



HAL
open science

Contributions to image retrieval in the wavelet transform domain

Amani Chaker

► **To cite this version:**

Amani Chaker. Contributions to image retrieval in the wavelet transform domain. Signal and Image Processing. Ecole Supérieure des Communications de Tunis 2016. English. NNT : . tel-01361579

HAL Id: tel-01361579

<https://theses.hal.science/tel-01361579>

Submitted on 7 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

présentée pour obtenir le titre de

DOCTEUR

DE L'ÉCOLE SUPÉRIEURE DES COMMUNICATIONS DE TUNIS

Discipline : Technologies de l'Information et de la Communication

par

Amani CHAKER

Contributions à l'indexation d'images dans le domaine transformé en ondelettes

Soutenue le 14 Mars 2016 devant le jury d'examen composé de :

M. Kais Ouni	Professeur à l'ENICarthage	Président
M. Frédéric Dufaux	Directeur de recherche CNRS au LTCI, Telecom ParisTech	Rapporteur
Mlle. Azza Ould Zaid	Professeure à l'ISI	Rapporteuse
M. Slim Mhiri	Maître de Conférences à l'ENSI	Examineur
Mme Amel Benazza-Benyahia	Professeure à SUP'COM	Directrice de thèse

Acknowledgments

I would like to take advantage of this opportunity to thank all those who supported me during this experience.

First, I would like to thank my Ph.D advisor, Professor Amel Benazza for giving me the opportunity to do a Ph.D under her supervision and her valuable insights, patience and guidance. I will forever be thankful to my co-supervisor Dr. Mounir Kaaniche for his endless support, appreciable help, insightful discussions and motivation. I am also grateful to Dr. Marc Antonini for giving me the opportunity to conduct a part of my work at I3S laboratory under his supervision. I am very pleased to have the opportunity to work with them.

I am very grateful to the jury members for their interest in my work. I thank Professor Kais Ouni for acting as the chairman and Dr. Frédéric Dufaux and Professor Azza Ouled Zaid for having accepted to review this thesis and giving me valuable comments. I thank Dr. Slim Mhiri to participate in the jury as examiner.

I address my gratitude to all the members of COSIM laboratory, Sup'Com and I3S laboratory, Nice Sophia Antipolis University, for the enjoyable work atmosphere they created.

I would like to thank EU Erasmus Mundus Alyssa programme for the financial support to do a part of my work at I3S Lab, particularly all administrative members for their welcome, help and availability.

A thought goes also to my friends Karima, Ahlem and Mariem.

Thanks to my fiance for his support and encouragement.

Last, but not least, I want to express my deep gratitude to my parents and brothers for their love and constant support. I owe them the success of my studies, and this thesis is also theirs.

Abstract

This thesis addresses the problem of images indexing and retrieval in the wavelet transform domain. In particular, two major issues are considered: the indexing of stereo images and the impact of quantization in still image retrieval schemes.

In the first part, we propose novel retrieval approaches devoted to stereo images which integrate the disparity information with the visual contents of stereo images. In the first strategy, the two views are processed separately through a univariate model. An appropriate bivariate model is employed to exploit the cross-view dependencies in the second method. In the third strategy, we resort to a multivariate model to further capture the spatial dependencies of wavelet subbands.

In the second part, different strategies are designed to improve the drop of retrieval performances resulting from the quantization of database or query images. First, we propose to operate on the quantized coefficients by applying a processing step that aims at reducing the mismatch between the bitrates of the model and the query images. As an alternative, we propose to recover the statistical parameters of original wavelet coefficients directly from the quantized ones. Then, we investigate different quantization schemes and we exploit inherent properties of each one in order to design an efficient retrieval strategy.

Contents

Introduction	15
1 Useful basics	21
1.1 Introduction	21
1.2 Content-Based Image retrieval systems	21
1.2.1 Principe of CBIR systems	21
1.2.2 Image features	22
1.2.3 Retrieval performance evaluation methods	28
1.2.4 Applications of CBIR	30
1.3 Compression techniques	30
1.3.1 Transformation	31
1.3.2 Quantization and entropy coding	36
1.4 Conclusion	38
2 Stereo images and disparity map	39
2.1 Introduction	39
2.2 Stereoscopic imaging	39
2.2.1 Image acquisition	39
2.2.2 Homologous points	40
2.2.3 Epipolar rectification	42
2.3 Stereo matching problem	43
2.3.1 Disparity information	43
2.3.2 Difficulties in stereo matching process	44
2.3.3 Stereo matching constraints	45
2.4 Overview of stereo matching approaches	46

2.4.1	Local approaches	47
2.4.2	Global approaches	49
2.5	Conclusion	50
3	Novel wavelet-based retrieval methods for stereo images	51
3.1	Introduction	51
3.2	Conventional wavelet-based CBIR System	52
3.2.1	Wavelet distribution model	52
3.2.2	Parameter estimation	53
3.2.3	Similarity measure	55
3.2.4	Intuitive extension to SI retrieval	55
3.3	Retained dense disparity estimation	56
3.3.1	Motivation	56
3.3.2	Retained disparity estimation method	56
3.4	Proposed disparity-based retrieval approaches through a univariate model	60
3.4.1	Stereo images-based modeling	61
3.4.2	Stereo images and disparity-based modeling	62
3.5	Proposed disparity-based retrieval approaches through bivariate model	64
3.5.1	Motivation	64
3.5.2	Bivariate generalized Gaussian model	64
3.5.3	Improved disparity-based retrieval strategies	68
3.6	Proposed retrieval approach through SIRV modeling	71
3.6.1	Motivation	71
3.6.2	SIRV-based model	72
3.6.3	Retrieval approach	74
3.7	Experimental results	75
3.7.1	Experimental setup	75
3.7.2	Univariate model-based retrieval approaches	76
3.7.3	Bivariate model-based retrieval approaches	81
3.7.4	Multivariate model-based retrieval approach	83
3.8	Conclusion	84

4 Retrieval approaches for wavelet-based quantized images	87
4.1 Introduction	87
4.2 Impact of compression on image retrieval performance	88
4.2.1 Motivation	88
4.2.2 Compression effect on image retrieval performance	89
4.3 Adopted methodology	92
4.4 Contributions to improve retrieval performance of JPEG2000 compressed images	93
4.4.1 Re-compression approach	93
4.4.2 Statistical-based approach	94
4.5 Design of appropriate retrieval approaches for various quantization schemes	98
4.5.1 Retrieval with moment preserving quantizers	98
4.5.2 Retrieval approach with distribution preserving quantizers	100
4.6 Experimental results	104
4.7 Conclusion	112
Conclusion and future work	115
A Subgradient projections and optimization algorithm	119
A.1 Subgradient projections	119
A.2 Optimization algorithm	120
B Orthogonal Polynomials	121
Bibliography	125

List of Figures

1.1	CBIR system architecture.	22
1.2	Different classes of texture features reported in the literature.	25
1.3	One dimensional wavelet decomposition over 2 resolution levels.	33
1.4	2D separable filter bank.	34
1.5	Uniform quantizer with deadzone with step size Δ	36
2.1	Stereoscopic imaging system.	40
2.2	Epipolar geometry.	42
2.3	Epipolar rectification.	42
2.4	Relation between the disparity and depth.	44
2.5	Illustration of occlusion effects: the point X is visible in the right image and absent in the left one.	46
2.6	Correlation-based stereo matching method.	48
3.1	The obtained disparity maps for the “Shrub” stereo image using (c) block-based DE method and (d) the considered DDE method.	60
3.2	The obtained disparity maps for the “Teddy” stereo image using (c) block-based DE method and (d) the considered DDE method.	61
3.3	Stereo images and disparity-based modeling approach.	63
3.4	Modeling the distribution of the horizontal detail subband at the second resolution level for three examples of disparity maps using the GG model.	63
3.5	Empirical bivariate histogram of the horizontal wavelet coefficients of the left and right images (denoted here by $w^{(l)}$ and $w^{(r)}$)(in blue) fitted with a BGG density (in red) for four different stereo images as well as their resulting KS measures.	69

3.6	Definition of the multivariate vector $\tilde{\mathbf{w}}_j^{(r,l)}$ when considering 3 neighbors in each view.	75
3.7	Some samples of right images for different classes in the database. From top to bottom: pentagon, urban area, flowerpots, mountains, pyramid, buildings and tree.	77
3.8	Retrieval performance in terms of precision and recall of the univariate approaches.	80
3.9	Impact of the weight values on the retrieval performances of the GG-RL-GG-D approach.	80
3.10	Retrieval performance in terms of precision and recall of the bivariate approaches.	81
3.11	Retrieval performance in terms of precision and recall of the univariate and the bivariate approaches.	82
3.12	Comparison between the multivariate and bivariate approaches.	84
3.13	Impact of the neighborhood size on the retrieval performance.	85
4.1	Original Lena image, its JPEG2000 compressed version at 0.15 bpp, Histograms of the diagonal wavelet coefficients at the third resolution level for : (a) the original and (b) compressed version of Lena image and (c) the difference between the two histograms.	90
4.2	Precision versus recall obtained for a query uncompressed images, when model images are : (a) uncompressed, compressed at $R_M =$ (b) 1.5 bpp, (c) 1 bpp, (d) 0.8 bpp, (e) 0.5 bpp, (f) 0.25 bpp.	91
4.3	Precision versus recall obtained under compression for a query compressed at 0.5 bpp.	91
4.4	Methodology to improve retrieval performance of lossy compressed images.	92
4.5	Diagram of a transformation-based DPQ.	101
4.6	Histograms of the diagonal wavelet coefficients at the second resolution level for an image taken from the VisTex dataset at high, medium and low bitrates.	103
4.7	Block diagram of the proposed retrieval system under a DPQ-based encoding scheme.	104
4.8	Some examples of images for different databases.	105

4.9	Retrieval performance in terms of precision and recall for UQ-based compressed images.	106
4.10	Wavelet-based compressed images of the DPE approach and the recompression one UQ-AR.	107
4.11	Retrieval performance in terms of precision and recall when using the MPQ-based coding scheme, the UQ-AR and DPE approaches.	109
4.12	Retrieval performance in terms of precision and recall of the MPQ-based coding scheme when the feature vectors are defined by retaining the second order moments of (i): <i>all</i> the wavelet subbands, (ii): <i>only</i> the wavelet subbands where $N_j \geq 2$	110
4.13	Retrieval performance in terms of precision and recall for the MPQ-based coding scheme and DPQ-based coding scheme.	111
4.14	Examples of retrieval results from compressed STex database images using UQ-AR approach. Query image (compressed at 0.1bpp) is on the top left corner; all other images (compressed at 0.5 bpp) are ranked in the order of similarity with the query image from left to right, top to bottom.	112
4.15	Examples of retrieval results from compressed STex database images using MPQ scheme. Query image (compressed at 0.1bpp) is on the top left corner; all other images (compressed at 0.5 bpp) are ranked in the order of similarity with the query image from left to right, top to bottom.	113
4.16	Examples of retrieval results from compressed STex database images using DPQ scheme. Query image (compressed at 0.1bpp) is on the top left corner; all other images (compressed at 0.5 bpp) are ranked in the order of similarity with the query image from left to right, top to bottom.	113
4.17	Retrieval performance in terms of precision and recall of the DCT-based quantized VisTex database images for the different retrieval approaches under the: (a) UQ-AR, (b) MPQ and (c) DPQ.	114

List of Tables

3.1	The considered approaches for SI retrieval.	79
3.2	ANMRR results of the univariate model-based retrieval approaches.	80
3.3	ANMRR results of the univariate and bivariate approaches.	82
3.4	ANMRR results of the multivariate and bivariate approaches.	84

Introduction

Thesis context

The recent advances in computer technologies, the development of digital acquisition systems and the emergence of different data compression standards have resulted in a great demand for storage and retrieval of the involved huge multimedia data. Content Based Image Retrieval (CBIR) systems were found to be an efficient tool for automatically managing still images for large databases based only on their visual contents [Smeulders *et al.*, 2000]. Typically, color, shape and texture features can be used to describe such contents for the indexing process. In conventional CBIR systems, these features are often extracted in the spatial domain [Yong, Thomas, 1999; Datta *et al.*, 2006]. However, with the increased size of the generated data, it becomes mandatory to compress them by using some image compression standards such as JPEG [Wallace, 1991] and JPEG2000 [Rabbani, Joshi, 2002] respectively based on Discrete Cosine Transform (DCT) and Wavelet Transform (WT). Thus, today's CBIR systems tend to involve images not only in their original versions, but also in their compressed forms. When the database (DB) images are available only on their compressed form, the classical methods operating in the spatial domain can be also applied to retrieve the resulting compressed images, they will require the decoding of all the database images, by applying the inverse transform, which will result in a significant computational overhead for feature extraction, especially for large image archives. Therefore, in order to design fast retrieval methods, other works have been developed by performing the feature extraction step in the transform domain [Mezaris *et al.*, 2004; Wang *et al.*, 2015]. Moreover, an additional benefit can also be drawn by exploiting the intrinsic sparsity of the coefficients resulting from the underlying transform. The most known retrieval approaches operate in the WT domain where features are directly extracted from the wavelet coefficients [Voulgaris, Jiang, 2001; Mandal *et al.*, 1996; Do, Vetterli, 2002;

Sakji-Nsibi, Benazza-Benyahia, 2009].

This thesis is concerned with two major issues in CBIR operating in the wavelet transform. The first one is related to Stereo Images (SI) and the second one to the impact of quantization in the CBIR schemes.

More precisely, stereo images belong to the huge multimedia data that are continuously produced. Indeed, the recent advances in acquisition and display technologies have allowed the widespread use of stereovision in various application fields such as telepresence in videoconferences [Tzovaras *et al.*, 1999], medical imaging [Field *et al.*, 2009], 3D TV [Min *et al.*, 2009], computer games and advanced graphic application [Yahav *et al.*, 2007] and 3D geographic information systems [Yang *et al.*, 2011].

Stereoscopic image display offers a simple way of presenting the depth information in a real world scene. Indeed, the disparity information which corresponds to the displacement that exists between the corresponding pixels of the left and right images, allows to provide the 3D-depth information of the scene. The increasing interest in SIs has led to the constitution of huge image databases. For example, the on board stereoscopic instrument of SPOT5 covers areas of $120 \text{ km} \times 60 \text{ km}$ and a single view can be represented by 1200×1200 pixels [Poli *et al.*, 2004]. Moreover, a single view of a scene acquired by the IKONOS satellite corresponds to 360 MB every 3 or 4 days. Hence, there is a strong need for both managing and storing these large amounts of stereo data [Yoon *et al.*, 2012].

On the other side, it is worth pointing out that most of the mono-view developed methods are well adapted for losslessly encoded images (i.e unquantized transformed coefficients). Thus, a particular attention should be paid to the effect of the quantization operation in the context of lossy data compression. Indeed, some studies have focused on the influence of lossy compression on the effectiveness of different still image retrieval approaches [Guldogan *et al.*, 2003; Schaefer, 2008; Edmundson, Schaefer, 2012a]. It was shown that compression adversely affects the image retrieval performance especially when the compression ratio of the query and that of database images (called also model images) are very different. Therefore, designing efficient indexing methods that account for the quantization step as well as the difference of qualities between the query and model images becomes a real challenge.

Objectives and contributions

The objectives of this thesis are twofold. The first one is to propose new retrieval approaches devoted to the case of stereo images. The second one is to improve the retrieval performance of wavelet based compressed images and design efficient indexing approaches robust against the quantization effects. The main contributions of our work are the following:

- In the first part, we focus on the retrieval of stereo images, and propose new methods which combine the visual contents of the stereo data with their corresponding disparity information. In this respect, we resort to a statistical modeling of the WT coefficients to extract the salient features. While the two views are processed separately through a univariate modeling in the first method, the second one exploits the correlation between the views by resorting to a bivariate modeling. We propose also a novel approach which exploits simultaneously the spatial and cross-view dependencies of the stereo images through multivariate modeling. For instance, in order to exploit the stereo images redundancies in the second and third approaches, we will resort to a smooth and dense disparity map.
 - In the second part, we propose some efficient techniques to improve the retrieval performance of lossy compressed (i.e. quantized) images. Indeed, on one hand, we start by focusing on the standard uniform quantizer and design salient features from the resulting quantized coefficients. To this end, two approaches have been developed. The first one consists in applying a preprocessing step before the indexing step in order to reduce the difference of qualities between the query and database images. The second one aims at recovering the statistical parameters of the original wavelet coefficients directly from the quantized ones.
On the other hand, we investigate other quantizers with some interesting properties that can be efficiently exploited in the retrieval process. To this end, we resort to two specific quantization schemes which can preserve the moments and probability distributions of the original wavelet coefficients, respectively. After that, we propose efficient retrieval approaches well adapted to the inherent properties of each quantizer.
-

Thesis outline

In Chapter 1, we first introduce the principle of content based image retrieval system. Then, we give a brief overview on image features reported in the literature for indexing purpose. We review the most important concepts behind lossy image compression techniques. In this respect, we pay a particular attention to lifting based coding methods.

In Chapter 2, we first introduce the main aspects relevant for understanding the stereoscopic imaging process. We also recall the most important techniques for estimating the disparity field. This step is a key issue for exploiting the cross-view redundancies and therefore achieving more efficient retrieval approach.

In Chapter 3, we present our contributions for the retrieval of stereo images. First, we describe the considered dense disparity algorithm. Then, we propose three different wavelet-based retrieval methods. While the first approach combines the visual contents of the stereo images with their corresponding disparity information using univariate model, the second one has been developed through bivariate statistical approach in order to take into account the similarity between the two views. The third approach exploits simultaneously the spatial and cross-view dependencies of the stereo images through multivariate statistical model.

In Chapter 4, we show how quantization can effect the performance of WT-based CBIR systems for some feature extraction methods. Then, we present new strategies for improving the retrieval performances of JPEG2000 compressed images. Finally, we investigate different quantization schemes and propose efficient retrieval approaches based on the inherent properties of each quantizer.

Finally, the main contributions of this reasearch work are summarized and some perspectives are provided.

Publications

Journal papers

- A. Chaker, M. Kaaniche, A. Benazza-Benyahia, and M. Antonini, Efficient transform-based image retrieval techniques under quantization effects, submitted in *Multimedia tools and applications*.
-

- A. Chaker, M. Kaaniche and A. Benazza-Benyahia, Disparity based stereo image retrieval through univariate and bivariate models, *Elsevier Signal Processing: Image Communication*, vol. 31, pp. 174-184, February 2015.

International Conference

- A. Chaker, M. Kaaniche, A. Benazza-Benyahia, and M. Antonini, An efficient statistical-based retrieval approach for JPEG2000 compressed images, in *European Signal Processing Conference, EUSIPCO*, 5 pages, Nice, France, September 2015.
 - A. Chaker, M. Kaaniche and A. Benazza-Benyahia, Multivariate statistical modeling for stereo image retrieval, *European Workshop on Visual Information Processing, EUVIP*, 6 pages, Paris, France, December 2014.
 - A. Chaker, M. Kaaniche and A. Benazza-Benyahia, Exploiting disparity information for stereo image retrieval, in *IEEE International Conference on Image Processing, ICIP*, 5 pages, Paris, France, October 2014.
 - A. Chaker, M. Kaaniche and A. Benazza-Benyahia, An efficient retrieval strategy for wavelet-based quantized images, in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 5 pages, Vancouver, Canada, May 2013.
 - A. Chaker, M. Kaaniche, and A. Benazza-Benyahia, An improved image retrieval algorithm for JPEG2000 compressed images, in *IEEE International Symposium on Signal Processing and Information Technology, ISSPIT*, 6 pages, Vietnam, December 2012.
-

Chapter 1

Useful basics

1.1 Introduction

In the last few years, CBIR has become one of the most active research areas. The development of huge and growing amounts of visual and multimedia data, and the advent of Internet underline the need to create thematic access methods that offer a quick and efficient interactive search.

This chapter gives a survey on CBIR systems. In the first part, we introduce the principle of CBIR and review different feature extraction approaches. In the second part, we present the most important concepts behind image compression techniques.

1.2 Content-Based Image retrieval systems

1.2.1 Principe of CBIR systems

The basic principle of an image retrieval system can be conceptually described by the framework depicted in Fig. 1.1. An image retrieval system consists of two elementary stages: feature extraction and the retrieval procedure. The feature extraction step aims at computing and storing relevant features (signatures) of the content of each image in the database. This may involve in some cases several intermediate steps such as preprocessing, image transformation and clustering process. This step is usually an off-line process, since it may not require user interaction or computational efficiency.

The retrieval procedure consists in finding the images in the database whose features are the closest to the query image according to a suitable similarity measure. Since the re-

trieval is performed on-line, its computation needs to be as fast as possible.

Thus, the retrieval performance depends critically on the efficiency of signature computation and the distance function.

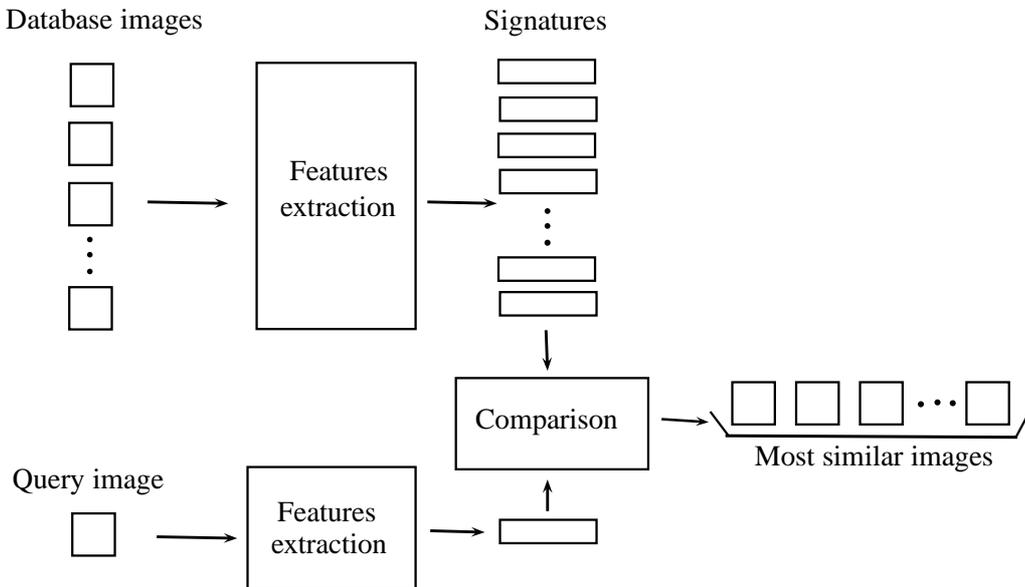


Figure 1.1: CBIR system architecture.

1.2.2 Image features

Image features are computed from visual cues contained in an image. We can distinguish two types of visual features: global and local ones. Global features describe the whole image without involving detailed local analysis, all pixels of the image are processed, while local features model localized areas in the image. The appropriate choice of the image feature depends on its ability to represent accurately the content of each image in the database according to the application area and the user need. For example, global features are recommended for homogenous images [Samadani *et al.*, 1993; Hara *et al.*, 1997], but local features are more appropriate to identify images having different objects with variable texture [Rao *et al.*, 2010]. Whether features are global or local, they belong generally to one of the three following categories: color descriptors, texture descriptors and shape descriptors.

In this work, we are only interested in shape- and texture-based features.

Shape features

Object shape features provide efficient information to identify objects. They contain a valuable information for image retrieval, since humans can recognize objects just from their shapes. Basically, shape features can be categorized into two classes contour- and region-based descriptors. Contour-based shape descriptors exploit only boundary information of object shapes while the region-based ones are computed from the entire region.

- Contour-based features:

Some features are computed directly from contour-based representations in the spatial domain such as area, circularity, eccentricity, major axis orientation [Zhang *et al.*, 2002]. To reduce boundary noise, a discrete curve evolution approach is presented in [Latecki, Lakämper, 2000] to simplify contours. Another shape presentation, called shape context, captures at a given reference point the distribution of the remaining points relative to it. This latter is compact and robust to several image transformations [Belongie *et al.*, 2002]. A shape retrieval approach based on token representation is presented in [Berretti *et al.*, 2000]. Shape boundary is firstly partitioned into tokens. Each token is described through a set of perceptually salient features. Token attributes are then organized into a tree index structure. This approach has been proved to be efficient in the presence of partially occluded objects, scaling and rotation. Different approaches have presented object shapes via Fourier descriptors [Zhang *et al.*, 2002; Datta *et al.*, 2006; Charmi *et al.*, 2008]. More precisely, these features represent the shape of the object in a frequency domain. Global shape features are captured from the first few low frequency coefficients, while finer features of the shape are extracted from higher frequency coefficients. These approaches allow information preserving and eliminate the noise sensitivity in the shape signature representations [Zhang *et al.*, 2002].

- Region-based features:

When the boundary information is insufficient for defining the object shape, important information in the shape interior should be exploited. Moments and function of moments have been well used as region-based features [Datta *et al.*, 2006]. The theory of moments including moment invariants and orthogonal moments provides useful alternative for series expansions for the representation of object shapes [Prokop,

Reeves, 1992; Vogel, Schiele, 2006; Ruberto, Morgera, 2008]. More precisely, moment invariants characterize images and they are independent of transformations such as rotation, translation and scaling. Orthogonal moments such as Zernike and Legendre moments [Prokop, Reeves, 1992] have been proven to be more robust in the presence of noise, and they are able to reduce the information redundancy in a set of conventional moments.

A complete set of Fourier-Mellin descriptors of gray-level image have been proposed in [Derrode *et al.*, 1999; Derrode, Ghorbel, 2004]. This set is commuted from the analytical Fourier-Mellin transform [Ghorbel, 1994], and it is invariant to translation, rotation and scale changes of the images. Moreover, this descriptor set has also proven to be stable and robust to shape distortions.

Texture features

Texture is one of the most important visual cue in identifying objects and describing the image content. There is no precise and general definition of texture. The number of different texture definitions by researchers, which depend upon their particular applications, explain this difficulty to define texture [Tamura *et al.*, 1978; Haralick, 1979]. Typically, texture is defined in the literature as a spatial repetition of one or more local patterns in a periodic manner and in different directions in space. However, this definition is inadequate since it does not take into account how the human visual system processes texture. Perceptually, texture has various discriminating properties such as uniformity, coarseness, bumpiness, fineness, regularity. Hence, a criterion reflecting the homogeneity of the texture can be used to distinguish between objects with different textures. It has been shown in [Julesz, 1962; Tuceryan, Jain, 1998] that the perception of texture has different dimensions due to physical variation in the scene which explains the fact that there is no general definition of texture. However, there are some intuitive properties of texture that are commonly accepted:

- A texture in an image can be perceived at different scales or levels of resolution [Gibson, 1950].
 - A texture is the spatial arrangement of pixel intensities (gray values).
 - If the number of primitive objects is relatively important in the region, the region is
-

perceived textured.

- The definition of a texture is related to the gray values in the spatial neighborhood whose size depends on the texture type or the size of the primitives forming the texture.

Texture information in an image can be extracted by building mathematical models which will be used as texture features. The related model parameters are used to characterize the images. It is worth noting that texture features have been employed in a variety of application fields such as aerial [Conners *et al.*, 1983], and medical image processing [Chen *et al.*, 1989], document processing [Wang, Srihari, 1989] and remote sensing [Haralick *et al.*, 1973] thanks to their close relation to the underlying semantics in these cases [Tuceryan, Jain, 1998].

Various techniques have been proposed for describing image texture [Kimchi, Palmer, 1982]. They can be classified into four major categories according to the characteristics of texture images [Tuceryan, Jain, 1998]: structural features, model based features, spectral features and statistical features (Fig. 1.2). In what follows, we will describe each category of features.

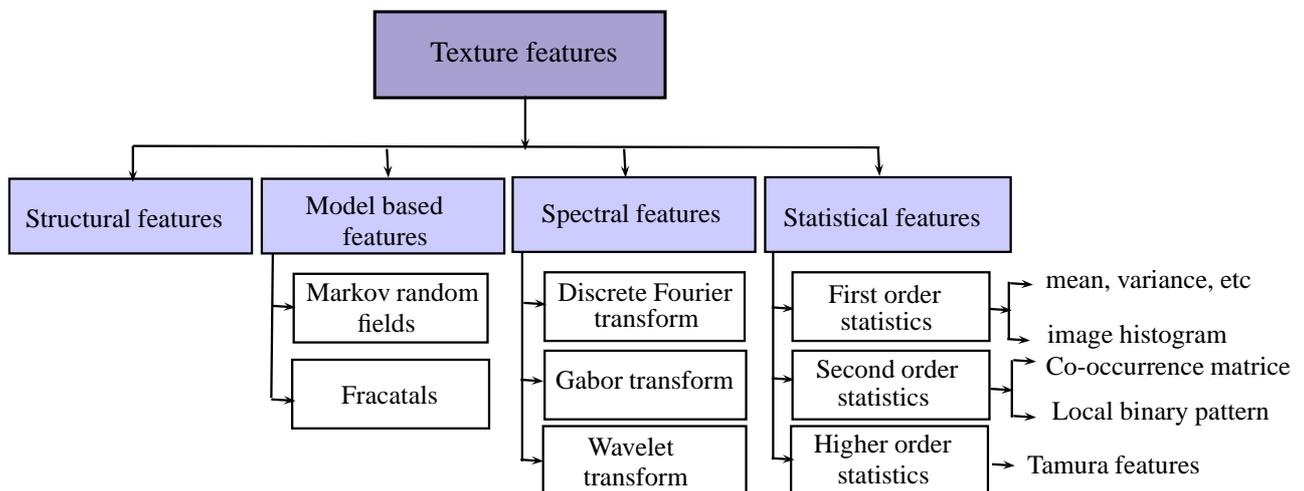


Figure 1.2: Different classes of texture features reported in the literature.

- Structural features:

Structural approaches consider textures as an almost regular repetition of some primitives [Haralick, 1979]. The structural models assume that textures are composed of

texture primitives. The texture is produced by the placement of these primitives according to certain rules. Morphological operations like opening and closing of interest point detectors are applied with the aim of finding and describing the arrangement of elements. However, texture patterns are not strictly periodic, the structural methods can not be applied in this case. Structural methods, in general, are very weak unless one is dealing with very regular textures [Tuceryan, Jain, 1998].

- Model based features:

Model-based texture methods are based on the building of an image model captured from textured images. The attributes derived from the models are used as image features. Model-based approaches make use of stochastic [Cross, Jain, 1983; Chellappa, Chatterjee, 1985] and fractal models [Pentland, 1984], to characterize an image texture by the use of stochastic model and generative image model, respectively. Several stochastic models have been proposed, among them the most popular being, autoregressive and Markov Random Field (MRF) models.

The autoregressive models assume relationships between groups of neighbouring pixels so that the pixel intensity is a weighted sum of the neighbouring pixel intensities [Castellano *et al.*, 2004]. They present the advantage of simplicity and efficiency in parameter estimation.

Markov random field is a probabilistic tool that is able to capture the local (spatial) contextual information in an image. These models assume that the intensity of each pixel in the image relies only on those of its neighbors [Blake, Zisserman, 1987]. MRF models have been employed in various image processing applications, such as texture synthesis [Cross, Jain, 1983], image segmentation [Speis, Healey, 1996] and image restoration [Geman, Geman, 1984].

Fractal models have also been proven to be very efficient for reflecting in modeling texture properties of natural images [Pentland, 1984; Materka *et al.*, 1998; Mandelbrot, 1983], especially for roughness and self-similarity quality surface and they are relatively insensitive to image scaling. However, they lack orientation selectivity and fail in characterizing local image structures.

- Spectral features:

These features represent texture images by using their spectral information. To this

respect, it is necessary to estimate the spectral frequency of pixel intensity. Discrete Fourier transform (DFT) was found to be suitable for highly periodic textures, since the DFT is an adequate tool for the analysis of periodic functions. Furthermore, texture descriptor invariant to translation, rotation and scale changes can be simply commuted in the DFT [Sim *et al.*, 2000]. Local spatial/frequency representations can be allowed by Gabor filters [Gabor, 1946] and wavelet transforms [Mallat, 1989]. The major disadvantage of the Gabor transform is that its output are not mutually orthogonal, which may result in a significant redundancy between texture features. In the last decade, wavelet theory has been widely used for texture retrieval purposes. Wavelet transforms offer flexible multiresolution and directional representations of an image consistent with the human visual system [Beck *et al.*, 1987], especially in the case of textured images. Once the multiresolution transformation is performed, the different statistical features can be extracted from the wavelet subbands such as the energy [Smith, Chang, 1994; Wouwer *et al.*, 1999], moments at several orders [Mandal *et al.*, 1996], statistical modeling of the distribution of the wavelet coefficients [Do, Vetterli, 2002; Sakji-Nsibi, Benazza-Benyahia, 2009; Sakji-Nsibi, Benazza-Benyahia, 2008; Lasmar, Berthoumieu, 2014], co-occurrence matrices [Kwitt, Uhl, 2008] and the correlation between all combinations of subband pairs on different channels [Van de Wouwer *et al.*, 1997]. Several extensions of the wavelet transform using multiscale and directional filter banks have been proposed to improve the treatment of oriented geometric image structures such as contourlet [Po, Do, 2006], curvelets [Dettori, Semler, 2007] and, grouplets [Peyré, 2010]. Several features were computed from the curvelet coefficients. The most used are a combination of mean, standard deviation, energy and entropy signal [Le Borgne, O'Connor, 2005]. The marginal and joint distributions were also calculated from the contourlet coefficients [Po, Do, 2006].

- Statistical features:

They describe the texture by its non-deterministic properties that reflect the distribution and relationship between the gray levels of an image.

- First order statistics derived from image histogram is the most basic approach describing texture. However, they only consider individual properties of pixels and do not take into account spatial relationships among them.
-

- Second-order statistical methods compute features of a texture using spatial relations of similar gray values. Indeed, gray level co-occurrence matrices [Haralick *et al.*, 1973] can provide valuable information about the neighboring pixels in a texture image. It is a matrix in which gray level pixels are considered in pairs with a relative distance d and orientation θ among them. The entry $G_{(d,\theta)}(i, j)$ of the matrix at position (i, j) is the number of occurrences of intensity values i and j separated by the specified displacement. Then, the set of the popular Haralick features [Haralick *et al.*, 1973] such as contrast, correlation, homogeneity and energy can be computed from these matrices. In addition, Local Binary Pattern (LBP) [Ojala *et al.*, 2002b] is a powerful invariant texture measure, derived from a general definition of texture in a local neighborhood. The intensity value of the current pixel is compared with that of its adjacent neighbors. If the intensity value of neighboring pixels is larger than the central pixel value, they are replaced by ones, or zero otherwise. The resulting eight bits are then interpreted as a natural number in the range of $[0, 255]$. An histogram is then built as an image descriptor.
- Higher order statistical features might be more accurate but they are not suitable for real-time applications due to their high computational complexity [Connors, Harlow, 1980]. For instance, Tamura features are designed according to psychological studies on the human visual perception of texture. Tamura features have six dimensions including coarseness, contrast, directionality, line-likeness, regularity and roughness [Tamura *et al.*, 1978]. The first three are more often used for image retrieval since they are more efficient in describing image textures [Liu *et al.*, 2007; Wang *et al.*, 2012].

1.2.3 Retrieval performance evaluation methods

CBIR systems search a number of images similar to a query from image database. Thus they differ from classification and recognition systems that can retrieve exact matches. The evaluation of the retrieval performance can be carried out by defining a quantitative objective metric. However, it is hard to quantify a measure of performance since the similarities between images are subjective and feature dependent, and there are no common standard image collection. Several quantitative metrics of performance are defined in the

literature [Müller *et al.*, 2001]. Most common and widespread evaluation measures used in image retrieval are precision $PR = \frac{R^r}{R^n}$ and recall $RC = \frac{R^r}{R^t}$ where

- R^r is the number of output documents considered as relevant.
- R^t is the total number of relevant documents in the database.
- R^n denotes the total number of documents retrieved.

Precision and recall are often presented as a precision versus recall graph. These graphs are increasingly used by the CBIR community [Müller *et al.*, 2001]. Several other measures are also derived from precision and recall such as $PR(10)$, $PR(30)$, $PR(R^t)$ which corresponds to the precision after the first 10, 30, R^t documents are retrieved, RC at $0.5PR$, recall at the rank where PR drops below 0.5, etc.

Rank is an important information that should be addressed for evaluating the retrieval performance. In this context, Average Normalized Modified Retrieval Rate (ANMRR) criterion is defined by MPEG-7 standard as a retrieval performance metric [Manjunath *et al.*, 2002]. ANMRR considers not only the recall and precision information, but also the rank information among the retrieved images.

Before describing the formulas of ANMRR, let $R^t(q)$, $K(q)$ and $R(k)$ be denoted as follows:

- $R^t(q)$ is the number of the ground truth images for a query q ;
- $K(q) = \min(4R^t(q), 2GTM)$, where GTM is the maximum of $R^t(q)$ over all queries;
- $R(k)$ represents the rank of an image k in retrieval results.

The rank of the k th ground truth images is defined as:

$$\text{Rank}(k) = \begin{cases} R(k) & \text{if } R(k) \leq K(q) \\ K(q) + 1 & \text{if } R(k) > K(q). \end{cases} \quad (1.1)$$

ANMRR is given by the following formulas [Manjunath *et al.*, 2002]:

$$\text{AVR}(q) = \sum_{k=1}^{R^t(q)} \frac{\text{Rank}(k)}{R^t(q)}, \quad (1.2)$$

$$\text{MRR}(q) = \text{AVR}(q) - 0.5 - \frac{R^t(q)}{2}, \quad (1.3)$$

$$\text{NMRR}(q) = \frac{\text{MRR}(q)}{K(q) + 0.5 - 0.5R^t(q)}, \quad (1.4)$$

$$\text{ANMRR} = \frac{1}{Q} \sum_{q=1}^Q \text{NMRR}(q). \quad (1.5)$$

where Q is the number of queries. NMRR and ANMRR are in the range of $[0, 1]$ and smaller values represent better retrieval performance.

In what concerns us, we have retained the most widespread measures in image retrieval works, namely PR and PC . We have also employed the ANMRR criterion to confirm results obtained by precision *versus* recall graph.

1.2.4 Applications of CBIR

The improvements of image acquisition systems in various areas such as remote sensing, astronomy, medical imaging and stereovision, have led to the production of huge amounts of scientific data that are available in digital form. For instance, the volumes of remotely sensed data gathered by the U.S. agencies NASA, NOAA, and USGS have dramatically grown resulting in about 18 Petabytes by 2010 [Ramapriyan, 2002]. The use of content-based image retrieval (CBIR) systems has become widespread as it allows a convenient and efficient data access by organizing images based only on their visual contents [Smeulders *et al.*, 2000]. Many CBIR systems both commercial and research have been built. The earliest commercial CBIR system is QBIC [Flickner *et al.*, 1995] was developed by IBM, it uses color, texture, shape descriptors and keywords. Virage [Bach *et al.*, 1996] uses color and texture descriptors. Moreover, a large number of academic retrieval systems have been developed by universities, companies and hospitals such as Photobook [Minka, Picard, 1996], Netra [Ma, Manjunath, 1999], and VisualSEEK [Smith, Chang, 1997] allow query by image content. Several CBIR applications have been developed not only for commercial and academic purposes but also for medical diagnostics [Long *et al.*, 2009; Müller *et al.*, 2004]. They are powerful to find similar images in various modalities acquired in different stages of the disease progression. Furthermore, CBIR becomes widely used on different hardware and platforms such as mobile phones [Quack *et al.*, 2004].

