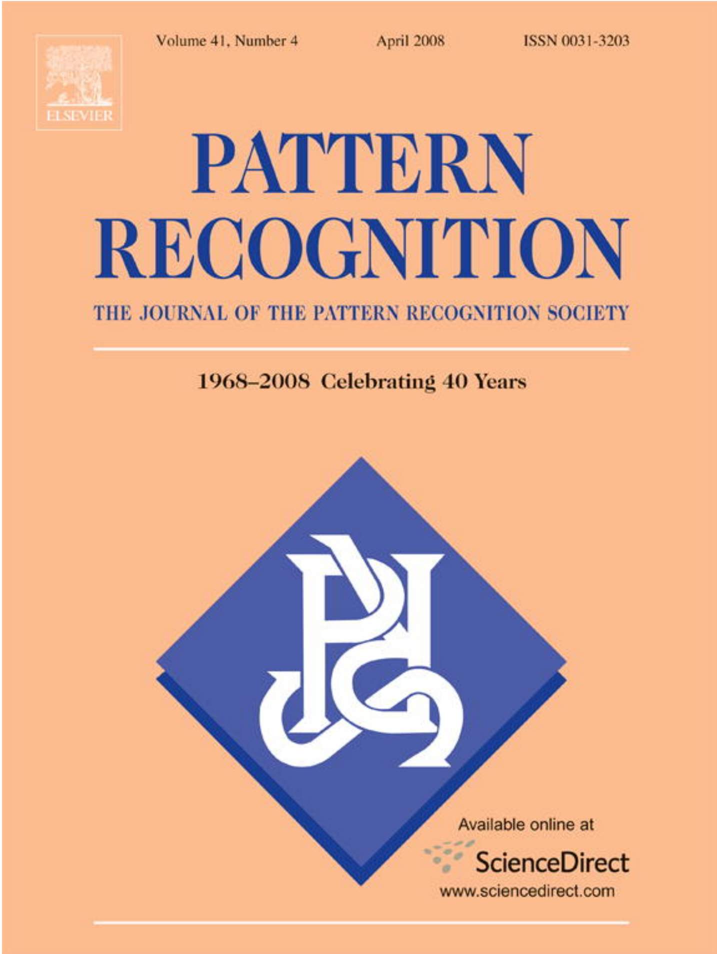
Volume 41, Number 4
April 2008
ISSN 0031-3203
ELSEVIER
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
1968–2008 Celebrating 40 Years
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect

Pattern Recognition 41 (2008) 1215–1229

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/pr

Combining local belief from low-level primitives for perceptual grouping

Nicolas Zlatoff, Bruno Tellez*, Atilla Baskurt

LIRIS, Villeurbanne, France

Received 10 November 2005; received in revised form 18 May 2007; accepted 2 July 2007

Abstract

Segmentation is usually unable to cope with artifacts due to slight change in lighting conditions or object occlusion for instance. That is why perceptual grouping is often used to overcome segmentation's lacks. This refers to the ability of human visual system to impose structure and regularity over signal-based data. Gestalt psychologists have exhibited some properties which are used during perceptual grouping, such as proximity, continuity, or symmetry. Then, some implementations of these have been proposed in computer vision. However, most of these works rely on contour-based primitives. Besides, they often use one single property to merge close regions, which may not be sufficiently robust. We propose a new framework for bottom-up perceptual grouping, which relies on a region-based segmentation. It allows us to use region or contour information, when it is the most suitable. Besides, we propose to trigger a grouping when several Gestalt properties support it. This could increase the robustness of perceptual grouping. We use Dempster–Shafer theory to combine the influence of several Gestalt properties over each grouping, as it is especially designed for this. We also present numerous promising results, which show the efficiency of our approach.

