assumption is only its existence. Provided that $j = 1, \ldots, C$ sensor measurements are available, the acquired values can be approximated by

$$Y_{r,j} \approx \sum_{c=1}^{C} d_{r,c} \int_{\Omega} E(\omega) \, s_c(\omega) \, R_j(\omega) \, d\omega \ ,$$

$$Y_r = B' d_r \ ,$$

where $d_r = [d_{r,1}, \ldots, d_{r,C}]^T$ and B' is a $C \times C$ matrix composed of the integral values. An image of the same scene illuminated with a different spectrum $\tilde{E}(\omega)$ is composed of

$$\tilde{Y}_{r,j} \approx \sum_{c=1}^{C} d_{r,c} \int_{\Omega} \tilde{E}(\omega) \, s_c(\omega) \, R_j(\omega) \, d\omega \ ,$$

$$\tilde{Y}_r = \tilde{B}' d_r \ ,$$

where $\tilde{B}'$ is a $C \times C$ matrix. Consequently, the two images $\tilde{Y}, Y$ acquired with different illumination brightness or spectrum can be transformed to each other by the linear transformation:

$$\tilde{Y}_r = B \, Y_r \qquad \forall r \ , \tag{4.3}$$

which is same for all the pixels. If we change the response functions of receptors $R_j(\omega)$ instead of changes of illumination spectrum $E(\omega)$, the derivation is almost the same and the formula (4.3) holds again.

Multiple illumination sources

The linear model (4.3) is valid even for several illumination sources with variable spectra provided that the spectra of all sources are the same and the positions of the illumination sources remain fixed. Let $S^{(p)}(r, \omega)$ denotes Lambertian reflectance coefficient corresponding to the p-th illumination and P is the number of illumination sources. The acquired values $Y_{r,j}$ can be expressed and approximated as

$$Y_{r,j} = \sum_{p=1}^{P} \int_{\Omega} E(\omega) \, S^{(p)}(r, \omega) \, R_j(\omega) \, d\omega \approx \sum_{p=1}^{P} \sum_{c=1}^{C} d_{r,c}^{(p)} \int_{\Omega} E(\omega) \, s_c(\omega) \, R_j(\omega) \, d\omega \ ,$$

$$Y_r = B' \sum_{p=1}^{P} d_r^{(p)} \ ,$$

where $d_{r,c}^{(p)}$ are respective coefficients from approximation (4.2). Consequently, the image acquired with a different illumination spectrum is expressed and related as

$$\tilde{Y}_r = \tilde{B}' \sum_{p=1}^{P} d_r^{(p)} \ ,$$

$$\tilde{Y}_r = \tilde{B}' B'^{-1} Y_r \ ,$$

where the second row complies with formula (4.3). However, illuminations with different spectra would break this relation (see Appendix A.1 for more details).

Natural illumination model

The surface reflectance can be further generalized to the natural model of Bidirectional Texture Function (BTF) (Dana et al., 1999), where the surface reflectance is function of surface position, wavelength, incoming and outgoing light directions. Let $L(r, \omega, v_i, v_o)$ is the surface reflectance, v_i is illumination direction and v_o viewing direction then equation (4.1) becomes

$$Y_{r,j} = \int_{\Omega} E(\omega) \, L(r, \omega, v_i, v_o) \, R_j(\omega) \, d\omega \ . \tag{4.4}$$

On the condition that Q is an arbitrary number of reflectance components in the reflectance model (e.g. Lambertian component, different isotropic or anisotropic spectacular components) and each component is separable in ω, the reflectance can be decomposed and approximated (Vacha et al., submitted):

$$L(r, \omega, v_i, v_o) = \sum_{q=1}^{Q} L^{(q)}(r, \omega, v_i, v_o) = \sum_{q=1}^{Q} \Lambda^{(q)}(r, v_i, v_o) \, \dot{S}^{(q)}(r, \omega) \tag{4.5}$$

$$\approx \sum_{q=1}^{Q} \Lambda^{(q)}(r, v_i, v_o) \sum_{c=1}^{C} d_{r,c}^{(q)} s_c(\omega) = \sum_{c=1}^{C} s_c(\omega) \sum_{q=1}^{Q} d_{r,c}^{(q)} \Lambda^{(q)}(r, v_i, v_o) \ ,$$

where the first row is the separation of ω variable, $\Lambda^{(q)}(r, v_i, v_o)$ is the q-th reflectance component at position r dependent on the angles, while $\dot{S}^{(q)}(r, \omega)$ is the reflectance dependent on ω. The second row is again approximation with optimal basis functions $s_c(\omega)$ (4.2). Substitution into (4.4) provides equations for the images with a different illumination spectrum:

$$Y_r = B' \sum_{q=1}^{Q} d_r^{(q)} \Lambda^{(q)}(r, v_i, v_o) = B' d_r' \qquad \forall r \ ,$$

$$\tilde{Y}_r = \tilde{B}' \sum_{q=1}^{Q} d_r^{(q)} \Lambda^{(q)}(r, v_i, v_o) = \tilde{B}' d_r' \qquad \forall r \ ,$$

which is in accordance with the linear model (4.3). For a fixed position r, the function $\sum_{q=1}^{Q} \Lambda^{(q)}(r, v_i, v_o)$ becomes the well-known Bidirectional Reflectance Distribution Function (BRDF) (Nicodemus et al., 1977). Obviously, the previous illumination model

includes simpler models as dichromatic reflection model (Shafer, 1985) or the well-known Phong reflection model.

The assumption of wavelength separability in (4.5) neglects effects, where the colour of surface depends on viewing or illumination angle. An example of the material with such effect is a furry textile, where the colour of fur is different from the colour of base textile. Consequently, we see either hairs or base textile depending on view angle and position of hairs.

Naturally, the linear model (4.3) includes all other colour models, which can be transformed linearly, i.e. CIE XYZ, opponent colours, Gaussian colour model (Geusebroek et al., 2003) when computed from RGB images, and YCbCr used in video coding.

Other illumination effects

We briefly review illumination related effects, which are not considered in the previous models and therefore they might be either approximated or completely neglected if they cannot fit in the linear model (4.3).

We considered opaque surfaces and their reflectance, which is the process when the incident light is immediately radiated without change of frequency. (However different frequencies are reflected or absorbed in different amount.) The BTF model includes interreflections and sub-surface light scattering, but they cannot be separated and examined separately. The previous illumination models do not account for polarisation effects.

Unlike reflectance, the fluorescence or phosphorescence is the process when the energy of incident light is absorbed and subsequently emitted at different wavelength. According to Kasha-Vavilov rule, the emitted wavelength does not depend on the excitation wavelength for most of fluorescent substances (Turro, 1978). However, different incident wavelength carries a different amount of energy, which results in different intensity of emitted light. The appearance of purely florescent surfaces under different illumination spectra can be represented by the linear model (4.3). Unfortunately, the transformation matrix B would be different for fluorescence and reflectance. Therefore a common matrix B cannot model the appearance change of both fluorescent and simple reflective surfaces in one image and it cannot model appearance of surfaces, which exhibit combination of fluoresce and reflectance.

4.2   Colour invariants

Colour invariant features are derived from the MRF textural representation on condition that images with different illumination are related by linear relation (4.3) and the transformation matrix B is regular (Vacha and Haindl, 2007a, 2010b). As we have shown, linear relation (4.3) comprises changes of colour and brightness of illumination source.

The colour invariants are derived for all three representations introduced in Chapter 3, we start with the derivation for the 3D CAR model, followed by 2D CAR and GMRF. In general, statistics, parameters and other variables corresponding to another illumination are denoted with ($\tilde{\cdot}$) accent.

4.2.1   3D causal autoregressive random field

Let us assume that two images Y, $\tilde{Y}$ with different illumination are related by (4.3). Consequently, the model data vectors (3.3) of 3D CAR model (3.4) are also related by the linear transformation

$$\tilde{Z}_r = \Delta\, Z_r \qquad \forall r \ , \tag{4.6}$$

where Δ is the $C\eta \times C\eta$ block diagonal matrix with blocks B on the diagonal. By substituting $\tilde{Y}_r$, $\tilde{Z}_r$ into the parameter estimate of 3D CAR model (3.6), we can derive corresponding statistics for different illumination, which are denoted with ($\tilde{\cdot}$) accent:

$$\tilde{V}_{yy(t-1)} \approx \sum_{r=1}^{t-1} B Y_r\, Y_r^{T} B^{T} = B V_{yy(t-1)} B^{T} \ , \tag{4.7}$$

$$\tilde{V}_{zz(t-1)} \approx \sum_{r=1}^{t-1} \Delta Z_r\, Z_r^{T} \Delta^{T} = \Delta V_{zz(t-1)} \Delta^{T} \ , \tag{4.8}$$

$$\tilde{V}_{zy(t-1)} \approx \sum_{r=1}^{t-1} \Delta Z_r\, Y_r^{T} B^{T} = \Delta V_{zy(t-1)} B^{T} \ . \tag{4.9}$$

The previous relations are approximations, not equations, because prior matrix V_0 do not follow the liner relation (4.3). It can be either corrected by modification of prior V_0 or simply neglected if enough data is available. Moreover, the relations (4.7) – (4.9) became equations for Least Square parameter estimate (3.9). Subsequently, the substitution into parameter estimates (3.5) and (3.7) produces following relations of model parameters for different illuminations:

$$
\begin{aligned}
\tilde{\hat{\gamma}}_{t-1}^{T} &\approx (\Delta^{T})^{-1} V_{zz(t-1)}^{-1} \Delta^{-1} \Delta V_{zy(t-1)} B^{T} \\
&= (\Delta^{T})^{-1} \hat{\gamma}_{t-1}^{T} B^{T} \ ,
\end{aligned}
\tag{4.10}
$$

$$
\begin{aligned}
\tilde{\lambda}_{t-1} &\approx B V_{yy(t-1)} B^{T} - B V_{zy(t-1)}^{T} \Delta^{T} (\Delta^{T})^{-1} V_{zz(t-1)}^{-1} \Delta^{-1} \Delta V_{zy(t-1)} B^{T} \\
&= B \left(V_{yy(t-1)} - V_{zy(t-1)}^{T} V_{zz(t-1)}^{-1} V_{zy(t-1)} \right) B^{T} \\
&= B \lambda_{t-1} B^{T} \ .
\end{aligned}
\tag{4.11}
$$

The same relation can by verified for the recursive parameter update (3.8):

$$
\begin{aligned}
\tilde{\hat{\gamma}}_{t}^{T} &= (\Delta^{T})^{-1} \hat{\gamma}_{t-1}^{T} B^{T} + \frac{(\Delta^{T})^{-1} V_{zz(t-1)}^{-1} \Delta^{-1} \Delta Z_t (B Y_t - B \hat{\gamma}_{t-1} \Delta^{-1} \Delta Z_t)^{T}}{1 + Z_t^{T} \Delta^{T} (\Delta^{T})^{-1} V_{zz(t-1)}^{-1} \Delta^{-1} \Delta Z_t} \\
&= (\Delta^{T})^{-1} \hat{\gamma}_{t-1}^{T} B^{T} + \frac{(\Delta^{T})^{-1} V_{zz(t-1)}^{-1} Z_t (Y_t + Z_t^{T} V_{zz(t-1)}^{-1} Z_t)^{T} B^{T}}{1 + Z_t^{T} V_{zz(t-1)}^{-1} Z_t}
\end{aligned}
$$

$$\approx (\Delta^T)^{-1} \left(\hat{\gamma}_{t-1}^T + \frac{V_{zz(t-1)}^{-1} Z_t (Y_t - \hat{\gamma}_{t-1} Z_t)^T}{1 + Z_t^T V_{zz(t-1)}^{-1} Z_t} \right) B^T$$

$$= (\Delta^T)^{-1} \hat{\gamma}_t^T B^T \quad .$$

Since $\hat{\gamma}_t$ is composed of submatrices A_s, the 3D CAR model parameters for different illuminations are related by

$$\tilde{A}_s = B\, A_s\, B^{-1} \;, \qquad \tilde{\lambda}_t = B\, \lambda_t\, B^T \;, \qquad \forall s \in I_r \,, \; \forall t \in I \;. \tag{4.12}$$

Colour invariant textural features

As a direct consequence of formulas (4.12), (4.3), (4.6), and (4.8), the following features can be proved to be colour invariant (Vacha and Haindl, 2007a):

1. trace: $\operatorname{tr} A_s \,, \; \forall s \in I_r$,

2. eigenvalues: $\nu_{s,j}$ of $A_s \,, \quad \forall s \in I_r \,, \; j = 1, \ldots, C$,

3. $\alpha_1 = 1 + Z_t^T V_{zz(t)}^{-1} Z_t$,

4. $\alpha_2 = \sqrt{\sum\limits_{\forall r \in I} (Y_r - \hat{\gamma}_t Z_r)^T \lambda_t^{-1} (Y_r - \hat{\gamma}_t Z_r)}$,

5. $\alpha_3 = \sqrt{\sum\limits_{\forall r \in I} (Y_r - \mu)^T \lambda_t^{-1} (Y_r - \mu)}$, μ is the mean value vector of Y_r ,

These colour invariants use linear relation (4.3), which could be considered too general for some applications, because it allows mutual swaps of sensors or spectral planes. In that case, matrix B can be restricted to a diagonal matrix, which models illumination change as multiplication of each spectral plane. For the diagonal B, the formula $B^{-1} A_s B$ do not change the diagonal elements of A_s. Therefore we can alternatively define invariants $\nu_{s,j}$:

2'. diagonals: $\nu_s = \operatorname{diag} A_s \,, \quad \forall s \in I_r$.

This alternative definition of $\nu_{s,j}$ should be preferred if the decorrelation of image spectral planes (K-L transformation) is employed before the estimation of texture model. Otherwise, the definition with eigenvalues would cancel the decorrelation effect.

We also provide an alternative definition of α_1, which is supposed to be more robust, because it do not depend on the single pixel neighbourhood Z_t :

3'. $\alpha_{1'} = \sqrt{\mu(Z_r)^T V_{zz(t)}^{-1} \mu(Z_r)}$, where $\mu(Z_r) = \sum\limits_{\forall r \in I} \frac{Z_r}{|I|}$ is the mean of data vector

obtained by the repetition of $\mu(Y_r)$: $\mu(Z_r) = [\mu(Y_r^T) : m = 1 \ldots |I_r|]^T$.

We did not use the sum of α_1 invariants computed at all pixel positions, because the formula $\sum_{t \in I} Z_t^T V_{zz(t)}^{-1} Z_t$ is approximately constant for the given neighbourhood and the number of analysed pixels.

Determinant based colour invariants

Additional colour invariant can be derived from determinants $|V_{yy(t-1)}|$, $|V_{zz(t-1)}|$, and $|\lambda_{t-1}|$, which relations implies from equations (4.7), (4.8), and (4.11):

$$|\tilde{V}_{yy(t-1)}| \approx |B\, V_{yy(t-1)}\, B^T| = |V_{yy(t-1)}|\,|B|^2 \ , \tag{4.13}$$

$$|\tilde{V}_{zz(t-1)}| \approx |\Delta\, V_{zz(t-1)}\, \Delta^T| = |V_{zz(t-1)}|\,|\Delta|^2 = |V_{zz(t-1)}|\,|B|^{2\eta} \ , \tag{4.14}$$

$$|\tilde{\lambda}_{t-1}| \approx |B\, \lambda_{t-1}\, B^T| = |\lambda_{t-1}|\,|B|^2 \ . \tag{4.15}$$

Consequently, the following formulas extend the previous set of colour invariants (Vacha and Haindl, 2010a):

6. $\beta_1 = \ln\left(\frac{\psi(r)^C}{\psi(t)^C}\, |\lambda_t|\,|\lambda_r|^{-1} \right)$,

7. $\beta_2 = \ln\left(\frac{\psi(r)^C}{\psi(t)^C}\, \left|V_{zz(t)}\right|\, \left|V_{zz(r)}\right|^{-1} \right)$,

8. $\beta_3 = \ln\left(\left|V_{zz(t)}\right|\, |\lambda_t|^{-\eta} \right)$,

9. $\beta_4 = \ln\left(\left|V_{zz(t)}\right|\, \left|V_{yy(t)}\right|^{-\eta} \right)$,

10. $\beta_5 = \operatorname{tr}\left\{ V_{yy(t)}\, \lambda_t^{-1} \right\}$.

Let us also consider the impact of illumination change on the pixel prediction probability $p\left(Y_t|Y^{(t-1)}\right)$ (3.11) and $\ln p\left(Y^{(t-1)}|M_\ell\right)$ (3.12) used in optimal model selection. After substitution of relations (4.3), (4.6), (4.8), (4.10), (4.14), and (4.15), we acquire:

$$p\left(\tilde{Y}_t|\tilde{Y}^{(t-1)}\right) = \frac{\Gamma\left(\frac{\psi(t)-C\eta+C+2}{2} \right)}{\Gamma\left(\frac{\psi(t)-C\eta+2}{2} \right) \pi^{\frac{C}{2}} \left(1 + Z_t^T \Delta^T (\Delta^T)^{-1} V_{zz(t-1)}^{-1} \Delta^{-1} \Delta Z_t \right)^{\frac{C}{2}}}$$

$$\frac{1}{|\lambda_{t-1}|^{\frac{1}{2}}\,|B|} \left(1 + \frac{\left(BY_t - B\hat{\gamma}_{t-1}\Delta^{-1}\Delta Z_t \right)^T (B^T)^{-1} \lambda_{t-1}^{-1} B^{-1}}{1 + Z_t^T \Delta^T (\Delta^T)^{-1} V_{zz(t-1)}^{-1} \Delta^{-1} \Delta Z_t} \right.$$

$$\left. \frac{\left(BY_t - B\hat{\gamma}_{t-1}\Delta^{-1}\Delta Z_t \right)}{1} \right)^{\frac{\psi(t)-C\eta+C+2}{2}}$$

$$= p\left(Y_t|Y^{(t-1)} \right) |B|^{-1}$$

$$\ln p\left(\tilde{Y}^{(t-1)}|M_\ell\right) = K_1(\psi(t-1)) - \frac{C}{2}\ln\left|V_{zz(t-1)}\right| - \frac{C}{2}\ln|B|^{2\eta} + \frac{C^2\eta}{2}\ln\pi$$
$$- \frac{\psi(t) - C\eta + C + 1}{2}\ln|\lambda_{t-1}| - \frac{\psi(t) - C\eta + C + 1}{2}\ln|B|^2$$
$$+ \sum_{m=1}^{C}\left[\ln\Gamma\left(\frac{\psi(t) - C\eta + C + 2 - m}{2}\right) - \ln\Gamma\left(\frac{\psi(0) - C\eta + C + 2 - m}{2}\right)\right]$$
$$= \ln p\left(Y^{(t-1)}|M_\ell\right) - (\psi(t) + C + 1)\ln|B|$$

Let us define:

11. $\beta_6 = \ln\left(\sum_{\forall r \in I}\frac{1}{|I|}p\left(Y_r|Y^{(r-1)}\right)\left|V_{yy(t)}\right|^{\frac{1}{2}}\right)$,

12. $\beta_7 = \ln\left(\ln p\left(Y^{(t)}|M_\ell\right) + (\psi(t+1) + C + 1)\ln\left|V_{yy(t)}\right|\right)$.

Alternatively, terms including function $\Gamma(x)$ can be omitted during computation of invariants β_6, β_7, which speeds up their computation. Since the terms with $\Gamma(x)$ function do not depend on pixel values, their omission almost does not effect the results (see experiment in Section 6.1.4). Instead of the logarithm, we can use an alternative normalisation of invariants $\beta_1 - \beta_5$ based on the geometric mean:

13. $\beta_8 = \left(\frac{\psi(r)^C}{\psi(t)^C}|\lambda_t|\,|\lambda_r|^{-1}\right)^{\frac{1}{2C}}$,

14. $\beta_9 = \left(\frac{\psi(r)^C}{\psi(t)^C}\left|V_{zz(t)}\right|\left|V_{zz(r)}\right|^{-1}\right)^{\frac{1}{2C\eta}}$,

15. $\beta_{10} = \left(\left|V_{zz(t)}\right|\,|\lambda_t|^{-\eta}\right)^{\frac{1}{2C}}$,

16. $\beta_{11} = \left(\left|V_{zz(t)}\right|\left|V_{yy(t)}\right|^{-\eta}\right)^{\frac{1}{2C}}$,

17. $\beta_{12} = \sqrt{\left|V_{yy(t)}\right|\,|\lambda_t|^{-1}}$.

Invariants $\alpha_1 - \alpha_3$, $\alpha_{1'}$, $\beta_3 - \beta_7$, $\beta_{10} - \beta_{12}$ are computed with $V_{zz(t)}$, $V_{yy(t)}$, λ_t estimates from all the image pixels, it means t equal to the last pixel position. However, they can be computed from actual estimates at each pixel position as well, which is useful in texture segmentation. Invariants β_1, β_2, β_8, and β_9 are computed from $V_{zz(r)}$, λ_r estimates at different positions r, t in the texture, e.g. first and last pixel position.

If the assumption of texture homogeneity is considered, the invariants β_1, β_2, β_8, and β_9 are necessary constant. Therefore, these invariants can be regarded as condensed indicators of texture homogeneity.

An intuitive interpretation of the other invariants is quite difficult. The invariants α_2, α_3 are based the statistic λ which is made illumination invariant. The statistic λ is used in the estimation of noise and actually it expresses the model ability to explain the

data. Furthermore, the invariants β_4, β_{11} are the ratios of correlations in the data vectors to correlations in the pixel vectors, which we consider to be a measure of dependency in the contextual neighbourhood.

4.2.2 2D causal autoregressive random field

Invariants for the 2D CAR model are formally same as the invariants for 3D CAR model, with the difference that they are computed for each spectral plane separately. It was shown that a set of 2D CAR models can be stacked to the form of 3D model (3.4), with restriction to diagonal matrices A_s. Additionally, the linear relation (4.3) have to be restricted to a diagonal matrix $B = \text{diag}[b_1, \ldots, b_C]$, because two dimensional models are not able to model interspectral relations.

For the 2D CAR model (3.13), relations of image value vectors Y_r and model data vectors Z_r can be expressed for each spectral plane separately, $\forall j = 1, \ldots, C$:

$$\tilde{Y}_{r,j} = b_j Y_{r,j} \ , \tag{4.16}$$

$$\tilde{Z}_{r,j} = b_j Z_{r,j} \ , \tag{4.17}$$

Thus, the statistics (3.15) for images with different illuminations are related as:

$$\tilde{V}_{yy(t-1),j} \approx \sum_{r=1}^{t-1} b_j Y_{r,j} {Y_{r,j}}^T b_j = V_{yy(t-1),j} \, b_j^2 \ , \tag{4.18}$$

$$\tilde{V}_{zz(t-1),j} \approx \sum_{r=1}^{t-1} b_j Z_{r,j} {Z_{r,j}}^T b_j = V_{zz(t-1),j} \, b_j^2 \ , \tag{4.19}$$

$$\tilde{V}_{zy(t-1),j} \approx \sum_{r=1}^{t-1} b_j Z_{r,j} {Y_{r,j}}^T b_j = V_{zy(t-1),j} \, b_j^2 \ , \tag{4.20}$$

so as their determinants:

$$|\tilde{V}_{yy(t-1),j}| \approx |V_{yy(t-1),j}| \, b_j^2 \ , \qquad |\tilde{V}_{zz(t-1),j}| \approx |V_{zz(t-1),j}| \, b_j^{2\eta} \ . \tag{4.21}$$

Subsequently, the substitution into parameter estimates (3.14) and (3.16) produces the following relations:

$$\begin{aligned} \tilde{\hat{\gamma}}_{t-1,j}^T &\approx V_{zz(t-1),j}^{-1} \, b_j^{-2} \, V_{zy(t-1),j} \, b_j^2 \\ &= \hat{\gamma}_{t-1,j}^T \ , \end{aligned} \tag{4.22}$$

$$\begin{aligned} \tilde{\lambda}_{t-1,j} &\approx V_{yy(t-1),j} \, b_j^2 - V_{zy(t-1),j}^T \, b_j^2 \, V_{zz(t-1),j}^{-1} \, b_j^{-2} \, V_{zy(t-1),j} \, b_j^2 \\ &= \lambda_{t-1,j} \, b_j^2 \ . \end{aligned} \tag{4.23}$$

Consequently, equations (4.12) hold again and matrices B, A_s are diagonal. Contrary to the 3D CAR model, model parameters A_s of 2D CAR are already colour invariant (see formula (4.22)) and it is not necessary to transform them into a illumination invariant form. However, this is a consequence of the stronger assumption of diagonal matrices A_s and diagonal matrix B, which are necessary for 2D model.

Colour invariant textural features

Formulas (4.16) – (4.23) produce the same set of colour invariants as the invariants for 3D model. Since the invariants $\alpha_1 - \alpha_3$, $\alpha_{1'}$, $\beta_1 - \beta_{12}$, belong to a single model, they are computed for each spectral plane separately. The following features are colour invariant (Vacha and Haindl, 2007a), $\forall j = 1, \ldots, C$:

1. trace: $\operatorname{tr} A_s$, $\forall s \in I_r$,

2. eigenvalues: $\nu_{s,j}$ of A_s or diagonals: $\nu_s = \operatorname{diag} A_s$, $\forall s \in I_r$,

3. $\alpha_{1,j} = 1 + Z_{t,j}^T \, V_{zz(t),j}^{-1} \, Z_{t,j}$, or
   $\alpha_{1',j} = \mu(Z_{r,j})^T \, V_{zz(t),j}^{-1} \, \mu(Z_{r,j})$, where $\mu(Z_{r,j}) = \sum\limits_{\forall r \in I} \frac{Z_{r,j}}{|I|}$.

4. $\alpha_{2,j} = \sqrt{\sum\limits_{\forall r \in I} \lambda_{t,j}^{-1} \left(Y_{r,j} - \hat{\gamma}_{t,j} Z_{r,j} \right)^2}$,

5. $\alpha_{3,j} = \sqrt{\sum\limits_{\forall r \in I} \lambda_{t,j}^{-1} \left(Y_{r,j} - \mu_j \right)^2}$, μ_j is the mean value of vector $Y_{r,j}$,

and corresponding invariants based on determinants:

$$\beta_{1,j} = \ln \left(\tfrac{\psi(r)}{\psi(t)} \, |\lambda_{t,j}| \, |\lambda_{r,j}|^{-1} \right) , \qquad \beta_{8,j} = \left(\tfrac{\psi(r)}{\psi(t)} \, |\lambda_{t,j}| \, |\lambda_{r,j}|^{-1} \right)^{\frac{1}{2}} ,$$

$$\beta_{2,j} = \ln \left(\tfrac{\psi(r)}{\psi(t)} \, \left|V_{zz(t),j}\right| \, \left|V_{zz(r),j}\right|^{-1} \right) , \quad \beta_{9,j} = \left(\tfrac{\psi(r)}{\psi(t)} \, \left|V_{zz(t),j}\right| \, \left|V_{zz(r),j}\right|^{-1} \right)^{\frac{1}{2\eta}} ,$$

$$\beta_{3,j} = \ln \left(\left|V_{zz(t),j}\right| \, |\lambda_{t,j}|^{-\eta} \right) , \qquad \beta_{10,j} = \left(\left|V_{zz(t),j}\right| \, |\lambda_{t,j}|^{-\eta} \right)^{\frac{1}{2}} ,$$

$$\beta_{4,j} = \ln \left(\left|V_{zz(t),j}\right| \, \left|V_{yy(t),j}\right|^{-\eta} \right) , \qquad \beta_{11,j} = \left(\left|V_{zz(t),j}\right| \, \left|V_{yy(t),j}\right|^{-\eta} \right)^{\frac{1}{2}} ,$$

$$\beta_{5,j} = V_{yy(t),j} \, \lambda_t^{-1} , \qquad \beta_{12,j} = \sqrt{\left|V_{yy(t),j}\right| \, |\lambda_{t,j}|^{-1}} ,$$

$$\beta_{6,j} = \ln \left(\sum_{\forall r \in I} \frac{1}{|I|} \, p \left(Y_{r,j} | Y^{(r-1),j} \right) \left|V_{yy(t),j}\right|^{\frac{1}{2}} \right) ,$$

$$\beta_{7,j} = \ln \left(\ln p \left(Y^{(t),j} | M_\ell \right) \, + \, (\psi(t+1) + 2) \ln \left|V_{yy(t),j}\right| \right) .$$

Again, invariants $\alpha_{1,j} - \alpha_{3,j}$, $\alpha_{1',j}$, $\beta_{3,j} - \beta_{7,j}$, $\beta_{10,j} - \beta_{12,j}$ are computed with $V_{zz(t),j}$, $V_{yy(t),j}$, $\lambda_{t,j}$ estimates from all the image pixels, it means t equal to the last pixel position.

The definition $\nu_s = \operatorname{diag} A_s$ (item 2. in the list above) should be used for the 2D CAR model preceded with K-L transformation, otherwise the order of K-L components would be mixed up. The reason is that computation of eigenvalues reorders the spectral planes and therefore the order of spectral planes in ν_s may not correspond for different $s \in I_r$. Alternatively, if the invariance to mutual swap of spectral planes is required, the invariants differing in spectral planes (e.g. $\alpha_{1,1}, \ldots, \alpha_{1,C}$) should be sorted according to their values.

If K-L transformation is used with the 2D CAR model, the invariance is provided to the transformed decorrelated values, for which diagonal matrix B is sufficient. However, the real illumination effects (modelled by (4.3)) occur before K-L transformation and they can result in different K-L components estimate. This might be either disadvantage or advantage, because we can exploit the ordering of K-L components, which are sorted according to the variance. Alternatively to K-L transformation, R, G, B colour values can be projected into opponent colour space, which components are independent with respect to human perception. The transformation into opponent colour space is also linear.

4.2.3   2D Gaussian Markov random field

Colour invariants, similar to those for 3D and 2D CAR models, can be derived also for GMRF model. As it was mention in Section 3.1.5, a set of GMRF models for different spectral planes can be stacked together to produce a 3D relation (3.4). We further assume that images Y, $\tilde{Y}$ with different illumination are linearly related, where transformation matrix B is again restricted to a diagonal matrix (4.16). The reason is that a set of two dimensional GMRF models is not able to model interspectral relations.

Substitution into GMRF parameter estimates (3.19) and (3.20) produce the following relations, $\forall j = 1, \ldots, C$:

$$\tilde{\hat{\gamma}}_j^T = \left[\sum_{\forall r \in I} b_j Z_{r,j} \, Z_{r,j}^T b_j \right]^{-1} \sum_{\forall r \in I} b_j Z_{r,j} \, Y_{r,j} b_j$$
$$= \hat{\gamma}_j^T \;, \tag{4.24}$$

$$\tilde{\hat{\sigma}}_j^2 = \frac{1}{|I|} \sum_{\forall r \in I} (b_j Y_{r,j} - \hat{\gamma}_j \, Z_{r,j} b_j)^2$$
$$= b_j^2 \, \hat{\sigma}_j^2 \;. \tag{4.25}$$

Alternatively, it can be expressed in matrix notation (3.23) as

$$\tilde{A}_s = B \, A_s \, B^{-1} = A_s \qquad \forall s \in I_r \;, \qquad\qquad \tilde{\hat{\Sigma}} = B \, \hat{\Sigma} \, B^T \;, \tag{4.26}$$

where all included matrices are diagonal. The model statistics (3.21) for different illuminations are related as

$$\tilde{V}_{yy,j} = \sum_{\forall r \in I} b_j Y_{r,j} \, Y_{r,j} b_j = V_{yy,j} \, b_j^2 \;, \qquad\qquad |\tilde{V}_{yy,j}| = |V_{yy,j}| \, b_j^2 \;, \tag{4.27}$$

$$\tilde{V}_{zz,j} = \sum_{\forall r \in I} b_j Z_{r,j} \, Z_{r,j}^T b_j = V_{zz,j} \, b_j^2 \;, \qquad\qquad |\tilde{V}_{zz,j}| = |V_{zz,j}| \, b_j^{2\eta} \;. \tag{4.28}$$

Colour invariant textural features

The colour invariants for GMRF model include $\operatorname{tr} A_s$ and ν_s derived for CAR models and modified version of α_2, α_3 (Vacha and Haindl, 2007a). They are consequences of relations (4.26), (4.3), (4.6), and they hold for $j = 1, \ldots, C$:

1. trace: $\operatorname{tr} A_s$, $\forall s \in I_r$,

2. eigenvalues: $\nu_{s,j}$ of A_s or diagonals: $\nu_s = \operatorname{diag} A_s$, $\forall s \in I_r$,

3. $\alpha_{2,j} = \sqrt{\frac{1}{|I|} \sum\limits_{\forall r \in I} \hat{\sigma}_j^{-2} \left(Y_{r,j} - \hat{\gamma}_j Z_{r,j}\right)^2}$,

4. $\alpha_{3,j} = \sqrt{\frac{1}{|I|} \sum\limits_{\forall r \in I} \hat{\sigma}_j^{-2} \left(Y_{r,j} - \mu_j\right)^2}$, μ_j is the mean value of $Y_{r,j}$.

Naturally, it possible to derive counterparts of $\beta_1 - \beta_{12}$ invariants for GMRF model (Vacha and Haindl, 2010a). Let as denote absolute value of determinant $\operatorname{abs} |V_{zz}|$, otherwise absolute values are denoted with the same symbol as determinant $|\cdot|$. The following invariants are similar to their 2D CAR counterparts with difference that $\operatorname{abs} |V_{zz}|$ have to be used instead of $|V_{zz}|$, because V_{zz} is not always positive definite in the GMRF model. Invariants β_7 and β_8 do not have their GMRF counterparts. Finally, the previous set of colour invariants for the GMRF model is extended with the following invariants, which are computed for each spectral plane j separately, $j = 1, \ldots, C$:

5. $\beta_{1,j} = \ln\left(\left|\hat{\sigma}_j^{2(I)}\right| \left|\hat{\sigma}_j^{2(I_1)}\right|^{-1} \right)$,

6. $\beta_{2,j} = \ln\left(\operatorname{abs}\left|V_{zz,j}^{(I)}\right| \operatorname{abs}\left|V_{zz,j}^{(I_1)}\right|^{-1} \frac{|I_1|^\eta}{|I|^\eta} \right)$,

7. $\beta_{3,j} = \ln\left(\operatorname{abs}\left|V_{zz,j}^{(I)}\right| \left|\hat{\sigma}_j^{2(I)}\right|^{-\eta} |I|^{-\eta} \right)$,

8. $\beta_{4,j} = \ln\left(\operatorname{abs}\left|V_{zz,j}^{(I)}\right| \left|V_{yy,j}^{(I)}\right|^{-\eta} \right)$,

9. $\beta_{5,j} = V_{yy,j}^{(I)} \hat{\sigma}_j^{-2(I)} |I|^{-1}$,

where $|V_{zz,j}^{(I)}|$ and $|V_{zz,j}^{(I_1)}|$ are statistics (4.28) computed on all image lattice I and its sublattice I_1, respectively. Analogically, $\hat{\sigma}_j^{2(I)}$ and $\hat{\sigma}_j^{2(I_1)}$ are estimates of σ_j^2 on I and I_1, respectively. Instead of the logarithm, we can use an alternative normalisation of invariants $\beta_{1,j} - \beta_{5,j}$ based on the geometric mean:

10. $\beta_{8,j} = \left(\left|\hat{\sigma}_j^{2(I)}\right| \left|\hat{\sigma}_j^{2(I_1)}\right|^{-1} \right)^{\frac{1}{2}}$,

11. $\beta_{9,j} = \left(\text{abs} \left| V_{zz,j}^{(I)} \right| \text{abs} \left| V_{zz,j}^{(I_1)} \right|^{-1} \frac{|I_1|^{\eta}}{|I|^{\eta}} \right)^{\frac{1}{2\eta}}$,

12. $\beta_{10,j} = \left(\text{abs} \left| V_{zz,j}^{(I)} \right| \left| \hat{\sigma}_j^{2(I)} \right|^{-\eta} |I|^{-\eta} \right)^{\frac{1}{2}}$,

13. $\beta_{11,j} = \left(\text{abs} \left| V_{zz,j}^{(I)} \right| \left| V_{yy,j}^{(I)} \right|^{-\eta} \right)^{\frac{1}{2}}$,

14. $\beta_{12,j} = \sqrt{ \left| V_{yy,j}^{(I)} \right| \left| \hat{\sigma}_j^{2(I)} \right|^{-1} |I|^{-1} }$.

The invariants $\beta_{1,j}$, $\beta_{2,j}$, $\beta_{8,j}$, $\beta_{9,j}$ were computed from $V_{zz,j}$, $\hat{\sigma}_j^2$ estimates in different regions I_1 , I in the texture, e.g. beginning and all image.

Similarly to the 2D CAR model, the definition $\nu_s = \text{diag}\, A_s$ (item 2.) should be used with the GMRF model preceded with K-L transformation. Additional invariance to mutual swap of spectral planes can be provided by sorting of corresponding invariants (e.g. $\alpha_{1,1}, \ldots, \alpha_{1,C}$).

4.3   Local intensity changes

All previous colour invariants were derived with the assumption of uniform illumination. We show that most of them are also invariant to locally constant intensity changes, which can be caused by cast shadows or objects with more textured planar surfaces.

Let us start with an auxiliary construction, which do not include intensity changes for now. We assume that a textured image is composed of n copies of the same small texture tile S, which is homogeneously illuminated. The tiles are placed side by side to cover the whole image lattice I (Fig. 4.1). Using the formula (3.5), the relation of parameter estimate on the tile $\hat{\gamma}^{(S)}$ and the estimate on the whole image $\hat{\gamma}^{(I)}$ is

$$\hat{\gamma}^{(S)} = \left(V_{zz}^{(S)} \right)^{-1} \left(V_{zy}^{(S)} \right) \approx \left(\sum_{r \in S} Z_r Z_r^T \right)^{-1} \left(\sum_{r \in S} Z_r Y_r^T \right) \ ,$$

$$\hat{\gamma}^{(I)} = \left(V_{zz}^{(I)} \right)^{-1} \left(V_{zy}^{(I)} \right) \approx \left(\sum_{r \in I} Z_r Z_r^T \right)^{-1} \left(\sum_{r \in I} Z_r Y_r^T \right)$$

$$\approx \left(n \sum_{r \in S} Z_r Z_r^T \right)^{-1} \left(n \sum_{r \in S} Z_r Y_r^T \right) \tag{4.29}$$

$$= \hat{\gamma}^{(S)} \ ,$$

where the first approximations ignores prior information and the last approximation (4.29) discards statistics at seams of the tiles. If the tiles are seamless the equation is precise. Consequently, if tiles are sufficiently large, the contribution of seam statistics decreases and parameter estimate $\hat{\gamma}$ is almost the same for the tile and the whole image.
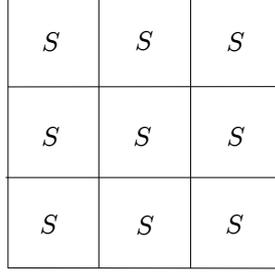
Figure 4.1: Example of image coverage with texture tiles S.