In what follows, we will review the basic concepts behind image compression techniques.

1.3 Compression techniques

Image compression algorithms are used to reduce the number of bits required to represent an image. They aim at representing information in a compact form. Based on the

requirement of image reconstruction, image compression approaches can be divided into two classes: *lossless* compression techniques, in which the reconstructed image is exactly identical to the original one and, *lossy* compression techniques which enable much higher compression with quality degradation of the reconstructed image. A typical lossy compression method can easily be represented in the form of a functional scheme composed of three blocks: transformation, quantization, and entropy encoding.

1.3.1 Transformation

In almost all images, there is typically a high redundancy between the neighboring pixels. The objective is then to reduce the amount of data by exploiting the spatial redundancies in the images. To this end, two fundamental compression techniques can be considered: predictive and transform ones. Although predictive coding methods are simple and can be effective, the key methods use orthogonal transform in order to generate transformed coefficients that present less mutual redundancies than pixels in the original spatial domain. Moreover, an appropriate transform should provide a new representation where the energy is concentrated on a few coefficients. The optimal transform for the data decorrelation is the discrete Karhunen-Loève Transform (KLT) [Hua, Liu, 1998] as it provides a strict diagonal autocorrelation matrix for the resulting coefficients. Even though it is optimal, it is not widely adopted because it involves a high computational load. Furthermore, as it is data-dependent, the Karhunen-Loève should be computed for each image. Although several fast algorithms have been proposed, the computational complexity is still very high [Hua, Liu, 1998]. It has been shown in [Feng, Effros, 2002], the KLT and its fast variants range in complexity requirement from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$ for vectors of dimension n . This has motivated to resort to other transforms that are sub-optimal (in terms of decorrelation) but whose computation is dramatically less expensive. Currently, the Discrete Cosine Transform (DCT) [Rao, Yip, 1990] and the Wavelet Transform (WT) [Mallat, 1999] are the main transforms widely used in image compression.

Discrete cosine transform

The DCT is a signal independent approximation of the KLT, recognized as an effective technique for image and video compression. It has been adopted by most image coding standards such as JPEG [Wallace, 1991], H.261 [Liou, 1991] and MPEG coding standard

[Le Gall, 1991]. More precisely, DCT is a block transform which consists firstly in partitioning the image into nonoverlapping adjacent blocks of size $N \times N$ blocks (typically, $N = 8$) and then DCT is applied independently to each block to separate the high and low frequency information. The DCT of an input pixel $x(i, j)$ in a current block, for $(i, j) \in \{0, \dots, N - 1\}^2$ is defined as:

$$X(u, v) = \frac{2}{N} C(u) C(v) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} x(i, j) \cos \frac{(2i+1)u\pi}{2N} \cos \frac{(2j+1)v\pi}{2N} \quad (1.6)$$

where

$$C(u) = \begin{cases} 1/\sqrt{2} & \text{for } u = 0 \\ 1 & \text{otherwise.} \end{cases}$$

DCT has an almost optimal energy concentration since the most significant information is contained in a few low-frequency coefficients. Indeed, among the N^2 coefficients, the continuous component $X(0, 0)$ (DC coefficients) contains the average block (up to a multiplicative factor), while the $(N^2 - 1)$ AC coefficients present the higher frequencies. In addition to the advantage of energy compacting, DCT reduces the correlation between the coefficients. It achieves acceptable compression results, and it gives a low memory implementation, since the encoding is performed on small individual blocks of pixels. However, the main limitation of block processing in the DCT is the appearance of blocking artifacts which lead to a degradation in coding performance, mainly at very low bitrate coding. Blocking artefact remains a major concern of DCT-based coding. Many research works have been done in order to eliminate its effects, but there is a considerable scope for further research.

Discrete wavelet transform

Unlike the DCT that transforms image from the spatial domain into the frequency domain, the discrete wavelet transform (WT) achieves a good tradeoff between space and frequency localization. Moreover, wavelet transform has been found to be one of the most powerful compression tools. For instance, it has been retained by JPEG2000 image coding standard as it achieves high compression ratio and a scalable decoding. The multiresolution analysis algorithm proposed in [Mallat, 1999] is recommended to process the wavelet transform of the image. The procedure starts with passing the input signal s_0 through a pair of filters with impulse response $\bar{h}_0(n) = h_0(-n)$ and $\bar{h}_1(n) = h_1(-n)$, called analysis filters. Note

that h_0 and h_1 are respectively a low-pass and high-pass filters. Then, the filtered signals are subsampled with a factor of 2 which results in a low-pass version s_1 and and respectively a high-pass version d_1 of the input signal. The low-pass version is the *approximation* subband, while the high-pass one correspond to the *detail* subband.

In practice, the decomposition procedure is iterated on the resulting approximation signal as illustrated in Fig. 1.3. The decomposition scheme corresponding to two-levels decomposition.

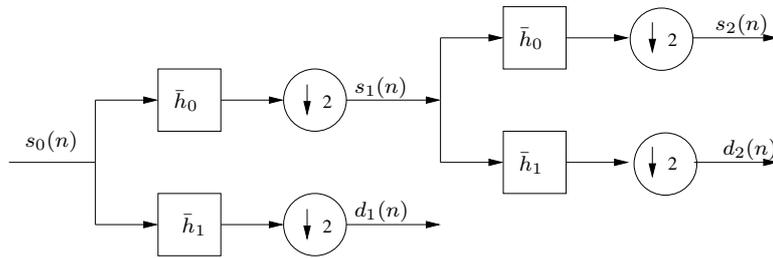


Figure 1.3: One dimensional wavelet decomposition over 2 resolution levels.

The approximation coefficients s_j at resolution level j are related to the approximation coefficients s_{j+1} and the detail ones at resolution $j + 1$ as follows:

$$\begin{aligned} s_{j+1}(n) &= \sum_{k=-\infty}^{+\infty} s_j(k)h_0(k - 2n) \\ d_{j+1}(n) &= \sum_{k=-\infty}^{+\infty} s_j(k)h_1(k - 2n). \end{aligned} \quad (1.7)$$

The reconstruction is straightforward. The outputs s_{j+1} and d_{j+1} are first upsampled by a factor 2, then passed through the synthesis filters with impulse responses g_0 and g_1 and, finally added together to form a reconstructed signal \tilde{s}_j :

$$\tilde{s}_j(n) = \sum_{k=-\infty}^{+\infty} s_{j+1}(k)g_0(n - 2k) + \sum_{k=-\infty}^{+\infty} d_{j+1}(k)g_1(n - 2k). \quad (1.8)$$

In the case of an image, such 1D decomposition is generally applied along the rows first then the columns (or inversely) in a separable way. Fig. 1.4 shows how an image $x_j(m, n)$ can be decomposed using separable filter bank. First, each row of the image is 1D transformed to obtain an approximation subband $s_{j+1}(m, n)$ and a detail subband $d_{j+1}(m, n)$. Then, each column of every subband is 1D transformed. Thus, four subbands are generated: an approximation subband $x_{j+1}^{(LL)}$ and three detail subbands $x_{j+1}^{(HL)}$, $x_{j+1}^{(LH)}$ and $x_{j+1}^{(HH)}$

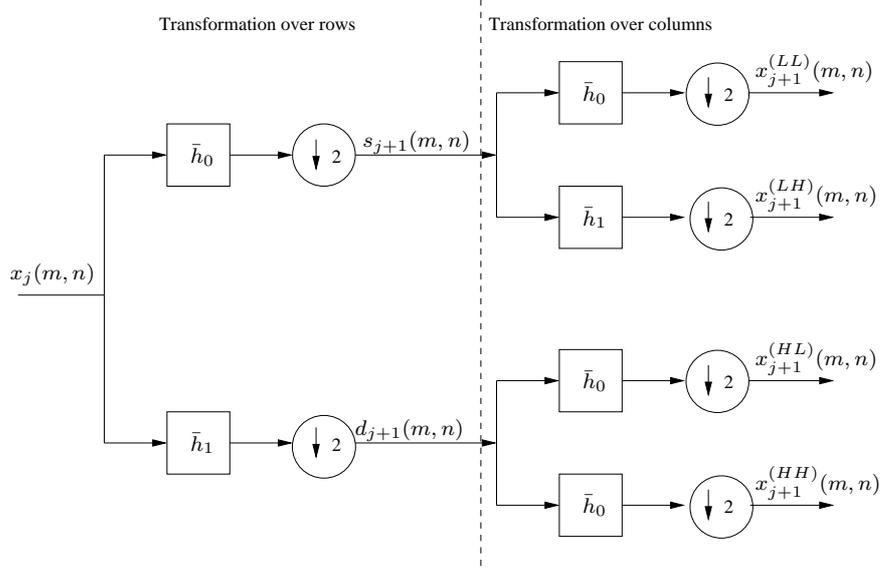


Figure 1.4: 2D separable filter bank.

corresponding respectively to the horizontal, vertical and diagonal orientations. This decomposition procedure can be applied on the approximation subband until the desired subband structure is obtained.

Lifting schemes

The Lifting Scheme (LS) is a flexible tool for computing the WT [Sweldens, 1996]. It was introduced to construct biorthogonal wavelets by a simple and reversible process. LS was found to be a very effective structure for encoding still and stereo images [Kaaniche *et al.*, 2009a], and it has been retained in the JPEG2000 image compression standard [Taubman, Marcellin, 2001]. A generic LS is performed in three steps namely *split*, *predict* and *update*.

- **Split:** The input 1D signal $s_j(n)$ is firstly divided into two subsets composed respectively of even $s_j(2n)$ and odd samples $s_j(2n + 1)$.
- **Predict:** Then, thanks to the local correlation, the samples of one subset (say the odd ones) are predicted from the neighboring even samples. Thus, the prediction error, referred to as detail signal, is computed as follows:

$$d_{j+1}(n) = s_j(2n + 1) - \mathbf{p}_j^\top \mathbf{s}_j(n) \quad (1.9)$$

where \mathbf{p}_j is the prediction vector, and $\mathbf{s}_j(n)$ is a reference vector containing the even image samples used in the prediction step.

- **Update:** The approximation signal $s_{j+1}(n)$ is generated by smoothing the even samples through some detail coefficients:

$$s_{j+1}(n) = s_j(2n) + \mathbf{u}_j^\top \mathbf{d}_{j+1}(n) \quad (1.10)$$

where \mathbf{u}_j is the update vector, and $\mathbf{d}_{j+1}(n)$ is a reference vector containing the detail coefficients used in the update step.

In the case of an image, such 1D-LS can be separately applied to the lines and, to the columns, as for the filter bank-based method. A multiresolution representation of the input image is obtained by recursively repeating these steps to the resulting approximation coefficients.

One of the interesting properties of the LS is its easy and intrinsic reversibility, it does not require any condition on the prediction and update operators. This flexibility has challenged researchers to develop various nonlinear wavelet transforms [Claypoole *et al.*, 2003; Combettes, Pesquet, 1998]. The reconstruction algorithm can be obtained from the forward transform simply by inverting the three step procedure which results in:

- **Undo update:**

$$s_j(2n) = s_{j+1}(n) - \mathbf{u}_j^\top \mathbf{d}_{j+1}(n). \quad (1.11)$$

- **Undo predict:**

$$s_j(2n+1) = d_{j+1}(n) + \mathbf{p}_j^\top \mathbf{s}_j(n). \quad (1.12)$$

- **Merge:**

$$\{s_j(n), n \in \mathbb{Z}\} = \{s_j(2n), n \in \mathbb{Z}\} \cup \{s_j(2n+1), n \in \mathbb{Z}\}. \quad (1.13)$$

Several examples of LS have been introduced in [Calderbank *et al.*, 1998]. Among them, the set of LSs denoted by (L, \tilde{L}) wavelet transforms is known to be effective for image coding, where L and \tilde{L} is the number of vanishing moments of the analyzes and the synthesis high pass filters, respectively. The conversion into an integer wavelet transform, which is essential for lossless compression, requires rounding steps. We present some examples of this group:

- (2, 2) transform

$$\begin{aligned} d_{j+1}(n) &= s_j(2n+1) - \lfloor \frac{1}{2}(s_j(2n) + s_j(2n+2)) + \frac{1}{2} \rfloor \\ s_{j+1}(n) &= s_j(2n) - \lfloor \frac{1}{4}(d_{j+1}(n-1) + d_{j+1}(n)) + \frac{1}{2} \rfloor. \end{aligned} \quad (1.14)$$

This transform is also known as the 5/3 transform. It has been retained for lossless compression mode of the JPEG2000 standard.

– (4, 2) transform

$$\begin{aligned} d_{j+1}(n) &= s_j(2n+1) - \lfloor \frac{9}{16}(s_j(2n) + s_j(2n+2)) - \frac{1}{16}(s_j(2n-2) + s_j(2n+4)) + \frac{1}{2} \rfloor \\ s_{j+1}(n) &= s_j(2n) + \lfloor \frac{1}{4}(d_{j+1}(n-1) + d_{j+1}(n)) + \frac{1}{2} \rfloor. \end{aligned} \quad (1.15)$$

In addition, 9/7 symmetric biorthogonal transform is another popular transform selected for the lossy compression mode of the JPEG 2000 standard.

$$\begin{aligned} \check{d}_{j+1}(n) &= s_j(2n+1) + \alpha(s_j(2n) + s_j(2n+1)) \\ \check{s}_{j+1}(n) &= s_j(2n) + \beta(\check{d}_{j+1}(n-1) + \check{d}_{j+1}(n)) \\ \tilde{d}_{j+1}(n) &= \check{d}_{j+1}(2n+1) + \gamma(\check{s}_{j+1}(n) + \check{s}_{j+1}(n+1)) \\ \tilde{s}_{j+1}(n) &= \check{s}_{j+1}(n) + \delta(\tilde{d}_{j+1}(n-1) + \tilde{d}_{j+1}(n)) \\ d_{j+1}(n) &= \frac{1}{\zeta} \tilde{d}_{j+1}(n) \\ s_{j+1}(n) &= \zeta \tilde{s}_{j+1}(n) \end{aligned} \quad (1.16)$$

where $\alpha = -1.586134342$, $\beta = -0.05298011854$, $\gamma = 0.8829110762$, $\delta = 0.4435068522$, $\zeta = 1.149604398$ are lifting coefficients.

1.3.2 Quantization and entropy coding

Quantization is a simple operation that converts a continuous set of values (or a large number of possible discrete values) to a relatively small discrete and finite set, which implies a loss of information. This loss is of course irreversible. Thus, the design of the quantizer has a significant impact on the compression ratio. We can define two main classes of quantization: scalar and vector quantization. A scalar quantization is applied to each input sample (i.e. transform coefficient) independently from the other input, whereas the vector quantization is applied to blocks of transform coefficients.

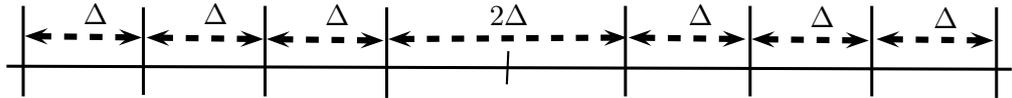


Figure 1.5: Uniform quantizer with deadzone with step size Δ .

The uniform quantizer with a central deadzone is a simple example to perform the scalar

quantization. Fig. 1.5 shows a uniform quantizer with a deadzone twice the quantization step-size. Let $x_j(m, n)$ the input sample, the output of the quantizer $\bar{x}_j(m, n)$ is given by:

$$\bar{x}_j(m, n) = \text{sign}(x_j(m, n)) \left\lfloor \frac{|x_j(m, n)|}{\Delta} \right\rfloor \quad (1.17)$$

where Δ is the quantization step size and $\lfloor \cdot \rfloor$ denotes the rounding operator. It is worth noting that a small (resp. high) Δ value corresponds to a high (resp. low) bitrate and, results in a high (resp. low) reconstructed subband quality.

After quantization, the finitely different values at the output of the quantizer must be encoded. The entropy encoding converts the quantized coefficients into a binary stream. JPEG specifies two entropy coding methods: Huffman coding [Huffman, 1952] and arithmetic coding [Rissanen, Langdon, 1979]. A variety of powerful and sophisticated algorithms have been developed for wavelet-based coding schemes. Among them, we cite the Embedded Zerotree Wavelet (EZW) coding [Shapiro, 1993], Set Partitioning in Hierarchical Tree (SPIHT) coding [Said *et al.*, 1996], and the Embedded Block Coding with Optimized Truncation (EBCOT) algorithm [Taubman, 2000] which is adapted in the JPEG2000 standard. These algorithms can achieve an excellent image quality while generating embedded bit-stream. A general overview on wavelet codecs can be found in [Fowler, Pesquet-Popescu, 2007].

EBCOT is the basic encoding tool of JPEG2000 [Taubman, Marcellin, 2001]. It is a block-based encoder organized around two passes: each subband is divided into small blocks called *codeblocks*, which are typically of size 32×32 or 64×64 . Then, an embedded bitstream is independently generated for each codeblock which results in a separately embedded and layered bit-stream organization. Thus, each codeblock bitstream is truncated in some fashion, and the truncated bitstreams are concatenated together to form the final bitstream. In JPEG2000, the method for codeblock bitstream truncation is a Lagrangian rate-distortion optimal approach referred to as Post-Compression Rate-Distortion (PCRD) optimization [Taubman, Marcellin, 2001]. An optimal truncation point for each code block is produced by processing this technique simultaneously across all of the codeblocks from the image. More precisely, the PCRD optimization distributes the total bite rate for the image spatially across the codeblocks in an optimal rate-distortion sense such that codeblocks with higher energy receive greater rate. Moreover, the constructed bitstream is organized in an optimal order such that the final bitstream is close to being rate-distortion optimal at many truncation points.

Bit allocation

In order to optimize the reconstruction quality of an image, the bit allocation process in wavelet based coders allows to efficiently distribute a given budget of bits among all subbands of a wavelet-based coded image [André *et al.*, 2007]. Generally, the bit allocation is expressed as an optimization problem based on Rate-Distortion (R-D) theory where the average distortion is minimized subject to a constraint on the available bitrate. The objective is to find the optimal quantization steps for the constrained minimization problem. The bit allocation algorithms in the literature can be classified into empirical and analytical based methods. Algorithms in the first category estimate the R-D curves by measuring a large number of R-D operating points for all possible quantization settings for each subband. The obtained R-D curves are often assumed to be both differentiable and convex [Sermadevi, Hemami, 2004]. Then, different iterative techniques are employed to find the optimal quantization parameters [Ortega, Ramchandran, 1998]. To this end, the standard Lagrangian optimization technique is well applied to solve the constrained minimization problem [Shoham, Gersho, 1988]. While this empirical method results in better optimization results, it is computationally intensive. To reduce the complexity, instead of estimating R-D curves for all quantization settings, the curves are generated by the computation of a few points and interpolating them using spline approximations in [André *et al.*, 2007]. Thus, to further overcome the complexity of these empirical methods, analytical approaches consist of finding closed-form analytical expressions of the R-D functions by assuming various input distributions and quantizer characteristics [Kaaniche *et al.*, 2014].

1.4 Conclusion

In this chapter, we have described the principle as well as the different stages involved in content based image retrieval systems. In this respect, we have reviewed shape and texture feature extraction techniques as well as different application fields of CBIR. Then, we have given some background on image compression tools by focusing on wavelet transforms, and particularly lifting schemes.

Chapter 2

Stereo images and disparity map

2.1 Introduction

A stereo system involves at least two cameras from which the depth information can be inferred by triangulation if corresponding (homologous) pixels are found. Hence, it is useful to recover the depth information of the underlying scene in order to exploit the 3D information during the retrieval procedure of Stereo Images (SI). To this end, the fundamental problem reduces to match the pixels from a pair of left and right views (if a binocular stereo system is considered). More precisely, the challenge is to estimate the difference in location between two corresponding pixels (called the disparity).

In this chapter, we focus on the stereo matching process. First, we will provide an introduction to stereo vision. Next, we will present the stereo correspondence problem and some constraints that can be used in order to simplify the matching process. Then, we will give an overview of the most commonly used disparity estimation methods.

2.2 Stereoscopic imaging

2.2.1 Image acquisition

The pinhole camera model is the simplest and the most used model for describing image acquisition. It has proven to be very efficient in approximating the geometry and optics of most modern cameras [Faugeras, 1993]. In this model, the image is rendered in perspective

projection that it is formed by a geometry mapping \mathcal{M} from 3D onto a 2D plane:

$$\begin{aligned} \mathcal{M} : \mathbb{R}^3 &\rightarrow \mathbb{R}^2 \\ (X, Y, Z) &\rightarrow (f \frac{X}{Z}, f \frac{Y}{Z}) \end{aligned} \quad (2.1)$$

where f is the focal distance of the camera and (X, Y, Z) are the 3D coordinates.

2.2.2 Homologous points

In a binocular stereoscopic imaging systems, two images of the scene are simultaneously acquired from two slightly different view angles by two cameras (see Fig. 2.1). The objective is to recover from the two images (left and right views) the third dimension (the depth) which has been lost during the image formation. The projections of any point M in the scene visible by the two cameras do not have the same coordinates in the two image planes. The position of this scene point can be retrieved if its two projections $m^{(l)}$ and $m^{(r)}$ onto the left and right images are identified as shown in Fig. 2.1. These points $m^{(l)}$ and $m^{(r)}$ are called homologous.

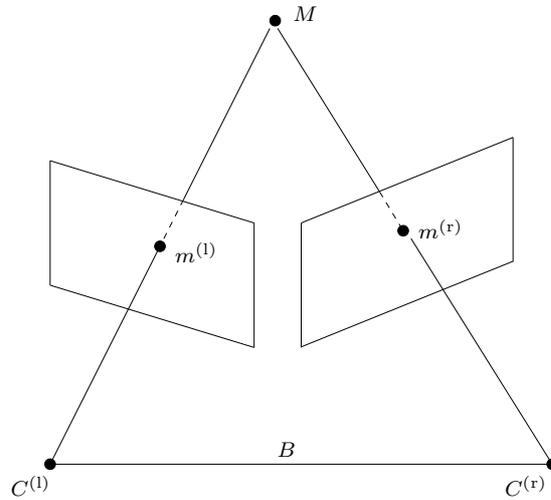


Figure 2.1: Stereoscopic imaging system.

Geometry

In order to establish the stereo matching problem, it is important to understand the two-view geometry, which describes the relation between the two views. This geometry, known as epipolar geometry, is illustrated in Fig. 2.2.

Let $C^{(l)}$ and $C^{(r)}$ be respectively the optical centers of the left and right cameras. As shown

in Fig. 2.2, a scene point M , their projection points $m^{(l)}$ and $m^{(r)}$ in the left and right images and $C^{(l)}$ and $C^{(r)}$ are co-planar and present the *epipolar plane* Π .

Suppose we only know $m^{(r)}$ and we have to search its corresponding point $m^{(l)}$. The plane Π is determined by the line joining the optical centers (called *baseline*) and the ray defined by $m^{(r)}$. From above, the ray must pass through $m^{(l)}$ and lies in Π . Thus, the point $m^{(l)}$ lies on the line of intersection $l^{(l)}$ of Π with the left image plane. This line is called the *epipolar line* $l^{(r)}$ associated to $m^{(r)}$.

Several entities involved in this geometry are illustrated in Fig. 2.2 and defined as following:

- The *epipolar plane* Π of a scene point M is the plane determined by M and the centers of projection of the two cameras $C^{(l)}$ and $C^{(r)}$.
- The *epipoles* are two points $e^{(l)}$ and $e^{(r)}$ obtained by the intersection of the baseline with the image plane.
- The *epipolar line* is the intersection of the epipolar plane with the image plane. All the epipolar lines in the image pass through a single point the epipole. The intersection of the epipolar plane with each of the two image planes are called the conjugate epipolar lines.

Given an image point $m^{(r)}$ lying on the right epipolar line $l^{(r)}$, its corresponding point $m^{(l)}$ must lie on the conjugate epipolar line $l^{(l)}$. This corresponds to the epipolar constraint which can be represented mathematically by a 3×3 matrix \mathbf{F} called the *fundamental matrix*. This matrix depends on the intrinsic and extrinsic parameters of the cameras [Zhang, Xu, 1997]. Note that the intrinsic parameters describe the geometric, optical and digital characteristics of the camera. They do not depend on the position and orientation of the camera. The extrinsic parameters define the orientation and position of the camera with respect to the world coordinate system [Faugeras, 1993]. A complete review of the techniques for estimating \mathbf{F} can be found in [Zhang, 1998]. Thus, such a constraint allows to reduce the correspondence research from 2D search problem to 1D one. In practice, this problem can be further simplified by using an appropriate processing known as *epipolar rectification*.

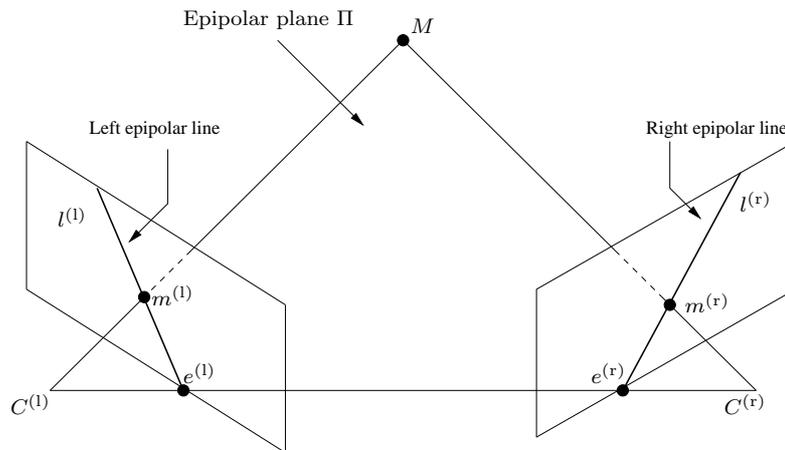


Figure 2.2: Epipolar geometry.

2.2.3 Epipolar rectification

Typically, cameras used in stereoscopic vision are rotated towards each other by a small angle. In this case, the epipolar lines lie at a variety of angles across the two images. This configuration is known as *converging camera configuration* [Dufaux *et al.*, 2013]. In contrast, if the optical axes of two cameras are parallel and the two cameras have the same intrinsic parameters the epipolar lines are horizontal and, will coincide with the image scan lines. This configuration is called *parallel camera configuration*.

The last configuration has several advantages. Indeed, it simplifies the correspondence search, which will be situated on parallel horizontal lines. Furthermore, since the epipolar lines lie on the image scan lines, two corresponding pixels are located at the same line in the two images as shown in Fig. 2.3.

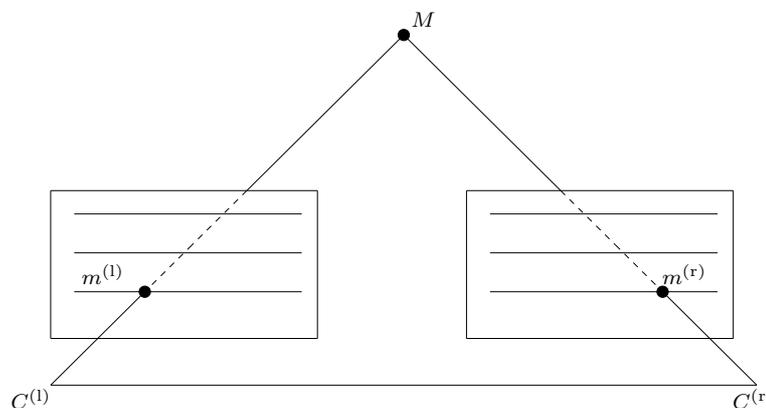


Figure 2.3: Epipolar rectification.

However, in practice most images are not perfectly aligned, thus a transformation process should be applied to images such that corresponding epipolar lines become colinear and coincide with the image scan lines. This procedure is called rectification. A survey of different rectification techniques proposed in the literature can be found in [Hartley, Gupta, 1993; Fusiello *et al.*, 2000b; Papadimitriou, Dennis, 1996].

2.3 Stereo matching problem

2.3.1 Disparity information

In the matching process, we consider that the mapping is performed from the right image (called “reference”) to the left one. For a given pixel $m^{(r)} = (x^{(r)}, y^{(r)})$ in the right image, the objective is to find its homologous pixel $m^{(l)} = (x^{(l)}, y^{(l)})$ in the left image, which corresponds to the same point in the scene. The difference between the coordinates of these pixels is called *disparity*.

Thus, the stereo correspondence problem is equivalent to find for a function u that associates a disparity vector to each pixel $(x^{(r)}, y^{(r)})$ in the right image as follows:

$$\begin{aligned} u : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ (x^{(r)}, y^{(r)}) &\mapsto u(x^{(r)}, y^{(r)}) = (x^{(l)} - x^{(r)}, y^{(l)} - y^{(r)}). \end{aligned} \quad (2.2)$$

If the stereo images are geometrically rectified, the disparity field is restricted to the horizontal component. Hence, in this case, the disparity is a scalar:

$$\begin{aligned} u_x : \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x^{(r)}, y^{(r)}) &\mapsto u_x(x^{(r)}, y^{(r)}) = x^{(l)} - x^{(r)}. \end{aligned} \quad (2.3)$$

It is important to note that there is an inverse relation between disparity and depth. Indeed, objects closer to (resp. far away from) the observer result in high (resp. low) disparity values. This information plays a crucial role in many application fields such as 3D reconstruction and stereo/multiview image compression [Kaaniche *et al.*, 2009b; Dricot *et al.*, 2015].

The depth information can be computed based on the triangulation method [Hartley, Sturm, 1997], as shown in Fig. 2.4. Recall that a 3D object point M is defined by its coordinates (X, Y, Z) where the Z component represents its depth and corresponds to the distance between the object and the cameras. By employing the similarity of the triangles

of hypotenuses $(C^{(l)}m^{(l)})$ and $(C^{(l)}M)$, we get:

$$\frac{x^{(l)}}{f} = \frac{X}{Z} \quad (2.4)$$

and,

$$\frac{x^{(r)}}{f} = \frac{X - B}{Z} \quad (2.5)$$

where B is the baseline and f is the camera focal length.

Consequently, we can express the disparity of a point M as a function of the depth Z between the corresponding 2D points as follows:

$$u_x(x^{(r)}, y^{(r)}) = x^{(l)} - x^{(r)} = \frac{B \cdot f}{Z}. \quad (2.6)$$

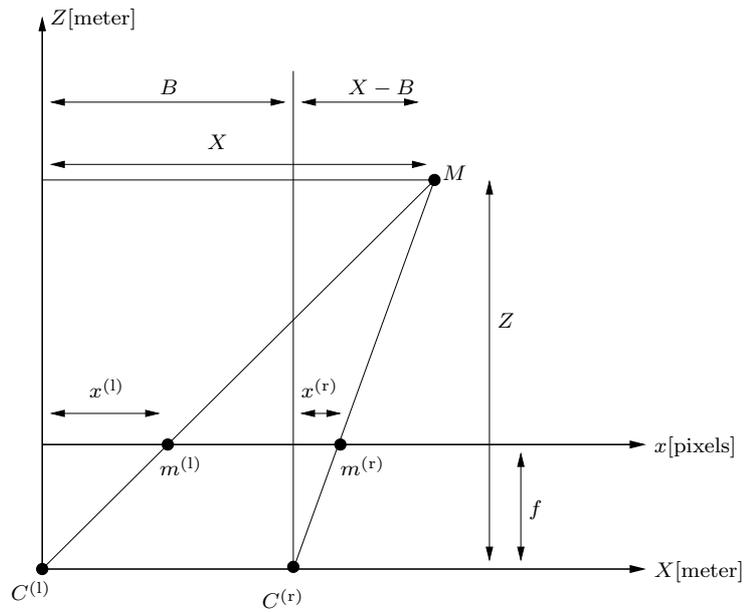


Figure 2.4: Relation between the disparity and depth.

2.3.2 Difficulties in stereo matching process

Disparity estimation is considered as a difficult problem in stereo vision, because of the presence of several sources of uncertainty such as illumination variations, depth discontinuities, and lack of texture [Dhond, Aggarwal, 1989].

- **Illumination variations:** In classical stereo matching process, the scene is assumed to be Lambertian and corresponding pixels have consistent intensities from different

viewpoints. However, in practice, illumination variations often occur due to slight differences in sensitivity of the cameras or shadows that appear in only one image. This illumination differences between two stereo images can lead to an undesirable disparity map. To reduce the effect of illumination change, a preprocessing step like the histogram equalization can be applied to the original stereo images.

- Occlusion problem: The different viewpoints of the cameras and the presence of discontinuities in the scene cause occlusions, which are points in one image that are absent in the other one. An example of an occlusion is shown in Fig. 2.5, where the point X is visible in the right image and absent in the left one. This makes the disparity assignment a very difficult task since occlusion areas have to be detected and discarded beforehand. A comparison of different occlusion detection methods can be found in [Egnal, Wildes, 2002]. For example, in [Weng *et al.*, 1992] occluded areas are detected and replaced with an interpolation based on the disparity of neighborhoods of the occluded area. A comparative study of five methods for occlusions detection was conducted in [Egnal, Wildes, 2002]. To reduce the sensitivity of the matching process to occlusions, a robust correlation measures such as normalized correlation and M-estimator have been proposed in [Oisel *et al.*, 2003]. To explicitly identify occluded regions, an iterative algorithm updates the match values, by diffusing search window among neighboring values and inhibiting others along similar lines of sight was developed in [Zitnick, Kanade, 2000].
- Textureless regions: Image regions which contain little or repetitive textures result in ambiguities in the matching process because of the presence of multiple possible matches. This ambiguity often leads to wrong correspondences. Homogeneous textured region are taken into account and detected using the gradient image in [Scharstein, Szeliski, 2002].

In conclusion, there are many factors that make stereo matching an ill-posed problem.

2.3.3 Stereo matching constraints

It is important to impose some matching constraints to solve the difficulties mentioned above. The most commonly used constraints are the following ones.



Figure 2.5: Illustration of occlusion effects: the point X is visible in the right image and absent in the left one.

- **Epipolar constraint:** This geometric constraint allows to reduce the matching problem from a 2D search problem to a 1D one by involving that the corresponding point of a given image point in one image, must be located on an epipolar line in the other one.
Moreover, when the stereo images are rectified, the search problem is further simplified since the horizontal scan lines reside on the same epipolar lines.
- **Uniqueness:** It imposes that one pixel in an image can have no more than one corresponding pixel in the other image. But, this constraint fails in the presence of transparent objects [Marr, Poggio, 1976].
- **Ordering:** It constrains the order of pixels along epipolar lines to be the same for the two images.
- **Smoothing constraint:** It imposes that disparity varies smoothly in the uniform areas (object interiors).

2.4 Overview of stereo matching approaches

As aforementioned, the requirement of disparity maps is motivated by several applications in stereo vision such as 3D reconstruction of the scene, the compression of multicomponents images. To this respect, a large number of stereo matching methods have been developed to produce accurate disparity maps from single stereo-pairs. Numerous techniques proposed

in the literature are summarized in [Scharstein, Szeliski, 2002; Brown *et al.*, 2003]. They are mainly categorized in two classes: local and global matching approaches.

2.4.1 Local approaches

Local methods seek to estimate disparity by comparing the similarity between two sets of pixels of two reference images. They can be subdivided into feature- and area-based approaches. Feature-based methods try to establish correspondences between some high-level features extracted from images such as edges [Ohta, Kanade, 1985], segments [Medioni, Nevatia, 1985] or curves [Schmid, Zisserman, 1998]. These methods are efficient and very fast since only few features are considered and less sensitive to photometric variations, as they represent geometric properties of a scene. However, they can not ensure the sparsity of the obtained disparity map. Thus, an interpolation step is required to produce a dense map [Konrad, Lan, 2000].

Unlike most feature-based methods, dense disparity fields are explicitly defined for area-based algorithms. These algorithms are efficient in highly textured areas but more sensitive to locally ambiguous regions like occlusion regions or textureless regions. This class of methods, known as correlation-based methods, uses pixels or regions to measure the similarity between stereo images [Alkanhal *et al.*, 1999].

It is worth noting that these techniques are highly popular and have been widely used in several applications such as motion estimation [Alkanhal *et al.*, 1999], image registration [Zitova, Flusser, 2003], etc.

In the context of stereo correspondence, the principle of these approaches is to compare the neighborhood of a pixel with the neighborhood of possible candidates located on the epipolar line associated to the other image, as illustrated in Fig. 2.6. More precisely, in order to find for a given pixel $m^{(r)}$ in the right image, the corresponding pixel in the left image, a window of pixels around $m^{(r)}$ is firstly selected. Then, this window is matched to a window in the left image by shifting the current window over a search area S . Finally, the shift which gives the best cost function is considered as a good match.

Generally, the cost function is a dissimilarity measure between the two windows. To this respect, different matching measures have been used in the literature. The most commonly used are the Sum of Square Differences (SSD), the Sum of Absolute Differences (SAD) and, the Normalized Cross-Correlation (NCC). NCC decreases the illumination variation

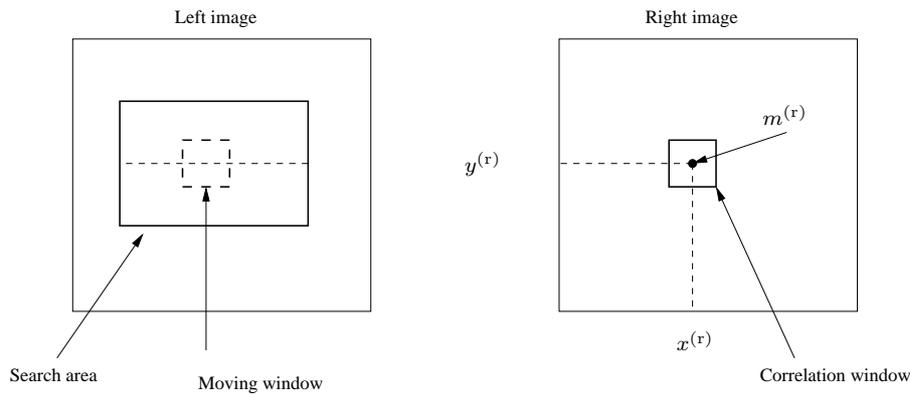


Figure 2.6: Correlation-based stereo matching method.

effect by compensating the differences of the gain and the bias between stereo images [Hirschmüller, Scharstein, 2007]. Other nonparametric measures have been presented in the literature [Zabih, Woodfill, 1994]. They are based on transformations which rely on the rank of the relative ordering of local intensity values within a local window and not on the intensity values themselves. For example, rank transform [Zabih, Woodfill, 1994] is the number of pixels in the local region centered at the pixel of interest whose intensity is less than the intensity of the center pixel.

The selection of the appropriate size of the search window is crucial for the effectiveness of local methods. Indeed, the basic correlation method uses a window with a fixed size and shape. However, this technique fails at object edges and boundaries. To overcome this problem, more relevant techniques have been proposed using adaptive window selection according to the intensity variation [Woo, Ortega, 2000; Veksler, 2001; Veksler, 2003; Ma *et al.*, 2009]. In [Veksler, 2001], Veksler estimates an appropriate window shape by optimizing over a large class of rectangular and non-rectangular windows. This approach is not efficient enough for real time implementation. To reduce the cost aggregation, Veksler uses in [Veksler, 2003] the integral image technique [Crow, 1984] for the window adjustment step. A multiple-window approach [Fusiello *et al.*, 2000a], uses for each pixel, nine distinct windows of the same size placed at different location, that are tried and the one providing highest correlation is retained. However, the number of windows is limited and cannot cover the whole range of different sizes needed. A difference-based adaptive searching windows approach is proposed in [Ma *et al.*, 2009]. The continuity of difference of stereo images is employed to compute the size of adaptive searching windows.

