

Contents

Tables	xvii
Figures	xix
Resumo	xxix
1 Introduction	1
1.1 Objectives	4
1.2 The Method	5
1.3 Theoretical Foundations	6
1.4 Applications	8
1.5 Contributions	9
1.6 Thesis Overview	10
2 Multiresolution Image Analysis	13
2.1 Low Level Processing in the HVS	13
2.2 Energy Filtering	15
2.2.1 Phase Congruency	17
2.3 Multiresolution Decomposition	18

2.3.1	Quadrature Filter Models	21
2.4	Frequency Feature Integration	26
2.4.1	Biologically Feasible Techniques	27
2.4.2	Composite Feature Learning	28
2.4.3	Unsupervised Region Classification	29
2.4.4	Unsupervised Feature Clustering	31
2.5	Feature-Tuned Detectors	32
3	Visual Pattern Detection	35
3.1	3D Filter Bank	37
3.1.1	3D Log Gabor Channels	37
3.1.2	Frequency Domain Sampling	40
3.1.3	Filter Bandwidths	42
3.2	Selection of Active Bands	43
3.3	Dissimilarity Measure Between Energy Maps	44
3.3.1	Information Theoretical Concepts	46
3.3.2	Entropy and Mutual Information	47
3.3.3	Normalized Measures of Mutual Information	49
3.4	Feature Clustering	50
3.5	Examples	55
4	Improving Dissimilarity Estimation	67
4.1	Measuring Feature Phase Alignment	68
4.2	Dissimilarity Measures	74
4.2.1	Statistical Measures of Dependence	74
4.2.1.1	Classical Dissimilarity Measures	74
4.2.1.2	Correlation Ratio	75

4.2.1.3	Other f-divergences	76
4.2.1.4	Correlation Coefficient	77
4.2.2	Attention Based Measures	78
4.2.2.1	The RGFF Dissimilarity Measure	78
4.2.2.2	Dissimilarity Estimation on Energy Maxima Maps	79
4.2.2.3	Dissimilarity Estimation on Regions of Interest	80
4.3	Computational Cost	81
4.4	Experimental Results	83
4.5	Discussion	83
5	Applications	97
5.1	Medical Image Segmentation Improvement	98
5.1.1	Active Model Initialization	99
5.1.2	Geodesic Active Models	100
5.1.3	Initialization with Visual Patterns	103
5.1.4	Segmentation Results	106
5.2	Other Applications	108
5.2.1	TMRI Data Enhancement	108
5.2.1.1	Examples	116
5.2.2	GPR Data Enhancement	117
5.2.2.1	Examples	118
6	Conclusions	125
6.1	Final Remarks and Future Work	127
	Notation and Abbreviation	131
	References	133

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
DEPARTMENT OF ELECTRONICS AND COMPUTER SCIENCE



Dissertation for the degree of Doctor of Philosophy in Physics

**DATA DRIVEN SYNTHESIS OF COMPOSITE-FEATURE
DETECTORS FOR 3D IMAGE ANALYSIS**

By:

Raquel Dosil Lago

Supervisors:

Xosé M. Pardo López

Xosé R. Fernández Vidal

Santiago de Compostela, April 2005.

Xosé M. Pardo López

Profesor Titular de Universidade

Área de Linguaxes e Sistemas Informáticos

Dept. de Electrónica e Computación,

Univ. de Santiago de Compostela

Xosé R. Fernández Vidal

Profesor Titular de Universidade

Área de Física Aplicada

Dept. de Física Aplicada

Univ. de Santiago de Compostela

FAN CONSTAR:

Que a memoria titulada "**Data Driven Synthesis of Composite-Feature Detectors for 3D Image Analysis**" foi realizada por **Dona Raquel Dosil Lago** baixo a nosa dirección no Departamento de Electrónica e Computación da Universidade de Santiago de Compostela e constitúe a Tese que presenta para optar ó grao de Doutora en Física.

Santiago de Compostela, Abril de 2005

Asinado: Xosé M. Pardo López

Codirector da Tese

Asinado: Xosé R. Fdez. Vidal

Codirector da Tese

Asinado: Diego Cabello Ferrer

Director do Departamento de

Electronica e Computación

Asinado: Raquel Dosil Lago

Doutoranda

Agradecementos

Nesta primeira liña da miña tese, debo e quero dálle-las gracias ós meus dous directores de tese, Xosé Manuel Pardo López e Xosé Ramón Fernández Vidal, por toda a dedicación e o esforzo que puxeron para que este traballo saíse o mellor posible, pero tamén pola confianza e o apoio ó longo destes anos.

Gracias tamén a Víctor Leborán. Primeiro pola colaboración nas ocasións que tivemos de traballar xuntos e, como non, por toda a axuda prestada en toda clase de cuestións informáticas. Non direi que esta tese non sería posible sen el, pero seguramente me levaría moito máis tempo.

Gracias a David García Pérez, por sentarte aí enfrente e axudarme cada día no que faga falla. David, canto sabes!

Gracias a Víctor Brea, polos consellos sobre os artigos, o LaTeX, a Universidade, a vida ... e polas risas.

Gracias a Diego Cabello e a Antonio Mosquera, polas ocasións que tiveron de colaborar con eles e tamén por tódalas facilidades que me deron para poder dedicarme a investigación neste grupo.

Gracias a Petia Radeva por los consejos y por las imágenes TMRI. Gracias en general toda la gente que he conocido en el Centre de Visió per Computador por las charlas, oficiales o no, de las que tanto he aprendido.

Gracias a Óscar García i Panyella, en primer lugar por las imágenes SPECT. Y en segundo lugar, pero no menos importante para mí, gracias Óscar y Eva, por hacerme

sentir como en casa y ser tan macos!

A Nando, a Pablo e a Julio: foi un pracer traballar convosco. Repetería encantada.

Gracias a Paula López pola plantilla de LaTeX e a axuda *on-line* dende Alemania.

Gracias a Manolo Penedo por preocuparse por como ia indo o meu traballo e asesorarme en tantas cousas.

A tódolos membros do Grupo de Visión Artificial, gracias por compartiren a súa experiencia comigo e máis por unha morea de pequenas cousas que o fan todo un pouco máis fácil.

Gracias a Fernando Vilariño, por facerme tantas preguntas, ou sexa, por facerme pensar. E pola amizade, claro.

Gracias Poal (Jordi González), por la ayuda y los consejos que me has ofrecido y por todas esas conversaciones sobre mil cosas. Gracias por darme ánimos.

Gracias especialmente á miña familia, a Manuela e Ignacio, meus pais, por darme todo. E gracias ós meus irmáns, Ito e Vane. A todos eles, gracias por existir.

Esta tese foi financiada polo proxecto de investigación TIC2000-0399-C02-02 do antigo Ministerio de Ciencia e Tecnoloxía e actual Ministerio de Educación e Ciencia e máis polos proxectos PGIDT99PXI20606B e PGIDIT04TIC206005PR da Xunta de Galicia.

"[...] we have learned that even natural
organs such as the eye need
the imagination to see"

Johann Wolfgang von Goethe,
Zur Farbenlehre (Theory of Colors), 1810

Non me importa repetir-me:
Ós meus pais

Contents

Tables	xvii
Figures	xix
Resumo	xxix
1 Introduction	1
1.1 Objectives	4
1.2 The Method	5
1.3 Theoretical Foundations	6
1.4 Applications	8
1.5 Contributions	9
1.6 Thesis Overview	10
2 Multiresolution Image Analysis	13
2.1 Low Level Processing in the HVS	13
2.2 Energy Filtering	15
2.2.1 Phase Congruency	17
2.3 Multiresolution Decomposition	18

2.3.1	Quadrature Filter Models	21
2.4	Frequency Feature Integration	26
2.4.1	Biologically Feasible Techniques	27
2.4.2	Composite Feature Learning	28
2.4.3	Unsupervised Region Classification	29
2.4.4	Unsupervised Feature Clustering	31
2.5	Feature-Tuned Detectors	32
3	Visual Pattern Detection	35
3.1	3D Filter Bank	37
3.1.1	3D Log Gabor Channels	37
3.1.2	Frequency Domain Sampling	40
3.1.3	Filter Bandwidths	42
3.2	Selection of Active Bands	43
3.3	Dissimilarity Measure Between Energy Maps	44
3.3.1	Information Theoretical Concepts	46
3.3.2	Entropy and Mutual Information	47
3.3.3	Normalized Measures of Mutual Information	49
3.4	Feature Clustering	50
3.5	Examples	55
4	Improving Dissimilarity Estimation	67
4.1	Measuring Feature Phase Alignment	68
4.2	Dissimilarity Measures	74
4.2.1	Statistical Measures of Dependence	74
4.2.1.1	Classical Dissimilarity Measures	74
4.2.1.2	Correlation Ratio	75

4.2.1.3	Other f-divergences	76
4.2.1.4	Correlation Coefficient	77
4.2.2	Attention Based Measures	78
4.2.2.1	The RGFF Dissimilarity Measure	78
4.2.2.2	Dissimilarity Estimation on Energy Maxima Maps	79
4.2.2.3	Dissimilarity Estimation on Regions of Interest	80
4.3	Computational Cost	81
4.4	Experimental Results	83
4.5	Discussion	83
5	Applications	97
5.1	Medical Image Segmentation Improvement	98
5.1.1	Active Model Initialization	99
5.1.2	Geodesic Active Models	100
5.1.3	Initialization with Visual Patterns	103
5.1.4	Segmentation Results	106
5.2	Other Applications	108
5.2.1	TMRI Data Enhancement	108
5.2.1.1	Examples	116
5.2.2	GPR Data Enhancement	117
5.2.2.1	Examples	118
6	Conclusions	125
6.1	Final Remarks and Future Work	127
	Notation and Abbreviation	131
	References	133

Tables

3.1	Values of several dependence measures in the cases of total independence and exact functional dependence	51
4.1	Normalized Mutual Information NI and associated distance values D_{NI} for instances of same-pattern –32, 31, in first row – and different-pattern frequency features. Filter’s responses are shown in 4.2. Their correlation coefficient ρ is also listed as a reference.	73
4.2	Asymptotical computational costs for each of the dissimilarity measures presented in section 4.2. Symbol δ represents all global measures listed in 4.2.1	82

Figures

2.1	Transfer function of the Gabor function. <i>Top</i> : Transfer function of the even component T^e of a Gabor function. The dotted lines represent the two Gaussian functions that contribute to the even part, one for the positive frequencies and the other for negative frequencies. As can be seen, the even part has a non null DC component. <i>Center</i> : Transfer function of the odd component T^o of a Gabor function multiplied by $i = \sqrt{-1}$. The dotted lines represent the two Gaussian functions that contribute to the odd part, one for the positive frequencies and the other for negative frequencies, which cancel out at the origin of frequencies frequency. <i>Bottom</i> : Transfer function of the Gabor function as the sum of the above even and odd components. .	25
2.2	Log Gabor transfer functions for various values of ρ_i and 2 octave bandwidth.	26
3.1	Flowchart of the visual pattern detection process.	36
3.2	Representation of the angular component of the filter in the (ϕ, θ) plane as iso-level curves. <i>Left</i> : $G(\phi, \theta; \phi_i, \theta_i)$. <i>Right</i> : $S(\alpha(\phi, \theta; \phi_i, \theta_i))$ <i>Top</i> : $(\phi_i = \pi/3, \theta_i = \pi/4)$. <i>Bottom</i> : $(\phi_i = 0, \theta_i = \pi/2)$	39
3.3	Representation of the angular component of the filter over a unit radius sphere as iso-level curves. <i>Left</i> : $G(\phi, \theta; \phi_i, \theta_i)$. <i>Right</i> : $S(\alpha(\phi, \theta; \phi_i, \theta_i))$ <i>Top</i> : $(\phi_i = \pi/3, \theta_i = \pi/4)$. <i>Bottom</i> : $(\phi_i = 0, \theta_i = \pi/2)$	39

- 3.4 Spatial domain representation of a 3D log Gabor filter with $\lambda_i = 16$ pixels, $\phi_i = 0$ and $\theta_i = 0$. Each function h is represented by the isosurfaces with levels $\pm 0.05 \cdot \max(h)$, positive in white color and negative in gray color. *Top Left*: Real component. *Top Right*: Imaginary component. *Bottom*: Amplitude. 40
- 3.5 *Right*: Sampling of the orientation space for $N_\phi = 6$. *Left*: Correspondent filter banks for $N_\phi = 6$ and $N_s = 4$, represented as the iso-surfaces of the filters' transfer functions verifying $T_i = \exp(-1/2)$. *Top*: Uniform orientation sampling. *Bottom*: Non-uniform orientation sampling. 42
- 3.6 2D example of radial median filtering using an image of a nematode. (a) Input image of a nematode. (b) $\hat{F} = \log(|\mathcal{F}(X)| + 1)$, where $|\mathcal{F}(X)|$ is the spectral energy. (c) Noise subtraction on $\hat{F}$, non null values depicted in white. (d) Radial median filtering, non null values in white. (e) Bands comprising non null values. (f) Associated active filters. 45
- 3.7 *Top*: Example of a dissimilarity matrix for the case of the nematode subbands in figure 3.1. *Center*: correspondent dendrogram obtained using Ward's method. Cutting the dendrogram at the level indicated with the dashed line produces three clusters. *Bottom*: V_{DB}^{MTS} for each partition on the previous hierarchy. 58
- 3.8 Nematode subband clusters resulting from applying Ward's method for hierarchical clustering and MST-based David-Boulding index for partition validation. The sum of subbands in each cluster gives place to the three visual patterns at the bottom of flowchart in figure 3.1 . 59
- 3.9 Synthetic test data set D1 represented by three cross sections showing slices normal to the image coordinate axis –the origin of coordinates is set in the image center–. It consists of two grating patterns, one with orientation $(\phi = 0, \theta = \pi/2)$ and the other with orientation $(\phi = \pi, \theta = \pi/3)$ 60

- 3.10 *Left column*: Filters from a bank with uniform angular frequency components sampling and log Gabor filters from equation 3.1, represented by isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$. *Right column*: Their response to data set D1. *Top*: Filter with $(\lambda_i = 4, \phi_i = 0, \theta_i = \pi/2)$. *Bottom*: Filter with $(\theta_i = 4, \phi_i = \pi, \theta_i = \pi/2)$ 60
- 3.11 *Left*: Filters with $(\lambda_i = 4, \phi_i = 0, \theta_i = \pi/2)$ and $(\theta_i = 4, \phi_i = \pi, \theta_i = \pi/2)$ from a bank with non-uniform angular components sampling and log Gabor filters from equation 3.1 –they are identical–, represented by isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$. *Right*: Their response to data set D1. 61
- 3.12 Volumetric data set D2 represented by orthogonal cross sections. . . . 61
- 3.13 Results obtained for image D2. *Left column*: Filters correspondent to each of the three clusters obtained represented by isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$. *Right column*: Cross sections of the volumes reconstructed from each of the clusters. Each volume represents the sum of the energies of all filters belonging to the correspondent cluster 62
- 3.14 *Top*: Image D3, represented by the slices normal to the image coordinate axis. *Remainder Rows*: Result of frequency feature clustering. *Left Column*: Filter clusters obtained, representing filter as the isosurfaces of level $\exp(-1/2)$ of their transfer functions. *Right Column*: Cross sections of their correspondent reconstructed visual patterns. . . 63
- 3.15 *Top Row*: Image D4, obtained from D3 by adding Gaussian noise of standard deviation $\sigma = 25$, represented by slices normal to the image coordinate axis. *Remainder Rows*: Visual patterns obtained for D4. Filter clusters are identical to those shown in 3.14. 64
- 3.16 Volumetric data set D5 represented by three orthogonal cross sections. *Left*: Shape of the inner region of the image. *Right*: Cross sections of the volume 65

- 3.17 Results obtained for image D5. *Left column*: Filters correspondent to each of the two clusters obtained, represented by isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$. *Right Column*: For each cluster, cross sections of the volumes reconstructed as the sum of energies of all filters in the cluster. 65
- 4.1 Example of clustering failure using normalized mutual information. Filters are represented as the iso-contours of level $\exp(-1/2)$ of their transfer functions. (a) 2D example image consisting of two orthogonal grating patterns and one line. (b) Filters corresponding to cluster 1. (c) Linear summation of the energies of filters in cluster 1. (d) Filters corresponding to cluster 2. (e) Linear summation of the energies of filters in cluster 2. 69
- 4.2 Energy of the responses of filters tuned to the orientations of different visual patterns in the example image from figure 4.1. (a) Filter 1: $(\phi = 0, \lambda = 4)$. (b) Filter 32: $(\phi = 2\pi/3, \lambda = 16)$. (c) Filter 31: $(\phi = 2\pi/3, \lambda = 4)$. (d) Filter 13: $(\phi = \pi/3, \lambda = 4)$ 70
- 4.3 Plots of pairs of energy values for each pixel location corresponding to filters from figure 4.2: (a) $\{1, 31\}$ (b) $\{32, 31\}$ (c) $\{13, 31\}$ 74
- 4.4 Percentage of correct (OK), incorrect (X) and indecisive (?) results for each distance. 84
- 4.5 Images used in the test. The first two rows correspond to 2D images and the remainder to 3D images, represented as the three cross-section of the volume passing through the image center. 85
- 4.6 *Top Row*: Cross-sections of a 3D synthetic image of two grating patterns differing in a phase shift. *Remainder Rows*: Result of frequency feature clustering using dissimilarity D_{NI} . Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Left Column*: The two clusters obtained. *Right Column*: Cross sections of their correspondent reconstructed visual patterns. 87

- 4.7 Result of frequency feature clustering of the image from 4.6, 1st row using dissimilarity D_ρ . *Left Column*: Resulting clusters. Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Right Row*: Cross sections of their correspondent reconstructed visual patterns. 88
- 4.8 *Top Row*: 2D image of a Brodatz texture. *Center Row*: Result of frequency feature clustering of the image above using dissimilarity D_{NI} . Filters are represented as the iso-contours of level $\exp(-1/2)$ of their transfer functions. *Bottom Row*: Reconstruction of the visual patterns associated to clusters in the 1st row. 89
- 4.9 *Top Row*: Result of frequency feature clustering of image in top of figure 4.8 using dissimilarity D_ρ . Filters are represented as the iso-contours of level $\exp(-1/2)$ of their transfer functions. *Bottom Row*: Reconstruction of the visual patterns associated to clusters in the 1st row. 90
- 4.10 *Top Row*: Result of frequency feature clustering of image in top of figure 4.8, using dissimilarity D_η . Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Bottom Row*: Reconstruction of the visual patterns associated to clusters in the 1st row. 91
- 4.11 (a) Result of frequency feature clustering of image in top of figure 4.8, using dissimilarity D_{RGFF} . Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. (b) Reconstruction of the visual patterns associated to clusters in the left column. 92
- 4.12 (a) Cross sections of a 3D image synthesized from three grating patterns with different orientations. (b) Energy map of the response to filter with $(\lambda = 4, \phi = 0, \theta = 0)$. (c) Image in figure b after non-maxima suppression. (d) Binary image obtained for image c by setting all maxima to white. (e) Image c after thresholding using a threshold selection that eliminates a fixed fraction of maxima depending of the scale of the filter. (f) Image c after thresholding using a threshold selection that eliminates a fixed fraction of the total energy of the image depending of the scale of the filter. 93

- 4.13 Visual pattern partitioning of image in figure 4.12.a using dissimilarity D_{NI} . *Left Column*: The two clusters obtained. Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Right Column*: Cross sections of their correspondent reconstructed visual patterns. 94
- 4.14 Visual pattern partitioning of image in figure 4.12.a using dissimilarity D_{NI}^* and threshold strategy as in figure 4.12.e. Other threshold strategies produce the fusion of filters associated to different gratings. *Left Column*: The three clusters obtained. Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Right Column*: Cross sections of their correspondent reconstructed visual patterns. 95
- 4.15 Influence of orientation overlapping parameter $\Delta\phi/N$. *Top Row*: Cross sections of the original 3D synthetic data. *Remainder Rows*: Result of visual pattern partitioning using dissimilarity measure D_ρ and $\Delta\phi/N = 1.2$. *Left Column*: The two clusters obtained. Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Right Column*: Cross sections of their correspondent reconstructed visual patterns. The reduction of orientation overlapping, using for example $\Delta\phi/N = 1.5$, produces the fragmentation of the sphere boundary in the first visual pattern (in second row). . . . 96
- 5.1 Scheme of the process of initialization of active models using matching with a priori global shape models. 101
- 5.2 Scheme of the process of initialization of geodesic deformable models using the presented visual pattern partitioning method. The visual pattern selected as initial model is selected manually. 105
- 5.3 (a) Cross sections of a 3D data set corresponding to a SPECT image of the left ventricle of the heart. (b, c) Clusters identified by the system. (d, e) Cross sections of the visual patterns correspondent to clusters in figures b and c respectively, reconstructed by linear summation of the half-wave rectified real part of the filters' responses. 109

5.4	Thresholding of image from figure 5.3.a. (a) Cross sections of original data with the intersections of isosurface of gray level 100 in white color. There is a hole in the ventricle wall. (b) 3D reconstruction of the surface in figure a. (c) Cross sections of original data with the intersections of isosurface of gray level 90 in white color. New artifacts appear in relation to previous thresholding. (d) 3D reconstruction of the surface in figure c.	110
5.5	Segmentation of image from figure 5.3.a. (a) Cross sections of original data with the intersections of the left ventricle segmentation at several time steps superimposed in white color, obtained using a square box as initial geodesic model. (b) 3D reconstruction of the surface correspondent to the segmentation in figure a with 60 iterations. (c) Cross sections of original data with the intersections of the left ventricle segmentation in white color, obtained using the visual patten in from image 5.3.e as initial geodesic model. (d) 3D reconstruction of the surface in figure c.	111
5.6	(a) Sample horizontal slices of a phantom T2-weighted MR image of the brain ventricles and the corpus callosum. (b) Sagittal section of the data set. (c, d) The two clusters generated by the method. (e, f) Reconstruction of the visual patterns correspondent to clusters in figures c and d respectively –image f is inverted– by linear summation of the filter’s responses real component of each cluster.	112
5.7	Thresholding of image from figure 5.6.a. (a) Thresholding using level 200, correspondent to the corpus callosum. Larger levels are set to 255 –white. (b) 3D reconstruction of the isosurface of level 200. (c) Thresholding using level 240, correspondent to the brain ventricles. Smaller levels are set to 0 –black. (d) 3D reconstruction of the isosurface of level 240.	113

- 5.8 Segmentation of 3D data set from figure 5.6 using different definitions of the initial model. *Top Row*: Using a square box. *Center Row*: Using visual patterns in 5.6.d. *Bottom Row*: Using visual pattern in 5.6.e. *Left Column*: Cross sections of the segmented data. The region outer to the final geodesic model has been set to white in the first and third rows and to black in the second to enhance visualization. The inner region presents the original data. *Right Column*: 3D reconstruction of the zero level sets. 114
- 5.9 Segmentation of a 3D real MR image of the brain ventricles and the corpus callosum. (a) Sample slices of the original data. (b, c) Slices of the two 3D visual patterns generated by the method, obtained by linear summation of the filter's responses real component of the selected cluster –image b is inverted. (d) Sample slices of the ventricles segmented using the visual pattern in image c as initial model. The outer region of the object is represented in white. (e) 3D reconstruction of the surface from image d. In the 3D reconstruction, the aspect ratio is 1:1:0.5 to provide a more realistic visualization, since the resolution in the z axis is lower. 115
- 5.10 Sample slices of the result of segmentation of the data set from 5.9.a using a square box as initialization. The outer region of the object has been represented in white. 116
- 5.11 (a) Example of $2D+t$ TMRI image of the left ventricle of the heart. (b, c) Clusters generated by the composite feature extraction method. (d, e) Reconstructed visual patterns. (f) Isosurface of the first visual pattern P at level $0.2 \cdot \max(P)$. (g) Phase of the tag pattern. White represents π radians and black $-\pi$ 120
- 5.12 (a) Example of $2D+t$ TMRI image of the left ventricle of the heart (b) Cluster generated correspondent to the tag pattern. (c) Reconstructed tag pattern. (d) Isosurface of the first tag pattern P at level $0.15 \cdot \max(P)$ (d) Phase of the tag pattern. White represents π radians and black $-\pi$ 121

- 5.13 a) GPR image of a soil with a buried LI-11 AP mine. (b, c) Composite feature detectors extracted. (d, e) Their respective detected visual patterns. (f) Isosurface 3D reconstruction of visual pattern in figure e. The ellipse is enclosing the reflection echoes of the AP mine. The arrow signals the apex of the reflection profile, where land mine is located. (g) Contour plot of a section of the thresholded visual pattern in figure e, superimposed to the original data. 122
- 5.14 (a) GPR image of a soil with a buried PFM-1 AP mine. (b, c, d, e) Composite feature detectors extracted. (f, g, h, i) Their respective detected visual patterns. (j) Isosurface 3D reconstruction of visual pattern in figure h. (k) Isosurface 3D reconstruction of visual pattern in figure k. 123
- 5.15 (a, b) Two views of a GPR image of an archaeological site, at different depths –origin of slice indexes is not at the center of the image here, but at origin of matrix indices. (c, d) Composite feature detectors extracted. (e, f) Two views of the first visual patterns. (g, h) Two views of the second visual pattern. 124

Resumo

O uso de modelos de representación de baixo nivel é de grande importancia en tarefas de análise de imaxes. O contido dunha imaxe pode ser representado a diferentes niveis de detalle. No nivel máis alto tense unha representación puramente conceptual da imaxe coma un todo, correspondente á interpretación da escena conforme a un contexto semántico dado. No nivel de detalle máis baixo están os datos de partida, unha colección de valores de intensidade medidos nun conxunto discreto de posicións no espazo. Entrambos, pódese xerar toda unha serie de representacións intermedias, como propiedades locais, puntos característicos, partes de obxectos ou obxectos. Canto máis baixo é o nivel de abstracción, máis preto está a representación dos datos e máis local e pormenorizada é. Canto máis alto é o nivel de abstracción, máis preto estará a descrición dos conceptos e máis global e estrutural a descrición.

As descricións de alto nivel dos datos obtéñense a partires das de baixo nivel mediante un proceso de modelaxe. Debe haber algunha idealización a priori dos obxectos que se quere identificar e describir. Un modelo é unha representación dos datos que establece unha serie de restricións sobre eles. Estas restricións poden ser inherentes ó propio modelo e reflecti-las súas propiedades ideais. Pensemos por exemplo nun modelo de *snake*, que permite unha certa flexibilidade na forma do obxecto pero impondo as condicións de continuidade e suavidade. Outra forma de restrinxi-la representación dun obxecto é introducindo información a priori sobre as súas propiedades. Un exemplo desta clase de modelado é a técnica de rexistro con modelos a priori, onde a forma do obxecto vén completamente predeterminada polo modelo e os únicos graos de liberdade son a posición e orientación na imaxe o, como moito, lixeiras deformacións locais.

Para obter unha representación de alto nivel do contido unha imaxe a partires

dun conxunto de datos de intensidade, moitas técnicas de análise de imaxes baséanse en representación de baixo nivel dos datos. O modelado dos datos pode levarse a cabo usando estratexias tanto *de abaixo cara arriba* (*bottom-up*) como *de arriba cara abaixo* (*top-down*). Nos procesos *bottom-up* ou baseados en datos, as descrições de alto nivel obtéñense buscando estrutura nas propiedades de baixo nivel e organizándoas en entidades máis complexa, como en (Ballard and Brown, 1982) por mencionar un exemplo. Neste caso, as suposicións que se adoitan para o modelado están implícitas nas regras que gobernan o proceso de organización. Aquí, a calidade da representación de baixo nivel é decisiva para o correcto funcionamento das seguintes etapas. Por outra banda, as técnicas *top-down* ou baseadas en modelos, parten de representacións de alto nivel dos datos e tentan axustalos á información de baixo nivel extraída da imaxe. Dependendo da flexibilidade do modelo será máis ou menos influenciable pola representación de baixo nivel.

As diversas técnicas de análise de imaxes usan diferentes fontes de información de baixo nivel, como contraste, simetría, curvatura, textura, orientación, fase, contido en frecuencia, etc. Moitas clases de detectores de baixo nivel foron desenvolto baseándose na análise destas propiedades locais. Entre eles, os máis sobranceiros son posiblemente contornos, rexións, esquinas, conxuncións e manchas (*blobs*). Habitualmente, a axeitada selección e detección das características a empregar como bloques constituíntes das representacións de alto nivel é crítica para o correcto funcionamento de todo o proceso. Parece amplamente aceptado que non hai unha única característica que dea lugar a unha representación completa da imaxe, senón que se han de empregar varias conxuntamente.

Certos patróns de baixo nivel pódense considerar como unha composición de características máis elementais. Pode resultar contraditorio falar de características compostas nunha etapa de baixo nivel. A expresión *patrón composto* non se refire aquí a unha representación máis abstracta, baseada en partes, obxectos, etc. Simplemente quere dicir que unha patrón estatístico presente nunha imaxe pódese definir como unha conxunción doutros patrons con estrutura estatística máis simple. Neste caso, o problema da representación pode ser abordado en dúas etapas, nas que primeiro se detectan as características elementais e logo se reorganizan para formaren as características compostas. Por exemplo, pódese pensar nunha esquina coma nun punto onde confluen dous bordes de tipo chanzo. Os bordes pódense caracterizar mediante unha función base coa escala e a orientación axeitadas e simetría impar. Un exemplo típico desto é a descomposición subbanda nun marco multirres-

olución. As funcións base en dita representación son usualmente patróns sintonizados a diferentes escalas, orientacións e fases. A descomposición multiresolución é unha ferramenta potente e xenérica, adecuada para unha grande variedade de aplicacións. De feito, esta é a clase de procesamento que ten lugar na visión biolóxica (Hubel, 1988). Non obstante, existe o problema de como combinar a información das diversas vistas para obter os patróns compostos.

Cando un patrón ten respostas a máis dun dos detectores multiresolución, unha estratexia común é seleccionar unha das vistas, tipicamente a que ten a maior resposta, como representante dese patrón, rexeitando as demais, como se fai en (Lindeberg, 1994). Deste xeito, o que se ten é unha representación parcial do patrón. No exemplo da esquina, esta solución implicaría quedarnos só cun dos dous bordes. Outras aproximacións realizan simplemente unha suma linear de tódalas características elementais, producindo un único *esquema principal* (*primal sketch*) que comprende tódolos patróns da imaxe (Morrone and Burr, 1988; Kovesi, 1996). Neste caso, o problema é cómo illar uns patróns compostos dos outros. Para xerar unha representación completa, cada patrón composto debería ser reconstruído por separado integrando unicamente a información das características elementais que o compoñen.

Tamén é común integrar a información das diferentes vistas buscando un conxunto de patróns predefinidos. Por exemplo, pódese considerar un conxunto de patróns, tanto elementais como compostos, que comprenda liñas, contornos e cruzamentos (*junctions*), como en (Rodrigues and du Buf, 2004). Entón, desenvólvese unha técnica para detectar cada un deles por separado, aplicando diferentes regras de acordo coas súas respectivas propiedades. Isto pode ser moi eficiente para aplicacións en dominios específicos, nas que unha cantidade reducida de patróns abonde. Pensemos, por exemplo, na análise e clasificación de sinais de tráfico. Outro exemplo típico, é reduci-lo conxunto de patróns a rexións de textura e ós seus contornos (Malik and Perona, 1990). Non obstante, para unha aplicación de propósito específico, tal representación sería limitada de máis.

En aplicacións de propósito xeral, a selección de características e o método de extracción en sí mesmo deberían ser conducidos por datos, del tal xeito que a representación se axustase ás particularidades da imaxe. A maior parte das solucións para este problema usan técnicas de aprendizaxe, extraendo coñecemento dunha base de datos, que despois pode ser aplicada á análise de novas imaxes (Wersing and Körner, 2003). Atópanse poucas investigacións sobre o desenvolvemento de sistemas para determinar as características máis representativas nunha imaxe e proporciona-

los correspondientes detectores usando só os datos da imaxe, sen usar información a priori.

Obxectivos

O propósito deste traballo é proporcionar un modelo de representación de datos volumétricos capaz de identificar e describir diferentes tipos de características, tanto simples como compostas, presentes nunha imaxe. Consideraremos só o caso de imaxes dixitais volumétricas en niveis de gris, ignorando información de cor, movemento ou estéreo. As características compostas de baixo nivel, que a partires de agora chamaremos *patrons visuais*, defínense aquí como as que poden ser descritas de forma apropiada por un ou máis características elementais. O conxunto de patróns visuais non debe estar predefinido, senón que se debe determinar para cada imaxe baseándonos só nas súas propiedades e sen usar coñecemento a priori. O método para a extracción da representación tampouco debe usar información a priori e debe ser xeral para calquera conxunto de características, sen introducir procedementos ad hoc nin propios dun certo dominio de aplicación. Para isto, faremos uso de diversas achegas en vision biolóxica.

Propoñemos o uso dunha estratexia de descomposición-integración. Isto implica, en primerio lugar, o desenvolvemento dun modelo para a descomposición dunha imaxe 3D nun conxunto de características elementais, que debe producir unha representación exhaustiva da imaxe. Con esta fin, adoitaremos un esquema multiresolución, no que as características elementais, que chamaremos características de frecuencia, están sintonizadas a diferentes frecuencias e orientacións. En segundo lugar, teremos que desenvolver un procedemento de integración para combina-las características elementais de modo que se xere unha vista completa de cada un dos patróns visuais específicos da imaxe. Esta etapa planeárase como un problema de agrupamento (*clustering*) non supervisado, no que os clusters determinan as características compostas.

Para acadar o agrupamento o resultado correcto, hase de definir unha medida de disparidade entre pares de características de frecuencia que capture as agrupacións correspondentes a patróns visuais relevantes. Polo tanto, dita distancia debe reflectir o criterio para que un patrón visual se considere relevante. Aquí, o criterio para ligar características está baseado no concepto de Congruencia de Fase (CF), i.e., aliña-

mento local da fase das compoñentes de frecuencia nun amplo rango de escalas. Este criterio foi proposto en (Morrone and Burr, 1988) como o que usa a visión biolóxica para integrar características elementais. Ademais do desenvolvemento do modelo de representación, buscamos determinar que relacións existen entre características baixo a suposición de CF. Para isto, realizaremos un estudo sobre un conxunto de medidas estatísticas de disparidade entre características. Deste estudo esperamos extraer conclusións sobre ditas relacións e atopar unha medida adecuada para o algoritmo de agrupamento.

Preténdese que o modelo desenvolto sirva como un modelo de representación de baixo nivel para ser aplicado cando non haxa unha aplicación específica ou como alternativa cando outras solucións non produzan resultados satisfactorios. Aínda así, queremos ilustra-las capacidades do modelo con aplicacións reais. Por ese motivo, centramos a atención na aplicación de mellora de modelos deformables para a segmentación de imaxes médicas. En particular, deseñárase unha técnica para a inicialización de modelos deformables mediante o uso do noso método de extracción de patróns visuais. Os patróns visuais illados na imaxe serán usados para definiren o estado inicial dun modelo de xeodésicos activos e así mellora-la converxencia ó obxecto anatómico de interese nun conxunto de datos volumétricos médicos.

O Método

O esquema da técnica desenvolta para chegar ós obxectivos planeados, xa foi adiantada na sección anterior. Implica o agrupamento dun conxunto de compoñentes de frecuencia –ás que a miúdo nos referiremos como características de frecuencia– de acordo coa súa habilidade para discriminar un certo patrón visual nunha imaxe. Deste xeito, o método xera un conxunto de detectores de características compostas adaptados ós patróns visuais específicos da imaxe baixo análise. O proceso pode ser descrito pola seguinte secuencia de pasos.

1. Selección dun subconxunto de canles de frecuencia no banco 3D con alto contido en información, chamadas canles *activas*, mediante a análise da amplitude do espectro da imaxe.
2. Detección das características de frecuencia usando os filtros correspondentes a canles activas

3. Cálculo das disparidades entre pares de características de frecuencia baseado na Información Mutua Normalizada (IMN) das súas enerxías locais.
4. Agrupamento das características de frecuencia a partires da matriz de disparidades.
5. Reconstrución dos patróns visuais asociados a cada grupo ou *cluster* sumando as correspondentes características de frecuencia.

A extracción de características de frecuencia é levada a cabo por medio dun banco de filtros log Gabor 3D, deseñados con simetría de rotación en torno a dirección dos fitros. Os parámetros dos filtros e a mostraxe do dominio de Fourier escóllense para producir un recubrimento exhaustivo e reduntante do espectro. O agrupamento faise usando un algoritmo xerárquico, que precisa só da matriz de disparidades para producir unha partición dos datos. A métrica usada reflicte o grao de aliñamento dos máximos de enerxía de pares de características e, polo tanto, o seu grao de congruencia de fase. Dita distancia é unha medida da dependencia estatística entre as enerxías das bandas activas.

Fundamentos Teóricos

Aínda que o traballo presentado nesta tese no pretende simular un proceso biolóxico de visión, si está inspirado en varias teorías relacionadas co comportamento do Sistema Visual Humano. Algunha das ideas extraídas destas teorías empréganse aquí para establece-las hipóteses de traballo nas que se basea esta investigación. O que segue é unha lista coas tres ideas chave tomadas como fundamento do noso esquema de representación.

Códificación Eficiente

A teoría do *Codificación Eficiente* sostén que o SVH está adaptado ó medio natural, de xeito que aproveita as invariancias das súas propiedades estatísticas para realizar unha codificación óptima (Field, 1987, 1993, 1994; Olshausen and Field, 1997; Turiel and Parga, 2000; Simoncelli et al., 1992; Simoncelli, 2003). O seu comportamento pódese modelar cunha transformada *wavelet* ou un banco de filtros (Hubel and Wiesel, 1968; Hubel, 1988; Marčelja, 1980;

Watson, 1987b, 2000; Graham, 1989; Field, 1993), que se axustan ós datos de experimentos psicofísicos e produce unha codificación compacta, aínda que redundante, das imaxes naturais. Os filtros do banco modelan o procesamento realizado no SVH por grupos de neuronas nas áreas corticais do cerebro, sensibles a frecuencia, orientación, movemento, etc. En (Field, 1987) propónse a función 2D log Gabor como modelo plausible para ditos sensores. Baseándonos nestas teorías, desenvolveremos un esquema de codificación 3D para realiza-la descomposición subbanda.

Integración Preatencional de Características

De acordo coa teoría da atención de integración de características (Treisman and Gelade, 1980), o SVH distingue entre dous tipos de características compostas: separables e integrais. As características separables, a diferenza das integrais, están formadas pola concurrencia de características máis simples asociadas a dimensións separables. Só poden ser integradas coa mediación de atención focal en serie, que precisa de información contextual, mentres que as características integrais détectanse automaticamente en etapas temperáns e en paralelo en todo o campo visual. Non se precisa de información a priori para detecta-las características integrais, xa que a interpretación contextual da imaxe faise en etapas de análise da imaxe superiores. Dito doutra forma, é posible extraer certas clases de características compostas en etapas de baixo nivel. Seguindo esta teoría, realizaremos a integración de características a baixo nivel asumindo as mínimas suposicións sobre os estícticos das escenas naturais.

Relevancia da Congruencia de Fase

O SVH percibe as características integrais en posicións onde se dá unha alta CF (Morrone and Owens, 1987; Morrone and Burr, 1988), i.e., posicións onde a fase local dun amplo rango de compoñentes de Fourier da imaxe toma o mesmo valor ou, o que é o mesmo, puntos onde a enerxía do estímulo visual é localmente máxima. Esta teoría sostén que o SVH realiza a detección de puntos con alta CF por medio de sensores de contraste organizados en parellas coa mesma orientación e escala pero en cadratura, isto é, cunha diferenza de fase de 90° , actuando coma estimadores da enerxía local. Esta hipótese concorda con estudos recentes que sosteñen que a atención inconsciente diríxese a puntos

da imaxe con alto contraste (Reinagel and Zador, 1999). Nese estudo, amósase que o alto contraste vai asociado á baixa correlación local en escenas naturais. Baixo a luz destas evidencias, o obxectivo da atención sería o de dirixi-lo centro da mirada, que ten maior resolución espacial que a visión parafoveal, ás rexións máis *informativas* da imaxe. Aquí, aplicaremos estas teorías ó deseño do criterio de integración de características. As nosas medidas de disimilaridade reflectirán o grao de aliñamento das respostas dos sensores de contraste, asignando así relevancia dos patróns visuais ó aliñamento en contraste

Baseándonos nestas tres ideas, é posible obter as características integrais dunha imaxe simulando a codificación sensorial do SVH e usando a CF como enlace entre características simples e compostas. Non é precisa información a priori nesta etapa, xa que a interpretación contextual da imaxe se realiza en etapas de alto nivel. Só a CF é precisa para guiar a integración de características.

Aplicacións

O método proposto aquí é potencialmente proveitoso para un amplo conxunto de aplicacións de análise de imaxe 3D, como análise de imaxe médica, análise de datos xeofísicos e inspección visual de materiais non invasiva e podería xeralizarse para a análise de patróns de movemento en secuencias de vídeo. No caso particular das imaxes médicas, unha representación de baixo nivel axeitada pode por si mesma ser de axuda ó especialista para diagnose e tratamento. Máis a miúdo, as representacións de baixo nivel son a base doutras aplicacións, como segmentación, rexistro multimodal, visualización ou morfometría anatómica.

Entre as técnicas de segmentación, os métodos de modelos activos (Caselles et al., 1997; Paragios and Deriche, 1998) destacan pola súa habilidade par integrar información de diferentes características. Non obstante, aínda hai moitas liñas de investigación abertas na procura de definicións máis apropiadas para as enerxías externas e para o modelo inicial, de maneira que o modelo optimizado se axuste o máis posible ó obxecto. Estes métodos derivan os potenciais de imaxe de propiedades locais. Como consecuencia, o estado inicial do modelo necesita estar moi preto da superficie do obxecto de interese para que a segmentación teña éxito. O illamento duns patróns visuais dos outros permite enfoca-la segmentación nun obxecto dado, evitando interferencias doutros patróns. Concretamente, aquí deseñamos unha aplicación para

a inicialización de modelos xeodéxicos activos, mellorando a converxencia cara ó obxecto de interese.

Por outra banda, outras aplicacións baseadas na extracción de características de baixo nivel podería beneficiarse desta técnica. Por completitude, aplicaremos o noso modelo a outras tarefas de análise de imaxes 3D. Concretamente, o modelo será aplicado ó realce de datos $2D+t$ de secuencias Resonancia Magnética Marcada (Tagged Magnetic Resonance, TMG), para o estudo de tensións no músculo cardíaco, e de datos 3D de Radar de Penetración de Solos (Ground Penetrating Radar, GPR) para a detección de obxectos enterrados, como minas antipersona ou xacementos arqueolóxicos.

Contribucións

O traballo presentado nesta tese é un intento de cubri-lo espazo entre os múltiples detectores de características simples para datos volumetricos 3D e as técnicas de procesamento de alto nivel. O método desenvolto ofrece tanto un modelo de representación como unha técnica para a detección de patróns visuais. Os patróns visuais, tal como se definen aquí, ofrecen unha descrición máis completa das características dunha imaxe que as producidas por detectores de características simples. Esta aproximación non se limita a unha descomposición multiresolución, senón que vai un paso máis aló e as múltiples vistas da imaxe son reorganizadas agrupando características correlacionadas xuntas. Todo o proceso está baseado en suposicións mínimas tomadas de teorías sobre o SVH. O método é totalmente conducido por datos; non se emprega información a priori.

Aínda que xa existen investigacións feitas nesta dirección, como por exemplo en (Bovik et al., 1990; Jain et al., 1997; Acharyya et al., 2003) e especialmente en (Rodríguez Sánchez et al., 2000; Chamorro Martínez et al., 2003; Manjunath and Chellappa, 1993), non temos coñemento de que ningunha aproximación deste tipo desenvolvida para datos volumetricos, o cal implica estende-las ideas tomadas das teorías do SHV de 2D a 3D. O SVH non manexa información 3D pura, senón proxeccións estéreo das escenas 3D. A xeralización a 3D non é trivial, posto que implica a redefinición da codificación sensorial tendo en conta unha nova dimensión pero mantendo un bo recubrimento espectral e sensibilidade á orientación e engadindo restricións, como a simetría de rotación en torno á dirección do filtro. O banco de

filtros 3D deseñado neste traballo ten en conta todos estes aspectos.

