University of Crete
Department of Computer Science

FO.R.T.H.
Institute of Computer Science

Content-based Image Retrieval via Alpha-Stable Modeling of Texture Information

(MSc. Thesis)

Georgios Tzagkarakis

Heraklion

November 2004

Department of Computer Science
University of Crete

Content-based Image Retrieval via Alpha-Stable Modeling of Texture Information

Submitted to the
Department of Computer Science
in partial fulfillment of the requirements for the degree of
Master of Science

November 16, 2004

© 2004 University of Crete & ICS-FO.R.T.H. All rights reserved.

Author: _____
Georgios Tzagkarakis
Department of Computer Science

Committee

Supervisor _____
Panagiotis Tsakalides
Associate Professor

Member _____
Georgios Tziritas
Professor

Member _____
Apostolos Traganitis
Professor

Accepted by:

Chairman of the
Graduate Studies Committee _____
Dimitris Plexousakis
Associate Professor

Heraklion, November 2004

Abstract

During the last decades, digital information is being gathered and stored at an explosive rate on large digital databases. This fact gives rise to the important issue of effectively and precisely searching and interacting with these collections. The purpose of an efficient content-based information retrieval system is to bring back all the relevant items from a specific database, given a user query. The term "*efficient*", refers to the ability to capture as much as possible significant information of an object of interest, using a description that is much smaller in size than the original object.

A great portion of these databases contains digital images, usually grouped in categories according to the specific application field. Recently, it has become evident from physiological studies of the human visual cortex that commonly used deterministic retrieval techniques, based on the energies of filter responses, are often inadequate in describing the image content. However, there is a strong motivation to search for more powerful *statistical* retrieval methods that can capture the intrinsic structure of digital images, represented by low level features such as color, shape and texture. The Gaussian distribution became a natural modeling tool because of its simplicity and the well studied properties. But recent studies, verified by experimental results, reveal that in most cases these statistics are highly *non-Gaussian* with heavier tails than the Gaussian, thus the previously developed retrieval methods based on a Gaussian assumption are unable to capture the true statistical behavior of the image features.

This thesis introduces the family of *Alpha-Stable* distributions as the modeling tool for designing a content-based image retrieval system exploiting the *texture* information. Texture plays an important role for the characterization of regions from digital images, since it carries information about their micro-structure and the distribution of the grey levels within these regions. It has been also shown that the performance of a retrieval system is improved by implementing a suitable pre-processing step, such as an image transformation, resulting in a simpler and more efficient statistical model. We build a hierarchical retrieval system, in terms of an increased retrieval performance, by adding possible inter-dependencies between the transform coefficients.

In particular, the first of the proposed methods constructs multi-resolution and multi-orientation image expansions using the standard 2-dimensional wavelet transform. Then, we represent the texture content of a given image by estimating the parameters of the marginal distribution at each orientation and scale, fitting the transform coefficients using a univariate Alpha-Stable distribution. The second method enhances the first one by extracting possible inter-dependencies between the transform coefficients at different orientations, which are tied up and jointly modeled as samples of a sub-Gaussian random process. Finally, the third method extracts inter-dependencies not only across orientations, but also across resolution scales. In order to achieve rotation invariance the wavelet transform is replaced by a steerable pyramid and the previous multivariate sub-Gaussian model is upgraded into a *steerable* model, appropriate for directly modeling a rotated image using the corresponding model of its original version. Since the family of Alpha-Stable distributions lacks finite variance, we apply a Gaussianization procedure which takes into account

the true non-Gaussian behavior as expressed by the so called *Fractional Lower-Order moments*, resulting in a set of almost jointly Gaussian pyramid coefficients. This makes the design of a rotation-invariant retrieval system simple and accurate, exploiting the heavy-tailed behavior of the marginal statistics of the original pyramid coefficients.

The efficiency of a retrieval system does not only depend on the set of features representing the texture information, but also on a suitable similarity function that measures how close to a given query each image in the database is. For each one of the proposed methods we construct appropriate similarity functions, exploiting the accuracy of the Alpha-Stable model in capturing the marginal and joint non-Gaussian behavior of the transform coefficients. We illustrate the improved performance of our models compared with the performance of recently developed statistical retrieval schemes, by applying them on real texture samples.

Περίληψη

Τις τελευταίες δεκαετίες, η ψηφιοποιημένη πλέον πληροφορία συγκεντρώνεται και αποθηκεύεται με καταιγιστικούς ρυθμούς σε μεγάλες ψηφιακές βάσεις δεδομένων. Αυτό το γεγονός οδηγεί στην ανάγκη εύρεσης μεθόδων για αποδοτική και ακριβή αναζήτηση και αλληλεπίδραση με αυτές τις συλλογές. Ο σκοπός ενός αποδοτικού συστήματος ανάκτησης πληροφορίας με βάση το περιεχόμενο, είναι η επιστροφή όλων των σχετικών αντικειμένων από μια συγκεκριμένη βάση, δεδομένης της εισόδου που δίνεται από το χρήστη. Ο όρος αποδοτικό αναφέρεται στην ικανότητα του συστήματος να συλλάβει όσο το δυνατόν περισσότερη σημαντική πληροφορία του αντικειμένου που ενδιαφέρει το χρήστη, χρησιμοποιώντας μια περιγραφή με πολύ μικρότερο μέγεθος από αυτό του αρχικού αντικειμένου.

Ένα σημαντικό ποσοστό αυτών των βάσεων περιέχει ψηφιακές εικόνες οι οποίες συνήθως ομαδοποιούνται σε κατηγορίες ανάλογα με το πεδίο εφαρμογής. Πρόσφατα, μετά από φυσιολογικές μελέτες του ανθρώπινου συστήματος όρασης, έχει γίνει εμφανές ότι οι έως τώρα χρησιμοποιούμενες ντετερμινιστικές τεχνικές ανάκτησης, βασισμένες στις ενέργειες των αποκρίσεων κατάλληλων φίλτρων, είναι συχνά ανεπαρκείς για την περιγραφή του περιεχομένου μιας εικόνας. Εντούτοις, υπάρχουν ισχυρά κίνητρα για την αναζήτηση πιο αποτελεσματικών στατιστικών μεθόδων ανάκτησης, οι οποίες μπορούν να συλλάβουν την εγγενή δομή των ψηφιακών εικόνων, η οποία αναπαρίσταται από χαρακτηριστικά χαμηλού επιπέδου, όπως το χρώμα, το σχήμα και η υφή. Η Γκαουσσιανή κατανομή έγινε το φυσικό εργαλείο μοντελοποίησης εξαιτίας της απλότητάς της και των καλά μελετημένων ιδιοτήτων της. Αλλά πρόσφατες μελέτες, επιβεβαιωμένες από πειραματικά αποτελέσματα, αποκάλυψαν ότι στις περισσότερες περιπτώσεις οι στατιστικές των παραπάνω χαρακτηριστικών είναι εξαιρετικά μη-Γκαουσσιανές με βαρύτερες ουρές από αυτές της Γκαουσσιανής, επομένως οι προηγούμενες μέθοδοι που στηρίζονταν στην Γκαουσσιανή υπόθεση δεν είναι ικανές να περιγράψουν την πραγματική στατιστική συμπεριφορά των χαρακτηριστικών μιας εικόνας.

Η παρούσα εργασία εισάγει την οικογένεια των Άλφα-Ευσταθών κατανομών ως το εργαλείο μοντελοποίησης για τη σχεδίαση ενός συστήματος ανάκτησης εικόνων με βάση το περιεχόμενο, εκμεταλλευόμενο την πληροφορία που σχετίζεται με την υφή. Η υφή παίζει σημαντικό ρόλο για το χαρακτηρισμό των περιοχών μιας ψηφιακής εικόνας, καθώς μεταφέρει πληροφορία για τη μικροδομή τους και την κατανομή των επιπέδων του γκρι σε αυτές τις περιοχές. Έχει δειχθεί επίσης ότι η επίδοση ενός συστήματος ανάκτησης βελτιώνεται με την εφαρμογή ενός κατάλληλου βήματος προεπεξεργασίας, όπως ένας μετασχηματισμός της εικόνας, καταλήγοντας σε ένα απλούστερο και πιο αποδοτικό στατιστικό μοντέλο. Κατασκευάζουμε ένα ιεραρχικό σύστημα ανάκτησης, αναφορικά με μια αυξημένη απόδοση ανάκτησης, λαμβάνοντας υπόψην πιθανές αλληλεξαρτήσεις μεταξύ των συντελεστών του μετασχηματισμού.

Ειδικότερα, η πρώτη από τις προτεινόμενες μεθόδους μετασχηματίζει την εικόνα σε πολλαπλές κλίμακες και κατευθύνσεις χρησιμοποιώντας τον διδιάστατο Κυματιδιακό μετασχηματισμό (Wavelet transform). Κατόπιν, αναπαριστούμε την πληροφορία της υφής για μια δοσμένη εικόνα εκτιμώντας τις παραμέτρους των περιθωριακών κατανομών σε κάθε κατεύθυνση και κλίμακα, προσεγγίζοντάς τις χρησιμοποιώντας μέλη από την οικογένεια των μονοδιάστατων Άλφα-Ευσταθών κατανομών. Η δεύτερη μέθοδος βελτιώνει την πρώτη λαμβάνοντας υπόψην πιθανές αλληλεξαρτήσεις μεταξύ των συν-

τελεστών σε διαφορετικές κατευθύνσεις, οι οποίοι ομαδοποιούνται και μοντελοποιούνται από κοινού σαν δείγματα μιας υπο-Γκαουσσιανής στοχαστικής διαδικασίας. Τέλος, η τρίτη μέθοδος εξάγει αλληλεξαρτήσεις όχι μόνο μεταξύ των διαφόρων κατευθύνσεων αλλά και μεταξύ των κλιμάκων. Για να επιτύχουμε ένα σύστημα αναλλοίωτο στην περιστροφή ο Κυματιδιακός μετασχηματισμός αντικαθίσταται από μια καθοδηγήσιμη πυραμίδα (steerable pyramid) και το προηγούμενο πολυδιάστατο υπο-Γκαουσσιανό μοντέλο αναβαθμίζεται σε ένα καθοδηγήσιμο μοντέλο κατάλληλο για την άμεση μοντελοποίηση μιας περιστραμμένης εικόνας χρησιμοποιώντας το αντίστοιχο μοντέλο της αρχικής (μη-περιστραμμένης) εικόνας. Καθώς η οικογένεια των Άλφα-Ευσταθών κατανομών στερείται πεπερασμένης διακύμανσης, εφαρμόζεται μια διαδικασία κανονικοποίησης η οποία λαμβάνει υπόψην την πραγματική μη-Γκαουσσιανή συμπεριφορά, όπως εκφράζεται από τις λεγόμενες Κλασματικές Χαμηλότερης τάξης Ροπές, καταλήγοντας σε μια πυραμίδα με σχεδόν από κοινού Γκαουσσιανούς συντελεστές. Αυτό κάνει τη σχεδίαση ενός συστήματος ανάκτησης, αναλλοίωτου στην περιστροφή, απλή και ακριβή λαμβάνοντας παράλληλα υπόψην τις, με βαριές ουρές, περιθωριακές στατιστικές των αρχικών συντελεστών της πυραμίδας.

Η αποδοτικότητα ενός συστήματος ανάκτησης δεν εξαρτάται μόνο από το σύνολο των χαρακτηριστικών που αναπαριστούν την πληροφορία υφής, αλλά επίσης από μια κατάλληλη συνάρτηση ομοιότητας η οποία μετράει πόσο κοντά σε μια δοσμένη εικόνα βρίσκεται κάθε εικόνα της βάσης. Για κάθε μια από τις προτεινόμενες μεθόδους κατασκευάζουμε κατάλληλες συναρτήσεις ομοιότητας οι οποίες εκμεταλλεύονται την ακρίβεια του Άλφα-Ευσταθούς μοντέλου στη σύλληψη της περιθωριακής και από κοινού μη-Γκαουσσιανής συμπεριφοράς των συντελεστών του μετασχηματισμού. Τέλος, επιδεικνύουμε τη βελτιωμένη απόδοση των προτεινόμενων μοντέλων συγκρίνοντας με την απόδοση πρόσφατα ανεπτυγμένων στατιστικών τεχνικών ανάκτησης, εφαρμόζοντάς τις σε πραγματικές εικόνες υφής.

Acknowledgements

This thesis is the result of a two-year working experience as graduate student at the research group for Telecommunications & Networks at the Institute of Computer Science (ICS) of the Foundation for Research and Technology Hellas (FO.R.T.H.). These years have been very instructive, giving rise to several interesting tasks more or less related to the research.

First of all I would like to thank Panagiotis Tsakalides, my supervisor, not only for his guidance and encouragement during this work, but also for trying to bridge the gap between the philosophy of a mathematician and that of an engineer. I think that my mathematical nature is still more potent!

I want also to thank my professors at the department of Mathematics of the University of Crete, for the meaningful discussions and their advice. I am also grateful to the felicitous suggestions and feedback from Baltasar Beferull-Lozano (EPFL), who somehow became my second supervisor, and to the help of Alin Achim for helping me at the beginning of my computer implementations.

The pleasant and creative research atmosphere at the group is of course also an outcome of all the other members of the group. Of these I want to give special thanks to C. Adamakis, V. Garganourakis, S. Papadakis (he also contributed in formatting part of the LaTeX document), G. Stamatakis and D. Triantafyllidou (in alphabetical order to avoid misunderstanding) for taking care of my transport from the institution to home saving my time, to V. Angelakis for sharing with me some of his profound knowledge on Photoshop, to A. Katranidou for being a congenial neighbour and bringing cookies and to a student outside the group, G. Pagonis, for the walks with deep discussions on just everything that is interesting in this world.

This work was supported by a graduate fellowship from the ICS (FO.R.T.H.), which I also truly acknowledge for providing the necessary technical equipment.

Finally, I would like to thank my family, for providing me with the ideal environment to grow up in, and exhorting me in changing for the better. Their constant support and encouragement have really brought me here.

I have had a lot of fun. Thank you all (I hope that I've not forgotten someone).

Contents

Abstract **v**

List of tables **xv**

List of figures **xvii**

1 Introduction **1**

1.1 Annotation-based Image Retrieval 1

1.2 Content-based Image Retrieval 2

1.3 Motivation 3

1.4 Structure of the thesis 5

I Background - Mathematical Foundations **7**

2 Background **9**

2.1 Introduction 9

2.2 Architecture of a CBIR system 11

2.3 Previous work on FE and SM 12

2.4 Probabilistic Image Retrieval 15

3 Wavelet Analysis **19**

3.1 Fourier Transform 19

3.1.1 Discrete Fourier Transform (DFT) 21

3.1.2 Short-Time Fourier Transform (STFT) 22

3.2 Wavelet Transform 24

3.2.1 The 1-Dimensional Wavelet Transform 24

3.2.2 The 2-Dimensional Discrete Wavelet Transform 29

3.2.3 Steerable Pyramids 31

3.2.4 Examples of wavelets 32

3.3 Statistical properties of the DWT coefficients 35

4 Alpha-Stable Models **37**
4.1 Introduction . . . 37
4.2 The family of Stable Distributions . . . 38
4.2.1 Univariate $S\alpha S$ Distributions . . . 38
4.2.2 Multivariate Stable Distributions . . . 41
4.2.3 Symmetric Stable Processes . . . 43
4.2.4 Covariations . . . 44

II Statistical Content-Based Image Retrieval **53**

5 CBIR via Univariate $S\alpha S$ distributions **55**
5.1 Introduction . . . 55
5.2 Statistical Modeling of Wavelet Coefficients . . . 56
5.3 $S\alpha S$ Modeling of the Wavelet Decompositions . . . 60
5.3.1 Feature Extraction . . . 60
5.4 Similarity Measurement between $S\alpha S$ distributions . . . 64
5.4.1 Relation to the KLD between Gaussian densities . . . 65
5.5 Experiments with real-world Texture Images . . . 66
5.5.1 Range of characteristic exponent α in $S\alpha S$ models . . . 68
5.5.2 Retrieval Performance . . . 68
5.6 Conclusions . . . 71

6 CBIR via Multivariate sub-Gaussian Distributions **75**
6.1 Introduction . . . 75
6.2 Joint sub-Gaussian Modeling of Wavelet Coefficients . . . 76
6.2.1 Feature Extraction . . . 77
6.2.2 Similarity Measurement . . . 81
6.3 Experimental Results . . . 83
6.3.1 Retrieval Performance . . . 83
6.4 Conclusions . . . 85

7 Rotation-Invariant CBIR via Joint sub-Gaussian Models **89**
7.1 Introduction . . . 89
7.2 Steerable Multivariate sub-Gaussian Model . . . 91
7.2.1 Feature Extraction . . . 94
7.2.2 Similarity Measurement . . . 95
7.3 Rotation-Invariant CBIR with Gaussianized Steerable Pyramids . . . 97
7.3.1 Variance-adaptive local modeling using multivariate sub-Gaussian distributions . . . 98
7.3.2 Gaussianization of the multivariate sub-Gaussian model . . . 99

7.3.3 Neighborhood Construction 100
7.3.4 Feature Extraction 102
7.3.5 Similarity Measurement 103
7.4 Experimental Results 106
7.4.1 Series I: Frobenius norm between steerable multivariate sub-Gaussian models 108
7.4.2 Series II: Rotation-invariant KLD between Gaussianized steerable pyramids 111
7.5 Conclusions 113

8 Conclusions and Future work **115**

III Appendices **117**

A Kullback-Leibler Divergence between Normalized $S\alpha S$ Characteristic Functions **119**

B Steerable multivariate sub-Gaussian model **123**
B.1 Rotation-Invariant Kullback-Leibler Divergence 124
B.1.1 Relation of equivalence between the covariance matrices corresponding to images I and I_θ 124
B.1.2 KLD $D(I\|Q_\theta)$ as a function of the KLD $D(I\|Q)$ 125
B.1.3 Rotation-Invariant KLD 127

List of Tables

4.1 Performance of the covariation estimators. 52

5.1 $S\alpha S$ modeling of wavelet subband coefficients of texture images from the VisTex database, using Daubechies' 4 filter and 3 decomposition levels. ML parameter estimates and 95% confidence intervals for the characteristic exponent α. 63
5.2 Index of each Texture Class. 68
5.3 Average retrieval rates (%) for the selected set of images in the top 15 matches, using different Daubechies' filters and for 3 decomposition levels. 70
5.4 Average retrieval rates (%) for the selected subset of images in the top 15 matches, using different Daubechies' filters and for 3 decomposition levels. 73

6.1 Optimal p parameter as a function of the characteristic exponent α. 82
6.2 Average retrieval rates (%) for the selected set of images in the top 15 matches, using different Daubechies' filters and for 3 decomposition levels. 84

7.1 Neighborhood shapes which were used in the measurement of the Gaussianization performance. 101
7.2 Index of each Texture Class. 107
7.3 Average retrieval rate (%) in the top 16 matches, for the set of 260 textures. . . . 109
7.4 Average retrieval rate (%) in the top 16 matches, for the subset of 200 textures. . . 110
7.5 Average retrieval rate (%) in the top 16 matches. 111
7.6 Average retrieval rate (%) in the top 16 matches. 113

List of Figures

1.1 An example of an image where it is almost impossible to describe all the contents. 2

2.1 An illustration of annotating images containing different objects. 10
2.2 The architecture of a typical CBIR system. 11
2.3 Example of color histograms. 13
2.4 Texture samples obtained from the Brodatz database. 14

3.1 The filtering process of the 1-D DWT. 29
3.2 A 3-Level 2-D DWT. 31
3.3 3-level wavelet decomposition of the image *Flowers.0006* using 'db4' filter. 31
3.4 Spectral decomposition performed by a steerable pyramid. 32
3.5 The filtering process implemented by a steerable pyramid. 33
3.6 Illustration of the structure of the steerable pyramid. (a) Test image, (b) Filter kernels used to construct the pyramid (These four filters which are rotated versions of a common filter, span the space of all rotations of that common filter), (c) Lowpass coefficient image, (d)-(f) Bandpass coefficient images. 34
3.7 Decomposition High-pass filters of the Daubechies' family. 34
3.8 2-D scaling and wavelet functions corresponding to the 1-D 'db4' wavelet. 35

4.1 Standard $S\alpha S$ density functions for different values of the characteristic exponent α, for the parameterization (4.2). 38
4.2 Tails of the standard $S\alpha S$ density functions for different values of the characteristic exponent α, for the parameterization (4.2). 40
4.3 Simulated $S\alpha S$ sequences for different values of the characteristic exponent. 41
4.4 Curves representing the standard deviation of the covariation estimation as a function of the parameter p for (a) the $\hat{\mathbf{c}}$ and (b) the $\hat{\mathbf{c}}^{FLOM}$ estimator. 50
4.5 Running sample variances for $\alpha = 1.4$ and different values of p, corresponding to (a) the $\hat{\mathbf{c}}$ and (b) the $\hat{\mathbf{c}}^{FLOM}$ estimator. 51

5.1 KLD between Gaussian PDFs with (a) the same mean and (b) the same standard deviation. 59

5.2 Normal probability plot of the horizontal subband at the first level of decomposition of the image *Flowers.0006*. 61
5.3 Modeling of the horizontal subband at the first level of decomposition of the *Flowers.0006* image with the $S\alpha S$ and the GGD. 62
5.4 Kullback-Leibler divergence between two normalized characteristic functions following parameterization (4.2), with (a) the same characteristic exponent and (b) the same dispersion. 66
5.5 Texture images from the VisTex and Brodatz databases that are used in the experiments. 67
5.6 Histogram of the estimated values for the characteristic exponent, α, for the set of 480 texture images of size 128×128. 69
5.7 Histogram of the estimated values for the characteristic exponent, α, for the selected subset of 320 texture images of size 128×128. 71
5.8 Retrieval performance according to the number of top matches considered. 72
5.9 Average percentages (%) of correct retrieval rate for each individual texture class using a 3-level wavelet transform with 'db4' filters. 74

6.1 Joint histograms for the Brodatz texture *1.2.04*, between the horizontal subband at the first level and: (a) the vertical and (b) diagonal subbands at the same level. Intensity corresponds to frequency of occurence. 77
6.2 Sub-Gaussian vector model for the wavelet subbands at the same decomposition level. 78
6.3 Curves representing the standard deviation of the covariation estimation as a function of the parameter p for $\alpha = 1.2$, 1.5, 1.8 and 25 dispersion pairs (γ_X, γ_Y), using the $\hat{\mathbf{c}}^{FLOM}$ estimator. 81
6.4 KLD between zero-mean multivariate Gaussian PDFs with covariance matrices of the form $\mathbf{R} = \sigma^2 I$. 83
6.5 Average percentages (%) of correct retrieval rate for each individual texture class using a 3-level wavelet transform with 'db4' filters. 86
6.6 Retrieval performance according to the number of top matches considered, using Daubechies' "db4" filter. 87

7.1 Relation between the covariation and underlying covariance matrices of signature $\mathcal{S}$, with the corresponding matrices of the transformed signature $\mathcal{S}_\theta$. 94
7.2 Quad-tree structure of the subband coefficients. 101
7.3 Histogram of the neighborhood indices in terms of resulting in the best Gaussianization performance for a set of 300 subbands. 102
7.4 (a) *Bark* physically rotated at 30 and 120 degrees, (b) $D(I,Q)(\theta)$ for $J = 4$. Notice that the minimum is achieved for $\theta^* = 90$ degrees, which is the exact relative angle between the two texture samples. 106

7.5 $D(I\|Q)(\theta)$ for $J = 4$. Notice that the minimum is achieved for $\theta^* = 90$ degrees, which is the exact relative angle between the two texture samples. 106
7.6 Texture images from the VisTex database, from left to right and top to bottom: Bark, Brick, Bubbles, Grass, Leather, Pigskin, Raffia, Sand, Straw, Water, Weave, Wood, Wool. 107
7.7 Histogram of the estimated values for the characteristic exponent, α, for the set of 260 texture images of size 256×256. 108
7.8 Average percentages (%) of correct retrieval rate for each individual texture class. 110
7.9 Retrieval performance according to the number of top matches considered. 110
7.10 Retrieval performance according to the number of top matches considered. 112
7.11 Retrieval performance according to the number of top matches considered. 113

Chapter 1

Introduction

During the last decades, information is being gathered and stored at an explosive rate on large digital databases. Examples include multimedia databases containing audio, images and video. This fact gives rise to a very important issue, that is, how to effectively and precisely search and interact with these, often unstructured, collections where exhaustive inspection is impossible (e.g. World Wide Web).

Focusing on a *digital image* database, we would desire to retrieve all the relevant images, similar to a user-provided query image. Traditional database technology results in the design of search engines oriented to text-based data structures. With the steady growth of computer power and the rapidly decaying cost of storage media, retrieving images from an unannotated image database, based on their visual content, is a challenging problem gaining the interest of the research community.

Automatically extracted features associated with the content of an image, representing semantically - meaningful information, is the goal of the current research on digital image databases. This is usually referred to as the *Semantic Gap*, that is, while the extracted visual features are low-level (e.g. color, shape), the concepts that are searched for, are often high-level (e.g. a sunset scene or a happy face).

1.1 Annotation-based Image Retrieval

While it is theoretically possible to annotate all the database images with text meta-data and rely on relational database architectures to retrieve the desired information, there are practical applications in which this solution is impossible. A main reason for this, is that "*an image equals thousands of words*" which means that the large amount of information contained in an image may be difficult or even impossible to be annotated (Figure 1.1). Besides, this process requires much human labeling work and there is still the problem of the various interpretations that can be given to an image. Some of these interpretations can even be unknown at annotation time, making it very difficult to predict the interpretation that a given user will have in mind during the retrieval procedure.

Figure 1.1: An example of an image where it is almost impossible to describe all the contents.

A major drawback of an annotation-based approach for image retrieval, is that the annotation is subject-related and biased, depending on the domain of application. This approach was also advocated primarily by the database research community. A representative system is the Cypress (Berkeley) [41], which employs a traditional file system to store the images and a relational database containing the meta-data that represent the content of each image. An advantage of such a system is the simplicity of design and use, as well as the scalability. A disadvantage is the lack of support for unanticipated requests.

1.2 Content-based Image Retrieval

Content-based image retrieval (CBIR) is the set of methods for retrieving relevant images from an image database using automatically-extracted image features. Features describing the content of an image, such as color distribution, shape and texture, are computed for both the database images and the query. Then, the similarity between two images is evaluated by comparing the corresponding features.

CBIR systems using low-level features need no special effort in capturing them, while systems using high-level features, such as object composition, have great potential to achieve advanced image retrieval, but the derivation of semantically-meaningful features is a big challenge. The extracted features are usually domain-specific and the development of these systems was primarily driven by the image processing community. A representative CBIR system is the PhotoBook (MIT) [90], developed for identifying and retrieving similar faces, which classifies images by shape and texture.

An integrated CBIR system can only result by the cooperation of scientists from the areas of database management, information retrieval, image processing, computer graphics and cognitive sciences. CBIR is critical in developing search engines in many application areas, including the following:

1. Multimedia delivery: for instance in mobile Internet applications the goal is to serve all

clients with various profiles, by delivering all image formats and resolutions. CBIR is also important for video indexing and retrieval, where a video sequence must be segmented into appropriate units (e.g. clips, key frames).

2. Medical diagnostics: CBIR is critical in the emerging fields of medical imaging and picture archiving and communication systems (PACS), for clinical diagnosis and decision making. A typical system belonging in this class is IRMA [31].

3. Space Image Retrieval: a characteristic example revealing the importance of a CBIR system in this field, is the fact that NASA's Earth Observing System (EOS) sends every day about 1 Terabyte of visual information.

4. Face/Fingerprint identification: such a system may be crucial for crime prevention and queen's peace.

5. Libraries, galleries and museums: storage of digitized paper collections in large databases in order to archive them as document image collections, emerge the need for designing CBIR systems in the form of Optical Character Recognition (OCR) and Document Image Analysis and Recognition (DIAR) systems. A typical example is the Alexandria Digital Library (ADL) project.

6. Remote sensing and geographic information systems (GIS): Webview is an example of such a CBIR system which ranks the increasing number of distributed geographical image databases based on their visual content (statistical meta-data). Based on a particular visual query, the system ranks the relevant image databases. The query is then guided to the respective databases and the relevant images in those databases are returned to users.

The above examples justify the great interest of the research community in solving the complex and challenging problem of designing, if possible, an integrated CBIR system with the ability to function not only in a specific, but in more of the above exciting scientific fields.

1.3 Motivation

Despite the different characteristics of the existing CBIR systems, there exists a common methodology of responding to a given query. In particular, in every CBIR system we can distinguish two major tasks, namely *Feature Extraction* (FE) and *Similarity Measurement* (SM). During the FE step, the system generates a set of features constituting the *signature*, to accurately represent the information content of a given query. This signature has to be much smaller in size than the original query, while capturing as much as possible of its information content. In the SM step, we construct an appropriate distance function which measures how close to the given query each image in the database is, by comparing their signatures.

Another common task in the majority of the CBIR systems is a pre-processing step which is applied before the FE step. This pre-processing is usually a transformation of each image in a

domain such as DFT, where the extracted features will be better correlated with image semantics. Depending on the application field, the features may adapt from system to system. For instance, many CBIR systems exploit the information about color, shape and texture or combinations of them.

Texture plays an important role for the characterization of regions from digital images. It carries information about the micro-structure of the regions and the distribution of the grey levels within such regions. A scheme for the retrieval based on the texture information should be capable of encoding the properties of the texture using a number of descriptors. In most of the previously developed CBIR systems, these descriptors were based on the assumption that the energy distribution in the frequency domain identifies a texture. So, they employed filtering approaches for texture extraction and specifically energy measurements at the output of filter banks as the extracted features. Then, the SM step consisted in employing a deterministic distance function, such as the Euclidean norm, between the vectors containing these energies.

In our work, we develop a CBIR system based on the texture information, however the texture descriptors are now represented by sets of statistical features, defining by this way the vectors to be used in the retrieval process. This statistical framework is based on the results of recent physiological studies on human texture perception, which state that "two homogeneous textures are difficult to discriminate if they produce similar marginal distribution of responses from a bank of filters". Thus, we apply an image transformation which makes statistical modeling easier. In this study we employ wavelet-based transforms, which result in a natural partition of the image spectrum into multiscale and oriented subbands. Then, each texture is modeled by the marginal densities of wavelet subband coefficients.

The important issue is the selection of the family of distributions which model the marginal densities of the wavelet subbands. In the past, the Gaussian distribution was the first candidate for modeling the transform coefficients. However, experimental results revealed that the Gaussian family is not suitable for providing an accurate modeling of the marginal distributions, as the coefficients seems to be drawn from densities with heavier, than the Gaussian, tails. As a consequence, many researchers introduced families such as the Cauchy, the Laplace and the generalized Gaussian, which resulted in better retrieval performance.

Although the above models improved the retrieval efficiency, in many cases they are not capable of capturing the significant texture information. This is due to the inherent property of sparseness characterizing the wavelet transforms. In particular, most of the subband coefficients have small value and only few of them are large. Thus, their distribution presents an impulsive nature with heavier tails, that is, with a greater probability of observing large coefficients. In the following study we show that these distributions are better approximated by members of the *Alpha-Stable* family that includes distributions with algebraic tails.

Regarding the SM step, we show that it is also of great importance for a CBIR system to have an improved performance. In the proposed statistical framework, the traditional approaches fail to exploit the extracted texture-specific information. For this reason, we introduce measures of

similarity, such as the Kullback-Leibler Divergence, between statistical texture descriptors.

Despite the fact that the model based on marginal Alpha-Stable distributions results in a better representation, it exploits only the marginal features of each individual subband. We enhance the capacity of this model by incorporating information between different subbands. This is achieved by introducing the family of multivariate sub-Gaussian distributions, which is characterized by the strong dependence between the different components and the heavy-tailed marginal behavior. Actually, the dependence across subbands is also an intrinsic property of the wavelet transform, thus the proposed joint sub-Gaussian modeling is expected to result in a further improvement regarding the retrieval performance.

An important property of a CBIR system is rotation invariance, that is, it must be able to retrieve the most relevant images even if the query image is rotated. We study this issue in the last part of this work, where we employ a locally adaptive joint sub-Gaussian model, combined with a Gaussianization procedure on the coefficients of a steerable pyramid. We replaced the standard wavelet transform with a steerable pyramid in order to achieve rotation and translation invariance. The Gaussianization procedure is justified by the fact that it results in a normalized transformation domain, where the modeling is simplified and we can also employ the second-order moments theory. Besides, we achieve rotation invariance by constructing a rotation-invariant version of the Kullback-Leibler Divergence.

1.4 Structure of the thesis

The present thesis deals with the problem of content-based image retrieval by exploiting the texture-specific information, which can be modeled via members of families including distributions with heavy (algebraic) tails. The thesis is organized as follows: In Chapter 2 we describe the architecture of a typical CBIR system along with the most commonly used retrieval methods. We also introduce the statistical framework in which we develop our proposed CBIR technique. In Chapter 3 we set out the basic theory concerning the wavelet analysis, as well as the theory of steerable pyramid transformations. In Chapter 4 we introduce the basic theory regarding the family of Alpha-Stable models.

In Chapter 5 we describe our proposed approach for CBIR using univariate Symmetric-Alpha-Stable distributions. In particular, we model the marginal distribution of the wavelet subband coefficients using members of the above family and the feature extraction step consists in estimating the model parameters using a Maximum-Likelihood estimator at each subband. The similarity measurement is performed by employing a suitably designed distance function between normalized characteristic functions.

In Chapter 6 we enhance the capacity of the univariate model by jointly modeling the subband coefficients at the same level or between adjacent levels. In this case, the extracted features are the so-called *covariation matrices* pertaining to lower-than-two order correlations. We compare two similarity functions, with the first based on the Frobenius norm of the difference between covariation matrices and the second based on the Kullback-Leibler Divergence between multivariate

Gaussian distributions.

In Chapter 7 we develop a rotation-invariant CBIR system based on a steerable pyramid image representation and a joint sub-Gaussian model. We describe the construction of a steerable model along with a rotation-invariant version of the distance based on Frobenius matrix norms. We also employ a local-variance adaptive model, combined with a Gaussianization procedure resulting in a simplified transformation domain, in terms of modeling the coefficients using well known tools from the Gaussian theory. A suitable similarity function for this model is a rotation-invariant version of the Kullback-Leibler Divergence between multivariate Gaussian distributions. Finally, we conclude with some remarks and we propose future research directions.

Part I

Background - Mathematical Foundations

Chapter 2

Background

2.1 Introduction

The search of large digital multimedia libraries, unlike the search of conventional text-based digital databases, cannot be realized by simply searching text annotations. Because of the detail in multimedia data, it is difficult to provide automatic annotation without human support. The challenging task is the design of completely automatic mechanisms that extract meaning from this data and characterize the information content in a compact and meaningful way.

Content-based information retrieval is a set of techniques for retrieving relevant information from a database on the basis of automatically-derived features, that accurately specify the information content of the items in the database. As we will focus our attention on digital image databases, we will refer to Content-based Image Retrieval (CBIR) as a technique for retrieving relevant images from an image database, by employing a set of automatically-derived features that describe the content of a given image.

CBIR for general image databases is a highly demanding and challenging problem. A major constraint is the large *size of the databases*, which leads traditional methods of interaction to failure. For instance, while it is relatively easy for a user to look over a few hundred thumbnail images searching for a specific query image, it is extremely hard to find the same query among several thousand images.

Another problem is the difficulty, of both people and machines, to understand the *"meaning" of an image*, which is subjective and depends on the experience of the viewer (cf. Figure 2.1). The problem of recognizing objects in pictures and translate image content into higher-level semantic terms, is one of the most difficult (and unsolved) problems in computer vision.

Image understanding is closely related to image semantic classification that assigns an image to a semantic class. Although it is impossible to recognize and isolate objects in general-purpose images, there are methods to distinguish certain semantic types of images. The semantic classification, as an initial stage of a CBIR system, may be useful since it can restrict the search to the class of images with a particular semantic type. Then, the retrieval process can be improved by tuning an appropriate matching scheme to the semantic class of the query image.

Image				
Original Annotation	people pool swimmers water	cars formula tracks wall	clouds mountain sky water	field foals horses mare
Automatic Annotation	water people swimmers pool	cars tracks wall formula	sky mountain clouds park	field horses foals mare

Figure 2.1: An illustration of annotating images containing different objects.

A third difficulty in the design of a general CBIR system, is the formulation of a query. Different users have different interests and this is reflected in the way they may formulate a query. For example, an astronomer may be interested in astronomical images which are similar to a specific query image, while a doctor may be interested in radiologic images with regions having certain patterns. This means that a CBIR system must be able to answer queries such as [50]:

1. *Color histogram query*: find all images with 20% green, 50% blue and 30% red.
2. *Color layout query*: find all images with more 40% blue on the top part, about 20% yellow in the upper right corner and at most 30% green at the bottom part.
3. *Shape query*: find all images with a yellow triangle.
4. *Query by example*: find all images that are similar to a given image.
5. *Object/Object relationship query*: find all images containing a chair, or find images with a chair next to a table.
6. *Place*: find images representing the island of Crete.

In general, there are three levels of abstraction at which image objects can be defined [20]: pixel, feature and semantic level. A pixel-level object is a connected subimage. A feature-level object is a connected region having uniform feature values (e.g. color, texture). Finally, a semantic-level object is a connected region to which we can assign a unique semantic content (e.g. house, mountain).

Queries of type 1-4 are more related to the raw pixel values and feature-level objects, while the last two queries are related to the semantics of an image. As we are moving to queries requiring higher-level information, they become more complex and difficult to automate. Although each higher level of abstraction improves retrieval efficiency by reducing the amount of information, there are corresponding losses of accuracy.

The existing CBIR systems have been designed for different application domains, they extract the content of an image computing different sets of features and employ diverse query mechanisms.

A natural consequence is the difficulty to establish criteria for evaluating and comparing all these systems. Two commonly used criteria for the evaluation of a CBIR system are *precision* and *recall*, defined as follows:

$$Precision = \frac{\#\ Retrieved\ relevant\ images}{\#\ Retrieved\ images} \tag{2.1}$$

$$Recall = \frac{\#\ Retrieved\ relevant\ images}{\#\ Relevant\ images} \tag{2.2}$$

However, the "relevance" used in the above definitions, is a subjective term depending on the user. Other commonly used methods are the provision of some examples of the retrieval results and their comparison with the results of other systems, and also the systematic evaluation starting with a small database containing a few distinct categories and then creating a larger database by adding new categories. Although these methods provide some useful indication about whether a system has a better performance than another one, they are not ideal in evaluating the performance of a system in real-world applications.

2.2 Architecture of a CBIR system

Many CBIR systems have been designed for applications in different domains. Some of them are the IBM QBIC system [38], the MIT PhotoBook [91], the WBIIS [129] and the Blobworld [19]. All these systems are built on a common basis.

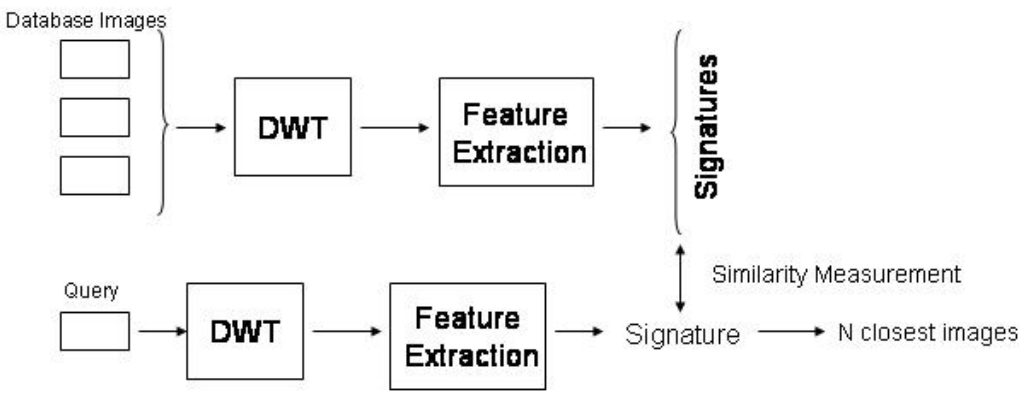


Figure 2.2: The architecture of a typical CBIR system.

The typical architecture of a CBIR system is shown in Figure 2.2. We can distinguish two major tasks, namely Feature Extraction (FE) and Similarity Measurement (SM). In the FE step, a set of features, constituting the so-called image signature, is generated to accurately represent the content of a given image, creating a better correlation of the pixel representation with image semantics. This set has to be much smaller in size than the original image, while capturing as much as possible of the image information. As a result, the original image database is now replaced by a signature database much smaller in size.

During the SM step, a distance function is employed which measures how close to a query

image each image in the database is, by comparing their signatures. This type of measurement is referred to as global similarity, since it is based on the overall properties of images. However, there are querying systems that retrieve based on a specific region in an image, by employing local similarity measures.

The retrieval procedure is roughly described by the following steps:

1. A query image is specified by a user and inserted in the retrieval system.
2. The FE module process the query, that is, it extracts a signature representing its content.
3. The SM module compares the query with each image in the database using the features stored in the signature database.
4. Sort the query results in descending order (starting from the most similar image)
5. Retrieve the (top N) relevant images from the image database and present them to the user.

The FE and SM steps are domain-specific, varying according to the application area. For example, a CBIR system adapted to biomedical images must, primarily, be able to extract local features (e.g. regions) rather than global features and employ local similarity measures, while a system designed for painting images may use global features (like color distribution) and the corresponding global similarity criteria.

2.3 Previous work on FE and SM

Extraction of features from pixel values and of semantics from features are lossy processes. Although each higher-level of abstraction improves retrieval efficiency, as it has been already mentioned, the extraction of image semantics is still an unsolved problem.

The FE step in traditional CBIR systems, is commonly based on pixel values of raw data. Typical low-level image features are the color, shape and texture.

Color Histogram

After selecting a color space in which we represent an image, a widely used method is to form the color histogram [112], which characterizes an image by its color distribution. For example, to represent a gray-scale image, we may generate 256 number of counting bins, each representing a value ranging from black ($= 0$) to white ($= 255$). This color histogram feature of 256 dimensions represents the global distribution of colors in an image, without containing any information about the location of the colors. Figure 2.3 shows the histograms for two color images in the R(ed)G(reen)B(lue) 3-dimensional color space. Thus, using similarity of histograms as a measure, two images may be considered to be very close to each other even though they have completely different semantics. The commonly used distances that define the SM step are the Euclidean distance and its variations [112, 51] and similarity measures derived from related optimality criteria, such as the Earth mover's distance [96]. Image retrieval based on color histogram is

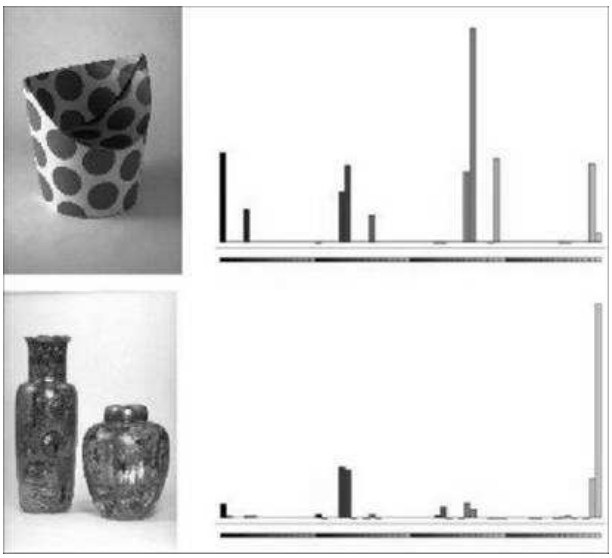


Figure 2.3: Example of color histograms.

sensitive to intensity variations and color distortions, and it does not take into consideration information about shape and texture.

Color Layout

A way to overcome the limitations of a histogram-based search, is to introduce a color layout approach. Traditional methods based on the color layout divide the image into equal-sized blocks and compute the average color in each block [5]. These values constitute the signature of the image. Then, image matching is performed using Euclidean distance or some of its variations. The color layout is, in fact, a low-resolution representation of the original image, so this approach does not perform well when the image contains high frequency information, such as abrupt color changes. For instance, if there are pixel values ranging from black to white in a block, the color layout approach will give a very smooth average value for this block, which does not represent its true color variation.

A different color layout approach is based on wavelet[1] coefficients, instead of averaging the color in each block. In this case, each image is decomposed in a set of low frequency and high frequency bands at different resolution levels. One of the first multiresolution approaches, presented in [56], makes use of the Haar wavelets to decompose each image. Then, the average color, the location and the signs of the M largest-magnitude coefficients constitute the signature components for each image. An appropriate "image querying metric" is constructed and the similarity between two images is, essentially, specified by the number of the common significant wavelet coefficients. A drawback of using the Haar wavelet family is that it cannot efficiently separate an image into low- and high-frequency bands. A later system, called WBIIS [129], uses the family of Daubechies' wavelets to achieve a more efficient and accurate representation of image content.

[1] Wavelets are introduced in Chapter 3

By adjusting the block sizes or the number of levels of wavelet decompositions, the resolution of a color layout representation can be refined. The finest color layout corresponds to a block size equal to one (pixel block). For a properly selected block size, the color layout retains information about color location, shape and texture. On the other hand, color layout representation is sensitive to shifting, scaling and rotation of the original image [129].

Region Analysis

In order to overcome the inadequacy of the above approaches, many CBIR systems use region-based methods representing an image at the object-level. The major module of such a system is the one that performs image segmentation, to decompose an image into regions corresponding to objects (in the ideal case). This is a higher-level image representation, as it attempts to approach the human perceptual system. The main difficulty in the design of a region-based CBIR system, is that the user's semantic understanding of an image is at a higher level than the region representation. As a consequence, we have to convert a semantic object into low-level features, such as color, texture, shape and location. For example, suppose that a user submits a query image containing a bird. Since the concept of bird cannot be explicitly given in region representation, the user must translate this concept into shape, color, texture or combinations of them.

Extraction of shape features [47, 57, 7] performs very well only in the case of typical geometric shapes (circles, triangles etc.). Texture is a well-studied property of image regions. Loosely speaking, texture characterizes spatially homogeneous regions consisted of repeated elements, often subject to some randomization in their location, size, color and orientation. Figure 2.4 shows examples of texture images. Many texture descriptors have been developed, including

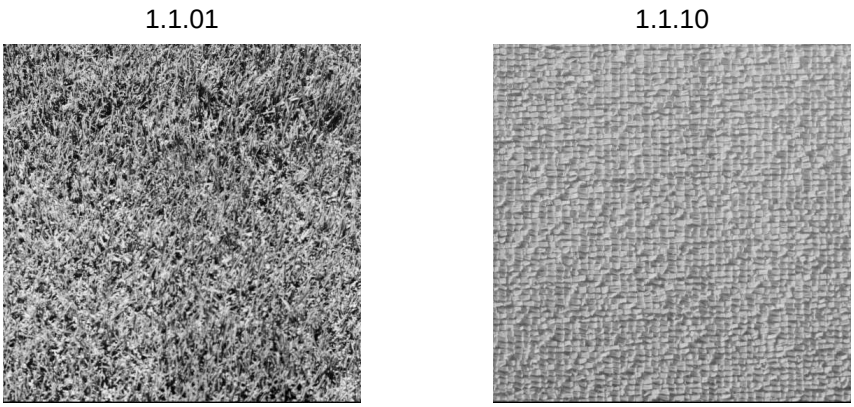


Figure 2.4: Texture samples obtained from the Brodatz database.

multiorientation filter banks and spatial Gabor filters [71, 42], as well as statistical descriptors like Markov random fields, in which the texture is characterized by statistical interactions within local neighborhoods [54, 28]. The joint statistics of wavelet coefficients were also employed in the construction of parametric texture models [107, 92] and in segmentation procedures [65].