2.4.2 Global approaches

Global methods aim at reducing the sensitivity to local regions in the image. To compute the disparity map u , global methods aim at minimizing a global energy function over the whole image area. Such a function is composed of two terms and has the following form:

$$E(u) = E_d(u) + \lambda E_s(u) \quad (2.7)$$

where $E_d(u)$ is the data term that measures the disagreement between corresponding pixels, $E_s(u)$ is a regularization term that characterizes the smoothness between neighboring pixel pairs and λ is a positive constant weighting the two terms.

The stereo correspondence problem reduces to a minimization procedure. To this end, different energy minimization algorithms have been proposed to solve Eq (2.7). The most commonly used approaches are dynamic programming [Veksler, 2005; Kim *et al.*, 2005; Sun, 2002], graph cuts [Boykov *et al.*, 2001; Kolmogorov, Zabih, 2001] and variational methods [Miled *et al.*, 2006a; Miled *et al.*, 2009b; Kosov *et al.*, 2009; Chaux *et al.*, 2013; Hiltunen *et al.*, 2012; Wang, Yang, 2011].

Dynamic programming is one of the oldest and still popular strategies for stereo matching. It exploits the ordering and smoothness constraints to optimize correspondences of all scanlines in the image independently. The matching costs of all pixels on a scanline describe the disparity search space. The method is performed in two steps: the first one consists in building the path cost matrix for each possible candidate pixel and the second step aims at finding the best possible matching pixel with the minimum path cost. Although dynamic programming method produces better results on occlusion boundaries, it is unable to enforce smoothness in both horizontal and vertical directions, which leads to the well-known streaking artifacts. To solve this problem, a multiple-label dynamic programming [Criminisi *et al.*, 2007] has been developed.

In order to cope with this problem of streaking effect, Roy and Cox [Roy, Cox, 1998] have employed graph cuts to compute a global optimal solution with a single minimum cut through a constructed graph. Then, different approaches based on graph cuts have been developed [Kolmogorov, Zabih, 2001; Kolmogorov, Zabih, 2006; Komodakis *et al.*, 2011]. The idea is to consider the stereo matching problem as a pixel labelling problem to find the minimum cut through a certain graph.

Variational approaches have also been very effective for producing consistent disparity map

while preserving the depth discontinuities [Miled *et al.*, 2009a; Cremers *et al.*, 2011; Wanner, Goldluecke, 2014; Chaux *et al.*, 2013]. Indeed, these techniques were introduced in image processing for restoration and denoising problems [Combettes, Pesquet, 2004] and have already been attracted much interest in the computer vision where they were first designed for the purpose of estimating a dense optical flow from a sequence of images [Nagel, Enkelmann, 1986]. Numerical studies on optical flow have shown that the variational optical flow methods are among the most powerful techniques, which motivates their extension for disparity map estimation [Alvarez *et al.*, 2002; Slesareva *et al.*, 2005]. Variational methods minimize the energy function (Eq. (2.7)) by solving the associated non-linear Euler-Lagrange equation [Alvarez *et al.*, 2002].

While dynamic programming and graph cuts methods operate in a discrete manner, variational techniques work in a continuous space. Therefore, they allow to produce a disparity field with ideally infinite precision. Since all these global methods achieve satisfactory results in certain situations [Miled *et al.*, 2006b], they are often implemented using numerical schemes which may be computationally intensive. Moreover, a suitable choice of the regularization weight λ may be a difficult task.

While most of the existing works operate in the spatial domain, some wavelet based disparity estimation methods have been recently attracted much attention due to the hierarchical and scale-space localization properties of the wavelet [Chan *et al.*, 2006; Kosov *et al.*, 2009; Mukherjee *et al.*, 2010].

2.5 Conclusion

In this chapter, we have presented the principal characteristics of stereoscopic imaging system. We have reviewed the most frequently used techniques for estimating the disparity field. This is a crucial step for exploiting the cross-view redundancies and therefore achieving more efficient SI retrieval system.

The different ways that can be employed to exploit the disparity information for the retrieval of the stereo data will be the objective of the next chapter.

Chapter 3

Novel wavelet-based retrieval methods for stereo images

3.1 Introduction

With the emergence of wide range of multiview applications such as video games with autostereoscopic display and 3DTV or telepresence in videoconferences [Tzovaras *et al.*, 1999], many views of the same scene have been generated which has led to an impending need for both managing and storing these large amounts of image data. In order to reduce the storage requirements, a lot of images are saved in a compressed format by applying a wavelet transform to the original ones according to the JPEG2000 image coding standard. Therefore, it becomes interesting to design a CBIR system operating in the wavelet domain in order to avoid *some* decoding steps at the retrieval process. To this end, different wavelet-based retrieval methods for mono-view images have been reported [Mandal *et al.*, 1996; Do, Vetterli, 2002]. It is important to note that most of these techniques are dedicated to the retrieval of single view images [Kiranyaz, Gabbouj, 2012]. However, to the best of our knowledge, there is only one research work developed for the context of SI [Feng *et al.*, 2011]. More precisely, the reported method consists of two steps: a conventional CBIR system is applied to *only* one view (for example the left one). Then, the obtained results are refined by comparing the histograms of the estimated disparity maps. However, such retrieval method presents a drawback as the visual contents of the right image are not directly exploited. To alleviate this shortcoming, we investigate in this work different techniques to improve the efficiency of a CBIR system. Our main contribution is to exploit

the dependencies between the two views through a dense and smooth disparity map. The remainder of this chapter is organized as follows. In Section 3.2, we give a brief description of a conventional CBIR system operating in the WT domain and the straightforward extension of this system to the context of SI is discussed. Then, we motivate our choice to resort to a smooth and dense disparity estimation (DDE) method. The retained method, developed in [Miled *et al.*, 2009a], will also be described in Section 3.3. After that, in Sections 3.4, 3.5 and 3.6, we describe the proposed disparity-based SI retrieval approaches based on univariate, bivariate and multivariate modeling, respectively. Finally, the performance of the proposed approaches is illustrated in Section 3.7 and some conclusions are drawn in Section 3.8.

3.2 Conventional wavelet-based CBIR System

3.2.1 Wavelet distribution model

In the case of mono-view images, a wavelet-based CBIR system consists of extracting salient features from the wavelet coefficients of the images. As aforementioned, a statistical framework could be adopted to model the wavelet coefficients of the different subbands. Indeed, histograms of wavelet coefficients have been successfully used for representing color and local geometric properties [Smeulders *et al.*, 2000]. They capture local features to characterize homogeneous textures [Zhang, Xu, 1997]. Since the wavelet coefficients histograms are often monomodal and located around 0, the Generalized Gaussian distribution is recommended to model the marginal distributions [Antonini *et al.*, 1992; Smeulders *et al.*, 2000]. Thus, in a given subband j , the wavelet coefficients are viewed as realizations of a continuous zero-mean random variable W_j whose probability density function f_j is given by a Generalized Gaussian distribution:

$$\forall \xi \in \mathbb{R} \quad f_j(\xi) = \frac{\beta_j}{2\alpha_j\Gamma(1/\beta_j)} e^{-(|\xi|/\alpha_j)^{\beta_j}} \quad (3.1)$$

where $\Gamma(z) \triangleq \int_0^{+\infty} t^{z-1} e^{-t} dt$ represents the Gamma function, α_j and β_j are respectively the scale and shape parameters. Note that in the special case when $\beta_j = 2$ (resp. $\beta_j = 1$), the GG distribution corresponds to the Gaussian (resp. the Laplace) distribution.

3.2.2 Parameter estimation

Several techniques have been proposed in the literature to estimate the parameters α_j and β_j [Roenko *et al.*, 2014]. We can cite the three more popular methods:

- Moment-based method [Sharifi, Leon-Garcia, 1995]

The empirical moments $m_1 = E[|W_j|]$ and $m_2 = E[|W_j|^2]$ can be employed to estimate the parameters α_j and β_j by using the following relationship

$$\frac{m_2}{m_1^2} = \frac{\Gamma(\frac{1}{\beta_j})\Gamma(\frac{3}{\beta_j})}{\Gamma^2(\frac{2}{\beta_j})}. \quad (3.2)$$

A look-up table can be used for different values of β_j in order to determine the optimal value from the pair (m_1, m_2) [Sharifi, Leon-Garcia, 1995]. Once $\hat{\beta}_j$ is computed, $\hat{\alpha}_j$ can be deduced by using:

$$\hat{\alpha}_j = m_1 \frac{\Gamma(\frac{1}{\hat{\beta}_j})}{\Gamma(\frac{2}{\hat{\beta}_j})}. \quad (3.3)$$

- Maximum likelihood algorithm [Do, Vetterli, 2002]

The likelihood function of wavelet coefficients $w_j = (w_j(1), \dots, w_j(L_j))$ assuming that they have independent component is defined as:

$$\mathcal{L}(w_j; \alpha_j, \beta_j) = \log \prod_{i=1}^{L_j} f_j(w_j(i)) \quad (3.4)$$

where L_j is the number of wavelet coefficients in the subband j .

Maximizing the likelihood function \mathcal{L} is carried out by first setting equal to zero the first partial derivatives of \mathcal{L} with respect to α_j and β_j , and then, solving the following likelihood equations:

$$\frac{\partial \mathcal{L}(w_j; \alpha_j, \beta_j)}{\partial \alpha_j} = -\frac{L_j}{\alpha_j} + \sum_{i=1}^{L_j} \frac{\beta_j |w_j(i)|^{\beta_j} \alpha_j^{-\beta_j}}{\beta_j} = 0 \quad (3.5)$$

$$\frac{\partial \mathcal{L}(w_j; \alpha_j, \beta_j)}{\partial \beta_j} = \frac{L_j}{\beta_j} + \frac{L_j \Psi(1/\beta_j)}{\beta_j^2} - \sum_{i=1}^{L_j} \left(\frac{|w_j(i)|}{\alpha_j} \right)^{\beta_j} \log \left(\frac{|w_j(i)|}{\alpha_j} \right) = 0 \quad (3.6)$$

where $\Psi(\cdot)$ is the digamma function [Abramowitz, Stegun, 1970].

Fix $\beta_j > 0$ then Eq. (3.5) has a unique, real and, positive solution

$$\hat{\alpha}_j = \left(\frac{\beta_j}{L_j} \sum_{i=1}^{L_j} |w_j(i)|^{\beta_j} \right)^{1/\beta_j}. \quad (3.7)$$

Substitute (3.7) into (3.6), the shape parameter $\hat{\beta}_j$ is the solution of the following transcendental equation:

$$1 + \frac{\Psi(1/\hat{\beta}_j)}{\hat{\beta}_j} - \frac{\sum_{i=1}^{L_j} |w_j(i)|^{\hat{\beta}_j} \log |w_j(i)|}{\sum_{i=1}^{L_j} |w_j(i)|^{\hat{\beta}_j}} + \frac{\log \left(\frac{\hat{\beta}_j}{L_j} \sum_{i=1}^{L_j} |w_j(i)|^{\hat{\beta}_j} \right)}{\hat{\beta}_j} = 0 \quad (3.8)$$

which can be solved numerically by using for instance the Newton-Raphson iterative algorithm [Kay, 1993] with the initial value β_j^0 obtained from the moment method.

- Entropy matching estimators [Aiazzi *et al.*, 1999; Prasad *et al.*, 2005]

This class of estimators are based on matching the entropy of GGD distribution with that of wavelet coefficient of a subband j (empirical data). Differential entropy \mathcal{H} of GGD signal is defined as:

$$\mathcal{H} = \log \left(\frac{2}{\beta_j} \sqrt{\frac{\Gamma^3(1/\beta_j)}{\Gamma(3/\beta_j)}} \right) + \frac{1}{\beta_j \log 2}. \quad (3.9)$$

Let $\mathcal{H}(W_j)$ is the entropy value of the wavelet coefficients obtained at the output of an optimum entropy constrained uniform quantizer with step size Δ . Then $\mathcal{H}(W_j) = \mathcal{H} - \log \Delta$ and substituting the expression for \mathcal{H} leads to

$$\mathcal{H} = \mathcal{H}(W_j) - \frac{1}{2} \log \frac{m_2}{\Delta^2}. \quad (3.10)$$

β_j can be obtained as:

$$\beta_j = \mathcal{H}^{-1} \left(\mathcal{H}(W_j) - \frac{1}{2} \log \frac{m_2}{\Delta^2} \right). \quad (3.11)$$

In [Varanasi, Aazhang, 1989; Roenko *et al.*, 2014], the evaluation of the accuracy of the estimates of GGD parameters on large and small size of samples of several classic statistical methods shows that the best accuracy is achieved by the maximum likelihood estimator, especially for heavy-tailed distribution (which is often the case for subband coefficients). The great advantage of the moment-based estimators is their relative simplicity from the computation point of view. It has been proved in [Do, Vetterli, 2002] that the maximum likelihood estimator requires typically around three iteration steps to obtain solutions with good accuracy degrees. The entropy matching estimators have been found to be biased for $\beta_j > 1$ because of the high steepness of entropic functions.

In our work, we have estimated the GGD model parameters at each subband using the maximum likelihood algorithm to obtain the best accurate estimation. Hence, the feature

vector of each image of the database is composed of the distribution parameters of all the detail subbands $(\alpha_j, \beta_j)_{1 \leq j \leq 3J}$, where J is the number of resolution levels used with the wavelet decomposition.

3.2.3 Similarity measure

Finally, for the different subbands j with $j \in \{1, \dots, 3J\}$, an appropriate metric should be defined in order to measure the similarity between the probability density function (pdf) f_j^{db} of an image in the database I^{db} and the pdf f_j^{q} of the query image I^{q} . To this respect, different divergences have been defined to measure the difference between two probability density functions in information theory [Cover, Thomas, 1991]. Among them, the Kullback-Leibler Divergence (KLD) is often invoked [Do, Vetterli, 2002; Lasmar, Berthoumieu, 2010; Yuan, Zhang, 2004; Naornita *et al.*, 2012] in image retrieval purpose. Indeed, it has several advantages:

- It is robust with respect to possibly noisy data.
- It is numerically stable, (in particular according to the data size).

Recall that KLD is defined using a paradigm from source-coding since it provides the degradation of coding efficiency when the code is designed based on the first probability while the symbols are generated according to the other one [Cover, Thomas, 1991]. Thanks to the closed form of the KLD for several distributions, the similarity measure between two images involves simple computation using only model parameters. In the case of the GG distribution, the KLD, denoted by \tilde{D}_{GG} , is expressed as:

$$\tilde{D}_{\text{GG}}(\alpha_j^{\text{db}}, \beta_j^{\text{db}} \parallel \alpha_j^{\text{q}}, \beta_j^{\text{q}}) \triangleq \log \left(\frac{\beta_j^{\text{db}} \alpha_j^{\text{q}} \Gamma(1/\beta_j^{\text{q}})}{\beta_j^{\text{q}} \alpha_j^{\text{db}} \Gamma(1/\beta_j^{\text{db}})} \right) - \frac{1}{\beta_j^{\text{db}}} + \left(\frac{\alpha_j^{\text{db}}}{\alpha_j^{\text{q}}} \right)^{\beta_j^{\text{q}}} \frac{\Gamma((\beta_j^{\text{q}} + 1)/\beta_j^{\text{db}})}{\Gamma(1/\beta_j^{\text{db}})} \quad (3.12)$$

where $(\alpha_j^{\text{db}}, \beta_j^{\text{db}})$ and $(\alpha_j^{\text{q}}, \beta_j^{\text{q}})$ represent the distribution parameters of f_j^{db} and f_j^{q} , respectively. Thus, the resulting similarity measure D_{GG} between the two images I^{db} and I^{q} is deduced as follows:

$$D_{\text{GG}}(I^{\text{db}}, I^{\text{q}}) = \sum_{j=1}^{3J} \tilde{D}_{\text{GG}}(\alpha_j^{\text{db}}, \beta_j^{\text{db}} \parallel \alpha_j^{\text{q}}, \beta_j^{\text{q}}). \quad (3.13)$$

3.2.4 Intuitive extension to SI retrieval

Let us now proceed to the retrieval problem in the case of database composed of stereo images. A straightforward solution consists in *separately* applying the aforementioned

conventional CBIR system to each view. More precisely, the retrieval procedure aims at comparing the left and right images of the query stereo pair $(I^{(l,q)}, I^{(r,q)})$ to those of the database $(I^{(l,db)}, I^{(r,db)})$. Thus, after extracting their corresponding feature vectors $(\alpha_j^{(l,q)}, \beta_j^{(l,q)})_{1 \leq j \leq 3J}$, $(\alpha_j^{(r,q)}, \beta_j^{(r,q)})_{1 \leq j \leq 3J}$, $(\alpha_j^{(l,db)}, \beta_j^{(l,db)})_{1 \leq j \leq 3J}$ and $(\alpha_j^{(r,db)}, \beta_j^{(r,db)})_{1 \leq j \leq 3J}$, the similarity criterion $D_{GG}^{(r,l)}$ can be simply obtained by computing the KL divergences defined on the right and left images:

$$D_{GG}^{(r,l)} = D_{GG}(I^{(r,db)}, I^{(r,q)}) + D_{GG}(I^{(l,db)}, I^{(l,q)}). \quad (3.14)$$

3.3 Retained dense disparity estimation

3.3.1 Motivation

The disparity information is related to the depth structure of the scene and, hence, it is expected to provide salient features for the stereo retrieval system. Thus, our goal is to enrich features extracted from stereo images by incorporating those of the disparity maps. Among the existing disparity estimation techniques, it is important to use an efficient method that ensures the smoothness property of the produced disparity map because such property allows us to interpret this map as an image, and therefore can undergo a wavelet decomposition in order to be efficiently exploited in the retrieval process of the stereo images.

As mentioned in the previous chapter, unlike the local approaches which can not ensure the accuracy of the disparity map and preserve the discontinuities of the object boundaries, global methods offer the advantage of producing a smooth disparity map. In particular, it has been shown that variational-based disparity estimation methods yield a consistent disparity map with ideally infinite precision while preserving the depth discontinuities [Miled *et al.*, 2006a]. This motivates us to generate the disparity maps of the stereo image database by using the global convex variational framework presented in [Miled *et al.*, 2009a]. This retained disparity estimation method will be described in what follows.

3.3.2 Retained disparity estimation method

The principle of the retained Dense Disparity Estimation (DDE) method is to formulate the matching problem as a convex programming problem within a global variational framework. Indeed, a convex quadratic objective function is minimized over the intersection of specific

constraint sets. Then, the resulting optimization problem is solved via a block-iterative algorithm.

Energy function

Let $I^{(l)}$ and $I^{(r)}$ be the left and right intensity images of a stereo pair. We assume that these images are geometrically rectified, so that the disparity field is limited to the horizontal component that will be denoted by u . DDE methods aim at finding for each pixel (x, y) in the right image $I^{(r)}$ the best corresponding pixel in the left image $I^{(l)}$, which amounts to search the disparity field $\hat{\mathbf{u}} = (\hat{u}(x, y))_{(x, y) \in \mathcal{D}}$ that minimizes a given cost functional $\tilde{\mathcal{J}}$. The most commonly used criterion is the sum of squared intensity differences:

$$\begin{aligned} \hat{\mathbf{u}} = (\hat{u}(x, y))_{(x, y) \in \mathcal{D}} &= \arg \min_{u \in \Omega} \tilde{\mathcal{J}}(u) \\ &= \arg \min_{u \in \Omega} \sum_{(x, y) \in \mathcal{D}} \left[I^{(r)}(x, y) - I^{(l)}(x + u(x, y), y) \right]^2 \end{aligned} \quad (3.15)$$

where $\mathcal{D} = \{1, \dots, N_l\} \times \{1, \dots, N_c\}$ is the image support and Ω is the range of candidate disparity values. Assuming that an initial estimate \bar{u} of u is available (by using for example a correlation-based method) and the magnitude difference of fields u and \bar{u} is small enough, the warped left image around \bar{u} can be approximated by a Taylor expansion:

$$I^{(l)}(x + u, y) \simeq I^{(l)}(x + \bar{u}, y) + \nabla I^{(l)}(x + \bar{u}, y)(u - \bar{u}) \quad (3.16)$$

where $\nabla I^{(l)}(x + \bar{u}, y)$ is the horizontal gradient of the warped left image. Using this linearization (3.16), the criterion $\tilde{\mathcal{J}}$ in (3.15) can be approximated by a quadratic functional \mathcal{J} which is convex in u :

$$\mathcal{J}(u) = \sum_{s \in \mathcal{D}} [r(s) - L(s)u(s)]^2 \quad (3.17)$$

where s denotes the spatial position (x, y) in each image, and

$$\begin{aligned} L(s) &= \nabla I^{(l)}(x + \bar{u}, y) \\ r(s) &= I^{(r)}(s) - I^{(l)}(x + \bar{u}(s), y) + \bar{u}(s)L(s). \end{aligned}$$

In order to guarantee the convergence of the DE algorithm [Combettes, 2003], the objective function J must be strictly convex. For this reason, an additive term has been introduced in the criterion \mathcal{J} as follows:

$$J(\mathbf{u}) = \sum_{s \in \mathcal{D}} [r(s) - L(s)v(s)]^2 + \alpha \sum_{s \in \mathcal{D}} [u(s) - \bar{u}(s)]^2, \quad \text{with } \mathbf{u} = (u(s))_{s \in \mathcal{D}} \quad (3.18)$$

where α is a positive real number.

Minimizing this objective function is an ill-posed problem as the components of L may locally vanish. To solve this problem, appropriate convex constraints should be added so as to formulate the problem within a set theoretic framework [Combettes, 1993]. In what follows, we will give some examples of possible constraints that could be defined.

Convex constraints

The main concern of set theoretic estimation is to find a feasible solution efficiently satisfying various constraints arising from prior knowledge. Generally, each constraint is represented by a closed convex set S_k with $k \in \{1 \dots K\}$, in a Hilbert image space \mathbb{H} . The intersection S of all the K sets $(S_k)_{1 \leq k \leq K}$ forms the family of admissible solutions [Miled *et al.*, 2006a; Miled *et al.*, 2009a]. Consequently, the disparity estimation problem is equivalent to find an acceptable solution in S which minimizes the underlying objective function \mathcal{J} :

$$\text{Find } \hat{u} \in S = \bigcap_{k=1}^K S_k \text{ such that } \mathcal{J}(\hat{u}) = \inf_{u \in S} \mathcal{J}(u). \quad (3.19)$$

The constraint sets can be represented as level sets:

$$\forall k \in \{1, \dots, K\}, \quad S_k = \{u \in \mathbb{H} \mid f_k(u) \leq \delta_k\} \quad (3.20)$$

where $f_k : \mathbb{H} \rightarrow \mathbb{R}$ is a continuous convex function for every $k \in \{1, \dots, K\}$ and $(\delta_k)_{1 \leq k \leq K}$ are real parameters such that $S = \bigcap_{k=1}^K S_k \neq \emptyset$.

In what follows, we will define the convex sets $(S_k)_{1 \leq k \leq K}$ of \mathbb{H} so as they reflect the expected properties of the disparity field to be estimated. We focus on two possible constraints ($K = 2$). The first one is the range of the disparity values. We can restrict the variation of the disparity values within a specific range $[u_{\min}, u_{\max}]$. Consequently, the corresponding constraint set S_1 is defined by:

$$S_1 = \{u \in \mathbb{H} \mid u_{\min} \leq u \leq u_{\max}\}. \quad (3.21)$$

Furthermore, in order to obtain a disparity map with smooth homogeneous areas while preserving edges, another constraint could be added. The total variation measure $\text{tv}(u)$, defined as the sum over \mathcal{D} of the norm of the spatial gradient of u [Rudin *et al.*, 1992], has been recently recognized to be an effective tool to recover smooth images in various

research fields. We recall that the total variation of the discrete disparity image \mathbf{u} is given by:

$$\begin{aligned}
tv(\mathbf{u}) &= \sum_{x=1}^{N_1-1} \sum_{y=1}^{N_c-1} \sqrt{|u(x+1, y) - u(x, y)|^2 + |u(x, y+1) - u(x, y)|^2} \\
&\quad + \sum_{x=1}^{N_1-1} |u(x+1, N_c) - u(x, N_c)| \\
&\quad + \sum_{y=1}^{N_c-1} |u(N_1, y+1) - u(N_1, y)|.
\end{aligned} \tag{3.22}$$

Hence, a total variation based regularization constraint involves to impose an upper bound τ on tv :

$$S_2 = \{u \in \mathbb{H} \mid tv(u) \leq \tau\}. \tag{3.23}$$

It should be noted that the positive constant τ can be estimated from a scale value of the total variation of the initial disparity map \bar{u} . It was shown in [Kaaniche, 2010] that lower scale values produce more smoothing disparity map.

Therefore, the disparity estimation problem can be formulated as a constrained optimization problem where the quadratic objective function \mathcal{J} in (3.19) is minimized over the feasibility set $S = S_1 \cap S_2$. Several optimization methods have been proposed to solve this problem [Boyle, Dykstra, 1986; Yamada *et al.*, 1998]. We employ a parallel block iterative algorithm developed in [Combettes, 2003]. This algorithm enables to achieve real-time performance and offers a great flexibility in the incorporation of several constraints. The optimization algorithm developed in [Miled *et al.*, 2009a] and some basic facts on subgradient projections are described in Appendix A.

Examples of estimated disparity maps

In order to show the benefit of using the retained DDE approach, we illustrate in Figures 3.1 and 3.2 the “Shrub” and “Teddy” stereo image pair, respectively, and their corresponding disparity maps provided by the block-based disparity estimation method and the retained DDE method. It can be noticed that the DDE method generates a smooth map while preserving the depth discontinuities.

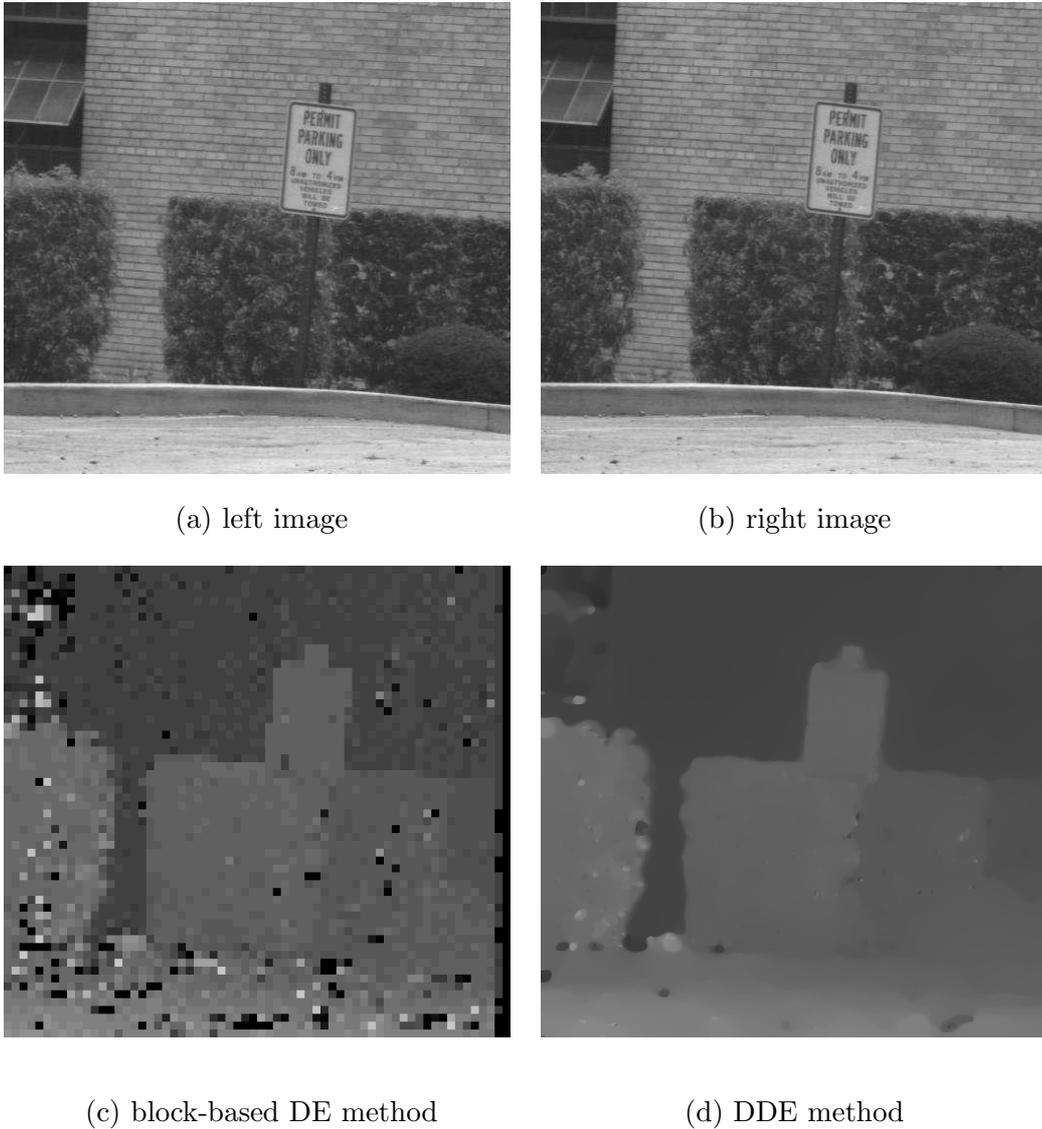


Figure 3.1: The obtained disparity maps for the “Shrub” stereo image using (c) block-based DE method and (d) the considered DDE method.

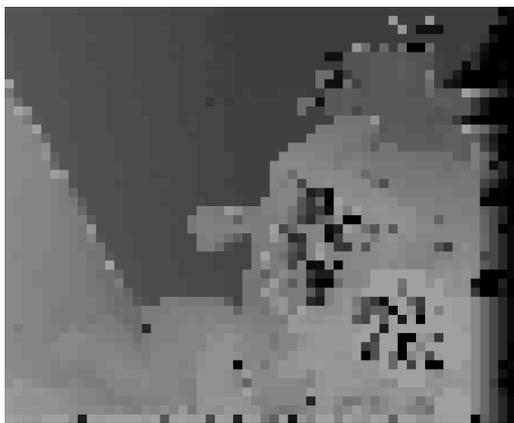
3.4 Proposed disparity-based retrieval approaches through a univariate model

The aforementioned intuitive strategy for indexing a SI pair is not so efficient since only the left and right images are used during the comparison process. Indeed, an important feature of the stereoscopic system, which corresponds to the estimated disparity map, has not been taken into account. A more efficient retrieval method could be designed by incorporating the disparity information in the feature vector. Thanks to the retained DDE method, the

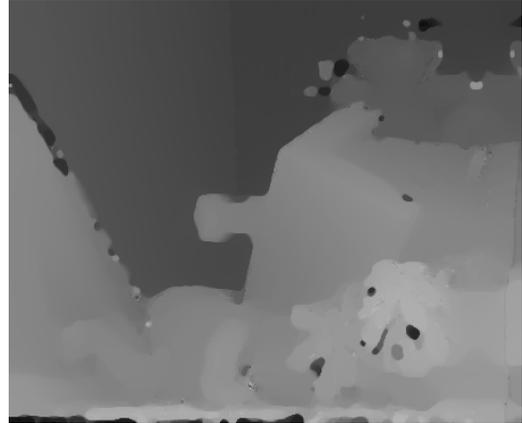


(a) left image

(b) right image



(c) block-based DE method



(d) DDE method

Figure 3.2: The obtained disparity maps for the “Teddy” stereo image using (c) block-based DE method and (d) the considered DDE method.

produced disparity map can be interpreted as an image, and therefore a wavelet transform can be applied to it in order to be exploited in the retrieval of stereo images. In what follows, we describe two novel disparity-based retrieval approaches relying on a univariate statistical model.

3.4.1 Stereo images-based modeling

In the first approach, the disparity is *implicitly* taken into account by computing the *compensated* left image in the wavelet domain $I_j^{(c)}$ from the multiresolution representation

of the left image $I_j^{(l)}$ as follows:

$$I_j^{(c)}(x, y) = I_j^{(l)}(x + u_j(x, y), y) \quad (3.24)$$

where the disparity u_j is obtained by sampling and dividing by 2^j the initial disparity field u :

$$u_j(x, y) = \frac{1}{2^j} u(2^j x, 2^j y). \quad (3.25)$$

Then, the GG distribution parameters of the different detail subbands of the right image $(\alpha_j^{(r)}, \beta_j^{(r)})_{1 \leq j \leq 3J}$ and the compensated left one $(\alpha_j^{(c)}, \beta_j^{(c)})_{1 \leq j \leq 3J}$ are extracted. Finally, the retrieval procedure for a given query stereo pair $(I^{(l,q)}, I^{(r,q)})$ aims at finding the best stereo pairs $(I^{(l,db)}, I^{(r,db)})$ that minimize the KL divergences, $D_{GG}^{(r,c)}$, defined on the right image and the compensated left one:

$$D_{GG}^{(r,c)} = D_{GG}(I^{(r,db)}, I^{(r,q)}) + D_{GG}(I^{(c,db)}, I^{(c,q)}) \quad (3.26)$$

where $I^{(c,q)}$ and $I^{(c,db)}$ represent respectively the compensated left images of the query and candidate stereo pairs.

3.4.2 Stereo images and disparity-based modeling

Unlike the first approach, the second one aims to exploit *explicitly* the disparity information by extracting a relevant signature from the disparity map and, combining it with the features defined previously on the SI pair. To this end, since a smooth disparity map is produced while preserving the depth discontinuities, we propose to apply a wavelet transform to the estimated disparity field. After performing an intensive experiments on a large data set of the estimated disparity maps, we have noticed that their wavelet coefficients can also be successfully modeled by a GG distribution. Fig. 3.3 shows a block diagram of the proposed approach.

Indeed, to objectively assess the appropriateness of the GG model, we have applied the Kolmogorov-Smirnov (KS) goodness-of-fit test [Massey, Frank, 1951; Justel *et al.*, 1997]. Note that the KS test is based on comparing the cumulative distribution functions. As an example, by taking three disparity maps and considering their horizontal detail coefficients, Fig. 3.4 shows the histograms of these coefficients (in blue) and the fitted GG distributions (in red) as well as their resulting KS measures. By performing this test on all the disparity

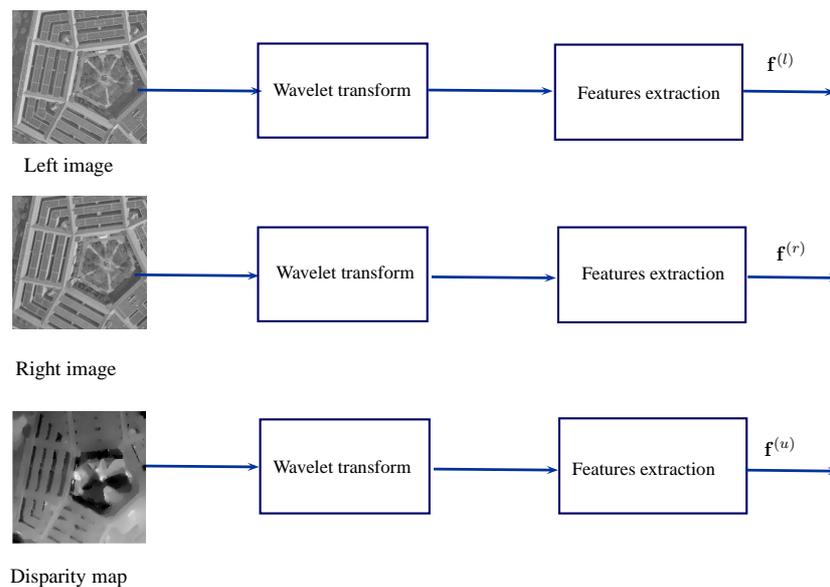


Figure 3.3: Stereo images and disparity-based modeling approach.

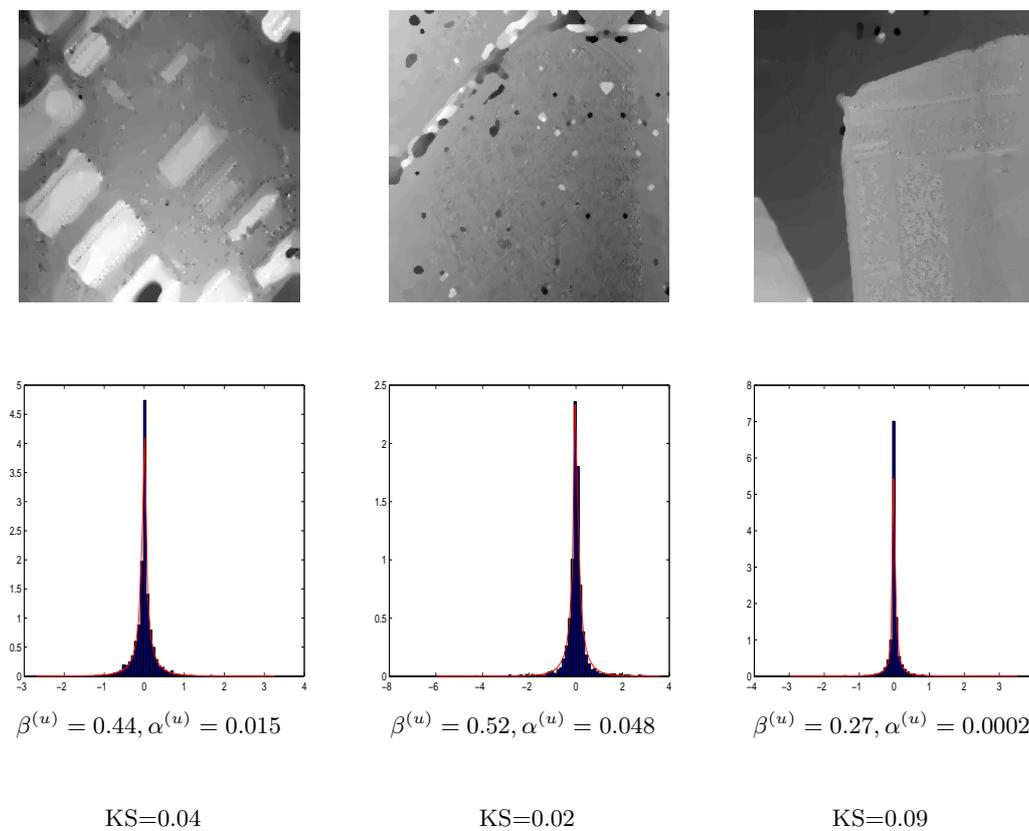


Figure 3.4: Modeling the distribution of the horizontal detail subband at the second resolution level for three examples of disparity maps using the GG model.

maps of the data set, the average of the KS values is equal to 0.1 which is very small. This confirms that the GG distribution is well-suited for modeling the disparity maps.

