Distances between Frequency Features for 3D Visual Pattern Partitioning

Abstract. In this paper we propose a technique for the decomposition of a 3D image into a set of low level patterns associated to phase congruency, which we call *visual patterns*. Those patterns have frequency components in a wide range of bands that are aligned in phase. The method involves clustering of the band-pass filtered versions of the image according to a measure of congruence in phase or, what is equivalent, alignment in the filter's responses energy maxima. This is achieved by defining a distance between the responses of pairs of filters and applying a hierarchical clustering analysis to the resulting distance matrix. To measure the degree of maxima alignment we propose a set of alternative distances and study their suitability. From this study we conclude that a measure of linear dependence between the local energy of filters' responses is more appropriate than a more general measure of dependence.

Keywords: low level feature extraction, visual pattern partitioning, multiresolution analysis, image dissimilarity estimation.

1 Introduction

Image analysis involves creating a low level image representation of image contents. In some cases 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. It seems to be widely accepted that there is not a unique feature that brings a complete representation of an image, but a set of different features should be used and that these features should correspond to patterns easily identifiable by the human visual system (HVS). There is also agreement in that low level image features representing discontinuities on local properties, like intensity, texture or phase, are detected by the HVS. Therefore, operators that detect these features can yield a meaningful representation of an image. However, HVS can only detect features in a domain of two spatial dimensions plus time. Although 3D spatial features can be inferred from stereo pairs of 2D images, the low level processing mechanisms performed by the cortical areas of brain in mammals are 2D. Despite of this, the algorithms that simulate those mechanisms in 2D can be extended to 3D for their application to the analysis of volumetric data.

1.1 Energy Based Feature Detectors

Much effort has been made in the field of image analysis to achieve good detectors for discontinuities on image properties. Canny (1986) stated that a good detector should

fulfill the criteria of good detection, good localization and single response. According to Owens et al. (1989) a feature detector should also be a projection. One of the techniques that verify these conditions is the energy based filtering, introduced by Morrone and Owens (1987) and widely studied by several authors (Morrone and Burr, 1988, Owens et al., 1989, Perona and Malik, 1990, Venkatesh and Owens, 1990). In this technique, image features are identified as the local maxima of the local energy of the image, which is calculated as the sum of the squared responses of a pair of filters in quadrature. This kind of operators outperforms previous linear filters (Canny, 1986, Marr and Hildreth, 1980) 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 projections.

Local energy maxima are also points of maximal phase congruency (PC), which measures the local degree of matching among the phase of Fourier components. Morrone and Owens (1987) sustain that the HVS perceives features at points of high PC. A good measure of PC must have into account the "amount" of frequencies that contribute to a feature to avoid, for instance, high responses to sine-wave signals, which are not perceived as features by the HVS. To this end, some techniques have employed broad bandwidth single-scale analysis (Morrone and Owens, 1987). However, most of them have opted for multiscale approaches (Morrone and Burr, 1988, Malik and Perona, 1990, Venkatesh and Owens, 1990, Kovesi, 1996). Multiresolution analysis is based on the decomposition of the image into a set of feature maps corresponding to different frequency bands and orientations. This is also consistent with the observed behavior of the HVS (Field, 1993).

To integrate the information from different frequency bands, most of the approaches combine their responses to produce one single feature map –or "primal sketch" (Marr and Hildreth, 1980) –of the image. Morrone and Burr (1988) developed a method to calculate local energy involving the sum of the responses of a bank of filters. Kovesi (1996) defined a new measure of PC based on statistical measures on the local phase of the responses of a set of 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.

An alternative solution is to determine what the spectral bands contributing to each image feature are. Separating the bands corresponding to each feature leads to a partition of the image into its most relevant low level features. This is the kind of analysis performed by Rodríguez-Sánchez et al. (1999) in the so-called RGFF model and extended in Chamorro-Martínez et al. (2003) and Dosil et al. (2004). In the RGFF, *visual patterns* or *integral features* are defined as patterns with alignment of frequency components in a set of local properties, not only energy but also entropy, contrast, etc. To isolate visual patterns the RGFF first uses a filter bank to decompose the image into its elementary frequency components. From now on we will call *frequency features* to the responses of the oriented band-pass filters in the bank. The information from different frequency bands is combined by grouping of similar frequency features together, using cluster analysis. To perform cluster analysis a suitable measure of dissimilarity between frequency features is necessary. Such distance must be small among those features belonging to the same visual pattern, i.e., it must depend on the degree of phase congruency or, equivalently, the degree of alignment among their local energy maxima.

1.2 Frequency Feature Dissimilarity Measurement

The dissimilarity employed by Rodríguez-Sánchez et al. (1999) and Chamorro-Martínez et al. (2003) to estimate the degree of alignment between two frequency features was inspired by biological processes, combining attention mechanisms and pooling of sensors' outputs. It 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.

On their part, Dosil et al. (2004) have chosen a distance based on the normalized mutual information of the local energy of filters' responses, which is less computationally expensive, less parameterized and less dependent on the performance of low level processes like non-maxima suppression and scale estimation. Mutual information is widely employed as a measure of image dissimilarity in different fields of application, with great popularity in medical image registration (Viola, 1995, Studholme et al., 1999).

However, we have observed that the behavior of this measure is not completely satisfactory. In some cases it produces the grouping of very dissimilar features together, like in the examples from Fig. 7 and Fig. 11. This is explained by the fact that mutual information treats intensity values in a qualitative fashion. It only considers their joint and marginal probabilities, not the difference between their values. As a result, it might take large values when the concurrence of weak and strong maxima takes place. To put it in another way, mutual information is an underconstrained measure of dependence because it makes no assumptions about the kind of functional dependence between the two images –see Roche et al. (2000) for a detailed explanation.

Thus, a measure that allows a general dependence between the images may not be the most appropriate in all applications. A measure of similarity that constrains the type of dependences between two images to those reflecting their true underlying relations should produce better results.

1.3 Our Approach

In this work we present a method for the decomposition of a 3D image into a set of low level features that brings a useful description of image contents for further use in high level applications. To this end, we employ a 3D filter bank, where frequency channels are represented by 3D log Gabor filters with rotational symmetry. Frequency features resulting from the application of this filter bank are classified into separate visual patterns by hierarchical cluster analysis.

Secondly, we study several image dissimilarity measures to find out which of them is the most appropriate to represent phase congruency. The set of distances has been extracted from the field of image registration, given that the task of registering two images involves the alignment of similar features together, which is a problem analogous to ours. In addition, we have considered the combination of the previous distances with attention mechanisms.

In section 2 the method for visual pattern partitioning is described. In section 3 we go into depth in the subject of comparing frequency features. The set of dissimilarity measures between pairs of filtered images is presented in subsection 3.1 and their

computational cost is analyzed in subsection 3.2. Section 4 presents an experimental study on the performance of these measures in the task of visual pattern partitioning. The results are discussed in section 5. Section 6 concludes the paper.

2 Visual Pattern Detection

As aforementioned, the method for visual pattern extraction consists of the decomposition of an image into 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 their spectral energy.
2. Calculation of the energy maps corresponding to the *active* filters' responses.
3. Measurement of dissimilarity between frequency features.
4. Hierarchical clustering of the frequency features based on the dissimilarity matrix.
5. Visual pattern reconstruction by linear summation of the energy of the filters in each cluster.

In the next subsections these procedures are detailed.