The Blobworld [19] is a typical system that combines color, texture and position features to

segment each image and perform the retrieval. The CBIR system architecture described in [20], also specifies image content at one or more of the above description levels (color, texture, objects) and organizes the data in a progressive manner which makes the search process more efficient, as it requires accessing only a small fraction of the data at each step.

Different similarity measures have been developed, depending on the selected features. If we treat region matching as template matching, the definition of a correlation coefficient results in retrieving the subimages that most closely approximate the template [20]. Texture matching is performed via a progressive algorithm or using the Bhattacharrya coefficient between multivariate kernel density estimates [42]. At the object level, an example of similarity measure is given in [109]. Similarity between images is measured by the distance between the Composite Region Template (CRT) descriptor matrices, which contain information about the relative ordering of regions/objects.

The CBIR systems employing region-based retrieval, have the drawback that an object is often partitioned into different regions, with none of them being representative for the object. This happens due to the great difficulty of achieving accurate segmentation. Queries based on regions often yield results that are unrelated to the query image.

Manjunath *et. al.* [77] present an overview of color and texture descriptors that have been approved for the MPEG-7 standard. The color descriptors include the above mentioned color structure histogram and color layout descriptor, combined with a histogram descriptor that is coded using the Haar transform and a dominant color descriptor. The texture descriptors include one that characterizes homogeneous texture regions and another that represents the local edge distribution.

2.4 Probabilistic Image Retrieval

Since the goal for any retrieval system is to minimize the probability of retrieval error, the retrieval problem can be formulated in a probabilistic framework. Then, images are represented as realizations of stochastic processes and the similarity measurement becomes the posterior probability of the query image under the probabilistic models associated with the image classes in the database.

Let $\mathcal{F}$ denote the feature space and $\vec{x} \in \mathcal{F}$ a feature vector. Let also $\mathcal{S} = \{1, \ldots, K\}$ be the set of class indicators associated with the image classes in the database. Denote the probability density function (PDF) of the query feature vectors by $p_q(\vec{x})$ and the PDF of class $i \in \mathcal{S}$ by $p_i(\vec{x})$. The design of a retrieval system in a probabilistic framework, consists of finding a map

$$g : \mathcal{F} \mapsto \mathcal{S} \tag{2.3}$$

These maps constitute the set of similarity functions.

The goal of a probabilistic CBIR system is the *minimization of the probability of retrieval error*, that is, the probability $P(g(\vec{x}) \neq s)$ that if we provide the system with a set of feature vectors $\vec{x}$ drawn from class s, the system will return images from a class $g(\vec{x})$ other than s. The following theorem holds:

Theorem 2.4.1 *Consider a feature space $\mathcal{F}$, a query set $\boldsymbol{X} = \{\vec{x}_1, \ldots, \vec{x}_N\}$ of N independent feature vectors with PDF $p_q(\vec{x})$, a set of K image classes with PDFs $p_i(\vec{x})$, $i = 1, \ldots, K$ and the set of similarity functions*

$$g : \mathcal{F} \mapsto \mathcal{S} \ .$$

Then, the optimal similarity function, i.e. the one minimizing the probability $P(g(\boldsymbol{X}) \neq s)$, is the Bayes or maximum a-posteriory (MAP) classifier

$$\begin{aligned} g^*(\boldsymbol{X}) &= arg \max_i P(s = i|\boldsymbol{X}) \\ &= arg \max_i p(\boldsymbol{X}|s = i)P(s = i) \end{aligned} \tag{2.4}$$

where $p(\boldsymbol{X}|s = i)$ is the likelihood for the i-th class and $P(s = i)$ its prior probability.

Proof. See [44], among many other books. ■

The minimum value of the probability of error is the *Bayes error* defined by [44]

$$L^* = 1 - E_{\mathbf{X}}\{\max_i P(s = i|\mathbf{X})\} \ . \tag{2.5}$$

Most of the similarity functions in current use for image retrieval are special cases of the MAP similarity function.

In general, image classes may have different prior probabilities. Under the assumption that all classes are a-priori equally likely, the MAP classifier reduces to the *Maximum Likelihood* (ML) classifier. Indeed, in a database with K equally likely classes, $P(s = i) = 1/K$, $i = 1, \ldots, K$, Eq. (2.4) becomes

$$g^*(\mathbf{X}) = arg \max_i p(\mathbf{X}|s = i) \ , \tag{2.6}$$

which is exactly the ML classifier. Assuming that the query consists of a set of N independent and identically distributed (i.i.d.) feature vectors $\mathbf{X} = \{\vec{x}_1, \ldots, \vec{x}_N\}$, the last equation can be written as

$$g^*(\mathbf{X}) = arg \max_i \frac{1}{N} \sum_{j=1}^{N} \log p(\vec{x}_j|s = i) \ . \tag{2.7}$$

When the number N is large, application of the Weak Law of Large Numbers to Eq. (2.7) results

in the following equations [27]

$$
\begin{aligned}
g^*(\mathbf{X}) &\overset{N\to\infty}{\longrightarrow} arg\max_i E_{p_q}\{\log p(\vec{x}|s=i)\} \\
&= arg\max_i \int p_q(\vec{x})\log p(\vec{x}|s=i)\,\mathrm{d}\vec{x} \\
&= arg\max_i \int p_q(\vec{x})\log p_i(\vec{x})\,\mathrm{d}\vec{x} \\
&= arg\min_i \int p_q(\vec{x})\log \frac{p_q(\vec{x})}{p_i(\vec{x})}\,\mathrm{d}\vec{x} \qquad (2.8) \\
&= arg\min_i D(p_q\|p_i) \qquad (2.9)
\end{aligned}
$$

where $D(p_q\|p_i)$ denotes the *Kullback-Leibler divergence* (KLD) (or *relative entropy*) between the two densities, $p_q(\cdot), p_i(\cdot)$.

Now, consider a database with K images and assume that each image corresponds to an image class, that is the set $\mathcal{S} = \{1,\ldots,K\}$ contains the indicators associated with the images in the database. The problem of retrieving the top M images similar to a given query image, can be formulated as a *Multiple Hypotheses Problem*. The query image I_q is represented by a data set, $\mathbf{X} = \{x_1,\ldots,x_N\}$, usually obtained after a pre-processing step, such as a transformation and each image in the database, I_i $(i = 1,\ldots,K)$, is assigned with a hypothesis H_i. Therefore, the problem of retrieving the top M images is converted to the problem of selecting, among the K possible hypotheses, the M best ones (with a ranking order) that describe the data $\mathbf{X}$ of the given query image.

Under the assumption that all hypotheses are a-priori equally likely, the optimum rule, i.e. with the minimum probability of retrieval error, is to select the hypotheses with the highest likelihoods among the K ones. Thus, the top M matches correspond to the M hypotheses, $H_{i_1}, H_{i_2},\ldots,H_{i_M}$ for which

$$
p(\mathbf{X}|H_{i_1}) \geq p(\mathbf{X}|H_{i_2}) \geq \cdots \geq p(\mathbf{X}|H_{i_M}) \geq p(\mathbf{X}|H_i) \qquad i \neq i_j\ (j=1,\ldots,M) \qquad (2.10)
$$

This is the well-known Maximum Likelihood (ML) selection rule.

The main drawback of the above setting is that its implementation requires M steps, each of which may be computationally inefficient depending on the size of the data set $\mathbf{X}$. A solution to this problem is to adopt a *parametric* approach. Then, each conditional PDF, $p(\mathbf{X}|H_i)$ is modeled by a member of a family of PDFs, denoted by $p(\mathbf{X};\boldsymbol{\theta}_i)$ where $\boldsymbol{\theta}_i$ is a set of parameters specifying the model. In this framework, the extracted signature for the image I_i is the *estimated model parameter* $\hat{\boldsymbol{\theta}}_i$, computed in the FE step. This set is usually of small size satisfying the computational constraints of a CBIR system. Let $\boldsymbol{\Theta}$ denote the space of model parameters and consider the data set $\mathbf{X}$ as an i.i.d. sequence from $p(\mathbf{X};\boldsymbol{\theta}_q)$, the model associated with the query image. Then for large N, implementation of Eq. (2.8) gives the optimal rule for retrieving the top M similar images to the given query:

1. Compute the KLDs between the query density $p(\mathbf{X};\boldsymbol{\theta}_q)$ and the density $p(\mathbf{X};\boldsymbol{\theta}_i)$ associated

with image I_i in the database, $\forall i = 1, \ldots, K$

$$D(p(\mathbf{X}; \boldsymbol{\theta}_q) \| p(\mathbf{X}; \boldsymbol{\theta}_i)) = \int p(x; \boldsymbol{\theta}_q) \log \frac{p(x; \boldsymbol{\theta}_q)}{p(x; \boldsymbol{\theta}_i)} \, \mathrm{d}x \ . \tag{2.11}$$

2. Retrieve the M images corresponding to the M smallest values of the KLD.

Under the same asymptotic assumption of a large N, if the FE step uses a consistent estimator, which ensures convergence of the estimated parameter $\hat{\boldsymbol{\theta}}$ to the true parameter $\boldsymbol{\theta}$, then Eq. (2.11) can be computed using the estimated model parameters $\hat{\boldsymbol{\theta}}_q$ and $\hat{\boldsymbol{\theta}}_i$. The ML estimator is a consistent estimator and for the query image it gives

$$\hat{\boldsymbol{\theta}}_q = arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \log p(\mathbf{X}; \boldsymbol{\theta}) \ . \tag{2.12}$$

We can also apply a *Chain Rule* [27], in order to combine the KLDs from multiple data sets. This rule states that the KLD between two joint PDFs, $p(\mathbf{X}, \mathbf{Y})$ and $q(\mathbf{X}, \mathbf{Y})$, is

$$D(p(\mathbf{X}, \mathbf{Y}) \| q(\mathbf{X}, \mathbf{Y})) = D(p(\mathbf{X}) \| q(\mathbf{X})) + D(p(\mathbf{Y}|\mathbf{X}) \| q(\mathbf{Y}|\mathbf{X})) \ . \tag{2.13}$$

Under the independent assumption between the data sets $\mathbf{X}, \mathbf{Y}$, Eq. (2.13) is simplified in the following form

$$D(p(\mathbf{X}, \mathbf{Y}) \| q(\mathbf{X}, \mathbf{Y})) = D(p(\mathbf{X}) \| q(\mathbf{X})) + D(p(\mathbf{Y}) \| q(\mathbf{Y})) \ . \tag{2.14}$$

Finally, the fact that KLD is a convex function permits the implementation of efficient retrieval schemes using multiscale representations [24]. The Wavelet Transform, introduced in the next chapter, belongs to this class of representations. Also, most of the current similarity functions in CBIR applications are sub-optimal special cases of the Bayesian classifier [126].

While the complexity of the MAP similarity function is linear in the cardinality N of the query data set $\mathbf{X}$, the complexity of the KLD is a function of the size of the parameter set $\hat{\boldsymbol{\theta}}$. Since this size is typically quite small, the optimal selection rule based on the minimization of KLD is computationally more efficient whenever it results in a closed-form expression. In general, the KLD does not have a closed form, except for some density families. This set of families includes models that have been widely applied to the CBIR problem, such as the Gaussian and the histogram. However, such models have important limitations in the context of a CBIR problem, as they may either be too simplistic to accurately model the densities describing the content of images or rely on assumptions that can compromise retrieval accuracy (e.g. independence between the components of the feature vectors). In Chapter 4, the Alpha-Stable family of densities is introduced as a candidate intending to overcome these limitations.

Chapter 3

Wavelet Analysis

3.1 Fourier Transform

Fourier analysis is one of the oldest subjects in mathematics and it is of great importance to both mathematicians and engineers, since it has many applications in different fields. Fourier analysis is well known for the (integral) Fourier transforms and Fourier series. A Fourier transform is the Fourier integral of a function f defined on the real line $\mathbb{R}$. When f represents an analog signal, then its domain of definition is called the continuous time-domain and the domain of definition of the Fourier transform F is called the frequency domain. The reason for this, is that the Fourier transform F of f describes the spectral behavior of the signal f in terms of frequency. A Fourier series expansion is a transformation of bi-infinite sequences to periodic functions. A digital signal is assumed to be a bi-infinite sequence and its domain of definition ($\mathbb{Z}$) is called the discrete time-domain. The Fourier series of the digital signal f again gives us the spectral information of the signal and the domain of definition is the real line $\mathbb{R}$, which is the frequency-domain.

For each p, $1 \leq p < \infty$, with $L^p(\mathbb{R})$ we denote the class of measurable functions on $\mathbb{R}$ such that the integral

$$\int_{-\infty}^{\infty} |f(x)|^p \, \mathrm{d}x$$

is finite ($< \infty$). Let $L^\infty(\mathbb{R})$ denote the class of almost everywhere (a.e.) bounded functions, that is, functions bounded everywhere except on sets of measure zero. If we supply $L^p(\mathbb{R})$ with the norm

$$\|f\|_p = \left\{ \int_{-\infty}^{\infty} |f(x)|^p \, \mathrm{d}x \right\}^{\frac{1}{p}} , \tag{3.1}$$

if $1 \leq p < \infty$ and

$$\|f\|_p = \sup_{-\infty < x < \infty} |f(x)| , \tag{3.2}$$

if $p = \infty$, then each of them is a Banach space. We may define the "inner product"

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x) \overline{g(x)} \, \mathrm{d}x , \tag{3.3}$$

$f, g \in L^2(\mathbb{R})$. With this inner product the Banach space $L^2(\mathbb{R})$ becomes a Hilbert space. We also know that $\langle f, f \rangle = \|f\|_2^2$, $f \in L^2(\mathbb{R})$.

The **Fourier transform (FT)** of a function $f \in L^1(\mathbb{R})$ is defined by

$$F(\omega) = (\mathcal{F}f)(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\,\mathrm{d}x \ . \tag{3.4}$$

(In the following we will use the notation F for the FT of a function f). The Fourier integral measures **"how much oscillations at the (radian) frequency ω there are in f"**. The FT is also a bounded and continuous function. The analysis coefficients $F(\omega)$ define the notion of global frequency ω in a signal. These coefficients are computed as inner products of the signal $f(t)$ with sinewave basis functions of infinite duration. This is why Fourier analysis works well if $f(t)$ is composed of a few stationary components. But in the non-stationary case any abrupt change in time is spread out over the whole frequency axis in $F(\omega)$, so the FT is not suitable.

The information we acquire by the integral, corresponds to all time instances, since the integration is from $-\infty$ to ∞ over time. As a consequence we have that no matter where in time the component with frequency ω appears, it will affect the result of the integration equally. So, the FT is suitable if we are interested in whether a frequency component exists or not and it gives no information about the time instant where it appears.

When we take the FT of a function (or a signal), what we are doing is multiplying the original function with a complex expression consisted of sines and cosines of frequency ω and then integrating which is a sort of infinite summation. If the value we will take as a result is large it means that a major portion of the function is composed of frequency ω, while if this value is small the function doe not have a major frequency component of ω and finally, if the value is zero the function does not contain a component of frequency ω.

Let $F(\omega) \in L^1(\mathbb{R})$, then the **inverse Fourier transform** of $F(\omega)$ is defined by

$$(\mathcal{F}^{-1}F)(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{ix\omega}\,\mathrm{d}\omega \ . \tag{3.5}$$

We have the following,

Theorem 3.1.1 *Let $f \in L^1(\mathbb{R})$ such that its Fourier transform $F \in L^1(\mathbb{R})$. Then $f(x) = (\mathcal{F}^{-1}F)(x)$ at every point x where f is continuous.*

Since we assume signals f with finite energy, which means that

$$\int_{-\infty}^{\infty} |f(x)|^2\,\mathrm{d}x < \infty$$

we want an extension of the Fourier transform for functions $f \in L^2(\mathbb{R})$. If $f \in L^2(\mathbb{R})$ but $f \notin L^1(\mathbb{R})$, we cannot calculate the Fourier transform with the previous formula because $f(x)e^{i\omega x}$ is not integrable. We will make the extension in $L^2(\mathbb{R})$ using the Fourier transforms of functions in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. The following theorem is valid:

Theorem 3.1.2 *Let $f, h \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Then*

$$\int_{-\infty}^{\infty} f(x)\overline{h(x)}\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(s)\overline{H(s)}\,\mathrm{d}s\ . \qquad (Parseval\ formula)$$

For $h = f$ we have that

$$\int_{-\infty}^{\infty} |f(x)|^2\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\infty}^{\infty} |F(s)|^2\,\mathrm{d}s\ . \qquad (Plancherel\ formula)$$

The previous theorem proves that the Fourier transform conserves inner products and norms in $L^2(\mathbb{R})$, up to a factor of 2π. $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, so we can find a sequence of functions in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ ($\{f_n\}_{n\in\mathbb{Z}}$) which converges to f:

$$\lim_{n\to\infty} \|f - f_n\| = 0\ .$$

that is, it is a Cauchy sequence. Moreover, $f_n \in L^1(\mathbb{R})$ so its Fourier transform F_n is well defined. Using the Plancherel formula we have that

$$\|F_n - F_p\| = \sqrt{2\pi}\|f_n - f_p\|$$

is arbitrarily small for n and p large enough, that is, $\{F_n\}_{n\in\mathbb{Z}}$ is also a Cauchy sequence. A Hilbert space is complete, which means that every Cauchy sequence converges to an element of the space. So, there is $F \in L^2(\mathbb{R})$ such that

$$\lim_{n\to\infty} \|F - F_n\| = 0\ .$$

By definition F is the Fourier transform of f. This extension to $L^2(\mathbb{R})$ satisfies the properties of the Fourier transform (in $L^1(\mathbb{R})$).

3.1.1 Discrete Fourier Transform (DFT)

We saw that the Fourier transform is used to describe the spectral content of an analog signal f with finite energy (that is, $f \in L^2(\mathbb{R})$). We also want to describe the spectral content in the case we have a digital (discrete) signal. A digital signal is represented by a sequence $\{f[k] = f(kT)\}_{k\in\mathbb{Z}} \in l^p(\mathbb{Z})$ sampling the values of the analog signal with period T, where $l^p(\mathbb{Z})$ are the sequence spaces of bi-infinite sequences $\{a_k\}$, $k \in \mathbb{Z}$, that satisfy $\|\{a_k\}\|_{l^p} < \infty$, where

$$\|\{a_k\}\|_{l^p} = \begin{cases} \left\{\sum_{k\in\mathbb{Z}} |a_k|^p\right\}^{\frac{1}{p}}\ , & 1 \le p < \infty, \\ \sup_k |a_k|\quad , & p = \infty. \end{cases}$$

The space $l^2(\mathbb{Z})$ is a Hilbert space with inner product

$$\langle\{a_k\}, \{b_k\}\rangle_{l^2} = \sum_{k\in\mathbb{Z}} a_k\bar{b}_k\ .$$

In order to represent the spectral content of a digital signal, we introduce the Discrete Fourier Transform (DFT) $\mathcal{F}^*$ as follows:

$$(\mathcal{F}^*\{f[k]\})(x) = \sum_{k\in\mathbb{Z}} f[k]e^{ikx} \ . \tag{3.6}$$

In other words, the DFT of $\{f[k]\}$ is the **Fourier series** with **Fourier coefficients** given by $\{f[k]\}$. For $\{f[k]\} \in l^1$, the series converges absolutely and uniformly for all $x \in \mathbb{R}$.

The Fourier coefficients can be written as inner products as follows:

$$f[n] = \langle F(\omega), e^{-i\omega n}\rangle = \frac{1}{2\pi}\int_0^{2\pi} F(\omega)e^{i\omega n}\,\mathrm{d}\omega \ . \tag{3.7}$$

The discrete sample values of the time function and the corresponding discrete spectrum constitute a DFT pair. The transform pair for the usual DFT definition of a N-point time sequence $f[n]$, $n = 0, ..., N-1$ and its N-point transform sequence $F[k]$, $k = 0, ..., N-1$, is given by (**Discrete Fourier Transform**)

$$F[k] = \sum_{n=0}^{N-1} f[n]e^{-i2\pi kn/N} \ , \quad k = 0, ..., N-1$$

(**Inverse Discrete Fourier Transform**)

$$f[n] = \frac{1}{N}\sum_{k=0}^{N-1} F[k]e^{i2\pi kn/N} \ , \quad n = 0, ..., N-1 \ .$$

There is a massive bibliography on the well studied properties of the Fourier transform. For the interested readers, some indicative books are [10, 37, 60, 85, 58].

3.1.2 Short-Time Fourier Transform (STFT)

Fourier transform is useful only when the signal is stationary. We can use it in the non-stationary case (when the signal is composed of spectral characteristics depending on time) if we are interested in what frequency components exist in the signal and we don't want any information about the time instants they occur. Now, what can we do if we want to have information both in the time and frequency domain? The natural problem when we have a non-stationary signal is to find a way, if possible, in order to discover what frequency components exist at each time instant.

The solution to this problem is to introduce a **two-dimensional** time-frequency representation $S_f(\tau, \omega)$ of a signal $f(t)$. But how such a representation could work? The general idea is that observing a signal through a "window" $g(t)$ of limited extend, centered at time location τ, it may be possible to assume that this portion of the signal (which is seen through the window) is stationary. So, the signal is divided into small enough portions, where each portion can be assumed to be stationary, using a window $g(t)$ with suitable width in order to ensure the stationarity of each portion.

This procedure can be realized as follows. First locate the window $g(t)$ to the beginning of the signal, that is, $g(t)$ is located at $\tau = 0$. Then, the signal $f(t)$ is multiplied by the window function (this happens when we "observe a signal through a window"). The result is a stationary signal (or at least it supposed to be) and now we can take the FT of it. After this, we have obtained a frequency representation of the first portion of the signal. Then we shift the window $g(t)$ to a new location τ_1, multiply the signal by it and finally take the FT of this second portion of the signal. We repeat these steps until the end of the signal. The Fourier transform of the windowed signal $f(t)\overline{g}(t-\tau)$ yields the short-time Fourier transform (STFT)

$$S_f(\tau,\omega) = \int_{-\infty}^{\infty} f(t)\overline{g}(t-\tau)e^{-i2\pi\omega t}\,\mathrm{d}t \tag{3.8}$$

which maps $f(t)$ into a two-dimensional function in a time-frequency plane (τ,ω). The STFT can be viewed as the projections (inner products) of the signal $f(t)$ onto a set of basis functions:

$$S_f(\tau,\omega) = \langle f, g_{\tau\omega}\rangle$$

where

$$g_{\tau\omega}(t) = g(t-\tau)e^{i2\pi\omega t}\ .$$

We want $\|g\| = 1$ so that $\|g_{\tau f}\| = 1$.

If the window function $g(t)$ is of finite energy (that is, $g(t) \in L^2(\mathbb{R})$), we can define the **inverse STFT** according to

$$f(t) = \frac{1}{E_g}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_f(\tau,\omega)g(t-\tau)e^{i2\pi t\omega}\,\mathrm{d}\tau\,\mathrm{d}\omega \tag{3.9}$$

with $E_g = \int_{-\infty}^{\infty} |\,g(t)|^2\,\mathrm{d}t$. The above equation expresses the signal as a decomposition on the family of the atoms $g_{\tau\omega}(t)$.

As a two-dimensional function we can view it in two ways, fixing each time the one of the two variables. If the variable τ is fixed, then STFT can be viewed as the windowing of the signal. Given the windowed signal around time τ we compute all "frequencies" ω of the STFT. On the other hand, at a given frequency ω_k STFT is interpreted as filtering the signal "at all times", using a bandpass filter with impulse response the window function modulated to that frequency. That is, STFT appears as an analysis of the signal by a continuous filter bank of uniform filters with constant bandwidth.

Each coefficient $\langle f, g_{\tau\omega}\rangle$ contains some information, which is represented in a time-frequency plane by a region whose location and width depends on the time-frequency spread of $g_{\tau\omega}$. The support of $g_{\tau\omega}$ corresponds to a Heisenberg box of constant area, which is independent of (τ,ω), centered at (τ,ω). The fact that the size of the boxes remains the same for all (τ,ω), means that the STFT has the same resolution across the time-frequency plane [25].

3.2 Wavelet Transform

As we saw above, the main drawback of the STFT is the fixed time-frequency resolution over the entire time-frequency plane, since STFT uses the same window at all frequencies. On the other hand, frequency is directly proportional to the number of cycles per unit time, which has as a result that it takes a narrow window in the time-domain to locate high-frequency components more precisely and a wide time-window to analyze low-frequency behavior. Hence, the STFT is not suitable for analyzing signals with both very high and very low frequencies. So, we would like to find a way to overcome the resolution limitation of the STFT, constructing a flexible time-frequency window which narrows when observing high-frequency components and widens when observing low-frequency components. The wavelet transform relative to a "basic wavelet", which is introduced in this section, provides such a window.

3.2.1 The 1-Dimensional Wavelet Transform

Continuous Wavelet Transform

We have seen the resolution limitation of the STFT, which uses a time-frequency window with unchanged width in observing a signal. The wavelet transform gives a solution to this problem providing a flexible time-frequency window, which automatically narrows when observing high-frequency components and widens when observing low-frequency components.

Definition 3.2.1 (1-D Continuous Wavelet Transform (CWT))
If $\psi(t) \in L^2(\mathbb{R})$ satisfies the "admissibility" condition

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} \, \mathrm{d}\omega < \infty \ , \tag{3.10}$$

then $\psi(t)$ is called a basic wavelet. The continuous wavelet transform (CWT) of a function $f(t)$, relative to a basic wavelet $\psi(t)$ is defined by

$$(W_\psi f)(b, \alpha) = \frac{1}{\sqrt{|\alpha|}} \int_{-\infty}^{\infty} f(t) \overline{\psi\Big(\frac{t-b}{\alpha}\Big)} \, \mathrm{d}t, \quad f \in L^2(\mathbb{R}) \ , \tag{3.11}$$

where $\alpha, b \in \mathbb{R}$ with $\alpha \neq 0$ and $\Psi(\omega)$ is the Fourier transform of $\psi(t)$

with b representing the time location parameter and α the scale parameter which is inversely proportional to the frequency, $\alpha = \frac{\omega_0}{\omega}$, with ω_0 denoting the central frequency of $\Psi(\omega)$ [25, 110, 73]. In the applications, $\psi(t)$ will be chosen such that its FT, $\Psi(\omega)$, has a sufficient decay so that due to the admissibility condition we require that

$$\Psi(0) = 0 \Leftrightarrow \int_{-\infty}^{\infty} \psi(t) \, \mathrm{d}t = 0 \ .$$

(This is why $\psi(t)$ is called a wavelet, that is, a "small wave")

By setting

$$\psi_{b,\,\alpha}(t) = \frac{1}{\sqrt{|\alpha|}}\,\psi\Big(\frac{t-b}{\alpha}\Big)\ , \tag{3.12}$$

the CWT becomes

$$(W_\psi f)(b,\alpha) = \int_{-\infty}^{\infty} f(t)\overline{\psi_{b,\,\alpha}}(t)\,dt = \langle f, \psi_{b,\,\alpha}\rangle\ . \tag{3.13}$$

This relation shows that the wavelet transform gives a measure of similarity between the basis functions (wavelets) and the function (or signal) itself. The similarity is in the sense of similar frequency content. The CWT coefficients given by Eqs. (3.11)and (3.13) measure how close to the wavelet, at the current scale, the signal is. Eq. (3.13) also represents the CWT as an expansion of the signal onto a set of basis functions (the wavelets), given by Eq. (3.12), which are scaled and translated versions of the basic (or mother) wavelet $\psi(t)$. If the signal has a major component of the frequency corresponding to the current scale, then the wavelet at the current scale will be similar (close) to the signal at the location where this frequency component exists. The corresponding CWT coefficient computed at this time-scale point will have a large value.

The basic wavelet, which corresponds to the scale $\alpha = 1$, contains only one modulation frequency. So, we have to rescale the wavelet (by dilating or compressing it changing the scale) in order to extract the frequency content of the signal at frequencies other than the frequency of the basic wavelet. This is one of the main differences with STFT, which uses an analysis window that contains a number of modulation frequencies. The second main difference is that the STFT uses an analysis window of fixed length at all frequencies, while the CWT uses an analysis window (the basic wavelet) with a length which depends on the frequency.

We can reconstruct any finite-energy signal $f(t) \in L^2(\mathbb{R})$ from its CWT values using the following inverse formula,

$$f(t) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{(W_\psi f)(b,\alpha)\}\,\psi_{b,\,\alpha}(t)\,\frac{\mathrm{d}\alpha\,\mathrm{d}b}{\alpha^2}\ . \tag{3.14}$$

The finiteness of the constant C_ψ (defined in Eq. (3.10) restricts the class of $L^2(\mathbb{R})$ functions $\psi(t)$ that can be used as basic wavelets.

Four basic properties of the CWT are the following:

P(1): The linearity of the inner product ensures the linearity of the CWT,

$$(W_\psi(f+g))(b,\alpha) = (W_\psi f)(b,\alpha) + (W_\psi g)(b,\alpha)\ . \tag{3.15}$$

P(2): If $(W_\psi f)(b,\alpha)$ is the CWT of $f(t)$, then $\tilde{f}(t) = f(t-b')$ has a CWT given by the following relation,

$$(W_\psi \tilde{f})(b,\alpha) = (W_\psi f)(b-b',\alpha)\ . \tag{3.16}$$

P(3): If $f(t)$ has CWT given by $(W_\psi f)(b,\alpha)$, then $f_s(t) = \frac{1}{\sqrt{s}}\,f\big(\frac{t}{s}\big)$ has CWT given by the relation

$$(W_\psi f_s)(b,\alpha) = (W_\psi f)\Big(\frac{b}{s},\,\frac{\alpha}{s}\Big)\ . \tag{3.17}$$

P(4): For the function $f(t) \in L^2(\mathbb{R})$ and its CWT $(W_\psi f)(b,\alpha)$, the following relation holds,

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |(W_\psi f)(b,\alpha)|^2 \, \frac{d\alpha \, db}{\alpha^2} \ . \tag{3.18}$$

For a more detailed study on the properties of the CWT, the interested reader is referred to [25, 73].

In applications, of course, we cannot implement a continuous wavelet transform, so we are interested in reconstructing the signal using samples on a discrete set of the time-scale plane (that is, b, α will be not in $\mathbb{R}$ anymore). For $\psi_{b,\alpha}(t)$ to cover the whole time-axis at a fixed discretized scale $\alpha = \alpha_0^m$, we have to choose the translation parameter $b = nb_0\alpha_0^m$. After this discretization process,

$$\alpha = \alpha_0^m, \ b = nb_0\alpha_0^m, \quad m, n \in \mathbb{Z}, \ \alpha_0 > 1, \ b_0 > 0 \ ,$$

we obtain a discretized family of wavelets

$$\psi_{n,m}(t) = \alpha_0^{-m/2} \, \psi(\alpha_0^{-m} t - nb_0) \tag{3.19}$$

(the normalization factor $\alpha_0^{-m/2}$ makes $\psi_{n,m}(t)$ of unit norm). The case in which $\alpha_0 = 2$ and $b_0 = 1$, is called dyadic and for such a case we can find orthonormal bases which are used in the reconstruction of a function from its CWT coefficients.

When this discretized set of wavelet functions constitutes an orthonormal basis for $L^2(\mathbb{R})$, we can reconstruct a function $f(t) \in L^2(\mathbb{R})$ using its CWT coefficients $\langle f(t), \psi_{n,m}(t) \rangle$ as a wavelet series

$$f(t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \langle f(t), \psi_{n,m}(t) \rangle \, \psi_{n,m}(t) \ . \tag{3.20}$$

Instead of using orthonormal wavelet families, we can reconstruct a function $f(t)$ using biorthogonal families of wavelets. In this case the wavelet used for the analysis is different from the one used for the synthesis (reconstruction) [73].

Multi-resolution Analysis

A fundamental concept in wavelet analysis is that of **multi-resolution analysis**. It is based on the idea of signal decompositions which rely on successive approximation. The given signal (or in general a function) will be decomposed in a coarse approximation plus added details. By applying the successive approximation recursively, the space of signals, $L^2(\mathbb{R})$, can be spanned by spaces of successive details at all resolutions. This happens due to the fact that, as the detail resolution goes to infinity, the approximation error goes to zero. We have the following definition [25]

Definition 3.2.2 *A multi-resolution analysis is a sequence $(V_j)_{j \in \mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ such that*

(i)

$$... \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset ...$$

(ii)

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$$

(iii)

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$$

(iv)

$$f(t) \in V_j \Leftrightarrow f(2^j t) \in V_0$$

(v)

$$f(t) \in V_0 \Leftrightarrow f(t-n) \in V_0 \ , \forall n \in \mathbb{Z}$$

(vi) *there exists a function* $\varphi(t) \in V_0$*, called a scaling function, such that the set* $\{\varphi(t-n)\}_{n\in\mathbb{Z}}$ *is an orthonormal basis in* V_0*.*

Note: Using $(iv), (v), (vi)$ we obtain that the set $\{\varphi_{n,j}(t) = 2^{-j/2}\varphi(2^{-j}t - n)\ ,\ n \in \mathbb{Z}\}$ is a basis for V_j.
Now consider W_j as the orthogonal complement of V_j in V_{j-1}, that is,

$$V_{j-1} = V_j \oplus W_j \ , \tag{3.21}$$

which leads to the decomposition (using (ii))

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j \ . \tag{3.22}$$

Also, using the property (iv) of the V_j spaces, there exists a same property for the W_j spaces,

$$f(t) \in W_j \ \Leftrightarrow \ f(2^j t) \in W_0 \ . \tag{3.23}$$

Hence, it is enough to find a function $\psi(t) \in W_0$, which will be the wavelet, so that $\{\psi(t-n)\ ,\ n \in \mathbb{Z}\}$ is a basis (orthonormal) of W_0. Then by the properties $(ii), (iii)$ and Eq. (3.22), the set

$$\psi_{n,j}(t) = 2^{-j/2}\psi(2^{-j}t - n), \quad n,\ j \in \mathbb{Z} \tag{3.24}$$

constitutes an orthonormal basis of $L^2(\mathbb{R})$ and by the property (3.23) $\{\psi_{n,j}\}_{n\in\mathbb{Z}}$ is an orthonormal basis for W_j. The scaling function $\varphi(t)$ has the characteristics of a lowpass filter, while the wavelet $\psi(t)$ has the characteristics of a highpass filter.

Concluding, we can give the following interpretation for the multi-resolution analysis:
By the property (ii) at the definition of multi-resolution analysis, every function $f \in L^2(\mathbb{R})$ can be approximated as closely as we want by an $f_N \in V_N$, for some $N \in \mathbb{Z}$. Since $V_{j-1} = V_j \oplus W_j$, f_N has a decomposition

$$f_N = f_{N+1} + g_{N+1} \ ,$$

where $f_{N+1} \in V_{N+1}$ is a coarsest approximation in the "approximation space" V_{N+1} and $g_{N+1} \in W_{N+1}$ is a detail in the "detail space" W_{N+1}. By repeating this process we obtain

$$f_N = g_{N+1} + g_{N+2} + \cdots + g_{N+M} + f_{N+M}$$

where $f_j \in V_j$ and $g_j \in W_j$. In general we can represent a function $f(t)$ as a succession of multi-resolution details and a coarsest approximation. Then following the inverse process, that is, adding these details to the coarsest approximation we can recover the original function f.

Discrete Wavelet Transform (DWT)

As we have already mentioned, we can discretize the continuous wavelet transform in order to reduce its redundancy. By discretizing the set of the indices n, m, we obtained a discrete set of wavelet functions $\psi_{n,m}$, which we used for the analysis and synthesis of a function $f(t)$. We computed the analysis coefficients

$$c_{n,m} = \langle f, \psi_{n,m} \rangle$$

and from these we recovered the original function

$$f(t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} c_{n,m}\, \psi_{n,m}(t) \ .$$

Considering the existence of a multi-resolution analysis produced by a scaling function $\varphi(t)$, we will implement the *Discrete Wavelet Transform* (DWT).

Since the approximation spaces V_j are getting larger as $j \to -\infty$ we can approximate $f(t)$ as closely as we desire, by choosing a small enough $j = -J\,(J > 0)$ and taking the projection of this function into V_{-J} using the basis functions $\{\varphi_{n,-J}(t)\}_{n\in\mathbb{Z}}$. The approximation coefficients are given by

$$a_0[n] = \langle f(t), \varphi_{n,-J}(t) \rangle$$

and we can use them to approximately recover $f(t)$,

$$f(t) \approx \sum_{n=-\infty}^{\infty} a_0[n] \varphi_{n,-J}(t) \ .$$

So, we can replace the original function $f(t)$ by its approximation in V_{-J} (or equivalently at **Level** 0), which is represented by the coefficients $a_0[n]$.

Using the property $V_{j-1} = V_j \oplus W_j$ we can further decompose the above approximation using

the basis functions $\{\varphi_{n,\,-J+1}(t)\}_{n\in\mathbb{Z}}$ in V_{-J+1} and $\{\psi_{n,\,-J+1}(t)\}_{n\in\mathbb{Z}}$ in W_{-J+1} obtaining

$$\begin{aligned} f(t) &= \sum_{n=-\infty}^{\infty} a_0[n]\,\varphi_{n,\,-J}(t) \\ &= \sum_{n=-\infty}^{\infty} a_1[n]\,\varphi_{n,\,-J+1}(t) + \sum_{n=-\infty}^{\infty} d_1[n]\,\psi_{n,-J+1}(t) \\ &= A_1(t) + D_1(T) \end{aligned}$$

where $A_1(t)$ is the approximation at **Level** 1 and $D_1(t)$ is the detail at the same level and the coefficients $a_1[n]$ and $d_1[n]$ the approximation and detail coefficients at **Level** 1, respectively. We can repeat the decomposition procedure for the approximation $A_j(t)$ at level j obtaining

$$\begin{aligned} f(t) &= A_1(t) + D_1(t) \\ &= A_2(t) + D_2(t) + D_1(t) \\ &= A_3(t) + D_3(t) + D_2(t) + D_1(t) \\ &= \cdots. \end{aligned}$$

With this procedure we go from a fine scale to a coarser scale.

We can verify [110, 73] that this decomposition procedure is a **filtering** followed by **downsampling** as Figure 3.1[1] displays.

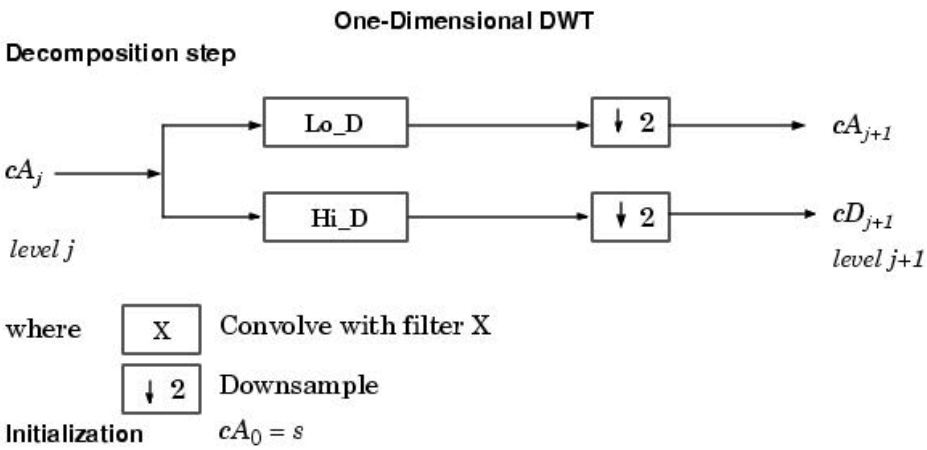


Figure 3.1: The filtering process of the 1-D DWT.

The above process can be reversed, that is, starting with two sets of coefficients $a_1[n], d_1[n]$ at scale level $(-J+1)$ we produce the coefficients $a_0[n]$ at scale level $-J$.

3.2.2 The 2-Dimensional Discrete Wavelet Transform

An extension of the 1-D DWT is available for the class of two-dimensional discrete functions. For instance, digital images belong to this class. In two dimensions, a two-dimensional scaling func-

[1]Obtained from http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet/dwt.html

tion, $\varphi(x,y)$, and three two-dimensional directionally sensitive wavelets, $\psi^H(x,y)$, $\psi^V(x,y)$, $\psi^D(x,y)$ are required. These functions can be written in a separable form in terms of a one-dimensional scaling function $\varphi(t)$ and its corresponding wavelet $\psi(t)$, as follows

$$\varphi(x,y) = \varphi(x)\varphi(y) \tag{3.25}$$

$$\psi^H(x,y) = \psi(x)\varphi(y) \tag{3.26}$$

$$\psi^V(x,y) = \varphi(x)\psi(y) \tag{3.27}$$

$$\psi^D(x,y) = \psi(x)\psi(y) \tag{3.28}$$

The directionally sensitive character of these wavelets, expresses their ability to measure functional variations (e.g. intensity or gray-level variations for images) along different directions. So, $\psi^H(x,y)$ measures variations along columns (e.g. horizontal edges), $\psi^V(x,y)$ localizes variations along rows (e.g. vertical edges) and $\psi^D(x,y)$ responds to variations along diagonals. Given the above separable functions, extension of the one-dimensional DWT to two dimensions is direct. In the 2-D case the basis functions at scale j are defined as follows [48]:

$$\varphi_{j,m,n}(x,y) = 2^{j/2}\varphi(2^j x - m, 2^j y - n) \tag{3.29}$$

$$\psi^i_{j,m,n}(x,y) = 2^{j/2}\psi^i(2^j x - m, 2^j y - n)\ , \qquad i \in \{H,V,D\} \tag{3.30}$$

Definition 3.2.3 (2-D Discrete Wavelet Transform) *The 2-D DWT of a function $f(x,y)$ of size $M \times N$ at a fixed scale j, is defined as follows:*

$$A_j(m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\varphi_{j,m,n}(x,y) \tag{3.31}$$

$$D^i_j(m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\psi^i_{j,m,n}(x,y)\ ,\ i \in \{H,V,D\} \tag{3.32}$$

where A_j denotes the set of approximation wavelet coefficients and D^i_j the set of detail coefficients across direction $i \in \{H,V,D\}$, at the given scale j (these four sets of coefficients are also called the **subbands** of the 2-D DWT for the given function $f(x,y)$). As in the 1-D DWT, we start from a small enough scale $j = -J(J > 0)$ (**Level** 0) and then we decompose the approximation in order to obtain the four sets of coefficients at the next scale $(-J+1)$ (**Level** 1). We iteratively repeat this decomposition process on the approximation coefficients at scale j (**Level** k), in order to calculate the four sets of coefficients at the next scale $(j+1)$ (**Level** (k+1)), as shown in Figure 3.2. The 2-D DWT is also invertible, that is, $f(x,y)$ can be recovered by adding the

details at the final approximation of the original function.

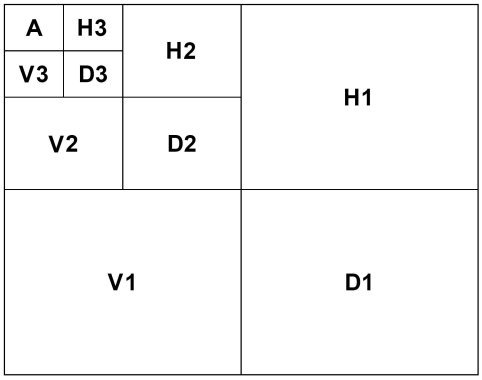


Figure 3.2: A 3-Level 2-D DWT.

As an example, Figure 3.3 displays the 3-level 2-D DWT using Daubechies' 4 filter (cf. section 3.2.4) for the image *Flowers.0006*.

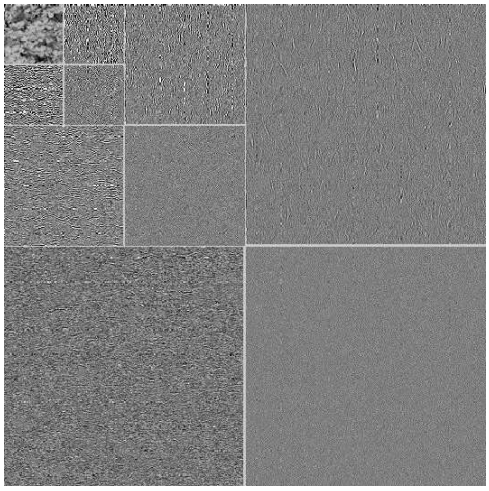


Figure 3.3: 3-level wavelet decomposition of the image *Flowers.0006* using 'db4' filter.

3.2.3 Steerable Pyramids

Although the 2-D DWT, in respect to a set of separable orthonormal wavelets, has become a very popular representation for multi-resolution image processing, we have to note some of its limitations. One of the major drawbacks of this representation is the lack of translation invariance. The content of wavelet subbands is unstable under transformations of the input 2-D signal, like translations, dilations and rotations. To overcome these problems, the class of the Shiftable Multi-Scale Transforms was introduced [105], which guarantee the translation invariance property (**shiftability**). Shiftability may be considered in different contexts, like "shiftability in position", "shiftability in orientation", "shiftability in scale" or even in a joint sense like "joint shiftability in 2-D position and orientation". An implementation of this class of transformations, the so called "Steerable Pyramid" [104], is a linear multi-scale, multi-orientation image decomposition which is defined in respect to a set of steerable basis functions.

Definition 3.2.4 (Steerable Basis) *A set of functions (filters) form a steerable basis if:*

(1) they are rotated copies of each other, and

(2) a copy of the function at any orientation may be computed as a linear combination of the basis functions.

An example of a steerable basis is a set of $N+1$ Nth-order directional derivative operators. The pyramid can be designed to produce any number of orientation subbands, k, more than the 3 fixed subbands obtained by applying the standard 2-D DWT. Figure 3.4 illustrates the spectral decomposition provided by a steerable pyramid.

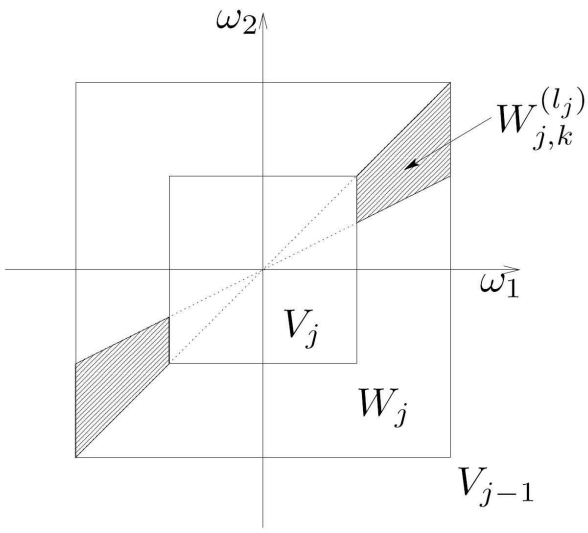


Figure 3.4: Spectral decomposition performed by a steerable pyramid.

The block diagram of a pyramid decomposition (both analysis and synthesis) is shown in Figure 3.5[2] and an illustration of a steerable pyramid is displayed in Figure 3.6[3].

3.2.4 Examples of wavelets

Two general classes of wavelet functions are the class of wavelets having closed form expressions and the class of those defined by iterative algorithms. The simplest example of a wavelet, expressed in closed-form, is the Haar wavelet given by

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2}, \\ -1, & \frac{1}{2} \leq t < 1, \\ 0, & otherwise. \end{cases}$$

[2]Obtained from [4].
[3]Obtained from [105].