Based on these observations, we select the distribution parameters of all the resulting detail subbands $(\alpha_j^{(u)}, \beta_j^{(u)})_{1 \leq j \leq 3J}$ to characterize the disparity map u . Therefore, at the retrieval step, the candidate stereo pairs of the database $(I^{(r,db)}, I^{(l,db)})$ that are similar to the query one $(I^{(r,q)}, I^{(l,q)})$ are determined by comparing the right and left images as well as their associated disparity maps $u^{(db)}$ and $u^{(q)}$. More precisely, we propose to define the similarity measure as follows:

$$\begin{aligned} D_{GG}^{(r,l,u)} &= a \cdot D_{GG}(I^{(r,db)}, I^{(r,q)}) + b \cdot D_{GG}(I^{(l,db)}, I^{(l,q)}) + c \cdot D_{GG}(u^{(db)}, u^{(q)}) \\ &= \sum_{j=1}^{3J} \left(a \cdot \tilde{D}_{GG}(\alpha_j^{(r,db)}, \beta_j^{(r,db)} \parallel \alpha_j^{(r,q)}, \beta_j^{(r,q)}) \right. \\ &\quad \left. + b \cdot \tilde{D}_{GG}(\alpha_j^{(l,db)}, \beta_j^{(l,db)} \parallel \alpha_j^{(l,q)}, \beta_j^{(l,q)}) \right. \\ &\quad \left. + c \cdot \tilde{D}_{GG}(\alpha_j^{(u,db)}, \beta_j^{(u,db)} \parallel \alpha_j^{(u,q)}, \beta_j^{(u,q)}) \right) \end{aligned} \quad (3.27)$$

where a , b and c are three positive weights.

To conclude this part, we should note that this first category of SI retrieval approaches are based on univariate model since, up to now, the right and left images are modeled separately without taking into account the cross-view dependencies.

3.5 Proposed disparity-based retrieval approaches through bivariate model

3.5.1 Motivation

Since the left and right views correspond to the same scene, and so have similar contents, the wavelet coefficients of both images could present strong statistical dependencies. For example, such dependencies between the two images have already been successfully exploited for SI compression purpose [Kaaniche *et al.*, 2009a; Woo, Ortega, 1997]. Therefore, a suitable statistical model should be employed to capture the dependencies across the wavelet coefficients of the left and right images.

3.5.2 Bivariate generalized Gaussian model

First, we should recall that recent works on color image retrieval [Sakji-Nsibi, Benazza-Benyahia, 2009; Kwitt *et al.*, 2011; Verdoolaege, Scheunders, 2011; Lasmar, Berthoumieu,

[2010] have shown that the distribution of the wavelet coefficients could be modeled by some specific multivariate models. The latter outperform the conventional univariate approaches in terms of accuracy since they account for both the spatial and the cross-channel dependencies. Motivated by these reasons, we propose in this work to resort to a bivariate model to further exploit the dependencies between the wavelet coefficients of the left and right images. Generally, the choice of the appropriate model should fulfill the following constraints:

- The retained model should reflect accurately the sparsity of the wavelet coefficients of each view and also the cross-view dependencies.
- The model structure should enable a straightforward estimation of the parameters from the coefficients in each subband.
- The bivariate model should allow an easy computation of a meaningful similarity measure. For example, knowing that the KLD has been widely used in mono-view image retrieval [Do, Vetterli, 2002; Lasmar, Berthoumieu, 2010; Yuan, Zhang, 2004; Naornita *et al.*, 2012], it would be interesting to select a bivariate model from which a closed form expression of the KLD could be derived in order to facilitate the retrieval stage by avoiding Monte-Carlo estimation procedures.

Based on the previous points, we find that using the well-known Bivariate Generalized Gaussian (BGG) distribution could be an appropriate way for characterizing the dependencies between the wavelet coefficients of the left and right images.

To introduce this model, let us denote by \mathbf{w}_j the bivariate vector composed of the wavelet coefficients of two correlated components, for each subband j . We assume that the set of coefficient vectors \mathbf{w}_j in each subband constitutes an independent identical distributed sample of a random vector \mathbf{W}_j .

Under the hypothesis of a zero-mean vector, the generic expression of the probability density function $f_{\mathbf{W}_j}$ of the BGG distribution is given by:

$$\forall \mathbf{w} \in \mathbb{R}^2, f_{\mathbf{W}_j}(\mathbf{w}; \beta_j, \mathbf{\Sigma}_j) = \frac{2}{\pi \Gamma(1 + \frac{1}{\beta_j}) 2^{1 + \frac{1}{\beta_j}}} |\mathbf{\Sigma}_j|^{-1/2} e^{\left(-\frac{1}{2}(\mathbf{w}^T \mathbf{\Sigma}_j^{-1} \mathbf{w})^{\beta_j}\right)}. \quad (3.28)$$

where $\beta_j > 0$ denotes the shape parameter, and $\mathbf{\Sigma}_j$ is a positive definite symmetric matrix (the scale matrix) of size 2×2 . where $\beta_j > 0$ denotes the shape parameter and, $\mathbf{\Sigma}_j$ is a

symmetric positive-definite matrix of size 2×2 (the scaling matrix). It is easy to show that the autocovariance matrix \mathbb{V}_j is related to $\mathbf{\Sigma}_j$ and, β_j as follows:

$$\mathbb{V}_j = \frac{2^{\frac{1}{\beta_j}-1} \Gamma(\frac{2}{\beta_j})}{\Gamma(\frac{1}{\beta_j})} \mathbf{\Sigma}_j \quad (3.29)$$

Hence, $\mathbf{\Sigma}_j$ is proportional to the covariance matrix. Thus, the problem reduces to estimate a 2×2 matrix $\mathbf{\Sigma}_j$ and the shape parameter β_j . These parameters can be estimated using the moment method [Gómez *et al.*, 2003] or the maximum likelihood criterion [Verdoolaeghe, Scheunders, 2011]. In [Kwitt *et al.*, 2009], it has been shown that moment method is numerically more stable and a computationally inexpensive way. In the following paragraph, we will describe both estimation techniques.

- **Moment estimation method**

The idea is to estimate the autocovariance matrix \mathbb{V}_j and Mardia's multivariate kurtosis coefficient [Zografos, 2008; Mardia, 1970] by their classical empirical estimates. First, the shape parameter $\hat{\beta}_j$ is determined and then used to compute $\hat{\mathbf{\Sigma}}_j$. We recall that Mardia's multivariate kurtosis coefficient γ_2 is defined as:

$$\gamma_2 = \mathbb{E} \left[(\mathbf{W}_j^\top \mathbf{\Sigma}_j \mathbf{W}_j)^2 \right] - 8. \quad (3.30)$$

It has a closed-form expression in the case of BGG distribution given by:

$$\gamma_2 = \frac{4\Gamma(\frac{1}{\beta_j})\Gamma(\frac{3}{2\beta_j})}{\Gamma^2(\frac{2}{\beta_j})} - 8. \quad (3.31)$$

Given that \mathbf{S}_j denotes the classic sample autocovariance matrix, γ_2 can be estimated by $\hat{\gamma}_2$

$$\hat{\gamma}_2(\mathbf{w}_j(1), \dots, \mathbf{w}_j(L_j)) = \frac{1}{L_j} \sum_{\ell=1}^{L_j} \left(\mathbf{w}_j(\ell)^\top \mathbf{S}_j^{-1} \mathbf{w}_j(\ell) \right)^2 - 8. \quad (3.32)$$

where L_j denotes the number of wavelet coefficients in each subband. By matching Eqs. (3.31) and (3.32), we can then derive the estimate $\hat{\beta}_j$. Using this result together with the expression for the variance \mathbb{V} , Eq. (3.29) allows to calculate $\mathbf{\Sigma}_j$. Then, the matrix $\mathbf{\Sigma}_j$ can be estimated from $\hat{\beta}_j$ and \mathbf{S}_j .

- **Maximum likelihood estimation**

The likelihood function $\tilde{\mathcal{L}}$ can be easily obtained from the definition of $f_{\mathbf{w}_j}$ given in Eq. (3.28) :

$$\tilde{\mathcal{L}}(\beta_j, \boldsymbol{\Sigma}_j; \mathbf{w}_j(1), \dots, \mathbf{w}_j(L_j)) = \prod_{l=1}^{L_j} \frac{2}{\pi \Gamma(1 + \frac{1}{\beta_j}) 2^{1 + \frac{1}{\beta_j}}} |\boldsymbol{\Sigma}_j|^{-1/2} e^{\left(-\frac{1}{2}(\mathbf{w}_j(l)^\top \boldsymbol{\Sigma}_j^{-1} \mathbf{w}_j(l))^{\beta_j}\right)} \quad (3.33)$$

Taking the logarithm leads to the following expression:

$$\begin{aligned} \check{\mathcal{L}}(\beta_j, \boldsymbol{\Sigma}_j; \mathbf{w}_j(1), \dots, \mathbf{w}_j(L_j)) &= L_j \log \Gamma(1) - L_j \log \Gamma\left(\frac{1}{\beta_j}\right) + L_j \log(\beta_j) \\ &\quad - L_j \log(\pi) - \frac{2L_j}{\beta_j} \log(2) - \frac{L_j}{2} \log(|\boldsymbol{\Sigma}_j|) - \frac{1}{2} \sum_{l=1}^{L_j} (\mathbf{w}_j(l)^\top \boldsymbol{\Sigma}_j^{-1} \mathbf{w}_j(l))^{\beta_j}. \end{aligned} \quad (3.34)$$

It can be used to compute the partial derivatives with respect to β_j and $\boldsymbol{\Sigma}_j$:

$$\begin{aligned} \frac{\partial}{\partial \beta_j} \check{\mathcal{L}}(\beta_j, \boldsymbol{\Sigma}_j; \mathbf{w}_j(1), \dots, \mathbf{w}_j(L_j)) &= \frac{1}{\beta_j} \left[L_j + \frac{2L_j}{\beta_j} \left(\log(2) + \Psi\left(\frac{1}{\beta_j}\right) \right) \right] \\ &\quad - \frac{1}{2} \sum_{l=1}^{L_j} \log(\mathbf{w}_j(l)^\top \boldsymbol{\Sigma}_j^{-1} \mathbf{w}_j(l)) (\mathbf{w}_j(l)^\top \boldsymbol{\Sigma}_j^{-1} \mathbf{w}_j(l))^{\beta_j} \end{aligned} \quad (3.35)$$

and,

$$\frac{\partial}{\partial \boldsymbol{\Sigma}_j} \check{\mathcal{L}}(\beta_j, \boldsymbol{\Sigma}_j; \mathbf{w}_j(1), \dots, \mathbf{w}_j(L_j)) = -\frac{L_j}{2} \boldsymbol{\Sigma}_j^{-1} + \frac{\beta_j}{2} (\mathbf{w}_j(l)^\top \boldsymbol{\Sigma}_j^{-1} \mathbf{w}_j(l))^{\beta_j-1} \boldsymbol{\Sigma}_j^{-1} \mathbf{w}_j(l) \mathbf{w}_j(l)^\top \boldsymbol{\Sigma}_j^{-1} \quad (3.36)$$

where Ψ denotes the digamma function. It is worth noting, that after setting the right-hand side of Eq. (3.36) to zero and performing some simple manipulations, we have:

$$\hat{\boldsymbol{\Sigma}}_j = \frac{\hat{\beta}_j}{L_j} \sum_{l=1}^{L_j} \mathbf{w}_j(l) \mathbf{w}_j(l)^\top (\mathbf{w}_j(l)^\top \hat{\boldsymbol{\Sigma}}_j^{-1} \mathbf{w}_j(l))^{\hat{\beta}_j-1} \quad (3.37)$$

which allows to employ a fixed-point iteration directly (e.g. Picard iteration). However, it is difficult to prove that Eq. (3.37) is a contraction to guarantee convergence of the fixed-point.

In [Kwitt, 2010], an alternative technique is employed, which is based on minimizing the negative log-likelihood $-\check{\mathcal{L}}(\beta_j, \boldsymbol{\Sigma}_j; \mathbf{w}_j(1), \dots, \mathbf{w}_j(L_j))$ by using a gradient descent approach. This corresponds to an optimization problem with non-linear constraints, since $\boldsymbol{\Sigma}_j$ must be symmetric positive definite and $\beta_j > 0$. We have the derivatives of the log-likelihood function with respect to β_j and $\boldsymbol{\Sigma}_j$ (Eqs. (3.35) and (3.36)). For the positive definiteness

criteria, the Sylvester criterion [Meyer, 2000] which requires that all leading principal minors of Σ_j are positive can be employed. This is a sufficient condition to guarantee positive definiteness.

Once the estimation problem is solved, it is logical to address the issue of the validity of the BGG model for SIs.

- **Validity of the BGG for SIs**

In order to validate the appropriateness of the BGG model, we have conducted the multivariate Kolmogorov-Smirnov (KS) test [Justel *et al.*, 1997] on the stereo images database. Indeed, by taking the horizontal detail subbands of the left and right images of four stereo pairs, Fig. 3.5 illustrates the empirical bivariate histograms (in blue) fitted with the BGG distribution (in red), and provides their associated KS measures. By repeating the same test on the whole set of stereo images in the database, an average KS value of about 0.09 is obtained. These results corroborate also that stereo wavelet subbands can be well modeled by a BGG distribution.

- **KLD for BGG distributions**

It is important to note that a closed form expression of the KLD is available for such BGG model. Indeed, for two zero-mean BGG distributions with parameters (β_j^q, Σ_j^q) and $(\beta_j^{db}, \Sigma_j^{db})$, the KLD is given by [Verdoolaege, Scheunders, 2011; Verdoolaege *et al.*, 2009]:

$$\begin{aligned} \tilde{D}_{\text{BGG}}(\beta_j^{db}, \Sigma_j^{db} \parallel \beta_j^q, \Sigma_j^q) = & \log \left[\frac{\Gamma\left(\frac{1}{\beta_j^q}\right)}{\Gamma\left(\frac{1}{\beta_j^{db}}\right)} 2^{\left(\frac{1}{\beta_j^q} - \frac{1}{\beta_j^{db}}\right)} \left(\frac{|\Sigma_j^q|}{|\Sigma_j^{db}|}\right)^{\frac{1}{2}} \frac{\beta_j^{db}}{\beta_j^q} \right] - \frac{1}{\beta_j^{db}} \\ & + \left[2^{\left(\frac{\beta_j^q}{\beta_j^{db}} - 1\right)} \frac{\Gamma\left(\frac{\beta_j^q + 1}{\beta_j^{db}}\right)}{\Gamma\left(\frac{1}{\beta_j^{db}}\right)} \times \left(\frac{\mu_1 + \mu_2}{2}\right)^{\beta_j^q} \right. \\ & \left. \times {}_2F_1\left(\frac{1 - \beta_j^q}{2}, \frac{-\beta_j^q}{2}; 1; \left(\frac{\mu_1 - \mu_2}{\mu_1 + \mu_2}\right)^2\right) \right] \end{aligned} \quad (3.38)$$

where μ_1 and μ_2 are the inverse of the eigenvalues of $(\Sigma_j^{db})^{-1}\Sigma_j^q$ and ${}_2F_1$ represents the Gauss hypergeometric function [Abramowitz, Stegun, 1970].

3.5.3 Improved disparity-based retrieval strategies

Now, we will describe three retrieval strategies based on the BGG distribution.

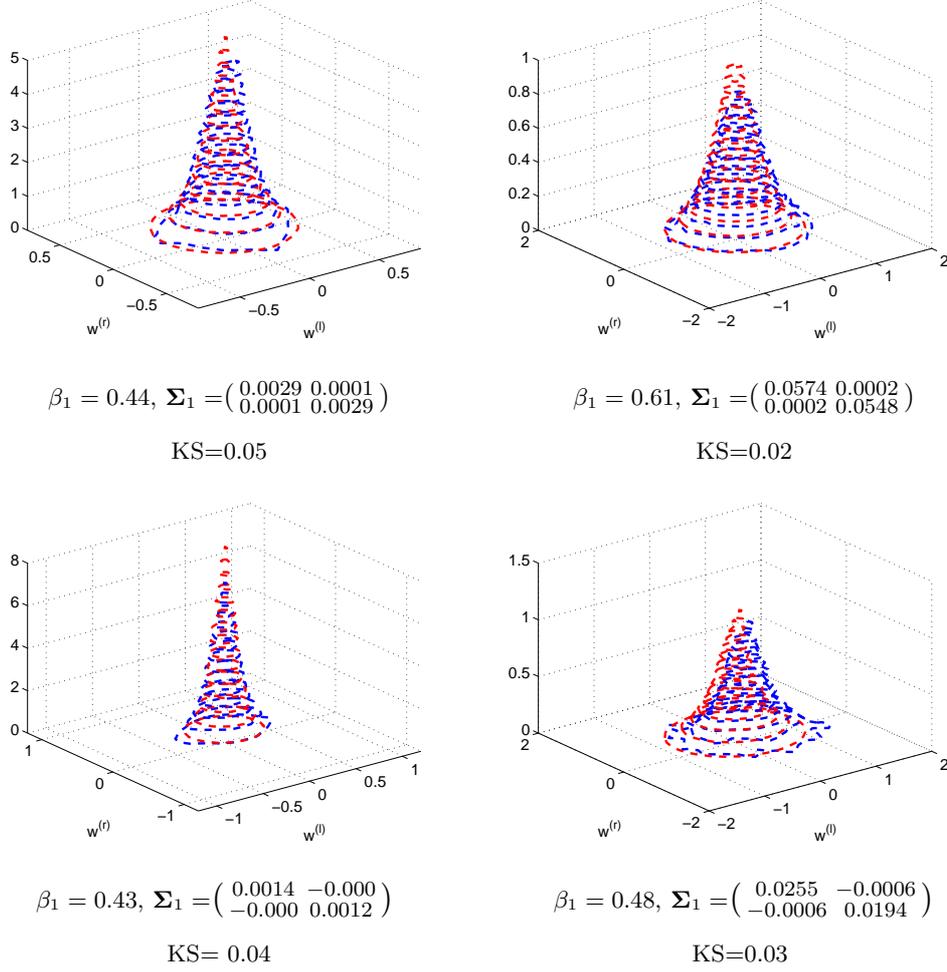


Figure 3.5: Empirical bivariate histogram of the horizontal wavelet coefficients of the left and right images (denoted here by $w^{(l)}$ and $w^{(r)}$) (in blue) fitted with a BGG density (in red) for four different stereo images as well as their resulting KS measures.

First strategy

Intuitively, in the first one, the bivariate vector, defined by $\mathbf{w}_j^{(r,l)} = \left(w_j^{(r)}, w_j^{(l)} \right)^\top$ is composed of the wavelet coefficients in the right and left images for each subband j . Let us denote by $\left(\Sigma_j^{(r,l)}, \beta_j^{(r,l)} \right)$ the distribution parameters of vector $\mathbf{w}_j^{(r,l)}$. Thus, in the indexing step, the comparison between the stereo pair in the database, characterized by its feature vector $\left(\beta_j^{(r,l,db)}, \Sigma_j^{(r,l,db)} \right)_{1 \leq j \leq 3J}$ and the query one parameterized by the feature $\left(\beta_j^{(r,l,q)}, \Sigma_j^{(r,l,q)} \right)_{1 \leq j \leq 3J}$ is achieved by computing the global KLD:

$$D_{\text{BGG}}^{(r,l)} = \sum_{j=1}^{3J} \tilde{D}_{\text{BGG}}(\beta_j^{(r,l,db)}, \Sigma_j^{(r,l,db)} \parallel \beta_j^{(r,l,q)}, \Sigma_j^{(r,l,q)}). \quad (3.39)$$

Second strategy

Since the highly similar pixels of the left and right images are located at different spatial positions identified by the disparity information, it would be more interesting to focus on the right image $I^{(r)}$ and the compensated left one $I^{(c)}$ by using their wavelet coefficients to build the bivariate vector $\mathbf{w}_j^{(r,c)} = \left(w_j^{(r)}, w_j^{(c)} \right)^\top$, for each subband j . Thus, in this second retrieval strategy, the feature vectors deduced from the right and compensated left images of the query stereo data $\left(\beta_j^{(r,c,q)}, \boldsymbol{\Sigma}_j^{(r,c,q)} \right)_{1 \leq j \leq 3J}$ will be compared to those of the database stereo pair $\left(\beta_j^{(r,c,db)}, \boldsymbol{\Sigma}_j^{(r,c,db)} \right)_{1 \leq j \leq 3J}$ by computing the following measure:

$$D_{\text{BGG}}^{(r,c)} = \sum_{j=1}^{3J} \tilde{D}_{\text{BGG}}(\beta_j^{(r,c,db)}, \boldsymbol{\Sigma}_j^{(r,c,db)} \parallel \beta_j^{(r,c,q)}, \boldsymbol{\Sigma}_j^{(r,c,q)}). \quad (3.40)$$

Third strategy

Although it is clear that stereo images contain nearly similar contents since they correspond to the same scene, there are some occluded areas due to the different viewpoints of the cameras and the presence of discontinuities in the scene. Generally, increasing the dependencies between the two components of the bivariate vector leads to an efficient retrieval procedure. For this reason, we propose in the third strategy to improve the previous one by taking into account the effect of the occlusions. Recall that the occluded areas are mainly located at the object boundaries of the SI. Let us denote by $w_j^{(r_{\text{ocl}})}$ and $w_j^{(c_{\text{ocl}})}$ the wavelet coefficients of the right and compensated left images resulting from the removal of the occluded regions. These coefficients will constitute the components of the bivariate vector $\mathbf{w}_j^{(r_{\text{ocl}},c_{\text{ocl}})} = \left(w_j^{(r_{\text{ocl}})}, w_j^{(c_{\text{ocl}})} \right)^\top$. After estimating their associated model parameters and building the feature vectors for the query stereo pair $\left(\beta_j^{(r_{\text{ocl}},c_{\text{ocl}},q)}, \boldsymbol{\Sigma}_j^{(r_{\text{ocl}},c_{\text{ocl}},q)} \right)_{1 \leq j \leq 3J}$ and the candidate one $\left(\beta_j^{(r_{\text{ocl}},c_{\text{ocl}},db)}, \boldsymbol{\Sigma}_j^{(r_{\text{ocl}},c_{\text{ocl}},db)} \right)_{1 \leq j \leq 3J}$, the global KLD is therefore obtained:

$$D_{\text{BGG}}^{(r_{\text{ocl}},c_{\text{ocl}})} = \sum_{j=1}^{3J} \tilde{D}_{\text{BGG}}(\beta_j^{(r_{\text{ocl}},c_{\text{ocl}},db)}, \boldsymbol{\Sigma}_j^{(r_{\text{ocl}},c_{\text{ocl}},db)} \parallel \beta_j^{(r_{\text{ocl}},c_{\text{ocl}},q)}, \boldsymbol{\Sigma}_j^{(r_{\text{ocl}},c_{\text{ocl}},q)}). \quad (3.41)$$

Finally, as it was performed with disparity-based retrieval approaches through univariate model, these three strategies should further incorporate the disparity information into their feature vector. In other words, in addition to the BGG distribution parameters of the right and left (or compensated left) images, it would be interesting to consider also the GG distribution parameters of the disparity maps $u^{(db)}$ and $u^{(q)}$ respectively associated to the

database and query SI. Consequently, during the comparison process between a query and a candidate stereo pair, we should add the measure $D_{\text{GG}}(u^{(\text{db})}, u^{(\text{q})})$ to each of the KL divergences given by Eqs. (3.39), (3.40) and (3.41).

3.6 Proposed retrieval approach through SIRV modeling

3.6.1 Motivation

In the previous sections, the strong similarities between the wavelet coefficients of the left and right images have been efficiently exploited by using the disparity information and through a bivariate model. In addition to the cross-view similarities, the spatial redundancies in each subband could be further exploited and not omitted as in the case of the statistical model BGG. For this reason, we propose in what follows to resort to an appropriate multivariate statistical model to better capture these two kinds of dependencies. We should note that different multivariate statistical models have already been reported in the literature [Kwitt *et al.*, 2011; Verdoolaege, Scheunders, 2011; Sakji-Nsibi, Benazza-Benyahia, 2009]. Among them, the Multivariate Generalized Gaussian (MGG) distribution was found to be an excellent descriptor for the heavy-tailed wavelet distributions in multi-component images, such as color or multispectral images. To introduce this model, let us denote by $\tilde{\mathbf{w}}_j$ the multivariate vector composed of the wavelet coefficients of B statistically dependent components, in each subband j . We assume that the set of coefficients vectors $\tilde{\mathbf{w}}_j$ constitutes an independent identical distributed sample of a random vector $\tilde{\mathbf{W}}_j$.

The probability density function $f_{\tilde{\mathbf{W}}_j}$ of the MGG distribution is given by:

$$\forall \mathbf{w} \in \mathbb{R}^n, f_{\tilde{\mathbf{W}}_j}(\mathbf{w}; \beta_j, \boldsymbol{\Sigma}_j) = \frac{n\Gamma(\frac{n}{2})}{\pi^{n/2}\Gamma(1 + \frac{n}{2\beta_j})2^{1 + \frac{n}{2\beta_j}}} |\boldsymbol{\Sigma}_j|^{-1/2} e^{\left(-\frac{1}{2}(\mathbf{w}^T \boldsymbol{\Sigma}_j^{-1} \mathbf{w})^{\beta_j}\right)}. \quad (3.42)$$

where n is the dimensionality of the probability space, $\beta_j > 0$ denotes the shape parameter and, $\boldsymbol{\Sigma}_j$ is a symmetric positive-definite matrix of size $n \times n$ (the scaling matrix). However, computing the KLD between two MGG distributions implies solving a multidimensional integral over the data space. Hence, it is difficult to find a closed-form expression of the KLD and, if we except the case of the BGG, it does not exist an analytic expression for the KLD between two MGG distributions. To overcome this problem, two solutions can be envisaged. The first one consists in adopting another similarity. For example, the Rao geodesic distance has been used to compute the similarity between such two distribution

models. However, this distance has been derived in [Verdoolaege, Scheunders, 2012] in the particular case of a fixed MGG shape parameter. Note that the case of variable shape parameter has been also studied, and *only* an approximation of this distance is obtained. To avoid the problem of choice of the similarity measure for such distributions, the second solution will consist in resorting to another appropriate multivariate statistical model that reflects the spatial and cross-views redundancies of the stereo data while allowing to easily compute a meaningful similarity measure. For this reason, we have preferred to select the Spherically Invariant Random Vectors (SIRV) model.

3.6.2 SIRV-based model

The SIRV model has been successfully used to model non-Gaussian stochastic processes for retrieval of textured mono-view images [Lasmar, Berthoumieu, 2010; Bombrun *et al.*, 2011a] and SAR image segmentation [Bombrun *et al.*, 2011b]. The SIRV model consists in expressing $\widetilde{\mathbf{W}}_j$ as the product of a zero-mean Gaussian vector \mathbf{G}_j , with covariance matrix $\mathbf{C}_j = \mathbb{E}[\mathbf{G}_j \mathbf{G}_j^\top]$, by the square root of a positive random variable S_j :

$$\widetilde{\mathbf{W}}_j = \sqrt{S_j} \mathbf{G}_j \quad (3.43)$$

where S_j and \mathbf{G}_j are assumed to be independent. As the couple (S_j, \mathbf{C}_j) could be defined up to a multiplicative constant, it is used to assume that $\text{tr}(\mathbf{C}_j) = B$. The probability density function (pdf) of $\widetilde{\mathbf{W}}_j$ denoted by $h_{\widetilde{\mathbf{W}}_j}$, could be deduced from that of S_j , h_{S_j} , thanks to Bayes' theorem [Rzangaswamy *et al.*, 1993]. Indeed, for every \mathbf{w} in \mathbb{R}^B , we have:

$$h_{\widetilde{\mathbf{W}}_j}(\mathbf{w}) \propto \int_0^{+\infty} \frac{1}{s^{\frac{B}{2}}} e^{\left(-\frac{\mathbf{w}^\top \mathbf{C}_j^{-1} \mathbf{w}}{2s}\right)} h_{S_j}(s) ds. \quad (3.44)$$

Although there is no systematically a closed form for $h_{\widetilde{\mathbf{W}}_j}$ (except for some special cases of h_{S_j}), the SIRV model is completely characterized by the covariance matrix \mathbf{C}_j and, the parameters of the multiplier h_{S_j} . At this level, it is important to emphasize that the estimation of the normalized \mathbf{C}_j could be performed independently from the multiplier. Indeed, in [Pascal *et al.*, 2008], and based on a sample of L_j realizations $\{\widetilde{\mathbf{w}}_j(1), \dots, \widetilde{\mathbf{w}}_j(L_j)\}$ of $\widetilde{\mathbf{W}}_j$, it has been found that the maximum likelihood estimate of \mathbf{C}_j could be approximated by the following matrix $\widehat{\mathbf{C}}_j$:

$$\widehat{\mathbf{C}}_j = \frac{B}{L_j} \sum_{\ell=1}^{L_j} \frac{\widetilde{\mathbf{w}}_j(\ell) \widetilde{\mathbf{w}}_j(\ell)^\top}{\widetilde{\mathbf{w}}_j(\ell)^\top \widehat{\mathbf{C}}_j^{-1} \widetilde{\mathbf{w}}_j(\ell)}. \quad (3.45)$$

Computing $\widehat{\mathbf{C}}_j$ reduces to iteratively search the fixed point of the function appearing at the right hand of Eq. (3.45). Then, at each position $\ell = 1, \dots, L_j$, the maximum likelihood estimate of the multiplier could be easily deduced as follows:

$$\hat{s}_j(\ell) = \frac{1}{B} \widetilde{\mathbf{w}}_j(\ell)^\top \widehat{\mathbf{C}}_j^{-1} \widetilde{\mathbf{w}}_j(\ell). \quad (3.46)$$

We have summarized the fixed-point estimation method in Algorithm 1.

Algorithm 1 Fixed point estimator

Require: $\widehat{\mathbf{C}}_j^0 = \frac{1}{L_j} \sum_{\ell=1}^{L_j} \widetilde{\mathbf{w}}_j(\ell) \widetilde{\mathbf{w}}_j(\ell)^\top$

Ensure: the covariance matrix $\widehat{\mathbf{C}}_j$ of the SIRV model and the multiplier values $\hat{s}_j(\ell)$, $\ell = 1, \dots, L_j$.

MaxNiter $\leftarrow 10$

$\epsilon \leftarrow 10^{-6}$

$k \leftarrow 0$

repeat

$$\widehat{\mathbf{C}}_j^{k+1} = \frac{B}{L_j} \sum_{\ell=1}^{L_j} \frac{\widetilde{\mathbf{w}}_j(\ell) \widetilde{\mathbf{w}}_j(\ell)^\top}{\widetilde{\mathbf{w}}_j(\ell)^\top (\widehat{\mathbf{C}}_j^k)^{-1} \widetilde{\mathbf{w}}_j(\ell)}$$

$k \leftarrow k + 1$

until $\frac{1}{2} \left(\text{tr}((\widehat{\mathbf{C}}_j^{k+1})^{-1} \widehat{\mathbf{C}}_j^k) + \log \frac{|\widehat{\mathbf{C}}_j^{k+1}|}{|\widehat{\mathbf{C}}_j^k|} - B \right) \leq \epsilon$ ou $k = \text{MaxNiter}$

$\hat{s}_j(\ell) = \frac{1}{B} \widetilde{\mathbf{w}}_j(\ell)^\top \widehat{\mathbf{C}}_j^{-1} \widetilde{\mathbf{w}}_j(\ell)$, $\ell = 1, \dots, L_j$

return $\widehat{\mathbf{C}}_j^{k+1}$, $\hat{s}_j(\ell)$, $\ell = 1, \dots, L_j$

Finally, the parameters of h_{S_j} could be derived from the sample $(\hat{s}_j(1), \dots, \hat{s}_j(L_j))$ by considering for instance the maximum likelihood criterion. To this respect, the selection of a suitable pdf for the multiplier is a key issue as it reflects the way of accounting for the non-Gaussianity of $\widetilde{\mathbf{W}}_j$. Several distributions h_{S_j} have been proposed for SAR images such as Gamma distribution [Quegan, Rhodes, 1993], the inverse Gamma [Freitas et al., 2005], and Fisher distributions [Bombrun et al., 2011b]. Recently, the Weibull distribution has been found to be an appropriate description of the texture in natural mono-view images [Lasmar, Berthoumieu, 2010; Bombrun et al., 2011a]. In this case, h_{S_j} is given by:

$$\forall s \in \mathbb{R}^+, \quad h_{S_j}(s) = \frac{a_j}{b_j} \left(\frac{s}{b_j} \right)^{a_j-1} e^{-\left(\frac{s}{b_j} \right)^{a_j}} \quad (3.47)$$

where $a_j > 0$ and $b_j > 0$ represent the shape and scale parameters, respectively. These parameters (a_j, b_j) can be easily estimated by (\hat{a}_j, \hat{b}_j) according to the maximum likelihood

criterion by numerically solving the following equations [Balakrishnana, Kateri, 2008]:

$$\frac{1}{\hat{a}_j} = \frac{\sum_{\ell=1}^{L_j} \hat{s}_j(\ell)^{\hat{a}_j} \log(\hat{s}_j(\ell))}{\sum_{\ell=1}^{L_j} \hat{s}_j(\ell)^{\hat{a}_j}} - \frac{1}{L_j} \sum_{\ell=1}^{L_j} \log(\hat{s}_j(\ell)), \quad (3.48)$$

$$\hat{b}_j = \left(\frac{1}{L_j} \sum_{\ell=1}^{L_j} \hat{s}_j(\ell)^{\hat{a}_j} \right)^{1/\hat{a}_j}. \quad (3.49)$$

Once the modeling step is performed, the resulting SIRV distribution parameters $(\mathbf{C}_j, a_j, b_j)_{1 \leq j \leq 3J}$ are used as a salient feature during the indexing process. It is important to note here that one can explicitly compute the KLD between two SIRV distributions parameterized by $(\mathbf{C}_j^{\text{db}}, a_j^{\text{db}}, b_j^{\text{db}})$ and $(\mathbf{C}_j^{\text{q}}, a_j^{\text{q}}, b_j^{\text{q}})$. The latter corresponds to the sum of the KLDs between the multiplier models and that obtained between the multivariate Gaussian processes, as the multiplier parameters S_j and the Gaussian processes \mathbf{G}_j are independent [Lasmar, Berthoumieu, 2010]:

$$\begin{aligned} D_{\text{SIRV}} \left(\mathbf{C}_j^{\text{db}}, a_j^{\text{db}}, b_j^{\text{db}} \parallel \mathbf{C}_j^{\text{q}}, a_j^{\text{q}}, b_j^{\text{q}} \right) &= \Gamma \left(\frac{a_j^{\text{q}}}{a_j^{\text{db}}} + 1 \right) \left(\frac{b_j^{\text{db}}}{b_j^{\text{q}}} \right)^{a_j^{\text{q}}} \\ &+ \log \left((b_j^{\text{db}})^{-a_j^{\text{db}}} a_j^{\text{db}} \right) - \log \left((b_j^{\text{q}})^{-a_j^{\text{q}}} a_j^{\text{q}} \right) + \log(b_j^{\text{db}}) \left(a_j^{\text{db}} - a_j^{\text{q}} \right) \\ &+ \frac{\gamma a_j^{\text{q}}}{a_j^{\text{db}}} - \gamma - 1 + \frac{1}{2} \left[\text{tr}((\mathbf{C}_j^{\text{q}})^{-1} \mathbf{C}_j^{\text{db}}) + \log \left(\frac{|\mathbf{C}_j^{\text{q}}|}{|\mathbf{C}_j^{\text{db}}|} \right) - B \right] \end{aligned} \quad (3.50)$$

where γ denotes the Euler-Mascheroni constant.

3.6.3 Retrieval approach

Once the SIRV model is defined, we describe now the proposed retrieval strategy for SI databases. Intuitively, a multivariate vector $\tilde{\mathbf{w}}_j^{(r,1)}$ can be defined by taking, for each subband j , the set of the wavelet coefficients of the left $w_j^{(l)}$ and right images $w_j^{(r)}$, located at the same spatial position (m, n) , as well as those of the neighboring pixels:

$$\tilde{\mathbf{w}}_j^{(r,1)} = \left(w_j^{(r)}, w_{j,p,q}^{(r)}, w_j^{(l)}, w_{j,p,q}^{(l)} \right)^\top \quad (3.51)$$

where for each $v \in \{l, r\}$, $w_{j,p,q}^{(v)}(m, n) = w_j^{(v)}(m + p, n + q)$ and $(p, q) \in (\mathbb{Z}^*)^2$ refers to the spatial support of the neighboring pixels. Note that the size of this multivariate vector depends clearly on the number of the neighboring pixels used in each subband to take into account the spatial redundancies. Indeed, if we denote this number by N , the number of the components of the multivariate vector is given by $B = 2(N + 1)$. Fig. 3.6 shows how

the multivariate vector $\tilde{\mathbf{w}}_j^{(r,l)}$ is defined when $N = 3$.

Since the highly similar pixels of the left and right images are located at different spatial positions which can be found thanks to the disparity map, it would be more interesting to define the multivariate vector from the wavelet coefficients of the right image $w_j^{(r)}$ and the compensated left one $w_j^{(c)}$. Thus, for each subband j , a new multivariate vector can be defined as follows:

$$\tilde{\mathbf{w}}_j^{(r,c)} = \left(w_j^{(r)}, w_{j,p,q}^{(r)}, w_j^{(c)}, w_{j,p,q}^{(c)} \right)^\top. \quad (3.52)$$

Therefore, during the indexing step, the feature vector of the right and compensated left images of the query stereo images $(\mathbf{C}_j^{(r,c,q)}, a_j^{(r,c,q)}, b_j^{(r,c,q)})$ will be compared to those of the database stereo pair $(\mathbf{C}_j^{(r,c,db)}, a_j^{(r,c,db)}, b_j^{(r,c,db)})$ by computing the global KLD:

$$\tilde{D}_{\text{SIRV}}^{(r,c)} = \sum_{j=1}^{3J} D_{\text{SIRV}}(\mathbf{C}_j^{(r,c,db)}, a_j^{(r,c,db)}, b_j^{(r,c,db)} \parallel \mathbf{C}_j^{(r,c,q)}, a_j^{(r,c,q)}, b_j^{(r,c,q)}). \quad (3.53)$$

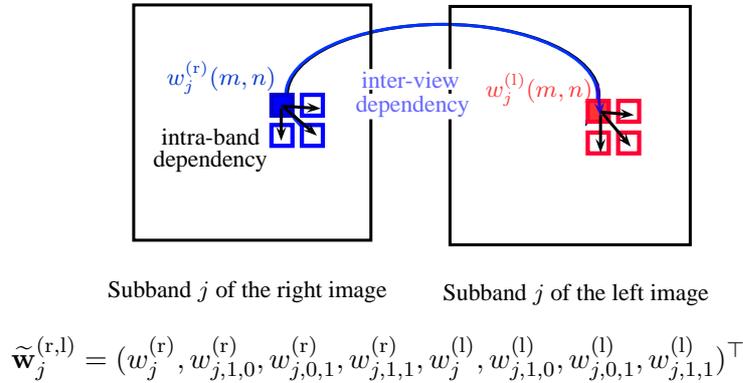


Figure 3.6: Definition of the multivariate vector $\tilde{\mathbf{w}}_j^{(r,l)}$ when considering 3 neighbors in each view.

