Object Description Based on Spatial Relations between Level-Sets

Mickaël Garnier, Thomas Hurtut, Laurent Wendling

Object Description Based on Spatial Relations between Level-Sets

Mickaël Garnier, Thomas Hurtut and Laurent Wendling
Université Paris Descartes, LIPADE
45 rue des Saints-Pères, 75006 Paris, France
Email: {firstname.lastname}@parisdescartes.fr

Abstract—**Object recognition methods usually rely on either structural or statistical description. These methods aim at describing different types of information such as the outer contour, the inner structure or texture effects. Comparing two objects then comes down to averaging different data representations which may be a tricky issue. In this paper, we introduce an object descriptor based on the spatial relations that structures object content. This descriptor integrates in a single homogeneous representation both shape information and relative spatial information about the object under consideration. We use this description in the context of image retrieval and show results on a butterfly image database compared with both GFD and dense SIFT descriptors. These results show that our method is more efficient to distinguish the objects where the spatial organization is a discriminative feature.**

I. INTRODUCTION

Many computer vision applications rely on the automatic description of an object image. For instance, object recognition and image classification usually use features that endeavour to describe different types of information such as the outer contour, inner structure or texture effects. These different image informations often lead to different types of data which can be tricky to combine and may lead to inhomogeneous mixing of data [8].

A. Our Contribution

In this paper, we introduce a new image representation that capture these heterogeneous information in a single homogeneous representation. Given a decomposition of an object image into several disjoint layers of pixels, representing the different patterns presented by the object, the key idea of our method is to encode the pairwise spatial relations between all these layers. When applied to each layer with itself, the spatial self-relations encode first order shape information whereas for two different layers, the spatial relations encode relative structure and texture aspects. We show that a simple image decomposition such as the quantized level sets of the image gives interesting results, thus preventing from considering complex segmentation techniques.

B. Related Works

1) Spatial Relations: A core aspect of our method is the encoding of the pairwise spatial relations. Literature in this domain can be structured in two main categories : qualitative and quantitative approaches. Qualitative approaches use symbolic relations such as positioning relations (left, right, below,

above, etc.) and topological relations (inside, outside, etc.), see for instance [7] [6]. In this paper, we seek to capture a precise description of possibly complex objects and to characterize both large-scale and low-scale directional relations. Depending on the content meaning, the object patterns may also contained unconnected subsets of pixels. Therefore, in our context the spatial relations cannot be summarized in a symbolic manner.

Quantitative approaches gather methods that precisely describe the relative positions between two binary objects. Fuzzy quantitative methods are popular in different application domains such as spatial reasoning in medical images [2] and handwritten symbol recognition [5]. These methods produce a fuzzy landscape per considered potential direction. Combining these landscapes in order to capture the omnidirectional spatial organisation of possibly sparse object is not obvious. In this paper, we build on a quantitative model called *force histogram* [10], thereafter noted F-histograms. This model straightforwardly handles sparse objects and summarize their relative position in every directions in a single histogram. Basically, a F-histogram between two objects is a circular distribution measuring the relative attraction between these objects along every desired directions.

2) Object Recognition: Aside from spatial relations reasoning, object recognition is an attested issue with many approaches offering good results irrespective of the spatial relations of the object patterns. Among them, the Generic Fourier Descriptors [16], hereafter named GFD, measure a shape descriptor. This descriptor belongs to the MPEG-7 standard and its key idea is to compute several polar Fourier transform of the image for several angular and radial frequencies, the descriptor being the normalized histogram composed by the values of these transforms.

Another classical method is the SIFT descriptor [9] which consists in a 128-dimensional vector containing a set of gradient orientation histograms computed. The keypoints where histograms are computed are either extracted using the SIFT detection algorithm or consist in a matrix of points regularly spaced on the image (dense SIFT). Around every keypoints, a neighbourhood, divided in 4x4 smaller areas, is considered. On each areas histograms with 8 intervals are then computed. The final descriptor for each keypoint consists in a concatenation and a normalisation of the 16 histograms.