2.1 3D Filter Bank

Visual patterns are represented as a combination of the responses of a set of band-pass filters tuned in to 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 (Field, 1987)

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

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. Here, the scattering function is a Gaussian on the angular distance α between the directions of the filter and the position vector of a given point $\mathbf{f}$ in the spectral domain (Faas and Van Vliet, 2003)

$$\alpha(\phi_i,\theta_i) = \arccos\left(\mathbf{f}\cdot\mathbf{v}/\|\mathbf{f}\|\right), \tag{2}$$

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

$$S(\phi,\theta;\phi_i,\theta_i) = S(\alpha(\phi_i,\theta_i)) = \exp\left(-\frac{\alpha(\phi_i,\theta_i)^2}{2\sigma_\alpha^2}\right). \tag{3}$$

The shape of S from equation (3) is depicted in Fig. 1. It can be seen that S from expression (3) has rotation symmetry.

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. Following this criterion, the filter bank has been designed using a number of orientations $N = 6$ in half equator, this is, the region with $\{\theta_i = 0; \phi \in [0, \pi]\}$, producing 23 3D orientations. The angular bandwidth is 25º. In the radial axis, four values have been taken, with wavelengths 4, 8, 16 and 32 pixels, and 2 octaves bandwidth. These settings yield 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 these cases, their responses approximately represent the average intensity level. Therefore, these bands are discarded and only the bands with 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.

Fig 1.

2.2 Selection of Active Bands

To decrease the computational cost, the number of filters is reduced by discarding filters with wavelengths greater than half the image size, roughly representing the average intensity, and with low information content, named non-*active*. The measure of information density is $E = \log(|F| + 1)$, where F is the Fourier transform of the image.

A band is *active* when it comprises any value of E over the maximum spectral noise. The maximum noise level is estimated as $m + x\sigma$, where m is the mean noise energy, σ is its standard deviation and $x \geq 0$. Here, m and σ have been measured in the band of frequencies greater than twice the largest of the bank's central frequencies. Assuming that the spectral noise is uniform and that it fits a Gaussian distribution, most of the spectral noise energy is eliminated by taking $x = 3$.

To eliminate remaining spurious noise "spots" an especial median operator is applied. 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, for a given pixel in the spectral domain, it only considers neighbors that are anterior or posterior in the radial direction. This eliminates isolated peaks but preserving the continuity of structures along scales. The expression of the radial median filter mask M of size $L \times L \times L$ is as follows

$$\mathrm{M}(q, p) = \begin{cases} 1 & \text{if } [(q-p) / \| q-p \|] = [p / \| p \|] \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where p and q are points in the image and mask domains respectively, $[\cdot]$ represents rounding to the nearest integer and the origin is placed in the image center. We choose a

mask size $L = 3$. Larger masks bring little improvement and make the filtering much slower, as the mask coefficient must be calculated for each pixel location. The behavior of this filtering is illustrated in Fig. 2.

Fig 2.

2.3 Feature Clustering

Our approach generates visual patterns by clustering of active bands. Dissimilarities between each pair of frequency features are computed to build a dissimilarity matrix. To determine the clusters from the dissimilarity matrix, 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 the this case, the input data are the energy maps, giving place to a space of as much as N^3 coordinates. Agglomerative methods are better suited, since they work directly with the dissimilarity matrix, not having into account data coordinates.

Hierarchical clustering has been applied using a complete-link algorithm. The distance between clusters is defined as the maximum of all pairwise distances between features in the two clusters, thus producing compact clusters. The number of clusters N_c that a hierarchical technique generates is an input parameter of the algorithm. 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. A modification of the Davies-Boulding index proposed by Pal and Biswas (1997) has proved to produce good results for our application. It is a graph-theory based index that measures the compactness of the clusters in relation to their separation.

3 Measuring Feature Phase Alignment

Phase alignment imposes a relationship among the energy of the responses of a subset of frequency channels, so that one frequency feature is, to some point, predictable from another frequency feature belonging to the same visual pattern. Given that statistical dependence is defined as the predictability of a random variable given knowledge of another one, then there exists some degree of statistical dependence among frequency features if they are aligned in phase. Consequently, it seems reasonable to expect statistical measures of dependence to produce good results when applied to frequency feature clustering.

The question now is what kind of measure is best suited to reflect phase alignment. The use of a measure that is invariant to unexpected or undesired types of dependence will prevent from grouping of features belonging to different patterns. On the other hand, an overconstrained measure could discard allowed relations. To determine which is the most appropriate measure one has to make some assumptions regarding the nature of the relations among the components of a 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, but total dependence is not possible, i.e., one response cannot

be perfectly predicted by the other, given that they are versions of the same image viewed through different sensors, this is, the filters' parameters are different.

The assumptions adopted for the present application are quite similar to those imposed in the field of image registration, like in multimodal medical image registration. The difference is that in image registration the alignment might not be produced among maxima, but among the intensity values associated to the same tissue in the various imaging modalities. Despite this difference, it is likely that the image distances employed in such application are also suitable for frequency feature comparison.

One of the most popular measures in image registration is mutual information, I, (Viola, 1995) and its normalized versions (Studholme et al. 1999). I is a measure of arbitrary statistical functional dependence. This means that this measure does not make any assumptions regarding the kind of functional relation between the intensity values in the two images. Another measure of this kind is the correlation ratio, η (Roche et al., 1998, 2000). Unlike I, η assumes that the joint and marginal probability density distributions of the intensity levels in the two images can be properly modeled by means of Gaussian functions.

These measures can be considered underconstrained, because they allow any kind of functional dependence. However, as posed before, the alignment should be produced only among maxima in the two images. A more restrictive measure is the correlation coefficient ρ. Nevertheless, this measure constrains the possible functional relation to be linear, so it could be too restrictive for our application.

Alternative measures can be obtained by modifying the previous measures so as to directly apply them over energy maxima maps instead of raw energy maps. In this way, it is ensured that the correspondence is produced between energy maxima pairs and not between other concurrent intensity values. This can be easily accomplished by applying non-maxima suppression to the energy of the frequency features before distance estimation. The weakness of this approach is that both the number and the location of energy maxima are strongly influenced by scale, so that there is not perfect match among maxima locations in different features of the same pattern. For this reasong, this measure is expected to disjoint the bands composing a visual pattern. This problem could be reduced considering also the regions of influence of each energy maxima. Even when the maxima of two energy maps do not match, the energy profiles in the maxima surroundings should have a similar trend. Furthermore, if the neighborhood of the maxima is not taken into account the notion of size of the spatial features is lost.

The previous approaches can be considered to introduce attention mechanisms, like those that take place in the HVS. Nevertheless, this is the only aspect in which the distance estimation emulates the biological process. On the other hand, the RGFF model presented in Rodríguez-Sánchez (1999) does combine the information from the attention points, simulating the pooling of the responses that is supposed to occur in the HVS (Quick, 1974, Graham, 1989). The main drawbacks of attention based measures are their dependence on the selection of threshold parameters and on the performance of local maxima detectors.

To study the suitability of the different alternatives, an experimental study on their performance has been realized. Next subsection presents a formal definition of the diverse distances used in the test. Section 4 presents the tests and its results and in

section 4.2 these results are discussed and the conclusions obtained from them are presented.

3.1 Dissimilarity between Energy Maps

3.1.1 Statistical Measures of Dependence

All of the dissimilarity measures presented here are derived from a similarity measure δ by applying to it a transformation to enhance intercluster distances, invert its range and limit it to the interval [0, 1]. What follows is the list of similarity measures δ followed by the distance functions D_δ derived from them. In our notation, X and Y represent two energy maps and M is the number of bins in histogram calculations.

a) Normalized Mutual Information

The mutual information I of two random variables is the reduction of uncertainty of the first variable brought by the knowledge of the second variable when Shannon's entropy H is used as the measure of uncertainty. It can also be considered the Kullback's divergence between the joint probability distribution of the two variables $P_{X,Y}$ and the predicted distribution in case of total statistical independence $P_{X\times Y} = P_X \times P_Y$.