3.7 Experimental results

3.7.1 Experimental setup

Since there are no SI databases publicly available to evaluate the performance of SI retrieval methods, we have built a database which can be downloaded from ¹. This database

¹<http://www-l2ti.univ-paris13.fr/~kaaniche/Download.html>

is composed of real SI pairs of size 248×248 taken from various sources. The images correspond to a variety of contents including natural scenes²,³ and⁴, man-made objects available at the Middlebury stereo vision website⁵, and SPOT5 scenes. According to their texture, these stereo images have been divided into 17 classes, with 40 pairs per class, such as wooded area, tree, bushes, mountains, urban area. Note that SI of the same class are taken from the same scene. An example of some right images of different classes in the database is shown in Fig. 3.7.

The retained DDE method, described in Section 3.3, has been used to generate the disparity maps. The retrieval performance is evaluated in terms of precision $PR = \frac{R^r}{R^n}$ and recall $RC = \frac{R^r}{R^t}$, where R^r is the number of output pairs considered as relevant, R^t is the total number of relevant pairs in the database and R^n denotes the number of returned pairs. A retrieved pair is considered as relevant if it belongs to the same category of the query pair. Note that the query images are taken from the database. Retrieval results are also evaluated in terms of ANMRR, which has been already defined in Section 1.2.3.

In order to show the benefits of using the disparity map in a SI retrieval system, three rounds of experiments are performed. The first one aims at illustrating the behavior of the univariate modeling-based retrieval approaches described in Section 3.4. The objective of the second one is to validate the interest of the bivariate modeling presented in Section 3.5. The third one aims to show the interest of the multivariate model described in Section 3.6. In what follows, we describe and discuss these experimental tests.

3.7.2 Univariate model-based retrieval approaches

In this part, we are interested in evaluating the methods related to the univariate model. The first one corresponds to the straightforward approach, where the GG distribution parameters of only the right and left images are compared. This method will be designated by GG-RL. The second one takes into account the disparity information in an implicit manner by computing features of the right view and the disparity compensated left view. This method will be denoted by GG-RDCL. The third one, designated by GG-RL-GG-D, is the second version of the proposed univariate model-based method where a new feature

²<http://www.mi.auckland.ac.nz/EISATS/>

³<http://vasc.ri.cmu.edu/idb/html/stereo/index.html>

⁴<http://vasc.ri.cmu.edu/idb/html/jisct/>

⁵<http://vision.middlebury.edu/stereo/>

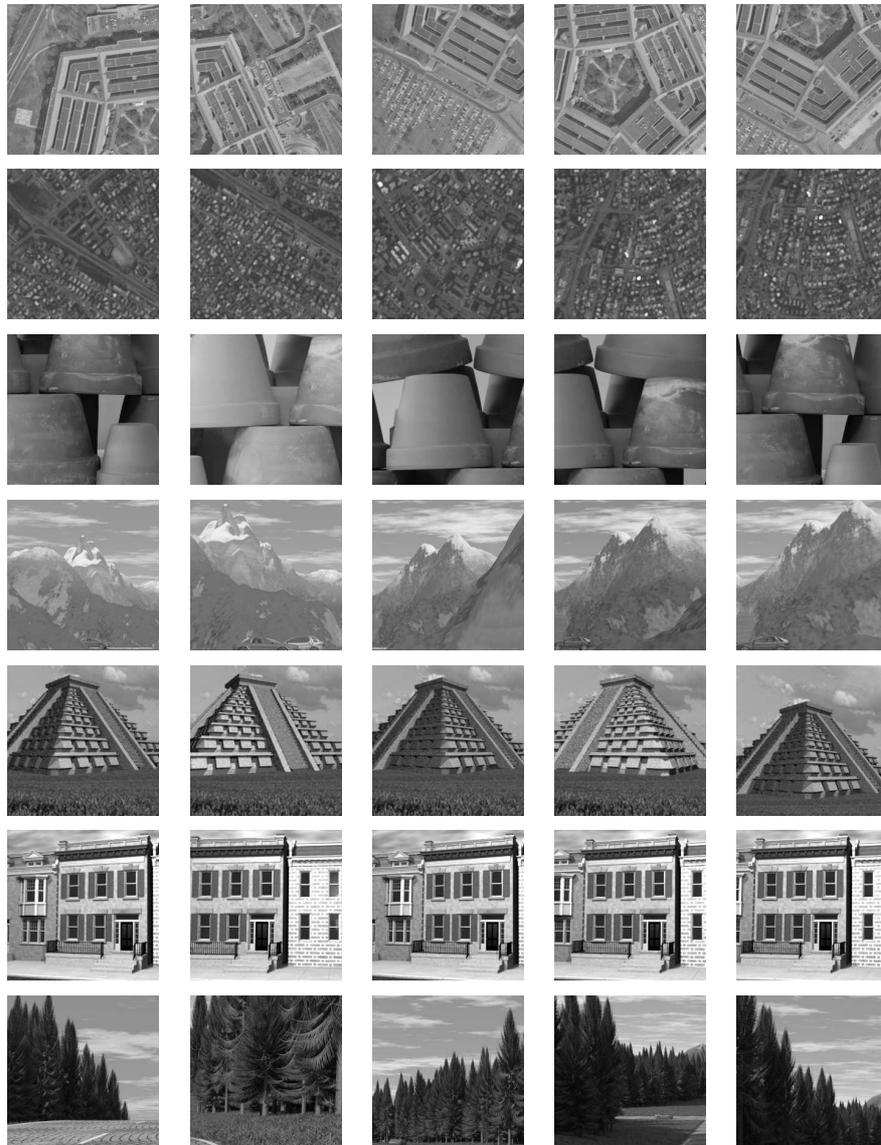


Figure 3.7: Some samples of right images for different classes in the database. From top to bottom: pentagon, urban area, flowerpots, mountains, pyramid, buildings and tree.

vector is defined by incorporating *simultaneously* the visual contents of the left and right images as well as the disparity information. We have also tested for comparison the recent state-of-the-art approach [Feng *et al.*, 2011]. Recall that its basic idea consists in using the disparity to refine the results provided by a conventional CBIR system. More precisely, the MPEG-7 edge histogram descriptor is employed for the left images and the diffusion distance is used to measure the similarity between the histograms of the disparity maps during the refinement step. Hereafter, this method will be designated by state-of-the-art [Feng *et al.*, 2011]. Moreover, since the developed approaches operate in the wavelet transform

domain, we have also proposed to apply the state-of-the-art method in the same domain. It is important to note here that designing a CBIR system operating in the wavelet domain may constitute an interesting feature in practice in the sense that the decoding procedure at the retrieval step becomes unnecessary when the images are saved in a compressed format before the design of the CBIR. To this end, the first step consists of comparing the feature vector of the query left image $(\alpha_j^{(l,q)}, \beta_j^{(l,q)})_{1 \leq j \leq 3J}$ to the database left images $(\alpha_j^{(l,db)}, \beta_j^{(l,db)})_{1 \leq j \leq 3J}$ using the KLD as a similarity measure. Then, the disparity features $(\alpha_j^{(u,db)}, \beta_j^{(u,db)})_{1 \leq j \leq 3J}$ are used in the re-ranking step which is applied to the first 10% of the returned images and the retrieval results are re-ordered. In what follows, this modified version of the state-of-the-art method, applied in the WT domain, will be designated by Mod-state-of-the-art-WT. Table 3.1 summarizes the different considered methods.

Fig. 3.8 provides the precision-recall plots of the various approaches. It indicates that using implicitly the disparity map by comparing the right image and the disparity compensated left one of the stereo pair outperforms the intuitive approach. Moreover, thanks to the second version of the proposed method where the disparity information is explicitly added to the feature vector, we achieve further improvements. Our approach becomes more performant than the other methods, and achieves a gain of about 1-12% in precision compared to the state-of-the-art method [Feng *et al.*, 2011]. Table 3.2 shows retrieval performance according to the ANMRR criterion. These results prove the outperformance of the GG-RL-GG-D approach (whose corresponding ANMRR value is 0.106) compared to the state-of-the-art method (whose corresponding ANMRR value is 0.245).

We should note here that different tests have been carried out to study the impact of the weights assigned to the disparity map and to both images on the retrieval performance of the GG-RL-GG-D approach. Since the left and right images correspond to the same scene and present very similar contents, we assume that the weights associated to these views are identical (i.e. $a = b$). Fig. 3.9 shows the precision-recall curves for different weight parameters.

Thus, it can be observed that selecting a very low or a high value of c leads to worse results. More generally, we conclude that weighting alike the features of the left, right and disparity images, by taking a value of c around $\frac{1}{3}$, allows to achieve good retrieval performance.

Therefore, these results corroborate that disparity gives additional cues for SI retrieval when it is combined with the visual contents of the two views.

Table 3.1: The considered approaches for SI retrieval.

Method	Description
State of the art	employs the disparity information to refine the results of the conventional CBIR system applied on one image.
Univariate model-based retrieval approaches	
GG-RL	computes and compares features of the left and right images (Subsection 3.2.4, Eq. (3.14)).
GG-RL-GG-D	the disparity has been exploited <i>explicitly</i> by incorporating simultaneously the features of the left and right images as well as those of the disparity map (Subsection 3.4.2, Eq. (3.27)).
GG-RDCL	the disparity has been exploited <i>implicitly</i> by computing features of the right image and the disparity compensated left one (Subsection 3.4.1, Eq. (3.26)).
Bivariate model-based retrieval approaches	
BGG-RL	computes features from the bivariate vector defined by using the wavelet coefficients of the right and left images (Subsection 3.5.3, Eq. (3.39)).
BGG-RDCL-1	computes features from the bivariate vector built from the wavelet coefficients of the right image and the disparity compensated left one (Subsection 3.5.3, Eq. (3.40)).
BGG-RDCL-2	is the improved version of BGG-RDCL-1 by taking into account the occlusion effect (Subsection 3.5.3, Eq. (3.41)).
BGG-RL-GG-D	concatenates features obtained from BGG-RL and those of the disparity map
BGG-RDCL-GG-D-1	incorporates features extracted from BGG-RDCL-1 and the disparity map
BGG-RDCL-GG-D-2	adds the disparity features to those extracted from BGG-RDCL-2
Multivariate model-based retrieval approach	
SIRV-RDCL-3N-1	exploits simultaneously the <i>spatial</i> and <i>cross-view</i> dependencies of the stereo images through SIRV modeling (Subsection 3.6.3, Eq. (3.53)).
SIRV-RDCL-3N-GG-D-1	incorporates features extracted from SIRV-RDCL-3N-1 and the GG parameters of the wavelet coefficients of the disparity map.

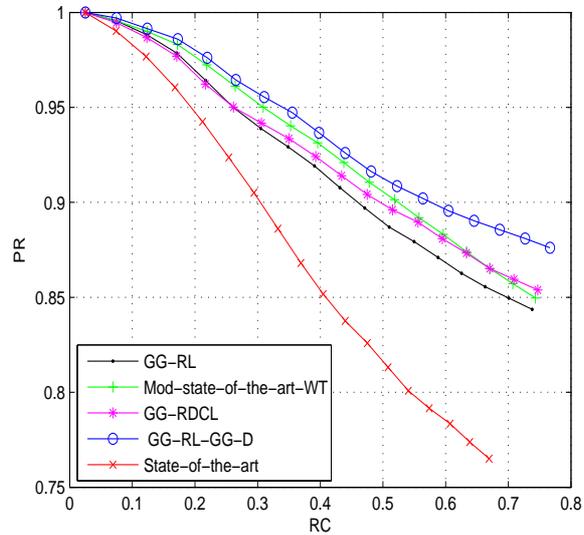


Figure 3.8: Retrieval performance in terms of precision and recall of the univariate approaches.

Table 3.2: ANMRR results of the univariate model-based retrieval approaches.

Method	State-of-the-art	GG-RL	GG-RDCL	GG-RL-GG-D
ANMRR	0.245	0.136	0.130	0.106

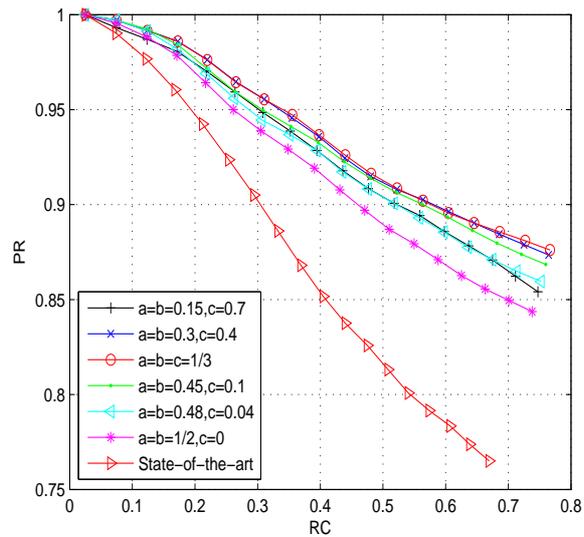


Figure 3.9: Impact of the weight values on the retrieval performances of the GG-RL-GG-D approach.

3.7.3 Bivariate model-based retrieval approaches

The second series of experiments is dedicated to the study of the effect of using a bivariate statistical model to capture the dependencies between the wavelet coefficients of a stereo pair and, to the illustration of the benefits of incorporating simultaneously the disparity features and the visual ones. To this end, we have also conducted the three following experiments related to the three retrieval strategies discussed in Subsection 3.5.3.

The first one, where the bivariate vector is constructed from the wavelet coefficients of the right and left images, is designated by BGG-RL. The second one, where the bivariate vector is defined by using the wavelet coefficients of the right image and the disparity compensated left one, is denoted by BGG-RDCL-1. The third approach, corresponding to the improved version of the previous one by taking into account the occlusion effect, is denoted by BGG-RDCL-2. By further adding the GG distribution parameters of the disparity map during the indexing step, these three methods will be respectively designated by BGG-RL-GG-D, BGG-RDCL-GG-D-1 and BGG-RDCL-GG-D-2.

Fig. 3.10 depicts the precision versus recall curves for these approaches.

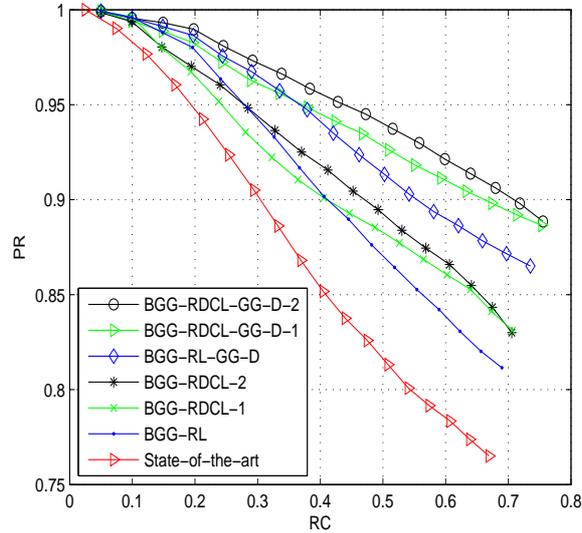


Figure 3.10: Retrieval performance in terms of precision and recall of the bivariate approaches.

It can also be noted that adding disparity improves the different bivariate modeling-based retrieval strategies. Moreover, these results show that taking into account the occlusion effect allows us to achieve the best retrieval performance.

Finally, we focus on comparing the performance of the proposed retrieval approaches based on bivariate and univariate models. It can be seen from Fig. 3.11 that the joint modeling of wavelet subbands BGG-RDCL-GG-D-1 achieves better retrieval performance compared to the univariate model-based approach. Further improvements are achieved when the bivariate model-based approach deals with the occlusion effect. Thus, compared to the state-of-the-art method [Feng *et al.*, 2011], the resulting gain reaches 15% in terms of precision-recall. Table 3.3 confirms that the best improvement are achieved by the bivariate model-based approach especially when dealing with the occlusion effect. The ANMRR value decreases by 1,34% when using the bivariate model (BGG-RDCL-GG-D-2 approach) compared to the univariate model (GG-RL-GG-D approach).

All these results confirm the effectiveness of incorporating the disparity information for stereo image retrieval purpose.

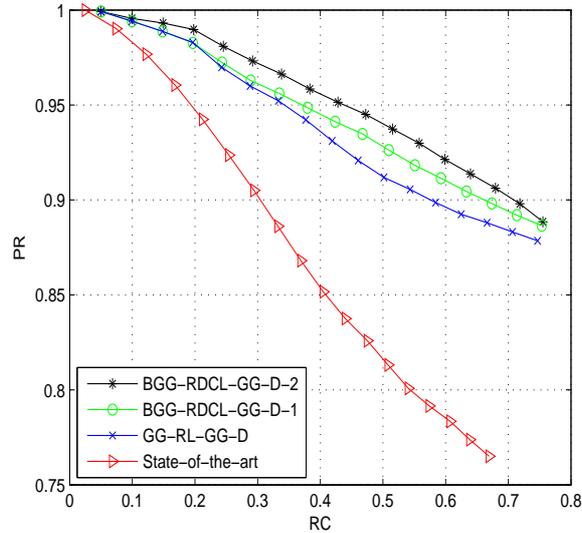


Figure 3.11: Retrieval performance in terms of precision and recall of the univariate and the bivariate approaches.

Table 3.3: ANMRR results of the univariate and bivariate approaches.

Method	State-of-the-art	GG-RL-GG-D	BGG-RDCL-GG-D-1	BGG-RDCL-GG-D-2
ANMRR	0.245	0.106	0.0951	0.0926

3.7.4 Multivariate model-based retrieval approach

In order to show the benefits of the multivariate SIRV modeling-based retrieval method, we will compare it to the previous retrieval approaches. To this end, the multivariate vector will be defined from the wavelet coefficients of the right and compensated left images as given by Eq. (3.52). To take into account the spatial dependency in each subband, 3 neighbors have been used by setting $(p, q) \in \{(1, 0), (0, 1), (1, 1)\}$ for the sake of simplicity. This method will be designated by SIRV-RDCL-3N-1. By adding the GG distribution parameters of the disparity map, this method will be noted by SIRV-RDCL-3N-GG-D-1. It can be seen from Fig. 3.12 that the SIRV modeling of wavelet subbands achieves better retrieval performance compared to the BGG model-based approach [Chaker *et al.*, 2014]. These results prove that exploiting both the spatial and cross-view dependencies of the stereo images through SIRV modeling improves retrieval performances of BGG model-based approach. Similar retrieval performance in terms of ANMRR values are presented in Table 3.4. It can also be noticed that adding disparity information explicitly improves the bivariate as well as multivariate modeling-based retrieval strategies. As shown in Table 3.4, the ANMRR values decrease by 4.43% , 1.16% when adding disparity information to the BGG model- and SIRV model-based approaches, respectively.

Now, we will study the impact of the size of the neighborhood on the retrieval performance. To this end, the SIRV modeling-based approach has also been implemented by using 8, 24, and 48 neighbors. Knowing that $(p, q) \neq (0, 0)$, the 8, 24 and 48 neighbors are obtained by varying p and q from -1 to 1 , -2 to 2 , and -3 to 3 , respectively. The related versions will be designated by SIRV-RDCL-8N-1, SIRV-RDCL-24N-1, and SIRV-RDCL-48N-1, respectively. Moreover, in order to show the benefits of exploiting the cross-view redundancies as well as the spatial dependencies, we consider also the case where the intra-band correlation is not taken into account. In others words, the multivariate vector is only composed of the wavelet coefficients of the right and compensated left image without adding the neighboring pixels. This method will be denoted by SIRV-RDCL-0N-1.

Fig. 3.13 shows the plots of the precision versus recall resulting from the use of different number of neighboring pixels. First, it is clear that combining the intra and inter-image redundancies outperforms the scheme where only the inter-view correlations are exploited. Moreover, it can be noticed that increasing the neighborhood size leads to a further improvement in the retrieval performance. However, it is worth pointing out that large

neighborhood size implies high computational complexity. Indeed, the complexity of the feature extraction step consists of the estimation of the covariance matrix $\hat{\mathbf{C}}_j$ and the computation of its inverse $\hat{\mathbf{C}}_j^{-1}$ which depends on its dimension $B \times B$, where $B = 2(N + 1)$ and N is the number of the neighboring pixels.

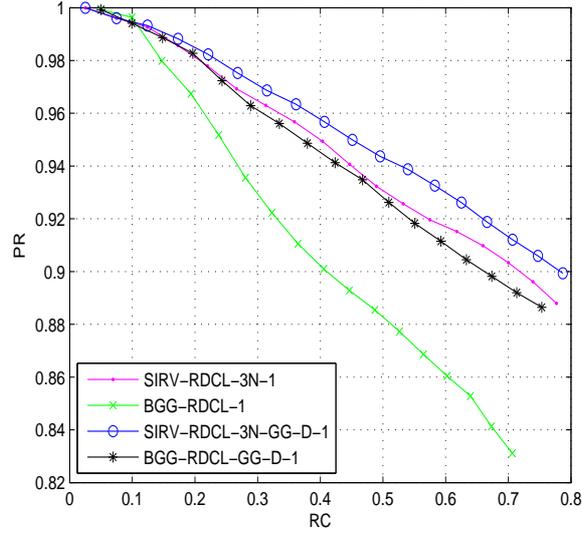


Figure 3.12: Comparison between the multivariate and bivariate approaches.

Table 3.4: ANMRR results of the multivariate and bivariate approaches.

Method	BGG-RDCL-1	BGG-RDCL-GG-D-1	SIRV-RDCL-3N-1	SIRV-RDCL-3N-GG-D-1
ANMRR	0.1394	0.0951	0.0992	0.0876

3.8 Conclusion

In this chapter, we have addressed the problem of indexing and retrieval of stereo images in the wavelet-transform domain. Our first contribution consists in employing dense disparity maps either implicitly or explicitly during the feature extraction step. The parameters of the generalized Gaussian distribution that model the detail subbands of each view are combined with those of the disparity map to build a salient feature of the stereo pair content. Our second contribution aims at resorting to an appropriate bivariate model that accounts for the cross-view dependencies. Finally, we have exploited both the dependencies

between the wavelet coefficients of the two views as well as those in the same subband through a suitable multivariate model. Experimental results indicate the good performance of the bivariate and multivariate approaches. As it has been described through this chapter, the features have been defined and extracted from the original wavelet coefficients. Such methods are interesting when the images are losslessly encoded. However, in the context of lossy compression, it becomes more difficult to define salient features from the quantized wavelet coefficients. This issue, studied with still images, will be the objective of the next chapter.

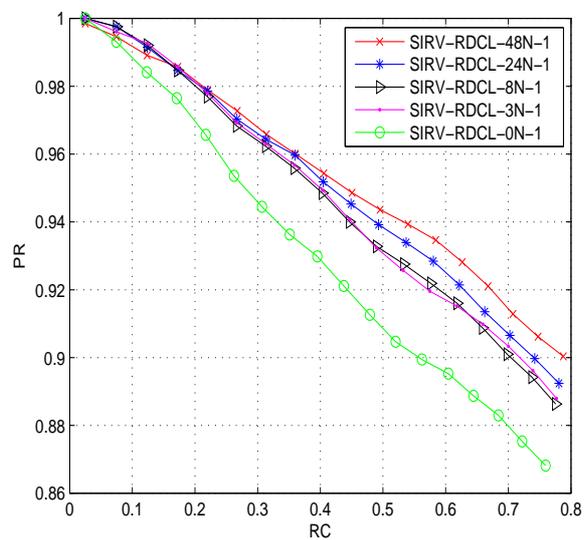


Figure 3.13: Impact of the neighborhood size on the retrieval performance.

Chapter 4

Retrieval approaches for wavelet-based quantized images

4.1 Introduction

The proliferation of images has led to very huge amount of data to be stored, managed and/or transmitted. In order to reduce the memory requirements, these images are stored in a compressed format. In this case, a CBIR system should be built for these coded versions. In this context, quantization may affect the image content and, hence the image features. Therefore, some cautions should be taken in the case of quantized images. Unlike most of the developed CBIR systems which deals with uncompressed images, this chapter addresses the problem of designing new retrieval schemes for lossy compressed images and, provides some novel solutions to this problem.

The remainder of this chapter is organized as follows. In Section 4.2, we investigate the compression effect on image retrieval performance. Then, we describe our adopted methodology in Section 4.3. In Section 4.4, we present new strategies to improve the retrieval performances of JPEG2000 compressed images. In Section 4.5, we describe the investigated quantization techniques as well as their corresponding proposed retrieval approaches. Finally, the performance of the different methods is illustrated in Section 4.6 and some conclusions are drawn in Section 4.7.

4.2 Impact of compression on image retrieval performance

4.2.1 Motivation

Many CBIR systems have been proposed in the literature to retrieve mono-view images using appropriate features. These features are often extracted in the spatial domain [Yong, Thomas, 1999; Datta *et al.*, 2006]. However, with the increased size of the generated data, it becomes mandatory to compress them by using some image compression standards such as JPEG [Wallace, 1991] and JPEG2000 [Rabbani, Joshi, 2002] based on DCT and WT, respectively. While classical methods operating in the spatial domain can be also applied to retrieve the resulting compressed images, they will require the decoding of all the database images, by applying the inverse transform, which will result in a significant computational overhead for feature extraction, especially for large image archives. Therefore, in order to design fast retrieval methods, other works have been developed by performing the feature extraction step in the transform domain [Mezaris *et al.*, 2004; Wang *et al.*, 2015]. Moreover, an additional benefit can also be drawn by exploiting the intrinsic sparsity of the coefficients resulting from the underlying transform. The most known retrieval approaches operate in the DCT and WT domains where features are directly extracted from the DCT coefficients [Ngo *et al.*, 2001; Feng, Jiang, 2003; Chang *et al.*, 2004; Climer, Bhatia, 2002] and the wavelet coefficients [Voulgaris, Jiang, 2001; Mandal *et al.*, 1996; Do, Vetterli, 2002; Sakji-Nsibi, Benazza-Benyahia, 2009]. The most popular and fast wavelet-based technique aims at retaining the first moments of the subband j as a salient feature [Mandal *et al.*, 1996], especially the energy E_j of the subband [Smith, Chang, 1994; Wouwer *et al.*, 1999]. Note that several statistical features such as inertia, entropy and local homogeneity, defined on the co-occurrence matrix, can also be used [Haralick *et al.*, 1973; Wouwer *et al.*, 1999]. Moreover, other works have been developed by resorting to a parametric modeling of the distribution of each wavelet subband [Mandal *et al.*, 1996; Wouwer *et al.*, 1999]. In particular, the generalized Gaussian distribution [Do, Vetterli, 2002] and the Gaussian mixture model [Yuan *et al.*, 2003] were found to be well-suited for modeling the wavelet coefficients in every subband. The related feature vector of the image is built by taking the distribution parameters of all the wavelet subbands. It is worth noting that the so far proposed statistical models concern the unquantized coefficients w_j considered as realizations of a continuous random variable. These models are no longer valid for quantized coefficients \bar{w}_j

which are samples of a discrete random variable. Most of the developed methods are well adapted for losslessly encoded images (i.e. *unquantized* transformed coefficients). However, a particular attention should be paid to the effect of the *quantization* operation in the context of lossy data compression.

4.2.2 Compression effect on image retrieval performance

A straightforward solution to CBIR system consists of directly comparing the query and model images as it is generally performed in the case of uncompressed data. However, it is clear that a lossy compressed image differs from its original. Since image features employed for CBIR are computed from the image content, compressed images may be very different from those defined on the uncompressed data.

A qualitative example is given in Fig. 4.1. We show a typical image (Lena) before and after applying JPEG2000 compression. It can be seen that the JPEG2000 compressed version contains visual degradations compared to its original. We also present the histograms of wavelet subbands descriptors which are often used in lossless CBIR systems [Mandal *et al.*, 1996; Wouwer *et al.*, 1999; Do, Vetterli, 2002] of the two images. As illustrated in Fig. 4.1, the difference between the two histograms shows a large deviation that can cause a degradation in the performance of the conventional CBIR techniques [Mandal *et al.*, 1996; Wouwer *et al.*, 1999; Do, Vetterli, 2002]. This was corroborated for all the considered images in the database. This problem has been addressed by earlier studies in [Schaefer, 2008; Edmundson, Schaefer, 2012b; Edmundson *et al.*, 2012], where it was concluded that, while slight compression has little effect on texture and color based CBIR systems, a significant degradation of the retrieval performance is shown at high compression levels. We confirm this by running retrieval experiments under different qualities of JPEG2000 compressed images. To this end, we select the Vistex database [vis,], which contains 40 textures images of size 512×512 . Each image is then divided into 16 non-overlapping images of size 128×128 , resulting in a collection of 640 images. The sixteen images belonging to the same family of a query image are considered as the relevant images for this query. These images (called model images) are compressed at bitrates R_M whose values are in the set $\{1.5, 1, 0.8, 0.5, 0.25, 0.1\}$ bpp. The retrieval performances are evaluated in terms of precision PR and recall RC . Recall that $PR = \frac{N^r}{N}$ and $RC = \frac{N^r}{N^t}$, where N^r is the number of output images considered as relevant, N^t represents the total number of relevant images

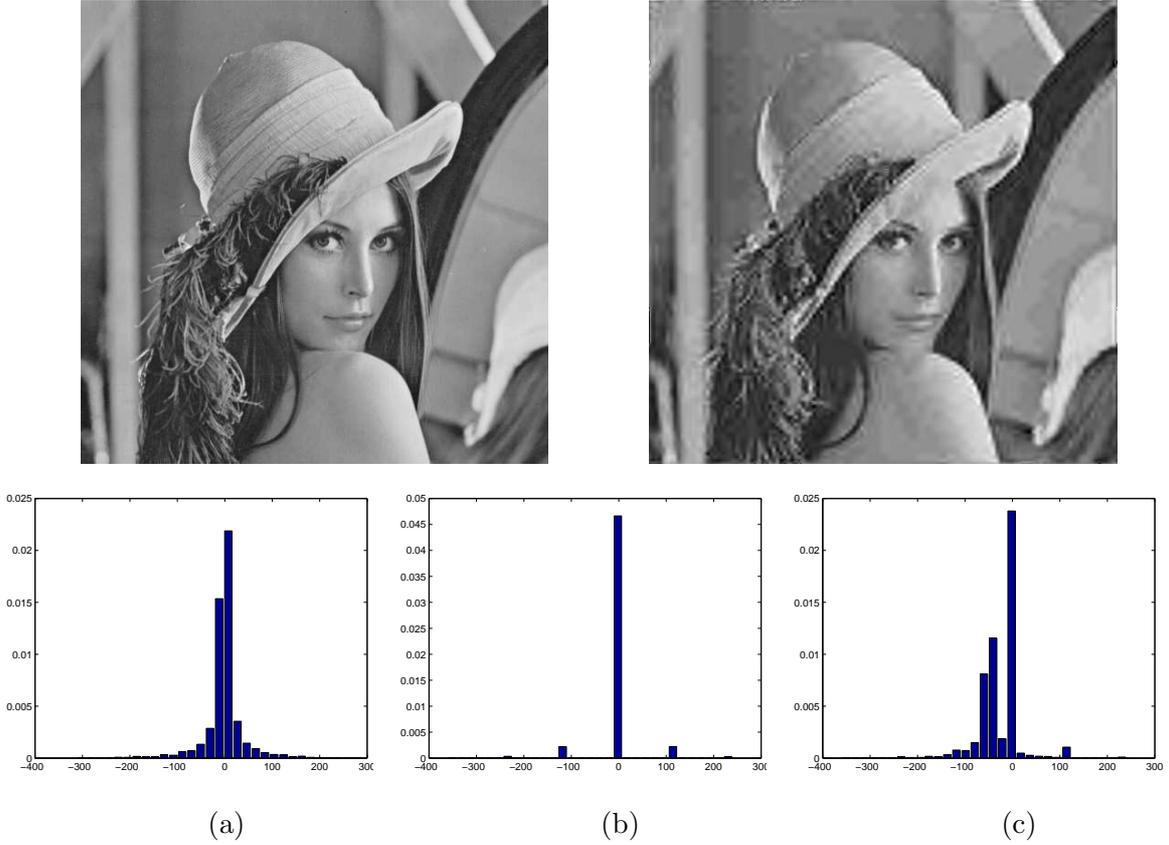


Figure 4.1: Original Lena image, its JPEG2000 compressed version at 0.15 bpp, Histograms of the diagonal wavelet coefficients at the third resolution level for : (a) the original and (b) compressed version of Lena image and (c) the difference between the two histograms.

in the database and N is the number of all returned images. We retain also the whole set of the energies $\left(E_j\right)_{j \in \{1, \dots, 3J+1\}}$ of all quantized subbands as a feature vector to compare the different images.

Fig. 4.2 provides the $PR - RC$ plots for different bitrates R_M . While the blue curve corresponds to the image retrieval algorithm performed on the uncompressed data (i.e on the original wavelet coefficients of the query and model images), the remaining ones are generated after compressing (i.e quantizing) the model images at the different bitrates R_M (the query one is kept as uncompressed). It can be noticed that a significant drop in the performances of the basic CBIR system occurs when model images are compressed at very low bitrates. To better understand the quantization effect on the image retrieval performance, we have considered another context of experiments in which the query and model images are both quantized. These results are illustrated in Fig. 4.3 where the query

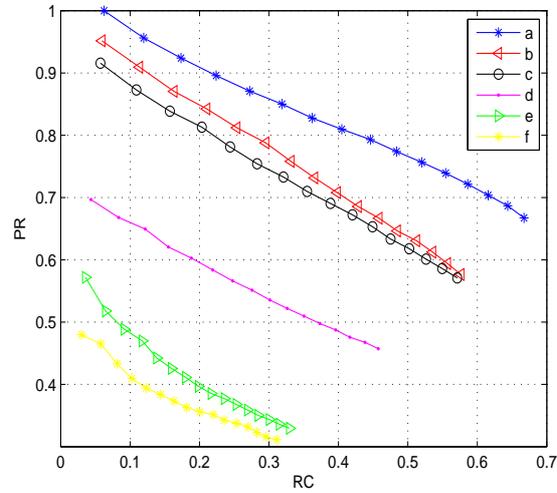


Figure 4.2: Precision versus recall obtained for a query uncompressed images, when model images are : (a) uncompressed, compressed at $R_M =$ (b) 1.5 bpp, (c) 1 bpp, (d) 0.8 bpp, (e) 0.5 bpp, (f) 0.25 bpp.

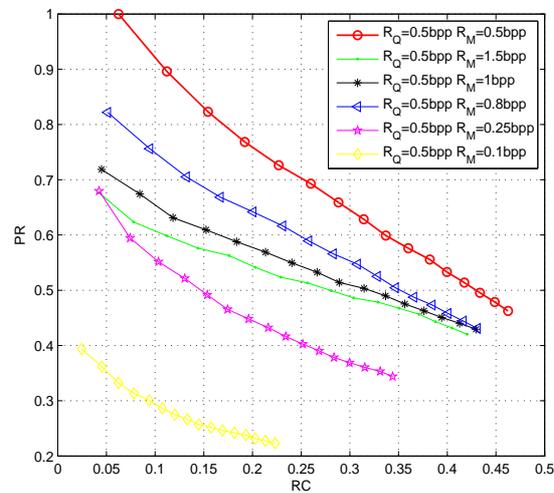


Figure 4.3: Precision versus recall obtained under compression for a query compressed at 0.5 bpp.

image is compressed at $R_Q = 0.5$ bpp and the model images are all compressed at R_M belonging to the set $\{1.5, 1, 0.8, 0.5, 0.25, 0.1\}$ bpp.

It can be observed that the best improvements are achieved when the query and model images are quantized at the same bitrate (i.e having similar qualities).

4.3 Adopted methodology

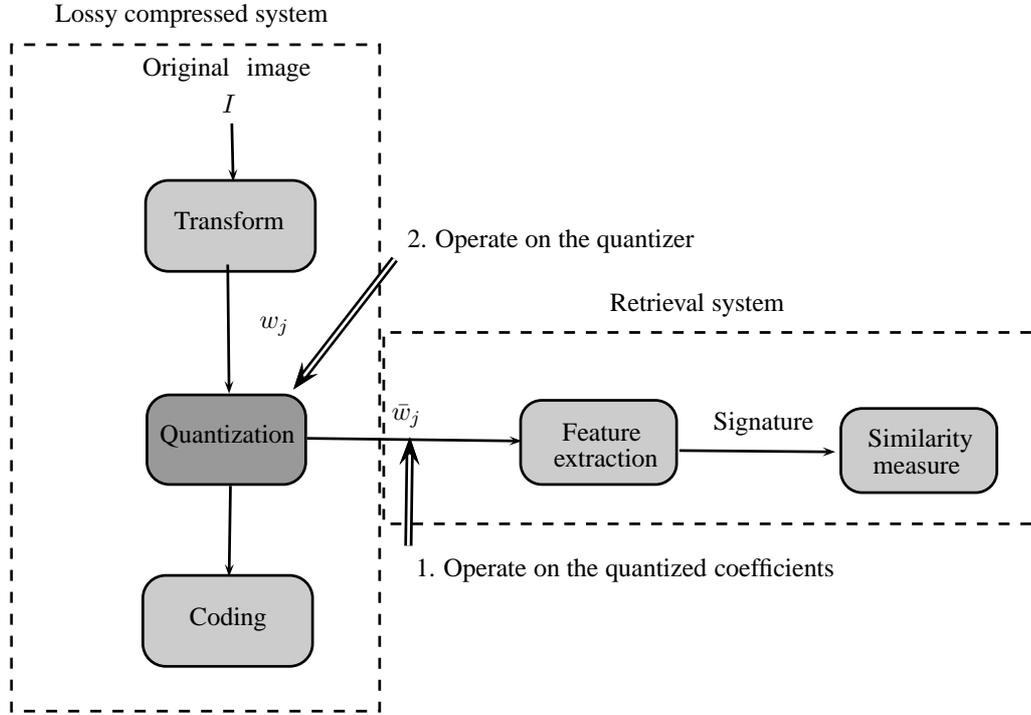


Figure 4.4: Methodology to improve retrieval performance of lossy compressed images.

The degradation of the retrieval performance are mainly due to the quantization itself and, the gap between the bitrates of the model and query image. Consequently, in order to avoid such drop in performance, it appears reasonable to proceed according to the following novel strategies, as depicted in Fig. 4.4.

- The first one operates on the quantized coefficients. To this end, we propose first to apply a processing step that aims at reducing the mismatch between the bitrates of the model and the query images. Then, we propose to recover the statistical parameters of original wavelet coefficients directly from the quantized ones using the maximum likelihood criterion.
- Another alternative consists in acting on the quantization operators by constraining it to preserve into the compressed domain some valuable information. More precisely, in our work, we propose to resort to two new quantizers which have the advantages to preserve the subband moments and distributions, respectively.

In the rest of the chapter, we will describe in more details the suggested solutions.