Keywords: Perceptual grouping; Segmentation; Gestalt laws; Pre-attentive vision; Dempster–Shafer

1. Introduction

Computer vision aims at extracting a symbolic description from signal-based raw data. It proceeds by a hierarchy of treatments, handling more and more abstract tokens. One important step is perceptual grouping, which refers to the ability of human visual system to impose structure and regularities over several stimuli. Hence, perceptual grouping aims at extracting salient patterns from images, which could be further handled by interpretative tasks in order to find semantic objects. Handling such patterns instead of pixels offers several advantages such as the reduction of computational complexity of further processes. Besides, it offers an intermediate level of description (shape, spatial relationships) for data, which is more suitable for object recognition tasks.

Among several approaches of perceptual grouping, the simplest step is known as segmentation. It consists in grouping pixels into various structures based on low-level descriptors such as color or texture [1–3]. However, those approaches are quite limited as they cannot extract heterogeneous patterns composed of several parts with very different low-level descriptors for each of them. As a lot of real objects are just of that kind, it is a strong limitation. Moreover, segmentation cannot cope with many artifacts due to slight changes in lighting conditions or object occlusions for instance.

That is why some other works have tried to apply other criteria for subsequent grouping. For instance, it may be relevant to group two regions with close and continuous borders, as those may be two parts of an object. Other criteria seem useful too, such as compactness, similarity or symmetry. Early in the 20th century, psychologists from Gestalt theory [4,5] have already formalized the importance of such properties for grouping. They have argued that vision proceeds by successive groupings, involving some basic properties: proximity, similarity, closure (compactness), continuity and symmetry for instance. See Fig. 1 for some examples.

Then, several works issued from computer vision have proposed implementations of those properties (among them, see Refs. [6,7]). They often rely on contour-based segmented image. Besides, they often use one single property to merge close

* Corresponding author. Tel.: +33 4 72 43 15 83.
E-mail address: bruno.tellez@liris.cnrs.fr (B. Tellez).

0031-3203/$30.00 © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.07.006

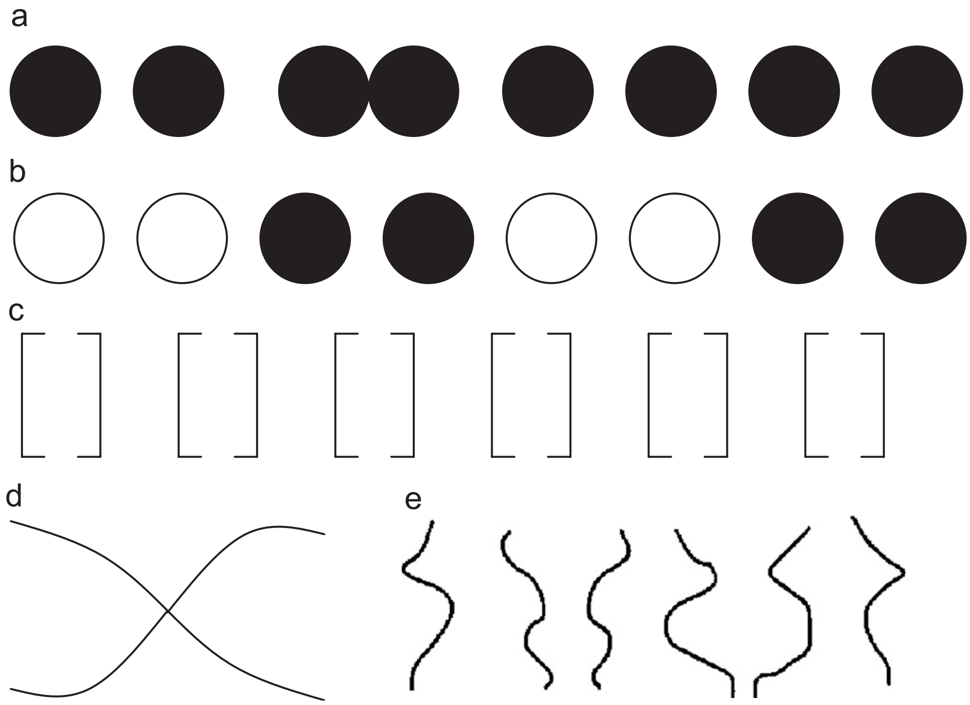

Fig. 1. Some Gestalt properties of grouping (adapted from Ref. [6]): proximity (a), similarity (b), closure (c), continuity (d), symmetry (e). Although segments in (c) are regularly distributed, they are grouped in order to form objects that tend to shape as close ones. Patterns in (d) are seen as two continuous lines intersecting rather than two contiguous curves with cusps.

