Pre-Attentive Vision: Combining Gestalt Laws with the Notion of Belief

N.Zlatoff B.Tellez A.Baskurt
LIRIS
Villeurbanne, FRANCE, 69622

Abstract

Pre-attentive vision aims at organizing raw data from images in more meaningful groups, with domain-independent criteria. We present a new framework for this process, at several levels of granularity. We consider a perceptual grouping occurs when multiple local evidence appear on its support. We rely on region-based segmented image, compute evidence from geometrical measurements, and combine evidence with Shafer's degree of belief. Several results are shown, both on artificial and real images, which show the relevancy of our method.

1. Introduction

Computer vision systems aim at extracting a symbolic description of images from signal-based raw data. To this end, they proceed by hierarchical treatments, iteratively handling more and more abstract objects. However, both the sensorial and the semantic gap make this goal so hard to reach. As a matter of fact, the former states the loss of information between the real 3D world and the image, while the latter refers to the non-obvious way from set of image descriptors to symbolic description.

Two main categories of processes could be found, though strongly linked [10]. On the one hand, top-down (attentive) processes need domain knowledge, in order to perform goal-oriented tasks. Such knowledge is of various forms and could be explicitly formalized, or directly integrated in control procedures. On the other hand, bottom-up (pre-attentive) processes, organize raw data into more meaningful groups with varied criteria, independent on domain. Hence, so-called intermediate objects are created, which are a first abstraction of signal-based data. These treatments allow to strongly reduce noise issued from previous steps (segmentation). Besides, they permit a partial reduction of sensorial gap by recovering primitives issued from a common underlying real object which were disconnected (for example by occlusions). Finally, pre-attentive processes also allow to reduce further computational complexity, for instance when matching extracted groups, instead of all the primitives, with objects from a database. In this article, we introduce a new framework for pre-attentive processes, which relies on a local interaction model of several laws. Besides, we use region- or contour-based primitives whenever they are the most appropriate.

2. Previous Work

Pre-attentive processes are inspired, to some extent, explicitly or not, from Gestalt theory introduced in the psychology field by Wertheimer, Kohler and Koffka [5]. These state that during perception, the several stimuli, acquire different new properties, depending on the whole in which they happen to be. Hence, visual perception proceeds by successive groupings, based on five laws: proximity, similarity, closure, symmetry and continuity (figure 1). There also exists one meta-law, that is the tendency of a process to realize the most regular, ordered and stable state possible.

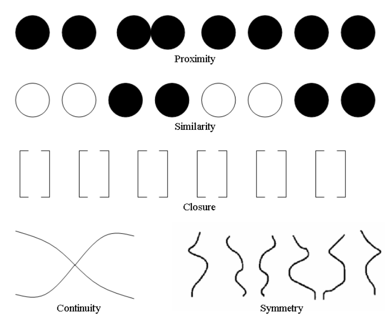

Figure 1: Gestalt laws of grouping

A complete survey of perceptual grouping techniques is presented by Sarkar [9] which classifies them according to two criteria: the considered dimension (2D, 2D1/2, 2D+time, 3D+time) and the kind of features used for grouping (signal, primitives, structural, assembly levels). Most of works deal with primitive and structural levels in 2D images. This means they try to group closed or single contours, curves, regions, polygons or corners. However, contour-based treatments are far more widespread compared with region-based ones [12]. Such a situation might be the consequence of the fact that almost all well-known examples of Gestalt laws deal with dots or lines. Nevertheless, as previously mentioned [5], the two kinds of primitives are fully adapted to Gestalt laws. More precisely, some are more adapted for some kinds of laws. For instance, the law of similarity is quite difficult to handle with contours. Region-based treatments also allow the definition of a local context in which to search for grouping hypotheses. Without it, the neighborhood of a primitive is quite difficult to define, hence leading to heuristics [1].

The SCERPO system [6] may be the first one which made intensive use of some Gestalt laws in order to organize previously extracted contours in so-called perceptual groupings. Their saliency are evaluated as the inverse of their prior probability of apparition in the image. Thus, perceptual groupings are considered to be useful since they are unlikely to have arisen by accident of viewpoint or position and therefore reflect a meaningful structure in the scene. This is known as the principle of non-accidentalness. In this view, only laws of proximity, colinearity and symmetry are considered, since all the other could lead to the organization of real-accidental features. Nevertheless, it seems to us a restriction to use only these laws at this stage. In the same view, Desolneux [3] introduces a systematic way, based on Helmholtz principles, for calculating groupings' saliency, but considering one law at a time.