Let us suppose that a modified image is composed of the same texture tiles S, where the ℓ-th texture tile is modified by the multiplication of all its pixels with some constant b_ℓ. This simulates locally constant intensity changes in the image. The parameter estimate become

$$
\begin{aligned}
\tilde{\hat{\gamma}}^{(I)} &= \left(\tilde{V}_{zz}^{(I)}\right)^{-1}\left(\tilde{V}_{zy}^{(I)}\right) \approx \left(\sum_{r\in I}\tilde{Z}_r\tilde{Z}_r^T\right)^{-1}\left(\sum_{r\in I}\tilde{Z}_r\tilde{Y}_r^T\right) \\
&\approx \left(\left(\sum_{\ell=1}^n b_\ell^2\right)\sum_{r\in S}Z_rZ_r^T\right)^{-1}\left(\left(\sum_{\ell=1}^n b_\ell^2\right)\sum_{r\in S}Z_rY_r^T\right) \\
&= \hat{\gamma}^{(S)}\ ,
\end{aligned}
\tag{4.30}
$$

where $\tilde{\hat{\gamma}}^{(I)}$, $\tilde{Z}_r$, $\tilde{Y}_r$, $\tilde{V}_{zz}^{(I)}$ are related to the illumination modified image. The last approximation again discards the seam statistics, which additionally includes local illumination changes. The previous assumption of the tile composed image can be further weaken. The image tiles could be even different on condition that the correlation statistics $\sum_{r\in S}Z_rZ_r^T$ and $\sum_{r\in S}Z_rY_r^T$ remain the same, which is implicated by the homogeneous property of textures; natural examples are stochastic textures. The equation (4.30) instantly implies that illumination invariants $\operatorname{tr}A_s$, $\nu_{s,j}$ are approximately invariant to local intensity changes. Analogically, it can be proved for the invariants α_2, $\beta_3-\beta_5$, and $\beta_{10}-\beta_{12}$ (see Appendix Section A.2).

In fact, the previous model of invariance to local intensity changes includes independence to size of texture sample (more data, not scale), it is a special case for $b_\ell=1$, $\ell=1,\ldots,n$. Almost all previously derived colour invariants: $\alpha_{1'}-\alpha_3$, $\beta_1-\beta_5$, $\beta_8-\beta_{12}$ comply with this independence to sample size. Exceptions are β_6 and β_7, which depend on texture sample size, because probabilities $p(\tilde{Y}_r|\tilde{Y}^{(r-1)})$ (3.11) and $\ln p(\tilde{Y}^{(t)}|M_\ell)$ (3.12) include non-linear functions of the number of previously analysed data, e.g. power to $\psi(t)$. Of course, a texture sample with sufficient size is required for a reliable estimation of the textural features and subsequent invariants.

The last consequence of equation (4.30) concerns the textural features without illumination invariants, described in Sections 3.1.3, 3.1.4, and 3.1.5. The part of feature

vector that include only textural features A_s, $\forall s \in I_r$ is invariant to simple linear brightness changes. However, illumination colour changes require the invariants derived in this section.

4.4 Discussion

Implementation of illumination invariant features is straightforward and efficient, because most of the used statistics are computed during the parameter estimation process. It is advantageous to compute invariants $\beta_1 - \beta_{10}$ from Cholesky decomposition of V_{zz}^{-1}, which is used in parameter estimation process and which allows efficient computation of $|V_{zz}|^{\frac{1}{n}}$ and $\ln|V_{zz}|$ without risk of overflow. Because the ordering of eigenvalues is arbitrary in general, the eigenvalues $\nu_{s,j}$ are ordered according to their absolute values for the particular s and $j = 1 \ldots C$.

To summarize the contribution of the previous colour invariants: they are used to transform MRF textural features (Section 3.1) into a colour invariant form. The proposed invariants are computed from the model parameter estimates, which are computed at each pyramid level of the Gaussian pyramid. Finally, the invariants from all pyramid levels are concatenated to form a common feature vector. Unfortunately, numeric scale of invariants (especially β_ℓ) may differ significantly, therefore the invariants should be used together with norms or classifiers that normalises the range of features, e.g. FC_3 (3.24).

Although generally, illumination invariance is a beneficial property of textural features, it is necessary to verify that the invariants have retained the dicriminability. Otherwise, the features would be useless despite their illumination invariance. The ability of the proposed invariants to discriminate among textures is confirmed in texture recognition tests in Sections 6.1.

In addition to illumination invariance, the features $\alpha_{1'} - \alpha_3$, $\beta_1 - \beta_{12}$ may be also rotation invariant, but it depends on the shape of used contextual neighbourhood I_r. The overall textural representation is not rotation invariant due to $\operatorname{tr} A_s$, $\nu_{s,j}$ features. Rotation invariant texture recognition and combined rotation and colour invariant features are introduced in the next chapter.

Chapter 5

Rotation Invariance

Our concern is the recognition of materials and their in-plane rotations, which are rotations with the rotation axis perpendicular to material surface. In principle, there are three approaches how to deal with different orientation of textures. The first way, known as exhaustive search, generates all possible rotations of the input texture and finds the best match over all its rotations. Although the method yields good and stable results, it is obviously not suitable for practical applications, because computational demands are extremely high. The second approach searches for a canonical representation (rotation) of textures and each texture is rotated into this particular orientation before classification or other texture processing. The method based on this paradigm is described in Section 5.1. Finally, the last group of methods uses texture description by features that do not change with texture rotation, such features are referred as rotation invariants. This kind of method is introduced in Section 5.2.

The rotation normalisation or construction of rotation invariants assume that the material rotation can be modelled as a rotation of its image. Unfortunately, this assumption does not apply for rough surfaces and illumination near surface plane (Chantler, 1995). However, we imagine a rotation of rough material as a two step process. In the first step, the material sample and the illumination source are rotated around the same axis, as they were firmly tied together. This step can be modelled as a simple image rotation and it is handled by the proposed rotation invariants. The second step consists of illumination rotation into its final position. This situation is supposed to be dealt with the proposed illumination invariants, despite the fact that they were derived with the assumption of fixed illumination position. The reason is that our experiments with natural surfaces (Sections 6.1.3 and 6.1.4) show that the derived illumination invariants are robust to change of illumination direction.

5.1 Orientation normalisation

The normalisation approach tries to find the rotation of each texture, which is uniquely determined and which can be robustly detected. The detection robustness is the key issue for practical applications of these kinds of methods. We detect the dominant texture
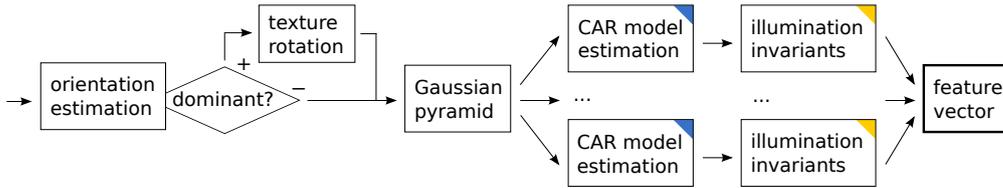
Figure 5.1: Texture analysis algorithm with orientation normalisation.

orientation by means of histogram of gradient orientations and we also estimate the significance of detected orientation (Vacha and Haindl, 2009). The similar algorithm is used to determine orientation of SIFT keypoints (Lowe, 2004).

The input texture is converted into a grey-scale image G and its gradient is estimated by means of differences:

$$\nabla G(r_1, r_2) = [G(r_1 + 1, r_2) - G(r_1 - 1, r_2), G(r_1, r_2 + 1) - G(r_1, r_2 - 1)] \ ,$$

at each pixel position $r = (r_1, r_2)$. Subsequently, a histogram of gradient orientations is computed. Each gradient orientation is weighted by its magnitude and the linear interpolation is used to assign the gradient weight into two adjacent bins. Since the textures are supposed to be homogeneous, gradients in some direction induce the gradients in the opposite direction. Therefore, their orientations are fitted into the interval $[0°, 180°)$, grater angles are decreased by $180°$ to fit in. We use histogram with 36 bins and 6 pass histogram smoothing (averaging of three adjacent bins) is applied after the histogram computation.

The detected dominant orientation is considered to be significant, if the height of the second highest peak in the histogram is lower than 80% of the highest peak. If there is no second highest peak, the sum of the highest peak bin and its two adjacent bins have to be grater than 150% of the expected value for three bins. The purpose of this preprocessing is detection of textures with a strong orientation and alignment of their rotation. The rotation of textures with less significant directions or nondirectional textures is not required, because they are aptly represented by the CAR model.

The overall algorithm scheme is depicted in Fig. 5.1. We start with the estimation of the dominant texture orientation. If the texture is significantly directional, it is rotated according to its dominant orientation. The texture is factorised into K levels of the Gaussian pyramid and subsequently each pyramid level is modelled by the 3D or 2D CAR model (Sections 3.1.3 and 3.1.4). The CAR model parameters are estimated and illumination invariants are subsequently computed from them (Section 4.2). Finally, the illumination invariants from all the models are concatenated into one feature vector.

As it was mentioned, possible drawback of any normalisation approach lies in the ambiguity of detected orientation. That is why we rotate textures only if a significant and unique orientation is detected. Another disadvantage is that orientation of gradients can be influenced by illumination direction. In the following section, we introduce an approach based on rotation invariants, which does not suffer from these drawbacks.

5.2 Rotation invariance

The rotation invariants are textural features that do not change with texture rotation. The important property of rotation invariants is how they retain their discriminability, because without sufficient discriminability the features would be useless despite their invariance. We propose two different methods for the rotation invariance of MRF features (Vacha et al., submitted). The first method computes rotation invariant features before the estimation of MRF parameters. While the second method build rotation invariants after the MRF parameter estimation by means of moment invariants.

5.2.1 Rotation autoregressive random model

The Rotation Autoregressive Random (RAR) model is inspired by the model of Kashyap and Khotanzad (1986), who estimated a regression model of pixel values and averages on concentric circles around these pixels. Although, this model is suitable for modelling of isotropic textures, it has difficulties with anisotropic texture properties. Our model uses multispectal images and extends the regression data with maximum and minimum from circular samples, which enables the model to capture some anisotropic texture properties.

The basic modelling equation is similar to (3.4):

$$Y_r = \gamma \, Z_r + \epsilon_r \;\; , \qquad Z_r = [Y_{r,\mathrm{mean}}^T \, , Y_{r,\mathrm{max}}^T \, , Y_{r,\mathrm{min}}^T : \forall s \in I_r^{\odot}]^T \;\; ,$$

with the difference in data vector Z_r. The vector Z_r is now composed of average $Y_{r,\mathrm{mean}}$ maximum $Y_{r,\mathrm{max}}$ and minimum $Y_{r,\mathrm{min}}$, all computed separately for each concentric circle in the neighbourhood $I_r^{\odot}$. The neighbourhood $I_r^{\odot}$ is composed of points sampled on concentric circles, e.g. with radii 1, 2, 3. The bilinear interpolation is used for interpolation of sampled points. For multispectral images, $Y_{r,\mathrm{mean}}$, $Y_{r,\mathrm{max}}$, and $Y_{r,\mathrm{min}}$ are $C \times 1$ vectors, where mean, maximum and minimum are computed for each spectral plane separately.

The RAR model is used either in 3D or 2D version, which are similar to 3D CAR (Section 3.1.3) or 2D CAR (Section 3.1.4). The differences in the contextual neighbourhood and the datavector Z_r cause that the parameter estimate $\hat{\gamma}_t$ cannot be computed using the analytical Bayesian estimate (3.5) or (3.14), anymore. Therefore we use the corresponding LS approximation (3.9) or its 2D version.

To achieve simultaneous rotation and colour invariance, the feature vector is composed of the colour invariants derived for CAR models, i.e. $\mathrm{tr}\, A_s$, $\nu_{s,j}$, $\alpha_1 - \alpha_3$, and additionally $\beta_1 - \beta_5$, $\beta_8 - \beta_{12}$ (β_6 and β_7 are not used because they are not valid for the RAR model).

5.2.2 Rotation moment invariants

The rotation moment invariants are used to describe anisotropic texture properties, which are only briefly captured by the RAR model. The CAR model parameters are estimated (Section 3.1.3 and 3.1.4) and the rotation moment invariants are computed

from the illumination invariants $\operatorname{tr} A_s$, $\nu_{s,j}$ (Section 4.2), according to their position in the unilateral neighbourhood I_r^u. Since the unilateral neighbourhood I_r^u covers only the upper half plane, the values are duplicated in the central symmetry to cover the entire plane, which is advantageous for the rotation invariance of moments. The moments are computed separately for each spectral plane of $\nu_{s,j}$. We also add moment invariants that describe interspectral relations of $\nu_{s,j}$.

Moment invariants

It is advantageous to compute the rotation invariants from complex moments, because they change more simply in rotation than other types of moments. The complex moment of the order $p + q$ of the function $f(r_1, r_2)$ is defined

$$c_{pq}^{(f)} = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} (r_1 + ir_2)^p (r_1 - ir_2)^q f(r_1, r_2) \, \mathrm{d}r_1 \, \mathrm{d}r_2 \ , \tag{5.1}$$

where i is an imaginary unit. We omit the superscript $(\cdot)^{(f)}$ if there is no danger of confusion. It follows from the definition that only the indices $p \geq q$ are meaningful because $c_{pq} = c_{qp}^*$ (the asterisk denotes complex conjugate).

The complex moment c'_{pq} after a rotation of function $f(r_1, r_2)$ equals

$$c'_{pq} = e^{-i(p-q)\theta} \cdot c_{pq} \ , \tag{5.2}$$

where θ is the rotation angle. Therefore a product of n complex moments to the powers of k_ℓ is

$$\prod_{\ell=1}^{n} c_{p_\ell q_\ell}^{k_\ell} \ , \tag{5.3}$$

which is invariant to rotation, if the sum of the first indices equals the sum of the second indices, i.e.

$$\sum_{\ell=1}^{n} k_\ell p_\ell = \sum_{\ell=1}^{n} k_\ell q_\ell \qquad \text{or} \qquad \sum_{\ell=1}^{n} k_\ell (p_\ell - q_\ell) = 0 \ .$$

The total number of moment invariants equals $\ddot{m} - \ddot{\tau}$, where $\ddot{m}$ is the number of real values of moments (complex moment with $p \neq q$ has two real values – real and imaginary parts) and $\ddot{\tau}$ is the number of transform parameters. In our case, the rotation has one parameter (the angle), i.e. $\ddot{\tau} = 1$ and the number of real values of moments of the order $p+q$ is $p+q+1$. Moment c_{00} is an exception (there should be no zeroth-order invariant according to this rule, but the transform parameter is dependent in this case and it is not counted). The set of invariants should be chosen to be independent, see Flusser et al. (2009); Flusser and Suk (2006) for more details and additional references.

In our case, we must consider the behaviour of the complex moments of symmetric functions. A function has so-called N-fold rotation symmetry (N-FRS), if it repeats itself when it rotates around its centroid by $2\pi\ell/N$ for all $\ell = 1, \ldots, N$. The central

symmetry is a special case of N-FRS, where $N = 2$. If $f(x, y)$ has N-FRS and $(p - q)/N$ is not an integer, then $c_{pq} = 0$. The reason is following: If f' is a rotated version of f by the angle $2\pi/N$, then, because of its symmetry, it must be the same as the original. Therefore, it must hold $c'_{pq} = c_{pq}$ for any p and q. At the same time, it follows from equation (5.2) that

$$c'_{pq} = e^{-2\pi i(p-q)/N} \cdot c_{pq} \ .$$

Since $e^{2\pi i \zeta} = 1$ for an integer ζ and $(p - q)/N$ is assumed not to be an integer, this equation can be fulfilled only if $c_{pq} = 0$.

Since our neighbourhood is centrally symmetric, we cannot use any odd-order moment. That is why we use these even-order rotation moment invariants:

1. zeroth order: c_{00}

2. second order: c_{11}, $c_{20}c_{02}$

3. fourth order: c_{22}, $c_{40}c_{04}$, $c_{31}c_{13}$

4. mixed order: $\mathcal{Re}(c_{40}c_{02}^2)$, $\mathcal{Re}(c_{31}c_{02})$.

We can utilize the fact that all colour channels are rotated together, by the same angle and construct joint colour rotation invariants

5. second order: $c_{20}^{(\ell)}c_{02}^{(j)}$,

where $\ell = 1$, $j = 2, \ldots, C$ are the individual colour channels. This full set of moment invariants is denoted as $\mathrm{m}_1(model)$. Since the high order moments tend to be numerically unstable, especially for roughly defined f, we also work with the reduced set of invariants denoted as $\mathrm{m}_2(model)$:

1. reduced set of moments: c_{00}, c_{11}, $c_{20}c_{02}$, c_{22}, and $c_{20}^{(1)}c_{02}^{(j)}$.

We have the matrix of discrete values instead of a continuous function here, therefore we must use a discrete approximation of the complex moments (5.1):

$$\hat{c}_{pq}^{(f)} = \sum_{r_1} \sum_{r_2} (r_1 + ir_2)^p (r_1 - ir_2)^q f(r_1, r_2) \ . \tag{5.4}$$

The bilinear interpolation of function $f(r_1, r_2)$ is used to enhance its resolution and precision of computed moments.

Combination with illumination invariants

Discrete complex moments $\hat{c}_{pq}$ are computed for invariants $\mathrm{tr}\, A_s$ and $\nu_{s,j}$, $j = 1, \ldots, C$ (defined in Section 4.2) according to their position in the unilateral neighbourhood I_r^u.

The input function f is correspondingly defined and made centrally symmetric as

$$f_A(r_1, r_2) = \begin{cases} \operatorname{tr} A_{(r_1, r_2)} & (r_1, r_2) \in I_r^u \\ \operatorname{tr} A_{(-r_1, -r_2)} & (-r_1, -r_2) \in I_r^u \\ 0 & \text{otherwise} \end{cases}, \qquad (5.5)$$

and for each spectral plane j as

$$f_{\nu, j}(r_1, r_2) = \begin{cases} \nu_{(r_1, r_2), j} & (r_1, r_2) \in I_r^u \\ \nu_{(-r_1, -r_2), j} & (-r_1, -r_2) \in I_r^u \\ 0 & \text{otherwise} \end{cases}. \qquad (5.6)$$

Subsequently, the previous set of moment invariants $(1. - 4.)$ is computed. The interspectral moment invariant $c_{20}^{(1)} c_{02}^{(j)}$ is computed only from multispectal function $f_{\nu, j}(r_1, r_2)$. Altogether, it makes 34 moment invariants for $C = 3$ and the full set $\mathrm{m}_1(model)$ (8 invariants from $f_A(r_1, r_2)$ + 8 invariants from $f_{\nu, j}(r_1, r_2)$ for each j + $c_{20}^{(1)} c_{02}^{(2)}$ + $c_{20}^{(1)} c_{02}^{(3)}$).

The moment invariants assume that a rotation of texture will result in rotated parameter estimates A_s. The estimation of A_s, according to (3.5), relays on datavector correlation and the relative positions s are used only in the forming datavector Z_r, no further. Therefore the reordering of the neighbourhood values (e.g. texture flip) will result in the same values of estimates A_s, but reordered accordingly. Therefore, we also expect that texture rotation by a factor of 90° will produce approximately rotated estimates, if computed on the square orthogonal lattice and with the hierarchical unilateral neighbourhood. However, an interpolation, which necessary follows an arbitrary texture rotation, influences the parameter estimates.

The illumination invariants α_1, α_2, α_3, and $\beta_1 - \beta_{12}$ (Section 4.2) are not associated with a position in the contextual neighbourhood, therefore the rotation invariant transformation is not needed, if they are computed with a model with suitable neighbourhood shape (the reasoning from the previous paragraph applies). Therefore the illumination invariants α_1, α_2, α_3, and $\beta_1 - \beta_{12}$, computed with hierarchical unilateral neighbourhood, can be added directly into the rotation invariant feature vector.

5.2.3 Texture analysis algorithm

The scheme of rotation invariant texture analysis is depicted in Fig. 5.2. The texture analysis algorithm starts with the factorisation of texture into K levels of the Gaussian down-sampled pyramid (it captures larger spatial relations), followed by modelling with two different MRF models. At first, each pyramid level is modelled by the RAR model and the illumination invariants are computed from its parameters. Secondly, each pyramid level is modelled by the CAR model. After the estimation of CAR model parameters, the illumination invariants and subsequently the rotation moment invariants are computed from the parameter estimates. Finally, the features from all the models are concatenated into one feature vector. Both RAR and CAR models exist either in 3D or in 2D version.
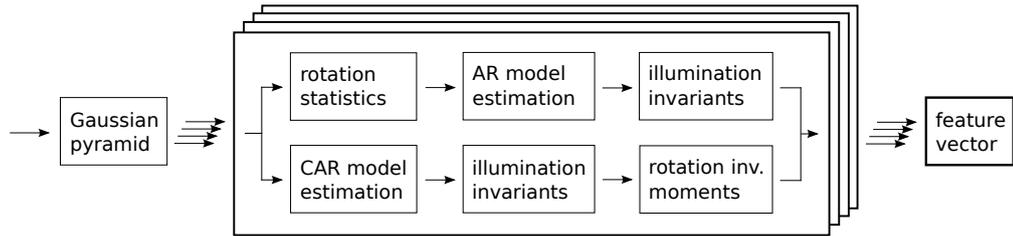
Figure 5.2: Texture analysis algorithm which combines illumination invariants with two approaches to rotation invariance. It is either a autoregressive model of rotation invariant statistics (RAR) in the upper line, or a causal autoregressive model followed by the computation of rotation moment invariants (m(CAR)) in the lower line.

In our experiments, we usually use $K = 4$ levels of Gaussian pyramid and the CAR models with the 6-th order hierarchical neighbourhood (cardinality $\eta = 14$), which corresponds to maximum radius 3 used in the RAR models. However, the optimal size of neighbourhood and pyramid depends on the size of input images, because the models require enough data for a reliable parameter estimation. The moment based features are composed of either a full or reduced set of invariants, in the results, they are denoted as "$m_1(model)$" or "$m_2(model)$", respectively.

It is advantageous to compute distance between two feature vectors using the fuzzy contrast (3.24), because it normalises different scales of moment features. However, the fuzzy contrast requires estimate of average and standard deviation of all features.

Chapter 6

Experimental Results

The proposed illumination invariant and rotation invariant textural features were tested in the task of natural and artificial material recognition under various circumstances. The experiments were conducted on five different textural databases, which differ in the variability of image acquisition conditions and include almost 25 000 of images in total.

In Section 6.1, we tested the performance of illumination invariant features in texture retrieval and texture classification tasks under various illumination conditions. In Sections 6.2 and 6.3, we tested the texture recognition with illumination variations in combination with different texture rotations and viewpoint positions. Such variations of acquisition conditions are usually encountered in an analysis of real-world scenes.

6.1 Illumination invariant features

The performance of the illumination invariant MRF features (proposed in Section 4.2) is demonstrated on three image databases, each with different variations of illumination conditions. At first, the Outex texture database was acquired with three illuminations with different spectra and only with slight differences in illumination positions, which complies with our theoretical assumptions. Secondly, a BTF database was acquired with a fixed illumination spectrum and with 91 different illumination directions, which drastically violates our restrictive assumption of fixed illumination position. Finally, the most difficult setup combined changes in illumination spectrum and direction, and also added slight viewpoint variation.

The textural representation is based on modelling by a MRF type of model, where models are either 2D CAR, 3D CAR, or GMRF. The models were computed over K levels of the Gaussian pyramids, which were built either directly on C spectral planes or on spectral planes decorrelated by K-L transformation (indicated with "-KL" suffix). Most of the experiments were computed with $K = 4$ pyramid levels. If the images were large enough, additional pyramid level was employed ($K = 5$), which is denoted with "↑" prefix.

method	experiment		
	i1	i2	i3
2D CAR-KL	**325**	**132**	**260**
2D CAR	**325**	**132**	**260**
3D CAR	295	108	236
GMRF-KL	310	120	248
Gabor features, RGB	144	144	144
Gabor features	48	48	48
Opponent Gabor features	**252**	**252**	**252**
Steerable pyramid features, RGB	2904	2904	2904
Gabor features, RGB, norm.	144	144	144
Gabor features, norm.	48	48	48
Opponent Gabor features, norm.	**252**	**252**	**252**
$\text{LBP}_{8,1+8,3}$, RGB	1536	1536	1536
$\text{LBP}_{16,2}^{u2}$, RGB	729	729	729
$\text{LBP}_{16,2}^{riu2}$, RGB	54	54	54
$\text{LBP}_{8,1+8,3}$	**512**	**512**	**512**
$\text{LBP}_{16,2}^{u2}$	243	243	243
$\text{LBP}_{16,2}^{riu2}$	18	18	18

Table 6.1: The size of feature vectors used in the experiments with illumination invariance.

The CAR models were tested with illumination invariant features: $\text{tr}\,A_s$, $\nu_{s,j}$, α_1, α_2, α_3, while the GMRF model was tested with features: $\text{tr}\,A_s$, $\nu_{s,j}$ and α_2, α_3 (see definitions in Section 4.2). The features $\nu_{s,j}$ were computed as diagonals of A_s if used with K-L transformation, otherwise the eigenvalues were used. Usually, we used the models with the 6-th order hierarchical contextual neighbourhood (cardinality $\eta = 14$), which is depicted in Fig. 3.2. Additionally, we compared three different distances of the feature vectors: L_1, $L_{0.2}$, and fuzzy contrast FC_3 (details in Section 3.2).

The proposed textural features were compared with the following alternatives: Gabor features (monochromatic), opponent Gabor features, steerable pyramid features, and the LBP features (see details in Sections 2.2.2, 2.2.4). The grey value based features as Gabor features and LBP were computed either on grey-scale images or separately on each spectral plane of colour images and concatenated (this is denoted with "RGB" suffix in the experiments). Moreover, Gabor features and opponent Gabor features were tested with and without normalisation of spectral channels, described in Section 2.3.1 and denoted with "norm." suffix. The mean and standard deviation of features (required by some dissimilarities) were estimated from all images.

The feature vector sizes used in the following Experiments i1 – i3 are compared in Tab. 6.1. The size of feature vectors in Experiment i4 is displayed within the experiment, because it additionally includes combinations of the invariant features β_ℓ.

6.1.1   Experiment i1 – Outex retrieval

In the first experiment, we demonstrate the performance of the proposed illumination invariant features on the Outex texture database (Ojala et al., 2002a), which contains images of 318 materials acquired under three different illuminations. The illumination sources were: 2856K incandescent CIE A light source – "inca", 2300K horizon sunlight – "horizon", and 4000K fluorescent TL84 – "tl84", the illumination positions were very close. All the images were acquired with a fixed camera position. Some example images are displayed in Fig. 6.1 and in Appendix Figs. B.2, B.3. The Outex texture database can be downloaded from (database Outex).

The tested task is illumination invariant image retrieval from the Outex texture database. In this task (Vacha and Haindl, 2007a, 2010b), a CBIR system tries to retrieve images similar to a given query image. The test set consisted of images acquired with 3 different illuminations for each material, without any rotation and with 100 dpi resolution, which is 954 images in total. All images were cropped to size 512×512 pixels. The relevant images to the query image were defined as images of the same material with the other two illuminations. Therefore there were 2 relevant images present in the test set for each query image, a total amount of 3 images was retrieved. The retrieval performance was measured using the recall rate (RR):

$$\mathrm{RR} = \frac{|\{\text{retrieved images}\} \cap \{\text{relevant images}\}|}{|\{\text{relevant images}\}|} \ . \tag{6.1}$$

We tested the image retrieval using every image from the test set and the averaged results are summarised in Tab. 6.2. Furthermore, the images were corrupted with an additive Gaussian noise to test noise robustness of the features.

The MRF models were computed with the sixth order hierarchical neighbourhood, which consisted in $\eta = 14$ neighbours (see Fig. 3.2), and $K = 5$ levels of the Gaussian pyramid, since the images were large enough. The size of feature vectors is displayed in Tab. 6.1. The best performance 94% was achieved with "↑ 2D CAR" model with FC_3 dissimilarity, which clearly presents their insensitivity to illumination spectrum variations. Moreover, the MRF features without K-L transformation performed better on noisy images than features with K-L transformation, which was deflected by uncorrelated noise. The LBP features also showed their illumination invariance property with 83% recall rate. However, their performance dropped down on noise corrupted images. These results also demonstrate that the spectral channel normalisation is essential for Gabor and steerable pyramid features, nevertheless, any variant of Gabor or steerable pyramid features did not perform satisfactory in this test.

Some examples of retrieved textures are presented in Fig. 6.1. The first two retrieved images are usually both correct, since the recall rate is about 94%. However, more interesting are image retrieved at further positions, where we can observe similar materials successfully recognised by the proposed features. In general comparison with LBP, the proposed MRF features with pyramids prefer overall structure contrary to micro patterns (such as lines) preferred by the LBP features. Additional examples of retrieved images are in Appendix Fig. B.3.
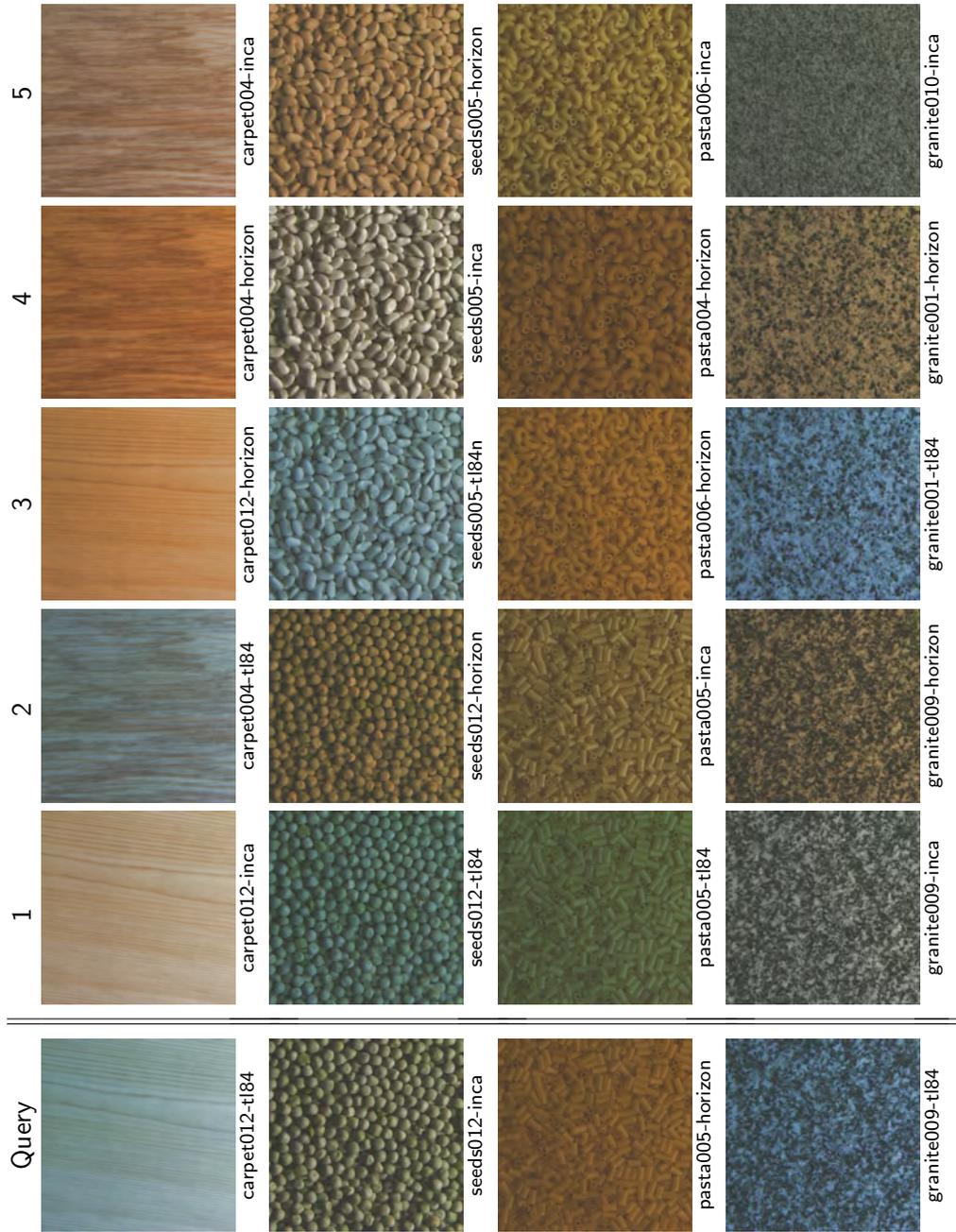
Figure 6.1: Experiment i1: Examples of illumination invariant retrieval from Outex texture database using "↑ 2D CAR-KL, FC_3" method. The query images are followed by retrieved images in order of similarity. From the left, query images are: carpet012-tl84, seeds012-inca, pasta005-horizon, granite009-tl84. The images of query materials acquired under different illumination spectra were successfully retrieved at positions $1-3$.

	added noise σ	
method	0	8
$\uparrow$ 2D CAR-KL L_1	83.4	60.8
$\uparrow$ 2D CAR-KL, $L_{0.2}$	87.8	69.4
$\uparrow$ 2D CAR-KL, FC_3	90.2	68.0
$\uparrow$ **2D CAR, FC$_3$**	**94.0**	**83.0**
$\uparrow$ 3D CAR, L_1	79.6	61.5
$\uparrow$ 3D CAR, $L_{0.2}$	79.1	58.8
$\uparrow$ 3D CAR, FC_3	**82.7**	**63.0**
$\uparrow$ 3D CAR-KL, FC_3	**83.1**	55.0
$\uparrow$ GMRF-KL, L_1	73.4	50.4
$\uparrow$ GMRF-KL, $L_{0.2}$	80.7	60.2
$\uparrow$ GMRF-KL, FC_3	81.3	55.9
Gabor features, RGB	14.0	13.4
Gabor features	42.8	42.4
Opponent Gabor features	38.8	30.5
Steerable pyramid features, RGB	19.4	18.9
Gabor features, RGB, norm.	40.4	27.5
Gabor features, norm.	**53.4**	**56.1**
Opponent Gabor features, norm.	46.9	37.8
Steerable pyramid features, RGB, norm.	41.2	39.4
LBP$_{8,1+8,3}$, RGB	51.5	20.0
LBP$^{u2}_{16,2}$, RGB	47.3	11.7
LBP$^{riu2}_{16,2}$, RGB	24.3	3.1
LBP$_{8,1+8,3}$	**83.1**	**50.3**
LBP$^{u2}_{16,2}$	80.6	40.8
LBP$^{riu2}_{16,2}$	61.5	21.3

Table 6.2: Experiment i1: Illumination invariant retrieval from the Outex texture database. The performance is measured as mean recall rate [%] with 3 retrieved images.

6.1.2 Experiment i2 – OUTEX_TC_00014

The second experiment (Vacha and Haindl, 2007a, 2010b) was performed on the Outex classification test set number 14 (Ojala et al., 2002a). It is also based on Outex database and it can be freely downloaded from (database Outex). The main differences to the Experiment i1 are: multiple training images, smaller resolution of images, and smaller subset of materials.

In this setup, 68 materials selected from the Outex database were treated in the following manner. Twenty subsamples with size 128×128 were extracted from each material image. The training set consisted of 10 samples per material, all illuminated with the 2586K incandescent CIE A light source. The test set consisted of 10 remaining subsamples for each material, all of them illuminated with the other two illuminants. Consequently, the training set consisted of 680 images, while the test set was composed of 1360 images. The classification was performed using the three Nearest Neighbours (3-NN) classifier as in Maenpaa et al. (2002); Pietikäinen et al. (2002).

The highest reported classification accuracy on the test set (Maenpaa et al., 2002) was 69% for "$\text{LBP}^{u2}_{16,2}$" features, which outperformed Gabor features with 66% accuracy (unfortunately our implementation of Gabor features reached only 54.5% in Tab. 6.3), the both features were computed on grey-scale images. Moreover, Pietikäinen et al. (2002) reported 68.4% accuracy for "$\text{LBP}_{8,1+8,3}$" also on grey-scale images, and 53.3% accuracy achieved by opponent Gabor features on colour images preceded by comprehensive colour normalisation.

In addition to the previously described experiment, we also degraded all images with an additive Gaussian noise (zero mean, variance σ^2). The experiment was performed directly on noisy images, without any noise removal method. The application of such method might increase classification accuracy, but only on condition that it would not introduce any artificial micro structures into the images.

The MRF models had to be restricted to the third order hierarchical neighbourhood ($\eta = 6$ neighbours) and $K = 4$ levels of Gaussian pyramid, due to small image sizes. As a consequence, the feature vectors of MRF features are about four times smaller than the vector of "$\text{LBP}_{8,1+8,3}$" features. The best results on the original test set were achieved with "$\text{LBP}_{8,1+8,3}$" on grey-scale images with 71.6% closely followed by "3D CAR-KL, FC_3" with 69.5% correct classification. However, the results changed dramatically with added noise, the "$\text{LBP}_{8,1+8,3}$" features dropped down to 38.6% showing their vulnerability to noise degradation. The MRF based features are not so noise sensitive, because Gaussian noise is an inherent part of the model and the Gaussian pyramid suppresses noise at its higher levels. In this experiment, Gabor features performed better than opponent Gabor features, especially on noisy images.

Quite surprisingly, the best results of the proposed features were achieved by "3D CAR-KL, FC_3" method, which comprise K-L transformation. The advantage of "3D CAR-KL" model is using features ν_s defined as diagonals instead of eigenvalues of matrices A_s used with "3D CAR". During the computation, the eigenvalues are ordered according to their absolute values, therefore relations between spectral planes for different indexes $s \in I_r$ are lost. On the other hand, the K-L transformation can be

method	added noise σ			
	0	2	4	8
2D CAR-KL, L_1	**67.6**	60.8	55.7	52.3
2D CAR-KL, $L_{0.2}$	66.3	60.5	55.2	51.0
2D CAR-KL, FC_3	**67.5**	**63.3**	55.8	51.0
2D CAR, FC_3	**67.5**	62.2	**61.0**	**56.6**
3D CAR, L_1	63.6	61.3	60.6	54.9
3D CAR, $L_{0.2}$	63.5	59.7	55.4	47.4
3D CAR, FC_3	65.3	60.4	58.0	51.3
3D CAR-KL, **FC_3**	**69.5**	**65.3**	**64.3**	**60.0**
GMRF-KL, L_1	61.5	57.0	51.1	46.1
GMRF, FC_3	64.3	59.9	58.1	53.8
Gabor features, RGB	37.5	37.0	36.2	35.6
Gabor features	44.3	43.3	43.2	41.3
Opponent Gabor features	50.7	49.3	45.3	37.3
Steerable pyramid features, RGB	37.5	35.9	34.9	32.6
Gabor features, RGB, norm.	57.0	59.9	60.3	57.1
Gabor features, norm.	**54.5**	**61.3**	**63.3**	**62.9**
Opponent Gabor features, norm.	**56.7**	55.8	54.3	47.9
$LBP_{8,1+8,3}$, RGB	66.8	56.6	48.8	36.7
$LBP_{16,2}^{u2}$, RGB	62.0	52.9	41.2	28.7
$LBP_{16,2}^{riu2}$, RGB	44.6	30.8	22.6	15.3
$LBP_{8,1+8,3}$	**71.6**	**62.2**	**54.6**	**38.6**
$LBP_{16,2}^{u2}$	67.6	60.4	49.8	33.0
$LBP_{16,2}^{riu2}$	56.9	45.2	34.2	19.7
Maenpaa et al. (2002):				
Gabor features	66			
$LBP_{16,2}^{u2}$	69			
Pietikäinen et al. (2002):				
Opponent Gabor features, norm.	53.3			
$LBP_{8,1+8,3}$	68.4			

Table 6.3: Experiment i2: The results [%] of the Outex classification test OU-TEX_TC_00014. The classification was performed using 3-NN classifier.

deflected by noise, which was demonstrated in the previous experiment.

An additional experiment is included in Appendix Section B.1, where the CAR and GMRF models were tested with and without features α_1, α_2, α_3. Moreover, due to the directionality of CAR models, we also tested the CAR models estimated in additional directions.