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

where

$$H(X) = \sum_i P_x(i) \log P_x(i) \quad \text{and} \quad H(X,Y) = \sum_{i,j} P_{x,y}(i,j) \log P_{x,y}(i,j),$$

Studholme proposed a normalized measure of image distance for medical image registration (Studholme et al., 1999). This measure is independent of the marginal entropies of the two images under comparison, so that it measures uncertainty reduction regardless of the uncertainty amount itself. The normalized similarity measure employed here is derived from Studholme's distance and has the following expression

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

This 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 high distances and reduces small ones, enhancing the compactness of the clusters.

$$D_{NI}(X,Y) = \left(1 - \sqrt{NI(X,Y)}\right)^2 \tag{5}$$

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.

b) Correlation Coefficient

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

$$\rho(X,Y) = \mathrm{Cov}(X,Y) \Big/ \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$$

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

$$D_\rho(X,Y) = \left(1 - \sqrt{(1+\rho(X,Y))/2}\right)^2 \tag{6}$$

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 over any divergence measure, which involves the discrete estimation of joint and/or marginal probabilities.

c) Correlation Ratio

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

$$\eta^2(X \mid Y) = 1 - \mathrm{Var}\big(X - \mathrm{E}(X \mid Y)\big) / \mathrm{Var}\,(X)$$

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

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 is inverted and transformed to accentuate inter-cluster distances.

$$D_\eta(X,Y) = 1 - \sqrt{\max\left(\eta^2(X \mid Y), \eta^2(Y \mid X)\right)} \tag{7}$$

d) Toussaint's Distance

Besides mutual information, other definitions of dependence in information theory exist (Basseville, 1996). As said before, the mutual information is equivalent to the Kullback's divergence when Shannon's entropy is used as a measure of uncertainty. The Toussaint's distance is an alternative measure of divergence that has proved to produce good results in the field of medical image registration (Sarrut and Miguet, 1999).

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

$$D_T(X,Y) = \left(1 - \sqrt{T(X,Y)/T_{\max}}\right)^2, \quad \text{with} \quad T_{\max} = 1 - 2/(M+1) \tag{8}$$

e) Lin's K-Divergence

The Lin's K-divergence is another example of a divergence measure (Basseville, 1996). It has also been tested with success in medical image registration (Sarrut and Miguet, 1999).

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

$$D_{Kdiv}(X,Y) = \left(1 - \sqrt{K_{div}(X,Y)/K_{div\ \max}}\right)^2, \text{ with } K_{div\ \max} = \log\left(2M/(M+1)\right) \tag{9}$$

3.1.2 Attention Based Measures

a) RGFF

The distance used in the RGFF model (Rodríguez-Sánchez et al., 1999, Chamorro-Martínez et al., 2003), which here we will call D_{RGFF}, is inspired in biological processes, combining attention mechanisms and Quick pooling of sensor outputs (Quick, 1974, Graham, 1989). 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 fixation point.

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

where Ω_X is the set of fixation points in X and μ_p is the weighted sum of differences between a set of local statistics at fixation points

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

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

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

$$D_{RGFF}(X,Y) = D_\beta^2(X,Y) + D_\beta^2(Y,X) \tag{10}$$

In our experiments the local statistics T_k are calculated weighting the contribution of each point in the neighborhood of a maximum with a Gaussian function on the distance to the maximum, so that the closer the neighbors, the more important the contribution in the calculation is. The local statistics used here are local energy and its local entropy.

b) Global Measures Over Energy Maxima Maps

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

$$D^*_\delta(X,Y) = D_\delta(X',Y') \tag{11}$$

where X 'and Y ' are respectively X and Y after thresholding followed by non-maxima suppression, i.e., after changing the intensity values to zero at points which are not a maximum. To determine if a point is an energy maximum its intensity is compared with the points immediately before and after in the filter's direction. The threshold must be carefully selected to avoid the elimination of relevant peaks.

c) Global Measures Over Regions of Interest

Again, a new set of dissimilarity measures $\hat{D}_\delta$ is obtained from a distance δ by using only the points belonging to the regions of interest. The region of interest of a frequency feature is defined as the union of all local maxima and their neighborhoods.

$$\hat{D}_\delta(X,Y) = D_\delta^2\left(X(\Gamma_X), Y(\Gamma_X)\right) + D_\delta^2\left(X(\Gamma_Y), Y(\Gamma_Y)\right) \tag{12}$$

where Γ_X and Γ_Y are the sets of points belonging to the region of interest of X and Y respectively.

The regions of interest are obtained in a two-stage process. Firstly, non-maxima suppression is applied to the energy maps. Secondly, the neighborhood of each maximum is determined as all points within a sphere of radius equal to the distance between the maximum and its nearest minimum. The region of interest of a frequency feature is the logical union of all neighborhoods of all maxima. This is the same approach used in the RGFF.

3.2 Computational Cost of Dissimilarity Estimation

One of the main advantages of global measures in relation to the RGFF measure is their lower computational cost. When applied to 3D images, the cost of the calculation based on local statistics on attention points strongly increases. Here, an analysis of the asymptotic computational cost of both 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 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 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 $O(F^2 \cdot N^3)$.

The modified distances D_δ^* and $\hat{D}_\delta$ have extra computational burden due to non-maxima suppression and estimation of regions of interest respectively. Non-maxima suppression has computational cost $O(F \cdot N^3)$, so that the asymptotical limit remains the same for D_δ than for D_δ^*. The calculation of the regions of interest, however, is more time consuming due to the calculus of the scales for each attention point, which is $O(F \cdot N^7)$. The total computational time is $O(\max(F \cdot N^7, F^2 \cdot N^6))$

In the case of the D_{RGFF}, 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 its nearest minimum. In large scale filters the neighborhood radius is of 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 would remain $O(F^2 \cdot N^6)$.

4 Results

To study the behavior of the visual pattern partitioning method and to compare the performance of the presented dissimilarity measures, the method described in section 2 is applied to a set of 48 test images employing each one of the distances. The test bench has been designed to comprise a large variety of low level image features, including grating patterns, textures, surfaces, lines, blobs and junctions. Besides the ability of the method to detect different types of features, we are also interested in observing the behavior of the method in the presence of different number of features, concurrence of features of different kinds, interference among features and presence of noise. Example cases of data sets presenting these characteristics can be seen in Fig. 3.

It is quite difficult to determine if the results obtained for a 3D data set are correct by visual inspection. For this reason, all the 3D images in the bench are synthetic data sets of size 64×64×64. 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.

Together with the 3D synthetic images, some 2D cases from natural images have been incorporated for the sake of completeness. They include images from biomedical applications 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

textured region. Additionally, they may appear patterns corresponding to texture boundaries. Examples of these can be seen in Fig. 3.(a) and (b).

Fig. 3

The results for the cases shown in Fig. 3 are presented in Fig. 4 and Fig. 5. We only present the best results for each case to illustrate the capabilities of the method. It can be seen that the method is able to identify and reconstruct different kinds of low level features present in an image. In particular, the examples in Fig. 5.k and Fig. 5.l show, respectively, how the method separates features of different kinds and features that interfere with each other.

Figs 4, 5

The results obtained with the different measures are summarized in Fig. 6. It can be seen that D_ρ has the best performance, followed by the D_{PSNR}. Underconstrained measures including the ones based on divergences D_{NI}, D_T, D_{kdiv} and the one derived from the correlation ratio, D_η, present a lower performance. This kind of measures systematically fails, for example, in separating grating patterns with different orientations. The use of attention based measures like D^*_ρ, $\hat{D}_\rho$, D^*_{NI} and $\hat{D}_{NI}$, worsens the overall effectiveness regarding the raw measures, although the analysis of the individual results shows that this modification means an improvement in certain cases. The D_{RGFF} measure has the additional problem of having an excessive computational cost.