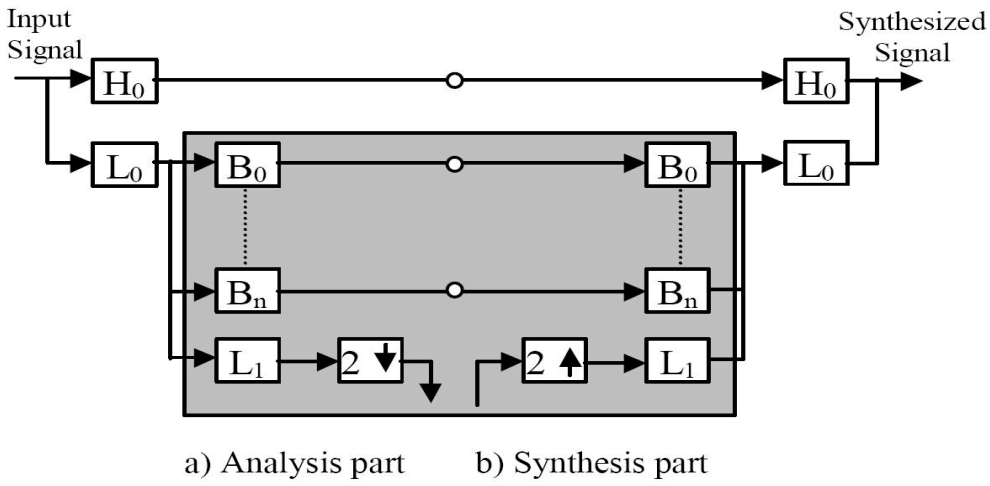


Figure 3.5: The filtering process implemented by a steerable pyramid.

It is compactly supported and the set $\psi_{n,m}(t) = 2^{-m/2}\,\psi(2^{-m}t-n)$, $n, m \in \mathbb{Z}$, is orthonormal (at a given scale, $\psi_{n,m}(t)$ and $\psi_{n',m}(t)$ have no common support). These symmetric basis functions are well localized in time, because their support is finite and as $m \to -\infty$ they are arbitrarily sharp in time, since the length goes to zero. A disadvantage is that the basis functions are not smooth, since they are not even continuous.

Another commonly used wavelet is the Morlet's wavelet, given by the formula

$$\psi(t) = e^{-t^2/2}\,\cos 5t$$

The corresponding set of basis functions $\psi_{n,m}(t)$ is not orthogonal and it has not compact support. We usually say that it has an essential support of $[-4, 4]$. This wavelet has no scaling function, but we use it because it has an explicit formula. It is symmetric and we use it in the continuous wavelet transform.

Daubechies' wavelets belong to the class of orthonormal wavelets with compact support which are constructed iteratively. They are defined by an iterative algorithm for computing function values dyadically using filter coefficients. The scaling function $\varphi(t)$, which is computed first, is given by

$$\varphi(t) = \lim_{k\to\infty} \varphi_k(t)\ ,$$

where $\varphi_k(t)$ is defined iteratively by

$$\begin{aligned}\varphi_0(t) &= \chi_{[-1/2,\,1/2]}\ ,\\ \varphi_k(t) &= \sqrt{2}\sum_n \alpha[n]\,\varphi_{k-1}(2t-n)\ ,\end{aligned}$$

where $\chi_{[-1/2,\,1/2]}$ is the characteristic function and $\alpha[n]$ are the filter coefficients. Then the wavelet function $\psi(t)$ with N vanishing moments is computed from the associated scaling function $\varphi_N(t)$

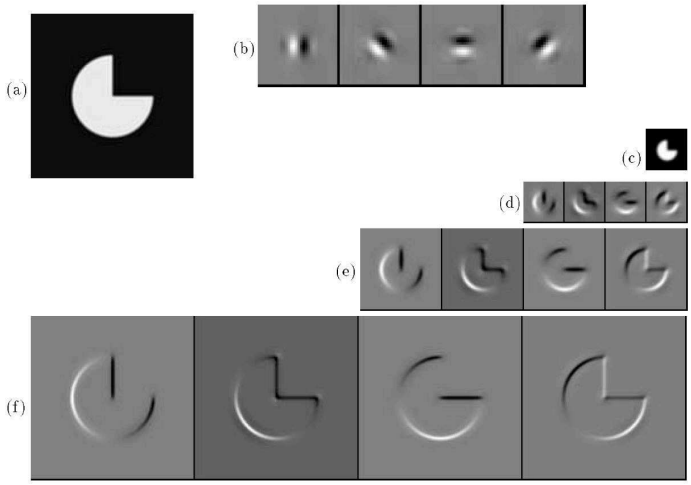


Figure 3.6: Illustration of the structure of the steerable pyramid. (a) Test image, (b) Filter kernels used to construct the pyramid (These four filters which are rotated versions of a common filter, span the space of all rotations of that common filter), (c) Lowpass coefficient image, (d)-(f) Bandpass coefficient images.

as follows

$$\psi(t) = \sum_{n=0}^{2N-1} (-1)^n \, \alpha_N[-n+1] \, \varphi_N(2t-n) \ .$$

These wavelets have a relatively small support of width $2N-1$, that is, $\alpha_n \neq 0$ only for $n = 1, 2, ..., 2N-1$. Daubechies' wavelets are not symmetric. Figure 3.7 shows some members of the Daubechies' wavelet family.

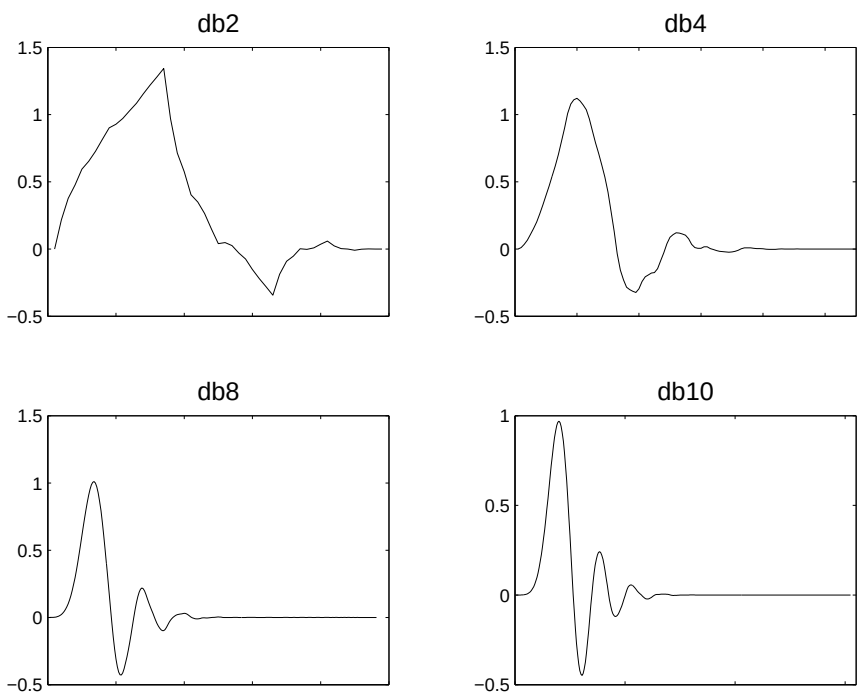


Figure 3.7: Decomposition High-pass filters of the Daubechies' family.

Note: A function $f(t)$ has N vanishing moments if

$$\int t^k \, f(t) \, \mathrm{d}t = 0 \ , \quad k = 0, ..., N-1 \ .$$

In the two-dimensional case, we can easily construct the four functions (one scaling function and three wavelets) due to their separable nature (Eqs. 3.25-3.28). Figure 3.8 displays these functions corresponding to the Daubechies' 4 ('db4') 1-D wavelet.

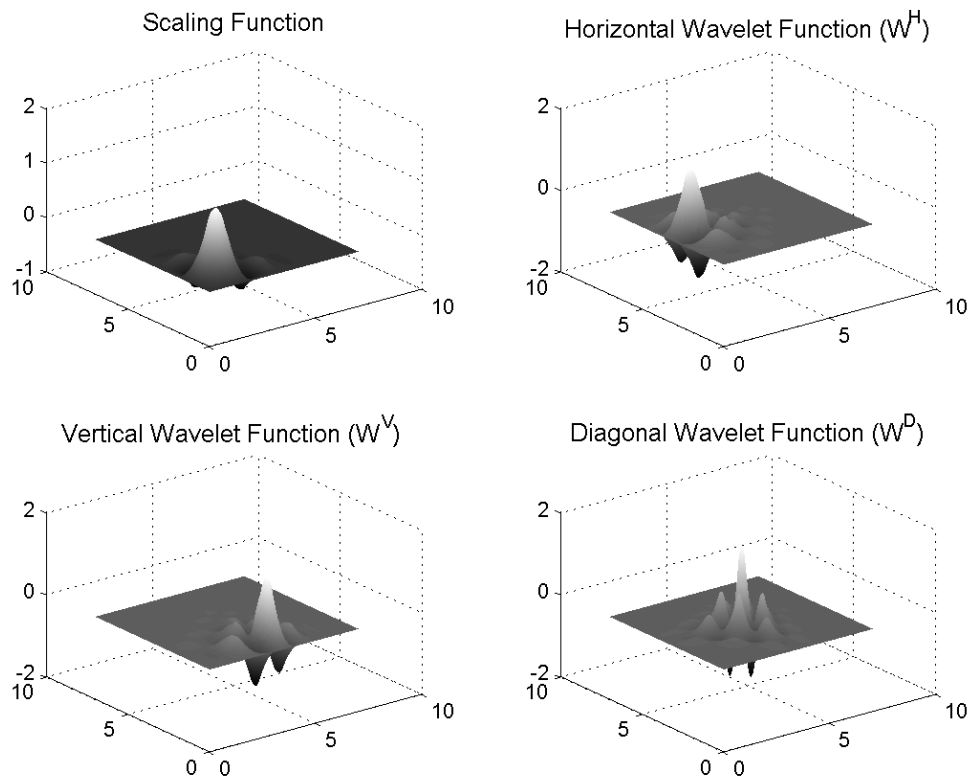


Figure 3.8: 2-D scaling and wavelet functions corresponding to the 1-D 'db4' wavelet.

3.3 Statistical properties of the DWT coefficients

In different applications, such as signal compression, it is often a very important issue to achieve high correlation between the specific components of the application at hand, usually the coefficients of a properly chosen transform. In the present case, we are interested in the correlation properties of the wavelet transform coefficients, and more specifically the wavelet coefficients of a 2-D DWT.

As it was mentioned in subsection (3.2.2), the application of a 2-D DWT to a 2-D discrete function (e.g. a digital image) is equivalent to a cascade implementation of two 1-D DWT's. After the wavelet transform is applied along one direction, the transform coefficients of different columns (or rows), but with the same dilation and translation, will form a (prediction) channel. The de-correlating capability of the DWT can be verified through evaluating the auto- and

cross-correlation of the wavelet coefficients among different channels. A detailed study on the correlation properties of these coefficients [55] shows that the statistical expectation of the approximation coefficients is proportional to the average of the original signal, whereas the corresponding expectation of the detail coefficients is almost zero. It is also proved that the autocorrelations of the approximation and detail coefficients are proportional to the autocorrelation of the original signal. On the other hand, the cross-correlation between approximation and detail coefficients at the same resolution level is equal to zero, which means that these coefficients are de-correlated. The detail coefficients at different resolution levels are also de-correlated.

The above observations concerned the correlation properties of the raw wavelet coefficients. A number of methods developed in the area of image processing, exploit the high correlation of the magnitudes of pairs of wavelet coefficients [103, 13]. This correlation is mainly due to image features, such as lines, edges and corners and arises between spatial neighbors and also between coefficients corresponding to adjacent scales and orientations. Also, the coefficient magnitudes are correlated, even when the raw coefficients are uncorrelated.

Chapter 4

Alpha-Stable Models

4.1 Introduction

The signal processing methods have traditionally been dominated by the Gaussian assumption. In many applications the Gaussian model is justified by the Central Limit Theorem. It has also been accepted as a useful tool, as it often leads to analytical solutions in many signal processing problems.

However, there are many phenomena in areas such as physics [81, 70], engineering [100, 99] and finance [76, 9, 39, 26], which are undoubtedly non-Gaussian. The Gaussian assumption may results in significant performance degradation for systems operating in a non-Gaussian environment. The common characteristic of phenomena belonging in the non-Gaussian class, is their impulsive nature. This fact is expressed by the heavy-tailed behavior of their density functions, which means that the probability of extremely large observations is non-negligible. As a result, the rate of decay in the tails of the distributions is less rapid than in the Gaussian case.

This heavy-tailed behavior of the distribution functions is described by the stable law, which is a direct generalization of the Gaussian distribution and includes the Gaussian as a limiting case. Stable family belongs to the class of heavy-tailed distributions. This characteristic of the stable distributions is one of the main reasons why they are suitable for modelling signals of impulsive nature.

The class of stable distributions was first characterized by Paul Lévy [64]. Since then, significant progress has been made in developing theory for stable processes [17, 97, 86, 16, 18, 80, 15, 14, 130, 98, 22, 36, 35]. Applications of the stable distributions in the fields of signal processing and communications can be also found in [101, 122, 123, 121, 69, 95, 131, 46, 114, 119, 116, 118, 117, 113, 115].

This chapter introduces the basic concepts of a statistical model based on a sub-class of the stable family, namely, the class of *Symmetric-α-Stable* ($S\alpha S$) *distributions*.

4.2 The family of Stable Distributions

4.2.1 Univariate $S\alpha S$ Distributions

The Symmetric-α-Stable ($S\alpha S$) distribution is best defined by its characteristic function as follows:

$$\phi(t) = \exp(\imath\delta t - \gamma|t|^{\alpha}) \ , \tag{4.1}$$

where α is the *characteristic exponent*, taking values $0 < \alpha \leq 2$, δ ($-\infty < \delta < \infty$) is the *location parameter*, and γ ($\gamma > 0$) is the *dispersion* of the distribution. The characteristic exponent is a shape parameter which controls the "thickness" of the tails of the density function. The smaller the α is, the heavier the tails of the $S\alpha S$ density function. The dispersion parameter determines the spread of the distribution around its location parameter, similar to the variance of the Gaussian. When $1 < \alpha \leq 2$, the location parameter δ equals the mean of the $S\alpha S$ distribution, while for $0 < \alpha \leq 1$, δ corresponds to its median.

A $S\alpha S$ distribution is called *standard* if $\delta = 0$, $\gamma = 1$. For the parameterization of a $S\alpha S$ random variable X, corresponding to the characteristic function of Eq. (4.1), it is true that if X follows a $S\alpha S$ distribution with parameters α, γ, δ ($X \sim f_{\alpha}(\gamma, \delta)$), then the variable $(X-\delta)/\gamma^{1/\alpha}$ is standard with characteristic exponent α. Examples of standard $S\alpha S$ density functions for different values of α are shown in Figure 4.1.

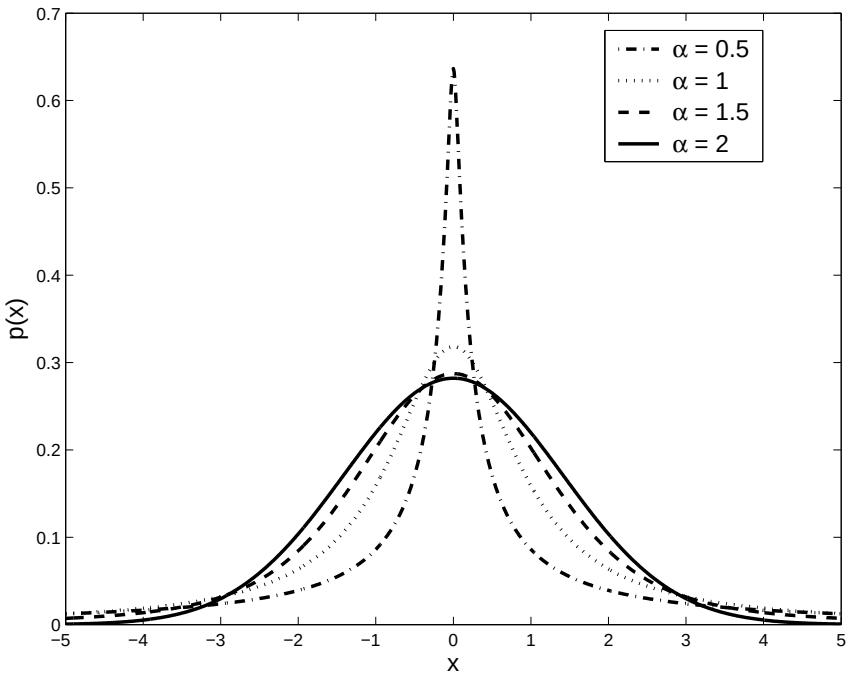


Figure 4.1: Standard $S\alpha S$ density functions for different values of the characteristic exponent α, for the parameterization (4.2).

There are multiple parameterizations of the general univariate stable distributions which are

useful for different problems. Two of them are the following [88]:

$$\phi(t) = \exp(\imath\delta t - \gamma^\alpha |t|^\alpha) \ , \tag{4.2}$$

$$\phi(t) = \exp(\imath\delta t - \frac{1}{\alpha}\,\gamma_*^\alpha |t|^\alpha) \ , \tag{4.3}$$

with $\gamma_* = \alpha^{1/\alpha}\gamma$. Parameterization (4.3) may be the most intuitive parameterization for users in applied fields, as it has a number of interesting properties [88]: for instance, the mode of a $S\alpha S$ density following this parameterization is at δ, and in the Gaussian case ($\alpha = 2$) γ_* is the standard deviation.

In general, no closed-form expressions exist for most density and distribution functions. Two important special cases of $S\alpha S$ densities with closed-form expressions are the Gaussian ($\alpha = 2$) and the Cauchy ($\alpha = 1$). Using parameterization (4.1) we have:

Gaussian

$$f_2(\gamma,\, \delta;\, x) = \frac{1}{\sqrt{4\pi\gamma}}\, e^{-\frac{(x-\delta)^2}{4\gamma}} \tag{4.4}$$

Cauchy

$$f_1(\gamma,\, \delta;\, x) = \frac{1}{\pi}\frac{\gamma}{\gamma^2 + (x-\delta)^2} \tag{4.5}$$

Note: Using parameterization (4.1), when $\alpha = 2$ (Gaussian) the dispersion γ is the half of the variance.

However, power series expansions of stable density functions are available. In particular, the standard $S\alpha S$ density function is given by [102]

$$f_\alpha(x) = \begin{cases} \frac{1}{\pi x}\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!}\,\Gamma(\alpha k+1)\, x^{-\alpha k}\, \sin(\frac{k\alpha\pi}{2}), & 0 < \alpha < 1 \\ \frac{1}{\pi(x^2+1)}, & \alpha = 1 \\ \frac{1}{\pi\alpha}\sum_{k=0}^{\infty} \frac{(-1)^k}{2k!}\,\Gamma\big(\frac{2k+1}{\alpha}\big)\, x^{2k}, & 1 < \alpha < 2 \\ \frac{1}{2\sqrt{\pi}}\, e^{-\frac{x^2}{4}}, & \alpha = 2 \end{cases} \tag{4.6}$$

where $\Gamma(\cdot)$ is the Gamma function defined by

$$\Gamma(x) = \int_0^\infty t^{x-1}\, e^{-t}\, \mathrm{d}t \ . \tag{4.7}$$

$S\alpha S$ densities have many common features with the Gaussian, they are smooth, unimodal, symmetric with respect to the median and bell-shaped. However, unlike the Gaussian density which has exponential tails, the stable densities have tails following an algebraic rate of decay which makes them heavier than the Gaussian tails, as shown in Figure 4.2.

Figure 4.3 displays some simulated $S\alpha S$ time series, demonstrating the fact that random variables following $S\alpha S$ distributions with small α values are highly impulsive.

Two of the most important properties of the stable distribution are the *stability property* and the *Generalized Central Limit Theorem.*

Stability property: if X_1, X_2, ..., X_n are independent following stable distributions with the

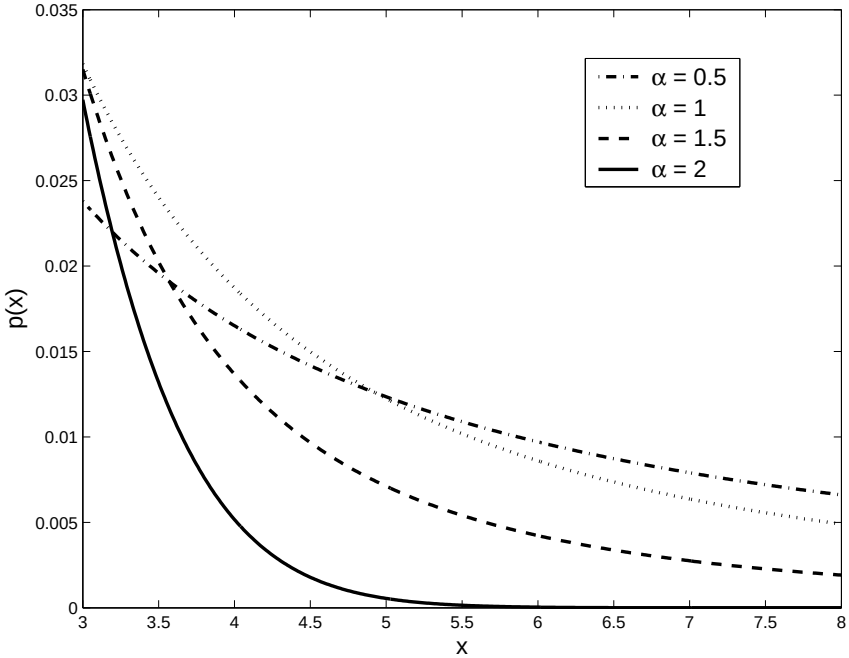


Figure 4.2: Tails of the standard $S\alpha S$ density functions for different values of the characteristic exponent α, for the parameterization (4.2).

same α, then all linear combinations $\sum_{k=1}^{n} a_k X_k$, for arbitrary constants a_k, are stable with the same *alpha*.

Generalized Central Limit Theorem: X is stable if and only if it is the limit in distribution of sums of the form

$$\frac{X_1 + \ldots + X_n}{a_n} - b_n$$

where X_1, X_2, ... is a sequence of i.i.d. random variables, and $\{a_n\}$, $\{b_n\}$ are sequences of positive and real numbers, respectively.

An important characteristic of the $S\alpha S$ distributions is the non-existence of the second-order moment, except for the Gaussian case ($\alpha = 2$). Instead, all moments of order p less than α do exist and are called the *Fractional Lower Order Moments* (FLOM's). The following theorem holds [97]

Theorem 4.2.1 *The FLOM's of a $S\alpha S$ random variable X, following parameterization (4.2) with zero location parameter and dispersion γ, are given by*

$$E\{|X|^p\} = \big(C(p,\alpha)\cdot\gamma\big)^p \qquad for\ 0 < p < \alpha \tag{4.8}$$

where

$$\big(C(p,\alpha)\big)^p = \frac{2^{p+1}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(-\frac{p}{\alpha}\right)}{\alpha\sqrt{\pi}\,\Gamma\left(-\frac{p}{2}\right)} = \frac{\Gamma\big(1-\frac{p}{\alpha}\big)}{\cos\big(\frac{\pi}{2}p\big)\Gamma(1-p)} \tag{4.9}$$

The practical implementation of a stable model is a non-trivial task due to the lack of analytical expressions for its probability density and distribution functions. DuMouchel [36] developed

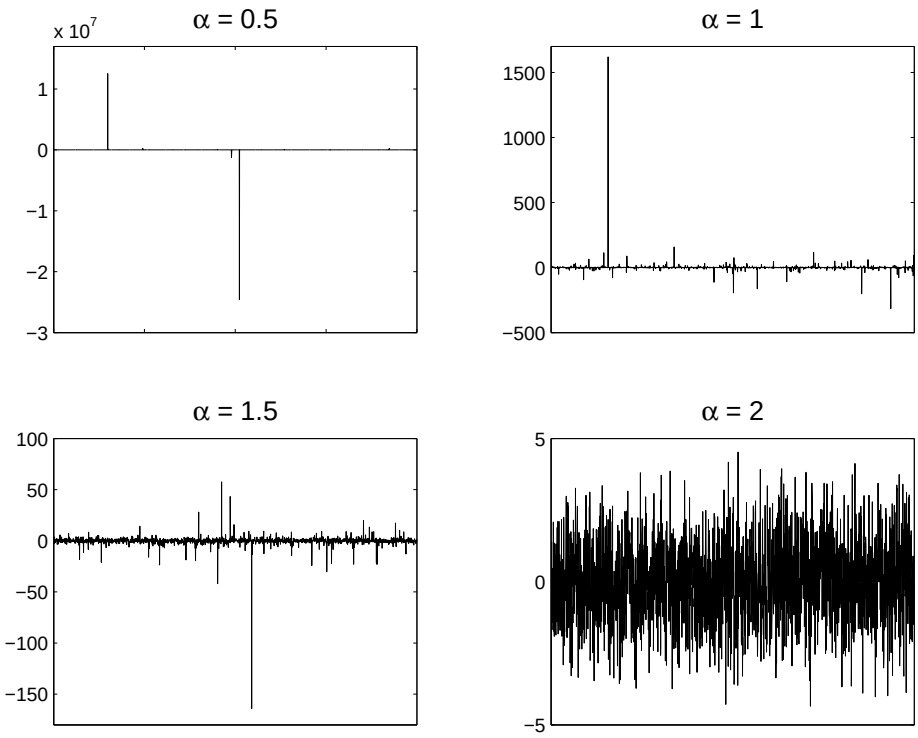


Figure 4.3: Simulated $S\alpha S$ sequences for different values of the characteristic exponent.

a (computationally intensive) procedure for approximating stable densities and distribution functions. Nolan [87] improved the approximation procedure by developing methods which make it possible to compute densities, distribution functions and quantiles for stable distributions quickly. An FFT-based algorithm for calculating the asymmetric stable distribution and density function can be found in [84] while a polynomial-based approximation procedure is described in [83].

A very important task in designing algorithms based on the $S\alpha S$ models, is the parameter estimation, that is, the estimation of the three parameters (α, γ, δ) describing the $S\alpha S$ distribution. Most of the conventional methods in mathematical statistics, such as the Maximum Likelihood (ML) estimation method, can not be used in the stable case, since these methods depend on an explicit form for the density. Nevertheless, there are suboptimal numerical methods that have been found useful and efficient in practical applications. Among them, there are methods based on sample quantiles [40] and sample characteristic functions [89, 61, 59, 62]. There are also Maximum Likelihood methods based on an approximation to the likelihood function [11, 87, 124, 45].

In the following chapters we assume a $S\alpha S$ distribution located around the origin (i.e. $\delta = 0$) and we estimate the $S\alpha S$ model parameters (α, γ) using the consistent Maximum Likelihood (ML) method described by Nolan [87], which gives reliable estimates and provides the most tight confidence intervals.

4.2.2 Multivariate Stable Distributions

Unlike the family of univariate stable distributions, the family of multivariate stable distributions forms a non-parametric set (except for the multivariate Gaussian case ($\alpha = 2$)). The definition

of the multivariate stable distribution is as follows:

Definition 4.2.1 *An n-dimensional distribution function $F(\vec{x})$, $\vec{x} \in \mathbb{R}^n$, is called stable if, for any i.i.d. random vectors $\vec{X}_1$, $\vec{X}_2$ with distribution function $F(\vec{x})$ and arbitrary constants a_1, a_2, there exist $a \in \mathbb{R}$, $\vec{b} \in \mathbb{R}^n$ and a random vector $\vec{X}$ with the same distribution function $F(\vec{x})$ such that*

$$a_1\vec{X}_1 + a_2\vec{X}_2 \stackrel{d}{=} a\vec{X} + \vec{b}$$

where $\stackrel{d}{=}$ means equality in distribution.

A multivariate stable distribution is determined by a location vector $\vec{\delta} \in \mathbb{R}^n$, the characteristic exponent $0 < \alpha \leq 2$ and a finite measure $\mu(\mathrm{d}\vec{s})$ on the unit sphere S_n of $\mathbb{R}^n$ [97]. In the multivariate case, a real random vector $\vec{X} = (X_1, \ldots, X_n)$ is $S\alpha S$, or the real random variables $X_1, \ldots, X_n$ are jointly $S\alpha S$, if their joint characteristic function is of the form

$$\phi(\vec{\theta}\,) = \exp\{\imath\vec{\theta}^{\,T}\vec{\delta} - \int_{S_n} |\,\vec{\theta}^{\,T}\vec{s}\,|^{\alpha}\,\mu(\mathrm{d}\vec{s})\} \tag{4.10}$$

with T denoting a transpose and where the spectral measure $\mu(\cdot)$ is symmetric, that is, $\mu(A) = \mu(-A)$ for any measurable set A on S_n.

A very important difference between multivariate Gaussian and stable distributions is that a stable random vector can not be whitened. Specifically, it is known that if $\vec{X}$ is a Gaussian vector, then it can be written as

$$\vec{X} = \mathbf{A}\vec{Y}$$

where $\mathbf{A}$ is a constant matrix and $\vec{Y}$ is a Gaussian random vector with independent components. In the stable case, it is shown that representation of even two stable variables with characteristic exponent $0 < \alpha < 2$ as the linear combination of a finite number of independent stable variables of the same α is impossible [98].

Sub-Gaussian random vectors

An important sub-class of multivariate $S\alpha S$ random vectors is the class of the so-called *α-sub-Gaussian* random vectors [97].

Definition 4.2.2 *Any vector $\vec{X}$ distributed as*

$$\vec{X} = A^{1/2}\,\vec{G} \tag{4.11}$$

where A is a positive $\frac{\alpha}{2}$-stable random variable and $\vec{G} = (G_1, G_2, \ldots, G_n)$ is zero-mean Gaussian random vector, independent of A, with covariance matrix $\boldsymbol{R}$, is called sub-Gaussian $S\alpha S$ random vector in $\mathbb{R}^n$ with underlying Gaussian vector $\vec{G}$.

This sub-class is often denoted by α-SG($\mathbf{R}$). An advantage of the sub-Gaussian $S\alpha S$ random vectors as a modelling tool, is the simple analytical expression of the corresponding characteristic

function. The following proposition holds:

Proposition 4.2.1 *A sub-Gaussian $S\alpha S$ random vector $\vec{X}$, as defined above, has characteristic function*

$$\begin{aligned}\phi(\vec{\theta}) &= \phi(\theta_1,\ldots,\theta_n) = E\{e^{\imath\langle\vec{\theta},\,\vec{X}\rangle}\} \\ &= E\{e^{\imath\sum_{k=1}^{n}\theta_k X_k}\} = \exp\{-|\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\theta_i\theta_j\boldsymbol{R}_{ij}|^{\alpha/2}\}\end{aligned} \tag{4.12}$$

where $\boldsymbol{R}_{ij} = E\{G_iG_j\}$, $i,j = 1,\ldots,n$, are the covariances of the underlying Gaussian random vector $(G_1,\ldots,G_n)$

where $\langle\cdot,\cdot\rangle$ denotes the inner product of two vectors. We also note that in the zero-mean case the covariance of two random variables equals their correlation, so the $n\times n$ matrix $\mathbf{R}$ can be also viewed as the correlation matrix of the underlying vector $\vec{G}$. Eq. (4.12) can be rewritten as follows

$$\phi(\vec{\theta}) = \exp\{-|\frac{1}{2}\,\vec{\theta}^{\,T}\mathbf{R}\vec{\theta}\,|^{\alpha/2}\}$$

and taking into account that the matrix $\mathbf{R}$ is non-negative definite, we can further simplify it as

$$\phi(\vec{\theta}) = \exp\{-(\frac{1}{2}\,\vec{\theta}^{\,T}\mathbf{R}\vec{\theta}\,)^{\alpha/2}\}\ . \tag{4.13}$$

The above proposition results in the next corollary [97]:

Corollary 4.2.1 *Let $\vec{X}$ be sub-Gaussian $S\alpha S$ random vector with underlying Gaussian vector $\vec{G}$. Then there is a one-to-one correspondence between the probability distribution of $\vec{G}$ and that of $\vec{X}$.*

It is shown [97] that a vector with independent non-zero components cannot be sub-Gaussian. The components of a sub-Gaussian $S\alpha S$ random vector are, in fact, strongly dependent.

4.2.3 Symmetric Stable Processes

A $S\alpha S$ stochastic process is defined as follows:

Definition 4.2.3 *A collection of random variables $\{X(t),\, t\in T\}$, where T is an arbitrary index set, is said to be a $S\alpha S$ stochastic process if for all $n\geq 1$ and distinct indices $t_1,\ldots,t_n\in T$, the random variables $X(t_1),\ldots,X(t_n)$ are jointly $S\alpha S$ with the same characteristic exponent α.*

The following theorem holds:

Theorem 4.2.2 *Let $\{X(t),\, t\in T\}$ be a stochastic process. Then, it is $S\alpha S$ if and only if all linear combinations*

$$\sum_{k=1}^{n} b_k X(t_k),\ n\geq 1,\ t_1,\ldots,t_n\in T,\ b_1,\ldots,b_n\in\mathbb{R}$$

are $S\alpha S$.

One of the basic types of stable processes that are commonly used in practice, is the type of sub-Gaussian processes.

Definition 4.2.4 *A stable process $\{X(t), t \in T\}$ is said to be an α-sub-Gaussian process (α-SG(R)), if for all $n \geq 1$ and distinct indices $t_1, \ldots, t_n \in T$, the random vector $(X(t_1), \ldots, X(t_n))$ has characteristic function given by*

$$\phi(\vec{\theta}\,) = \exp\{-(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\theta_i\theta_j R(t_i, t_j))^{\alpha/2}\} \tag{4.14}$$

where $R(u, v)$ is a positive definite function, and $1 < \alpha \leq 2$.

When $\alpha = 2$, $X(t)$ is a zero-mean Gaussian process with covariance function $R(u, v)$. A sub-Gaussian process is *stationary* if and only if $R(u, v) = R(u - v) = R(v - u)$. Sub-Gaussian processes can be also viewed as variance mixtures of Gaussian processes [16]. In particular, if $X(t)$ is α-SG(R), then it can be expressed as

$$X(t) = A^{1/2}\, G(t) \tag{4.15}$$

where A is a positive $\frac{\alpha}{2}$-stable random variable and $G(t)$ a zero-mean Gaussian process, independent of A, with covariance function $R(u, v)$. Note again, that one of the important properties about sub-Gaussian random variables is that they cannot be independent.

4.2.4 Covariations

The concept of covariance plays a fundamental role in the second-order moment theory. As it was mentioned above, covariances do not exist in the family of $S\alpha S$ random variables, due to their infinite variance. Instead, a quantity called *covariation*, which under certain constraints plays an analogous role for $S\alpha S$ random variables to the one played by covariance for Gaussian random variables, has been proposed.

Definition 4.2.5 *Let X and Y be jointly $S\alpha S$ random variables with $1 < \alpha \leq 2$, zero location parameters and dispersions γ_X and γ_Y respectively. Then for all $1 < p < \alpha$, the covariation of X with Y is defined by*

$$[X, Y]_\alpha = \frac{E\{XY^{<p-1>}\}}{E\{|Y|^p\}}\|Y\|_\alpha^\alpha \tag{4.16}$$

where for any complex number z and $a \geq 0$ we use the notation

$$z^{<a>} = |z|^{a-1}\bar{z}$$

with $\bar{z}$ denoting complex conjugation. The symbol $\|Y\|_\alpha$ denotes the scale parameter of Y and is explained further down.

Some of the useful properties of the covariation are the following:

P(1) (Linearity in the first argument): If X_1, X_2, Y are jointly $S\alpha S$ then

$$[aX_1 + bX_2, Y]_\alpha = a[X_1, Y]_\alpha + b[X_2, Y]_\alpha \tag{4.17}$$

for any constants $a, b \in \mathbb{R}$.

P(2) (Scaling): If X, Y are jointly $S\alpha S$ and $a, b \in \mathbb{R}$, then

$$[aX, bY]_\alpha = ab^{<\alpha-1>}[X, Y]_\alpha \tag{4.18}$$

P(3) (Pseudo-linearity in the second argument): If X, Y_1, Y_2 are jointly $S\alpha S$ and Y_1, Y_2 are independent, then

$$[X, aY_1 + bY_2]_\alpha = a^{<\alpha-1>}[X, Y_1]_\alpha + b^{<\alpha-1>}[X, Y_2]_\alpha \tag{4.19}$$

for any constants $a, b \in \mathbb{R}$.

P(4): If X, Y are jointly $S\alpha S$ and independent, then

$$[X, Y]_\alpha = 0 \ .$$

In general, the converse is not true (for example, let $(X, Y) = (A^{1/2}G_1, A^{1/2}G_2)$ be sub-Gaussian vector with G_1, G_2 i.i.d. standard normal random variables. Then, $[X, Y]_\alpha = [Y, X]_\alpha = 0$, but X and Y are dependent).

P(5): Let U, V be i.i.d. $S\alpha S$ random variables and a_i, b_i $(i = 1, 2)$ real constants. For the $S\alpha S$ random variables X, Y defined as $X = a_1U + b_1V$, $Y = a_2U + b_2V$ we have

$$[X, Y]_\alpha = \gamma_U \, a_1 a_2^{<\alpha-1>} + \gamma_V \, b_1 b_2^{<\alpha-1>} \tag{4.20}$$

The *covariation coefficient* of X with Y, for all $1 < p < \alpha$, is defined by

$$\lambda_{XY} = \frac{[X, Y]_\alpha}{[Y, Y]_\alpha} = \frac{E\{XY^{<p-1>}\}}{E\{|Y|^p\}} \tag{4.21}$$

Note the asymmetric nature of the covariation and the covariation coefficient. However, we can define a *symmetric covariation coefficient* as follows:

$$Corr_\alpha(X, Y) = \lambda_{XY}\lambda_{YX} = \frac{[X, Y]_\alpha \, [Y, X]_\alpha}{[Y, Y]_\alpha \, [X, X]_\alpha} \tag{4.22}$$

Let $\mathcal{S}_\alpha$ be a linear space of jointly $S\alpha S$ random variables. When $1 < \alpha \leq 2$, the covariation induces a norm on $\mathcal{S}_\alpha$.

Definition 4.2.6 *The covariation norm of $X \in \mathcal{S}_\alpha$, $1 < \alpha \leq 2$, is defined by*

$$\|X\|_\alpha = ([X, X]_\alpha)^{1/\alpha}$$

Depending on the parameterization of the characteristic function of a $S\alpha S$ distribution, the following proposition holds:

Proposition 4.2.2 *If $X \sim f_\alpha(\gamma, 0)$ with $\alpha > 1$, then*

$$\|X\|_\alpha = \gamma^{1/\alpha} \ , \quad for\ parameterization\ (4.1)$$

and

$$\|X\|_\alpha = \gamma \ , \quad for\ parameterization\ (4.2)$$

For any two jointly $S\alpha S$ random variables X, Y (with $1 < \alpha \leq 2$), the covariation satisfies the following "Cauchy-Schwartz" inequality:

$$|[X, Y]_\alpha| \leq \|X\|_\alpha \|Y\|_\alpha^{\alpha-1} \ .$$

Estimation of Covariations

In general, it is difficult to find analytical expressions for the covariations (or covariation coefficients). In practical applications it is important to have unbiased and efficient estimators of these quantities. The covariation coefficient λ_{XY} was defined as a fraction of the covariation $[X, Y]_\alpha$ and the scale parameter $[Y, Y]_\alpha$, which can be estimated by the methods described above. So, if we have estimated the covariation coefficient λ_{XY}, we can find an estimation of $[X, Y]_\alpha$ by multiplying λ_{XY} and $[Y, Y]_\alpha$.

Focusing our attention on estimating the covariation coefficient λ_{XY}, we choose two methods that gave promising results by employing Monte-Carlo simulations. The first method, called the *Fractional Lower Order Moment (FLOM) Estimator*, is very simple and computationally efficient and gives very good results especially when α is close to 2. The second method, called the *Screened Ratio (SR) Estimator*, exhibits a better performance for small values of α and is strongly consistent.

In the following, these two estimators are given under the assumption of jointly $S\alpha S$ random variables X, Y with $\alpha > 1$, and a set of *independent* observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ [102].

FLOM Estimator: For some $1 \leq p < \alpha$, the estimation of λ_{XY} is given by

$$\hat{\lambda}_{FLOM} = \frac{\sum_{i=1}^{n} X_i |Y_i|^{p-1}\, sign(Y_i)}{\sum_{i=1}^{n} |Y_i|^p} \tag{4.23}$$

where

$$sign(x) = \begin{cases} 1 & for\ x > 0 \\ 0 & for\ x = 0 \\ -1 & for\ x < 0 \end{cases} \tag{4.24}$$

Screened Ratio Estimator: For arbitrary constants c_1, c_2, with $0 < c_1 < c_2 \leq \infty$ the SR estimator of λ_{XY} is given by

$$\hat{\lambda}_{SR} = \frac{\sum_{i=1}^{n} X_i Y_i^{-1} \chi_{Y_i}}{\sum_{i=1}^{n} \chi_{Y_i}} \tag{4.25}$$

with the random variable χ_Y defined as follows:

$$\chi_Y = \begin{cases} 1 & if\ c_1 < |Y| < c_2 \\ 0 & otherwise. \end{cases}$$

We usually choose $c_2 = \infty$ and c_1 be a relatively "small" number.

In subsequent chapters, we will be particularly interested in estimating covariations between sub-Gaussian random variables. Consider the sub-Gaussian random vector $\vec{X} = A^{1/2}\,\vec{G}$, where A is a positive $\frac{\alpha}{2}$-stable random variable and $\vec{G} = (G_1, G_2, \ldots, G_n)$ is zero-mean Gaussian random vector, independent of A, with covariance matrix $\mathbf{R}$. Then, the covariations $[X_i, X_j]_\alpha$, $i, j = 1, \ldots, n$ are given by [97]:

$$[X_i, X_j]_\alpha = 2^{-\frac{\alpha}{2}}\,[\mathbf{R}]_{ij}\,[\mathbf{R}]_{jj}^{\frac{(\alpha-2)}{2}} \tag{4.26}$$

(covariation is symmetric, i.e. $[X_i, X_j]_\alpha = [X_j, X_i]_\alpha$ if $[\mathbf{R}]_{ii} = [\mathbf{R}]_{jj}$.)

In the following, we will estimate these covariations using parameterization (4.2) of the characteristic function. Then, by combining Eq. (4.16) and proposition (4.2.2) we obtain

$$[X, Y]_\alpha = \frac{E\{XY^{<p-1>}\}}{E\{|Y|^p\}}\,\gamma_Y^\alpha\ . \tag{4.27}$$

Besides, by applying theorem (4.2.1) on Y we have

$$\gamma_Y = \frac{\left(E\{|Y|^p\}\right)^{1/p}}{C(p,\alpha)}\ . \tag{4.28}$$

Substitution of Eq. (4.28) into Eq. (4.27) results in the following relations:

$$\begin{aligned} [X, Y]_\alpha &= \frac{E\{XY^{<p-1>}\}}{E\{|Y|^p\}} \cdot \frac{\left(E\{|Y|^p\}\right)^{\alpha/p}}{\left(C(p,\alpha)\right)^\alpha} = \\ &= \left(C(p,\alpha)\right)^{-\alpha} \cdot E\{XY^{<p-1>}\} \cdot \left(E\{|Y|^p\}\right)^{\frac{\alpha}{p}-1} \end{aligned} \tag{4.29}$$

where the constant $C(p, \alpha)$ is given by Eq. (4.9).

Let the vectors $\{\vec{X}^1, \vec{X}^2, \ldots, \vec{X}^N\}$ constitute a set of N independent realizations of a sub-Gaussian (α-SG($\mathbf{R}$)) process, where $\vec{X}^k = (X_1^k, X_2^k, \ldots, X_n^k)$, $k = 1, \ldots, N$. Then, Eq. (4.29)

results in the following estimator of the covariations:

$$\hat{\mathbf{c}}_{ij} = [\widehat{X_i, X_j}]_\alpha = (C(p,\alpha))^{-\frac{\alpha}{p}} \left[\frac{1}{N}\sum_{k=1}^{N} \vec{X_i^k}\,(\vec{X_j^k})^{<p-1>}\right]\left[\frac{1}{N}\sum_{k=1}^{N} |\vec{X_j^k}|^p\right]^{\frac{\alpha}{p}-1} \tag{4.30}$$

where $\vec{X_i^k}$ denotes the i-th component of the vector $\vec{X^k}$. Another covariation estimator is obtained from the combination of Eq. (4.23) with Eq. (4.27) as follows

$$\hat{\mathbf{c}}_{ij}^{FLOM} = \frac{\sum_{k=1}^{N} \vec{X_i^k}\,|\vec{X_j^k}|^{p-1}\, sign(\vec{X_j^k})}{\sum_{k=1}^{N} |\vec{X_j^k}|^p}\,\gamma_{X_j}^{\alpha} \tag{4.31}$$

where we assume that the dispersion γ_{X_j} has been estimated using the ML estimator. In the stable case, we define the *covariation matrix* $\mathbf{C}$ with elements the covariations. Then, the estimated covariation matrix $\hat{\mathbf{C}}$, is the matrix with elements $[\hat{\mathbf{C}}]_{ij} = \hat{\mathbf{c}}_{ij}$ (or $\hat{\mathbf{c}}_{ij}^{FLOM}$).

Now, we can estimate the elements $[\mathbf{R}]_{ij}$ of the underlying covariance matrix using Eq. (4.26), resulting in the following estimators:

$$[\hat{\mathbf{R}}]_{jj} = \left(2^{\frac{\alpha}{2}}\,[\hat{\mathbf{C}}]_{jj}\right)^{\frac{2}{\alpha}}, \quad [\hat{\mathbf{R}}]_{ij} = 2^{\frac{\alpha}{2}}\,\frac{[\hat{\mathbf{C}}]_{ij}}{[\hat{\mathbf{R}}]_{jj}^{\frac{(\alpha-2)}{2}}} \tag{4.32}$$

which are consistent and asymptotically normal.

The estimation of the covariations (or covariation coefficients) requires the specification of the arbitrary parameter p. We compute the optimal p as a function of the characteristic exponent α, by optimizing the extracted information. In this case, the optimization is equivalent to finding the value of p that minimizes the standard deviation of the estimators, for different values of $\alpha > 1$. For this purpose we studied the influence of the p parameter to the performance of the covariation estimators given by Eqs. (4.30), (4.31) via Monte-Carlo simulations. We generated two real $S\alpha S$ $(1 < \alpha \leq 2)$ random variables, X and Y, as follows:

$$\begin{aligned} X &= a_1 U + b_1 V \\ Y &= a_2 U + b_2 V \ , \end{aligned}$$

where U and V are independent, standard $(\gamma_U = \gamma_V = 1)$ $S\alpha S$ random variables and $\{a_i,\, b_i,\, i = 1,2\}$ are real coefficients. We randomly selected these coefficients to be equal to

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} = \begin{bmatrix} 0.32 & -1.7 \\ -2.45 & 0.44 \end{bmatrix} .$$

From Eq. (4.20) it follows that the true covariation of X with Y is

$$[X,Y]_\alpha = a_1 a_2^{<\alpha-1>} + b_1 b_2^{<\alpha-1>} \ .$$

We generated $N = 5000$ independent samples of U and V and calculated the covariation estimator by means of Eqs. (4.30), (4.31) for different values of p in the range $(0, 2]$. We ran $K = 1000$ Monte-Carlo simulations for different values of $\alpha \in (1, 2]$, comparing the performance of the above two covariation estimators. Figure 4.4 displays the standard deviation of the two estimators $\hat{\mathbf{c}}(p)$, $\hat{\mathbf{c}}^{FLOM}(p)$ as a function of the parameter p and for different values of α.

