Data-Driven Synthesis of Composite-Feature Detectors for 3D Image Analysis

Abstract. Most image analysis techniques are based upon low level descriptions of the data. It is important that the chosen representation is able to discriminate as much as possible among independent image features. In particular, this is of great importance in segmentation with deformable models, which must be guided to the target object boundary avoiding other image features. In this paper, we present a multiresolution method for the decomposition of a volumetric image into its most relevant *visual patterns*, which we define as features associated to local energy maxima of the image. The method involves the clustering of a set of predefined band-pass energy filters according to their ability to segregate the different features in the image. In this way, the method generates a set of composite-feature detectors tuned to the specific visual patterns present in the data. Clustering is accomplished by defining a distance metric between the frequency features that reflects the degree of alignment of their energy maxima. This distance is related to the mutual information of their responses' energy maps. As will be shown, the method is able to isolate the frequency components of independent visual patterns in 3D images. We have applied this composite-feature detection method to the initialization of active models. Among the visual patterns detected, those associated to the segmentation target are selected by user interaction to define the initial state of a geodesic active model. We will demonstrate that this initialization technique facilitates the evolution of the model to the proper boundary.

1 Objectives

Computer vision systems in general usually involve the use of some low level representation of image contents. In many occasions the appropriate selection and detection of features to be the building blocks of higher level tasks is critical for the correct performance of the whole process. In some applications, low level features by themselves can be helpful to assist the physician in

diagnosis and treatment. More often, low level patterns are the basis to other applications, like segmentation, multimodal image registration, visualization or anatomical morphometry. In particular, segmentation techniques handle multiple cues of information to delineate the various objects of interest appearing in an image. Among segmentation techniques, active model methods [1, 2] stand out due to their ability to integrate information from different features. 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 optimal 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.

In this paper we present a method for the low level representation of 3D images with the aim of facilitating higher level analysis processes. Our goal is to obtain a decomposition of an image into its most relevant low level features. We define relevant low level features, which we will call *visual patterns*, as those associated to the local energy maxima of the image [3]. Isolating visual patterns from each other allows focusing of high level analysis processes on a given target object, avoiding the interference with other features. In particular, we think that this kind of representation could enhance the performance of the various active model segmentation techniques. The combination of our low level representation and active models could be tackled either integrating low level information into the definition of image potentials or in an initialization stage. We will demonstrate that the visual patterns isolated by our method are useful to place the initial model near the target object. Nevertheless, other applications based on the extraction of low level features could take advantage of this technique. For example, there are cases in which precise boundary detection is not necessary and the energy of the detected visual patterns can be simply thresholded. As an illustration of other possible uses of our method, we will apply it to the detection of microcalcifications of small size in X-Ray mammograms.

The technique developed to arrive to these objectives 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. Clustering is accomplished by defining a distance metric between the frequency features that reflects the degree of alignment of their energy maxima. Such distance is related to the mutual information [4] of their responses' energy maps.

The next section introduces the theoretical fundamentals of our method and the decisions taken in our specific implementation are justified. The method for visual pattern extraction is detailed in section 3. The 3D filter bank definition is introduced in subsection 3.1 and the selection of its most relevant channels in 3.2. The measure of dissimilarity between filter responses, based on mutual information, is defined in subsection 3.3. Subsection 3.4 presents the clustering method employed to group visual patterns. The integration with a geodesic active model is explained in section 4. In section 5, we firstly illustrate the behavior of the method with synthetic 3D images. Then, we show some medical image segmentation results, obtained using our method together with a geodesic active model. In section 6, we present the conclusions and future work.

2 Introduction

It is difficult to define what the most relevant low level information in an image is. Some authors agree in that these features should be easily identifiable by the Human Visual System (HVS), as, for example, discontinuities on local properties, like intensity, texture or phase [5, 6]. Morrone and Owens [3] sustain that the HVS perceives features at points where the local energy is maximal or, what is the same, points of maximal phase congruency (PC), which measures the local degree of matching among the phase of Fourier components.

For this reason, the method presented here for the decomposition of 3D images into visual patterns is based on the multiresolution analysis of the images by means of energy based filtering. Energy filters were introduced by Morrone and Owens [3] and widely studied by several authors [7, 8, 9, 10]. They are scalable operators that measure local energy maxima as the sum of the squared responses of a pair of filters in quadrature, which verify Canny's criteria

of good detection, good localization and single response [11]. They also fulfill the requirement stated by Owens et al. of being a projection [7]. This kind of operators outperforms previous linear filters [11, 12] in many aspects: they do not mark edges in sine-wave signals, detect features that are a mixture of odd and even intensity profiles, do not suffer from multiple responses and are a projection.

To study the degree of alignment in phase along scales, many approaches use multiscale analysis [9, 13, 14, 15]. This is also consistent with the behavior observed in the HVS [16]. To integrate the information from different scales and orientations, most solutions combines it to produce one single feature map –or "primal sketch" [12] –of the image, like in the energy filtering of Morrone and Burr [9] or in the PC measure defined by Kovesi [14]. An alternative solution is to determine what the spectral bands contributing to each image feature are. If we can separate the bands correspondent to each feature, we will be able to reconstruct them individually. This is the kind of analysis performed by the RGFF model [15]. It defines *visual patterns* or *integral features* as patterns with alignment in a set of local statistics along wide frequency ranges. To isolate visual patterns the RGFF first decomposes the image into elementary features using a filter bank. Then, it groups similar bands together employing a measure of dissimilarity based on local statistical properties of the filter's response energy maps. These properties are measured only at points that are candidates to belong to a relevant image feature. These points are called *fixation points* or *attentional points* because their use implies a simulation of the process of attention [17] and are defined as local energy maxima [3].

In the RGFF, frequency channels are represented by 2D log Gabor filters. This function is considered well suited to represent the behavior of the mammalian visual cortical cells [16]. In the frequency domain, the channel is the product of two terms: a log Gabor [5] of the radial frequency component and a Gaussian of the angular frequency. The spread of the Gaussian determines orientation selectivity. To extend the log Gabor filter to 3D a second angular frequency coordinate must be taken into account. Recently, Chamorro et al. [18] redesigned the RGFF for its application to 3D data from video sequences to detect low level motion patterns. This extension –the same as in [19] – is accomplished by adding a third term dependent on the

elevation component of frequency that is also a Gaussian of the elevation component of frequency. The drawback of this approach is that the so-defined filters lack of rotational symmetry. The product of the two angular terms is a 2D Gaussian if a 2D Cartesian –flat– domain is considered, but not in the 2D spherical domain. On their part, Faas and van Vliet [20] define the orientation signature as a Gaussian on the angular distance with rotational symmetry, although radial part of their filter is not a log Gabor function.

To select the set of orientations of the bank, the usual solution is to sample the angular coordinates uniformly [19]. However, this sampling is uniform in the Cartesian domain, but not in a spherical domain, producing an overrepresentation of the region near $\theta = \pm \pi/2$. An alternative is to choose the directions of vectors pointing at the vertices of regular polyhedrons centered at the origin [21] or some other fixed distribution [18], but this solution lacks of flexibility in the selection of the number of filters. The solution of Faas and van Vliet [20] 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 on the unit sphere to obtain the orientations. The number of orientations is controlled by imposing a finer/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 *x*-*y* plane. This is important because many medical imaging modalities are axial and the spatial resolution in the axis direction is usually lower than the resolution in the slices. Therefore, a special attention must be paid to orientation in the *x*-*y* plane.

Based on the previous considerations, we have developed our approach for the partition of volumetric data into visual patterns, which consists of the decomposition of the image into a series of frequency features and the subsequent reconstruction of the main visual patterns of the image by cluster analysis of the frequency features. To represent frequency features we use a bank of 3D log Gabor filters with rotational symmetry. The bank is designed using a non-uniform sampling scheme of the orientation space, with constant arc-length between filters with same elevation and neighboring azimuth. To accomplish frequency feature clustering, a new measure of dissimilarity between the responses of the filters in the bank has been introduced,

based on the normalized mutual information of their energy maps. This measure reduces the computational burden of attentional measures. It is a global measure that reflects the degree of coincidence of similar structures in the two maps. Grouping of frequency features is accomplished from a dissimilarity matrix by hierarchical cluster analysis.

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 *active* filters in the 3D filter bank –i.e., channels with high information content, by analyzing the spectral energy map– and generation of their responses.
2. Measurement of dissimilarity between the energy maps of pairs of filter responses based on their normalized mutual information.
3. Hierarchical clustering of the features based on the dissimilarity matrix.

In the next subsection, the filter bank employed to represent the elementary features of an image is presented. The first step of the sequence of processes, the technique to select the most relevant filters in the bank, is described in subsection 3.2.

3.1 Bank of 3D multiscale multiorientation filters

Visual patterns are represented as a combination of the responses of a set of band-pass filters tuned in different scales and orientations. The filters' transfer function T is designed as the product of separable factors R and S in the radial and angular components respectively, such that $T = R \cdot S$. The radial term R is given by the log Gabor function

$$R(\rho;\rho_i) = \exp\left(-\frac{\left(\log\left(\rho/\rho_i\right)\right)^2}{2\left(\log\left(\sigma_\rho/\rho_i\right)\right)^2}\right), \tag{1}$$

where σ_ρ is the standard deviation and ρ_i the central radial frequency.

