

Dissimilarity Measures for Visual Pattern Partitioning*

Raquel Dosil¹, Xosé R. Fdez-Vidal² and Xosé M. Pardo¹

¹ Dep. de Electrónica e Computación, Univ. de Santiago de Compostela,
Campus Universitario Sur, s/n, 15782, Santiago de Compostela, Spain
rdosil@usc.es, pardo@dec.usc.es

<http://www-gva.dec.usc.es/grupo/grupo.htm>

² Escola Politécnica Superior, Univ. de Santiago de Compostela,
Campus Universitario, s/n, 27002, Lugo, Spain
faxose@usc.es

<http://www.lugo.usc.es/~ffeeeps>

Abstract. We define a *visual pattern* as an image feature with frequency components in a range of bands that are aligned in phase. A technique to partition an image into its visual patterns involves clustering of the band-pass filtered versions of the image according to a measure of congruence in phase or, equivalently, alignment in the filter's responses energy maxima. In this paper we study some measures of dissimilarity between images and discuss their suitability to the specific task of misalignment estimation between energy maps.

1 Introduction

The identification and extraction of relevant low level features in an image is of great importance in image analysis. Field states that meaningful features present some degree of alignment in the phase of its spectral components [1]. In the RGFF representational model introduced in [2] and extended in [3, 4], such features are called *visual patterns* and defined as patterns with alignment in a set of local statistics along wide frequency ranges. These methods can detect a wide variety of features, like textures, grating patterns, blobs and symmetric and antisymmetric discontinuities in intensity, texture, and phase. They share a common scheme consisting of the decomposition of the image into elementary features using a bank of log Gabor filters followed by the clustering of these features according to some measure of dissimilarity among them.

The distance used in [2] and [3] is inspired in biological processes. It combines attention mechanisms and pooling of sensor outputs. Attention points are identified as energy maxima. On their part, Dosil et al. [4] use a distance based on the normalized mutual information of the filter's response energy, 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 I is

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

widely employed as a measure of image dissimilarity in various fields of application, with great popularity in medical image registration [5, 6]. However, we have observed that the behavior of I is not completely satisfactory for filter clustering. In some cases it groups very dissimilar frequency features. This is due to that I treats intensity values qualitatively, increasing with the concurrence of weak and strong maxima. I is an underconstrained measure of dependency since it makes no assumptions about the kind of functional relation between the images –see [7] for a detailed explanation.

Then again, a measure that allows a generic dependency between the images may not be the most appropriate in all applications. In the specific case of filter's responses energy maps it seems that the kind of dependency that best reflects the relation between features belonging to the same visual pattern is linear functional. A measure of similarity that constrains the allowed relations between two images to a linear transformation is the correlation coefficient. To test this assumption, here we make a comparison among a series of dissimilarity measures, including distances based on correlation coefficient, mutual information and the original measure proposed in the RGFF.

In the next section the set of dissimilarity measures between pairs of filtered images is presented. In section 3, the method for visual pattern partitioning is described. Section 4 presents an experimental study on the performance of these measures in the task of visual pattern partitioning. Section 5 presents the conclusions derived from it.

2 Dissimilarity between Energy Maps

All measures presented here are derived from a similarity measure δ by applying to it a transformation to enhance intercluster distances, invert its range and map it to the interval $[0, 1]$. What follows is the list of proximities δ and their correspondent distances D_δ . X and Y represent energy maps and M is number of bins in an histogram.

a) *Normalized mutual information* [4, 5]

If H stands for entropy, then

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

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

b) *Correlation ratio* η [7, 8, 9]

$$\eta^2(X | Y) = 1 - \text{Var}(X - E(X | Y)) / \text{Var}(X)$$

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

c) *Correlation coefficient*

This measure has into account the sign of the correlation coefficient, so that an image and its inverse have maximum distance

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

$$D_\rho(X, Y) = \left(1 - \sqrt{(1 + \rho(X, Y))/2}\right)^2 \quad (3)$$

e) *Toussaint's distance* [9, 10]

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

$$D_T(X, Y) = \left(1 - \sqrt{T(X, Y)/T_{\max}}\right)^2, \quad \text{with } T_{\max} = 1 - 2/(M + 1) \quad (4)$$

f) *Lin's K divergence* [9, 10]

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

$$D_{Kdiv}(X, Y) = \left(1 - \sqrt{K_{div}(X, Y)/K_{div \max}}\right)^2, \quad \text{with } K_{div \max} = \log(2M/(M + 1)) \quad (5)$$

g) *Dissimilarity measures on energy maxima.*

A new dissimilarity measure D_δ^* is obtained from each D_δ as follows

$$D_\delta^*(X, Y) = D_\delta(X', Y') \quad (6)$$

where X' and Y' are respectively X and Y after non-maxima suppression. Maxima are determined by comparing each point with its neighbors in the filter's direction.