In Figure 4.5 we plot the running sample variance estimates (or *Variograms*) σ_k^2 defined as follows:

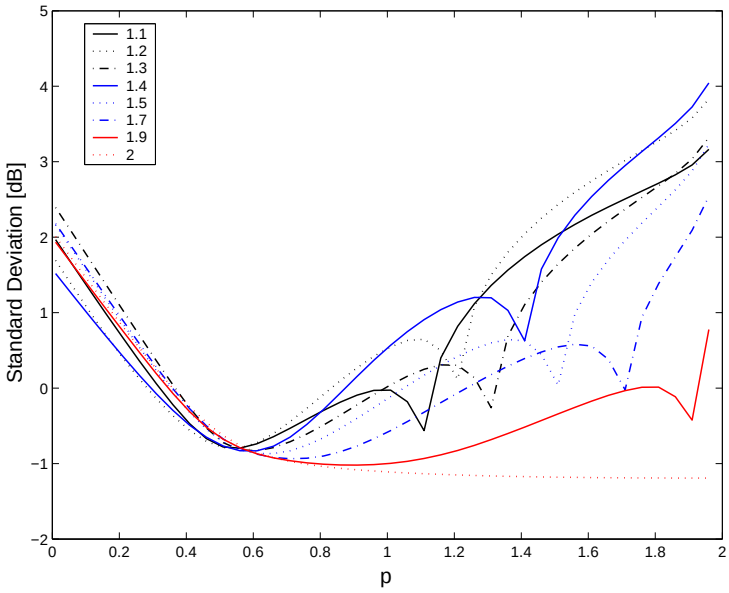
$$\sigma_k^2 = \frac{1}{k-1} \sum_{i=1}^{k} (\hat{\mathbf{c}}^{(i)}(p) - \bar{\hat{\mathbf{c}}}^{(k)}(p))^2, \; k = 1, \ldots, K \tag{4.33}$$

where

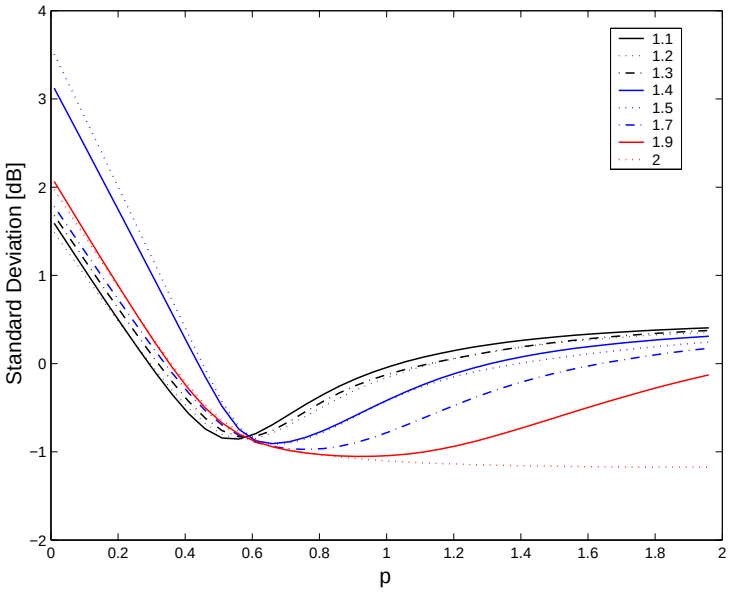
$$\bar{\hat{\mathbf{c}}}^{(k)}(p) = \frac{1}{k} \sum_{i=1}^{k} \hat{\mathbf{c}}^{(i)}(p) \; .$$

with $\hat{\mathbf{c}}^{(i)}(p)$ denoting the covariation estimation at the i-th Monte-Carlo iteration. The same definition holds for the variogram of $\hat{\mathbf{c}}^{FLOM}(p)$. If the set of the covariation estimates $\{\hat{\mathbf{c}}^{(i)}(p)\}_{i=1,\ldots,K}$ (or $\{\hat{\mathbf{c}}^{FLOM,(i)}(p)\}_{i=1,\ldots,K}$) has finite variance, the variogram will converge to a finite value.

Table 4.1 shows comparative results on the performance of the two estimators. We include the mean of the estimators, the standard deviation in parentheses and the value of p for which the smallest standard deviation is achieved by the estimators. We observe that, in general, the two estimators have comparable performance, with the $\hat{\mathbf{c}}^{FLOM}(p)$ estimator resulting in a smaller standard deviation for the most of the α values. Thus, in our experimental evaluations we will mainly rely on this estimator. Besides, in Figure 4.4(a) we observe that each curve corresponding to the $\hat{\mathbf{c}}$ estimator for a given value of α, exhibits a peak for p near the corresponding value of α. This is due to the fact that the exponent of the third factor in Eq. (4.30) vanishes when $p = \alpha$, which somehow results in a loss of the information provided by the lower-order moments of the second variable X_j. This observation gives also cause for selecting the second estimator, $\hat{\mathbf{c}}^{FLOM}$, in the experiments.



(a)



(b)

Figure 4.4: Curves representing the standard deviation of the covariation estimation as a function of the parameter p for (a) the $\hat{\mathbf{c}}$ and (b) the $\hat{\mathbf{c}}^{FLOM}$ estimator.

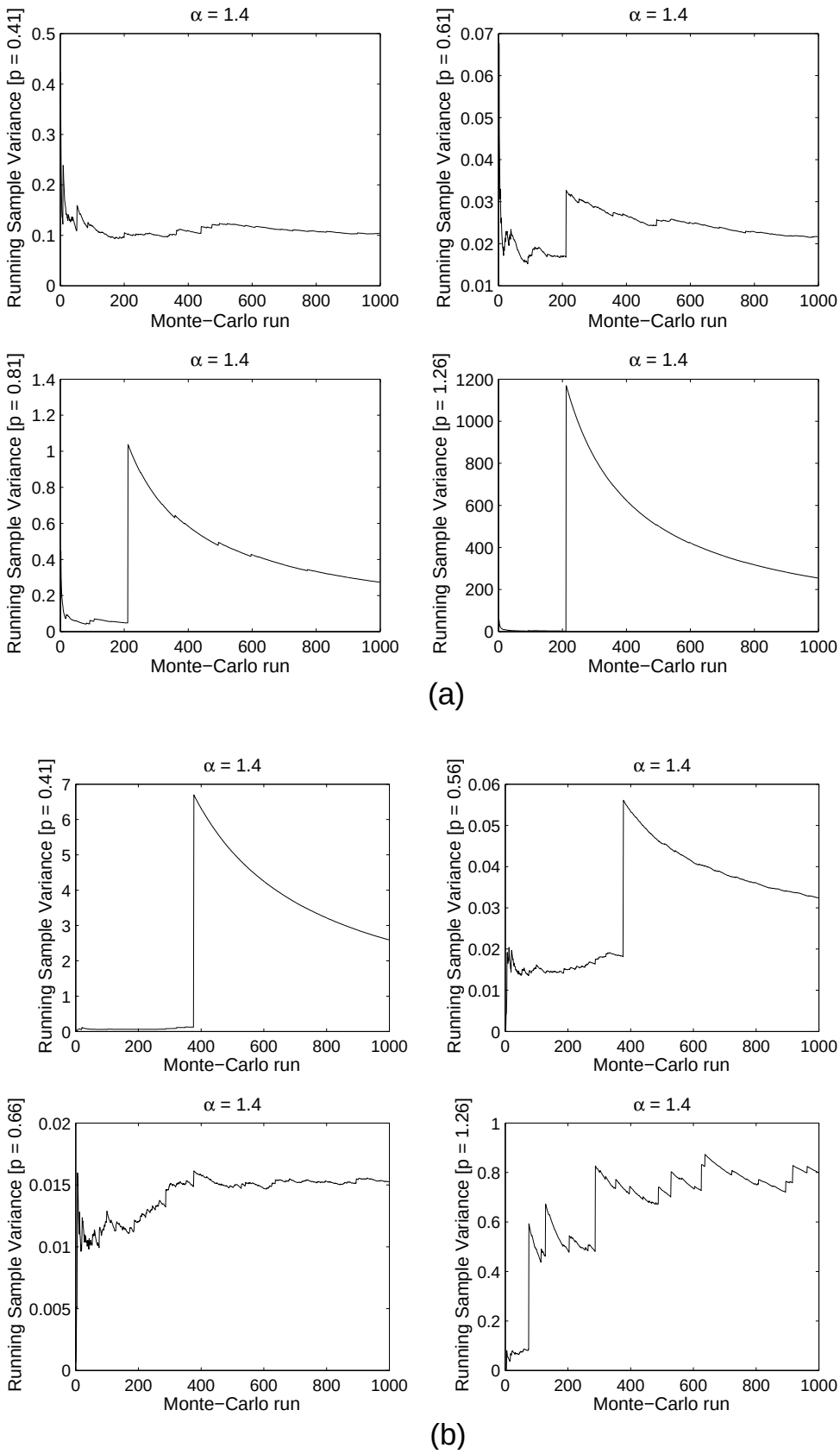


Figure 4.5: Running sample variances for $\alpha = 1.4$ and different values of p, corresponding to (a) the $\hat{\mathbf{c}}$ and (b) the $\hat{\mathbf{c}}^{FLOM}$ estimator.

α	$\hat{\mathbf{c}}(p)$	$\hat{\mathbf{c}}^{FLOM}(p)$		True $[X,Y]_\alpha$
1.1	-1.9129 (0.1604)	-1.9158 (0.1397)	(p=0.56)	-1.916
1.2	-1.8211 (0.1607)	-1.8203 (0.1390)	(p=0.57)	-1.8254
1.3	-1.7476 (0.1476)	-1.755 (0.1481)	(p=0.59)	-1.7476
1.4	-1.6913 (0.1472)	-1.6814 (0.1234)	(p=0.66)	-1.6821
1.5	-1.6292 (0.1354)	-1.6201 (0.1218)	(p=0.68)	-1.6285
1.7	-1.5614 (0.1156)	-1.5533 (0.1062)	(p=0.76)	-1.5561
1.9	-1.5252 (0.0954)	-1.5109 (0.0886)	(p=0.91)	-1.5288
2	-1.5324 (0.0640)	-1.5301 (0.0668)	(p=2)	-1.532

Table 4.1: Performance of the covariation estimators.

Part II

Statistical Content-Based Image Retrieval

Chapter 5

Wavelet-Based Texture Retrieval using Univariate $S\alpha S$ Distributions

5.1 Introduction

During the last decades, digital images are being gathered and stored at an explosive rate on large digital image databases. This fact gives rise to a very important issue, that is, how to effectively and precisely search and interact with these collections. We would desire to retrieve all the relevant images from a specific database, given a user query. For this purpose, retrieving images from an unannotated image database, based on their visual content, is a challenging problem.

In section 2.2 we introduced the typical architecture of a Content-Based Image Retrieval (CBIR) system, where there are two major tasks. The first one is the Feature Extraction (FE), where a set of features, constituting the so-called image signature, is generated to accurately represent the content of a given image. This set has to be much smaller in size than the original image, while capturing as much as possible of the image information. The second task is Similarity Measurement (SM), where a distance function is employed using their signatures, which measures how close to a query image each image in the database is, in order to retrieve the "most relevant" images.

The set of all possible images is immense and yet only a small portion of these are likely to be encountered in a real setting. It has been proven difficult to characterize this set of images using either deterministic or statistical models. A commonly used approach is to develop adaptive systems according to the specific application area, that is, systems which take into consideration the common characteristics of images belonging to the same category, e.g. industrial automation databases, biomedical and astronomical databases, etc.

In section 2.3 we mentioned typical low-level image features, such as color, shape, layout and texture, which are commonly used in CBIR applications. In this work, we focus on the use of *texture information* for image retrieval. Loosely speaking, the class of images that we commonly call *texture images* includes images that are spatially homogeneous and consist of repeated elements, often subject to some randomization in their location, size, color and orientation. Previously de-

veloped texture extraction methods are based on filtering approaches, measuring energies at the output of filter-banks. The basic assumption for these approaches is that the energy distribution in the frequency domain identifies a texture. In the SM step, the ultimate goal is to find similarity (distance) functions that are close to the human texture perception. Many retrieval systems use simple norm-based distances, like Euclidean distance, on the extracted image signatures as a similarity measure.

In this study we consider the tasks of FE and SM in a joint statistical framework, as presented in section 2.4. Following this approach, the FE step becomes a Maximum Likelihood (ML) estimator of the model parameters fitting the given image data, while in the SM step we implement the Kullback-Leibler divergence (KLD) as a measure of similarity between model parameters. In this setting, optimal retrieval is asymptotically achieved. Using this statistical approach, an extension of the energy-based methods for texture retrieval is to model each texture by the marginal densities of the coefficients obtained by applying a transformation on it. This is based on the results of resent psychological research on human texture perception, which suggests that two homogeneous textures are often difficult to discriminate if they produce similar marginal distributions of responses from a filter-bank [8].

In this chapter the goal is the design of a CBIR method that takes into consideration the actual heavy-tailed behavior of the wavelet subband coefficients, modeling their marginal distribution using a symmetric alpha-stable prior distributiom.

The outline of this chapter is as follows. In the next section we justify the design of a CBIR system in a transform domain and specifically in the wavelet domain and we briefly review the generalized Gaussian model. In section 5.3 we introduce the family of univariate $S\alpha S$ distributions in the context of wavelet-based texture retrieval, while in section 5.4 we study the construction of a suitable similarity measure. Finally, section 5.5 presents experimental results on simulated data and on a large texture image database indicating a decreased probability of retrieval error, concluding with some discussion.

5.2 Statistical Modeling of Wavelet Subband Coefficients

The development of retrieval models in a transform-domain is based on the observation that often a linear, invertible transform re-structures the image, resulting in a set of transform coefficients whose structure is "simpler" to model. Real-world images are characterized by a set of "features", such as edges, ridges and lines. For such images, the wavelet transform is a powerful modeling tool [72].

The following properties of the wavelet transform justify the implementation of CBIR methods in wavelet domain.

1. *Locality*: Each wavelet coefficient represents image content localized in both space and frequency.

2. *Multiresolution*: The wavelet transform decomposes each image at a nested set of scales.

3. *Edge detection*: Wavelet functions operate as local edge detectors, representing the edges by large wavelet coefficients at the corresponding locations.

Consequently, we can closely approximate an image using just a few, large amplitude, wavelet coefficients.

The wavelet transform provides a natural arrangement of the wavelet coefficients into multiscale and oriented subbands representing the horizontal, vertical and diagonal edges (cf. section 3.2.2). Traditional approaches computed the mean absolute value and the L_2 norm of the wavelet subbands as texture features. That is, given the wavelet coefficients $x_1^k, x_2^k, \ldots, x_N^k$ at the k-th subband of a given image I, the following two values correspond to the extracted features:

$$f_k^{(1)} = \frac{1}{N}\sum_{i=1}^{N} |x_i^k| \ , \tag{5.1}$$

$$f_k^{(2)} = \left(\frac{1}{N}\sum_{i=1}^{N} |x_i^k|^2\right)^{1/2} \ , \quad k = 1, \ldots, J \tag{5.2}$$

where J denotes the number of subbands. Then, the signature of image I was represented by the vector $\mathcal{S}_I = (f_1^{(1)}, f_1^{(2)}, \ldots, f_J^{(1)}, f_J^{(2)})$ and the similarity between two images I and Q was measured by the Euclidean distance of their corresponding signatures $\mathcal{S}_I$ and $\mathcal{S}_Q$, respectively.

Manjunath and Ma [78] represent image content using a set of Gabor wavelet features for texture analysis. In particular, the feature vector, which is used to represent the region for classification and retrieval, contains the means and standard deviations of the magnitudes of the transform coefficients and the similarity function is defined as the sum of the absolute value of normalized differences between corresponding features. A texture descriptor based on a multiresolution decomposition using Gabor wavelets is also proposed in [132]. The descriptor consists of a perceptual browsing component, which provides a quantitative characterization of the texture's structuredness and directionality, and a similarity retrieval component that characterizes the distribution of texture energy in different subbands and supports similarity retrieval. This representation is also quite robust to illumination variations. A texture description that is translation invariant is detailed in [125]. The analysis uses an overcomplete wavelet decomposition constituting a tight frame [25], characterizing the texture by a set of channel variances estimated at the output of the corresponding filter bank.

On the other hand, working in a statistical framework a texture is modeled by the marginal densities of wavelet subband coefficients. Consider the simple model where the marginal distributions belong to the Gaussian family. Then, given the wavelet coefficients $x_1^k, x_2^k, \ldots, x_N^k$ at the k-th subband of a given image I, we assume that they are drawn from a Gaussian distribution with mean μ and standard deviation σ,

$$p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.3}$$

In this case, the extracted features of the k-th subband are the estimated model parameters, i.e.,

the sample mean and the sample variance values,

$$\hat{\mu}_k = \frac{1}{N}\sum_{i=1}^{N} x_i^k \,, \tag{5.4}$$

$$\hat{\sigma}_k = \Big(\frac{1}{N-1}\sum_{i=1}^{N}(x_i^k - \hat{\mu}_k)^2\Big)^{1/2} \,. \tag{5.5}$$

The distance between the k-th subbands of two images I and Q, is defined as the KLD between the corresponding Gaussian distributions. Let $(\hat{\mu}_{k,I}, \hat{\sigma}_{k,I})$ and $(\hat{\mu}_{k,Q}, \hat{\sigma}_{k,Q})$ denote the model parameters of the k-th subband of images I and Q, respectively. The KLD has the following closed form expression:

$$\begin{aligned} D(p(\cdot\,;\hat{\mu}_{k,I},\hat{\sigma}_{k,I})\|p(\cdot\,;\hat{\mu}_{k,Q},\hat{\sigma}_{k,Q})) &= \ln\big(\frac{\hat{\sigma}_{k,Q}}{\hat{\sigma}_{k,I}}\big) - \frac{1}{2} + \\ &+ \frac{1}{2\hat{\sigma}_{k,Q}^2}\big(\hat{\sigma}_{k,I}^2 + \hat{\mu}_{k,I}^2 + \hat{\mu}_{k,Q}^2 - 2\hat{\mu}_{k,I}\hat{\mu}_{k,Q}\big) \end{aligned} \tag{5.6}$$

Assuming that wavelet coefficients in different subbands are independent, Eq. (2.14) dictates to define the overall similarity distance between images I and Q (where I is the query image) as the sum of the KLDs between corresponding pairs of subbands:

$$D(I\|Q) = \sum_{k=1}^{J} D(p(\cdot\,;\hat{\mu}_{k,I},\hat{\sigma}_{k,I})\|p(\cdot\,;\hat{\mu}_{k,Q},\hat{\sigma}_{k,Q})) \tag{5.7}$$

Figure 5.1 displays the expression for the KLD between univariate Gaussian densities when they have the same mean value and when they have the same standard deviation. We observe that the minimum KLD is achieved when the two PDFs are equal, i.e., they have equal parameters.

Until recently, wavelet coefficients have been modeled either as independent Gaussian or as jointly Gaussian [68, 82]. The above mentioned properties of the wavelet transform, yield that the wavelet transforms of real-world images tend to be sparse, resulting in a large number of small coefficients and a small number of large coefficients. This property is in conflict with the Gaussian assumption, giving rise to peaky and heavy-tailed *non-Gaussian* marginal distributions of the wavelet subband coefficients [12, 103, 127].

Experimental results have proven that the *generalized Gaussian density* (GGD) is a suitable member of the class of non-Gaussian distributions for modeling the marginal behavior of the wavelet coefficients [74, 23, 12]. Liapis and Tziritas [66] explore computationally tractable image retrieval mechanisms based on a combination of texture and color features. Texture features are extracted using Discrete Wavelet Frames (DWF) analysis, while 2-D or 1-D histograms of the *CIE Lab* chromaticity coordinates are used as color features. The 1-D histograms of the *a, b* coordinates were modeled according to the generalized Gaussian distribution. The similarity measure defined on the feature distribution is based on the Bhattacharrya distance.

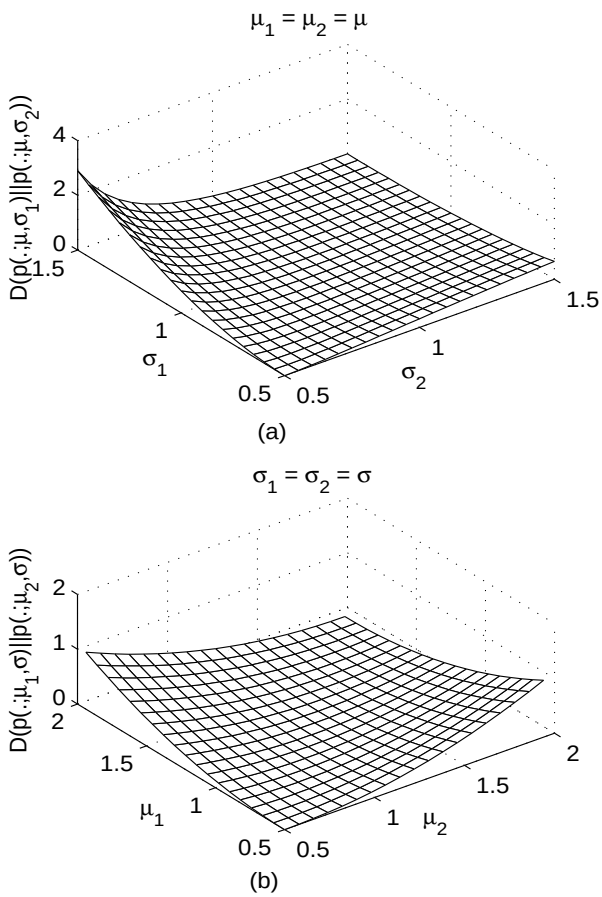


Figure 5.1: KLD between Gaussian PDFs with (a) the same mean and (b) the same standard deviation.

A GGD is defined as:

$$p(x;\alpha,\,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)}\; e^{-\left(\frac{|x|}{\alpha}\right)^{\beta}}\;, \tag{5.8}$$

where $\Gamma(\cdot)$ is the Gamma function, $\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, \mathrm{d}t, \quad x > 0$. Here α is a scale parameter modeling the width of the PDF peak, while β is a shape parameter, inversely proportional to the decreasing rate of the peak. The Gaussian and Laplacian PDFs are special cases of a GGD model corresponding to $\beta = 2$ and $\beta = 1$, respectively.

Recently, the GGD models have been introduced in a statistical framework for texture retrieval in CBIR applications, by jointly considering the two problems of FE and SM [34]. In this approach, the FE step becomes an ML estimator for the model parameters $(\alpha,\,\beta)$ of the image data and the SM step computes the Kullback-Leibler distance between the model parameters. The KLD between two GGDs has a closed form expression, given by:

$$D(p(\cdot;\alpha_1,\beta_1)\|p(\cdot;\alpha_2,\beta_2)) = \ln\!\left(\frac{\beta_1\alpha_2\Gamma(1/\beta_2)}{\beta_2\alpha_1\Gamma(1/\beta_1)}\right) + \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2}\frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1} \tag{5.9}$$

where $(\alpha_1,\,\beta_1)$ and $(\alpha_2,\,\beta_2)$ are the scale and the shape parameters' pair, of each PDF respectively. With the assumption that wavelet coefficients in different subbands are independent, the overall similarity distance between images I and Q (where I is the query image) is again defined as the

sum of the KLDs between corresponding pairs of subbands. That is, if we denote $(\alpha_{k,I}, \beta_{k,I})$ and $(\alpha_{k,Q}, \beta_{k,Q})$ as the model parameters of the k-th subband of images I and Q, then the overall KLD is defined as:

$$D(I\|Q) = \sum_{k=1}^{J} D(p(\cdot\,;\alpha_{k,I}, \beta_{k,I})\|p(\cdot\,;\alpha_{k,Q}, \beta_{k,Q})) \tag{5.10}$$

where J is the number of subbands.

So far forth, we have seen the reasons that justify the use of a non-Gaussian model for the approximation of the heavy-tailed behavior of the wavelet subband coefficients, with respect to a CBIR system design. In the next section, we will introduce the family of symmetric alpha-stable ($S\alpha S$) distributions, in the context of a CBIR application, as a more suitable modeling tool for the statistical characterization of the marginal distributions of the wavelet coefficients.

5.3 $S\alpha S$ Modeling of the Wavelet Decompositions

In this section, we address the CBIR problem in a statistical framework by employing the family of univariate $S\alpha S$ distributions (cf. section 4.2.1). The approach taken is similar to the method presented in the previous section in that

(i) texture retrieval is performed in the wavelet transform domain where wavelet coefficients in each subband are independently modeled, and

(ii) texture similarity is measured by means of the Kullback-Leibler distances (KLD) between model parameters.

The added value of the present method is twofold: first, we show that an often better approximation for the marginal density of coefficients at a particular subband produced by various types of wavelet transforms may be obtained by alpha-stable models; second, we provide similarity measures between two $S\alpha S$ densities, which by definition do not have a closed-form expression.

The $S\alpha S$ model is suitable for describing signals with heavier distribution tails than what is assumed by exponential families, like the GGD. Indeed, the $S\alpha S$ density follows an algebraic rate of decay that depends on the value of the characteristic exponent: $P(X > x) \sim c_\alpha x^{-\alpha}$.

In recent work, it has been shown that successful image processing algorithms can achieve both noise reduction and feature preservation if they take into consideration the actual heavy-tailed behavior of the signal and noise densities [1, 2]. Specifically, it has been demonstrated that the subband decompositions of various types of images, including ultrasound and SAR, have non-Gaussian statistics that are best described by families of distributions with algebraic tails, such as the alpha-stable. Then, appropriate Bayesian processors were designed that exploited these statistics and had significantly improved performance in real data.

5.3.1 Feature Extraction

The $S\alpha S$ distribution, which does not have a closed-form expression except for the Cauchy and Gaussian cases, is best defined by its characteristic function, which can take different forms

depending on the choice of the parameterization (cf. Eqs. (4.1-4.3)). In the FE step, the image is decomposed into several scales through a multiresolution analysis employing the 2-D wavelet transform [75]. Our proposed method is based on the accurate modeling of the tails of the marginal distribution of the wavelet coefficients at each subband: the wavelet subband coefficients in various scales are modeled as symmetric alpha-stable ($S\alpha S$) random variables.

Working in such a CBIR statistical framework, the extracted features for each subband are the estimated model parameters, that is, the value of the characteristic exponent α ($0 < \alpha \leq 2$) and the value of the dispersion γ ($\gamma > 0$). We estimate the pair (α, γ) using the consistent maximum likelihood (ML) method described by Nolan [87], which gives reliable estimates and provides the most tight confidence intervals.

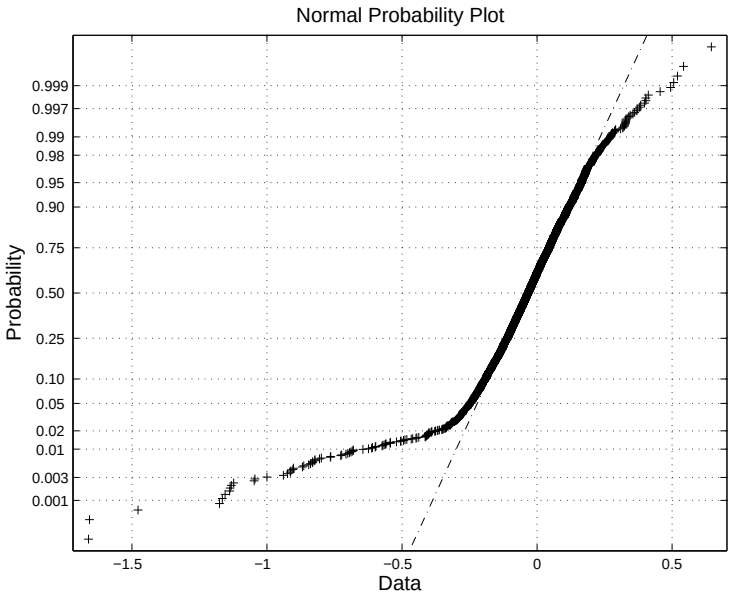


Figure 5.2: Normal probability plot of the horizontal subband at the first level of decomposition of the image *Flowers.0006*. The "+" marks correspond to the empirical probability density vs. the data value for each point in the sample. Since the marks do not follow the straight Gaussian line, the normality assumption is violated for this data.

In our data modeling study, the statistical fitting proceeds in two steps: first, we assess whether the data deviate from the normal distribution and we determine if they have heavy tails by employing normal probability plots [21]. Then, we check if the data is in the stable domain of attraction by estimating the characteristic exponent, α, directly from the data and by providing the related confidence intervals. As further stability diagnostics, we employ the amplitude probability density (APD) curves ($P(|X| > x)$ that give a good indication of whether the $S\alpha S$ fit matches the data near the mode and on the tails of the distribution.

As an example, the normal probability plot corresponding to the vertical subband at the first level of decomposition of the *Flowers.0006* image is shown in Figure 5.2. The plot provides strong evidence that the underlying distribution is not normal, since the "+" marks, corresponding to the empirical probability versus the data value for each point in the sample, are in a curve that does not follow the straight Gaussian line. While non-Gaussian stable densities are heavy-tailed,

not all heavy-tailed distributions are stable. Hence, we further assess the stability of the wavelet coefficients at the given subband using APD curves. Figure 5.3 compares the $S\alpha S$ and GGD fits for the selected subband. Clearly, the $S\alpha S$ density is superior to the GGD, following more closely both the mode and the tail of the empirical APD than the exponentially decaying GGD.

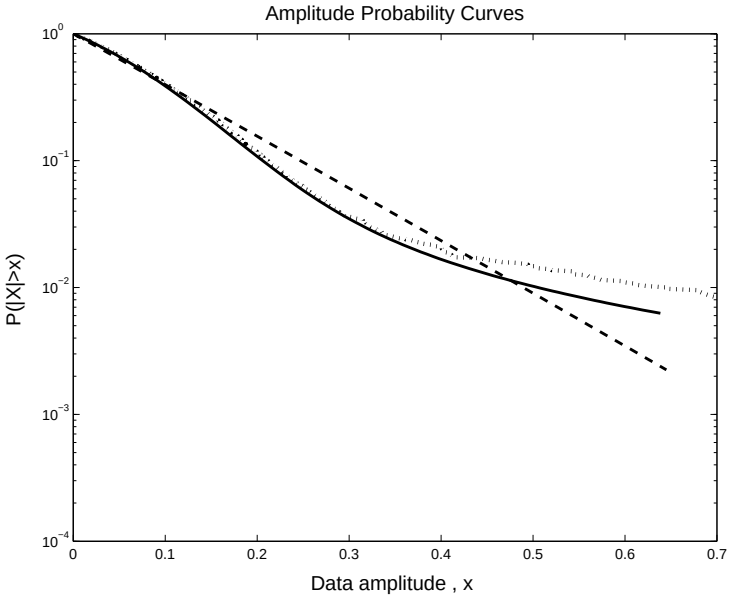


Figure 5.3: Modeling of the horizontal subband at the first level of decomposition of the *Flowers.0006* image with the $S\alpha S$ and the GGD depicted in solid and dashed lines, respectively. The estimated parameters for the $S\alpha S$ distribution have the values $\alpha = 1.76$, $\gamma = 0.08$ while the GGD has parameters $\alpha = 0.11$ and $\beta = 1.02$. The dotted line denotes the empirical APD.

For every image in a set of 10 randomly chosen textures (real-world 512×512 natural scene images) obtained from the MIT Vision Texture (VisTex) database, we applied a 3-level wavelet decomposition using Daubechies' 4 ('db4') basis wavelet. We modeled the coefficients of each subband using the $S\alpha S$ parameterization given by Eq. (4.2) (we use parameterizations (4.2), (4.3) in order to be in agreement with Nolan's ML estimator):

$$\phi(t) = \exp(\imath\delta t - \gamma^{\alpha}|t|^{\alpha}) \ .$$

Table 5.1 shows the ML estimates of the characteristic exponent α together with the corresponding 95% confidence intervals. It can be observed that the confidence intervals depend on the decomposition level. In particular, they become wider as the level increases since the number of samples used for estimating the $S\alpha S$ parameters gets smaller. This table also demonstrates that the coefficients of different subbands and decomposition levels exhibit various degrees of non-Gaussianity, with values of α varying between 0.9 and 2.

As a conclusion, with only two parameters for the $S\alpha S$ distribution, we can accurately capture the texture information expressed by the marginal distribution of wavelet coefficients in a subband. This significantly reduces the storage of the image features, as in a L-level wavelet decomposition the total number of parameters required to represent an image equals $L \times 3 \times 2 = 6L$

IMAGE	Level	Image Subbands		
		Horizontal	Vertical	Diagonal
Bark.0010	1	1.6006 ± 0.0608	1.6847 ± 0.0588	1.6811 ± 0.0579
Brick.0001		1.614 ± 0.0567	1.577 ± 0.0564	1.8965 ± 0.0310
Buildings.0004		1.6811 ± 0.0394	1.6847 ± 0.0478	1.6006 ± 0.0467
Fabric.0000		2.000 ± 0.0068	1.2703 ± 0.0535	1.3224 ± 0.0556
Fabric.0010		1.2299 ± 0.0578	1.3671 ± 0.0586	1.1757 ± 0.0539
Flowers.0006		1.76 ± 0.0416	1.7011 ± 0.0441	1.9868 ± 0.0254
Food.0009		1.6263 ± 0.0619	1.3390 ± 0.0550	1.3862 ± 0.0562
Grass.0001		1.8796 ± 0.0472	1.8532 ± 0.0530	1.7911 ± 0.0557
Metal.0004		1.3202 ± 0.0571	1.2287 ± 0.0543	1.4023 ± 0.0589
Stone.0003		1.5908 ± 0.0542	1.6883 ± 0.0499	1.7464 ± 0.0502
Bark.0010	2	1.8552 ± 0.0976	1.8585 ± 0.1101	1.8506 ± 0.1072
Brick.0001		1.3115 ± 0.1059	1.5395 ± 0.1036	1.8500 ± 0.1008
Buildings.0004		0.9219 ± 0.0893	1.1866 ± 0.1197	1.1513 ± 0.0986
Fabric.0000		2.000 ± 0.0067	1.1712 ± 0.0998	1.3493 ± 0.1138
Fabric.0010		1.6775 ± 0.1203	1.6109 ± 0.1152	1.5576 ± 0.1139
Flowers.0006		1.3349 ± 0.0974	1.4248 ± 0.1004	1.9001 ± 0.0577
Food.0009		1.9900 ± 7.7883e-008	1.4647 ± 0.1074	1.7503 ± 0.1009
Grass.0001		1.9209 ± 0.0727	1.8588 ± 0.0999	1.9900 ± 0.0727
Metal.0004		1.6802 ± 0.1205	1.5050 ± 0.1193	1.6905 ± 0.1198
Stone.0003		1.8693 ± 0.0834	1.5097 ± 0.1034	1.6586 ± 0.1167
Bark.0010	3	1.7232 ± 0.2256	2.0000 ± 0.1623	1.7926 ± 0.2486
Brick.0001		1.2276 ± 0.2427	1.3680 ± 0.2048	1.9900 ± 0.2191
Buildings.0004		1.2201 ± 0.2010	2.0000 ± 0.1090	1.0141 ± 0.1813
Fabric.0000		2.0000 ± 0.0977	1.4985 ± 0.2463	1.8748 ± 0.1678
Fabric.0010		1.8654 ± 0.2059	2.0000 ± 0.1031	1.9047 ± 0.1840
Flowers.0006		1.6437 ± 0.2220	1.5734 ± 0.2100	1.8515 ± 0.2116
Food.0009		2.0000 ± 0.0059	1.6770 ± 0.2223	1.9900 ± 0.2456
Grass.0001		2.0000 ± 0.3758	1.8626 ± 0.1772	2.0000 ± 0.2413
Metal.0004		1.7873 ± 0.2179	1.8887 ± 0.1448	1.8631 ± 0.1817
Stone.0003		2.0000 ± 0.0968	1.2785 ± 0.2235	1.5003 ± 0.2263

Table 5.1: $S\alpha S$ modeling of wavelet subband coefficients of texture images from the VisTex database, using Daubechies' 4 filter and 3 decomposition levels. ML parameter estimates and 95% confidence intervals for the characteristic exponent α.

and this number is independent of the original image size.

5.4 Similarity Measurement between $S\alpha S$ distributions

When considering the alpha-stable family, the KLD between two Cauchy distributions ($S\alpha S$ with $\alpha = 1$) is calculated to be

$$D(p(\cdot;\gamma_1)\|p(\cdot;\gamma_2)) = \ln\Big(\frac{\gamma_1}{\gamma_2}\Big) + 2\frac{\gamma_1^2}{\gamma_2}\cdot\ln\Big(1+\frac{\gamma_2}{\gamma_1}\Big) - \gamma_1\cdot\ln 4 \tag{5.11}$$

where

$$p(x;\gamma) = \frac{\gamma}{\pi(x^2+\gamma^2)} \tag{5.12}$$

is a symmetric Cauchy density function with zero mean. Unfortunately, there is no closed-form expression for the KLD between two general $S\alpha S$ distributions, which are not Cauchy or Gaussian. To address this problem, we could employ numerical techniques for the computation of the KLD between two, numerically approximated, $S\alpha S$ densities. This approach could be adequate when the image is decomposed in a small number of levels, because the number of the approximated densities is proportional to the number of these levels. In any other case, this method would significantly increase the computational burden.

In order to avoid the increased computational complexity of a numerical scheme, we applied the KLD on the normalized versions of the corresponding characteristic functions (CF). Due to the one-to-one correspondence between a $S\alpha S$ density and its associated characteristic function, we expect that the KLD between normalized CFs will be a good similarity measure between $S\alpha S$ distributions.

The normalization procedure for a characteristic function results in its transformation into a valid probability density function. If $\phi(t)$ is a CF corresponding to a $S\alpha S$ distribution, then the function

$$\hat{\phi}(t) = \frac{\phi(t)}{c} \tag{5.13}$$

is a valid density function when

$$c = \int_{-\infty}^{\infty}\phi(t)\,\mathrm{d}t\ .$$

For convenience, let $\phi_1(t)$, $\phi_2(t)$ and $\phi_3(t)$ correspond to the three parameterizations (4.1), (4.2) and (4.3), respectively. Then, assuming that the $S\alpha S$ densities are centered at zero, i.e., $\delta = 0$, the corresponding normalized CFs are defined as:

$$\hat{\phi}_i(t) = \frac{\phi_i(t)}{c_i}, \qquad i = 1,2,3 \tag{5.14}$$

with

$$c_1 = \frac{2\Gamma\left(\frac{1}{\alpha}\right)}{\alpha\,\gamma^{1/\alpha}}, \quad c_2 = \frac{2\Gamma\left(\frac{1}{\alpha}\right)}{\alpha\,\gamma}, \quad c_3 = \frac{2\Gamma\left(\frac{1}{\alpha}\right)}{\alpha^{1-1/\alpha}\,\gamma_*}\ . \tag{5.15}$$

The promising result of this procedure is that we have obtained closed-form expressions for

the KLD between normalized CFs, related with $S\alpha S$ densities. By employing the KLD between a pair of normalized CFs for each one of the above three parameterizations and after some manipulation, we obtain the following closed-form expressions:

$$D_1(\hat{f}_1\|\hat{g}_1) = \ln\Big(\frac{c_{\hat{g}_1}}{c_{\hat{f}_1}}\Big) - \frac{1}{\alpha_{\hat{f}_1}} + \frac{2\gamma_{\hat{g}_1}\Gamma\Big(\frac{\alpha_{\hat{g}_1}+1}{\alpha_{\hat{f}_1}}\Big)}{c_{\hat{f}_1}\alpha_{\hat{f}_1}\gamma_{\hat{f}_1}^{(\alpha_{\hat{g}_1}+1)/\alpha_{\hat{f}_1}}} \tag{5.16}$$

$$D_2(\hat{f}_2\|\hat{g}_2) = \ln\Big(\frac{c_{\hat{g}_2}}{c_{\hat{f}_2}}\Big) - \frac{1}{\alpha_{\hat{f}_2}} + \Big(\frac{\gamma_{\hat{g}_2}}{\gamma_{\hat{f}_2}}\Big)^{\alpha_{\hat{g}_2}} \cdot \frac{\Gamma\Big(\frac{\alpha_{\hat{g}_2}+1}{\alpha_{\hat{f}_2}}\Big)}{\Gamma\Big(\frac{1}{\alpha_{\hat{f}_2}}\Big)} \tag{5.17}$$

$$D_3(\hat{f}_3\|\hat{g}_3) = \ln\Big(\frac{c_{\hat{g}_3}}{c_{\hat{f}_3}}\Big) - \frac{1}{\alpha_{\hat{f}_3}} + \Big(\frac{\gamma_{\hat{g}_3}}{\gamma_{\hat{f}_3}}\Big)^{\alpha_{\hat{g}_3}} \cdot \frac{\alpha_{\hat{f}_3}^{\alpha_{\hat{g}_3}/\alpha_{\hat{f}_3}}\,\Gamma\Big(\frac{\alpha_{\hat{g}_3}+1}{\alpha_{\hat{f}_3}}\Big)}{\alpha_{\hat{g}_3}\,\Gamma\Big(\frac{1}{\alpha_{\hat{f}_3}}\Big)} \tag{5.18}$$

where $(\alpha_{\hat{f}_i}, \gamma_{\hat{f}_i})$, $(\alpha_{\hat{g}_i}, \gamma_{\hat{g}_i})$ are the parameters of the CFs, $f_i(t)$ and $g_i(t)$, respectively and $c_{\hat{f}_i}$, $c_{\hat{g}_i}$ are the corresponding normalizing factors according to the selected parameterization $i = 1, 2, 3$.

As an illustration, Figure 5.4 shows the $D_2(\cdot\,\|\,\cdot)$ version of the KLD when the two densities have the same characteristic exponent and when they have the same dispersion parameter. Notice that the substitution of the appropriate normalizing factors in Eq. (5.17) results in a KLD which is a function of three variables: the ratio of the two dispersions $\gamma_{\hat{f}_2}/\gamma_{\hat{g}_2}$ and the two characteristic exponents $\alpha_{\hat{f}_2}$ and $\alpha_{\hat{g}_2}$.

5.4.1 Relation to the KLD between Gaussian densities

Consider the case when the characteristic exponent $\alpha = 2$, i.e., we are modeling the wavelet subband coefficients using a zero-mean Gaussian density. The extracted feature corresponding to each subband is the sample variance σ^2 (L_2 norm), while from a $S\alpha S$ point of view this feature is the estimated dispersion γ, which is equal to $2\sigma^2$ under parameterization (4.2).

From Eq. (5.6), the KLD between two zero-mean Gaussian densities p_1, p_2 is

$$D(p_1\|p_2) = \ln\Big(\frac{\sigma_2}{\sigma_1}\Big) - \frac{1}{2} + \frac{1}{2}\Big(\frac{\sigma_1}{\sigma_2}\Big)^2 \tag{5.19}$$

while from Eq. (5.17), the KLD between their corresponding normalized CFs with dispersions γ_1, γ_2 is

$$\begin{aligned} D_2(\hat{\phi}_1\|\hat{\phi}_2) &= \ln\Big(\frac{\gamma_1}{\gamma_2}\Big) - \frac{1}{2} + \frac{1}{2}\Big(\frac{\gamma_2}{\gamma_1}\Big)^2 \\ &= \ln\Big(\frac{2\sigma_1}{2\sigma_2}\Big) - \frac{1}{2} + \frac{1}{2}\Big(\frac{2\sigma_2}{2\sigma_1}\Big)^2 \\ &= \ln\Big(\frac{\sigma_1}{\sigma_2}\Big) - \frac{1}{2} + \frac{1}{2}\Big(\frac{\sigma_2}{\sigma_1}\Big)^2 \\ &\overset{(5.19)}{=} D(p2\|p1) \end{aligned} \tag{5.20}$$

We observe that the role of an image as a query is interchanged in the last two expressions. That

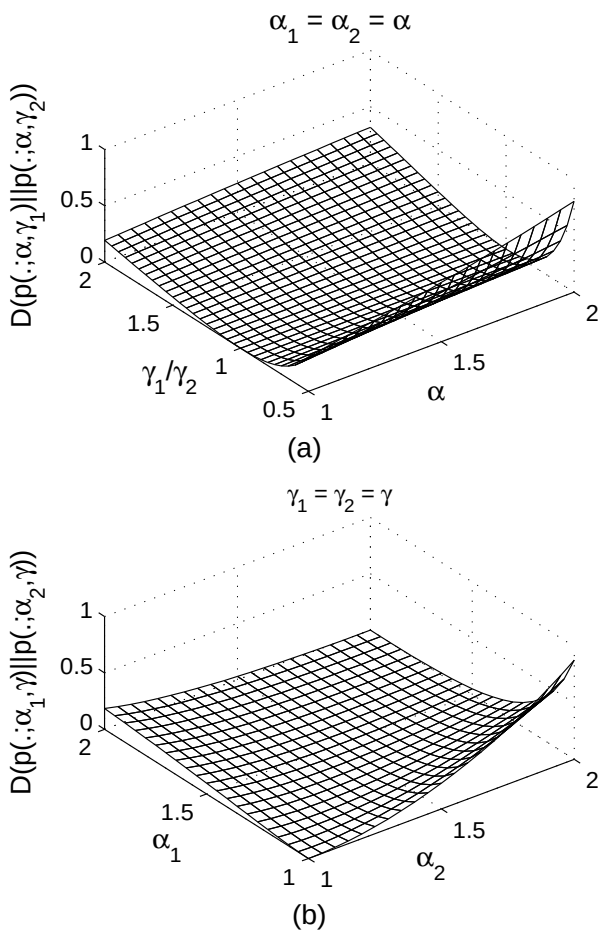


Figure 5.4: Kullback-Leibler divergence between two normalized characteristic functions following parameterization (4.2), with (a) the same characteristic exponent and (b) the same dispersion.

is, the KLD between normalized CFs when the query image is I_1, is equal to the KLD between Gaussian distributions with image I_2 to be the query.

5.5 Experiments with real-world Texture Images

In this experiment, we evaluated the performance of the proposed CBIR technique by applying it on the set of 30 gray-scale texture images obtained from the VisTex and Brodatz databases, shown in Figure 5.5 (in the following study we may use the index of a class instead of its name, as shown in Table 5.2). Then, these real-world 512×512 natural scene images were divided into 16 128×128 non-overlapping subimages, resulting in a database of 480 textures. Besides, to reduce the effect of bias in evaluation, since our similar images might share the same mean or variance, each subimage was normalized to zero mean and unit variance before applying the wavelet decomposition.