Fig 6

5 Discussion

As mentioned in section 3 and verified in the results from section 4, distances based on measures of arbitrary dependence cause errors because they take small values when dependences other than the ones expected appear. This can be clearly seen in the example of Fig. 7. When applying the method to the image in Fig. 3.n. using D_{NI}, the clustering algorithm is not able to separate the texture inside the circle from the other one. Instead, it decomposes the texture of the outer region into its vertical and horizontal components. It is easy to understand why when we observe some of the frequency features of the image, which are presented in Fig. 8. Although features with labels 12 and 13 belong to different visual patterns, their mutual information is high, $NI(12,13)=0.379$. Each feature is highly predictable from the other since they are almost inverse images. Then, $D_{NI}(12,13)=0.148$. This value is comparable to the distance between features 11 and 12, both associated to the outer visual pattern which is $D_{NI}(11,12)=0.126$. Hence, frequency features associated to different visual patterns are grouped together.

The problem illustrated in the previous example is common to all the distances based on measures of general statistical dependence, like Toussaint's distance, Lin's K-divergence or correlation ratio. That is why we consider these measures here as

underconstrained for the task in hands, which is in agreement with the assumption presented in section 3 regarding the kind of dependence among frequency features.

In section 3 we proposed two possible solutions for this problem: the use of a different measure which introduces constraints on the type of dependence, like the correlation coefficient and the use of attention mechanisms. Regarding the correlation coefficient, we have seen that its overall performance is the best. In the particular example of data the set in Fig. 3.n, the results, presented in Fig. 9 do not reproduce the previously reported problem. Here, $\rho(12, 13) = 0.863$, while $\rho(12, 11) = -0.816$. As the proposed distance D_ρ constraints the type of dependence to linear functional with positive slope, by having the sign of ρ into account, then D_ρ (12, 13) = 0.001 and $D_\rho(12,11)=0.486$. Phase alignment within the components of a visual pattern is therefore better represented using D_ρ.

Fig 7, Fig 8 and Fig 9

In relation to attention based measures, we were expecting that such distances could enhance the results obtained with underconstrained measures under the assumption that the elimination of weak maxima could avoid dependences not caused by phase alignment. This is illustrated in Fig. 10 with a 2D example. Frequency features with labels 13 and 31, in Fig. 10.c and d respectively, correspond to different visual patterns. However, they have energy maxima in the same locations: strong energy maxima in feature 13 are aligned with weak maxima in feature 31 and vice versa. The joint histogram of these two features, in Fig. 10.f, can be modeled by three elliptical Gaussians: one for the mapping between backgrounds and two for the mappings between strong and weak maxima. This does not mean a linear dependence but a composition of several linear trends. The correlation coefficient is not sensitive to this kind of dependence, but it reflects only pure linear dependences like the one appearing between features 31 and 32, associated to the same visual pattern –Fig. 10.g shows their joint histogram. On the other hand, normalized mutual information calculated from joint histogram in Fig. 10.f takes a relatively large value compared to other pairs from same or different visual patterns –see Table 1. The application of non maxima suppression and thresholding –see Fig. 10.f, g, h and i– before the estimation of *NI* leads to the joint histogram in Fig. 10.l, which in turn, produces a reduction on the mutual information regarding other feature pairs–see Table 1.

Table 1.

As can be seen in Fig. 11, in the previous 2D example, both D_ρ and D^*_{NI} outperform D_{NI} as was expected. However, the summary of the results for the whole test bench shows that attention based distances produce worse results in general. This is caused by the thresholding step. There is not a robust procedure to estimate energy thresholds. The thresholding technique used here consists in determining the energy level such that a given percentage of the total energy of the feature is preserved. This is working well in many cases but not in the totality of the test bench.

Fig. 10.

6 Conclusions

Visual pattern partitioning makes reference to the process of isolation of the constituent low level features that are perceptually relevant in an image. It involves the clustering of the frequency components of the image according to some distance reflecting the degree of alignment between them. In this paper we have presented a method for the visual pattern partitioning of 3D images that consists of the multiresolution decomposition of the image and the subsequent clustering of frequency components based on a measure of frequency feature distance. The examples presented indicate that the developed system is capable of discriminating among various types of low level features.

We have also discussed the suitability of a set of dissimilarity measures to this task. The distances selected have been taken from the field of multimodal medical image registration because the goals in the two applications are quite similar. The dissimilarities have been tested with a set of 2D and 3D images. The results obtained have shown that the correlation coefficient distance solves some of the problems observed with other measures and improves the original measure proposed in the RGFF in speed and performance.

Acknowledgments

The authors desire to acknowledge the Xunta de Galicia for their financial support of this work by means of the research project PGIDIT04TIC206005PR.

References

Basseville, M., 1996. Information: Entropies, divergences et moyennes. Technical Report 1020, IRISA, 35042 Rennes Cedex France.

Canny, J., 1986. A Computational Approach to Edge Detection. IEEE. Trans. on Pattern Analysis and Machine Intelligence. Vol. 8, pp. 679-698.

Chamorro-Martínez, J., Fdez-Valdivia, J.A., García, J.A., Martínez-Baena, J., 2003. 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.

Dosil, R., Fdez-Vidal, X.R., Pardo, X.M., 2004. Multiresolution Approach to Visual Pattern Partitioning of 3D Images, in Campilho, A., Kamel, M., (Eds.), LNCS 3211: Image Analysis and Recognition. Porto (Portugal), pp. 655-663.

Faas, F.G.A., Van Vliet, L.J., 2003. 3D-Orientation Space; Filters and Sampling. Scandinavian Conference on Image Analysis, pp. 36-42.

Field, D.J., 1987. 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.

Field, D.J., 1993. 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. Clarendon Press, Oxford, pp. 151-193.

Graham, N.V.S., 1989. Visual Pattern Analyzers. Oxford University Press.

Kovesi, P.D., 1996. Invariant Measures of Image Features from Phase Information, The University or Western Australia.
http://www.cs.uwa.edu.au/pub/robvis/theses/PeterKovesi/

Malik, J., Perona., P., 1990. Preattentive Texture Discrimination with Early Vision Mechanisms. J. Opt. Soc. Am. A. Vol. 7(5), pp. 923-932.

Marr, D., Hildreth, E., 1980. Theory of Edge Detection. Proc. R. Soc. Lond. B. Vol. 207, pp. 187-217.

Morrone, M.C., Burr, D.C., 1988. Feature Detection in Human Vision: a Phase-Dependent Energy Model. Proc. R. Soc. Lond. B. Vol. 235, pp. 221-245.

Morrone, M.C., Owens, R.A., 1987. Feature Detection from Local Energy. Pattern Recognition Letters. Vol. 6, pp. 303-313.

Owens, R., Venkatesh, S., Ross, J., 1989. Edge Detection is a Projection. Pattern Recognition Letters. Vol. 9, pp. 233-244.

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

Perona, P., Malik, J., 1990. Detecting and Localizing Edges Composed of Steps, Peaks and Roofs. Third Int. Conf. on Computer Vision, pp. 52-57.

Quick, R.F., 1974. A vector magnitude model of contrast detection. Kybernetic. Vol. 16, 65-67.

Roche, A., Malandain, G., Ayache, N., 2000. Unifying Maximum Likelihood Approaches in Medical Image Registration. Int. J. of Imaging Systems and Technology. Vol. 11, pp. 71-80.

Roche, A., Malandain, G., Pennec, X., Ayache, N., 1998. The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration, in LNCS 1496: MICCAI’98. Springer-Verlag, pp. 115-1124.

Rodríguez-Sánchez, R., García, J.A., Fdez-Valdivia, J., Fdez-Vidal, X.R., 1999. 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.