In this paper, we compare our proposed descriptor to these two generic and widely used methods. Our aim is to show
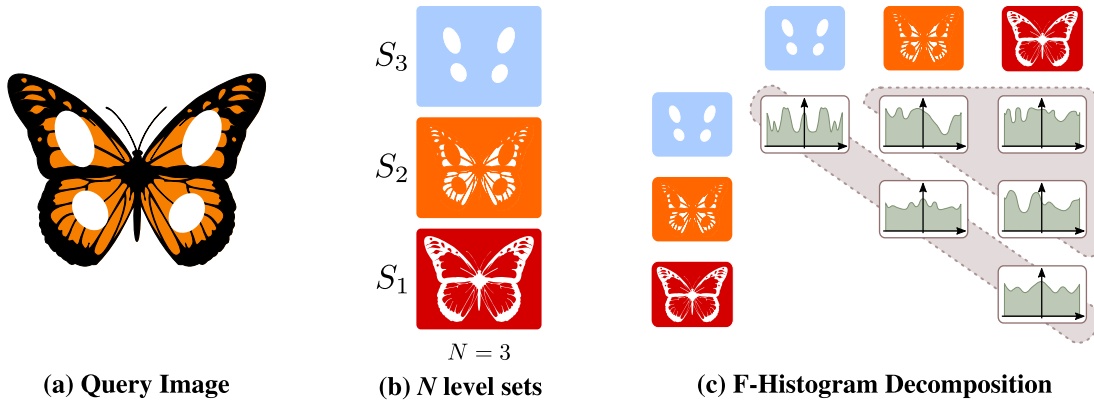
(a) Query Image   **(b) N level sets**   **(c) F-Histogram Decomposition**

Fig. 1. **Overview of the proposed descriptor**. A query image (a) is first decomposed into its N grey level sets S_i (b), $N = 3$ on this Figure. Based on the set of S_i, we then compute all F-histograms between every pairs (S_i, S_j) for all $1 \leq i \leq j \leq N$ inside a symmetric table (c). Each F-histogram in this table represents the description of the spatial relations between each considered pair of layers. Note that the diagonal part of the table represents self-relations, hence encoding shape information. The upper triangular of this symmetric structure embeds pairwise relations information between different layers.

the interest of our proposed model considering the inner spatial relations to efficiently recognize objects, compared to methods such as the GFD, that encodes exclusively shape information and as the SIFT, that models the organisation of local information.

C. Method Overview

Our goal in this paper is to demonstrate the descriptive strength of the inner pairwise spatial relations of an object content. We therefore assume that the objects we consider have been preprocessed using a background-foreground segmentation step. That is, the background pixels are not considered in our computations. As summarized in Figure 1b, our method first decomposes the object into N layers S_i of its level sets. We then build a table formed by all the F-histograms computed between every pairs $(S_i, S_j), \forall i \in \{1..N\}, \forall j \in \{i..N\}$ (see Figure 1c). Consequently, this structure contains two types of information encoded in a single homogeneous representation. First the upper triangular encodes the pairwise spatial relations between every different layers. The diagonal encodes self spatial relations, which naturally models shape information of each considered layer.

II. F-HISTOGRAM DECOMPOSITION

In this section, in order to keep the paper self-contained, we first recall the definition of a F-histogram between two binary objects. Then, we will introduce the proposed object image representation based on a table of F-histograms.

A. F-Histograms

Originally, force histograms were introduced to solve the problem of measuring the fuzzy relative direction between two objects [10]. Basically, these are circular histograms measured along the directions $\theta \in [0, 2\pi[$. Looking for the principal mode of this histogram is somehow equivalent to find the best θ that support the proposition "the first object is in direction θ from the second one".

To compute such a F-histogram, the objects are immersed in a space where an attractive force φ_r operates. The definition of this force can vary widely, depending on the feature searched inside the objects. In order to have an intuitive representation of the involved spatial relations, this attractive force φ_r is typically defined by a gravitational force based on the pairwise point distance:

$$\forall(x, y) \in \mathbb{R}^2 \times \mathbb{R}^2, \varphi_r(x, y) = \frac{1}{(d_{xy})^r} \tag{1}$$

where d_{xy} is the Euclidean distance between two points x and y.