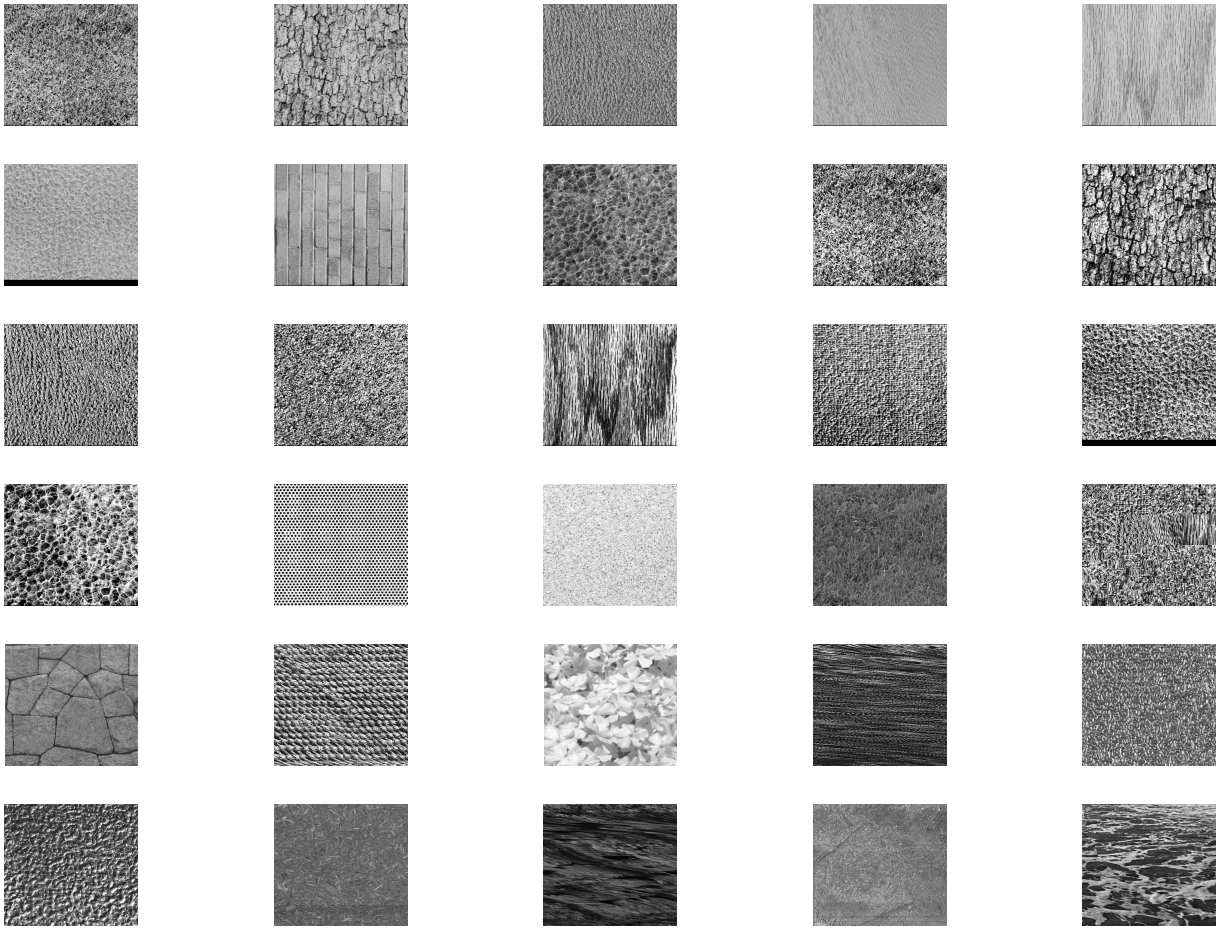


Figure 5.5: Texture images from the VisTex and Brodatz databases that are used in the experiments. From left to right and top to bottom: (**Brodatz**) 1.1.01, 1.1.02, 1.1.06, 1.1.08, 1.1.09, 1.1.11, 1.1.12, 1.1.13, 1.2.01, 1.2.02, 1.2.06, 1.2.07, 1.2.09, 1.2.10, 1.2.11, 1.2.13, 1.5.02, 1.5.04, 1.5.07, texmos1.p512, (**VisTex**) Brick.0004, Fabric.0009, Flowers.0006, Food.0004, Metal.0002, Metal.0004, Misc.0002, Stone.0003, Stone.0004, Water.0004

Index	Texture Class	Index	Texture Class
1	1.1.01	16	1.2.13
2	1.1.02	17	1.5.02
3	1.1.06	18	1.5.04
4	1.1.08	19	1.5.07
5	1.1.09	20	texmos1.p512
6	1.1.11	21	Brick.0004
7	1.1.12	22	Fabric.0009
8	1.1.13	23	Flowers.0006
9	1.2.01	24	Food.0004
10	1.2.02	25	Metal.0002
11	1.2.06	26	Metal.0004
12	1.2.07	27	Misc.0002
13	1.2.09	28	Stone.0003
14	1.2.10	29	Stone.0004
15	1.2.11	30	Water.0004

Table 5.2: Index of each Texture Class.

In this experiment, we employed the standard 2-D wavelet transform with three decomposition levels using the Daubechies' orthogonal filters ('dbN'). From every image in the database, two $S\alpha S$ parameters were estimated from each of the 9 subbands (except for the approximation subband) using Nolan's ML estimator. As a result, each texture image is represented by the $2 \times 9 = 18$ model parameters that capture texture-specific information.

5.5.1 Range of characteristic exponent α in $S\alpha S$ models

The key parameter α defines the degree of non-Gaussianity as deviations from the value $\alpha = 2$, which corresponds to the Gaussian case. So, it may be of interest to know the typical range for the values of α in $S\alpha S$ distributions modeling texture images.

The modeling of the 30 database images decomposed in 3 levels using 'db4' filters, produced values of the α parameter in the range between 0.4 and 2, as it is shown in the histogram of the estimated values (cf. Figure 5.6). We observe that, for this set of images, a large fraction of the wavelet coefficients follows statistics close to Gaussian ($1.8 \leq \alpha \leq 2$). These coefficients generally belong to the coarsest decomposition level. On the other hand, a considerable portion of the coefficients (belonging to the first two decomposition levels) exhibits a clear heavy-tailed behavior ($\alpha \leq 1.8$).

5.5.2 Retrieval Performance

In this system, a query texture sample is anyone of the 480 images (of size 128×128) in the database. In the ideal case, the 15 closest textures will belong to the same class as the class of the query texture, that is, the same original 512×512 image.

In practice, a certain number $N_{closest} < 16$ will be obtained, giving a certain percentage $100 \cdot N_{closest}/16$. The performance for a given class is measured in terms of the average percentage

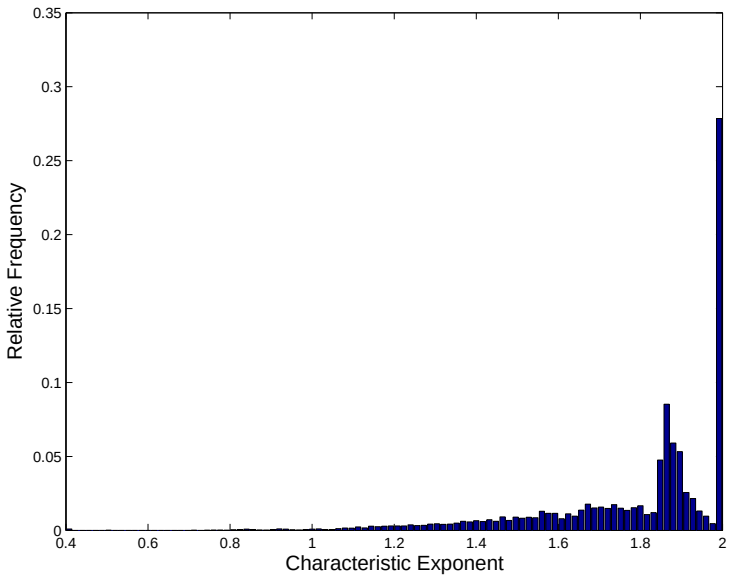


Figure 5.6: Histogram of the estimated values for the characteristic exponent, α, for the set of 480 texture images of size 128×128.

that is obtained after using as a query all the texture samples belonging to that class. The system's performance is evaluated in terms of the average rate of retrieving relevant images as a function of the number of top retrieved images.

We applied a 3-level wavelet transform using members of the Daubechies' 'dbN' family of orthogonal wavelets. Also, the three decomposition levels are justified by the fact that for an input image of size 128×128 the size of each subband at the coarsest level equals 16×16, that is already a small sample size for a robust estimation of the $S\alpha S$ parameters.

Table 5.3 shows the comparison in performance in average percentages of retrieving relevant images in the top 15 matches, with different filters and decomposition levels. $L_1 + L_2$ denotes the method which uses the energy-based texture features obtained by Eqs. (5.1, 5.2) and yields the same number of features per image as the statistical model (2 features per subband). In this case, the normalized Euclidean distance (nED) is chosen to be the similarity measure between the feature vectors of two images. We also applied the nED between the feature vectors with components the estimated $S\alpha S$ parameters. The proposed CBIR scheme uses the combination of a $S\alpha S$ model with the corresponding KLD as a similarity function, compared with the GGD & KLD model.

We observe that in the case of decomposing each image using 'db2' filter, the average retrieval rate using 3 levels is less than the retrieval rate using 2 levels. This is due to the uncertainty in the estimation of the characteristic exponents. We can also see that the first two decomposition levels contain the most discrimination information, since the improvement in the retrieval rates obtained by including the third level is relatively small. Another conclusion is that the retrieval performance does not only depend on the extracted features, but also on the choice of a suitable similarity measure, justifying the use of a probabilistic similarity function.

We also tested the retrieval performance in a subset of 320 texture images with fairly heavy-

Filters	Methods			
	$L_1 + L_2$ & nED	GGD & KLD	$S\alpha S$ & nED	$S\alpha S$ & D_2
db2	54.375	63.724	60.326	60.794
db4	53.177	62.995	58.359	58.633
db8	52.227	59.466	55.729	57.044
db10	52.643	59.714	53.698	56.628

(a) One-level Decomposition

Filters	Methods			
	$L_1 + L_2$ & nED	GGD & KLD	$S\alpha S$ & nED	$S\alpha S$ & D_2
db2	59.779	68.763	67.174	67.695
db4	59.206	66.953	65.43	65.495
db8	60.339	64.232	59.74	63.854
db10	60.169	63.945	59.049	63.529

(b) Two-level Decomposition

Filters	Methods			
	$L_1 + L_2$ & nED	GGD & KLD	$S\alpha S$ & nED	$S\alpha S$ & D_2
db2	61.536	69.01	66.159	67.604
db4	61.953	66.758	63.646	67.656
db8	61.901	65.469	59.922	66.029
db10	61.745	65.026	59.974	66.406

(c) Three-level Decomposition

Table 5.3: Average retrieval rates (%) for the selected set of images in the top 15 matches, using different Daubechies' filters and for 3 decomposition levels.

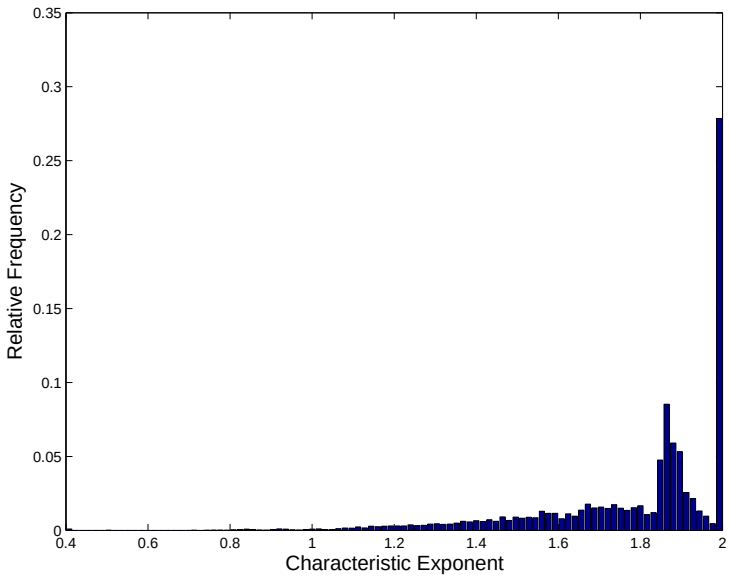


Figure 5.7: Histogram of the estimated values for the characteristic exponent, α, for the selected subset of 320 texture images of size 128×128.

tailed wavelet decompositions, obtained from the set of 20 512×512 textures with indices $\{2, 3, 5, 7, 9, 10, 12, 14, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30\}$. The corresponding wavelet decompositions had estimated feature values ($S\alpha S$ parameters α and γ) falling further apart than the range of the associated estimation confidence intervals. One should expect that the tighter the confidence intervals, the better the estimation of the features, and the better the retrieving performance. Figure 5.7 shows the histogram of the estimated characteristic exponent values.

Indeed, Table 5.4 shows that the $S\alpha S$ & D_2 variant of the $S\alpha S$ & KLD method has an improved retrieval rate than the GGD & KLD method. We observed that large confidence intervals in the estimation of the characteristic exponent values, permit the values of α between two distinct images to fall very close to each other. Hence, inaccuracies in the feature extraction process impede the similarity measurement task of texture retrieval.

Figure 5.8 shows the average percentages of retrieving relevant subimages as a function of the number of top matches, using 'db8' filter. We can see that the proposed scheme converges faster than the other methods. Figure 5.9 shows the comparison between the four retrieval methods on each texture class using a 3-level wavelet transform with the 'db4' filters. All the methods extract the same number of features ($2 \times 9 = 18$) from each image.

5.6 Conclusions

In this chapter we presented a $S\alpha S$ statistical model in the context of CBIR. We demonstrated that in many cases the modeling of wavelet decompositions of texture images using univariate $S\alpha S$ distributions is superior than that obtained using other non-Gaussian distributions. The construction of a suitable similarity function results in an increased retrieval performance, for a selected set of textures from VisTex and Brodatz databases.

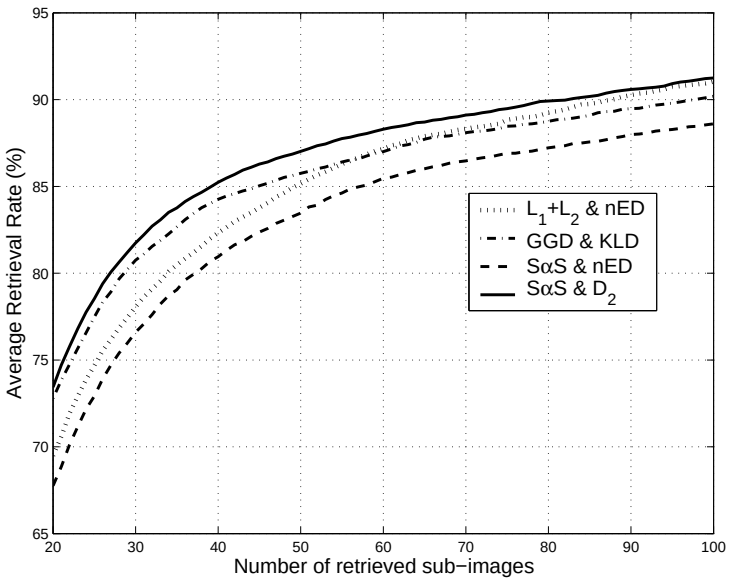


Figure 5.8: Retrieval performance according to the number of top matches considered.

The experimental results showed that an increased retrieval performance is not only due to a set of texture-specific extracted features, but also together with a suitable similarity measure. Finally, we observed that in most textures the predominant discrimination information lies in the first two decomposition levels, since the inclusion of the third level yields a little improvement in terms of the retrieval rates. This is also due to the fact that the widest confidence intervals for the estimated α values occur in the third decomposition level.

Filters	Methods			
	$L_1 + L_2$ & nED	GGD & KLD	$S\alpha S$ & nED	$S\alpha S$ & D_2
db2	62.522	73.872	69.683	75.846
db4	63.433	73.22	68.121	71.81
db8	60.829	67.556	64.431	68.641
db10	62.261	68.121	60.742	67.904

(a) One-level Decomposition

Filters	Methods			
	$L_1 + L_2$ & nED	GGD & KLD	$S\alpha S$ & nED	$S\alpha S$ & D_2
db2	69.618	79.839	78.95	82.14
db4	70.074	78.646	78.45	79.102
db8	72.287	74.718	70.833	76.519
db10	71.701	74.783	69.444	76.801

(b) Two-level Decomposition

Filters	Methods			
	$L_1 + L_2$ & nED	GGD & KLD	$S\alpha S$ & nED	$S\alpha S$ & D_2
db2	73.155	79.536	76.953	82.183
db4	72.721	77.496	76.324	81.011
db8	73.503	75.846	71.332	79.47
db10	73.568	76.128	71.072	79.926

(c) Three-level Decomposition

Table 5.4: Average retrieval rates (%) for the selected subset of images in the top 15 matches, using different Daubechies' filters and for 3 decomposition levels.

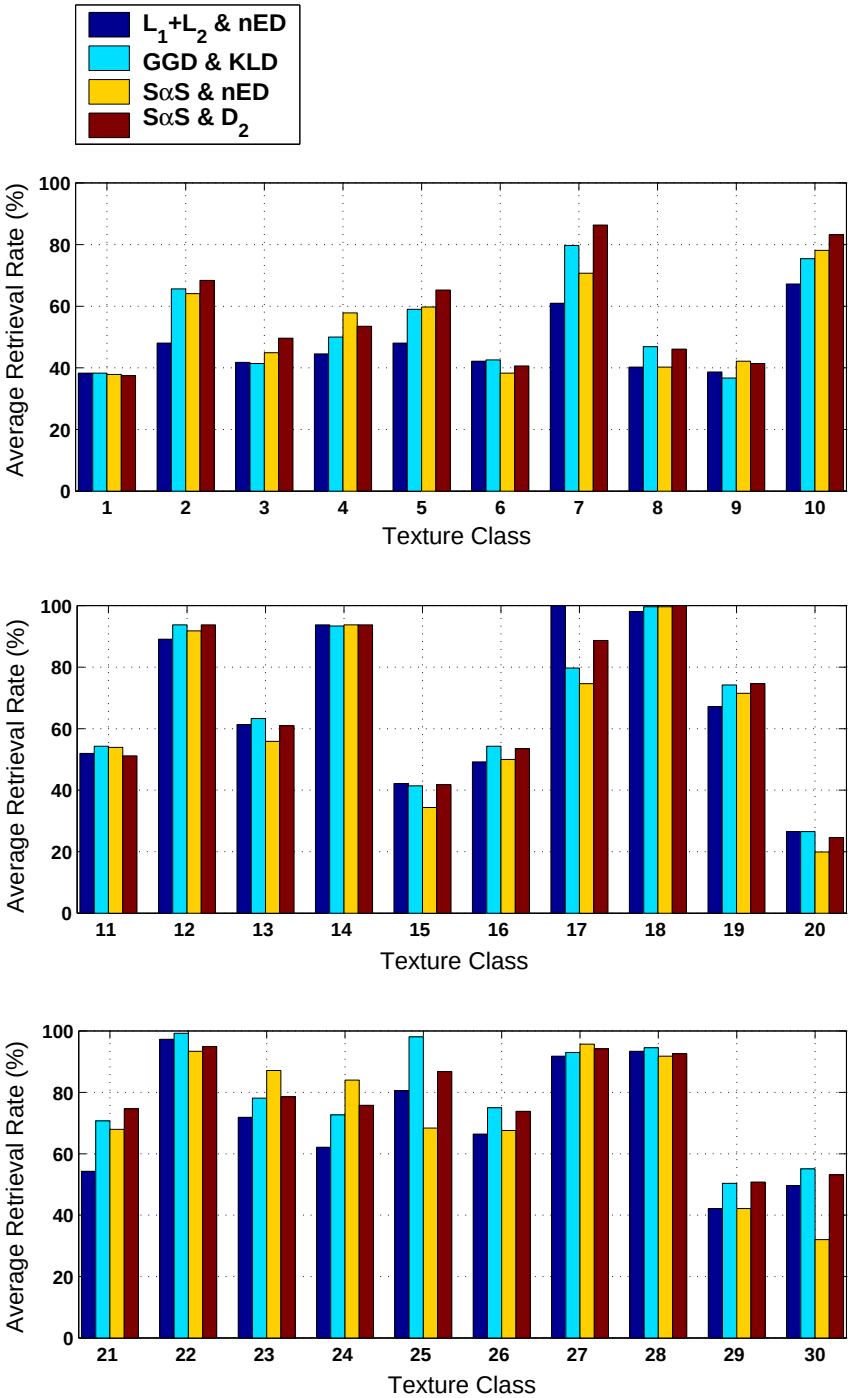


Figure 5.9: Average percentages (%) of correct retrieval rate for each individual texture class using a 3-level wavelet transform with 'db4' filters.

Chapter 6

Wavelet-Based Texture Retrieval via Multivariate sub-Gaussian Distributions

6.1 Introduction

In the previous chapter, we presented a method for wavelet-based retrieval of texture images. The basic assumption was that the modeling of marginal distributions of the wavelet subband coefficients using univariate $S\alpha S$ densities, provided an efficient representation of the texture information. We also illustrated that taking this heavy-tailed behavior into consideration resulted in an increased retrieval performance.

Although we obtained an improved retrieval scheme having low complexity, however, this approach does not take into account the possible interdependencies between different subbands (orientations) of a given image, which can be employed in order to provide a more accurate representation of the texture image profile. In section 3.3 we briefly reviewed the correlation properties of wavelet transform coefficients at different subbands and resolution levels. Several authors exploited a joint statistical characterization of the wavelet subband coefficients, developing algorithms by setting different constraints on the correlation between the raw coefficients, their magnitudes, and other statistics, in applications such as texture synthesis [93], image denoising [94] and compression [12].

The theory of Markov random fields has enabled a new generation of statistical texture models, in which the full model is characterized by statistical interactions within local neighborhoods [28]. Recently, a new framework for statistical signal processing based on wavelet-domain hidden Markov models has been proposed [30]. It provides an attractive modeling of both the non-Gaussian statistics and the persistence across scales property of a wavelet decomposition. The capacity of this model in capturing the inter-subband and inter-scale dependencies, can be enhanced by grouping wavelet coefficients and modeling them using multidimensional hidden Markov models in the wavelet domain [32].

In the present chapter, we proceed by considering the wavelet subband coefficients at each decomposition level as samples of a sub-Gaussian random process. The sub-Gaussian model and the associated fractional lower-order statistics have been used in the past to develop data-adaptive algorithms for signal detection in impulsive interference [120, 121]. Efficient methods for the fast estimation of the model parameters have also been provided [124]. Within the framework of sub-Gaussian processes, we use the notion of covariation instead of the second-order correlation, in order to extract possible interdependencies between wavelet coefficients at different image orientations.

This chapter is organized as follows. In the next section we extent the univariate $S\alpha S$ model, by jointly considering the wavelet coefficients at different orientations, in order to capture their cross-dependencies. Working in a statistical framework, we also describe the corresponding Feature Extraction (FE) and Similarity Measurement (SM) steps. In the section 6.3 we provide an illustration of the effectiveness of the above methods in a texture retrieval application, concluding in section 6.4.

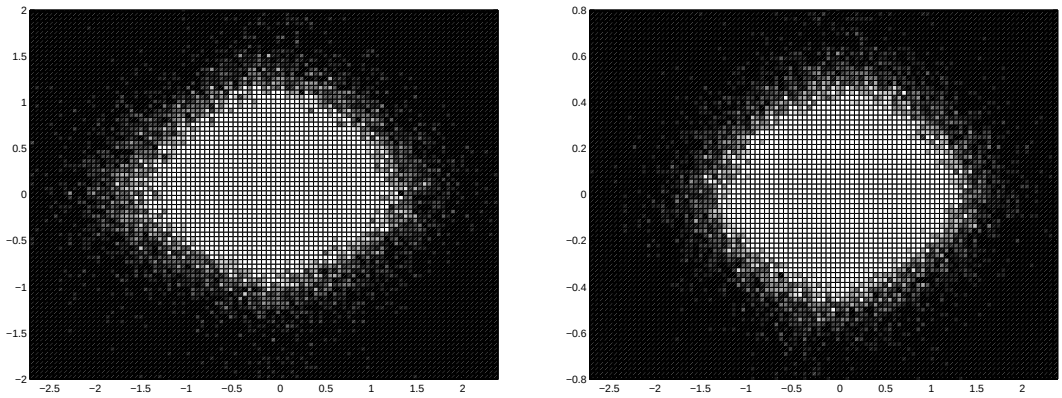
6.2 Joint sub-Gaussian Modeling of Wavelet Coefficients

In this section, we enhance the capability of the univariate $S\alpha S$ model, introduced in the previous chapter, which modeled the marginal distributions of the wavelet subband coefficients. We proceed by using a contiguous statistical model that captures both wavelet subband marginal distributions and inter-subband correlations. Experimental results show the importance of the cross-correlation of each subband with other orientations at the same decomposition level in characterizing the texture information [108].

In the proposed scheme, we tie up the wavelet coefficients at the same location and level but different subbands (orientations), to form a sub-Gaussian vector (cf. definition 4.2.2). This modeling of the subband coefficients preserves the heavy-tailed behavior of their marginal distributions. Notice that, by definition, each component of a sub-Gaussian vector is a $S\alpha S$ random variable and remember that we have already assessed the effectiveness of a $S\alpha S$ density function for the approximation of the empirical density near the mode and on the tails (cf. section 5.3). Besides, the components of a sub-Gaussian vector are highly dependent and this makes the joint sub-Gaussian model appropriate for capturing the cross-dependencies between different subbands, since around "features" in an image, such as edges and lines, the wavelet coefficients at all subbands are correlated in the sense that they have high probability of being significant. Figure 6.1 displays the joint histograms for pairs of coefficients at the same level and different orientations of the Brodatz texture *1.2.04*. In this case we can see that the joint distribution is almost Gaussian.

For convenience, we repeat the definition of a so-called α-sub-Gaussian process (α-SG($\mathbf{R}$)), which is a variance mixture of Gaussian processes:

Definition: Let $\{G(t), t \in T\}$ be a Gaussian process with covariance function R(u,v) and



(a) (b)

Figure 6.1: Joint histograms for the Brodatz texture *1.2.04*, between the horizontal subband at the first level and: (a) the vertical and (b) diagonal subbands at the same level. Intensity corresponds to frequency of occurence.

$A \sim S_{\alpha/2}((\cos\frac{\pi\alpha}{4})^{2/\alpha}, 1, 0)$ be a positive $\frac{\alpha}{2}$-stable random variable where $\alpha < 2$. Assume that the random variable A is independent of $\{G(t), t \in T\}$. The $S\alpha S$ process $\{X(t) = A^{1/2}G(t), t \in T\}$ is a sub-Gaussian process with an underlying Gaussian process $\{G(t), t \in T\}$. The finite dimensional projections, $(X(t_1), ..., X(t_d))$, $d \geq 1$, are α-SG($\mathbf{R}$) random vectors with underlying covariance matrix $\mathbf{R}$.

More precisely, we extract the interdependencies between pairs of subbands at a given level l by utilizing their joint statistics. For this purpose, we construct the set of vectors with each component corresponding to a subband and consider it as a sample of an α-sub-Gaussian process, α-SG($\mathbf{R}^l$), with $\mathbf{R}^l$ denoting the underlying covariance matrix corresponding to the subbands at the lth-level. In the standard 2-D DWT with 3 subbands at each decomposition level, H: horizontal, V: vertical, D: diagonal, we define these vectors as $\vec{X} = [X_H, X_V, X_D]^T$ (see Figure 6.2). Then, for the subbands at level l, we assume that the vectors $\vec{X}_k = [X_{H,k}, X_{V,k}, X_{D,k}]^T$, $k = 1, ..., N$, are samples taken from an α-SG($\mathbf{R}^l$) process, where N is the number of pixels of each of the subbands and the subscript k goes through all the spatial locations of the subband.

The notion of covariance between two random variables plays an important role in the second-order moment theory. However, covariances do not exist in the family of $S\alpha S$ random variables, due to the lack of finite variance. Instead, a quantity called *covariation*, which under certain constraints plays an analogous role for $S\alpha S$ random variables to the one played by covariance for Gaussian random variables, has been reviewed in section 4.2.4.

6.2.1 Feature Extraction

Applying the FE step of the proposed method at the l-th decomposition level of a given image, we proceed in two steps: first we estimate the characteristic exponent of the sub-Gaussian process. This value can be estimated on the basis of the observation that each component in the sub-Gaussian vector is a $S\alpha S$ random variable of the same characteristic exponent, α, as the vector [97]. For instance, we could estimate the α parameter for a given level l, using all the

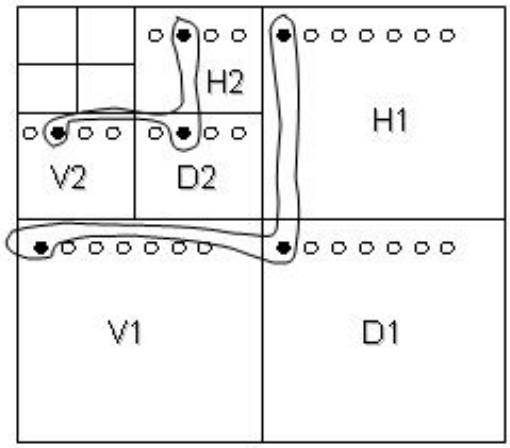


Figure 6.2: Sub-Gaussian vector model for the wavelet subbands at the same decomposition level.

wavelet subband coefficients at this level, $\{\{X^l_{H,k}\},\ \{X^l_{V,k}\},\ \{X^l_{D,k}\}\},\ k = 1, ..., N$. Another way to estimate α is to estimate its value at each individual subband and then use the mean of these estimated characteristic exponents as the value of α corresponding to the given decomposition level. The mean operator acts like the above merging of all subband coefficients into a single vector, which is then used to estimate α.

However, averaging in these two ways is going to ignore the inherent variability between each different pair of subbands (either obtained from different orientations and/or scales), since two distinct images with very different α values may have the same mean, which means that they are supposed to have the same multivariate distribution, which is obviously undesired. We overcome this problem by referring to the following intuitive arguments:

- the covariation between two $S\alpha S$ random variables assumes that they have the same characteristic exponent
- from Eq. (4.31), we also observe that in the expression of the covariation estimator the free p parameter affects the second variable (as an exponent)
- the intuitive idea is to use the estimated characteristic exponent, α, corresponding to the second variable/subband in order to estimate the covariation: for instance, in order to estimate $[X, Y]_\alpha$ we first estimate α from Y and then assume that X follows a distribution with the same α, while in order to estimate $[Y, X]_\alpha$ we estimate α from X and then assume that Y also follows a distribution with that α. This procedure exploits the differences between the subbands, regarding their distribution.

The second step of the proposed method consists of the estimation of the covariation matrix $\mathbf{C}^l$, with elements the covariations between the components of the α-SG($\mathbf{R}^l$) vectors. Starting from the definition of the covariation between two $S\alpha S$ random variables, the elements $[\mathbf{C}^l]_{mn} = [X^l_m, X^l_n]_\alpha$ of the covariation matrix $\mathbf{C}^l$ can be estimated from Eq. (4.31), which results in a

smaller standard deviation compared with the estimator given by Eq. (4.30) (for convenience we set $X_1 = X_H$, $X_2 = X_V$, $X_3 = X_D$).

Repeating this estimation process for each decomposition level, we define the signature $\mathcal{S}$ of an image I to be the set of the *estimated covariation matrices*:

$$I \mapsto \mathcal{S} = \{\mathbf{C}_I^1, \mathbf{C}_I^2, \ldots, \mathbf{C}_I^L\}$$

where $\mathbf{C}_I^l$ is the covariation matrix of the l-th decomposition level and L the number of levels. In the general case of L levels and J wavelet subbands per level, the total size of the above signature equals:

$$size(\mathcal{S}) = L \cdot J^2 \ ,$$

and this is due to the fact that the covariation of two $S\alpha S$ random variables is not symmetric.

Besides, corollary 4.2.1 states that if $\vec{X}$ is an α-sub-Gaussian random vector with underlying Gaussian vector $\vec{G}$, then there is a one-to-one correspondence between the probability distribution of $\vec{G}$ and that of $\vec{X}$. Since the probability distribution of the underlying Gaussian vector is determined by its covariance matrix, we may define a signature $\mathcal{S}_G$ of an image I as the set of the *estimated underlying covariance matrices*, computed from the corresponding covariation matrices at each decomposition level:

$$I \mapsto \mathcal{S}_G = \{\mathbf{R}_I^1, \mathbf{R}_I^2, \ldots, \mathbf{R}_I^L\}$$

where $\mathbf{R}_I^l$ is the underlying covariance matrix of the l-th decomposition level. Note that due to the symmetric property of the covariance matrix, the total size of this signature for L levels and J wavelet subbands per level equals:

$$size(\mathcal{S}_G) = \frac{J(J+1)}{2} \cdot L$$

Notice that this size is equal to the size of the signature obtained by employing a univariate $S\alpha S$ model. For instance, in the case of a 3-level wavelet decomposition of an image I, we saw that modeling the marginal distributions of the subbands results in a set of $3 \times 6 = 18$ features (6 model parameters per level). This is also the size of the signature $\mathcal{S}_G$, since each covariance matrix is determined by 6 elements and we have 3 such matrices. The advantage of using $\mathcal{S}_G$ is that it contains more texture-specific information. Specifically, the diagonal elements of a covariance matrix $\mathbf{R}_I^l$ contain information about the marginal distribution of each wavelet subband, while the off-diagonal elements express the interdependencies across subbands at the l-th level.

The above observations state that if two distinct images I, Q are similar in terms of their corresponding multivariate α-sub-Gaussian distributions, they should also be similar in terms of the associated underlying multivariate Gaussian distributions. So, we can define similarity measures between either the signatures $\mathcal{S}$ or the signatures $\mathcal{S}_G$.

Estimate optimal p parameter

In section 4.2.4, we saw that the estimation of covariations and consequently the estimation of covariation matrices, depends on a free parameter p ($0 \leq p \leq 2$) which, in turn, depends on the value of the characteristic exponent α. After implementing a number of Monte-Carlo simulations, for different values of α, we concluded that the $FLOM$ covariation estimator, given by Eq. (4.31), results in an increased performance compared with that given by Eq. (4.30), in terms of a small standard deviation. The Monte-Carlo simulations estimated the covariation between two $S\alpha S$ random variables X and Y.

In the present chapter and in the following one, we employ the multivariate sub-Gaussian model and we need to estimate the covariations between the components of sub-Gaussian vectors. We repeated the Monte-Carlo simulations, as described in section 4.2.4, using two sub-Gaussian random variables:

$$\begin{aligned} X &= A^{1/2} G_X \\ Y &= A^{1/2} G_Y \end{aligned}$$

By definition, X and Y can be viewed as $S\alpha S$ random variables with dispersion γ_X and γ_Y, respectively. We generate a sample of a sub-Gaussian random variable by first generating a sample A drawn from a $S_{\alpha/2}((\cos\frac{\pi\alpha}{4})^{2/\alpha}, 1, 0)$ distribution and then by generating a sample G drawn from a zero-mean Gaussian distribution with variance $2\gamma^2$, which is viewed as $S_2(\gamma, 0, 0)$ (with $\gamma = \gamma_X$ or $\gamma = \gamma_Y$ whether the Gaussian part G corresponds to the variable X or Y, respectively).

Figure 6.3 displays the curves representing the standard deviation of the $FLOM$ covariation estimator as a function of p, for three values of α and 25 pairs of dispersions (γ_X, γ_Y), with the dispersions ranging in the interval (0, 3.5), as estimated from the wavelet subbands of the selected images used in our experiments.

For each α, we observe that all the curves are minimized in a common interval on the p-axis, and actually their optimal value for p is close to each other. We repeated the procedure for $\alpha = 1 : 0.05 : 2$ and for a given α_i we defined the optimal p_i as the mean of the optimal p values of its corresponding 25 curves. In the section of the experimental results at the end of this chapter, as well as at the next chapter, we use these values for the optimal p as a function of α, shown in the Table 6.1, which is used as a lookup table in order to find the optimal p for every $1 < \alpha \leq 2$ by interpolating these values.

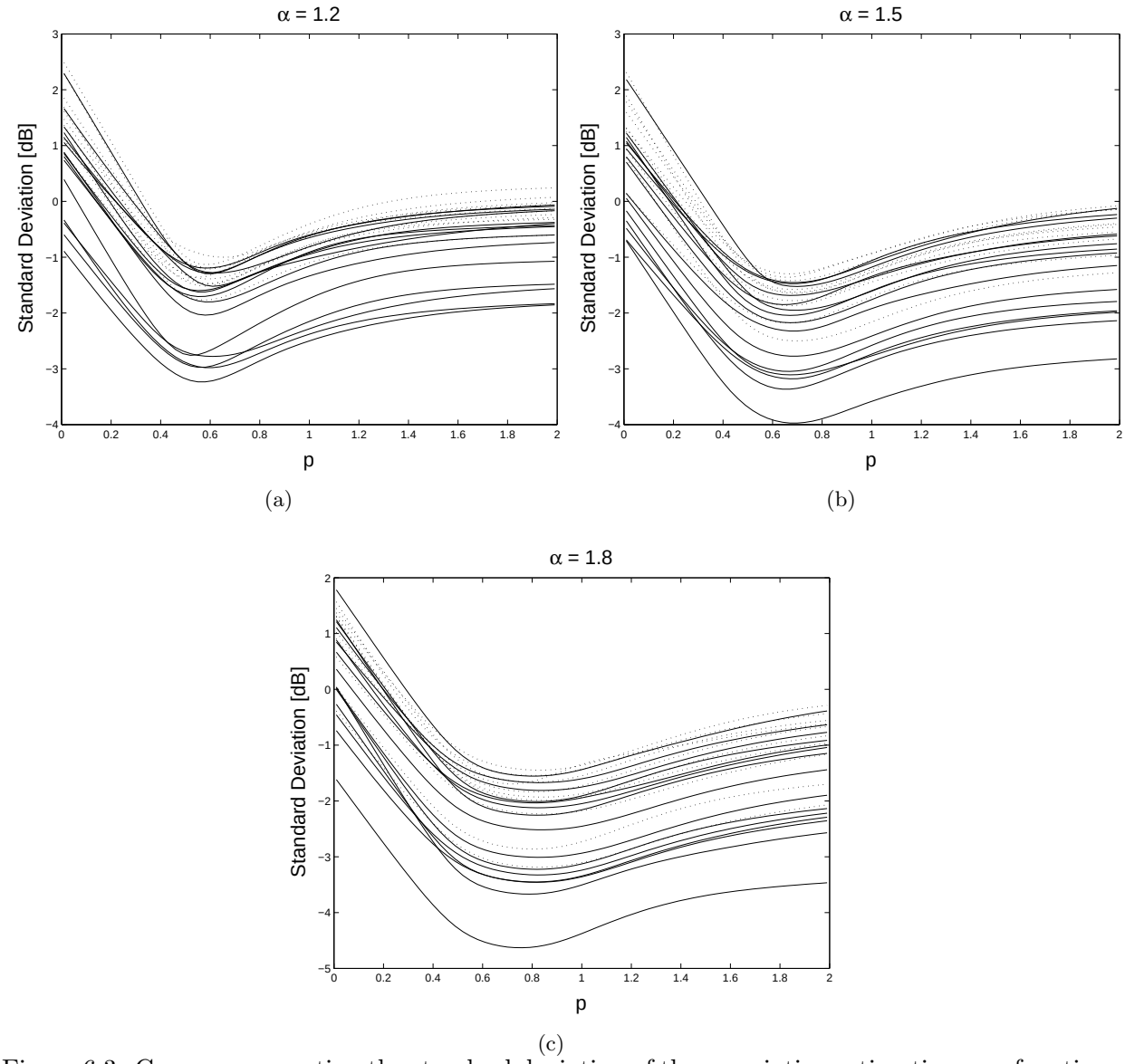


Figure 6.3: Curves representing the standard deviation of the covariation estimation as a function of the parameter p for $\alpha = 1.2, 1.5, 1.8$ and 25 dispersion pairs (γ_X, γ_Y), using the $\hat{\mathbf{c}}^{FLOM}$ estimator.

6.2.2 Similarity Measurement

When the FE step consists in representing the texture information of an image using the covariation matrix for each decomposition level, a way to measure the distance between two corresponding levels of two distinct images, is to take a matrix norm of the difference of the two covariation matrices. Assuming independency across levels, the overall distance between two images I and Q (where I is the query image) is given as the sum of the squared distances between the corresponding decomposition levels:

$$D(I, Q) = \sum_{l=1}^{L} \|\mathbf{C}_I^l - \mathbf{C}_Q^l\|^2 \tag{6.1}$$

α	Optimal p	α	Optimal p
1	0.52	1.5	0.69
1.05	0.54	1.55	0.71
1.1	0.56	1.6	0.72
1.15	0.57	1.65	0.74
1.2	0.58	1.7	0.76
1.25	0.59	1.75	0.79
1.3	0.61	1.8	0.81
1.35	0.62	1.85	0.84
1.4	0.64	1.9	0.88
1.45	0.66	1.95	0.93
		2	0.8

Table 6.1: Optimal p parameter as a function of the characteristic exponent α.

where L is the number of decomposition levels and $\|\cdot\|$ may be anyone of the commonly used matrix norms. In the experiments we will use the Frobenius norm ($\|\cdot\|_F$) of an $m \times n$ matrix $\mathbf{A}$, which gives an indication of the matrix "magnitude" and is defined as the square root of the sum of the squares of its elements:

$$\|\mathbf{A}\|_F \triangleq \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(\mathbf{A}_{ij})^2} = \sqrt{tr(\mathbf{A}\mathbf{A}^H)}\ , \tag{6.2}$$

where $tr(\cdot)$ denotes the matrix trace and $\mathbf{A}^H$ is the conjugate transpose.

If the FE step consists in representing the texture-specific information using the signature containing the underlying covariance matrices, the similarity between two images is measured by employing the Kullback-Leibler Divergence (KLD) between multivariate Gaussian distributions. In particular, let I^l, Q^l be the l-th decomposition level of two images I, Q, with underlying mean vectors $\vec{\mu}_{I^l}$, $\vec{\mu}_{Q^l}$ (i.e. the mean vectors of the underlying Gaussian vectors) and underlying covariance matrices $\mathbf{R}_{I^l}$, $\mathbf{R}_{Q^l}$, respectively. The KLD between these two levels is given by:

$$D(I^l\|Q^l) = \frac{1}{2}\Big((\vec{\mu}_{Q^l} - \vec{\mu}_{I^l})^T\mathbf{R}_{Q^l}^{-1}(\vec{\mu}_{Q^l} - \vec{\mu}_{I^l}) + tr(\mathbf{R}_{I^l}\mathbf{R}_{Q^l}^{-1} - I) - \ln|det(\mathbf{R}_{I^l}\mathbf{R}_{Q^l}^{-1})|\Big) \tag{6.3}$$

with $det(\cdot)$ denoting the determinant of a matrix.

By definition, the underlying Gaussian vector of an α-sub-Gaussian vector has zero mean. Thus, the above equation can be simplified resulting in the following KLD between I^l, Q^l:

$$D(I^l\|Q^l) = \frac{1}{2}\Big(tr(\mathbf{R}_{I^l}\mathbf{R}_{Q^l}^{-1} - I) - \ln|det(\mathbf{R}_{I^l}\mathbf{R}_{Q^l}^{-1})|\Big) \tag{6.4}$$

Making an assumption of independence across levels, the overall KLD between images I, Q is given by the sum:

$$D(I\|Q) = \sum_{l=1}^{L} D(I^l\|Q^l) \tag{6.5}$$

Figure 6.4 displays the above KLD for diagonal covariance matrices of the form $\mathbf{R} = \sigma^2 I$, for

different values of σ.

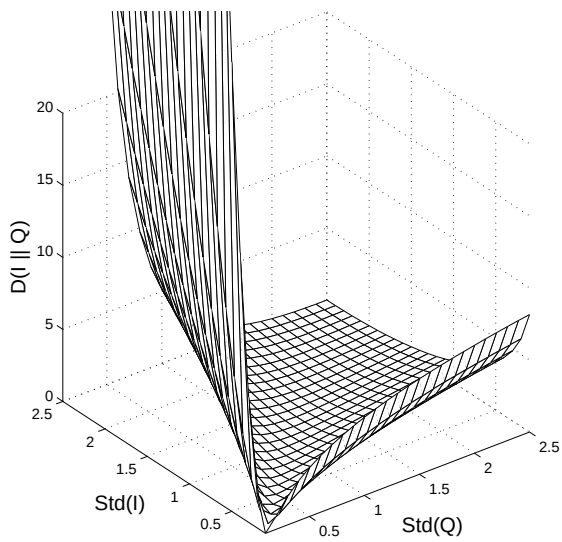


Figure 6.4: KLD between zero-mean multivariate Gaussian PDFs with covariance matrices of the form $\mathbf{R} = \sigma^2 I$.

6.3 Experimental Results

In this experiment, we evaluated the performance of the proposed CBIR method by applying it on the same set of texture samples, obtained from the Brodatz and VisTex collections, that has been used in section 5.5 (cf. Figure 5.5). As before, these real-world 512×512 natural scene images were divided into 16 128×128 non-overlapping subimages, each of which was normalized to zero mean and unit variance before applying the wavelet decomposition. We employed the standard 2-D wavelet transform with three decomposition levels using the Daubechies' orthogonal filters ('dbN'), for different values of N.

In the following retrieval experiments, each subimage in the database is simulated as a query and the relevant images for each query are defined as the other 15 subimages from the same original 512×512 image. The system's performance is evaluated in terms of the average rate of retrieving relevant images as a function of the number of top retrieved images.

6.3.1 Retrieval Performance

For this series of experiments, we used the 3-level wavelet transform with Daubechies' filters. As we concluded in section 5.5, the 3 decomposition levels are enough since most of the texture-specific information of the selected texture images is concentrated in these levels.

Each image I in the database is associated with a signature $\mathcal{S}(I)$ containing the three estimated covariation matrices. That is, each image is represented by 27 parameters (9 covariations per level). The similarity measurement between two distinct images is performed using the sum of squared Frobenius norms given by Eq. (6.1).

We also conducted a second experiment, where the signature of each image I ($\mathcal{S}_G(I)$) contains the 3 underlying covariance matrices. The advantage of this representation is that each image is described by 18 parameters (6 covariances per level), since the covariance matrices are symmetric. The similarity between two images is measured by employing the KLD between zero-mean multivariate Gaussian densities, given by Eq. (6.5).

Table 6.2 shows the average retrieval performance on the database of 480 texture images by implementing the univariate $S\alpha S$ model and the corresponding KLD, covariation matrices combined with Frobenius norm and underlying covariance matrices with the associated KLD, to characterize wavelet coefficients and measure their similarity.

	Methods		
Filters	$S\alpha S$ & D_2	$\mathcal{S}$ & Frobenius	$\mathcal{S}_{\mathcal{G}}$ & KLD
db4	58.633	52.031	50.716
db8	57.044	50.208	61.302

(a) One-level Decomposition

	Methods		
Filters	$S\alpha S$ & D_2	$\mathcal{S}$ & Frobenius	$\mathcal{S}_{\mathcal{G}}$ & KLD
db4	65.495	54.831	67.214
db8	63.854	54.818	69.01

(b) Two-level Decomposition

	Methods		
Filters	$S\alpha S$ & D_2	$\mathcal{S}$ & Frobenius	$\mathcal{S}_{\mathcal{G}}$ & KLD
db4	67.656	46.419	70.195
db8	66.029	44.089	70.286

(c) Three-level Decomposition

Table 6.2: Average retrieval rates (%) for the selected set of images in the top 15 matches, using different Daubechies' filters and for 3 decomposition levels.

Figure 6.5 details the retrieval performance on each individual texture class using a 3-level wavelet transform with the 'db4' filters, while Figure 6.6 shows the average percentages of retrieving relevant subimages as a function of the number of top matches. The three curves correspond to the following combinations: **1.** $\mathcal{S}_{\mathcal{G}}$ & KLD, where the signature consists of the underlying covariance matrices estimated using the lower-than-two order covariations and the similarity function is given by Eq. (6.5), **2.** $S\alpha S$ & D_2, where the signature contains the ML estimated parameters of the univariate $S\alpha S$ model and the similarity function is given by Eq. (5.17) and **3.** 2nd order & KLD, where the signature contains the *sample covariance* matrices estimated *directly from the wavelet subband coefficients.*

From Table 6.2, we observe that the significant texture-specific information is contained in the first two decomposition levels. We can also see that when we use the signature $\mathcal{S}$ combined with the Frobenius matrix norm, the inclusion of the coarsest level in the similarity function reduces the retrieval performance. A reason for this, is that due to the small sample size at the

third level the estimation of the covariations is not quite accurate, since the free p parameter depends on α which has large confidence intervals in that level. So, two distinct images with different covariation matrices at the first two levels may have similar covariation matrices at the lowest level, which brings them "closer". On the other hand, we can see that combining the signature $\mathcal{S}_{\mathcal{G}}$ with the KLD between multivariate Gaussian densities, results in an improved retrieval performance compared with that of the univariate $S\alpha S$ model.

Figure 6.6 shows that the second order statistics, as expressed by the sample covariance matrices, are not adequate in capturing the true distribution of the wavelet coefficients and consequently they are inadequate in fully extracting the texture-specific image content. On the other hand, the lower order statistics, even in the univariate $S\alpha S$ case, are proven to be superior in exploiting the texture-specific information. Another important conclusion of this experiment is that the retrieval performance does not only depend on the Feature Extraction step, but also on the choice of the Similarity Measure. In particular, a statistical similarity function, such as KLD, is superior compared with a deterministic one, such as the Frobenius norm.