6.1.3 Experiment i3 – Bonn BTF

The third experiment (Vacha and Haindl, 2008a, 2010b) was designed to test the feature robustness against illumination direction changes, which are in contradiction with our theoretical assumptions.

The experiment was performed on BTF texture images, which are from the University of Bonn BTF database (Meseth et al., 2003) and consist of fifteen BTF colour measurements: ceiling, corduroy, two fabrics, walk way, foil, floor tile, pink tile, impalla, proposte, pulli, wallpaper, wool, and two lacquered wood textures (see Fig. B.4). Ten of these measurements are now publicly available (database Bonn BTF). Each BTF mate-
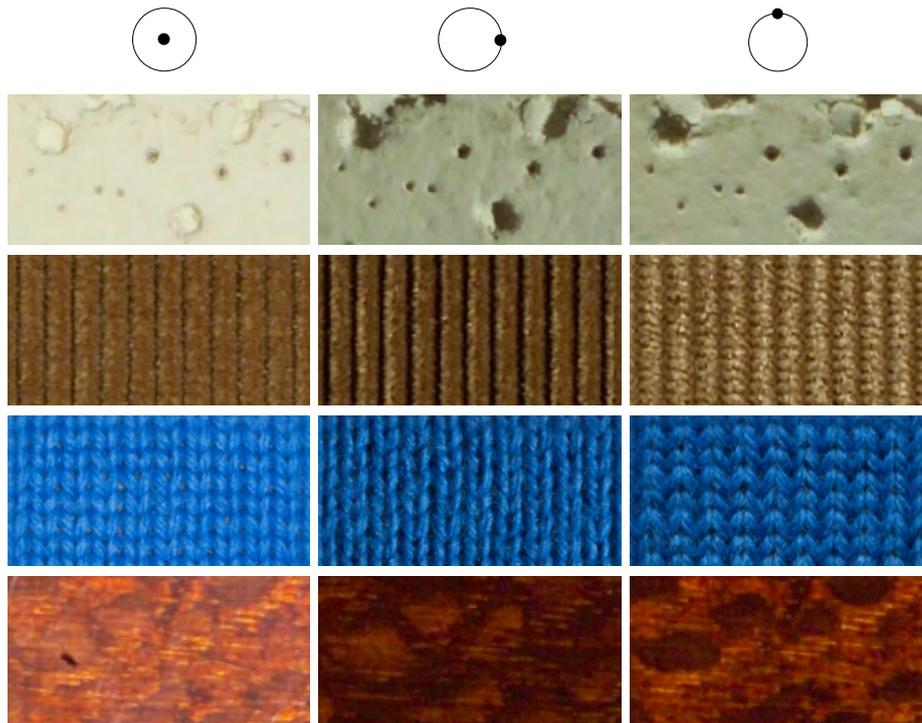
Figure 6.2: Effects of illumination direction changes on selected Bonn BTF material samples (rows from top): ceiling, corduroy, wool, lacquered wood1. Columns from the left consist of illumination with declination angle: 0°, 60°, 60° with different azimuth angle.
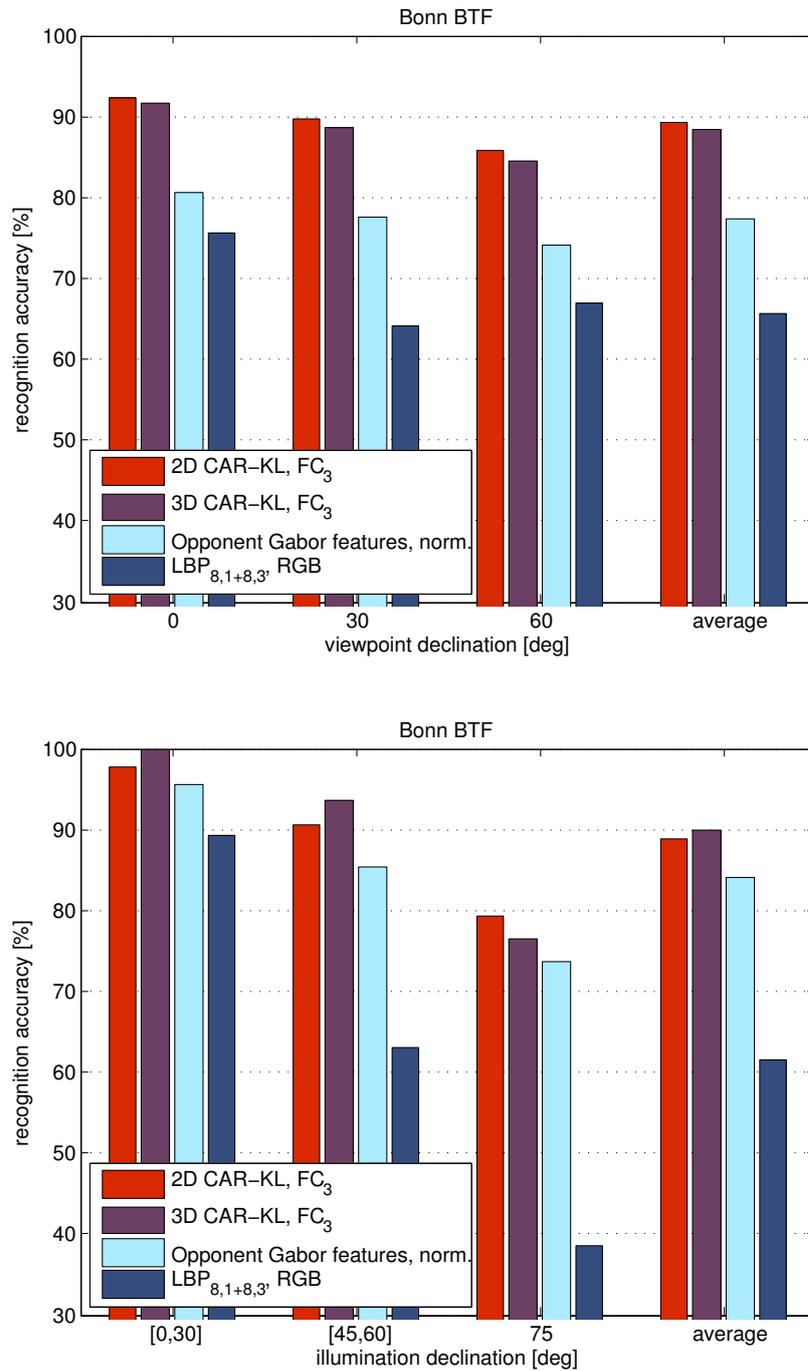
Figure 6.3: Experiment i3a: Accuracy of material recognition [%] on the Bonn BTF database, using a single training image per material. On the top, training images were randomly selected within the three image sets. In the bottom, training images were fixed to perpendicular illumination and the results are grouped by illumination of test images.

rial is measured in 81 illumination and 81 camera positions as an RGB image ($C = 3$). Examples of material appearance under varying illumination direction are shown in Fig. 6.2 and Appendix Figs. B.5, B.6. We prepared three image sets, which included all illumination positions, but differed in selected viewpoint positions. The declination angle of viewpoint position from the surface normal was 0°, 30°, and 60°, successively, in-plane texture rotation was not included. Each set consisted in $15 \times 81 = 1215$ images, all cropped to the same size 256×256 pixels.

The proposed features were compared with the same alternative features as in the previous experiments. The MRF models were computed with the sixth order hierarchical neighbourhood ($\eta = 14$ neighbours, see Fig. 3.2) and $K = 4$ levels of the Gaussian pyramid, the size of feature vectors is listed in Tab. 6.1.

The experiment contains two parts: i3a and i3b. The first one focuses on the classification with a single training image per material, while the second part consists of retrieval of similar texture images.

Results

In the first part of this experiment, a single training image per each material was randomly selected and the remaining images were classified using the Nearest Neighbour (1-NN) classifier. The results were averaged over 10^5 random selections of training images. The experiment was performed separately on each of the three image sets differing in viewpoint position, and the results were averaged again.

The best results are depicted on the top of Fig. 6.3, and the exact values of classification accuracy are displayed in Tab. 6.4. It can be observed that the best performance 90.3% was achieved with "2D CAR-KL, L_1" method, closely followed with the same model with FC_3 dissimilarity. The best alternative features were opponent Gabor features with the average performance 77.4%, the best of LBP features achieved 65.6%. Standard deviation was bellow 4% for Gabor features and LBP features, and below 3% for CAR and GMRF models. Although the LBP features are invariant to brightness changes, these results demonstrate their inefficiency to handle illumination direction variations. Rotation invariant LBP features are more capable, however rotating illumination cannot be modelled as a simple image rotation. For the MRF features, the worst classification were for ceiling and fabric2 materials. The ceiling material was misclassified as floor tile (for illumination near surface), and fabric2 was sometimes misclassified as fabric1, since they have very similar structures.

Furthermore, we explored how the performance depends on the light source declination from the surface normal. Only the image set with the viewpoint fixed at 0° declination was used and the single training sample per each material was selected, so that all the training samples were illuminated with 0° declination angle (perpendicular illumination), the other 1200 images were classified. The results depicted in Tab. 6.5 and in the bottom of Fig. 6.3 show that the recognition accuracy decreases as the illumination position of test samples move away from the training sample position. The best results were achieved by "3D CAR-KL, FC_3" with the average classification 89.9%, similar results 88.8% were achieved by "2D CAR, FC_3" method.

method	viewpoint declination angle			
	$0°$	$30°$	$60°$	average
2D CAR-KL, L_1	92.4	91.1	87.5	**90.3**
2D CAR-KL, $L_{0.2}$	91.8	89.5	85.8	89.0
2D CAR-KL, **FC_3**	**92.3**	**89.6**	**85.7**	**89.2**
2D CAR, FC_3	88.7	87.3	82.9	86.3
3D CAR, L_1	87.4	84.3	78.9	83.5
3D CAR, $L_{0.2}$	89.2	85.7	81.0	85.3
3D CAR, FC_3	89.8	86.1	80.2	85.4
3D CAR-KL, **FC_3**	**91.4**	**88.7**	**84.5**	**88.2**
GMRF-KL, L_1	89.6	86.3	81.0	85.6
GMRF-KL, $L_{0.2}$	87.1	83.7	79.6	83.5
GMRF-KL, FC_3	86.5	82.6	78.7	82.6
Gabor features, RGB	71.7	64.6	60.1	65.5
Gabor features	69.8	62.9	55.6	62.8
Opponent Gabor features	**82.5**	77.7	71.7	77.3
Steerable pyramid features, RGB	72.3	65.5	60.4	63.1
Gabor features, RGB, norm.	60.1	58.1	57.9	58.7
Gabor features, norm.	50.8	50.1	51.3	50.7
Opponent Gabor features, norm.	**80.5**	**77.6**	**74.2**	**77.4**
$LBP_{8,1+8,3}$, RGB	**65.7**	**64.2**	**67.0**	**65.6**
$LBP_{16,2}^{u2}$, RGB	62.5	61.6	64.6	62.9
$LBP_{16,2}^{riu2}$, RGB	**68.4**	60.7	57.4	62.2
$LBP_{8,1+8,3}$	61.2	61.1	65.4	62.6
$LBP_{16,2}^{u2}$	55.7	56.3	60.7	57.6
$LBP_{16,2}^{riu2}$	58.6	52.1	52.5	54.4

Table 6.4: Experiment i3a: Accuracy of material recognition [%] on the Bonn BTF database, using a single training image per material. The results were averaged over 10^5 random selections of training images. The columns contain results for three image sets differing in viewpoint position, the averages are in the last column.

method	light source declination			
	$[0°, 30°]$	$[45°, 65°]$	$75°$	average
2D CAR-KL, L_1	96.3	87.5	78.3	86.7
2D CAR, FC_3	96.7	91.6	78.1	88.7
2D CAR-KL, $L_{0.2}$	96.7	85.4	78.3	85.8
2D CAR-KL, FC_3	**97.8**	**90.5**	**79.4**	**88.8**
3D CAR, L_1	97.8	89.6	75.6	87.2
3D CAR, $L_{0.2}$	97.8	91.2	72.8	87.2
3D CAR, FC_3	99.3	93.6	76.7	89.8
3D CAR-KL, **FC_3**	**100**	**93.7**	**76.4**	**89.9**
GMRF-KL, L_1	95.9	82.6	65.3	80.4
GMRF-KL, $L_{0.2}$	94.4	82.3	68.3	80.8
GMRF-KL, FC_3	93.3	86.5	72.2	83.7
Gabor features, RGB	**96.3**	71.4	28.9	64.2
Gabor features	95.2	64.7	34.7	62.6
Opponent Gabor features	95.6	83.9	50.0	76.4
Steerable pyramid features, RGB	90.7	69.5	36.1	64.3
Gabor features, RGB, norm.	81.9	49.1	19.4	47.6
Gabor features, norm.	81.9	38.8	13.9	41.0
Opponent Gabor features, norm.	**95.6**	**85.3**	**73.6**	**84.1**
$\text{LBP}_{8,1+8,3}$, RGB	**89.3**	**63.0**	**38.6**	**61.6**
$\text{LBP}_{16,2}^{u2}$, RGB	84.4	51.4	35.6	54.1
$\text{LBP}_{16,2}^{riu2}$, RGB	84.4	44.6	31.9	49.8
$\text{LBP}_{8,1+8,3}$	86.3	57.4	38.3	58.2
$\text{LBP}_{16,2}^{u2}$	79.3	50.7	34.7	52.3
$\text{LBP}_{16,2}^{riu2}$	74.1	36.8	16.7	39.2

Table 6.5: Experiment i3a: Accuracy of material recognition [%] on the Bonn BTF database with training images fixed to the perpendicular illumination. The performance is grouped for different intervals of illumination declination angles of test images, the last column is average for all test images. Viewpoint declination angle was $0°$.

method	Bonn BTF		Bonn BTF public	
	RR_{88}	MAP	RR_{88}	MAP
2D CAR-KL, L_1	88.1	**91.0**	93.3	95.2
2D CAR-KL, **FC$_3$**	**88.1**	90.5	**95.1**	**96.5**
3D CAR-KL, L_1	81.4	85.2	89.0	91.7
3D CAR-KL, FC_3	84.3	86.9	93.2	95.0
Opponent Gabor features	75.6	79.9	81.2	85.5
Opponent Gabor features, norm.	75.3	80.3	78.8	83.7
LBP$_{8,1+8,3}$, RGB	65.4	69.2	73.9	76.9

Table 6.6: Experiment i3b: Retrieval of similar textures from the Bonn BTF database. The results are evaluated by mean recall rate for 88 retrieved images and mean average precision (RR_{88} and MAP) [%]. The last two columns contain result with the image sets restricted to the publicly available material measurements.

In the second part of the experiment, denoted as i3b, we tested a retrieval of similar texture images from Bonn BTF database. The performance was evaluated using recall rate (6.1) and average precision. The average precision (AP) is defined as the average of precisions computed at position of every relevant retrieved image:

$$AP = \frac{\sum_{\ell=1}^{N} PR_\ell \cdot rel(\ell)}{|\{relevant\ images\}|} \quad , \tag{6.2}$$

$$PR_\ell = \frac{|\{relevant\ images\ retrieved\ at\ position\ \ell\ or\ less\}|}{\ell} \quad , \tag{6.3}$$

where N is the size of image database, $rel(\ell) = 1$ if a relevant image is retrieved at ℓ-th position and $rel(\ell) = 0$ otherwise. The retrieval was preformed for every image in a image set and means of RR and AP were computed.

The experiment was performed separately on the image sets with three different viewpoint positions and the results were averaged again. The final results are displayed in Tab. 6.6, where the last two columns contains the results on the image sets restricted to the publicly available BTF measurements: ceiling, corduroy, walkway, floor tile, pink tile, impalla, proposte, pulli, wallpaper, and wool (see Fig. B.4). The best results were achieved by "2D CAR-KL, FC_3" method with more than 10% improvement to alternative methods. The other methods from Tab. 6.4 are not displayed, since the results were worse.

6.1.4 Experiment i4 – ALOT

In this experiment (Vacha and Haindl, 2010a), we tested the proposed features in the recognition of natural and artificial materials under various light conditions, which is needed in a real scene analysis. We focused on the feature robustness under changing

illumination spectrum and direction and we also included slight viewpoint changes, which were limited to declination angle variations. Additionally, the feature vectors based on MRF features were extended with β_ℓ invariants (see definition in Section 4.2).

The images of materials are from the recently created Amsterdam Library of Textures (ALOT) (Burghouts and Geusebroek, 2009b). The ALOT is a BTF database containing an extraordinary collection of 250 materials, each acquired with varying viewpoint and illumination positions, and one additional illumination spectrum. Most of the materials have rough surfaces, so the movement of light source changes the appearance of materials. Moreover, the significant height variation of some materials (e.g. leaves) causes large and variable cast shadows, which make the recognition even more difficult. Example images from ALOT texture database are shown in Figs. 1.2, 6.6, however, this experiment do not include texture rotations. The ALOT database can be downloaded from (database ALOT).

This experiment consists of two parts: i4a and i4b. In the part i4a, we used one half of the dataset of Burghouts and Geusebroek (2009b) to exclude multiple texture rotations. It consisted of images of the first 200 materials divided into parameter tuning, training, and test sets (3×1200 images). Let c stands for camera, l for light, i for reddish illumination, and r for optional material rotation. The tuning set consisted in samples $c\{1,4\}l\{1,4,8\}r60°$; the training set was defined as $c\{1,4\}l\{1,4,8\}$ and the test set contained setups $c\{2,3\}l\{3,5\}$, $c3l2$, and $c1i$. We cropped all the images to the same size 1536×660 pixels. The recognition accuracy was evaluated on the test set images, where the nearest neighbour (1-NN) classifier was trained on 4 images per material, randomly selected from the training set.

In the part i4b, we used images of all 250 materials, with all light setups, no rotations and cameras 1 and 3, which is 14 images per material. One training image per material was randomly selected and the others were classified with the 1-NN classifier. This test was performed separately for images from the camera 1 and 3, the results were averaged (2×1750 images in total). As a consequence this experiment did not include recognition under viewpoint variation, which is in contrast with the part i4a.

The proposed illumination invariant features were computed at $K = 5$ levels of the Gaussian pyramid, using the 6-th order hierarchical neighbourhood ($\eta = 14$ neighbours). The proposed features were again compared with the most frequented features as in the previous experiments. Additionally, we also included recently published LBP-HF features (Ahonen et al., 2009). The mean and standard deviation of features, which are required by the FC_3 dissimilarity and dissimilarity of Gabor features and opponent Gabor features, were estimated on the parameter tuning set of Experiment i4a and on all images in Experiment i4b.

Results

Both parts of the experiment were computed for 10^3 random selections of training images and average classification results are shown in Tab. 6.7. Standard deviations were below 0.5% and 1.4% for part i4a and i4b, respectively. The best results were achieved with "↑ 2D CAR-KL $\beta_1 - \beta_{12}$, FC_3" method, which includes all proposed invariants

74

method	experiment		
	i4a	i4b	size
↑ 2D CAR-KL, FC_3	48.5	67.2	325
↑ **2D CAR-KL $\beta_1 - \beta_{12}$, FC$_3$**	**55.7**	**69.9**	505
↑ 2D CAR-KL $\alpha_{1'}, \beta_1 - \beta_{12}, FC_3$	**56.3**	69.1	505
↑ 2D CAR-KL $\beta_1 - \beta_7, FC_3$	54.4	69.8	430
↑ 2D CAR-KL $\beta_6 \beta_7, FC_3$	51.4	68.6	355
↑ 2D CAR-KL $\beta_6 - \beta_{12}, FC_3$	54.8	**70.0**	430
↑ 3D CAR-KL, FC_3	51.2	65.1	295
↑ 3D CAR no eigenvalues, opponent, FC_3	49.6	**69.1**	295
↑ 3D CAR no eigenvalues, FC_3	46.6	66.0	295
↑ 3D CAR, FC_3	47.4	65.2	295
↑ 3D CAR-KL $\beta_1 - \beta_{12}, FC_3$	56.3	**68.6**	355
↑ 3D CAR-KL $\alpha_{1'} \beta_1 - \beta_{12}, FC_3$	**56.7**	68.2	355
↑ GMRF-KL, FC_3	36.6	52.2	310
↑ GMRF-KL $\beta_1 - \beta_5, \beta_8 - \beta_{12}, FC_3$	47.7	59.0	460
Gabor features, RGB	**44.6**	34.0	144
Opponent Gabor features	41.8	**53.1**	252
LBP$_{8,1+8,3}$	32.8	39.8	512
LBP$_{8,1+8,3}$, RGB	**41.2**	45.6	**1536**
LBP$^u_{16,2}$, RGB	38.6	43.4	729
LBP$^{riu2}_{8,1+24,3}$, RGB	34.2	42.6	108
LBP-HF$_{8,1+24,3}$	32.6	**50.0**	340

method	i4a	i4b	size
2D CAR-KL, FC_3	40.2	60.8	260
2D CAR-KL $\beta_1 - \beta_{12}, FC_3$	**46.3**	**64.3**	404
3D CAR, FC_3	39.2	59.2	236
3D CAR $\beta_1 - \beta_{12}, FC_3$	43.1	60.5	284
3D CAR-KL $\beta_1 - \beta_{12}, FC_3$	**47.7**	**63.0**	284

Table 6.7: Experiment i4: Accuracy of material recognition [%] on the ALOT texture database, with additional β_ℓ colour invariants. The values were averaged over 10^3 random selections of training images and the last column contains feature vector sizes. The bottom table displays the results without the additional level of Gaussian pyramid ($K = 4$).
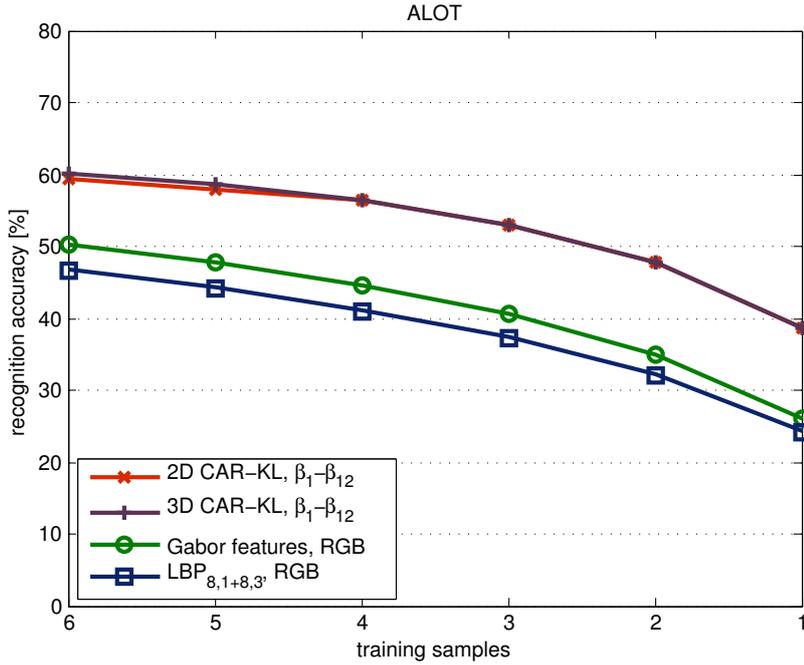
Figure 6.4: Experiment i4b: Accuracy of material recognition [%] on the ALOT texture database for different numbers of random training images per material. The values were averaged over 10^3 random selections of training images.

and ↑ stands for five pyramid levels. Improvements to alternative features are 11%, 17% and the addition of invariants β_ℓ is responsible for 7%, 3% increase of the performance, however, the feature vector with β_ℓ invariants is substantially longer. Tab. 6.7 also displays performance of "2D CAR-KL" model with different groups of the illumination invariants β_ℓ.

We also experimented with invariants β_6, β_7 computed without $\Gamma(x)$ function terms (see formulae (3.11), (3.12)), which saves 20% of overall computation time since the evaluation of $\Gamma(x)$ is very time consuming. The performance differences were below the rounding error in Tab. 6.7 as expected, so that we suggest this implementation. Additionally, the alternative definition $\alpha_{1'}$ was tested instead of invariant α_1 (see Section 4.2.1 for definitions). The results denoted as "↑ 2D CAR-KL $\alpha_{1'}$, $\beta_1 - \beta_{12}$, FC_3" show performance increase in Experiment i4a, but decrease in i4b, therefore additional experiments are required to decide which definition is advantageous. Finally, the bottom of Tab. 6.7 displays the results with $K = 4$ level of Gaussian pyramid in order to make the results comparable with the previous experiments. As expected, the results confirmed that higher pyramid levels are advantageous for high resolution textures, because they are able to capture larger spatial relations.

A further comparison of the correct classification in the part i4a is shown in Fib. 6.4, which displays the progress for different numbers of training samples. The performance

superiority of the CAR features is maintained for all numbers of training samples and it is even more significant as the number of training samples decreases. It is obvious that the results of 2D and 3D models are almost identical. The reason is that the employed K-L transformation is used in conjunction with the features ν_s computed as the diagonals of the parameter matrices A_s instead of eigenvalues of A_s (see Section 4.2.1 for details). This effectively pushes the 3D model into the 2D form and throws away any interspectral information.

Moreover, the 3D CAR-KL performed better than the version without K-L transformation. We speculate that it was caused by avoiding the computation of eigenvalues, because after the computation of eigenvalues it is not guaranteed that the spectral plane components in the features ν_s will correspond for different $s \in I_r$. To test this hypothesis we used ν_s computed as the diagonals and no K-L transformation, the results are denoted as "↑ 3D CAR no eigenvalues, FC_3" in Tab. 6.7. Although the correspondence of spectral planes was maintained, the results did not improved the original "↑ 3D CAR, FC_3", therefore another properties are important. One of them is that too much off-diagonal information was thrown away in "↑ 3D CAR no eigenvalues, FC_3". This is confirmed by transforming the image into the opponent colour model, denoted as "↑ 3D CAR no eigenvalues, opponent, FC_3", which improved the results, because this transformation partly decorrelates spectral planes. In comparison with the other decorrelation by means of K-L transformation "↑ 3D CAR-KL, FC_3", the results are better in Experiment i4b, but worse in i4a. We conclude that it would be advantageous to design some modification of the features that would preserve the correspondence of spectral planes and simultaneously it neither discard interspectral information of matrices A_s nor rely on a decorrelation.

The final remark compares the results of Experiments i4b and i4b. Although Experiment i4b used only a single training image per material, the results are about 15% better than in Experiment i4a. The reason is that Experiment i4a included a viewpoint variation, which was even grater in the test set than in the training set. Moreover, the hold-out methodology used in Experiment i4a produces a lower bound on classification accuracy.

The proposed features were approximately 1.5× slower than "$\text{LBP}^{riu2}_{8,1+24,3}$" and 4× faster than Gabor features.

6.1.5 Discussion

To summarise the previous experiments, we conclude that the proposed illumination invariant features confirmed their invariance to changes of illumination colour and brightness. They are also considerably robust to changes of illumination direction and image degradation with an additive Gaussian noise. Most importantly, the illumination invariants retained the discriminability and outperformed the alternative textural features in the texture recognition tests. A summary of the tested recognition conditions is displayed in Tab. 6.8.

The overall best results were achieved with the illumination invariants based on the 2D CAR model with K-L transformation followed by the results of 3D CAR model.

	Experiment					
	i1	i2	i3a	i3b	i4a	i4b
texture database	Outex		Bonn BTF		ALOT	
experiment conditions:						
illumination spectrum	$+$	$+$	$-$	$-$	$+$	$+$
illumination direction	$-$	$-$	$+$	$+$	$+$	$+$
viewpoint azimuth	$-$	$-$	$-$	$-$	$-$	$-$
viewpoint declination	$-$	$-$	$-$	$-$	$+$	$-$
experiment parameters:						
image size (bigger)	512	128	256	256	1536	1536
number of materials	318	68	15	15, 10	200	250
result tables	6.2	6.3	6.4, 6.5	6.6	6.7	6.7

Table 6.8: Parameters of experiments with illumination invariance and comprised variations of recognition conditions.

However, we suppose that the 3D CAR model would benefit from its generality if larger textures are available. We suggest using the CAR models with the sixth order hierarchical neighbourhood and four levels of Gaussian pyramid as it produced steadily good results. The same parameters are also used in the experiments through the following sections. Naturally, the results on large textures can be improved with additional levels of the pyramid and the third order hierarchical neighbourhood is more suitable for textures smaller than 128×128.

The most of the discriminative information is concentrated in the invariants ν_s and $\operatorname{tr} A_s$, however, the addition of invariants $\alpha_1 - \alpha_3$, $\beta_1 - \beta_{12}$ still improves the performance. We expect that feature selection methods (Liu and Yu, 2005; Somol et al., 2010) can be used to evaluate a mutual dependency or redundancy of the features and to improve the classification results. The definition of features ν_s as diagonals of the matrices A_s is preferred to eigenvalues, because it preserves the ordering according to image planes and it should be accompanied with some decorrelation of spectral planes.

Moreover, the fuzzy contrast FC_3 outperformed the other tested dissimilarities of feature vectors. Mean and standard deviation of features, which are required by fuzzy contrast, can be estimated with a sufficiently precision on a small subset of images. This is confirmed in the next experiments with rotation invariance, where these statistics are estimated on a fragment of dataset or a specially defined tuning set. If such estimate is not available, we suggest using L_1 norm without β_ℓ invariants. We also suppose that it is possible to estimate means and deviations of features, which can be generally used in texture comparison, however, additional experiments are required to confirm this hypothesis.

Additionally, the proposed illumination invariants are also fast to compute and the feature vector has a reasonable low dimension. A disadvantage is that a reliable estimation of the MRF parameters requires a sufficient size of training data. An interactive

demonstrations of the performance of the proposed features are available online (Vacha and Haindl, 2008b, 2007b).

The rest of the chapter presents experiments with methods, which deal with illumination changes in combination with a rotation of textures.

6.2 Rotation normalisation and illumination invariant features

This section presents experimental results of the method which combines illumination invariants with a normalisation of texture orientation. The texture is rotated in accordance with its estimated dominant orientation (see Section 5.1 for more details). The experiment (Vacha and Haindl, 2009) tests the robustness of this normalisation approach under varying illumination direction and also limited viewpoint changes, which are compensated by texture rotation.

We evaluated the material recognition on the CUReT database (Dana et al., 1999) to be able to compare our results with alternative approaches even if this database is already overcome by BTF measurements from the University of Bonn. CUReT is the first BTF database and it consists of 61 real-world materials captured under different combinations of viewing and illumination directions. The dataset provided by Varma and Zisserman (2005) consists of 61 materials, each with 92 images differing in viewpoint and illumination positions, the resolution is 200×200 pixels. The dataset of Varma and Zisserman (2005) can be downloaded from (dataset CUReT).

We followed the experimental setup of Burghouts and Geusebroek (2009b), where the classification accuracy was tested with randomly selected training samples and the SVM classifier. The number of training samples per material decreased from 8 to 1. The mean and standard deviation of correct classification was computed over 10^3 repetitions (random selections of training images). On the contrary, we used only the simple 1-NN classifier.

Additionally to the published results (Burghouts and Geusebroek, 2009b) of rotation and illumination invariant MR8 features, we compared also the performance of some other most frequented features. It is worth to note that Gabor features, opponent Gabor features, and LBP are rotation variant, only exception is rotation invariant "$\text{LBP}_{16,2}^{riu2}$". The LBP features were also computed on texture images transformed into opponent colour space, which is denoted with "opponent" suffix in the results.

The CAR features were computed for $K = 4$ levels of Gaussian pyramid, using the 6-th order hierarchical neighbourhood. We also experimented with combination of two models, where each level of Gaussian pyramid was modelled by two models with different neighbourhood I_r. In that case, illumination invariants for both models were included in the final feature vector. We used 6-th and 3-th order hierarchical neighbourhoods, which consisted in $\eta = 14$ and $\eta = 6$ neighbours, respectively.

Results

Fig. 6.5 shows classification accuracy averaged over 10^3 random repetitions. It is directly comparable to the results of Burghouts and Geusebroek (2009b), where the correct classification with MR8-LINC monotonously decreased, approximately, from 75% to 45% for 8 to 1 training samples. The best performance was achieved with the combination of two models "2D CAR-KL 6+3 rot. norm., FC_3" and it went from 87% to 53%

method	accuracy	size
2D CAR-KL rot. norm., L_1	75.6	260
2D CAR-KL rot. norm., FC_3	75.1	260
2D CAR-KL 6+3 rot. norm., L_1	77.0	392
2D CAR-KL 6+3 rot. norm., FC$_3$	**77.6**	**392**
3D CAR rot. norm., L_1	69.7	236
3D CAR rot. norm., FC_3	67.6	236
3D CAR 6+3 rot. norm., L_1	72.4	344
3D CAR 6+3 rot. norm., FC_3	**72.6**	344
Gabor features	61.7	144
Opponent Gabor features	**68.7**	252
LBP$_{8,1+8,3}$, RGB	**70.9**	**1536**
LBP$^{u2}_{16,2}$, RGB	68.7	729
LBP$^{riu2}_{16,2}$, RGB	64.2	54
LBP$_{8,1+8,3}$	66.9	512
LBP$_{8,1+8,3}$, opponent	57.4	1536
LBP$^{u2}_{16,2}$, opponent	69.7	729
Burghouts and Geusebroek (2009b):		
MR8	58	600
MR8-NC	54	600
MR8-INC	60	600
MR8-LINC	**67**	600
MR8-SLINC	57	600

Table 6.9: Accuracy of material recognition [%] on the CUReT dataset, using 4 random training images per texture. The values were averaged over 10^3 random selections of training images and the last column contains feature vector sizes.
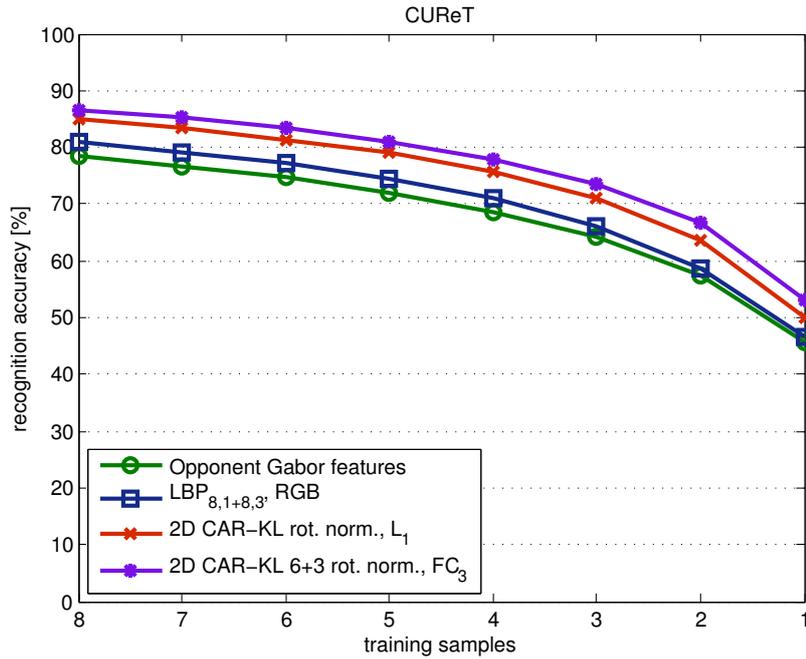
Figure 6.5: Accuracy of material recognition on the CUReT dataset for different numbers of random training images, mean values were computed over 10^3 repetitions.

with standard deviations from 0.6% to 1.5%. The suffix "rot. norm." stands for the rotation normalisation by means of dominant orientation estimation. More details are displayed in Tab. 6.9, which compares classification accuracy of different features, all standard deviations were below 1%. The best performance 77,6% was again achieved with "2D CAR-KL 6+3 rot. norm., FC_3" method, closely followed by a single model "2D CAR-KL rot. norm., L_1". The best alternative features were "LBP$_{8,1+8,3}$" with average performance 70.9% and four times longer feature vector.

Unfortunately, illumination direction changes can influence the detection of dominant orientation and result in possible misclassification, especially for rough textures. This difficulty is solved by rotation invariants features, which outperformed this normalisation approach and whose results are presented in the following section.

6.3   Rotation and illumination invariant features

We present the performance of the proposed method which combines illumination invariant CAR features with rotation invariance (described in Section 5.2). The comparison (Vacha et al., submitted) was performed on four different texture databases in three experimental setups. The first experiment is focused on the robustness of textural features under varying illumination and viewpoint positions, which resembles real-world scenes with natural materials. In the second experiment, we tested features under varying illumination spectrum and texture rotation, which simulates different day light or artificial illuminations. In the third experiment, our results were compared with other recently published features.

The proposed features were again computed on $K = 4$ levels of Gaussian pyramid, which were built either directly on C spectral planes or on spectral planes decorrelated by K-L transformation (indicated with "-KL" suffix). The CAR models were estimated with the 6-th order hierarchical neighbourhood ($\eta = 14$ neighbours), which corresponds to maximum radius 3 used in the RAR models. The moment based features are composed of either a full set of invariants "$m_1(model)$" or reduced set of invariants "$m_2(model)$", both sets are defined in Section 5.2.2. The feature vector sizes are displayed together with the results in each experiment. Finally, the feature vectors are compared in fuzzy contrast FC_3 (see Section 3.2), since the normalisation of different feature scales is necessary. The feature means and standard deviations, which are required by fuzzy contrast, were estimated either on a parameter tuning set or on a training set if the tuning set was not available.

The proposed features were compared with the following illumination and rotation invariant features: MR8-NC and MR8-LINC (which were reported with the best performance from MR8-* texton methods in Burghouts and Geusebroek (2009b)), $LBP_{P,R}^{riu2}$ and LBP-HF features. The details of these methods are described in Sections 2.2.5, 2.2.4.

6.3.1   Experiment $\varrho 1$ – ALOT, CUReT

In the first experiment in this section, we followed the experimental setup of Burghouts and Geusebroek (2009b) and evaluated the texture recognition accuracy on CUReT (Dana et al., 1999) and ALOT (Burghouts and Geusebroek, 2009b) datasets.

As it was mentioned, the ALOT library is a BTF database containing a collection of 250 natural and artificial materials, each acquired with varying viewpoint and illumination positions, plus one illumination spectrum. Most of the materials have rough surfaces, which result in significant variations of their appearance, including variable cast shadows (see example images in Figs. 1.2, 6.6). The ALOT database is available for download at (database ALOT).

The dataset of Burghouts and Geusebroek (2009b) consisted of images of the first 200 materials divided into parameter tuning, training, and test sets, each with 2400 images. Let c stands for camera, l for light, i for reddish illumination, and r for material rotation. The parameter tuning set consisted in samples with setups $c\{1,4\}l\{1,4,8\}r\{60°, 180°\}$; the training set contained images with $c\{1,4\}l\{1,4,8\}r\{0°, 120°\}$ and finally, the test

Figure 6.6: Example materials from the ALOT dataset and their appearance for different camera and light conditions. The two columns on the right are acquired from viewpoint with declination angle $60°$ from the surface macro-normal.

set was defined as $c3l2r\{0°, 120°\}$, $c\{2, 3\}l\{3, 5\}r\{0°, 120°\}$, $c2l2r0°$, and $c1ir0°$. Additionally, we cropped all the images to the same size 1536×660 pixels.

The CUReT database also consists of real-world materials acquired with different combinations of viewing and illumination directions. The dataset provided by Varma and Zisserman (2005) contained of 61 materials, each with 92 samples differing in viewpoint and illumination positions; image resolution was 200×200 pixels. This dataset is freely available and it can be downloaded from (dataset CUReT). Since the dataset did not define any parameter tuning set, we defined it as the subset of training set which contained the first four samples of each material.