Outra contribución deste traballo é o estudo de diferentes medidas para a estimación da CF entre características de frecuencia. Algunhas das medidas estudadas son amplamente usadas para comparar imaxes en técnicas de rexistro ou en recuperación de información. O seu comportamento neses campos xa foi estudado, mais a clase de imaxes implicadas nesta aplicación, i.e., un conxunto de versións subbanda dunha imaxe, ten unha serie de particularidades que deben se consideradas. Ademais destas medidas, derívase delas un novo conxunto de distancias nun intento de simular un proceso de atención. A adecuación das distancias téstase analizando o axuste entre os patróns visuais esperados e os producidos polo modelo con cada unha delas. Esta experimentación serve tamén para extraer conclusións sobre a relación estatística entre imaxes subbanda.

A aplicabilidade do detector de características compostas desenvolvido amósase incorporándoo a un método de segmentación de imaxes médicas 3D, concretamente, cunha nova técnica semiautomática de inicialización de modelos activos xeodésicos. Nesta solución non se necesita defini-lo modelo a priori manualmente nin usar información a priori, senón que se identifica con algún dos patróns visuais extraídos da imaxe. Un ou máis dos patróns visuais resultantes son seleccionados como unha descrición aproximada do obxectivo da segmentación. O uso desta técnica mellora os resultados e acelera a converxencia da segmentación.

Organización da Tese

Esta tese está organizada en 6 capítulos. No segundo capítulo preséntanse as bases teóricas do método. En primeiro lugar, introducimos unha revisión dalgúns aspectos sobre o procesamento de información visual no SVH. O SVH realiza unha descomposición multiresolución da información visual en etapas temperás de procesamento. Á dita descomposición supónselle que está optimizada para reflecti-lo carácter disperso dos estímulos visuais naturais e pode ser modelada por filtros de enerxía. Etapas posteriores de procesamento son responsables da integración de características de baixo nivel para producir representacións de máis alto nivel. Na sección 2.4 repásanse algunhas aproximacións á integración de características. Na subsección 2.5 repasamos ó obxectivos do noso traballo en relación con outros traballos.

A técnica proposta detállase no capítulo 3. Na sección 3.1 descríbese a descomposición multiresolución empregada. O procesamento 3D de datos volumétricos require ter en conta unha segunda dimensión de orientación para representa-las canles de frecuencia e impón o uso dun esquema de mostraxe non uniformes nas coordenadas angulares. As nosas solucións a estas cuestións preséntanse nesta sección. Na sección 3.2 propoñemos unha técnica para selecciona-las características máis relevantes da imaxe, reducindo a carga computacional e mellorando a eficacia. Na sección 3.3 presentaremos unha primeira aproximación á estimación da disparidade entre características de frecuencia. A CF impón unha dependencia estatística entre as distribucións dos máximos de enerxía de dúas imaxes subbanda. Por iso, unha medida baseada no concepto de Información Mutua da teoría da información é probable que produza bos resultados. O algoritmo de agrupamento aplicado á matriz de distancias descríbese na seccións 3.4. A sección 3.5 presenta algúns exemplos do comportamento do método con imaxe sintéticas.

O uso da Información Mutua para reflecti-la dependencia entre características de frecuencia é só unha entre unha longa lista de medidas de dependencia estatística. No capítulo 4 propoñemos un abano de medidas alternativas para valida-los resultados obtidos coa Información Mutua. Compárase a eficiencia das distintas medidas conforme ó seu custe computacional na sección 4.3 e en función da súa efectividade cun banco de imaxe de proba na sección 4.4. Os resultados experimentais discútnense na sección 4.5.

No capítulo 5 preséntanse varias aplicacións do método. A primeira sección está centrada na aplicación do modelo de representación na mellora da segmentación con modelos activos. Alí, presentaremos a estratexia de inicialización de modelos xeodésicos activos. Na sección 5.2 propóñense outras aplicacións, basicamente técnicas de realce de datos 3D de diferentes modalidades de imaxe. Neste caso o realce resulta de eliminar patróns visuais que non son de interese.

No último capítulo preséntanse as conclusións xerais derivadas do traballo desenvolvido. Baseándonos nesas conclusións, proporemos novas liñas de continuidade desta investigación para perfeccionármolo traballo realizado e cubrir novas aplicacións.

CHAPTER 1

Introduction

The employment of a complete and efficient representation of an image at a low level of abstraction is of great importance in image analysis tasks. Image contents can be represented at different levels of detail. At the top level we have a fully conceptual representation of the image as a whole, related to the interpretation of the scene according to a given semantic context. At the bottom we have the raw data, a collection of intensity values measured on sample spatial positions. In between, one can produce a series of intermediate representations, like local properties, features, parts and objects. The lowest the level of representation, the closer to the data and the more local and detailed the description. The highest the level of abstraction, the closer to the concepts and the more global and structural the description.

High level descriptions of the data are obtained from low level representations through a process of modelling. There must be some a priori idealization of the objects one wants to identify and describe. A model is a representation of the data which establish a series of constraints on them. These constraints may be inherent to the model itself and reflect its ideal properties. Let's think of a *snake* model, which permits a certain flexibility on the object's shape, but imposes continuity and a certain degree of smoothness to its boundary. Another way of constraining the representation of an object is by introducing prior knowledge about its properties. An example of this kind of modelling is the model matching technique, where the shape of the object is fully predetermined and the only degrees of freedom are position and orientation in the image or, at most, slight local deformations.

High level representations of an image commonly make use of low level represen-

tations of the data as first stages of the abstraction process. Data modelling may be accomplished using either a *bottom-up* or a *top-down* strategy. In a *bottom-up* or data-based process, high level descriptions are obtained by searching for structure in the low level features and organizing them into more complex entities, like in (Ballard and Brown, 1982) to mention an example. Here, the modelling assumptions are implicit in the rules that govern such organization. The quality of the low level representation is decisive in the success of the subsequent processes. *Top-down* or model-based techniques, on the other hand, start from high-level representations of the data and try to fit them to the extracted low level information. Depending on its flexibility –degrees of freedom– the model is more or less influenced by the low level representation.

The multitude of analysis techniques use different low level information sources, like intensity, contrast, symmetry, curvature, texture, orientation, frequency and phase contents, etc. Many kinds of low level feature detectors have been developed based on the analysis of these local properties. Among them, the most outstanding are possibly the detectors of boundaries, regions, corners, junctions and blobs. Usually, the appropriate selection and detection of the image features to be the building blocks of higher level representations is critical for the correct performance of the whole process. It seems to be widely accepted that there is not a unique feature type that brings a complete representation of an image, but a set of different feature types must be used.

Many of those patterns can be obtained as a composition of elementary features. It may seem contradictory to talk about composite features in a low level stage. The expression composite feature does not refer to a more abstract representation, based on objects, object parts or so. We simply mean that a statistical pattern present in an image can be defined as the conjunction of other patterns with simpler statistics. Hence, the problem of low level representation can be tackled as a two step process, where first the elementary features are detected and then they are organized to form the composite feature. For example, a corner can be thought of as a point where the concurrence of two step edges takes place. The two step edges can be characterized by two of the basis functions with the proper scale and orientation and odd symmetry. A typical example of this is subband decomposition in a multiresolution framework. The basis functions for such a representation are usually patterns tunable on different scales, orientations and phases. Multiresolution decomposition is a powerful and general tool, suitable for a wide variety of applications. In fact,

this is the kind of processing that takes place in biological vision (Hubel, 1988). Nevertheless, some decomposition scheme is used for low level representation, the problem arises of how to combine the information from the diverse views generated to represent the composite pattern.

When a pattern has responses to more than one of the multiresolution detectors, a common solution is to select one of the extracted views, typically the one with the highest response, to be the representative of that pattern, rejecting the others, as in (Lindeberg, 1994). In this way, what one has is a partial representation of a pattern. In the example of the junction, the use of this solution would imply to represent the junction by only one of the two step edges instead of both. Other approaches simply perform a linear summation of all features, producing one single primal sketch comprising all patterns in the image (Morrone and Burr, 1988; Kovessi, 1996). Now, the problem is how to isolate patterns from each other. To obtain a complete representation, each composite pattern should be reconstructed separately by integrating the information from its constituent elementary features.

It is also common to integrate information from different views by searching for a predefined set of patterns. For instance, one can consider a set of elementary and composite features comprising lines, edges and junctions (Rodrigues and du Buf, 2004). Then, a technique for detecting each of them is developed separately, applying different rules according to their respective properties. This can be very efficient for specific application domains, with limited allowed features. Let us think, for example, in the analysis and classification of traffic signs. Another typical example, is to limit the set of features to textured regions and their boundaries (Malik and Perona, 1990). Nonetheless, for a general purpose application, such representations will produce partial descriptions of images.

In a general purpose application, the selection of features and the extraction method itself should be data-driven, in order that the representation fits image characteristics. The main of the approaches to solve this problem use learning techniques to extract knowledge from a data base, which later can be applied to analyze new images (Wersing and Körner, 2003). Little research is found on the development of a system to determine the most representative features in an image and to provide their correspondent detectors using only the image data, without using prior information.

1.1 Objectives

The aim of this work is to provide a low level representation model for volumetric data sets that is able to identify and describe different types of simple and composite features present in an image. We will consider only the case of volumetric gray level digital images and ignore color, video or stereo data. Composite low level features, which hereafter will be referred to as *visual patterns*, are defined here as those that can be properly described by one or more of the elementary features. The set of visual patterns must not be predefined, but must be determined for each image based only on its properties and without using a priori knowledge. The method for the extraction of the representation must neither use prior information and must be general for any set of features, without introducing ad hoc or domain-specific procedures. To this end, we will take advantage of different biological vision findings to inspire this work.

We propose to use a decomposition-integration strategy. This implies, in the first place, to develop a model for the decomposition of a 3D image into a set of elementary features, which must yield a comprehensive representation of the image. To this end, we will adopt a multiresolution scheme, where the elementary features, which we will refer to as frequency features, are tuned in frequency and orientation. Secondly, we will have to develop an integration procedure to combine the information from the elementary features in order to generate a complete view of each of the specific composite features in the image. This stage will be stated as an unsupervised clustering problem, where clusters determine composite features.

In order for the clustering to arrive to the proper result, a dissimilarity measure between pairs of frequency features must be defined that captures the conjunctions of features corresponding to relevant visual patterns. Hence, such a distance must embody a criterion to decide what is a relevant conjunction of features. Here, the criterion to link features is based on the concept of Phase Congruence, i.e., local alignment in phase of frequency components through a wide range of scales. This criterion has been defended as being the one used in biological vision for feature integration (Morrone and Burr, 1988). Besides the development of the representation model, we seek to determine the relationship between frequency features under the assumption of Phase Congruence. To this end, we will perform a study on a set of statistical measures of feature dissimilarity. From this study we expect to draw some conclusions about the mentioned relations and to find a suitable measure for

the clustering method.

The model to be developed is intended to serve as a general purpose low level representation to be applied when no specific purpose application is available or as alternative model when other solutions do not offer completely satisfactory results. Nevertheless, we desire to illustrate the capabilities of the method in some real applications. Hence, in this work the attention is focused in the enhancement of deformable models for medical image segmentation. In particular, we will design a technique for the initialization of deformable models using our visual pattern extraction method. The visual patterns isolated from the image will be used to define the initial state of a geodesic active model, to improve convergence to the target anatomical object in a volumetric data set from medical imaging.

1.2 The Method

The outline of the technique developed to arrive to the objectives established in section 1.1, has already been disclosed in that section. It involves the clustering of a set of frequency components –which we will often refer to as frequency features, obtained by means of a 3D filter bank– according to their ability to discriminate some given visual pattern in an image. In this way, the method generates a set of composite-feature detectors tuned to the specific visual patterns present in the data. The process can be described as a sequence of steps.

1. Selection of a subset of bands in the 3D filter bank with high information content, named *active* channels, by analyzing the spectral amplitude of the image.
2. Generation of frequency features using filters correspondent to active bands.
3. Measurement of dissimilarity between pairs of frequency features based on the normalized mutual information of their local energies.
4. Clustering of frequency features based on the dissimilarity matrix.
5. Reconstruction of visual patterns correspondent to each cluster by linear summation of frequency features.

The whole process is illustrated in figure 3.1 using a 2D example image for the sake of clarity.

Frequency feature extraction is accomplished by means of a bank of 3D log Gabor filters designed with rotational symmetry. The filter parameters and the sampling of the Fourier domain are determined to produce a comprehensive and redundant coverage of the spectrum. Clustering is accomplished using a hierarchical algorithm, which needs only a distance matrix to produce a partition of the data. The distance metric between the frequency features reflects the degree of alignment of their energy maxima and, hence, the degree of phase congruence. Such a distance is a measure of the statistical dependence of the energy of the active bands. A study on the adequacy of a set of statistical measures of dependence between energy maps is performed in order to provide the best clustering results.

1.3 Theoretical Foundations

Although the work presented in this thesis is not intended to simulate any biological process of vision, it has been inspired in several theories concerning the behavior of the Human Visual System (HVS). Some ideas extracted from this theories are used here to establish the working assumptions in which the research is based. What follows is a list with the three key ideas employed as the foundations of our representations scheme.

Efficient Coding

The *Efficient Coding* theory states that the HVS is adapted to the natural environment, so that it takes advantage of the invariances on its statistical properties to perform an optimal codification (Field, 1987, 1993, 1994; Olshausen and Field, 1997; Turiel and Parga, 2000; Simoncelli et al., 1992; Simoncelli, 2003). Its behavior can be modelled as a set of self-similar wavelet transforms or filter bank (Hubel and Wiesel, 1968; Hubel, 1988; Marčelja, 1980; Watson, 1987b, 2000; Graham, 1989; Field, 1993), which fits the experimental psychophysical evidence and yields a compact, although redundant, coding of natural images. The filters in the bank model the processing realized in the HVS by arrangements of neurons in the cortical areas of the brain, sensitive to frequency, orientation, motion, etc. In (Field, 1987), the 2D log Gabor

function is proposed as a plausible model for such sensors. From this theories, we will develop a 3D codification scheme to perform image decomposition.

Preattentive Feature Integration

According to the feature-integration theory of attention (Treisman and Gelade, 1980), the HVS distinguish between two types of composite features: separable and integral. Separable features, unlike integral features, are formed as a conjunction of simpler features from different sources of information or separable dimensions. They can only be integrated through mediation of serial focal attention, which needs of contextual information, while integral features are detected automatically in an early stage and in parallel across the visual field. No prior information is needed to detect integral features, since contextual interpretation of the image is performed in higher stages of the analysis pathway. To put it in another way, it is possible to extract some kinds of composite features in early stages of processing. Following this theory, we will perform feature integration at a low level stage by taking minimal assumptions about natural scene image statistics.

Relevance of Phase Congruence

The HVS perceives integral features as locations of high Phase Congruence (PC) (Morrone and Owens, 1987; Morrone and Burr, 1988), i.e., locations where local phase of a wide range of Fourier components of the image take the same value or, what is equivalent, points where the energy of the visual stimulus is locally maximum. This theory holds that the HVS performs detection of points with high phase congruence by means of contrast sensors organized in oriented and scaled pairs in quadrature, this is, with a phase shift of 90° , acting as estimators of local energy. This hypothesis is in agreement with recent studies that sustain that attention is unconsciously directed to image locations with high contrast (Reinagel and Zador, 1999). In that work, it is reported that high contrast is associated to low local correlation in natural scenes. Under the light of these evidences, the goal of attention would be to direct the center of gaze, with higher spatial resolution than parafoveal vision, to highly informative regions of the image. Here, we will apply this theories to design the integration criterion. Our dissimilarity measures will reflect the degree of alignment of the contrast detector responses, thus assigning relevance

of visual patterns to contrast and alignment.

Based on these three key points it is possible to derive integral features from an image by simulating the sensory coding of the HVS and using phase congruence as the link between simple and composite features. No prior information is needed at this stage, since contextual interpretation of the image is performed in high level stages of image analysis. PC alone is used to guide feature integration.

1.4 Applications

The method proposed here is potentially profitable in a wide range of 3D image applications, like medical image analysis, geophysical data analysis and non invasive visual inspection of materials and could be generalizable for the analysis of motion patterns from video sequences. In the particular case of medical image analysis, a suitable low level image representation can on its own be helpful to assist the physician in diagnosis and treatment. More often, low level representations are the basis to other applications, like segmentation, multimodal image registration, visualization or anatomical morphometry.

Among segmentation techniques, active models (Caselles et al., 1997; Paragios and Deriche, 1998) stand out due to their ability to integrate information from different cues. However, there are still many open research lines related to active models in the direction of finding the most appropriate definition of the external energies and the initial model, so that the optimized model resembles the true object shape as much as possible. All these methods derive image potentials from local features. Therefore, the initial state of the model needs to be close to the desired object boundary for the segmentation to succeed. Isolating visual patterns from each other allows the focusing of segmentation on a given target object, avoiding the interference of patterns. Specifically, we have designed an application for the initialization of geodesic active models using our composite-feature detector scheme. Visual patterns isolated from the image are used to define the initial state of a geodesic active model, improving convergence to the shape of the target object.

Nevertheless, other applications based on the extraction of low level features could take advantage of this technique. For the sake of completeness, we will apply our model to other 3D image analysis tasks. Specifically, the model will be applied for the enhancement of $2D+t$ data from Tagged Magnetic Resonance Imag-

ing (TMRI) sequences of the left ventricle of the heart, to study the strains in the ventricle wall muscle, and 3D Ground Penetrating Radar (GPR) imaging for the detection of buried objects, like anti-personnel mines or archeological sites.

1.5 Contributions

The work presented in this thesis is an attempt to fill the gap between the many simple feature detectors for 3D volumetric data and high level processing techniques. The developed method offers both a representation model and an extraction technique for 3D integral features or visual patterns. Visual patterns, as defined here, offer a more complete view of image features than the ones produced by simple-feature detectors. This approach is not limited to a multiresolution decomposition, but it goes a step further and the multiple views of the image are rearranged to group correlated information together. The whole process is based on minimal assumptions taken from theories about the HVS. The method is completely data-driven; no kind of prior information is used.

Although some research has been already done in this direction, like for example in (Bovik et al., 1990; Jain et al., 1997; Acharyya et al., 2003) and especially in (Rodríguez Sánchez et al., 2000; Chamorro Martínez et al., 2003; Manjunath and Chellappa, 1993), we don't have knowledge of any approach of this kind developed for volumetric data, which implies the extension of some ideas extracted for the HVS from 2D to 3D. The HVS does not deal with pure 3D information, but with stereo projections of 3D scenes. The generalization of the biological coding processes is not trivial, since it implies the redefinition of the sensory coding to take into account a new dimension but maintaining a good spectral coverage, orientation sensitivity and adding new constraints, like rotational symmetry around the filter's direction. The 3D filter bank designed in this work takes into account all these aspects.

Another contribution of this work is the study on efficient measures for PC estimation between frequency features. Some of the studied measures are widely applied to compare images for registration or for image retrieval. Their behavior in those fields have already been studied. However, the kind of images implicated in the present application, a set of subband versions of a given image, has a series of particularities that have to be considered. In addition to those measures, a new set of distances is derived from them in an attempt to simulate the biological

process of attention. The suitability of the distances is tested analyzing the fitness of the composite features resulting from visual pattern extraction with the expected ones. This experimentation has also served to take conclusions about the statistical relation between subband images.

The applicability of the developed composite-feature detector is proved by its incorporation to a 3D medical image segmentation method, specifically, a novel semiautomatic technique for the initialization of a geodesic active model. This solution uses neither manual nor prior definition of the initial model, but it is selected from the visual pattern extraction output. One or more of the composite-features identified in the image are used as a rough description of the segmentation target. The use of this technique improves segmentation results and speeds up the convergence of the method.

1.6 Thesis Overview

This thesis is organized in 6 chapters. In the second chapter the theoretical basis of the proposed method are presented. Firstly, we introduce a review on some findings about the processing of visual information in HVS in which this work is inspired. The HVS seems to perform a multiresolution decomposition of visual information in the early stages of processing. Such decomposition is supposed to be optimized to reflect the sparseness of natural visual stimuli and can be modelled by energy filtering methods. Higher stages of processing in the HVS must be responsible for the binding of low level features in order to produce higher level representations of the elements in an image. Some approaches for feature integration are reviewed in section 2.4. In subsection 2.5, we come back to the goals of our work in relation to the other cited approaches.

The proposed technique is described in detail in chapter 3. In section 3.1 the multiresolution decomposition strategy is described. The processing of 3D volumetric data requires taking into account a second orientation dimension to represent frequency channels and imposes the use of non uniform sampling schemes in the angular coordinates. Our solutions for these issues will be proposed in this section. In section 3.2 we propose a technique to select the most relevant features to represent the visual patterns in the image in order to reduce computational burden and enhance performance. In section 3.3 we present a first approach for the estimation

of frequency feature dissimilarity. Phase Congruence imposes a statistical dependence between the energy maxima distribution of two subband images. Hence, a measure based on information theoretical concept of Mutual Information is likely to produce good results. The clustering algorithm applied to the dissimilarity matrix is described in section 3.4. Section 3.5 presents some examples on the performance of the method with synthetic images.

The use of Mutual Information to reflect the dependence between frequency features, is just one among a large amount of measures for statistical dependence. In chapter 4, we propose a series of alternative measures in order to validate the results obtained with Mutual Information. The performance of the different measures is compared according to their computation costs in section 4.3 and according to their effectiveness in a test bench in section 4.4. Experimental results are discussed in section 4.5.

Several applications of the method are explained in chapter 5. In particular, the first section is centered in the application of the representation model to the improvement of segmentation with active models. There, we will present a strategy for the initialization of active models –specifically geodesic active models– in order to ensure convergence to the target patterns and to reduce convergence time. This will be accomplished by generating the initial state of the model from the visual patterns that best represent the target object. Nonetheless, other applications are proposed in section 5.2, which mainly imply enhancement of 3D data from different image modalities and with different purposes. Enhancement results from eliminating non desired visual patterns in the image.

In the last chapter we present the general conclusions derived from the developed work. Based on those conclusions, we will propose new lines for future work in order to round off the model and cover more applications.

CHAPTER 2

Multiresolution Image Analysis

2.1 Low Level Processing in the HVS

The first attempts to learn how the human brain processes visual information took place in the 60's. Hubel and Wiesel (1962; 1968) recorded the responses of neurons from the retina, the lateral geniculate nucleus and the striate cortex –or primary visual cortex– of the brain to a wide set of visual stimuli. They found that such neurons are organized hierarchically, following a topographic representation, this is, the connections between neurons in one level of processing and neurons in the next one define a mapping that preserves spatial contiguity. Hence, the set of neurons in one stage that *feed* a single neuron in the following stage, which together with their synapses conform a *receptive field*, define a well localized area.

The receptive field of a cell can be inferred by identifying the visual stimulus that gives place to the strongest cell response. In their experiments, Hubel and Wiesel, discovered a high degree of specialization in the cells involved in low level visual processing. According to their receptive field response profiles, they classified neurons into four main types: center-surround cells, simple cells, complex cells and hypercomplex or end-stopped cells. They are supposed to process visual information in a hierarchical manner: simple cell's receptive fields are formed by as a combination of center-surround cells' responses, complex cells receive inputs from simple cells and so on.

Center-surround cells, present circularly symmetric receptive fields, with either

an excitatory central region and an inhibitory surround or the opposite. Such cells are known to computational vision researchers as *Mexican-hat* operators. In words of Hubel, they do not seem to be exerting any profound transformation of the visual information. Simple cells have elongated receptive fields, with two or three parallel regions, with alternating inhibitory and excitatory behavior. Consequently, they respond best to elongated contrast stimuli, namely bars and edges. They are very selective in location, orientation and size of the input stimulus. Their possible preferred orientations seem to sweep the whole range and they appear in a wide range of sizes. Complex Cells are also orientation and size selective but they are less sensitive to location. Their response does not depend on the exact position of the stimulus in their receptive field, which is wider than in simple cells. Many complex cells are also selective in motion and even in the direction of movement, in the sense that they respond only when the stimulus moves in one direction and not when it moves in the opposite direction. Finally, hypercomplex cells, also called end-stopped cells, present lateral inhibition besides the flanking polarity responsible for orientations selectivity. They seem to be very specialized feature detectors. Some respond to moving bars of a specific length and some to moving corners or angles.

The studies accomplished by Hubel and Wiesel suggested that the visual system of mammals acts as a set of spatial feature detectors of different kinds, like lines, edges and corners. Nevertheless, diverse results from psychophysical and neurophysiological experimentations generated a debate on whether the visual cortical areas of the brain perform a spatial or a Fourier analysis of the visual stimuli. Some authors claimed that the visual system is performing a Fourier analysis of visual information (Campbell and Robson, 1968; Campbell et al., 1969; DeValois et al., 1978). Campbell et al. showed that cortical cells are narrowly tuned to the spatial frequency of inputs, and had typical bandpass responses, responding only to inputs in a specific frequency range.

In (Marčelja, 1980) both explanations were unified. Marčelja suggested that the receptive fields of simple cells can be considered to perform simultaneous localization in space and spatial frequency by means of local spatial frequency analysis. Based on this assumption and on the evidence for linear spatial summation across the receptive field, he proposed the Gabor filter (Gabor, 1946) as a mathematical model of the response of simple cells, arguing that achieve the lower limit of joint uncertainty in space and spatial frequency and, hence, they are optimal in information transmission.

Local spatial frequency representation is thought to be performed at various levels of spatial and frequency resolution. The evidence of different sizes of the receptive field of the neurons of the retina and visual cortex (Hubel, 1988), links to scale space theory (Witkin, 1983; Koenderink, 1984; ter Haar Romeny, 2003; Florack et al., 1994; Lindeberg, 1994). It is also known that the spatial bandwidths of simple cells vary in a wide range (Webster and DeValois, 1985) and that they decrease with spatial frequency (Anderson and Burr, 1987). Such characteristics have been computationally simulated by convolving the input image with a bank of geometrically scaled filters –filters with spatial bandwidths that decrease with spatial frequency– with tuned frequencies, orientations and bandwidths (Watson, 2000). Particularly, Field (1993) developed a model for visual processing of mammals based on multiresolution coding using wavelet transforms.

2.2 Energy Filtering

Several research studies (Pollen and Ronner, 1981; Watson, 1987b) evidenced the presence of pairs of adjacent simple cells matched in preferred orientation and spatial frequency but having a quadrature (90°) phase relationship within their common receptive-field area. Watson pointed that these pairs could be building the receptive fields of complex cells by non linear mechanisms and that they may be responsible for motion sensing.

Several authors have modelled the behavior of complex cortical cells as the envelope or energy of the responses of oriented and scaled quadrature pairs of filters, which simulate the responses of simple cells (Adelson and Bergen, 1985; Field, 1987). Heitger et al. (1992) proposed a plausible hierarchical scheme for the receptive profiles of the visual cortex neurons. Simple cells would be *built* by linear summation of responses from center-surround cells in the lateral geniculate nucleus whose receptive fields fall along a line across the retina, as pointed by Hubel (1988). Complex cells responses would be the result of non-linear –quadratic– summation of the responses of two simple cells in quadrature phase. End-stopped cells would be the result of pooling complex cell responses so as to define their first and second derivatives to represent respectively asymmetric –single-stopped– and symmetric –double stopped– hypercomplex cells.

A feature detector that resembles the behavior of complex cortical cells is the

energy based filtering, introduced in (Morrone and Owens, 1987) and widely studied by several authors (Owens et al., 1989; Perona and Malik, 1990; Morrone and Burr, 1988; Venkatesh and Owens, 1990). In this technique, image features are identified as the maxima of the local energy of the image. This kind of operators outperforms previous linear filters in many aspects. Unlike, for instance, Canny's maxima of derivative of Gaussian (Canny, 1986) or Marr-Hildreth's zero-crossings of Laplacian of Gaussian (Marr and Hildreth, 1980), energy based filters do not mark edges in sine-wave signals, detect features that are mixtures of step –odd– and delta –even– intensity profiles –like Mach bands– and are a projection. Furthermore, this type of filters fulfill Canny's criteria of good detection, good localization and single response and proposed a set of quantitative measures for these properties (Canny, 1986). They also verify Owens' requirement of being a projection (Owens et al., 1989). A comparison of energy-based with zero-crossing feature detection is presented in (Ronse, 1990).

The energy E of a signal X can be computed as the sum of the squared signal and its Hilbert transform $\mathcal{H}(X)$ (Morrone and Owens, 1987; Granlund and Knutsson, 1995).

$$E(x) = \sqrt{X^2(x) + \mathcal{H}(X)^2(x)} \quad (2.1)$$

Nevertheless, the calculus of E using the Hilbert transform is computationally expensive. In (Morrone and Burr, 1988; Owens et al., 1989) a new definition of local energy $\hat{E}$ was introduced which reduced computational cost. It is calculated as the sum of the squared responses of a pair of filters in quadrature, this is, a pair of filters with the same scale and orientation but with a phase shift of 90° . This is achieved by using an even filter, h^e , tuned to detect symmetric patterns, and an odd filter h^o , sensitive to antisymmetric patterns.

$$\hat{E}(x) = \sqrt{(X(x) * h^e(x))^2 + (X(x) * h^o(x))^2} \quad (2.2)$$

The equivalence between maxima in E and $\hat{E}$ was demonstrated based on the fact that a quadrature pair of filters are the Hilbert transform of each other (Venkatesh and Owens, 1990).

Another advantage of the use of quadrature filter pairs, the local phase of the signal can be easily derived from the responses of the quadrature pair.

$$\arg(x) = \arctan 2((X(x) * h^e(x)), (X(x) * h^o(x))) \quad (2.3)$$

This information can be used to classify features according to their symmetry (Venkatesh and Owens, 1990; Perona and Malik, 1990).

The use of quadrature filter pairs allow to analyze a signal at different levels of resolution and, hence, the relevance of a spatial feature is measured in relation to a local context, relatively independent on the presence of other features in the image. This is achieved by taking h^e and h^o with spectral amplitude localized in a given frequency and orientation. The selection of suitable sampling of frequency and orientation dimensions and the appropriate selection on their respective bandwidths leads to a multiresolution decomposition scheme resembling the one that takes place in the HVS.

Perona and Malik (1990) arrived at the same formulation as Morrone, not through the study and simulation of the HVS, but using a different reasoning: the computational criterion of localizing composite edges –mixture of step, peak and roof profiles– exactly . They conclude that linear filters are not suitable for exact edge detection while the use of the norm of a response vector composed of a pair of even and odd symmetric linear filters guaranties exact detection. They also derived a series of optimality criteria for evaluating quadratic filters for edge detection in a manner similar to Canny. However, they do not constrain the filter pair to have the same scale.

2.2.1 Phase Congruency

The energy of a signal is closely related to its Phase Congruence (PC). Phase congruence measures the local degree of alignment among the Fourier components of an image. It can be estimated by measuring the mean deviation of the local phase of the Fourier components from the phase of the resultant.

Given the Fourier series expansion of a signal $X(x)$,

$$X(x) = \sum_n A_n \cos(2\pi nx + \phi_{n_0}) = \sum_n \cos(\phi_n(x)) \quad (2.4)$$

where $A_n(x)$ represents the amplitude of the n^{th} cosine component, ϕ_{n_0} its phase offset and $\phi_n(x)$ is the local phase of the Fourier component in position x . Then the phase congruency of the signal $X(x)$ is defined as (Morrone and Owens, 1987)

$$PC(x) = \max_{\bar{\phi} \in [0, 2\pi]} \frac{\sum_n A_n \cos(\phi_n(x) - \bar{\phi}(x))}{\sum_n A_n} \quad (2.5)$$

where $\phi_n(x)$ is estimated in a given vicinity of x related to the scale. Maxima in PC is approximately equivalent to minima in the weighted variance of local phase angles (Kovesi, 1996).

In (Venkatesh and Owens, 1990), it was demonstrated that energy is equal to phase congruency scaled by the sum of the Fourier amplitudes.

$$E(x) = PC \sum_n A_n \quad (2.6)$$

Therefore, searching for peaks in the local energy is equivalent to look for points where the phase congruence (PC) is maximum. If we consider the Fourier components as vectors with length A_n and angle $\phi_n(x)$, when all the Fourier components are locally in phase they maximally contribute to the resultant and, hence, the energy is locally maximum, since the amplitudes A_n are independent on x .

The success of the energy model in predicting several psychophysical effects, like Mach bands or illusory contours, has led Morrone and Owens (1987) to the hypothesis that the HVS perceives features at points of high PC. The sensitivity of the HVS to PC has also been studied by other authors (Oppenheim and Lim, 1981; Ross et al., 1989; Fleet, 1992; du Buf, 1994). The calculus of energy maxima can be employed as an alternative way of estimating PC. This is the mechanism that the human visual mechanisms is supposed to employ to detect features at a low level of representation. In a multiresolution scheme, points of high PC can be identified with points where there is concurrence of local maxima on different subband images.

2.3 Multiresolution Decomposition

There exist many approaches to accomplish a multiresolution decomposition of an image. In the vision research field, one of the most popular techniques is the octave filter bank (Field, 1987; Morrone and Burr, 1988; Daugman, 1988). It produces a set of subband versions of a signal by convolving it with a family of band-pass filters covering the frequency spectrum. The set of filters is chosen with center frequencies and bandwidths such that the spectral domain is roughly decomposed in an octave or dyadic subband tree structure. The octave subband structure can be considered to recursively split a band of low frequencies in two. As a result, each band pass filter has double central frequency and bandwidth that the immediately

previous one. Since the amplitude spectrum of natural images statistically decays with the inverse of the frequency, octave band filter banks produce subband images with roughly the same energy. In 2D the filter bank is also selective in orientation. Usually, each frequency band is split in several angular bands with equal angular bandwidth.

This kind of decomposition can be considered a special case of Discrete Wavelet Transform (DWT), since it is a complete family of functions that can be derived by translation –modulation– and dilation of the same function –the mother wavelet. In two dimensions, the octave filter bank can be thought of as an anisotropic or oriented wavelet, generated from an elementary function by dilation, translation and rotation. Therefore, they are not true 2D wavelets, in the usual sense, which would have the same orientation. Nevertheless, the term wavelet is commonly restricted to the specific case of affine wavelets, where the mother wavelet and the sampling schemes is chosen to produce optimally compact representations with orthonormal basis.

Many of the band-pass filters used by the vision research community to represent the receptive fields of visual cortical cells, like Gaussians derivative, Difference of Gaussians (DoG) (Field, 1994), Gabor (Marčelja, 1980; Daugman, 1985), log Normal (Morrone and Burr, 1988; Granlund and Knutsson, 1995), log Gabor (Field, 1987) or Cauchy filter (Klein and Levi, 1985) –Poisson distribution– do not yield orthogonal basis. They are not linearly independent. Hence, there always exists some degree of redundancy in the representation –the subband images are not uncorrelated. Moreover, the sampling schemes are not as restrictive as affine wavelet schemes. Center frequencies and bandwidths are chosen more freely in order to fit the experimental data. Another implication of non-orthogonality is that the wavelet coefficients cannot be determined by simply projecting the original image over the filter’s functions.

Another point against these elementary functions is that they are not strictly local, they stretch out to infinity. Non compact support is not biologically possible. Even considering the case of digital implementations of the filters, they extend to the whole image domain and have to be truncated. However, they only pretend to be a mathematical idealization of the receptive fields. For example, in the case of a Gabor function, its gaussian envelope has 99.7% of the area within three standard deviations. Although the Gabor functions are not strictly local, their Gaussian envelope causes their greatest sensitivity to be concentrated near the center of that

envelope.

Although these representations are not a basis, they are a *frame* (Daubechies et al., 1986) as long as the spatial and frequency uncertainties verify the relation $\Delta x \Delta u = 1/m$ and $\Delta y \Delta v = 1/n$ with $n, m = 2, 3, 4$, (Daubechies et al., 1986) –consistent with the uncertainty principle, adapted from Heisenberg ideas to communication theory by Gabor (Gabor, 1946). Frames are *complete* representations of a signal in either spatial and spectral domains, which means that both domains are partitioned into a set of subspaces in a jointly exhaustive fashion, this is, without gaps. To put it in another way, the intersections of the subspaces defined by each of the functions is the whole image domain. In a complete representation there is no information loss. The original signal can be reconstructed from the subband decomposition, although the obtention of the transform coefficients for non-orthogonal frames is very difficult. In a non-orthogonal representation, the summation of the energies of the subband images is not equal to the energy of the original image. It only can be ensured to be between the so-called *frame bounds*. The frame bounds are usually expressed as fractions A and B of the total energy of the original image. When $A = B = 1$ the representation is orthonormal. When $1 < A < B$ the representation is complete but redundant. The closer A and B to 1, the *tighter* the representation and the less redundant.

The lack of orthogonality of the octave filter bank has led various authors to search for alternative decomposition schemes, like affine discrete wavelets (Mallat, 1989), generalized Gabor schemes (Bastiaans, 1981; Porat and Zeevi, 1989), the Cortex Transform (Watson, 1987b,a) or quadrature mirror filter banks (Woods and O’Neal, 1986; Simoncelli and Adelson, 1990). Using these approaches the original signal can be easily exactly reconstructed. The possibility of exactly reconstruct the original signal is of great importance in applications like image compression.

However, the HVS does not seem to perform the reconstruction of the original image. Neither low level feature extraction needs to produce a reconstruction of the image. Furthermore, a certain amount of redundancy can be convenient since it reduces the effects of noise. Indeed, it is known that the HVS produces a redundant codification of the visual information. As noted in (Daugman, 1988), nonorthogonal representations frequently appear in biological sensory and motor systems. In (Morrone and Burr, 1988) a primal sketch of image features is constructed by linear summation of filter’s responses. The DC component is not taken into account. In (Kovesi, 1996) the summation is normalized and weighted by spread estimation to

obtain a PC map. Although the reconstruction by linear summation of filter responses is never exact due to redundancy, the effect of non-orthogonality can be reduced by carefully choosing the filter parameters. The goal is to obtain an efficient low level representation, not a compact and reversible representation of the image and filter bank schemes have fast implementations in spatial and/or Fourier domains.

2.3.1 Quadrature Filter Models

One of the most popular models of simple cells receptive fields in the striate cortex of mammals is the Gabor function. It has been widely employed as the elementary basis function of filter bank schemes for multiresolution image representation and analysis. It is a Gaussian function modulated by a complex exponential. Its one-dimensional expression of the impulse response of the Gabor filter for a given modulating frequency ρ_i and scale σ is the following.

$$h(x; \rho_i) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \cdot \exp(ix\rho_i). \quad (2.7)$$

Thus, it has an even real part h^e and an odd imaginary part $i \cdot h^o$. Thus, one single filter $h = h^e + i \cdot h^o$, with $i = \sqrt{-1}$, simultaneously encodes the even and odd components of the image or, to put it in another way, the local energy and the local phase. In the frequency domain it is a Gaussian centered at the modulation frequency. The selection of Gabor filters to model simple cells receptive field is justified in (Marčelja, 1980) and in (Daugman, 1985). Daugman derived the two-dimensional analog of Gabor's Uncertainty Principle and showed that the Gabor function is the one with minimal uncertainty in both spatial and Fourier domains and, hence, with maximum attainable information. Minimizing the uncertainties in both domains implies that the function is well localized in both domains. Other choices of simple cell functions, like the Derivative of Gaussians wavelet will result in good spatial localization but poor localization in frequency. The selection of a codification with minimal joint entropy conveys the underlying assumption that the simple cell receptive fields are optimized to maximize information transmission, producing a *compact code* which allows the input to be represented with a reduced number of cells with minimal error (Barlow, 1960).

Field (1987, 1994) claims that the efficiency of a code depends on the statistics of the input and the goal of the code. In his opinion, the HVS would perform an

efficient coding optimized to the statistics of natural images. He claims, however, that maximum attainable information could not be the goal of the human sensory coding and that Gabor code is quite inefficient for a wide variety of images. He and other authors (Srivastava et al., 2003; Simoncelli, 2005) sustain that the most efficient code may not be a compact code, but a *sparse distributed* one, where all cells in the code have an equal response probability across the class of images but have a low response probability for any single image. The response histograms of such codes are sparse, i.e., show high kurtosis. The probability distribution that models this behavior is the *leptokurtotic* or *supergaussian*, with sharp peak at zero and a long tail. This means that for most inputs only a few set of cells is probable to be active or, in other words, that a given cell is inactive for most inputs and is only rarely activated. This seems to fit visual cortex cells' properties. In such a code, the dimensionality is not reduced, which is also in accordance with nature, since the visual cortex of the brain contains a large number of cells in relation to other areas. In contrast to the compact code, in this approach the goal is to maximize the entropy of the response histograms by minimizing statistical dependence between units. Physiological evidence for sparse and independent representations is presented in (Vinje and Gallant, 2000; Bell and Sejnowski, 1997).

A great deal of research has been developed in order to test the sparse coding scheme with promising results. Olshausen and Field (1997) derived a set of basis functions for image coding under the a criterion of optimal complete sparse representation. The derived basis functions are well localized, sensitive to orientation and tuned to different spatial frequency bands, resembling the receptive fields of the striate cortex cells. They also suggested that other hypothetical objectives of efficient coding, such as statistical independence may produce similar results. Indeed, several authors (Bell and Sejnowski, 1997; Hoyer, 2002) claimed that simple cell receptive fields are optimized for statistical independence. The results obtained by Hoyer on deriving simple cells receptive fields from Independent Component Analysis (ICA) under the constraint of linearity leads to second order uncorrelated functions, but not independent, which means that there exist higher order redundancy. They also derived complex cells receptive fields by Independent Subspace Analysis (ISA), which assumes that the processing units are not linearly independent but form clusters of highly correlated unit with inter-cluster independence. They followed the model of sum of the squares of the responses of simple cell models as in the energy model. This led to receptive fields highly correlated in orientation and spatial frequency but independent in phase.

The first decomposition scheme proposed by Field (1987), was self-similar bank of filters like the one proposed previously. He related the success of these transforms in coding natural scenes to its suitability to capture the statistical redundancy of natural scene images, that lays in the scale invariance of spectral amplitude and in phase congruence –scale invariance of spectral phase. He also proposed the use of filters other than Gabors to produce a sparse coding with highly kurtotic response histograms. Specifically, he suggested the use of Difference of Gaussians (DoG) or log Gabor functions (Field, 1994). Both present high *leptokurtotic* response histograms. The DoG filter –or *Mexican hat* filter– is well suited to model the receptive fields of lateral geniculate cells and retinal ganglion cells (Marr and Hildreth, 1980). The log Gabor filter resembles the receptive fields of simple cells, which have symmetric profiles in a logarithmic frequency scale. In the frequency domain, when represented in a logarithmic scale, it is a Gaussian. The log Gabor function is quite similar to the log Normal function, which had been already proposed in (Morrone and Burr, 1988) to fit the experimental data collected in (Anderson and Burr, 1985, 1987). Logarithmic scale Gaussians have been also used for other authors (Haglund and Fleet, 1994; Granlund and Knutsson, 1995; Kovessi, 1996; Rodríguez Sánchez et al., 1999; Chamorro Martínez et al., 2003). It has the following expression in the frequency domain

$$T(\rho; \rho_i) = \exp \left(-\frac{(\log(\rho/\rho_i))^2}{2(\log(\sigma_\rho/\rho_i))^2} \right). \quad (2.8)$$

Due to the logarithm, it has no analytic expression in the spatial domain.