h) *RGFF dissimilarity measure* [2]

For each energy map X , the set of its maxima Ω_X is determined. For each p in Ω_X a vector T^p of length Q of local statistics is measured. Then, for a given $\beta > 0$

$$D_\beta(X, Y) = \frac{1}{\text{Card}(\Omega_X)} \left(\sum_{p \in \Omega_X} |\mu_p(X, Y)|^\beta \right)^{1/\beta}, \quad \mu_p(X, Y) = \sum_{k=1}^Q \frac{1}{\omega_k} d(T_k^p(X), T_k^p(Y)).$$

$$D_{RGFF}(X, Y) = D_\beta^2(X, Y) + D_\beta^2(Y, X). \quad (7)$$

where ω_X is the maximum T_k over all Ω_X and all X . The local statistics they employ are local phase, normalized local energy and its entropy, contrast and standard deviation.

2.2 Computational Cost of Dissimilarity Estimation

One of the main advantages of global measures in relation to the RGFF measure is their lower computational cost. In the following, an analysis of the asymptotic computational cost of the presented 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 $O(F^2 \cdot N^3)$.

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 maxima 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 $O(F^2 \cdot N^6)$.

3 Visual Pattern Partitioning Methodology

Visual pattern partitioning of a 3D image consists of the next sequence of steps:

1. Selection of *active* bands—with high information content
2. Calculation of the energy maps correspondent to the *active* filters' responses
3. Measure of dissimilarity between pairs of energy maps
4. Hierarchical clustering of the energy maps based on the dissimilarity matrix
5. Visual pattern reconstruction by linear summation of cluster energy maps.

In the next subsections these procedures are detailed.

3D Filter Bank

The filters' transfer function T is designed as the product of separable factors R and S in the radial and angular components respectively with expressions

$$R(\rho; \rho_i) = \exp \left\{ - \log^2 (\rho / \rho_i) / (2 \log^2 (\sigma_\rho / \rho_i)) \right\}, \quad (8)$$

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

$$S(\phi, \theta; \phi_i, \theta_i) = S(\alpha) = \exp \left\{ - \alpha^2 / (2 \sigma_\alpha^2) \right\}, \text{ with } \alpha(\phi_i, \theta_i) = \text{acos} (\mathbf{f} \cdot \mathbf{v} / \|\mathbf{f}\|), \quad (9)$$

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, σ_α is the angular standard deviation and \mathbf{f} the point in the frequency space in Cartesians.

In our configuration elevation is sampled uniformly, while azimuth is non-uniformly sampled by maintaining equal arc-length between adjacent azimuth values over the unit radius sphere. The bank has been designed using 4 elevations –only one hemisphere is needed due to symmetry– and 6 azimuths to sample half the $z=0$ plane, yielding 23 orientations with angular bandwidth of 25° . In the radial axis, 4 values have been taken with wavelengths 4, 8, 16 and 32 and 2 octave bandwidth.

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 image Fourier transform.

A band is *active* if 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 that double the largest of the bank's central frequencies and $x = 3$.

To eliminate remaining spurious noise “spots” a radial median operator is applied, which 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. In this work the mask size is taken to $L = 3$.

Feature Clustering

Here, hierarchical clustering has been chosen to group features, using a complete-link algorithm, where the distance between clusters is defined as the maximum of all pairwise intercluster distances, thus producing compact groups. The number of clusters N_c is an input parameter of the algorithm. The usual strategy to determine the optimal N_c is to run the algorithm for each possible N_c and evaluate the quality of each resulting partition according to a given validity index. Here, the modified Davies-Boulding index, introduced in [9] 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.

4 Results

To compare the performance of the presented dissimilarity measures, the visual pattern partitioning method described in section 3 has been applied using each of them to a set of test images. The test bench is composed of 32 images, 13 of them 2D and the other 19 3D. While it is quite easy to determine if the results obtained for a 2D image are correct by visual inspection, this is more difficult for 3D images. For this reason, all the 3D images in the bench are synthetic. The correctness of the results is deter-

mined by comparing them with the design specifications. The result must contain one cluster for each visual pattern in the image and their frequency bands must match the expected ones. 2D cases are either synthetic images or natural images with clearly identifiable visual patterns or images synthesized as a collage of natural Brodatz textures. In this last type the result must contain one cluster for each texture. Additionally, they may appear patterns correspondent to texture boundaries.

The results obtained are summarized in Fig. 1. The measures have been sorted by percentage of correct results. It can be seen that D_ρ has the best performance, followed by D_{NI} . In general the results are not very good due to the complexity of the task of matching same-pattern frequency features, which present strong differences – these results can not be extrapolated to other applications, like image registration.