To achieve orientation selectivity, the angular component is usually defined as a scattering function centered at the filter's direction. A common approach is to define *S* as the product of two Gaussians in each of the spherical angular coordinates, azimuth ϕ and elevation θ [18, 19]

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

where (ϕ_i, θ_i) is the filter orientation and σ_ϕ and σ_θ are the standard deviations. Considering σ_ϕ = σ_θ, this expression represents an isotropic attenuation regards the central direction, as long as ϕ and θ are treated as Cartesian coordinates. However, they actually are spherical coordinates, so that *S* is projected over a curved plane and the Gaussian is no longer isotropic.

A better choice is to define *S* as a Gaussian on the angular distance α between the directions of the filter and the position vector of a given point **f** in the spectral domain [20]

$$\alpha(\phi_i,\theta_i)=\operatorname{acos}\left(\mathbf{f}\cdot\mathbf{v}/\|\mathbf{f}\|\right), \tag{3}$$

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

$$S(\phi,\theta;\phi_i,\theta_i)=S(\alpha)=\exp\left(-\frac{\alpha^2}{2\sigma_\alpha^2}\right). \tag{4}$$

The shape of *S* from equations (2) and (4) is depicted in Fig. 1 and Fig. 2. Fig. 1 shows that expression (2) has rotation symmetry when represented in the plane, but not in the sphere, while expression (4) yields the same response for points at the same angular distance from the filters direction.

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. In our configuration, elevation is sampled uniformly 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. The filter bank

configurations resulting from uniform and non-uniform sampling of the angular frequency space are shown in Fig. 3.

Following this criterion, the filter bank has been designed using four elevations –only one hemisphere is needed due to symmetry– and six azimuths to sample half the $\theta_i = 0$ plane, yielding 23 orientations. The angular bandwidth is 25º, but it is changed to 20º when more orientation selectivity is needed. In the radial axis, four values have been taken, with wavelengths 4, 8, 16 and 32, and 2 octaves bandwidth. With four frequencies and 23 orientations, in the most general case the bank is composed of 92 filters. These settings give place to a bank with wide spectral coverage and high redundancy.

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 these cases, filter responses approximately represent the average intensity level. Therefore, these bands are discarded and only the highest frequencies are considered. In the case of images with different sizes in each image axis direction, the projections of the wavelength in each direction are studied for each filter in a band.

The response of a filter is calculated directly in the spectral domain as the scalar product of its transfer function and the Fourier transform of the image. The spatial domain representation of the response is computed as the inverse Fourier transform of that product. This way of filtering is very fast when using Fast Fourier Transform and Inverse Fast Fourier Transform algorithms.

3.2 Selection of active bands

For all filters in the bank, their responses to the input data must be obtained and their pair-wise dissimilarities calculated. 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 transformed spectral energy density $E = \log(|F| + 1)$ is used as a measure of the information associated to a channel, where F is the Fourier transform of the image.

The energy density map E reflects the preferred orientations and frequency bands plus a background of spectral noise. A 2D example can be seen in Fig. 4.b. In [15], the classification of the bands into active and non-active is realized 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 [16]. 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. We assume that the main contribution to the spectral energy at the highest frequencies is due to noise and that the amplitude spectrum of noise is typically flat [14]. Hence, the mean energy of the noise, m, and its dispersion, σ, are estimated as the average and the standard deviation of the energy in a band of high frequencies, whose cut-off frequency is double the maximum of the bank's center frequencies. Such band comprises all orientations. The maximum expected value of the noise is estimated as $m + 3\sigma$–three standard deviations is a pessimistic estimation to eliminate most of the noise energy without eliminating clumps of energy caused by visual patterns.

Still, subtracting the maximum noise level from the energy map is not sufficient to "clean" the background. Some spurious values remain, as shown in Fig. 4.c. To eliminate them, some kind of filtering can be applied that takes into account the characteristics of spectral energy maps. Relevant features in the spatial domain are translated to the spatial-frequency domain into alignment through scales [16]. Therefore, this kind of structures must be preserved in the filtering process.

Standard median filters produce the elimination of thin linear 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 of the radial median filter mask M of size $N \times N \times N$ is as follows

$$\mathrm{M}(q,p) = \begin{cases} 1 & \text{if } [(q-p)/\|q-p\|] = [p/\|p\|] \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

where p and q are points in the image and mask domains respectively, $[\cdot]$ represents rounding to the nearest integer and the origin is in the image center. In this work the mask size is taken to $N = 3$. The behavior of this filtering is illustrated in Fig. 4.d.

3.3 Dissimilarity measure between energy maps

In the RGFF, dissimilarity estimation involves 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.

$$D_\beta(X,Y) = \frac{1}{Card(\Omega_X)} \left(\sum_{p \in \Omega_X} \left| \mu_p(X,Y) \right|^\beta \right)^{1/\beta}, \tag{6}$$

where X and Y are two frequency features, Ω_X is the set of fixation points in X and μ_p is the weighted sum of differences d between statistics at fixation points

$$\mu_p(X,Y) = \sum_{k=1}^{Q} \frac{1}{\omega_k} d\left(T_k^p(X), T_k^p(Y) \right) \tag{7}$$

where T_k^p is the k^{th} elements of the vector T^p of local statistics measured at fixation point p and ω_k is the maximum value of T_k^p over all fixation points in all energy maps. The local statistics they employ are local phase, normalized local energy and local statistics of the normalized local energy, like entropy, contrast and standard deviation. Difference d is computed as a regular subtraction for all local features except from local phase, if employed, which is additionally mapped to the interval $(-\pi,\pi]$.

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

$$D_{RGFF}(X,Y) = D_{\beta}^{2}(X,Y) + D_{\beta}^{2}(Y,X) \tag{8}$$

This distance is computationally very expensive, as will be seen in the next subsection. Furthermore, it is highly parameterized and highly dependent on the performance of low level processes like non-maxima suppression and scale estimation.

To reduce the computational burden of the dissimilarity estimation, in this work the measure of dissimilarity between pairs of energy maps has been defined as a function of their mutual information MI [22]. MI captures the amount of concurrence of values in the two signals, giving an idea of the amount of shared information.

$$\mathrm{MI}(X,Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y),$$

where H stands for entropy and X and Y are the energy maps of the responses of two of the filters in the bank.

Since it depends on the individual information amounts of each energy map, it must be normalized to represent the shared information in relation to the total information quantity. Studholme proposed a normalized measure of statistical dependence for image distance estimation in the field of medical image registration [4]. Although Studholme called this measure the Normalized Mutual Information, is 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.

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

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 (NMI), is an alternative form of normalization of mutual information.

$$\mathrm{NMI}(X,Y) = 2 \cdot \frac{\mathrm{MI}(X,Y)}{\mathrm{H}(X)+\mathrm{H}(Y)} = 2 \cdot \frac{\mathrm{H}(X)+\mathrm{H}(Y)-\mathrm{H}(X,Y)}{\mathrm{H}(X)+\mathrm{H}(Y)} = 2\left(1-\frac{1}{\mathrm{NE}(X,Y)}\right), \tag{9}$$

NMI values range from zero that means no information in common, to one, representing images with the same information contents. To transform NMI into a measure of dissimilarity instead of similarity, its range must be inverted. It is also convenient to apply transformations that equalize its range. The distance measure used here is

$$D_{NMI}(X,Y) = \left(1-\sqrt{\mathrm{NMI}(X,Y)}\right)^2. \tag{10}$$

This transformation enlarges high differences and shortens low ones. This improves the performance of clustering, because generally D tends to be very small since most of the energy values correspond to the background.

An important advantage of NMI is that it is invariant to contrast changes. This is convenient because high frequency components of visual patterns have higher contrast responses than low frequency ones [16], but they should be grouped together if they have the same location, i.e., there is phase congruency. Other advantages of the NMI are its simplicity compared with the measure used in the RGFF, where decisions are to be taken about the values of a great amount of parameters, and its lesser time consumption.

3.4 Computational cost of dissimilarity estimation

In this subsection an analysis of the asymptotic computational cost of the two previous dissimilarity estimation approaches is presented. Let us suppose that the input data are a volume of dimensions $N \times N \times N$, that our filter bank consists of F filters and that the number of bins used for histogram calculations is M. The calculus of NMI 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 dissimilarity calculation is $O(N^3)$. This must be done for each of the $F(F-1)$ pairs of filters, resulting in a computational cost of $O(F^2 \cdot N^3)$.

In the case of the RGFF distance, the cost of the dissimilarity calculations is $O(F^2 \cdot N^6)$. This is due to the calculus of the neighborhood of each attention point and the local statistics on it. The neighborhoods are related to the scales of each maximum and are defined as the distance from each energy maximum to the nearest minimum. In high scale filters, the neighborhood radius is in the order of the image size. Hence, this calculations are $O(N^3)$ and must be done for each attention point, i.e., $O(N^3)$ times, and for each filter pair, i.e., $O(F^2)$ times. Even if the points of each neighborhood where stored, what would have a memory cost of $O(F \cdot N^6)$, the calculus of the local statistics differences maintains a total cost $O(F^2 \cdot N^6)$. Therefore, the cost of the calculus of D_{RGFF} strongly increases when applied to 3D images.

3. 5 Feature clustering

To group filter responses a hierarchical clustering method has been chosen. Other clustering techniques, like k-means, are not adequate due to the nature of our data. Divisive clustering methods deal with the input data themselves, treating them as coordinate vectors in a multidimensional space. In this case, the input data are the energy maps, giving place to a space of as much as N^3 coordinates. The centroid of a cluster calculated as the average is the one that minimizes the distances to the elements of the cluster as long as one accepts the usual Euclidean metric as a proper dissimilarity measure, which is not the case. Hence one should have to estimate the centroid as the feature that minimizes, for example, the overall NMI-based distance to all cluster elements. Agglomerative methods are better suited, since they work directly with the dissimilarity matrix. They do not consider cluster elements as vectors in a coordinate space, with their compactness associated to geometrical proximity, but they treat them as graph nodes, with their distances being edge weights.