Besides all these considerations regarding the statistics of natural images, the log Gabor filter presents several advantages in relation to Gabor filters. Gabor filters are not actually in quadrature as are non-null for negative frequencies and non-zero DC component. This is evident from the Fourier domain representation of the Gabor function. As a result of linearity property of the Fourier transform, $\mathcal{F}(h) = \mathcal{F}(h^e + i \cdot h^o) = \mathcal{F}(h^e) + i \cdot \mathcal{F}(h^o) = T^e + i \cdot T^o = T$. As can be seen in figure 2.1, the transfer function of the even symmetric component results from the contributions of two Gaussians, one for positive frequencies and the other for negative frequencies. They overlap at the low frequencies, producing a non-null DC component. Provided that the odd symmetric component of the transfer function T^o is pure imaginary, then the transfer function T is real. Most contributions to negative frequencies from the two components cancel out, but still T has a tail towards the negative frequencies which increases with the bandwidth of the filter. In order to keep a small DC component, the bandwidth of a Gabor cannot be

arbitrarily enlarged.

As can be observed in figure 2.2, this problem does not appear with the log Gabor function, which has no DC component independently of the filters bandwidth –to be precise, its value at zero frequency from equation 2.8 is an indeterminate, but in the limit tends to zero. The transfer function of the log Gabor filter in equation 2.8 can be thought of as the sum of one real even symmetric component T^e and an imaginary odd symmetric component T^o multiplied by the imaginary unit i , just like in the case of the Gabor function. Unlike with Gabors, the negative and positive frequency parts of T^e and T^o do not have tails towards the zero frequency and T^e and T^o completely cancel out in the negative frequencies. Hence, its real and odd parts are in quadrature. As a consequence, a smaller number of filters with wider bandwidths is necessary to cover the spectral domain.

Another advantage of the log Gabor relies in its long tail towards the high frequencies. Since natural images present scale invariance, this is, present spectral amplitude profiles that decays with the inverse of the frequency (Field, 1993), then a filter that presents a similar behavior should be able to properly encode those images (Kovesi, 1996). According to Kovesi, the Gabor function over-represents low frequencies and under-represents high frequencies. Finally, as explained before, neither Gabor not log Gabor filter banks form an orthogonal basis, but they are a frame. Haglund92 compared the Gabor and log Normal frame bounds, concluding that log Normal is tighter than Gabor in equal conditions. This is supposed to hold for log Gabor, which are almost identical, apart from the definition of the bandwidths.

The main drawback of the log Gabor filter is that it has not an analytical expression in the spatial domain. However, design and filtering can be efficiently accomplished in the Fourier domain. Convolution in the spatial domain is equivalent to the product in the Fourier domain. The response to a log Gabor in the spatial domain is obtained by Fast Fourier Transform of the image, product with the filter transfer function and Fast Inverse Fourier Transform of the multiplication result. Considering the transfer function T of the log Gabor filter as $T = T^e + i \cdot T^o$, the real part of the filtered image can be identified as $X(x) * h^e$ and the imaginary part as $X(x) * h^o$.

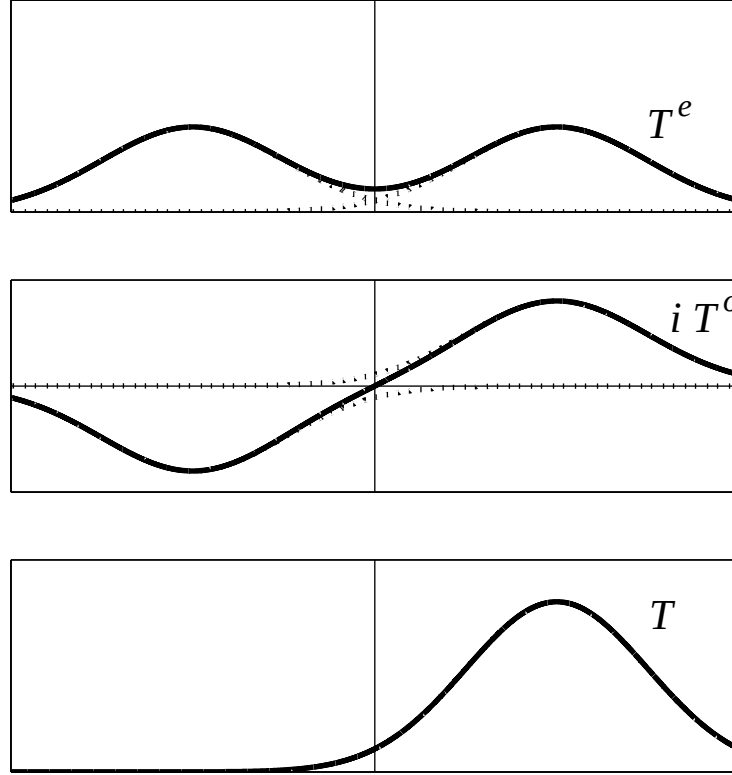


Figure 2.1: Transfer function of the Gabor function. *Top:* Transfer function of the even component T^e of a Gabor function. The dotted lines represent the two Gaussian functions that contribute to the even part, one for the positive frequencies and the other for negative frequencies. As can be seen, the even part has a non null DC component. *Center:* Transfer function of the odd component T^o of a Gabor function multiplied by $i = \sqrt{-1}$. The dotted lines represent the two Gaussian functions that contribute to the odd part, one for the positive frequencies and the other for negative frequencies, which cancel out at the origin of frequencies frequency. *Bottom:* Transfer function of the Gabor function as the sum of the above even and odd components.

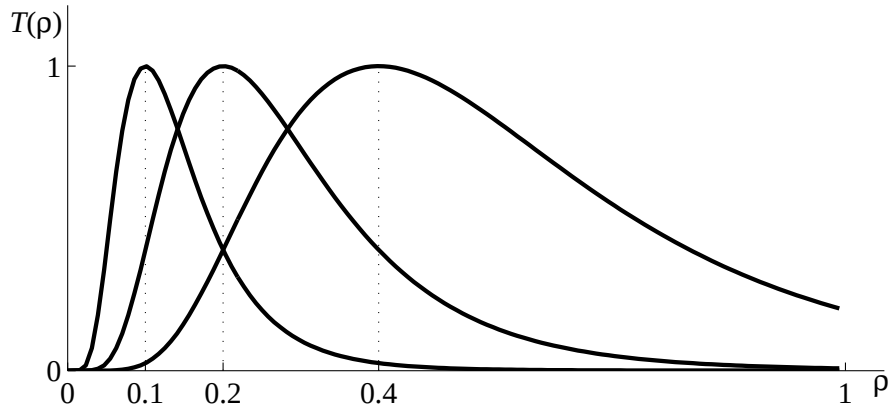


Figure 2.2: Log Gabor transfer functions for various values of ρ_i and 2 octave bandwidth.

2.4 Frequency Feature Integration

To integrate the information from different scales and orientations, most of the energy based approaches combine the information of the different bands to produce one single *primal sketch* (Marr and Hildreth, 1980) of the image, i.e., a single map of all the features present in an image. Morrone and Burr (1988) developed a method to calculate local energy that involved the sum of the responses of a bank of filters covering the visible spectrum. Kovessi (1996) defined a new measure of PC as the weighted mean absolute deviation from the weighted mean of phase from the responses of a set of 2D log Gabor filters. Additionally, he incorporated factors related to the spread of filter responses to enlarge the PC of features with contributions from broad ranges of frequencies.

The generation of a primal sketch of the image features by means of energy estimation is a possible way of combining information from a set of frequency features. However, with the information provided by multiresolution decomposition, it is already possible to extract a higher level representation of the image, where distinct features can be identified and isolated. For example, using phase information it is possible to classify image features in relation to their symmetry (Venkatesh and Owens, 1990; Perona and Malik, 1990). More sophisticated features that are a composition of the previously defined elementary features, like textures, junctions, etc, could be derived by introducing some binding criterion.

It is known that several cues of visual information, which are processed inde-

pendently at early stages, are then integrated at high levels of processing with the mediation of attention mechanisms, which can be guided by saliency, contextual information, memory, etc (Treisman and Gelade, 1980; Wolfe, 1998, 2000). However, it is also known that there exist features that are non-separable and are integrated preattentively, this is, concurrence of low level non-separable features produces their binding without the mediation of attention (Treisman and Gelade, 1980). Moreover, more recent studies suggest that feature that were supposed separable, do present some degree of synergy during preattentive processing (Kubovy et al., 1999). Particularly, it has been long reported that information from different scales and orientations is combined preattentively in the case of textures. (Julesz, 1981, 1986; Bergen and Julesz, 1983; Beck, 1982; Beck et al., 1983).

2.4.1 Biologically Feasible Techniques

Several authors have tried to accomplish preattentive feature integration by simulating the behavior of the next cells in the visual processing hierarchy. So far, we have talked about multiresolution decomposition schemes that simulate the receptive field of simple and complex cell. The next step in the hierarchy, these cells interact producing more sophisticated receptive fields sensitive to composite features, like end-stopped cells, that integrates concurrent responses of cells with different sizes or orientation to perform end-stopping.

Malik and Perona (1990) designed a method for preattentive texture discrimination. Firstly, the image is decomposed by a bank of even symmetric linear filters. Suppression of weak responses is then accomplished by spatial inhibition of responses. Detection of texture contours is realized using wide odd-symmetric filters. Authors sustain that there are signals that are indistinguishable when seen with quadrature filter pairs and that using only even filters is better, while odd filters are not part of the texture detection pathway and are used only to detect texture boundaries.

A similar technique is presented in (Manjunath and Chellappa, 1993), but now using a Gabor wavelet to represent elementary features. Again, local competitive interaction within spatial neighbors within each orientation and scale are performed to reduce noise and illumination effects. Line-end and corner detection is accomplished by interscale interactions, this is, the responses of complex cell of different sizes are combined to produce end inhibition. This is followed by a stage of bound-

ary completion by grouping of the responses of complex and hypercomplex cells with similar orientations. This step is also responsible for the detection of illusory contours (Treisman and Gelade, 1980). Edge-based contours –real and illusory– are identified with local maxima of the output and texture boundaries with the maxima of the derivative of the output.

In (Heitger et al., 1998), even and odd symmetric end-stopped receptive fields are defined for single or double sided end-stopping detection respectively. These end-stopped operators are constructed by combinations of adjacent complex cells of different sizes. Boundary completion is accomplished here by convolving hypercomplex cell's responses with club-shaped grouping fields of two types: *para* grouping fields, which collect responses with orientation parallel its, and *ortho* grouping field, which collect responses of cells orthogonal to it. *Para* grouping fields tends to produce the continuation of existing features, like corners, while *ortho* grouping field tends to form contours from line-ends. In (Rodrigues and du Buf, 2004), the operation defined in (Heitger et al., 1998) are employed for junction classification

2.4.2 Composite Feature Learning

The approaches described in the previous subsection are derived from the study and simulation of biological process. Other authors, instead, have tried to follow an inverse reasoning. They have hypothesized about the goals of visual coding and have tried to optimize some processing architecture to fit the stated requirements. The results obtained can be compared to experimental data to test the correctness of the working assumptions.

One example of such strategy is presented in (Hoyer, 2002; Hoyer and Hyvärinen, 2002). As mentioned before, Hoyer realized unsupervised learning of simple cell receptive fields by ICA and complex cells receptive fields by a hierarchical ISA model. Encouraged, with the results obtained, this research goes further and tries to derive the receptive fields of more sophisticated cells, which they call *contour cells*, by adding a 3rd layer in the hierarchical model. In this occasion, the bottom layers of the hierarchy do not play an active part in the optimization process and are considered known. Then, contour cells receptive fields are learned using ICA and the constraints of non-negativity and sparse distribution of the responses. Although the network is linear, the inferred functions are not linear due to noise and overcompleteness of the basis. The properties of the learned functions resemble the

behavior of end-stopped cells in that they are selective for position, orientation and contour length.

A much more sophisticated methodology is presented in (Wersing and Körner, 2003). Authors develop a recognition technique following a feed-forward neural architecture. The processing hierarchy consists three main layers. The first layer, can be considered composed of orientation sensitive cells, similar simple cells in the visual cortex. Cells in the second layer, are called *combination cells*, i.e., cells sensitive to combinations of cells from previous layer with shift invariance, which is achieved using a share-weighting architecture. Cells in the final layer are view-tuned cells, sensitive to a whole view of a presented object. The receptive field from the cells in the different layers are learned using various techniques, like PCA, sparse invariant coding optimization or supervised training.

2.4.3 Unsupervised Region Classification

Among non biologically feasible strategies, one of the most common is the unsupervised classification of image points into a set of spatial features, mainly regions with homogeneous texture properties. With small variations, this technique involves the decomposition of the image into a set of N_f frequency features using some multiresolution family of functions. Each image point is characterized by a vector of length N_f containing the values of the frequency features. A metric on this N_f -dimensional space is defined to estimate distances between pairs of positions. Then, clustering of spatial positions into uniform regions is performed by minimization of the overall distance. The main divergences between these techniques relies on the definition of texture distances and the clustering algorithm employed.

The typical example of such segmentation technique appears in (Jain et al., 1997). Here, image decomposition is performed using a bank of Gabor filters, but considering only the real part. Therefore, the only contribution to energy comes from even symmetric features. Image pixels are segmented using a square-error clustering algorithm.

In (Hofmann et al., 1998), the local description of texture and clustering is not performed in a point-wise manner but block-wise. The spatial domain is decomposed into small blocks. The texture features at each block are the histograms of a set of Gabor frequency features. For each frequency channel, its histogram at each pair

of blocks is compared by taking the maximum difference among energy bins. The distance between two block is the maximum histogram difference among frequency features. Block clustering is made by deterministic annealing.

In (Malik et al., 1999), descriptors for both texture and contours are combined. There, textons are redefined as frequently concurrent combinations of frequency features, represented by the outputs to a bank of Gaussian derivative filters –again, only even symmetric features are considered for texture description. Texture is described at each spatial position by measuring texton histograms in a neighborhood of that position. Contours are described by oriented quadrature filter pairs outputs, similar to those defined in (Perona and Malik, 1990). The distance between two spatial positions represents the chi-square distance between texton histograms weighted by a measure that decreases with the presence of a strong contour. Clustering is accomplished by the Normalized Cut algorithm (Shi and Malik, 1997).

One of the few methods developed for the treatment of 3D data $-2D+t$, indeed – is described in (Nestares et al., 2000). It is a technique for the segmentation of target object from background in noisy video sequences. Feature decomposition is accomplished using a bank of third order Gaussian derivative filters, which present even symmetry. The bank of filters decomposes the 3D frequency space, with two spatial frequency dimensions and one temporal frequency dimension, in a set of scale-orientation channels. Orientations with temporal-frequency component represent direction of movement. Local descriptions of texture and texture motion are based on the estimation of energy. Clustering is preceded by a feature selection stage, where the most representative features are selected. These are identified as features that present histograms with long tails, since a small number of high energy pixel reveals the presence of relevant information. Clustering is accomplished using the k-means algorithm.

In (Ontrup et al., 2004), the clustering process has the additional interest of being simulating the Gestalt laws of proximity and similarity. This is accomplished using a Competitive Layer Model to group image pixels, where the interaction function has two terms: one reflecting spatial proximity of the pixels and the other the similarity between their texture descriptors. The texture distance represents the pooling of differences between the components of a feature vector, which are basically the local mean and the local standard deviation of frequency feature energy.

2.4.4 Unsupervised Feature Clustering

A different perspective consists in assume that the various integral features present in an image have frequency contributions from different bands, so that isolating the main bands that differentiate each feature from the others leads to a decomposition of the image into those integral features. This would involve performing a clustering or classification, not over image pixels, but over the multiresolution frequency features. The set of filters that generates a clusters of frequency features can be considered a set of detectors for those composite features and the sum of the clustered features, i.e., the composite feature, can be thought of as the response to the cluster of filters.

This is the kind of analysis performed in the RGFF model, developed in (Rodríguez Sánchez et al., 1999). It defines visual patterns or integral features as patterns with *congruence in statistical structure*. Visual pattern detection is realized by identifying the frequency bands with alignment in various local statistics along wide frequency ranges. To isolate visual patterns, the RGFF first decomposes the image into elementary features using a bank of log Gabor filters tuned at different scales and orientations. Then, these components are grouped together by cluster analysis according to a measure of dissimilarity among them.

The distance measure used in the clustering algorithm reflects the degree of disparity between the vector of statistics describing each frequency feature, that may include one or more of the following: local phase, normalized local energy and its entropy, standard deviation and contrast. This set of properties is evaluated only at points that are candidates to belong to a relevant image feature: the local maxima of each energy map. These points are called *fixation points* or *attentional points*, for their use implies a simulation of the process of attention. They are identified as local energy maxima –contrast guided attention. First, a local distance measure is evaluated at each point as the weighted sum of the differences between each of the statistics. The total distance between two frequency features is the obtained by Quick pooling (Quick, 1974) of the contributions of each fixation point. As energy maxima do not coincide at the same locations of a pair frequency features, the dissimilarity is evaluated first at the maxima of one of them and then in the other. The final distance is the squared sum of both. Clusters are determined using a partitionial clustering algorithm

A 3D version $-2D+t-$ of this method has been developed in (Chamorro Martínez

et al., 2003) for the analysis of motion patterns from video sequences. The method is basically the same, except that the feature decomposition is achieved using a bank of 3D filters. Such filters are designed from separable components in spherical coordinates. The angular part of the filter represents both orientation in the static 2D plane and direction of motion.

2.5 Feature-Tuned Detectors

The approaches for feature binding cited in the previous section present some drawbacks. The main inconvenience is that they are limited to some given specific types of combination strategies. Region classification techniques are mostly limited to texture segmentation. Although contour information can be derived determining texture boundaries, it's clear that important low level cues of information are not taken into account. A further problem is the point-wise treatment of the image when clustering. This usually leads to multiple clusters for disconnected textures and lack of robustness. In the case of biologically feasible and composite learning approaches, only some specific types of hypercomplex cells are represented. It is known that the degree of specialization of those cells is high (Hubel, 1988). Therefore, such techniques also provide a partial representation of the image. In the case of composite-feature learning approaches, an additional drawback is that they learn receptive field functions using sets of training images, this is, prior information on the domain is needed.

In this work we pursue a completely data-driven and unsupervised method. This means that we seek to decide the appropriate feature detectors from the information implicit in the data, not in a predefined manner, and that no *a priori* information should be used. To our purpose, the most appropriate strategy is that of feature clustering, based on the concept of phase alignment. As aforementioned, it seems plausible that the HVS perceives features where Fourier components are locally in phase. Wherever a pair of bands are contributing to energy constructively – independently of the sign or particular value of phase – a feature is perceived. Our composite feature detector integrates information from bands which present a high degree of PC through the image. No *a priori* knowledge is employed. The combination of frequency features may produce textures, edges, blobs, junctions, etc. Unlike the HVS, here frequency features are treated as a whole, not locally at each cell's receptive field. Besides computational efficiency, considering the frequency features

globally provides robustness, specially in textured images, where textured regions do not appear spatially disconnected.

To represent PC, here we will use a measure of statistical dependence between each of frequency features, like correlation or mutual information. Alignment of energy maxima across scales must produce a high degree of redundancy between frequency bands contributing to it. Later on, we will study and compare several measures of dependence that could properly reflect the overall PC of two frequency features. This will help us to draw some conclusions about the type of statistical dependence between subband images caused by PC.

Although the RGFF model already performs feature clustering, the distance employed there lead to a quite different interpretation of the process. If the only statistic considered in the RGFF is local phase, and considering that the dissimilarity measure is evaluated only at points that are local energy maxima, this approximation can be considered to be performing some sort of overall phase congruency measurement. When using other statistics, the interpretation of the distance is not clear. The use of the normalized local energy could also be considered some kind of phase congruence estimation, since statistics differences are only evaluated at energy maxima location. However, the distance employed does not take into account the degree of coincidence of maxima, in which the notion of PC relies.

CHAPTER 3

Visual Pattern Detection

As aforementioned, the method for visual pattern extraction consists of the decomposition of the image in a set of frequency channels and their grouping according to some dissimilarity measure. The process can be described as a sequence of steps:

1. Selection of a subset of bands in the 3D filter bank with high information content, named *active* channels, by analyzing the spectral amplitude of the image.
2. Generation of frequency features using filters correspondent to active bands.
3. Measurement of dissimilarity between pairs of frequency features based on the normalized mutual information of their local energies.
4. Hierarchical clustering of frequency features based on the dissimilarity matrix.
5. Reconstruction of visual patterns correspondent to each cluster by linear summation of frequency features.

The whole process is illustrated in figure 3.1 using a 2D example image for the sake of clarity. In the next section we describe the filter bank employed for subband decomposition. The remainder sections describe in detail the different stages of processing involved in detecting the composite-features of a 3D image.

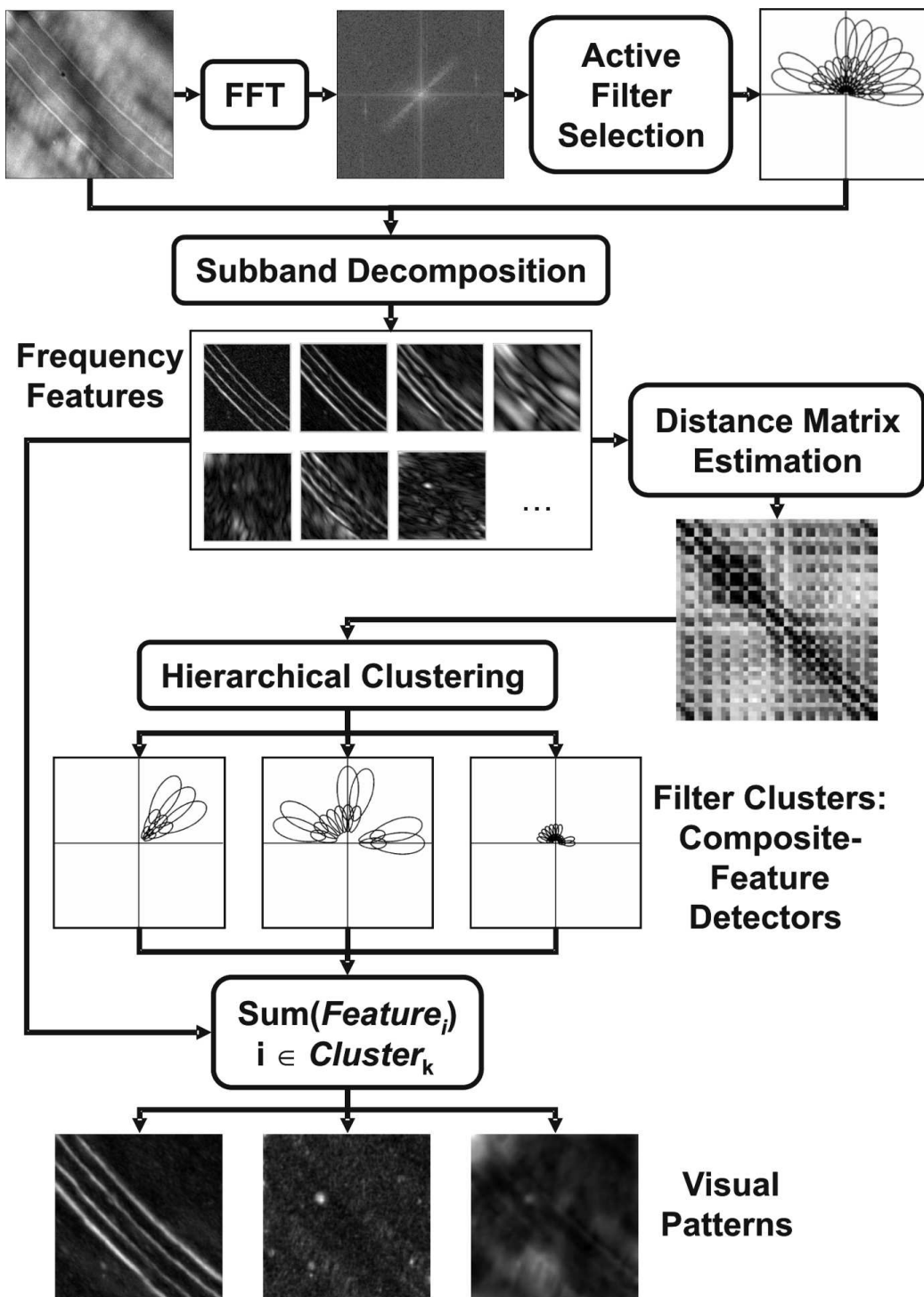


Figure 3.1: Flowchart of the visual pattern detection process.

3.1 3D Filter Bank

When implementing a filter bank, some decisions must be made regarding the following aspects: what is the most suitable number of channels, which filter type to use to model them and which portion of the spectrum should each channel capture. This last issue refers to the values of the central frequency and bandwidth parameters of the filters. In this section the specific spectral information coding chosen for our application is presented. It begins with the definition of the 3D log Gabor filters in spherical space-frequency coordinates, built up as the product of two separable factors, one standing for the radial component and the other for the angular part. Then, the partition scheme for the frequency space is depicted.

3.1.1 3D Log Gabor Channels

As introduced in section 2.3.1, one of the functions proposed to model energy filtering in the HVS is the log Gabor function. As was introduced in section 2.3.1, such function fit the response profiles of complex cells in the striate visual cortex and properly capture the statistical properties of natural images. Based on the log Gabor filter, here we design a 3D frequency channel with orientation selectivity. Its transfer function T is designed as the product of separable factors R and S in the radial and angular dimensions respectively, such that $T = R \cdot S$. The radial term R is given by the log Gabor function, in equation 2.8. It represents a Gaussian function in a logarithmic scale of spatial frequencies, where ρ_i is the central radial frequency and σ_ρ is the standard deviation. Its shape in a normal scale can be seen in figure 2.2.

The use of separable terms in polar coordinates is a common strategy to extend a filter from one to two dimensions providing it with orientation selectivity (Granlund and Knutsson, 1995). To extend the log Gabor filter to two dimensions, the angular component is usually represented by a scattering function centered at the filter's direction, like for example a Gaussian function (Field, 1993; Kovessi, 1996). The spread of the Gaussian controls orientation selectivity. The so-defined 2D log Gabor filter is considered well suited to represent the behavior of the mammalian visual cortical cells (Field, 1993).

To extend the log Gabor filter to 3D a second angular frequency coordinate must be taken into account. In (Chamorro Martínez et al., 2003), the 2D log Gabor filter

is extended to 3D for its application to 3D data from video sequences with the purpose of low level motion pattern detection. This extension –similar to that in (Yu et al., 2003)– is accomplished by adding a third term dependent on the elevation component of frequency. Then, the angular part of the filter S is the product of two Gaussians in each of the spherical angular coordinates, azimuth ϕ and elevation θ

$$G(\phi, \theta; \phi_i, \theta_i) = \exp \left(-\frac{(\phi - \phi_i)^2}{2\sigma_\phi^2} - \frac{(\theta - \theta_i)^2}{2\sigma_\theta^2} \right)$$

where (ϕ_i, θ_i) is the filter orientation and σ_ϕ and σ_θ are the standard deviations. The drawback of this approach is that the so-defined filters lack of rotational symmetry. Considering $\sigma_\phi = \sigma_\theta$, this expression represents an isotropic attenuation regards the central direction, as long as ϕ and θ are treated as Cartesian coordinates –a flat domain–. However, they actually are spherical coordinates, so that S is projected over the sphere and the Gaussian is no longer isotropic.

A rotational symmetric orientation signature was proposed in (Andersson, 1992; Granlund and Knutsson, 1995) as the cosine of the angle α formed by the directions of the filter and the position vector of a given point $\mathbf{f}$ in the spectral domain, risen to a power $l \geq 2$, with $l \in \mathbb{Z}$. A better choice, is to define S as a Gaussian on the angular distance α , so that the angular bandwidth is easily controlled by the standard deviation of the Gaussian (Faas and van Vliet, 2003)

$$\alpha(\phi, \theta; \phi_i, \theta_i) = \arccos(\mathbf{f} \cdot \mathbf{v} / \|\mathbf{f}\|)$$

where $\mathbf{v} = (\cos(\phi_i) \cos(\theta_i), \cos(\phi_i) \sin(\theta_i), \sin(\theta_i))$ is a unit vector in the filter's direction and $\mathbf{f}$ is expressed in Cartesian coordinates. Then, for a given angular standard deviation σ_α

$$S(\phi, \theta; \phi_i, \theta_i) = G(\alpha(\phi, \theta; \phi_i, \theta_i)) = \exp \left(-\frac{\alpha^2}{2\sigma_\alpha^2} \right).$$

The shapes of $G(\phi, \theta; \phi_i, \theta_i)$ and $S(\alpha)$, are depicted in figures 3.2 and 3.3. It can be seen that $G(\phi, \theta; \phi_i, \theta_i)$ has rotational symmetry when represented in the plane, but not in the sphere, while $S(\alpha)$ takes the same value for points at the same angular distance from the filters direction.

As aforementioned, the log Gabor filter does not have an analytic expression in the spatial domain. It's impulse response, illustrated in figure 3.4, can be numerically estimated as the inverse discrete Fourier transform from the previously specified

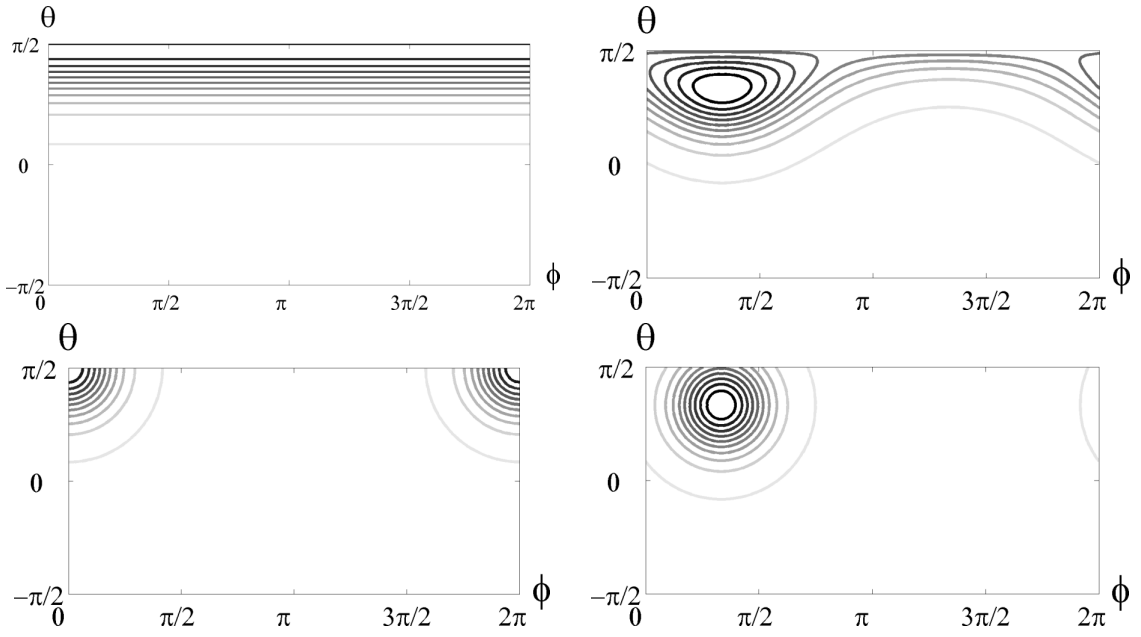


Figure 3.2: Representation of the angular component of the filter in the (ϕ, θ) plane as iso-level curves. *Left:* $G(\phi, \theta; \phi_i, \theta_i)$. *Right:* $S(\alpha(\phi, \theta; \phi_i, \theta_i))$ *Top:* $(\phi_i = \pi/3, \theta_i = \pi/4)$. *Bottom:* $(\phi_i = 0, \theta_i = \pi/2)$

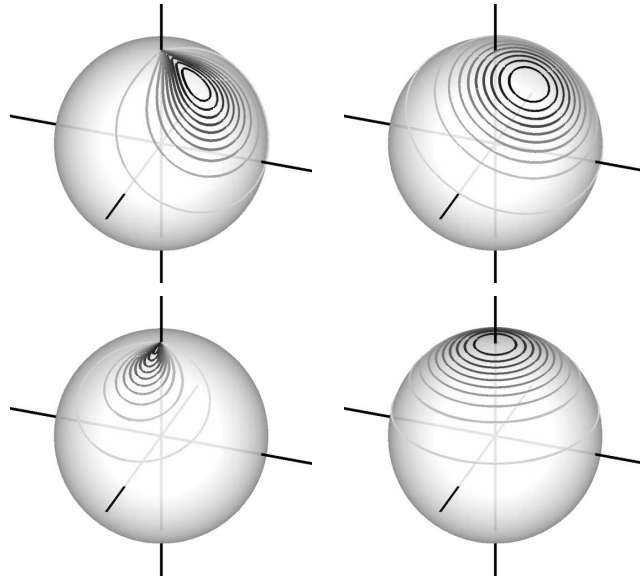


Figure 3.3: Representation of the angular component of the filter over a unit radius sphere as iso-level curves. *Left:* $G(\phi, \theta; \phi_i, \theta_i)$. *Right:* $S(\alpha(\phi, \theta; \phi_i, \theta_i))$ *Top:* $(\phi_i = \pi/3, \theta_i = \pi/4)$. *Bottom:* $(\phi_i = 0, \theta_i = \pi/2)$

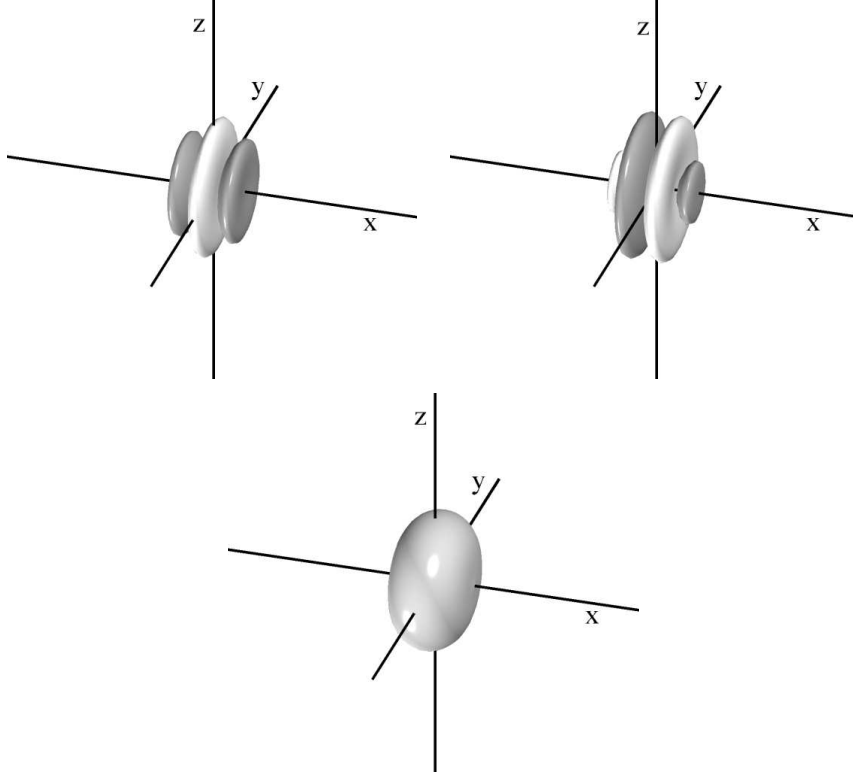


Figure 3.4: Spatial domain representation of a 3D log Gabor filter with $\lambda_i = 16$ pixels, $\phi_i = 0$ and $\theta_i = 0$. Each function h is represented by the isosurfaces with levels $\pm 0.05 \cdot \max(h)$, positive in white color and negative in gray color. *Top Left:* Real component. *Top Right:* Imaginary component. *Bottom:* Amplitude.

transfer function. Here, image filtering is performed in the Fourier domain. The spectral representation of the image is first obtained by means of the Fast Fourier Transform (FFT). The convolution in the spatial domain is equivalent to the product in the spatial frequency domain. Hence, the subband image is obtained as the inverse FFT of the product of the image spectrum and the filter's transfer function.

3.1.2 Frequency Domain Sampling

The complete 3D bank is composed of a number of the above described filters to tile the frequency domain, selecting a number of wavelengths and orientations and tuning the bandwidths to cover the spectrum properly. To select the set of orientations of the bank, the usual solution is to sample the angular coordinates

uniformly (Yu et al., 2003). However, this sampling is uniform in the Cartesian domain, but not in a spherical domain, producing an over-representation of the region near $\theta = \pm\pi/2$. An alternative solution is to choose the directions of vectors pointing at the vertices of regular polyhedrons centered at the origin (Granlund and Knutsson, 1995) or some other fixed distribution (Chamorro Martínez et al., 2003), but this solution lacks of flexibility in the selection of the number of filters. The solution in (Faas and van Vliet, 2003) consists of firstly defining a rough sampling of the orientation space, as the vertices of an icosahedron, secondly imposing an hexagonal grid on each of its faces and, finally, projecting the grid onto the unit sphere to obtain the orientations. The number of orientations is controlled by imposing a finer or coarser grid. Still, this approach does not permit enough control over the orientations on the equator of the sphere, this is, the orientations that fall in the xy plane. This is important because many medical imaging modalities are axial and the spatial resolution in the axial direction is usually lower than the resolution in the slices. Therefore, orientation in the xy plane must be given special attention.

In our configuration, elevation is uniformly sampled while the number of azimuth values decreases with elevation in order to keep the *density* of filters constant. This is achieved by maintaining equal arc-length between adjacent azimuth values over the unit radius sphere instead of taking uniform angular distances. Then, if the selected arc length is s , the number of azimuths for a given elevation is L/s , where L is the length of the parallel $\{\rho = 1, \theta = \theta_i, \phi \in [0, 2\pi]\}$. The arc-length s is determined from the number of azimuths N_ϕ in the curve $\{\rho = 1, \theta_i = 0, \phi \in [0, \pi]\}$ as $s = N_\phi/\pi$. For example, setting $N_\phi = 6$ and number of elevations $N_\theta = 4$ in the interval $[0, \pi/2]$ produces a total number of orientations of $N_o = 23$. It has to be remarked that only positive elevations and half equator are sampled to take advantage of symmetry. The filter bank configurations resulting from uniform and non-uniform sampling of the angular frequency space are shown in figure 3.5. The radial frequency dimension is tiled by octave bands, centered at wavelengths $\lambda_i = \lambda_{\min} \cdot 2^{i-1}$, with $i = 1, \dots, N_s$ and N_s being the number of scales. The radial frequency bandwidth is set to 2 octaves, yielding a highly redundant bank.

Small images must be given an especial treatment. It may happen that some of the bands in the bank have wavelengths larger than half the image size in the filter's direction. In these cases, filter responses approximately represent the average intensity level. Therefore, these bands are discarded and only the highest frequencies are considered.

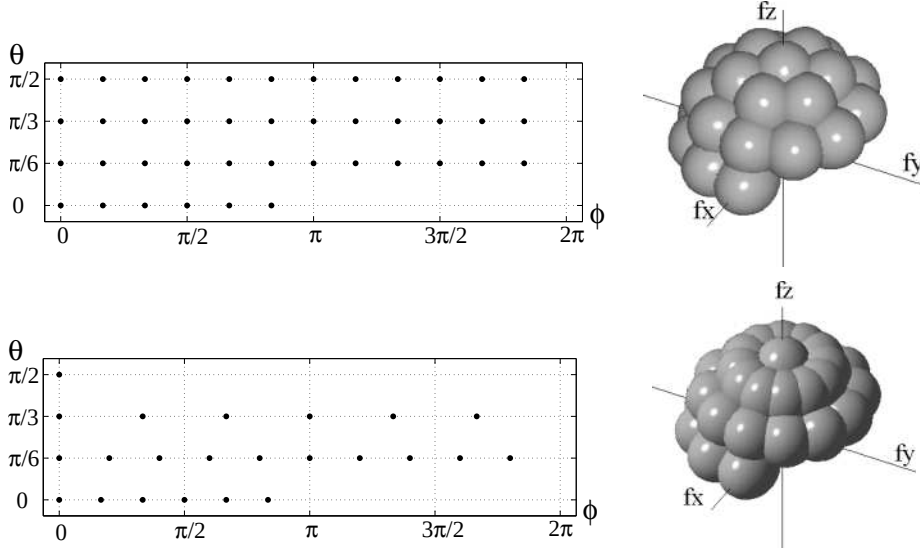


Figure 3.5: *Right:* Sampling of the orientation space for $N_\phi = 6$. *Left:* Correspondent filter banks for $N_\phi = 6$ and $N_s = 4$, represented as the iso-surfaces of the filters' transfer functions verifying $T_i = \exp(-1/2)$. *Top:* Uniform orientation sampling. *Bottom:* Non-uniform orientation sampling.

3.1.3 Filter Bandwidths

To set constant relative width in the radial part of the filters, the ratio between the radial transfer function standard deviation and the central frequency is fixed in equation 2.8. This is a common approach since its introduction by Gabor (1946), which ensures that the different channels provide roughly equal information. In a Gabor filter, the ratio between the standard deviation and the center frequency, σ_ρ/ρ_i , is determined from the chosen bandwidth in octaves Bw .

$$\sigma_\rho^g = \frac{\rho_i}{\sqrt{2 \ln 2}} \frac{2^{Bw} - 1}{2^{Bw} + 1} \quad (3.1)$$

In our experiments Bw is usually 2 octaves.

To obtain the ratio σ_ρ/ρ_i for the log Gabor function, the following expression is applied (Kovesi, 1996)

$$\sigma_\rho/\rho_i = \exp(-\sigma_\rho^g/\rho_i) \quad (3.2)$$

The width of the transfer function angular part can be determined by either controlling the aspect ratio of the filter (Daugman, 1985) or the desired overlapping

degree in the angular dimension (Kovesi, 1996). This is the solution adopted here. The angular separation between neighboring orientations in the xy plane –without considering the elevation– for a total of N_ϕ orientations in that plane is $\Delta\phi = \pi/N$. Fixing $\Delta\phi/\sigma_\alpha = 1$ ensures that the iso-set of the angular transfer function S with level $\exp(-1/2)$ covers as much spectral area as possible without overlapping its neighbors. In our implementation, we have taken $\Delta\phi/\sigma_\alpha = 1.2$ and $N = 6$, which leads to $\sigma_\alpha = 25^\circ$.

3.2 Selection of Active Bands

In the most general case, the filter bank consists of 92 filters. Their responses to the input data must be obtained and their pairwise dissimilarities computed. To decrease the high computational cost of such tasks, the number of filters can be reduced. Firstly, filters with wavelengths greater than half the image size are discarded, as they approximately represent the average intensity level. Secondly, filters with low information content can be eliminated. To this end, the magnitude $\hat{F} = \log(|\mathcal{F}(X)| + 1)$ is used as a measure of the information associated to a channel, with $|\mathcal{F}(X)|$ being the amplitude of the Fourier transform of the image.

The energy density map $\hat{F}$ reflects the preferred orientations and frequency bands plus a background of spectral noise. A 2D example can be seen in figure 3.6.b obtained from the image in figure 3.6.a. In (Rodríguez Sánchez et al., 1999), the classification of the bands into active and non-active is accomplished by applying cluster analysis to the total energy comprised by each channel. The problem with this solution is that the energy values of the activated filters range a wide interval, following a geometric decay as frequency increases (Field, 1993). This makes the definition of a class of activated patterns very difficult.

A simpler solution has been chosen here, involving the characterization, detection and elimination of non active channels, instead of active ones, through statistical methods. Non active bands are associated to the background noise. A band is active if it comprises any energy value over the estimated maximum level of the spectral noise. The maximum noise level is estimated as $m + x\sigma$, where m is the mean noise energy, s is its standard deviation and $x > 0$. Here, $x = 3$ has been considered enough to eliminate the main noise contribution and m and s have been measured in a band of high frequencies. The cutoff frequency of such band is set to $2 \max(\rho_i)$

with ρ_i being the center frequencies of the filter's in the bank.