regions. We propose in this article a new framework for region-based perceptual grouping, which needs the support of several Gestalt properties in order to trigger a grouping. This leads to a strong increase of robustness, as it prevents one single property alone from merging two regions.

We use Dempster–Shafer theory [8] in order to combine the influences of several properties in each hypothesis, since this formalism offers a flexible and efficient model for such combination. Indeed, we suppose that each Gestalt "law" could be modelized as an isolated mechanism dedicated to a specific feature evaluation. Combining the Gestalt laws with the Dempster–Shafer theory, we could manage important and natural laws for perception, easy and quick to evaluate and to combine, like a network of features that can be activated to extract complex objects.

The image indexing and retrieval schemes generally consider heterogeneous natural color image scenes. Our paper addresses this problematic. So we considered that a color-based segmentation will be more convenient regarding to our merging process. Furthermore, we integrate contour information to evaluate laws like continuity or symmetry. We can say that our approach is a hybrid one: region based for color, texture and compactness and contour based for continuity and symmetry.

The scope of this article will be as follows: next section will review related works on perceptual grouping. Section 2 presents the model for perceptual grouping, and basics of Dempster–Shafer formalism. After detailing Gestalt measurements in Section 3, we present several results in Section 4.

1.1. Related work

Basically, perceptual grouping could be divided into two kinds of treatments, though strongly related. On the one hand, top-down (attentive) processes need external knowledge in order to perform goal-oriented tasks. Such knowledge could be of various kinds. For instance Bayesian networks have been used by Sarkar and Boyer [9], so that to infer the presence or absence of geometric patterns from different clues (corners, ribbons, curves, etc.). External knowledge should also take the form of shape models. For instance, Sclaroff and Liu [10] group regions by using statistical shape models to inforce prior probabilities on global deformations for each class (e.g. fish, leaf, etc.). In the same way, Forsyth and Fleck [11] group skin-colored regions according to some pre-defined patterns, in order to recognize human nudes. Generally speaking, it is usually quite difficult to generalize from such systems, as they often rely on several ad hoc steps.

On the other hand, bottom-up (pre-attentive) processes handle only signal-based data, without any additional knowledge. One should keep in mind that an efficient vision system obviously needs to integrate both attentive and pre-attentive components. Even if we focus in this article on pre-attentive processes, we are currently working on further treatments, that involve structural shape models, in order to ensure verification and interpretation of patterns extracted by perceptual grouping.

Sarkar and Boyer [7] present a very extensive survey on perceptual grouping. They argue that a lot of approaches make use of Gestalt principles such as proximity, similarity, closure, continuity and symmetry. Let first note that most of these approaches focus on contour-based primitives, which seems to us a limitation. As a matter of fact, region-based structures may be more suitable to quantify some properties such as similarity (thanks to color or texture features) or closure-compactness (with shape). That is why we rely on a region-based segmented image for perceptual grouping. This allows us to use information from both region and contour, when it is the most appropriate.

Lowe [6] may be the first one who made use of some Gestalt properties in order to perform perceptual groupings. His main contribution is the formalization of the non-accidentalness principle, so as to compute the significance of a grouping: it is inversely proportional to its prior probability of occurrence. For instance, as three points are very unlikely to be aligned in a noisy image, if such an alignment occurs, it may reflect a salient structure in the image. Moisan and Desolneux [12] formalized an analogous strategy to infer groupings' saliency based on Helmholtz principle. We make also use of the principle of non-accidentalness in our framework.