Sarrut, D., and Miguet, S., 1999. Similarity Measures for Image Registration. 1st European Workshop on Content-Based Multimedia Indexing, Toulouse, France, pp. 263-270.

Studholme, C., Hill, D.L.G., Hawkes, D.J., 1999. An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. Pattern Recognition. Vol. 32, pp. 71-86.

Venkatesh, S., Owens, R., 1990. On the Classification of Image Features. Pattern Recognition Letters. Vol. 11, pp. 339-349.

Viola, P.A., 1995. Alignment by Maximization of Mutual Information. Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Technical Report 1548 (1995).

Tables

Filter Pair	NI	D_{NI}	ρ	D_{ρ}	NI^*	D^*_{NI}
{32, 31}	0.3146	0.1752	0.8391	0.0017	0.5239	0.0763
{1, 31}	0.1157	0.5559	–0.0096	0.0878	0.0146	0.7729
{13, 31}	0.2041	0.3621	–0.1630	0.1247	0.0079	0.8302

Table 1. Values of the normalized mutual information, correlation coefficient and their corresponding dissimilarities for each feature pair in the last row of Fig. 10. Asterisk indicates that measures are taken over maxima maps.

Figure Captions

Fig 1. *Left*: Representation of S ($\phi_i = \pi/3$, $\theta_i = \pi/4$) in the (ϕ, θ) plane by isolevel curves. *Right*: Representation of S ($\phi_i = \pi/3$, $\theta_i = \pi/4$) over a unit radius sphere by isolevel curves.

Fig 2. 2D example of radial median filtering. (a) Input image 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 3. Examples of the different types of images used in the text. Cases in first row correspond to 2D images of (a) biomedical 2D data –a nematode in the example– and (b) 2D synthetic images composed from natural textured images –Brodatz textures. The remainder cases are 3D synthetic images represented by two cross-sections of the volume comprising different kinds of visual patterns: (c) regions with different grating patterns, (d) textures, (e) odd symmetric surfaces, (f) even symmetric surfaces, (g) surfaces caused by a phase shift, (h) lines, (i) blobs, (j) junctions –star-shaped junctions in this case–, (k) mixture of various features of different types, (l) superimposed features and (m, n) textures composed by grating patterns of different orientations and scales.

Fig 4. The best results obtained for example cases from (a) to (g) in Fig. 3. The second subscript is the index for the set of resulting visual patterns. All results correspond to the dissimilarity measure based on the correlation coefficient, except from (c) and (g), which can be obtained, for example, using D_{RGFF}.

Fig 5. The best results obtained for example cases from (h) to (m) in Fig. 3. The second subscript is the index for the set of resulting visual patterns. All results correspond to the dissimilarity measure based on the correlation coefficient, except from (i) and (h), which can be obtained, for example, using D_{RGFF}.

Fig 6. Percentage of correct (OK), incorrect (X) and indecisive (?) results for each distance.

Fig 7. The two visual patterns obtained for image in Fig. 3.m using D_{NI}. *Left*: Clusters of filters represented by isosurfaces of level $\exp(-1/2)$ of the filters' transfer function. *Right*: Patterns associated to the previous clusters.

Fig 8. For image in Fig. 3.m, *Left*: frequency feature with label 13, with $\{\lambda_{13}=16, \phi_{13}=2\pi/3, \theta_{13}=0\}$, associated to the inner region, *Middle*: frequency feature 11, with $\{\lambda_{11}=4, \phi_{11}=0, \theta_{11}=0\}$, associated to the outer region and *Right*: frequency feature 12, with $\{\lambda_{12}=8, \phi_{12}=0, \theta_{12}=0\}$, associated to the outer region.

Fig 9. The two visual patterns obtained for image in Fig. 3.m using D_ρ *Left*: Clusters represented by isosurfaces of level exp(−1/2) of the filters' transfer function. *Right*: Patterns associated to the previous clusters.

Fig 10. (a) 2D test image and some of its frequency features: (b) frequency feature with label 13, with $\{\lambda_{13}=4, \phi_{13}=\pi/3\}$, (c) frequency feature 31, with $\{\lambda_{31}=4, \phi_{31}=2\pi/3\}$, (d) frequency feature 32, with $\{\lambda_{32}=16, \phi_{32}=2\pi/3\}$ and (e) frequency feature 1, with $\{\lambda_1=4, \phi_1=0\}$. (f, g) Frequency features 13 and 31 respectively after non maxima suppression and binarization. (h, i) Frequency features 13 and 31 after thresholding of energy maxima and binarization. The bottom row shows the joint histograms for the energies of the following couples of frequency features from the previous row: (j) {31, 13} (k) {31, 32} and (l) {31, 13} from energy maps after non maxima suppression and thresholding.

Fig 11. (a, b) The two filter clusters obtained for image in Fig. 10.a using the dissimilarity D_{NI}, represented as the sets of level exp(−1/2) of the filters' transfer function. (c, d) Visual patterns associated to previous clusters. (e, f, g) The three filter clusters obtained for image in Fig. 10.a using dissimilarities D_ρ and D^*_{NI}, represented as the sets of level exp(−1/2) of the filters' transfer function. (h, i, j) Visual patterns associated to clusters in previous row.

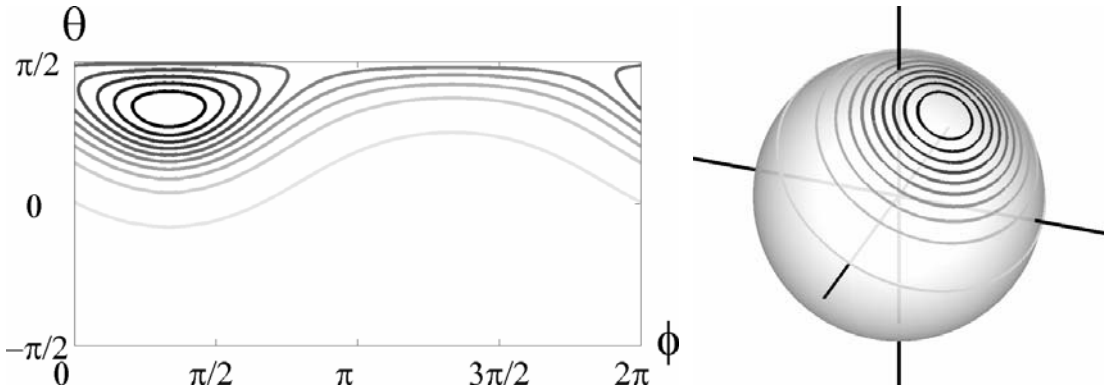
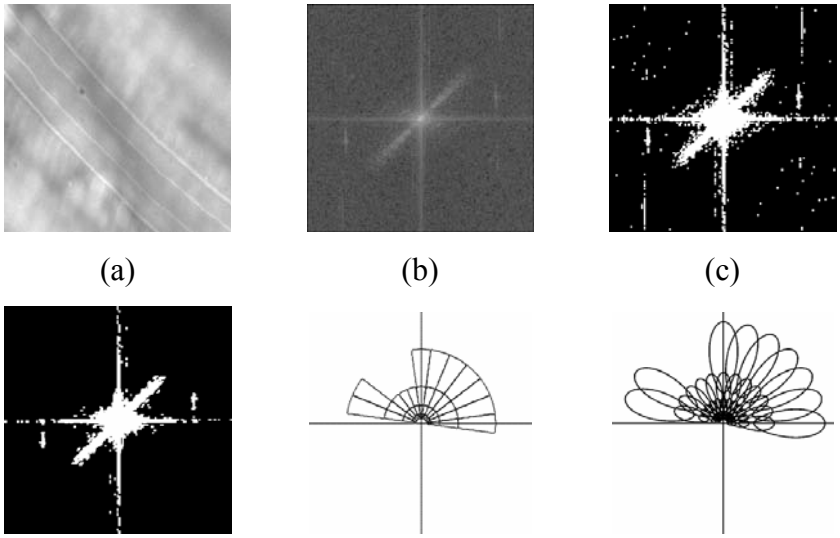


Fig. 1.