Still, subtracting the maximum noise level from the energy map is not sufficient to *clean* the background. Some spurious values remain, as shown in figure 3.6.c. To eliminate them, some kind of filtering can be applied that takes into account the characteristics of spectral energy maps. Under our working hypothesis, relevant features in the spatial domain are translated into the spatial-frequency domain to alignment through scales. Therefore this kind of structures must be preserved in the filtering process. Standard median filters produce the elimination of thin lineal structures. To avoid this, we have designed a radial median operator. The difference with an ordinary median filter is that, given a certain pixel in the energy map, it only considers neighbors that are anterior or posterior in the radial direction to calculate the median. This eliminates isolated peaks but preserving the continuity of structures along scales. The expression for the radial median filter mask M with size $l \times l \times l$ is as follows

$$M(p, q) = \begin{cases} 1 & \text{if } [(q - p)/\|q - p\|] = [p/\|p\|] \\ 0 & \text{otherwise} \end{cases},$$

where p and q are points in the spectrum and mask domains respectively, $[\cdot]$ represents rounding to the nearest integer and the origin is set at the image center. In this work the mask size is fixed to $l = 3$ pixels. The behavior of this filtering is illustrated in figure 3.6.d. Figure 3.6.e shows the bands comprising non-null values after radial median filtering and 3.6.f their associated filters.

3.3 Dissimilarity Measure Between Energy Maps

We have so far defined visual patterns as composite frequency features whose elementary components present a high degree of phase alignment. As cited in chapter 2, the phase congruence of two subband images is directly related to the degree of alignment between the structure of their local maxima maps. Therefore, a measure that reflects the statistical dependence between the local energy of two subband images could be a good dissimilarity measure for the clustering of frequency features. One such measure is the mutual information, which has been successfully used for image comparison in various fields of application, like image registration (Viola, 1995; Studholme et al., 1999; Butz and Thiran, 2001) or image retrieval (Rubner et al., 2001). Here, we present a measure of local subband energy dissimilarity based

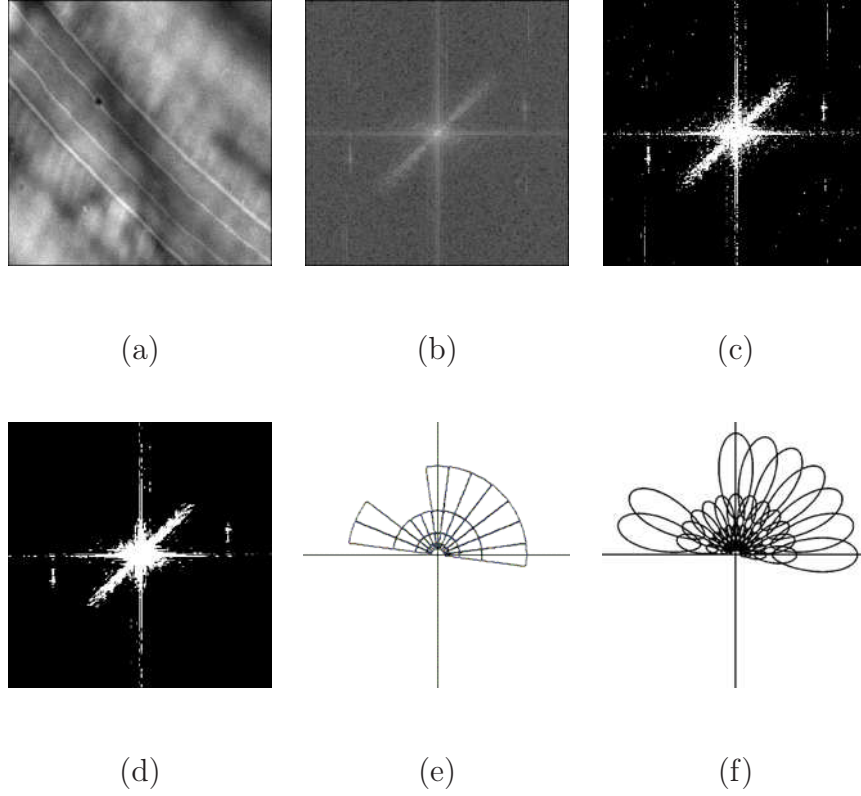


Figure 3.6: 2D example of radial median filtering using an image of a nematode. (a) Input image of a nematode. (b) $\hat{F} = \log(|\mathcal{F}(X)| + 1)$, where $|\mathcal{F}(X)|$ is the spectral energy. (c) Noise subtraction on $\hat{F}$, non null values depicted in white. (d) Radial median filtering, non null values depicted in white. (e) Bands comprising non null values. (f) Associated active filters.

on their mutual information. In chapter 4 we will present alternative measures for the estimation of frequency feature dissimilarity and we will study their properties and suitability. Before presenting normalized mutual information distance, we will introduce some concepts and fundamentals of information theory.

3.3.1 Information Theoretical Concepts

Here, we will get into the questions of how a random variable and the relation between random variables can be characterized (Papoulis, 1991). We will employ the same symbols to represent the random variables than for any instantiation of them. For example, we will use the same symbol X for the variable representing the intensity values of an image and for a specific image.

A variable X is called *random* when the set of allowed values Ω_X , usually named *events*, is known but the specific value of a future sample cannot be predicted with certainty. The *probability* P of an event $x_i \in \Omega_X$ is the proportion of times that x_i is expected to occur in a sufficiently large set of samples. A variable X is not random when there exists some $x_i \in \Omega_X$ verifying $P(x_i) = 1$, with $P(X) = 0$ for any other event.

In some situations the probability distribution of a variable can be known a priori, like in the case of the result of throwing a dice. If this is not possible, it can be estimated from the event's *frequencies* from previous samples. The *frequency* of an event is the proportion of times it has happened in previous samples, i.e., the normalized count of that event. The vector $f(X)$ of frequencies for each $x_i \in \Omega_X$ is the *histogram* of the set of samples. When dealing with continuous variables, the domain of allowed events Ω_X is divided into a set M intervals called *bins*. The histogram is calculated as a vector whose components are the frequencies of all the events within each bin.

The probability distribution or its estimation by means of the histogram completely characterizes a random variable. Now, we are interested in characterizing the relation existing between two random variables. This can be done by determining the probability of concurrence of pairs of values of the two variables. This is called the *joint probability distribution* $P(X, Y)$. Again, this probability can be estimated from a sample set by means of the *joint histogram* $f(X, Y)$. It is a matrix whose elements are the frequencies of each pair of events (x_i, y_j) with $x_i \in \Omega_X$ and $y_j \in \Omega_Y$.

From the joint distribution one can compute the *marginal probability distributions* $P(X)$ and $P(Y)$ as follows.

$$P(X) = \sum_{y_j \in \Omega_Y} P(X, Y = y_j)$$

and analogously for $P(Y)$.

When two random variables are related to each other they are said to be *dependent*. In that case, a variable X can be predicted from the other variable Y with a given probability $P(X|Y)$, named *conditional probability distribution*. $P(X = x_i|Y = y_j)$ tells us the probability that X takes value x_i if we know that variable Y takes value y_j . The joint and conditional probability distributions are related by the following expressions.

$$P(X, Y) = P(X|Y) \cdot P(Y) = P(Y|X) \cdot P(X)$$

From these expressions one can derive the Bayes' rule to derive the conditional probability distribution $P(Y|X)$ from $P(X|Y)$

$$P(Y|X) = P(X|Y) \frac{P(Y)}{P(X)}$$

When there is a functional dependence between two random variables, this is, $X = F(Y)$, X can be certainly predicted from Y . Then, $P(X = F(Y)|Y) = 1$ and $P(X = F(Y), Y) = P(Y)$, with both probabilities being zero when $X \neq F(Y)$. On the other hand, two random variables are *independent* then $P(X|Y) = P(X)$, i.e., the probability of X depends only on X , and their joint probability is the product of their marginal probabilities $P(X, Y) = P(X) \cdot P(Y)$. It is also possible that two random variables are related to some point but not certainly predicted from each other. Then, the type of dependence is not functional but it presents some degree of randomness.

3.3.2 Entropy and Mutual Information

So far, we have defined the conditions under which a variable X and a relation between two variables X and Y are random. Now, we are interested in quantifying the degree of randomness of a variable and the degree of randomness of the relation between two variables. To this end we must introduce the concept of entropy. Note

that, from now on, we will use a simplified notation to express the probability of an event as $P_X(i) = P(X = x_i)$. We will also omit to specify the domain of event indices. We will assume that i in $P_X(i)$ is such that $x_i \in \Omega_X$.

Shannon's entropy is a statistic that summarizes randomness. It is defined, for discrete random variables, as the negative expectation of the log probability

$$H(X) = -\mathbb{E}_X[\log(P(X))] = -\sum_i \log(P_X(i))P_X(i),$$

assuming that $0 \log(0) = 0$.

It can be easily proven that, if X is not random, then $H(X) = 0$. On the other hand, when all events are equally probable, H takes its maximum, which is $\log(N)$, with N being the size of the sample set. Using base 2 in the logarithm, H is measured in *bit* units. An interesting interpretation of entropy was posed by Shannon, who demonstrated that, on average, the minimum number of bits to code a variable X is equal to $H(X) + 1$. That is why entropy is often called the *information* content of a variable.

To measure the randomness of the relation between two variables, we can define a measure of entropy for conditional and joint probabilities. Thus, the *conditional entropy* is the measure of the randomness of X given knowledge of Y and is defined as the expectation of the log conditional probability.

$$H(X|Y) = -\mathbb{E}_X[\mathbb{E}_Y[\log(P(Y|X))]] = -\sum_i \sum_j P_{X,Y}(i|j) \log(P_{X,Y}(i|j)).$$

Analogously, the *joint entropy* is defined as the expectation of the joint probability distribution.

$$H(X, Y) = -\mathbb{E}_X[\mathbb{E}_Y[\log(P(Y, X))]] = -\sum_i \sum_j P_{X,Y}(i, j) \log(P_{X,Y}(i, j)).$$

The following identity can be easily derived from the relations in previous section.

$$H(X, Y) = H(X|Y) + H(Y).$$

It is also straightforward to derive the entropy conditions for functional dependent and independent variables. In the case that X and Y are related as $Y = F(X)$, then $H(X|Y) = 0$, while in the case of independence $H(X|Y) = H(X)$.

Another well-known measure of dependence between two variables is the Mutual Information I . It represents the reduction of uncertainty of the first variable brought

by the knowledge of the second variable. It is defined from marginal and conditional entropies by the following equivalent expressions.

$$I(X, Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) = I(Y, X)$$

It can also be expressed by means of the joint entropy as follows.

$$I(X, Y) = H(X) + H(Y) - H(X, Y)$$

Its expression for discrete variables as a function of joint and marginal probabilities is the following.

$$I(X, Y) = \sum_{i,j} P_{X,Y}(i, j) \log \frac{P_{X,Y}(i, j)}{P_X(i)P_Y(j)},$$

In this case, the condition for total independence of the two variables is $I(X, Y) = 0$, while exact functional dependence is produced when $I(X, Y) = H(X) = H(Y)$, provided that the marginal probability distributions are identical for $Y = F(X)$.

3.3.3 Normalized Measures of Mutual Information

Conditional and joint entropies of two variables can be low either because they are strongly dependent or because they are little random. Therefore, they cannot be considered good indicators for statistical dependence. The same happens with mutual information. It can be low either because the two variables present a very weak relation or because the marginal entropies are small.

Studholme proposed a normalized measure of statistical dependence for image distance estimation in the field of medical image registration (Studholme et al., 1999). Although Studholme called this measure the Normalized Mutual Information, it has been also referred to as the Normalized Entropy (NE). Here, we will use this last name for Studholme's measure to distinguish from our proposal for the normalization of mutual information.

$$NE(X, Y) = \frac{H(Y) + H(X)}{H(X, Y)} = 1 + \frac{I(X, Y)}{H(X, Y)}$$

This measure is independent of the marginal entropies of the two compared images, so that it measures uncertainty reduction regardless of the uncertainty amount itself. It ranges from 1, for the case of total independence, to 2, representing exact functional dependence.

The similarity measure employed here to represent the statistical dependence between two subband images, which we will call the Normalized Mutual Information (NI), is an alternative form of normalization of mutual information.

$$NI(X, Y) = \frac{2 \cdot I(X, Y)}{H(Y) + H(X)}.$$

It is very similar to that presented in (Strehl, 2002), except from that we normalize by the sum of the marginal probabilities of the two frequency features involved instead of the maximum of all possible pairs of maps. It is related to the NE as follows.

$$NI(X, Y) = 2 \cdot \left(1 - \frac{1}{NE(X, Y)}\right).$$

NI is a similarity measure with values in the interval $[0, 1]$. To transform it into a distance, its range must be inverted. To this end, the following transformation is applied, which also increases large distances and reduces small ones, enhancing the compactness and separation of the clusters. Therefore, the actual distance employed is the following.

$$D_{NI}(X, Y) = \left(1 - \sqrt{NI(X, Y)}\right)^2. \quad (3.3)$$

This measure has a strong dependence on the bin size chosen in the histogram calculations needed to estimate the probability densities. It has been observed that a small number of bins, of $O(16)$ produce the best results and incidentally speed up the calculations.

The values for the diverse measures of dependence presented are shown in table 3.1 for the cases of total independence and exact functional dependence. It can be seen that NE , NI and D_{NI} are independent on the marginal entropies.

3.4 Feature Clustering

The choice on the clustering criterion should depend on the statistical structure of the data. However, it is very difficult to know a priori the underlying structure of a data set when one can not visualize the patterns as points in a low dimensional space or when the distance between patterns has not a clear geometrical interpretation. This is the case with frequency features and the NI distance.

Measure	Independence	Functional Dependence
$H(X Y)$	$H(X)$	0
$H(X, Y)$	$H(X) + H(Y)$	$H(Y) = H(X)$
$I(X, Y)$	0	$H(X) = H(Y)$
$NE(X, Y)$	1	2
$NI(X, Y)$	0	1
$D_{NI}(X, I)$	1	0

Table 3.1: Values of several dependence measures in the cases of total independence and exact functional dependence

A widely applied criterion to define clusters in diverse algorithms is minimum variance by minimization of the square error. The typical clustering method fulfilling such criterion is the k-means algorithm. It is a partitional method that iteratively moves observations from one cluster to another in order to minimize the average distance to the cluster centroid. Methods of this kind deal with the input data themselves, not only the distance matrix, since centroids and dispersions must be calculated in each iteration. In the case of frequency features, they would be considered as observations in a feature space with as many dimensions as the image size. At each iteration, the centroids must be recomputed as the mean vector of all observations in a given cluster. The distances from each observation to the cluster's centroid have also to be recomputed at each stage using the specified metric. The mean pattern the optimal representative of the cluster elements if the Euclidean distance properly reflects the dissimilarity between patterns, in which case it is an estimation of the cluster's geometrical center. However, this may not always be the case.

Therefore we are interested in using a clustering method that uses only information from the dissimilarity measure chosen to represent feature disparity. Here, the method employed for clustering of frequency features is hierarchical clustering. A hierarchical clustering method is a procedure to transform a distance matrix into a sequence of *nested partitions* of the data. A *partition* of a data set is a collection of non-overlapping subsets that comprises all the data. A partition $\mathcal{A}$ is *nested* into a partition $\mathcal{B}$ if every component of $\mathcal{A}$ is a subset of a component of $\mathcal{B}$. Hierarchical clustering methods generate a sequence of partitions in which each partition is nested into the next partition in the sequence. Thus, it produces a hierarchy of clusterings ranging from the completely disjoint partition, in which there is one

cluster for each single datum, to a partition with one single cluster comprising the whole data set.

One interesting characteristic of hierarchical methods is that the sequence of partitions is obtained using only the information provided by the distance matrix, following some given strategy to decide how to link nested clusters in consecutive levels of the hierarchy. These properties make this method very interesting to our application. On the one hand, one must not deal with the features themselves as vectors in a multidimensional space –with as much dimensions as the image size. On the other hand, once the *linkage* strategy is decided, only N_f partitions are possible, where N_f is the number of data, so that only N_f partitions are to be compared in order to select the best.

The algorithm employed to produce the hierarchical clustering is the Matrix Updating Algorithm (King, 1967; Johnson, 1967; Jain and Dubes, 1988). It is an agglomerative algorithm, this is, it starts with the completely disjoint partition and, at each new iteration, it decides which clusters from the old partitions are to be merged to generate the new partition. The two clusters merged are those with minimal inter-cluster distance $\mathcal{D}$. The method for estimating the inter-cluster distances from the pair-wise distances between their constituent elements determines the linkage strategy. The general algorithm, valid for any linkage strategy, is described in the next sequence of steps.

1. Begin with the disjoint partition, where each pattern is a cluster. Set sequence number $m \leftarrow 0$. Initialize the distance matrix $\mathcal{D}$ with the pair-wise distances between patterns.
2. Find the least dissimilar pair of clusters (r) and (s) by taking the minimum in the distance matrix $\mathcal{D}$.
3. Merge clusters (r) and (s) into a single cluster to form the next partition. Increment sequence number $m \leftarrow m + 1$.
4. Update the distance matrix $\mathcal{D}$ by deleting the rows and columns corresponding to clusters (r) and (s) and adding one row and one column for the newly formed cluster (r, s) . The distance between the new cluster (r, s) and a given old cluster (k) has the following general expression.

$$\begin{aligned} \mathcal{D}[(k), (r, s)] = & \alpha_r \mathcal{D}(k, r) + \alpha_s \mathcal{D}(k, s) + \beta \mathcal{D}(r, s) + \\ & \gamma |\mathcal{D}(k, r) - \mathcal{D}(k, s)|. \end{aligned} \quad (3.4)$$

5. If all objects are in one cluster, stop. Else, go to step 2.

The specific form of $\mathcal{D}[(k), (r, s)]$ determines the linkage strategy, this is, the rule to decide when two clusters must be merged. The most common linkage strategies are single-link and complete-link. Single-link consists in merging at each stage the two clusters that have the minimal distance between their nearest elements. In the matrix updating algorithm, single-link is equivalent to define $\mathcal{D}[(k), (r, s)] = \min(\mathcal{D}(k, r), \mathcal{D}(k, s))$. This is achieved by setting $(\alpha_r, \alpha_s, \beta, \gamma) = (1/2, 1/2, 0, -1/2)$. Complete-link consists in merging at each stage the two clusters that have the minimal distance between their farthest elements. In the matrix updating algorithm, complete-link is equivalent to define $\mathcal{D}[(k), (r, s)] = \max(\mathcal{D}(k, r), \mathcal{D}(k, s))$. This is achieved by setting $(\alpha_r, \alpha_s, \beta, \gamma) = (1/2, 1/2, 0, 1/2)$.

Complete-link clustering ensures that the farthest elements of one clusters are as close as possible, thus produces very compact clusters. This is not guaranteed in single-link clustering, which tends to produce *straggly* clusters. Amid both extremes, other link strategies have been proposed (Jain and Dubes, 1988). Among them, Ward's linkage method, employed here, has the particular interest of producing minimum variance partitions. It merges, at each stage, the pair of clusters that minimizes the increment in the square error of the whole partition, measured as the sum of the clusters dispersions. Several comparative studies conclude that Ward's method, also called minimum variance method, outperforms other hierarchical clustering methods (Milligan and Shilling, 1985; Jain and Dubes, 1988). Ward's inter-cluster distance results from taking the following coefficients in the general expression.

$$\alpha_r = \frac{n_r + n_k}{n_r + n_s + n_k}, \quad \alpha_s = \frac{n_s + n_k}{n_r + n_s + n_k}, \quad \beta = \frac{-n_k}{n_r + n_s + n_k}, \quad \gamma = 0 \quad (3.5)$$

Replacing these values in equation 3.4 gives rise to the next distance updating rule.

$$\mathcal{D}^{Ward}[(k), (r, s)] = \frac{(n_r + n_k)\mathcal{D}(k, r) + (n_s + n_k)\mathcal{D}(k, s) - n_k\mathcal{D}(r, s)}{n_r + n_s + n_k} \quad (3.6)$$

Figure 3.7 shows an example of clustering hierarchy produced using Ward's linkage. The data set corresponds to the set of subband images generated from the nematode image in figure 3.1. The top row of figure 3.7 shows the dissimilarity matrix for the subband energies of the nematode image using the D_{NI} distance. The plot presented in the center row of the figure is called *dendrogram* and represents

the hierarchy of clusters obtained from the dissimilarity matrix using the Ward's method. A *dendrogram* is a special type of tree structure, where leaves represent the data, nodes indicate the fusion of two clusters and branch heights are proportional to the distance between the two clusters merged. The heights of two merged branches are useful to visually give some indication of the degree of cohesion of the newly formed cluster. Selecting a given partition in the hierarchy is represented by cutting the dendrogram at that level. The resulting clusters are those merged at a level below the cutting line.

The question now is at which level to cut the dendrogram or, in other words, which is the optimal partition in the hierarchy. To this end, it is necessary to evaluate the quality of each resulting configuration according to a given validity index. Here, the criterion to determine the quality of a partition is related to the separation and cohesion of the clusters generation. Many validity indices have been proposed to reflect these properties, like for instance, the Davies-Boulding index V_{DB} (Davies and Boulding, 1979). Given a partition at a level with a number N_c of clusters, this index has the next expression.

$$V_{DB}(N_c) = \frac{1}{N_c} \sum_{r=1}^{N_c} \max_{s, s \neq r} \left\{ \frac{\sigma_r + \sigma_s}{S_{rs}} \right\} \quad (3.7)$$

where σ_r is a measure of the dispersion among the elements of cluster r and S_{rs} is a measure of the separation between clusters r and s .

As aforesaid, here we are interested in deriving the clustering result using only the matrix distance. The typical measures of dispersion of a set of patterns involve the measurement of their average distance to their mean. In the present implementation the intra-cluster dispersions have been estimated using a graph-theoretical measure introduced in (Pal and Biswas, 1997). It is computed as the maximum edge length in the Minimum Spanning Tree (MST) of the elements of a cluster. The MST is defined as the minimal possible subgraph connecting all the vertices. Let $E_r^{MST} = \{e_{ri}^{MST}\}$ be the set of edges of the MST computed on cluster r , with $1 \leq r \leq N_c$, $1 \leq i \leq |E_r|$, where $|\cdot|$ gives the number of elements. Then

$$\sigma_r^{MST} = \max_i (e_{ri}^{MST}, i = \{1, 2, \dots, |E_r|\}) \quad (3.8)$$

. Again, instead of calculating the separation between two clusters as the distance between their centroids, we take the mean between the pair-wise dissimilarities D

of their components.

$$S_{rs} = \frac{1}{|r||s|} \sum_{i \in r, j \in s} D_{i,j} \quad (3.9)$$

Then, the modified Davies-Boulding index V_{DB}^{MST} is defined as follows.

$$V_{DB}^{MST}(N_c) = \frac{1}{N_c} \sum_{r=1}^c \max_{s, s \neq r} \left\{ \frac{\sigma_r^{MST} + \sigma_s^{MST}}{S_{rs}} \right\} \quad (3.10)$$

As an example, the values of V_{DB}^{MST} for the hierarchy of partitions obtained for the frequency features of the nematode image are plotted at the bottom of figure 3.7. In this example the maximum value is obtained for $N_c = 3$. The partition with three clusters is presented in figure 3.8. The filters correspondent to each feature cluster and the visual pattern reconstructed from them are those presented in the flowchart in figure 3.1.

3.5 Examples

In this section, the behavior of the described method is illustrated with some examples. The cases have been designed to illustrate the performance of the chosen bank configuration, the capability of the system to identify and extract visual patterns of different types and its robustness to noise. All example data are digital images of size $64 \times 64 \times 64$ pixels and with 256 gray intensity levels. The method has been applied to these images using the filter bank settings detailed in section 3.1.2. Images D1, D2 and D3 have been processed using angular bandwidth of 20° while D2, D6 and D7 take 25° . We must warn that, since all example data are synthetic images, the radial median filters is not applied to avoid the elimination of important features. This is necessary due to the nature of synthetic images, in which a visual pattern can be composed of finite set of aligned harmonics, instead of a continuum of frequencies. If the radial median filter is applied, isolated pulses in the frequency domain are deleted.

In the first example, a comparison is made between the responses produced by the filters of our 3D bank and filters from the bank proposed in (Chamorro Martínez et al., 2003; Yu et al., 2003), this is, a bank with uniform angular frequency sampling and 3D log Gabor filters using equation 3.1. Both approaches are applied to the data set D1, in figure 3.9, which is a synthetic volume that contains two grating patterns

with orientations $(\phi = 0, \theta = \pi/2)$ and $(\phi = \pi, \theta = \pi/3)$ respectively. These orientations coincide with two of the bank's filters directions, one normal to the z axis and the other adjoining that one. Figure 3.10 shows two filters from the bank of uniformly distributed asymmetric filters, with elevation component $\theta_i = \pi/2$ and their responses to D1. It can be seen that the filter with $(\lambda_i = 4, \phi_i = 0, \theta_i = \pi/2)$ – where λ_i stands for center wavelength – offers a partial representation of the direction normal to the z plane. The lack of rotational symmetry and the over-representation of frequencies with high elevation component produce *orange slice* shaped filters. Actually, a filter with $\theta_i = \pi/2$ should not be selective in the azimuth component, but in this case six filters like the one in figure 3.10, sweeping the azimuth coordinate, would be necessary to completely represent the $\theta_i = \pi/2$ direction. This filter produces no response to the pattern with $(\phi = \pi, \theta = \pi/3)$, which is quite close to $(\phi = 0, \theta = \pi/2)$. On the other hand, the filter with $(\lambda_i = 4, \phi_i = \pi, \theta_i = \pi/2)$ produces a larger response to the grating with $\theta_i = \pi/3$ than to the one with $\theta_i = \pi/2$. If this analysis is repeated for filters from the bank proposed in this paper, one can see from figure 3.11 that both filters produce the same response, as it was expected. The response is non null for both patterns and it is stronger in the case of the pattern normal to the z axis than in the other case.

In the second example, in figure 3.12, the data set D2 is a synthetic volume that consists of three regions with different texture properties. In the most inner region, the main frequency components have orientation $(\phi = 3\pi/4, \theta = 0)$ while in its surrounding region the main orientations are $(0, 0)$, $(\pi/2, 0)$ and $(0, \pi/2)$. The texture of the most outer region is mainly originated by noise. The system groups the filters into three clusters, each correspondent to one of the three regions. In figure 3.13 the resulting clusters are visualized together with the correspondent visual patterns, represented as the sum of the energies of all filters in the cluster. It can be seen that filters in the first and second clusters comprise orientations similar to those composing the two most inner regions in D1, while the third cluster collects the high frequency bands of the remainder orientations that contribute only to noise.

The purpose of the third example is twofold: to demonstrate the robustness of the method and its ability to deal with different types of features. To this end, the data set D3 has been designed, which contains three types of visual patterns: texture, phase discontinuity and intensity level discontinuity, which in this case is a plane, this is, a symmetric edge. Subsequently, new images have been created from D3 by adding Gaussian noise of different standard deviations. Top row at figure 3.14

shows the volume D3 and the images at the top row of figure 3.15 show an example of noise corrupted data, named D4, obtained from D3 by adding Gaussian noise of standard deviation $\sigma = 25$. The results obtained are very similar for D3 and all the noisy versions of it. Figures 3.14 and 3.15 show the results for D3 and D4. In both cases, the method generates three clusters: one representing the texture pattern, one correspondent to the phase shift and one for the plane. It has to be pointed that in some cases, the method produces a fourth cluster containing low frequency components contributing to the background texture. Nevertheless, this fact does not depend on the noise standard deviation, but on the particular distribution of the noise signal.

The ability to group bands with different orientations that contribute to a single object's shape has special interest in many applications. In the fourth example, the data set D5, in figure, 3.16, contains a region with a characteristic shape immersed in a textured medium. The image has been corrupted with Gaussian noise of standard deviation $\sigma = 10$. The results obtained are shown in figure 3.17. As can be seen, the visual patterns with orientations predominant in the object's shape are isolated from those that mainly contribute to the background.

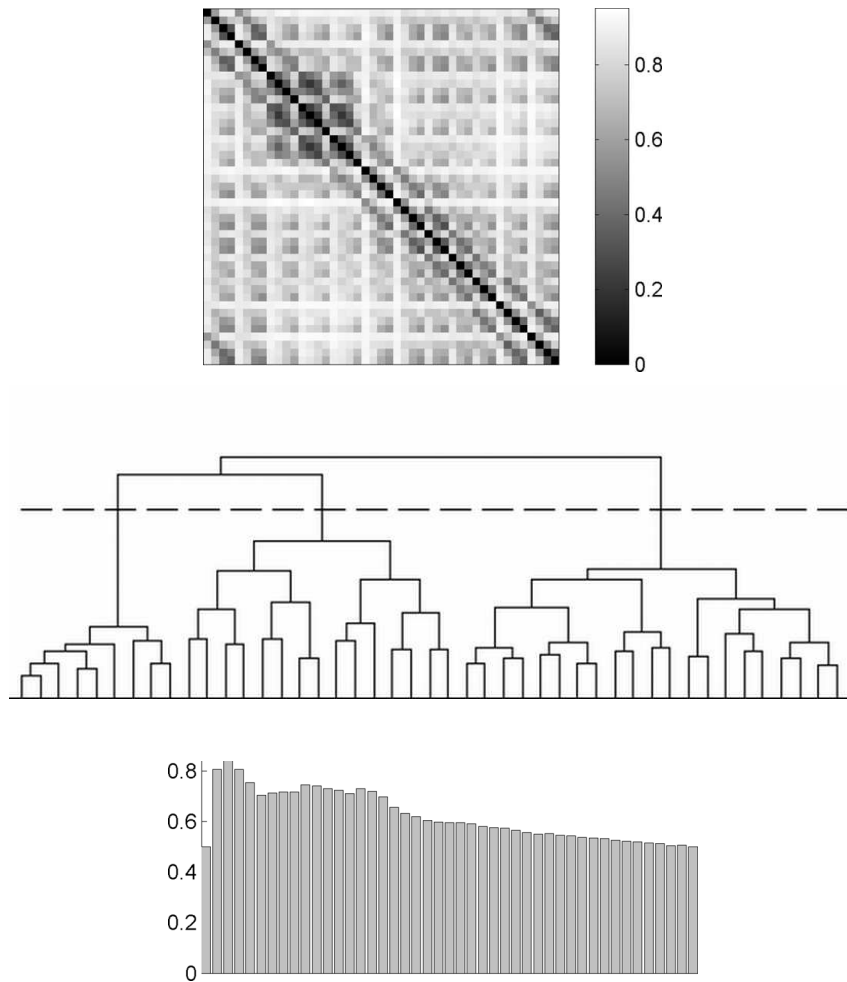


Figure 3.7: *Top:* Example of a dissimilarity matrix for the case of the nematode subbands in figure 3.1. *Center:* correspondent dendrogram obtained using Ward's method. Cutting the dendrogram at the level indicated with the dashed line produces three clusters. *Bottom:* V_{DB}^{MTS} for each partition on the previous hierarchy.

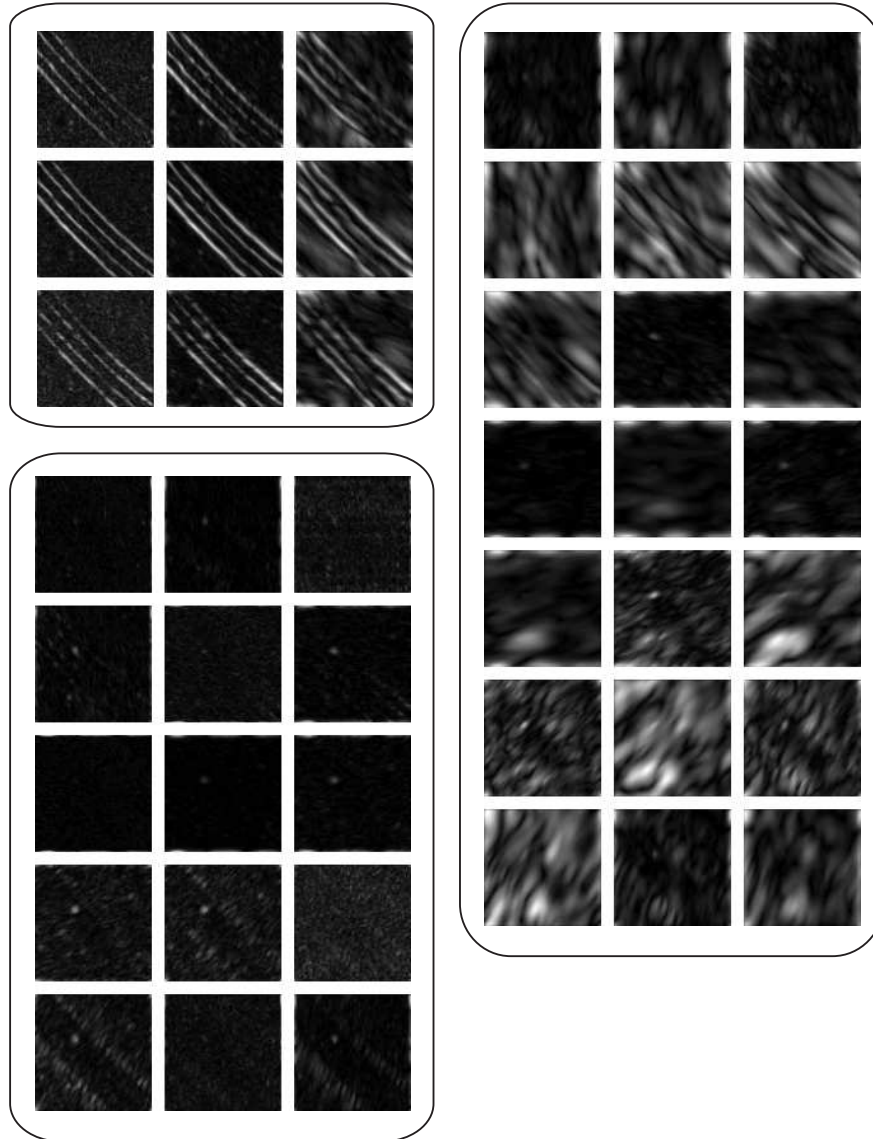


Figure 3.8: Nematode subband clusters resulting from applying Ward’s method for hierarchical clustering and MST-based David-Boulding index for partition validation. The sum of subbands in each cluster gives place to the three visual patterns at the bottom of flowchart in figure 3.1

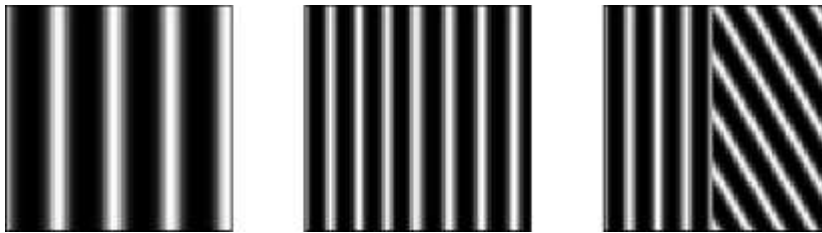


Figure 3.9: Synthetic test data set D1 represented by three cross sections showing slices normal to the image coordinate axis –the origin of coordinates is set in the image center–. It consists of two grating patterns, one with orientation $(\phi = 0, \theta = \pi/2)$ and the other with orientation $(\phi = \pi, \theta = \pi/3)$.

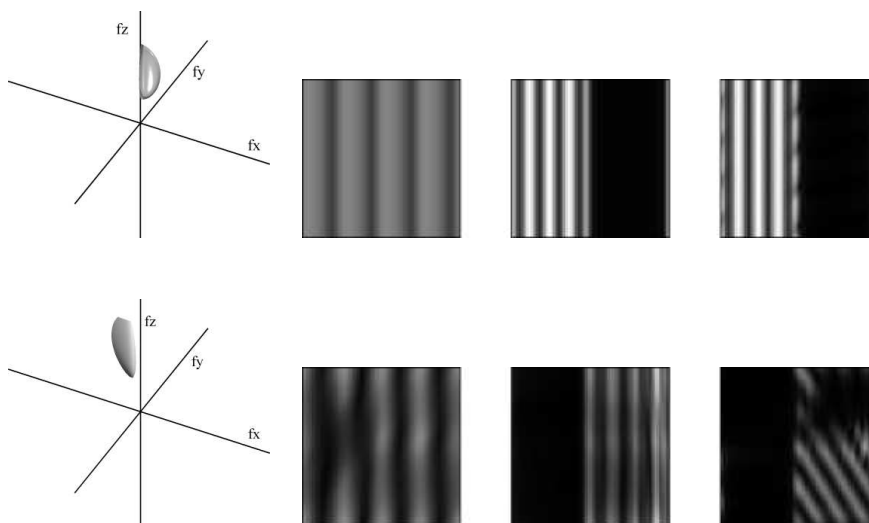


Figure 3.10: *Left column:* Filters from a bank with uniform angular frequency components sampling and log Gabor filters from equation 3.1, represented by isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$. *Right column:* Their response to data set D1. *Top:* Filter with $(\lambda_i = 4, \phi_i = 0, \theta_i = \pi/2)$. *Bottom:* Filter with $(\theta_i = 4, \phi_i = \pi, \theta_i = \pi/2)$.

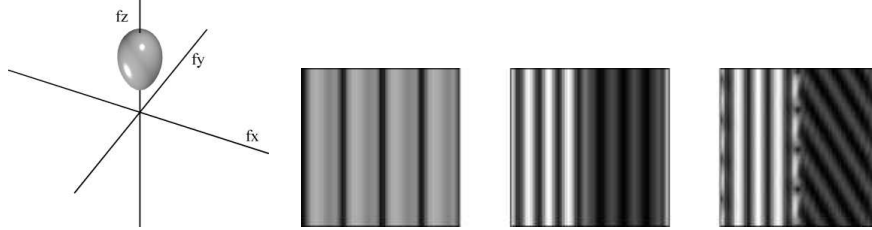


Figure 3.11: *Left:* Filters with $(\lambda_i = 4, \phi_i = 0, \theta_i = \pi/2)$ and $(\theta_i = 4, \phi_i = \pi, \theta_i = \pi/2)$ from a bank with non-uniform angular components sampling and log Gabor filters from equation 3.1 –they are identical–, represented by isosurfaces of the filter’s transfer functions verifying $T_i = \exp(-1/2)$. *Right:* Their response to data set D1.

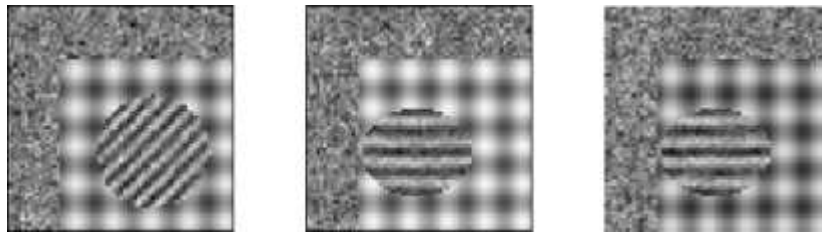


Figure 3.12: Volumetric data set D2 represented by orthogonal cross sections.

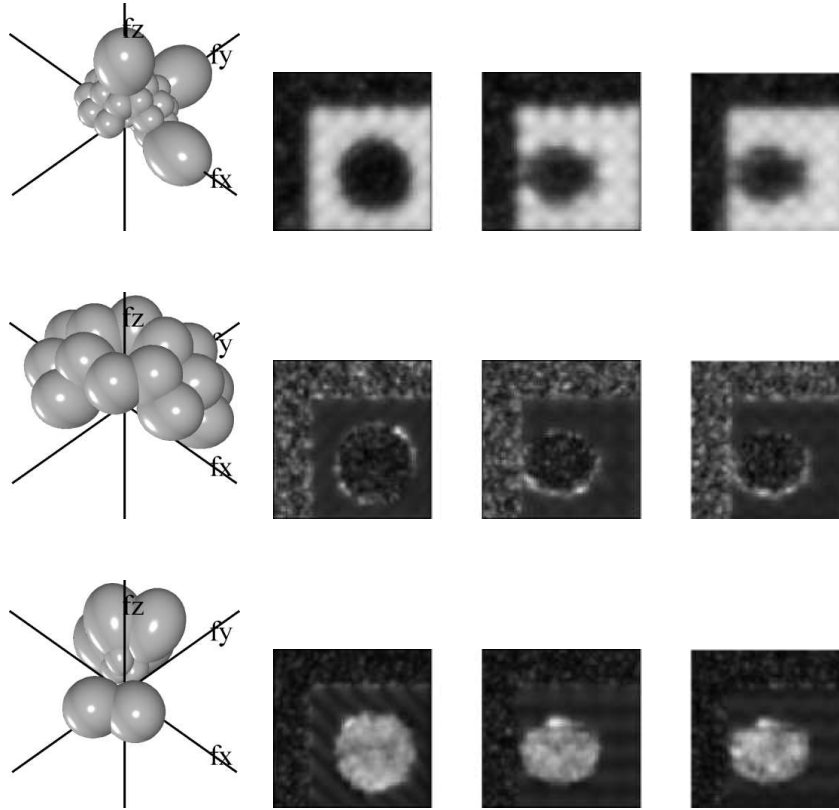


Figure 3.13: Results obtained for image D2. *Left column:* Filters correspondent to each of the three clusters obtained represented by isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$. *Right column:* Cross sections of the volumes reconstructed from each of the clusters. Each volume represents the sum of the energies of all filters belonging to the correspondent cluster

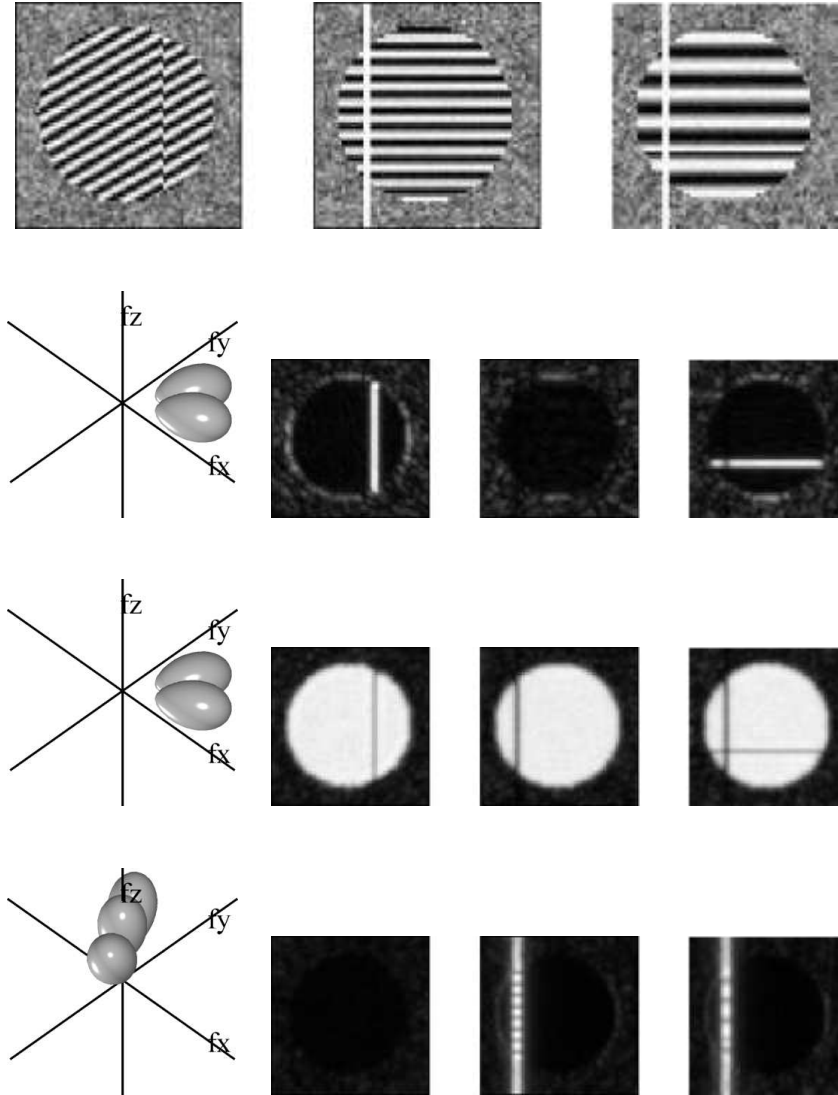


Figure 3.14: *Top:* Image D3, represented by the slices normal to the image coordinate axis. *Remainder Rows:* Result of frequency feature clustering. *Left Column:* Filter clusters obtained, representing filter as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Right Column:* Cross sections of their correspondent reconstructed visual patterns.