The F-histogram value along a direction θ between two objects A and B corresponds to the global force exerted by A with regard to B in the direction θ. In other words, this force $\mathcal{F}_r^{AB}(\theta)$ is the integral sum of the infinitesimal forces $\varphi_r(a, b)$ where $(a, b) \in A \times B$ and the vector **ab** is along direction θ. Due to computational considerations, this global force is calculated on a set $\mathcal{C}_\theta$ of θ-oriented longitudinal cuts of the objects. Each of these longitudinal cut is built upon a straight line δ_θ along direction θ, and is composed of two sets c_A and c_B of possibly disjoint segments (see Figure 2):

$$c_A = \delta_\theta \cap A$$
$$c_B = \delta_\theta \cap B$$

The force exerted along δ_θ by every points of c_A with regards to every points of c_B can be written as:

$$f^\theta(\delta_\theta) = \int_{c_A} \int_{c_B} \varphi_r(a - b) \, \mathrm{d}b \, \mathrm{d}a \tag{2}$$

The F-histogram value between A and B along the direction θ is then:

$$\mathcal{F}_r^{AB}(\theta) = \sum_{\delta_\theta \in \mathcal{C}_\theta} f^\theta(\delta_\theta)$$

$$= \sum_{\mathcal{C}_\theta} \int_{\delta_\theta \cap A} \int_{\delta_\theta \cap B} \varphi_r(a-b) \, \mathrm{d}b \, \mathrm{d}a \quad (3)$$
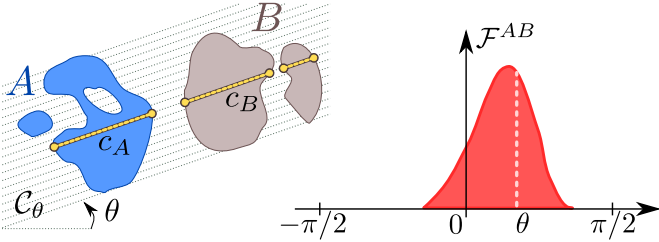

Fig. 2. The global attractive force between A and B along the direction θ is the integral sum of infinitesimal forces computed on longitudinal cuts (c_A, c_B).

According to the value given to r, the resulting F-histogram can vary widely, giving more or less importance to closer objects. For $r = 0$ the F-histogram gives the same influence to distant organized structures from any local point. In this specific case, a histogram of forces is somehow equivalent to the previously proposed *histogram of angles* [11]. Yet practically, it has been shown that histograms of angles are not isotropic, not robust to rasterization and more computationally expensive [10]. Moreover, histogram of angles do not handle overlapping objects which will appear in our situation when computing the histogram between the same layers of a decomposed object. Note that F-histograms are naturally translation invariant, symmetric and isotropic [10].

B. F-Histogram Decomposition

Unlike the classical mode searching use of a F-histogram [10], we propose to consider the whole F-histogram as a complete signature of the relative positions between two possibly sparse objects, and therefore benefit from the whole information contained in the F-histogram.

In order to capture the inner spatial relations that structure the object, we first have to *break* this object up into multiple parts. Object segmentation is still an open research issue and no generic robust algorithm exists yet. Besides, our goal is different here since we aim at cutting out the object into its subparts. Therefore instead of using a complex segmentation algorithm, we choose a more pragmatic approach by decomposing a greyscale object image Q with scalar values $Q(x) \in [0;1[$, into its N intensity level sets S_i (Figure 1b). That is:

$$\forall i \in [1;N], S_i = \left\{ x \in \mathbb{R}^2, \frac{i-1}{N} \leq Q(x) < \frac{1}{N} \right\} \quad (4)$$

This underlying assumption is also supported by the fact that mathematical morphology has shown that image contours locally coincide with level-set borders [13], [3]. By doing so, we thus adopt an oversegmentation, yet following the inner

contours and whose behaviour will not vary from an image to another.