(a) (b) (c)

(d) (e) (f)

Fig. 2.

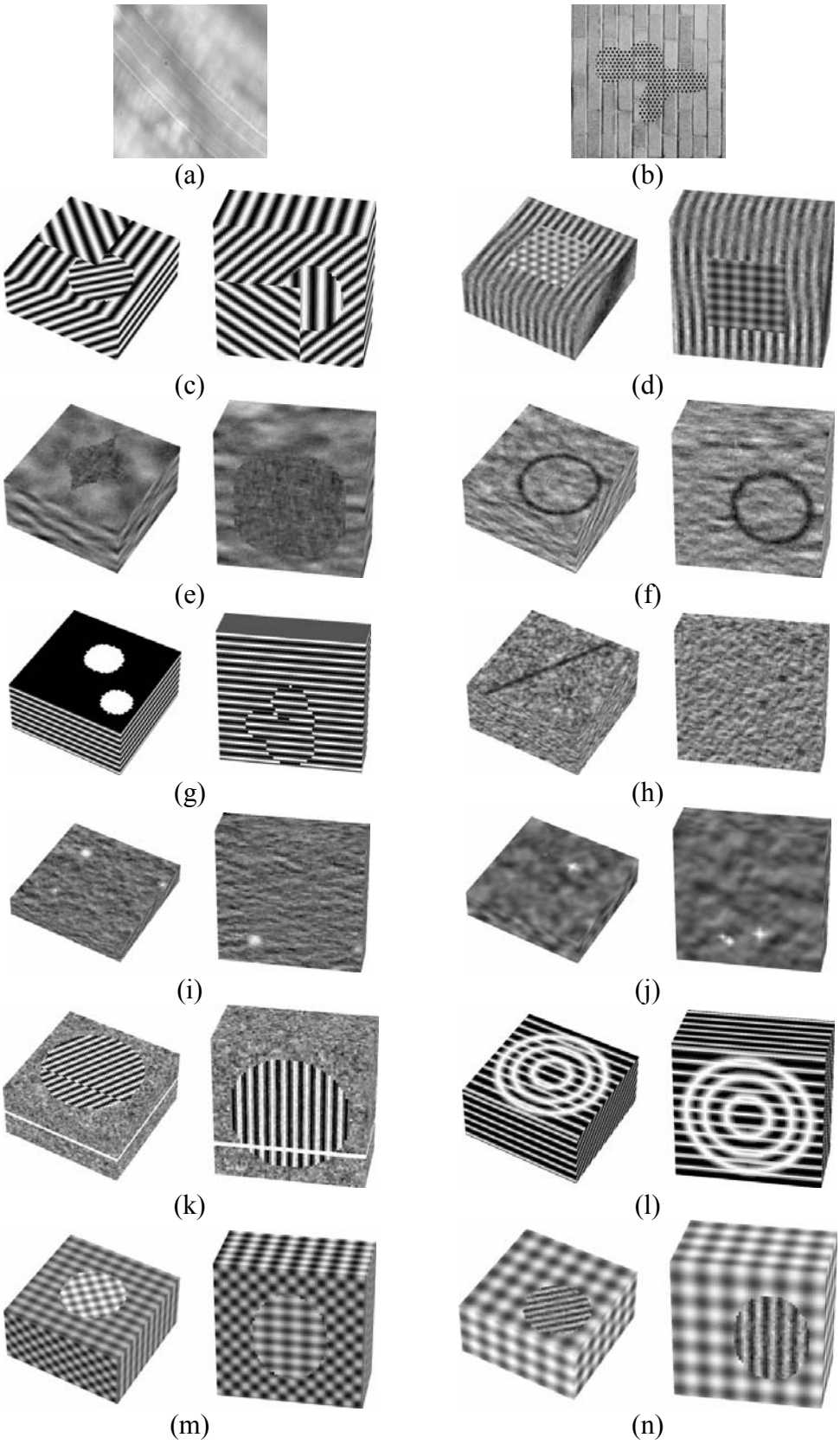


Fig. 3.

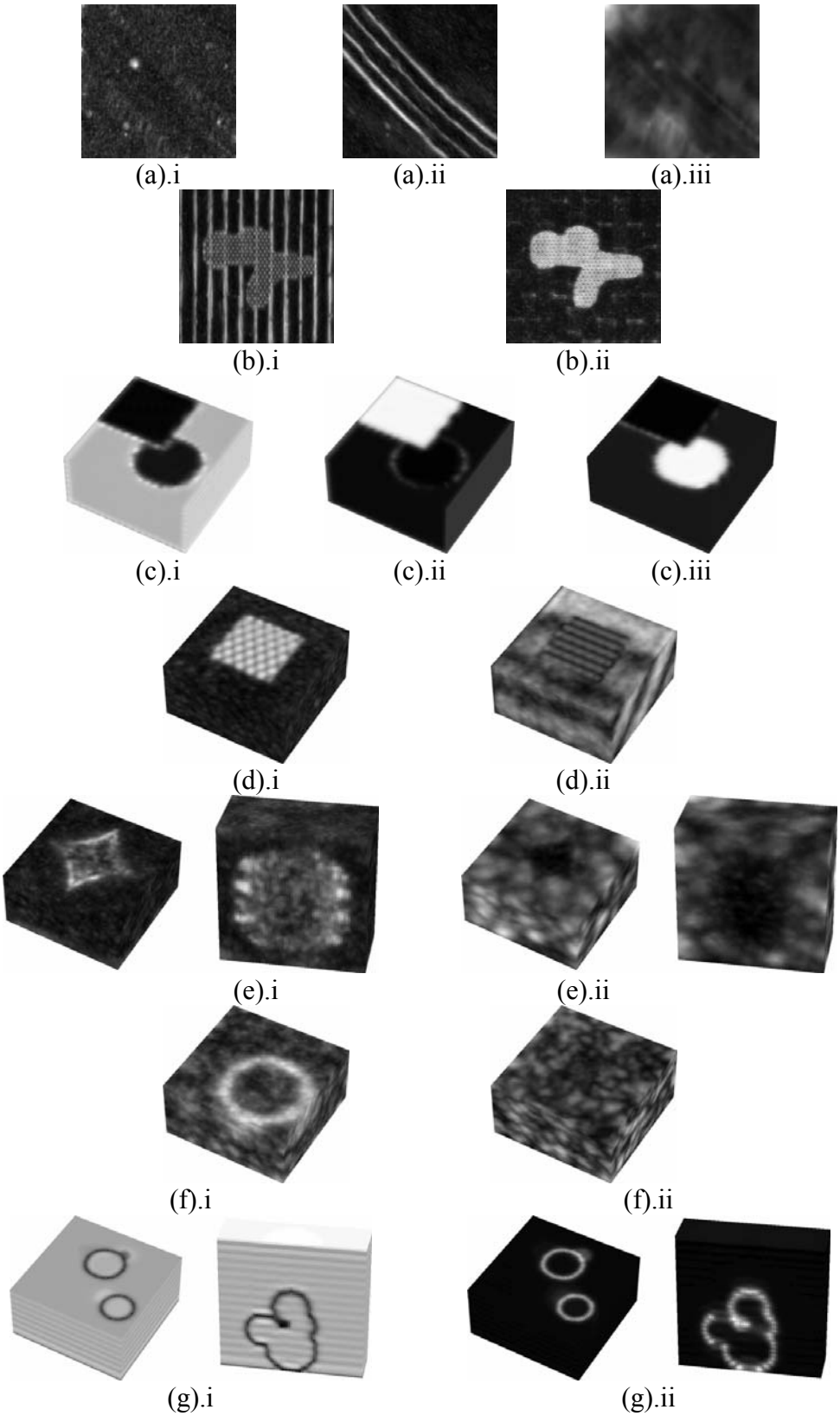
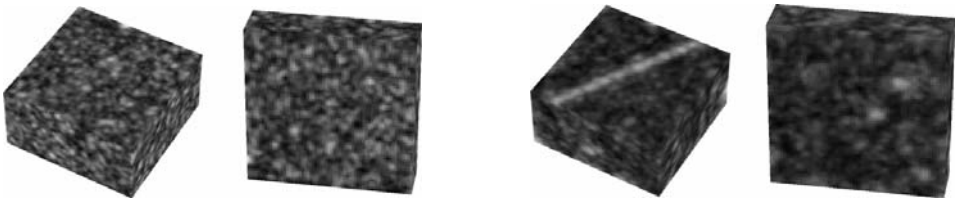