Fig. 2 shows one example result for a 3D image that presents diverse visual patterns: a grating pattern, a plane –even feature– and a phase change –odd feature. In this specific case all the distances produce the correct result, except from D_T and D_{NI}^* which do not detect the phase change.

The remainder figures illustrate the improvements brought by the use of the D_ρ distance in relation to the other measures. Fig. 3 shows an example of 2D synthetic image. It can be seen that the D_{NI}^* distance groups orthogonal grating patterns together while D_ρ separates them. This is caused by the non null response of the filters to patterns with orientation orthogonal to it. Given that NI does not consider the magnitude of the difference between the responses of the filters, the resulting dissimilarity is small. Fig. 4 shows a similar example with a natural image of a Brodatz texture.

Fig. 5 and Fig. 6 present other cases were D_ρ corrects D_{NI} results. In Fig. 5 mutual information is not able to separate the texture inside the circle. Instead it decomposes the texture of the outer region into its vertical and horizontal components. In the example of Fig. 6 the results for D_{NI} are not shown since they are a total of 7 clusters, as the different components of each region have not been correctly integrated.

5 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 consists of the clustering of the frequency components of the image according to some distance

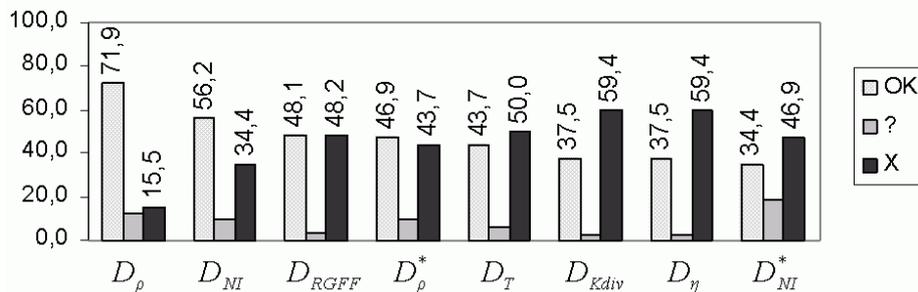


Fig. 1. Percentage of correct (OK), incorrect (X) and indecisive (?) results for each distance.

reflecting the degree of alignment between them. In this paper we have discussed the suitability of a set of dissimilarity measures to this task.

We have planted the assumption that the kind of dependency that appears between the frequency components of the same visual pattern is a linear functional one. This explains the incorrect results obtained with measures based on mutual information, other information divergences and correlation ratio. Upon this assumption we predict that a measure based on the correlation coefficient should yield better results.

To test this hypothesis the dissimilarities have been tested with a set of 2D and 3D images. The results obtained have shown that the correlation coefficient distance solves the problems observed with mutual information and other global distances and improves the original measure proposed in the RGFF in speed and performance.

References

1. Field, D.J.: 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 (1993) 151-193
2. Rodríguez-Sánchez, R., García, J.A., Fdez-Valdivia, J., Fdez-Vidal, X.R.: The RGFF Representational Model: A System for the Automatically Learned Partition of "Visual Patterns" in Digital Images, *IEEE Trans. Pattern Anal. Mach. Intell.*, **21**(10) (1999) 1044-1073
3. Chamorro-Martínez, J., Fdez-Valdivia, J.A., García, J.A., Martínez-Baena, J.: A frequency Domain Approach for the Extraction of Motion Patterns, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Hong Kong, **3** (2003) 165-168
4. Dosil, R., Fdez-Vidal, X. R. and Pardo, X. M.: Multiresolution Approach to Visual Pattern Partitioning of 3D Images. In Campilho, A. and Kamel, M., (eds.) LNCS 3211: Image Analysis and Recognition, Porto, Portugal (2004) 655-663
5. Studholme, C. Hill, D. L. G. and Hawkes, D. J.: An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. *Pattern Recognition*, **32** (1999) 71-86
6. Viola, P. A.: Alignment by Maximization of Mutual Information. Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Technical Report 1548 (1995)
7. Roche, A., Malandain, G. and Ayache, N.: Unifying Maximum Likelihood Approaches in Medical Image Registration. *Int. J. of Imaging Systems and Technology*, **11** (2000) 71-80
8. Roche, A., Malandain, G., Pennec, X. and Ayache, N.: The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. In LNCS 1496: MICCAI'98, Springer-Verlag (1998) 115-1124
9. Sarrut, D. and Miguet S. Similarity Measures for Image Registration. 1st European Work-