Sarkar [10] and Ackermann [1][12] introduce hierarchies of treatment, so as to use several laws. Several groupings are performed from raw data to assembly level, each of them handling tokens issued from previous step, according to one law at a time. Global control can be super-imposed with graph-based heuristics [10] or with Markov Random Fields [1] for instance. However, as previously noted by Desolneux [3], Gestalt laws have complex interactions among them, so that it seems impossible to decide in which order they may be used. Besides, laws often cooperate in favor of a grouping.

That is why other works model global interaction between laws. Hence, Murino [7] uses a graph with straight segments as nodes and geometrical relationships as edges. An energy function is computed, based on geometrical relationships between segments, and then minimized, in order to find the most stable state of this Markov Random Fields' model. Geometrical constraints used are colinearity, symmetry and junctions. Urago [13] relies on the same approach. One can also mentions Kang [4] for his iterative use of fuzzy logics in order to decide whether several geometric segment-based objects should be grouped or not. However, one law is used at a time as a unique relation between two segments, even if interaction is modeled from a more global point of view.

In order to handle more complex local interactions between several laws, Vasseur [14] exhibits a belief-computing system for grouping contours, based on Dempster-Shafer theory. Nevertheless, model used leads to conflicting hypotheses which are then impossible to integrate. Besides each belief value is computed in an empirical way, with fixed threshold, making the whole process difficult to control.

3. System Overview

We present a pre-attentive grouping module, which relies on an efficient interaction model between Gestalt laws. More precisely, we consider a perceptual grouping occurs when multiple local evidence appear on support of it. Each Gestalt law is used as a potential local evidence, and thus as a potential grouping activator. We use the Dempster-Shafer theory [11] in order to model the degree of belief commited in one grouping hypothesis under the influence of each Gestalt law. Then, it is possible to derive a combined belief for this grouping hypothesis, considering all Gestalt laws, by computing the orthogonal sum of belief functions. Hence, several laws are at work in order to trigger one local grouping.

Beliefs commited in one hypothesis are statistically computed, given the set of all measurements in the image, hence avoiding the use of static thresholds. We use as a primitive for grouping a region-based segmented image on which we extract a Region Adjacency Graph (RAG). It allows us to use both region or contour clues whenever it is the most adapted for a given law. In RAG, nodes represent segmented regions while an edge is instantiated between two nodes when corresponding regions are adjacent. Thus, each edge represents a grouping hypothesis. Then, the process involves four steps:

1. The local measurements of Gestalt properties for each edge of the graph

2. The normalization of these values, leading to one belief function for each hypothesis, considering one Gestalt law.

3. The combination of the belief functions, in order to compute, for each hypothesis, a combined belief value.

4. The reduction of the graph, based on the combined belief values, and considering more global constraints.

Next section will present fundamentals of Dempster-Shafer theory, before detailing the four steps in section 5. Then, section 6 presents some results on color-based primitives.

4. Demspter-Shafer theory

Formalized by Shafer in [11], based on Dempster's work, it is a probabilistic theory of evidence, allowing a flexible modeling of uncertainty. Thus, it is well-designed when working with incomplete data. It strongly differs from other probabilistic approach in so far that it does not model for one hypothesis its probability of occurrence but rather the belief commited in its realization. More precisely, classical theories engage a $1-x$ probability in event $\bar{A}$ each time they engage x in A. However, Shafer considers that, when dealing with belief, it should be possible allowing a portion x_1 of belief in A while another portion x_2 in $\bar{A}$ with $x_1+x_2 \leq 1$. Besides, Shafer's work strongly focuses on combination of distinct bodies of evidence, which is especially well-suited for our needs.

4.1. Basic definitions and notations

Let Θ be a finite set of mutually exclusive hypotheses $\{H_1, H_2, , H_n\}$, called frame of discernment. The set of all subsets of Θ is denoted 2^Θ. We call belief function a function $Bel : 2^\Theta \rightarrow [0\ 1]$ which satisfies the three following conditions:

(i) $Bel(\oslash) = 0$

(ii) $Bel(\Theta) = 1$

(iii) for every integer k and every collection $A_1, A_2, , A_k$ of subset of Θ:

$$Bel(\bigcup_{i=0}^{k} A_i) \geq \sum_{\substack{I \subset \{1,...,k\} \\ I \neq \oslash}} (-1)^{card(I)+1} Bel(\bigcap_{i \in I} A_i)$$

For each subset A of Θ, $Bel(A)$ could be interpreted as one's degree of belief that the truth lies in A. When building a belief function, we distribute the whole belief on each hypothesis H_i or more generally on a subset A of Θ when uncertainty prevents us from being more precise. Given two or more belief functions on the same frame of discernment, the Dempster's rule of combination allows to compute their orthogonal sum, which is still a belief function, representing the combined belief on the frame of discernment.