Sarkar and Boyer [13] propose to go one step further by using a hierarchical framework of grouping, based on contour segmentation. Here, tokens of increasingly complexity are iteratively grouped, thanks to a static hierarchy of treatments. Similar approaches can be found [14–16] with different global control imposed on groupings such as Markov Random Fields (MRF) [14] or fuzzy logics [15]. However, those works always extract static, pre-defined structures, which is a strong limitation. Besides, they use one single Gestalt property at a time in order to characterize a grouping, although several properties may be at work [12]. Taking this information into account can lead to a strong increase in robustness, since it prevents one single property alone from triggering a grouping. We model such cooperation in our framework.

In order to handle more complex interactions, without any prior hierarchy of treatments, Murino et al. [17] model in a graph a set of contours linked together by grouping hypothesis. A MRF is then used in order to find the most stable state. However, one Gestalt property is used as a potential activator for each grouping hypothesis. Idrissi et al. [18] use a heuristic in order to reduce an analogous graph, with properties of proximity, similarity and closure at work for each hypothesis. However, they use a weighted sum to combine properties' effects and thus do not handle precisely the interaction. The same holds for works from Luo and Guo [19], who uses a MRF with a greedy algorithm for grouping.

In fact, classical Bayesian theories handle with difficulties interactions of several properties on one given hypothesis. As a matter of fact, properties are likely to contradict each other and it is therefore difficult to combine their influence. On the contrary, Dempster–Shafer theory [8] is especially well-suited for such needs, as it allows to handle belief rather than probabilities. The main difference between belief and probability relies in the fact that a portion of belief could be committed into one hypothesis without committing the remainder into the hypothesis' negation. Hence, several criteria may not contradict each other. Vasseur et al. [21] make use of Dempster–Shafer formalism for perceptual grouping on contour primitives. However, their model is just a Bayesian view of Dempster–Shafer theory and still leads to conflicting sets of hypotheses. On the contrary, our framework does model cooperation between Gestalt properties and prevents conflict between them from jamming the groupings.

1.2. Framework for perceptual grouping

In a classical scheme with three levels of treatment (low level, intermediate and high level), we focus on intermediate level to fill the semantic gap. Our system uses perceptual features to build complex objects (required in high level) from low-level regions proposed by segmentation methods at low level. We propose a cooperative framework for perceptual grouping. We rely on a region-based segmented image, and we impose several Gestalt properties to be at work in order to trigger a grouping. This model leads to a strong increase of robustness as it prevents one single property from triggering a grouping.

Grouping hypotheses are generated from a Region Adjacency Graph (RAG) where a vertex stands for a region while an edge is instantiated between two regions when those are adjacent. Hence, edges from RAG represent grouping hypotheses. Then, each Gestalt property leads to a partial belief for grouping on each hypothesis. Those beliefs are then combined together in order to assign each hypothesis with a global belief for grouping. Finally, RAG is reduced, based on those global beliefs. The result of this process is a tree of regions where each node defines a composite object (leafs stand for the regions stemming from the segmentation process). The subtree associated to a node captures the various steps of the merging process: in a perceptual point of view, all the regions that contribute to the emergence of a relevant object. Fig. 2 shows an example of an image processing architecture and the details of the perceptual grouping process we propose.

2. Model for perceptual grouping

As several Gestalt properties characterize each grouping hypothesis, we need to combine their influences in order to derive a global confidence on each hypothesis. A lot of works uses here a weighted sum. However, this way is not robust to local error: if only one property does not match, the global belief strongly increases. That is why we make use of Dempster–Shafer theory, that is specially designed for combining several points of view over one hypothesis [8]. Here, the hypothesis is the potential grouping between two adjacent regions R_i and R_j. Each Gestalt property could be seen as a different point of view over the hypothesis, while the theory allows us to derive from them a combined view point, which takes into account all the others.