4.4 Contributions to improve retrieval performance of JPEG2000 compressed images

4.4.1 Re-compression approach

Although uniform quantizer allows a good preservation of the energy of the subbands encoded at high bitrate, it fails to guarantee this property for the subbands encoded at low bitrate [Rabbani, Joshi, 2002]. As it has been shown in Section 4.2.2, and according to many experiments studying the effect of quantization in the DCT domain [Edmundson, Schaefer, 2012b], and WT domain [Chaker *et al.*, 2012], it has been found that the best retrieval performance is obtained when the query and model images are compressed at the same bitrate and have similar qualities. For these reasons, instead of *directly* extracting the features from the quantized query and model images, it would be interesting to add a pre-processing step that constrains these images to have similar qualities. Since the resulting quantization error at low bitrate is more important than that obtained at high bitrate, it appears preferable to apply the pre-processing step on the high quality image. Thus, the main idea is to keep the lower quality image unchanged and transform the high quality one to the lower quality. More precisely, the following transform is firstly applied on the finely quantized subbands $\bar{w}_j(i)$ with the quantization step q_j :

$$\tilde{w}_j(i) = \begin{cases} (\bar{w}_j(i) + \gamma) q_j & \text{if } \bar{w}_j(i) > 0 \\ (\bar{w}_j(i) - \gamma) q_j & \text{if } \bar{w}_j(i) < 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.1)$$

where $0 \leq \gamma < 1$ is a reconstruction parameter chosen by the decoder. Indeed, choosing $\gamma = 0.5$ corresponds to a mid-point reconstruction as it is widely used in many encoding strategies [Rabbani, Joshi, 2002]. It should be noted here that this step (4.1) is known in the context of compression as “inverse” quantization process. Then, the reconstructed subband coefficients $\tilde{w}_j(i)$ are re-quantized at the same bitrate of the lower quality image. Once the model and query images are processed, the energy-based features are extracted from their quantized subbands. Finally, the normalized Euclidean distance [Wouwer *et al.*, 1999] can be used as a similarity measure to compare the resulting feature vectors.

4.4.2 Statistical-based approach

Motivation

It is worth noting that the previous recompression approach for the retrieval of compressed images presents some limitations. First, when the images are quantized at very low bitrates, the related quantization error increases and thus, the recompression of the finely quantized coefficients technique will negatively affect the feature relevance. Moreover, it implies higher computational cost due to the reconstruction and requantization operations which will be performed on all the database images when these images have higher quality than that of the query one. To overcome these drawbacks, it would be interesting to extract salient features *directly* from the quantized wavelet coefficients. To this end, by adopting a statistical model-based approach, we propose to estimate the original distribution parameter from the quantized coefficients, and use it to compare the images during the indexing step. In what follows, the parameter estimation method from the quantized coefficients as well as the retrieval strategy will be described.

Estimation of the original parameter distribution

As mentioned before, the GG law, given by Eq. (3.1), has been extensively used to model the distribution of the wavelet coefficients.

Thus, by assuming that the random variable $w_j(i)$ with $j \in \{1, \dots, 3J\}$ and $i \in \{1, \dots, L_j\}$ is distributed according to (3.1), the probability of the quantized value $\bar{w}_j(i)$ is given by:

$$P(\bar{w}_j(i)) = \int_{\bar{w}_j(i) - \frac{q_j}{2}}^{\bar{w}_j(i) + \frac{q_j}{2}} \frac{\beta_j}{2\alpha_j \Gamma(1/\beta_j)} e^{-(|\xi|/\alpha_j)^{\beta_j}} d\xi \quad (4.2)$$

where α_j and β_j are the GG parameters. It can be shown that Eq. (4.2) leads to [Kaaniche *et al.*, 2014]:

$$P(\bar{w}_j(i)) = \frac{1}{2} \left[Q_{1/\beta_j} \left(\frac{\bar{w}_j(i) + \frac{q_j}{2}}{\alpha_j} \right)^{\beta_j} - Q_{1/\beta_j} \left(\frac{\bar{w}_j(i) - \frac{q_j}{2}}{\alpha_j} \right)^{\beta_j} \right] \quad (4.3)$$

where Q_a , with $a \in \mathbb{R}_+^*$, is the normalized incomplete Gamma function:

$$\forall \xi \in \mathbb{R}, \quad Q_a(\xi) = \frac{1}{\Gamma(a)} \int_0^\xi \theta^{a-1} e^{-\theta} d\theta. \quad (4.4)$$

Therefore, in order to estimate α_j and β_j from the quantized coefficients, we need to maximize the likelihood function which will be expressed as:

$$\begin{aligned}
 \mathcal{L}(\bar{w}_j, \alpha_j, \beta_j) &= \log\left(\prod_{i=1}^{L_j} P(\bar{w}_j(i))\right) \\
 &= \sum_{i=1}^{L_j} \log(P(\bar{w}_j(i))) \\
 &= \sum_{i=1}^{L_j} \log \frac{1}{2} \left[Q_{1/\beta_j}\left(\frac{\bar{w}_j(i)+\frac{q_j}{2}}{\alpha_j}\right)^{\beta_j} - Q_{1/\beta_j}\left(\frac{\bar{w}_j(i)-\frac{q_j}{2}}{\alpha_j}\right)^{\beta_j} \right]
 \end{aligned} \tag{4.5}$$

Thus, the obtained likelihood function depends on the incomplete Gamma function, with parameter $1/\beta_j$ and arguments (α_j, β_j) . Such expression makes the computation of derivatives of the likelihood function with respect to α_j and β_j difficult and will result in very complicated equations.

To deal with this problem, we have preferred to resort to the Laplacian distribution (i.e particular case of GG, when $\beta_j = 1$) which has been used for subband image coding [Birney, Fischer, 1995; Price, Rabbani, 1999]. Indeed, the Laplacian distribution allows an optimal trade-off between the simplicity and tractability of the model and modeling accuracy. The related probability density function \tilde{f}_j is given by:

$$\forall \xi \in \mathbb{R}, \quad \tilde{f}_j(\xi) = \frac{\lambda_j}{2} e^{-\lambda_j |\xi|}. \tag{4.6}$$

where $\lambda_j > 0$ denotes the distribution parameter. It can be estimated as the average of the original coefficients according to the maximum likelihood criterion:

$$\hat{\lambda}_j = \frac{L_j}{\sum_{i=1}^{L_j} |w_j(i)|}. \tag{4.7}$$

However, since we assume in this work that the original versions of the query and model images are not available as they are compressed, we propose to estimate λ_j from the quantized coefficients \bar{w}_j by considering again the maximum likelihood criterion. First, the coefficient values of quantized subband will be distributed according to the discrete Laplace distribution:

$$\bar{f}_j(\bar{w}_j(i)) = \begin{cases} 1 - e^{-\frac{\lambda_j q_j}{2}} & \text{if } \bar{w}_j(i) = 0 \\ \frac{1}{2}(1 - e^{-\lambda_j q_j}) e^{-\lambda_j |\bar{w}_j(i)| + \frac{\lambda_j q_j}{2}} & \text{otherwise.} \end{cases} \tag{4.8}$$

Thus, $\hat{\lambda}_j$ can be estimated by maximizing over λ_j the log-likelihood function of $h_j(\bar{w}_j)$

given by:

$$\bar{\mathcal{L}}_j(\lambda_j; \bar{w}_j) = \log\left[\prod_{i=1}^{L_j} \bar{f}_j(\bar{w}_j(i))\right] = \sum_{i=1}^{L_j} \log[\bar{f}_j(\bar{w}_j(i))] \quad (4.9)$$

hence,

$$\hat{\lambda}_j = \arg \max_{\lambda_j} \{\bar{\mathcal{L}}_j(\lambda_j; \bar{w}_j)\}. \quad (4.10)$$

Substituting (4.8) in (4.10) leads to:

$$\begin{aligned} \hat{\lambda}_j = & \arg \max_{\lambda_j} \left\{ \sum_{\bar{w}_j(i)=0} \log(1 - e^{-\frac{\lambda_j q_j}{2}}) \right. \\ & \left. + \sum_{\bar{w}_j(i) \neq 0} \log \frac{1}{2} (e^{-\lambda_j |\bar{w}_j(i)| + \frac{\lambda_j q_j}{2}}) (1 - e^{-\lambda_j q_j}) \right\}. \end{aligned} \quad (4.11)$$

The derivative of (4.11) with respect to λ_j leads to:

$$(L_j q_j + 2S_j) e^{-\lambda_j q_j} + L_j^0 q_j e^{-\frac{\lambda_j q_j}{2}} + L_j^1 q_j - 2S_j = 0 \quad (4.12)$$

where $S_j = \sum_{\bar{w}_j(i) \neq 0} |\bar{w}_j(i)|$ and, L_j^0 , L_j^1 are respectively the number of wavelet coefficients quantized to zero and non-zero values and, $L_j = L_j^0 + L_j^1$.

Using the substitution:

$$t_j^2 = e^{-\lambda_j q_j}, \quad (4.13)$$

Eq. (4.12) is then a second order polynomial with solution :

$$t_j = \frac{-L_j^0 q_j + \sqrt{(L_j^0 q_j)^2 - 4(L_j q_j + 2S_j)(L_j^1 q_j - 2S_j)}}{2L_j q_j + 4S_j}. \quad (4.14)$$

The estimated parameter $\hat{\lambda}_j$ is computed by substituting (4.12) in:

$$\hat{\lambda}_j = -\frac{2}{q_j} \log(t_j). \quad (4.15)$$

It should be noted that, at very low bitrates where the wavelet coefficients are coarsely quantized to zero, they could not carry any useful information: the value of t_j will tend to zero, and so $\hat{\lambda}_j$ will tend to infinity. This problem has also already been reported in [Ichigaya *et al.*, 2006]. In what follows, we will address this issue in the context of image retrieval.

Proposed retrieval strategy

Due to the estimation problem of the distribution parameter at very low bitrate, the feature vector should be appropriately defined according to the bitrate of the different wavelet

subbands. More precisely, before building this vector, it is necessary to check if $\hat{\lambda}_j$ is a finite number. Based on many experiments, it can be observed that the coarse-scale (resp. fine-scale) subbands represent large (resp. small) percentage of the total bitrate. Indeed, the cases where $\hat{\lambda}_j$ is equal to infinity occur especially at low bitrates in the higher frequency subbands, whereas at middle and high bitrates, $\hat{\lambda}_j$ is often a finite number. Thus, the feature vector is built by taking the $\hat{\lambda}_j$ of the different wavelet subbands while omitting those whose values are equal to infinity. As a result, the resulting feature vectors of the query and model images may have different sizes when they are compressed at different bitrates. For this reason, during the indexing step, we propose to adjust the descriptor vector dimension of images compressed at high bitrate to the size of that obtained for images compressed at low bitrate by omitting the estimated parameter of the high frequency subbands.

Once the feature vectors are generated, the retrieval procedure can be applied. The objective is to search the top candidate database images whose feature vectors $(\hat{\lambda}_j^{\text{db}})_{1 \leq j \leq K}$ (where K is the length of the feature vector) are closer to that of the query one $(\hat{\lambda}_j^{\text{q}})_{1 \leq j \leq K}$, according to a given similarity measure. More precisely, the database image I^{db} and the query one I^{q} are compared by computing the following measure:

$$D(I^{\text{db}}, I^{\text{q}}) = \sum_{j=1}^K \tilde{D}(\hat{\lambda}_j^{\text{db}} \parallel \hat{\lambda}_j^{\text{q}}) \quad (4.16)$$

where $\tilde{D}(\hat{\lambda}_j^{\text{db}} \parallel \hat{\lambda}_j^{\text{q}})$ represents the Kullback-Leibler divergence that measures the dissimilarity of the two Laplacian distributions of parameters $\hat{\lambda}_j^{\text{db}}$ and $\hat{\lambda}_j^{\text{q}}$. It is given by [Do, Vetterli, 2002]:

$$D(\hat{\lambda}_j^{\text{db}} \parallel \hat{\lambda}_j^{\text{q}}) = \log \left(\frac{\hat{\lambda}_j^{\text{db}}}{\hat{\lambda}_j^{\text{q}}} \right) + \frac{\hat{\lambda}_j^{\text{q}}}{\hat{\lambda}_j^{\text{db}}} - 1. \quad (4.17)$$

It can be concluded that JPEG2000 compression, based on uniform quantization, allows good preservation of the energy of the subbands encoded at high bitrates. However, it fails to guarantee this property for the subbands encoded at low bitrates.

To avoid the limitations of the two previous approaches, we have also proposed to relax the constraint of employing an uniform quantizer for encoding the database images, and to resort to other quantizers having some interesting properties that can be exploited for retrieval purpose.

4.5 Design of appropriate retrieval approaches for various quantization schemes

Motivated by the fact that popular indexing methods in the transform domain employ statistical parameters of the resulting subbands as salient features, efficient retrieval techniques will be designed by resorting to quantization schemes that preserve the statistical properties of the source regardless of the bitrate of the query and database images.

4.5.1 Retrieval with moment preserving quantizers

Design of a Moment Preserving Quantizer (MPQ)

Before addressing the MPQ scheme as well as its appropriate retrieval approach, it is worth noting that MPQ is a generalization of block truncation coding, which turned out to be advantageous in image coding [Delp, Mitchell, 1979]. Without loss of generality, we assume that the image subband coefficients w_j to be quantized are viewed as realizations of a random variable W_j whose probability distribution function is denoted by F_{W_j} . A N_j -level MPQ can be characterized in terms of a set of N_j representation levels $\{y_{j,1}, y_{j,2}, \dots, y_{j,N_j}\}$ and $(N_j - 1)$ decision levels $\{s_{j,1}, s_{j,2}, \dots, s_{j,N_j-1}\}$, with $y_{j,1} \leq s_{j,1} \leq y_{j,2} \leq \dots \leq s_{j,N_j-1} \leq y_{j,N_j}$. The idea behind a MPQ is to find the partition and the reconstruction points of the quantizer so that the first $(2N_j - 1)$ moments of W_j are preserved. More precisely, the design of a N_j -level MPQ requires the knowledge of the first $(2N_j - 1)$ moments of W_j as well as the probability distribution function F_{W_j} for which the recursion relationships among the moment can be exploited. Indeed, by considering a wavelet decomposition performed over J resolution levels, for each subband j , with $j \in \{1, \dots, 3J + 1\}$, the N_j -level MPQ can be designed by solving the following system of equations:

$$m_j^n \triangleq \mathbb{E}[W_j^n] = \int_{-\infty}^{+\infty} w_j^n dF_{W_j}(w_j) = \sum_{i=1}^{N_j} y_{j,i}^n P_{j,i} \quad (4.18)$$

for every $n = 1, 2, \dots, 2N_j - 1$, where

- $s_{j,0} \triangleq -\infty$,
- $s_{j,N_j} \triangleq +\infty$
- $P_{j,i} \triangleq F_{W_j}(s_{j,i}) - F_{W_j}(s_{j,i-1})$.

Equation (4.18) is a form of the Gauss-Jacobi mechanical quadrature [Szegő, 1939]. The representation levels $y_{j,i}$ of the N_j -level MPQ are the zeros of the N_j -th degree orthogo-

nal polynomial associated with F_{W_j} . Note that the polynomials can be generated using the standard recursion relation for any three consecutive orthogonal polynomials [Jackson, 2004]. The $P_{j,i}$ are Christoffel numbers and, $s_{j,i}$ and $y_{j,i}$ alternate by the separation theorem of Chebyshev-Markov-Stieltjes [Szegő, 1939]. More technical details about orthogonal polynomials, Christoffel numbers and, the separation theorem can be found in Appendix B. As shown in Eq. (4.18), the design of the MPQ depends on the probability density function of the input subband w_j . To this end, we propose to model the subband coefficients by using the Generalized Gaussian (GG) distribution which has been extensively employed to model the unquantized WT coefficients [Do, Vetterli, 2002] and also the DCT ones [Muller, 1993]. Its probability density function f_j is defined in Eq. (3.1).

The cumulative distribution function (c.d.f.) F_{W_j} of a GG law is expressed as [Nadarajah, 2005]:

$$F_{W_j}(\xi) = \begin{cases} \frac{1}{2} Q_{\frac{1}{\beta_j}} \left(\left(-\frac{\xi}{\alpha_j} \right)^{\beta_j} \right) & \text{if } \xi \leq 0 \\ 1 - \frac{1}{2} Q_{\frac{1}{\beta_j}} \left(\left(\frac{\xi}{\alpha_j} \right)^{\beta_j} \right) & \text{if } \xi > 0. \end{cases} \quad (4.19)$$

Moreover, for a GG distribution, the n^{th} -order moment used in Eq. (4.18) will be expressed as follows:

$$m_j^n = \begin{cases} \alpha_j^n \frac{\Gamma(\frac{n+1}{\beta_j})}{\Gamma(1/\beta_j)} & \text{if } n \text{ is even} \\ 0 & \text{otherwise.} \end{cases} \quad (4.20)$$

Once the principle of the MPQ is addressed for a given subband j , the objective is to define the different N_j values for all the subbands. To this end, a bit allocation among the subbands is carried out in order to compute the levels $(N_1, N_2, \dots, N_{3J+1})$ based on a rate-distortion criterion. More precisely, the average distortion is minimized subject to a constraint on the total available bitrate. The Lagrangian optimization approach will also be used to solve this constrained minimization problem [Shoham, Gersho, 1988].

Indexing strategy with adaptive feature extraction

Unlike the conventional method involving an uniform quantizer where the energy (i.e the second order moments) of all the subbands are retained as salient descriptors, a particular attention should be paid to select the subbands j according to the assigned number of quantization level N_j . Indeed, based on its definition, a N_j -level MPQ is known to preserve the first $(2N_j - 1)$ moments of the input subband w_j . Therefore, it is first necessary to check for each subband if $N_j \geq 2$ in order to guarantee the preservation of the second

order moment of the j^{th} subband. Then, only the subbands which satisfy this condition are selected and their energies (*i.e.* second order moments m_j^2) are taken to build the feature vector. Generally, it is well known that the coarse-scale (resp. fine-scale) subbands represent large (resp. small) percentage of the total bitrate. For instance, after applying the bit allocation algorithm among the different subbands, it can be observed that the cases where the second order moment is not preserved (*i.e.* $N_j = 1$) occur frequently at low bitrates in the higher frequency subbands, whereas at middle and high bitrates, N_j is often greater than or equal to 2. Consequently, we propose to build the feature vector by computing

- the second order moment of *only* the low frequency subbands at low bitrate,
- the second order moment of *all* the subbands at middle and high bitrates.

It is worth pointing out that the resulting feature vectors of the query and model images may have different sizes when we encounter images compressed at different bitrates. As a result, during the indexing step, the descriptor vector dimension of images compressed at high bitrate will be adjusted to the size of that obtained with images compressed at low bitrate by omitting the second order moments of the high frequency subbands.

4.5.2 Retrieval approach with distribution preserving quantizers

Design of a Distribution Preserving Quantizer (DPQ)

While the MPQ presents the advantage of preserving a certain statistical moments of the source, its main limitation is that the rate of the MPQ and the number of moments it can preserve are mutually dependent. For example, as aforementioned, at very low bitrate, the N_j value is often equal to 1 which means that the energy of the subband is not preserved. To alleviate this drawback, it has been recently developed a DPQ scheme where the probability distribution of the source, and so *all* statistical properties can be preserved at *any* bitrate [Li, 2011]. This scheme has shown a superior performance over conventional quantizers in audio coding [Li *et al.*, 2010; Li, 2011] and multiple description coding [J. Klejsa, Kleijn, 2013].

While a DPQ can be designed in different ways, it was shown that the optimal one (in terms of rate-distortion performance) can be achieved by using a dithered quantizer followed by a non-linear transformation [Li *et al.*, 2010]. This latter, referred to as transformation-based

DPQ, is illustrated in Fig. 4.5.

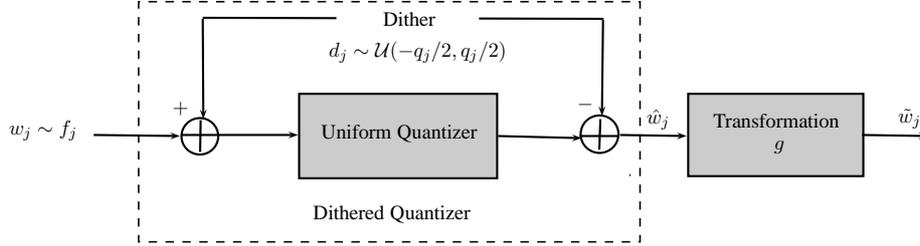


Figure 4.5: Diagram of a transformation-based DPQ.

Indeed, as defined in Section 4.5.1, let us emphasize that the subband coefficients w_j to be quantized are modeled by a GG distribution with probability density function f_j and c.d.f F_{W_j} given respectively by Eqs. (3.1) and (4.19). As it can be seen in Fig. 4.5, a dither d_j , uniformly distributed over $[-q_j/2, q_j/2]$, is added to w_j before quantization and subtracted from the quantized signal, yielding \hat{w}_j . Finally, a transformation g is performed on \hat{w}_j to produce the final reconstruction signal \tilde{w}_j while preserving the distribution of the source. Let us assume that \hat{w}_j is also a realization of a random variable \widehat{W}_j . By choosing a uniformly distributed dither d_j , it is possible to check that its characteristic function ϕ_{d_j} satisfies the following conditions:

$$\phi_{d_j}(k/q_j) = 0, \text{ for } k = \pm 1, \pm 2, \dots \quad (4.21)$$

Indeed, Schuman has shown that the conditions in (4.21) are equivalent to guarantee that the quantization error $e_j = \hat{w}_j - w_j$ is statistically independent of the source [Schuchman, 1964]. Consequently, under the Schuman's conditions, \widehat{W}_j follows a continuous probability distribution and its c.d.f $F_{\widehat{W}_j}$ is related to F_{W_j} as follows:

$$F_{\widehat{W}_j}(\hat{w}_j) = \frac{1}{q_j} \int_{-\frac{q_j}{2}}^{\frac{q_j}{2}} F_{W_j}(\hat{w}_j - \tau) d\tau. \quad (4.22)$$

Thus, the transformation g consists in applying firstly the above operation $F_{\widehat{W}_j}$ on \hat{w}_j and then performing $F_{W_j}^{-1}$ on the obtained result in order to retrieve the source (i.e subband) distribution:

$$g(\hat{w}_j) = F_{W_j}^{-1} \left(\frac{1}{q_j} \int_{-\frac{q_j}{2}}^{\frac{q_j}{2}} F_{W_j}(\hat{w}_j - \tau) d\tau \right) \quad (4.23)$$

where the inverse c.d.f $F_{W_j}^{-1}$ is given by:

$$F_{W_j}^{-1}(\xi) = \begin{cases} -\alpha_j(Q_{\frac{1}{\beta_j}}^{-1}(2\xi))^{\frac{1}{\beta_j}} & \text{if } \xi \leq 1/2 \\ \alpha_j(Q_{\frac{1}{\beta_j}}^{-1}(2(1-\xi)))^{\frac{1}{\beta_j}} & \text{if } \xi > 1/2 \end{cases} \quad (4.24)$$

with $Q_{\frac{1}{\beta_j}}^{-1}$ denotes the inverse function of the normalized incomplete Gamma function. To illustrate the interest of the transformation g in preserving the distribution of the input signal, Fig. 4.6 illustrates the histograms of the diagonal detail coefficients, before and after transformation, for a given image taken from the standard VisTex dataset [vis,]. It can be noticed that at high bitrate, the probability distribution of the output of the dithered quantizer \hat{w}_j is similar to that of the input subband w_j , and hence, it is slightly modified after the transformation. However, at low and middle bitrates, the probability distribution of the output of the dithered quantizer \hat{w}_j follows a nearly uniform distribution, and it is modified by the transformation to become very close to that of the input subband w_j .

Indexing strategy

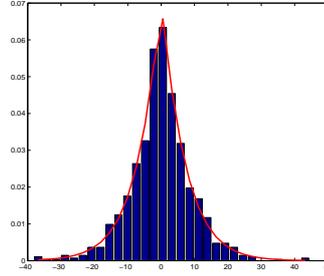
An efficient retrieval strategy which can be applied in the context of a DPQ-based coding scheme is depicted in Fig. 4.7.

Indeed, when examining the structure of a DPQ-based coder, it can be seen that the block involving the subtraction of the dither followed by the transformation g corresponds to the decoding part which aims at reconstructing the subbands \tilde{w}_j while maintaining the statistical properties of the original subbands w_j . From this point of view, it would be interesting in the DPQ-based coding context to define the features from the synthesized subbands \tilde{w}_j . Since the main advantage of DPQ is its independency of the bitrate, the feature will be extracted from all the subbands. To this end, during the indexing step, we propose to model the synthesized detail subbands \tilde{w}_j by using the GG distribution and, the estimated scale and shape parameters $(\tilde{\alpha}_j, \tilde{\beta}_j)_{1 \leq j \leq 3J}$ are used as a salient features to form the descriptor vector. After that, a similarity measure should be defined to compare the query image I^q and the database one I^{db} characterized respectively by the feature vectors $(\tilde{\alpha}_j^q, \tilde{\beta}_j^q)_{1 \leq j \leq 3J}$ and $(\tilde{\alpha}_j^{db}, \tilde{\beta}_j^{db})_{1 \leq j \leq 3J}$. This can be done by computing the Kullback-Leibler divergence given by:

$$KLD(I^{db}, I^q) \triangleq \sum_{j=1}^{3J} \left(\log \left(\frac{\tilde{\beta}_j^{db} \tilde{\alpha}_j^q \Gamma(1/\tilde{\beta}_j^q)}{\tilde{\beta}_j^q \tilde{\alpha}_j^{db} \Gamma(1/\tilde{\beta}_j^{db})} \right) - \frac{1}{\tilde{\beta}_j^{db}} + \left(\frac{\tilde{\alpha}_j^{db}}{\tilde{\alpha}_j^q} \right)^{\tilde{\beta}_j^q} \frac{\Gamma((\tilde{\beta}_j^q + 1)/\tilde{\beta}_j^{db})}{\Gamma(1/\tilde{\beta}_j^{db})} \right). \quad (4.25)$$

To conclude this section, it is important to emphasize that the three proposed retrieval

Original wavelet coefficients distribution

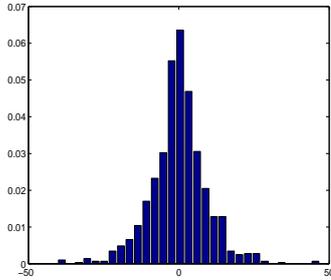


$$w_j: (\alpha_j = 7.40, \beta_j = 1.08)$$

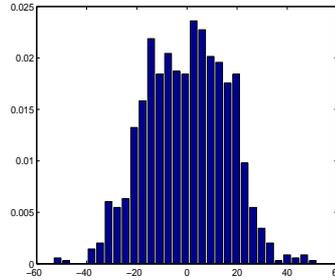
(a) high rate (1.5 bpp)

(b) medium rate (0.5 bpp)

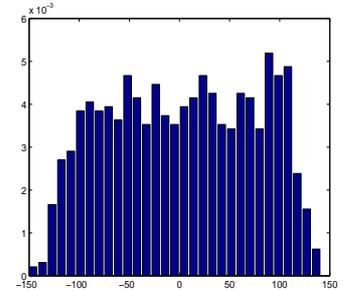
(c) low rate (0.1 bpp)



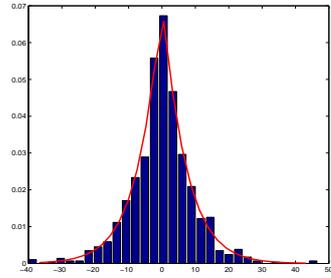
$$\hat{w}_j$$



$$\hat{w}_j$$

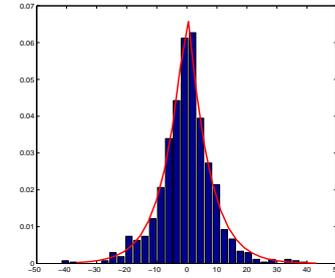


$$\hat{w}_j$$



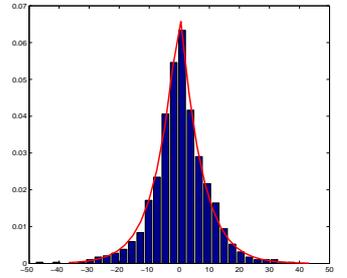
$$\tilde{w}_j = g(\hat{w}_j)$$

$$(\tilde{\alpha}_j = 6.88, \tilde{\beta}_j = 1.03)$$



$$\tilde{w}_j = g(\hat{w}_j)$$

$$(\tilde{\alpha}_j = 7.05, \tilde{\beta}_j = 1.08)$$



$$\tilde{w}_j = g(\hat{w}_j)$$

$$(\tilde{\alpha}_j = 7.40, \tilde{\beta}_j = 1.09)$$

Figure 4.6: Histograms of the diagonal wavelet coefficients at the second resolution level for an image taken from the VisTex dataset at high, medium and low bitrates.

approaches adapted to the different image quantization schemes constitute a general framework which is independent of the transform domain used to encode the database images. Indeed, while it is well known that the UQ has been employed to quantize the DCT and wavelet coefficients, the MPQ- and DPQ-based methods require to model the transformed coefficients. To this end, we have retained the GG distribution which has been successfully

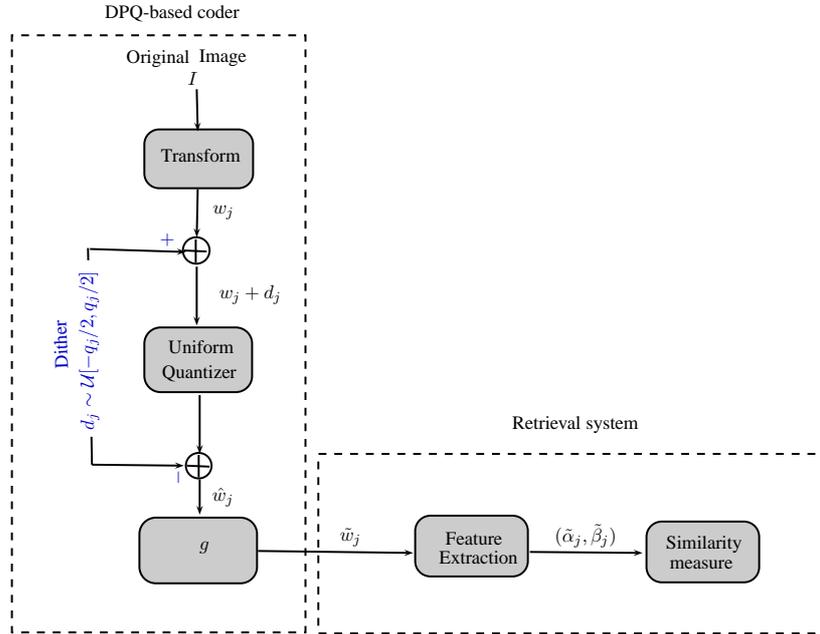


Figure 4.7: Block diagram of the proposed retrieval system under a DPQ-based encoding scheme.

used to model the DCT and wavelet coefficients [Do, Vetterli, 2002; Muller, 1993]. Therefore, the developed methods with MPQ and DPQ are appropriate for both DCT and WT domains.

4.6 Experimental results

Experiments were conducted on three different well known datasets of texture images. The first one corresponds to the Vision Texture (VisTex) database [vis,]. The dataset ¹ consists of different natural surfaces that have quite easily distinguishable texture properties. The visual properties of the 40 selected texture images, of size 512×512 , do not change too much over the image. Each image was split into 16 non-overlapping images which results in a collection of 640 images of size 128×128 . The second one is the Outex database [Ojala *et al.*, 2002a]. In particular, we select the Outex test suite, referred to as Outex_TR_00000, which was created for retrieval purpose. These test sets contain 319 different grayscale textures ², with 20 samples for each texture. The third one is the Salzburg Texture (STex)

¹<http://vismod.media.mit.edu/vismod/imagery/VisionTexture/>

²<http://www.outex.oulu.fi/>

database [Kwitt, Meerwald,] which corresponds to a large collection of 476 images with different textures³. Each image of size 512×512 is divided into 16 non-overlapping subimages of size 128×128 , which results in a database of 7,616 images. Some samples of images taken from the different databases are shown in Fig. 4.8.

The retrieval performances are evaluated in terms of precision and recall. Note that relevant images for each query correspond to all the images from the same category. Besides, in the experiments, each image in the dataset will be considered as a query image for the complete database.

In the first part of our experiments, we focus on the case of wavelet-based quantized images. To this respect, the 9/7 lifting scheme is applied to all the database images over $J = 3$ resolution levels. Thus, the number of resulting subbands for each image is equal to 10. Once the wavelet representations of the different images are generated, we proceed to the different quantization techniques in order to evaluate their associated retrieval approaches.

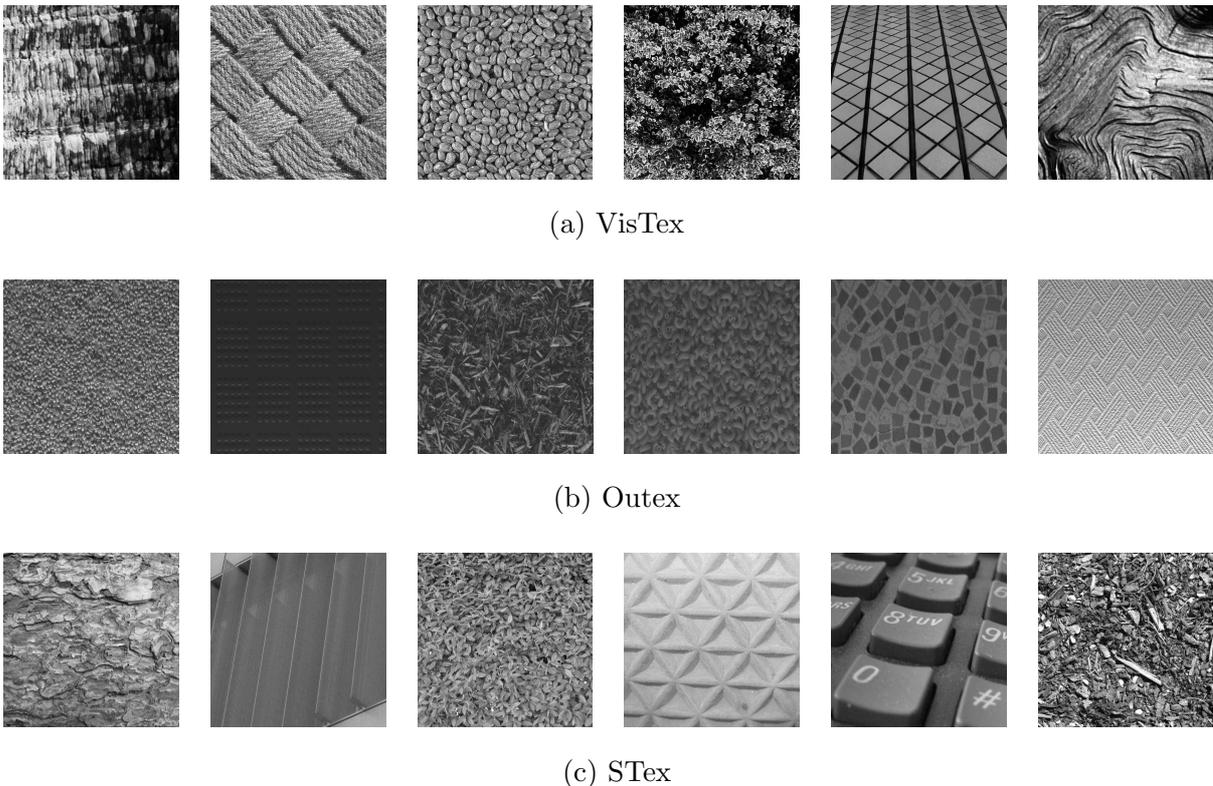


Figure 4.8: Some examples of images for different databases.

³<http://www.wavelab.at/sources>

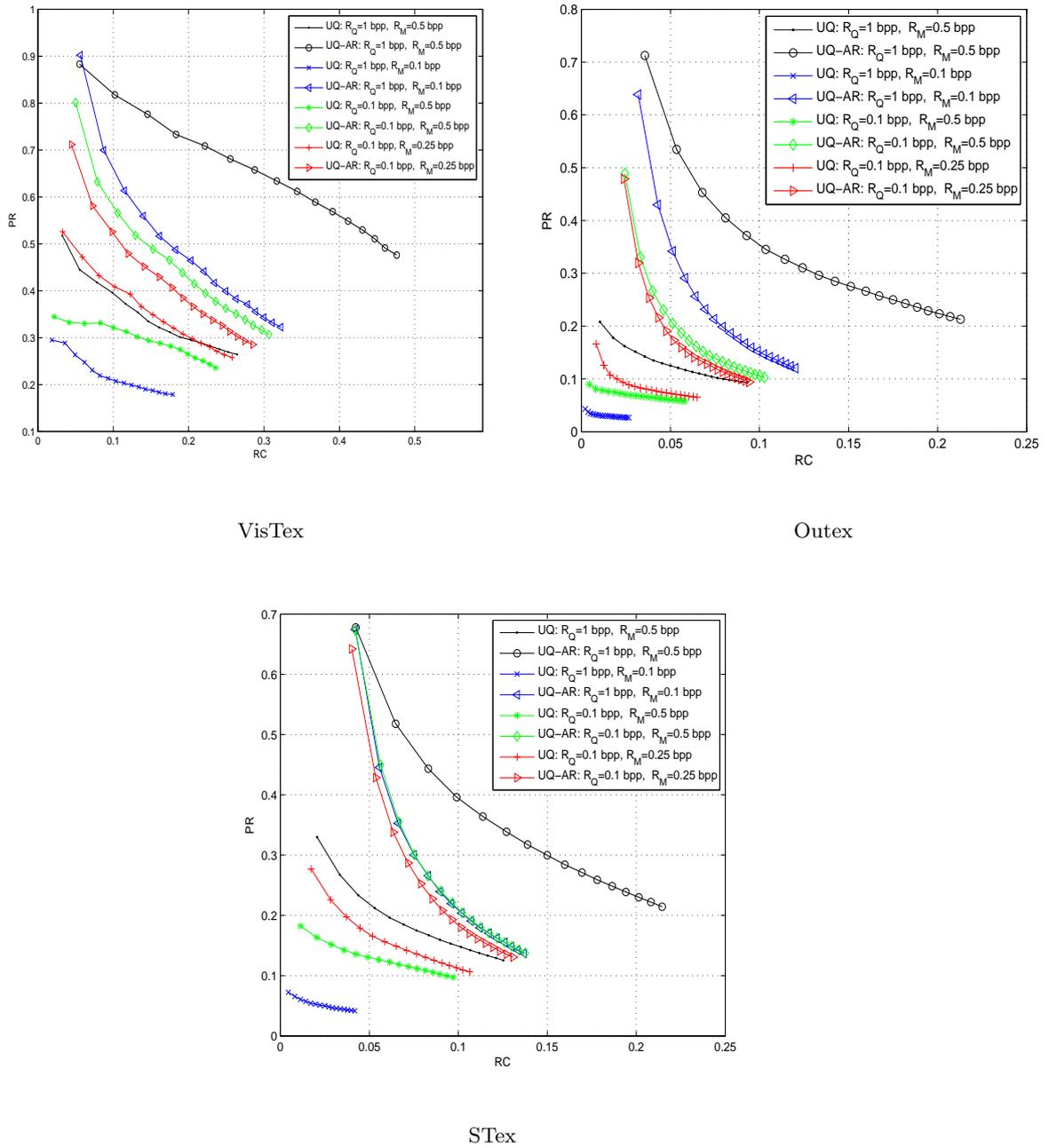


Figure 4.9: Retrieval performance in terms of precision and recall for UQ-based compressed images.