6.4 Conclusions

In this chapter, we enhanced the capacity of the univariate $S\alpha S$ model by exploiting the interdependencies between subband coefficients at each decomposition level. This is achieved by tying up the wavelet subbands at the same level into a sub-Gaussian vector. The joint sub-Gaussian modeling preserves the heavy-tailed behavior of the marginal distribution of each subband, as well as the strong statistical dependence across subbands at a given decomposition level.

The experimental results verified that the achievement of an increased retrieval performance does not only depend on the set of the extracted features, representing the image content, but also on the choice of a suitable measure of similarity between image signatures. In particular, the application of a statistical similarity function, such as KLD between multivariate densities, presents an increased discriminating power compared with that of a deterministic similarity function, such as a matrix norm. It is also confirmed that the lower-than-two order statistics are preferable in exploiting the texture-specific information, instead of using second order statistics.

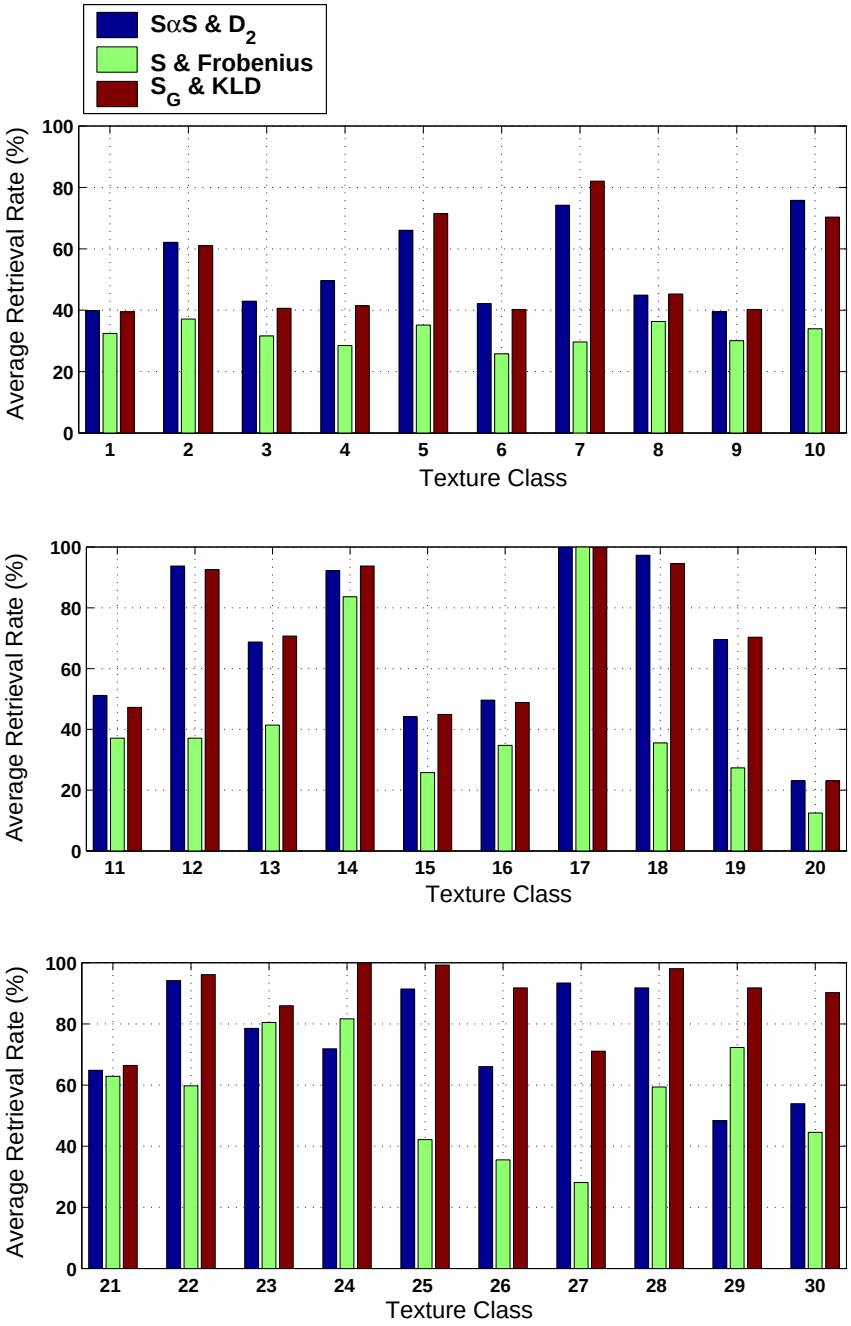


Figure 6.5: Average percentages (%) of correct retrieval rate for each individual texture class using a 3-level wavelet transform with 'db4' filters.

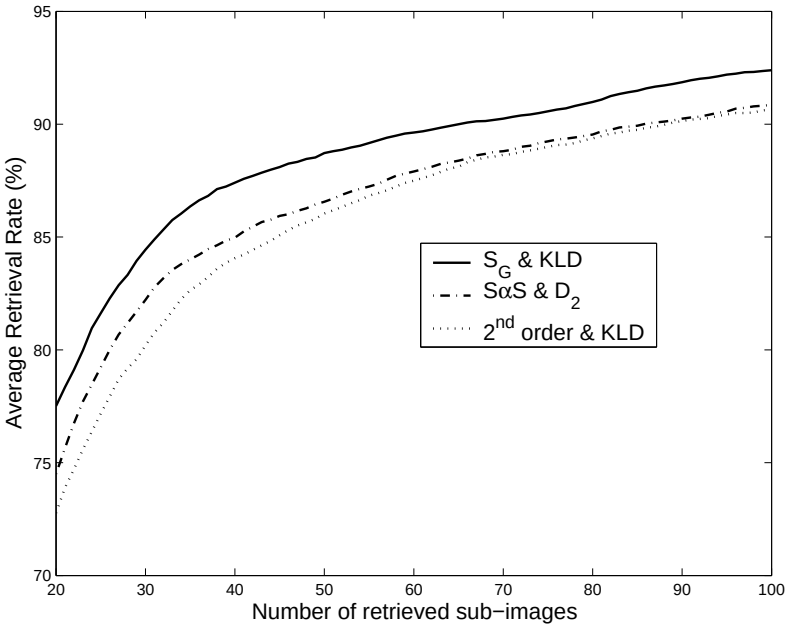


Figure 6.6: Retrieval performance according to the number of top matches considered, using Daubechies' "db4" filter.

Chapter 7

Rotation-Invariant CBIR via Joint sub-Gaussian Modeling of Steerable Pyramids

7.1 Introduction

In the previous two chapters, we introduced the univariate $S\alpha S$ model and the multivariate sub-Gaussian model in the context of a wavelet-based texture retrieval application. Both of these models have the drawback that they are sensitive to the orientation of the given query image. On the other hand, a very important property of a CBIR system is rotation invariance, that is, given a query image the system must be able to retrieve the relevant images in the case of a database containing images along with rotated versions of them.

Actually, the problem of insensitivity under rotations, that the above methods present, is due to the fact that the standard wavelet transform lacks the translation and rotation invariant properties. This results in a mismatch of the retrieval process when the image orientation varies. In fact, the wavelet coefficients of the rotated image will be completely different compared with the wavelet coefficients of its original version.

A way to overcome this problem is to replace the wavelet transform with a steerable pyramid (cf. section 3.2.3). A steerable pyramid results in a linear multi-scale, multi-orientation image decomposition produced by a set of orientation filters generated by a set of basis functions (directional derivative operators). Steerable pyramids are overcomplete and possess the desired properties of translation and rotation invariance.

Rotation invariance is a topic that has been previously pursued by various researchers. Greenspan *et. al.* [49] and Haley and Manjunath [52] used rotation-invariant structural features obtained via multiresolution Gabor filtering, where rotation invariance is achieved by using the magnitude of a discrete Fourier transform (DFT) in the rotation dimension. The last two authors also achieve rotation invariance by transforming Gabor into rotation-invariant features (using autocorrelation and DFT magnitudes) and by utilizing rotation-invariant statistics of ro-

tation dependent features [53]. Recently, a rotation-invariant image retrieval system based on steerable pyramids was proposed by Beferull-Lozano *et al.* [6]. In this system, the correlation matrices between the basic orientations at each level of the pyramid are chosen as the energy-based texture features. Then, the similarity measure between two images is defined as the minimum Frobenius norm, over all possible rotations θ, of the difference between the correlation matrix of the original (query) image and that of each image in the database. Mao and Jain [79] presented a multiresolution simultaneous autoregressive (MR-SAR) model for texture classification and segmentation. A multivariate rotation-invariant SAR (RISAR) model is introduced which is based on the circular autoregressive (CAR) model.

A second category of methods achieving rotation invariance includes the implementation of a Hidden Markov Model (HMM) on the subband coefficients of the transformed image. Wu and Wei [133] obtained rotation-invariant features, first by converting the 2-D texture image to an 1-D signal via spiral sampling and then by applying a subband decomposition to the 1-D signal, followed by an HMM on the subband coefficients. Do and Vetterli [32] derived a steerable rotation-invariant statistical model by enhancing a recently introduced technique on wavelet-domain HMM [30], for better dealing with images by incorporating the dependency of steerable pyramid coefficients across orientations. Liu and Picard [67] exploited the effectiveness of the 2-D Wold decomposition of homogeneous random fields, in order to extract features that represent perceptual properties which can be described as "periodicity", "directionality" and "randomness", approximating what are indicated to be the three most important dimensions of human texture perception.

The above rotation-invariant CBIR techniques could be classified in two categories. The first category includes the techniques whose Feature Extraction step consists of computing rotation-invariant texture features, while the Similarity Measurement step consists of applying a common similarity function, such as the Euclidean distance and the KLD. The second category includes the techniques whose Feature Extraction step consists of estimating the parameters of a so-called *steerable model* and then by applying a rotation-invariant version of a known similarity function (e.g. KLD between Gaussian densities) during the Similarity Measurement step.

In this chapter, we describe two approaches belonging to the second category, that is, of designing a steerable model and then deriving a rotation-invariant similarity function. The first approach is based on the joint sub-Gaussian modeling of a steerable pyramid coefficients, incorporating dependence across angles and scales. We derive analytical expressions relating the fractional lower-order statistics of a rotated image with that of its original version. In the second approach we apply a Gaussianization procedure on the steerable pyramid coefficients, by jointly considering them as samples of a multivariate sub-Gaussian distribution, viewed as a special case of a Gaussian Scale Mixture (GSM). After the Gaussianization step, we can use an analytical expression for the KLD between two multivariate Gaussian densities, avoiding the not-so-good Monte-Carlo method, usually used to approximate the KLD in the non-Gaussian case [32].

Also, HMMs require the use of an Expectation-Maximization (EM) algorithm, which in some

cases may not converge, for the estimation of the model parameters (hidden state variables and statistics of a Gaussian mixture). On the other hand, our proposed method incorporates dependence across space, orientations and scales, combined with an efficient way of estimating the multiplier of the multivariate sub-Gaussian model. Besides, by including some of the next lower scale coefficients in the joint sub-Gaussian model, we insert the same first-order Markovian dependence as in HMMs, but in a simpler way.

The rest of the chapter is organized as follows: in the next section we describe a steerable multivariate sub-Gaussian model, by deriving analytical expressions that give the covariations of the pyramid subband pairs of an image oriented at an angle θ, from the corresponding covariations of its original version, without having to re-extract them from the rotated image. We also describe similarity measures which are appropriate in the development of a rotation-invariant CBIR system. Section 7.3 describes a Gaussianization procedure on the pyramid coefficients which results in a normalized pyramid decomposition, making the subsequent steps of Feature Extraction and Similarity Measurement easier and efficiently computed. We derive a rotation-invariant version of the Kullback-Leibler Divergence which is used as the similarity function. In the section 7.4 we illustrate the implementation of the two approaches on a set of texture images including rotated version of them at different angles. Finally, section 7.5 summarizes the retrieval results and concludes with some discussion and future work.

7.2 Steerable Multivariate sub-Gaussian Model

The desired rotation invariance property can be achieved by considering a FE step based on the subbands in multiple scales and orientations, obtained from a steerable pyramid [106, 43]. In the case of a database containing images along with rotated versions of them, we are interested in finding features which are as "steerable" as possible, that is, given the features of an image oriented at an angle ϕ, we should be able to obtain the features corresponding to the same image rotated at an angle θ, without having to re-extract the features from the rotated image[1].

Let $c^l(x_k, \phi)$ represent the value of a transform coefficient at a spatial location x_k, orientation ϕ and level l. In a steerable pyramid with J basic orientations (subbands) and L levels, at each level l, given the J basic coefficients

$$\{c^l(x_k, \phi_1), c^l(x_k, \phi_2), ..., c^l(x_k, \phi_J)\}$$

the transform coefficient $c^l(x_k, \phi)$ for any angle ϕ is given by [6]:

$$c^l(x_k, \phi) = \sum_{i=1}^{J} f_i(\phi) c^l(x_k, \phi_i) \quad \forall \phi, \quad l = 1, ..., L \tag{7.1}$$

where $\{f_1(\phi), f_2(\phi), ..., f_J(\phi)\}$ is the set of J steering functions.

[1]Through the next sections, we consider counter-clockwise rotation

Let $\mathbf{R}^l$ and $\mathbf{R}^l_\theta$ denote the sampled correlation matrices, with elements given by the correlations between pairs of subbands (at a given decomposition level l) of the original image I and its rotated version I_θ, respectively. Proposition 1 in [6] states that these two matrices are equivalent, i.e., they can be written in the form:

$$\mathbf{R}^l_\theta = \mathbf{F}(\theta)\mathbf{R}^l\mathbf{F}^T(\theta) \tag{7.2}$$

where

$$\mathbf{F}(\theta) = \begin{bmatrix} f_1(\phi_1-\theta) & f_2(\phi_1-\theta) & \cdots & f_J(\phi_1-\theta) \\ f_1(\phi_2-\theta) & f_2(\phi_2-\theta) & \cdots & f_J(\phi_2-\theta) \\ \vdots & \vdots & \vdots & \vdots \\ f_1(\phi_J-\theta) & f_2(\phi_J-\theta) & \cdots & f_J(\phi_J-\theta) \end{bmatrix} \tag{7.3}$$

In the following study the J basic angles are taken to be equispaced, which makes $\mathbf{F}(\theta)$ an orthogonal matrix for any θ ($\mathbf{F}(\theta)\mathbf{F}^T(\theta) = I_{J\times J} \Rightarrow \mathbf{F}^T(\theta) = \mathbf{F}^{-1}(\theta) = \mathbf{F}(-\theta)$), and thus, $\mathbf{R}^l$ and $\mathbf{R}^l_\theta$ become orthogonally equivalent. It can be easily shown, that an equivalence relation also holds for the covariance matrices corresponding to images I and I_θ (see Appendix B.1.1 for a proof).

Under a joint sub-Gaussian assumption, the coefficients of the J basic orientations (subbands) at a given level l are modeled as joint sub-Gaussian vectors, α-SG($\mathbf{R}^l$) (see definition 4.2.2). Let

$$X^l_k = \{c^l(x_k,\phi_1), c^l(x_k,\phi_2), ..., c^l(x_k,\phi_J)\},\ k = 1,\ldots,N,\ l = 1,\ldots,L$$

denote the vector containing the J basic coefficients at the spatial location k and level l, where N is the number of coefficients per subband and L the number of decomposition levels.

The pyramid coefficients at a given subband follow a sub-Gaussian marginal distribution. So, the coefficients corresponding to the basic orientation ϕ_i at level l can be expressed as:

$$c^l(x_k,\phi_i) = A^{1/2}c^l_G(x_k,\phi_i), \quad i = 1,...,J \tag{7.4}$$

where $c^l_G(x_k,\phi_i)$ is the Gaussian part of the α-SG($\mathbf{R}^l$) vector. From Eq. (7.1), the transform coefficient at any angle ϕ is given by:

$$\begin{aligned} c^l(x_k,\phi) &= \sum_{i=1}^{J} f_i(\phi)A^{1/2}\, c^l_G(x_k,\phi_i) \\ &= A^{1/2}\sum_{i=1}^{J} f_i(\phi)c^l_G(x_k,\phi_i) \\ &= A^{1/2}\, c^l_G(x_k,\phi) \end{aligned} \tag{7.5}$$

Eq. (7.5) shows that the pyramid subband coefficients of a rotated image at an angle ϕ, are also sub-Gaussian random variables with the same characteristic exponent with that of the corresponding subband of the original image, and a Gaussian part which is the rotated version of the

original Gaussian part at the same angle ϕ.

If we let $\mathbf{R}^l$ and $\mathbf{R}^l_\theta$ denote the underlying covariance matrices (corresponding to the underlying Gaussian random vector) of the l-th decomposition level of the original image I and its rotated version I_θ, respectively, then Eq. (7.2) indicates that they are equivalent.

In the section 6.2.1 we derived the signature $\mathcal{S}$ for a given image I, under the assumption of a multivariate sub-Gaussian model. In this case, the model parameters are the estimated covariation matrices. In the following, we derive the corresponding multivariate sub-Gaussian model on a steerable pyramid, if the input image is rotated by an angle θ. For convenience we omit the superscript l of the notation $\mathbf{R}^l$, denoting the lth-level underlying covariance matrix. The relation between the covariation matrix $\mathbf{C}$ and the underlying covariance matrix $\mathbf{R}$ is given by Eq. (4.26):

$$[\mathbf{C}]_{ij} = 2^{-\frac{\alpha}{2}}\,[\mathbf{R}]_{ij}\,[\mathbf{R}]_{jj}^{(\alpha-2)/2}\ .$$

This results in the following relation between the elements of the covariation matrix $\mathbf{C}_\theta$ of the rotated image and the corresponding elements of the underlying covariance matrix $\mathbf{R}_\theta$:

$$[\mathbf{C}_\theta]_{ij} = 2^{-\frac{\alpha}{2}}\,[\mathbf{R}_\theta]_{ij}\,[\mathbf{R}_\theta]_{jj}^{(\alpha-2)/2}\ . \tag{7.6}$$

The combination of Eqs. (7.2), (7.6) yields the following proposition, which shows that for the multivariate sub-Gaussian model on a steerable pyramid, if the input image is rotated, then the new model can be obtained by a simple transformation of the original multivariate sub-Gaussian model.

Proposition 7.2.1 *Suppose that $\mathcal{S} = \{\mathbf{C}^1, \mathbf{C}^2, \ldots, \mathbf{C}^L\}$ is the signature obtained from a multivariate sub-Gaussian modeling of a steerable pyramid for a given homogeneous texture image. Then the corresponding model for the rotated version of that texture by θ is $\mathcal{S}_\theta = \{\mathbf{C}^1_\theta, \mathbf{C}^2_\theta, \ldots, \mathbf{C}^L_\theta\}$ where*

$$\mathbf{C}^l_\theta = \mathbf{M}^l\ .*\ D(\mathbf{M}^l), \quad l = 1, \ldots, L \tag{7.7}$$

with

$$\mathbf{M}^l = \mathbf{F}(\theta)\,\tilde{\mathbf{C}}^l\,\mathbf{F}^T(\theta), \tag{7.8}$$

$$\tilde{\mathbf{C}}^l = \begin{bmatrix} \frac{[\mathbf{C}^l]_{11}}{[\mathbf{C}^l]_{11}^{(\alpha-2)/\alpha}} & \frac{[\mathbf{C}^l]_{12}}{[\mathbf{C}^l]_{22}^{(\alpha-2)/\alpha}} & \cdots & \frac{[\mathbf{C}^l]_{1J}}{[\mathbf{C}^l]_{JJ}^{(\alpha-2)/\alpha}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{[\mathbf{C}^l]_{J1}}{[\mathbf{C}^l]_{11}^{(\alpha-2)/\alpha}} & \frac{[\mathbf{C}^l]_{J2}}{[\mathbf{C}^l]_{22}^{(\alpha-2)/\alpha}} & \cdots & \frac{[\mathbf{C}^l]_{JJ}}{[\mathbf{C}^l]_{JJ}^{(\alpha-2)/\alpha}} \end{bmatrix} = \frac{1}{2}\,\mathbf{R}, \tag{7.9}$$

$$D(\mathbf{M}^l) = \begin{bmatrix} \{diag(\mathbf{M}^l)\}\ .\wedge\left(\frac{\alpha-2}{2}\right) \\ \vdots \\ \{diag(\mathbf{M}^l)\}\ .\wedge\left(\frac{\alpha-2}{2}\right) \end{bmatrix}, \tag{7.10}$$

where $\mathbf{R}$ is the underlying Gaussian matrix corresponding to the multivariate sub-Gaussian model

of the original image and with $(.*)$, $(.\wedge)$ *denoting element-by-element multiplication and element-by-element involution, respectively, and* $diag(\mathbf{M}^l)$ *is a row vector containing the main diagonal of the square matrix* $\mathbf{M}^l$. *The dimension of all the above matrices equals* $J \times J$, *where* J *is the number of orientations at each decomposition level.*

Proof. The detail of the proof is described in Appendix B. ■

The relation between the elements of the signatures $\mathcal{S}$ and $\mathcal{S}_\theta$, as well as between the underlying covariance matrices is shown in Figure 7.1.

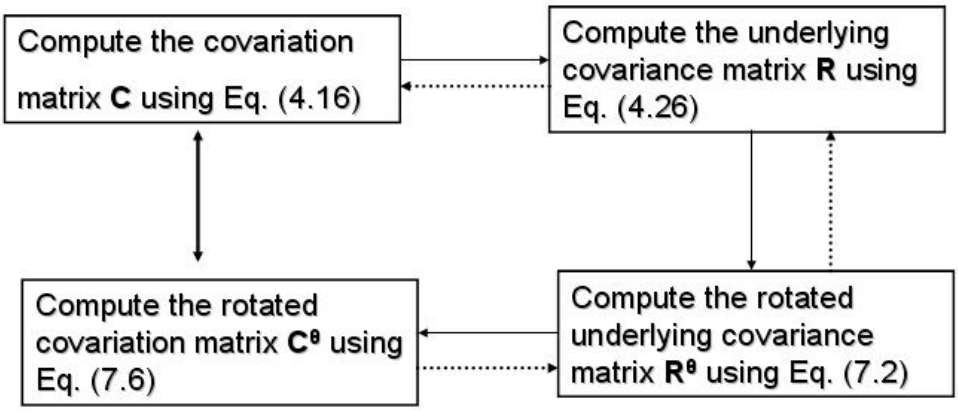


Figure 7.1: Relation between the covariation and underlying covariance matrices of signature $\mathcal{S}$, with the corresponding matrices of the transformed signature $\mathcal{S}_\theta$.

7.2.1 Feature Extraction

Under the multivariate sub-Gaussian model on the steerable pyramid subbands at a given decomposition level, the Feature Extraction step consists in estimating the $J \times J$ covariation matrix at each decomposition level. Thus, for a given image I decomposed in L levels, the corresponding signature $\mathcal{S}$ is given by the set of the L estimated covariation matrices:

$$I \mapsto \mathcal{S} = \{\mathbf{C}_I^1, \mathbf{C}_I^2, \ldots, \mathbf{C}_I^L\} \ .$$

The total size of this signature equals $L \cdot J^2$.

The signature $\mathcal{S}$ contains only the across orientation dependence at a given decomposition level. We may consider the across levels dependence by estimating the covariation matrices between consecutive levels (this is equivalent to a first-order Markovian dependence used in an HMM). In this case the signature of an image I is the following:

$$I \mapsto \mathcal{S}_\mathcal{E} = \{\mathbf{C}_I^1, \mathbf{C}_I^2, \ldots, \mathbf{C}_I^L, \mathbf{C}_I^{1\to 2}, \mathbf{C}_I^{2\to 3}, \ldots, \mathbf{C}_I^{(L-1)\to L}\} \ , \tag{7.11}$$

where $\mathbf{C}_I^{l\to(l+1)}$ denotes the covariation matrix corresponding to the subbands at levels l and $(l+1)$. In particular, the element $[\mathbf{C}_I^{l\to(l+1)}]_{ij}$ is equal to the covariation of the i-th subband

at level l with the j-th subband at the next lower level $(l+1)$. Obviously, the signature $\mathcal{S}_{\mathcal{E}}$ contains more texture-specific information than the signature $\mathcal{S}$, since it exploits not only the inter-subband but also the inter-level dependencies. On the other hand, the cost of using $\mathcal{S}_{\mathcal{E}}$ is an increased computational complexity, since its size equals

$$size(\mathcal{S}_{\mathcal{E}}) = L \cdot J^2 + (L-1) \cdot J^2 = (2L-1) \cdot J^2$$

Extract inter-level covariations

The construction of the enhanced signature $\mathcal{S}_{\mathcal{E}}$ requires the computation of covariations between subbands at different levels. Using the standard pyramid decomposition, we move from level l to the next lower level $(l+1)$ by subsampling the output of the low-pass filter (cf. Figure 3.5). As a result, the subbands at the $(l+1)$-th level are half in size than those of the l-th level, which is undesired since the covariation estimation includes summations between vectors of equal length. A way to overcome this difficulty is to upsample and interpolate the subbands at the coarser level, in order to equate their dimensions with the dimensions of the subbands at the previous finer decomposition level. Subsampling is good for compression and for saving memory and complexity when performing the decomposition. However, subsampling introduces some aliasing, which is not good for extracting features.

Instead of subsampling the output of the filters, we can upsample the filters and of course, every time we filter, we just do not subsample their outputs. At the same time, it allows us to keep the same number of coefficients across scales, which favors the computation of correlation (covariation) matrices across scales with the different orientations. There are two ways to perform this procedure, one in the space domain, convolving as usual the image and the filter, and another one where the implementation is in the Fourier domain, simply by multiplying the DFT of the image with the DFT of the filter and then taking the inverse DFT to obtain the subband coefficients. We follow the frequency-domain approach since, although both implementations should give approximately the same values, the frequency-domain implementation is more exact as it does not suffer from the finite-length constraint that is imposed on the filters for the convolution.

7.2.2 Similarity Measurement

In the development of a rotation-invariant CBIR system, except for a steerable model we have to specify an appropriate similarity measure. One way to measure the similarity between the images is to proceed as in [6], thus the distance between two images I, Q (where I is the query) is defined in the following proposition:

Proposition 7.2.2 *Under the steerable multivariate sub-Gaussian model represented by signatures in the form of $\mathcal{S}$, the distance between two images I and Q, where I is the query, is given by:*

$$D(I,Q) = \min_{\theta} \sum_{l=1}^{L} \|\mathbf{C}_I^l - \tilde{\mathbf{C}}_Q^l\| \tag{7.12}$$

where

$$\tilde{\mathbf{C}}_Q^l = 2^{-\alpha/2} \cdot \left(\tilde{\mathbf{R}}_Q^l \; .* \, D(\tilde{\mathbf{R}}_Q^l)\right) \; ,$$

with

$$\tilde{\mathbf{R}}_Q^l = \mathbf{F}(-\theta)\mathbf{R}_Q^l\mathbf{F}^T(-\theta) \; ,$$

and

$$\mathbf{R}_Q^l = 2 \cdot \left(\mathbf{C}_Q^l \; .* \, \bar{D}(\mathbf{C}_Q^l)\right) \; .$$

For any square matrix M, the operators $D(\cdot)$ and $\bar{D}(\cdot)$ are defined as:

$$D(\mathbf{M}) = \begin{bmatrix} \{diag(\mathbf{M})\} \; .\wedge \left(\frac{\alpha-2}{2}\right) \\ \vdots \\ \{diag(\mathbf{M})\} \; .\wedge \left(\frac{\alpha-2}{2}\right) \end{bmatrix} , \tag{7.13}$$

$$\bar{D}(\mathbf{M}) = \begin{bmatrix} \{diag(\mathbf{M})\} \; .\wedge \left(\frac{2-\alpha}{\alpha}\right) \\ \vdots \\ \{diag(\mathbf{M})\} \; .\wedge \left(\frac{2-\alpha}{\alpha}\right) \end{bmatrix} , \tag{7.14}$$

In the above expressions, $\mathbf{R}_Q^l$ is the l-th underlying covariance matrix corresponding to the multivariate sub-Gaussian model of image Q and with $(.*)$, $(.\wedge)$ denoting element-by-element multiplication and element-by-element involution, respectively, and $diag(\mathbf{M})$ is the row vector containing the main diagonal of $\mathbf{M}$. The dimension of all the above matrices equals $J \times J$, where J is the number of orientations at each decomposition level.

Proof. In order to prove the expression for the matrix $\tilde{\mathbf{C}}_Q^l$, we have to reverse the procedure shown in Figure 7.1, that is, to follow the dashed lines. Notice that the covariation matrix $\mathbf{C}_Q^l$ corresponds to image Q, which is supposed to be a rotated version of image I. So, we can imagine $\mathbf{C}_Q^l$ belonging to the steerable model, i.e. it should be viewed as the matrix given by Eq. (7.7). First we obtain the underlying covariance matrix $\mathbf{R}_Q^l$ from $\mathbf{C}_Q^l$ using Eq. (4.26). In particular, by solving first with respect to $[\mathbf{R}_Q^l]_{jj}$ and then with respect to $[\mathbf{R}_Q^l]_{ij}$ we have

$$[\mathbf{R}_Q^l]_{ij} = 2\,[\mathbf{C}_Q^l]_{ij}\,([\mathbf{C}_Q^l]_{jj})^{(2-\alpha)/\alpha}$$

which in matrix notation is written as

$$\mathbf{R}_Q^l = 2\left(\mathbf{C}_Q^l \; .* \, \bar{D}(\mathbf{C}_Q^l)\right) \; .$$

Then, by applying Eq. (7.2) on $\mathbf{R}_Q^l$ we obtain

$$\tilde{\mathbf{R}}_Q^l = \mathbf{F}(-\theta)\mathbf{R}_Q^l\mathbf{F}^T(-\theta) \; ,$$

and finally, we obtain the desired relation by employing Eq. (4.26) in matrix notation,

$$\tilde{\mathbf{C}}_Q^l = 2^{-\alpha/2} \cdot \left(\tilde{\mathbf{R}}_Q^l \,.*\, D(\tilde{\mathbf{R}}_Q^l)\right) .$$

■

The idea resulting in this similarity function is as follows: consider I to be the query image and Q to be an image in a database. In the SM step, we measure the closeness of I and Q by computing the distance between their signatures. As we may have a database of images along with rotated versions of them, we assume that Q is a rotated version of the given query I. So, the signature of image Q, $\mathcal{S}(Q)$, corresponds to the steerable model. For instance, in the case we consider only inter-subband dependencies, it is of the form

$$\mathcal{S}(Q) = \{\mathbf{C}_{Q,\theta}^1, \mathbf{C}_{Q,\theta}^2, \ldots, \mathbf{C}_{Q,\theta}^L\}$$

for an unknown angle θ. Before measuring the similarity between I and Q we have to align them, that is, to rotate clockwise the signature of Q. This clockwise rotation is expressed by the term $\tilde{\mathbf{C}}_Q^l$. Since we do not know the angle which gives the best alignment, we have to search over a set of possible rotations and this is expressed by the minimization operation over θ.

If we represent the texture information using the signature $\mathcal{S}_\mathcal{E}$ instead of $\mathcal{S}$, the distance between two images I and Q is measured using Eq. (7.12), by taking the difference between the corresponding matrices in the arrangement shown in (7.11). The only difference is in the number of summands, since the enhanced signature $\mathcal{S}_\mathcal{E}$ contains $2L-1$ terms, so we have to replace L with $2L-1$.

Under a Gaussian assumption, Eq. (7.12) is simplified and takes the following form [6]

$$D(I,Q) = \min_\theta \sum_{l=1}^{L} \|\mathbf{C}_I^l - \mathbf{F}(-\theta)\mathbf{C}_Q^l\mathbf{F}^T(-\theta)\| \tag{7.15}$$

7.3 Rotation-Invariant CBIR with Gaussianized Steerable Pyramids

In the previous section, we described a steerable multivariate sub-Gaussian model exploiting the pyramid subband coefficients dependence across orientations and scales. In this section, we enhance the capacity of the above model by incorporating dependence not only across orientations and scales, but also across space. We achieve this by defining an appropriate neighborhood for each coefficient, which is then modeled as a sub-Gaussian random vector. This joint sub-Gaussian modeling is followed by a Gaussianization procedure, which results in a steerable pyramid whose coefficients are jointly Gaussian.

The justification for this Gaussianization step, stems from the fact that the normalized transform domain is simply modeled yielding standard extracted features, such as covariances between

pairs of subbands. Another advantage is that the similarity measurement can be performed using an analytical expression for the Kullback-Leibler Divergence between two multivariate Gaussian distributions, avoiding computationally complex methods, such as the Monte-Carlo method. Besides, the normalized pyramid allows to perform easily steerability in the feature space.

7.3.1 Variance-adaptive local modeling using multivariate sub-Gaussian distributions

The local dependencies of the coefficients at a given subband and the associated marginal distributions, can be modeled using a homogeneous random field with a spatially changing variance. This requirement can be realized by modulating each coefficient (node of the field) with hidden scaling random variables.

A representative example of such a field is the *Gaussian Scale Mixture* (GSM) [3], which is the product of a Gaussian random vector and a hidden scalar random variable (multiplier) taking discrete or continuous values. By definition, a random vector $\vec{X}$ follows a GSM distribution if and only if it can be written as

$$\vec{X} \stackrel{d}{=} \sqrt{A}\,\vec{G} \tag{7.16}$$

where $\vec{G}$ is a zero-mean Gaussian random vector and A a positive scalar variable independent of $\vec{G}$ ($\stackrel{d}{=}$ denotes equality in distribution). GSM models have been applied in describing image content and in problems such as image denoising and detection [128, 94, 29, 111].
Two basic assumptions are made in order to reduce the dimensionality of these models:

(i) the probability structure is defined *locally*. In particular, the probability density of a coefficient when conditioned on a set of neighbors, is independent of the coefficients outside the neighborhood,

(ii) all such neighborhoods obey the same distribution (*spatial homogeneity*).

The construction of a global probabilistic model for images, based on these local descriptions, needs the specification of a *neighborhood structure* for each subband coefficient and the *distribution of the multipliers*. In this section, we extract the interdependencies between coefficients at different subbands and levels by utilizing their joint statistics, expressed by a GSM model in which the multiplier A is drawn from a $S\alpha S$ distribution. This is exactly the case of a multivariate α-sub-Gaussian model.

The construction of a suitable neighborhood is studied below. As an example of a neighborhood for a subband coefficient, except of the coefficient itself, we can define its neighbors to be the $m \times m$ adjacent coefficients at the same subband, the J-1 coefficients of the other subbands at the same spatial location and the 1 coefficient at the same subband of the next lower level, placed at the corresponding location. Then, the vector representing this neighborhood is supposed to be a sample of an α-sub-Gaussian process.

Let $\vec{X}_k^{l,i} = [c^l(x_k, \phi_i), x_1, x_2, ..., x_{P-1}]$ denote the P-dimensional neighborhood of the coefficient $c^l(x_k, \phi_i)$ at the spatial position k ($k = 1, \ldots, N$), orientation ϕ_i ($i = 1, \ldots, J$) and level

l ($l = 1, \ldots, L$). This neighborhood is supposed to be drawn of an α-SG($\mathbf{R}^l$) process, where $\mathbf{R}^l$ is the underlying covariance matrix. Notice that although the superscript l refers to the level including the coefficient which is Gaussianized, the estimation of $\mathbf{R}^l$ takes also into account coefficients from the next lower level $(l+1)$.

7.3.2 Gaussianization of the multivariate sub-Gaussian model

An important property of a GSM model is that the probability density of a P-dimensional vector $\vec{X}$, drawn from this model, is Gaussian when conditioned on A. Combining this property with Eq. (7.16), it is clear that the normalized vector $\vec{X}/\sqrt{A}$ follows a Gaussian distribution.

The probability density of $\vec{X}$ conditioned on A is given by:

$$p(\vec{X}|A) = \frac{\exp(-\vec{X}^T(A\mathbf{R})^{-1}\vec{X}/2)}{(2\pi)^{P/2}\ |A\mathbf{R}|^{1/2}} \ . \tag{7.17}$$

From Eq. (7.17), it can be seen that the Maximum Likelihood estimator for the multiplier A is

$$\hat{A}(\vec{X}) = \frac{\vec{X}^T\mathbf{R}^{-1}\vec{X}}{P} \tag{7.18}$$

where the estimator is viewed as a function of $\vec{X}$ to emphasize the assumption of locality. This simplifies the computational procedure for the Gaussianization of the steerable pyramid subband coefficients, as we assume that the multipliers associated with different neighborhoods are estimated independently, even though the neighborhoods are overlapping. In our implementation, we estimate the underlying covariance matrix $\mathbf{R}^{l,i}$, corresponding to the basic orientation ϕ_i at the lth-level, by employing the neighborhoods of all coefficients at the given orientation ($\vec{X}_k^{l,i}$, $k = 1, \ldots, N$). Under the assumption of a GSM model, to which the multivariate sub-Gaussian model belongs, the proposed procedure has the advantage of resulting in a computationally efficient way to estimate the hidden multiplier A and normalize the subband coefficients.

Notice that a Gaussian Mixture model (GMM), which is a commonly used method as we mentioned in the introduction (e.g. [32]), is also a particular case of a GSM since it corresponds to the case where the multiplier is discrete. However, using a GMM, it is not possible to perform such a normalization step and we usually have to resort to an HMM, where we need to use complicated Expectation-Maximization algorithms to estimate the multipliers, which are nested through scales in a Markovian manner.

Summarizing, the Gaussianization of a given image proceeds as follows:

1. Decompose the given image into L levels and J orientations per level, via a steerable pyramid.

2. For each decomposition level l, $l = 1, \ldots, L$:

   ▶ For each orientation ϕ_i, $i = 1, \ldots, J$ (except the low-pass residual), at the l-th decomposition level:

i) Estimate the underlying covariance matrix $\mathbf{R}^{l,i}$ using Eq. (4.32).

ii) For each coefficient $c^l(x_k, \phi_i),\ k = 1, \ldots, N$:

- Construct the corresponding neighborhood $\vec{X}_k^{l,i}$.
- Estimate the associated multiplier $\hat{A}_k^{l,i}(\vec{X}_k^{l,i})$ using Eq. (7.18).
- Compute the normalized coefficient $\tilde{c}^l(x_k, \phi_i) = c^l(x_k, \phi_i)/\sqrt{\hat{A}_k^{l,i}}$.

From Eq. (7.18), it is obvious that the estimation accuracy for the multiplier depends on the accurate estimation of the underlying covariance matrix and the neighborhood structure.

7.3.3 Neighborhood Construction

The Gaussianization procedure is obviously affected by the construction of a suitable neighborhood for each subband coefficient. There is a trade-off between the computational complexity and the neighborhood size. This can be seen from Eq. (7.18), where the estimated stable multiplier depends on the inverse of the underlying covariance matrix $\mathbf{R}$. The computational complexity increases as the neighborhood size increases, since the estimation time of $\mathbf{R}$ depends on the dimension $(P \times P)$ of the estimated covariation matrix, where P denotes the neighborhood size. It is clear that is not possible to construct all possible neighborhoods for each subband coefficient in order to select the optimal neighborhood.

In this section we examine the performance of the Gaussianization procedure with respect to different neighborhoods, taking also into consideration the computational limitations. For this purpose, we implement the Gaussianization process using the neighborhoods shown in Table 7.1. In this table we use the following notation:

1. (l, ϕ_i): X, means that X is the neighborhood at the level l and orientation ϕ_i, centered at the coefficient $c^l(x_k, \phi_i)$.

2. $\{(l, \phi_j) | j = 1, \ldots, J,\ j \neq i\}$: Y, means that for each one of the rest $(J-1)$ orientations at level l, Y is the neighborhood centered at the coefficient $c^l(x_k, \phi_j), \jmath = 1, \ldots, J,\ j \neq i$.

3. $(l+1, \phi_i)$: Z, means that Z is the neighborhood at the next lower level $l+1$ and orientation ϕ_i centered at the corresponding spatial location.

4. Finally, the addition operator $(1.+2.+3.)$ means that the overall neighborhood of $c^l(x_k, \phi_i)$ is the concatenation of the above components.

For a given subband coefficient $c^l(x_k, \phi_i)$, the formation of some of the above neighborhoods requires the inclusion of coefficients at the corresponding spatial location of the subbands at the next coarser scale. Applying the standard steerable pyramid decomposition, the dimension of the subbands as we move from one level to the next coarser is halved. In this case, the "corresponding spatial location" is in terms of a quad-tree structure, where a coefficient at the coarser level is associated with 4 coefficients at the same subband of the previous finer scale,

Index	Neighborhood Structure for a given $c^l(x_k, \phi_i)$	Size (P)
1	(l, ϕ_i): (3x3) + $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1 + $(l+1, \phi_i)$: 1	$J+9$
2	(l, ϕ_i): (3x3) + $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1	$J+8$
3	(l, ϕ_i): (3x3)	9
4	(l, ϕ_i): 4 (cross shape (c.s.))	5
5	(l, ϕ_i): 4 (c.s.) + $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1 + $(l+1, \phi_i)$: 1	$J+5$
6	(l, ϕ_i): 4 (c.s.) + $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1	$J+4$
7	(l, ϕ_i): 4 (c.s.) + $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 4 (c.s.) + + $(l+1, \phi_i)$: 4 (c.s.)	$5(J+1)$
8	(l, ϕ_i): (5x5) + $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1 + $(l+1, \phi_i)$: 1	$J+25$
9	(l, ϕ_i): (5x5) + $\{(l, \phi_j) \vert j = 1, \ldots, J,\ j \neq i\}$: 1	$J+24$
10	(l, ϕ_i): (5x5)	25

Table 7.1: Neighborhood shapes which were used in the measurement of the Gaussianization performance.

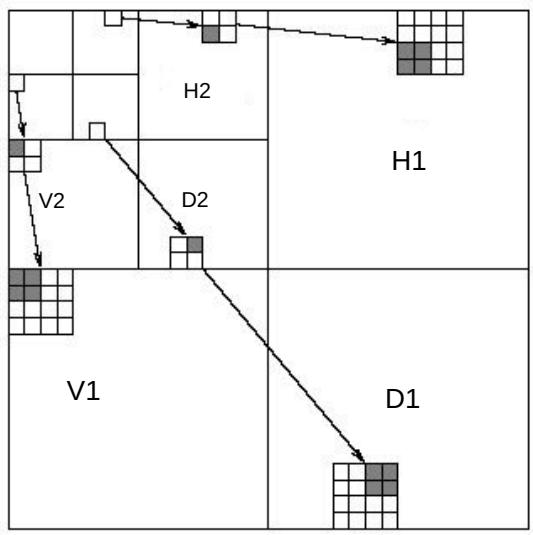


Figure 7.2: Quad-tree structure of the subband coefficients.

as shown in Figure 7.2[2]. A better way to associate coefficients between adjacent levels is to use the frequency-domain implementation of the steerable pyramid, without subsampling the output of the filters, as described in subsection 7.2.1, resulting in subbands with equal size at all decomposition levels. Then, the coefficient at the corresponding spatial location of the next coarser subband is simply $c^{(l+1)}(x_k, \phi_i)$.

We tested the performance of the Gaussianization procedure with respect to the neighborhood structure, by implementing that on a set of randomly selected textures, which were decomposed in 3 levels and 4 orientations per level. After the Gaussianization, at each subband we calculated the relative entropy between the histogram (with 256 bins) and the Gaussian PDF fitting the

[2]Obtained from [33].

normalized subband coefficients, as a fraction of the histogram entropy:

$$\frac{\Delta H}{H} = \frac{\sum_{k=1}^{256} h(x_k)\ \log\left(\frac{h(x_k)}{g(x_k)}\right)}{-\sum_{k=1}^{256} h(x_k)\ \log(h(x_k))} \tag{7.19}$$

where $h(x_k)$ is the probability density of the center x_k of the k-th bin, as estimated from the histogram, and $g(x_k)$ is the corresponding value of the fitting Gaussian PDF with parameters estimated from the normalized coefficients. The best choice for the neighborhood structure is that resulting in the smallest fraction.

Figure 7.3 displays the histogram of the neighborhood indices (cf. Table 7.1), for the subbands of the following 5 textures selected from the Brodatz database: **1.** 1.1.01, **2.** 1.1.04, **3.** 1.1.08, **4.** 1.2.08, **5.** 1.5.04, along with their rotations at 30, 60, 90 and 120 degrees. For a 3-level pyramid decomposition with 4 orientations per level, we obtain a set of $25 \times 3 \times 4 = 300$ subbands. The vertical axis is the relative frequency of each neighborhood shape (horizontal axis), whose selection resulted in the smallest fraction (7.19) for the above subbands. For the given set of textures, the

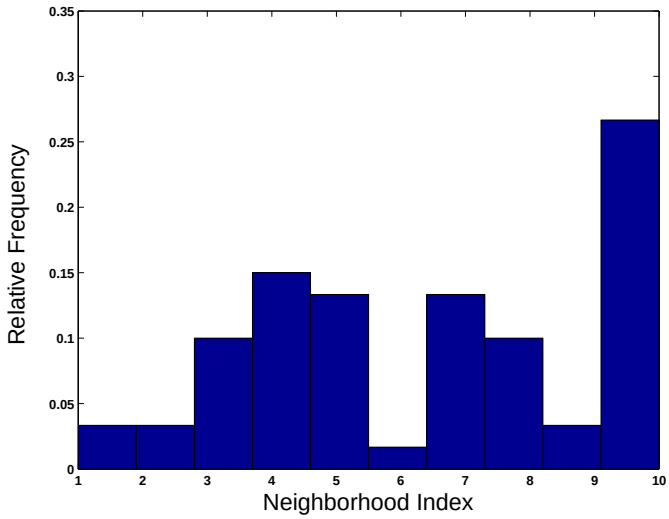


Figure 7.3: Histogram of the neighborhood indices in terms of resulting in the best Gaussianization performance for a set of 300 subbands.

choice of the 10th neighborhood shape results in the best Gaussianization performance for most of the pyramid subbands. In the subsequent experimental evaluation we will mainly rely on this neighborhood for performing the Gaussianization procedure.

7.3.4 Feature Extraction

Following the normalization procedure, the marginal and joint statistics of the coefficients at adjacent spatial positions, orientations and levels are close to the Gaussian distribution. In the Feature Extraction step, we compute the $J \times J$ covariance matrix at each decomposition level.

Thus, for a given image I decomposed in L levels, the corresponding signature $\mathcal{S}^{\mathcal{G}}$ is given by

the set of the L covariance matrices:

$$I \mapsto \mathcal{S}^{\mathcal{G}} = \{\mathbf{\Sigma}_I^1, \mathbf{\Sigma}_I^2, \ldots, \mathbf{\Sigma}_I^L\}$$

where $\mathbf{\Sigma}_I^l$ is the covariance matrix of the l-th decomposition level. Due to the symmetric property of the covariance matrix, the total size of the above signature equals

$$size(\mathcal{S}^{\mathcal{G}}) = \frac{J(J+1)}{2} \cdot L$$

This signature size is almost the half of the size corresponding to the signature $\mathcal{S}$ described in section 7.2.1, while maintaining the same amount of information between pairs of subbands at each decomposition level.