The originality of our contribution is to consider various properties to group regions based on Dempster–Shafer Theory. The idea is to estimate and use the relative influence of each criterion. In our approach, each criterion has the same importance and can contribute to adapt the merging process to the image content. For a given class of images like medical images,

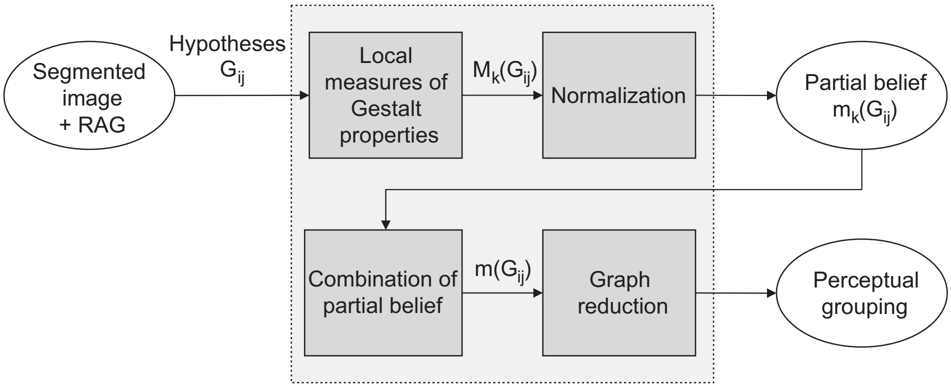

Fig. 2. Overview of an image processing architecture (left) and details of the perceptual grouping framework (right).

the gray level similarity of connected regions will be dominant. In other context like urban photos (including buildings, streets, etc.), it will be the continuity and/or the symmetry. By considering all the features with the same importance, the proposed system does not need to solve the difficult weighting problem of features combination.

Further details on Dempster–Shafer formalism are presented in the next paragraphs. We will present only concepts strictly related to our needs. As a matter of fact, a detailed presentation of Dempster–Shafer theory is far beyond the scope of this article.

2.1. Basics of Dempster–Shafer theory

Basically, Dempster–Shafer is a probabilistic theory that models beliefs over hypotheses [8].

Let Θ be a set of mutually exclusive hypotheses $\{H_1, H_2, \ldots, H_n\}$, called frame of discernment. The set of all subsets of Θ is denoted 2^{Θ}.

We call *basic probability assignment* (bpa) a function $m : 2^{\Theta} \longrightarrow [0; 1]$ satisfying:

$$m(\oslash) = 0, \tag{1}$$

$$\sum_{A \subset \Theta} m(A) = 1. \tag{2}$$

For each part A from Θ, $m(A)$ represents the belief someone exactly commits in A. Basically, bpa m could be seen as an extended probability function. When building a bpa, one must commit belief over the set of hypotheses H_i from Θ or over a subset A of Θ when uncertainty prevents us from being more precise. For instance, given $\Theta = \{H_1, \overline{H_1}\}$, a classical probability function p would say that event H_1 has a probability x of occurrence. This implies that event $\overline{H_1}$ has a probability of $1 - x$.

In Dempster–Shafer theory, one should say that even if he believes H_1 to the extent x (that is $m(H_1) = x$), he does not believe $\overline{H_1}$ to the extent $1 - x$ (that is $m(\overline{H_1}) = 0$ for instance). Therefore, in order to satisfies equation above, one has: $m(\Theta) = 1 - x$, which is called uncertainty.

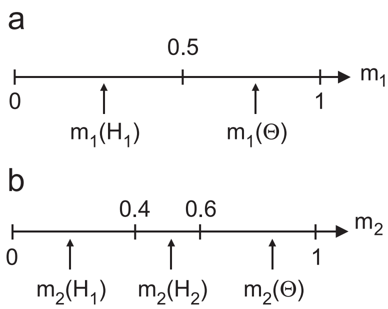

Fig. 3. Example of two bpa over the set $\Theta = \{H_1, H_2\}$. Note that m_1 commits belief over H_1 and uncertainty Θ while a classical probability function would have committed belief over H_1 and $\overline{H_1}$.