In the setup of Burghouts and Geusebroek (2009b), the classification accuracy was tested with randomly selected training samples from the training set and the SVM classifier. The number of training samples per material decreased from 8 to 1, mean and standard deviation of classification accuracy were computed over 10^3 repetitions (random selections of training images). We differ only in the classifier, where the simple 1-NN was employed instead of SVM.
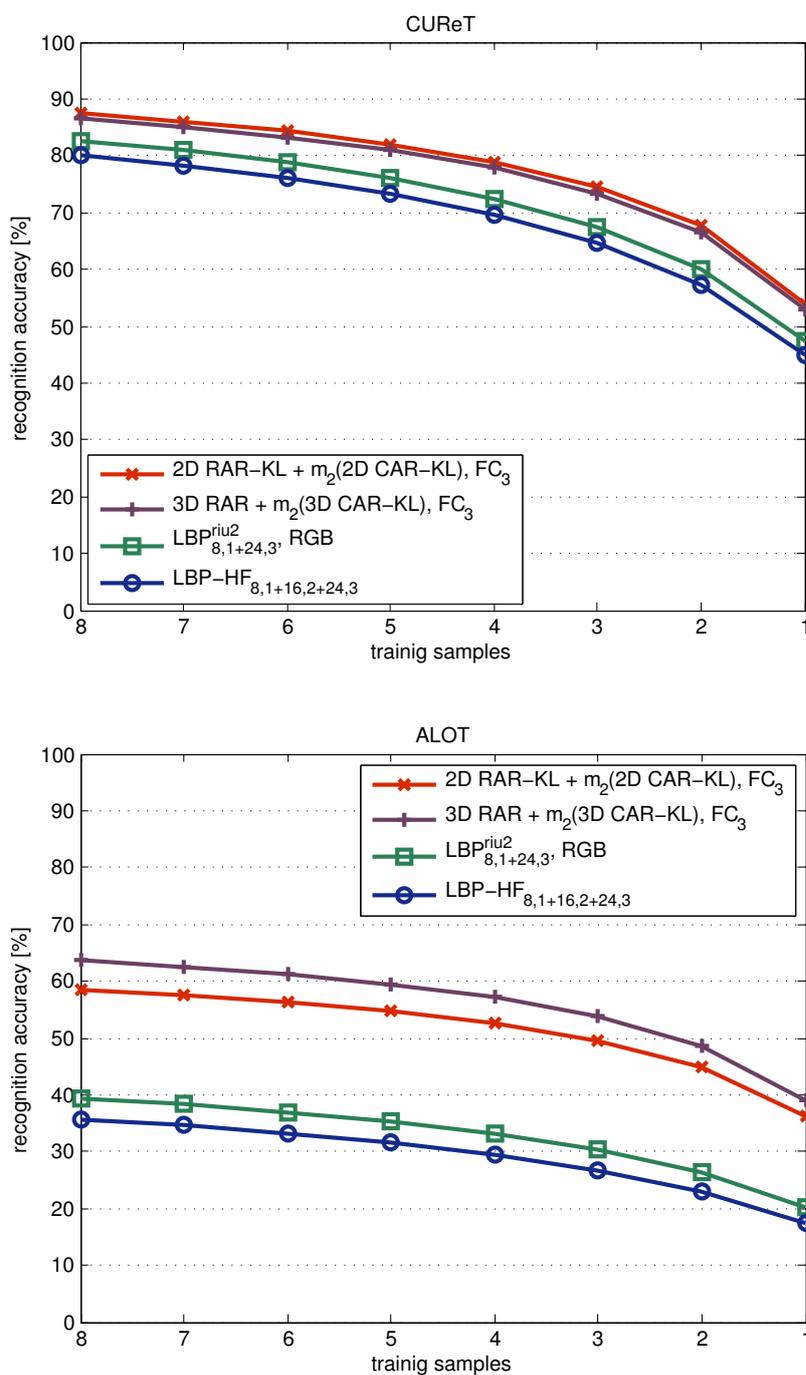
Figure 6.7: Experiment $\varrho 1$: Accuracy of material recognition [%] for CUReT and ALOT datasets, using different numbers of random training images per material. The values were averaged over 10^3 random selections of training images.
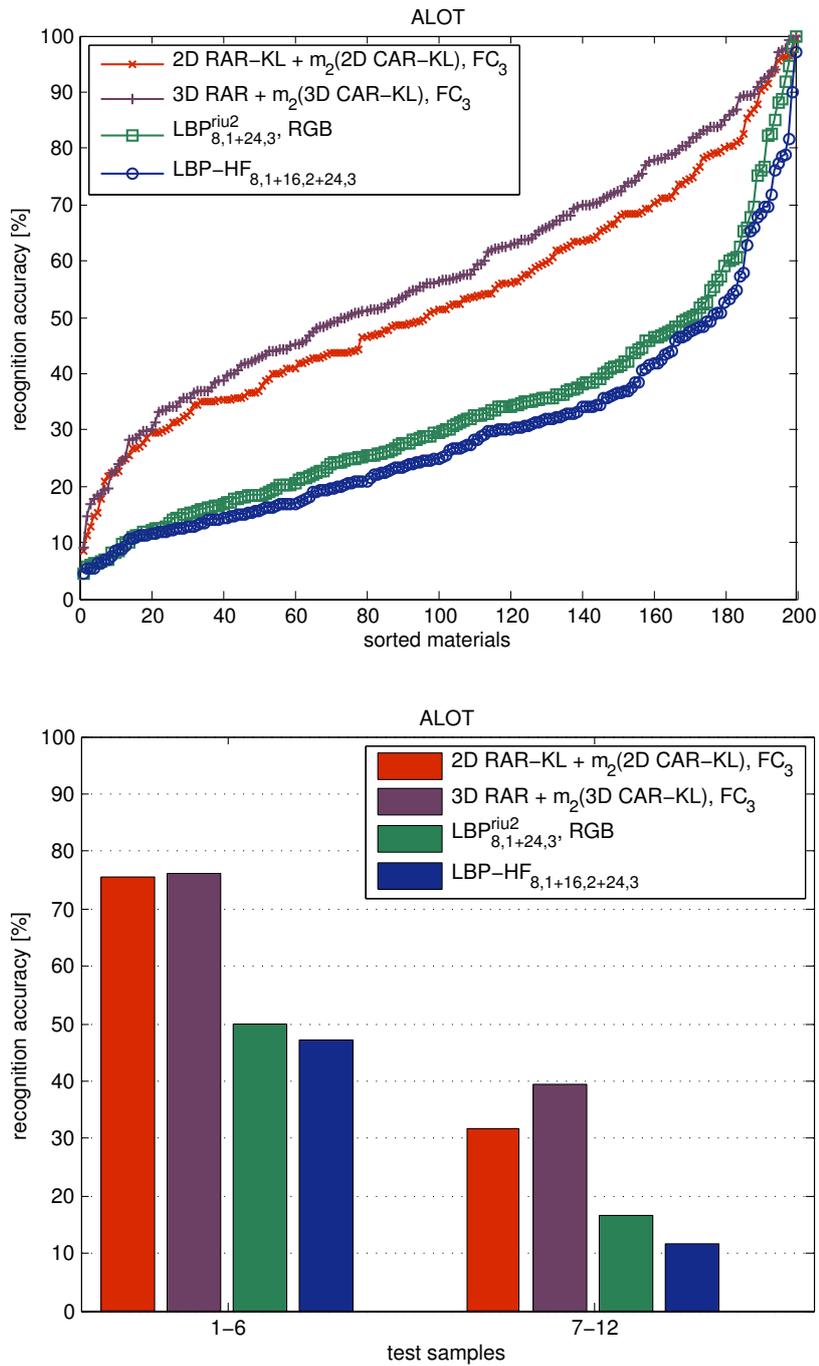
Figure 6.8: Experiment $\varrho 1$: Accuracy of material recognition [%] for the ALOT dataset, using 4 training samples per material. On the top, there is the recognition accuracy per material, where the materials were sorted by their recognition accuracy. In the bottom, the accuracy is grouped by camera position of test samples: top (1-6), from side (7-12).

Results

The results of correct classification and the progression for different number of training samples are displayed in Fig. 6.7. Standard deviations for the CUReT is below 0.7%, 1%, and 1.6% for 8, 4, and 1 samples, respectively and for the ALOT dataset, they are below 0.4%, 0.5% and 0.6% for the same number of samples. The graphs in Fig. 6.7 are directly comparable to the results of Burghouts and Geusebroek (2009b), where the best classification accuracy monotonously decreased from 75% to 45% for MR8-LINC on the CUReT and from 40% to 20% for MR8-NC on the ALOT dataset.

The more detailed comparison is displayed in Tab. 6.10, which includes also the separate results of our two approaches to rotation invariance. The best results were achieved with the combination of these two approaches "3D RAR + m_1(3D CAR-KL), FC_3" on ALOT, and its 2D version on CUReT, both closely folowed by variants with reduced moment set m_2. They performed significantly better than LBP and MR8-* alternatives on both datasets. On the ALOT dataset, the proposed features surpassed the best alternative by more than 20%. This remarkable improvement was probably achieved by the combination of colour invariance and robustness to local intensity changes. The performance difference was maintained for all numbers of training images. Moreover, the 3D model outperformed its 2D counterpart on the ALOT dataset, since large textures provided enough training data for a precise estimation of interspectral relations.

The recognition accuracy per material is displayed in Fig. 6.8, where the materials are sorted according to their recognition accuracy. This graph implies that the ALOT dataset includes some very easily recognisable materials as well as extremely difficult ones. It is worth to note that one half of the ALOT test set is acquired with camera 3, which is closer to the material surface and which viewpoint declination angle is more extreme than declinations of cameras used in the training set. (Example images from camera 3 are in two columns on the right in Figs. 1.2, 6.6). As result, the classification accuracy for these side viewed images is approximately half of the accuracy for the images from top camera positions, or even worse for LBP features as shown in Fig. 6.8. The reason is that none of the compared features are invariant to perspective projective transformation.

Moreover, large texture size in the ALOT database enabled us to experiment with an additional level of the Gaussian pyramid ($K = 5$). This additional level with lower resolution captures larger spatial relations in textures, which is confirmed with a significant performance increase in the ALOT column in Tab. 6.10 – bottom table. The CUReT column in the same table displays that the additional pyramid levels may decrease the performance when the images do not provide enough data.

Finally, the results on the CUReT dataset (Tab. 6.10, Fig. 6.7) are directly comparable with the results of rotation normalisation method displayed in Tab. 6.9 and Fig. 6.5. The results of the rotation invariants are slightly better than the results of rotation normalisation approach. The experiment on ALOT dataset can be also very roughly compared with Experiment i4a (Tab. 6.7, Fig. 6.4), which has similar experiment setup, but excluding one half of images to avoid texture rotations.

method	CUReT	ALOT	size
2D RAR-KL, FC_3	63.2	45.3	180
m_1(2D CAR-KL), FC_3	75.1	38.8	172
m_2(2D CAR-KL), FC_3	76.4	37.1	108
2D RAR-KL + m_1(2D CAR-KL), FC_3	**79.6**	**53.4**	352
2D RAR-KL + m_2(2D CAR-KL), FC_3	79.0	52.6	288
3D RAR, FC_3	61.9	46.8	156
m_1(3D CAR), FC_3	57.4	26.0	148
m_1(3D CAR-KL), FC_3	70.5	41.1	304
m_2(3D CAR-KL), FC_3	72.6	39.2	84
3D RAR + m_1(3D CAR-KL), FC_3	**77.9**	**58.3**	304
3D RAR + m_2(3D CAR-KL), FC_3	**77.9**	57.1	240
$LBP_{8,1+8,3}$, RGB	70.9	32.0	1536
$LBP^{riu2}_{8,1+24,3}$, RGB	**72.4**	**33.2**	108
$LBP^{riu2}_{8,1+24,3}$	66.6	24.3	36
$LBP\text{-}HF_{8,1+24,3}$	69.1	29.9	340
$LBP\text{-}HF_{8,1+16,2+24,3}$	69.6	29.4	448
Burghouts and Geusebroek (2009b):			
MR8-NC	54	**36**	600
MR8-LINC	**67**	30	600

method	CUReT	ALOT	size
↑ 2D RAR-KL + m_1(2D CAR-KL), FC_3	**78.5**	61.6	440
↑ 3D RAR + m_1(3D CAR-KL), FC_3	74.7	**65.3**	380

Table 6.10: Experiment $\varrho1$: Accuracy of material recognition [%] on CUReT and ALOT datasets, using 4 random training images per material. The values were averaged over 10^3 random selections of training images. The bold values highlight the best results in groups and the last column consists of feature vector sizes. The bottom table displays the results with one additional level of Gaussian pyramid ($K = 5$).

The average analysis time for large ALOT images was 20 s for "2D RAR-KL", 11 s for "m_1(2D CAR-KL)", and 10 s for "$LBP^{riu2}_{8,1+24,3}$, RGB" features, all computed on AMD Opteron 2.1 GHz. The analysis of small CUReT images spent 0.8 s, 0.5 s, and 0.4 s of CPU time per image, respectively.

6.3.2 Experiment $\varrho 2$ – OUTEX_TC_00012

In the second experiment, we demonstrate the performance of the proposed rotation invariant features on the Outex database (Ojala et al., 2002a), which consists of material images acquired under three illuminations with different spectra. Some example images are displayed in Fig. 6.1. The experiment was performed on the classification test OUTEX_TC_00012 (Ojala et al., 2002a). The experimental setup slightly resembles Experiment i2 (OUTEX_TC_00014), however, texture rotations were added and only about one fourth of materials was used. The test is freely available for download at (database Outex).

In this setup, 24 materials were selected from the Outex database, subsequently 20 subsamples with size 128×128 were extracted for each of 9 rotations of each material. The training set consisted in 20 subsamples per material, with 0° rotation, illuminated with the "inca" light source. On the other hand, the each of two test sets consisted of 20 subsamples per material, with all 9 rotations. The first test set contained subsamples illuminated with "horizon", while the second one contained images with "tl84" light. Consequently, the train set consisted of 480 images, while the test sets were composed of 4320 images each. Although this setup (Ojala et al., 2002a) was designed with focus on colour invariance, all images are available only in the grey-scale. Therefore, this setup

method	average	size
2D RAR-KL, FC_3	87.5	48
m_1(2D CAR-KL), FC_3	64.6	44
m_2(2D CAR-KL), FC_3	68.1	28
2D RAR-KL + m_1(2D CAR-KL), FC_3	87.6	92
2D RAR-KL + m_2(2D CAR-KL), FC_3	**89.6**	60
$LBP^{riu2}_{8,1+24,3}$	87.6	36
Ojala et al. (2002b):		
$LBP^{riu2}_{8,1+24,3}$	87.2	36
Ahonen et al. (2009):		
$LBP^{riu2}_{8,1+24,3}$	88.3	36
LBP-HF$_{8,1+24,3}$	91.7	340
LBP-HF$_{8,1+16,2+24,3}$	**92.5**	448

Table 6.11: Experiment $\varrho 2$: The results [%] of the Outex classification test OUTEX_TC_00012 averaged over both test sets. The last column consists of feature vector sizes.

disables an exploitation of interspectral dependences, which are the key properties in illumination spectrum invariance.

We used the 1-NN classifier instead of 3-NN used by Ojala et al. (2002b), since the performance differences were negligible. The averages of correct classification on both test sets are displayed in Tab. 6.11. The recently published results of "LBP-HF$_{8,1+16,2+24,3}$" are slightly better than the proposed features, however the feature vector of "LBP-HF$_{8,1+16,2+24,3}$" is almost five times longer. The proposed features suffered from grey-scale input images, which are not, in our opinion, suitable for testing of colour invariance.

6.3.3 Experiment $\varrho3$ – KTH

The third experiment with rotation invariance compares the performance of the proposed features on the KTH-TIPS2 database (Caputo et al., 2005), which includes samples with different scales and rotations. Because the scale and rotation variations are included in the training set, the invariance is not a key issue.

The KTH-TIPS2 database contains 4 samples of 11 materials categories, each sample consists of images with 4 different illuminations, 3 in-plane rotations and 9 scales.

method	average	size
2D RAR-KL, FC_3	58.6	180
m_1(2D CAR-KL), FC_3	59.6	172
m_2(2D CAR-KL), FC_3	59.1	108
2D RAR-KL + m_1(2D CAR-KL), FC_3	**63.2**	352
2D RAR-KL + m_2(2D CAR-KL), FC_3	63.0	288
3D RAR, FC_3	58.8	156
m_1(3D CAR), FC_3	49.6	148
m_1(3D CAR-KL), FC_3	58.7	148
m_2(3D CAR-KL), FC_3	57.8	84
3D RAR + m_1(3D CAR-KL), FC_3	**65.0**	304
3D RAR + m_2(3D CAR-KL), FC_3	**65.0**	240
LBP$_{8,1+8,3}$, RGB	**56.0**	**1536**
LBP$^{riu2}_{8,1+24,3}$, RGB	54.1	108
LBP$^{riu2}_{8,1+24,3}$	49.6	36
Ahonen et al. (2009):		
LBP$^{riu2}_{8,1+24,3}$	50.7	36
LBP-HF$_{8,1+24,3}$	54.2	340
LBP-HF$_{8,1+16,2+24,3}$	**54.6**	448

Table 6.12: Experiment $\varrho3$: Accuracy of material classification [%] on KTH-TIPS2 database averaged over 10^4 random training set selections. The last column consists of feature vector sizes.

The illumination conditions consist in 3 different directions plus 1 image with different spectrum. There are 4572 images in total and their resolution is varying around 200×200 pixels. The KTH-TIPS2 database can be freely downloaded from (database KTH-TIPS2).

We followed the experimental setup of Ahonen et al. (2009), where the 1-NN classifier was trained with one random sample ($4 \times 3 \times 9$ images) per material category. The remaining images (3×108 per category) were used for testing. This was repeated for 10^4 random partitioning to training and test sets. Since the setup did not define any parameter tuning set, we defined it as the subset of training set which contained the first sample of each material category.

The average classification accuracy for different features is compared in Tab. 6.12, where standard deviations were 2% or below. Although, a large variety of training image conditions allowed non-invariant features to perform comparably, still the proposed features took advantage of their invariance and outperformed alternatives by more than 10%.

6.3.4 Discussion

The previous experiments were designed to closely resemble real-life conditions of a material recognition. The tests were performed on 4 different texture databases, which included almost 300 natural and artificial materials and which were acquired with various conditions of viewpoint, illumination colour and direction. A summary of the tested recognition conditions is displayed in Tab. 6.13.

The experiments confirmed that the proposed illumination invariants were successfully integrated with two constructions of rotation invariants: either modelling of rotation invariant statistics (RAR model) or moment invariants computed from direction sensi-

	Experiment			
	$\varrho 1$		$\varrho 2$	$\varrho 3$
texture database	CUReT	ALOT	Outex	KTH-TIPS2
experiment conditions:				
illumination spectrum	−	+	+	+
illumination direction	+	+	−	+
viewpoint azimuth	+	+	−	−
viewpoint declination	+	+	−	−
experiment parameters:				
image size (bigger)	200	1536	128	200
number of materials	61	200	24	11
result tables	6.10	6.10	6.11	6.12

Table 6.13: Parameters of experiments with combined illumination and rotation invariance, including variations of recognition conditions.

tive model parameters (m(CAR) model). As the overall best method we suggest the combination "3D RAR + m_2(3D CAR-KL)" or its 2D counterpart if less training data are available. The proposed features outperformed leading alternative features as MR8-*, LBP^{riu2} and LBP-HF.

In all experiments with rotation invariance, we included the 2D CAR model with $K = 4$ level of Gaussian pyramid and 6-th order hierarchical neighbourhood so that the results were comparable. Naturally, the performance on large textures can be improved by additional pyramid levels as it was demonstrated in Experiment $\varrho1$.

It is worth to note that, from the theoretical point of view, the employed rotational invariants are invariant only to image rotation. However, in our experiments we tested the feature robustness to real rotation of materials including rough ones, whose appearance depends on orientation to the light source and therefore they cannot be modelled as a simple image rotation.

Finally, from the statistical point of view, Experiments i2, i4a, $\varrho1$, $\varrho2$ used the hold-out estimation of classification accuracy. This estimation is based on strictly separated training and test sets and it produces a lower bound on classification accuracy. On the other hand, the methodology in Experiments i3, i4b, $\varrho3$ is somewhere between the hold-out and the leave-one-out estimation, which yields an upper bound. The leave-one-out exploits all but one training samples, while we used only a single or few training samples per material.

An interactive demonstration of the proposed methods and their perfomance on ALOT textures is available online (Vacha and Haindl, 2010d).

Chapter 7

Applications

The proposed textural features were applied in various fields, which range from decoration industry to psychophysical studies and a medical application.

Firstly, we present the content-based tile retrieval system (Vacha and Haindl, 2010c), which was built on the proposed colour invariant textural features, supplemented with colour histograms and LBP features. This computer-aided tile consulting system retrieves tiles from digital tile catalogues, so that the retrieved tiles have as similar pattern and/or colours to the query tile as possible. The system can be exploited in many ways: A user can take a photo of old tile lining and find a suitable replacement of broken tiles from recent production. Or during browsing of digital tile catalogues, the system can offer another tiles that "you may like" based on similar colours or patterns, which could be integrated into an internet tile shop. Or tiles can be clustered according to visual similarity and, consequently, digital catalogues can be browsed through the representatives of visually similar groups (Chen et al., 2005). An user would start with general groups, browse to specific design styles and further to particular tiles. In all previous cases, the system would benefit from its robustness to illumination changes and possible noise degradation. Finally, the performance of the system was verified on a large commercial tile database in a visual psychophysical experiment.

The second application (Haindl et al., 2009) integrated the proposed colour invariants into the unsupervised texture segmentation method by Haindl and Mikeš (2006); Mikeš (2010), which works with multispectral textures and unknown number of classes. The performance of the presented method was tested on the large illumination invariant benchmark from the Prague Segmentation Benchmark (Haindl and Mikeš, 2008) using 21 frequently used segmentation criteria and compared favourably with an alternative segmentation method. Segmentation is the fundamental process of computer vision and its performance critically determines results of many automated image analysis systems. The segmentation applications (Mikeš, 2010) include: remote sensing, defect detection, mammography, and cultural heritage applications. Finally, the segmentation can be employed in extension of previously mentioned tile retrieval system to a general CBIR system.

In the third application (Filip et al., 2010), the proposed textural features were successfully used as statistical descriptors of subtle texture degradations. The features were markedly correlated with the psychophysical measurements and therefore they can be used for automatic detection of subtle texture changes on rendered surfaces in accordance with human vision. Such degradation descriptors are beneficial for compression methods, where the compression parameters have to be set so that the compression is efficient and visual appearance changes remain negligible. The proposed descriptors were targeted to compression of view- and illumination-dependent textures, which depend on massive measured data of BTF and therefore their compression is inevitable. The descriptors allow automatic tuning of compression parameters to a specific material so that subsequent BTF based rendering methods can deliver realistic appearance of materials (Filip and Haindl, 2009; Havran et al., 2010).

Finally, the proposed textural features were applied (Kolář and Vacha, 2009) to analysis of images of retinal nerve fibers (RNF) layer, which texture changes indicate gradual loss of the RNF that it is one of glaucoma symptoms. The early stage detection of RNF losses is desired since the glaucoma is the second most frequent cause of permanent blindness in industrial developed countries. It was shown that the proposed textural features can be used for discrimination between healthy and glaucomatous tissue and therefore they may be used as a part of feature vector in Glaucoma Risk Index, as described in Bock et al. (2007) or in a screening program.

The second, third, and fourth applications were developed jointly with colleagues from Pattern recognition department and DAR research centre.

7.1   Content-based tile retrieval system

Ceramic tile is a decoration material, which is widely used in the construction industry. Tiled lining is relatively long-lived and labour intensive, hence a common problem to face is how to replace damaged tiles long after they are out of production. Obvious alternative to costly and laborious complete wall retiling is finding of the tile replacement from recent production which is as similar to the target tiles as possible. Tiles can differ in size, colours or patterns. We are interested in automatic retrieval of tiles as the alternative to usual slow manual browsing through digital tile catalogues and the subsequent subjective sampling. Manual browsing suffers from tiredness and lack of concentration problems, leading to errors in grading tiles. Additionally, gradual changes and changing shades due to variable light conditions are difficult to detect for humans. The presented computer-aided tile consulting system retrieves tiles from a tile digital database so that the retrieved tiles are maximally visually similar to the query tile. A user can demand either similar patterns, colours or a combination of both. Although this section is concerned with the problem of automatic computer-aided content-based retrieval of ceramic tiles, the modification for defect detection or product quality control is straightforward.

Textures are important clues to specify surface materials as well as design patterns. Without textural description the recognition is limited to different modifications of colour histograms only and it produces unacceptably poor retrieval results. Therefore image retrieval systems (e.g. Chen et al. (2005); Snoek et al. (2008)) employed combination of various textural and colour features. A tile classifier (Ar and Akgul, 2008) used veins, spots, and swirls resulting from the Gabor filtering to classify marble tiles. The verification was done using manual measurement from a group of human experts. The method neglected spectral information and assumed oversimplified normalized and controlled illumination in a scanner. Similar features were used for a detection of tile defects (Monadjemi, 2004).

Unfortunately, the appearance of natural materials is dependent on illumination colour or direction, which variations are inevitable, unless all images are acquired in a strictly controlled environment. One of solutions is a texture representation by means of illumination invariant features. Popular choices are LBP features (Ojala et al., 2002b; Ahonen et al., 2009), which are, however, very noise sensitive. Or illumination invariant extensions (Burghouts and Geusebroek, 2009b) of MR8 texton representation of Varma and Zisserman (2005).

We presented (Vacha and Haindl, 2010c) a tile retrieval system, which takes advantage of a separate representation of colours and texture. The performance of tile retrieval system was evaluated in a visual psychophysical experiment.

7.1.1   Tile analysis

Analysis of tile image is separated into two independent parts: colour analysis and texture analysis. An advantage of this separation is ability to search for tiles with similar colours, texture, or both — according to the user preference.

The colours are represented by histograms, which discard any spatial relations. On the other hand, the texture analysis is based on spatial relation modelling by means of the 2D CAR model, which is followed by computation of colour invariants introduced in Section 4.2. Colour invariants are employed instead of texture analysis of grey-scale images, because colour invariants are able to distinguish among structures with same luminance. Other important properties of the proposed colour invariants are: invariance to brightness changes, robustness to illumination direction variation, and robustness to an image degradation with Gaussian noise. These properties were confirmed by the experiments in Section 6.1, where the proposed colour invariants outperformed alternative textural features such as Gabor features or LBP.

Colour histograms

Colour information is represented by means of cumulative histograms (Stricker and Orengo, 1995), which are computed for each spectral plane separately. The cumulative histogram is defined as the distribution function of the image histogram, the i-th bin H_i is computed as

$$H_i = \sum_{\ell \leq i} h_\ell \ , \tag{7.1}$$

where h_ℓ is the ℓ-th bin of the ordinary histogram. The distance between two cumulative histograms is computed in L_1 metric.

2D CAR textural analysis

The texture analysis is based on the underlying representation with the efficient 2D CAR model (Section 3.1.4). The model parameters are estimated and subsequently transformed into the colour invariants, which characterize the texture (see details in Section 4.2.2).

At the beginning, the spectral planes of a tile image are decorrelated with K-L transformation and Gaussian pyramid with $K = 4$ levels is built. Subsequently each pyramid level is modelled by the set of C 2D CAR models with the sixth order hierarchical neighbourhood I_r, $|I_r| = 14$ neighbours. The following colour invariant features are employed:

$$\operatorname{tr} A_s \ , \ \nu_{s,j} \ , \qquad \alpha_{1,j} \ , \ \alpha_{2,j} \ , \ \alpha_{3,j} \qquad \qquad \forall s \in I_r \ , \ j = 1 \ldots C \ ,$$

where C is the number of image spectral planes. Moreover, the 2D CAR models analyse a texture in some fixed movement direction, therefore additional directions are employed to capture supplementary texture properties. The texture is analysed in three orthogonal directions: row-wise, column-wise top-down, and column-wise bottom-up. Finally, the colour invariants from all pyramid levels and directions are concatenated into a common feature vector.

This textural representation is based on the homogeneity assumption, which is an inherent property of all textures. Unfortunately, some tiles contain insets or other violations of the homogeneity assumption. Therefore the textural features are additionally
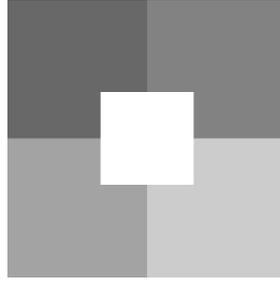
Figure 7.1: Partition of tile image into five regions. The texture is analysed in the whole image and separately in these regions.

computed on each of five tile regions depicted in Fig. 7.1. The dissimilarity of two tiles T, S is combined from dissimilarity of whole images and dissimilarities of corresponding image regions, which utilise fuzzy contrast FC_3 (3.24). Consequently, the overall dissimilarity of tiles $D(T, S)$ is defined as

$$D(T, S) = \text{Norm}\left(\sum_{\ell=1}^{5} FC_3\left(T_\ell, S_\ell\right)\right) + \text{Norm}\left(FC_3\left(T, S\right)\right) , \quad (7.2)$$

$$\text{Norm}(FC_3\left(T, S\right)) = \frac{FC_3\left(T, S\right) - \mu(FC_3)}{\sigma(FC_3)} , \quad (7.3)$$

where T_ℓ, S_ℓ are the ℓ-th regions of images T, S, respectively. *Norm* is dissimilarity normalisation, where $\mu(FC_3)$ and $\sigma(FC_3)$ are mean and standard deviation of FC_3 dissimilarities of all images. In practice, $\mu(FC_3)$ and $\sigma(FC_3)$ could be estimated on a subset of dataset, since the precise estimation is not necessary. This textural tile representation is denoted as "2D CAR-KL 3x" in the results.

7.1.2   Experiment

The textural part of the proposed tile representation was evaluated in a visual psychophysical experiment, where the quality of retrieved images was evaluated by volunteers. The results were compared to alternative textural representation by "LBP$_{8,1+8,3}$" features (see definition in Section 2.2.4).

The experiment was conducted on the dataset of 3301 tile images downloaded from an internet tile shop.[1] All images were resampled to the common size 300×400 pixels, the aspect ratio of rectangular images was maintained and the bigger side was resized to match the size. Thirty-four volunteers (26 males, 8 females) participated in our test. Age of participants ranged from nineteen to sixty, but majority was below forty. About one half of the participants were specialist in the field of image processing. The test was administered over the Internet using a web application so each participant used

---

[1] http://sanita.cz

their own computer in their environment. This setup is plausible, because we focused on significant, first glance differences, which are unlikely to be influenced by test conditions.

The test was composed of subsequent steps, in each step a query image and four test images were displayed. These four test images were composed of two images retrieved by the 2D CAR method and two retrieved by LBP as the most similar to the query image, they were presented in a random order. Participants were instructed to evaluate quality of the retrieved images according to structural/textural similarity with the query image, regardless of colours. There were four ranks available: similar = 3, quite similar = 2, little similar = 1, dissimilar = 0. Subjects were also instructed that they should spend no more than one or two seconds per one test image. Because the presented system has been intended to be a real-life application, we did not provide any examples of similar or dissimilar images, but we let people to judge the similarity according to their own subjective opinion.

The query images were once randomly selected and remained same for all participants in one run. Moreover, the query images were presented in a fixed order, so that the participants were not influenced by different knowledge of previous images. The first three query images were selected manually and were not counted in the results. The reason was to allow subjects to adjust and stabilise their evaluation scale.

The test was performed in two runs, where a single run consisted of the same query and test images evaluated with different subjects. The first run consisted of 66 valid steps evaluated with 23 subjects, while the second one contained 67 valid steps ranked by 11 subjects. The evaluation of one subject was removed due to significant inconsistency with the others (correlation coefficient = 0.4). (The definition of correlation coefficient is in equation (7.7) in the following section.) Average correlation coefficients of subjects' evaluations were 0.64 and 0.73 for the first and the second run, respectively, which implies certain consistency in subjects' similarity judgements.

Results

The experimental results are presented in Tab. 7.1, which shows average ranks and standard deviations of retrieved images. The distribution of given ranks is displayed in Fig. 7.2. It can be seen that the performance of both methods is comparable and successful. About 76% of retrieved images were considered to be similar or quite similar and only 12% were marked as dissimilar. More than two thirds of the participants ranked the retrieved tiles as quite similar or better on average, as can be seen in Fig. 7.3. Different subjects' means in Fig. 7.3 show that the level of perceived similarity is subjective and a personal adaptation would be beneficial. Unfortunately, such an adaptation is not always possible since it requires a user feedback.

As expected, the further analysis of the collected data revealed that LBP and 2D CAR methods prefer different aspects of structural similarity. The LBP method is better with regular images that contain several distinct orientations of edges, while the 2D CAR model excels in modelling of stochastic patterns. Moreover, LBP describes any texture irregularities in contrast to 2D CAR model, which enforces homogeneity and small irregularities are ignored as errors or noise. Both approaches are plausible and it depends on

	2D CAR-KL 3x	LBP$_{8,1+8,3}$
run 1	2.21 ± 0.64	2.22 ± 0.65
run 2	2.23 ± 0.62	2.21 ± 0.57

Table 7.1: Quality of texture retrieval methods as evaluated by subjects. The table contains average ranks (0 = dissimilar – 3 = similar) and corresponding standard deviations.
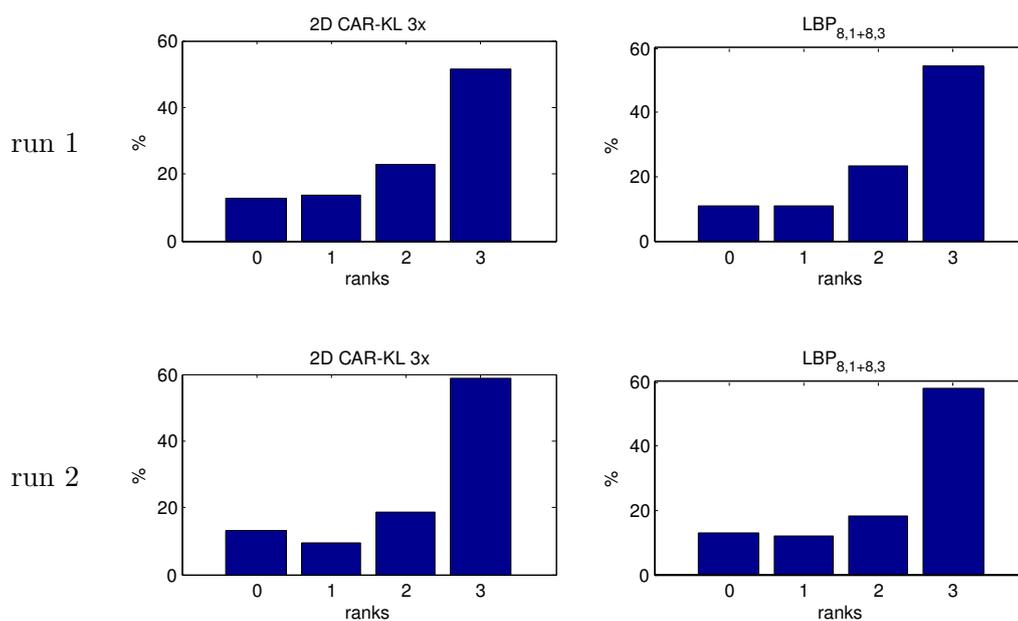
Figure 7.2: Histogram of ranks (0 = dissimilar – 3 = similar) given by participants. The first row shows histograms for the first test run, while the second row for the second run.
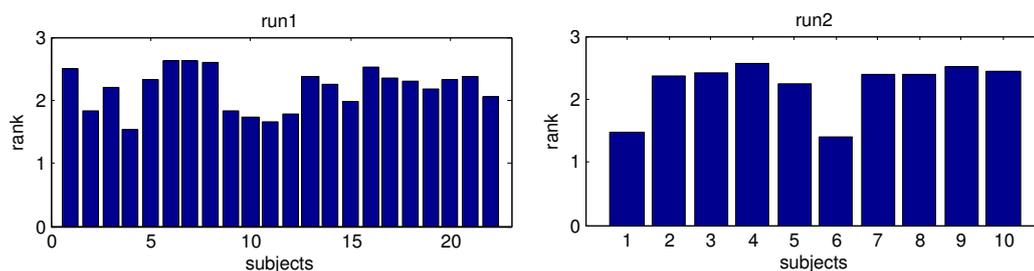
Figure 7.3: Distribution of average ranks given by participants in the first and the second test run.

99

Figure 7.4: Examples of similar tiles retrieved by our system, which is available online at `http://cbir.utia.cas.cz/tiles/`. The query image, on the left, is followed by two images with similar colours and texture ("2D CAR-KL 3x" features). The images are from the internet tile shop `http://sanita.cz` .

a subjective view, which approach should be preferred. Moreover, the 2D CAR features are more robust to changes of illumination direction, which was demonstrated in the experiments in Section 6.1.

Based on the previous evaluation, we decided to benefit from the both tested textural representations and include them into our retrieval system. The final retrieval result is consequently composed of images with colour similarity, texture similarity according to "2D CAR-KL 3x", and texture according to "$LBP_{8,1+8,3}$" features.

7.1.3 Conclusion

In this application, the proposed colour invariant textural features were incorporated into a tile retrieval system, which takes advantage of separate representation of colours and texture. Moreover, the performance of the proposed features was successfully evaluated in a psychophysical experiment, where the relevance of retrieved tiles was rated by more that thirty volunteers. Finally, an interactive demonstration of our tile retrieval system is available online at (Vacha and Haindl, 2010e), some examples of retrieved tiles are shown in Fig. 7.4

The presented retrieval system is not limited to tile images and it can be used with other kinds of images, where the structure is important property, e.g. textiles/cloths and wallpapers.

7.2 Illumination invariant unsupervised segmenter

Segmentation is the fundamental process which affects the overall performance of many automated image analysis systems. Image regions, homogeneous with respect to some usually textural or colour measure, which result from a segmentation algorithm, are analysed in subsequent interpretation steps. Texture-based image segmentation is an area of intense research activity in recent years and many algorithms were published in consequence of all this effort. These methods are usually categorised (Reed and du Buf, 1993) as region-based, boundary-based, or as a hybrid of the two. Different published methods are difficult to compare because of lack of a comprehensive analysis together with accessible experimental data, however available results indicate that the ill-defined texture segmentation problem is still far from being satisfactorily solved. Spatial interaction models and especially MRF-based models are increasingly popular for texture representation: Reed and du Buf (1993); Kashyap and Khotanzad (1986); Haindl (1991), etc. Several researchers dealt with the difficult problem of unsupervised segmentation, where the number of regions is unknown: Panjwani and Healey (1995); Manjunath and Chellapa (1991); Andrey and Tarroux (1998); Haindl (1999); Haindl and Mikeš (2004); Hoang et al. (2005) or Haindl and Mikeš (2006); Mikeš (2010).

Realistic remote sensing, outdoor, security, and many others applications of these segmenters often have to deal with variable illumination of the segmented scene. Therefore we integrated the proposed illumination invariant features into the unsupervised segmenter of Haindl and Mikeš (2006) to a produce illumination invariant texture segmenter (Haindl et al., 2009).

7.2.1 Texture segmentation algorithm

The texture is modelled by the 3D CAR model (see Section 3.1.3), which is adaptively estimated at each pixel position. The proposed colour invariants (see Section 4.2.1), are computed from the model parameters and the segmentation is performed by clustering of these colour invariants.