The signature $\mathcal{S}^{\mathcal{G}}$ contains only the across orientation, second-order dependence at a given decomposition level. We may enhance the capacity of $\mathcal{S}^{\mathcal{G}}$ by considering the across levels dependence as expressed by the covariance matrices between consecutive levels. In this case the signature of an image I is the following:

$$I \mapsto \mathcal{S}_{\mathcal{E}}^{\mathcal{G}} = \{\mathbf{\Sigma}_I^1, \mathbf{\Sigma}_I^2, \ldots, \mathbf{\Sigma}_I^L, \mathbf{\Sigma}_I^{1\to 2}, \mathbf{\Sigma}_I^{2\to 3}, \ldots, \mathbf{\Sigma}_I^{(L-1)\to L}\} \ , \tag{7.20}$$

where $\mathbf{\Sigma}_I^{l\to(l+1)}$ denotes the covariance matrix corresponding to the subbands at levels l and $(l+1)$. In particular, the element $[\mathbf{\Sigma}_I^{l\to(l+1)}]_{ij}$ is equal to the covariance of the i-th subband at level l with the j-th subband at the next lower level $(l+1)$. As before, the enhanced signature $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$ contains more texture-specific information than the signature $\mathcal{S}^{\mathcal{G}}$ at the cost of an increased computational complexity, since its size equals

$$size(\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}) = \frac{J(J+1)}{2} \cdot L + \frac{J(J+1)}{2} \cdot (L-1) = \frac{J(J+1)}{2} \cdot (2L-1)$$

In order to overcome the problem of different dimensions when we extract subband inter-dependencies between adjacent levels, we follow the procedure described in subsection 7.2.1, by upsampling the filters and working in the frequency domain.

7.3.5 Similarity Measurement

After the Gaussianization procedure, we model the distribution of each decomposition level, as well as the joint distribution between consecutive levels in the case of using $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$, using a Multivariate Gaussian density (MvGD). The similarity between two images is measured by employing the KLD between MvGDs. Consider the case in which the texture information of each image is expressed using the signature $\mathcal{S}^{\mathcal{G}}$, that is, each decomposition level is associated with a covariance matrix.

In particular, let I^l, Q^l be the l-th decomposition level of two images I, Q, following MvGDs with mean vectors $\vec{\mu}_I^l$, $\vec{\mu}_Q^l$ and covariance matrices $\mathbf{\Sigma}_I^l$, $\mathbf{\Sigma}_Q^l$, respectively. The KLD between these

two levels is given by:

$$D(I^l\|Q^l) = \frac{1}{2}\Big((\vec{\mu}_Q^l - \vec{\mu}_I^l)^T(\mathbf{\Sigma}_Q^l)^{-1}(\vec{\mu}_Q^l - \vec{\mu}_I^l) + tr(\mathbf{\Sigma}_I^l(\mathbf{\Sigma}_Q^l)^{-1} - I) - \\ - \ln|\mathbf{\Sigma}_I^l(\mathbf{\Sigma}_Q^l)^{-1}|\Big) \tag{7.21}$$

Making an assumption of independence between scales, the overall KLD between images I, Q is given by the sum:

$$D(I\|Q) = \sum_{l=1}^{L} D(I^l\|Q^l) \tag{7.22}$$

In this problem, we deal with image databases which may contain rotated versions of a given image. Replacing coefficients $c^l(x_k, \phi_i)$, in Eq. (7.1), with their normalized versions $\tilde{c}^l(x_k, \phi_i)$ and combining it with Eq. (7.2), we obtain the following relation of equivalence:

$$\mathbf{\Sigma}_{Q_\theta}^l = \mathbf{F}(\theta)\mathbf{\Sigma}_Q^l\mathbf{F}^T(\theta) \tag{7.23}$$

where $\mathbf{\Sigma}_Q^l$, $\mathbf{\Sigma}_{Q_\theta}^l$ are the l-level covariance matrices of image Q and its rotated version at an angle θ, Q_θ, respectively. Besides, substituting Eq. (7.23) into Eq. (7.21), we obtain the KLD between I^l and rotated Q_θ^l as a function of the KLD between I^l and original Q^l (see Appendix B.1.2).

Consider I to be the query image and $\tilde{Q} = Q_\phi$ to be a counter-clockwise rotation, by an angle ϕ, of the original image Q in the database. In a real application, of course, the value of ϕ is unknown. Thus, the distance between the lth-levels of I and $\tilde{Q}$ (I^l and $\tilde{Q}^l$, respectively) is defined as the minimum KLD between I^l and $\tilde{Q}_{-\theta}^l$, where the minimization is over a set of rotations Θ. The notation $\tilde{Q}_{-\theta}$ means a clockwise rotation of image $\tilde{Q}$ in order to find the best alignment with image I.

By noticing that

$$\begin{aligned} \vec{\mu}_{\tilde{Q}_{-\theta}}^l &= \mathbf{F}(-\theta)\vec{\mu}_{\tilde{Q}}^l \\ \mathbf{\Sigma}_{\tilde{Q}_{-\theta}}^l &= \mathbf{F}(-\theta)\mathbf{\Sigma}_{\tilde{Q}}^l\mathbf{F}^T(-\theta)\ , \end{aligned}$$

and substituting Eq. (7.23) into Eq. (7.21), results in the following proposition:

Proposition 7.3.1 *The KLD between the l-th levels of an image I and a clockwise rotation of image $\tilde{Q}$ by an angle θ (I^l and $\tilde{Q}_{-\theta}^l$, respectively), is given by*

$$D(I^l\|\tilde{Q}_{-\theta}^l) = \Bigg[\frac{1}{2}\Big(-\big(\mathbf{V}_l(\theta) + \mathbf{V}_l^T(\theta)\big) + \vec{v}_l^T(\theta)(\mathbf{\Sigma}_{\tilde{Q}}^l)^{-1}\vec{v}_l(\theta) + \\ + tr\big(\mathbf{\Sigma}_I^l\mathbf{F}^T(\theta)(\mathbf{\Sigma}_{\tilde{Q}}^l)^{-1}\mathbf{F}(\theta) - I\big)\Big)\Bigg] - \\ - \frac{1}{2}\big(\ln(|\mathbf{\Sigma}_I^l||(\mathbf{\Sigma}_{\tilde{Q}}^l)^{-1}|) - (\vec{\mu}_{\tilde{Q}}^l)^T(\mathbf{\Sigma}_{\tilde{Q}}^l)^{-1}\vec{\mu}_{\tilde{Q}}^l\big) \tag{7.24}$$

where

$$\begin{aligned}\mathbf{V}_l(\theta) &= (\vec{\mu}^l_{\tilde{Q}})^T(\mathbf{\Sigma}^l_{\tilde{Q}})^{-1}\mathbf{F}(\theta)\vec{\mu}^l_I \\ \vec{v}_l(\theta) &= \mathbf{F}(\theta)\vec{\mu}^l_I\end{aligned}$$

Proof. The detail of the proof is described in Appendix B.1.3. ■

Finally, the overall KLD between I and $\tilde{Q}$ is defined as:

$$D(I\|\tilde{Q}) = \min_{\theta\in\Theta}\sum_{l=1}^{L} D(I^l\|\tilde{Q}^l_{-\theta}) \tag{7.25}$$

Substituting Eq. (7.24) into Eq. (7.25) and taking into account that the subbands of a steerable pyramid have zero means, we have the following proposition:

Proposition 7.3.2 *Suppose that* $\mathcal{S}^{\mathcal{G}}(I) = \{\mathbf{\Sigma}^1_I, \mathbf{\Sigma}^2_I, \ldots, \mathbf{\Sigma}^L_I\}$ *and* $\mathcal{S}^{\mathcal{G}}(\tilde{Q}) = \{\mathbf{\Sigma}^1_{\tilde{Q}}, \mathbf{\Sigma}^2_{\tilde{Q}}, \ldots, \mathbf{\Sigma}^L_{\tilde{Q}}\}$ *are the signatures corresponding to the normalized coefficients of the steerable pyramids for two given homogeneous textures* I *and* $\tilde{Q}$*, respectively. The rotation-invariant KLD between the two textures takes the following form:*

$$\begin{aligned}D(I\|\tilde{Q}) &= \min_{\theta\in\Theta}\Big[\frac{1}{2}\sum_{l=1}^{L} tr\big(\mathbf{\Sigma}^l_I\mathbf{F}^T(\theta)(\mathbf{\Sigma}^l_{\tilde{Q}})^{-1}\mathbf{F}(\theta)\big)\Big] - \frac{JL}{2} - \\ &\quad -\frac{1}{2}\sum_{l=1}^{L}\ln(|\mathbf{\Sigma}^l_I||(\mathbf{\Sigma}^l_{\tilde{Q}})^{-1}|)\end{aligned} \tag{7.26}$$

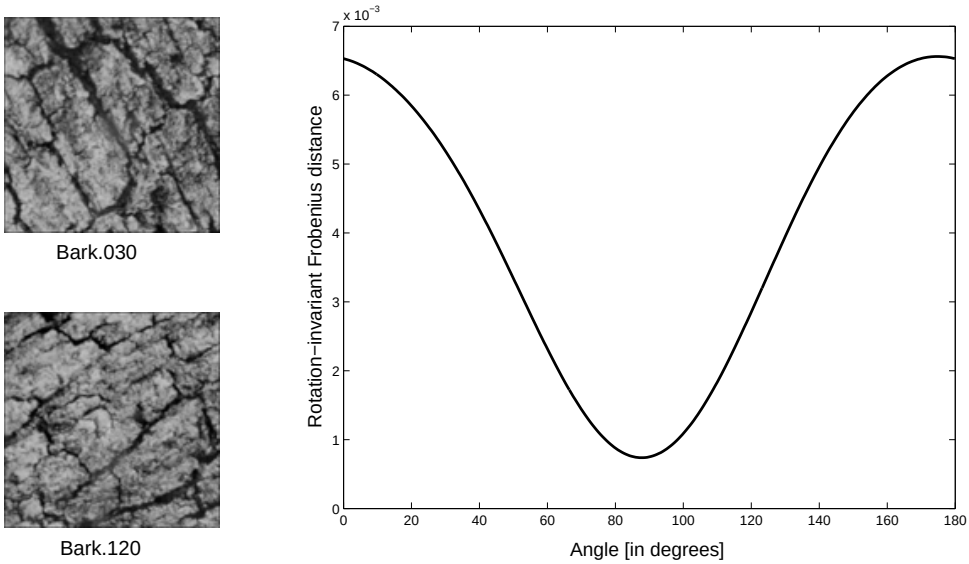
where J *is the number of orientations and* L *the number of decomposition levels.*

Remember that in the above derivations we have assumed equispaced basic orientations, which makes the matrix $\mathbf{F}(\theta)$ to be orthogonal for any angle θ ($\mathbf{F}(\theta)\mathbf{F}^T(\theta) = I_{J\times J} \Rightarrow \mathbf{F}^T(\theta) = \mathbf{F}^{-1}(\theta) = \mathbf{F}(-\theta)$).

Note that in the last proposition each image is mapped to the signature $\mathcal{S}^{\mathcal{G}}$. If we employ the enhanced signature $\mathcal{S}^{\mathcal{G}}_{\mathcal{E}}$ instead of $\mathcal{S}^{\mathcal{G}}$, the derivation of the KLD between two distinct textures I and $\tilde{Q}$ is straightforward by implementing the above proposition on the signatures $\mathcal{S}^{\mathcal{G}}_{\mathcal{E}}(I)$ and $\mathcal{S}^{\mathcal{G}}_{\mathcal{E}}(\tilde{Q})$, with the only difference being with respect to the number of summands. In particular, we have to replace L with the size of $\mathcal{S}^{\mathcal{G}}_{\mathcal{E}}$ which is equal to $(2L-1)$.

Notice that when I and $\tilde{Q}$ are two rotated versions of the same image, the angle θ^* for which the minimum is achieved in Eq. (7.26) (as well as in (7.12)) should be close to the relative angle between I and $\tilde{Q}$, that is, the angle one needs to clockwise rotate I in order to get $\tilde{Q}$. Thus, a way to evaluate the performance of the above rotation-invariant KLD, is to verify whether the estimated angle θ^* is actually close to the real relative angle between two physically rotated versions of the same image. Besides, it may also be useful in many practical applications to find out approximately this relative angle. Figure 7.4 illustrates this by showing the function

$D(I,Q)(\theta)$ given by Eq. (7.12), for the case where I and Q are rotated versions of *Bark* texture sample from the Brodatz database, while Figure 7.5 shows the function $D(I\|Q)(\theta)$ given by Eq. (7.26).



(a) (b)

Figure 7.4: (a) *Bark* physically rotated at 30 and 120 degrees, (b) $D(I,Q)(\theta)$ for $J = 4$. Notice that the minimum is achieved for $\theta^* = 90$ degrees, which is the exact relative angle between the two texture samples.

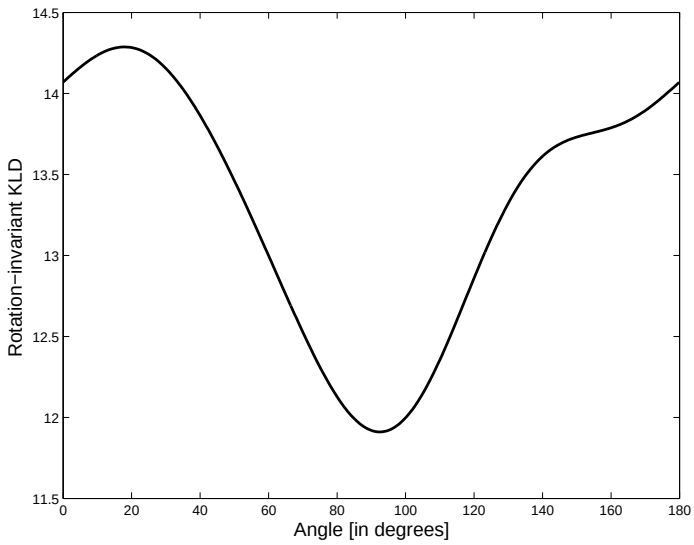


Figure 7.5: $D(I\|Q)(\theta)$ for $J = 4$. Notice that the minimum is achieved for $\theta^* = 90$ degrees, which is the exact relative angle between the two texture samples.

7.4 Experimental Results

In order to evaluate the efficiency, the retrieval schemes proposed in sections 7.2, 7.3 were applied on a set of 13, 512×512 texture images (cf. Figure 7.6). In the following study we may also use the index of a class instead of its name, as shown in Table 7.2.

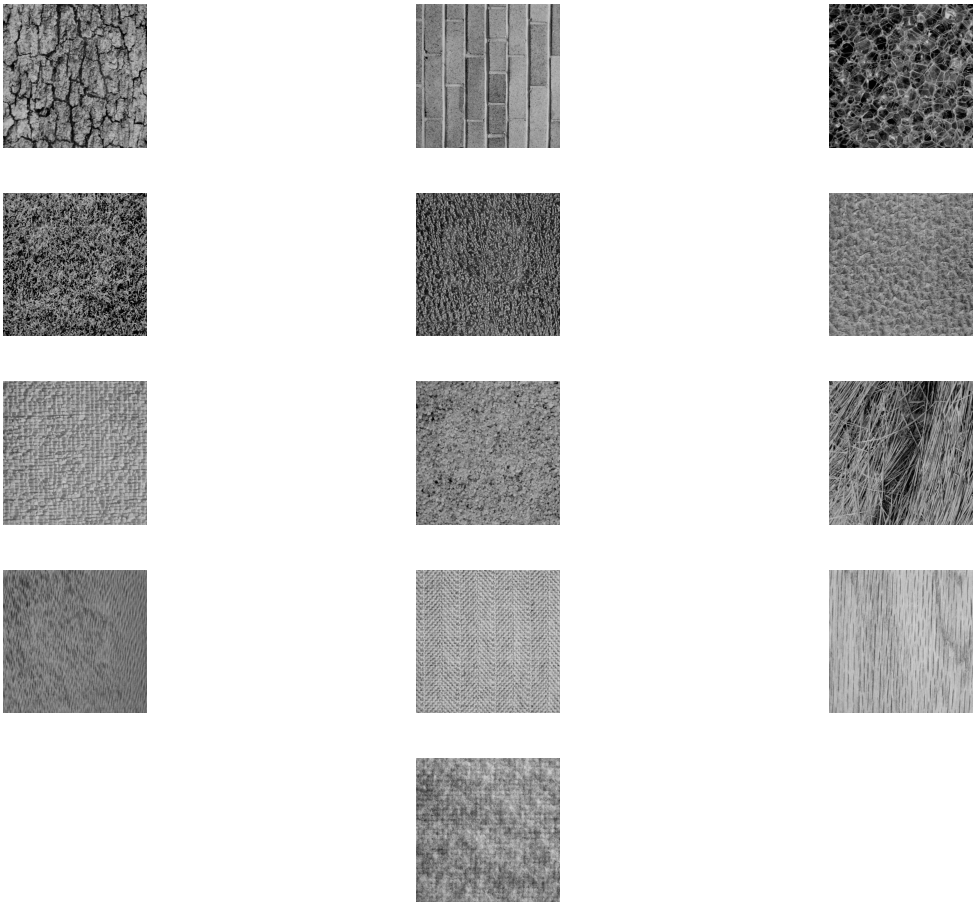


Figure 7.6: Texture images from the VisTex database, from left to right and top to bottom: Bark, Brick, Bubbles, Grass, Leather, Pigskin, Raffia, Sand, Straw, Water, Weave, Wood, Wool.

Index	Texture Class	Index	Texture Class
1	Bark	8	Sand
2	Brick	9	Straw
3	Bubbles	10	Water
4	Grass	11	Weave
5	Leather	12	Wood
6	Pigskin	13	Wool
7	Raffia		

Table 7.2: Index of each Texture Class.

Each of them was physically rotated at 30, 60, 90 and 120 degrees, resulting in a set of 65 texture samples. Then, we divided each image into 4, 256×256 non-overlapping subimages constructing a database with a total of $4 \times 65 = 260$ textures. We implemented a 3-level steerable pyramid decomposition, by employing the following oriented basis (steering) functions:

$$f_1(\theta) = \frac{1}{2}[\cos(\theta) + \cos(3\theta)] \qquad f_2(\theta) = f_1\left(\frac{\pi}{4} - \theta\right)$$

$$f_3(\theta) = f_1\left(\frac{\pi}{2} - \theta\right) \qquad f_4(\theta) = f_1\left(\frac{3\pi}{4} - \theta\right)$$

with basic angles $\phi_1 = 0$, $\phi_2 = \pi/4$, $\phi_3 = \pi/2$, $\phi_4 = 3\pi/4$, resulting in 4 oriented subbands at each level. The histogram of the estimated characteristic exponent values for the 260 textures is shown in Figure 7.7.

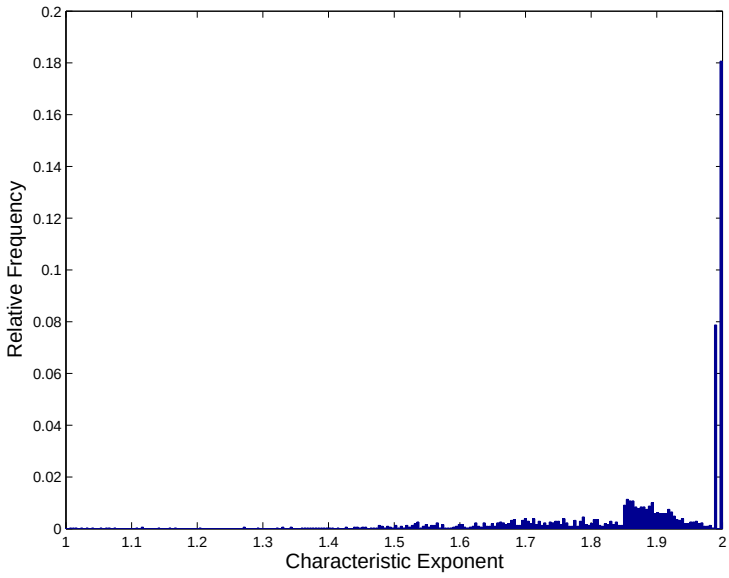


Figure 7.7: Histogram of the estimated values for the characteristic exponent, α, for the set of 260 texture images of size 256×256.

We conducted two series of experiments. In the first one, we represented the texture-specific content of each image by estimating the covariation matrices at each level or between consecutive levels and then we measured the similarity between two images by employing the Frobenius norm, as described in section 7.2. In the second series, we applied the Gaussianization procedure on the subband coefficients, described in section 7.3 and then we used the rotation-invariant version of the KLD, given by Eq. (7.26), as a similarity function. In both cases, the query is anyone of the subimages corresponding to the original, non-rotated images. The relevant images for each query are defined as the other 16 subimages corresponding to the rotated versions of the same original image.

7.4.1 Series I: Frobenius norm between steerable multivariate sub-Gaussian models

In this experiment we evaluate the performance of the steerable multivariate sub-Gaussian model combined with the rotation-invariant version of the Frobenius distance. The signature $\mathcal{S}$, containing the estimated covariation matrices between pairs of subbands at the same decomposition level, is a subset of the enhanced signature $\mathcal{S}_{\mathcal{E}}$ given by Eq. (7.11). Thus, we extract the signature $\mathcal{S}_{\mathcal{E}}$ from each texture sample and then, the similarity measurement in the case we were supposed to use signature $\mathcal{S}$ is given from Eq. (7.12) by simply applying the first L matrices of $\mathcal{S}_{\mathcal{E}}$.

The components of $\mathcal{S}_{\mathcal{E}}$ that contain the inter-scale dependencies, that is, the matrices $\mathbf{C}_I^{l\to(l+1)}$, are computed using the approach described in section 7.2.1, by employing a steerable pyramid without subsampling when moving from one level to the next coarser. The advantages of this

representation is that the subbands at all levels are of equal dimensions, which simplifies the numerical implementation and the most important, that the estimation of the model parameters (α, γ) is more accurate due to the increased number of pyramid coefficients, as opposed to the standard steerable pyramid which subsamples between consecutive levels. Notice that starting from an image of dimension $M \times M$ and applying an L-level standard pyramid decomposition, the dimension of the subbands at the L-th (coarsest) level is equal to $(\frac{M}{2^L} \times \frac{M}{2^L})$. On the other hand, by applying a pyramid without subsampling across levels, the dimension of the L-th level subbands equals $(\frac{M}{2} \times \frac{M}{2})$, which is also the dimension of the subbands at the first decomposition level.

For the applied 3-level pyramid decomposition, the signatures $\mathcal{S}$ and $\mathcal{S}_{\mathcal{E}}$ contain 3 and 5 matrices, respectively. Table 7.3 shows the average retrieval rates for the proposed multivariate sub-Gaussian steerable model using the two kinds of signatures, $\mathcal{S}$ and $\mathcal{S}_{\mathcal{E}}$, as well as the average retrieval rate of the method whose signature, of size equal to 3, contains the sample covariance matrices between pairs of subbands at the same level, combined with the corresponding similarity function given by Eq. (7.15).

Methods		
Gaussian signature & Frobenius (Eq. (7.15))	$\mathcal{S}$ & Frobenius (Eq. (7.12))	$\mathcal{S}_{\mathcal{E}}$ & Frobenius (Eq. (7.12))
86.75	78.6	78.8

Table 7.3: Average retrieval rate (%) in the top 16 matches, for the set of 260 textures.

We observe that for the selected texture images, the Gaussian model results in an increased retrieval performance in comparison with the sub-Gaussian model. This should be expected, since a large portion of the estimated characteristic exponent values is close to the Gaussian ($\alpha = 2$), as shown in Figure 7.7. Regarding the proposed method, we can see that the enhanced signature $\mathcal{S}_{\mathcal{E}}$ results in a slightly better retrieval performance than the signature $\mathcal{S}$. Of course, someone would expect a higher increase of the average retrieval rate in the case of $\mathcal{S}_{\mathcal{E}}$. The reason for this is the not so strong correlation between subbands at adjacent levels, for the selected set of textures. This can be verified by examining the off-diagonal elements of the $\mathbf{C}_I^{l \to (l+1)}$ matrices, which have relatively small amplitudes compared with the corresponding elements of the $\mathbf{C}_I^l$ matrices. Due to the increased computational complexity, the enhanced signature should be employed when a stronger inter-scale correlation is apparent.

Figure 7.8 shows the average percentages of correct retrieval rate for each one of the 13 selected texture classes, for the Gaussian signature and the sub-Gaussian signatures $\mathcal{S}$ and $\mathcal{S}_{\mathcal{E}}$. We observe that the Gaussian model is far better only for 3 classes (2, 9, 12), which are exactly those with the greatest portion of characteristic exponent values near or equal to 2. We omit the 3 "undesired" classes and repeat the above experiment for the remaining set of 10 original, non-rotated textures. The previous set of 260 texture subimages (along with their rotations), is reduced to a set of 200 texture samples. Table 7.4 shows the corresponding average retrieval

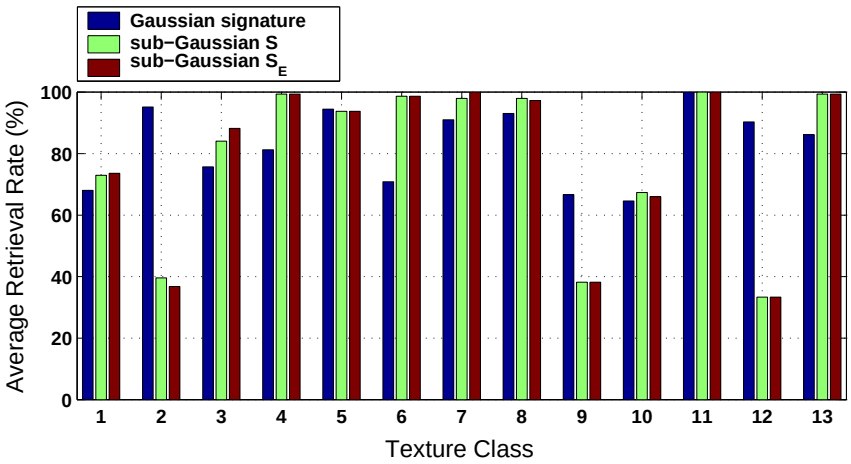


Figure 7.8: Average percentages (%) of correct retrieval rate for each individual texture class.

rates, while Figure 7.9 displays the average percentages of retrieving relevant subimages as a function of the number of top matches.

Methods		
Gaussian signature & Frobenius (Eq. (7.15))	$\mathcal{S}$ & Frobenius (Eq. (7.12))	$\mathcal{S}_{\mathcal{E}}$ & Frobenius (Eq. (7.12))
83.75	93.05	93.55

Table 7.4: Average retrieval rate (%) in the top 16 matches, for the subset of 200 textures.

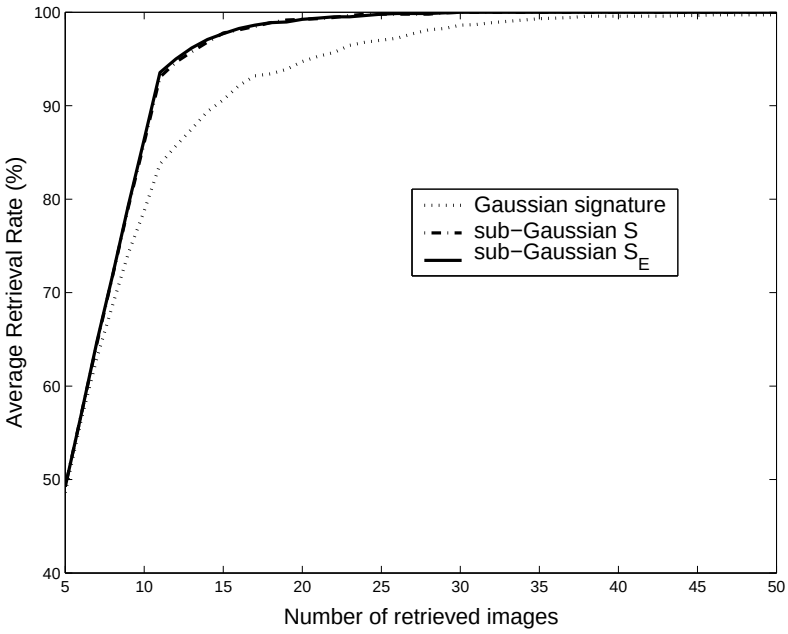


Figure 7.9: Retrieval performance according to the number of top matches considered.

As it was expected, there is a significant improvement of the average retrieval performance when the pyramid coefficients follow a heavy-tailed non-Gaussian distribution. From Figure 7.9 it is clear that the proposed multivariate sub-Gaussian steerable model outperforms the Gaussian

one, in terms of the convergence rate of the retrieval method. The reason for which the curves corresponding to the $\mathcal{S}$ and $\mathcal{S}_{\mathcal{E}}$ signatures overlap, is the relatively weak correlation between the subband coefficients at consecutive decomposition levels, thus the inclusion of information about the inter-scale dependencies gives us a small improvement compared with the increased computational complexity of using $\mathcal{S}_{\mathcal{E}}$.

7.4.2 Series II: Rotation-invariant KLD between Gaussianized steerable pyramids

In this series of experiments we evaluate the performance of the retrieval scheme which employs the signatures $\mathcal{S}^{\mathcal{G}}$ and $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$ as the set of extracted features, containing intra- and inter-scale dependencies (underlying covariance matrices between pairs of subbands at the same and at adjacent decomposition levels), and the rotation invariant KLD as the similarity measure. We compare the performance of this retrieval scheme with that obtained by minimizing the Frobenius norm of the differences between the corresponding covariance matrices ([6]) after the Gaussianization procedure, as well as with the performance obtained by minimizing the corresponding Frobenius norm between sample covariance matrices (Gaussian assumption) without applying the Gaussianization step. These 3 retrieval methods are implemented on the set containing the 13 texture samples.

In the first experiment we perform the Gaussianization procedure using the first neighborhood shape as shown in Table 7.1. We chose this neighborhood because of its medium size, which is preferable for a fast implementation, and also because it has a low relative frequency in the histogram shown in Figure 7.3, which means that it is not as good as other neighborhood shapes (such as 10) in terms of a small fraction $\Delta H/H$. However, we would like to test the performance of our retrieval system even in this case. Besides, the neighborhood is constructed using the quad-tree structure of the subband coefficients, described in section 7.3.3. The signature $\mathcal{S}^{\mathcal{G}}$ is constructed at the feature step and the rotation-invariant KLD is employed as a measure of similarity.

Table 7.5 shows the comparison in performance in average percentages of retrieving relevant images in the top 16 matches, by employing the above 3 retrieval schemes, while Figure 7.10 shows the average percentages of retrieving relevant subimages as a function of the number of top matches.

Methods		
Non-Gaussianized & Frobenius	Gaussianized & Frobenius	Gaussianized & KLD
88.62	89.26	94.01

Table 7.5: Average retrieval rate (%) in the top 16 matches.

Comparing the average retrieval rates in the first two entries of the table, we conclude that

once again the lower-order statistics are better approximations of the joint statistics between coefficients at adjacent orientations and scales, than the second order moments. Of course, both methods employ the covariance matrices between pairs of subbands, but in the first scheme ((Non-Gaussianized & Frobenius)) we estimate the sample covariances using the raw subband coefficients, while in the second scheme ((Gaussianized & Frobenius)) the covariances are estimated after the implementation of the Gaussianization procedure, which exploits lower-order moments, since the estimation is based on covariations.

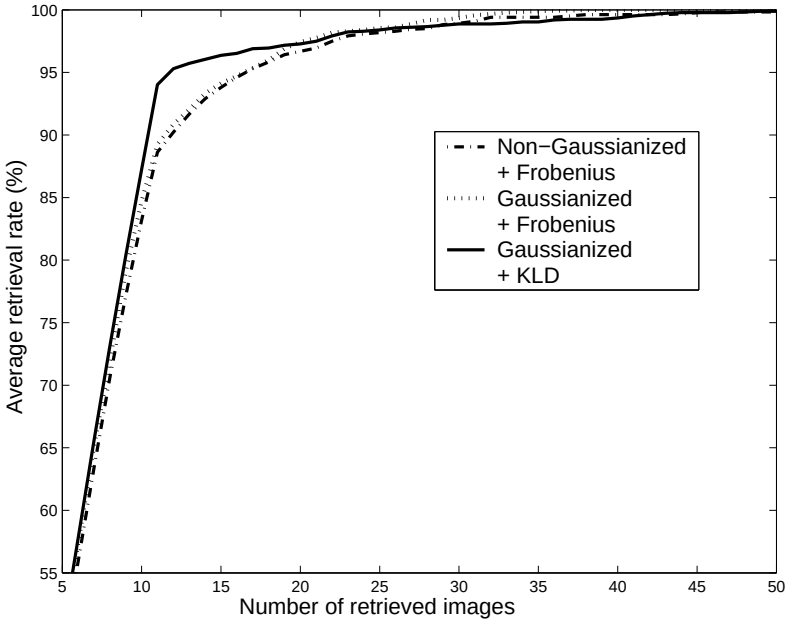


Figure 7.10: Retrieval performance according to the number of top matches considered.

On the other hand, the comparison between the second and third rate verifies our certainty that a statistical similarity function (KLD) is preferable than a deterministic one (Frobenius norm), for the same set of extracted features. Besides, Figure 7.10 shows that the proposed scheme converges faster than the other two retrieval schemes, while asymptotically the three methods achieve 100% average retrieval performance. The two curves corresponding to the methods that employ the Frobenius norm as the similarity measure, are almost identical. This is due to the fact that a large portion of the estimated characteristic exponent values are close to the Gaussian ($\alpha = 2$) (cf. Figure 7.7), so the Gaussianization of subbands with $\alpha \approx 2$ actually preserves their original normal distribution.

In the second experiment we perform the Gaussianization procedure using the last neighborhood shape (10), since it results in the best Gaussianization performance for most of the 300 constructed subbands. We employ the enhanced signature $\mathcal{S}_{\mathcal{E}}^{\mathcal{G}}$ and a 3-level steerable pyramid without subsampling across levels. Table 7.6 shows the comparison in performance in average percentages of retrieving relevant images in the top 16 matches, by employing the above 3 retrieval schemes, while Figure 7.11 shows the average percentages of retrieving relevant subimages as a function of the number of top matches.

From Table 7.6, we observe that the retrieval performance is improved. This is consistent

Methods		
Non-Gaussianized & Frobenius	Gaussianized & Frobenius	Gaussianized & KLD
88.62	95.40	95.62

Table 7.6: Average retrieval rate (%) in the top 16 matches.

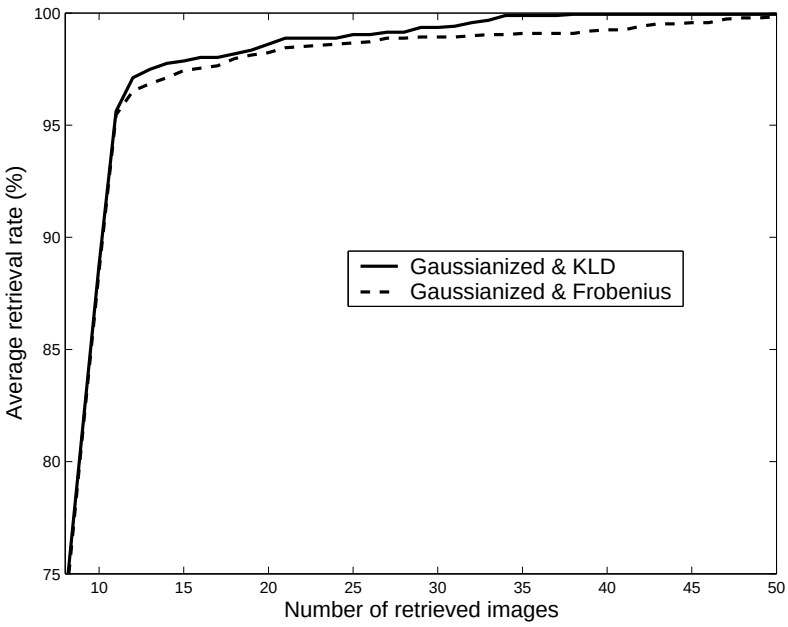


Figure 7.11: Retrieval performance according to the number of top matches considered.

with the theory, since the use of a steerable pyramid decomposition without subsampling results in a more accurate estimation of the $S\alpha S$ parameters (α, γ), which is crucial for an accurate estimation of the covariations between pairs of subbands. Besides, Eq. (7.19) gives a valid criterion for selecting the optimal neighborhood shape, since the selection of the 10-th shape according to this rule resulted in an increased average retrieval performance. From Figure 7.11 we can also observe that the convergence rate using the rotation-invariant Frobenius norm (Eq. (7.15)) is very close to the rate corresponding to the rotation-invariant KLD. This is due to the improved Gaussianization performance, compared with that of the first experiment, which results in an increased performance of the Frobenius norm that is best suited for Gaussian distributions. Thus, the more efficient the Gaussianization procedure, the more closely to a joint Gaussian the distribution of the subband coefficients is, which yields an improved performance for the Frobenius norm. However, the statistical similarity function (KLD) remains superior than the deterministic one (Frobenius norm).

7.5 Conclusions

In this chapter, we studied the design of a rotation-invariant CBIR system based on a multivariate sub-Gaussian ($S\alpha S$) model. The rotation invariance was achieved by deriving a steerable model based on fractional lower-order statistics (covariations) and constructing a rotation-invariant

version of the Frobenius norm as a similarity measure. The experimental results showed an increased average retrieval performance in comparison with the performance of previous methods based on second-order statistics (covariances).

In the second part of the chapter, we exploited the variance adaptation by applying a local Gaussianization procedure on the subband coefficients, resulting in a set of almost jointly Gaussian coefficients. Then we extracted the intra- and inter-scale dependencies by estimating the covariance for pairs of subbands at the same and at adjacent decomposition levels. This process is different from estimating the covariances using directly the original coefficients, since the Gaussinization procedure takes into account the actual heavy-tailed behavior of the coefficients. Therefore, we simplify the steerable model which is reduced to a second-order model with well known properties, exploiting at the same time the non-Gaussian statistics of the pyramid subbands. We verified that the efficiency of the Gaussianization process depends on the neighborhood shape and we derived a criterion for selecting the optimal neighborhood. The similarity measurement step was carried out by constructing a rotation-invariant version of the KLD between multivariate Gaussian densities. The experimental results showed an increased average retrieval performance in comparison with the sub-Gaussian steerable model. The reason for this, is that the retrieval efficiency does not only depend on the choice of the extracted features, but also on the choice of the appropriate similarity function. In general, we concluded that a statistical similarity function, such as KLD, is preferable than a deterministic one, such as the Frobenius norm.

Chapter 8

Conclusions and Future work

In the present work we described the design of an image retrieval system based on a statistical description of texture information. We built a hierarchical model, in the sense that we started by incorporating only marginal information modeled via univariate $S\alpha S$ distributions. Then, we enhanced the capacity of this method by exploiting the inter-dependencies between subbands at the same decomposition level, by applying a joint sub-Gaussian model (fractional lower-order statistics). We illustrated the increased retrieval efficiency of the multivariate approach, compared with the performance of the univariate method, as well as with the performance of methods that use second order statistics. We further improved the performance of the joint sub-Gaussian model by incorporating, not only the dependencies across orientations, but also the dependencies across scales.

Finally, we derived a rotation-invariant retrieval scheme by forming a steerable model based on covariations between the subbands of a steerable pyramid which are modeled as sub-Gaussian random variables, and then by constructing a rotation-invariant version of the Kullback-Leibler Divergence (KLD) between multivariate Gaussian probability density functions. In order to use this similarity measure, we applied a Gaussianization procedure on the above steerable model, obtaining subband coefficients which are almost jointly Gaussian. The experimental results showed an increased average retrieval performance compared with previous methods extracting second order moments (covariances) between the original subband coefficients.

An important observation from all the above retrieval schemes is that, for a given set of extracted features, the average retrieval performance is highly dependent on the choice of the similarity measure. In particular, we illustrated the superior performance of statistical similarity functions, such as KLD, in comparison with deterministic ones, such as matrix norms.

Now, we describe future research directions which could further result in an improved retrieval system with decreased probability of retrieval error. First of all, the main assumption throughout the present work was the stationary behavior of texture content. That is, we assumed that the distribution of the subband coefficients, which is closely related with the texture-specific information, is invariable within each subband. Instead, we could consider a *non-stationary* approach by permitting a *locally adapted* distribution, that is, by spatially adapting the characteristic ex-

ponent and the dispersion parameters. This can also be used in segmentation applications, since the different "objects" contained in a picture can be viewed as local intensity variations. The main drawback that we have to overcome in a spatially adapted $S\alpha S$ model, is the small sample size of the local neighborhoods, which results in large confidence intervals for the ML estimators of the model parameters.

In Chapters 6 and 7 the proposed retrieval methods were based on the estimation of covariations. The specification of a free p parameter was crucial in order to achieve small standard deviation of the covariation estimators. The optimal p value depends on triads of the form $(\alpha, \gamma_X, \gamma_Y)$. For a given value of the characteristic exponent α, we defined the optimal p as the mean of the optimal values corresponding to different pairs of (γ_X, γ_Y). Although we showed that, for a given α and different pairs of dispersion, these optimal values are in a relatively tight interval, however we expect that the performance of the covariation estimator will be increased by estimating the optimal p values for a quite "dense" 3-D grid $(\alpha, \gamma_X, \gamma_Y)$ and then use 3-D interpolation to obtain the optimal p corresponding to a specific triad $(\alpha_f, \gamma_1, \gamma_2)$.

Regarding the task of similarity measurement between two distinct images, we assumed independence across scales resulting in an overall similarity function that is written as the sum of partial distances between corresponding pairs of subbands. We could further improve the power of the similarity measure by dropping the assumption of independence and considering some kind of chain rule for the KLD between two images [63].

In Section 7.3 we evaluated the performance of a Gaussianization procedure applied on the subband coefficients of a steerable pyramid. This procedure is applied locally and its performance depends on the choice of the neighborhood for each coefficient. We tested the Gaussianization efficiency for a set of (10) neighborhood "shapes". Then, we applied our retrieval method using the optimal neighborhood, where the optimality is in terms of the smallest deviation (KLD) of the subband histograms from the Gaussian density with parameters estimated from the subband coefficients. We preserved the same neighborhood shape across subbands, for every image in the database. Obviously, it is impossible to check the Gaussianization performance for all possible neighborhood formations. However, we expect an increased Gaussianization performance and consequently an increased retrieval performance, by adapting (in some efficient way) the optimal neighborhood shape across subbands and images. Of course, this would result in an increased computational complexity, since we should test the performance for different neighborhoods before deciding which is the optimal.

Closing this chapter, we should refer to another factor which must be taken into account in order to improve the capabilities of a retrieval system, namely the presence of noise. One way to solve this problem, is to apply an image denoising scheme before the implementation of the retrieval process. A second way to overcome this problem or to degrade its influence, is to search for noise-invariant models in proportion to the rotation-invariant methods, in the case of lower-order statistics. If this is possible, we could adapt the retrieval schemes presented in the this work, making them robust to noisy rotations.

Part III

Appendices

Appendix A

Kullback-Leibler Divergence between Normalized $S\alpha S$ Characteristic Functions

In the section we derive the Kullback-Leibler Divergence (KLD) between the normalized characteristic functions of two univariate $S\alpha S$ densities following parameterization (4.1). It is known that if

$$c = \int_S g(x)\,\mathrm{d}x, \quad g(x) \geq 0 \; on \; S$$

then,

$$f(x) = \frac{g(x)}{c}$$

is a valid probability density function (PDF) on S. Thus, we have to compute the constant c for the characteristic function given by parameterization (4.1):

$$\begin{aligned}
c = \int_{\mathbb{R}} \phi(t)\,\mathrm{d}t \quad &= \quad \int_{\mathbb{R}} e^{-\gamma|t|^\alpha}\,\mathrm{d}t = \int_{-\infty}^{\infty} e^{-\gamma|t|^\alpha}\,\mathrm{d}t \\
&= \quad 2\int_0^\infty e^{-\gamma t^\alpha}\,\mathrm{d}t \overset{u=t^\alpha}{=} 2\int_0^\infty e^{-\gamma u}\,\frac{1}{\alpha}\,(u^{1/\alpha})^{1-\alpha}\,\mathrm{d}u \\
&= \quad \frac{2}{\alpha}\int_0^\infty e^{-\gamma u}\,\frac{\gamma^{\frac{1}{\alpha}-1}}{\gamma^{\frac{1}{\alpha}-1}}\,u^{\frac{1}{\alpha}-1}\,\mathrm{d}u \\
&\overset{v=\gamma u}{=} \quad \frac{2}{\alpha\gamma^{\frac{1}{\alpha}-1}}\int_0^\infty e^{-v}\,v^{\frac{1}{\alpha}-1}\,\frac{1}{\gamma}\,\mathrm{d}v \\
&= \quad \frac{2}{\alpha\,\gamma^{1/\alpha}}\,\Gamma\Big(\frac{1}{\alpha}\Big)
\end{aligned}$$

By definition, the function

$$\hat{\phi}(t) = \frac{\phi(t)}{c}$$

is a valid PDF, so we can employ the KLD in order to measure the probabilistic distance between two normalized characteristic functions, $\hat{\phi}_1(t)$, $\hat{\phi}_2(t)$.