Hierarchical clustering [23] has been applied using a complete-link algorithm. It is an iterative algorithm that starts with a completely disjoint partition of the data, i.e., each element is one cluster. At each iteration, distances between clusters are calculated and the two nearest clusters are merged to form the next partition of the hierarchy. To define distances between clusters from distances between elements, different approaches have been proposed [23]. Here,

inter-cluster distance is defined as the maximum of all pair-wise distances between features in the two clusters, thus producing compact clusters.

Once the hierarchy of partitions is generated, the resulting clustering must be selected from it by taking the appropriate number of clusters N_c. When this parameter is not known a priori, the usual strategy to determine the N_c is to run the algorithm for each possible N_c and evaluate the quality of each resulting configuration according to a given validity index. The selected N_c is the one that produces the cluster partitioning with the highest validity index. In this work, a modification of the MTS Davies-Bouldin index [24] has proved to produce good results. It is a graph-theory based index that measures the compactness of the clusters in relation to their separation.

$$V_{DB}^{MTS}(N_c) = \frac{1}{N_c}\sum_{r=1}^{N_c} \max_{s,s\neq r}\left\{\frac{\sigma_r^{MTS}+\sigma_s^{MTS}}{S_{rs}}\right\}, \quad \text{with} \quad r,s = \{1,...,N_c\}. \tag{11}$$

where σ_r^{MTS} is the maximum edge length of the Minimum Spanning Tree of cluster r. In the original measure from [24], S_{rs} is taken as the distance between centroids of the clusters r and s. Here, to avoid estimation of centroids, S_{rs} is the average of the dissimilarities between elements of two different clusters r and s.

4 Active Model Initialization

A geodesic active model represents objects using an implicit function defined in the whole image domain, where the surface of the objects corresponds to its zero level set. We will not get here into the description of the geodesic deformable model. 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, of course, more precise segmentations could be achieved by combining the initialization process with active models which exploit the knowledge about the application domain. The concrete implementation used here is the one described in [25], applying the same stopping criterion based on a decreasing function of the gradient modulus of the original image. Neither balloon forces nor *a priori* knowledge on the

domain are incorporated. Usually, the model is initialized defining a window function enclosing all objects or by user interaction.

In the present application, the initial model has been defined choosing among the visual patterns generated by our partitioning method the one that best represents the target object. User interaction is necessary for the selection of this visual pattern. The initial state of the geodesic active model is defined from the selected cluster by simply rescaling the intensity levels of its response. The response is taken as the linear summation of the real part of the responses of its filters, instead of the squared sum of the real and imaginary parts, which would produce a less precise initialization.

In Fig. 5 the initialization and segmentation processes are illustrated using an example case in 2D to simplify visualization. It is a 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 calculated from the original image as a function of its gradient modulus. Therefore, the model stops evolving when it arrives to a strong local maximum of this energy, so that the segmentation without initialization produces a large amount of contours in the background texture. The 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.

5 Results

In this section, some results are presented to illustrate the ability of the system to identify and extract visual patterns of different types, its robustness to noise and its usefulness in medical image applications. The first block of results has been obtained using a set of synthetic 3D data designed containing different kinds of visual patterns. The correctness of the results is

determined by comparing them with the design specifications. The second block comprises several examples of analysis of 2D and 3D medical images aided by the presented visual pattern decomposition method, including segmentation with geodesic deformable models and microcalcification detection in mammograms. 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 2.1. Images D1, D2 and D3 have been processed using angular bandwidth of 20º while the remainder 3D images use 25º. The 2D filter bank employed to process the mammograms uses 12 orientations with angular bandwidth of 12.5º. It must be remarked that, in the case of synthetic images, the radial median filter 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 the presented 3D bank and filters from the bank proposed by Yu [19], this is, a bank with uniform angular frequency sampling and 3D log Gabor filters using equation (2). Both approaches are applied to the data set D1 (Fig. 6), which is a synthetic volume of 64×64×64 size that contains two patterns of parallel planes 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. Fig. 7 shows two filters from Yu's bank with elevation component $\theta_i = \pi/2$ and their responses to D1. It can be seen that filter ($\lambda_i = 4$, $\phi_i = 0$, $\theta_i = \pi/2$) –where λ is the wavelength– represents the direction normal to the z plane only partially. The lack of rotational symmetry and the overrepresentation 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 Fig. 7, sweeping the azimuth coordinate, would be necessary to fully 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 Fig. 8 that both filters produce the same response, as 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, the data set D2 (Fig. 9) is a synthetic volume of 64×64×64 size, which consists of three regions with different texture properties. In the most inner region, the main frequency component has orientation ($\phi = 3\pi/4$, $\theta = 0$) while in its surrounding region the main components are (0, 0), ($\pi/2$, 0) and (0, $\pi/2$). The texture of the outer region is mainly originated by noise. The system groups the filters into three clusters, each correspondent to one of the three regions. In Fig. 10 the resultant 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 cluster comprise orientations similar to those composing the two inner regions in D2, 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 containing three types of visual patterns: texture, phase discontinuity and intensity level discontinuity, which in this case is a plane, this is, a symmetric edge. This image is also 64×64×64 size. Subsequently, new images have been created from D3 by adding Gaussian noise of different standard deviations. Fig. 11 shows the volume D3 and an example of noise corrupted data, named D4, obtained from D3 by adding Gaussian noise of standard deviation $\sigma = 25$. The results obtained have been the very similar for D3 and all the noisy versions of it. Fig. 12 shows 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 must 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.

In the field of medical imaging, the ability in grouping bands with different orientations that contribute to a single object's shape has special interest. In the fourth example, the data set D5 (Fig. 13) contains a region with a characteristic shape immersed in a textured medium. The image, of size 64×64×64, has been corrupted with Gaussian noise of standard deviation $\sigma = 10$. The results obtained are shown in Fig. 14. 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.

The first segmentation example corresponds to real 3D medical imaging data. Image D6 is a volumetric data set of size 32×64×20 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 to the acquisition process itself. The goal is to isolate the ventricle wall from the background. The image and the results obtained are presented in Fig. 15. The method detects two visual patterns, one gathering the background artifacts and the other one mainly representing the ventricle wall, which is the one selected to initialize the active model. The segmentation result, in Fig. 16, shows that both the outer and inner side of the ventricle wall have been correctly captured, since the initial surface was already very close to them.

The second segmentation example is the 3D data set D7, of size 71×107×35, correspondent to a phantom T2-weighted MR image of the lateral ventricles of the brain. Fig. 17 shows the result of visual pattern decomposition and Fig. 18 presents the segmentation results. Although it is quite difficult to analyze the results in such complex 3D data, it can be seen that the largest response values for the second visual pattern, in Fig. 17(f), occur in the region occupied by the lateral ventricles. The first visual pattern, in Fig. 17(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 Fig. 17(b) as the arched region on top of the ventricles. If

each of these two visual patterns is used to initialize one geodesic active model, both structures can be independently segmented, as can be seen in Fig. 18.

The third segmentation example is the 3D data set D8, of size 87×42×44, correspondent to a real CT image of the brain. In this imaging modality, the brain ventricles produce a very weak signal. 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 in the original, as can be observed in Fig. 19(c). The second visual pattern, in Fig. 19(d), captures the corpus callosum contour and other brain sulci. The segmentation obtained using the first cluster as initial geodesic model captures the surface of the ventricles – see Fig. 19(a)(b) –, while the segmentation obtained using a square box, slightly smaller than the image volume, as initialization can not overcome the other structures present in the image – see Fig. 20.

The last three cases, in Fig. 21, are examples of microcalcification detection in 2D X-Ray mammograms. The mammograms D9, D10 and D11 have size 570×540, 170×284 and 563×468 respectively. Again, in these images the target patterns have intensity and contrast similar to other structures present in the data. What makes the difference among the calcifications and the remainder tissues is the frequency content of each to them. In each example case, the method produces one visual pattern clearly representing the microcalcifications. In these cases, segmentation is easily achieved by simple thresholding the cluster energy. The energy thresholds have been obtained ad hoc for each example.

6 Conclusions

In this paper, we have presented a method for the isolation of visual patterns from 3D images, consisting of the decomposition of the image into a set of frequency features by means of a 3D filter bank and the grouping of similar features together by cluster analysis. Although other authors have already developed similar methods, in this work the extension to 3D has been accomplished introducing a different dissimilarity measure in the cluster analysis stage that reduces the computational burden of the process. The measure is based on the normalized

mutual information of the energy of the frequency features. Other advantages of this measure are that it is less parameterized and less complex.

The examples presented indicate that the developed system is capable of discriminating among various types of low level features and that the results are stable under corruption by different noise levels. The results obtained from the medical imaging field show that the method is suitable for its use in combination with segmentation techniques. Specifically, we deal with the combination of this technique with an active surface model in order to separate different objects in volumetric medical images. The initialization of the model using the extracted visual patterns leads to better results than those obtained without initialization, provided that image features belonging to non target objects are removed. The starting configuration of the geodesic model is kept away for non-desired potential minima and maxima in the external energy, which otherwise would stop the active model evolution towards the desired object's boundary.