Whenever a portion of belief is commited in a hypothesis A, it is also commited in every hypothesis implied by A, that is, from the frame of discernment, in every subset B like $A \subset B$. Consequently, when we commit a portion of belief in A, we have to precise the proportion commited in A exclusively and in subsets of A. Practically, we define a belief function by making use of a basic probability assignment (bpa).

A function $m : 2^\Theta \rightarrow [0\ 1]$ is called basic probability assignment, if and only if: $m(\oslash) = 1$ and $\sum_{A \subset \Theta} m(A) = 1$. $m(A)$ is called probability mass of A and represents the belief one commits exactly in A. The associated belief function, Bel, is obtained by:

$$Bel(A) = \sum_{B \subset A} m(B) \tag{1}$$

A subset A of Θ is called focal element if and only if $m(A) > 0$.

4.2. Demspter's rule of combination

Let m_1 and m_2 be basic probability assignments associated with belief functions Bel_1 and Bel_2 respectively, over the same frame of discernment Θ. $A_1, A_2, , A_k$ are focal elements of Bel_1 while $B_1, B_2, , B_n$ are those of Bel_2. If $\sum_{A_i \cap B_j = \oslash} m_1(A_i) m_2(B_j) < 1$, then the function $m : 2^\Theta \rightarrow [01]$ defined by

$$\begin{cases} m(\oslash) = 0 \\ m(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \oslash} m_1(A_i) m_2(B_j)} \end{cases}$$

for all non-empty $C \subset \Theta$ is a basic probability assignment. Its associated belief function is called orthogonal sum of Bel_1 and Bel_2 and denoted $Bel_1 \oplus Bel_2$. The number $\sum_{A_i \cap B_j = \oslash} m_1(A_i) m_2(B_j)$ is called conflict measure and tends towards 1 when belief functions tend to be incompatible (which means they are defined over hypotheses sets that are incompatible).

5. Model for perceptual grouping

We rely on the Region Adjacency Graph (RAG) of a region-based segmented image. This allows defining a context of grouping for each region, considering the law of proximity. Then, for each adjacency between regions R_i and R_j, we consider the hypothesis G_{ij}: "R_i and R_j are grouped".

The scope of the process is to compute for each hypothesis G_{ij}, several belief functions Bel_k (via bpa m_k) corresponding to Gestalt laws of grouping: similarity, closure, continuity and symmetry. Then, these belief functions are combined according to Dempster's rule, in order to compute the total belief in the hypothesis G_{ij}. Finally, graph reduction is performed, handling global constraints. Basic probability assignement are derived from local measurements M_k of Gestalt properties.

One of the originalities of our approach is to consider that a perceptual grouping occurs when multiple evidence appear on support of it. Each Gestalt law is used as a potential evidence, and thus as a potential grouping activator. Consequently, for each grouping hypothesis, total belief will be distributed over two hypotheses: G_{ij} and Θ (total ignorance). No Gestalt property could be directly interpreted so that G_{ij} is denied. On the contrary, if one does not tend to support G_{ij}, this can only increase the ignorance about the grouping.

5.1. Local measurements

Proximity property is directly handled by the use of a RAG. It first relies on a 1-neighborhood for each node during the local measurements of Gestalt properties, since those are processed for each edge. Nevertheless, such a neighborhood will be further expanded during the reduction and the integration of global inconsistencies (See section 5.4).

Similarity property is thought from the point of view of the descriptors used during segmentation step. As a matter of fact, each segmented region R_i, handles a set of descriptors $d_{i,k}$ that take homogeneous values among each of its pixels. Hence, we define the similarity measure as a classical distance in Euclidean space:

$$M_1(G_{ij}) = \left(\sum_k (d_{i,k} - d_{j,k})^2 \right)^{1/2}$$

When the descriptors used are color-based, we rely on the CIE Lab color space. In this way, the Euclidean distance reflects explicitly the perceptual distance between the two sets of colors. $M_1(G_{ij})$ tends towards 0 when the sets of descriptors for the two regions tend to be identical.