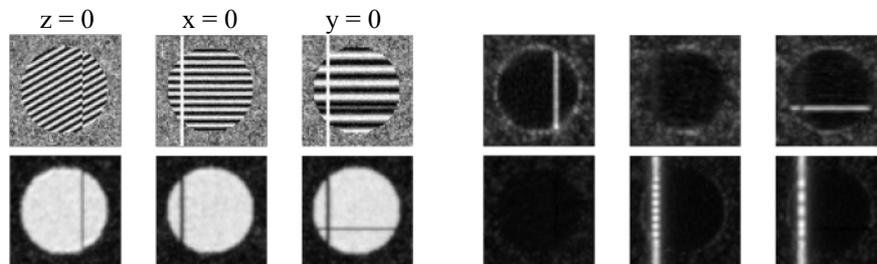


Fig. 2. *Top left:* Cross sections of the original 3D data. *Remainder:* Sections of the three patterns isolated using D_ρ

- shop on Content-Based Multimedia Indexing, Toulouse, France (1999) 263-270
10. Basseville, M.: Information: Entropies, divergences et moyennes. Technical Report 1020, IRISA, 35042 Rennes Cedex France (1996)
11. Pal, N.R., Biswas, J.: Cluster Validation Using graph Theoretic Concepts. Pattern Recognition, **30**(6) (1997) 847-857

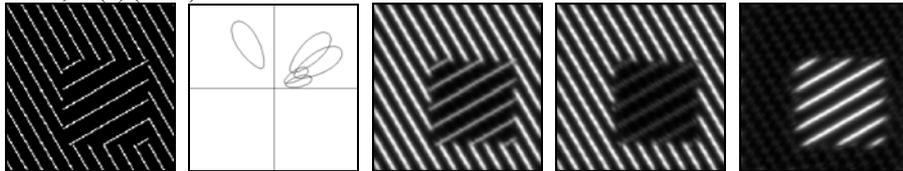


Fig. 3. From left to right: Original image. One of the clusters obtained with D_{NB}^* , represented by the $e^{-1/2}$ level curves of the filters' transfer function. Pattern associated to the previous cluster. The two patterns obtained with D_ρ

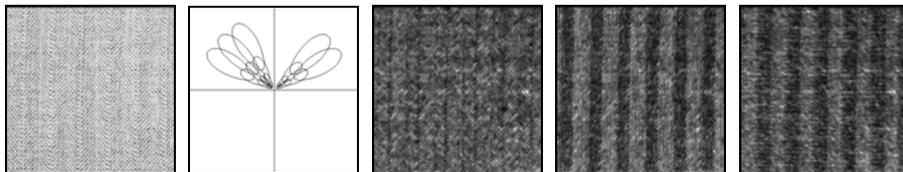


Fig. 4. From left to right: Original image. One of the clusters obtained with D_{NB} , represented by the $e^{-1/2}$ level curves of the filters' transfer function. Pattern associated to the previous cluster. The two patterns obtained with D_ρ

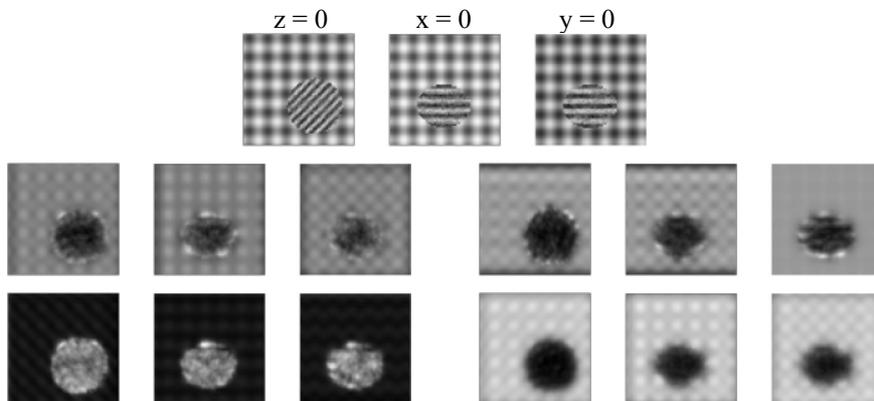


Fig. 5. Top: Cross sections of the original 3D data. Middle: Sections of the two patterns isolated using D_{NI} . Bottom: Sections of the two patterns isolated using D_ρ

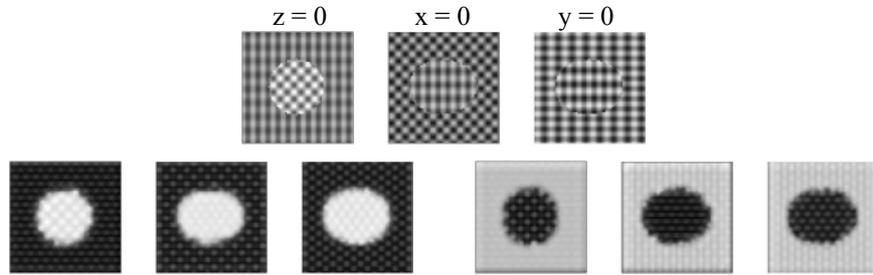


Fig. 6. *Top:* Cross sections of the original 3D data. *Bottom:* The two patterns isolated using D_p