This approach needs of user interaction to determine which of the resulting visual pattern corresponds to the target objects. Therefore, the method, as it is, is rather more useful for 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. Present and future work points to the direction of obtaining a fully automated initialization method.

It also has to be said that, although the method is quite successful in dealing with simple structures, it has problems with images containing complex visual patterns. One of the main weaknesses of the method is that the congruency of frequency features with different orientations belonging to an object boundary is small and it is mostly due to the redundancy of the bank representation. Another drawback is its incapability to distinguish between features with the same frequency composition, like, for example, the closed contours of two different objects. The method, as is defined, is helpful for some kinds of medical images, but not as a general purpose technique. We think that the way of improving the method is the introduction of *a priori* information. Future work will also point to this direction.

References

1. M. Kass, A. Witkin, D. Terzopoulos. Snakes: Active Contour Models. Int. Journal of Computer Vision, Vol. 55(4), pp. 321-331, January 1988.
2. V. Caselles, R. Kimmel, G. Sapiro. Geodesic Active Contours. Int. Journal on Computer Vision, Vol. 22(1), pp. 61-79, 1997.
3. M. C. Morrone, R. A. Owens. Feature Detection from Local Energy. Pattern Recognition Letters, Vol. 6(5), pp. 303-313, December 1987
4. C. Studholme, D. L. G. Hill, D. J. Hawkes. An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. Pattern Recognition, Vol. 32, pp. 71-86, 1999.
5. D. J. Field. Relations between the Statistics of Natural Images and the Response Properties of Cortical Cells. J. Opt. Soc. Am. A, Vol. 4(12), pp. 2379-2394, December 1987
6. B. S. Manjunath, R. Chellapa. A Unified Approach to Boundary Perception: Edges, Textures and Illusory Contours. IEEE Trans. on Neural Networks, Vol. 4(1), pp. 96-108, 1993.
7. R. Owens, S. Venkatesh, J. Ross. Edge Detection is a Projection. Pattern Recognition Letters, Vol. 9(4), pp. 233-244, May 1989
8. P. Perona, J. Malik. Detecting and Localizing Edges Composed of Steps, Peaks and Roofs. IEEE Third Int. Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 52-57, December 1990
9. M. C. Morrone, D. C. Burr. Feature Detection in Human Vision: a Phase-Dependent Energy Model. Proc. R. Soc. Lond. B, Vol. 235 , pp. 221-245, 1988
10. S. Venkatesh, R. Owens. On the Classification of Image Features. Pattern Recognition Letters, Vol. 11(5), pp. 339-349, May 1990
11. J. Canny. A Computational Approach to Edge Detection. IEEE. Trans. on Pattern Analysis and Machine Intelligence, Vol. 8(6), pp. 679-698, November, 1986
12. D. Marr, E. Hildreth. Theory of Edge Detection. Proc. R. Soc. Lond. B, Vol. 207, pp. 187-217, 1980

13. J. Malik, P. Perona. Preattentive Texture Discrimination with Early Vision Mechanisms. J. Opt. Soc. Am. A, Vol. 7(5), pp. 923-932, 1990
14. P. D. Kovesi. Invariant Measures of Image Features from Phase Information, PhD. Thesis, The University or Western Australia, May 1996 http://www.cs.uwa.edu.au/pub/robvis/theses/PeterKovesi/
15. R. Rodríguez-Sánchez, J. A. García, J. Fdez-Valdivia, X. R. Fdez-Vidal. The RGFF Representational Model: A System for the Automatically Learned Partition of "Visual Patterns" in Digital Images, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 21(10), pp. 1044-1073, October 1999
16. D. J. Field. Scale–Invariance and self-similar "wavelet" Transforms: An Analysis of Natural Scenes and Mammalian Visual Systems. In: Farge, M., Hunt, J.C.R., Vassilicos, J.C. (eds.): Wavelets, fractals and Fourier Transforms, Oxford, Clarendon Press, 1993, chapter 9, pp. 151-193
17. A. M. Treisman, G. Gelade. A Feature-Integration Theory of Attention. Cognitive Psychology, Vol. 12, pp. 97-136, 1980.
18. J. Chamorro-Martínez, J. Fdez-Valdivia, J. A. García, J. Martínez-Baena. A frequency Domain Approach for the Extraction of Motion Patterns, in IEEE International Conference on Acoustics, Speech and Signal Processing, Hong Kong, Vol. 3, pp. 165-168, April 2003
19. W. Yu, G. Sommer, K. Daniilidis. Three dimensional orientation signatures with conic kernel. Image and Vision Computing, Vol. 21(5) (2003) 447-458
20. F. G. A. Faas, L. J. van Vliet. 3D-Orientation Space; Filters and Sampling. in: J. Bigun, T. Gustavsson (eds.), LNCS: Scandinavian Conference on Image Analysis, vol. 2749, Springer Verlag, Berlin, 2003, pp. 36-42, 2003.
21. G. H. Granlund, H. Knutsson. Signal Processing for Computer Vision. Boston, Kluwer Academic Publishers, 1995
22. P. A. Viola. Alignment by Maximization of Mutual Information. Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Technical Report 1548, 1995
23. A. Jain, R. Dubes. Algorithms for Clustering Data. Prentice Hall, New Jersey, 1988

24. N. R. Pal, J. Biswas. Cluster Validation Using graph Theoretic Concepts. Pattern Recognition, Vol. 30(6), pp. 847-857, June 1997

25. J. Weickert, G. Kühne. Fast Methods for Implicit Active Contour Models. In: S. Osher, N. Paragios (eds.): Geometric Level Set Methods in Imaging, Vision and Graphics, Springer, New York, pp. 43-58, 2003.

Figure Captions

Fig 1. Representation of S in the (ϕ, θ) plane as isolevel curves. *Left*: S from equation (2). *Right*: S from equation (4). *Top*: $(\phi_i = \pi/3, \theta_i = \pi/4)$. *Bottom*: $(\phi_i = 0, \theta_i = \pi/2)$.

Fig 2. Representation of S over a unit radius sphere as isolevel curves. *Left*: S from equation (2). *Right*: S from equation (4). *Top*: $(\phi_i = \pi/3, \theta_i = \pi/4)$. *Bottom*: $(\phi_i = 0, \theta_i = \pi/2)$.

Fig 3. Filter bank representation by plotting the isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$. *Left*: Filter bank with uniform angular frequency sampling. *Right*: Non-uniform sampling.

Fig 4. 2D example of radial median filtering. (a) Input image in spatial domain of a nematode. (b) $E = \log(|F| + 1)$, where $|F|$ is the spectral energy. (c) Noise subtraction on E, 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.

Fig 5. 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.

Fig 6. 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 $(\phi = \pi, \theta = \pi/3)$

Fig 7. *Left column*: Filters from a bank with uniform angular frequency components sampling and log Gabor filters from equation (2), 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 $(\lambda_i = 4, \phi_i = \pi, \theta_i = \pi/2)$.

Fig 8. *Left*: Filters with $(\lambda_i = 4, \phi_i = 0, \theta_i = \pi/2)$ and $(\lambda_i = 4, \phi_i = \pi, \theta_i = \pi/2)$ from a bank with non-uniform angular components sampling and log Gabor filters from equation (4), represented

by isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$. *Right*: Their response to data set D1 –they have the same transfer function.

Fig 9. Volumetric data set D2 represented by orthogonal cross sections.

Fig 10. 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.

Fig 11. *Top*: Image D3 and *Bottom*: image D4, obtained from D3 by adding Gaussian noise of standard deviation 25. Both images are represented by the slices normal to the image coordinate axis

Fig 12. Results obtained for images D3 and D4. *Left column*: the filters correspondent to each of the three clusters obtained, represented by isosurfaces of the filter's transfer functions verifying $T_i = \exp(-1/2)$, are the same for the two images. *Right*: for each cluster, cross sections of the volumes reconstructed for image D3 (*top*) and D4 (*bottom*).

Fig 13. 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.

Fig 14. 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*: For each cluster, cross sections of the volumes reconstructed as the sum of energies of all filters in the cluster.

Fig 15. (a) Cross sections of 3D data set D6, 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 (b) and (c) respectively, reconstructed by linear summation of the half-wave rectified real part of the filters' responses.

Fig 16. Segmentation of data set D6 (a) using a square box as initial model, at three different evolution states and (b) using visual pattern from (d) (e)

Fig. 15.(e) as initial model. Segmentations are represented by cross sections of the resulting surface, in white color, superimposed to the original data (c) 3D reconstruction of the surface in (b).

Fig 17. (a) Sample horizontal slices of the original data set D7, a phantom T2-weighted MR image of the brain ventricles and the corpus callosum. (b) Sagittal section of data set D7 (c, d) The two clusters generated by the method. (e, f) Reconstruction of the visual patterns correspondent to clusters in (c) and (d) respectively –image (f) is inverted– by linear summation of the filter's responses real component of each cluster.

Fig 18. Segmentation of 3D data set D7 (a, b) Segmentations obtained using visual patterns in Fig. 16.(d) and (e) respectively as initialization of the geodesic model. The outer region of the segmented objects is represented in white and black respectively. (c, d) 3D reconstructions of the surfaces of the segmentations in (a) and (b) respectively.

Fig 19. Segmentation of 3D data set D8, from MR images of slices 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).

Fig 20. Sample slices of the result of segmentation of the data set D8 using a square box as initialization. The outer region of the object has been represented in white.