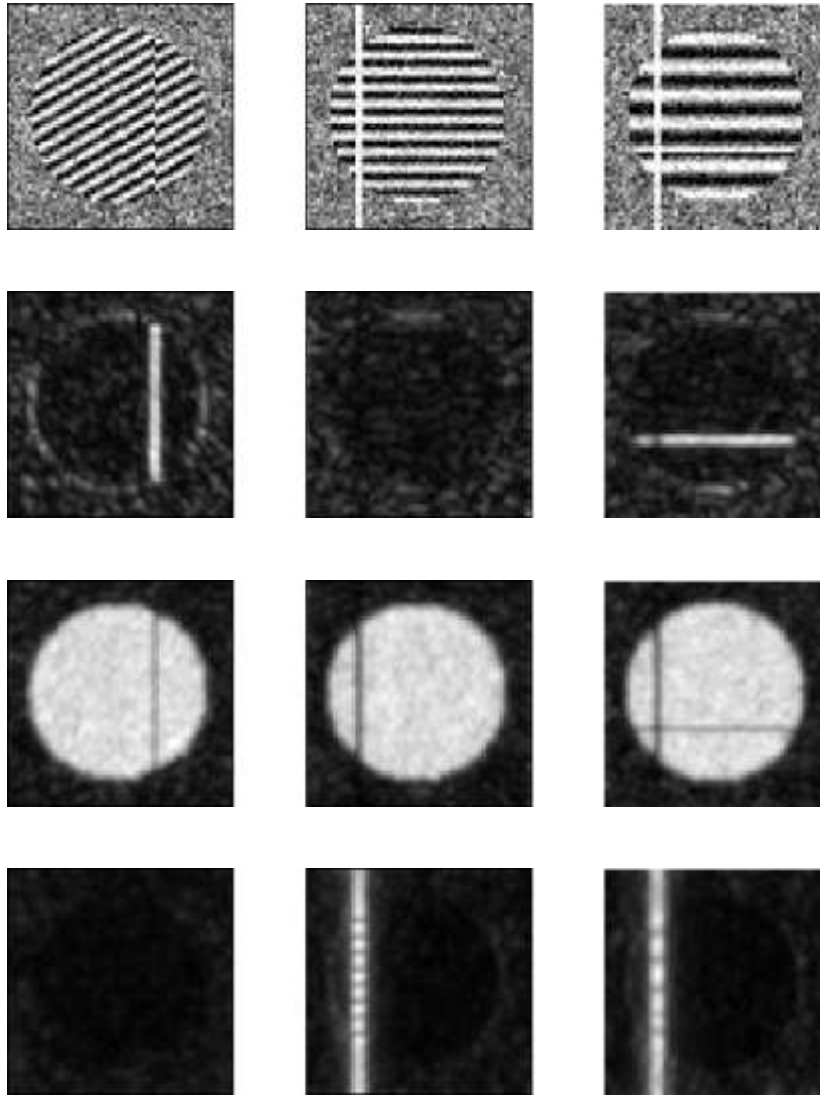


Figure 3.15: *Top Row:* Image D4, obtained from D3 by adding Gaussian noise of standard deviation $\sigma = 25$, represented by slices normal to the image coordinate axis. *Remainder Rows:* Visual patterns obtained for D4. Filter clusters are identical to those shown in 3.14.

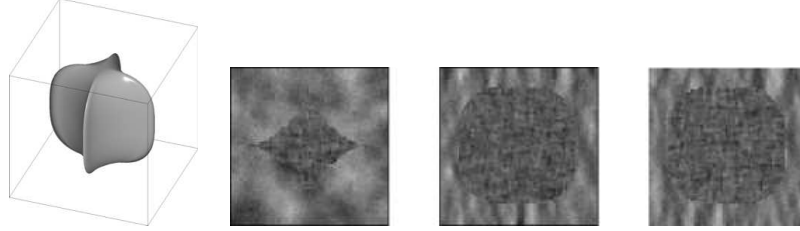


Figure 3.16: Volumetric data set D5 represented by three orthogonal cross sections. *Left:* Shape of the inner region of the image. *Right:* Cross sections of the volume

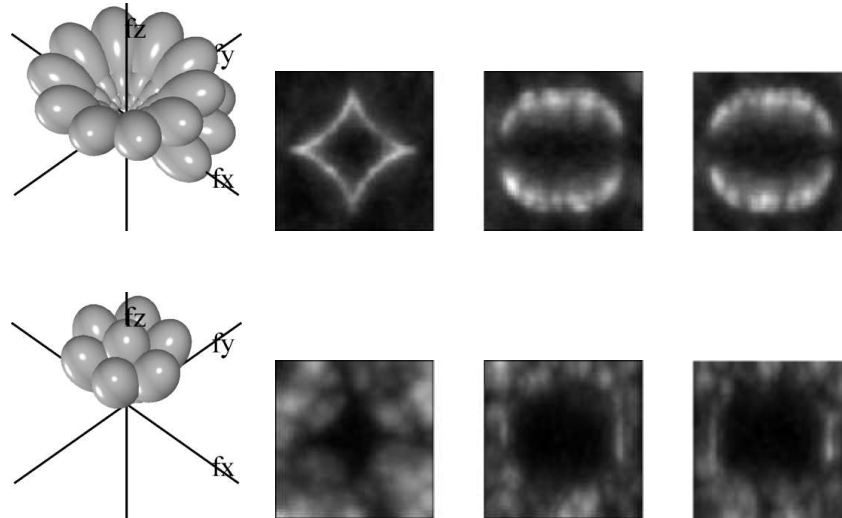


Figure 3.17: Results obtained for image D5. *Left column:* Filters correspondent to each of the two clusters obtained, represented by isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$. *Right Column:* For each cluster, cross sections of the volumes reconstructed as the sum of energies of all filters in the cluster.

CHAPTER 4

Improving Dissimilarity Estimation

In the previous chapter, the general description of the visual pattern partitioning method has been presented. In this first look, the normalized mutual information measure has been chosen to estimate filter's responses dissimilarity. There are several reasons for this selection. Mutual Information has great popularity in the field of image registration and, therefore, there is a great deal of research regarding it. It highly improves the computational cost of distances based on attention processes. Other advantages of the distance based on Mutual Information in relation to attention-based measures are that it is less parameterized and it does not depend on the performance of low level processes, like non-maxima suppression and scale estimation.

Nevertheless, in our experiments we have observed that the results are not completely satisfactory. In particular, there is a systematic error produced in a specific type of images with apparently very well distinguishable patterns: when using the distance based on mutual information, the composite-feature detector is not able to separate two regions, each one containing a grating pattern with orientation orthogonal to the other one.

This chapter is dedicated to the resolution of the problems arisen with the mutual information distance. To this end, a comparison with alternative measures is accomplished. In section 4.1 we analyze the meaning of mutual information and hypothesize about what a measure of dissimilarity between frequency features must represent to reflect phase congruence. Such analysis leads to the hypothesis that the linear functional dependency is the one that best approximates the relation-

ship between their intensity maps. As a consequence, we propose the correlation coefficient as a suitable distance for our application. To test this hypothesis, the performance of mutual information, the correlation coefficient and other alternative measures is studied. In section 4.2, the measures used in the comparison are described. Then, we will compare the presented distances in terms of computational cost and effectiveness. The analysis of the asymptotical computational time of each measure is presented in 4.3. In section 4.4 the experiment carried out to compare the performance of the different distances and its results are presented. Results are discussed in section 4.5.

4.1 Measuring Feature Phase Alignment

To initiate the discussion on how to measure the phase alignment between frequency features, this section begins with an illustrative example of the problems encountered using normalized mutual information. Figure 4.1.a shows an instance of the kind of images with which this measure produces incorrect results. The HVS easily identifies three visual patterns in this image: two grating patterns with orthogonal orientations separated by a thin line. The visual pattern partitioning method is expected to isolate them, producing, for each pattern, one cluster with its constituent frequency bands. However, the results obtained, presented in figures 4.1.b, c, d and e, show that this is not the case. For some reason, the frequency bands corresponding to the two grating patterns are grouped together in the same cluster.

The explanation for this behavior is related to the fact that, given a channel tuned to the orientation of one of the grating patterns, its response to the other grating pattern is not null, as can be seen in figure 4.2. Mutual information takes into account the frequency of the concurrence of each pair of intensity values in the two images, but not their proximity. As a result, it takes high values when the concurrence of weak and strong maxima takes place.

Hence, the question arises about the adequacy of I to the representation of phase congruence. As was introduced in section 2.1, it is likely that preattentive feature detection is accomplished by the HVS detecting points of locally maximum phase congruence. This is equivalent to identify locations where there is alignment on the local maxima of the energy of a range of subband versions of the image.

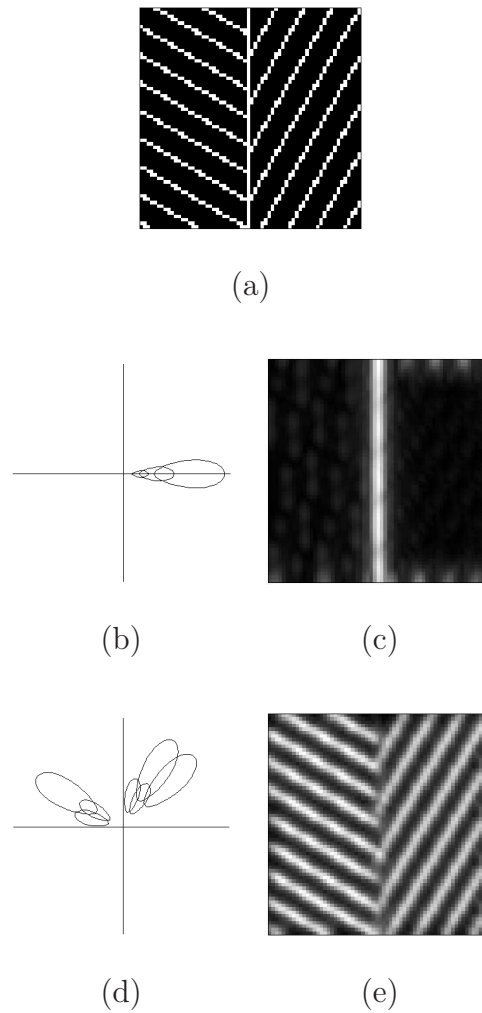


Figure 4.1: Example of clustering failure using normalized mutual information. Filters are represented as the iso-contours of level $\exp(-1/2)$ of their transfer functions. (a) 2D example image consisting of two orthogonal grating patterns and one line. (b) Filters corresponding to cluster 1. (c) Linear summation of the energies of filters in cluster 1. (d) Filters corresponding to cluster 2. (e) Linear summation of the energies of filters in cluster 2.

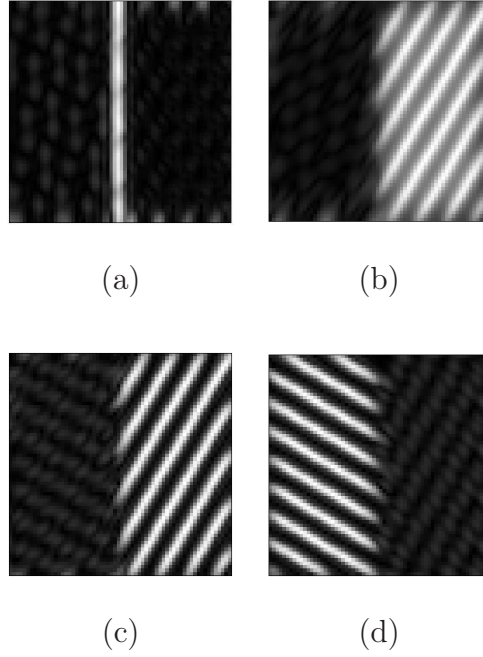


Figure 4.2: Energy of the responses of filters tuned to the orientations of different visual patterns in the example image from figure 4.1. (a) Filter 1: ($\phi = 0, \lambda = 4$). (b) Filter 32: ($\phi = 2\pi/3, \lambda = 16$). (c) Filter 31: ($\phi = 2\pi/3, \lambda = 4$). (d) Filter 13: ($\phi = \pi/3, \lambda = 4$).

Consequently, the presence of a visual pattern imposes a relationship among a subset of subband images, which means that there exists some degree of dependence among them.

Dependence can be defined as the predictability of a variable given knowledge of another one. From the point of view of information theory, dependence is considered the reduction of uncertainty of one random variable brought by the knowledge of a second random variable. When uncertainty is estimated by means of the Shannon entropy H and uncertainty reduction is measured using the Kullback's divergence K , what one has is the definition of mutual information.

Then, what I is measuring is statistical dependence in a general sense. Then again, other measures could be employed, using other definitions of uncertainty, other definitions of uncertainty reduction, or even more restrictive types of dependence. Basseville proposes a number of alternative definitions for entropy and divergence besides H and K (Basseville, 1996). The measure of information gain

proposed by Roche is restricted to the case when the conditional probability densities can be modelled by a Gaussian function (Roche et al., 1998, 2000). A much more restrictive measure of dependency is the correlation coefficient, which reflects the degree to which the two variables are related by a linear function.

The question now is what kind of measure is best suited to reflect phase alignment and produce the best results in feature clustering. The use of a measure that is invariant to unexpected or undesired types of dependence will prevent from incorrect grouping of features. On the other hand, an overconstrained measure could discard allowed relations. To determine what is the most appropriate measure one has to make some assumptions regarding the nature of the relations among the components of the same visual pattern. Here, the following *a priori* assumptions are adopted: the energy values of two frequency features belonging to the same cluster are linked; the link increases with energy maxima alignment; total dependence is not possible, as one response cannot be perfectly predicted by the other, given that they are versions of the same image *viewed* through different sensors.

Now, returning to the question about what is the measure that best reflects those assumptions, for the application in hands mutual information is considered an underconstrained measure of dependence. This is due to that it makes no assumptions regarding the kind of statistical functional relationship that exists among the frequency components of a visual pattern. For example, an image is totally predictable given its negative –in fact their mutual information is maximum– but that situation does not reflect phase alignment. There must be alignment between bright locations in the two images. This is also evident from the example presented at the beginning of this section. Table 4.1 shows the distance values obtained using mutual information for several frequency features. It is clear that the distance value obtained for the two orthogonal grating patterns is not small enough to split them into two different clusters.

For this reason, we pose the hypothesis that, even when the dependence among frequency features belonging to the same visual pattern is not exactly linear, linearity is the most appropriate kind of dependence to model the true link and should enhance the results obtained with an underconstrained measure. Hence, we propose a measure based on the correlation coefficient to represent the distance between frequency features. Such measure must take into account the sign of the correlation coefficient to constrain the possible linear functions to have positive slope, thus reflecting that maxima must be aligned with maxima.

The plots presented in figure 4.3 reinforce this hypothesis. There we observe, for three feature pairs, the mapping between the energy values of one of them into the other's. A pair of energy values from the two maps at the same location is marked with a dot. The first case, in figure 4.3.a, shows the mapping correspondent to one feature belonging to the vertical line and the other to one of the grating patterns. As can be seen, the points are distributed forming one quadrant of a circle –negative energy values are not allowed–, so that the correlation is near zero –see table 4.1–. The second case, in figure 4.3.b, shows the mapping for two features belonging to the same grating pattern but with different scales. Here, the linear tendency is evident. The last example in figure 4.3.c corresponds to one feature from each of the two orthogonal grating patterns. It has been already seen that channels tuned to one of the patterns have non-null response to the other one. That is why the mapping is approximately a composition of three Gaussians: the background of each frequency feature is mapped into the other's, giving place to a circle of low energies and the weak peaks in one feature are mapped into strong peaks into the other, producing the two elongated ellipses that separately show a linear tendency. The mutual information measure is able to detect this kind of dependence whilst the correlation ratio is relatively low –and, in this case, negative– in comparison to the first case –see table 4.1.

Besides all aforementioned, there could be other ways of improving the performance of the dissimilarity estimation. For example, it has been said earlier that visual patterns correspond to energy maxima alignment across frequency bands, so one possibility is to directly measure dissimilarity over energy maxima maps instead of raw energy maps. This will reduce the influence of scale. This can be easily accomplished by non-maxima suppression. On the other hand, when dissimilarity measures are applied over energy maxima maps, important differences due to scale are not being considered. It is well-known that scale enlargement produces an effect of displacement of the energy maxima. Two nearby maxima are probable to be detected as one single peak when viewed through a low spatial-resolution sensor. As a consequence, the use of energy maxima maps might fail to link high and low frequency channels belonging to the same visual pattern. An alternative is to consider also the regions of influence of each energy maximum. Even when maxima of two energy maps do not match, the energy profiles in the surroundings of each maximum should have similar tendencies. Furthermore, if the neighborhood of the maxima is not taken into account the notion of size of the spatial features is lost.

Filter Pair	NI	D_{NI}	ρ
{32,31}	0.3146	0.1752	0.8391
{ 1,31}	0.1157	0.5559	-0.0096
{13,31}	0.2041	0.3621	-0.1630

Table 4.1: Normalized Mutual Information NI and associated distance values D_{NI} for instances of same-pattern –32, 31, in first row – and different-pattern frequency features. Filter’s responses are shown in 4.2. Their correlation coefficient ρ is also listed as a reference.

These last modifications of the measures can be considered to simulate a context-free attention mechanism like those that take place in the visual system of mammals. In this case, the fixation points are represented as the local maxima of the energy, thus associating relevance of the spatial features to phase congruence, as suggested in (Morrone and Owens, 1987; Reisfeld, 1996).

Despite everything, the problem with weak maxima remains. In an ideal situation, one channel would not have any response to a pattern with orientation very different to its and, consequently, the mutual information would also produce a good result in cases like the example presented in figure 4.1. Another way of improving the measurement of phase alignment would be to eliminate weak maxima by simply thresholding the energy maps or the energy maxima maps.

The problems one is probable to find when applying all those alternative implementations are practical issues. How to determine appropriate thresholds? Are there thresholds or methods for calculating them that are valid for all images and even for all features? Is always contrast –energy– related to the relevance of an energy peak? How then to determine fixation points? What is the best way to determine their scales? Is the reduction of influence of the scales profitable? etc. To test the posed hypothesis and to study the benefits of the different alternatives, an experimental study on their performance has been carried out.

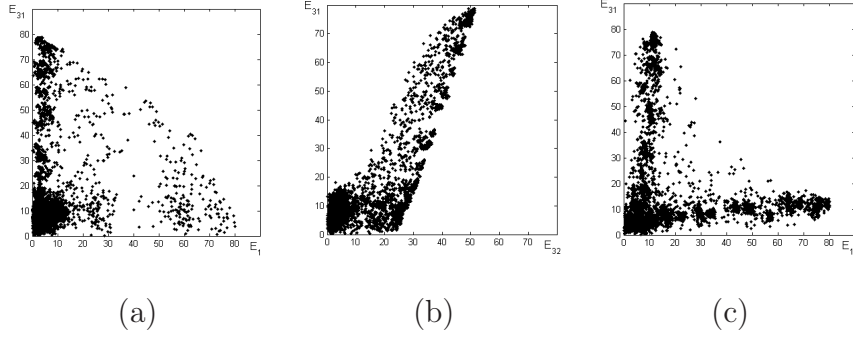


Figure 4.3: Plots of pairs of energy values for each pixel location corresponding to filters from figure 4.2: (a) $\{1, 31\}$ (b) $\{32, 31\}$ (c) $\{13, 31\}$

4.2 Dissimilarity Measures

This section presents a formal definition of the diverse distances used in the test. Subsection 4.4 presents the test and its results and, in subsection 4.5, these results are discussed and the conclusions obtained from them are presented.

4.2.1 Statistical Measures of Dependence

In this subsection we describe some commonly employed measures of the statistical dependence between two images which we have chosen for our study besides NI .

4.2.1.1 Classical Dissimilarity Measures

A pair of classical measures to compare images, the Root Mean Square Error (RMSE) and the Peak Signal to Noise Ratio (PSNR) are included in the experimental test for the sake of completeness. They will also serve as reference measures. The expression of PSNR in decibels is calculated from the RMSE as follows

$$PSNR = 20 * \log_{10} \left(1 / RMSE(\tilde{X}, \tilde{Y}) \right), \quad (4.1)$$

where $\tilde{X}$ and $\tilde{Y}$ are obtained from X and Y by mapping the range of the two images to $[0, 1]$ –zero corresponds to the minimum of the two images and one to their maximum.

The measure RMSE is already a measure of dissimilarity. In the experiment it has been employed as it is.

$$D_{RMSE}(X, Y) = RMSE(X, Y) = \frac{1}{|\Omega|} \sqrt{\sum_{i \in \Omega} (x_i - y_i)^2}, \quad (4.2)$$

where Ω is the image domain.

On the other hand, PSNR is a measure of proximity with values ranging from zero to ∞ . Its correspondent distance function is obtained with the following transformation.

$$D_{PSNR} = 1/(1 + PSNR). \quad (4.3)$$

The advantage of D_{PSNR} regarding D_{RMSE} is that the first is a normalized measure, independent on the intensity ranges

4.2.1.2 Correlation Ratio

The correlation ratio was proposed in (Roche et al., 1998) as an image dissimilarity estimation for multimodal medical image registration. Like mutual information, it is a measure of functional dependency, but it is restricted to the case when the conditional probability densities of the image intensities can be modelled as Gaussians (Roche et al., 2000). The correlation ratio of X conditioned to Y is defined as

$$\eta^2(X|Y) = 1 - \frac{Var(X - \mathbb{E}(X|Y))}{Var(X)} = \frac{Var(\mathbb{E}(X|Y))}{Var(X)}.$$

It represents the amount of the uncertainty of X that is predicted by Y in relation to the total uncertainty of X . The magnitude η^2 can be considered as a measure of mutual information when the variance is taken as a measure of uncertainty.

What follows is an efficient algorithm for the estimation of η^2 which does not require the calculus of the joint histogram (Roche et al., 1998).

$$\eta^2(X|Y) = 1 - \frac{1}{N\sigma^2} \sum N_i \sigma_i^2,$$

where

$$\sigma_i^2 = \frac{1}{N_i} \sum_{p \in \Omega_i} (X(p)^2 - \mu_i^2),$$

$$\mu_i = \frac{1}{N_i} \sum_{p \in \Omega_i} X(p),$$

and $\Omega_i = \{p \in \Omega, Y(p) = i\}$ is the iso-set of Y with constant value i and N_i is its size. In this way, η is estimated as the uniformity in the X values to which each iso-set of Y is projected. However, it does not take into account the difference between a value from Y and the mean of its projected values in X . Therefore, the same problems encountered with NI are expected. It also depends on the discretization of the values in Y to determine the center bin of each iso-set, although it is more robust to the selection of this parameter than NI .

The correlation ratio is not a symmetric measure. One signal may tell us much about another one but the opposite may not be true –let’s think of one image and a degraded version of it corrupted by noise. As $\eta(X|Y) \neq \eta(Y|X)$, the symmetric measure is obtained by taking the maximum of both. Again, it is a similarity measure, so that its range must be inverted and transformed to accentuate inter-cluster distances.

$$D_\eta(X, Y) = 1 - \sqrt{\max(\eta^2(X|Y), \eta^2(Y|X))}. \quad (4.4)$$

4.2.1.3 Other f-divergences

In general a divergence is any functional of two probability distributions that measures the uncertainty reduction between them. To study the dependence between two variables X and Y , the two compared probabilities must be the observed joint distribution $P(X, Y)$ and the computed joint distribution in case of independence $P(X \times Y) = P(X) \times P(Y)$. A divergence evaluated from these probability distributions is a measure of the reduction of uncertainty of the first variable brought by the knowledge of the second variable. Many alternative definitions of divergence are possible, depending of the specific measure of uncertainty employed and the functional selected to represent its reduction.

One such divergence is the Mutual Information I of two random variables. It is the divergence that uses Shannon’s entropy H as a measure of uncertainty and Kullback’s divergence K as the specific functional to represent the reduction of H . It is defined as follows. Nevertheless, other divergences are possible (Basseville, 1996). Here, a pair of distances derived from two of these divergencies are considered. These divergences are of the type of f-divergences, which have the general form

$$\mathbf{I}_f(P, Q) = \mathbb{E}_Q \left(f \left(\frac{p}{q} \right) \right) = \int f \left(\frac{p(\lambda)}{q(\lambda)} \right) q(\lambda) d\lambda, \quad (4.5)$$

where f is continuous and convex in $[0, +\infty)$. It is also imposed that $f(1) = 0$ –to ensure $I_f(P, P) = 0$ – and $f''(1) > 0$. $I_f(P, Q)$ is a measure of similarity between two signals X and Y when P is their observed joint distribution $P(X, Y)$ and Q is the expected distribution in case of total independence $P(X \times Y) = P(X) \times P(Y)$.

The mutual information I is the f -divergence with

$$f(u) = u \log(u). \quad (4.6)$$

Here, two of these divergences have been included in the comparison, encouraged by the good results obtained with them in (Sarrut and Miguet, 1999). They are the Toussaint's distance T and the Lin's K-divergence K_{div} . The Toussaint's distance is the f -divergence resulting from taking

$$f(u) = u \frac{(u-1)}{(u+1)}, \quad (4.7)$$

$$T(X, Y) = \sum_{i,j} P_{x,y}(i, j) - \frac{2P_{x,y}(i, j)P_x(i)P_y(j)}{P_{x,y}(i, j) + P_x(i)P_y(j)},$$

$$D_T(X, Y) = \left(1 - \sqrt{T(X, Y)/T_{max}}\right)^2, \quad (4.8)$$

where $T_{max} = 1 - 2/(N+1)$.

In the Lin's K-divergence, f takes the following expression

$$f(u) = u \log(u) - u \log\left(\frac{1+u}{2}\right), \quad (4.9)$$

$$K_{div}(X, Y) = \sum_{i,j} P_{X,Y}(i, j) \log \frac{2P_{X,Y}(i, j)}{P_{X,Y}(i, j) + P_X(i)P_Y(j)},$$

$$D_{K_{div}}(X, Y) = \left(1 - \sqrt{K_{div}(X, Y)/K_{div \ max}}\right)^2, \quad (4.10)$$

where $K_{div \ max} = \log(2N/(N+1))$.

4.2.1.4 Correlation Coefficient

The correlation coefficient is a measure of the degree of linear functional dependence. It is defined as follows.

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sum (X_i - \mu_X)(Y_i - \mu_Y)}{\sqrt{\sum (X_i - \mu_X)^2 \sum (Y_i - \mu_Y)^2}}.$$

The values of ρ are in the range $[-1, 1]$. When $|\rho| = 1$ the data set perfectly fits a linear function and its sign reflects the sign of the slope of the function. The lower $|\rho|$, the lesser the linear tendency. In the present application the sign of the coefficient turns out to be very useful to constrain the possible relationships to positive slope linear functions, thus preventing from the clustering of features with concurrence of maxima and minima.

$$D_\rho(X, Y) = \left(1 - \sqrt{(1 + \rho(X, Y)) / 2}\right)^2. \quad (4.11)$$

This distance function takes values in the range $[0, 1]$. The minimum value corresponds to perfect linear dependence with positive slope and the maximum corresponds to the case of perfect fit with negative slope, like, for example, an image and its inverse.

This measure does not depend on the selection of any parameter. This is an advantage with regard to any divergence measure, which involve the discrete estimation of joint and/or marginal probabilities.

4.2.2 Attention Based Measures

The measures presented in this subsections are also statistical measures of dependence. The reason to present them in a different subsection is that they all have the particularity of introducing some attention mechanism. Only a set of fixation points and their neighborhoods contribute to the final estimation through a given pooling strategy.

4.2.2.1 The RGFF Dissimilarity Measure

The distance used in the RGFF (Rodríguez Sánchez et al., 2000; Chamorro Martínez et al., 2003), which here we will call D_{RGFF} , is inspired in biological processes, combining attention mechanisms (Morrone and Owens, 1987; Reissfeld, 1996) and Quick pooling of sensor outputs (Quick, 1974; Graham, 1989; Watson, 2000). It consists of the measurement of a set of local statistics in a neighborhood of the attention points, located at energy maxima. The distance is obtained by applying a beta-norm function over the weighted sum of the differences between the statistics of each filter's response. This distance is computationally very expensive, highly parameterized and highly dependent on the performance of low level processes

like non-maxima suppression and scale estimation. Its calculation is described in the following expressions.

Quick pooling is accomplished by applying a β -norm to the contributions of each attention point p

$$D_\beta(X, Y) = \frac{1}{Card(\Omega_X)} \left(\sum_{p \in \Omega_X} |\mu_p(X, Y)|^\beta \right)^{1/\beta}, \quad (4.12)$$

where Ω_X is the set of fixation points in X and μ_p is the weighted sum of differences between statistics at attention points, with the following expression

$$\mu_p(X, Y) = \sum_{k=1}^Q \frac{1}{\omega_k} d(T_k^p(X), T_k^p(Y)), \quad (4.13)$$

where T_k^p are the elements of the vector T^p of local statistics measured at fixation point p , ω_k is the maximum value of T_k over all fixation points in all energy maps and Q is the length of T^p .

As $D_\beta(X, Y) \neq D_\beta(Y, X)$, the symmetric measure is obtained as follows

$$D_{RGFF}(X, Y) = D_\beta^2(X, Y) + D_\beta^2(Y, X). \quad (4.14)$$

In our experiments β equals to 2 in equation 4.12. The neighborhood of an attention point for $p \in \Omega_X$, named $\Gamma(p, \sigma_p)$ is defined as all positions with distance to the maximum lesser than its scale σ_p , which is the distance between the maximum and the nearest minimum. The local statistics T_k^p in equation 4.13 are estimated by weighting the contribution of each point in $\Gamma(p, \sigma_p)$ with a Gaussian function on the distance to the maxima, so that the nearer the neighbors, the more important the contribution in the calculation. The local statistics used here are the normalized energy and its local entropy, although other statistics could be considered, like phase or standard deviation and contrast of the normalized energy.

4.2.2.2 Dissimilarity Estimation on Energy Maxima Maps

A new set of dissimilarity measures D_δ^* can be defined from each of the distances δ presented in section 4.2 by applying them over the thresholded energy maxima maps of the filter's responses.

$$D_\delta^*(X, Y) = D_\delta(X', Y'), \quad (4.15)$$

where X' and Y' are respectively X and Y after thresholding followed by non-maxima suppression, i.e., changing the intensity values to zero at points which are not maxima. The direction used in non-maxima suppression to search for maxima previous and next neighbors is the direction of the filter. The threshold must be carefully selected to avoid the elimination of relevant peaks. In our experiments, the threshold energy is calculated such that a given percentage p of the total amount of maxima is preserved. This percentage depends on the density of maxima d , measured as the number of maxima in relation to the image size, as $p = 1/(1 + kd)$, with $k \in \mathbb{R}$ and $k > 0$. The reason for this dependence is that small scale subband images usually present a larger amount of maxima for a given image than large scale filters, but they also contain a larger percentage of low contrast maxima in natural images, due to noise or to the non null response of a filter to structures with orientations different to its. In the results presented in next section, k takes value 50.

4.2.2.3 Dissimilarity Estimation on Regions of Interest

Again, a new set of dissimilarity measures $\hat{D}_\delta$ is obtained from the distances δ from section 4.2, this time by using only the points belonging to the regions of interest. This approach differs from the previous one, not only in maintaining unaltered the energy values in the maxima neighborhoods, but also in that the background points do not contribute to the distance estimation.

$$\hat{D}_\delta(X, Y) = D_\delta^2(X(\Gamma_X), Y(\Gamma_X)) + D_\delta^2(X(\Gamma_Y), Y(\Gamma_Y)) \quad (4.16)$$

where Γ_X is the mask defined by the set of points belonging to the region of interest of any maximum in an energy map X . A mask Γ_X is obtained in a two-stage process. Firstly, the set energy maxima of X , Ω_X , is determined in the same manner as in the dissimilarity estimation over energy maxima maps, described in subsection 4.2.2.2. Secondly, for each maxima $p \in \Omega_X$ its neighborhood $\Gamma(p, \sigma_p)$ is defined as all positions with distance to the maximum lesser than its scale σ_p , which is the distance between the maximum and the nearest minimum. Then, Γ_X , is the union of the regions of interest associated to each attention point p .

$$\Gamma_X = \bigcup_{p \in \Omega_X} \Gamma(p, \sigma_p).$$

Using this strategy, the dissimilarity measurement is less influenced by the incorrect location of fixation points.

4.3 Computational Cost

One of the main advantages of global measures in relation to attention-based measures is their lower computational cost, specially when neighborhood estimations are needed. When applied to 3D images, the computational time of scale and local statistics estimation strongly increases. Here, an analysis of the asymptotic computational cost of the different measures presented in previous section takes place. The conclusions are summarized in table 4.2.

Let us suppose that the input data are a volume of size $N \times N \times N$, that our filter bank consists of N_f filters and that the number of bins used for histogram calculations is M . The calculus of ρ is $O(N^3)$ while the estimation of NI , η , T and K_{div} involves the construction of the joint histogram of the two maps, which is $O(N^3)$, and the posterior accumulation of the contributions of each bin in the histogram, which is $O(M^2)$. Supposing that N and M are of the same order of magnitude, the cost of the calculation of each of these dissimilarities is $O(N^3)$. Dissimilarities must be calculated for each of the $N_f(N_f - 1)$ pairs of filters, resulting in a computational cost of $O(N_f^2 \cdot N^3)$. The modified distance D_δ^* has extra computational burden due to non-maxima suppression and estimation of regions of interest. Non-maxima suppression has a cost of $O(N_f \cdot N^3)$ and, thus, the asymptotical limit remains the same for D_δ^* than for D_δ .

Dissimilarity estimation on regions of interest, however, is more time consuming due to the calculus of the scales for each attention point. The algorithm can be decomposed in two steps: 1) detection of attention points and estimation of their scales and 2) for each pair of filters determine the points belonging to regions of interest and estimate D_δ on them. The following pseudo-code describes the algorithm employed in the first step

For each attention point	$O(N^3)$
For $r = 1$ to maximum radius	
Find the neighborhood of radius r	$O(r^3)$
If there is a minimum in the neighborhood	$O(r^3)$
Set r as the scale of this point	
Break	

Measure	Asymptotical Cost
D_δ	$O(N_f^2 \cdot N^3)$
D_δ^*	$O(N_f^2 \cdot N^3)$
$\hat{D}_\delta$	$O(\max(N_f \cdot N^7, N_f^2 \cdot N^6))$
D_{RGFF}	$O(\max(N_f \cdot N^7, N_f^2 \cdot N^6))$

Table 4.2: Asymptotical computational costs for each of the dissimilarity measures presented in section 4.2. Symbol δ represents all global measures listed in 4.2.1

```

end if

end for

end for

```

If R is the maximum radius, the computational cost of the inner loop in this algorithm is $\sum_{r=1}^R r^3 = (R(R+1)/2)^2 = O(R^4)$. Assuming that, in the worst case, $O(R) = O(N)$, then the cost of the complete module is $O(N^7)$. The second step has a cost of $O(N_f^2 \cdot N^6)$. For each attention point, all positions at a distance lesser than the correspondent scale have to be determined. The asymptotic cost of this task is proportional to the number of attention points and their neighborhood size, resulting $O(N^6)$, which predominates over the cost of calculating $\hat{D}_\delta$ on the regions of interest. If the local differences are stored, then the computational cost of the third module is reduced to $O(N_f^2 \cdot N^3)$, but memory requirements increase in $O(N_f \cdot N^7)$.

The calculus of D_{RGFF} can be summarized in a sequence of three steps: 1) calculus of fixation points and scales for each feature, 2) for each filter pair, calculus of local statistics differences $d(T_k^p(X), T_k^p(Y))$ and their weighting factors ω_k and 3) for each filter pair calculate D_{RGFF} using equations 4.13, 4.12 and 4.14, which is $O(N_f^2 \cdot N^3)$. First step is identical to the first step in $\hat{D}_\delta$ calculus. The second step has also a cost of $O(N_f^2 \cdot N^6)$ due to the sequenced calculation of the neighborhoods of attention points and their local statistics. Hence, the asymptotical computational cost is the same for D_{RGFF} and $\hat{D}_\delta$.

4.4 Experimental Results

To compare the performance of the dissimilarity measures under consideration, the visual pattern partitioning method described in the previous chapter has been applied to a set of test images employing each one of the distance functions. The test bench is composed of 32 images, 13 of them 2D and the other 19 3D. While it is quite easy to determine whether the results obtained for a 2D image are correct by visual inspection, for 3D images this is more difficult. For this reason, all the 3D images in the bench are synthetic. The correctness of the results is determined by comparing them with the design specifications. The result must contain one cluster for each visual pattern present in the image and the frequency bands composing them must coincide with the expected ones. In the 2D case, the bench contains both synthetic images, natural images with clearly identifiable visual patterns, and images synthesized as a collage of natural Brodatz textures. In this last case the result must contain one cluster for each texture. Additionally, they may appear patterns correspondent to texture boundaries.

The results obtained are summarized in figure 4.4. It can be seen that D_ρ has the best performance, followed by D_{NI} . The application of these measures to energy maxima maps or using regions of interest worsens the overall effectiveness, although the analysis of the individual results shows that this variation improves them in certain kinds of images. In general the results are not very good due to the complexity of the task of matching same-pattern frequency features, which present strong variability. These results can not be extrapolated to other applications, like image registration.

4.5 Discussion

The results of the comparative study show that the correlation coefficient generally outperforms the other measures. This fact points to the conclusion that linearity reflects quite well the dependence existing between same-pattern frequency features. Nevertheless, D_ρ worsens the results produced by D_{NI} in some specific cases, like in the example of figures 4.6 and 4.7. In such situations the simplification that linearity implies may be too strong.

The mutual information measure also produces acceptable results. However, it

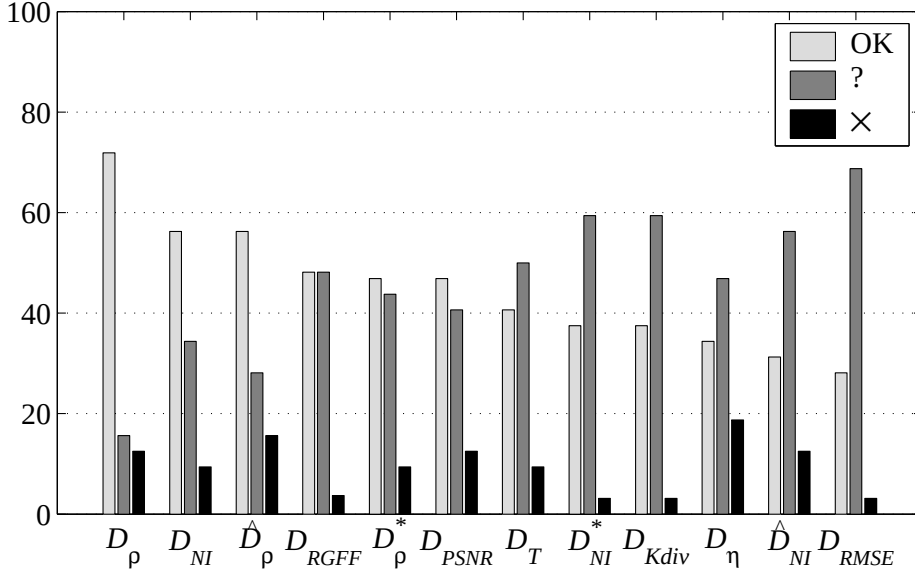


Figure 4.4: Percentage of correct (OK), incorrect (X) and indecisive (?) results for each distance.

systematically fails, for example, to separate grating patterns with different orientations. This is caused by the non null response of the filters to patterns with orientation orthogonal to it. Therefore, D_{NI} between two of such patterns is small, provided that NI does not consider the magnitude of the difference between the responses of the filters. This problem is observed in the example of figure 4.8. The use of the correlation coefficient does not suffer from this inconvenience, as seen in figure 4.9.

The correlation ratio gives very poor results. One of the sources of error is that, like the NI , it treats energy values in a semi-qualitative manner: it takes into account the dispersion of values of X inside the region defined as an iso-set in Y , but it does not consider the difference between a value from Y and the mean of its projected values into X . Consequently, the same error as with NI occurs with grating patterns of different orientations, as seen in figure 4.10. Besides, in many other cases D_η tends to produce too many clusters independently of the histogram number of bins used in estimation of probabilities.

The distances based on alternative measures of divergence do not improve NI . It must be taken into account that these measures are normalized using its theoretical maximum. T and $Kdiv$ can be small either because there is little uncertainty reduction or simply because the two images have little uncertainty in relation to the

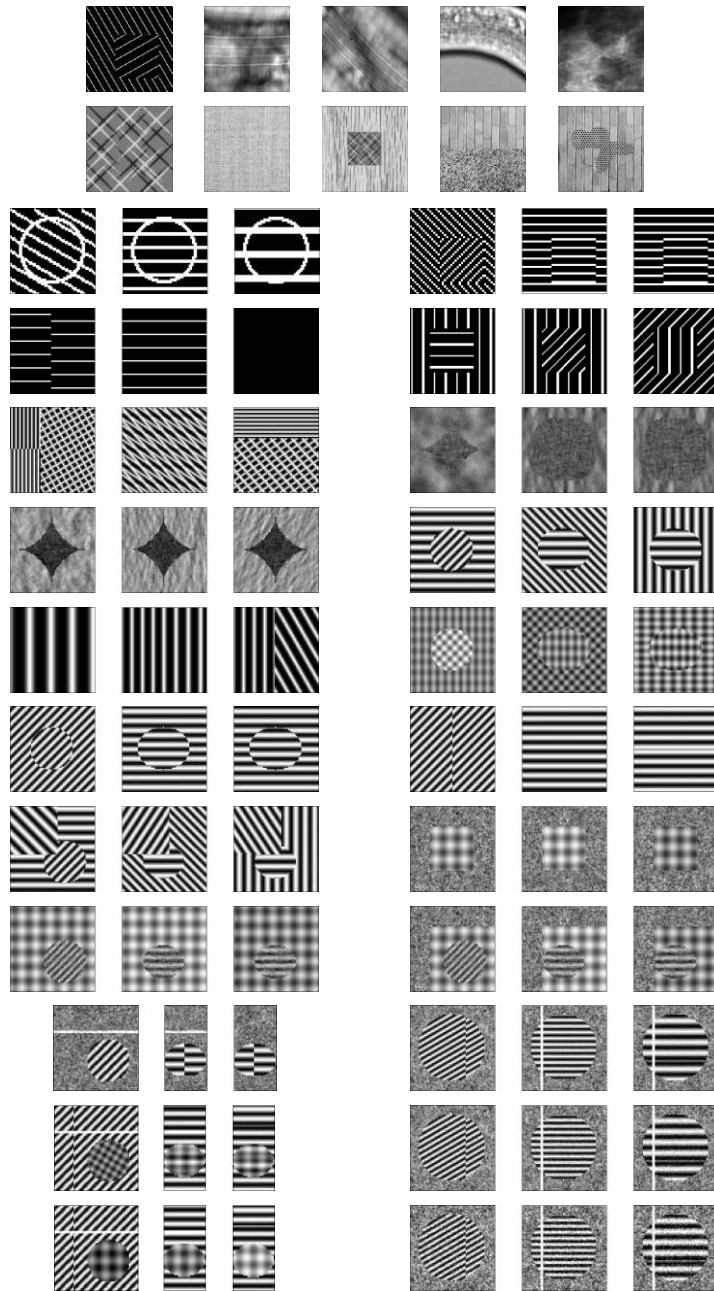


Figure 4.5: Images used in the test. The first two rows correspond to 2D images and the remainder to 3D images, represented as the three cross-section of the volume passing through the image center.

allowed maximum. Conversely, NI is normalized by the marginal entropies, so that the measure of divergence is independent on the amount of uncertainty of the two features.