Based on this level-sets decomposition, we then compute all F-Histograms of every pair (S_i, S_j). These F-histograms encode first-order **shape** information for each layer when $i = j$ and second-order **spatial relations** information when $i \neq j$, both in the same mathematical formalism (Figure 1c). Considering the force φ_r used during these computation, one may have a twofold strategy. First, the self $(i = j)$ F-histograms along the table diagonal are computed with a force φ_0. The natural overlapping induced by a level set with itself indeed lead to infinite forces when using $\varphi_{r \neq 0}$. Dealing with the spatial relations information contained in $(i \neq j)$ F-histograms, the φ_2 force (gravitational case) can be used. This choice is preferable in this specific case, since it models relative spatial relations, as suggested in [10].

The FHD of an object image Q is thus defined as:

$$FHD(Q) = \left\{ \mathcal{F}_0^{S_i S_i} \right\}_{\forall i \in \{1..N\}}$$
$$\cup \left\{ \mathcal{F}_2^{S_i S_j} \right\}_{\forall (i,j) \in \{1..N\}^2, j>i} \quad (5)$$

This decomposition sums up $N(N+1)/2$ F-histograms, made of N shape descriptors elements (diagonal), and $N(N-1)/2$ relative spatial relations elements (upper triangular).

Let notice that due to the invariant properties of F-histograms, FHD are naturally translation invariant and symmetric. Depending on the application requirements, one can make FHD scale invariant by normalizing the histogram surfaces by the object surface. Finally, F-Histograms being isotropic [10], rotation invariant can be pursued by estimating the principal mode of the FHD or by minimizing the distance between globally shifted FHDs.

C. FHD Matching

In order to test the FHD descriptor on recognition and retrieval tasks, a dissimilarity measure is needed. The F-histograms are first normalized by their surface in order to give an equal potential contribution to each F-histogram of the FHD. The distance between a query image Q and a target image T is then defined as:

$$\mathcal{D}(Q,T) = \alpha \times \mathcal{D}_{shape}(Q,T)$$
$$+ (1-\alpha) \times \mathcal{D}_{spatial}(Q,T) \quad (6)$$

where :

$$\mathcal{D}_{shape}(Q,T) = \frac{1}{N} \sum_{i=1}^{N} d_{\chi^2}\left(\mathcal{F}_0^{S_i S_i}(Q), \mathcal{F}_0^{S_i S_i}(T)\right) (7)$$

$$\mathcal{D}_{spatial}(Q,T) = \frac{2}{N(N-1)} \times$$
$$\sum_{i=1}^{N} \sum_{j=i+1}^{N} d_{\chi^2}\left(\mathcal{F}_2^{S_i S_j}(Q), \mathcal{F}_2^{S_i S_j}(T)\right)(8)$$

where N is the number of layers of the FHD and α is the weight level given to the shape information compared to the spatial relations information. We use a chi-square distance to

measure the distance between two single F-histograms, defined as:

$$d_{\chi^2}(a,b) = \sum_{i=1}^{i_{max}} \frac{(a(i) - b(i))^2}{a(i) + b(i)} \tag{9}$$

III. EXPERIMENTAL RESULTS

Several experimentations have been conducted using different sets of parameters for the FHD compared to two classical recognition methods, the Generic Fourier Descriptors (GFD) and the dense SIFT descriptor (dSIFT).

A. Image Database

Our experiments have been conducted on a database made from a subset of the Peale Collection [1]. This database is composed of 318 greyscale butterfly images grouped in 28 classes along the butterfly species. The typical height of an image is 640 pixels. All these images being over an homogeneous background, we first preprocessed the database by easily segmented the image backgrounds using a simple *magic wand* thresholding technique. Samples of this database are shown in Figure 3 and the subset we use is available online[1]. Butterflies are a typical case wherein inner spatial relations are a distinguishing feature making the wings patterns a direct link with the species. Another interesting application domain one can think about is botany taxonomy (flowers, mushrooms, . . .).