Hence, Shafer's bpa allow a finer modeling of belief over a set of hypotheses than classical probabilities do. Fig. 3 shows two examples of bpa, m_1 and m_2, over the set of two hypotheses $\Theta = \{H_1, H_2\}$.

When a bpa m commits belief into a subpart A from Θ, one has $m(A) > 0$ and A is called a focal element of m.

Note that, strictly speaking, Dempster–Shafer theory makes a difference between *exact* belief ($m(A)$) and *total* belief (via other kind of functions). However, we do not emphasize this distinction here, as it is pointless for our model.

2.2. Combining bpa

Dempster's rule of combination allows one to combine several bpa over the same set of hypotheses, in order to infer a new bpa that takes into account all the influences of the others.

For the examples from Fig. 3, the combination could be illustrated by Fig. 4. The combination leads to six sub-cases. The belief committed into each of them is represented by the corresponding area. Hence, for hypothesis H_2, one has

$$m(H_2) = m_2(H_2) m_1(\Theta). \tag{3}$$

This is formalized by Dempster's rule of combination: let m_1 and m_2 be two bpa, over the same frame of discernment Θ. Let also denote A_i the focal elements of m_1 while B_j those of m_2.

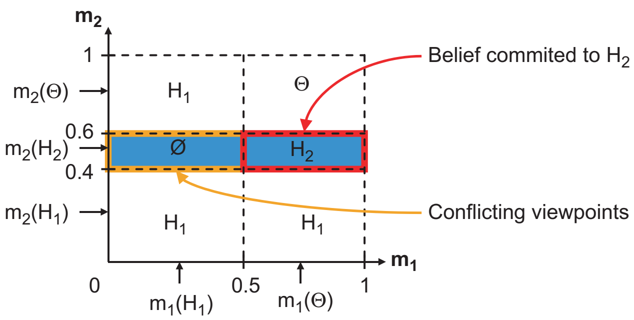

Fig. 4. Combination of the two bpa from Fig. 3. Belief committed on each subcase is represented by the corresponding area. Since all H_i are mutually exclusives, combination may lead to some subcases that correspond to no hypothesis ($\emptyset$). On the contrary, when both bpa agree on one hypothesis (H_1), combination tends to reinforce it.

We call orthogonal sum of m_1 and m_2, the bpa m defined on any subset C from Θ by

$$m(\emptyset) = 0, \quad m(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}. \tag{4}$$

The orthogonal sum of m_1 and m_2 represents the total belief one has over Θ, given both m_1 and m_2.

The number $k = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)$ is called weight of conflict. It corresponds to the case when bpa contradict each other, that is, when they commit belief over hypotheses whose intersection boils down to $\emptyset$. In this case, results from Eq. (4) are unlikely to be representative, as the conflict may represent a large portion of belief. If $k = 1$, conflict is total and the orthogonal sum of m_1 and m_2 does not exist.

Hence, for the two bpa from Fig. 3, one has, as a result of combination:

$$k = 0.6 \cdot 0.2 = 0.12, \tag{5}$$

$$m(H_1) = \frac{0.5 \cdot 0.4 + 0.5 \cdot 0.4 + 0.5 \cdot 0.4}{1 - 0.12} = 0.68, \tag{6}$$

$$m(H_2) = \frac{0.5 \cdot 0.2}{1 - 0.12} = 0.11, \tag{7}$$

$$m(\Theta) = \frac{0.5 \cdot 0.4}{1 - 0.12} = 0.23. \tag{8}$$

2.3. Handling conflicting viewpoints during perceptual grouping

For our purpose, each bpa corresponds to the influence of one Gestalt property. For each edge of the RAG, we consider $\Theta = \{G_{ij}, \overline{G_{ij}}\}$ where G_{ij} stands for the hypothesis *grouping region i and region j*. Hence, $\overline{G_{ij}}$ stands for the hypothesis *not grouping region i and region j*.