Fig 21. Microcalcification detection in CT mammograms by simple thresholding of the selected visual pattern energy. *Left column*: From top to bottom, original 2D data sets D9, D10 and D11. Center column: for each image in left column, energy of the correspondent selected visual pattern. *Right column*: Energy maps from center column after thresholding, with thresholds 0.6, 0.6 and 0.8 times the maximum energy values in each case respectively.

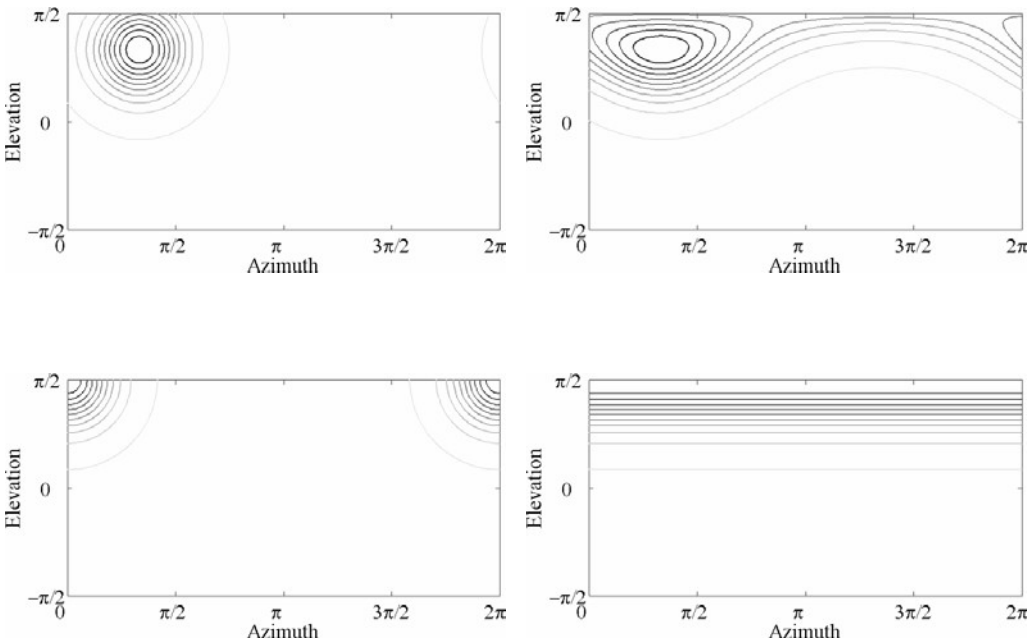


Fig. 1

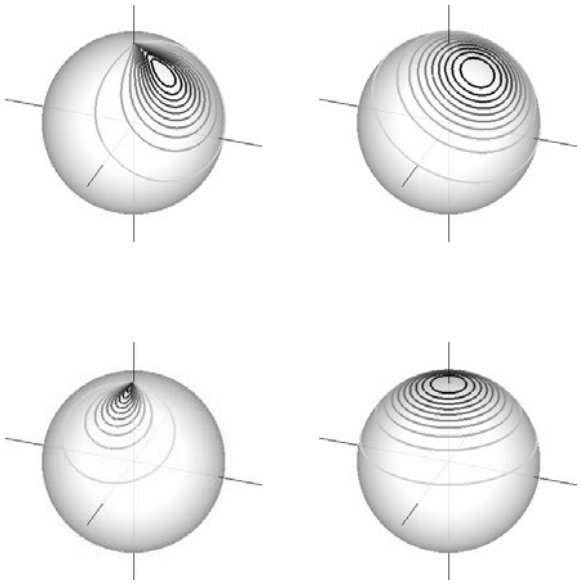


Fig. 2

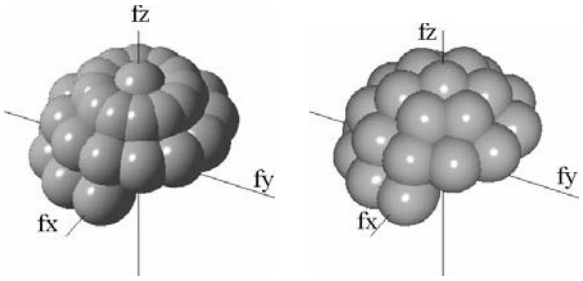
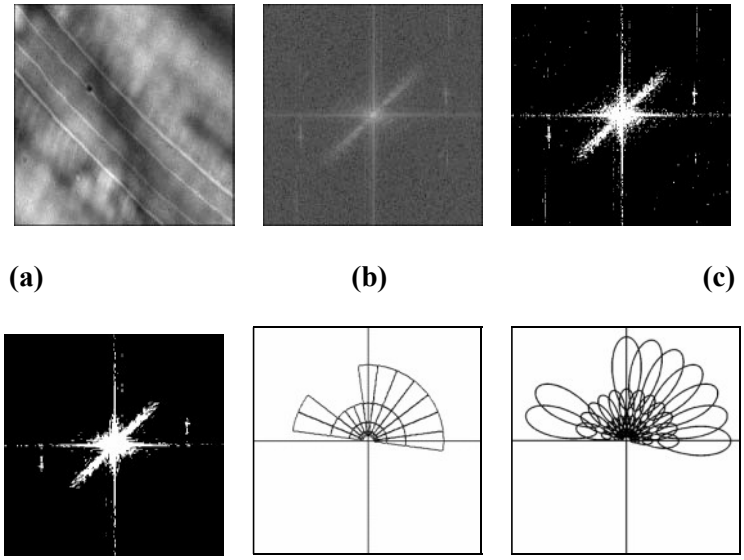


Fig. 3



(a) **(b)** **(c)**

(d) **(e)** **(f)**

Fig. 4

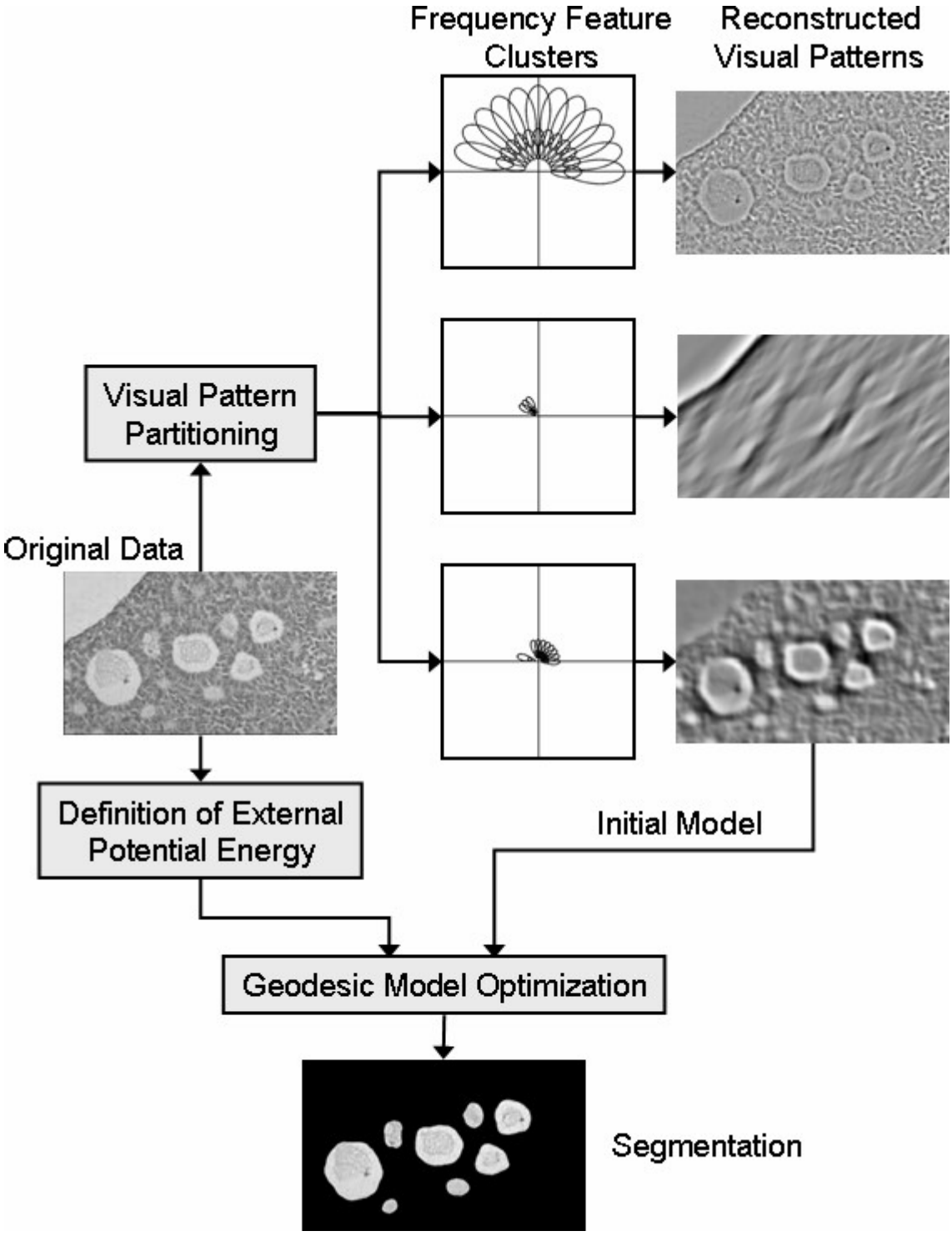

Frequency Feature Clusters
Reconstructed Visual Patterns
Visual Pattern Partitioning
Original Data
Definition of External Potential Energy
Initial Model
Geodesic Model Optimization
Segmentation