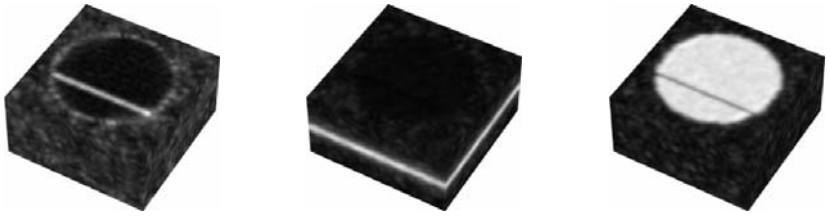
Fig. 4.



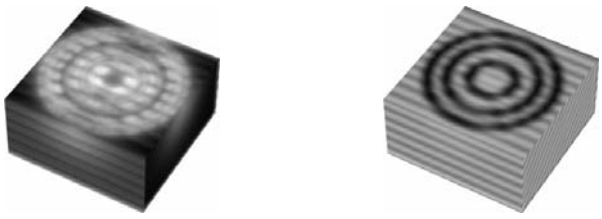
(h).i (h).ii

(i).i (i).ii

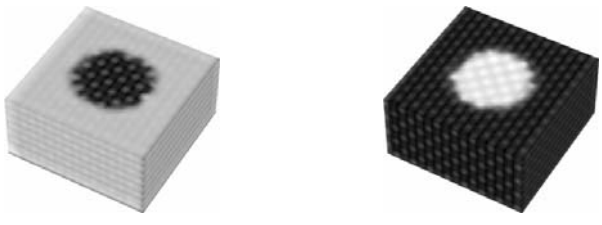
(j).i (j).ii



(k).i (k).ii (k).iii



(l).i (l).ii



(m).i (m).ii

Fig. 5.

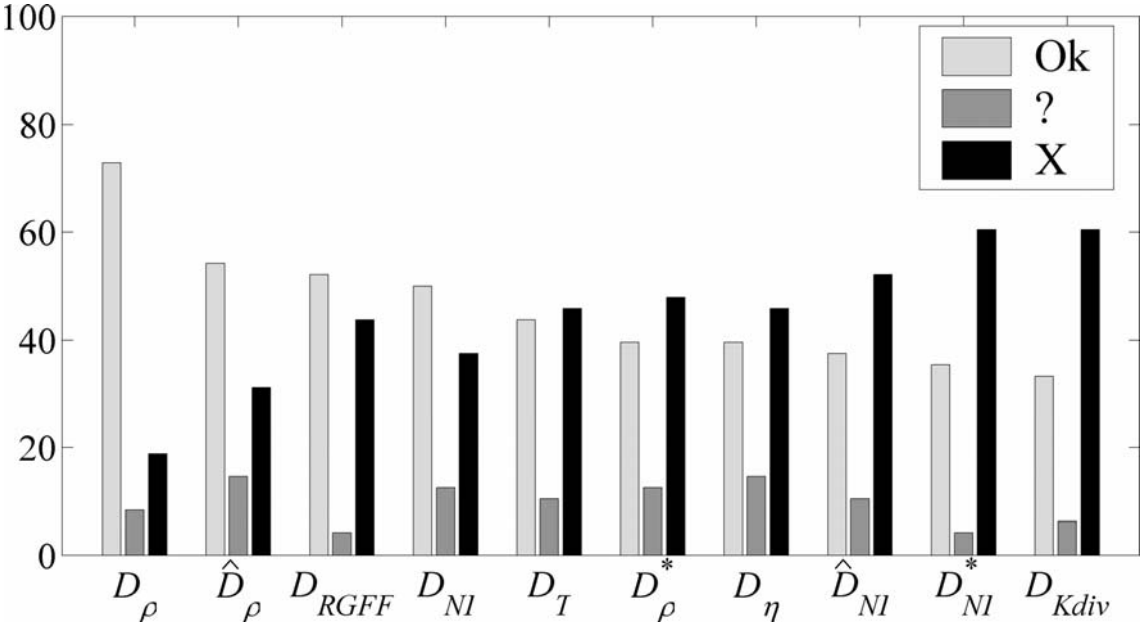

100
80
60
40
20
0
Ok
?
X
D_{ρ}
$\hat{D}_{\rho}$
D_{RGFF}
D_{Nl}
D_{T}
D^{*}_{ρ}
D_{η}
$\hat{D}_{Nl}$
D^{*}_{Nl}
D_{Kdiv}

Fig. 6.

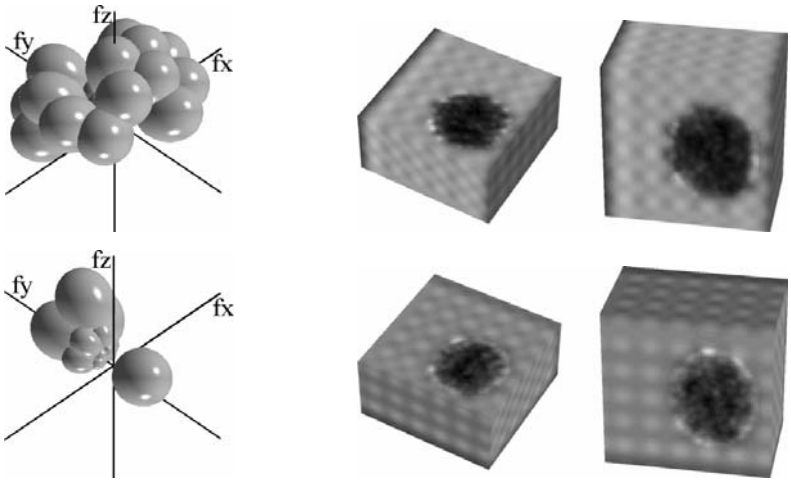


Fig. 7.

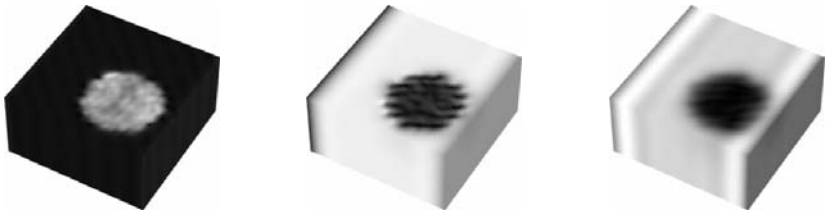


Fig. 8.