First, Fig. 4.9 provides the retrieval results for the three compressed databases VisTex, Outex and STex when the query and model images are respectively uniformly quantized at different bitrates R_Q and R_M . The intuitive approach consists in considering the energy values of all the quantized wavelet subbands for each image. This method is denoted in

Fig. 4.9 by UQ since the features are extracted directly from the output of the Uniform Quantizer (UQ). Retrieval results prove that the retrieval performances are affected by an UQ-based compression scheme, especially when query and model images have different qualities. To improve these results, a pre-processing step that consists in re-quantizing the high quality image to the low one is added before extracting the energy-based features. This method is designated in Fig. 4.9 by UQ-AR (since features are extracted After Re-quantization). Let us recall here that this approach corresponds to the state-of-the-art method that has been recently introduced in the DCT domain in [Edmundson, Schaefer, 2012b], and then, we have extended it to the WT domain in [Chaker *et al.*, 2012]. Thus, we can see from Fig. 4.9 that the retrieval performances of the UQ after re-quantization step are significantly improved. Obviously, the improvement is more important when the difference of qualities between the query and model images increases (i.e. when $|R_Q - R_M|$ is high).

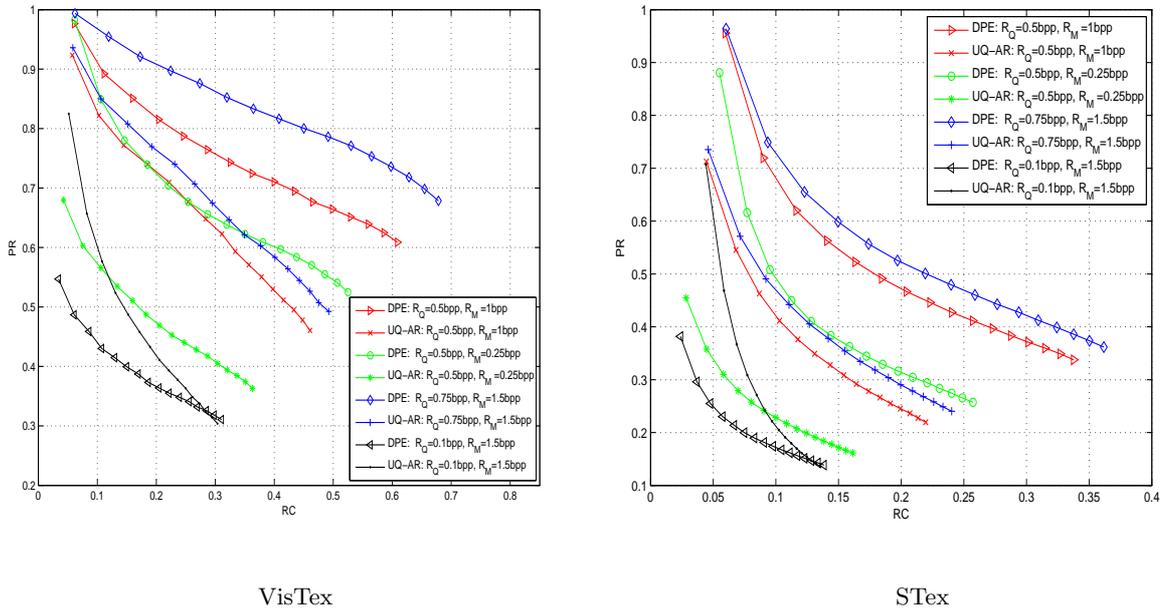


Figure 4.10: Wavelet-based compressed images of the DPE approach and the recompression one UQ-AR.

Always in the case of JPEG 2000 compressed images, the proposed retrieval approach based on the distribution parameter estimation, which is denoted in what follows by DPE, will be compared to the UQ-AR method. Thus, as shown in Fig. 4.10, DPE approach outperforms the retrieval method based on the recompression technique when images are compressed

at middle and high bitrates. However, DPE approach fails when images are compressed at very low bitrate. This can be explained by the limitation of DPE mentioned in Section 4.4.2. Let us recall that it has been shown that, at very low bitrate, the wavelet coefficients are coarsely quantized to zero and therefore the estimated distribution parameter will tend to infinity. After that, the UQ-AR and DPE approaches will be compared to the case when the images are encoded by using the moment preserving quantizer (MPQ). Fig. 4.11 provides the plots of the precision versus recall when the query and model images are quantized at different bitrates. It can be seen that MPQ-based retrieval approach outperforms the retrieval method based on an UQ followed by the pre-processing step and DPE approaches. The gain in performances becomes much more important when the query or model images are compressed at low and middle bitrates.

We should emphasize here on the benefits of the proposed adaptive feature extraction method, in the context of MPQ-based coding scheme, for the definition of the feature vector described in Section 4.5.1. This can be shown in Fig. 4.12 where the feature vectors are built by considering the two following cases:

- (i): the second order moments of *all* the quantized wavelet subbands are considered,
- (ii): the second order moments of *only* the quantized wavelet subbands where $N_j \geq 2$ are selected.

The obtained results confirm the interest of adapting the energy-based feature vector according to the N_j level of the different subbands.

Finally, the retrieval approach under the MPQ is compared to the proposed method developed with a DPQ-based coding scheme. Fig. 4.13 depicts the retrieval results when the model images and the query one are quantized at different bitrates. It can be observed that the proposed retrieval approach under the DPQ achieves a significant improvement compared to that proposed under the MPQ. For instance, while the latter is still inefficient when the database images is compressed at low bitrates, the image retrieval system based on DPQ is very efficient and its performances are almost *independent* of the bitrate. Furthermore, as shown in Fig. 4.13, DPQ-based retrieval approach gives similar results to standard wavelet-based approach, used generally in the context of retrieval of losslessly encoded database images, where the GGD parameters are extracted from the original wavelet subbands.

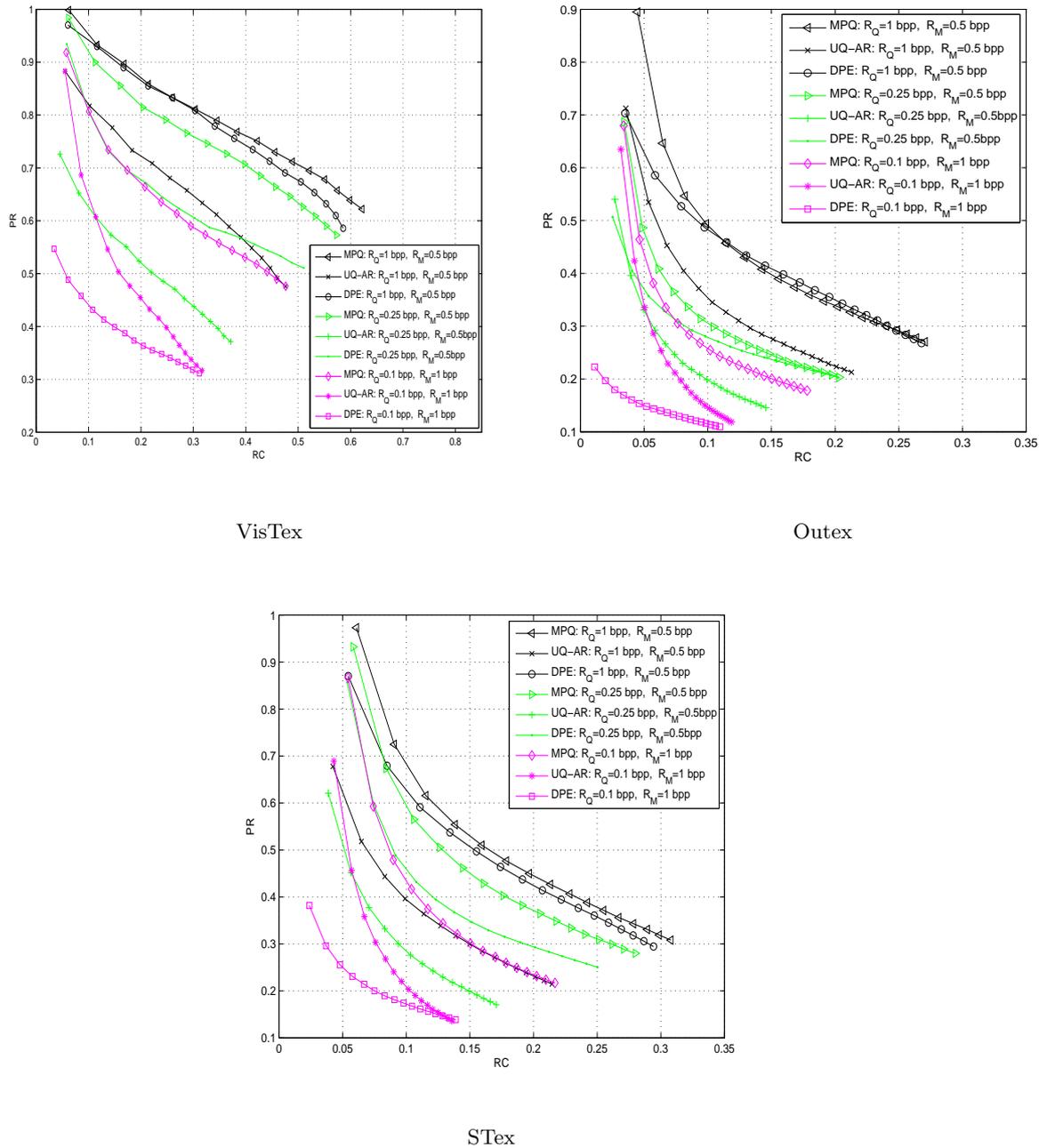


Figure 4.11: Retrieval performance in terms of precision and recall when using the MPQ-based coding scheme, the UQ-AR and DPE approaches.

Figures 4.14, 4.15 and 4.16 show some examples of retrieval results of the compressed STex texture image database to demonstrate the capability the UQ-AR approach, the MPQ-based approach and the DPQ-based approach, respectively. The query is “fabric” compressed at 0.1 bpp, all database images are compressed at 0.5 bpp using the corre-

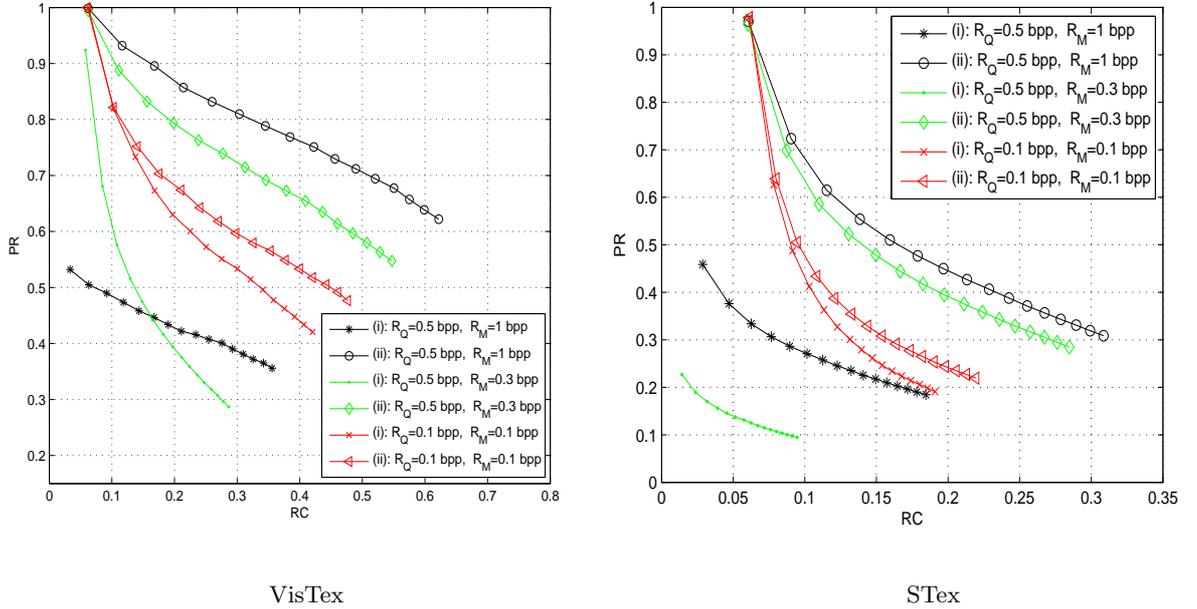
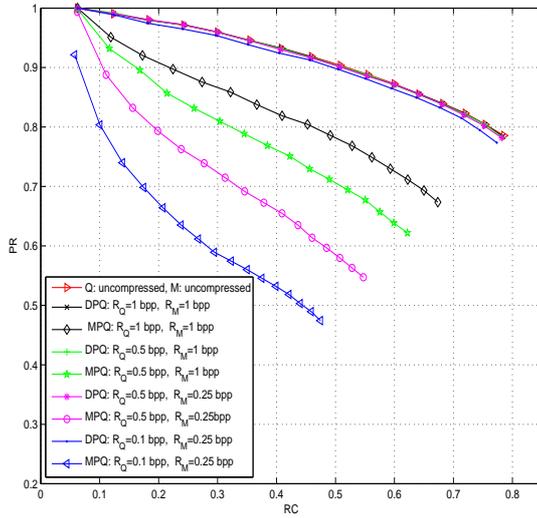


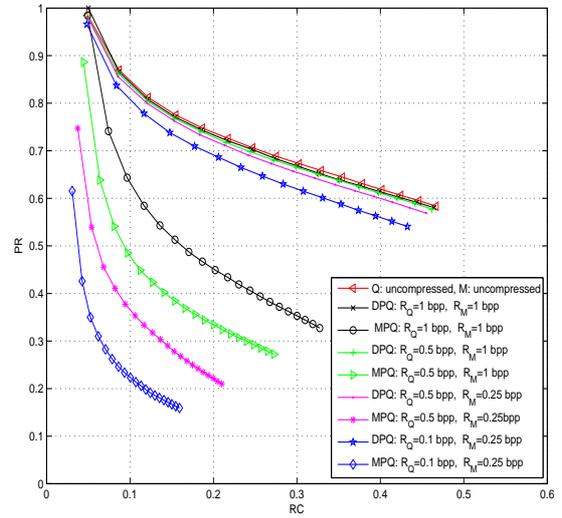
Figure 4.12: Retrieval performance in terms of precision and recall of the MPQ-based coding scheme when the feature vectors are defined by retaining the second order moments of (i): *all* the wavelet subbands, (ii): *only* the wavelet subbands where $N_j \geq 2$.

sponding compression scheme. It can be shown from Fig. 4.14, that only three of the top twenty retrieved images are relevant. This result shows that when query is compressed at very low bitrate, the retrieval system based on UQ-AR approach fails. From Fig. 4.15, we can see that in the case of MPQ-based retrieval approach eight correct matches are returned in the top twenty retrieved images. Thus, MPQ-based approach has shown little improvement in performance when compared to the UQ-AR approach. In Fig. 4.16, the DPQ-based retrieval system perfectly retrieves all images of the same “fabric”, and all relevant images are correctly ranked as the top matches following by images of similar textures.

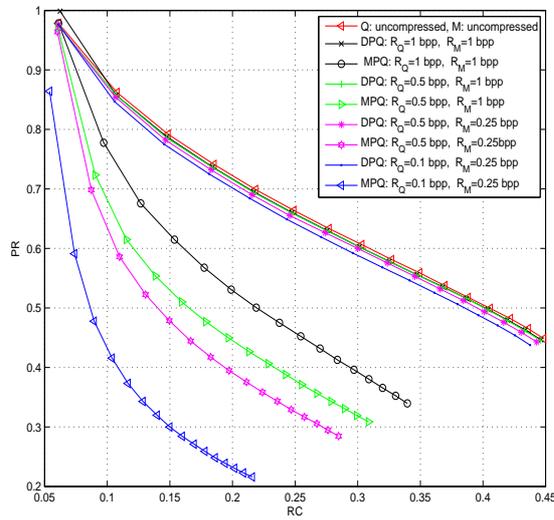
For completeness sake, the performance of the developed retrieval approaches under the different quantizers has also been compared when the images are quantized in the DCT domain. Thanks to the rearrangement technique applied to the DCT coefficients to form structures similar to wavelet subbands [Au *et al.*, 2007], the same salient features used previously with the wavelet-based quantized images will also be extracted from the DCT-based quantized images. As it can be seen in Fig. 4.17, the behavior of the precision-recall curves are similar to those obtained with the wavelet-based quantized images. Indeed, the



VisTex



Outex



STex

Figure 4.13: Retrieval performance in terms of precision and recall for the MPQ-based coding scheme and DPQ-based coding scheme.

MPQ-based retrieval approach outperforms the state-of-the-art one based on UQ followed by the re-quantization technique. Most importantly, the proposed retrieval system based on the DPQ scheme leads to the best results and its performances remain independent of the bitrate (i.e. the qualities of the query and model images).

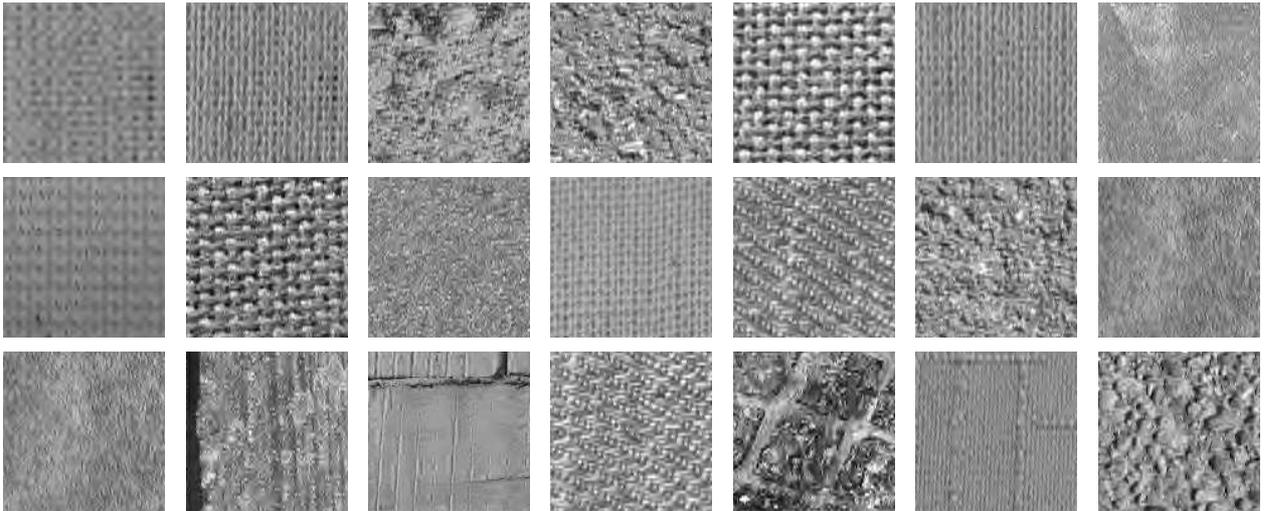


Figure 4.14: Examples of retrieval results from compressed *S*Tex database images using UQ-AR approach. Query image (compressed at 0.1bpp) is on the top left corner; all other images (compressed at 0.5 bpp) are ranked in the order of similarity with the query image from left to right, top to bottom.

4.7 Conclusion

In this chapter, we have investigated different quantization techniques and proposed efficient retrieval methods adapted to the inherent properties of each quantizer. More precisely, the standard uniform quantizer, the moment preserving quantizer and the new introduced distribution preserving quantizer are considered. While the proposed retrieval approaches under the first and second quantizers may be somewhat limited for images compressed at *low* quality, it has been shown that the retrieval approach under the DPQ leads to good results whatever the quality of the query and model images of the database.

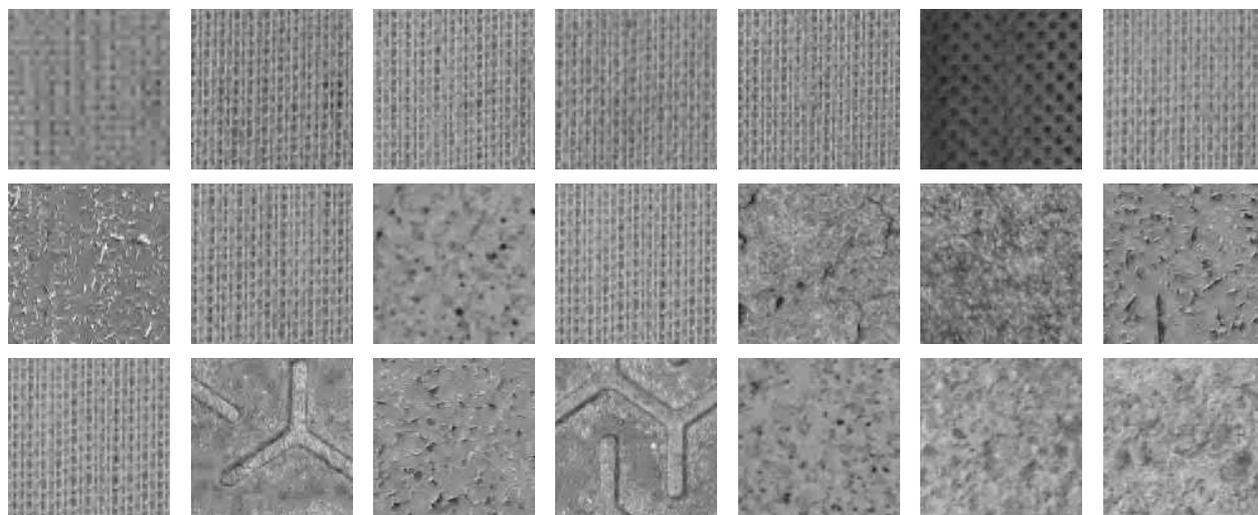


Figure 4.15: Examples of retrieval results from compressed STex database images using MPQ scheme. Query image (compressed at 0.1bpp) is on the top left corner; all other images (compressed at 0.5 bpp) are ranked in the order of similarity with the query image from left to right, top to bottom.

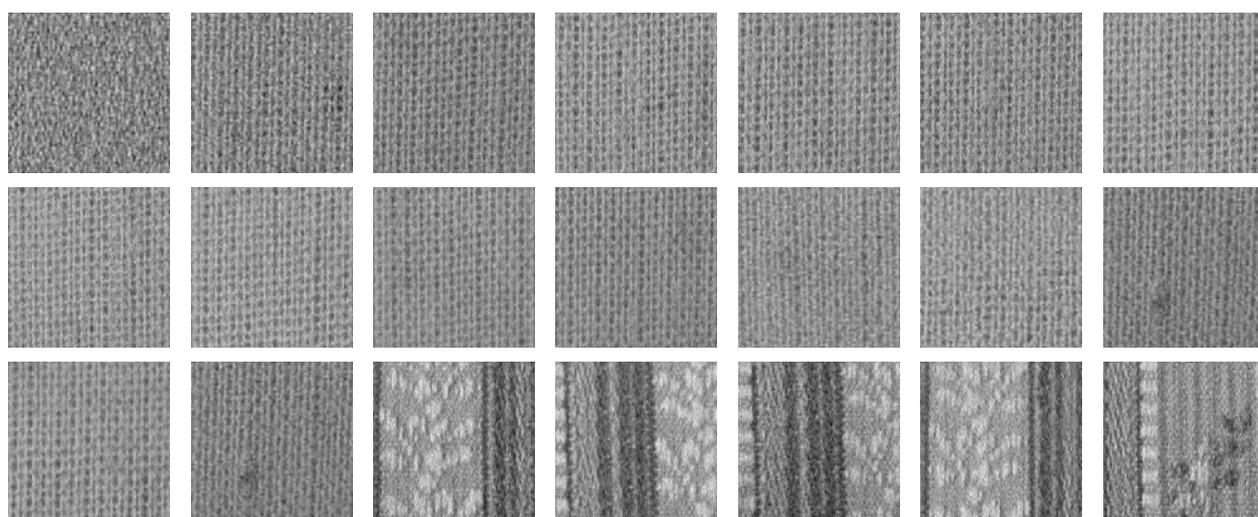


Figure 4.16: Examples of retrieval results from compressed STex database images using DPQ scheme. Query image (compressed at 0.1bpp) is on the top left corner; all other images (compressed at 0.5 bpp) are ranked in the order of similarity with the query image from left to right, top to bottom.

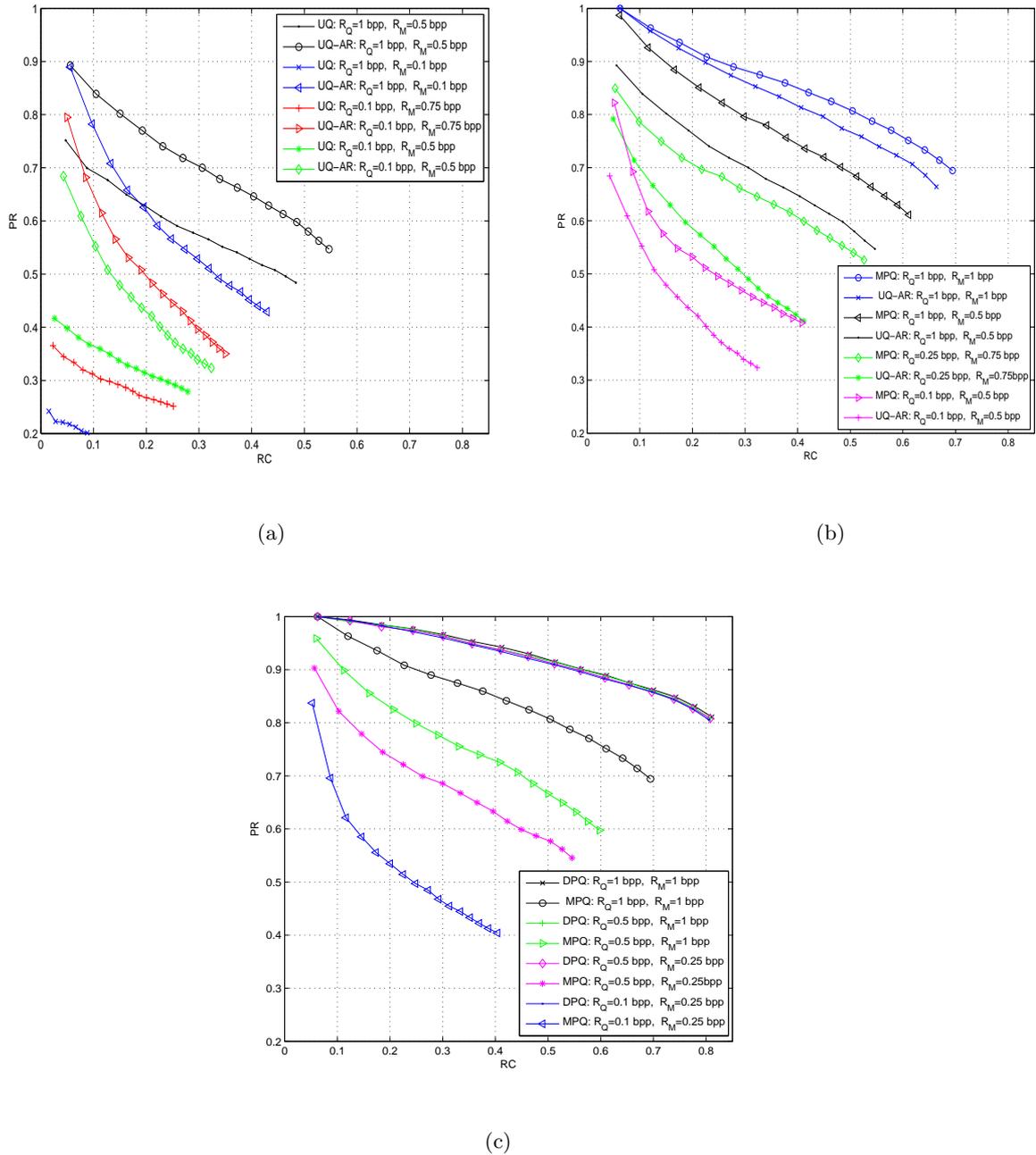


Figure 4.17: Retrieval performance in terms of precision and recall of the DCT-based quantized VisTex database images for the different retrieval approaches under the: (a) UQ-AR, (b) MPQ and (c) DPQ.

Conclusion and future work

This thesis has been devoted to the design of new and efficient retrieval schemes in the wavelet transform domain. Two important issues in CBIR have been investigated. The first one concerns the stereo images which play a crucial role in different application fields. The second one is related to the quantization effect on the performance of CBIR systems since much images are presented in the compressed format. The performances obtained by the different proposed approaches were evaluated on different texture images databases. Retrieval results show the relevance and robustness of the proposed methods.

The main contributions of this PhD are the followings:

- **Methods for content-based stereo image retrieval.** Unlike the state of the art approach which consists of applying the conventional CBIR system to *only* one view (for example the left one) and refining the obtained results by comparing features extracted from the estimated disparity maps, the proposed retrieval schemes exploit simultaneously the visual contents of the left and right images as well as their associated disparity fields. To this respect, we have studied different ways to account for the disparity in order to extract relevant features. More precisely, two principal stereo image retrieval approaches, operating in the WT domain, have been developed. The distribution parameters of wavelet subbands of each image have been captured by using the generalized Gaussian statistical model, and the estimated parameters are employed in the retrieval process. More precisely, in the first scheme, the two views are processed separately through a univariate modeling. Despite its simplicity, this scheme does not exploit the strong statistical correlation between the wavelet coefficients of two views. Thus, in order to capture the dependencies across the wavelet coefficients of the left and right images, we have resorted to a bivariate modeling in the second retrieval scheme. Moreover, to exploit simultaneously the *spatial* and
-

cross-view dependencies of the stereo images, we have employed multivariate modeling based on SIRV model. Experimental results have indicated the outperformance of combining the disparity information with the bivariate model.

One perspective of this work is to investigate the effect of compression of the stereo images on the retrieval performance. Furthermore, the proposed stereo images retrieval approaches can be extended to the context of multiview data.

- **Retrieval strategies for JPEG 2000 compressed images.** We have shown that quantization has a negative impact on the performance of WT-based CBIR system. To reduce an eventual drop in performances, we have firstly proposed a recompression strategy that imposes on query and model images to have similar qualities before feature extraction. Similar recompression strategy has also been investigated in the context of DCT-based CBIR system. This approach is simple and robust. However, it presents also some shortcomings. Indeed, the reconstruction and the requantization steps lead to a computational overhead at the retrieval stage, particularly if the model images have a quality higher than that of the query image. To alleviate this drawback, we have proposed an alternative strategy which consists of recovering the statistical parameters of the original wavelet coefficients directly from the quantized ones. To this end, we have considered the Laplacian distribution to model the distribution of the wavelet coefficients. Experiments, conducted on three different datasets of texture images, have shown the benefits of using the recompression technique and the retrieval approach based on the distribution parameter estimation.

- **Appropriate retrieval approaches for various quantization schemes.** We have investigated different quantization schemes that preserve statistical features of the input signal and propose for each one an efficient retrieval approach. More precisely, the moment preserving quantizer and the distribution preserving quantizer have been considered. The characteristics of each quantizer have been then taken into account in order to design an appropriate retrieval strategy, and so improve the drop of retrieval performances resulting from the quantization effect. Experimental results, carried out on three standard texture databases, show the benefits which can be drawn from the proposed retrieval approaches.

An important area of research will consist in designing other sophisticated features

more robust against quantization effects. Last but not least, further investigations deserve to be conducted concerning the design of retrieval strategies that account for some visual degradations such as blurring, low contrast corrupting the model or the query images.

Appendix A

Subgradient projections and optimization algorithm

A.1 Subgradient projections

We will briefly recall the basic elements on subgradient projection which are necessary for describing the optimization algorithm. More details can be found in [Combettes, 2003]. Let S_k be a nonempty closed and convex subset of the Hilbert image space \mathbb{H} with scalar product $\langle \cdot | \cdot \rangle$ and Euclidean norm $\| \cdot \|$. For every point \mathbf{u} in \mathbb{H} , there exists a unique point $P_k(\mathbf{u}) \in S_k$ called the projection of \mathbf{u} onto S_k such that $\| \mathbf{u} - P_k(\mathbf{u}) \| = d_{S_k}(\mathbf{u})$, where $d_{S_k}(\mathbf{u}) = \inf_{\mathbf{v} \in S_k} \| \mathbf{u} - \mathbf{v} \|$. We assume that S_k is given by Eq. (3.20), where f_k is a continuous convex function. For every $\mathbf{u} \in \mathbb{H}$, f_k has at least one subgradient at \mathbf{u} , i.e., a vector $\mathbf{t}_k \in \mathbb{H}$ such that

$$\forall \mathbf{z} \in \mathbb{H}, \quad \langle \mathbf{z} - \mathbf{u} | \mathbf{t}_k \rangle + f_k(\mathbf{u}) \leq f_k(\mathbf{z}). \quad (\text{A.1})$$

The set of all subgradients of f_k at \mathbf{u} is the subdifferential of f_k at \mathbf{u} and is denoted by $\partial f_k(\mathbf{u})$. If f_k is differentiable at \mathbf{u} , then $\partial f_k(\mathbf{u}) = \{ \nabla f_k(\mathbf{u}) \}$. Fix $\mathbf{u} \in \mathbb{H}$ and a subgradient $\mathbf{t}_k \in \partial f_k(\mathbf{u})$. The subgradient projection of \mathbf{u} onto S_k is given by:

$$G_k(\mathbf{u}) = \begin{cases} \mathbf{u} - \frac{f_k(\mathbf{u}) - \delta_k}{\| \mathbf{t}_k \|^2} \mathbf{t}_k \tilde{\mathbf{d}}_i & \text{if } f_k(\mathbf{u}) > \delta_k \\ \mathbf{u} & \text{otherwise.} \end{cases} \quad (\text{A.2})$$

It is worth noting that computing $G_k(\mathbf{u})$ requires only the availability of a subgradient and is, so, often much easier than computing the exact projection $P_k(\mathbf{u})$ onto S_k . But, when

the projection is simple to compute, one can obviously use it as a subgradient projection. Eventually, we will need to compute the subgradient projections onto the sets S_1 and S_2 .

A.2 Optimization algorithm

The optimization algorithm used to estimate the disparity field u is described as follows [Miled *et al.*, 2009a]:

1. Fix $\epsilon \in]0, 1/K[$ and set $i = 0$. Compute \mathbf{u}_0 as:

$$u_0(s) = (L(s)r(s) + \alpha\bar{u}(s))/(L(s)^2 + \alpha)$$

2. Take a nonempty index set $\mathbb{K}_i \subseteq \{1, \dots, K\}$.
3. For every index $k \in \mathbb{K}_i$, set $\mathbf{a}_{i,k} = G_k(\mathbf{u}_i) - \mathbf{u}_i$, where $G_k(\mathbf{u}_i)$ is a subgradient projection of \mathbf{u}_i onto S_k as in Eq.(A.2).
4. Choose weights $\{\zeta_{i,k}\}_{k \in \mathbb{K}_i} \subset]\epsilon, 1[$ such that $\sum_{k \in \mathbb{K}_i} \zeta_{i,k} = 1$.
Set $\mathbf{z}_i = \sum_{k \in \mathbb{K}_i} \zeta_{i,k} \mathbf{a}_{i,k}$ and $\mathbf{k}_i = \sum_{k \in \mathbb{K}_i} \zeta_{i,k} \|\mathbf{a}_{i,k}\|^2$.
5. If $\mathbf{k}_i = 0$, exit iteration. Otherwise set:

- $\mathbf{b}_i = \mathbf{u}_0 - \mathbf{u}_i$,
- \mathbf{c}_i such that

$$c_i(s) = (L(s)^2 + \alpha)b_i(s)$$

- \mathbf{d}_i such that

$$d_i(s) = z_i(s)/(L(s)^2 + \alpha)$$

- $\tilde{k}_i = k_i / \langle \mathbf{d}_i, \mathbf{z}_i \rangle$.

6. Choose $\lambda_i \in [\epsilon\tilde{k}_i, \tilde{k}_i]$ and set $\tilde{\mathbf{d}}_i = \lambda_i \mathbf{d}_i$.
7. Set $\pi_i = -\langle \mathbf{c}_i, \tilde{\mathbf{d}}_i \rangle$, $\mu_i = \langle \mathbf{b}_i, \mathbf{c}_i \rangle$, $\nu_i = \lambda_i \langle \tilde{\mathbf{d}}_i, \mathbf{z}_i \rangle$ and $\rho_i = \mu_i \nu_i - \pi_i^2$.
8. Set:

$$\mathbf{u}_{i+1} = \begin{cases} \mathbf{u}_i + \tilde{\mathbf{d}}_i & \text{if } \rho_i = 0, \pi_i > 0, \\ \mathbf{u}_0 + (1 + \pi_i/\nu_i)\tilde{\mathbf{d}}_i & \text{if } \rho_i > 0, \pi_i \nu_i \geq \rho_i, \\ \mathbf{u}_i + \frac{\nu_i}{\rho_i}(\pi_i \mathbf{b}_i + \nu_i \tilde{\mathbf{d}}_i) & \text{if } \rho_i > 0, \pi_i \nu_i < \rho_i, \end{cases} \quad (\text{A.3})$$

9. Increment i and go to step (2).

Appendix B

Orthogonal Polynomials

While there is a variety orthogonal polynomials presented in the literature, this section describe a review of only the basic concepts needed to understand the MPQ. For further details about orthogonal polynomials, the reader is referred to [Szegő, 1939].

Let the probability distribution function $F(x)$ be a fixed non-decreasing real-valued function with infinitely many points of increase in the finite or infinite interval $[a, b]$. The Lebesgue-Stieltjes integral of a class of functions $g(x)$ which are measured with respect to $F(x)$ is given by:

$$\int_a^b |g(x)|^2 dF(x) \quad (\text{B.1})$$

is finite is known as $L_F^2(a, b)$.

Let π_n be the set of all polynomials $\{\rho(x)\} \in L_F^2(a, b)$ with degree less than or equal to n .

Hence π_n is a subspace of $L_F^2(a, b)$.

Definition A: Let the probability distribution function $F(x)$ having moments

$$m_n = \int_a^b x^n dF(x); n = 0, 1, 2, \dots \quad (\text{B.2})$$

that exist and finite. If we orthogonalize the set of functions $\{x^n\}$, $n = 0, 1, 2, \dots$ using a Gram-Schmidt procedure we can obtain a set of orthogonal polynomials $p_0(x), p_1(x), \dots, p_n(x), \dots$,

$$\langle p_n(x), p_m(x) \rangle = \int_a^b p_n(x)p_m(x)dF(x) = k_n\delta_{nm} \quad (\text{B.3})$$

where δ_{nm} is the Kronecker delta. $p_n(x)$ is an orthogonal polynomial of degree n where the coefficient of x^n positive.

A similar definition holds if $F(x)$ admits a density function $f(x)$. The set $\{p_n(x)\}$ of orthogonal polynomials is said to be associated with $F(x)$. We shall use the notation

$\{\Psi_n(x)\}$ to denote the normalized set of $\{p_n(x)\}$. If $F(x)$ has only a finite number of points of increase then a finite system of polynomials is obtained.

Theorem 1: The orthonormal polynomials satisfy the three-term recurrence relationship:

$$x\Psi_n(x) = \frac{a_n}{a_{n+1}}\Psi_{n+1}(x) + \left(\frac{b_n}{a_n} - \frac{b_{n+1}}{a_{n+1}}\right)\Psi_n(x) + \frac{a_{n-1}}{a_n}\Psi_{n-1}(x) \quad (\text{B.4})$$

where a_n is the coefficient of x^n in $\Psi_n(x)$ and b_n is the coefficient of x^{n-1} in $\Psi_n(x)$ [Jackson, 2004].