Fig. 5

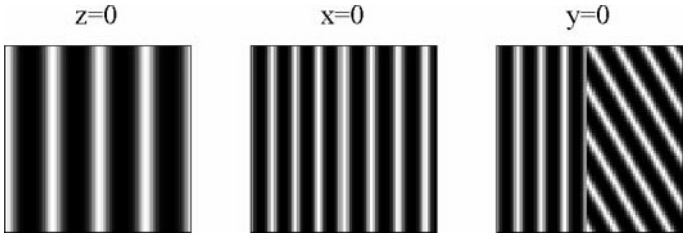


Fig. 6

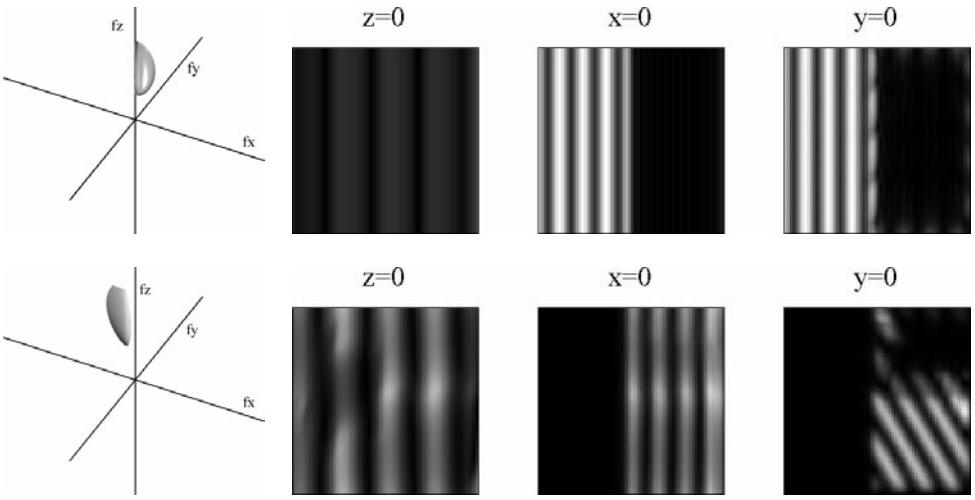


Fig. 7

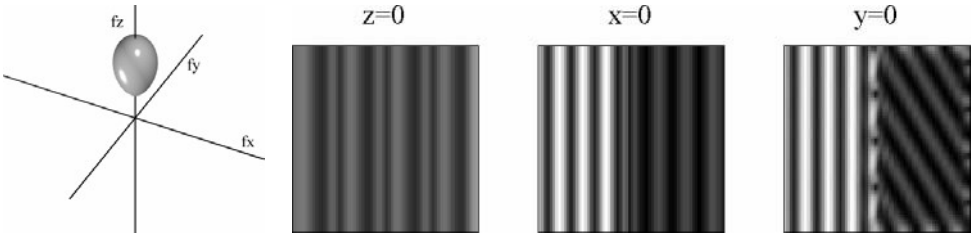


Fig. 8

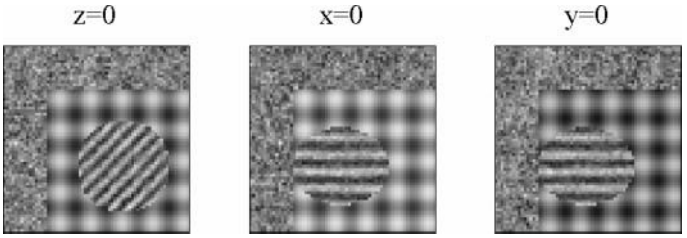


Fig. 9

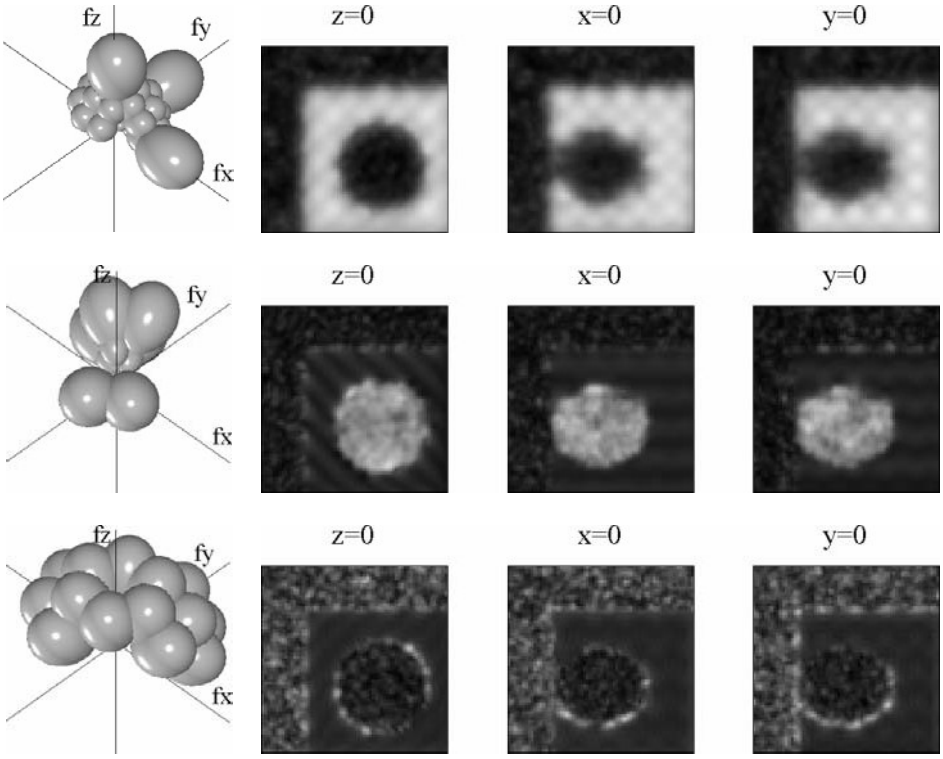


Fig. 10

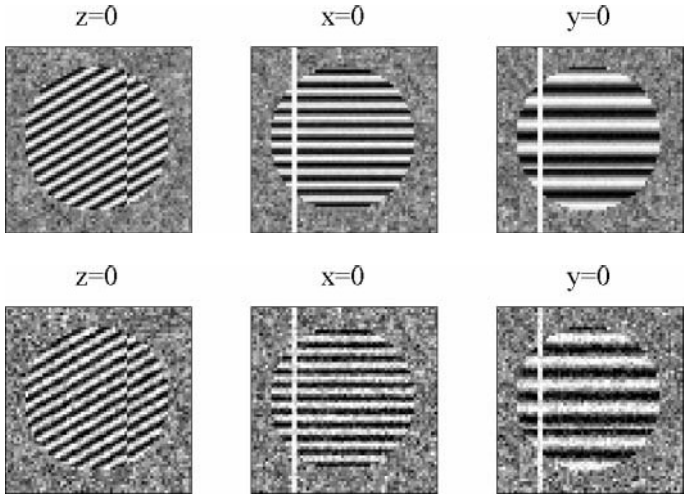


Fig. 11

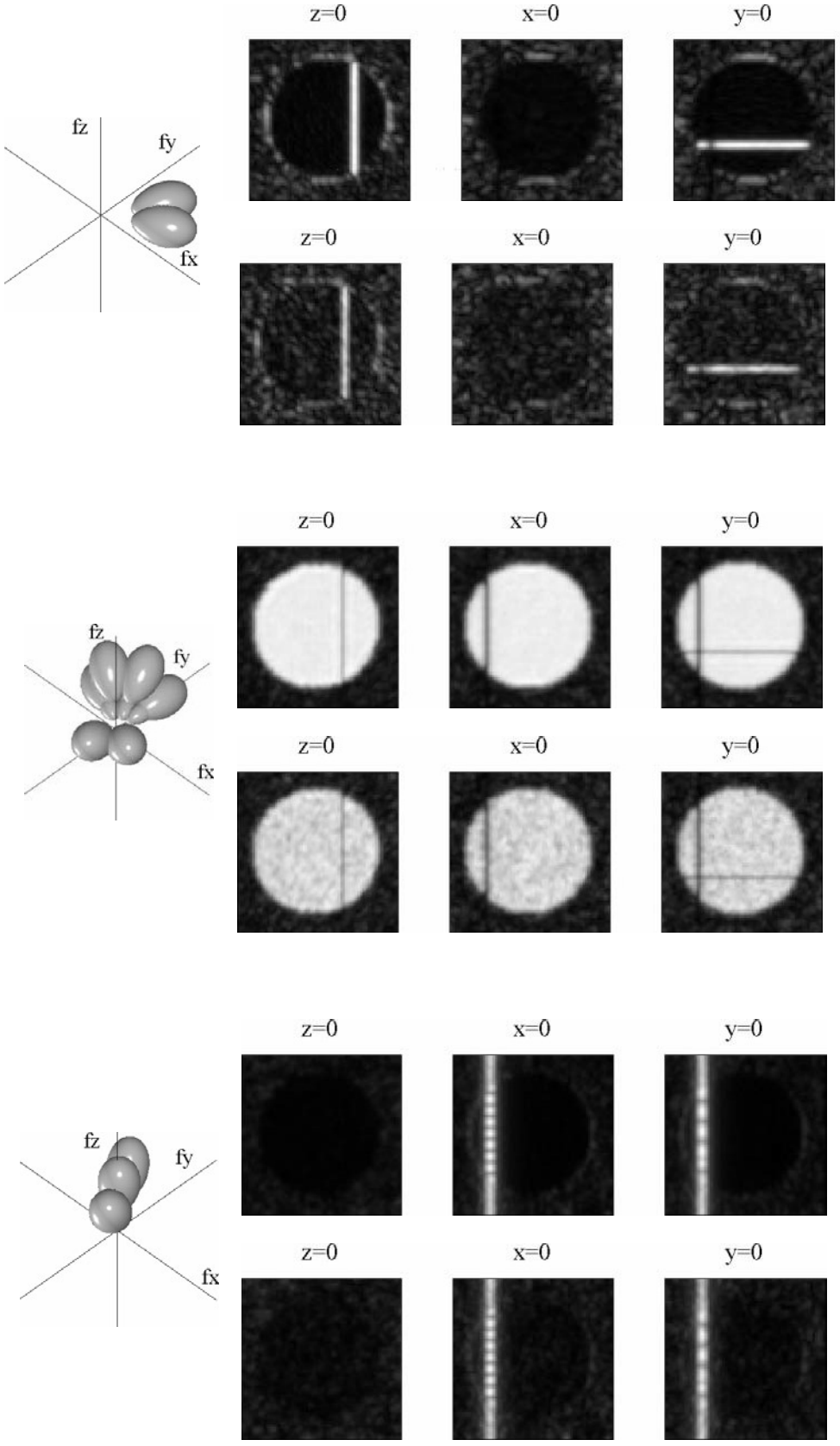

fz
fy
fx
z=0
x=0
y=0
z=0
x=0
y=0
fz
fy
fx
z=0
x=0
y=0
z=0
x=0
y=0
fz
fy
fx
z=0
x=0
y=0