The RGFF distance does not suffer from the problem arisen with NI –see figure 4.11– for two reasons: it applies thresholding in the selection of fixation points and the discrepancy between local statistics is calculated as a difference, so that matching of weak and strong maxima is penalized. However, it produces worse results than D_ρ and D_{NI} . This is mainly due to an incorrect selection of thresholds, local statistics or any other of the many parameters involved in the calculations. Additionally, the computational cost of this measure is excessive.

Neither D_δ^* nor $\hat{D}_\delta$ overcome their correspondent measures taken over the raw data, D_δ , with $\delta = \{NI, \rho\}$. There are specific examples in which the use of the modified measure D_δ^* , $\hat{D}_\delta$ or both imply an improvement, but globally this is not true. Success strongly relies on the selection of the energy threshold. A small variation on this parameter can dramatically changes the result. For example, in the case presented in figures 4.12, 4.13 and 4.14, D_{NI} fails to detect the three grating patterns, while D_{NI}^* succeeds precisely when then thresholds are such that weak maxima are eliminated. However it is very difficult to find a general method to select the appropriate threshold for each feature. Some trials have been made to determine the thresholds using the energy histograms that have produced excellent result in many cases, but a universal rule has not been found.

Apart from the adequacy of each measure, the results are also dependent on other factors, mainly the selection of the filter-bank parameters and the performance of the active filter selection. Some of the wrong results in this experiment can be easily corrected by adjusting the parameters that control the filter’s transfer function overlapping, both in frequency and orientation. For instance, the method succeeds in isolating the sphere shape in case shown in figure 4.15, but a decrease in the orientation overlapping produces the disintegration of the boundary. Another source of error is the incorrect selection of active filters. The most critical case appears when a relevant frequency channel is discarded in this stage, although this is not likely to happen using the current selection method, which is quite conservative. What is most probable to happen is that not all of the low energy channels had been eliminated. Under these circumstances, those frequency features may join the main clusters or generate a new cluster or produce the proliferation of new clusters with low information content. None of these cases has been considered incorrect in the

presented study.

To summarize, D_ρ brings the best performance in our experiment. It seems to be the most appropriate measure for a general purpose application. Nevertheless, other measures, like D_{NI} may bring also good results. The selection of the measure may depend on the kind of images in a specific application. Given an application with low variability in the images involved, it may be possible to enhance the results of D_ρ or D_{NI} by using attention mechanisms, by determining application specific threshold parameters. However, for a general purpose application, we conclude that the use of measures that are closer to the biological processes is not profitable if crucial information, like contrast thresholds values or the specific attention mechanisms that take place in the HVS, are not known.

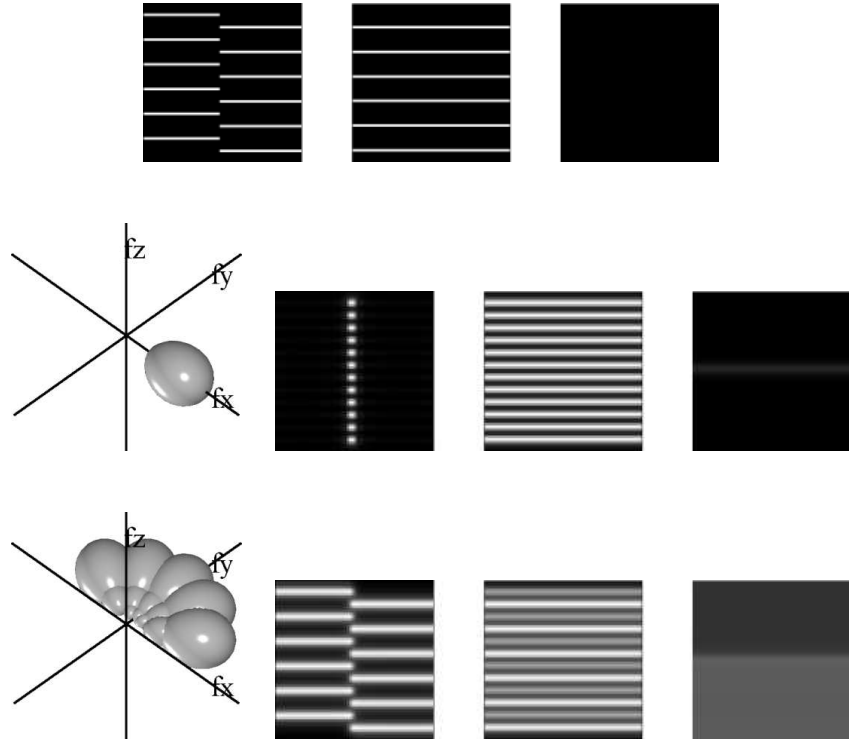


Figure 4.6: *Top Row:* Cross-sections of a 3D synthetic image of two grating patterns differing in a phase shift. *Remainder Rows:* Result of frequency feature clustering using dissimilarity D_{NI} . Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Left Column:* The two clusters obtained. *Right Column:* Cross sections of their correspondent reconstructed visual patterns.

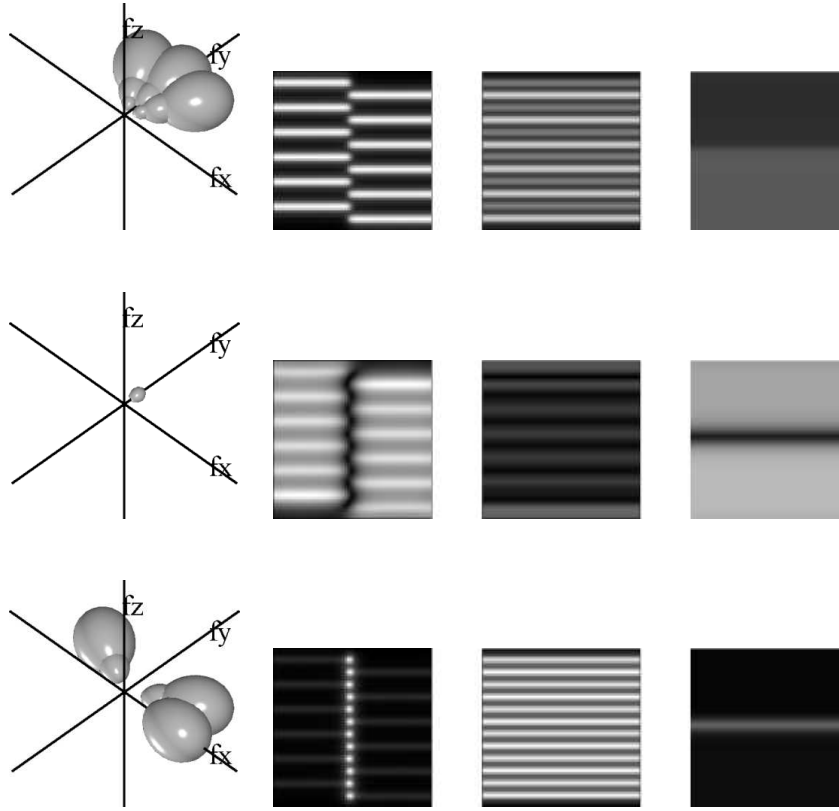


Figure 4.7: Result of frequency feature clustering of the image from 4.6, 1st row using dissimilarity D_ρ . *Left Column:* Resulting clusters. Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Right Row:* Cross sections of their correspondent reconstructed visual patterns.

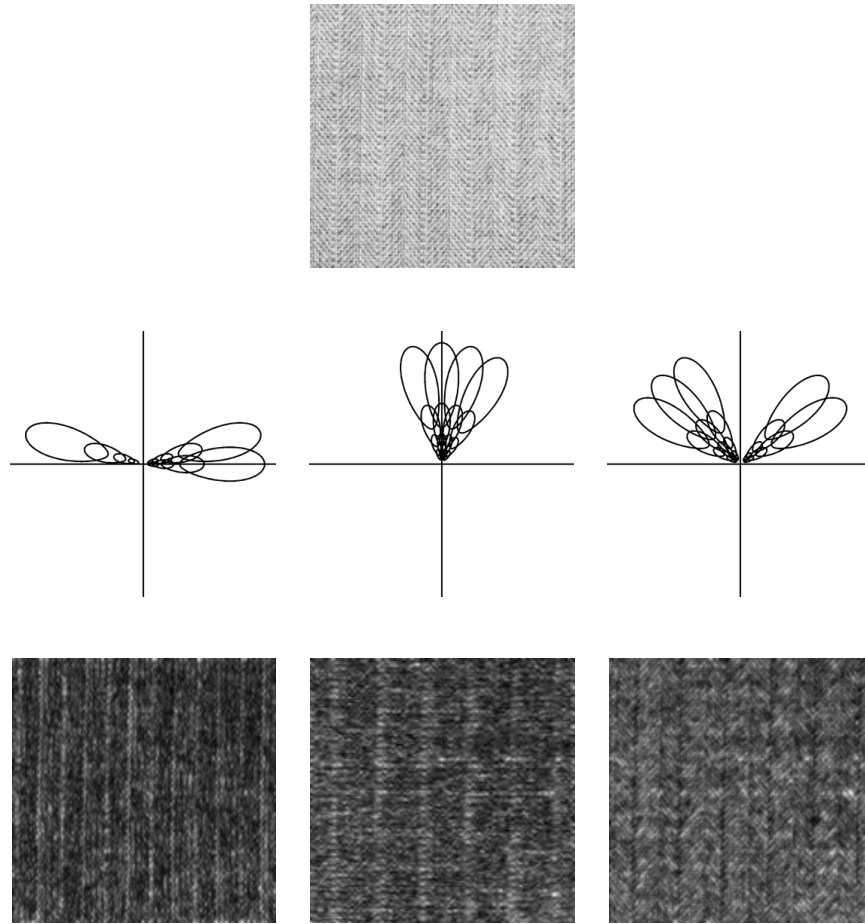


Figure 4.8: *Top Row:* 2D image of a Brodatz texture. *Center Row:* Result of frequency feature clustering of the image above using dissimilarity D_{NI} . Filters are represented as the iso-contours of level $\exp(-1/2)$ of their transfer functions. *Bottom Row:* Reconstruction of the visual patterns associated to clusters in the 1st row.

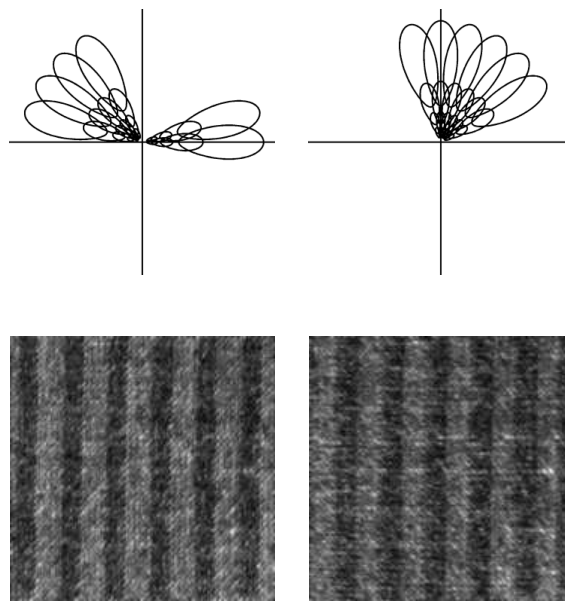


Figure 4.9: *Top Row:* Result of frequency feature clustering of image in top of figure 4.8 using dissimilarity D_ρ . Filters are represented as the iso-contours of level $\exp(-1/2)$ of their transfer functions. *Bottom Row:* Reconstruction of the visual patterns associated to clusters in the 1st row.

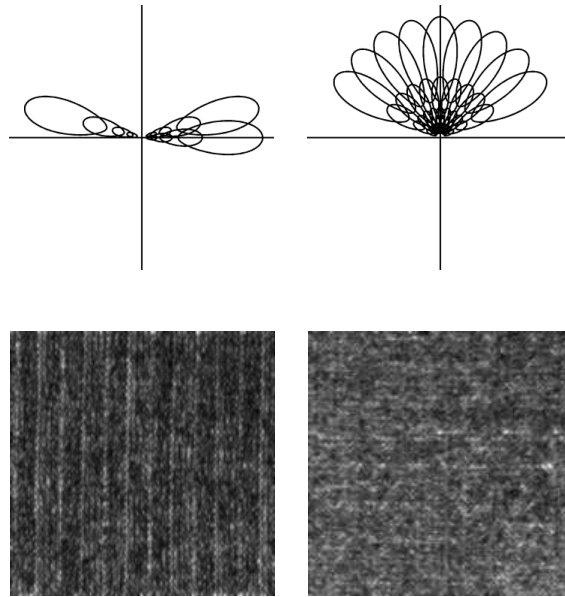


Figure 4.10: *Top Row:* Result of frequency feature clustering of image in top of figure 4.8, using dissimilarity D_η . Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Bottom Row:* Reconstruction of the visual patterns associated to clusters in the 1st row.

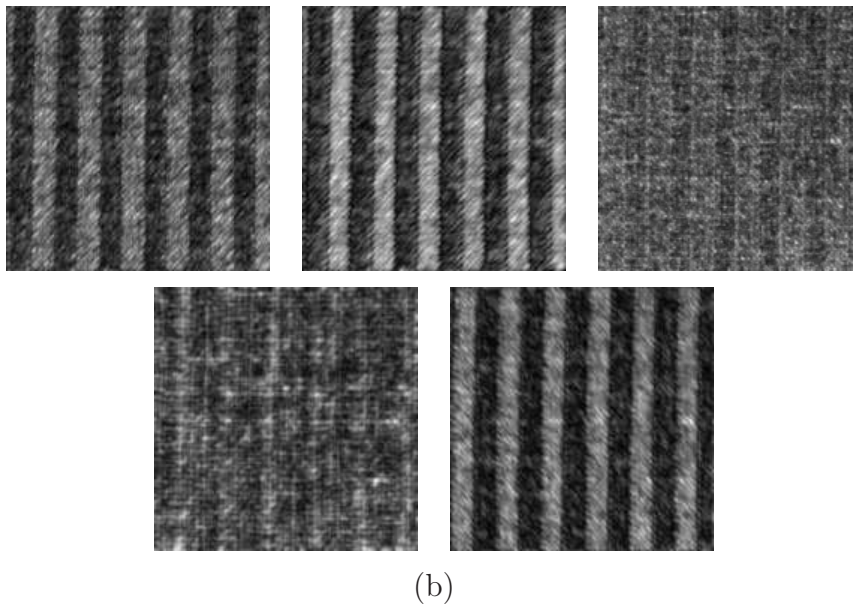
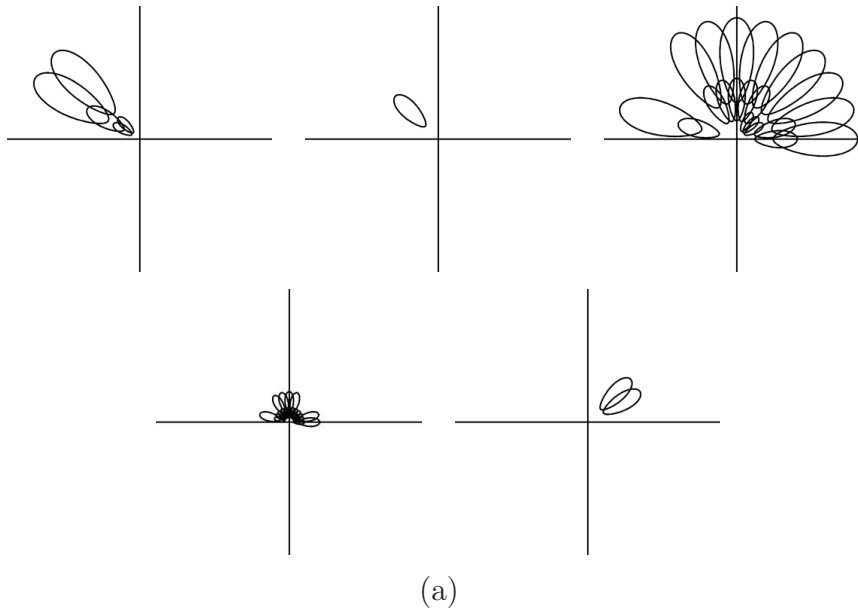


Figure 4.11: (a) Result of frequency feature clustering of image in top of figure 4.8, using dissimilarity D_{RGFF} . Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. (b) Reconstruction of the visual patterns associated to clusters in the left column.

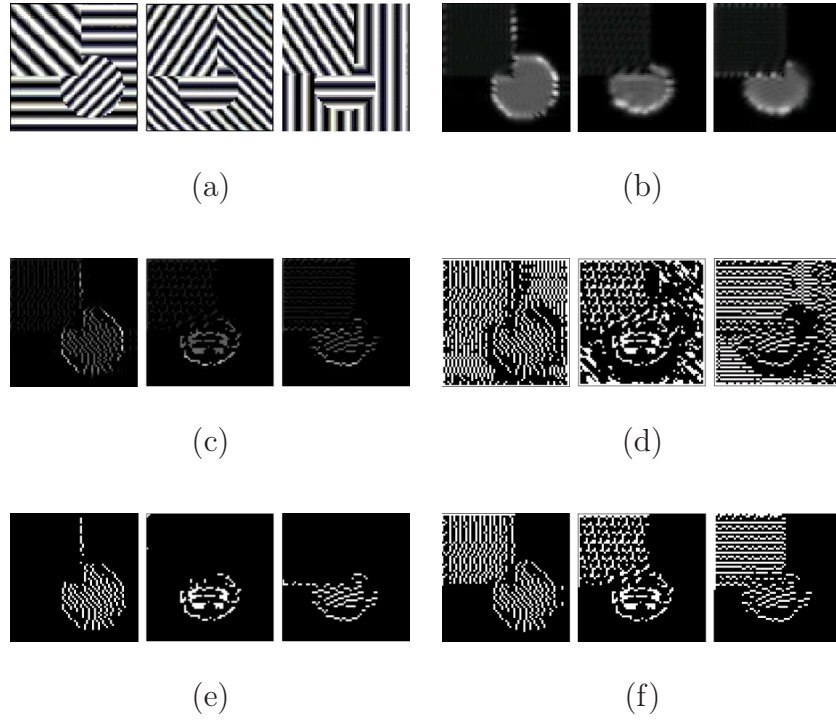


Figure 4.12: (a) Cross sections of a 3D image synthesized from three grating patterns with different orientations. (b) Energy map of the response to filter with $(\lambda = 4, \phi = 0, \theta = 0)$. (c) Image in figure b after non-maxima suppression. (d) Binary image obtained for image c by setting all maxima to white. (e) Image c after thresholding using a threshold selection that eliminates a fixed fraction of maxima depending of the scale of the filter. (f) Image c after thresholding using a threshold selection that eliminates a fixed fraction of the total energy of the image depending of the scale of the filter.

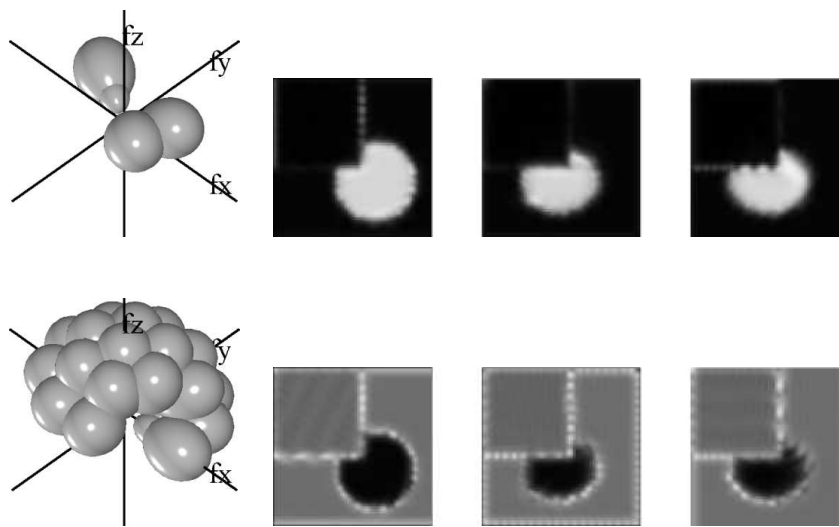


Figure 4.13: Visual pattern partitioning of image in figure 4.12.a using dissimilarity D_{NI} . *Left Column:* The two clusters obtained. Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Right Column:* Cross sections of their correspondent reconstructed visual patterns.

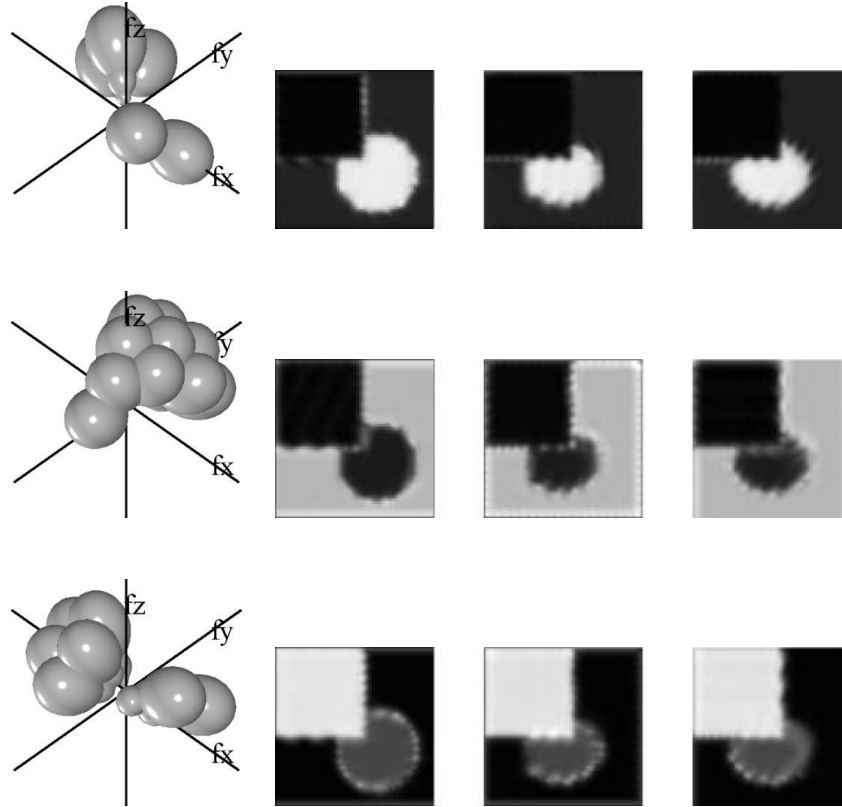


Figure 4.14: Visual pattern partitioning of image in figure 4.12.a using dissimilarity D_{NI}^* and threshold strategy as in figure 4.12.e. Other threshold strategies produce the fusion of filters associated to different gratings. *Left Column:* The three clusters obtained. Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Right Column:* Cross sections of their correspondent reconstructed visual patterns.

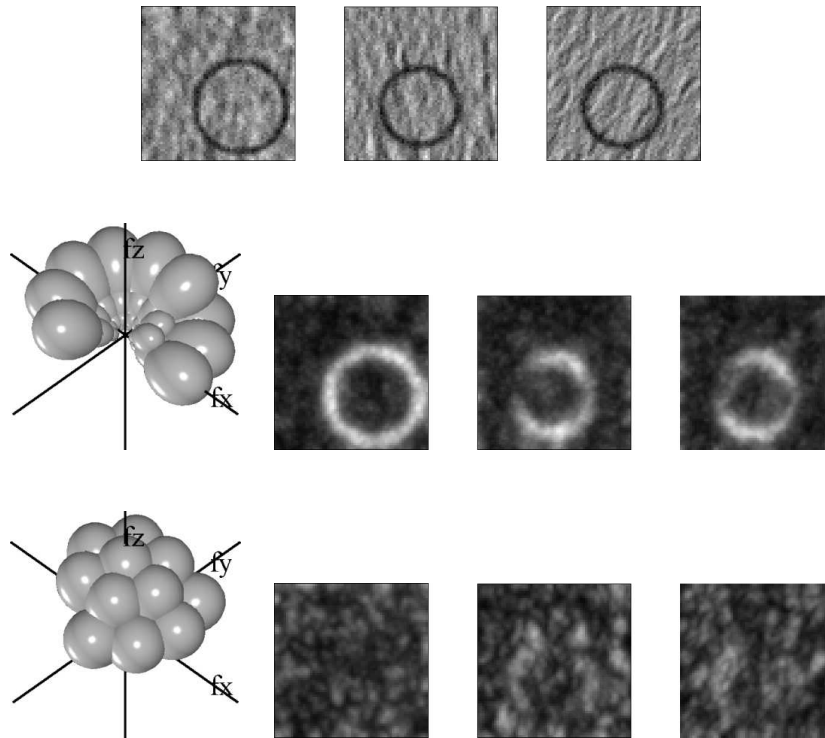


Figure 4.15: Influence of orientation overlapping parameter $\Delta\phi/N$. *Top Row:* Cross sections of the original 3D synthetic data. *Remainder Rows:* Result of visual pattern partitioning using dissimilarity measure D_ρ and $\Delta\phi/N = 1.2$. *Left Column:* The two clusters obtained. Filters are represented as the iso-surfaces of level $\exp(-1/2)$ of their transfer functions. *Right Column:* Cross sections of their correspondent reconstructed visual patterns. The reduction of orientation overlapping, using for example $\Delta\phi/N = 1.5$, produces the fragmentation of the sphere boundary in the first visual pattern (in second row).

CHAPTER 5

Applications

The presented representation model has been developed in the framework of a more ambitious project. We are working towards a complete and coherent system for analysis and reconstruction of 3D data. The system will comprise a set of stages implying a top-down strategy. It can be described as the sequence of initialization, deformation and surface reconstruction of an implicit representation model. The surface reconstruction stage is currently quite mature (Leborán et al., 2003; Leborán and Pardo, 2003; Leborán et al., 2004). We have also developed several works on initialization by matching with a priori shape models (Dosil et al., 2000, 2001a; Rodríguez et al., 2002; Dosil and Pardo, 2003). However, some problems appear with that solution.

This chapter is mainly focused on the application of the composite feature representation model for the initialization of active models to improve segmentation. In its subsections, we first make an introduction to segmentation with active models and then we tackle the problem of initialization, our previous approach to it and the difficulties encountered with it. Before presenting the new proposal for active model initialization with visual patterns, the specific active model employed, the geodesic active model, is described. Finally, the initialization is presented and illustrated by some examples. In the second section of this chapter we propose alternative applications of the presented representation model.

5.1 Medical Image Segmentation Improvement

Low level feature detection has an important area of application in 3D medical imaging. In this field the input is a volumetric data set representing a certain property of the matter –depending on the acquisition method– obtained from anatomical organs. The diverse analysis methods employed in medical image make use of the different kinds of low level features. In some cases, low level features by themselves can be helpful to the physicians for diagnosis. More often, the low level patterns are the basis to other applications, like segmentation, multimodal image registration, visualization or measurement of morphometric anatomical properties.

In particular, segmentation techniques handle different cues of information to delineate the several objects of interest appearing in an image. They are traditionally classified into region based and boundary based, depending on the type of information they use to obtain the final surface or volume model. Examples of boundary based methods are boundary completion and deformable surface models. Region based methods use region information like intensity or texture. In this class are included simple thresholding methods, region growing and region based deformable models. Nevertheless, a desired property of a segmentation technique is that it can take advantage of different sources of information to enhance its power to discriminate among the objects present in an image and, thus, optimally determine their shape.

Among segmentation techniques, active model methods (Kass et al., 1988; McInerney and Terzopoulos, 1996; Caselles et al., 1997) stand out due to their ability to integrate information from different features. The concept of deformable or active model refers to any model that changes its configuration to fit the image data but obeying a set of shape constraints to guaranty properties like continuity, smoothness, etc. There are many techniques that follow this definition, with diverse model representations and optimizations algorithms. Deformable models can describe by the region they occupy in space or its surface and, to this end, many representations can be used, like finite elements, B-splines, level sets, etc. The most commonly applied optimization techniques are energy minimization, parametric surface evolution and front propagation of a implicit surface. There is an interrelation among these formulations (Caselles et al., 1997). They all can be seen as a system subject to a set of forces originated by some potential energies, some of them internal, representing image constraints, and some external, originated by image features. The optimal

model is the one that reaches the equilibrium among the forces or, from a different point of view, the minimum of the system's total energy.

5.1.1 Active Model Initialization

Most active model methods derive image potentials from local features. Therefore, the initial state of the model needs to be close to the desired object boundary for the segmentation to succeed. A typical solution is to automatically define a window function enclosing all objects. However, the presence of non target objects may deflect model evolution. Furthermore, many types of active models are not able to converge to complex shapes if this imply large deformations, like in the case of concavities or sharp cusps. Therefore, the model should start from state that is already quite close to the target object in location and shape. To avoid this, the model is usually initialized by user interaction. This is difficult to realize with 3D solid data, where one can not see all the data at a time, but has to figure out image contents from slice views.

In (Dosil et al., 2000, 2001a; Rodríguez et al., 2002; Dosil and Pardo, 2003), we have already proposed an automatic technique for the initialization of model position and shape by matching with prior models. In those works, a method for matching of low level features with a priori shape models was suggested, for subsequent deformation by means of active models. The process is illustrated in figure 5.1. The idea of the technique is to benefit from either global a priori information, local image information and local shape constraints. Model matching allows for reutilization of domain specific information by fitting a global –structural– shape model to the low level information of a new image instance. This provides the method with robustness. Segmentation by deformable models does the rest, by locally adapting itself to image features but imposing local shape constraints.

In the aforementioned work on model matching, prior shape was represented by a Constructive Solid Geometry (CSG) model, based on superquadric implicit functions with global deformations. These shape primitives offer an interesting balance between simplicity and flexibility. Thus, a priori information is relative only to global shape. The CSG model was fitted to boundary information on the image. The model was optimized by minimization of a distance between the model and boundary points. To eliminate the influence of non target boundaries, surface patches were classified according to shape, contrast and size properties, associated

to patch average curvatures, gradient and area respectively. Those properties were introduced in the evaluation of the distance function as weights on the contributions of each point belonging to a given patch. A method for robust estimation of differential local properties of the image was developed, which eliminated noise influence in the calculations, by means of corner and gradient preserving anisotropic diffusion filtering (Dosil et al., 2001b; Dosil and Pardo, 2003). However, in complex shapes it is difficult to discriminate between patches belonging to the target object. The low level features belonging to different objects appear together in the same feature map. Average differential properties are not enough to discriminate among the patches belonging to different objects. Moreover, prior information is not always available.

On the contrary, the presented low level representation method produces a disjoint representation of visual patterns, separated from each other. We will show that the visual patterns isolated by our method can be used to place the initial model near the target object. To this end, we have designed an initialization procedure consisting on manually selecting, among the results generated by visual pattern extraction, the ones that best represent the object to be segmented and defining the initial model from them. We will only evaluate the contribution of our approach to the solution of the initialization problem, not the performance of the segmentation model itself, although, naturally, segmentation precision also depends on the employed active model and the definition of image potentials. Of course, more robust methods can be achieved when information on the application domain can be exploited. Still, the proposed solution can be applied when no a priori information is available and here we focus in this situation. However, the presented low level representation is compatible with prior model matching, as we will propose for future work.

5.1.2 Geodesic Active Models

To accomplish medical image segmentation, here we have chosen the use of an implicit representation for object boundaries, where the surface is defined as the zero-level set of an implicit function. Implicit active models present important advantages regarding parametric representations. The problem of surface re-sampling when stretching, shrinking, merging and splitting is avoided. Implicit representations

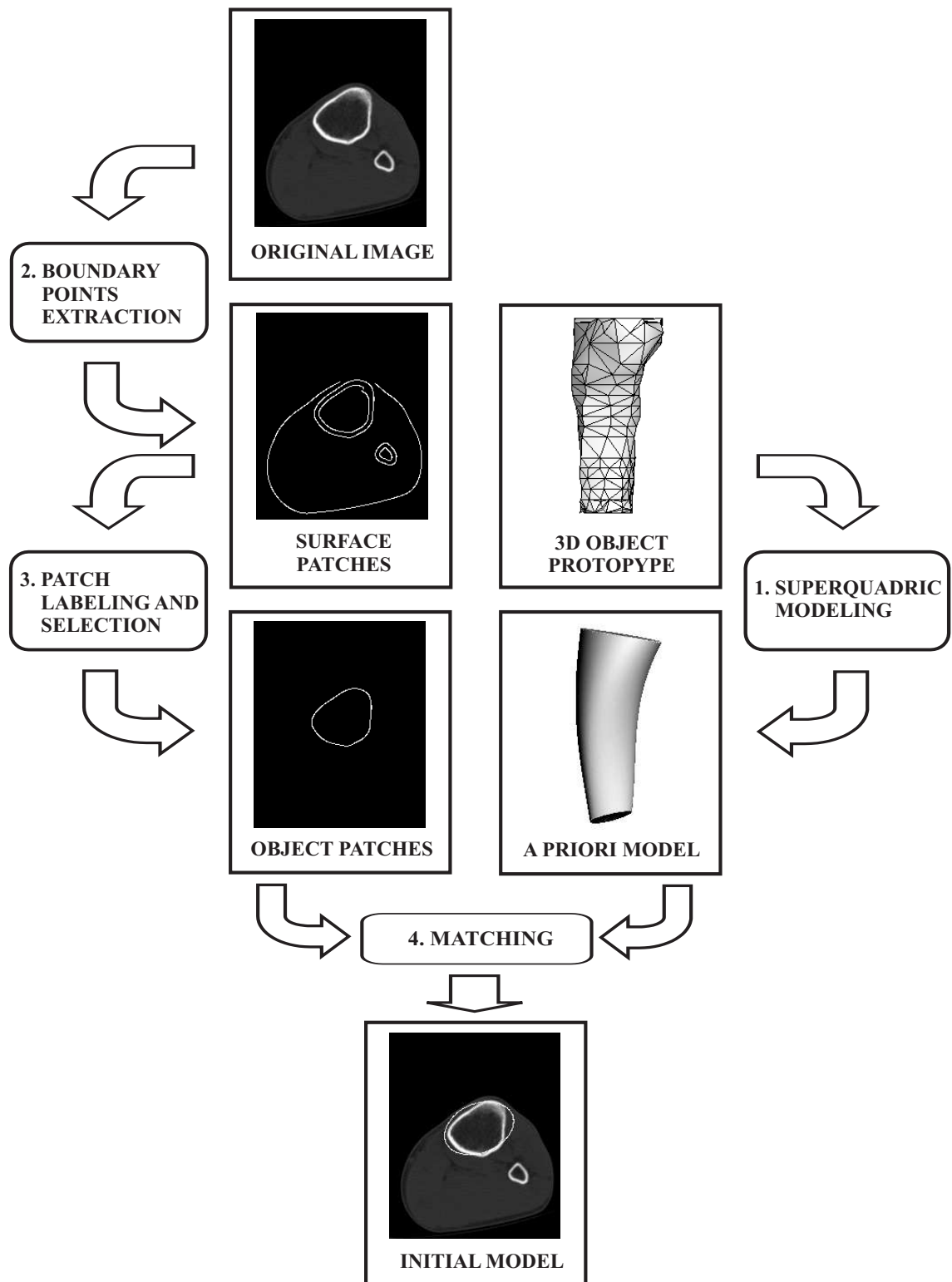


Figure 5.1: Scheme of the process of initialization of active models using matching with a priori global shape models.

allow for the simultaneous detection of inner and outer surfaces of an object and naturally manage topological changes. Inner and outer regions are determined by the sign of the implicit function. Another reason to choose an implicit representation is to facilitate surface reconstruction for visualization by using the software developed in our research team for fast triangulation of implicit surfaces (Leborán et al., 2003; Leborán and Pardo, 2003; Leborán et al., 2004).

The optimization model employed for segmentation is the geodesic active model (Caselles et al., 1997). The method can be formulated as an initial value problem of finding the surface of minimal weighted Lebesgue measure –in this case the area– in a Riemannian space with a metric derived from the image content. Instead of minimizing the Euclidean definition of area, it minimizes a different definition of area that takes into account image characteristics, like edges. Area minimization by the steepest descent method gives place to a parametric surface evolution equation in partial derivatives, like that in (Kass et al., 1987). The implicit formulation results from embedding the parametric surface into an implicit function as its zero level set.

If $S(t) = S(v, w, t)$ is the parametric surface, with $v, w \in [0, 1]$ being the natural parameters and $u(\mathbf{x}, t)$ is the implicit function, defined in the image domain Ω , then the evolution of the surface is determined from the evolution of the zero-level set of implicit function, i.e., $S(t) = \{\mathbf{x} \in \Omega : u(\mathbf{x}, t) = 0\}$. The equations governing the evolution of the implicit function are the following.

$$\begin{aligned} u(\mathbf{x}, 0) &= u_0(\mathbf{x}) \quad \text{on } \Omega \\ \frac{\partial u}{\partial t} &= |\nabla u| \left(\operatorname{div} \left(g(X) \frac{\nabla u}{|\nabla u|} \right) + c \cdot g(X) \right) \quad \text{on } \Omega \times (0, \infty) \end{aligned}$$

where c is real constant, $g(X)$ is a function with values in the interval $[0, 1]$ that decreases in the presence of relevant image features and

$$u_0(\mathbf{x}) = \begin{cases} d(\mathbf{x}, S(t=0)) & \text{if } \mathbf{x} \text{ is inside } S_0 \\ 0 & \text{if } \mathbf{x} \in S(t=0) \\ -d(\mathbf{x}, S(t=0)) & \text{if } \mathbf{x} \text{ is outside } S(t=0) \end{cases}$$

and $d(\mathbf{x}, S(0))$ is the distance between a point in Ω and the initial curve $S(0)$.

Provided that the curvature κ is defined as

$$\kappa = \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) = \frac{u_{xx}u_y^2 - 2u_xu_yu_{xy} + u_{yy}u_x^2}{(u_x^2 + u_y^2)^{3/2}},$$

then, the partial differential equation for u can be expressed as a sum of the contributions of a curvature term and a gradient term.

$$\frac{\partial u}{\partial t} = g(X)|\nabla u|\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) + \nabla g(X) \cdot \nabla u + c \cdot g(X) \cdot |\nabla u| \quad (5.1)$$

$$= g(X)|\nabla u| \cdot (\kappa + c) + \nabla g(X) \cdot \nabla u. \quad (5.2)$$

The first term on the right can be identified as the *geometric active model*, where $g \cdot (\kappa + c)$ represents the velocity of the evolving surface. κ can be interpreted as a geometry dependent velocity. In the diffusion filtering framework, it can be also consider a diffusion term with a regularizing effect (Kornprobst et al., 1997). Its effect is also equivalent to the internal forces in a thin-plate-membrane spline model, also called *snake* (Kass et al., 1987). Constant c represents a constant velocity or *advection* velocity in the geometric active model and is equivalent to a balloon force in the snake model. Factor $g(X)$ has the effect of stopping the surface at the desired feature locations.

The second term is the image dependent term, that pushes the level-set to image features. It is analogous to external forces in the snake formulation. This term did not appear in the geometric active model, which made necessary the use of a constant velocity term to approach the level-set to the object boundary. This the use of a feature attraction term this is no longer necessary. However, if the model is initialized far away from the image features to be segmented, the attraction term may not have enough influence over the level-set. As a result, the constant velocity term is often used to compensate for the lack of an initialization stage.

5.1.3 Initialization with Visual Patterns

In the present application, the initial model has been defined from choosing, among the visual patterns generated by our partitioning method, the one that best represents the target object. The selection of this visual pattern is done manually by the user. Visual patterns are constructed from composite-feature detectors by computing the linear summation of the real or half-wave rectified real part, depending on the case, of the responses of its filters. The use of the sum of amplitudes, i.e., the squared sum of the real and imaginary parts, would produce less precise initializations. The initial state of the geodesic active model is defined from the selected cluster by simply re-scaling the intensity levels of its response to the interval $[-1, 1]$.

This may seem an arbitrary criterion. However, when the selected visual patterns assign high energy values to the target patterns, such a re-scaling results in a quite acceptable approximation of the target surface. The active model will do the rest.

The concrete implementation of the geodesic active model used here is the one described in (Weickert and Kühne, 2003). We do not employ balloon forces either, since with the initialization they are no longer needed, so then $c = 0$. We also apply the same stopping criterion based, on a decreasing function of the smoothed gradient modulus of the original image.

$$g(X) = \frac{1}{1 + |\nabla G_\sigma * X(\mathbf{x})|^2 / \lambda^2},$$

where ∇G_σ is the derivative of Gaussian operator at scale σ , which here equals 2 and λ real positive parameter that weights the estimated gradient. It can be considered a gradient threshold. If $|\nabla G_\sigma * X(\mathbf{x})| \gg \lambda$ then $g(X) \rightarrow 1$ and the stopping factor has no effect. If $|\nabla G_\sigma * X(\mathbf{x})| \ll \lambda$ then $g(X) \rightarrow 0$ and, hence, $\partial u / \partial t = 0$ so the steady state is achieved.

In figure 5.2 the initialization and segmentation processes are illustrated using an example case in 2D to simplify visualization. It is an histology image of granulosa cell tumors of ovary. The target objects are immersed in a textured background and have intensity levels and contrast very close to other structures in the image. The stopping criterion of the model is computed from the original image as a function of its gradient modulus. Therefore, the model stops evolving when it arrives to a strong local maxima of the energy, so that the segmentation without initialization produces a large amount of contours in the background texture. The composite-feature partitioning method produces three visual patterns for this image. The system is acting as blob, texture and line detector at the same time. The cluster of filters selected for initialization of the geodesic model is the one acting as a blob detector. It contains large scale filters, eliminating high frequencies correspondent to texture details and also orientations relative to the sample tissue contour. Since the geodesic model starts placed very near the contour of the target object, it easily evolves to the correct local maxima of the external energy.

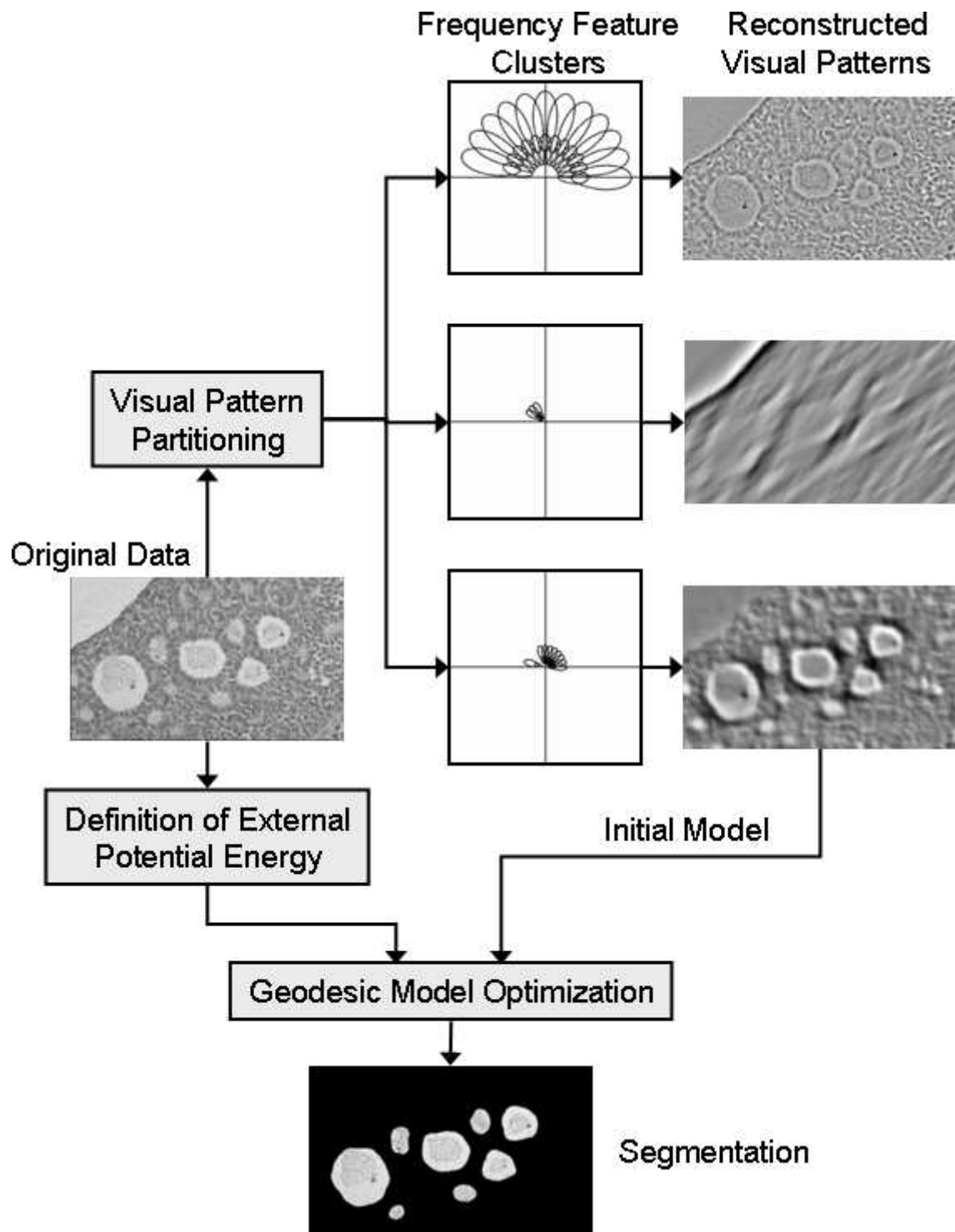


Figure 5.2: Scheme of the process of initialization of geodesic deformable models using the presented visual pattern partitioning method. The visual pattern selected as initial model is selected manually.