According to [5], Closure property tends to favor the perception of simple, closed and regular object with smooth contours. It has often been reduced in contour-based approaches only to closed objects. We consider as a reference shape an ellipse, and we introduce:

$$M_2(G_{ij}) = \left| 1 - \frac{area(R_i + R_j)}{area(ellipse(R_i + R_j))} \right|$$

where:

- $R_i + R_j$ represents the region issued from the merging of regions R_i and R_j.
- $ellipse(R_i + R_j)$ stands for the ellipse which has the same second order moments as $R_i + R_j$.

$M_2(G_{ij})$ tends towards 0 when $R_i + R_j$ tends to shape as an ellipse with the same second order moments.

On the contrary to both previous properties, continuity and symmetry far more rely on contours of regions than on regions only. We decided to unify them in one single law, considering they both rely on the same kind of notion that is the orientation difference between primitive segments. The main difference is that continuity need two segment to be close, while symmetry can handle more distant ones. We use a polygonal approximation of regions contours, based on a recursive approximation [8]. Then, orientation θ_{s_i} is extracted for each segment s_i and a global measure is set:

$$M_3(G_{ij}) = Min_{\substack{s_i \in \sigma_i \\ s_j \in \sigma_j}} \left(\left| \theta_{s_i} - \theta_{s_j} \right| \alpha_{s_i} \alpha_{s_j} \beta_{s_i,s_j} \right)$$

where

- σ_{s_i} stands for the set of segments issued from polygonal approximation of R_i
- $\alpha_{s_i} = \frac{Max_{s_i \in \sigma_i}(l_{s_i})}{l_{s_i}}$ with l_{s_i} the length of segment s_i. α_{s_i} is used as a corrective parameter ($\alpha_{s_i} > 1$), which models s_i's relevance in region R_i. It tends to discard symmetries that involve short segments, compared to R_i's size (See figure 2(c)).
- $\beta_{s_i,s_j} = \frac{Max(l_{si}, l_{sj})}{Min(l_{si}, l_{sj})}$. β_{si} is used as a corrective parameter ($\beta_{si} > 1$) which grants that a symmetry is detected among two segments with almost the same size. It prevents the detection of symmetry like one shown in figure 2(d)).

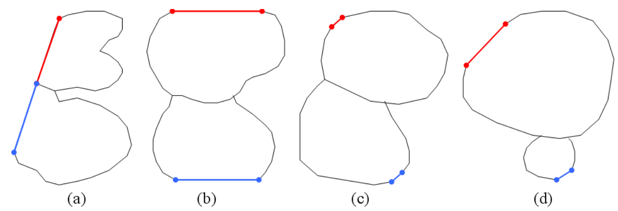

Figure 2: Examples of allowed (a,b) or discarded (c,d) symmetries

5.2. Computation of belief functions

At this stage, several measures $M_k(G_{ij})$ for k=1..3 have been extracted. These tend towards 0 when corresponding Gestalt property tends to occur in the hypothesis. We need now a process in order to compute belief functions for each measurement on each edge of the graph, over the frame of discernment $\{G_{ij}, \Theta\}$.

Considering the principle of non-accidentalness, stating that perceptual groupings occur when their prior probability of apparition decreases, we consider that the mean value $\overline{M_k(G)}$ of a measurement cannot be regarded as salient. Consequently, such a value leads to a basic probability mass of 0 for $m_k(G_{ij})$. Then, a linear regression is handled:

$$\begin{cases} m_k(G_{ij}) = \frac{\overline{M_k(G)} - M_k(G_{ij})}{\overline{M_k(G)}} \; if \; M_k(G_{ij}) < \overline{M_k(G)} \\ m_k(G_{ij}) = 0 \; else \\ m_k(\Theta) = 1 - m_k(G_{ij}) \end{cases}$$

5.3. Combination of belief functions

Belief functions (or, more precisely, basic probability assignments) are iteratively combined, according to Dempster's rule. Figure 3 illustrates the mechanism for the combination of two belief functions Bel_k and Bel_l, with m_k and m_l as corresponding basic probability assignment.