Theorem 2: The zeros of real orthogonal polynomials are real, simple and if $[a, b]$ is a finite interval, the zeros lie in $[a, b]$.

Theorem 3: Let $z_1 < z_2 < \dots < z_n$ be the zeros of $\Psi_n(x)$; $z_0 = a$ and $z_{n+1} = b$. So, there is exactly one zero of $\Psi_{n+1}(x)$ in each interval $[z_i, z_{i+1}]$, $i = 0, 1, 2, \dots, n$.

Theorem 4: Between two zeros of $\Psi_n(x)$, there is at least one zero of $\Psi_m(x)$, $m > n$.

Theorem 5 (Gauss-Jacobi Mechanical Quadrature): Let $z_1 < z_2 < \dots < z_n$ denote the zeros of $\Psi_n(x)$, there exist real numbers $\lambda_1, \lambda_2, \dots, \lambda_n$ such that

$$\int_a^b \rho(x)dF(x) = \sum_{i=1}^n \lambda_i \rho(z_i) \quad (\text{B.5})$$

whenever $\rho(x) \in \pi_{2n-1}$. Thus the numbers λ_i can be determined using uniquely the distribution $F(x)$ and the integer n . λ_i are referred to us as the Christoffel numbers. The λ_k are independent of $\rho(x)$ [Szegő, 1939].

This result is generally used in numerical integration where ρ is replaced by a general function $g(x) \in L_F^2(a, b)$.

Theorem 6: The Christoffel numbers, λ_k are positive and they satisfy

$$\sum_{k=1}^n \lambda_k = \int_a^b dF(x) = F(b) - F(a) = 1. \quad (\text{B.6})$$

and given by

$$\lambda_k^{-1} = \sum_{m=0}^n \Psi_m^2(z_k), \quad ; k = 1, 2, \dots, n. \quad (\text{B.7})$$

From the obtained result that prove the positiveness of the λ_k 's and the previously cited properties of $F(x)$, there exists numbers $q_1 < q_2 < \dots < q_{n-1}$, $a < q_1$, $q_{n-1} < b$ such that

$$\begin{aligned} \lambda_k &= F(q_k) - F(q_{k-1}), \quad ; k = 1, 2, \dots, n \\ q_0 &= a \\ q_n &= b \end{aligned} \quad (\text{B.8})$$

Note that the q_k 's are not in general uniquely determined.

Theorem 7 (Separation Theorem of Chebyshev-Markov-Stieltjes):

The zeros $z_1 < z_2 < \cdots < z_n$ alternate with the numbers $q_1 < q_2 < \cdots < q_n$ that is

$$q_k < z_k < q_{k+1}. \quad (\text{B.9})$$

Hence, Theorem 6 could be written as

$$\int_a^b \rho(x) dF(x) = \sum_{k=1}^n \rho(z_k) (F(x_k) - F(x_{k-1})). \quad (\text{B.10})$$

The proofs of the theorem are described in [Szegő, 1939].

Bibliography

- MIT vision and modeling group. vision texture. <http://vismod.www.media.mit.edu>. [Online].
- ABRAMOWITZ, M., STEGUN, I. A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publishing Inc., New York.
- AIAZZI, B., ALPARONE, L., BARONTI, S. (1999). Estimation based on entropy matching for generalized gaussian PDF modeling. *IEEE Signal Processing Letters*, 6(6):138–140.
- ALKANHAL, M., TURAGA, D., CHEN, T. (1999). Correlation based search algorithms for motion estimation. *In Picture Coding Symposium*, vol. 99, pages 99–102.
- ALVAREZ, L., DERICHE, R., SANCHEZ, J., WEICKERT, J. (2002). Dense disparity map estimation respecting image discontinuities: A PDE and scale-space based approach. *Journal of Visual Communication and Image Representation*, page 13:3â€š21.
- ANDRÉ, T., CAGNAZZO, M., ANTONINI, M., BARLAUD, M. (2007). JPEG2000-compatible scalable scheme for wavelet-based video coding. *EURASIP Journal on Image and Video Processing*, 2007.
- ANTONINI, M., BARLAUD, M., MATHIEU, P., DAUBECHIES, I. (1992). Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1(2):205–220.
- AU, K. M., LAW, N.-F., SIU, W.-C. (2007). Unified feature analysis in JPEG and JPEG 2000-compressed domains. *Pattern recognition*, 40(7):2049–2062.
- BACH, J. R., FULLER, C., GUPTA, A., HAMPAPUR, A., HOROWITZ, B., HUMPHREY, R., JAIN, R. C., SHU, C. F. (1996). Virage image search engine: an open framework for image management. *In Electronic Imaging: Science & Technology*, pages 76–87. International Society for Optics and Photonics.
-

-
- BALAKRISHNANA, N., KATERI, M. (2008). On the maximum likelihood estimation of parameters of Weibull distribution based on complete and censored data. *Statistics and Probability Letters*, 78:2971–2975.
- BECK, J., SUTTER, A., IVRY, R. (1987). Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing*, 37(2):299–325.
- BELONGIE, S., MALIK, J., PUZICHA, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522.
- BERRETTI, S., DEL BIMBO, A., PALA, P. (2000). Retrieval by shape similarity with perceptual distance and effective indexing. *IEEE Transactions on Multimedia*, 2(4):225–239.
- BIRNEY, K. A., FISCHER, T. R. (1995). On the modeling of DCT and subband image data for compression. *IEEE Transactions on Image Processing*, 4(2):186–193.
- BLAKE, A., ZISSERMAN, A. (1987). *Visual reconstruction*, vol. 2. MIT press Cambridge.
- BOMBRUN, L., LASMAR, N., BERTHOUMIEU, Y., VERDOOLAEGE, G. (2011a). Multivariate texture retrieval using the SIRV representation and the geodesic distance. In *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 865–868, Prague, C. Republic.
- BOMBRUN, L., VASILE, G., GAY, M., TOTIR, F. (2011b). Hierarchical segmentation of polarimetric SAR images using heterogeneous clutter models. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2):726–737.
- BOYKOV, Y., VEKSLER, O., ZABIH, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239.
- BOYLE, J. P., DYKSTRA, R. L. (1986). A method for finding projections onto the intersection of convex sets in hilbert spaces. In *Advances in order restricted statistical inference*, pages 28–47. Springer.
-

-
- BROWN, M. Z., BURSCHKA, D., HAGER, G. D. (2003). Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008.
- CALDERBANK, A. R., DAUBECHIES, I., SWELDENS, W., YEO, B. L. (1998). Wavelet transforms that map integers to integers. *Applied and computational harmonic analysis*, 5(3):332–369.
- CASTELLANO, G., BONILHA, L., LI, L. M., CENDES, F. (2004). Texture analysis of medical images. *Clinical radiology*, 59(12):1061–1069.
- CHAKER, A., KAANICHE, M., BENAZZA-BENYAHIA, A. (2012). An improved image retrieval algorithm for JPEG 2000 compressed images. In *IEEE International Symposium on Signal Processing and Information Technology*, pages 1–6, Ho Chi Minh City, Vietnam.
- CHAKER, A., KAANICHE, M., BENAZZA-BENYAHIA, A. (2014). Exploiting disparity information for stereo image retrieval. In *IEEE International Conference on Image Processing*, Paris, France.
- CHAN, W. L., CHOI, H., BARANIUK, R. G. (2006). Multiscale image disparity estimation using the quaternion wavelet transform. In *IEEE International Conference on Image Processing*, pages 1229–1232, Atlanta, GA, USA.
- CHANG, C. C., CHUANG, J. C., HU, Y. S. (2004). Retrieving digital images from a JPEG compressed image database. *Image and Vision Computing*, 22(6):471–484.
- CHARMI, M. A., DERRODE, S., GHORBEL, F. (2008). Fourier-based geometric shape prior for snakes. *Pattern Recognition Letters*, 29(7):897–904.
- CHAUX, C., EL-GHECHE, M., FARAH, J., PESQUET, J. C., PESQUET-POPESCU, B. (2013). A parallel proximal splitting method for disparity estimation from multicomponent images under illumination variation. *Journal of mathematical imaging and vision*, 47(3):167–178.
- CHELLAPPA, R., CHATTERJEE, S. (1985). Classification of textures using gaussian markov random fields. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(4):959–963.
-

- CHEN, C. C., DAPONTE, J. S., FOX, M. D. (1989). Fractal feature analysis and classification in medical imaging. *IEEE Transactions on Medical Imaging*, 8(2):133–142.
- CLAYPOOLE, R. L., DAVIS, G. M., SWELDENS, W., BARANIUK, R. G. (2003). Nonlinear wavelet transforms for image coding via lifting. *IEEE Transactions on Image Processing*, 12(12):1449–1459.
- CLIMER, S., BHATIA, S. K. (2002). Image database indexing using JPEG coefficients. *Pattern Recognition*, 35(11):2479–2488.
- COMBETTES, P. L. (1993). The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208.
- COMBETTES, P. L. (2003). A block-iterative surrogate constraint splitting method for quadratic signal recovery. *IEEE Transactions on Signal Processing*, 51(7):1771–1782.
- COMBETTES, P. L., PESQUET, J. C. (1998). Convex multiresolution analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1308–1318.
- COMBETTES, P. L., PESQUET, J. C. (2004). Image restoration subject to a total variation constraint. *IEEE Transactions on Image Processing*, 13(9):1213–1222.
- CONNERS, R. W., HARLOW, C. A. (1980). A theoretical comparison of texture algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (3):204–222.
- CONNERS, R. W., MCMILLIN, C. W., LIN, K., VASQUEZ-ESPINOSA, R. E. (1983). Identifying and locating surface defects in wood: Part of an automated lumber processing system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):573–583.
- COVER, T. M., THOMAS, J. (1991). *Elements of Information Theory*. John Wiley and Sons, Inc., New York.
- CREMERS, D., POCK, T., KOLEV, K., CHAMBOLLE, A. (2011). Convex relaxation techniques for segmentation, stereo and multiview reconstruction.
- CRIMINISI, A., BLAKE, A., ROTHER, C., SHOTTON, J., TORR, P. H. S. (2007). Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming. *International Journal of Computer Vision*, 71(1):89–110.
-

-
- CROSS, G. R., JAIN, A. K. (1983). Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):25–39.
- CROW, F. C. (1984). Summed-area tables for texture mapping. *ACM SIGGRAPH computer graphics*, 18(3):207–212.
- DATTA, R., JOSHI, D., LI, J., WANG, J. Z. (2006). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 39:5.
- DELP, E. J., MITCHELL, O. R. (1979). Image compression using block truncation coding. *IEEE Transactions on Communications*, 27(9):1335–1342.
- DERRODE, S., DAOUDI, M., GHORBEL, F. (1999). Invariant content-based image retrieval using a complete set of fourier-mellin descriptors. In *IEEE International Conference on Multimedia Computing and Systems*, vol. 2, pages 877–881.
- DERRODE, S., GHORBEL, F. (2004). Shape analysis and symmetry detection in gray-level objects using the analytical fourier–mellin representation. *Signal processing*, 84(1):25–39.
- DETTORI, L., SEMLER, L. (2007). A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography. *Computers in Biology and Medicine*, 37(4):486–498.
- DHOND, U. R., AGGARWAL, J. K. (1989). Structure from stereo—a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510.
- DO, M. N., VETTERLI, M. (2002). Wavelet-based texture retrieval using generalized gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 11(2):146–158.
- DRICOT, A., JUNG, J., CAGNAZZO, M., PESQUET, B., DUFAUX, F., KOVÁCS, P. T., ADHIKARLA, V. K. (2015). Subjective evaluation of super multi-view compressed contents on high-end light-field 3d displays. *Signal Processing: Image Communication*.
- DUFAUX, F., PESQUET-POPESCU, B., CAGNAZZO, M. (2013). *Emerging technologies for 3D video: creation, coding, transmission and rendering*. John Wiley & Sons.
-

-
- EDMUNDSON, D., SCHAEFER, G. (2012a). Fast JPEG image retrieval using optimised huffman tables. *In International Conference on Pattern Recognition (ICPR)*, pages 3188–3191.
- EDMUNDSON, D., SCHAEFER, G. (2012b). Recompressing images to improve image retrieval performance. *In IEEE International Conference on Acoustics, Speech and Signal Processing*, page 4 pages, Kyoto, Japan.
- EDMUNDSON, D., SCHAEFER, G., CELEBI, M. E. (2012). Robust texture retrieval of compressed images. *In IEEE International Conference on Image Processing*, pages 2421–2424, Orlando, USA.
- EGNAL, G., WILDES, R. P. (2002). Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1127–1133.
- FAUGERAS, O. (1993). *Three-dimensional computer vision: a geometric viewpoint*. MIT press.
- FENG, G., JIANG, J. (2003). JPEG compressed image retrieval via statistical features. *Pattern Recognition*, 36(4):977–985.
- FENG, H., EFFROS, M. (2002). On the rate-distortion performance and computational efficiency of the Karhunen-Loeve transform for lossy data compression. *IEEE Transactions on Image Processing*, 11(2):113–122.
- FENG, Y., REN, J., JIANG, J. (2011). Generic framework for content-based stereo image/video retrieval. *Electronics letters*, 47(2):97–98.
- FIELD, M., CLARKE, D., STRUP, S., SEALES, W. B. (2009). Stereo endoscopy as a 3D measurement tool. *In International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5748–5751.
- FLICKNER, M., SAWHNEY, H., NIBLACK, W., ASHLEY, J., HUANG, Q., DOM, B., GORKANI, M., HAFNER, J., LEE, D., PETKOVIC, D. *et al.* (1995). Query by image and video content: The qbic system. *Computer*, 28(9):23–32.
-

-
- FOWLER, J. E., PESQUET-POPESCU, B. (2007). An overview on wavelets in source coding, communications, and networks. *EURASIP Journal on Image and Video Processing*, 2007(1):060539.
- FREITAS, C. C., FRERY, A. C., CORREIA, A. H. (2005). The polarimetric G distribution for SAR data analysis. *Environmetrics*, 16(1):13–31.
- FUSIELLO, A., ROBERTO, V., TRUCCO, E. (2000a). Symmetric stereo with multiple windowing. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(08):1053–1066.
- FUSIELLO, A., TRUCCO, E., VERRI, A. (2000b). A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22.
- GABOR, D. (1946). Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441.
- GEMAN, S., GEMAN, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741.
- GHORBEL, F. (1994). A complete invariant description for gray-level images by the harmonic analysis approach. *Pattern recognition letters*, 15(10):1043–1051.
- GIBSON, J. J. (1950). The perception of the visual world.
- GÓMEZ, E., GÓMEZ-VILLEGAS, M. A., MARIN, J. M. (2003). A survey on continuous elliptical vector distributions. *Revista Matemática Complutense*, 36(1):345–361.
- GULDOGAN, E., GULDOGAN, O., KIRANYAZ, S., CAGLAR, K., GABBOUJ, M. (2003). Compression effects on color and texture based multimedia indexing and retrieval. In *IEEE International Conference on Image Processing*, pages 9–12, Barcelona, Spain.
- HARA, Y., HIRATA, K., TAKANO, H., KAWASAKI, S. (1997). Hypermedia navigation and content-based retrieval for distributed multimedia databases. In *Proceedings of the sixth NEC research symposium on Multimedia computing*, pages 133–148. Society for Industrial and Applied Mathematics.
-

- HARALICK, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804.
- HARALICK, R. M., SHANMUGAM, K., DINSTEIN, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621.
- HARTLEY, R., GUPTA, R. (1993). Computing matched-epipolar projections. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 549–555.
- HARTLEY, R. I., STURM, P. (1997). Triangulation. *Computer vision and image understanding*, 68(2):146–157.
- HILTUNEN, S., PESQUET, J. C., PESQUET-POPESCU, B. (2012). Comparison of two proximal splitting algorithms for solving multilabel disparity estimation problems. *In the 20th European Signal Processing Conference*, pages 1134–1138.
- HIRSCHMÜLLER, H., SCHARSTEIN, D. (2007). Evaluation of cost functions for stereo matching. *In IEEE Conference on Computer Vision and Pattern Recognition, 2007*, pages 1–8.
- HUA, Y., LIU, W. (1998). Generalized Karhunen-Loeve transform. *IEEE Signal Processing Letters*, 5(6):141–142.
- HUFFMAN, D. A. (1952). A method for the construction of minimum redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101.
- ICHIGAYA, A., KUROZUMI, M., HARA, N., NISHIDA, Y., NAKASU, E. (2006). A method of estimating coding PSNR using quantized DCT coefficients. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(2):251–259.
- J. KLEJSA, G. Zhang, M. L., KLEIJN, W. B. (2013). Multiple description distribution preserving quantization. *IEEE Transactions on Signal Processing*, 61(24):6410–6422.
- JACKSON, D. (2004). *Fourier series and orthogonal polynomials*. Dover Publications.
- JULESZ, B. (1962). Visual pattern discrimination. *IRE Transactions on Information Theory*, 8(2):84–92.
-

-
- JUSTEL, A., PEÑA, D., ZAMAR, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & Probability Letters*, 35(3):251–259.
- KAANICHE, M. (2010). *Adaptive vector lifting schemes and applications to stereo image coding*. Thèse de doctorat, Télécom ParisTech.
- KAANICHE, M., BENAZZA-BENYAHIA, A., PESQUET-POPESCU, B., PESQUET, J.-C. (2009a). Vector lifting schemes for stereo image coding. *IEEE Transactions on Image Processing*, 18(11):2463–2475.
- KAANICHE, M., FRAYSSE, A., PESQUET-POPESCU, B., PESQUET, J. C. (2014). A bit allocation method for sparse source coding. *IEEE Transactions on Image Processing*, 23(1):137–152.
- KAANICHE, M., MILED, W., PESQUET-POPESCU, B., BENAZZA-BENYAHIA, A., PESQUET, J. C. (2009b). Dense disparity map representations for stereo image coding. *In IEEE International Conference on Image Processing*, pages 725–728.
- KAY, S. M. (1993). *Fundamentals of statistical signal processing*. Englewood Cliffs, NJ: Prentice Hall.
- KIM, J. C., LEE, K. M., CHOI, B. T., LEE, S. U. (2005). A dense stereo matching using two-pass dynamic programming with generalized ground control points. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pages 1075–1082.
- KIMCHI, R., PALMER, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 8(4):521.
- KIRANYAZ, S., GABBOUJ, M. (2012). *Content-based management of multimedia databases*. Lambert Academic Publishing.
- KOLMOGOROV, V., ZABIH, R. (2001). Computing visual correspondence with occlusions using graph cuts. *In IEEE International Conference on Computer Vision*, vol. 2, pages 508–515.
- KOLMOGOROV, V., ZABIH, R. (2006). Graph cut algorithms for binocular stereo with occlusions. *In Handbook of Mathematical Models in Computer Vision*, pages 423–437.
-

- KOMODAKIS, N., PARAGIOS, N., TZIRITAS, G. (2011). MRF energy minimization and beyond via dual decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):531–552.
- KONRAD, J., LAN, Z. D. (2000). Dense-disparity estimation from feature correspondences. In *Electronic Imaging*, pages 90–101. International Society for Optics and Photonics.
- KOSOV, S., THORMÄHLEN, T., SEIDEL, H.-P. (2009). Accurate real-time disparity estimation with variational methods:. In *International Symposium on Visual Computing*, vol. 5875, pages 796–807, Las Vegas, United States.
- KWITT, R. (2010). *Statistical Modeling in the Wavelet Domain and Applications*. Thèse de doctorat, University of Cambridge, Cambridge, United Kingdom.
- KWITT, R., MEERWALD, P. Salzburg texture image database. <http://www.wavelab.at/sources>. [Online].
- KWITT, R., MEERWALD, P., UHL, A. (2009). Blind DT-CWT domain additive spread-spectrum watermark detection. In *International Conference on Digital Signal Processing*, pages 1–8.
- KWITT, R., MEERWALD, P., UHL, A. (2011). Efficient texture image retrieval using copulas in a Bayesian framework. *IEEE Transactions on Image Processing*, 20(7):2063–2077.
- KWITT, R., UHL, A. (2008). Color wavelet cross co-occurrence matrices for endoscopy image classification. In *3rd International Symposium on Communications, Control and Signal Processing*, pages 715–718.
- LASMAR, N., BERTHOUMIEU, Y. (2010). Multivariate statistical modeling for texture analysis using wavelet transforms. In *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 790–793, Dallas, TX, USA.
- LASMAR, N., BERTHOUMIEU, Y. (2014). Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms. *IEEE Transactions on Image Processing*, 23(5):2246–2261.
- LATECKI, L. J., LAKÄMPER, R. (2000). Shape similarity measure based on correspondence of visual parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1185–1190.
-

-
- LE BORGNE, H., O'CONNOR, N. (2005). Natural scene classification and retrieval using ridgelet-based image signatures. *In Advanced Concepts for Intelligent Vision Systems*, pages 116–122. Springer.
- LE GALL, D. (1991). Mpeg: A video compression standard for multimedia applications. *Communications of the ACM*, 34(4):46–58.
- LI, M. (2011). Distribution preserving quantization. *Ph.D. dissertation, KTH Royal Institute of Technology*.
- LI, M., KLEJSA, J., KLEIJN, W. B. (2010). Distribution preserving quantization with dithering and transformation. *IEEE Signal Processing Letters*, 17(12):1014–1017.
- LIU, M. (1991). Overview of the p× 64 kbit/s video coding standard. *Communications of the ACM*, 34(4):59–63.
- LIU, Y., ZHANG, D., LU, G., MA, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262–282.
- LONG, L. R., ANTANI, S., DESERNO, T. M., THOMA, G. R. (2009). Content-based image retrieval in medicine: Retrospective assessment, state of the art, and future directions. *International journal of healthcare information systems and informatics: official publication of the Information Resources Management Association*, 4(1):1.
- MA, W. Y., MANJUNATH, B. S. (1999). Netra: A toolbox for navigating large image databases. *Multimedia systems*, 7(3):184–198.
- MA, X., HEXIN, C., YAN, Z. (2009). Stereo image coding method using stereo matching with difference-based adaptive searching windows. *In IEEE International Workshop on Imaging Systems and Techniques*, pages 373–376.
- MALLAT, S. (1989). Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12):2091–2110.
- MALLAT, S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press.
- MANDAL, M. K., ABOULNASR, T., PANCHANATHAN, S. (1996). Image indexing using moments and wavelets. *IEEE Transactions on Consumer Electronics*, 42(3):557–565.
-

-
- MANDELBROT, B. B. (1983). *The fractal geometry of nature*, vol. 173. Macmillan.
- MANJUNATH, B. S., SALEMBIER, P., SIKORA, T. (2002). *Introduction to MPEG-7: multi-media content description interface*, vol. 1. John Wiley & Sons.
- MARDIA, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530.
- MARR, D., POGGIO, T. (1976). Cooperative computation of stereo disparity. *Science*, 194(4262):283–287.
- MASSEY, J., FRANK, J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78.
- MATERKA, A., STRZELECKI, M. *et al.* (1998). Texture analysis methods—a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, pages 9–11.
- MEDIONI, G., NEVATIA, R. (1985). Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31(1):2–18.
- MEYER, C. D. (2000). *Matrix analysis and applied linear algebra*. Society for Applied and Industrial Mathematics.
- MEZARIS, V., KOMPATSIARIS, I., BOULGOURIS, N. V., STRINTZIS, M. G. (2004). Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(5):600–621.
- MILED, W., PESQUET, J. C., PARENT, M. (2006a). Disparity map estimation using a total variation bound. *In Canadian Conference on Computer and Robot Vision*, pages 48–55.
- MILED, W., PESQUET, J. C., PARENT, M. (2006b). Wavelet-constrained regularization for disparity map estimation. *In 14th European Signal Processing Conference*, pages 1–5.
- MILED, W., PESQUET, J.-C., PARENT, M. (2009a). A convex optimization approach for depth estimation under illumination variation. *IEEE Transactions on Image Processing*, 18(4):813–830.
-

-
- MILED, W., PESQUET-POPESCU, B., CHÉRIF, W. (2009b). A variational framework for simultaneous motion and disparity estimation in a sequence of stereo images. *In IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 741–744.
- MIN, D., KIM, D., YUN, S., SOHN, K. (2009). 2D/3D freeview video generation for 3DTV system. *Signal Processing: Image Communication*, 24(1):31–48.
- MINKA, T. P., PICARD, R. W. (1996). Interactive learning with a society of models. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 447–452.
- MUKHERJEE, D., WANG, G., WU, Q. M. J. (2010). Stereo matching algorithm based on curvelet decomposition and modified support weights. *In IEEE International Conference on Acoustics Speech and Signal Processing*, pages 758–761.
- MULLER, F. (1993). Distribution shape of two-dimensional DCT coefficients of natural images. *Electronics Letters*, 29(22):1935–1936.
- MÜLLER, H., MICHOUX, N., BANDON, D., GEISSBUHLER, A. (2004). A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International journal of medical informatics*, 73(1):1–23.
- MÜLLER, H., MÜLLER, W., SQUIRE, D. M., MARCHAND-MAILLET, S., PUN, T. (2001). Performance evaluation in content-based image retrieval: overview and proposals. *Pattern Recognition Letters*, 22(5):593–601.
- NADARAJAH, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, 32(7):685–694.
- NAFORNITA, C., BERTHOUMIEU, Y., NAFORNITA, I., ISAR, A. (2012). Kullback-Leibler distance between complex generalized Gaussian distributions. *In European Signal Processing Conference*, pages 1850–1854, Bucharest, Romania.
- NAGEL, H. H., ENKELMANN, W. (1986). An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5):565–593.
- NGO, C. W., PONG, T. C., CHIN, R. T. (2001). Exploiting image indexing techniques in DCT domain. *pattern Recognition*, 34(9):1841–1851.
-

-
- OHTA, Y., KANADE, T. (1985). Stereo by intra-and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2): 139–154.
- OISEL, L., MEMIN, E., MORIN, L., GALPIN, F. (2003). 1D dense disparity estimation for 3D reconstruction. *IEEE Transactions on Image Processing*, 12:1107–1119.
- OJALA, T., MAENPAA, T., PIETIKAINEN, M., VIERTOLA, J., KYLLONEN, J., HUOVINEN, S. (2002a). Outex-new framework for empirical evaluation of texture analysis algorithms. *In International Conference on Pattern Recognition*, vol. 1, pages 701–706.
- OJALA, T., PIETIKÄINEN, M., MÄENPÄÄ, T. (2002b). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987.
- ORTEGA, A., RAMCHANDRAN, K. (1998). Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine*, 15(6):23–50.
- PAPADIMITRIOU, D. V., DENNIS, T. J. (1996). Epipolar line estimation and rectification for stereo image pairs. *IEEE Transactions on Image Processing*, 5(4):672–676.
- PASCAL, F., CHITOUR, Y., OVARLEZ, J.-P., FORSTER, P., LARZABAL, P. (2008). Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis. *IEEE Transactions on Signal Processing*, 56(1):34–48.
- PENTLAND, A. P. (1984). Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):661–674.
- PEYRÉ, G. (2010). Texture synthesis with grouplets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):733–746.
- PO, D. D., DO, M. N. (2006). Directional multiscale modeling of images using the contourlet transform. *IEEE Transactions on Image Processing*, 15(6):1610–1620.
- POLI, D., ZHANG, L., GRUEN, A. (2004). SPOT-5/HRS stereo images orientation and automated DSM generation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35(B1):421–432.
-

-
- PRASAD, R., SARUWATARI, H., SHIKANO, K. (2005). Estimation of shape parameter of GGD function by negentropy matching. *Neural processing letters*, 22(3):377–389.
- PRICE, J. R., RABBANI, M. (1999). Biased reconstruction for JPEG decoding. *IEEE Signal Processing Letters*, 6(12):297–299.
- PROKOP, R. J., REEVES, A. P. (1992). A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP: Graphical Models and Image Processing*, 54(5):438–460.
- QUACK, T., MÖNICH, U., THIELE, L., MANJUNATH, B. S. (2004). Cortina: a system for large-scale, content-based web image retrieval. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 508–511.
- QUEGAN, S., RHODES, I. (1993). Statistical models for polarimetric SAR data. In *IEE Seminar on Texture Analysis in radar and sonar*, pages 1–8, London, UK.
- RABBANI, M., JOSHI, R. (2002). An overview of the JPEG2000 still image compression standard. *Signal Processing: Image Communication*, 17(1):3–48.
- RAMAPRIYAN, H. K. (2002). *Satellite imagery in Earth science applications*. Image Databases: Search and Retrieval of Digital Imagery, Edited by V. Castelli, L. D. Bergman, John Wiley & Sons.
- RAO, K. R., YIP, P. (1990). *Discrete cosine transform: algorithms, advantages, applications*. Academic press.
- RAO, S. R., MOBAHI, H., YANG, A. Y., SASTRY, S. S., MA, Y. (2010). Natural image segmentation with adaptive texture and boundary encoding. In *Computer Vision—ACCV 2009*, pages 135–146. Springer.
- RISSANEN, J., LANGDON, G. G. (1979). Arithmetic coding. *IBM Journal of research and development*, 23(2):149–162.
- ROENKO, A. A., LUKIN, V. V., DJUROVIC, I., SIMEUNOVIC, M. (2014). Estimation of parameters for generalized gaussian distribution. In *International Symposium on Communications, Control and Signal Processing*, pages 376–379.
-

- ROY, S., COX, I. J. (1998). A maximum-flow formulation of the n-camera stereo correspondence problem. *In International Conference on Computer Vision*, pages 492–499.
- RUBERTO, C. D., MORGERA, A. (2008). Moment-based techniques for image retrieval. *In International Workshop on Database and Expert Systems Application*, pages 155–159.
- RUDIN, L. I., OSHER, S., FATEMI, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268.
- RZANGASWAMY, M., WEINER, D., OTURK, A. (1993). Non-Gaussian random vector identification using spherically invariant random processes. *IEEE Transactions on Aerospace and Electronic Systems*, 29(1):111–124.
- SAID, A., PEARLMAN, W. *et al.* (1996). A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(3):243–250.
- SAKJI-NSIBI, S., BENAZZA-BENYAHIA, A. (2008). Indexing of multichannel images in the wavelet transform domain. *In IEEE International Conference on Communication Technologies: Theory & Practice*, pages 1–6, Damascus, Syria.
- SAKJI-NSIBI, S., BENAZZA-BENYAHIA, A. (2009). Copula-based statistical models for multicomponent image retrieval in the wavelet transform domain. *In IEEE International Conference on Image Processing*, pages 253–256, Cairo, Egypt.
- SAMADANI, R., HAN, C. S., KATRAGADDA, L. K. (1993). Content-based event selection from satellite images of the aurora. *In IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, pages 50–59. International Society for Optics and Photonics.
- SCHAEFER, G. (2008). Does compression affect image retrieval performance? *International Journal of Imaging Systems and Technology*, 18(2-3):101–112.
- SCHARSTEIN, D., SZELISKI, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3):7–42.
- SCHMID, C., ZISSERMAN, A. (1998). The geometry and matching of curves in multiple views. *In Computer Vision—ECCV'98*, pages 394–409. Springer.
-

-
- SCHUCHMAN, L. (1964). Dither signals and their effect on quantization noise. *IEEE Transactions on Communication Technology*, 12(4):162–165.
- SERMADEVI, Y., HEMAMI, S. S. (2004). Convexity results for a predictive video coder. *In Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, vol. 2, pages 1713–1717.
- SHAPIRO, J. M. (1993). Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462.
- SHARIFI, K., LEON-GARCIA, A. (1995). Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(1):52–56.
- SHOHAM, Y., GERSHO, A. (1988). Efficient bit allocation for an arbitrary set of quantizers. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(9):1445–1453.
- SIM, D. G., KIM, H. K., OH, D. (2000). Translation, scale, and rotation invariant texture descriptor for texture-based image retrieval. *In International Conference on Image Processing*, vol. 3, pages 742–745.
- SLESAREVA, N., BRUHN, A., WEICKERT, J. (2005). Optic flow goes stereo: A variational method for estimating discontinuity-preserving dense disparity maps. *In Pattern Recognition*, pages 33–40.
- SMEULDERS, A. W. M., WORRING, M., SANTINI, S., GUPTA, A., JAIN, R. (2000). Content-based image retrieval at the end of the early years. *Journal of Visual Communication and Image Representation*, 12:1349–1380.
- SMITH, J. R., CHANG, S. F. (1994). Transform features for texture classification and discrimination in large image databases. *In IEEE International Conference on Image Processing*, vol. 3, pages 407–411, Austin, TX, USA.
- SMITH, J. R., CHANG, S. F. (1997). Querying by color regions using the visualseek content-based visual query system. *Intelligent multimedia information retrieval*, 7(3):23–41.
- SPEIS, A., HEALEY, G. (1996). An analytical and experimental study of the performance of markov random fields applied to textured images using small samples. *IEEE Transactions on Image Processing*, 5(3):447–458.
-

-
- SUN, C. (2002). Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques. *International Journal of Computer Vision*, 47(1-3):99–117.
- SWELDENS, W. (1996). The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and computational harmonic analysis*, 3(2):186–200.
- SZEGÖ, G. (1939). *Orthogonal polynomials*, vol. 23. Amer Mathematical Society.
- TAMURA, H., MORI, S., YAMAWAKI, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8(6):460–473.
- TAUBMAN, D. (2000). High performance scalable image compression with EBCOT. *IEEE transactions on Image Processing*, 9(7):1158–1170.
- TAUBMAN, D., MARCELLIN, M. (2001). *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, Norwell, MA, USA.
- TUCERYAN, M., JAIN, A. K. (1998). Texture analysis. *The handbook of pattern recognition and computer vision*, 2:207–248.
- TZOVARAS, D., KOMPATSIARIS, I., STRINTZIS, M. G. (1999). 3D object articulation and motion estimation in model-based stereoscopic videoconference image sequence analysis and coding. *Signal Processing: Image Communication*, 14(10):817–840.
- Van de WOUWER, G., LIVENS, S., SCHEUNDERS, P., VAN DYCK, D. (1997). Color texture classification by wavelet energy correlation signatures. *In Image Analysis and Processing*, pages 327–334.
- VARANASI, M. K., AAZHANG, B. (1989). Parametric generalized gaussian density estimation. *The Journal of the Acoustical Society of America*, 86(4):1404–1415.
- VEKSLER, O. (2001). Stereo matching by compact windows via minimum ratio cycle. *In IEEE International Conference on Computer Vision*, vol. 1, pages 540–547.
- VEKSLER, O. (2003). Fast variable window for stereo correspondence using integral images. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pages I–556.
-

- VEKSLER, O. (2005). Stereo correspondence by dynamic programming on a tree. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pages 384–390.
- VERDOOLAEGE, G., ROSSEEL, Y., LAMBRECHTS, M., SCHEUNDERS, P. (2009). Wavelet-based colour texture retrieval using the Kullback-Leibler divergence between bivariate generalized Gaussian models. *In IEEE International Conference on Image Processing*, pages 265–268.
- VERDOOLAEGE, G., SCHEUNDERS, P. (2011). Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination. *International Journal of Computer Vision*, 95(3):265–286.
- VERDOOLAEGE, G., SCHEUNDERS, P. (2012). On the geometry of multivariate generalized gaussian models. *Journal of Mathematical Imaging and Vision*, 43(3):180–193.
- VOGEL, J., SCHIELE, B. (2006). Performance evaluation and optimization for content-based image retrieval. *Pattern Recognition*, 39(5):897–909.
- VOULGARIS, G., JIANG, J. (2001). Texture-based image retrieval in wavelets compressed domain. *In International Conference on Image Processing*, vol. 2, pages 125–128.
- WALLACE, G. K. (1991). The JPEG still picture compression standard. *Communications of the ACM*, 34(4):30–44.
- WANG, C.-Y., ZHANG, X., SHAN, R., ZHOU, X. (2015). Grading image retrieval based on DCT and DWT compressed domains using low-level features. *Journal of Communications*, 10(1):64–73.
- WANG, D., SRIHARI, S. N. (1989). Classification of newspaper image blocks using texture analysis. *Computer Vision, Graphics, and Image Processing*, 47(3):327–352.
- WANG, L., YANG, R. (2011). Global stereo matching leveraged by sparse ground control points. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3033–3040.
- WANG, S., QI, C., CHENG, Y. (2012). Application of tamura texture feature to classify underwater targets. *Applied Acoustics*, 2:135–139.
-

-
- WANNER, S., GOLDLUECKE, B. (2014). Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):606–619.
- WENG, J., AHUJA, N., HUANG, T. S. (1992). Matching two perspective views. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (8):806–825.
- WOO, W., ORTEGA, A. (1997). Stereo image compression based on disparity field segmentation. In *SPIE Conference on Visual Communications and Image Processing*, vol. 3024, pages 391–402.
- WOO, W., ORTEGA, A. (2000). Overlapped block disparity compensation with adaptive windows for stereo image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(2):194–200.
- WOUWER, G. V., SCHEUNDERS, P., DYCK, D. V. (1999). Statistical texture characterization from discrete wavelet representations. *IEEE Transactions on Image Processing*, 8(4):592–598.
- YAHAV, G., IDAN, G. J., MANDELBOUM, D. (2007). 3D imaging camera for gaming application. In *International Conference on Consumer Electronics*, pages 1–2.
- YAMADA, I., OGURA, N., SAKANIWA, K. (1998). Quadratic optimization of fixed points of nonexpansive mappings in hubert space. *Numerical Functional Analysis and Optimization*, 19(1-2):165–190.
- YANG, S., HUANG, G., ZHAO, Z., WANG, N. (2011). Extraction of topographic map elements with SAR stereoscopic measurement. In *International Symposium on Image and Data Fusion*, pages 1–4.
- YONG, R., THOMAS, S. H. (1999). Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10:39–62.
- YOON, K., KIM, Y., PARK, J. H., DELGADO, J., YAMADA, A., DUFAUX, F., TOUS, R. (2012). JPSearch: New international standard providing interoperable framework for image search and sharing. *Signal Processing: Image Communication*, 27(7):709–721.
-

- YUAN, H., ZHANG, X. P. (2004). Texture image retrieval based on a Gaussian mixture model and similarity measure using a Kullback divergence. *In IEEE International Conference on Multimedia and Expo*, pages 1867–1870, Taipei, Taiwan.
- YUAN, H., ZHANG, X. P., GUAN, L. (2003). Content-based image retrieval using a Gaussian mixture model in the wavelet domain. *In SPIE, Visual Communications and Image Processing*, Lugano, Switzerland.
- ZABIH, R., WOODFILL, J. (1994). Non-parametric local transforms for computing visual correspondence. *In Computer Vision ECCV'94*, pages 151–158. Springer.
- ZHANG, D., LU, G. *et al.* (2002). A comparative study of fourier descriptors for shape representation and retrieval. *In Proc. 5th Asian Conference on Computer Vision*. Citeseer.
- ZHANG, Z. (1998). Determining the epipolar geometry and its uncertainty: A review. *International journal of computer vision*, 27(2):161–195.
- ZHANG, Z., XU, G. (1997). A general expression of the fundamental matrix for both perspective and affine cameras. *In Proceedings of the Fifteenth international joint conference on Artificial intelligence-Volume 2*, pages 1502–1507.
- ZITNICK, C. L., KANADE, T. (2000). A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):675–684.
- ZITOVA, B., FLUSSER, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000.
- ZOGRAFOS, K. (2008). On Mardia’s and Song’s measures of kurtosis in elliptical distributions. *Journal of Multivariate Analysis*, 99(5):858–879.
-