Fig. 3. Sample images from our database (reproduced with the permission of The Academy of Natural Sciences, Philadelphia) [1]

B. Method Settings

In order to assess the descriptive strength of our Force Histogram Decomposition (FHD), we compare them wih the Generic Fourier Descriptors (GFD) and the Dense Scale Invariant Feature Transform (dSFIT). We present in this section all the settings we use for these methods in our experiments. The cross validation is made using a leave-one-out method, that is, for every object image, the remaining of the dataset serves as the training data. This method has been favoured due to the size of our database, leaving other approaches less statistically significant.

[1]http://www.math-info.univ-paris5.fr/~mgarnier/dicta2012/

1) FHD settings: The FHD are tested with several parameter sets, evaluating the gain from multiple forces histograms and the weighting between shape information and spatial relations parts of the feature. The robustness of F-histograms to the directions quantization has been studied in [10]. In order to avoid any binning effects, all the FHD are computed along 180 directions, regularly spanning the $[0, 2\pi]$ interval with a 2 degrees step. In Equation (6), we test α values in $\{0, 0.2, 0.4, 0.5, 0.6, 0.8, 1\}$. Let notice that for $\alpha = 0$ the shape descriptor is suppressed to promote the spatial relations and vice versa for $\alpha = 1$.

2) GFD Settings: The GFD are based on the Polar Fourier defined as:

$$PF(\rho, \theta) = \sum_x \sum_y I(x,y) \times e^{\left[2j\pi\left(\frac{r(x,y)}{R}\rho + v(x,y)\theta\right)\right]} \tag{10}$$

where $r(x,y)$ and $v(x,y)$ are respectively the radius and angle of the polar coordinates of the point (x,y), I is the intensity function and the parameters ρ and θ are bounded: $0 \le \rho < R$ and $0 \le \theta < T$ with R and T respectively the the radial and angular resolutions. Finally, the GFD is written:

$$GFD(m,n) =$$

$$\left\{ \frac{|PF(0,0)|}{M_{11}}, \frac{|PF(0,1)|}{|PF(0,0)|}, ..., \frac{|PF(m,n)|}{|PF(0,0)|} \right\} \tag{11}$$

where m and n are the radial and angular frequencies and M_{11} is the order 1 moment. In our experiments, the GFD are computed on the object images with $\rho = 4$ and $\theta = 9$, thus giving a signature of 37 bins, as suggested in [15].

3) dSIFT settings: The dSIFT are extracted with a step of 16 pixels and at several scales, 4 and 8 giving both local information and a more global one. The recognition is then made using a classical pair-wise image matching. For every points of the query image, the matching algorithm searches for the best matching point in the target image, if the resulting match gives a good contrast, this point vote goes to the target image. The finally matched image is the one with the higher votes. Although this pairwise image matching protocol is time consuming, it has been chosen due to its similarity with the matching used for the GFD and the FHD. In our experiments, we use the VL_Feat library SIFT implementation [14].

C. Results and Discussion

Mean computational times for the processing of one image using $N = 4$ on an Intel CPU Xeon 3.0 GHz are the following. The FHD computation is approximately 2.3 seconds using a C programming implementation. Querying on the butterflies database using an unoptimised Matlab implementation takes around 5.2 seconds. Let notice that the overall complexity is $\mathcal{O}(N^2)$.

1) Recognition Rates: The recognition rates of the FHD are shown in Table I. The Table II shows the recognitions rates for thebest set of parameters $\{\alpha = 0.8, N = 4\}$, as opposed to the GFD and the dSIFT performance rates.

$N\backslash\alpha$	0	0.2	0.4	0.5	0.6	0.8	1
1	3.46	29.9	29.9	29.9	29.9	29.9	29.9
2	12.6	32.7	36.2	37.4	38.1	39.6	41.5
4	45.9	53.5	53.5	53.1	53.8	**56.3**	44.3
8	49.7	52.2	52.8	53.1	52.2	49.7	41.8
16	41.8	45.3	46.9	47.5	48.4	51.9	47.2

descriptor	FHD	GFD	dSIFT
recognition %	56.3	28.6	43.4

Fig. 4. Precision-recall curves over the whole database using the best sets of parameters for the FHD, the GFD and the dSIFT descriptors.