Let $\phi_1(t)$, $\phi_2(t)$ be two characteristic functions with normalizing factors c_1 and c_2, respectively, corresponding to two $S\alpha S$ densities with parameters (α_1, γ_1) and (α_2, γ_2). The KLD between their normalized versions, $\hat{\phi}_1(t)$, $\hat{\phi}_2(t)$ is computed as follows:

$$\begin{aligned}
D_1(\hat{\phi}_1 \| \hat{\phi}_2) &= \int_{-\infty}^{\infty} \hat{\phi}_1(t) \ln \frac{\hat{\phi}_1(t)}{\hat{\phi}_2(t)} \, \mathrm{d}t = \int_{-\infty}^{\infty} \frac{\phi_1(t)}{c_1} \ln \frac{c_2 \, \phi_1(t)}{c_1 \, \phi_2(t)} \, \mathrm{d}t \\
&= \int_{-\infty}^{\infty} \frac{\phi_1(t)}{c_1} \Big(\ln \frac{c_2}{c_1} + \ln \frac{\phi_1(t)}{\phi_2(t)} \Big) \, \mathrm{d}t \\
&= \int_{-\infty}^{\infty} \frac{\phi_1(t)}{c_1} \ln \frac{c_2}{c_1} \, \mathrm{d}t + \int_{-\infty}^{\infty} \frac{\phi_1(t)}{c_1} \ln \frac{\phi_1(t)}{\phi_2(t)} \, \mathrm{d}t \\
&= \ln \frac{c_2}{c_1} \underbrace{\int_{-\infty}^{\infty} \frac{\phi_1(t)}{c_1}}_{=1} \mathrm{d}t + \frac{1}{c_1} \int_{-\infty}^{\infty} \phi_1(t) \ln \frac{\phi_1(t)}{\phi_2(t)} \, \mathrm{d}t \\
&= \ln \frac{c_2}{c_1} + \frac{1}{c_1} \underbrace{\int_{-\infty}^{\infty} \phi_1(t) \ln \frac{\phi_1(t)}{\phi_2(t)} \, \mathrm{d}t}_{(I)}
\end{aligned}$$

$$\begin{aligned}
(I) &= \int_{-\infty}^{\infty} e^{-\gamma_1 |t|^{\alpha_1}} \ln \frac{e^{-\gamma_1 |t|^{\alpha_1}}}{e^{-\gamma_2 |t|^{\alpha_2}}} \, \mathrm{d}t = \int_{-\infty}^{\infty} e^{-\gamma_1 |t|^{\alpha_1}} \ln e^{-\gamma_1 |t|^{\alpha_1} + \gamma_2 |t|^{\alpha_2}} \, \mathrm{d}t \\
&= \int_{-\infty}^{\infty} e^{-\gamma_1 |t|^{\alpha_1}} \left(-\gamma_1 |t|^{\alpha_1} + \gamma_2 |t|^{\alpha_2} \right) \mathrm{d}t \\
&= \underbrace{\int_{-\infty}^{\infty} e^{-\gamma_1 |t|^{\alpha_1}} \left(-\gamma_1 |t|^{\alpha_1} \right) \mathrm{d}t}_{(A)} + \underbrace{\int_{-\infty}^{\infty} e^{-\gamma_1 |t|^{\alpha_1}} \left(\gamma_2 |t|^{\alpha_2} \right) \mathrm{d}t}_{(B)}
\end{aligned}$$

Notice that the integral (B) is in a general form. That is, if we compute (B), then the value of (A) is obtained by setting $\alpha_2 = \alpha_1$ and $\gamma_2 = \gamma_1$. The integral (B) is computed as follows:

$$\begin{aligned}
(B) &= \int_{-\infty}^{\infty} e^{-\gamma_1 |t|^{\alpha_1}} \left(\gamma_2 |t|^{\alpha_2} \right) \mathrm{d}t = 2 \int_{0}^{\infty} e^{-\gamma_1 t^{\alpha_1}} \, \gamma_2 t^{\alpha_2} \, \mathrm{d}t \\
&= 2\gamma_2 \int_{0}^{\infty} e^{-\gamma_1 t^{\alpha_1}} \, t^{\alpha_2} \, \mathrm{d}t \\
&\overset{u=\gamma_1 t^{\alpha_1}}{=} 2\gamma_2 \int_{0}^{\infty} e^{-u} \Big[\Big(\frac{u}{\gamma_1} \Big)^{1/\alpha_1} \Big]^{\alpha_2} \frac{1}{\gamma_1 \, \alpha_1} \Big[\Big(\frac{u}{\gamma_1} \Big)^{1/\alpha_1} \Big]^{1-\alpha_1} \mathrm{d}u \\
&= \frac{2\,\gamma_2}{\gamma_1 \, \alpha_1} \int_{0}^{\infty} e^{-u} \Big(\frac{u}{\gamma_1} \Big)^{\frac{\alpha_2+1}{\alpha_1} - 1} \mathrm{d}u \\
&= \frac{2\,\gamma_2}{\alpha_1 \, \gamma_1^{(\alpha_2+1)/\alpha_1}} \int_{0}^{\infty} e^{-u} \, u^{\left(\frac{\alpha_2+1}{\alpha_1} \right) - 1} \mathrm{d}u \\
&= \frac{2\,\gamma_2}{\alpha_1 \, \gamma_1^{(\alpha_2+1)/\alpha_1}} \, \Gamma\Big(\frac{\alpha_2+1}{\alpha_1} \Big)
\end{aligned}$$

Thus, the integral (A) equals

$$(A) = -\frac{2}{\alpha_1^2 \, \gamma_1^{1/\alpha_1}} \, \Gamma\Big(\frac{1}{\alpha_1} \Big)$$

Finally, the combination of the above expressions results in the KLD between normalized char-

acteristic functions:

$$
\begin{aligned}
D_1(\hat{\phi}_1 \| \hat{\phi}_2) &= \ln \frac{c_2}{c_1} + \frac{1}{c_1} \Big[-\frac{2}{\alpha_1^2 \, \gamma_1^{1/\alpha_1}} \, \Gamma\Big(\frac{1}{\alpha_1}\Big) + \frac{2 \, \gamma_2}{\alpha_1 \, \gamma_1^{(\alpha_2+1)/\alpha_1}} \, \Gamma\Big(\frac{\alpha_2+1}{\alpha_1}\Big) \Big] \\
&= \ln \frac{c_2}{c_1} - \frac{c_1}{c_1 \, \alpha_1} + \frac{1}{c_1 \, \alpha_1} \, \frac{2 \, \gamma_2}{\gamma_1^{(\alpha_2+1)/\alpha_1}} \, \Gamma\Big(\frac{\alpha_2+1}{\alpha_1}\Big) \\
&= \ln \frac{c_2}{c_1} - \frac{1}{\alpha_1} + \frac{2 \, \gamma_2}{c_1 \, \alpha_1 \, \gamma_1^{(\alpha_2+1)/\alpha_1}} \, \Gamma\Big(\frac{\alpha_2+1}{\alpha_1}\Big)
\end{aligned}
$$

which is the desired result given by Eq. (5.16). We can easily obtain Eqs. (5.17), (5.18) following an analogous procedure.

Appendix B

Steerable multivariate sub-Gaussian model

In this section we give the proof for the proposition 7.2.1 which derives a steerable model based on the covariation matrices of sub-Gaussian vectors. In particular, we show how to obtain the lth-level covariation matrix for a rotated image I, decomposed in multiple orientations and scales using a steerable pyramid, if we have already computed the corresponding lth-level covariation matrix of its original version.

Let $\mathbf{C}^l$ and $\mathbf{R}^l$ denote the lth-level covariation and underlying covariance matrix, respectively and $c^l(x_k, \phi)$ represent the value of a transform coefficient at a spatial location x_k, orientation ϕ and level l. For convenience, we omit the superscript l in the following study. The relation between the two matrices, $\mathbf{C}^l$ and $\mathbf{R}^l$, is given by:

$$[\mathbf{C}]_{ij} = 2^{-\frac{\alpha}{2}} \, [\mathbf{R}]_{ij} \, [\mathbf{R}]_{jj}^{(\alpha-2)/2} \, . \tag{B.1}$$

This results in the following relation between the elements of the covariation matrix $\mathbf{C}_\theta$ of the rotated image and the corresponding elements of the underlying covariance matrix $\mathbf{R}_\theta$:

$$[\mathbf{C}_\theta]_{ij} = 2^{-\frac{\alpha}{2}} \, [\mathbf{R}_\theta]_{ij} \, [\mathbf{R}_\theta]_{jj}^{(\alpha-2)/2} \tag{B.2}$$

The matrices $\mathbf{R}$, $\mathbf{R}_\theta$ are equivalent, i.e., they can be written in the form:

$$\mathbf{R}_\theta = \mathbf{F}(\theta)\mathbf{R}\mathbf{F}^T(\theta) \tag{B.3}$$

where $\mathbf{F}(\theta)$ is given by Eq. (7.3). Expanding (B.3) results in the following equivalent relation:

$$\begin{aligned} [\mathbf{R}_\theta]_{ij} &= \sum_{m=1}^{J} \Big(\sum_{n=1}^{J} f_n(\phi_i - \theta) \, [\mathbf{R}]_{nm} \Big) \, f_m(\phi_j - \theta) \\ &= \sum_{m=1}^{J} \sum_{n=1}^{J} f_n(\phi_i - \theta) \, f_m(\phi_j - \theta) \, [\mathbf{R}]_{nm}, \quad i \neq j \end{aligned} \tag{B.4}$$

and

$$[\mathbf{R}_\theta]_{jj} = \sum_{m=1}^{J}\sum_{n=1}^{J} f_n(\phi_j - \theta)\, f_m(\phi_j - \theta)\, [\mathbf{R}]_{nm} \tag{B.5}$$

for the diagonal elements of the underlying covariance matrix, where J is the number of basic orientations $\{\phi_1, \ldots, \phi_J\}$. Solving Eq. (B.1) with respect to $[\mathbf{R}]_{ij}$ we have

$$[\mathbf{R}]_{ij} = 2\,[\mathbf{C}]_{ij}\,[\mathbf{C}]_{jj}^{(2-\alpha)/\alpha} \tag{B.6}$$

Combining Eqs. (B.2), (B.5) we proceed as follows:

$$\begin{aligned}
[\mathbf{C}_\theta]_{ij} &= 2^{-\frac{\alpha}{2}} \Big(\sum_{k=1}^{J}\sum_{l=1}^{J} f_l(\phi_i - \theta)\, f_k(\phi_j - \theta)\, [\mathbf{R}]_{lk}\Big) \cdot \\
&\quad \Big(\sum_{m=1}^{J}\sum_{n=1}^{J} f_n(\phi_j - \theta)\, f_m(\phi_j - \theta)\, [\mathbf{R}]_{nm}\Big)^{(\alpha-2)/2} \\
&= 2^{-\frac{\alpha}{2}} \Big(\sum_{k=1}^{J}\sum_{l=1}^{J} f_l(\phi_i - \theta)\, f_k(\phi_j - \theta)\, 2\,[\mathbf{C}]_{lk}\; [\mathbf{C}]_{kk}^{(2-\alpha)/\alpha}\Big) \cdot \\
&\quad \Big(\sum_{m=1}^{J}\sum_{n=1}^{J} f_n(\phi_j - \theta)\, f_m(\phi_j - \theta)\, 2\,[\mathbf{C}]_{nm}\, [\mathbf{C}]_{mm}^{(2-\alpha)/\alpha}\Big)^{(\alpha-2)/2} \\
&= \Big(\sum_{k=1}^{J}\sum_{l=1}^{J} f_l(\phi_i - \theta)\, f_k(\phi_j - \theta)\, [\mathbf{C}]_{lk}\; [\mathbf{C}]_{kk}^{(2-\alpha)/\alpha}\Big) \cdot \\
&\quad \Big(\sum_{m=1}^{J}\sum_{n=1}^{J} f_n(\phi_j - \theta)\, f_m(\phi_j - \theta)\, [\mathbf{C}]_{nm}\, [\mathbf{C}]_{mm}^{(2-\alpha)/\alpha}\Big)^{(\alpha-2)/2}
\end{aligned}$$

which, in matrix notation, yields exactly the desired Eq. (7.7).

B.1 Rotation-Invariant Kullback-Leibler Divergence

In this section we give the proof of two results regarding the KLD between an image I and a rotated version of an image Q by an angle θ, Q_θ. The first result refers to the expression of the KLD between I and Q_θ, denoted as $D(I\|Q_\theta)$, as a function of the KLD between I and the original image Q, while the second result pertains to a rotation-invariant version of the KLD between multivariate Gaussian densities.

B.1.1 Relation of equivalence between the covariance matrices corresponding to images I and I_θ

First of all we derive the relation of equivalence between the lth-level covariance matrix of an image I ($\mathbf{\Sigma}^l$), with the corresponding matrix of its rotated version by an angle θ, I_θ ($\mathbf{\Sigma}_\theta^l$). The elements of these two matrices are the covariances between pairs of subbands at the l-th decomposition level. For simplicity we omit the superscript l in the following study.

We begin by relating the corresponding lth-level mean vectors $\vec{\mu}$, $\vec{\mu}_\theta$. Expressing their components in terms of the lth-level transform coefficients $c(x_k, \phi)$ we have:

$$\vec{\mu}_i = \frac{1}{N_l}\sum_{k=1}^{N_l} c(x_k, \phi_i) \tag{B.7}$$

$$\begin{aligned}\vec{\mu}_{\theta,i} &= \frac{1}{N_l}\sum_{k=1}^{N_l} c(x_k, \phi_i - \theta) \\ &= \frac{1}{N_l}\sum_{k=1}^{N_l}\sum_{m=1}^{J} f_m(\phi_i - \theta)c(x_k, \phi_m) = \sum_{m=1}^{J} f_m(\phi_i - \theta)\Big(\frac{1}{N_l}\sum_{k=1}^{N_l} c(x_k, \phi_m)\Big) \\ &= \sum_{m=1}^{J} f_m(\phi_i - \theta)\,\vec{\mu}_m\ , \quad i = 1, \ldots, J \end{aligned} \tag{B.8}$$

Eq. (B.8) can be rewritten in matrix notation as follows:

$$\vec{\mu}_\theta = \mathbf{F}(\theta)\,\vec{\mu} \tag{B.9}$$

By definition, the covariance of two random variables X, Y equals:

$$Cov\{X, Y\} = E\{XY\} - \mu_X\mu_Y$$

where $E\{\cdot\}$ denotes the expectation operator. From this equation, we obtain the relation between the covariance matrix $\mathbf{\Sigma}$ and the correlation matrix $\mathbf{R}$:

$$\mathbf{\Sigma} = \mathbf{R} - \vec{\mu}\vec{\mu}^{\,T} \tag{B.10}$$

For the covariance matrix of the rotated image I_θ we have:

$$\begin{aligned}\mathbf{\Sigma}_\theta &= \mathbf{R}_\theta - \vec{\mu}_\theta\vec{\mu}_\theta^{\,T} \overset{(7.2),(B.9)}{=} \mathbf{F}(\theta)\,\mathbf{R}\,\mathbf{F}^T(\theta) - \mathbf{F}(\theta)\,\vec{\mu}\,(\mathbf{F}(\theta)\,\vec{\mu})^T \\ &= \mathbf{F}(\theta)\,\mathbf{R}\,\mathbf{F}^T(\theta) - \mathbf{F}(\theta)\,\vec{\mu}\,\vec{\mu}^{\,T}\,\mathbf{F}^T(\theta) \\ &= \mathbf{F}(\theta)\,\underbrace{[\mathbf{R} - \vec{\mu}\,\vec{\mu}^{\,T}]}_{\mathbf{\Sigma}}\,\mathbf{F}^T(\theta) \\ &= \mathbf{F}(\theta)\,\mathbf{\Sigma}\,\mathbf{F}^T(\theta)\end{aligned} \tag{B.11}$$

which proves the equivalence between the two covariance matrices.

B.1.2 KLD $D(I\|Q_\theta)$ as a function of the KLD $D(I\|Q)$

In the following we derive an expression which gives the KLD between image I and the rotated image Q_θ, as a function of the KLD between I and the original Q, under the assumption that their subband coefficients at a given decomposition level l follow a multivariate Gaussian distribution.

Let $(\vec{\mu}_I,\ \mathbf{\Sigma}_I)$, $(\vec{\mu}_Q,\ \mathbf{\Sigma}_Q)$ and $(\vec{\mu}_{Q,\theta},\ \mathbf{\Sigma}_{Q,\theta})$ denote the mean vector-covariance matrix pairs for the images I, Q and Q_θ, respectively, at a given decomposition level l (index l is omitted for

simplicity). The KLD between the lth-levels of I and Q, modeled using multivariate Gaussian densities, is given by:

$$D(I\|Q) = \frac{1}{2}\Big((\vec{\mu}_Q - \vec{\mu}_I)^T \mathbf{\Sigma}_Q^{-1}(\vec{\mu}_Q - \vec{\mu}_I) + tr(\mathbf{\Sigma}_I \mathbf{\Sigma}_Q^{-1} - I) - \ln|det(\mathbf{\Sigma}_I \mathbf{\Sigma}_Q^{-1})|\Big) \tag{B.12}$$

The corresponding expression for the KLD between the lth-levels of I and Q_θ, as a function of $D(I\|Q)$, is derived as follows:

$$\begin{aligned}
D(I\|Q_\theta) &= \frac{1}{2}\Big((\vec{\mu}_{Q,\theta} - \vec{\mu}_I)^T \mathbf{\Sigma}_{Q,\theta}^{-1}(\vec{\mu}_{Q,\theta} - \vec{\mu}_I) + tr(\mathbf{\Sigma}_I \mathbf{\Sigma}_{Q,\theta}^{-1} - I) - \\
&\quad - \ln|det(\mathbf{\Sigma}_I \mathbf{\Sigma}_{Q,\theta}^{-1})|\Big) \\
&\overset{(B.9),(B.11)}{=} \frac{1}{2}\Big((\mathbf{F}(\theta)\vec{\mu}_Q - \vec{\mu}_I)^T (\mathbf{F}(\theta)\mathbf{\Sigma}_Q \mathbf{F}^T(\theta))^{-1}(\mathbf{F}(\theta)\vec{\mu}_Q - \vec{\mu}_I) + \\
&\quad + tr(\mathbf{\Sigma}_I(\mathbf{F}(\theta)\mathbf{\Sigma}_Q \mathbf{F}^T(\theta))^{-1} - I) - \ln|det(\mathbf{\Sigma}_I(\mathbf{F}(\theta)\mathbf{\Sigma}_Q \mathbf{F}^T(\theta))^{-1})|\Big) \\
&\overset{(*)}{=} \frac{1}{2}\Big((\vec{\mu}_Q^T \mathbf{F}^T(\theta) - \vec{\mu}_I^T)((\mathbf{F}^T(\theta))^{-1}\mathbf{\Sigma}_Q^{-1}\mathbf{F}^{-1}(\theta))(\mathbf{F}(\theta)\vec{\mu}_Q - \vec{\mu}_I) + \\
&\quad + tr(\mathbf{\Sigma}_I(\mathbf{F}^T(\theta))^{-1}\mathbf{\Sigma}_Q^{-1}\mathbf{F}^{-1}(\theta) - I) - \ln|det(\mathbf{\Sigma}_I \mathbf{\Sigma}_Q^{-1})|\Big) \\
&= \frac{1}{2}\Big(\vec{\mu}_Q^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q - \vec{\mu}_Q^T \mathbf{\Sigma}_Q^{-1}\mathbf{F}^{-1}(\theta)\vec{\mu}_I - \vec{\mu}_I^T(\mathbf{F}^T(\theta))^{-1}\mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q + \\
&\quad + \vec{\mu}_I^T(\mathbf{F}^T(\theta))^{-1}\mathbf{\Sigma}_Q^{-1}\mathbf{F}^{-1}(\theta)\vec{\mu}_I + tr(\mathbf{\Sigma}_I(\mathbf{F}^T(\theta))^{-1}\mathbf{\Sigma}_Q^{-1}\mathbf{F}^{-1}(\theta) - I) - \\
&\quad - \ln|det(\mathbf{\Sigma}_I \mathbf{\Sigma}_Q^{-1})|\Big) \quad (I)
\end{aligned}$$

In the last expression we add/remove the appropriate terms in order to construct the $D(I\|Q)$. We proceed as follows:

$$\begin{aligned}
(I) &= \frac{1}{2}\Big(\underbrace{\vec{\mu}_Q^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q} - \vec{\mu}_Q^T \mathbf{\Sigma}_Q^{-1}\mathbf{F}^{-1}(\theta)\vec{\mu}_I + \vec{\mu}_Q^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_I - \underbrace{\vec{\mu}_Q^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_I} - \\
&\quad - \vec{\mu}_I^T(\mathbf{F}^T(\theta))^{-1}\mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q + \vec{\mu}_I^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q - \underbrace{\vec{\mu}_I^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q} + \\
&\quad + \vec{\mu}_I^T(\mathbf{F}^T(\theta))^{-1}\mathbf{\Sigma}_Q^{-1}\mathbf{F}^{-1}(\theta)\vec{\mu}_I + \underbrace{\vec{\mu}_I^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_I} - \vec{\mu}_I^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_I + \\
&\quad + tr(\mathbf{\Sigma}_I(\mathbf{F}^T(\theta))^{-1}\mathbf{\Sigma}_Q^{-1}\mathbf{F}^{-1}(\theta) - I) + \underbrace{tr(\mathbf{\Sigma}_I \mathbf{\Sigma}_Q^{-1} - I)} - tr(\mathbf{\Sigma}_I \mathbf{\Sigma}_Q^{-1} - I) - \\
&\quad - \underbrace{\ln|det(\mathbf{\Sigma}_I \mathbf{\Sigma}_Q^{-1})|}\Big) \quad (II)
\end{aligned}$$

By an appropriate grouping of the terms in equation (II) and noticing that the underbraced terms

constitute $D(I\|Q)$ we have the desired relation between the two KLDs, $D(I\|Q)$ and $D(I\|Q_\theta)$:

$$\begin{aligned} D(I\|Q_\theta) \overset{(II)}{=} & \; D(I\|Q) + \frac{1}{2}\Big(\overbrace{\vec{\mu}_Q^T \mathbf{\Sigma}_Q^{-1}(\vec{\mu}_I - \mathbf{F}^{-1}(\theta)\vec{\mu}_I)}^{\mathbf{V}} + \overbrace{(\vec{\mu}_I - \mathbf{F}^{-1}(\theta)\vec{\mu}_I)^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q}^{\mathbf{V}^T} + \\ & + (\mathbf{F}^{-1}(\theta)\vec{\mu}_I)^T \mathbf{\Sigma}_Q^{-1}(\mathbf{F}^{-1}(\theta)\vec{\mu}_I) - \vec{\mu}_I^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_I + \\ & + tr(\mathbf{\Sigma}_I(\mathbf{F}^T(\theta))^{-1}\mathbf{\Sigma}_Q^{-1}\mathbf{F}^{-1}(\theta) - \mathbf{\Sigma}_I\mathbf{\Sigma}_Q^{-1})\Big) \\ \overset{(*)}{=} & \; D(I\|Q) + \frac{1}{2}\Big(\mathbf{V} + \mathbf{V}^T + \vec{\mu}_I^T \mathbf{F}(\theta)\mathbf{\Sigma}_Q^{-1}\mathbf{F}^T(\theta)\vec{\mu}_I - \vec{\mu}_I^T \mathbf{\Sigma}_Q^{-1}\vec{\mu}_I + \\ & + tr(\mathbf{\Sigma}_I[\mathbf{F}(\theta)\mathbf{\Sigma}_Q^{-1}\mathbf{F}^T(\theta) - \mathbf{\Sigma}_Q^{-1}])\Big) \end{aligned}$$

(*) [Notice that we have chosen equispaced basic orientations for the construction of the steerable pyramid, which makes matrix $\mathbf{F}(\theta)$ to be orthogonal, that is,

$$\mathbf{F}(\theta)\mathbf{F}^T(\theta) = I \Rightarrow \mathbf{F}^{-1}(\theta) = \mathbf{F}^T(\theta) \ .$$

By the properties of steering functions $f_i(\cdot)$ the following relations hold:

$$\mathbf{F}^{-1}(\theta) = \mathbf{F}(-\theta), \quad (\mathbf{F}^T(\theta))^{-1} = \mathbf{F}^T(-\theta) \ . \ ]$$

B.1.3 Rotation-Invariant KLD

In this section we derive a rotation-invariant version of the KLD between multivariate Gaussian densities. Assume that we want to measure the similarity between two images I and Q, where Q is a counter-clockwise rotated version of I by an angle ϕ. Both images are decomposed via a steerable pyramid in L levels and J orientations per level. Let I^l and Q^l denote the lth-level of I and Q, respectively, modeled using a J-dimensional Gaussian density.

The total distance between images I and Q is defined as:

$$D(I\|Q) = \min_\theta \sum_{l=1}^{L} D(I^l\|Q^l_{-\theta}) \tag{B.13}$$

where $D(I^l\|Q^l_{-\theta})$ is the KLD between the lth-levels of image I and a clockwise rotation of image Q. The purpose of this clockwise rotation is that we have to align I and Q before measuring their similarity. As the true angle ϕ is unknown, we must search for the best alignment angle of a set of possible rotations, and this justifies the min operator.

We focus on the computation of $D(I^l\|Q^l_{-\theta})$ in order to find a closed-form expression for Eq. (B.13). Let $(\vec{\mu}_{I^l}, \mathbf{\Sigma}_{I^l})$ and $(\vec{\mu}_{Q^l}, \mathbf{\Sigma}_{Q^l})$ denote the mean vector-covariance matrix pairs at the lth decomposition level of I and Q, respectively, omitting the index l for convenience in the following study. From Eqs. (B.9), (B.11), the mean vector and the covariance matrix corresponding

to $Q^l_{-\theta}$ are given by:

$$\vec{\mu}_{Q_{-\theta}} = \mathbf{F}(-\theta)\vec{\mu}_Q \tag{B.14}$$

$$\mathbf{\Sigma}_{Q_{-\theta}} = \mathbf{F}(-\theta)\mathbf{\Sigma}_Q\mathbf{F}^T(-\theta) \tag{B.15}$$

Combining Eqs. (B.12), (B.14) and (B.15) we obtain the expression for the KLD $D(I^l\|Q^l_{-\theta})$, as follows:

$$\begin{aligned}
D(I^l\|Q^l_{-\theta}) &= \frac{1}{2}\Big((\mathbf{F}(-\theta)\vec{\mu}_Q - \vec{\mu}_I)^T\,(\mathbf{F}(-\theta)\mathbf{\Sigma}_Q\mathbf{F}^T(-\theta))^{-1}\,(\mathbf{F}(-\theta)\vec{\mu}_Q - \vec{\mu}_I) + \\
&\quad + tr(\mathbf{\Sigma}_I\,(\mathbf{F}(-\theta)\mathbf{\Sigma}_Q\mathbf{F}^T(-\theta))^{-1} - I) - \ln|det(\mathbf{\Sigma}_I(\mathbf{F}(-\theta)\mathbf{\Sigma}_Q\mathbf{F}^T(-\theta))^{-1}|\Big) \\
&= \frac{1}{2}\Big((\vec{\mu}_Q^T\mathbf{F}^T(-\theta) - \vec{\mu}_I^T)\,(\mathbf{F}^T(\theta)\mathbf{\Sigma}_Q^{-1}\mathbf{F}(\theta))\,(\mathbf{F}(-\theta)\vec{\mu}_Q - \vec{\mu}_I) + \\
&\quad + tr(\mathbf{\Sigma}_I\,(\mathbf{F}(-\theta)\mathbf{\Sigma}_Q\mathbf{F}^T(-\theta))^{-1} - I) - \ln|det(\mathbf{\Sigma}_I\mathbf{\Sigma}_Q^{-1})|\Big) \\
&= \frac{1}{2}\Big(\vec{\mu}_Q^T\mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q - \vec{\mu}_Q^T\mathbf{\Sigma}_Q^{-1}\mathbf{F}(\theta)\vec{\mu}_I - \vec{\mu}_I^T\mathbf{F}^T(\theta)\mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q + \vec{\mu}_I^T\mathbf{F}^T(\theta)\mathbf{\Sigma}_Q^{-1}\mathbf{F}(\theta)\vec{\mu}_I + \\
&\quad + tr(\mathbf{\Sigma}_I\,(\mathbf{F}(-\theta)\mathbf{\Sigma}_Q\mathbf{F}^T(-\theta))^{-1} - I) - \ln|det(\mathbf{\Sigma}_I\mathbf{\Sigma}_Q^{-1})|\Big) \\
&= \frac{1}{2}\Big(\vec{\mu}_Q^T\mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q - \overbrace{\vec{\mu}_Q^T\mathbf{\Sigma}_Q^{-1}\mathbf{F}(\theta)\vec{\mu}_I}^{\mathbf{V}(\theta)} - \overbrace{\vec{\mu}_I^T\mathbf{F}^T(\theta)\mathbf{\Sigma}_Q^{-1}\vec{\mu}_Q}^{\mathbf{V}^T(\theta)} + \vec{\mu}_I^T\mathbf{F}^T(\theta)\mathbf{\Sigma}_Q^{-1}\overbrace{\mathbf{F}(\theta)\vec{\mu}_I}^{\vec{r}(\theta)} + \\
&\quad + tr(\mathbf{\Sigma}_I\,(\mathbf{F}^T(\theta)\mathbf{\Sigma}_Q^{-1}\mathbf{F}(\theta))) - tr(I) - \ln|det(\mathbf{\Sigma}_I\mathbf{\Sigma}_Q^{-1})|\Big) \quad (III)
\end{aligned}$$

In the last equation, we group the terms that depend on θ and take into account that $tr(I) = J$, since I denotes here the $J \times J$ identity matrix. Substituting (III) in Eq. (B.13) and applying the min operator only on those terms which depend on θ, we obtain a closed-form expression for the total KLD $D(I\|Q)$:

$$\begin{aligned}
D(I\|Q) &= \min_\theta\left[\sum_{l=1}^{L}\frac{1}{2}\Big(-(\mathbf{V}_l(\theta) + \mathbf{V}_l^T(\theta)) + \vec{r}_l^T(\theta)\,\mathbf{\Sigma}_{Q^l}^{-1}\,\vec{r}_l(\theta) + tr(\mathbf{\Sigma}_{I^l}\,\mathbf{F}^T(\theta)\,\mathbf{\Sigma}_{Q^l}^{-1}\,\mathbf{F}(\theta))\Big)\right] + \\
&\quad + \sum_{l=1}^{L}\frac{1}{2}\Big(\vec{\mu}_{Q^l}^T\mathbf{\Sigma}_{Q^l}^{-1}\vec{\mu}_{Q^l} - \ln|det(\mathbf{\Sigma}_{I^l}\mathbf{\Sigma}_{Q^l}^{-1})| - \underbrace{J}_{=tr(I)}\Big)
\end{aligned}$$

where

$$\mathbf{V}_l(\theta) = \vec{\mu}_{Q^l}^T\mathbf{\Sigma}_{Q^l}^{-1}\,\mathbf{F}(\theta)\vec{\mu}_{I^l}$$

$$\vec{r}_l(\theta) = \mathbf{F}(\theta)\vec{\mu}_{I^l}$$

Assuming zero-mean Gaussian densities, that is, $\vec{\mu}_{I^l} = \vec{0}$ and $\vec{\mu}_{Q^l} = \vec{0}$, the expression of the total

KLD can be further simplified:

$$\begin{aligned} D(I\|Q) &= \min_{\theta}\Big[\frac{1}{2}\sum_{l=1}^{L} tr(\mathbf{\Sigma}_{I^l}\,\mathbf{F}^T(\theta)\,\mathbf{\Sigma}_{Q^l}^{-1}\,\mathbf{F}(\theta))\Big] - \frac{JL}{2} - \\ &\quad - \frac{1}{2}\sum_{l=1}^{L}\ln|det(\mathbf{\Sigma}_{I^l}\mathbf{\Sigma}_{Q^l}^{-1})| \end{aligned} \tag{B.16}$$

which is exactly the expression of Eq. (7.26).

Bibliography

[1] A Achim, A Bezerianos, and P Tsakalides. Novel Bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Trans. Med. Imag.*, 20:772–783, Aug. 2001.

[2] A Achim, P Tsakalides, and A Bezerianos. SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosc. and Rem. Sens.*, 41(8):1773–1784, Aug. 2003.

[3] D Andrews and C Mallows. Scale mixtures of normal distributions. *J. Royal Stat. Soc.*, 36:99–, 1974.

[4] S Areepongsa, D Park, and K R Rao. Invariant features for texture image retrieval using steerable pyramid.

[5] J Ashley, M Flickner, J Hafner, D Lee, W Niblack, and D Petkovic. The query by image content (qbic) system. In *Proceedings of the 1995 ACM SIGMOD Interl. Conf. on Management of data*, pages 475–, New York, U.S., 1995. ACM Press.

[6] B Beferull-Lozano, H Xie, and A Ortega. Rotation-invariant features based on steerable transforms with an application to distributed image classification.

[7] S Belongie, J Malik, and J Puzicha. Matching shapes. *In International Conference on Computer Vision, Vancouver, Canada*, 2001.

[8] J R Bergen and E H Adelson. Theories of visual texture perception. In D Regan Ed., editor, *Spatial Vision*. CRC press, 1991.

[9] P Bidarkota and J H McCulloch. Real stock returns: Non-normality, seasonality, and volatility persistence, but no predictanility. *Working Paper*, Sept. 1996.

[10] R N Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill, 1986.

[11] B W Brorsen and S R Yang. Maximum likelihood estimates of symmetric stable distribution parameters. *Commun. Statist.-Simul.*, 19:1459–1464, 1990.

[12] R W Buccigrossi and E P Simoncelli. Image compression via joint statistical characterization in the wavelet domain. Technical Report 414, Munich, Germany, 1997.

[13] R W Buccigrossi and E P Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. Image Proc.*, 8(12):1688–1701, Dec. 1999.

[14] S Cambanis. Complex symmetric stable variables and processes. In P Sen, editor, *Contributions to Statistics: Essays in Honor of Norman L. Johnson*, pages 63–79. North-Holland, New York, 1983.

[15] S Cambanis, C D Hardin, and A Weron. Ergodic properties of stationary stable processes. *Stochastic Processes and their Applications*, 24:1–18, 1987.

[16] S Cambanis and G Miller. Linear problems in pth order and stable processes. *SIAM J. Appl. Math.*, 41:43–69, 1981.

[17] S Cambanis, G Samorodnitsky, and M S Taqqu, editors. *Stable Processes and Related Topics*. Birkhauser, Boston, 1991.

[18] S Cambanis and A R Soltani. Prediction of stable processes: Spectral and moving average representations. *Z. Wahrsch verw. Gebiete*, 66:593–612, 1984.

[19] C Carson, S Belongie, H Greenspan, and J Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8), August 2002.

[20] V Castelli, L D Bergman, I Kontoyiannis, C S Li, J T Robinson, and J J Turek. Progressive search and retrieval in large image archives.

[21] J Chambers, W Cleveland, B Kleiner, and P Tukey. *Graphical Methods for Data Analysis*. Wadsworth, 1983.

[22] J M Chambers, C L Mallows, and B W Stuck. A method of simulating stable random variables. *Journal of the American Statistical Association*, 71:340–344, 1976.

[23] S G Chang, Bin Yu, and M Vetterli. Lossy compression and wavelet thresholding for image denoising. *submitted to IEEE Trans. Image Processing*.

[24] J Y Chen, C A Bouman, and J P Allebach. Multiscale branch and bound algorithm image database. *Proc SPIE*, 3022:133–144, 1997.

[25] C K Chui. *An Introduction to Wavelets*. Academic Press, 1992.

[26] A Consiglio, I Massabo, and S Ortobelli. Non-gaussian distribution for var calculation: an assessment for the italian market. *IFAC*, 2001.

[27] T Cover and J Thomas. *Elements of Information Theory*. John Wiley, 1991.

[28] G Cross and A Jain. Markov random field texture models. *IEEE Trans Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.

[29] M Crouse, R Nowak, and R Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on Signal Processing*, 1997.

[30] M S Crouse, R D Nowak, and R G Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. on Signal Processing (Special issue on Wavelets and Filterbanks)*, pages 886–902, April 1998.

[31] J Dahmen, T Theiner, D Keysers, H Ney, T Lehmann, and B Wein. Classification of radiographs in the Image Retrieval in Medical Applications-system (IRMA).

[32] M N Do and M Vetterli. Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models. *Submitted to IEEE Trans. on Multimedia*, April 2001.

[33] M N Do and M Vetterli. Rotation invariant texture characterization and retrieval using steerable wavelet domain hidden markov models. *IEEE Trans. on Multimedia*, 4:517–527, Dec. 2002.

[34] M N Do and M Vetterli. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Trans. Image Processing*, 11:146–158, Feb. 2002.

[35] W H DuMouchel. Stable distributions in statistical interference: 2. Information from stably distributed samples. *J. Amer. Statist. Assoc.*, 70:386–393, 1970.

[36] W H DuMouchel. *Stable Distributions in Statistical Inference*. PhD thesis, Dept. of Statistics, Yale University, 1971.

[37] H Dym and H P McKean. *Fourier Series and Integrals*. Academic Press, 1972.

[38] C Faloutsos, R Barber, M Flickner, J Hafner, W Niblack, D Petkovic, and W Equitz. Efficient and effective querying by image content. *Journal of Intelligent Information Systems Integrating Artificial Intelligence and Database Technologies*, 3(3-4):231–262, July 1994.

[39] E Fama. The behavior of stock market prices. *Journal of Business*, 38:34–105, 1965.

[40] E F Fama and R Roll. Parameter estimates for symmetric stable distributions. *Journal Amer. Statist. Assoc.*, 66:331–338, June 1971.

[41] D A Forsyth, J Malik, M M Fleck, H Greenspan, T Leung, S Belongie, C Carson, and C Bregler. Finding pictures of objects in large collections of images. *In Proc. ECCV 96 Workshop on Object Rep.*, 1996.

[42] R Francois, R Fablet, and C Barillot. Robust statistical registration of 3d ultrasound images using texture information.

[43] W T Freeman and E H Adelson. The design and use of steerable filters. *IEEE Trans. Patt. Anal. Mach. Intell.*, 13(9):891–906, 1991.

[44] K Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, 1990.

[45] P G Georgiou and C Kyriakakis. Maximum likelihood parameter estimation under impulsive conditions, a sub-gaussian signal approach. *IEEE Trans. Signal Proc.*

[46] P G Georgiou, P Tsakalides, and C Kyriakakis. Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise. *IEEE Trans. Multimedia*, 1:291–301, Sep. 1999.

[47] Y Gong, H Zhang, H C Chuan, and M Sakauchi. An image database system with content capturing and fast image indexing abilities. *In Proceedings of the International Conference on Multimedia Computing and Systems*, pages 121–130, 1994.

[48] R C Gonzalez and R E Woods. *Digital Image Processing, 2nd Edition.* Prentice Hall, 2002.

[49] H Greenspan, S Belongie, R Goodman, and P Perona. Rotation invariant texture recognition using a steerable pyramid. *in Proc. Int. Conf. on Pattern Recognition*, pages 162–167, 1994.

[50] A Gupta and R Jain. Visual information retrieval. *Communications of the ACM*, 40(5), May 1997.

[51] J Hafner, H Sawhney, W Equitz, M Flickner, and W Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736, July 1995.

[52] G M Haley and B S Manjunath. Rotation invariant texture classification using modified gabor filters. *in Proc. IEEE Int. Conf. Image Processing, Washington, DC*, Oct 1995.

[53] G M Haley and B S Manjunath. Rotation-invariant texture classification using a complete space-frequency model. *IEEE Trans. Image Processing*, 8:255–269, Feb 1999.

[54] M Hassner and J Sklansky. The use of markov random fields as models of texture. *Comp. Graphics Image Proc.*, 12:357–370, 1980.

[55] J Huang. Study on the correlation properties of wavelet transform coe•cients and the applications in a neural network-based hybrid image coding system. *CISST'03, Las Vegas, USA*, June 22-26 2003.

[56] C E Jacobs, A Finkelstein, and D H Salesin. Fast multiresolution image querying. *Proceedings of SIGGAPH '95 in Computer Graphics Proceedings Annual Conference Series*, pages 277–286, August 1995.

[57] A Jain and A Vailaya. Image retrieval using color and shape. *Pattern Recognition Journal*, 29:1233–1244, August 1996.

[58] D W Kammler. *A First Course in Fourier Analysis.* Prentice Hall, Upper Saddler River, NJ, 2000.

[59] S M Kogon and D B Williams. On characteristic function based stable distribution parameter estimation techniques. In R J Adler, R E Feldman, and M S Taqqu, editors, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* Birkhauser, Boston, MA, 1998.

[60] T W Korner. *Fourier Analysis.* Cambridge University Press, 1988.

[61] I A Koutrouvelis. Regression-type estimation of the parameters of stable laws. *J. Amer. Statist. Assoc.*, 75(372):918–928, Dec. 1980.

[62] I A Koutrouvelis. An iterative procedure for the estimation of the parameters of stable laws. *Commun. Statist.-Simula.*, 10(1):17–28, 1981.

[63] S Kullback. *Information Theory and Statistics.* Dover, 1997.

[64] P Lévy. *Calcul des Probabilités.* Paris: Gauthier-Villars, 1925.

[65] J Li and R M Gray. Text and picture segmentation by the distribution analysis of wavelet coefficients. *Int. Conf. Image Processing, Chicago*, 3:790–794, Oct 1998.

[66] S Liapis and G Tziritas. Color and texture image retrieval using chromaticity histograms and wavelet frames. *IEEE Trans. on Multimedia*, 6(5):676–686, Oct. 2004.

[67] F Liu and R W Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Trans. Patt. Anal. Machine Intell.*, 18:722–733, July 1996.

[68] S LoPresto, K Ramchandran, and M Orchard. Wavelet image coding via rate-distortion optimized adaptive classification. 1997.

[69] X Ma and C L Nikias. On blind channel identification for impulsive signal environments: A method-of-moments approach. Technical Report USC-SIPI, University of Southern California, September 1994.

[70] F W Machell, C S Penrod, and G E Ellis. Statistical characteristic of ocean acoustic noise processes. pages 29–57, 1989.

[71] J Malik and P Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of Optical Soc. Am.*, 7(5).

[72] S Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[73] S Mallat. *A Wavelet Tour of Signal Processing, 2nd Edition.* Academic Press, 1998.

[74] S G Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, July 1989.

[75] S G Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11:674–692, July 1989.

[76] B Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963.

[77] B Manjunath, J R Ohm, V Vasudevan, and A Yamada. Color and texture descriptors. *IEEE Trans. Circuits and Syst. Video Tech.*, 11:703–715, June 2001.

[78] B S Manjunath and W Y Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Patt. Anal. Machine Intell.*, 18:837–842, Aug. 1996.

[79] J Mao and A Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, 25:173–188, Feb. 1992.

[80] E Masry and S Cambanis. Spectral density estimation for stationary stable processes. *Stochastic Processes and their Applications*, 18:1–31, 1984.

[81] D Middleton. Statistical-physical models of electromagnetic interference. *IEEE Trans. Electromagn. Compat.*, EMC-19(3):106–127, 1977.

[82] M K Mihcak, I Kozintsev, K Ramchandran, and P Moulin. Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Processing Letters*, 6(12):300–303, 1999.

[83] S Mittnik and T Doganoglu. An approximation procedure for asymmetric stable paretian densities. *Inst. of Stat. and Econometrics, Christian Albrechts Univ. at Kiel.*

[84] S Mittnik and T Doganoglu. Computing the probability density function of the stable paretian distribution. *Inst. of Stat. and Econometrics, Christian Albrechts Univ. at Kiel.*

[85] N Morrison. *Introduction to Fourier Analysis.* New York: Wiley, 1994.

[86] C L Nikias and M Shao. *Signal Processing with Alpha-Stable Distributions and Applications.* John Wiley and Sons, New York, 1995.

[87] J P Nolan. Numerical calculation of stable densities and distribution functions. *Commun. Statist.-Stochastic Models*, 13:759–774, 1997.

[88] John P. Nolan. Parameterizations and modes of stable distributions. *Statistics & Probability Letters*, (38):187–195, 1998.

[89] A S Paulson, E W Holcomb, and R A Leitch. The estimation of the parameters of the stable laws. *Biometrika*, 62:163–170, 1975.

[90] A Pentland, R W Picard, and S Sclaroff. Photobook: tools for content-based manipulation of image databases. *in Proc. SPIE, 7-8 February, San Jose*, 2185:34–47, 1994.

[91] A Pentland, R W Picard, and S Sclaroff. Photobook: tools for content-based manipulation of image databases. *Proc SPIE, San Jose*, 2185:34–47, February 7-8 1994.

[92] J Portilla and E P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–71, Dec. 2000.

[93] J Portilla and E P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40:49–71, Dec. 2000.

[94] J Portilla, V Strela, M Wainwright, and E Simoncelli. Image denoising using gaussian scale mixtures in the wavelet domain. *Technical Report, TR2002-831, New York University*, Sept. 2002.

[95] S Resnick. Heavy tail modeling and teletraffic data. *Preprint*, 1995.

[96] Y Rubner, C Tomasi, and L Guibas. A metric for distributions with applications to image databases. *International Conference on Computer Vision, Bombay, India*, 1998.

[97] G Samorodnitsky and M S Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, New York, 1994.

[98] M Schilder. Some structure theorems for the symmetric stable laws. *The Annals of Mathematical Statistics*, 41:412–421, 1970.

[99] S C Schwartz and J B Thomas. Detection in a non-gaussian environment. pages 93–105, 1984.

[100] J Seo, S Cho, and K Feher. Impact of non-gaussian impulsive noise on the performance of high-level qam. *IEEE Trans. Electromagn. Compat.*, EMC-31:177–180, May 1989.

[101] M Shao and C L Nikias. On symmetric stable models for impulsive noise. Technical Report USC-SIPI-231, University of Southern California, Feb. 1993.

[102] M Shao and C L Nikias. Signal processing with fractional lower order moments: Stable processes and their applications. *Proc. IEEE*, 81:986–1010, 1993.

[103] E P Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf on Signals, Systems and Computers*, pages 673–678, Pacific Grove, CA, 1997. IEEE Computer Society.

[104] E P Simoncelli and W T Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. *2nd Annual IEEE International Conference on Image Processing, Washington, DC*, October 1995.

[105] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger. Shiftable multi-scale transforms. *IEEE Trans. Information Theory*, 38(2):587–607, March 1992.

[106] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger. Shiftable multi-scale transforms. *IEEE Trans. Inform. Theory*, 38(2), 1992.

[107] E P Simoncelli and J Portilla. Texture characterization via joint statistics of wavelet coefficient magnitudes. *Proceedings of Fifth International Conference on Image Processing, Chicago, IL*, I, October 4-7 1998.

[108] E P Simoncelli and J Portilla. Texture characterization via joint statistics of wavelet coefficient magnitudes. *in Proc. of IEEE Int. Conf. on Image Processing*, 1998.

[109] J R Smith and C S Li. Image classification and querying using composite region templates. *Journal of Computer Vision and Image Understanding*, 75(1-2):165–174, 1990.

[110] G Strang and T Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, 1996.

[111] V Strela, J Portilla, and E Simoncelli. Image denoising using a local Gaussian scale mixture model in the wavelet domain. In *Proc SPIE, 45th Annual Meeting*, San Diego, 2000.

[112] M Swain and D Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[113] P Tsakalides F Trinci and C L Nikias. Radar CFAR processors in Pearson-distributed heavy-tailed. *IEEE Trans. Aerosp. Electron. Syst.*, March 29 1999. Submitted for publication consideration, pp. 21.

[114] P Tsakalides. *Array Signal Processing with Alpha-Stable Distributions.* PhD thesis, University of Southern California, Los Angeles, California, December 1995.

[115] P Tsakalides and C L Nikias. Broad-band beamforming with alpha-stable distributions. In *Twenty-Ninth Annual Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, October 1995.

[116] P Tsakalides and C L Nikias. The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments. *IEEE Trans. Signal Processing*, 44:1623–1633, July 1996.

[117] P Tsakalides and C L Nikias. Robust space-time adaptive processing (STAP) in non-gaussian clutter environments. *IEE Proceedings, Radar, Sonar, Nav.*, 146:84–94, April 1999.

[118] P Tsakalides, R Raspanti, and C L Nikias. Angle/doppler estimation in heavy-tailed clutter backgrounds. *IEEE Trans. Aerosp. Electron. Syst.*, 35:419–436, April 1999.

[119] P Tsakalides, P Reveliotis, and C L Nikias. Scalar quantization of heavy-tailed signals. *IEE Proc. - Vision, Imag. Sign. Proc.*, 147(5):475–484, Oct. 2000.

[120] G A Tsihrintzis and C L Nikias. Data adaptive algorithms for signal detection in subgaussian impulsive interference. *IEEE Trans. Signal Processing.* submitted on January, 1996, pp. 25.

[121] G A Tsihrintzis and C L Nikias. Robust signal detection and classification based on fractional lower-order statistics. *IEEE Trans. Signal Processing.* submitted on April, 1995, pp. 36.

[122] G A Tsihrintzis and C L Nikias. Advanced statistical signal processing for radar clutter environments. Technical Report USC-SIPI, University of Southern California, Nov. 1994.

[123] G A Tsihrintzis and C L Nikias. Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an alpha-stable process. *IEEE Trans. Comm.*, 43:904–914, Mar. 1995.

[124] G A Tsihrintzis and C L Nikias. Fast estimation of the parameters of alpha-stable impulsive interference. *IEEE Trans. Signal Processing*, June 1996.

[125] M Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. on Image Processing*, 4:1549–1560, November 1995.

[126] N Vasconcelos and A Lippman. A unifying view of image similarity. *Proceedings of 5-th International Conference on Pattern Recognition, Barcelona, Spain*, Sept. 2000.

[127] M Wainwright, O Schwartz, and E Simoncelli. Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons. In R Rao, B Olshausen, and M Lewicki, editors, *Statistical Theories of the Brain.* MIT Press, (in press), 2001.

[128] M J Wainwright and E P Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, *Adv. Neural Information Processing Systems (NIPS*99)*, volume 12, pages 855–861, Cambridge, MA, 2000. MIT Press.

[129] J Z Wang, G Wiederhold, O Firschein, and X W Sha. Content-based image indexing and searching using daubechies' wavelets. *Int. Journal of Digital Libraries*, 1(4):311–328, 1998.

[130] A Weron. Stable processes and measures; A survey. In D Szynal and A Weron, editors, *Probability Theory on Vector Spaces III*, pages 306–364. Springer's LNM, Lublin, 1983.

[131] W Willinger, M S Taqqu, W E Leland, and D V Wilson. Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements. *Stat. Science*, 10:67–85, 1995.

[132] P Wu, B S Manjunath, S Newsam, and H D Shin. A texture descriptor for browsing and similarity retrieval. *Signal Processing: Image Communication*, 16:33–43, 2000.

[133] W R Wu and S C Wei. Rotation and gray-scale transfrom-invariant texture classification using spiral resampling, subband decomposition and hidden markov model. *IEEE Trans. Image Processing*, 5:1423–1434, Oct 1996.