5.1.4 Segmentation Results

In this section some results of medical image segmentation are presented to illustrate the benefits of initialization with visual patterns. All test data are digital images with 256 gray intensity levels. The method has been applied to these images using the filter bank settings detailed in section 3.1, with $N_s = 4$, $\lambda_{\min} = 4$ and $N_\phi = 6$. This gives place to wavelengths $\lambda_i = 4, 8, 16, 32$, $\sigma_\alpha = 25^\circ$ and a total number of orientations $N_o = 23$. The dissimilarity measure used is the correlation based distance D_ρ defined in 4.2.1.4.

The first segmentation example corresponds to real 3D medical imaging data. It is a volumetric data set of size $32 \times 64 \times 20$ from Single Photon Emission Computerized Tomography (SPECT) of the left ventricle of the heart. This modality of imaging is quite blurry and presents textured background due to the surrounding tissues and also to the acquisition process itself. The goal is to isolate the ventricle wall from the background. The image and the results obtained with the visual pattern extraction method are presented in figure 5.3. The method detects two visual patterns. One of them gathers the background artifacts, including the ramp surfaces produced by the image acquisition method. The other one mainly represents the ventricle wall and will be the one selected to initialize the active model.

Figure 5.4 is presented to show the impossibility of correctly segment the ventricle wall just by simple thresholding, as it might be thought. More sophisticated segmentations techniques are needed. However, the geodesic active model without initialization is not able to correctly segment the ventricle wall either, as can be seen on top of figure 5.5. The model can not evolve to the wall cavity without overcome the external wall face. Images on bottom of figure 5.5 the result of geodesic active model segmentations using the visual pattern in figure 5.3.d to generate the initial model. Such visual pattern is already a good approximation of the ventricle wall, so the model converges fast and captures both internal and external faces of the wall without any problem.

The second segmentation example is a 3D data of size $71 \times 107 \times 35$, correspondent to a phantom T2-weighted MR image of the lateral ventricles of the brain. 5.6 shows the image and the result of visual pattern decomposition. Although it is quite difficult to analyze the results in such complex 3D data, it can be seen that the first visual pattern, in figure 5.6.e, enhances the corpus callosum, a structure that produces a very weak signal in the T2-weighted MR imaging modality, but can

be identified in the sagittal section in figure 5.6.b as the arched region on top of the ventricles. The second visual pattern, in 5.6.f, has its largest response in the region occupied by the lateral ventricles.

Again, we present thresholding results to clarify the complexity of the segmentation problems. The result of thresholding at the intensity level of the corpus callosum is shown on top of figure 5.7. The corpus callosum has the same level of response to the T2-weighted MR imaging that the lateral regions it interconnects and the septum between ventricles. Therefore, intensity alone is not enough to discriminate it. Note that the surface in figure 5.7.b correspondent to ventricles and to cingulate gyrus are actually a hole in the segmented region, represented in white in figure 5.7.a. The result of thresholding at the intensity level of the corpus callosum is shown on bottom of figure 5.7. As can be seen, such segmentation captures also the fissures separating the two brain hemispheres besides the ventricles. Gradient information is not sufficient to segment the parts of that image. The segmentation using a geodesic active model without initialization, guided by image gradients, is shown on top of figure 5.8. The result captures parts of the corpus callosum, the ventricles and the caudate nucleus. Each of these structures has strong gradients in their external surfaces that attract the geodesic active model with similar strength. As a conclusion neither intensity nor contrast are sufficient to discriminate the different parts of the image.

The results of geodesic active model segmentation with initialization is shown in center and bottom rows of figure 5.8. Each of the two visual patterns extracted for that image is used to initialize one geodesic active model respectively. As a result, both the corpus callosum and the lateral brain ventricles can be independently segmented, although the segmentation of the corpus callosum captures also the fornix of the brain, which has similar frequency composition.

The third segmentation example is a 3D data set of size $87 \times 42 \times 44$, correspondent to a real CT image of the brain. In this imaging modality the brain ventricles produce a very weak signal. The corpus callosum has the same intensity level as all surrounding tissues apart from the ventricles, including the cingulate gyrus. The only differentiation now the sulci between them which in this particular image presents low contrast. Visual pattern decomposition applied to this image produces two clusters, one of them mainly representing the brain ventricles and attenuating other structures with the same intensity level and contrast, as can be seen in figure 5.9.c. The second visual pattern, in figure 5.9.b, captures the cor-

pus callosum contour and other brain sulci. The segmentation obtained using the second cluster, in figure 5.9.c, as initial geodesic model captures the surface of the ventricles, while the model obtained for a square box initialization is not able to overcome other brain structures.

5.2 Other Applications

As has been repeated during this dissertation, the developed model is intended for general purposes. In this section we propose other applications for our composite-feature representation model and illustrate them with some examples. The results presented here have not been combined with higher level analysis techniques, although further processing could be applied in all cases to obtain more robust solutions. The two applications proposed can be basically considered pattern enhancement tools.

5.2.1 TMRI Data Enhancement

Tagged Magnetic Resonance Imaging (TMRI) is an image modality to study the motion of material points on the left ventricle of the heart. MR tagging uses a special pulse sequence to spatially modulate the longitudinal magnetization of the subject, prior to acquiring image data. Over many heartbeats in a single breath hold, enough data can be acquired to reconstruct an image sequence in which the tag pattern is deformed by the underlying motion of the heart. This process is being applied both to 2D and 3D MR data, producing $2D+t$ and $3D+t$ data sets respectively. The tagging process may produce saturated planes in the image, orthogonal to the image axis, or sinusoidal patterns, called SPATial Modulation of Magnetization (SPAMM) tag patterns.

Besides the patterns themselves, when using SPAMM tag patterns, one important feature to extract for motion analysis is the phase of the sinusoidal signal. By tracking constant phase points one can obtain the trajectories of material points in the ventricle wall (Osman et al., 2000). To obtain phase information the frequency components of the sinusoidal signal are isolated in the spectrum and then the pattern is reconstructed. Phase information is obtained from the complex reconstructed pattern, producing the so called HARmonic Phase (HARP) image. The character-

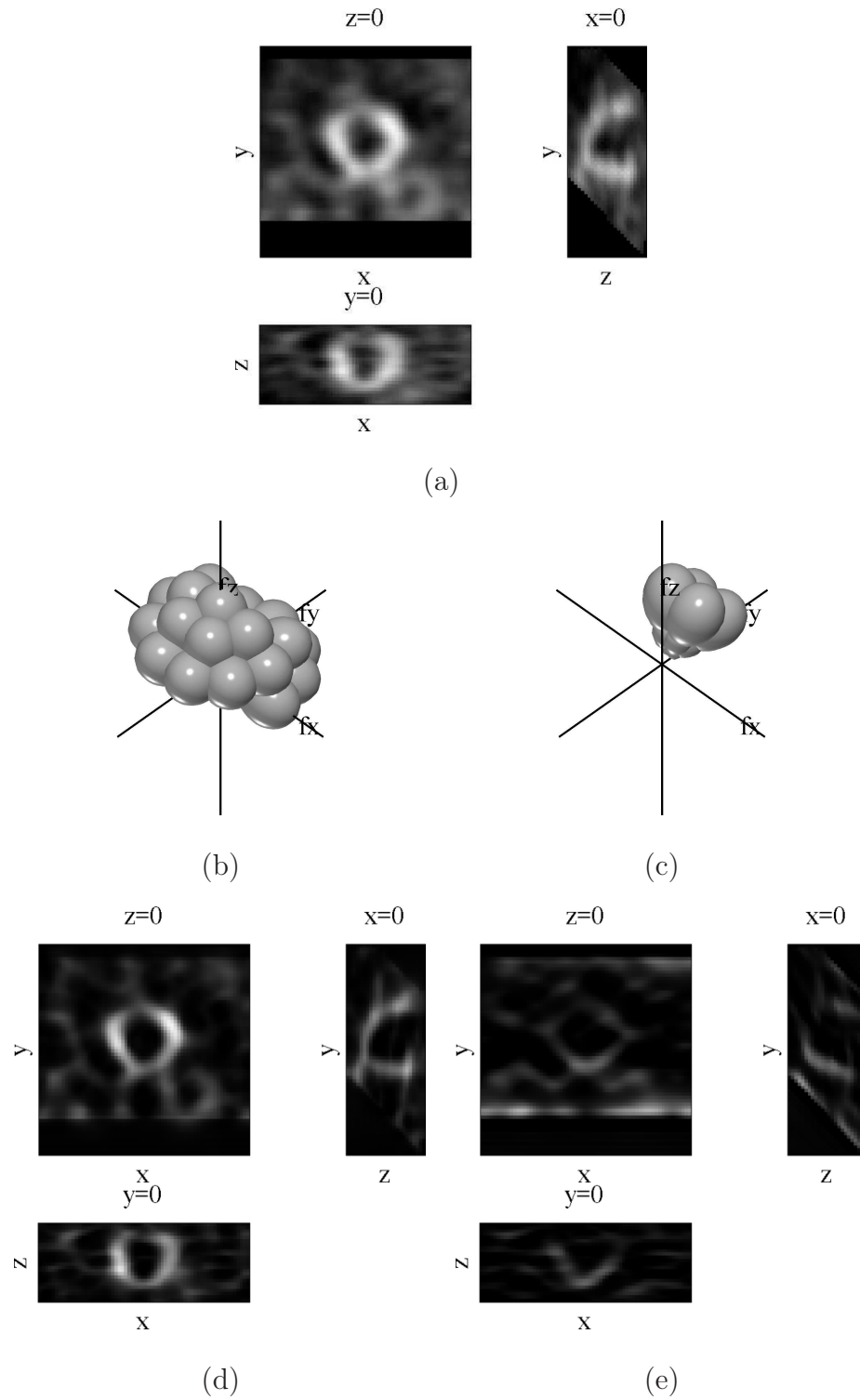


Figure 5.3: (a) Cross sections of a 3D data set corresponding to a SPECT image of the left ventricle of the heart. (b, c) Clusters identified by the system. (d, e) Cross sections of the visual patterns correspondent to clusters in figures b and c respectively, reconstructed by linear summation of the half-wave rectified real part of the filters' responses.

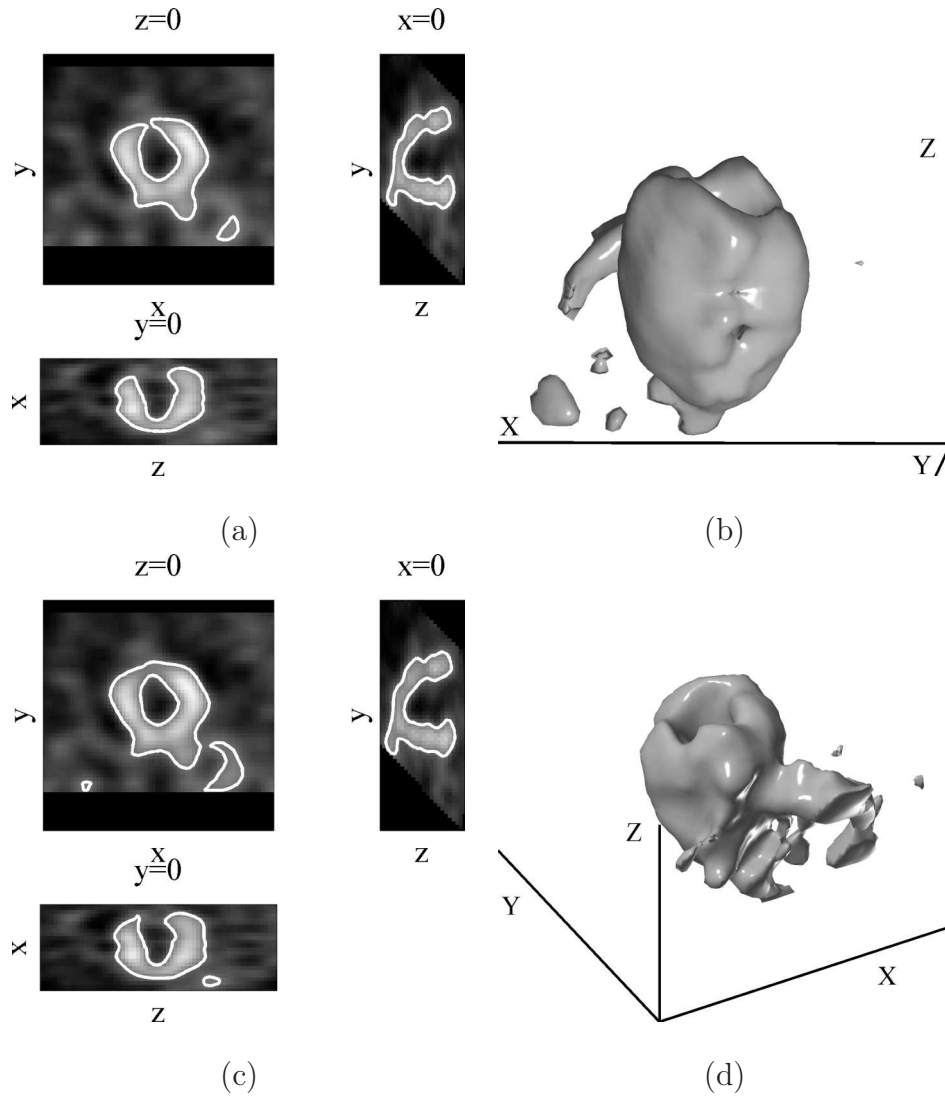


Figure 5.4: Thresholding of image from figure 5.3.a. (a) Cross sections of original data with the intersections of isosurface of gray level 100 in white color. There is a hole in the ventricle wall. (b) 3D reconstruction of the surface in figure a. (c) Cross sections of original data with the intersections of isosurface of gray level 90 in white color. New artifacts appear in relation to previous thresholding. (d) 3D reconstruction of the surface in figure c.

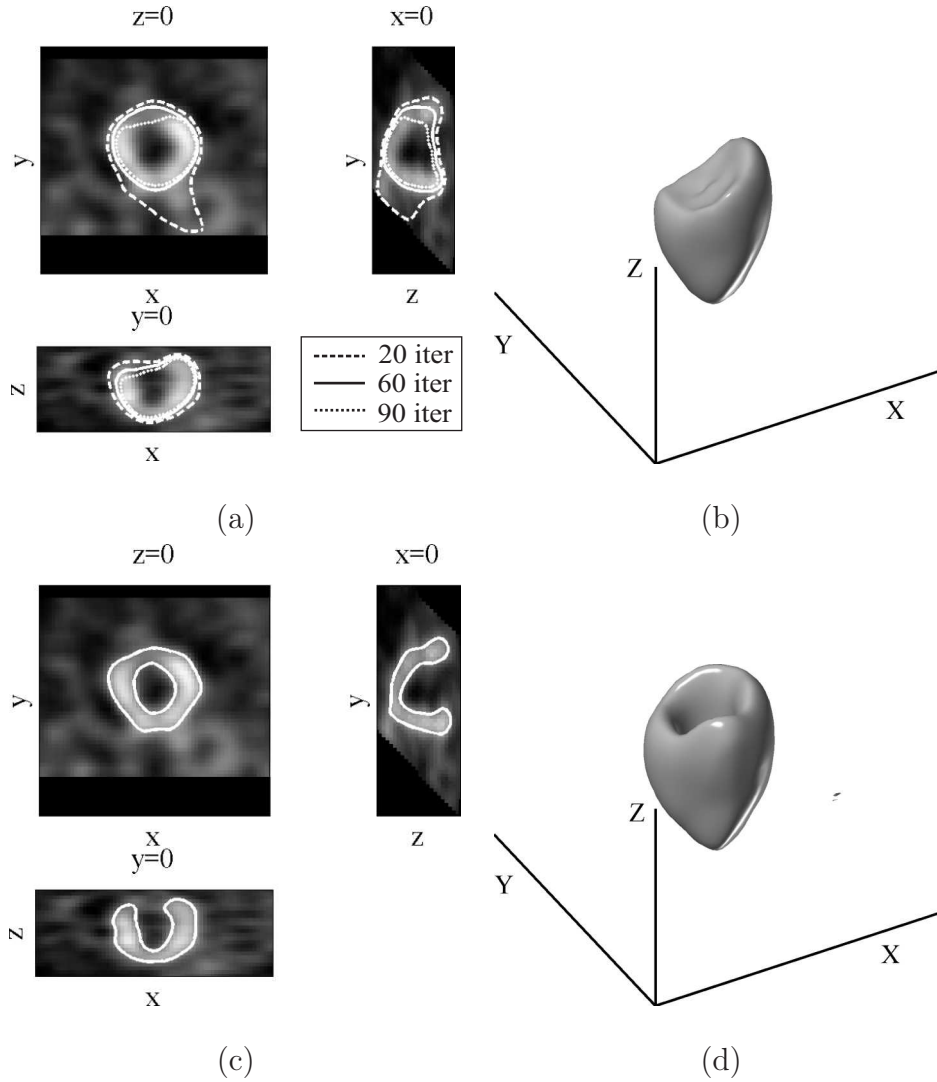


Figure 5.5: Segmentation of image from figure 5.3.a. (a) Cross sections of original data with the intersections of the left ventricle segmentation at several time steps superimposed in white color, obtained using a square box as initial geodesic model. (b) 3D reconstruction of the surface correspondent to the segmentation in figure a with 60 iterations. (c) Cross sections of original data with the intersections of the left ventricle segmentation in white color, obtained using the visual patten in from image 5.3.e as initial geodesic model. (d) 3D reconstruction of the surface in figure c.

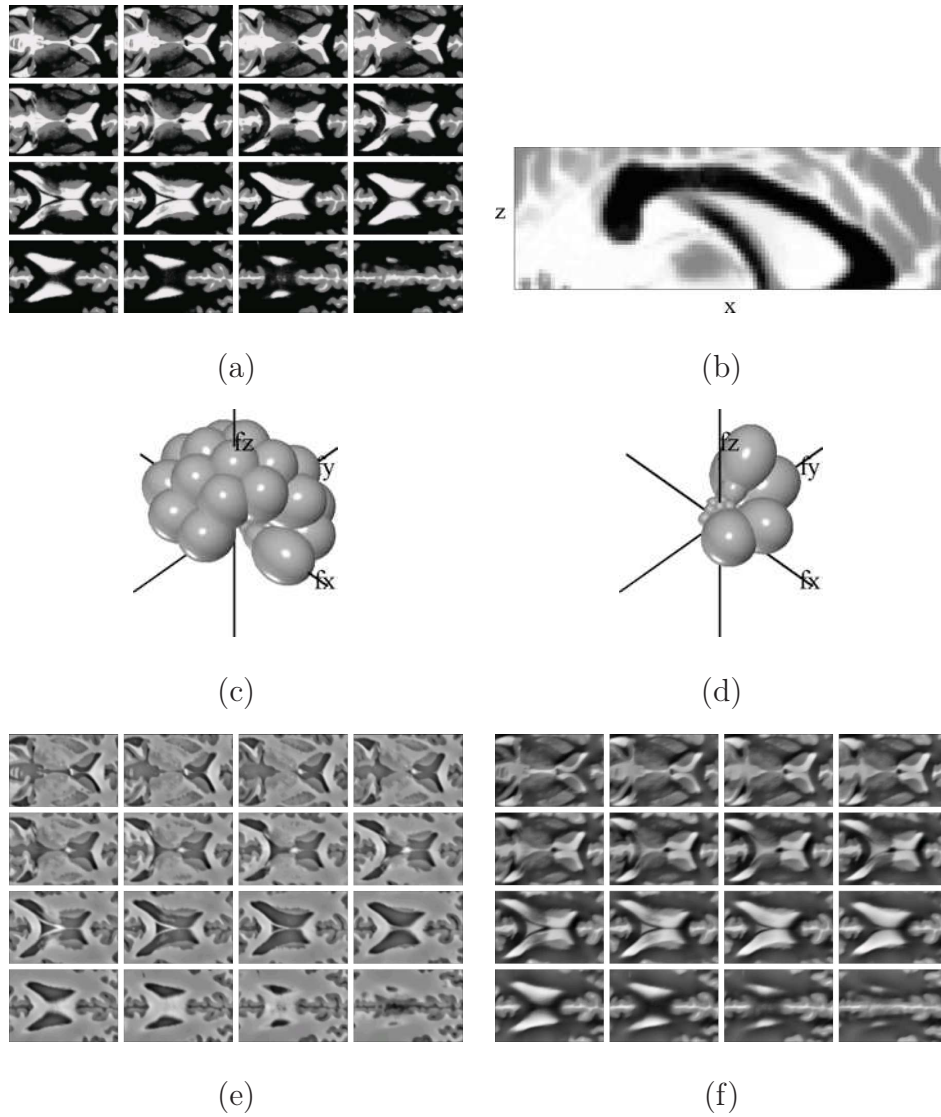


Figure 5.6: (a) Sample horizontal slices of a phantom T2-weighted MR image of the brain ventricles and the corpus callosum. (b) Sagittal section of the data set. (c, d) The two clusters generated by the method. (e, f) Reconstruction of the visual patterns correspondent to clusters in figures c and d respectively –image f is inverted– by linear summation of the filter’s responses real component of each cluster.

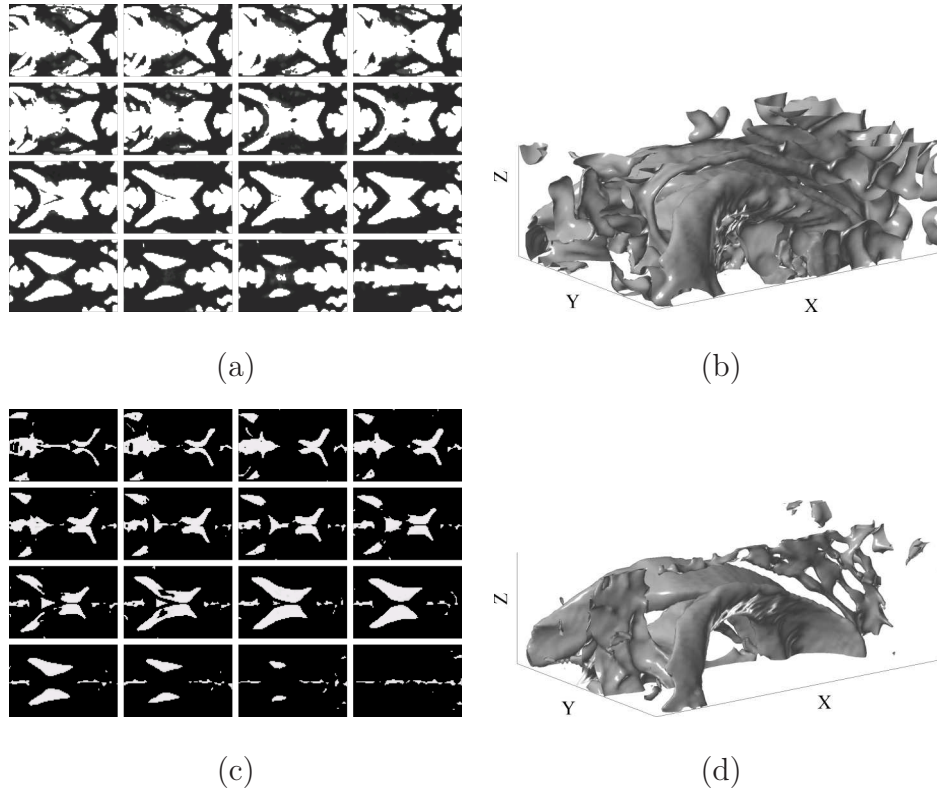


Figure 5.7: Thresholding of image from figure 5.6.a. (a) Thresholding using level 200, correspondent to the corpus callosum. Larger levels are set to 255 –white. (b) 3D reconstruction of the isosurface of level 200. (c) Thresholding using level 240, correspondent to the brain ventricles. Smaller levels are set to 0 –black. (d) 3D reconstruction of the isosurface of level 240.

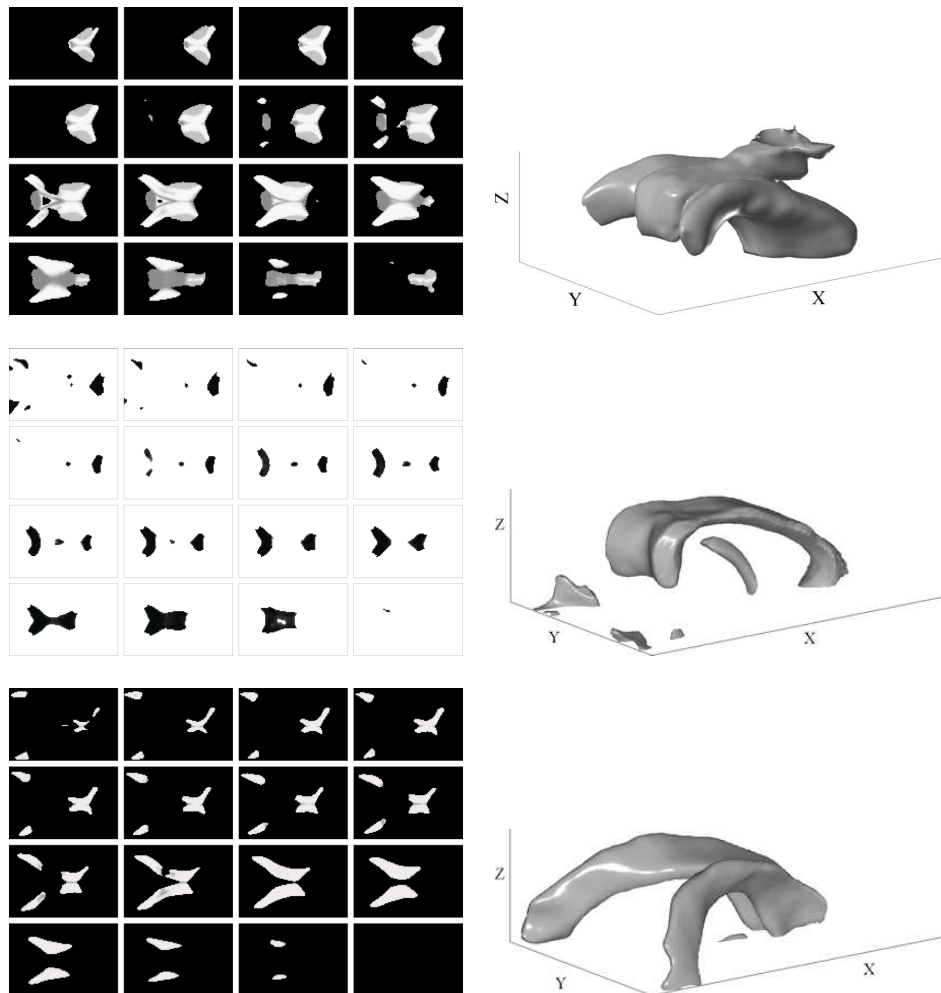


Figure 5.8: Segmentation of 3D data set from figure 5.6 using different definitions of the initial model. *Top Row:* Using a square box. *Center Row:* Using visual patterns in 5.6.d. *Bottom Row:* Using visual pattern in 5.6.e. *Left Column:* Cross sections of the segmented data. The region outer to the final geodesic model has been set to white in the first and third rows and to black in the second to enhance visualization. The inner region presents the original data. *Right Column:* 3D reconstruction of the zero level sets.

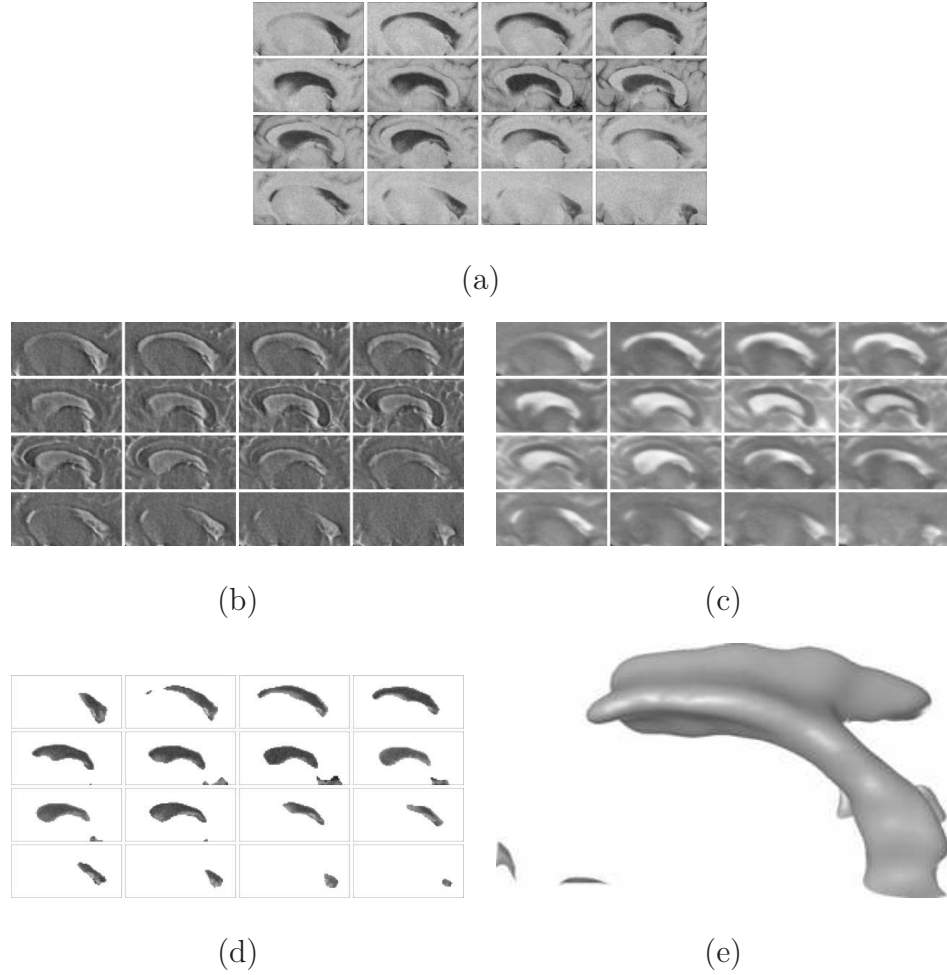


Figure 5.9: Segmentation of a 3D real MR image of the brain ventricles and the corpus callosum. (a) Sample slices of the original data. (b, c) Slices of the two 3D visual patterns generated by the method, obtained by linear summation of the filter's responses real component of the selected cluster –image b is inverted. (d) Sample slices of the ventricles segmented using the visual pattern in image c as initial model. The outer region of the object is represented in white. (e) 3D reconstruction of the surface from image d. In the 3D reconstruction, the aspect ratio is 1:1:0.5 to provide a more realistic visualization, since the resolution in the z axis is lower.

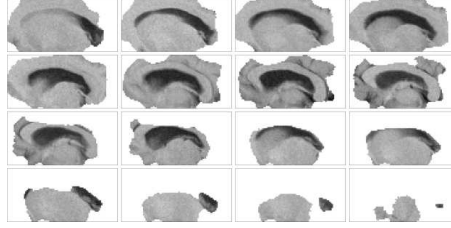


Figure 5.10: Sample slices of the result of segmentation of the data set from 5.9.a using a square box as initialization. The outer region of the object has been represented in white.

istics of this process made it very suitable to accomplish using our representation model.

5.2.1.1 Examples

Figure 5.11.a shows an example of 2D+ t TMRI data of a section of the left ventricle of the heart. The visual pattern extraction method produces two clusters, represented in figures 5.11.b and c. The reconstructed visual patterns, calculated as the sum of half-wave rectified real components of cluster's filter responses, as shown in figures 5.11.d and e. It can be seen that one of the visual pattern captures the tag patterns while the other captures boundary information.

We have reconstructed in figure 5.11.e the isosurfaces of the first visual pattern with a level that is a 20% of its maximum. The vertical axis represents time. From this simple reconstruction of the ventricle wall one can get an idea of the motion patterns involved. The contraction movement was already evident from the 2D slices. Note also the lateral bending stripes in the upper side of the ventricle wall. They indicate that there exists a rotation around the center of the ventricle cavity. This rotation is known to be produced with different angles at different axial cuts, thus producing a twisting of the wall. This twisting can not be accounted in regular MRI imaging.

Figure 5.11.f presents a reconstruction of the phase of the first of the two visual patterns obtained. To calculate phase, the response to the cluster is firstly recalculated as the summation of complex filter's responses. The phase of the complex visual pattern obtained is calculated. In the representation of figure 5.11.f, white color represents a phase of π radians and black represents $-\pi$ radians. The image presents a *blocky* aspect because the patterns are not purely SPAMM, but are quite

saturated. However, in the xt and yt cuts the lines of constant phase can be clearly observed. They represent the trajectories of material points.

A second example is shown in figure 5.12. In this case the visual pattern extraction method has produced five clusters. Only the one correspondent to the tag pattern is show in the figure. The remainder visual patterns are low frequency composite features correspondent to diverse parts of the ventricle wall and a visual pattern collecting the high frequency components with orientation parallel to the time axis. The isosurface representation and the phase estimation from the extracted tag pattern are similar to the ones in the previous example.

5.2.2 GPR Data Enhancement

Ground Penetrating Radar (GPR) imaging is widely employed to explore the properties of soils and to search for buried objects in many field of applications, like geophysical studies, archeology or land demining. It works by emitting an electromagnetic wave into the ground using an antenna. GPR systems usually operate in the microwave region. Buried objects, as well as the air-ground interface, cause reflections of the emitted energy, which are detected by the receiver antenna and associated circuitry. What matters for the detection of objects in a background medium is the difference between the electromagnetic properties of the target –in particular the dielectric constant– and those on the background. Metal is a perfect radar reflector and is visible on profiles as a distinct hyperbola. However, unlike metal detectors, GPR imaging detects also many kinds of non-metallic materials.

The typical GPR reflection profile produces a two-dimensional image of a vertical slice through the ground. Sweeping the GPR system across an area of terrain allows to produce a 3D image of the soil. Reflection profiles are extremely useful in a raw format, since they are acquired in almost real-time. However, processing of the acquired data is usually necessary to clean up the data for more definitive interpretations, and to remove the patterns produced by the multiple sources of noise and artifacts. The main noise contribution comes from interferences with electromagnetic field of electronic devices, which produces horizontal banding that obscures reflections from the ground. Those bands near the top of the profiles –collected in the first few nanoseconds of the time window– are caused by antenna ringing as well as the recording of the initial radar pulse, as the antenna attempts to couple the radar energy with the ground.

Reflection profiles also contain high-angle reflections that are the tails of hyperbolas generated from buried rocks and walls. Although these reflections can be helpful in visual interpretation, because they show the tops of walls and individual rocks, they tend to produce blurry maps at lower depths. The tails of the hyperbolas should be removed, leaving only the reflections at the apex, which represent the true location where the top of the feature is buried.

In this section we present the results of applying our visual pattern extraction method to the enhancement of GPR images. We will see how typical noise patterns and artifacts are directly removed with our representation model. Specifically, we have applied the model to images of buried Anti-Personnel (AP) mines –or landmines– (Bruschini et al., 1998) and archeological sites (Conyers and Goodman, 1997).

5.2.2.1 Examples

The first example data, in image 5.13, corresponds to a GPR image of a terrain with a buried AP mine of type LI-11. This is a small metallic type of mine. It is buried at 1 cm depth. In the image of the raw data, in figure 5.14.a, the reflection profiles of buried object are very difficult to observe due the interfering noise horizontal band patterns. Observing the visual patterns isolated one can clearly distinguish the horizontal noise bands in the first one and the vertical reflection echoes in the second one, that are evident now. The 3D reconstruction of the second visual pattern shows that there exists more that one buried object that could be a mine. In this case it is a controlled medium, so it is known that the mine is placed at 1 cm depth. The sequence of echoes caused by the mine is marked with an ellipse in figure 5.13.f. The apex, signaled by an arrow, indicated the mine position. The appearing of false positives is very frequent, due to the presence of rocks or debris, which have a response profile similar to that of the mine. Hence, reliable mine detection systems involve image fusion, combining information from GPR, metal detectors, etc.

In demining applications, the GPR imaging modality has the advantage of being sensitive also to plastic mines. Second example data, in image 5.14, corresponds to a GPR image of a terrain with a buried AP mine of type PFM-1. This kind of mine is a small air delivered plastic mine with low metallic signature. The mine is buried at 1 cm depth. In the image of the raw data, in figure 5.14.a, the hyperbolic reflection profile of the mine is clearly seen. The composite feature extraction method

isolates four visual patterns here. The first visual pattern, in image 5.14.f, captures the typical air-ground interface band. The second visual pattern, in image 5.14.g captures high frequency horizontal noise. The third visual pattern, in image 5.14.h, captures the hyperbola tails of the reflection profile of the mine. This is clearer in the yz plane and in the 3D reconstruction of image 5.14.j. Finally, the fourth visual pattern, in image 5.14.i, isolates the apices of the reflection echoes of the mine. A 3D reconstruction is presented in figure 5.14.k.

The last example, in figure 5.15, corresponds to a GPR image of an archeological site, located in Miami. Again, our representation model separated the horizontal noise bands from the buried structures reflection profiles, which in this case correspond to walls.