The 3D CAR model with exponential data forgetting (Haindl and Šimberová, 1992) was used in order that the model was able to adapt to different textures. The model is estimated in four directions of movement (top-down, bottom-up, rightward, leftward) and colour invariants for all directions are used to form parameter space Θ. We use colour invariants:

$$\operatorname{tr} A_s , \ \nu_{s,j} , \qquad \forall s \in I_r , \ j = 1 \ldots C \ ,$$

where I_r is model contextual neighbourhood and C is the number of image spectral planes. Finally, pixel colour is transformed into Lab colour space and a, b components are added into parameter space Θ.

Mixture based segmentation

Multi-spectral texture segmentation is done by the algorithm of Haindl and Mikeš (2006); Mikeš (2010) applied to parameter space based on the colour invariants. The segmen-

tation is performed by clustering in the parameter space Θ defined on the lattice I, Θ_r is parameter vector computed at pixel location $r \in I$. We assume that this parametric space can be represented by means of the Gaussian mixture model (GM) with diagonal covariance matrices. The Gaussian mixture model for parametric representation Θ is as follows:

$$p(\Theta_r) = \sum_{\ell=1}^{L} p_\ell \, p(\Theta_r \,|\, \mu_\ell, \Sigma_\ell) \;, \tag{7.4}$$

$$p(\Theta_r \,|\, \mu_\ell, \Sigma_\ell) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_\ell|^{\frac{1}{2}}} \exp -\frac{(\Theta_r - \mu_\ell)^T \Sigma_\ell^{-1} (\Theta_r - \mu_\ell)}{2} \;, \tag{7.5}$$

where L is the number of components, p_ℓ is component weight and μ_ℓ, Σ_ℓ are mean and standard deviation parameters of the ℓ-th component, $d = |\Theta|$ is the dimension of feature space. The mixture model equations (7.4), (7.5) are solved using a modified Expectation Maximisation (EM) algorithm. The algorithm is initialised using μ_ℓ, Σ_ℓ statistics estimated from the corresponding rectangular subimages obtained by regular division of the input texture mosaic.

After the initialisation, the steps of the EM algorithm are alternating. The components with smaller weights than a fixed threshold are eliminated a merged to the most similar component according to Kullback Leibler divergence. The algorithm stops when either the likelihood function has negligible increase or the maximum iteration number threshold is reached.

The pixels are assigned to the clusters according to the highest component probabilities. A local smoothing is applied so pixel at location r is assigned to the ℓ-th cluster if

$$\pi_{r,\ell} = \max_\ell \sum_{s \in I_r^\square} w_s \, p(\Theta_{r-s} \,|\, \mu_\ell, \Sigma_\ell) \;,$$

where w_s are fixed distance-based weights, $I_r^\square$ is a rectangular neighbourhood. If $\pi_{r,\ell}$ is lower than a threshold, the pixel at location r is unclassified. In the post-processing step, the regions with similar statistics are merged and clusters smaller than a given threshold are attached to their larger neighbours with the highest similarity value. More details on the segmentation algorithm are presented in Haindl et al. (2009) or Mikeš (2010).

7.2.2 Experimental results

The algorithm was tested on mosaics of natural colour textures from the Prague Texture Segmentation Data-Generator and Benchmark (Haindl and Mikeš, 2008). This benchmark consists of randomly generated layouts of mosaic, which are filled with randomly selected colour textures from the large Prague colour texture database. The benchmark also provides ranking of segmentation results according to various segmentation criteria. There are implemented 27 most frequented evaluation criteria (Haindl and Mikeš, 2008)

	Benchmark – colour with illumination invariance			
	AR3D+EM ii	AR3D+EM	HGS E	HGS C
↑ *CS*	**40.70**	34.14	9.55	9.17
↓ *OS*	53.02	53.33	19.30	**12.80**
↓ *US*	16.76	**13.29**	30.05	37.48
↓ *ME*	**13.96**	20.12	39.72	38.41
↓ *NE*	**14.85**	20.57	39.64	35.36
↓ *O*	35.17	**31.53**	56.44	68.87
↓ *C*	91.72	95.34	60.20	**51.63**
↑ *CA*	**59.15**	57.87	40.20	35.81
↑ *CO*	**65.72**	64.76	53.61	50.70
↑ *CC*	86.36	**87.17**	62.45	60.67
↓ *I.*	**34.28**	35.24	46.39	49.30
↓ *II.*	3.83	**3.52**	12.11	16.15
↑ *EA*	**68.26**	68.15	51.44	46.22
↑ *MS*	**56.91**	57.23	34.80	28.32
↓ *RM*	5.89	**4.78**	12.93	16.63
↑ *CI*	71.32	**71.40**	54.22	50.03
↓ *GCE*	**14.34**	16.99	25.36	21.31
↓ *LCE*	**7.62**	8.64	16.69	12.23
↓ *dM*	16.58	**14.64**	29.18	38.39
↓ *dD*	**19.82**	20.27	29.21	29.82
↓ *dVI*	15.80	16.75	13.98	**12.61**

Table 7.2: Comparison of segmentation results according to benchmark criteria (see Tab. 7.3). The proposed "AR3D+EM ii" algorithm is compared with three alternative methods. The arrows indicate whether higher or lower criterion value is better.

CS	= correct segmentation	I.	= type I error
OS	= over-segmentation	II.	= type II error
US	= under-segmentation	EA	= mean class accuracy estimate
ME	= missed error	MS	= mapping score
NE	= noise error	CI	= comparison index
O	= omission error	GCE	= global consistency error
C	= commission error	LCE	= local consistency error
CA	= class accuracy	dM	= Mirkin metric
CO	= recall - correct assignment	dD	= Van Dongen metric
CC	= precision - object accuracy	dVI	= variation of information
RM	= root mean square proportion estimation error		

Table 7.3: List of the most frequented criteria used in evaluation of segmentation results.
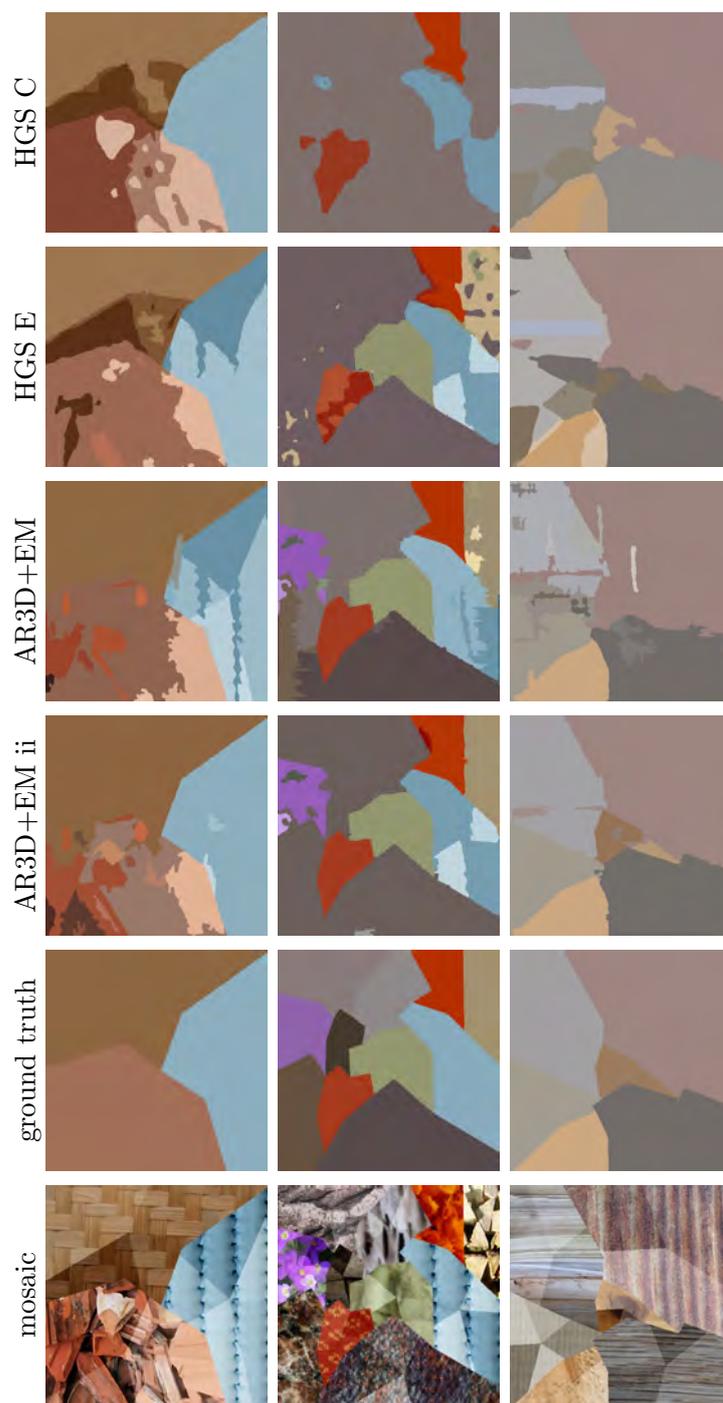
Figure 7.5: Selected experimental texture mosaics from the Prague Texture Segmentation Data-Generator and Benchmark (`http://mosaic.utia.cas.cz`), followed by ground truth and the corresponding segmentation results.

categorised into four groups: region-based (5+5), pixel-wise (12), consistency measures (2) and clustering comparison criteria (3).

The proposed algorithm, denoted as "AR3D+EM ii", is compared with its non-illumination invariant version "AR3D+EM" (Haindl and Mikeš, 2006) and the HGS method (Hoang et al., 2005) either in its fully illumination invariant version "HGS-C" or the non-illumination invariant version "HGS-E". The HGS segmenter combines the k-means clustering with region merging step. It uses a Gabor-Gaussian spatial-colour texture representation and its illumination invariant C version uses features derived from the Gabor filters applied to colour invariants of Geusebroek et al. (2001).

Tab. 7.2 compares the overall benchmark performance, the results of our algorithm demonstrate very good performance in all criteria with the exception of oversegmentation tendency and slightly worse variation of information criterion. The important correct region segmentation criterion is four times better than for the HGS method, undersegmentation is low as well as missed and noise errors. The proposed illumination invariant segmenter outperformed its non–invariant counterpart as expected, however the same conclusion cannot be claimed for the HGS method.

Fig. 7.5 shows three selected 512×512 experimental benchmark mosaics created from three to eleven natural colour textures. The last four columns demonstrate comparative results from two alternative methods, both in illumination invariant and non-invariant versions, respectively. The third column demonstrates robust behaviour of the proposed algorithm but also show infrequent algorithm failures producing oversegmentation for some textures. Such failures can be reduced by a more elaborate postprocessing step. The "HGS-C" (Hoang et al., 2005), "HGS-E" (Hoang et al., 2005) algorithms performed worse on these data as can be seen in the last two columns of Fig. 7.5, some areas are undersegmented while other parts of the mosaics are oversegmented.

The resulting segmentation results are promising even if we could compare only one illumination invariant alternative method. However, the older, non-invariant version "AR3D+EM" was compared with 22 other leading unsupervised segmenters with very good results in the extensive verification (Mikeš, 2010; Haindl and Mikeš). The results can be easily further improved by an appropriate more elaborate postprocessing.

7.2.3 Conclusion

The proposed illumination invariants were successfully integrated into the unsupervised texture or image segmentation algorithm with unknown number of classes. Although the algorithm uses the random field type of model, it is very fast, because it uses efficient recursive parameter estimation of the model and therefore is much faster than the usual MCMC estimation approach. Typically, segmentation methods suffer from many application dependent parameters, which have to be experimentally estimated. The presented algorithm requires only a size of contextual neighbourhood and two additional thresholds. The performance was demonstrated on the extensive benchmark with natural texture mosaics, where the proposed algorithm was favourably compared with the previous illumination variant version and the alternative HGS segmentation algorithm.

In future research, we are going to utilise the segmentation algorithm in a CBIR system, where the natural images will be segmented into regions of homogeneous textures, which will be, subsequently, represented by the proposed colour invariant features.

7.3 Psychophysical evaluation of texture degradation descriptors

Advanced graphics applications such as virtual interior design, cultural heritage digitization, etc. require considerable effort to render the appearance of real-world accurately. When it comes to photo-realistic appearance of materials there is no other way than to use view- and illumination-dependent measurements of real materials. Such measurements can be represented by means of BTF (Dana et al., 1999). Seven-dimensional BTFs represent challenging data due to theirs massive size and thus have high processing and rendering expenses. A number of approaches to BTF compression and modelling have been published in the past as shown in survey Filip and Haindl (2009).

However, the disadvantage of most of the compression methods is that they have fixed parameters regardless of the type of sample being compressed. There have been attempts to use data on visual perception for improvement of texture data compression. Filip et al. (2008) applied a psychophysical study to obtain a perceptually important subset of view- and illumination-dependent images and thus consequently reduced the amount of data to be processed. On the other hand, Guthe et al. (2009) used standard contrast sensitivity in cone response space together with a psychometric difference for improvement of the data compression. Interactions of human gaze fixation with different surface textures have also been analysed (Filip et al., 2009). Although these approaches provide pioneering introductions of perceptual methods for improvement of texture compression, they are not suitable for evaluation of subtle visual compression effects.

The main motivation of this study (Filip et al., 2010) was to find a *computational texture descriptor having responses highly correlated with human vision*. Such a descriptor could be used for comparison of rendered images resulting from original data and data parametrised by compression methods. Based on the responses from the descriptor the methods could iteratively adapt their parameters to automatically achieve an optimal visual performance. We tested a set of descriptors based on either perceptually motivated measures or standard texture features used in texture retrieval and recognition applications. The performance of the descriptors was evaluated by a psychophysical experiment on a group of twelve subjects.

7.3.1 Test data design

To evaluate suitability of the degradation descriptors we designed a set of testing images. Each image contained a rendered cube whose three visible faces were textured with BTF material samples. We used five different material samples depicted in Fig. 7.6, which were measured by Meseth et al. (2003) (each sample comprised 81 illuminations $\times$ 81 view directions, i.e. 6561 texture images with resolution 256×256 pixels).

The cube faces were modified in a way that all three visible faces (top, left, right) contained different geometry. To test a range of shapes that occur in the real-world we

| alu | fabric | leather | wood | wool |
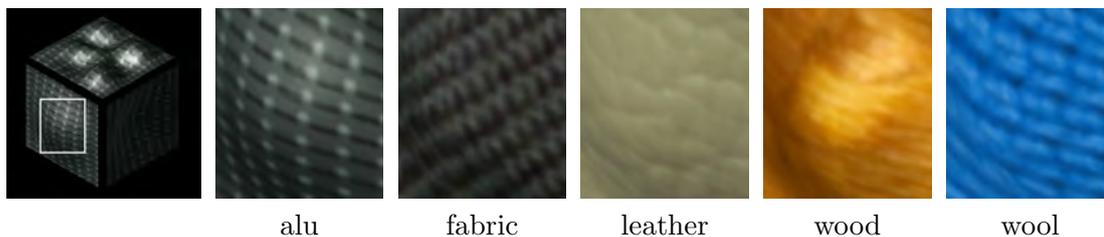
Figure 7.6: Examples of five tested BTF material samples shown on a region of one test image.

| 1-L | 1-R | 2-L | 2-R | 3-L | 4-L | 5-R |
| FIB | FIB | BFI | BFI | IBF | IRB | RBI |

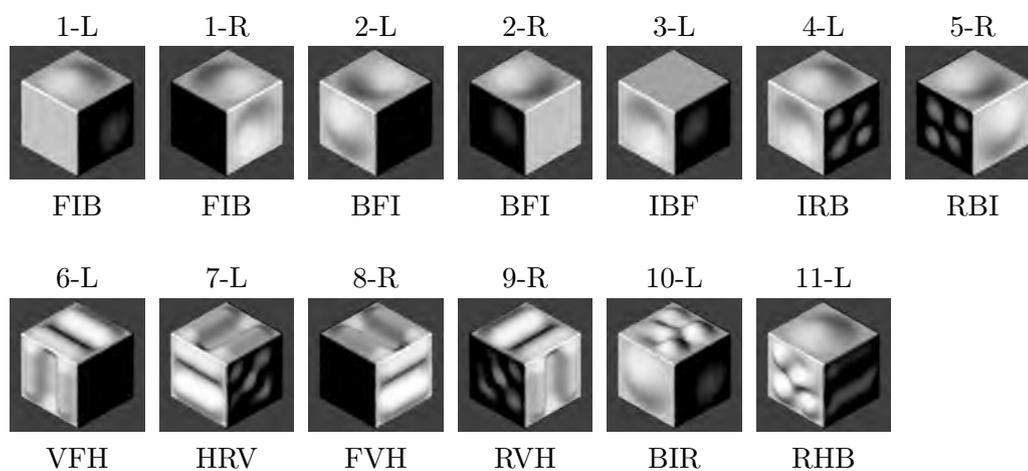| 6-L | 7-L | 8-R | 9-R | 10-L | 11-L |
| VFH | HRV | FVH | RVH | BIR | RHB |

Figure 7.7: Tested combinations of cube face shapes and illumination direction. The configuration of face shapes is displayed below images, while the illumination direction (R, L) is above images.
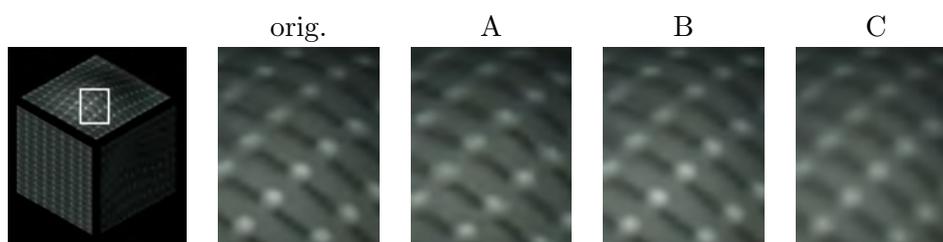

| orig. | A | B | C |

Figure 7.8: Degradation of material sample *alu* for different filters (A, B, C).

used different shapes for each cube face:

I	–	wide indent,	**R**	–	random bumps,
B	–	wide bump,	**F**	–	flat face,
H	–	horizontal waves,	**V**	–	vertical waves,

For illumination we used directional light from two positions:

L	–	left side,	**R**	–	right side.

both positions were above the cube and they were chosen to guarantee similar distribution of light across the top and left/right faces in a single cube. Not all combinations of test cube faces were used in the experiment as this would result in too high a number of test images. We used only eleven different configurations selected in a way to allow us to compare the most interesting combinations of face geometry. Additionally, not all these configurations were illuminated from both directions as shown in Fig. 7.7.

To simulate possible effects of texture compression we used three filters introducing artificial degradation to the original data, they were:

A	–	illumination/view directions downsampling to 50%
B	–	spatial filtering (averaging by kernel 3×3)
C	–	spatial filtering (averaging by kernel 5×5).

The proposed filters introduced only very subtle differences (see Fig. 7.8) between the original and the modified data and forced subjects to perform extensive visual search, which allowed us to collect detailed gaze data.

Finally, for 13 combinations of cube face shapes & illuminations and 5 material samples, we obtained 65 test images for each degradation. These images were used for testing of texture descriptors described in the following section and also to generate stimuli in the validation experiment in Section 7.3.4.

7.3.2   Texture degradation descriptors

Degradation descriptors were used for comparison of degraded images with their original counterparts. We always compared images with the same material, cube face shapes, and illumination direction, the only differences were faint degradation artefacts. Therefore, we do not require the descriptors to be view or illumination invariant. Since some of state of the art textural features are illumination invariant, we included them as well. The tested descriptors can be principally divided into those which are translation invariant and those which are not.

Translation variant measures

These descriptors are based on perceptually motivated measures of image quality assessment measures computed in pixel-wise manner in a local neighbourhood.

The first is ***Visual Difference Predictor*** (VDP) (Mantiuk et al., 2004), which simulates low level human perception for known viewing conditions (in our case: display

size 37×30 cm, resolution 1280×1024 pixels, observer's distance 0.7 m) and thus it is sufficient for our task of perceptually plausible detection of subtle texture degradation artefacts. The VDP provides percentage of pixels that differ with probability $p > 75\%$ or $p > 95\%$ from all pixels in the compared images. To ensure consistency with other descriptors, we set the VDP output to $(1 - p)$, i.e. giving interval (0,1), where for an output 1 the images are the same.

The **Structure Similarity Index Metric** (SSIM) (Wang et al., 2004) is an empirical measure, which compares in ability to VDP. SSIM measures the local structure similarity in a local neighbourhood of an image window, we used 11×11 pixels. The basic idea of SSIM is to separate the task of similarity measurement into comparisons of luminance, contrast, and structure. These independent components are then combined into one similarity function

$$SSIM(T,S) = \frac{(2\mu(T)\mu(S) + K_1)(2\sigma(T,S) + K_2)}{(\mu(T)^2 + \mu(S)^2 + K_1)(\sigma(T)^2 + \sigma(S)^2 + K_2)} \quad , \tag{7.6}$$

whose formulation should be qualitatively consistent with human perception of difference. In equation (7.6), $\mu(T), \mu(S)$, $\sigma(T), \sigma(S)$ and $\sigma(T,S)$ are mean values, standard deviations, and mutual variance of values in the local neighbourhood of compared images T and S. K_1, K_2 are specific non-zero constants. The valid range of SSIM for a single pixel is $[-1, 1]$, with higher values indicating higher similarity. When the local neighbourhood is evaluated for each pixel we obtain the SSIM difference of two images as a mean value of SSIM values across all pixels.

Translation invariant features

The **Gabor features** are statistics of Gabor filter responses, which can be considered as orientation and scale tunable edge and line detectors. An extension to colour textures is *opponent Gabor features*, which analyse also relations between spectral channels. The details on both Gabor features are described in Section 2.2.2. Since our implementation involves FFT, the both Gabor features were computed only in the square cuts of each cube face.

The **Local Binary Patterns** (LBP$_{P,R}$), which are histograms of texture micro patterns, are described in Section 2.2.4 and they are illumination invariant. We tested features "LBP$_{8,1+8,3}$", which were computed either on grey-scale images or on each spectral plane separately (denoted with "RGB" suffix) and concatenated to form the feature vector.

The proposed **CAR textural features** factorise the image into K pyramid levels and model each level separately by the CAR model. The CAR model parameters are estimated and transformed into colour invariants, which characterise the texture (see details in Section 4.2). The CAR models were estimated with the sixth order hierarchical neighbourhood I_r, which is depicted in Fig. 3.2. The following colour invariant features were employed:

$$\operatorname{tr} A_s \, , \ \nu_{s,j} \, , \qquad \alpha_1 \, , \ \alpha_2 \, , \ \alpha_3 \qquad\qquad \forall s \in I_r \, , \ j = 1 \dots C \, ,$$
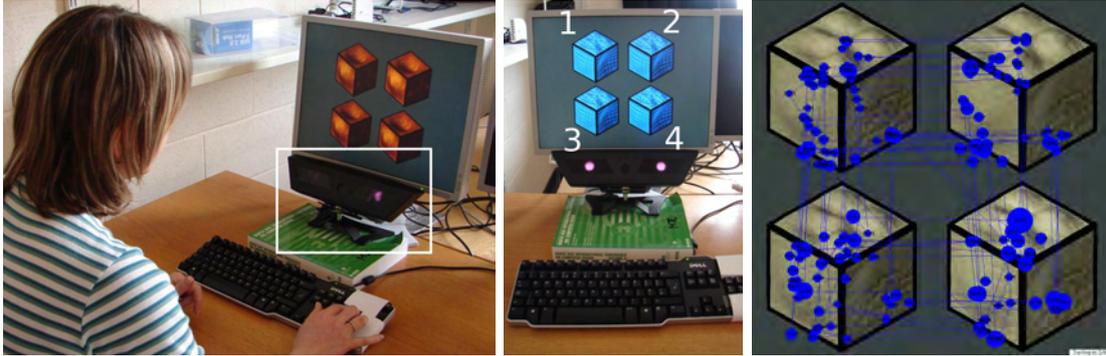
Figure 7.9: From left: setup of the psychophysical experiment with the eye-tracker high-lighted, presentation of stimulus image from subject's view, and a typical gaze fixation pattern.

where C is number of image spectral planes. In case of 2D CAR models, the invariants α_1, α_2, α_3 were estimated for each spectral plane separately and ν_s were computed as diagonals of A_s instead of eigenvalues so the spectral planes were not reordered. From the same reason, the K-L transformation was not used as well.

Alternatively, to image factorisation by the Gaussian pyramid (GP), we experimented with the Gaussian-Laplacian pyramid (GLP), which is computed from GP in the following way. Each pyramid level of GP is up-sampled and the difference to its adjacent upper level is computed. The top pyramid levels in GLP and GP are the same. Moreover, because the CAR models analyse a texture in some fixed movement direction, additional directions were employed to capture supplementary texture properties: row-wise, column-wise top-down and column-wise bottom-up. This representation is denoted in the results as "3D CAR 3x" or "2D CAR 3x" depending on the used model.

Finally, three dissimilarities of the feature vectors were compared: L_1, $L_{0.2}$ norms, and fuzzy contrast FC_3 (see Section 3.2). Although the CAR models theoretically assume texture homogeneity, they can be still used as statistical descriptors of textured surfaces, and so we expect their ability to detect the degradation artefacts.

It is important to note that all the previous textural features are not invariant to texture deformation, which is caused by different shapes. Therefore, the features are always compared between the same surface shapes only.

7.3.3 Psychophysical experiment

We performed a visual search experiment in order to investigate subjects' ability to identify introduced visual degradations. We also recorded their gaze fixations in order to analyse relations between their decisions and their fixations statistics.

For experimental stimuli we used static images of size 1000×1000 pixels, displaying four cubes, described in Section 7.3.1, in individual quadrants (see Fig. 7.9-middle).

We used this layout of stimuli to avoid the central bias in fixations reported in Tatler (2007), i.e. observers have a tendency to fixate the central area of the screen. In each quadruple, three cubes showed the original data rendering and the remaining one showed rendering with degraded texture. The positions of the cubes were random. Examples of stimuli are shown in Fig. 7.9. The edges of the cubes were set to black to mask potentially salient texture seams. The background and the remaining space on the screen were set to dark grey. Fig. 7.7 shows the 13 configurations of cube and illumination direction that were used. Together with five BTF texture samples, and three degradation filters, the total number of stimuli was 195 (13×5×3).

Twelve paid observers (three females, nine males) participated in the experiment. All were students or university employees, were less than 35 years of age, and had normal or corrected to normal vision. All were naive with respect to the purpose and design of the experiment.

The participants were shown the 195 stimuli in a random order and asked to identify which of the cubes had a surface texture slightly different from the remaining three cubes. A stimulus was shown until one of four response keys, identifying the different cube, was pressed. There was a pause of one second between stimulus presentations, and participants took on average around 90 minutes to perform the whole experiment, which was split into four sessions. All stimuli were presented on a calibrated 20.1" NEC2090UXi LCD display (60 Hz, resolution 1600×1200, color temperature 6500 K, gamma 2.2, luminance 120 cd/m^2). The experiment was performed in a dark room. Participants viewed the screen at a distance of 0.7 m so each sphere in a pair subtended approximately 9° of visual angle. Subjects' gaze data was recorded during the experiment using a Tobii x50 infrared-based binocular eye-tracking device as shown in Fig. 7.9. The device was calibrated for each subject individually and provided the locations and durations of fixations at a rate of 50 samples per second. The shortest fixation duration to be recorded was set to 100 ms.

Results

On average, the subjects were able to find the modified cube in 67% of the stimuli, which was surprisingly high in relation to the random chance level 25%, given the subtle changes introduced by filters used (see Fig. 7.8). Informal interviews after the experiment revealed that the subjects were certain in less than 50% of stimuli and for the rest they believed that they were only guessing the right answer. The obtained rates suggest that in the difficult cases they often successfully relied on low level visual perception. The accuracy of responses for individual filters is shown in Fig. 7.10-a and reveals that modifications introduced by the filter **A** are the hardest to spot while the smoothing by filter **C** is the most apparent. This was expected, since smoothing effect is uniform and generally more apparent that the slight illumination and view direction dependent changes in reflectance caused by reduction of directions (filter **A**). While success rates across textures were quite similar for smoothing filters **B** and **C**, their values for filter **A** varied much more.
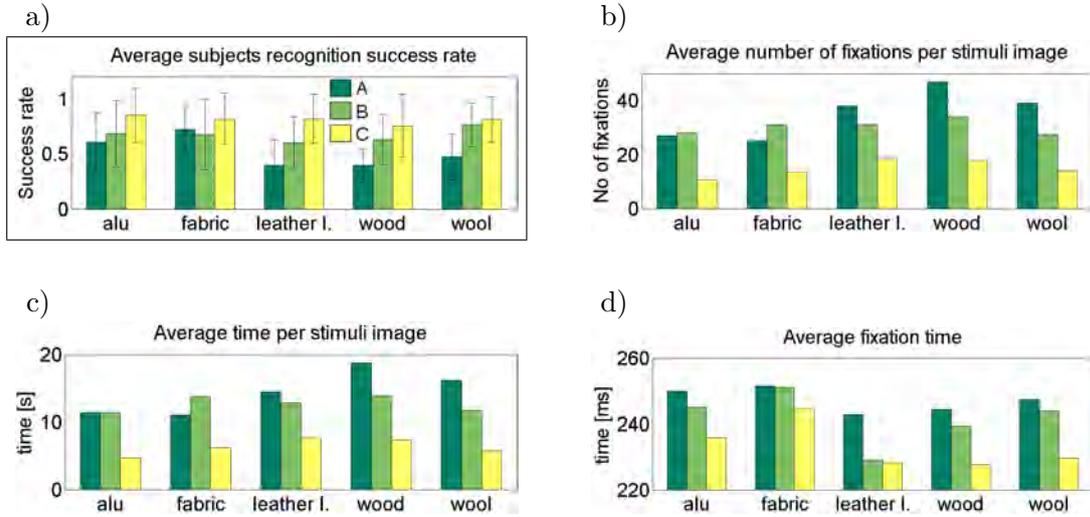
a)

b)

c)

d)

Figure 7.10: Subjects' average (a) recognition success rate, (b) number of fixations per stimuli, (c) time spend on stimuli, (d) fixation duration, all displayed for individual degradations (A, B, C) and tested materials. Error-bars represent twice the standard error across subjects, different cube face shapes and illuminations.

Twelve subjects performed 62 916 fixations longer than 100 ms. Average fixation duration was 242 ms. Each stimulus was on average fixated for 11 s by means of 26 fixations. Figs. 7.10-b,c,d show subjects' gaze fixation statistics as (b) average number of fixations per stimuli, (c) average time spent fixating stimuli, and (d) average fixation time. The (b) and (c) statistics are highly inversely correlated with subjects' response accuracies Fig. 7.10-a, with correlation coefficients $R_{(b)} = -0.904$ and $R_{(c)} = -0.930$, respectively. The figures also reveal apparent differences between the tested material samples. For materials *leather* and *wood*, the subjects were less successful in identification of the cube with degraded texture; they fixated the stimuli longer, and made significantly more fixations, which were shorter than those on the other materials. We suspect that a lower local texture contrast in these materials makes detection of degradation artefacts more difficult.

7.3.4 Perceptual evaluation and discussion

In this section we evaluate performance of the proposed descriptors by comparison with subjects' responses obtained from the psychophysical experiment. The evaluation was based on computation of correlation coefficient

$$R(U,V) = \frac{\mathbb{E}[(U - \mu(U))(V - \mu(V))]}{\sigma(U)\sigma(V)} \ , \tag{7.7}$$

where U, V are compared data vectors, i.e. subjects responses and descriptor responses, and μ and σ are their means and variances. The descriptors were employed to compute
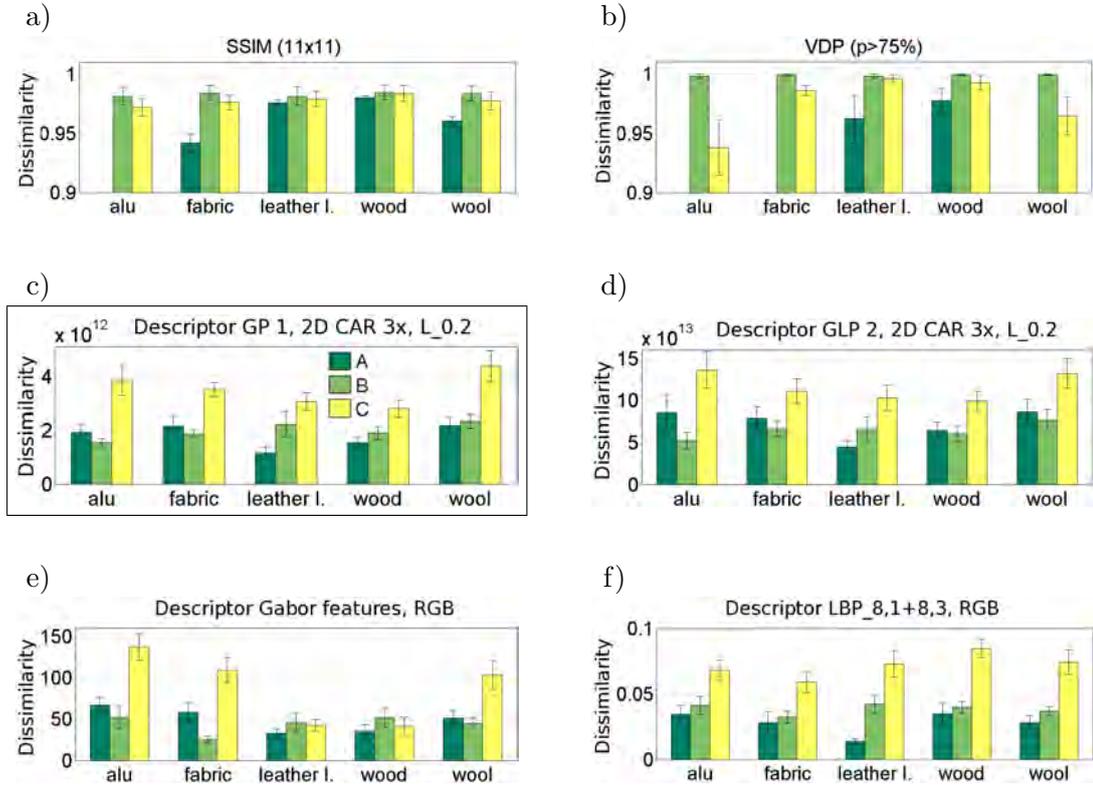
Figure 7.11: Dissimilarity of the degraded images (A, B, C) with the original image according to descriptors. The best correlation with human perceived dissimilarity (Fig. 7.10-a) was achieved with 2D CAR features, without pyramid, using $L_{0.2}$ distance (framed). Each features rely on different dissimilarity measure, therefore the values in graphs are not directly comparable.

dissimilarity of original images and images obtained for each degradation method; the results were averaged across different cube orientations and illumination directions. Finally, we computed correlation of dissimilarity according to the descriptors with subjects' recognition success rate (Fig. 7.10-a), which indicates visibility of differences.

The overall results are shown in Tab. 7.4. We observe low performance of SSIM and VDP descriptors, which can be caused by 1) their translation non-invariance so they give high responses even to a slight, perceptually insignificant, planar shift of texture caused by the filter **A** (this is most apparent for sample *alu* in Fig. 7.8), and 2) their lower sensitivity to the very subtle degradations that were tested. We also observe that although the 3D CAR has a slightly shorter feature vector than its 2D variant, it did not achieve the same performance. It was probably caused by computation of eigenvalue features in the 3D CAR model, this computation may mix the correspondence of spectral planes between features (see details in Section 4.2.1). The table also shows the

method	correlation	size
SSIM, 11×11	**0.125**	
VDP, p>75%	**0.107**	
VDP, p>95%	0.097	
LBP$_{8,1+8,3}$	0.610	512
LBP$_{8,1+8,3}$, RGB	**0.712**	1536
Gabor features	0.569	48
Gabor features, RGB	**0.578**	144
Opponent Gabor features	0.322	252

method	correlation			size
	L_1	$L_{0.2}$	FC_3	
GP 1, 2D CAR 3x	0.677	**0.787**	0.777	195
GP 2, 2D CAR 3x	0.644	0.752	0.710	390
GP 1, 3D CAR 3x	0.581	0.542	0.550	177
GP 2, 3D CAR 3x	0.573	0.552	0.517	354
GLP 2, 2D CAR 3x	0.638	**0.714**	0.654	390
GLP 3, 2D CAR 3x	0.620	0.677	0.648	585
GLP 2, 3D CAR 3x	0.573	0.362	0.360	354
GLP 3, 3D CAR 3x	0.475	0.439	0.422	531

Table 7.4: Correlation coefficient (R) of the tested degradation descriptors with data obtained from the psychophysical experiment. The best variant of each descriptor type is highlighted. The size of feature based descriptors is displayed in the last column.

CAR model comparison with different feature vector distances. While for 2D CAR the best performance was achieved with $L_{0.2}$ norm, for 3D CAR the best results were with L_1 norm. A high correlation with the psychophysical results was achieved by descriptors based on 2D CAR model and LBP features.

Fig. 7.11 shows performance of the best combination of parameters for each type of descriptor, which are highlighted in Tab. 7.4. Generally, the best results were obtained for 2D CAR model without any pyramid (GP 1), where the difference of the feature vectors was evaluated using $L_{0.2}$ norm. Additionally, the CAR features enable to adjust pyramid type and size with regards to the type and intensity of degradation. The LBP features (Fig. 7.11-f) are fast to compute and also had quite high correlation with the human judgements, however, they were little worse in representation of the trend of values across the samples present in Fig. 7.10-a. We tested also other variants of LBP features such as "LBP$_{24,3+8,1}^{riu2}$" and "LBP$_{16,2}^{u2}$", however, their descriptive abilities were clearly worse than of those shown in Tab. 7.4.

Both, best perception correlated, 2D CAR and LPB features are illumination invariant, therefore additional measures of brightness and contrast should be used if the degradation might cause such changes.

7.3.5 Conclusion

The results showed that the dissimilarity measure based on the proposed 2D CAR colour invariant textural features is related with human perception of subtle texture degradation. The human judgements of texture degradation were obtained from a psychophysical study with rendered surfaces with view- and illumination-dependent textures. The proposed features were best, out of tested descriptors, for the automatic prediction of human perceived similarity of these subtle texture differences. This highly demanded property can be used as automatic feedback for optimisation of the visual performance of texture compression and rendering methods.

7.4 Texture analysis of the retinal nerve fiber layer in fundus images

Glaucoma is the second most frequent cause of permanent blindness in industrial developed countries. During glaucoma progression, the degeneration of retinal ganglia cells, axons and gliocells proceeds and causes their damage. If not diagnosed in early stage, the damage of the optical nerve and other structures becomes permanent, which in the final stage may lead to blindness. One of the glaucoma symptoms is the gradual loss of the RNF, which has been proved of high diagnostic value. The RNF atrophy is indicated as texture changes in colour or grey-scale retinal photographs. Therefore, there has been a high effort to use these retinal images to evaluate the RNF since 1980 (Peli et al., 1989). However until now, there is no routinely used method for RNF quantification (based only on colour photography), although an increasing effort in this field is noticeable (Lee et al., 2004; Hayashi et al., 2007).

Although, the expansion of new technologies in diagnosis process is increasing, the diagnosis based only on colour fundus photographs is important, because digital colour fundus camera became a standard tool for fundus examination (Tuulonen et al., 2000; Kolář and Jan, 2008) and the resolution of acquired image is sufficient for detection of fine structures created by retinal nerve fibres.

In this section, we present application (Kolář and Vacha, 2009) of the proposed textural features based on 2D CAR model, without computation of illumination invariants. The textural features were applied to detection of early stage glaucoma, which was recognised via texture changes in RNF.