z=0
x=0
y=0

Fig. 12

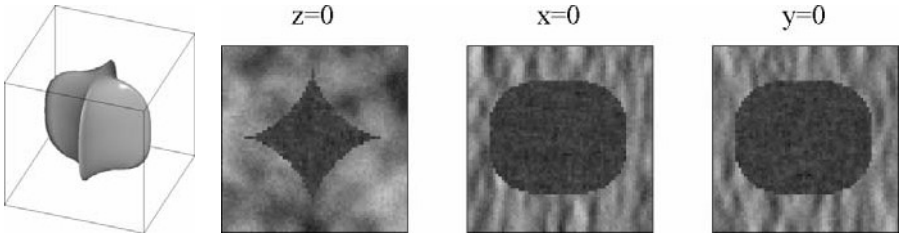


Fig. 13

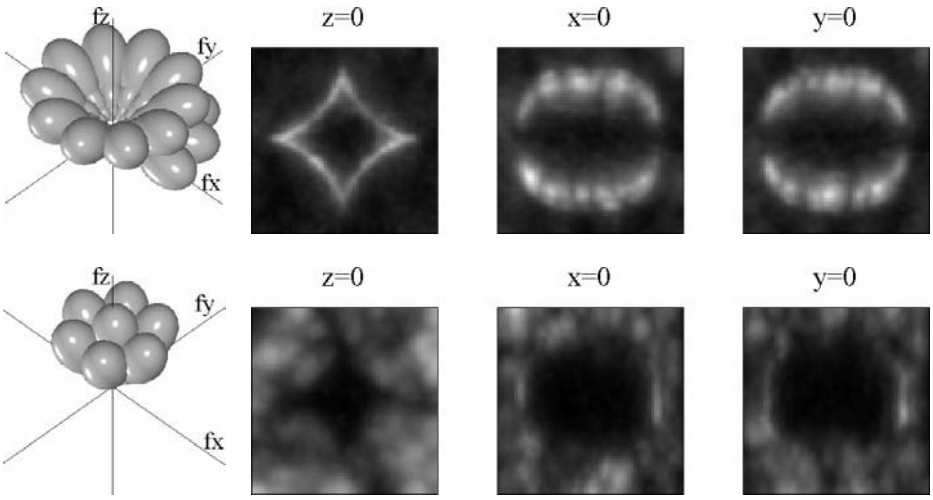
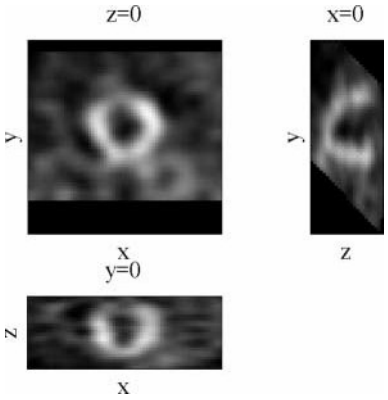
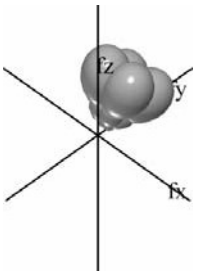
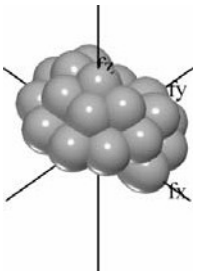


Fig. 14

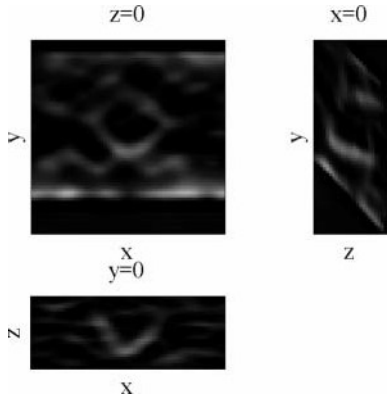


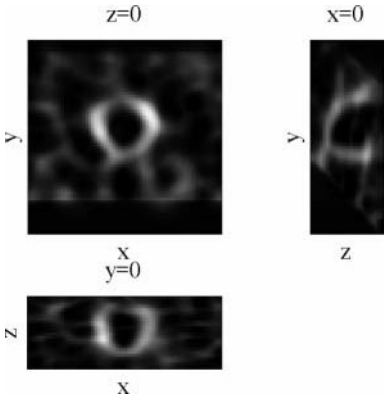
(a)





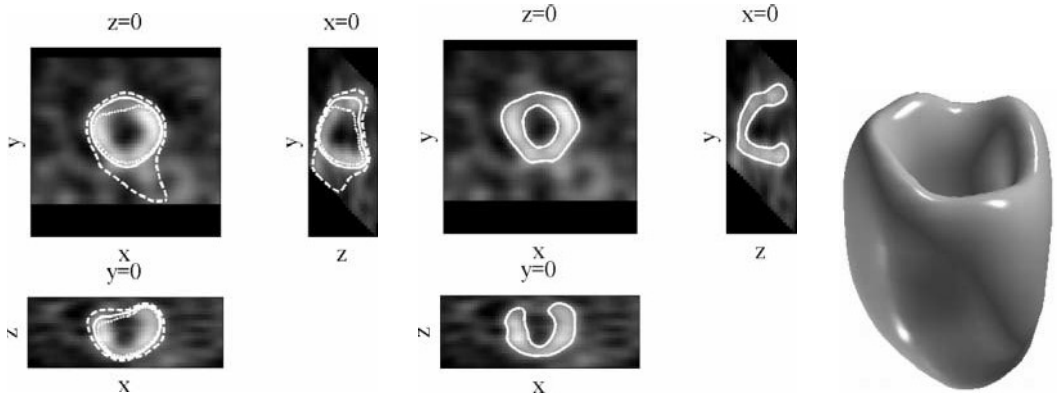
(b) (c)





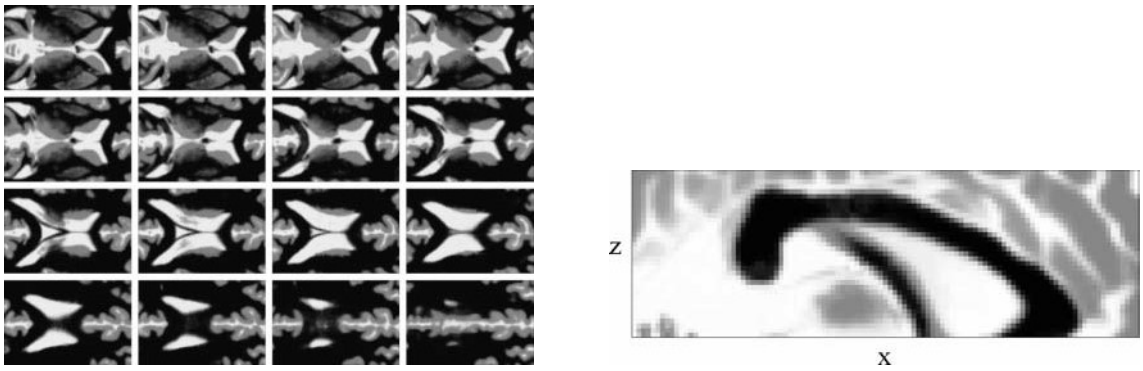
(d) (e)