The rates in Table I approximately double from $N = 2$ to $N \leq 4$ which show the interest of encoding inner layers information. $N = 1$ is indeed equivalent to only considering the whole object shape and thus gives results close to the GFD, see Table II. A second observation is that the values' increase tend to slow while N increases and start to decrease for $N \leq 8$. This fact is first related to our database. Butterflies present homogeneous patterns with a rather limited range of luminance: black, white and one to three grey values, pleading to adopt an optimal value of $N = 4$ or $N = 8$. Secondly, the χ^2 distance compare F-histograms from the same combination of indices (i, j) for both Q and T objects. This histogram-to-histogram distance is thus neither illumination invariant nor contrast invariant. More sophisticated distances such as inter-histograms distances should be investigated in order to tackle this issue.

The columns $\alpha = 0$ and $\alpha = 1$ in Table I show respectively the results with only the spatial relations descriptor and with only the shape description of every layer. These two sets of results point out the interest of extracting the information from both shape and spatial relations.

2) Precision-Recall Tests: Since we used a classified database, for each query, we can also compute the precision-recall curve, classically defined as follows. Consider a query image belonging to a class of size $\mathcal{C}$. For a given number W of images returned by this query, W_{tp} is defined as the number of returned images belonging to the same class as the query (*true positives*). The precision-recall curve is then obtained by plotting the ratio $P = W_{tp}/W$ (*precision rate*) as a function of $R = W_{tp}/C$ (*recall rate*) [4]. Averaged precision-recall curves over the whole database using the best parameters of the different compared methods are shown in Figure 4.

3) Qualitative Retrieval Results: Several comparative retrieval results are shown in Figure 5. These results show that the FHD are more efficient to distinguish the butterflies where the spatial organization is a discriminative feature, compared to the dSIFT and the GFD.

The GFD focuses on the global shape and is thus not able to correctly discriminate two butterflies having the same global contour, yet being from different species due to different wing patterns (see for instance second column, third row of Figure 5). On the first query of the same Figure, and contrary to the two other methods, the FHD manages to capture the spatial organisation of the sparse black stripes.
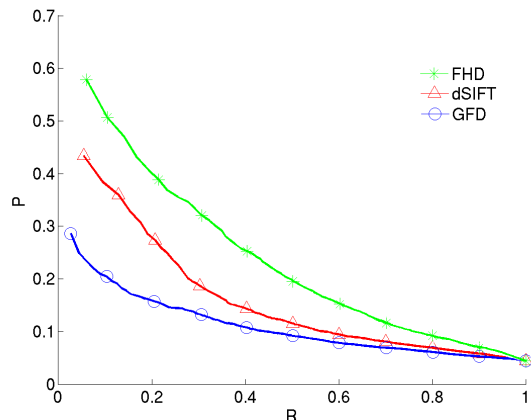
The dSIFT are actually not meant to be used with a classical matching but rather with a learning and a bag of word for instance. In this experiments, numerous matches are made between keypoints in homogeneous areas with very low distances overwhelming relevant matches thus giving descriptors with a very low variance all over the database hence some of the poor visual results. Despite this, the dense SIFT have been prefered to the classical SIFT keypoints detection. Since our database present objects after a background segmentation, most of the keypoints using classical SIFT remain located on the contrasted outer border. Such an approach lead to results similar to the GFD.

4) FHD Noise Robustness: The robustness of our descriptor is also evaluated on the same image database altered with noise. To do this, we choose the set of parameters giving the best recognition results and apply them with the same protocol but on a noised version of the database. We used two different kind of noise, that are speckle and Gaussian noise, with increasing variances, as shown in Figure 6. Speckle noise tests assess the robustness of the F-histograms to possibly low quality parsing of the level sets. It also shows that the FHD are more sensitive to Gaussian noise, as the recognition rates drops along with the variance increase.