7.4.1 Data

The database contained 16 colour images of glaucomatous eyes with focal RNF loss and 14 colour images of healthy eyes. They were acquired by fundus camera (Canon CF-60UDi with digital camera Canon D20) in JPEG format with very low compression. Green and blue channels were averaged to create intensity images, since the red component does not carry any information from the RNF layer. The information is contained only in the corresponding green-blue wavelengths reflected from eye. The size of the images was 3504×2336 pixels with a large field of view (60°).

One image from that database with RNF loss is shown in Fig. 7.12, depicting the main structures and also two areas with RNF losses. Optical nerve head (ONH) is a place where blood vessels and RNF enters or leaves the inner eye. The macula is the place with the highest concentration of retinal ganglia cells. The RNF runs mainly from the ONH to macula with the highest concentration in radial direction. The RNF losses appear as a darker area, which is caused by decreasing number of the nerve fiber bundles and lowering the reflection of incident light. As the reflection depends on the optical properties of the examined eye, the brightness is not a reliable feature for RNF description. The RNF are represented as a striation pattern, which creates texture – the neural fibers are locally oriented in parallel, which causes their lightly stripy appearance.
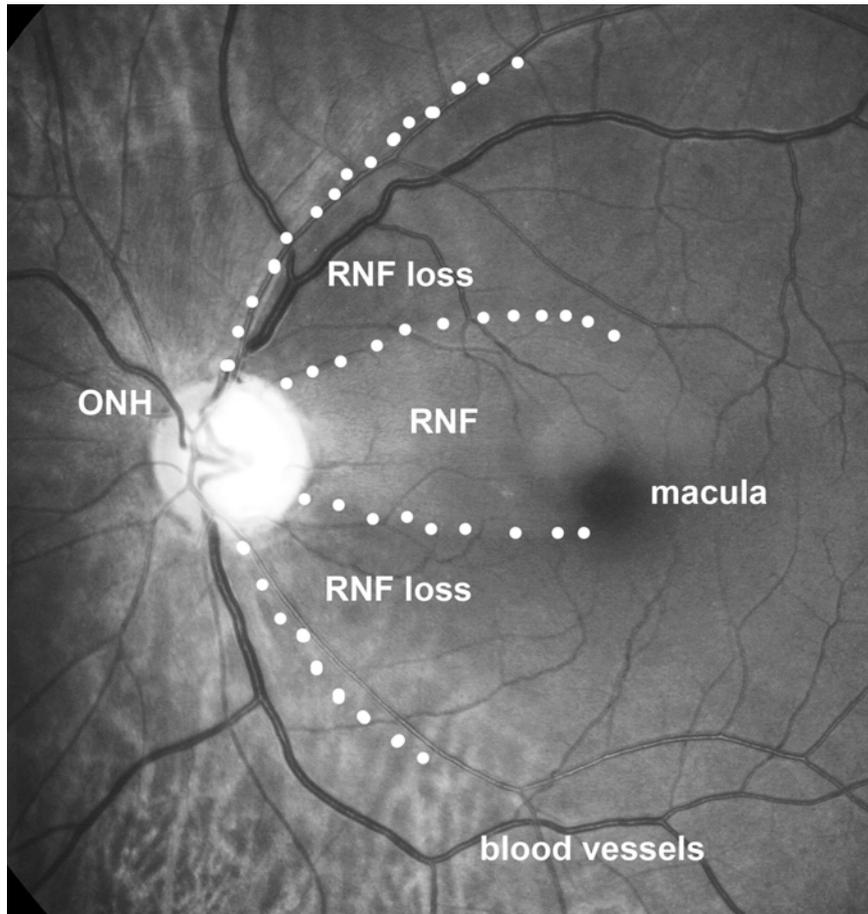
118

Figure 7.12: One image from our database showing macula, optic nerve head (ONH), blood vessels and areas with/without retinal nerve fibers (RNF).

The testing of the textural features was done on the small square samples (41×41 pixels), which were cut from the retinal images. The image samples were divided into the following classes:

A – *area containing RNF* from patients with glaucoma (304 samples)
B – *area without RNF* from patients with glaucoma (176 samples)
C – *control group* – patients without glaucoma (227 samples)

The size of these samples was selected in order to span a sufficiently large region with RNF striation. The maximum size was limited by the blood vessels and other anatomical and pathological structures in the retinal image. Their positions were selected in close surrounding around ONH, not exceeding double radius of the ONH from ONH border. Pixel values in all samples were normalized into range 1 to 64 in order to eliminate different illumination conditions caused by the different optical eye properties.

7.4.2 Method

Texture representation

The texture is characterised by 2D CAR textural features, described in Section 3.1.4, without computation of illumination invariants. The proposed textural features are estimated on a gradient image, which is estimated by means of differences:

$$\nabla G(r_1, r_2) = [G(r_1 + 1, r_2) - G(r_1 - 1, r_2), G(r_1, r_2 + 1) - G(r_1, r_2 - 1)] \ ,$$

where G is grey-scale image and $r = (r_1, r_2)$ is a pixel position composed of row and column index. Subsequently, the Gaussian pyramid is built and each level is modelled by the 2D CAR model, with the third order hierarchical contextual neighbourhood I_r, depicted in Fig. 3.2. The textural features are composed of the model parameters:

$$A_s, \quad \forall s \in I_r \,, \qquad\qquad \hat{\Sigma} \ ,$$

defined in Section 3.1.4. The texture is analysed in three orthogonal directions: row-wise, column-wise top-down and column-wise bottom-up. Finally, the features from all model directions and pyramid levels are concatenated into a common feature vector f.

The sample images were analysed with only $K = 2$ levels of pyramid and the third order neighbourhood, because small resolution of image samples did not allow analysis of larger texture relations. It is worth to recall that the features A_s are invariant to brightness changes (see formula 4.30), which is advantageous since optical eye properties may differ across humans.

Feature selection and classification

The feature selection was performed in the sense of Maximum Relevance and Minimum Redundancy (MRMR) approach introduced by Peng et al. (2005). This scheme is based on two concepts: a good feature should have maximum relevance to target class and minimum redundancy to already selected features. These two properties can be described by mutual information.

The features selection works in an iterative manner. At first, the feature with the highest mutual information is selected. The rest features are selected incrementally, e.g. previously selected features stay in a new subset and the new feature is selected in such way, that it maximizes MRMR criterion (Ding and Peng, 2003). This approach was applied to all combination of classes (A – C, B – C, A – B), providing different features for each pair. The best features selected for each pair of classes were subsequently tested with a classifier to evaluate the classification accuracy. Two classifiers were used: the nonlinear SVM classifier and linear Ho-Kashyap classifier.

The SVM (Cortes and Vapnik, 1995; Cristianini and Taylor, 2000) is a classifier that maximizes the geometrical margin between the considered classes. The ν-SVM variant, in implementation by Chang and Lin (2001), was used with the radial basis non-linear transformation (variance parameter equal to 0.5). The classifier was tested in the range

of the penalization parameter ν and it was observed that for ν changing from 0.2 to 0.7 the classification error was almost constant. Value $\nu = 0.5$ was used for the next tests.

Ho-Kashyap classifier (Duda et al., 2001) combines perceptron and Least Mean Squares (LMS) classification. This classifier ensures that either separating hyperplane is computed (if available) or LMS optimal solution is found from the training data.

The random k-fold cross-validation method was utilized to test the performance of these two classifiers. This procedure took randomly selected k-folds for training (200 samples) and different k-folds for testing (100 samples). The training and testing samples were run 100 times for different sets to evaluate the classification error.

7.4.3 Results

The features from the feature vector f were sorted according to the criteria defined by MRMR approach. The five most relevant features for each pair of classes (B – C, A – C, and A – B) are displayed in Tab. 7.5. The features f_{19} and f_7 appeared in all cases at the top of the ordered sequences. These parameters correspond to relative positions $s = [-1, 0]$ and $s = [-2, 0]$ in model neighbourhood I_r (see Fig. 3.2).

	class B – C	class A – C	class A – B
Indices of features in feature vector f	3, 19, 7, 37, 31, 22, 25	19, 7, 5, 61, 37, 31, 10	45, 9, 11, 24, 19, 7, 58

Table 7.5: The seven best textural features based on 2D CAR model as ordered by MRMR approach.

The classifiers were used successively for these best features, starting with the most relevant feature, then adding the second most relevant feature, etc. The results of classification are presented in Tab. 7.6 for different combinations of classes and two tested classifiers. The best results for each pair of classes are boldfaced marked.

It can be seen that Ho-Kashyap outperformed SVM classifier in all cases on our dataset. Moreover, the classification error decreased with the number of features used for classification as expected. The best results were achieved with number of features between 5 and 7, depending on the classified classes. Using more features increased classification error (for Ho-Kashyap classifier) or caused error fluctuation with changing standard deviation.

The best selected features can be also used for further improvement of already efficient estimation of the 2D CAR model. The contextual neighbourhood I_r can be restricted to positions of the selected features so the neighbourhood size would decrease as well as the computation time.

Finally, Fig. 7.13 presents the feature space for features f_{19} and f_7 as an example of appropriate features for all combination of classes. Class B and C created a well separated clusters, which means that we can differentiate between the regions with RNF layer losses and healthy tissue (which was given by healthy population). The cluster
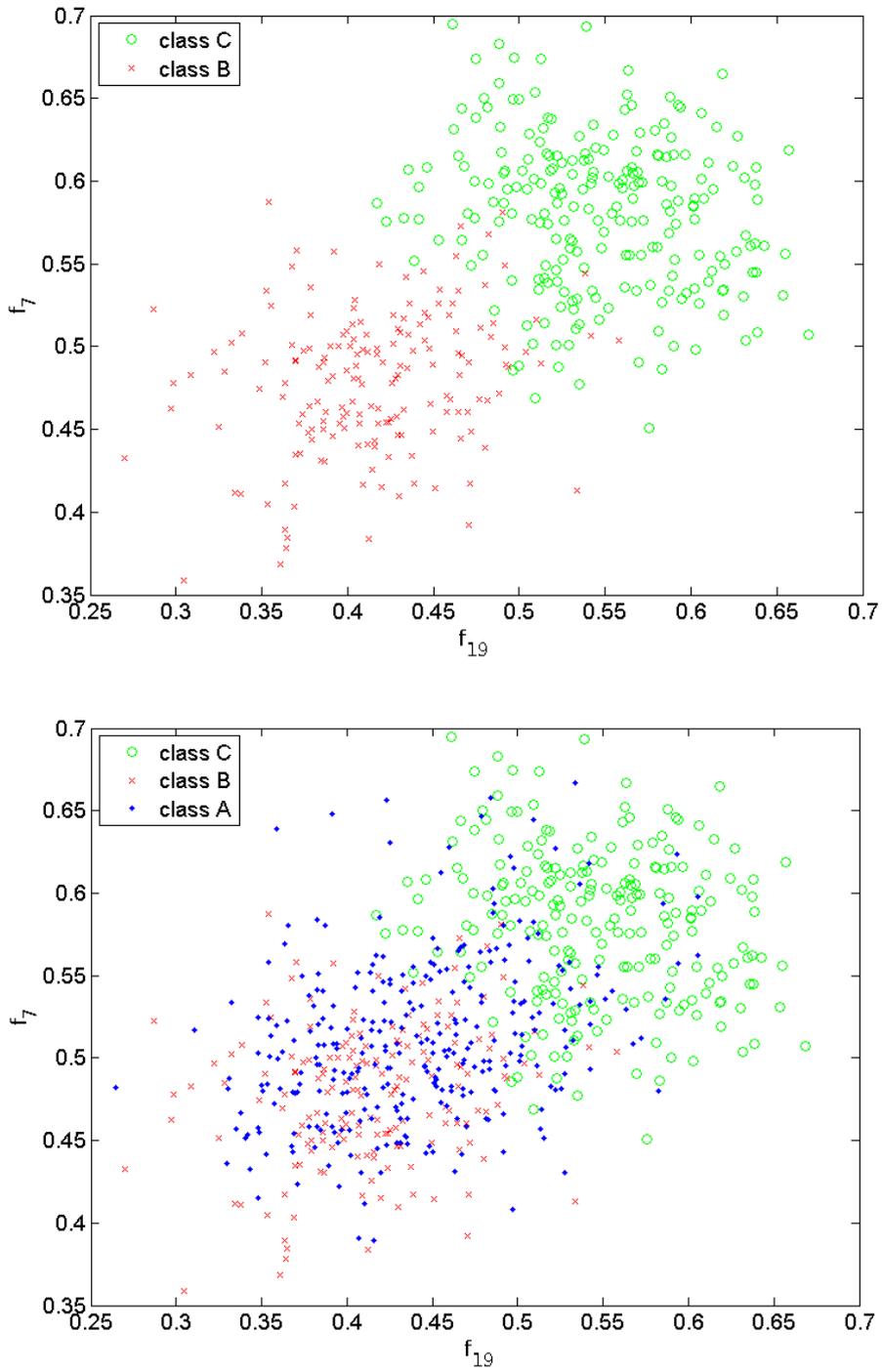
Figure 7.13: Feature space for features f_{19} and f_7, corresponding to relative positions $s = [-1, 0]$ and $s = [-2, 0]$, respectively, in the contextual neighbourhood I_r.

no. of features	classifier	class B – C	class A – C	class A – B
1	Ho-Kashyap	9.21 ± 2.71	17.75 ± 3.57	31.65 ± 3.63
	SVM	28.40 ± 28.71	40.89 ± 18.96	49.78 ± 14.76
2	Ho-Kashyap	6.69 ± 2.32	14.09 ± 3.31	28.49 ± 3.53
	SVM	8.67 ± 2.54	31.59 ± 21.49	44.65 ± 13.36
3	Ho-Kashyap	4.80 ± 2.15	13.45 ± 3.10	26.70 ± 3.84
	SVM	7.28 ± 2.90	20.73 ± 13.10	39.26 ± 12.99
4	Ho-Kashyap	4.78 ± 1.89	13.20 ± 3.23	24.50 ± 3.77
	SVM	7.73 ± 3.09	19.91 ± 13.51	39.47 ± 12.77
5	**Ho-Kashyap**	4.96 ± 1.82	13.21 ± 3.15	$\mathbf{23.40 \pm 3.72}$
	SVM	7.80 ± 2.81	19.54 ± 15.16	39.58 ± 12.21
6	**Ho-Kashyap**	$\mathbf{3.97 \pm 1.91}$	13.59 ± 3.40	23.75 ± 3.42
	SVM	6.31 ± 3.26	17.25 ± 9.24	35.47 ± 12.31
7	**Ho-Kashyap**	4.09 ± 2.03	$\mathbf{11.58 \pm 2.83}$	23.63 ± 3.72
	SVM	6.4 ± 2.82	13.40 ± 4.34	34.59 ± 12.93

Table 7.6: Classification errors [%] of RNF images using two classifiers with the most relevant features from the 2D CAR textural representation (different features were selected for each pair of classes). Boldfaced numbers indicate the best classification results.

for class A is also well separated with respect to class C. Clusters for classes A and B overlaps each other for these features.

7.4.4 Conclusion

The presented results indicate that the textural features based on 2D CAR model can be used for detection of the focal losses in RNF layer. The classification error for our dataset reached 3.97% for discrimination between regions from healthy tissue and regions from tissue affected by RNF losses.

The proposed 2D CAR features may be used as a part of feature vector in Glaucoma Risk Index, as described in Bock et al. (2007). These features can be also used in the screening program together with other features, based on different texture analysis methods (Kolář et al., 2008; Gazárek et al., 2008), which uses a large database of healthy eyes as a control (reference) group.

Chapter 8

Conclusions

We proposed several illumination invariant textural representations, which are based on the modelling of local spatial relations. The texture characteristics are modelled by 2D/3D CAR or GMRF models, which are special types from the Markovian model family and which allow a very efficient estimation of their parameters, without the demanding Monte Carlo minimisation. We derived the novel illumination invariants, which enable to extract the textural representation invariant to brightness, illumination colour/spectrum and which are simultaneously approximately invariant to local intensity changes. These illumination invariants were extended to be simultaneously illumination and rotation invariant. On top of that, the experiments with the proposed invariant textural features showed their robustness to illumination direction variations and the image degradation with an additive Gaussian noise.

The experimental evaluation was performed on five different textural databases: Outex, Bonn BTF, CUReT, ALOT, and KTH-TIPS2, which include images of real-world materials acquired at various conditions. The experiments were designed to closely resemble real-life conditions and the proposed features confirmed their ability to recognise materials in variable illumination conditions and also different viewpoint directions. Our methods do not require any knowledge of acquisition conditions and the recognition is possible even with a single training image per material, if substantial scale variation or perspective projection is not included. The proposed representation outperformed other state of the art textural representations (among others opponent Gabor features, LBP, LBP-HF, and MR8-LINC), only LBP features performed slightly better in two tests with small texture samples. Although, LBP features are nowadays very popular and effective in many situations, they turned out to be very sensitive to noise degradation and illumination direction variations.

The proposed methods for evaluation of textural similarity are also related to the human perception of textures, according to the performed psychophysical experiments. They were either the low level perception of texture degradations or the subjective ranking of tile similarity.

The presented applications included the content based tile retrieval system, which is able to find tiles with similar textures or colours and, consequently, to ease browsing

of digital catalogues. The proposed invariants were also integrated into a segmentation algorithm, in order that computer vision applications can analyse images regardless of illumination conditions. In computer graphics, the features were used for texture degradation description, which opens utilisation in an optimisation of texture compression methods. Last but not least, we applied our textural features in medical imaging and presented their ability to recognise a glaucomatous tissue in retina images.

The results of the invariant texture retrieval or recognition can be reviewed online in our interactive demonstrations[1] so as the presented tile retrieval system[2].

8.1 Future research

Despite the encouraging results presented in this thesis, we still see many possible improvements of the proposed methods as well as feasible applications:

(a) Creating texture-based image representation, which would characterise an image by the invariant textural features computed from homogeneous regions, which would be extracted by the illumination invariant segmenter. This would be an advantageous extension for current CBIR systems based on colours and SIFT features.

(b) Representation of complex textures by means of either a compound model or a combination of models.

(c) Modification of the invariants for 3D CAR model so that it retains the correspondence of spectral planes and simultaneously it does not require the decorrelation, e.g. with a joint diagonalisation (Iferroudjene et al., 2009).

(d) Illumination invariant representation and recognition of dynamic textures.

(e) Thorough evaluation of mutual dependency and redundancy of the features with feature selection methods.

(f) Robustness to other acquisition conditions, namely, reasonable affine transformation as the approximation of projective transformation.

(g) Parallel implementation of the proposed methods.

A long therm objective is a retrieval from a large medical database, where the texture analysis methods can be successfully exploited. Particularly, we intend to study dermatological images, which would create an online automated dermatology consulting system provided that we will have access to relevant medical images.

---

[1]`http://cbir.utia.cas.cz`, `http://cbir.utia.cas.cz/rotinv/`
[2]`http://cbir.utia.cas.cz/tiles/`

Appendix A

Illumination Invariance

A.1 Multiple illumination sources

Let us assume that a textured Lambertian surface is illuminated with two uniform illuminations with different positions and spectra. The notation follows formula (4.1), additionally, $E'(\omega)$ denotes the spectral power distribution of the second illumination and $S'(r, \omega)$ is the Lambertian reflectance coefficient at the position r, again corresponding to the second illumination. The value acquired by the j-th sensor at the location r can be expressed and approximated with formula (4.2) as

$$
\begin{aligned}
Y_{r,j} &= \int_{\Omega} E(\omega)\, S(r, \omega)\, R_j(\omega)\, d\omega + \int_{\Omega} E'(\omega)\, S'(r, \omega)\, R_j(\omega)\, d\omega \ , \\
Y_{r,j} &= \sum_{c=1}^{C} d_{r,c} \int_{\Omega} E(\omega) s_c(\omega) R_j(\omega) d\omega + \sum_{c=1}^{C} d'_{r,c} \int_{\Omega} E'(\omega) s_c(\omega) R_j(\omega) d\omega \\
Y_r &= B' d_r + B'' d'_r \ , \\
\tilde{Y}_r &= \tilde{B}' d_r + \tilde{B}'' d'_r \ ,
\end{aligned}
$$

The linear model (4.3) is valid no more. The model is valid only for synchronised change of spectra:.

$$
\tilde{Y}_r = \tilde{B} \left(B' d_r + B'' d'_r \right) \ .
$$

A.2 Invariance to local intensity changes – 3D CAR

Let us start with an auxiliary construction, which do not include intensity changes for now. We assume that a textured image is composed of n copies of the same small texture tile S, which is homogeneously illuminated. The tiles are placed side by side to cover the whole image lattice I (Fig. 4.1). Statistics computed on the tile S are denoted with superscript $(\cdot)^{(S)}$, while statistics computed on the all image lattice I are denoted with

superscript $(\cdot)^{(I)}$. Relations of the statistics used in model parameter estimation (3.6) are following:

$$V_{yy}^{(I)} \approx \sum_{r \in I} Y_r \, Y_r{}^T \approx n \sum_{r \in S} Y_r \, Y_r{}^T = n V_{yy}^{(S)} \ , \tag{A.1}$$

$$V_{zz}^{(I)} \approx \sum_{r \in I} Z_r \, Z_r{}^T \approx n \sum_{r \in S} Z_r \, Z_r{}^T = n V_{zz}^{(S)} \ , \tag{A.2}$$

$$V_{zy}^{(I)} \approx \sum_{r \in I} Z_r \, Y_r{}^T \approx n \sum_{r \in S} Z_r \, Y_r{}^T = n V_{zy}^{(S)} \ . \tag{A.3}$$

The first approximation in the rows neglects the model prior information, while the second one discards statistics at seams of the tiles. Later on, we will show that colour invariants $\alpha_{1'} - \alpha_3$, $\beta_1 - \beta_5$, $\beta_8 - \beta_{12}$ (introduced in Section 4.2.1) are independent on this enlargement of texture sample. It will be proved as a special case of the invariance to local intensity changes, which follows.

Let us suppose that a modified image is composed of the same texture tiles S, where the ℓ-th texture tile is modified by the multiplication of all its pixels with some constant b_ℓ. This simulates locally constant intensity changes in the image. Accent ($\tilde{\cdot}$) denotes statistics and parameters related to the image with local intensity changes. The relations (A.1 – A.3) become:

$$\tilde{V}_{yy}^{(I)} \approx \sum_{r \in I} \tilde{Y}_r \, \tilde{Y}_r^T \approx \left(\sum_{\ell=1}^{n} b_\ell^2 \right) \sum_{r \in S} Y_r \, Y_r{}^T = k_2 V_{yy}^{(S)} \ , \tag{A.4}$$

$$\tilde{V}_{zz}^{(I)} \approx \sum_{r \in I} \tilde{Z}_r \, \tilde{Z}_r^T \approx \left(\sum_{\ell=1}^{n} b_\ell^2 \right) \sum_{r \in S} Z_r \, Z_r{}^T = k_2 V_{zz}^{(S)} \ , \tag{A.5}$$

$$\tilde{V}_{zy}^{(I)} \approx \sum_{r \in I} \tilde{Z}_r \, \tilde{Y}_r^T \approx \left(\sum_{\ell=1}^{n} b_\ell^2 \right) \sum_{r \in S} Z_r \, Y_r{}^T = k_2 V_{zy}^{(S)} \ , \tag{A.6}$$

where constant $k_2 = \left(\sum_{\ell=1}^{n} b_\ell^2 \right)$. Consequently, model parameter estimates (3.5, 3.7) are related:

$$\tilde{\hat{\gamma}}^{(I)} = \left(\tilde{V}_{zz}^{(I)} \right)^{-1} \tilde{V}_{zy}^{(I)} \approx \left(k_2 V_{zz}^{(S)} \right)^{-1} k_2 V_{zy}^{(S)} = \hat{\gamma}^{(S)} \ , \tag{A.7}$$

$$\tilde{\lambda}^{(I)} \approx k_2 V_{yy}^{(S)} - k_2 \left(V_{zy}^{(S)} \right)^T \left(k_2 V_{zz}^{(S)} \right)^{-1} k_2 V_{zy}^{(S)} = k_2 \lambda^{(S)} \ . \tag{A.8}$$

The invariants $\alpha_1 - \alpha_3$, $\alpha_{1'}$ (defined in Section 4.2.1) are related with following formulas, where the approximation neglects statistics at seams of the tiles:

$$\tilde{\alpha}_1^{(I)} = 1 + \tilde{Z}_t^T \left(\tilde{V}_{zz}^{(I)} \right)^{-1} \tilde{Z}_t = 1 + k_2^{-1} b_n^2 Z_t^T \left(V_{zz}^{(S)} \right)^{-1} Z_t \ , \tag{A.9}$$

$$
\begin{aligned}
\tilde{\alpha}_{1'}^{(I)} &= \left(\sum_{\forall r \in I} \frac{\tilde{Z}_r}{|I|} \right)^T \left(\tilde{V}_{zz}^{(I)} \right)^{-1} \left(\sum_{\forall r \in I} \frac{\tilde{Z}_r}{|I|} \right) \\
&\approx \left(\sum_{\ell=1}^n \frac{b_\ell}{n} \sum_{\forall r \in S} \frac{Z_r}{|S|} \right)^T k_2^{-1} \left(V_{zz}^{(S)} \right)^{-1} \left(\sum_{\ell=1}^n \frac{b_\ell}{n} \sum_{\forall r \in S} \frac{Z_r}{|S|} \right) \\
&= k_2^{-1} \left(\sum_{\ell=1}^n \frac{b_\ell}{n} \right)^2 \left(\sum_{\forall r \in S} \frac{Z_r}{|S|} \right)^T \left(V_{zz}^{(S)} \right)^{-1} \left(\sum_{\forall r \in S} \frac{Z_r}{|S|} \right)
\end{aligned} \tag{A.10}
$$

$$
\begin{aligned}
\tilde{\alpha}_2^{(I)} &= \sqrt{ \sum_{\forall r \in I} \left(\tilde{Y}_r - \tilde{\hat{\gamma}}^{(I)} \tilde{Z}_r \right)^T \left(\tilde{\lambda}^{(I)} \right)^{-1} \left(\tilde{Y}_r - \tilde{\hat{\gamma}}^{(I)} \tilde{Z}_r \right) } \\
&\approx \sqrt{ \sum_{\ell=1}^n \sum_{\forall r \in S} \left(b_\ell Y_r - \hat{\gamma}^{(S)} b_\ell Z_r \right)^T \left(k_2 \lambda^{(S)} \right)^{-1} \left(b_\ell Y_r - \hat{\gamma}^{(S)} b_\ell Z_r \right) } \\
&= \sqrt{ k_2^{-1} \left(\sum_{\ell=1}^n b_\ell^2 \right) \sum_{\forall r \in S} \left(Y_r - \hat{\gamma}^{(S)} Z_r \right)^T \left(\lambda^{(S)} \right)^{-1} \left(Y_r - \hat{\gamma}^{(S)} Z_r \right) } \\
&= \alpha_2^{(S)} \ ,
\end{aligned} \tag{A.11}
$$

$$
\begin{aligned}
\tilde{\alpha}_3^{(I)} &= \sqrt{ \sum_{\forall r \in I} \left(\tilde{Y}_r - \tilde{\mu}^{(I)} \right)^T \left(\tilde{\lambda}^{(I)} \right)^{-1} \left(\tilde{Y}_r - \tilde{\mu}^{(I)} \right) } \\
&\approx \sqrt{ \sum_{\ell=1}^n \sum_{\forall r \in S} \left(b_\ell \tilde{Y}_r - \frac{\sum_l b_l}{n} \mu^{(S)} \right)^T \left(k_2 \lambda^{(S)} \right)^{-1} \left(b_\ell \tilde{Y}_r - \frac{\sum_l b_l}{n} \mu^{(S)} \right) } \\
&= \sqrt{ \sum_{\ell=1}^n \sum_{\forall r \in S} \left(\tilde{Y}_r - \frac{\sum_l b_l}{n b_\ell} \mu^{(S)} \right)^T \left(\lambda^{(S)} \right)^{-1} \left(\tilde{Y}_r - \frac{\sum_l b_l}{n b_\ell} \mu^{(S)} \right) } \ ,
\end{aligned} \tag{A.12}
$$

Therefore only α_2 is invariant to local intensity changes. Terms $\alpha_{1'}$, α_3 are at least independent on texture sample size, which is the consequence of substitution $b_\ell = 1$, $\ell = 1, \ldots, n$ into (A.10), (A.12). The illumination invariant α_1 is neither robust to size of texture sample nor translation invariant.

It is worth to note, why we did not define $\alpha_{1'}$ by means of α_1 computed at all pixel positions: $-|I| + \sum_{\forall t \in I} (1 + Z_t^T V_{zz(t)}^{-1} Z_t)$, which would be invariant to local inten-

sity changes. The reason is that such definition would be very close to the expression $\sum_{\forall r \in I} Z_r^T (\sum_{\forall t \in I} Z_t Z_t^T)^{-1} Z_r$, which is constant for given I (regardless the pixel values) and therefore useless.

Invariance to local intensity changes of colour invariants β_3 , β_4 , β_{10} , β_{11} (defined in Section 4.2.1) is a direct consequence of the following relations, which implies from (A.8), (A.4), (A.5):

$$\left|\tilde{\lambda}^{(I)}\right| \approx \left|k_2 \lambda^{(S)}\right| = k_2^C \left|\lambda^{(S)}\right| \; , \tag{A.13}$$

$$\left|\tilde{V}_{yy}^{(I)}\right| \approx \left|k_2 V_{yy}^{(S)}\right| = k_2^C \left|V_{yy}^{(S)}\right| \; , \tag{A.14}$$

$$\left|\tilde{V}_{zz}^{(I)}\right| \approx \left|k_2 V_{zz}^{(S)}\right| = k_2^{C\eta} \left|V_{zz}^{(S)}\right| \; . \tag{A.15}$$

Similarly, local intensity invariance for β_5 and β_{12} implies from formulas (A.8), (A.4). Terms β_1 , β_2 , β_8 , and β_9 are only independent on texture sample size (see analogous derivation for GMRF in the next section). Unfortunately, β_6 and β_7 depend on texture sample size, because probabilities $p(Y_r|Y^{(r-1)})$ (3.11) and $\ln p(Y^{(t)}|M_\ell)$ (3.12) include non-linear functions of the number of previously analysed data, e.g. power to $\psi(t)$.

As it was mentioned, all invariants to local intensity changes are simultaneously independent on texture sample size, which is the consequence of substitution $b_\ell = 1$, $\ell = 1, \ldots, n$. The previously derived properties of $\alpha_{1'} - \alpha_3$, $\beta_1 - \beta_5$, and $\beta_8 - \beta_{12}$ applies for 2D CAR model as well (C is set to $C = 1$), where they are computed for each spectral plane separately.

A.3   Invariance to local intensity changes – GMRF

Let us again assume that a textured image is composed of n small texture tiles S placed side by side to cover the whole image lattice I (Fig. 4.1). The statistics of tile S are denoted with superscript $(\cdot)^{(S)}$, while statistics computed on the all image lattice I are denoted with superscript $(\cdot)^{(I)}$. Relations of the statistics (3.21) used in model parameter estimation are following, $j = 1, \ldots, C$:

$$V_{yy,j}^{(I)} = \sum_{r \in I} Y_{r,j} Y_{r,j} \approx n \sum_{r \in S} Y_{r,j} Y_{r,j} = n V_{yy,j}^{(S)} \; , \tag{A.16}$$

$$V_{zz,j}^{(I)} = \sum_{r \in I} Z_{r,j} Z_{r,j}^T \approx n \sum_{r \in S} Z_{r,j} Z_{r,j}^T = n V_{zz,j}^{(S)} \; , \tag{A.17}$$

$$V_{zy,j}^{(I)} = \sum_{r \in I} Z_{r,j} Y_{r,j} \approx n \sum_{r \in S} Z_{r,j} Y_{r,j} = n V_{zy,j}^{(S)} \; . \tag{A.18}$$

The approximation discards statistics at seams of the tiles. Again, independence on this enlargement of texture sample will be shown as a special case of the following invariance to local intensity changes.

We assume that the image with locally constant intensity changes is composed of the same texture tiles S, where the ℓ-th texture tile is modified by the multiplication of all

its pixels and spectral planes with some constant b_ℓ. Accent ($\tilde{\cdot}$) denotes statistics and parameters related to the image with local intensity changes. The relations of statistics (A.16) – (A.18) become:

$$\tilde{V}_{yy,j}^{(I)} = \sum_{r \in I} \tilde{Y}_{r,j} \tilde{Y}_{r,j} \approx \left(\sum_{\ell=1}^{n} b_\ell^2 \right) \sum_{r \in S} Y_r Y_r = k_2 V_{yy,j}^{(S)} \ , \tag{A.19}$$

$$\tilde{V}_{zz,j}^{(I)} = \sum_{r \in I} \tilde{Z}_{r,j} \tilde{Z}_{r,j}^T \approx \left(\sum_{\ell=1}^{n} b_\ell^2 \right) \sum_{r \in S} Z_{r,j} Z_{r,j}^T = k_2 V_{zz,j}^{(S)} \ , \tag{A.20}$$

$$\tilde{V}_{zy,j}^{(I)} = \sum_{r \in I} \tilde{Z}_{r,j} \tilde{Y}_{r,j} \approx \left(\sum_{\ell=1}^{n} b_\ell^2 \right) \sum_{r \in S} Z_{r,j} Y_{r,j} = k_2 V_{zy,j}^{(S)} \ , \tag{A.21}$$

where constant $k_2 = \left(\sum_{\ell=1}^{n} b_\ell^2 \right)$. Consequently, model parameter estimates (3.22, 3.20) are related:

$$\begin{aligned} \tilde{\hat{\gamma}}_j^{(I)} &= \left(\tilde{V}_{zz,j}^{(I)} \right)^{-1} \left(\tilde{V}_{zy,j}^{(I)} \right) = \left(k_2 V_{zz,j}^{(S)} \right)^{-1} \left(k_2 V_{zy,j}^{(S)} \right) \\ &= \hat{\gamma}_j^{(S)} \ , \end{aligned} \tag{A.22}$$

$$\begin{aligned} \left(\tilde{\hat{\sigma}}_j^{(I)} \right)^2 &= \frac{1}{|I|} \sum_{\forall r \in I} (\tilde{Y}_{r,j} - \tilde{\hat{\gamma}}_j^{(I)} \tilde{Z}_{r,j})^2 \\ &\approx \frac{1}{n|S|} \sum_{\ell=1}^{n} \sum_{\forall r \in S} (b_\ell Y_{r,j} - \hat{\gamma}_j^{(S)} Z_{r,j} b_\ell)^2 \\ &= \frac{k_2}{n} \left(\hat{\sigma}_j^{(S)} \right)^2 \end{aligned} \tag{A.23}$$

The invariants $\alpha_{2,j}$, $\alpha_{3,j}$ for the GMRF model (introduced in Section 4.2.3) are related with the following formulas, where the approximation again discards statistics at seams of the tiles, $j = 1, \ldots, C$:

$$\begin{aligned} \tilde{\alpha}_{2,j}^{(I)} &= \sqrt{\frac{1}{|I|} \sum_{\forall r \in I} \left(\tilde{\hat{\sigma}}_j^{(I)} \right)^{-2} \left(\tilde{Y}_{r,j} - \tilde{\hat{\gamma}}_j^{(I)} \tilde{Z}_{r,j} \right)^2} \\ &\approx \sqrt{\frac{1}{n|S|} \sum_{\ell=1}^{n} \sum_{\forall r \in S} \frac{n}{k_2} \left(\hat{\sigma}_j^{(S)} \right)^{-2} \left(b_\ell Y_{r,j} - \hat{\gamma}_j^{(S)} Z_{r,j} b_\ell \right)^2} \\ &= \alpha_{2,j}^{(S)} \ , \end{aligned} \tag{A.24}$$

$$\tilde{\alpha}_{3,j}^{(I)} = \sqrt{\frac{1}{|I|} \sum_{\forall r \in I} \left(\tilde{\tilde{\sigma}}_j^{(I)}\right)^{-2} \left(\tilde{Y}_{r,j} - \tilde{\mu}_j^{(I)}\right)^2}$$

$$\approx \sqrt{\frac{1}{n|S|} \sum_{\ell=1}^{n} \sum_{\forall r \in S} \frac{n}{k_2} \left(\hat{\sigma}_j^{(S)}\right)^{-2} \left(b_\ell Y_{r,j} - \frac{\sum_l b_l}{n} \mu_j^{(S)}\right)^2}$$

$$\approx \sqrt{\frac{1}{|S|} \sum_{\ell=1}^{n} \sum_{\forall r \in S} \left(\hat{\sigma}_j^{(S)}\right)^{-2} \left(Y_{r,j} - \frac{\sum_l b_l}{n b_\ell} \mu_j^{(S)}\right)^2} \quad , \tag{A.25}$$

Again, $\alpha_{2,j}$ is invariant to local intensity changes, while $\alpha_{3,j}$ is only independent on texture sample size (the consequence of substitution $b_\ell = 1, \ \ell = 1, \ldots, n$).

Invariance to local intensity changes of colour invariants $\beta_{3,j} - \beta_{5,j}$, $\beta_{10,j} - \beta_{12,j}$ (defined in Section 4.2.3) is a consequence of the following relations, which implies from (A.23), (A.19), (A.20):

$$\left|\tilde{\tilde{\sigma}}_j^{2(I)}\right| \approx \frac{k_2}{n} \left|\hat{\sigma}_j^{2(S)}\right| \quad , \qquad \left|\tilde{V}_{yy,j}^{(I)}\right| \approx k_2 \left|V_{yy,j}^{(S)}\right| \quad , \qquad \left|\tilde{V}_{zz,j}^{(I)}\right| \approx k_2^\eta \left|V_{zz,j}^{(S)}\right| \quad . \tag{A.26}$$

Derivation for $\beta_{1,j} - \beta_{5,j}$ is following, where S_1 is the first placement of tile S:

$$\beta_{1,j}^{(I)} = \ln\left(\left|\tilde{\tilde{\sigma}}_j^{2(I)}\right| \left|\tilde{\tilde{\sigma}}_j^{2(S_1)}\right|^{-1}\right)$$

$$\approx \ln\left(\frac{k_2}{n b_1^2} \left|\hat{\sigma}_j^{2(S)}\right| \left|\hat{\sigma}_j^{2(S)}\right|^{-1}\right) \quad ,$$

$$\beta_{2,j}^{(I)} = \ln\left(\operatorname{abs}\left|\tilde{V}_{zz,j}^{(I)}\right| \operatorname{abs}\left|\tilde{V}_{zz,j}^{(S_1)}\right|^{-1} \frac{|S_1|^\eta}{|I|^\eta}\right)$$

$$\approx \ln\left(\frac{k_2^\eta}{b_1^{2\eta}} \operatorname{abs}\left|V_{zz,j}^{(S)}\right| \operatorname{abs}\left|V_{zz,j}^{(S)}\right|^{-1} n^{-\eta}\right) \quad ,$$

$$\beta_{3,j}^{(I)} = \ln\left(\operatorname{abs}\left|\tilde{V}_{zz,j}^{(I)}\right| \left|\tilde{\tilde{\sigma}}_j^{2(I)}\right|^{-\eta} |I|^{-\eta}\right)$$

$$\approx \ln\left(k_2^\eta \operatorname{abs}\left|V_{zz,j}^{(S)}\right| \frac{n^\eta}{k_2^\eta} \left|\hat{\sigma}_j^{2(S)}\right|^{-\eta} n^{-\eta}|S|^{-\eta}\right) = \beta_{3,j}^{(S)} \quad ,$$

$$\beta_{4,j}^{(I)} = \ln\left(\operatorname{abs}\left|\tilde{V}_{zz,j}^{(I)}\right| \left|\tilde{V}_{yy,j}^{(I)}\right|^{-\eta}\right)$$

$$\approx \ln\left(k_2^\eta \operatorname{abs}\left|V_{zz,j}^{(S)}\right| k_2^{-\eta} \left|V_{yy,j}^{(S)}\right|^{-\eta}\right) = \beta_{4,j}^{(S)} \quad ,$$

$$\begin{aligned}
\beta_{5,j}^{(I)} &= \tilde{V}_{yy,j}^{(I)} \; \tilde{\hat{\sigma}}_j^{\,-2(I)} \; |I|^{-1} \\
&\approx k_2 \, V_{yy,j}^{(I)} \; \frac{n}{k_2} \; \hat{\sigma}_j^{\,-2(S)} \; n^{-1} \, |S|^{-1} = \beta_{5,j}^{(S)} \quad ,
\end{aligned}$$

and derivation for $\beta_{8,j}$ – $\beta_{12,j}$ can be done analogically. Terms $\beta_{1,j}$, $\beta_{2,j}$, $\beta_{8,j}$, and $\beta_{9,j}$ are only independent on texture sample size (the consequence of substitution $b_\ell = 1$, $\ell = 1, \ldots, n$).