Hence, the basic probability assignment of $Bel_k \oplus Bel_l$, denoted m, is defined as follows:

$$\begin{cases} m(G_{ij}) = s_l s_k + s_l(1 - s_k) + s_k(1 - s_l) \\ m(\Theta) = (1 - s_l)(1 - s_k) \end{cases}$$

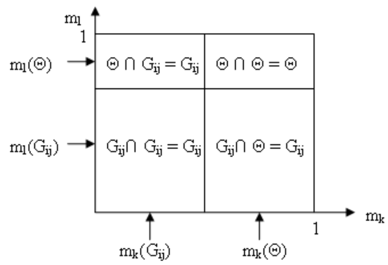

Figure 3: Dempster's rule of combination

with $s_l = m_l(G_{ij})$ and $s_k = m_k(G_{ij})$ Associated belief function can be derived with equation (1). Note that because of our frame of discernment, we have no conflict between our hypotheses. Besides, we can remark $m(G_{ij}) > s_k$ and $m(G_{ij}) > s_l$. This implies that evidence in favor of G_{ij}'s realization tend to reinforce themselves, thus leading to a system that extract perceptual grouping when multiple evidence appear on support of it.

5.4. Global constraints integration

Graph reduction consists in grouping all nodes separated by an edge whose belief in G_{ij} is more than a belief value. This one is linked to the granularity of the process: the more it will be, the less the regions will be perceptually grouped. However, local evidence is bound to create inconsistency from a more global point of view. We therefore propose a reduction process that handles two concurrent mechanisms: global constraints integration and evidence propagation. While the latter tends to activate several grouping hypotheses, the former judges them from a more global view and prevents inconsistent nodes from being grouped together. More precisely, the process consists in ranking edges thanks to their decreasing belief values and in iteratively grouping corresponding nodes of the graph. In doing so, two distinct cases appear, when considering for current nodes i and j, the configuration they take with any third node from their 1-neighborhood: either those 3 nodes form a complete sub-graph of the RAG or a planar sub-graph. Since first case could lead to global inconsistency, the lowest belief value of sub-graph's edges is transmitted to the resulting edge when merging nodes. Hence, it acts as global constraints integration. An example of such case is presented in figure 4. While nodes i and j are merged, there may be a kind of inconsistency between belief value b_{ik} and b_{jk} that respectively express the belief for G_{ik} and G_{jk}.

Second case does not handle any inconsistency. Thus, we propagate belief value to the resulting node, in order to keep this edge as a potential grouping activator (See figure 5). This is justified by the fact that local properties that have lead to the belief b_{jl} are still active between merged nodes and node l.

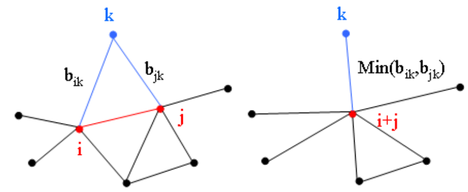

Figure 4: Global constraint integration

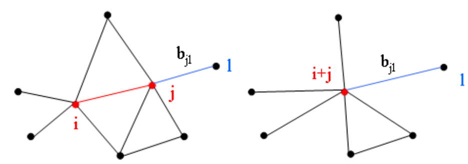

Figure 5: Evidence propagation

6. Results

We first present some results on non-natural images, in order to show the relevancy of our descriptors. Tests are then performed on color-based segmented images, with technique described in [2]. All our results involve one single extraction of Gestalt measurements over the whole image, and then several reductions. In first line of figure 6, little bright squares have color descriptors very different from black ones, compared to the enclosing rectangle. However, bright and black squares are grouped together, thanks to the continuity/symmetry law. As a matter of fact, this law finds its most significant measurement for those squares and therefore triggers corresponding grouping. If the combined belief is set to lower, grouping increases and leads to a single rectangle on the right, thanks to a combined similarity-continuity-symmetry evidence, while on the left, the enclosing rectangle is grouped with the background according to

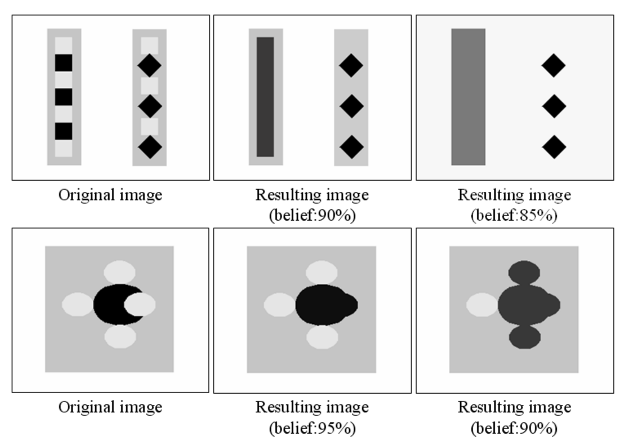

Figure 6: Basic examples

similarity law.