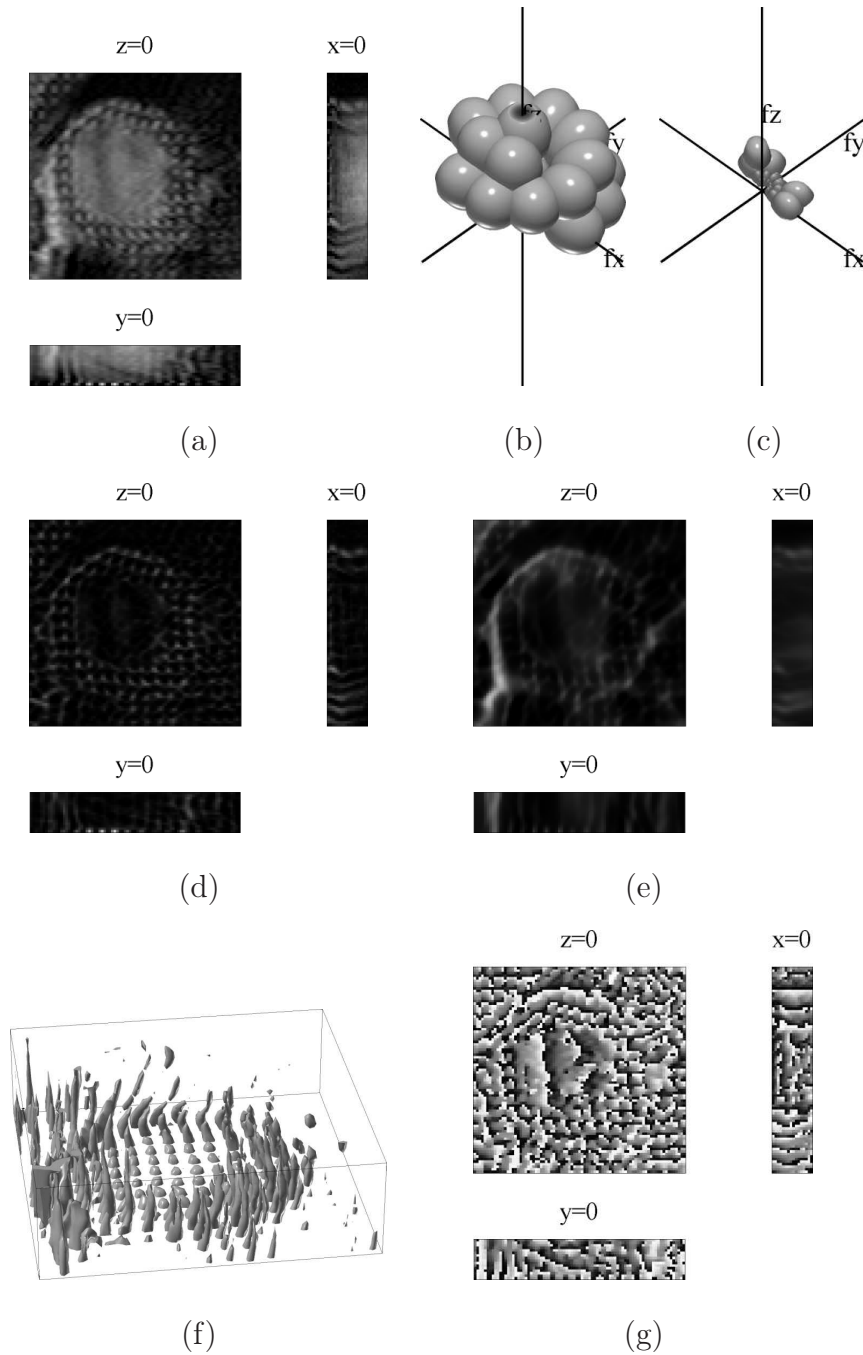


Figure 5.11: (a) Example of 2D+t TMRI image of the left ventricle of the heart. (b, c) Clusters generated by the composite feature extraction method. (d, e) Reconstructed visual patterns. (f) Isosurface of the first visual pattern P at level $0.2 \cdot \max(P)$. (g) Phase of the tag pattern. White represents π radians and black $-\pi$

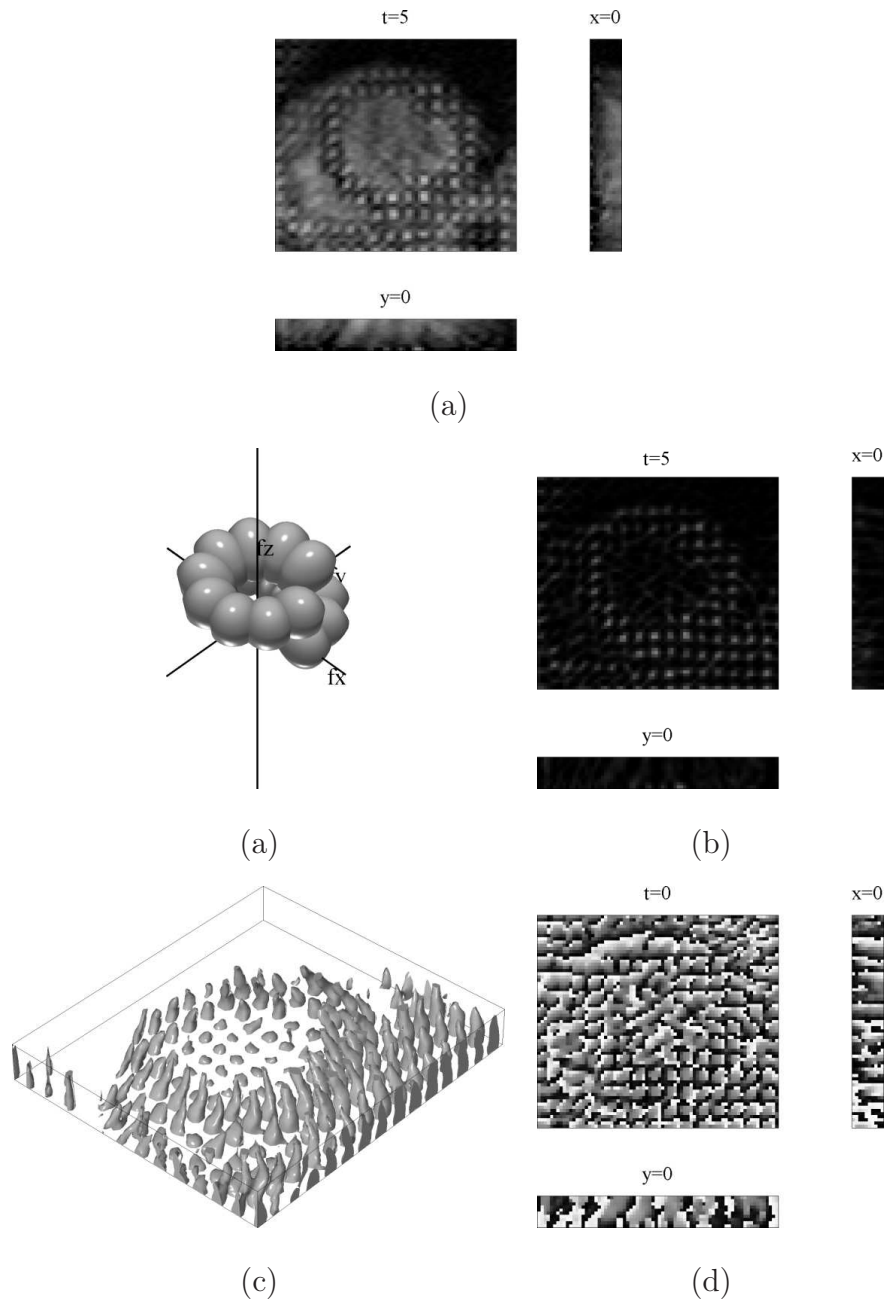


Figure 5.12: (a) Example of 2D+t TMRI image of the left ventricle of the heart
 (b) Cluster generated correspondent to the tag pattern. (c) Reconstructed tag pattern. (d) Isosurface of the first tag pattern P at level $0.15 \cdot \max(P)$ (d) Phase of the tag pattern. White represents π radians and black $-\pi$

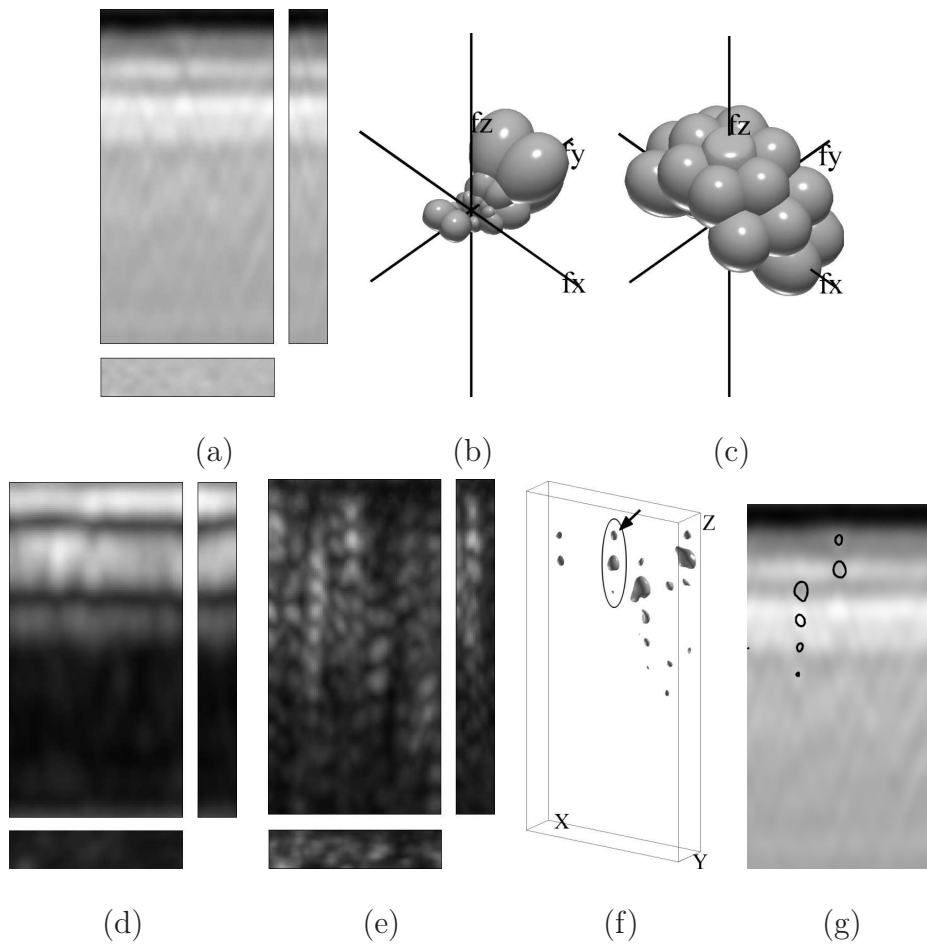


Figure 5.13: a) GPR image of a soil with a buried LI-11 AP mine. (b, c) Composite feature detectors extracted. (d, e) Their respective detected visual patterns. (f) Isosurface 3D reconstruction of visual pattern in figure e. The ellipse is enclosing the reflection echoes of the AP mine. The arrow signals the apex of the reflection profile, where land mine is located. (g) Contour plot of a section of the thresholded visual pattern in figure e, superimposed to the original data.

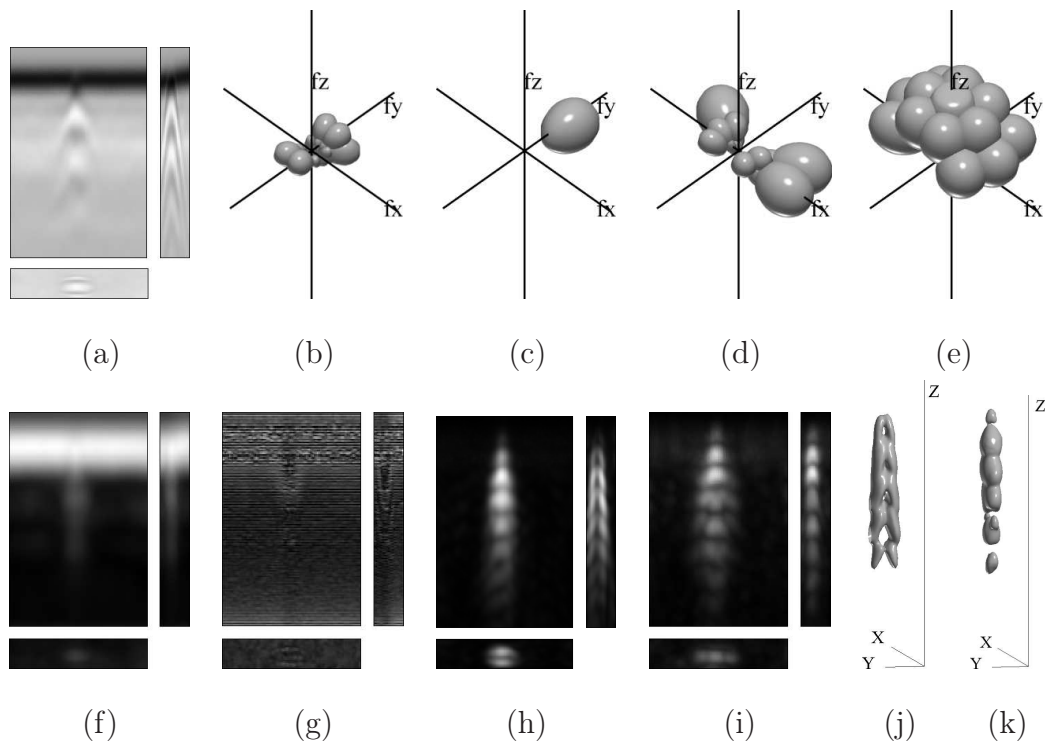


Figure 5.14: (a) GPR image of a soil with a buried PFM-1 AP mine. (b, c, d, e) Composite feature detectors extracted. (f, g, h, i) Their respective detected visual patterns. (j) Isosurface 3D reconstruction of visual pattern in figure h. (k) Isosurface 3D reconstruction of visual pattern in figure k.

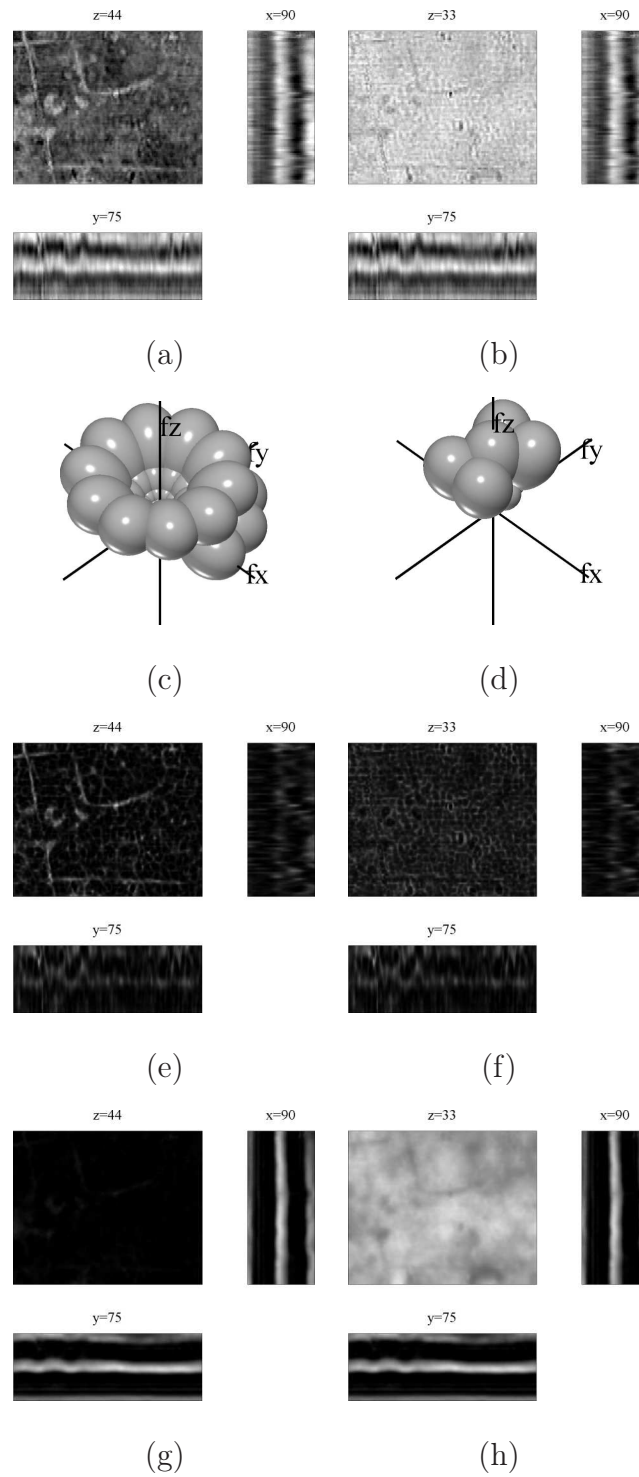


Figure 5.15: (a, b) Two views of a GPR image of an archaeological site, at different depths –origin of slice indexes is not at the center of the image here, but at origin of matrix indices. (c, d) Composite feature detectors extracted. (e, f) Two views of the first visual patterns. (g, h) Two views of the second visual pattern.

CHAPTER 6

Conclusions

In this work we have presented a low level representation model for volumetric data, based on composite low level features, which we have named visual patterns (Dosil et al., 2004, 2005a) – further work under revision in (Dosil et al., 2005d). The method for the unsupervised isolation of those patterns is founded in the notion of Phase Congruence, which allows to identify and integrate relevant composite features by multiresolution analysis. The method involves subband decomposition of the image using a 3D filter bank and integration of subband images associated to composite features by cluster analysis. Clustering is accomplished using a dissimilarity measure that reflects the degree of phase alignment between subband images. A study on the relationship between the energy profiles of subband version of the same image has been realized. The performance of several dissimilarity measures to reflect that relationship has been studied (Dosil et al., 2005b) – further work under revision in (Dosil et al., 2005c). The model has been employed to 3D data enhancement and in combination with higher level processes, namely, as the initialization stage of a geodesic active model segmentation technique. In the following paragraphs we list the main conclusions from this dissertation.

- Although 3D multiresolution decomposition is not novel, the filter bank proposed in this work takes advantage of interesting properties of several former solutions to provide a biologically inspired bank with good performance. Particularly, the 2D log Gabor filter, which fits the response profiles of cortical cells, has been combined with a rotationally symmetric orientation signature. The proposed sampling scheme produces even coverage of both the scale space

and the orientation space.

- For the unsupervised integration of subband features, here we have studied the use of a criterion of Phase Congruence. PC allows to detect composite features and to integrate its constituent simple features driven by the image data alone, without the use of a priori knowledge. The resulting composite feature detection method is of general purpose, i.e., is not tuned to predefined specific domain features. Unlike the criterion of statistical congruence, phase congruence is simpler but more robust. The examples presented indicate that phase congruence is able to discriminate among various types of low level composite features and that the results are stable under corruption by different noise levels.
- To estimate the degree of misalignment between maxima of two frequency features, we have presented a study on the performance of a set of statistical measures. All of the studied functions are global dissimilarity measures that reflect statistical dependence between the structure of the two subband energy maps. Global measures are less computationally expensive and less parameterized than those obtained by pooling of local measures. The results of the performance test have led to interesting conclusions on the statistical properties of the relations between subband versions of an image. We have shown that overconstrained measures produce, in general, better clustering results than underconstrained measures. The selection of the correlation coefficient is in agreement with statistical studies that reveal that the responses of complex cells –which are here modelled by the energy filters– are highly correlated in scale and orientation (Hoyer, 2002).
- The applicability of the proposed representation model has been demonstrated by its in combination with an active models for volumetric medical image segmentation. The composite features extracted by our model are used to define the initial state of a geodesic active model, so that the model starts from a state that is already a rough approximation in location and shape to the target object. The cooperation of both models leads to improved segmentations regarding those obtained without initialization, overcoming the typical problems caused by the local evolution of geodesic models. The elimination of non desired visual patterns minimizes their interference in the evolution of the model to the target object. The selected visual patterns are also close to the target object in shape, so that problems with cavities and sharp cusps

are avoided. Additionally, the initialization process reduces the computational cost the active model, since less iterations are necessary for convergence. Other applications, mainly 3D data enhancement, have also produced promising results.

Naturally, the proposed model has all the limitations of a low level representation and the introduction of high level processes and specific domain application knowledge should, of course, be profitable. We think that the characteristics of the model made it suitable as a sophisticated preprocessing tool for a wide variety of applications.

6.1 Final Remarks and Future Work

In the decomposition of visual patterns an important point is that a visual pattern may be relevant in some certain semantic context, but might also be decomposed into other visual patterns that, at the same time, are relevant in a different semantic context. Let us think of a GPR image of a soil containing a buried object. We consider that the pattern correspondent to the buried object is relevant for demining applications and that the remainder patterns correspond to the background. All the same, the background is the relevant pattern in seismic stratigraphic studies and the different patterns composing it are individually relevant. In that context –but not in the demining application– it would be interesting to be able to discriminate also between those patterns. What determines that background features are grouped or not, is the relative magnitude of their dissimilarities, in relation to their dissimilarities to other features. This is a typical issue concerning clustering analysis. This is often tackled using multiresolution clustering, where the a hierarchy of clusters is cut a different depths –or maximum dissimilarities– at different regions in the feature space. A research line for future work points to the direction of improving clustering by introducing these concepts.

In any case, the clustering process by itself can not take semantic decisions. The presented method means simply a low level model and the produced representations are just building blocks for higher level processed. From here, new integration stages are necessary to continue combining the information from successive layers of representation, thus emulating the hierarchical representation strategy of the HVS. An interesting way of doing so, following again the biological paradigm, is by

means of attention processes, guided both by statistical relevance and contextual information –memory and purpose. Future work will be mainly centered in the search and study of further integration stages.

Regarding initialization of active models, the method proposed here has more discrimination power than the former solution, but, on the other hand, the model matching initialization method has the advantage of providing the segmentation process with a priori information, bringing robustness. Hence, it seems a good idea to make a combination of both solutions that profits from their respective advantages. Instead of using visual patterns to directly initialize the active model, they could be employed to fit the prior shape model. The disjoint representation of visual patterns, separated from each other, offers the same advantages to model matching than it did to active models. The elimination of the visual patterns not belonging to the object should produce more robust matching results. One of the lines considered for future work is the introduction of the intermediate model matching step in the object reconstruction pipeline.

We are also weighing up the possibility of taking advantage of the visual pattern and a priori shape information in the definition of image potentials for the active models. One possibility we are considering is to use the prior shape model, not just as the initial state, like we did in previous works, but also as a constraint term in active model potentials, as we have already experimented in other works (Pardo et al., 2004). This last approach avoids excessive deviation of the model from a prototype model by penalizing large differences between the prototype and the current evolution state. In that work, the new image and the training set of images are supposed to be already registered among them, so no previous matching would be necessary. Combining the model matching initialization with the constrained active model, replacing the prototype model by the initial model, allows the two-fold goal of initialization and global shape constraint.

Either if visual patterns are used directly for initialization of active models or are used for model matching, the problem of the selection of the proper visual pattern among the results must be solved. The current solution needs of the user interaction. In some situations, manual visual pattern selection is necessary only once, like in the segmentation of large sets of images of the same domain, since the information about the frequency bands composing the target visual patterns in an image can be reused for other instance images of the same objects. Nevertheless, we are interested in achieving a fully automatic system. Another line for future work

points to the direction of automating visual pattern selection.

Notation and Abbreviations

1D	One-Dimensional
2D	Two-Dimensional
2D+ t	Two-Dimensions Plus Time
3D	Three-Dimensional
α	Angular distance between two directions
θ	Elevation component of frequency
λ	Wavelength
ρ	Radial component of frequency
ϕ	Azimuth component of frequency
Ω_X	Domain of definition of a variable X
AP	Anti-personnel
D	Image Dissimilarity
$\mathcal{D}$	Inter-Cluster Distance
DC	Offset component of the spectrum of a signal
DWT	Discrete Wavelet Transform
E	Local Energy
$\mathbb{E}$	Mathematical expectation
f	Histogram
F	Bijjective Application
$\mathcal{F}$	Fourier Transform
G	Gaussian function
DoG	Difference of Gaussians

GPR	Ground Penetrating Radar
h	Impulse response of a filter
H	Shannon's Entropy
$\mathcal{H}$	Hilbert Transform
HVS	Human Visual System
I	Mutual Information
ICA	Independent Component Analysis
ISA	Independent Subspace Analysis
K	Kullback's divergence
N^3	Size of a volumetric data set
N_θ	Number of sample elevations
N_ϕ	Number of sample azimuths
N_c	Number of clusters
N_f	Number of frequency features
N_o	Number of sample orientations
N_s	Number of sample frequencies or scales
NE	Normalized Entropy
NI	Normalized Mutual Information
MR	Magnetic Resonance
MST	Minimum Spanning Tree
P	Probability
PC	Phase Congruence
PCA	Principal Component Analysis
PSNR	Peak Signal to Noise Ratio
R	Radial separable component of a transfer function
RGFF	Representation model developed by Rodriguez-Sanchez et al. (1999)
RMSE	Root Mean Square Error
S	Angular separable component of a transfer function
SPECT	Single Photon Emission Computerized Tomography
T	Transfer function of a filter
TMRI	Tagged Magnetic Resonance Imaging
X, Y	Random variables

References

- Acharyya, M., De, R. and Kundu, M. “Extraction of Features Using M-Band Wavelet Packet Frame and their Neuro-Fuzzy Evaluation for Multitexture Segmentation.” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25, 12, 1639–1644, 2003.
- Adelson, E. and Bergen, J. “Spatiotemporal Energy Models for the Perception of Motion.” *Journal of the Optical Society of America A*, 2, 2, 284–299, 1985.
- Anderson, S. and Burr, D. “Spatial and Temporal Selectivity of the Human Motion Detection System.” *Vision Research*, 25, 1147–1154, 1985.
- Anderson, S. and Burr, D. “Receptive Field Sizes of Human Motion Detectors.” *Vision Research*, 27, 621–635, 1987.
- Andersson, M. *Controllable Multidimensional Filters in Low Level Computer Vision*. Ph.D., Linköping University, Sweden, Linköping, Sweden, 1992. Dissertation No 282, ISBN 91-7870-981-4.
- Ballard, D. and Brown, C. *Computer Vision*. Prentice Hall, Englewoods Clifs, New Jersey, 1982.
- Barlow, H. “The Coding of Sensory Messages.” In W. Thorpe and O. Zangwill (eds.), *Current Problems in Animal Behaviour*, pp. 331–360. Cambridge University Press, Cambridge, 1960.
- Basseville, M. “Information: Entropies, Divergences et Moyennes.” Tech. rep. 1020, IRISA, 35042 Rennes Cedex France, 1996.

- Bastiaans, M. "A Sampling Theorem for the Complex Spectrogram and Gabor's Expansion of a Signal in Gaussian Elementary Signals." *Opt. Engineer.*, 20, 594–598, 1981.
- Beck, J. "Textural Segmentation." In J. Beck (ed.), *Organization and Representation in Perception*. Erlbaum, Hillsdale, N.J., 1982.
- Beck, J., Prazdny, K. and Rosenfeld, A. "A Theory of Textural Segmentation." In J. Beck, B. Hope and A. Rosenfeld (eds.), *Human and Machine Vision*, pp. 1–38. Academic Press, New York, 1983.
- Bell, A. and Sejnowski, T. "The 'Independent Components' of Natural Scenes are Edge Filters." *Vision Research*, 37, 23, 3327–3338., 1997.
- Bergen, J. and Julesz, B. "Rapid Discrimination of Visual Patterns." *IEEE Trans. Syst. Man Cyber.*, 13, 857–863, 1983.
- Bovik, A., Clark, M. and Geisler, W. "Multichannel Texture Analysis Using Localized Spatial Filters." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12, 1, 55–73, 1990.
- Bruschini, C., Gros, B., Guerne, F., Pièce, P. and Carmona, O. "Ground Penetrating Radar and Induction Coil Sensor Imaging for Antipersonnel Mines Detection." *Journal of Applied Geophysics*, 40, 1–3, 59–71, 1998.
- Butz, T. and Thiran, J. "Affine Registration with Feature Space Mutual Information." In W. Niessen and M. Viergever (eds.), *LNCS 2208: Medical Image Computing and Computer-Assisted Intervention*, pp. 14–17. Springer, 2001.
- Campbell, F., Cooper, G. and Enroth-Cugell, C. "The Spatial Selectivity of Gratings." *J. Physiol. London*, 203, 223–235, 1969.
- Campbell, F. and Robson, J. "Application of Fourier Analysis to the Visibility of Gratings." *J. Physiol. London*, 197, 551–566, 1968.
- Canny, J. "A Computational Approach to Edge Detection." *IEEE. Trans. on Pattern Analysis and Machine Intelligence*, 8, 679–698, 1986.
- Caselles, V., Kimmel, R. and Sapiro, G. "Geodesic Active Contours." *Int. Journal on Computer Vision*, 22, 1, 61–79, 1997.

- Chamorro Martínez, J., Fernández Valdivia, J., García, J. and Martínez Baena, J. "A Frequency Domain Approach for the Extraction of Motion Patterns." *IEEE International Conference on Acoustics, Speech and Signal Processing*, 3, 165–168, 2003.
- Conyers, L. and Goodman, D. *Ground Penetrating Radar: An Introduction for Archaeologists*. Albania Press, London, New Delhi, 1997.
- Daubechies, I., Grossman, A. and Meyer, Y. "Painless Nonorthogonal Expansions." *Journal of Mathematical Physics*, 27, 1271–1283, 1986.
- Daugman, J. "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters." *J. Opt. Soc. Am. A*, 2, 7, 1160–1169, 1985.
- Daugman, J. "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression." *IEEE Trans. on Acoustics Speech and Signal Processing*, 36, 7, 1169–1179, 1988.
- Davies, D. and Boulding, D. "A Cluster Separation Measure." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1, 4, 224–227, 1979.
- DeValois, R., Albrecht, D. and Thorell, L. "Cortical Cells: Bar and Edge Detectors or Spatial Frequency Filters." In S. Cool and E. Smith III (eds.), *Frontiers of Visual Science*. Springer-Verlag, New York, 1978.
- Dosil, R., Fdez-Vidal, X. and Pardo, X. "Multiresolution Approach to Visual Pattern Partitioning of 3D Images." In A. Campilho and M. Kamel (eds.), *LNCS 3211: Image Analysis and Recognition*, pp. 655–663. Springer-Verlag, Berlin Heidelberg, 2004.
- Dosil, R., Fdez-Vidal, X. and Pardo, X. "Decomposition of 3D Medical Images Into Visual Patterns." *IEEE Trans. on Biomedical Engineering*, 2005a. To appear.
- Dosil, R., Fdez-Vidal, X. and Pardo, X. "Dissimilarity Measures for Visual Pattern Partitioning." In J. Marques, N. Pérez de la Blanca and P. Pina (eds.), *LNCS 3523: Pattern Recognition and Image Analysis*, pp. 287–294. Springer-Verlag, Berlin Heidelberg, 2005b.

- Dosil, R., Fdez-Vidal, X. and Pardo, X. "Distances Between Frequency Features for 3D Visual Pattern Partitioning." *Pattern Recognition Letters*, 2005c. Under revision.
- Dosil, R. and Pardo, X. "Generalized Ellipsoids and Anisotropic Filtering for Segmentation Improvement in 3D Medical Imaging." *Image and Vision Computing*, 21, 4, 325–343, 2003.
- Dosil, R., Pardo, X. and Fdez-Vidal, X. "Data-Driven Synthesis of Composite-Feature Detectors for 3D Image Analysis." *Image and Visión Computing*, 2005d. Under revision.
- Dosil, R., Pardo, X. and Mosquera, A. "Inicialización de Superficies Deformables Mediante Elipsoides Generalizados." In *XI Congreso Español de Informática Gráfica*, pp. 235–248. Girona, 2001a.
- Dosil, R., Pardo, X., Mosquera, A. and Cabello., D. "Curvature Dependent Diffusion for Feature Detection in 3D Medical Images." In M. Hamza (ed.), *Visualization, Imaging and Image Processing (VIIP2001)*, pp. 371–376. ACTA Press, Calgary, 2001b.
- Dosil, R., Pardo, X., Mosquera, A. and Vilarino., F. "Modelado 3D Aplicado a Segmentación." *X Congreso Español de Informática Gráfica*, pp. 393–394, 2000.
- du Buf, J. "Ramp Edges, Mach Bands and the Functional Significance of the Simple Cell Assembly." *Biological Cybernetics*, 70, 5, 449–461, 1994.
- Faas, F. and van Vliet, L. "3D-Orientation Space; Filters and Sampling." In *Scandinavian Conference on Image Analysis*, pp. 36–42, 2003.
- Field, D. "Relations Between the Statistics of Natural Images and the Response Properties of Cortical Cells." *Journal of the Optical Society of America A*, 4, 12, 2379–2394, 1987.
- Field, D. "Scale-Invariance and Self-Similar 'wavelet' Transforms: An Analysis of Natural Scenes and Mammalian Visual Systems." In M. Farge, J. Hunt and J. Vassilicos (eds.), *Wavelets, Fractals and Fourier Transforms*, pp. 151–193. Clarendon Press, Oxford, 1993.
- Field, D. "What is the Goal of Sensory Coding." *Neural Computation*, 6, 559–601, 1994.

- Fleet, D. *Measurement of Image Velocity*. Kluwer Academic Publishers, Massachusetts, 1992.
- Florack, L., ter Haar Romeny, B., Koenderink, J. and Viergever, M. “Linear Scale-Space.” *Journal of Mathematical Imaging and Vision*, 4, 4, 325–351, 1994.
- Gabor, D. “Theory of Communication.” *J. Inst. Electr. Eng.*, 93, 429–457, 1946.
- Graham, N. *Visual Pattern Analyzers*. Oxford University Press, 1989.
- Granlund, G. and Knutsson, H. *Signal Processing for Computer Vision*. Kluwer Academic Publishers, Boston, 1995.
- Haglund, L. and Fleet, D. “Stable Estimation of Image Orientation.” In *Proceedings of the IEEE International Conference on Image Processing*, pp. 68–72, 1994.
- Heitger, F., Rosenthaler, L., von der Heydt, R., Peterhans, E. and Kubler, O. “Simulation of Neural Contour Mechanisms: From Simple to End-Stopped Cells.” *Vision Research*, 32, 5, 963–981, 1992.
- Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L. and Kübler, O. “Simulation of Neural Contour Mechanisms: Representing Anomalous Contours.” *Image and Vision Computing*, 16, 407–421, 1998.
- Hofmann, T., Puzicha, J. and Buhmann, J. “Unsupervised Texture Segmentation in a Deterministic Annealing Framework.” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20, 8, 803–818, 1998.
- Hoyer, P. *Probabilistic Models of Early Vision*. Ph.D., Computer Science Dep. Helsinki University of Technology, 2002.
- Hoyer, P. and Hyvärinen, A. “Sparse Coding of Natural Contours.” *Neurocomputing*, 44–46, 459–466, 2002.
- Hubel, D. *Eye, Brain and Vision*. Freeman, New York, 1988.
- Hubel, D. and Wiesel, T. “Receptive Fields, Binocular Interaction and Functional Architecture in the Cat’s Visual Cortex.” *J. Physiol. London*, 160, 106–154, 1962.
- Hubel, D. and Wiesel, T. “Receptive Fields and Functional Architecture of Monkey Striate Cortex.” *J. Physiol.*, 195, 215–244, 1968.

- Jain, A. and Dubes, R. *Algorithms for Clustering Data*. Prentice Hall, New Jersey, 1988.
- Jain, A., Ratha, N. and Lakshmanan, S. "Object Detection Using Gabor Filters." *Pattern Recognition*, 30, 2, 295–309, 1997.
- Johnson, S. "Hierarchical Clustering Schemes." *Psychometrika*, 32, 241–254, 1967.
- Julesz, B. "Textons, the Elements of Texture Perception and their Interations." *Nature*, 290, 91–97, 1981.
- Julesz, B. "Texton Gradients: The Texton Theory Revisited." *Biol. Cybern.*, 54, 245–251, 1986.
- Kass, M., Witkin, A. and Terzopoulos, D. "Snakes: Active Contour Models." *International Journal of Computer Vision*, 1, 4, 321–331, 1987.
- Kass, M., Witkin, A. and Terzopoulos, D. "Snakes: Active Contour Models." *Int. Journal of Computer Vision*, pp. 321–331, 1988.
- King, B. "Step-Wise Clustering Procedures." *Journal of the American Statistical Association*, 69, 86–101, 1967.
- Klein, S. and Levi, D. "Hyperacuity Thresholds of 1 Sec: Theoretical Predictions and Empirical Validation." *Journal of the Optical Society of America*, 2, 7, 1170–1190, 1985.
- Koenderink, J. "The Structure of Images." *Biological Cybernetics*, 50, 363–370, 1984.
- Kornprobst, P., Deriche, R. and Aubert, G. "Image Sequence Restoration: A PDE Based Coupled Method for Image Restoration and Motion Segmentation." Tech. rep. 3308, INRIA, Sophia Antipolis, 1997.
- Kovesi, P. *Invariant Measures of Image Features from Phase Information*. Ph.D., The University of Western Australia, 1996. [Http://www.cs.uwa.edu.au/pub/robvis/theses/PeterKovesi/](http://www.cs.uwa.edu.au/pub/robvis/theses/PeterKovesi/).
- Kubovy, M., Cohen, D. and Hollier, J. "Feature Integration That Routinely Occurs Without Focal Attention." *Psychonomic Bulletin & Review*, 6, 2, 183–203, 1999.

- Leborán, V., Dosil, R. and Pardo, X. "Smooth Surface Reconstruction from Points and Normals Using Implicit Surfaces." In *XIII Congreso Español de Informática Gráfica*, pp. 203–216. A Coruña, 2003.
- Leborán, V., Dosil, R. and Pardo, X. "Anisotropic Diffusion Applied to Surface Reconstruction of Implicit Surfaces." In *XIV Congreso Español de Informática Gráfica*, pp. 256–270. Sevilla, 2004.
- Leborán, V. and Pardo, X. "Implicit Surface Reconstruction and Uniform Sampling by Striped Marching Triangles." In H. Hager, M. Chover and D. Tost (eds.), *EUROGRAPHICS*, pp. 89–96. The Eurographics Association, 2003.
- Lindeberg, T. *Scale-Space Theory in Computer Vision*. Kluwer Academic, Dordrecht, The Netherlands, 1994.
- Malik, J., Belongie, S., Shi, J. and Leung, T. "Textons, Contours and Regions: Cue Integration in Image Segmentation." In *Int. Conf. Computer Vision (ICCV)*, Vol., pp. 918–926. Corfu, Greece, 1999.
- Malik, J. and Perona, P. "Preattentive Texture Discrimination with Early Vision Mechanisms." *Journal of the Optical Society of America A*, 7, 5, 923–932, 1990.
- Mallat, S. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 11, 7, 674–693, 1989.
- Manjunath, B. and Chellappa, R. "A Unified Approach to Boundary Perception: Deges, Textures, and Illusory Contours." *IEEE Trans. on Neural Networks*, 4, 1, 96–108, 1993.
- Marr, D. and Hildreth, E. "Theory of Edge Detection." *Proceedings of the Royal Society of London B*, 207, 187–217, 1980.
- Marčelja, S. "Mathematical Description of the Responses of Simple Cortical Cell." *J. Opt. Soc. Am.*, 70, 11, 1297–1300, 1980.
- McInerney, T. and Terzopoulos, D. "Deformable Models in Medical Image Analysis: A Survey." *Medical Image Analysis*, 1, 2, 91–108, 1996.
- Milligan, G. and Shilling, D. "Asymptotic and Finite-Sample Characteristics of Four External Criterion Measures." *Multivariate Behavioral Research*, 20, 97–109, 1985.

- Morrone, M. and Burr, D. "Feature Detection in Human Vision: A Phase-Dependent Energy Model." *Proceedings of the Royal Society of London B*, 235, 221–245, 1988.
- Morrone, M. and Owens, R. "Feature Detection from Local Energy." *Pattern Recognition Letters*, 6, 303–313, 1987.
- Nestares, O., Miravet, C., Santamaria, J. and Navarro, R. "Automatic Enhancement of Noisy Image Sequences Through Local Spatio-Temporal Spectrum Analysis." *Optical engineering*, 39, 1457–1469, 2000.
- Olshausen, B. and Field, D. "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?" *Vision Research*, 37, 3311–3325, 1997.
- Ontrup, J., Wersing, H. and Ritter, H. "A Computational Feature Binding Model of Human Texture Perception." *Cognitive Processing*, 5, 1, 32–44, 2004.
- Oppenheim, A. and Lim, J. "The Importance of Phase in Signals." *Proceedings of the IEEE*, 69, 529–541, 1981.
- Osman, M., Mc Veigh, E. and Prince, J. "Imaging Heart Motion Using Harmonic Phase MRI." *IEEE Trans on Medical Imaging*, 19, 3, 186–202, 2000.
- Owens, R., Venkatesh, S. and Ross, J. "Edge Detection is a Projection." *Pattern Recognition Letters*, 9, 233–244, 1989.
- Pal, N. and Biswas, J. "Cluster Validation Using Graph Theoretic Concepts." *Pattern Recognition*, 30, 6, 847–857, 1997.
- Papoulis, A. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 3rd ed., 1991.
- Paragios, N. and Deriche, R. "Active Regions for Texture Segmentation." Tech. rep. 3440, INRIA, 1998.
- Pardo, X., Leborán, V. and Dosil, R. "Integrating Prior Shape Models Into Level-Set Approaches." *Pattern Recognition Letters*, 25, 6, 631–639, 2004.
- Perona, P. and Malik, J. "Detecting and Localizing Edges Composed of Steps, Peaks and Roofs." In *Third Int. Conference on Computer Vision*, pp. 52–57, 1990.
- Pollen, D. and Ronner, S. "Phase Relationships Between Adjacent Simple Cells in the Cat." *Science*, 212, 1409–1411, 1981.

- Porat, M. and Zeevi, Y. "The Generalized Gabor Scheme of Image Representation in Biological and Machine Vision." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10, 4, 452–468, 1989.
- Quick, R. "A Vector Magnitude Model of Contrast Detection." *Kybernetik*, 16, 65–67, 1974.
- Reinagel, P. and Zador, A. "Natural Scene Statistics Ant the Centre of Gaze." *Network: Computation in Neural Systems*, 10, 1–10, 1999.
- Reisfeld, D. "The Constrained Phase Congruency Feature Detector: Simultaneous Localization, Classification and Scale Determination." *Pattern Recognition Letters*, 17, 11, 1161–1169., 1996.
- Roche, A., Malandain, G. and Ayache, N. "Unifying Maximum Likelihood Approaches in Medical Image Registration." *Int. J. of Imaging Systems and Technology*, 11, 71–80, 2000.
- Roche, A., Malandain, G., Pennec, X. and Ayache, N. "The Correlation Ratio as a New Similarity Measure for Multidimodal Image Registration." In *LNCS 1496: MICCAI'98*, pp. 1115–1124. Springer Verlag, 1998.
- Rodrigues, J. and du Buf, J. "Visual Cortex Frontend: Integrating Lines, Edges, Keypoints and Disparity." In A. Campilho and M. Kamel (eds.), *LNCS 3211: Image Analysis and Recognition (ICIAR 2004)*, Vol., pp. 664–671. Springer-Verlag, Berlin, 2004.
- Rodríguez, P., Dosil, R., Pardo, X. and Leborán, V. "Automatic Generation of Initial Surfaces for Implicit Snakes." In J. Villanueva (ed.), *Visualization, Imaging and Image Processing (VIIP2002)*, pp. 113–118. ACTA Press, Calgary, 2002.
- Rodríguez Sánchez, R., García, J., Fernández Valdivia, J. and Fernández Vidal, X. "The RGFF Representational Model: A System for the Automatically Learned Partition of "Visual Patterns" in Digital Images." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21, 10, 1044–1073, 1999.
- Rodríguez Sánchez, R., García Soria, J., Fernández Valdivia, J. and Fernández Vidal, X. *Representación de Imágenes Digitales Basada en Patrones Visuales. Modelos Computacionales y Aplicaciones*. RA-MA Editorial, Madrid, 2000.

- Ronse, C. "A Twofold Model of Edge and Feature Detection." Working Document 65, Philips Research Laboratory, 1990.
- Ross, J., Morrone, M. and Burr, D. "The Conditions under which Mach Bands are Visible." *Vision Research*, 29, 6, 699–715, 1989.
- Rubner, Y., Puzicha, J., Tomasi, C. and Buhmann, J. "Empirical Evaluation of Dissimilarity Measures for Color and Texture." *Computer Vision and Image Understanding*, 84, 25–43, 2001.
- Sarrut, D. and Miguet, S. "Similarity Measures for Image Registration." In *First European Workshop on Content-Based Multimedia Indexing*, pp. 263–270. Toulouse, France, 1999.
- Shi, J. and Malik, J. "Normalized Cuts and Image Segmentation." In *Normalized Cuts and Image Segmentation*, pp. 731–737, 1997.
- Simoncelli, E. "Vision and the Statistics of the Visual Environment." *Current Opinion in Neurobiology*, 13, 144–149, 2003.
- Simoncelli, E. "Statistical Modeling of Photographic Images." In A. Bovik (ed.), *Handbook of Image and Video processing, 2nd Edition*, chap. 7. Academic Press, 2005.
- Simoncelli, E. and Adelson, E. "Nonseparable Extensions of Quadrature Mirror Filters to Multiple Dimensions." *Proceedings of the IEEE*, 78, 4, 652–664, 1990.
- Simoncelli, E., Freeman, W., Adelson, E. and Heeger, D. "Shiftable Multi-Scale Transforms." *IEEE Trans. on Information Theory*, 38, 2, 587–607, 1992.
- Srivastava, A., Lee, A., Simoncelli, E. and Zhu, S. "On Advances in Statistical Modeling of Natural Scenes." *Journal of Mathematical Imaging and Vision*, 18, 17–33, 2003.
- Strehl, A. *Relationship-Based Clustering and Cluster Ensembles for High-Dimensional Data Mining*. Ph.D., The University of Texas, Austin, 2002.
- Studholme, C., Hill, D. and Hawkes, D. "An Overlap Invariant Entropy Measure of 3D Medical Image Alignment." *Pattern Recognition*, 32, 71–86, 1999.

- ter Haar Romeny, B. *Front-End Vision and Multiscale Image Analysis: Introduction to Scale-Space Theory*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 2003.
- Treisman, A. and Gelade, G. "A Feature-Integration Theory of Attention." *Cognitive Psychology*, 12, 97–136, 1980.
- Turiel, A. and Parga, N. "The Multifractal Wavelet Filter of Natural Images." *Physical Review Letters*, 85, 3325–3328, 2000.
- Venkatesh, S. and Owens, R. "On the Classification of Image Features." *Pattern Recognition Letters*, 11, 339–349, 1990.
- Vinje, W. and Gallant, J. "Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision." *Science*, 287, 5456, 1273–1276, 2000.
- Viola, P. "Alingment by Maximization of Mutual Information." Tech. rep. 1548, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1995.
- Watson, A. "The Cortex Transform: Rapid Computation of Simulated Neural Images." *Computer Vision, Graphics and Image Processing*, 39, 311–327, 1987a.
- Watson, A. "Efficiency of a Model Human Image Code." *J. Opt. Soc. Am. A*, 4, 12, 2401–2417, 1987b.
- Watson, A. "Visual Detection of Spatial Contrast Patterns: Evaluation of Five Simple Models." *Journal of the Optics Society of America*, 6, 1, 12–36, 2000.
- Webster, M. and DeValois, R. "Relationship Between Spatial-Frequency and Orientation Tuning of Striate-Cortex Cells." *Journal of the Optical Society of America A*, 2, 2, 322–341, 1985.
- Weickert, J. and Kühne, G. "Fast Methods for Implicit Active Contour Models." In S. Osher and P. Paragios (eds.), *Geometric Level Set Methods in Imaging, Vision and Graphics*, pp. 43–58. Springer, New York, 2003.
- Wersing, H. and Körner, E. "Learning Optimized Features for Hierarchical Models of Invariant Object Recognition." *Neural Computing*, 15, 7, 1559–1588, 2003.
- Witkin, A. "Scale-Space Filtering." In *Int. Joint Conference of Artificial Intelligence*, pp. 1019–1023. Karlsruhe, Germany, 1983.

-
- Wolfe, J. “Visual Search.” In H. Pashler (ed.), *Attention*, pp. 13–39. Psychology Press, Hove, East Sussex, 1998.
- Wolfe, J. “Visual Attention.” In K. De Valois (ed.), *Seeing, 2nd Edition*, pp. 335–386. Academic Press, San Diego, CA, 2000.
- Woods, J. and O’Neal, S. “Subband Coding of Images.” *IEEE Trans. on Acoustics, Speech and Signal Processing*, 34, 1278–1288, 1986.
- Yu, W., Sommer, G. and Daniilidis, K. “Three Dimensional Orientation Signatures with Conic Kernel.” *Image and Vision Computing*, 21, 5, 447–458, 2003.