As it was mentioned, all invariants to local intensity changes are simultaneously independent on texture sample size.

Appendix B

Additional Experiments

This chapter contains results of two additional experiments and some examples of material appearance in variable illumination conditions. The first experiments in the appendix comprise additional results for Experiment i2 (Section 6.1.2), including subsets of the proposed invariants and the textural representation without these invariants. The second experiment is recognition of grey-scale images from Bonn BTF database, using the textural representation without the invariants.

B.1   Experiment i2 – Outex_TC_0014

This experiment extended Experiment i2 described in Section 6.1.2. The experimental setup was identical, but different combinations of textural features were tested. The CAR model based features were additionally tested with and without invariants α_1, α_2, α_3 in the feature vectors. Similarly, the GMRF based features were tested with and without invariants α_2, α_3 (see Section 4.2 for definitions). Moreover, we included the textural features without computation of illumination invariants. They were defined in Section 3.1 and they are denoted by "noII" suffix in the results. Finally, due to the directionality of CAR models, we also tested CAR models estimated in three orthogonal directions: row-wise, column-wise top-down and column-wise bottom-up; this is denoted by "3x" suffix in the results.

The results are displayed in Tab. B.1, which also includes correct classification of images degraded by an additive Gaussian noise. It can be seen that illumination invariants outperformed textural features without illumination invariance, and the addition of invariants α_1, α_2, α_3 further improved the performance. The CAR models computed in three orthogonal directions did not achieved significant improvement in this experiment, although the feature vector was three times longer. Finally, the results of dissimilarities L_1 and FC_3 were comparable in this experiment.

	added noise σ				
method	0	2	4	8	size
2D CAR-KL noII, L_1	64.0	60.4	54.1	47.3	72
2D CAR-KL α_1, L_1	64.9	58.7	56.5	49.4	108
2D CAR-KL $\alpha_1\alpha_3$, L_1	67.6	60.8	55.7	52.3	120
2D CAR-KL $\alpha_1\alpha_2\alpha_3$, L_1	**67.6**	**60.8**	**55.7**	**52.4**	132
2D CAR-KL 3x α_1, L_1	64.9	58.8	53.3	50.9	324
2D CAR-KL 3x $\alpha_1\alpha_3$, L_1	67.0	61.0	55.1	51.6	360
2D CAR-KL 3x $\alpha_1\alpha_2\alpha_3$, L_1	**67.5**	**61.0**	55.1	51.8	396
2D CAR-KL $\alpha_1\alpha_2\alpha_3$, $L_{0.2}$	66.3	60.5	55.2	51.0	132
2D CAR-KL $\alpha_1\alpha_2\alpha_3$, FC_3	**67.5**	**63.3**	**55.8**	**51.0**	132
2D CAR-KL 3x $\alpha_1\alpha_2\alpha_3$, FC_3	**68.8**	61.8	55.1	50.1	396
3D CAR α_1, L_1	58.9	54.8	53.0	47.6	100
3D CAR $\alpha_1\alpha_3$, L_1	63.5	61.2	60.5	54.7	104
3D CAR $\alpha_1\alpha_2\alpha_3$, L_1	63.6	**61.3**	**60.6**	**54.9**	108
3D CAR 3x α_1, L_1	60.6	54.8	53.3	48.9	300
3D CAR 3x $\alpha_1\alpha_3$, L_1	65.4	61.2	60.7	55.7	312
3D CAR 3x $\alpha_1\alpha_2\alpha_3$, L_1	**65.5**	**61.3**	**60.8**	**55.9**	324
3D CAR $\alpha_1\alpha_2\alpha_3$, $L_{0.2}$	63.5	59.7	55.4	47.4	108
3D CAR $\alpha_1\alpha_2\alpha_3$, FC_3	65.3	60.4	58.0	51.3	108
3D CAR 3x $\alpha_1\alpha_2\alpha_3$, FC_3	**67.2**	**63.1**	**60.4**	**52.5**	324
3D CAR-KL 3x α_1, L_1	62.3	57.0	52.3	48.7	300
3D CAR-KL 3x $\alpha_1\alpha_3$, L_1	65.8	63.3	61.0	56.1	312
3D CAR-KL 3x $\alpha_1\alpha_2\alpha_3$, L_1	65.7	**63.2**	**61.0**	**56.1**	324
GMRF-KL noII, L_1	56.7	54.4	48.5	43.0	72
GMRF-KL, L_1	58.2	53.5	47.3	44.6	96
GMRF-KL α_3, L_1	61.5	57.1	51.2	46.1	108
GMRF-KL $\alpha_2\alpha_3$, L_1	61.5	57.0	51.1	46.1	120

Table B.1: Experiment i2: Results [%] of the Outex classification test Outex_TC_0014, the classification was performed using the 3-NN classifier. The columns also contain classification on the images degraded with an additive Gaussian noise with variance σ^2. The last column contains size of feature vectors.

B.2   Experiment i5 – Bonn BTF grey

This experiment (Haindl and Vacha, 2006) was performed with the textural features proposed in Section 3.1, without computation of illumination invariants. The features were tested in the task of natural material recognition under varying illumination direction. The setup is similar to Experiment i3 (Section 6.1.3), but only materials with larger resolution were used and seven of them were selected for this experiment.

The proposed textural features were estimated on a gradient image, because it partly reduced effects of illumination variations. Firstly, the input image was converted into a grey-scale image G and the gradient image was estimated by means of differences:

$$\nabla G(r_1, r_2) = [G(r_1 + 1, r_2) - G(r_1 - 1, r_2), G(r_1, r_2 + 1) - G(r_1, r_2 - 1)] \ ,$$

at each pixel position $r = (r_1, r_2)$. Subsequently, the Gaussian pyramid with $K = 4$ levels was built and each level was modelled by either 2D CAR or GMRF model, with the fifth hierarchical neighbourhood ($\eta = 12$ neighbours). The textural features were model parameters A_s, $\forall s \in I_r$ (see details in Section 3.1) estimated separately on both spectral planes of the gradient image at all its pyramid levels, finally, they were all concatenated into a common feature vector. The distance between two feature vectors was measured using L_1 norm. This method, which is denoted with "noII$_2$" suffix in the results, was compared with two alternative textural representations: Gabor features and steerable pyramid based features (see Section 2.2 for details).

The experiment was performed on BTF material measurements from University of Bonn BTF database (Meseth et al., 2003), where natural materials were measured with 81 different illumination directions. The seven materials with resolution 512×512 pixels were selected, which was 567 images in total. The materials used in the experiment are displayed in Fig. B.1. We tested the correct classification of these materials, where the 1-NN classifier was trained with a single image per material and the other 560 images were classified. Additionally, we evaluated the performance of illumination invariant texture retrieval.

The percentage of correct classification is displayed in Tab. B.2, where the single training image per material was selected to be perpendicularly illuminated and the test images were grouped according to declination angle of illumination direction. It can be seen that 2D CAR and GMRF based textural features are more robust to illumination direction changes than Gabor features or steerable pyramid based features. Moreover, Tab. B.3 displays correct classification averaged over 10^5 random selections of a single training image per material, as well as mean recall rates (6.1). The improvement to alternative features was about 9% and the proposed features were approximately two times faster than the fastest compared alternative.

Figure B.1: High resolution material measurements from Bonn BTF database which were used in Experiments i5.

method	$[0°, 30°]$	$[45°, 65°]$	$75°$	average
2D CAR noII$_2$, L_1	84.1	73.3	67.2	73.9
GMRF noII$_2$, L_1	92.8	**80.5**	**69.0**	**79.8**
Gabor features	**97.6**	75.2	24.4	64.9
Steerable pyramid features	82.5	49.2	27.4	50.2

Table B.2: Experiment i5: Accuracy of material recognition [%], with the single training image per material fixed to the perpendicular illumination (declination angle $0°$). The test images are grouped according to the declination angle of illumination direction. The last column is the average over all test images.

method	classif.	RR$_{88}$	RR$_{100}$	size
2D CAR noII$_2$, L_1	81	80	82	96
GMRF noII$_2$, L_1	**85**	**84**	**85**	**96**
Gabor features	71	70	72	48
Steerable pyramid features	77	75	77	968

Table B.3: Experiment i5: Accuracy of material recognition (classif.) [%] and mean recall rate (RR$_n$) [%] for n textures retrieved. The last column contains size of feature vectors.

B.3 Example images

This section contains example images from Outex and Bonn BTF texture databases, which show variation of natural material appearance in different illumination conditions. Materials illuminated with different illumination spectra are displayed in Fig. B.2 and some examples of illumination invariant retrieval are shown in Fig. B.3, all from Outex texture database. On the other hand, effects of illumination direction changes are presented on images from Bonn BTF databse. Fig. B.5 shows materials illuminated with different declination angle, while Fig. B.5 displays changes of appearance for different azimuthal angle.
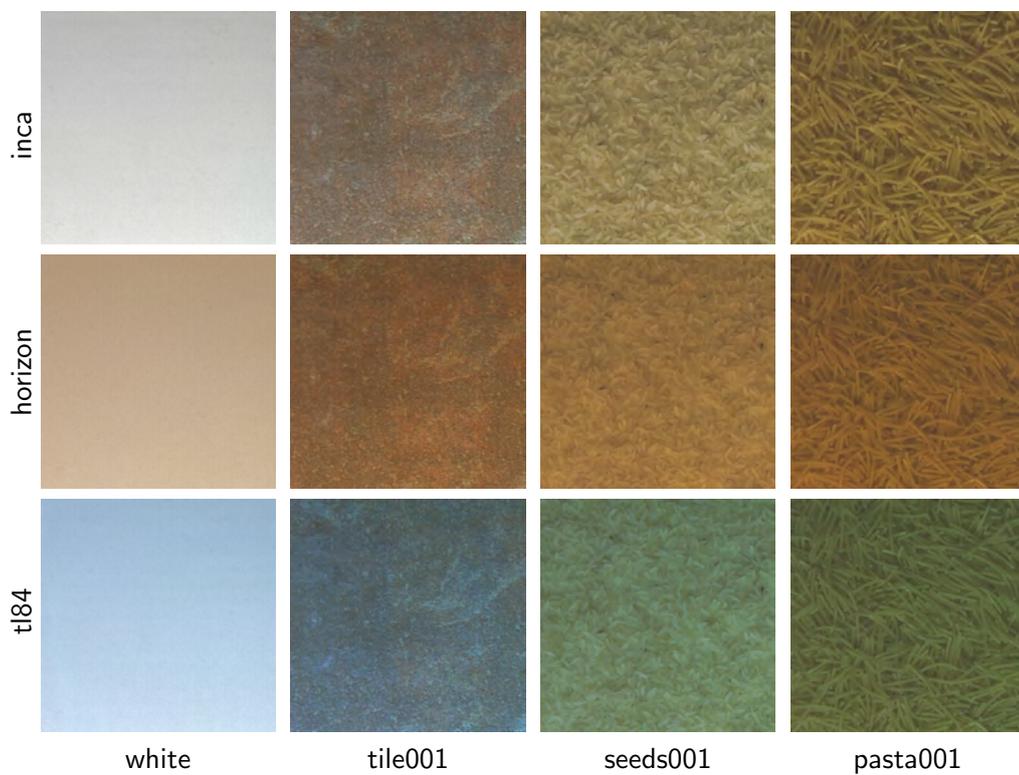
Figure B.2: Apperance of selected materials from Outex texture database. The materials, displayed in columns, are illuminated with different light sources in each row. From top, illuminants are: incandescent CIE A, horizon sunlight, and fluorescent TL84. The first column contains images of the reference white paper.
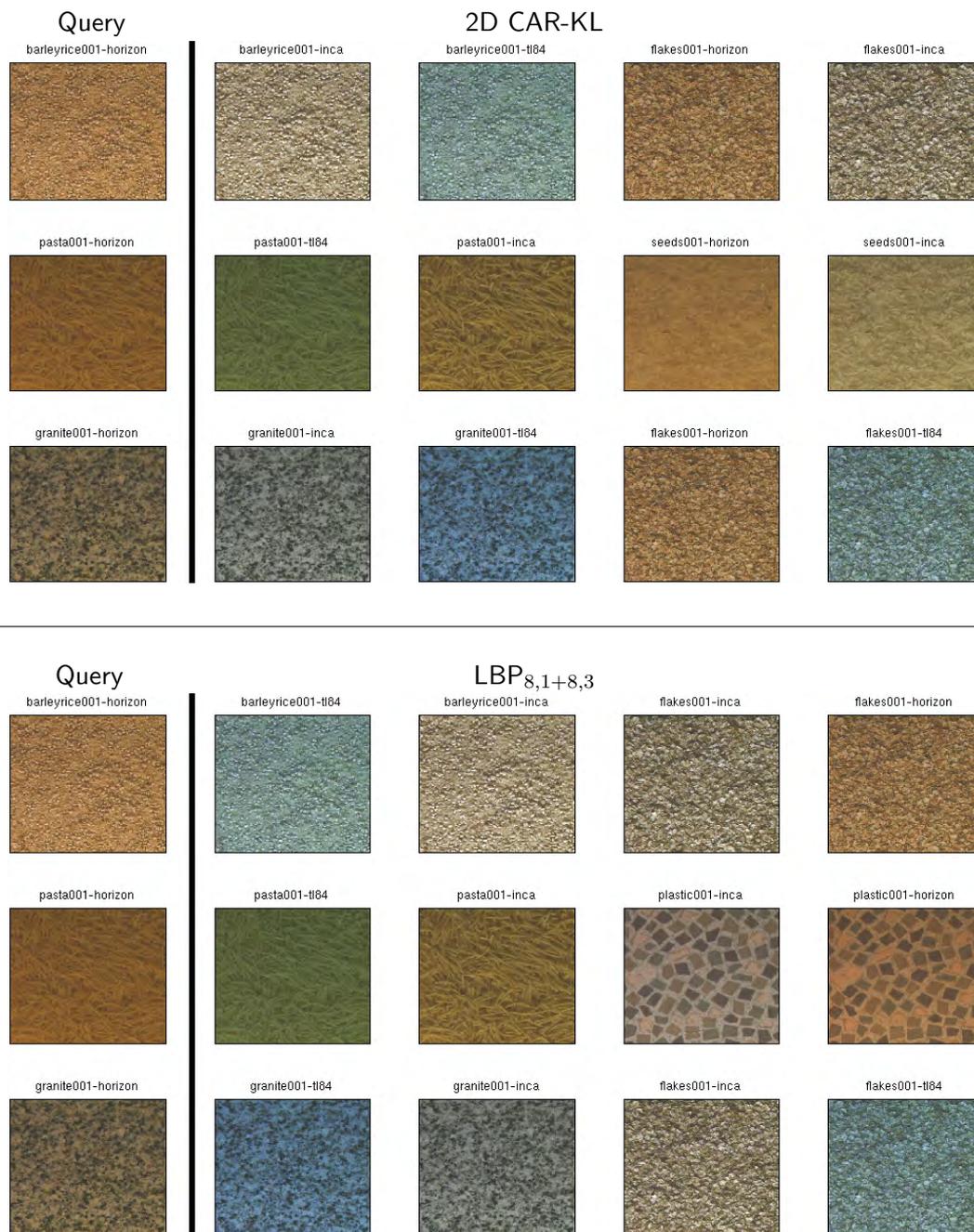
Query   2D CAR-KL

barleyrice001-horizon | barleyrice001-inca | barleyrice001-tl84 | flakes001-horizon | flakes001-inca

pasta001-horizon | pasta001-tl84 | pasta001-inca | seeds001-horizon | seeds001-inca

granite001-horizon | granite001-inca | granite001-tl84 | flakes001-horizon | flakes001-tl84

Query   LBP$_{8,1+8,3}$

barleyrice001-horizon | barleyrice001-tl84 | barleyrice001-inca | flakes001-inca | flakes001-horizon

pasta001-horizon | pasta001-tl84 | pasta001-inca | plastic001-inca | plastic001-horizon

granite001-horizon | granite001-tl84 | granite001-inca | flakes001-inca | flakes001-tl84

Figure B.3: Experiment i1: Illumination invariant image retrieval from Outex texture database. The query images are followed by retrieved images with either the proposed "2D CAR-KL" features or "LBP$_{8,1+8,3}$" computed on grey-scale images. We can observe that both features recognised visual similarity of barley-rice, flakes and granite (the first and third rows).

140

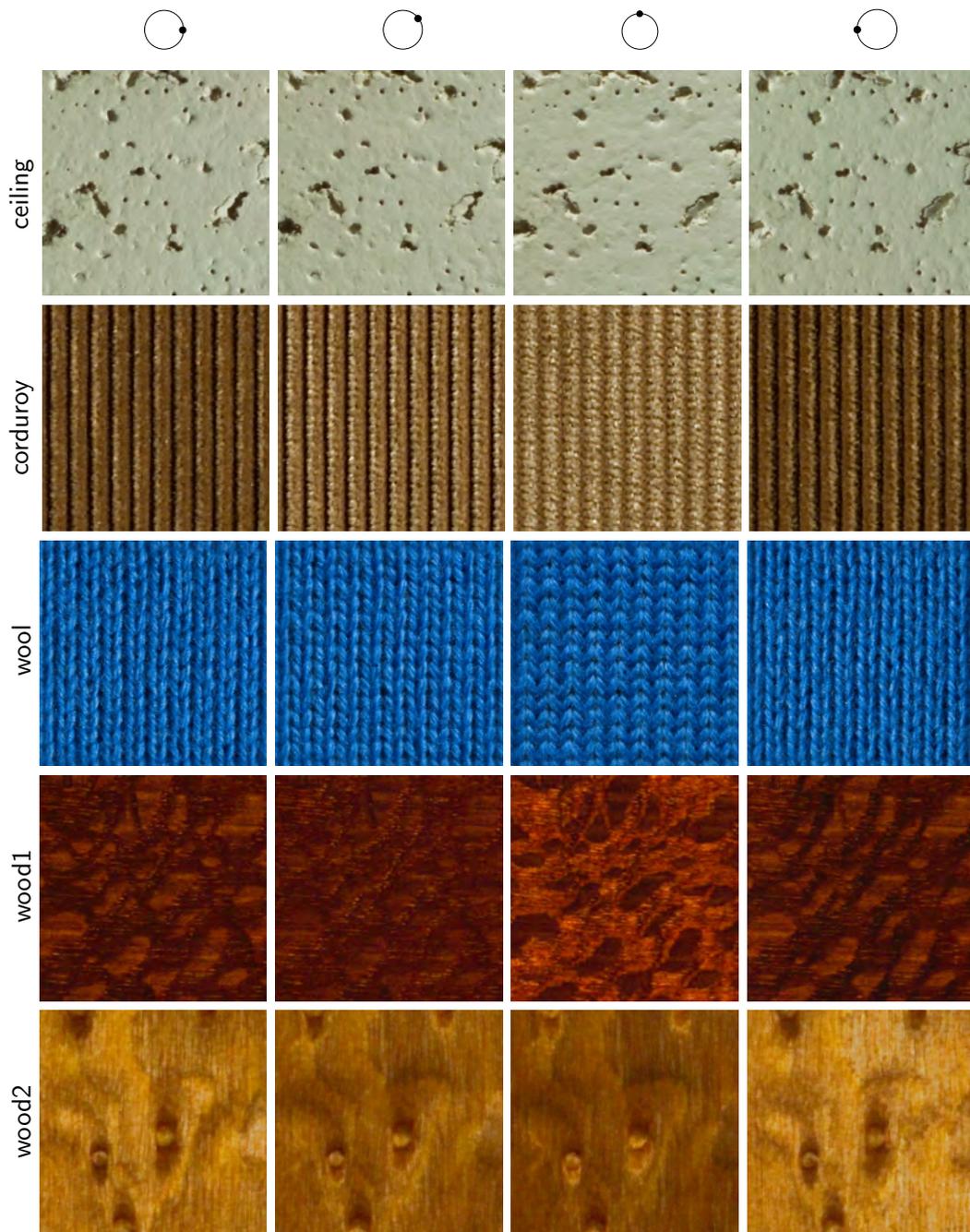Figure B.4: Material measurements from Bonn BTF database used in Experiment i3.
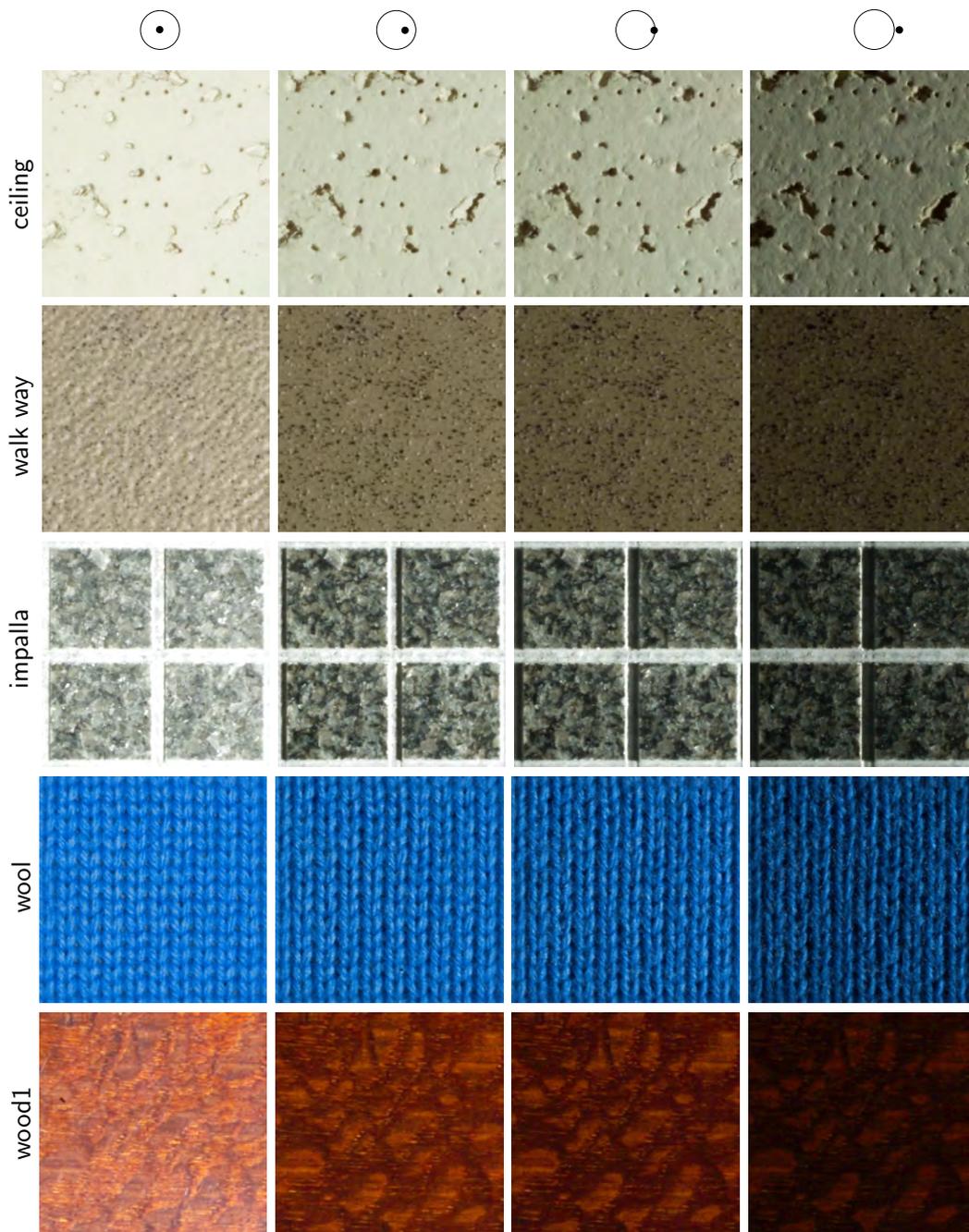
Figure B.5: Appearance of selected materials from Bonn BTF database under illumination with varying declination angle. The columns from left were illuminated with the following declination angles: 15°, 45°, 60°, and 75° from the surface macro-normal.

Figure B.6: Appearance of selected materials from Bonn BTF database under illumination with varying azimuthal angle. The columns from left are illuminated with the following azimuthal angles: 0°, 36°, 90°, and 180°; the declination angle was fixed to 60° from the surface macro-normal.

Appendix C

Demonstrations

We developed a simple CBIR system to demonstrate the capabilities of the proposed illumination and rotation invariant textural features (Sections 4.2, 5.2).

The demonstration usecase is following: At first, a user selects a query image for the content-based image retrieval. The query image can be any image from the provided database. When the retrieve action is triggered, the system retrieves the given number of images, which are visually most similar according to the used method (features) and displays the result images. Optionally, the user can change the number of retrieved images and the method used for similarity judgement.

C.1   Online demonstrations

The demonstration application is a web based application, which implements the previously described CBIR usecase. The application consists of two main pages. The input page allows the user to select a query image, while the result page shows retrieved images.

The input page (Fig. C.1) consists of thumbnail images from the image database and the parameters of retrieval. After the selection of the query image by the left click, the retrieval is performed and the result images are displayed. The "`settings`" button allows to change the number of retrieved images or the method (features) used for retrieval.

The result page (Fig. C.2) revises the query image and retrieval parameters in its upper part, while the thumbnails of retrieved images are displayed in the lower part of the page. The user can select one of the result images as the query image for the next retrieval task. The "`< input`" button returns to the input page.

On the settings page, the user can enable comparison of two methods used for the texture description and similarity judgement. Subsequently, the result page is split into left and right parts, which include results from the respective methods (see Fig. C.2).

Alternatively, the same application is used for the exploration of classification performance. In this case, the input page consists of test images and the result page shows

---

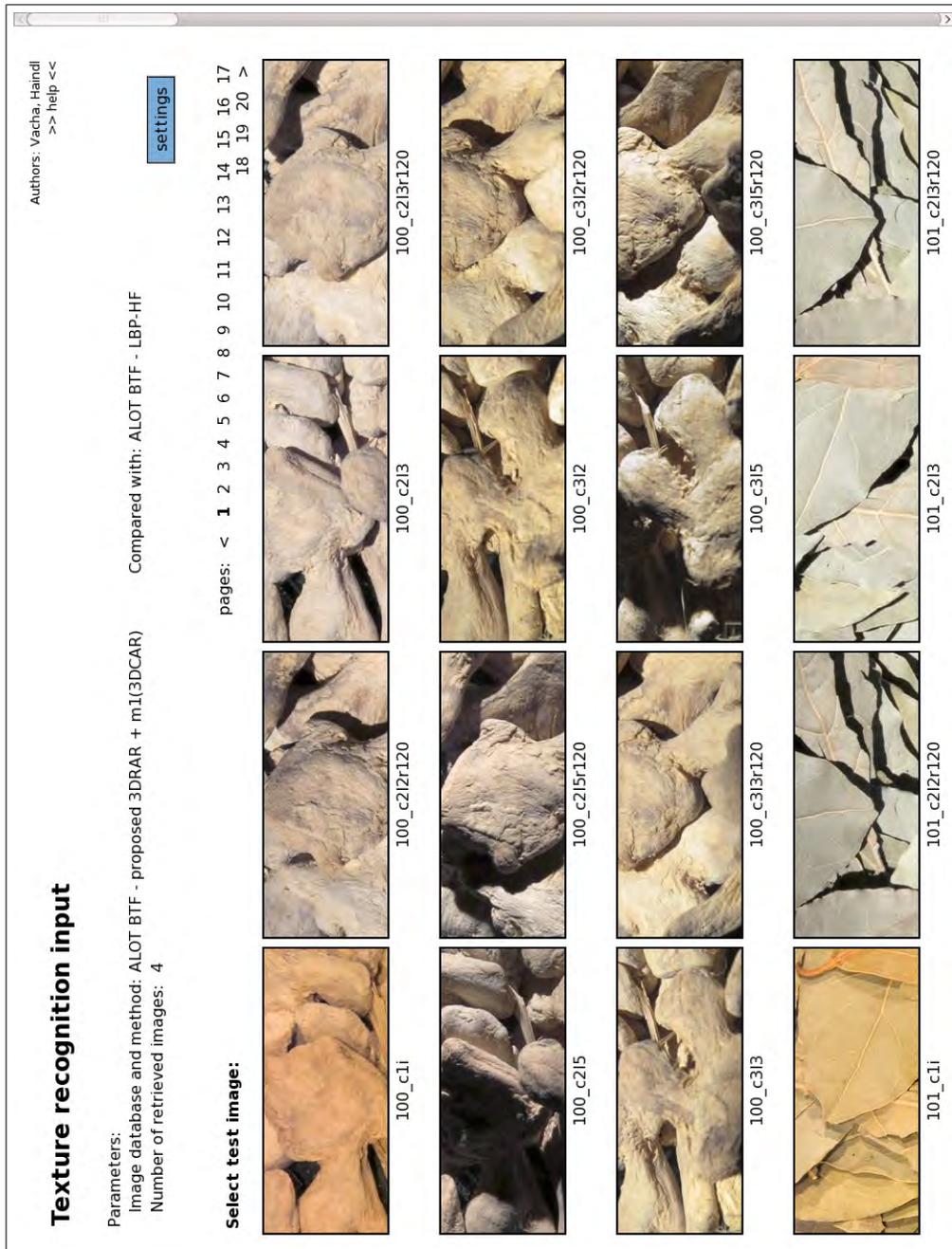[1]Java, J2EE, Java runtime environment, JRE, JSP are registered trademarks of Sun Microsystems.

Figure C.1: Input page of the online demonstration. Left click on the image thumbnail triggers the retrieval of similar images. Optionally, a user can change the number of retrieved images or the used method by pressing the "settings" button.

Figure C.2: Result page of the online demonstration. The query is revised in the upper part and the most similar images are displayed below that. This screenshot show comparison of two methods, their results are displayed in left and right parts of the page.

URL	experiment
description	

`http://cbir.utia.cas.cz/retrievalWebDemoOutex/`	Section 6.1.1
Retrieval of similar textures from the Outex database. It contains materials illuminated with three different spectra. Alternatively, noise degraded Outex images can be selected.	

`http://cbir.utia.cas.cz/retrievalWebDemoCuret/`	Section 6.2
Retrieval of similar textures from the CUReT database. It contains materials with various illumination and viewpoint directions.	

`http://cbir.utia.cas.cz/retrievalWebDemoSanita/`	Section 7.1
Retrieval of similar tiles from Sanita.cz catalogue. The system retrieves tiles with similar colours or texture.	

`http://cbir.utia.cas.cz/retrievalWebDemoAlot/`	Section 6.3.1
Rotation invariant classification on the ALOT database. The input page contains test images and the result page shows the nearest training images.	

Table C.1: List of online demonstrations, all of them utilise the proposed illumination invariant textural features.

the nearest training images. This possibility is depicted in screenshot Fig. C.2. Available online demonstrations are listed in Tab. C.1.

Concerning implementation details, the demonstrations are implemented as web applications according to Java 2 Platform Enterprise Edition (J2EE) standard. The presentation layer is composed of Java Servlets and Java Server Pages (JSP), while the business layer contains Java objects. The demonstrations are running at servlet container Apache Tomcat version 6.0, however, any J2EE 5 compliant container can be utilised.

C.2 Standalone application

Similarly to the previous web application, we developed a standalone desktop application (Vacha and Haindl, 2007b), which also implements the previous CBIR usecase. It again consists of two screens: input screen and result screen.

The input screen (Fig. C.3) consists of thumbnail images from the image database and fields for optional settings as number of retrieved images. The left click selects the query image and performs the retrieval. In the "`image database`" menu, the user can choose from different methods (features) used for the evaluation of image similarity.

The result screen (Fig. C.4) revises the query image and retrieval parameters, followed by the retrieved images. The user can select one of the result images as the query image
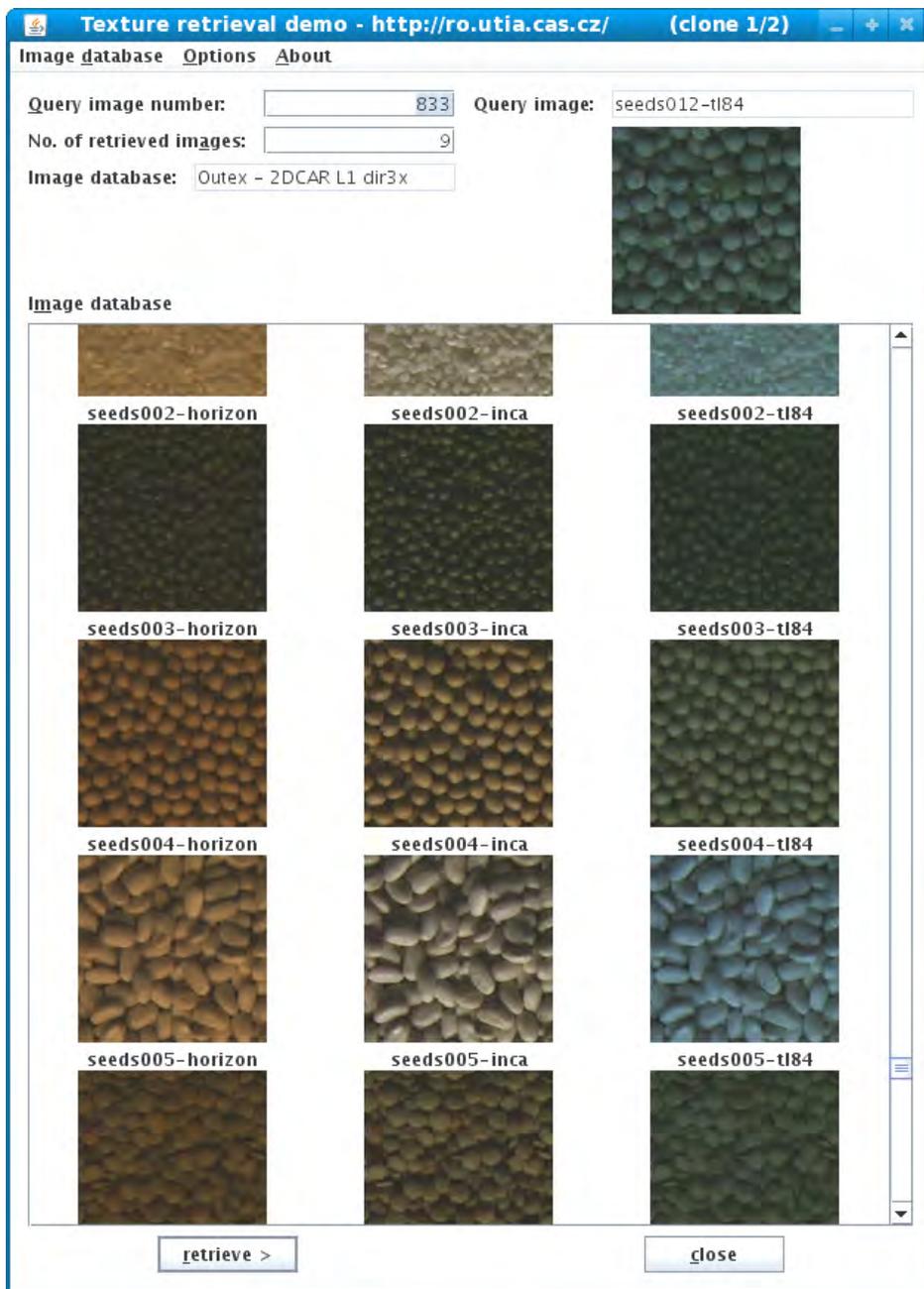
Figure C.3: Input screen of the desktop demonstration. Left click on an image thumbnail triggers the retrieval. Optionally, a user can change the number of retrieved images or the method used for similarity judgement.

Figure C.4: Result screen of the desktop demonstration. The query image and retrieval parameters are revised in the upper part, while the retrieved images are shown in the rest of the screen. Additional information is provided in tooltips.

for the next retrieval task (right click on the image and choose the "`retrieve`" item from the pop-up menu). The "`< input`" button returns to the input screen.

A comparison of different methods for texture description and similarity judgement can be enabled by adding the command line parameter "`--clones 2`", which creates two application windows. Subsequently, different methods can be selected in each application window. These windows are synchronised so that the retrieval with the same query image is performed simultaneously in both windows, but the results differ according to the selected methods.

A captured video of demonstration usage is available online.[2]

System requirements

Our CBIR application requires a computer with installed Java Runtime Environment (JRE) version 6 or later, which is freely available for download.[3] Moreover, the demonstration needs 300 MB of free disk space, 1 GHz processor, and 0.5 GB of RAM. The demonstration application was tested on operation systems GNU Linux and Windows XP.

---

[2]`http://ro.utia.cas.cz/demos/civr-demo.html`
[3]`http://java.com`
[4]Windows is a registered trademark of Microsoft Corporation.

Bibliography

T. Ahonen, J. Matas, C. He, and M. Pietikäinen. Rotation invariant image description with local binary pattern histogram Fourier features. In A.-B. Salberg, J. Y. Hardeberg, and R. Jenssen, editors, *Proceedings of the 16th Scandinavian Conference on Image Analysis, SCIA 2009*, volume 5575 of *Lecture Notes in Computer Science*, pages 61–70. Springer-Verlag, 2009. ISBN 978-3-642-02229-6. doi: 10.1007/978-3-642-02230-2_7.

S. Alvarez, A. Salvatella, M. Vanrell, and X. Otazu. Perceptual color texture codebooks for retrieving in highly diverse texture datasets. In *Proceedings of the 20th International Conference on Pattern Recognition, ICPR 2010*, pages 866–869. IEEE, 23-26 August 2010. doi: 10.1109/ICPR.2010.218.

A. Amir, J. Argillander, M. Campbell, A. Haubold, G. Iyengar, S. Ebadollahi, F. Kang, M. R. Naphade, A. P. Natsev, J. R. Smith, J. Tešić, and T. Volkmer. IBM research TRECVID-2005 video retrieval system. In *NIST TRECVID-2005 Workshop*, Gaithersburg, November 2005.

P. Andrey and P. Tarroux. Unsupervised segmentation of Markov random field modeled textured images using selectionist relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):252–262, March 1998. doi: 10.1109/34.667883.

I. Ar and Y. Akgul. A generic system for the classification of marble tiles using Gabor filters. In *Proceedings of the 23rd International Symposium on Computer and Information Sciences, ISCIS 2008*, pages 1–6, 27-29 October 2008. doi: 10.1109/ISCIS.2008.4717915.

Y. H. Bai, C. Park, and Y. Choe. Relative advantage of touch over vision in the exploration of texture. In *Proceeding of the 19th International Conference on Pattern Recognition, ICPR 2008*, pages 1–4. IEEE, 8-11 December 2008. doi: 10.1109/ICPR.2008. 4760961. URL `http://figment.cse.usf.edu/~sfefilat/data/papers/MoAT4.2. pdf#page=1#page=1`.