IV. CONCLUSION

We proposed in this paper a feature based on spatial relations for grayscale object recognition. Based on a homogeneous stack of F-histograms, it naturally embeds both absolute and relative spatial informations about the considered object. It thus encodes information both on outer and inner contours and information on the spatial relations that organize the underlying grey levels of the object. We showed that a simple level-sets quantization is sufficient to capture enough information to discriminate highly structured images.

Fig. 5. Eight retrieval results on a 318 butterflies public database, obtained with the different compared descriptors: the proposed F-Histogram Decomposition (FHD), the Generic Fourier Descriptor (GFD) and the dense SIFT descriptor (dSIFT). The FHD well succeeds to capture both the inner relative spatial organisation of the grey levels and absolute shape of the different patterns composing each butterfly.

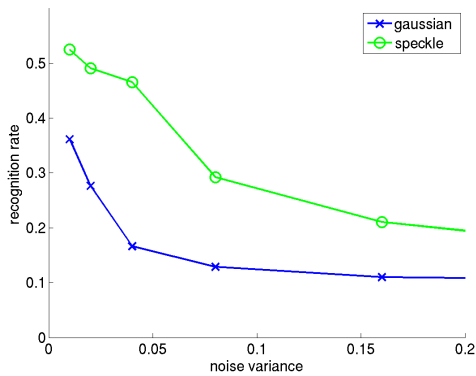Fig. 6. Recognition rates on the butterfly database altered with either Gaussian or speckle noise.

The limitations of our approach are twofold. First we presently do not encore the color content. Since this information might be of prime importance in several applications, this future work is one of our very next goals. By using bin-to-bin distances, another aspect that has been eluded is the circularity of F-histograms. Circular distances such as CEMD [12] could be investigated here although normalization constraints raise several issues.

REFERENCES

[1] Philadelphia Academy of Natural Sciences. http://clade.ansp.org/entomology/, 2012.
[2] I. Bloch. Fuzzy spatial relationships for image processing and interpretation : a review. *Image and Vision Computing*, 23(2):89–110, 2005.
[3] F. Cao, J.L. Lisani, J.-M. Morel, P. Musé, and F. Sur. *A theory of shape identification*. Lecture Notes in Mathematics. Springer, 2008.
[4] A. del Bimbo. *Visual information retrieval*. Morgan Kufman Publishers, 1999.
[5] A. Delaye and E. Anquetil. Fuzzy relative positioning templates for symbol recognition. In *Proc. ICDAR*, 2011.
[6] M. J. Egenhafer and J. Herring. Categorizing binary topological relations between regions, lines, and points in geographic databases. Technical report, University of Maine, 1991.
[7] J. Freeman. The modelling of spatial relations. *Computer Graphics and Image Processing*, 4(2):156–171, 1975.
[8] J.M. Jolion. Feature similarity. In *Principles of visual information retrieval*, pages 122–162. Springer Verlag, 2001.
[9] David G. Lowe. Object recognition from local scale-invariant features. In *Proc. IEEE ICCV*, pages 1150–1157, 1999.
[10] P. Matsakis and L. Wendling. A new way to represent the relative position between areal objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(7):634–643, 1999.
[11] K. Miyajima and A. Ralescu. Spatial organization in 2D images. In *Proc. IEEE Int. Conf. on Fuzzy Systems*, pages 100–105, 1994.
[12] J. Rabin, J. Delon, and Y. Gousseau. Circular Earth Mover's Distance for the comparison of local features. In *Proc. IEEE ICPR*, pages 1–5, 2008.
[13] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
[14] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.
[15] D. Zhang. Image retrieval based on shape. PhD Thesis, Monash University, 2002.
[16] D. Zhang and G. Lu. Shape based image retrieval using generic fourier descriptors. *Signal Processing: Image Communication*, 17(10):825–848, 2002.