In second line of figure 7, bright ellipses' descriptors are very close to those of foreground, compared to black ellipse's ones. Nevertheless, the black ellipse is grouped with the bright one on the right, thanks to closure law. If the combined belief is set to lower, two other bright ellipses are grouped into the black one, according to closure parameter.

Figure 6(d) displays results on a real image, issued from the GoodShot database. Perceptual grouping allows significantly reducing the number of regions, from 250 to 76. Besides, resulting regions are perceptually relevant as they tend to make emerge semantic objects. Results on (c) are obtained with the similarity law only, and show the advantages of our combination of several laws. Note that if belief in (c) is set to lower, bright regions are grouped together, leading to the loss of the elephants' silhouettes.

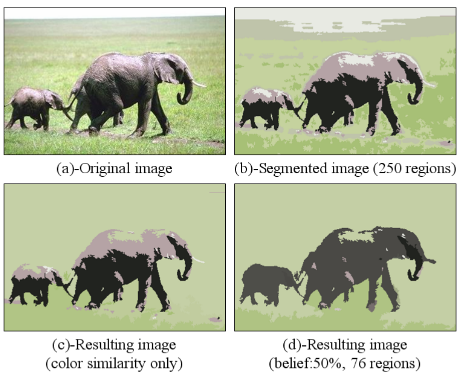

Figure 7: Example of results on real image

7. Conclusion and future work

We have presented a new framework for pre-attentive vision, in combining several Gestalt laws in a multiple evidence triggering way. Besides, our system relies on dynamic thresholds for the computing of belief functions and is thus easy to control. Several results both on artificial and real images have shown the relevancy of our model. Further work will be focused on an optimization of our system. We are currently working on an additional descriptor which would be an implementation of the Gestalt meta-law. At this time, we are investigating a curvature-based descriptor. In addition, special attention should be put on the iterative use of our tool. As a matter of fact, results shown in this paper involve one single extraction of Gestalt properties followed by several graph reductions. It would be fruitful to make several iterations of the same process.

Finally, we aim at using pre-attentive descriptions in order to extract an image's signature for indexing. In doing so, we hope handling more meaningful descriptors than those currently used in content-based indexing.

References

[1] F. Ackermann, A. Mamann, S. Posch et al. "Perceptual Grouping of Contour Segments using Markov Random Fields", *Pattern Recognition and Image Analysis*, Vol. 7(1), pp. 11-17, 1997.

[2] D. Comaniciu, P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation", In Proc. *Of Conference on Computer Vision and Pattern Recognition*, pp. 750-757, 1997.

[3] A. Desolneux, J.M. Morel, "A Grouping Principle and Four Applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25(4), pp. 508-513, 2003.

[4] H.B. Kang, E.L. Walker, "Multilevel Grouping: Combining Bottom-Up and Top-Down Reasoning for Object Recognition", In Proc. Of *International Conference on Pattern Recognition*, Vol. A, pp. 559-562, 1994.

[5] K. Koffka, *Principles of Gestalt Psychology*, Harcourt, New-York, 1935.

[6] D. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer, Boston, 1985.

[7] V. Murino, C.S. Regazzoni, G.L. Foresti, "Grouping as a Searching Process for Minimum-Energy Configurations of Labelled Random Fields", *Computer Vision and Image Understanding*, Vol. 64(1), pp. 157-174, 1996.

[8] T. Pavlidis, S.L. Horowitz, "Segmentation of Plane Curves", *IEEE Transaction on Computers*, Vol. 23(8), pp. 860-870, 1974.

[9] S. Sarkar, K.L. Boyer, "Perceptual Organization in Computer Vision: A Review and a Proposal for a Classificatory Structure", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23(2), pp. 388-399, 1993.

[10] S. Sarkar, K.L. Boyer, "A Computational Structure for Preattentive Perceptual Organization : Graphical Enumeration and Voting Method", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 24(2), pp. 246-267, 1994.

[11] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[12] D. Schluter, S. Posch, "Combining contour and region information for perceptual grouping", In Proc. Of *DAGM-Symposium, Informatik Aktuell*, pp. 393-401, Springer, 1998.

[13] S. Urago, J. Zerubia, M. Berthod, "A Markovian Model for Contour Grouping", *Pattern Recognition*, Vol. 28(5), pp. 683-693, 1995.

[14] P. Vasseur, C. Pegard, E.M. Mouaddib et al. "Perceptual Organization Approach based on Dempster-Shafer Theory", *Pattern Recognition*, Vol. 32(8), pp. 1449-1462, 1999.