M. Batko, F. Falchi, C. Lucchese, D. Novak, R. Perego, F. Rabitti, J. Sedmidubsky, and P. Zezula. Building a web-scale image similarity search system. *Multimedia Tools and Applications*, 47(3):599–629, 2010. ISSN 1380-7501. doi: 10.1007/s11042-009-0339-z.

Bibliography

R. Bock, J. Meier, G. Michelson, L. G. Nyúl, and J. Hornegger. Classifying glaucoma with image-based features from fundus photographs. In *Proceedings of the 29th DAGM conference on Pattern recognition*, pages 355–364. Springer-Verlag, 2007. ISBN 978-3-540-74933-2. doi: 10.1007/978-3-540-74936-3_36.

A. Bosch, A. Zisserman, and X. Muoz. Scene classification using a hybrid generative/discriminative approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4):712–727, April 2008. ISSN 0162-8828. doi: 10.1109/TPAMI.2007.70716.

A. Bovik. Analysis of multichannel narrow-band filters for image texture segmentation. *IEEE Transactions on Signal Processing*, 39(9):2025–2043, 1991. doi: 10.1109/78.134435.

P. Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover Publications, New York, 1966.

G. J. Burghouts and J.-M. Geusebroek. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 113(1):48–62, 2009a. doi: 10.1016/j.cviu.2008.07.003.

G. J. Burghouts and J.-M. Geusebroek. Material-specific adaptation of color invariant features. *Pattern Recognition Letters*, 30:306–313, 2009b. doi: 10.1016/j.patrec.2008.10.005. URL http://www.science.uva.nl/research/publications/2009/BurghoutsPRL2009.

P. J. Burt. Fast algorithms for estimating local image properties. *Computer Vision, Graphics, and Image Processing*, 21(3):368–382, 1983. ISSN 0734-189X. doi: 10.1016/S0734-189X(83)80049-8.

B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. In *Proceedings of the 10th IEEE International Conference on Computer Vision, ICCV 2005*, pages 1597–1604. IEEE, 17-21 October 2005. ISBN 0-7695-2334-X. doi: 10.1109/ICCV.2005.54.

C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

M. Chantler. Why illuminant direction is fundamental to texture analysis. *IEE Proceedings - Vision, Image and Signal Processing*, 142(4):199–206, August 1995. ISSN 1350-245X. doi: 10.1049/ip-vis:19952065.

M. J. Chantler, M. Petrou, A. Penirsche, M. Schmidt, and G. McGunnigle. Classifying surface texture while simultaneously estimating illumination direction. *International Journal of Computer Vision*, 62(1-2):83–96, 2005. doi: 10.1007/s11263-005-4636-3.

154

S. A. Chatzichristofis, K. Zagoriz, Y. S. Boutalis, and N. Papamarkos. Accurate image retrieval based on compact composite descriptor and relevance feedback information. *Iternational Journal of Pattern Recognition and Artificial Intelligence*, 24(2):207–244, 2010. doi: 10.1142/S0218001410007890.

H. F. Chen, P. N. Belhumeur, and D. W. Jacobs. In search of illumination invariants. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2000*, volume 1, pages 254–261. IEEE, 13-15 June 2000. doi: 10.1109/CVPR. 2000.855827.

Y. Chen, J. Z. Wang, and R. Krovetz. CLUE: Cluster-based retrieval of images by unsupervised learning. *IEEE Transactions on Image Processing*, 14(8):1187–1201, August 2005. doi: 10.1109/TIP.2005.849770.

D. Chetverikov and R. M. Haralick. Texture anisotropy, symmetry, regularity: Recovering structure and orientation from interaction maps. In *Proceedings of th 6th British Machine Vision Conference, BMVC 1995*, pages 57–66, 1995. URL `http://www.bmva.org/bmvc/1995/bmvc-95-005.pdf`.

C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. ISSN 0885-6125. doi: 10.1023/A:1022627411411.

N. Cristianini and J. S. Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, 2000. URL `http://www.support-vector.net/`.

K. Dana, B. Van-Ginneken, S. Nayar, and J. Koenderink. Reflectance and Texture of Real World Surfaces. *ACM Transactions on Graphics*, 18(1):1–34, 1999. ISSN 0730-0301. doi: 10.1145/300776.300778.

R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, 2008. ISSN 0360-0300. doi: 10.1145/1348246.1348248.

H. Deng and D. A. Clausi. Gaussian MRF rotation-invariant features for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7): 951–955, July 2004. ISSN 0162-8828. doi: 10.1109/TPAMI.2004.30.

T. Deselaers, D. Keysers, and H. Ney. Features for image retrieval: an experimental comparison. *Information Retrieval*, 11(2):77–107, 2008. ISSN 1386-4564. doi: 10. 1007/s10791-007-9039-3.

C. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. In *Proceedings of IEEE Bioinformatics Conference, CSB 2003*, pages 523–528, 11-14 August 2003. doi: 10.1109/CSB.2003.1227396.

A. Diplaros, T. Gevers, and I. Patras. Combining color and shape information for illumination-viewpoint invariant object recognition. *IEEE Transactions on Image Processing*, 15(1):1–11, January 2006. ISSN 1057-7149. doi: 10.1109/TIP.2005.860320.

Bibliography

O. Drbohlav and M. Chantler. Illumination-invariant texture classification using single training images. In *Proceedings of the 4th international workshop on texture analysis and synthesis, Texture 2005*, pages 31–36, 2005. URL `http://www.macs.hw.ac.uk/texture2005/programme/papers/049.pdf`.

D. M. Drucker and G. K. Aguirre. Different spatial scales of shape similarity representation in lateral and ventral loc. *Cerebral Cortex*, 19(10), October 2009. ISSN 1047-3211. doi: 10.1093/cercor/bhn244.

D. M. Drucker, W. T. Kerr, and G. K. Aguirre. Distinguishing conjoint and independent neural tuning for stimulus features with fMRI adaptation. *Journal of Neurophysiology*, 101(6):3310–3324, June 2009. ISSN 0022-3077. doi: 10.1152/jn.91306.2008.

R. Duda, P. E. Hart, and D. Stork. *Pattern Classification*. John Wiley and Sons, 2001. ISBN 0-471-05669-3.

S. Fazekas, T. Amiaz, D. Chetverikov, and N. Kiryati. Dynamic texture detection based on motion analysis. *International Journal of Computer Vision*, 82(1):48–63, 2009. ISSN 0920-5691. doi: 10.1007/s11263-008-0184-y.

J. Filip and M. Haindl. Bidirectional texture function modeling: A state of the art survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1921–1940, 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.246.

J. Filip, M. Chantler, P. Green, and M. Haindl. A psychophysically validated metric for bidirectional texture data reduction. *ACM Transactions on Graphics*, 27(5):138, December 2008. doi: 10.1145/1457515.1409091. URL `http://staff.utia.cas.cz/filip/projects/pertex`.

J. Filip, M. J. Chantler, and M. Haindl. On uniform resampling and gaze analysis of bidirectional texture functions. *ACM Transactions on Applied Perception (TAP)*, 6 (3):1–15, 2009. ISSN 1544-3558. doi: 10.1145/1577755.1577761.

J. Filip, P. Vacha, M. Haindl, and P. R. Green. A psychophysical evaluation of texture degradation descriptors. In E. R. Hancock, R. C. Wilson, T. Windeatt, I. Ulusoy, and F. Escolano, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 6218 of *Lecture Notes in Computer Science*, pages 423–433. Springer-Verlag, 2010. doi: 10.1007/978-3-642-14980-1_41.

G. D. Finlayson. *Coefficient color constancy*. PhD thesis, Simon Fraser University, 1995.

G. Finlyason and R. Xu. Illuminant and gamma comprehensive normalisation in log*RGB* space. *Patterm Recognition Letters*, 24:1679–1690, 2002. doi: 10.1016/S0167-8655(02)00324-0.

R. W. Fleming, R. O. Dror, and E. H. Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5):347–368, 2003. ISSN 1534-7362. doi: 10.1167/3.5.3.

M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: the QBIC system. *Computer*, 28(9):23–32, September 1995. ISSN 0018-9162. doi: 10.1109/2.410146.

J. Flusser and T. Suk. Rotation moment invariants for recognition of symmetric objects. *IEEE Transactions on Image Processing*, 15(12):3784–3790, 2006. doi: 10.1109/TIP. 2006.884913.

J. Flusser, T. Suk, and B. Zitová. *Moments and Moment Invariants in Pattern Recognition*. Wiley, Chichester, 2009.

J. Gazárek, J. Jan, and R. Kolář. Detection of neural fibre layer in retina images via textural analysis. In *Proceedings of Analysis of Biomedical Signals and Images, Biosignal 2008*, pages 1–7, 2008.

J.-M. Geusebroek and A. W. M. Smeulders. A six-stimulus theory for stochastic texture. *International Journal of Computer Vision*, 62:7–16, 2005. doi: 10.1007/ s11263-005-4632-7. URL http://www.science.uva.nl/research/publications/ 2005/GeusebroekIJCV2005a.

J.-M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts. Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12): 1338–1350, December 2001. ISSN 0162-8828. doi: 10.1109/34.977559.

J.-M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and T. Gevers. Color constancy from physical principles. *Pattern Recognition Letters*, 24(11):1653–1662, 2003. doi: 10.1016/S0167-8655(02)00322-7. URL http://www.science.uva.nl/ research/publications/2003/GeusebroekPRL2003.

G. L. Gimel'Farb. *Image Textures and Gibbs Random Fields*. Kluwer, 1999.

M. Guthe, G. Müller, M. Schneider, and R. Klein. BTF-CIELab: a perceptual difference measure for quality assessment and compression of BTFs. *Computer Graphics Forum*, 28(1):101–113, 2009.

E. Hadjidemetriou, M. Grossberg, and S. Nayar. Multiresolution histograms and their use for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):831–847, July 2004. ISSN 0162-8828. doi: 10.1109/TPAMI.2004.32.

M. Haindl. Texture synthesis. *CWI Quarterly*, 4(4):305–331, December 1991.

M. Haindl. Texture segmentation using recursive Markov random field parameter estimation. In K. Bjarne and J. Peter, editors, *Proceedings of the 11th Scandinavian Conference on Image Analysis, SCIA 1999*, pages 771–776. Pattern Recognition Society of Denmark, June 1999. ISBN 87-88306-42-9.

Bibliography

M. Haindl and V. Havlíček. Prototype Implementation of the Texture Analysis Objects. Technical Report 1939, ÚTIA AV ČR, Praha, 1997.

M. Haindl and S. Mikeš. Unsupervised texture segmentation using multispectral modelling approach. In Y. Tang, S. Wang, D. Yeung, H. Yan, and G. Lorette, editors, *Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006*, pages 203–206. IEEE, 20-24 August 2006. doi: 10.1109/ICPR.2006.1148.

M. Haindl and S. Mikeš. Texture segmentation benchmark. In B. Lovell, D. Laurendeau, and R. Duin, editors, *Proceedings of the 19th International Conference on Pattern Recognition, ICPR 2008*, pages 1–4. IEEE, 8-11 December 2008. ISBN 978-1-4244-2174-9. doi: 10.1109/ICPR.2008.4761118.

M. Haindl and S. Mikeš. The Prague texture segmentation datagenerator and benchmark. URL `http://mosaic.utia.cas.cz/`.

M. Haindl and S. Mikeš. Model-based texture segmentation. In A. Campilho and M. Kamel, editors, *Image Analysis and Recognition*, volume 3212 of *Lecture Notes in Computer Science*, pages 306–313. Springer-Verlag, 2004. ISBN 3-540-23240-0. doi: 10.1007/b100438.

M. Haindl and S. Šimberová. *Theory & Applications of Image Analysis*, chapter A Multispectral Image Line Reconstruction Method, pages 306–315. World Scientific Publishing Co., Singapore, 1992. ISBN 981-02-0945-2.

M. Haindl and S. Šimberová. A multi-model image line reconstruction. In V. Hlaváč and R. Šára, editors, *Proceeding of the 6th International Conference on Computer Analysis of Images and Patterns, CAIP 1995*, volume 970 of *Lecture Notes in Computer Science*, pages 735–740. Springer-Verlag, 1995. ISBN 3-540-60268-2. doi: 10.1007/3-540-60268-2_373.

M. Haindl and P. Vacha. Illumination invariant texture retrieval. In Y. Tang, S. Wang, D. Yeung, H. Yan, and G. Lorette, editors, *Proceeding of the 18th International Conference on Pattern Recognition, ICPR 2006*, volume 3, pages 276–279. IEEE, 20-24 August 2006. doi: 10.1109/ICPR.2006.678.

M. Haindl, S. Mikes, and P. Vacha. Illumination invariant unsupervised segmenter. In *Proceeding of IEEE International Conference on Image Processing, ICIP 2009*, pages 4025–4028. IEEE, 7-10 November 2009. doi: 10.1109/ICIP.2009.5413753.

G. M. Haley and B. S. Manjunath. Rotation-invariant texture classification using a complete space-frequency model. *IEEE Transactions on Image Processing*, 8(2):255–269, February 1999. doi: 10.1109/83.743859.

G. M. Haley and B. S. Manjunath. Rotation-invariant texture classification using modified Gabor filters. In *Proceeding of IEEE International Conference on Image Processing, ICIP 1995*, volume 1, pages 262–265. IEEE, 23-26 October 1995. doi: 10.1109/ICIP.1995.529696.

R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, May 1979. ISSN 0018-9219. doi: 10.1109/PROC.1979.11328.

V. Havran, J. Filip, and K. Myszkowski. Bidirectional texture function compression based on multi-level vector quantization. *Computer Graphics Forum*, 29(1):175–190, 2010. doi: 10.1111/j.1467-8659.2009.01585.x.

Y. Hayashi, T. Nakagawa, Y. Hatanaka, A. Aoyama, M. Kakogawa, T. Hara, H. Fujita, and T. Yamamoto. Detection of retinal nerve fiber layer defects in retinal fundus images using Gabor filtering. volume 6514, page 65142Z. SPIE, 2007. doi: 10.1117/12.710181.

C. He, T. Ahonen, and M. Pietikainen. A Bayesian local binary pattern texture descriptor. In *Proceeding of the 19th International Conference on Pattern Recognition, ICPR 2008*, pages 1–4. IEEE, 8-11 December 2008. doi: 10.1109/ICPR.2008.4761100. URL `http://figment.cse.usf.edu/~sfefilat/data/papers/MoCT1.3.pdf#page=1#page=1`.

G. Healey and L. Wang. The illumination-invariant recognition of color texture. In *Proceedings of the 5th IEEE International Conference on Computer Vision, ICCV 1995*, pages 128–133. IEEE, 20-23 June 1995. doi: 10.1109/ICCV.1995.466796.

M. A. Hoang, J.-M. Geusebroek, and A. W. Smeulders. Color texture measurement and segmentation. *Signal Processing*, 85(2):265–275, 2005. doi: 10.1016/j.sigpro.2004.10.009.

J. Huang, S. Kumar, M. Mitra, W. Zhu, and R. Zabih. Image indexing using color correlograms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 1997*, pages 762–768. IEEE, 17-19 June 1997. doi: 10.1109/CVPR.1997.609412.

R. Iferroudjene, K. A. Meraim, and A. Belouchrani. A new jacobi-like method for joint diagonalization of arbitrary non-defective matrices. *Applied Mathematics and Computation*, 211(2):363 – 373, 2009. ISSN 0096-3003. doi: 10.1016/j.amc.2009.01.045.

D. Jacobs, P. Belhumeur, and R. Basri. Comparing images under variable illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 1998*, pages 610–617. IEEE, 23-25 June 1998. doi: 10.1109/CVPR.1998.698668.

D. W. Jacobs, D. Weinshall, and Y. Gdalyahu. Classification with nonmetric distances: Image retrieval and class representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6):583–600, June 2000. ISSN 0162-8828. doi: 10.1109/34.862197.

K. Jafari-Khouzani and H. Soltanian-Zadeh. Radon transform orientation estimation for rotation invariant texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):1004–1008, 2005. doi: 10.1109/TPAMI.2005.126.

Bibliography

A. Jain and G. Healey. A multiscale representation including opponent colour features for texture recognition. *IEEE Transactions on Image Processing*, 7(1):124–128, January 1998. ISSN 1057-7149. doi: 10.1109/83.650858.

J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6): 1233–1258, 1987.

B. Julesz. Early vision and focal attention. *Reviews of Modern Physics*, 63(3):735–772, July 1991. doi: 10.1103/RevModPhys.63.735.

B. Julesz. Visual pattern discrimination. *IRE Transactions on Information Theory*, pages 84–92, February 1962. doi: 10.1109/TIT.1962.1057698.

B. Julesz, E. Gilbert, and J. Victor. Visual discrimination of textures with identical third-order st atistics. *Biological Cybernetics*, 31:137–140, 1978. doi: 10.1007/BF00336998.

J.-K. Kamarainen, V. Kyrki, and H. Kälviäinen. Invariance properties of Gabor filter-based features-overview and applications. *IEEE Transactions on Image Processing*, 15(5):1088–1099, May 2006. ISSN 1057-7149. doi: 10.1109/TIP.2005.864174.

R. L. Kashyap and A. Khotanzad. A model-based method for rotation invariant texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(4):472–481, July 1986. ISSN 0162-8828. doi: 10.1109/TPAMI.1986.4767811.

R. Kolář and J. Jan. Detection of glaucomatous eye via color fundus images using fractal dimensions. *Radioengineering*, 17(3):109–114, September 2008.

R. Kolář and P. Vacha. Texture analysis of the retinal nerve fiber layer in fundus images via Markov random fields. In O. Dössel and W. C. Schlegel, editors, *World Congress on Medical Physics and Biomedical Engineering*, volume 25/11 of *IFMBE Proceedings*, pages 247–250. Springer-Verlag, 7-12 September 2009. ISBN 978-3-642-03890-7. doi: 10.1007/978-3-642-03891-4_66.

R. Kolář, D. Urblánek, and J. Jan. Texture based discrimination of normal and glaucomatous retina. In *Proceedings of Analysis of Biomedical Signals and Images, Biosignal 2008*, pages 1–5, 2008.

A. Laine and J. Fan. Texture classification by wavelet packet signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1186–1191, November 1993. ISSN 0162-8828. doi: 10.1109/34.244679.

S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8): 1265–1278, August 2005. ISSN 0162-8828. doi: 10.1109/TPAMI.2005.151.

S. Lee, K. Kim, J. Seo, D. Kim, H. Chung, K. Park, and H. Kim. Automated quantification of retinal nerve fiber layer atrophy in fundus photograph. In *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEMBS 2004*, volume 1, pages 1241–1243, 1-5 September 2004. doi: 10.1109/IEMBS.2004.1403394.

Y.-B. Lee, U. Park, and A. K. Jain. Pill-id: Matching and retrieval of drug pill imprint images. In *Proceedings of the 20th International Conference on Pattern Recognition, ICPR 2010*, pages 2632–2635. IEEE, 23-26 August 2010. doi: 10.1109/ICPR.2010.645.

T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1): 29–44, 2001. doi: 10.1023/A:1011126920638.

W. H. Leung and T. Chen. Trademark retrieval using contour-skeleton stroke classification. In *Proceedings of IEEE International Conference on Multimedia and Expo, ICME 2002*, volume 2, pages 517–520, 2002. doi: 10.1109/ICME.2002.1035662.

M. Lew and N. Huijsmans. Information theory and face detection. In *Proceedings of the 13th International Conference on Pattern Recognition, ICPR 1996*, volume 3, pages 601–605. IEEE, 25-29 August 1996. doi: 10.1109/ICPR.1996.547017.

M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2(1):1–19, 2006. ISSN 1551-6857. doi: 10.1145/1126004.1126005.

S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 2009. ISBN 9781848002784.

S. Liao, M. W. K. Law, and A. C. S. Chung. Dominant local binary patterns for texture classification. *IEEE Transactions on Image Processing*, 18(5):1107–1118, May 2009. doi: 10.1109/TIP.2009.2015682.

S. Liapis and G. Tziritas. Color and texture image retrieval using chromaticity histograms and wavelet frames. *IEEE Transactions on Multimedia*, 6(5):676–686, October 2004. ISSN 1520-9210. doi: 10.1109/TMM.2004.834858.

F. Liu. Image and video modelling and understanding. In *Fourth DELOS workshop Image Indexing and Retrieval*, pages 19–28. European Research Consortium for Informatics and Mathematics ERCIM, August 1997.

F. Liu and R. W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):722–733, July 1996. ISSN 0162-8828. doi: 10.1109/34.506794.

H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):491–502, 2005. ISSN 1041-4347. doi: 10.1109/TKDE.2005.66.

D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. doi: 10.1023/B:VISI.0000029664.99615.94.

W. Y. Ma and B. S. Manjunath. Texture features and learning similarity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 1996*, pages 425–430. IEEE, 18-20 June 1996. doi: 10.1109/CVPR.1996.517107.

T. Maenpaa, M. Pietikainen, and J. Viertola. Separating color and pattern information for color texture discrimination. In *Proceeding of the 16th International Conference on Pattern Recognition, ICPR 2002*, volume 1, pages 668–671. IEEE, 11-15 August 2002. doi: 10.1109/ICPR.2002.1044840.

B. Manjunath and R. Chellapa. Unsupervised texture segmentation using Markov random field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:478–482, 1991.

B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715, June 2001. ISSN 1051-8215. doi: 10.1109/76.927424.

B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996. ISSN 0162-8828. doi: 10.1109/34.531803.

R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Visible difference predicator for high dynamic range images. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 3, pages 2763–2769, 10-13 October 2004. doi: 10.1109/ICSMC.2004.1400750.

J. Mao and A. K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, 25(2):173–188, 1992. ISSN 0031-3203. doi: 10.1016/0031-3203(92)90099-5.

D. H. Marimont and B. A. Wandell. Linear models of surface and illuminant spectra. *Journal of the Optical Society of America*, 9:1905–1913, 1992.

M. Masotti and R. Campanini. Texture classification using invariant ranklet features. *Pattern Recognition Letters*, 29(14):1980–1986, 2008. ISSN 0167-8655. doi: 10.1016/j.patrec.2008.06.017.

G. McGunnigle and M. Chantler. Rotation invariant classification of rough surfaces. *IEE Proceedings - Vision, Image and Signal Processing*, 146(6):345–352, December 1999. ISSN 1350-245X. doi: 10.1049/ip-vis:19990707.

J. Meseth, G. Müller, and R. Klein. Preserving realism in real-time rendering of bidirectional texture functions. In *OpenSG Symposium 2003*, pages 89–96. Eurographics Association, Switzerland, April 2003.

S. Mikeš. *Image Segmentation*. PhD thesis, Charles University in Prague, Prague, 2010.

K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005. ISSN 0162-8828. doi: 10.1109/TPAMI.2005.188.

A. Mojsilovic, J. Kovacevic, J. Hu, R. J. Safranek, and S. K. Ganapathy. Matching and retrieval based on the vocabulary and grammar of color patterns. *IEEE Transactions on Image Processing*, 9(1):38–54, 2000. doi: 10.1109/83.817597.

A. Monadjemi. *Towards efficient texture classification and abnormality detection*. PhD thesis, University of Bristol, 2004.

T. Mäenpää and M. Pietikäinen. Classification with color and texture: jointly or separately? *Pattern Recognition*, 37(8):1629–1640, 2004. ISSN 0031-3203. doi: 10.1016/j.patcog.2003.11.011.

F. Nicodemus, R. J.C., J. Hsia, I. Ginsburg, and T. Limperis. Geometrical considerations and nomenclature for reflectance. *NBS Monograph 160, National Bureau of Standards, U.S. Department of Commerce, Washington, D. C.*, pages 1–52, October 1977.

P. P. Ohanian and R. C. Dubes. Performance evaluation for four classes of textural features. *Pattern Recognition*, 25(8):819–833, 1992. ISSN 0031-3203. doi: 10.1016/0031-3203(92)90036-I.

T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996. ISSN 0031-3203. doi: 10.1016/0031-3203(95)00067-4.

T. Ojala, T. Mäenpää, M. Pietikäinen, J. Viertola, J. Kyllönen, and S. Huovinen. Outex - new framework for empirical evaluation of texture analysis algorithms. In *Proceeding of the 16th International Conference on Pattern Recognition, ICPR 2002*, volume 1, pages 701–706. IEEE, 11-15 August 2002a. doi: 10.1109/ICPR.2002.1044854.

T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, Jul 2002b. ISSN 0162-8828. doi: 10.1109/TPAMI.2002.1017623.

V. Ojansivu, E. Rahtu, and J. Heikkila. Rotation invariant local phase quantization for blur insensitive texture analysis. In *Proceeding of the 19th International Conference on Pattern Recognition, ICPR 2008.*, pages 1–4. IEEE, 8-11 December 2008. doi: 10.1109/ICPR.2008.4761377. URL `http://figment.cse.usf.edu/~sfefilat/data/papers/ThCT4.3.pdf#page=1#page=1`.

Bibliography

S. Padilla, O. Drbohlav, P. R. Green, A. Spence, and M. J. Chantler. Perceived roughness of 1/fˆbeta noise surfaces. *Vision Research*, 48(17):1791–1797, 2008. ISSN 0042-6989. doi: 10.1016/j.visres.2008.05.015.

D. Panjwani and G. Healey. Markov random field models for unsupervised segmentation of textured color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):939–954, October 1995. ISSN 0162-8828. doi: 10.1109/34.464559.

G. Paschos, I. Radev, and N. Prabakar. Image content-based retrieval using chromaticity moments. *IEEE Transactions on Knowledge and Data Engineering*, 15(5):1069–1072, September-October 2003. ISSN 1041-4347. doi: 10.1109/TKDE.2003.1232264.

E. Pekalska and R. P. W. Duin. Dissimilarity representations allow for building good classifiers. *Pattern Recognition Letters*, 23(8):943–956, 2002. ISSN 0167-8655. doi: 10.1016/S0167-8655(02)00024-7.

E. Pekalska, R. P. W. Duin, and P. Paclík. Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*, 39(2):189–208, 2006. ISSN 0031-3203. doi: 10.1016/j.patcog.2005.06.012.

E. Peli, T. R. Hedges, and B. Schwartz. Computer measurement of retinal nerve-fiber layer striations. *Applied Optics*, 28(6):1128–1134, March 1989. doi: 10.1364/AO.28.001128.

H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, August 2005. ISSN 0162-8828. doi: 10.1109/TPAMI.2005.159.

R. Phan and D. Androutsos. Content-based retrieval of logo and trademarks in unconstrained color image databases using color edge gradient co-occurrence histograms. *Computer Vision and Image Understanding*, 114(1):66–84, 2010. ISSN 1077-3142. doi: 10.1016/j.cviu.2009.07.004.

M. Pietikäinen, T. Maenpaa, and J. Viertola. Color texture classification with color histograms and local binary patterns. In *Proceedings of the 2nd Workshop on Texture Analysis in Machine Vision*, pages 109–112, June 2002. URL `http://www.cee.hw.ac.uk/~texture2002/papers/ab002.pdf`.

J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–70, 2000. ISSN 0920-5691. doi: 10.1023/A:1026553619983.

C.-M. Pun and M.-C. Lee. Log-polar wavelet energy signatures for rotation and scale invariant texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):590–603, May 2003. ISSN 0162-8828. doi: 10.1109/TPAMI.2003.1195993.

164

T. Randen and J. H. Husoy. Multichannel filtering for image texture segmentation. *Optical Engineering*, 33(8):2617–2625, 1994. ISSN 0091-3286. doi: 10.1117/12.177115.

T. Randen and J. H. Husøy. Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):291–310, April 1999. ISSN 0162-8828. doi: 10.1109/34.761261.

A. R. Rao and G. L. Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Research*, 36(11):1649–1669, June 1996. doi: 10.1016/0042-6989(95)00202-2.

G. Ravichandran and M. M. Trivedi. Circular-Mellin features for texture segmentation. *IEEE Transactions on Image Processing*, 4(12):1629–1640, 1995. doi: 10.1109/83.475513.

T. R. Reed and J. M. H. du Buf. A review of recent texture segmentation and feature extraction techniques. *CVGIP–Image Understanding*, 57(3):359–372, May 1993. doi: 10.1006/ciun.1993.1024.

C. J. Rivero-Moreno and S. Bres. Texture feature extraction and indexing by Hermite filters. In *Proceeding of the 17th International Conference on Pattern Recognition, ICPR 2004*, volume 1, pages 684–687. IEEE, 23-26 August 2004. doi: 10.1109/ICPR.2004.1334266.

S. Santini and R. Jain. Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):871–883, September 1999. ISSN 0162-8828. doi: 10.1109/34.790428.

F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In *Proceedings of the 8th IEEE International Conference on Computer Vision, ICCV 2001*, volume 2, pages 636–643. IEEE, 7-14 July 2001. doi: 10.1109/ICCV.2001.937686.

L. Semler and L. Dettori. Curvelet-based texture classification of tissues in computed tomography. In *Proceeding of IEEE International Conference on Image Processing, ICIP 2006*, pages 2165–2168. IEEE, 8-11 October 2006. doi: 10.1109/ICIP.2006.312873.

S. A. Shafer. Using color to seperate reflection components. *COLOR research and application*, 10(4):210–218, Winter 1985. doi: 10.1002/col.5080100409.

J. D. J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 81(1):2–23, 2009. doi: 10.1007/s11263-007-0109-1.

E. Simoncelli. Statistical models for images: compression, restoration and synthesis. In *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems*

Computers, 1997, volume 1, pages 673–678, 2-5 November 1997. doi: 10.1109/ACSSC.1997.680530.

E. Simoncelli and J. Portilla. Texture characterization via joint statistics of wavelet coefficient magnitudes. In *Proceeding of IEEE International Conference on Image Processing, ICIP 1998*, volume 1, pages 62–66. IEEE, 4-7 October 1998. doi: 10.1109/ICIP.1998.723417.

J. Sivic and A. Zisserman. Video data mining using configurations of viewpoint invariant regions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2004*, volume 1, pages 488–495. IEEE, 27 June - 2 July 2004. doi: 10.1109/CVPR.2004.1315071.

A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000. ISSN 0162-8828. doi: 10.1109/34.895972.

J. R. Smith and S.-F. Chang. VisualSEEk: a fully automated content-based image query system. In *Proceedings of the fourth ACM international conference on Multimedia, MULTIMEDIA 1996*, pages 87–98, New York, NY, USA, 1996. ACM. ISBN 0-89791-871-1. doi: 10.1145/244130.244151.

C. G. M. Snoek, K. E. A. van de Sande, O. de Rooij, B. Huurnink, J. van Gemert, J. R. R. Uijlings, J. He, X. Li, I. Everts, V. Nedovic, M. van Liempt, R. van Balen, M. de Rijke, J.-M. Geusebroek, T. Gevers, M. Worring, A. W. M. Smeulders, D. Koelma, F. Yan, M. A. Tahir, K. Mikolajczyk, and J. Kittler. The MediaMill TRECVID 2008 semantic video search engine. In P. Over, G. Awad, R. T. Rose, J. G. Fiscus, W. Kraaij, and A. F. Smeaton, editors, *TRECVID*. National Institute of Standards and Technology (NIST), 2008. URL `http://www-nlpir.nist.gov/projects/trecvid/`.

P. Somol, J. Novovičová, and P. Pudil. *Pattern Recognition, Recent Advances*, chapter Efficient Feature Subset Selection and Subset Size Optimization, pages 75–98. In-Teh, Vukovar, Croatia, 2010. ISBN 978-953-7619-90-9. URL `http://sciyo.com/books/show/title/pattern-recognition-recent-advances`.

D. M. Squire, W. Müller, H. Müller, and T. Pun. Content-based query of image databases: inspirations from text retrieval. *Pattern Recognition Letters*, 21(13-14):1193–1198, 2000. ISSN 0167-8655. doi: 10.1016/S0167-8655(00)00081-7. Selected Papers from The 11th Scandinavian Conference on Image Analysis.

M. A. Stricker and M. Orengo. Similarity of color images. volume 2420, pages 381–392. SPIE, 1995. doi: 10.1117/12.205308.

P.-H. Suen and G. Healey. The analysis and recognition of real-world textures in three dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):491–503, May 2000. ISSN 0162-8828. doi: 10.1109/34.857005.

A. T. Targhi, J.-M. Geusebroek, and A. Zisserman. Texture classification with minimal training images. In *Proceedings of the 19th International Conference on Pattern Recognition, ICPR 2008*, pages 1–4. IEEE, 8-11 December 2008. doi: 10.1109/ICPR.2008. 4761388. URL `http://figment.cse.usf.edu/~sfefilat/data/papers/TuAT1.6.pdf#page=1#page=1`.

B. W. Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):1–17, November 2007. ISSN 1534-7362. doi: 10.1167/7.14.4.

N. J. Turro. *Modern Molecular Photochemistry*. Benjamin/Cummings, Menlo Park, CA, USA, 1978.

A. Tuulonen, H. Alanko, P. Hyytinen, J. Veijola, T. Seppanen, and P. Airaksinen. Digital imaging and microtexture analysis of the nerve fiber layer. *Journal of Glaucoma*, 9 (1):5–9, February 2000. ISSN 1057-0829.

P. Vácha. Texture similarity measure. In J. Safrankova, editor, *WDS'05 Proceedings of Contributed Papers: Part I - Mathematics and Computer Sciences*, pages 47–52, Prague, 2005. Charles University in Prague Faculty of Mathematics and Physics, Matfyzpress.

P. Vacha and M. Haindl. Image retrieval measures based on illumination invariant textural MRF features. In N. Sebe and M. Worring, editors, *Proceedings of ACM International Conference on Image and Video Retrieval, CIVR 2007*, pages 448–454. ACM, 9-11 July 2007a. ISBN 978-1-59593-733-9. doi: 10.1145/1282280.1282346.

P. Vacha and M. Haindl. Demonstration of image retrieval based on illumination invariant textural MRF features. In *Proceedings of ACM International Conference on Image and Video Retrieval, CIVR 2007*, pages 135–137. ACM, 9-11 July 2007b. ISBN 978-1-59593-733-9. doi: 10.1145/1282280.1282305.

P. Vacha and M. Haindl. Illumination invariants based on Markov random fields. In *Proceedings of the 19th International Conference on Pattern Recognition, ICPR 2008*, pages 1–4. IEEE, 8-11 December 2008a. doi: 10.1109/ICPR.2008.4761375. URL `http://figment.cse.usf.edu/~sfefilat/data/papers/ThCT4.1.pdf#page=1#page=1`.

P. Vacha and M. Haindl. Illumination invariant and rotational insensitive textural representation. In *Proceeding of IEEE International Conference on Image Processing, ICIP 2009*, pages 1333–1336. IEEE, 7-10 November 2009. doi: 10.1109/ICIP.2009.5413578.

P. Vacha and M. Haindl. Natural material recognition with illumination invariant textural features. In *Proceedings of the 20th International Conference on Pattern Recognition, ICPR 2010*, pages 858–861. IEEE, 23-26 August 2010a. doi: 10.1109/ICPR.2010.216.

Bibliography

P. Vacha and M. Haindl. *Pattern Recognition, Recent Advances*, chapter Illumination Invariants Based on Markov Random Fields, pages 253–272. In-Teh, Vukovar, Croatia, 2010b. ISBN 978-953-7619-90-9. URL `http://sciyo.com/books/show/title/pattern-recognition-recent-advances`.

P. Vacha and M. Haindl. Content-based tile retrieval system. In E. R. Hancock, R. C. Wilson, T. Windeatt, I. Ulusoy, and F. Escolano, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 6218 of *Lecture Notes in Computer Science*, pages 434–443. Springer-Verlag, 2010c. doi: 10.1007/978-3-642-14980-1_42.

P. Vacha and M. Haindl. Interactive demonstration of combined illumination and rotational invariants., 2010d. URL `http://cbir.utia.cas.cz/rotinv/`.

P. Vacha and M. Haindl. Interactive demonstration of illumination invariants., 2008b. URL `http://cbir.utia.cas.cz/`.

P. Vacha and M. Haindl. Interactive demonstration of contend-based tile retrieval system., 2010e. URL `http://cbir.utia.cas.cz/tiles/`.

P. Vacha, M. Haindl, and T. Suk. Colour and rotation invariant textural features based on Markov random fields. (submitted).

K. van de Sande, T. Gevers, and C. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582 –1596, sep. 2010. ISSN 0162-8828. doi: 10.1109/TPAMI.2009.154.

J. van de Weijer, T. Gevers, and J.-M. Geusebroek. Edge and corner detection by photometric quasi-invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):625–630, 2005. doi: 10.1109/TPAMI.2005.75. URL `http://www.science.uva.nl/research/publications/2005/vandeWeijerTPAMI2005`.

M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1-2):61–81, 2005. doi: 10.1007/s11263-005-4635-4.

M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2032–2047, November 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.182.

F. M. Vilnrotter, R. Nevatia, and K. E. Price. Structural analysis of natural textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(1):76–89, January 1986. ISSN 0162-8828. doi: 10.1109/TPAMI.1986.4767754.

Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13 (4):600–612, April 2004. ISSN 1057-7149. doi: 10.1109/TIP.2003.819861.

168

C.-H. Wei, Y. Li, W.-Y. Chau, and C.-T. Li. Trademark image retrieval using synthetic features for describing global shape and interior structure. *Pattern Recognition*, 42(3): 386–394, 2009. ISSN 0031-3203. doi: 10.1016/j.patcog.2008.08.019.

Y. Xu, H. Ji, and C. Fermuller. A projective invariant for textures. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2006*, volume 2, pages 1932–1939. IEEE, 17-22 June 2006. doi: 10.1109/CVPR.2006.38.

J. Yang and M. Al-Rawi. Illumination invariant recognition of three-dimensional texture in color images. *Journal of Computer Science and Technology*, 20(3):378–388, 2005. ISSN 1000-9000. doi: 10.1007/s11390-005-0378-5.

J. Yellot, John I. Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture. *Journal of the Optical Society of America A*, 10(5):777–793, May 1993. ISSN 0740-3232. doi: 10.1364/JOSAA.10.000777.

G. Zhao and M. Pietikainen. 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, June 2007. ISSN 0162-8828. doi: 10.1109/ TPAMI.2007.1110.

Texture database ALOT.
URL `http://staff.science.uva.nl/~mark/ALOT/`.

Texture database Bonn BTF.
URL `http://btf.cs.uni-bonn.de`.

Texture database Brodatz.
URL `http://sipi.usc.edu/database/`.

Texture dataset CUReT.
URL `http://www.robots.ox.ac.uk/~vgg/research/texclass/`.

Texture database KTH-TIPS2.
URL `http://www.nada.kth.se/cvap/databases/kth-tips/`.

Texture database MeasTex.
URL `http://www.texturesynthesis.com/meastex/meastex.html`.

Texture database Microsoft Textiles.
URL `http://research.microsoft.com/en-us/um/people/antcrim/data_objrec/ msrc_materialsimagedatabase.zip`.

Texture database of Univeristy of Maryland.
URL `http://www.cfar.umd.edu/~fer/High-resolution-data-base/hr_ database.htm`.

Texture database Outex.
URL `http://www.outex.oulu.fi`.

Bibliography

Texture database PhoTex.
   URL `http://www.macs.hw.ac.uk/texturelab/resources/databases/Photex/`.

Texture database UIUC.
   URL `http://www-cvr.ai.uiuc.edu/ponce_grp/data/`.

Texture database VisTex.
   URL `http://vismod.media.mit.edu/vismod/imagery/VisionTexture/`.

Implementation LBP, LBP-HF.
   URL `http://www.ee.oulu.fi/mvg/page/downloads`.

170

Index