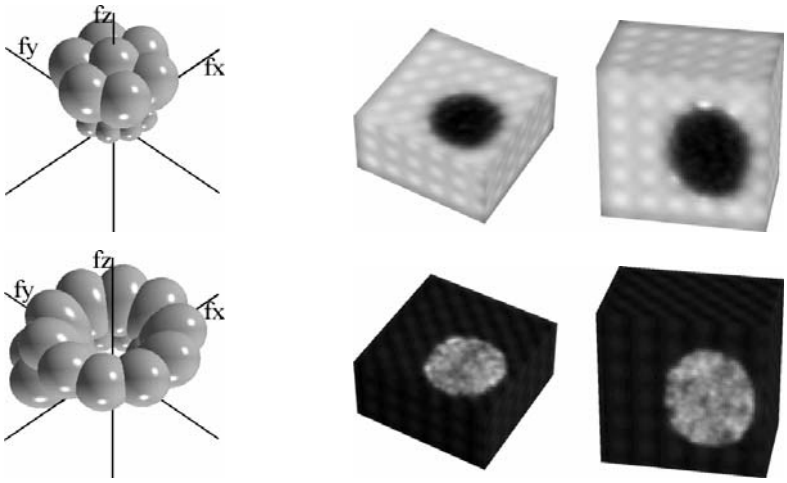


Fig. 9.

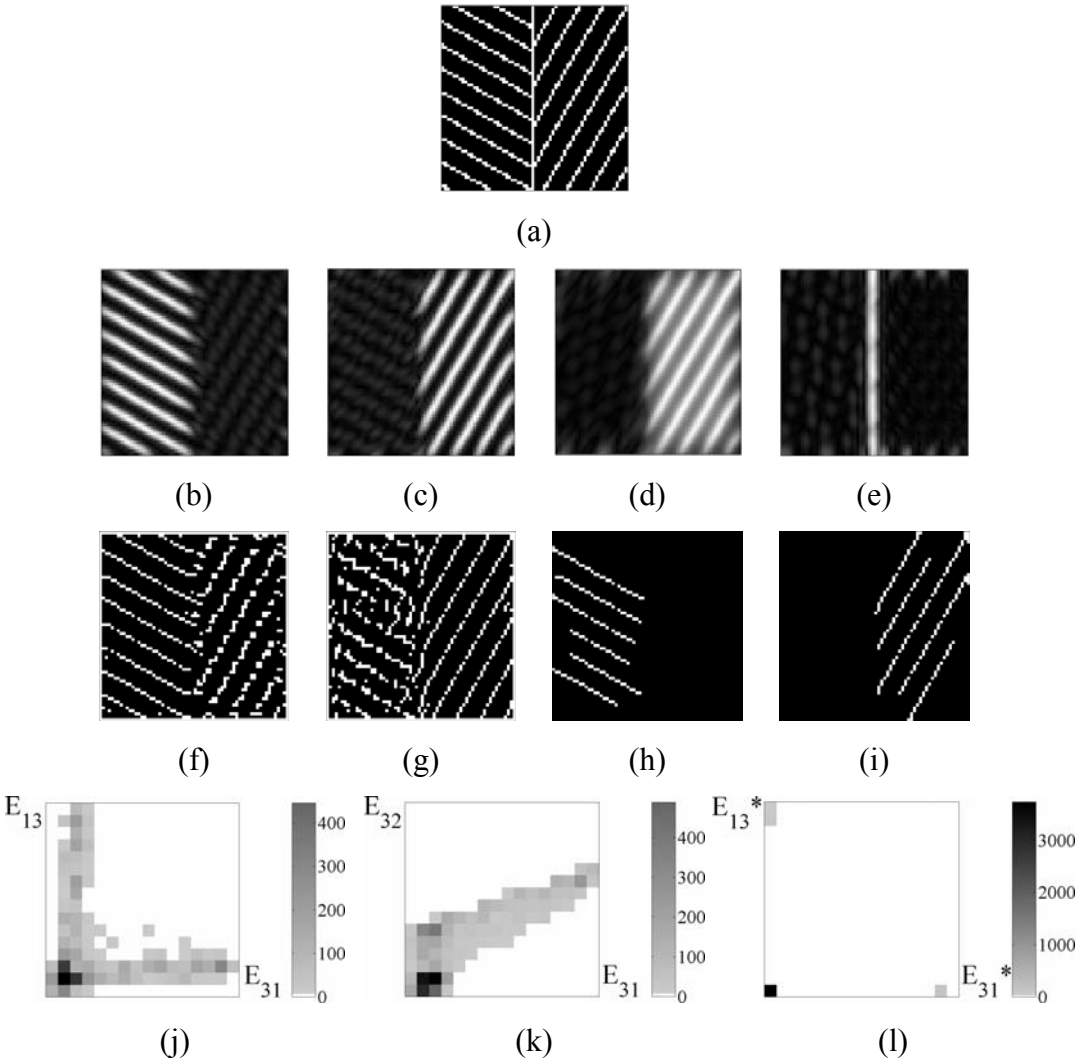

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
E13
E31
400
300
200
100
0
E32
E31
400
300
200
100
0
E13*
E31*
3000
2000
1000
0
(j)
(k)
(l)

Fig. 10.

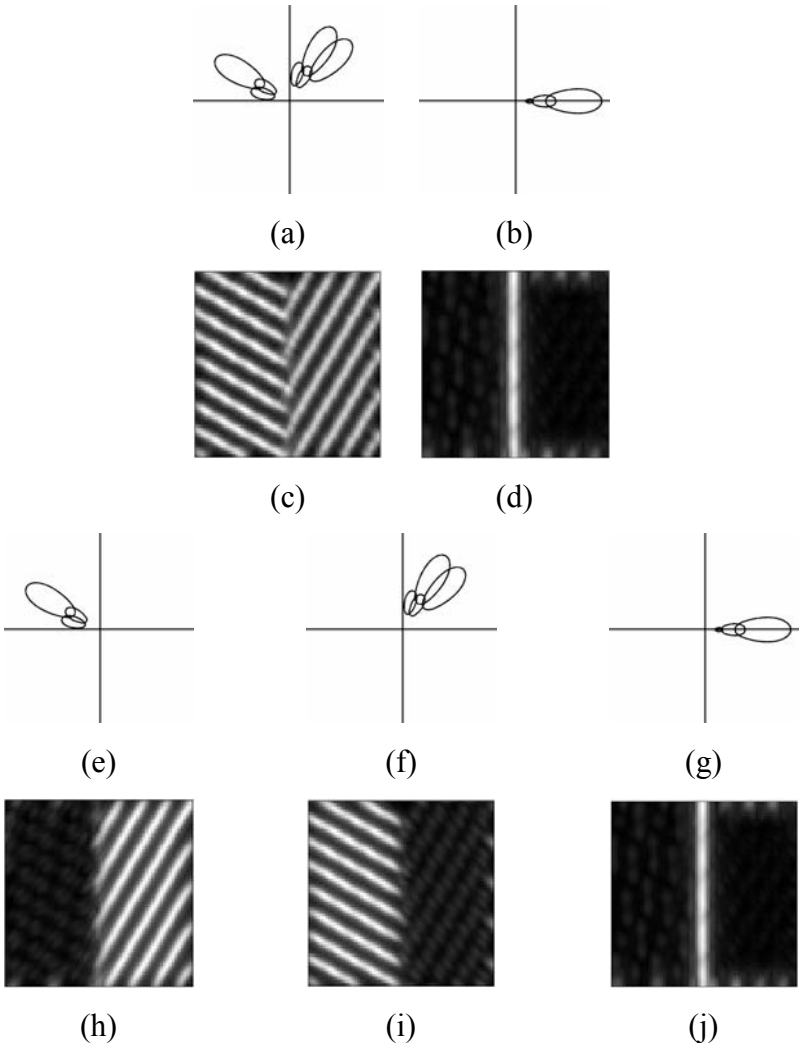


Fig. 11.