Fig. 15



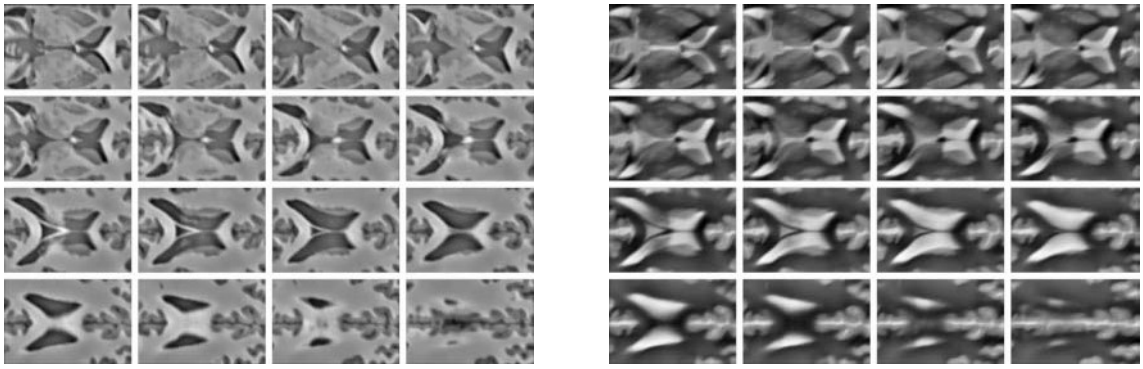
(a) (b) (c)

Fig. 16



(a) (b)

(c) (d)



(e) (f)

Fig. 17

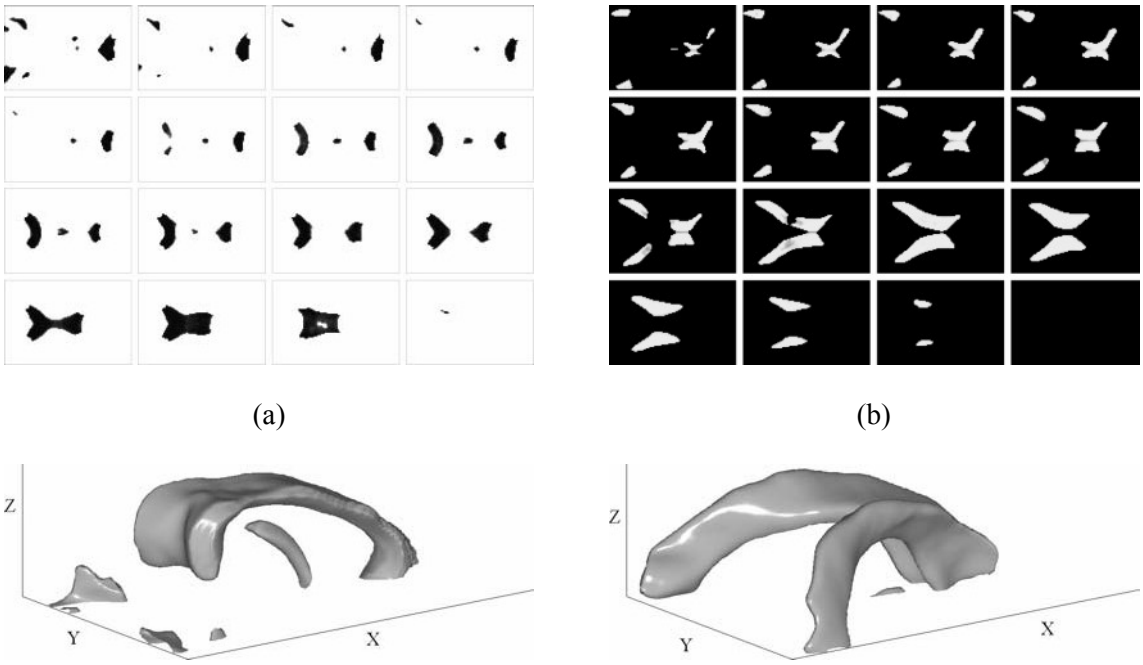
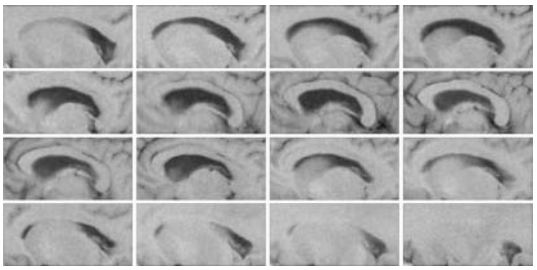
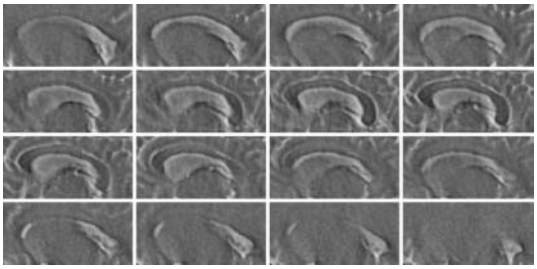


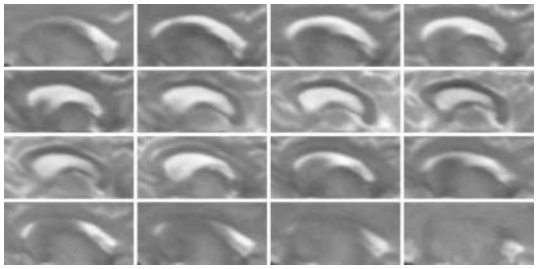
Fig. 18



(a)

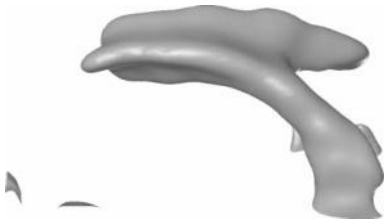


(b)



(c)

(d)



(e)

Fig. 19

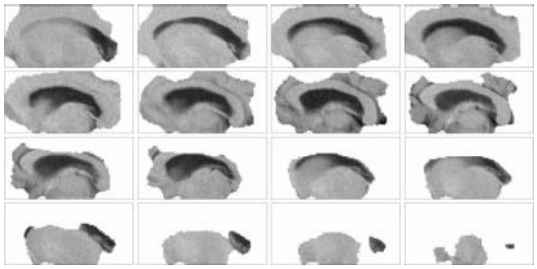


Fig. 20

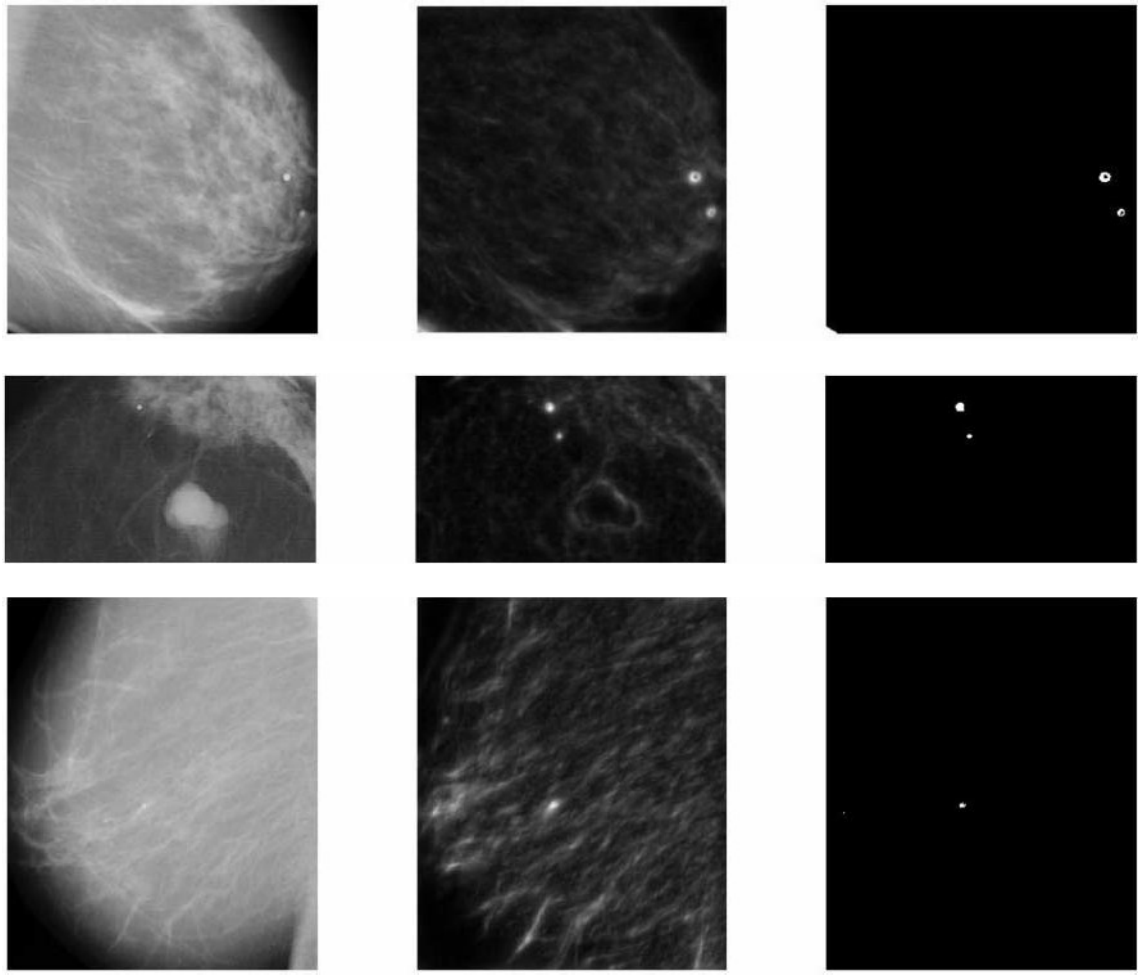


Fig. 21