Each bpa commits belief over two hypotheses: G_{ij} and uncertainty Θ (see Fig. 5(a)).

Note that, due to the choice of our focal elements, there is no conflict among our hypotheses ($k = 0$). Hence, results from combination are always representative. This is very different from classical Bayesian theories where belief would have been committed both to G_{ij} and to $\overline{G_{ij}}$. In this case, conflicting

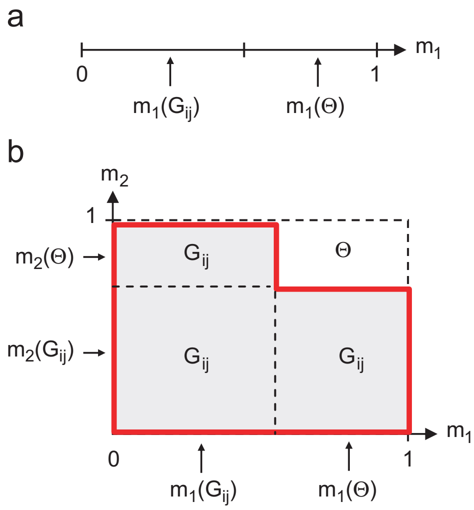

Fig. 5. Interaction model between two Gestalt properties, represented by bpa m_1 and m_2. $\Theta = \{G_{ij}, \overline{G_{ij}}\}$, where G_{ij} stands for *grouping regions i and j*. One property cannot contradict with another one, but could only increase uncertainty Θ. Besides, beliefs in one given hypothesis from different properties tend to reinforce themselves.

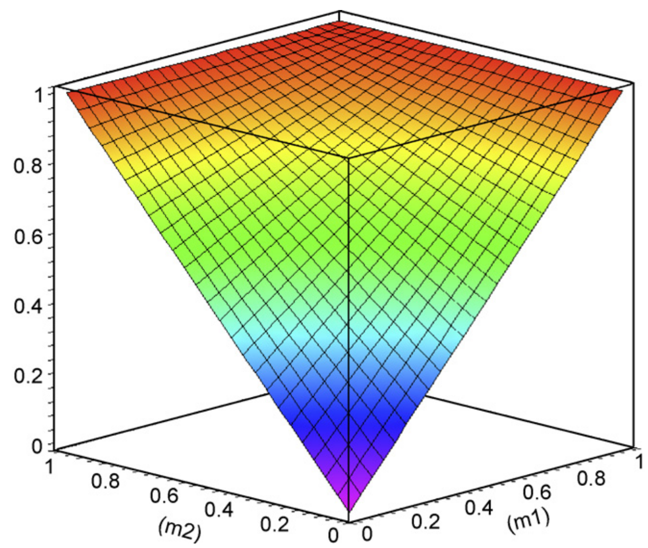

Fig. 6. Dempster's rule of combination as a function of two variables: bpa m_1 and m_2.

hypotheses may have appeared which would result in a breakdown for the grouping. The combination of two bpa of that kind is illustrated in Fig. 5(b).

In application of Dempster's rule:

$$m(G_{ij}) = m_1(G_{ij}) + m_2(G_{ij})(1 - m_1(G_{ij})). \tag{9}$$

Fig. 6 represents $m(G_{ij})$ as a function of the two variables $m_1(G_{ij})$ and $m_2(G_{ij})$.

Note that

$$m(G_{ij}) > m_1(G_{ij}) \quad \text{and} \quad m(G_{ij}) > m_2(G_{ij}). \tag{10}$$

That means that beliefs from several bpa over one grouping hypothesis tend to reinforce themselves. Hence, when different

Gestalt properties are activated from one grouping hypothesis they act in favor of the grouping in a cooperative way. Numerical examples of several combinations are shown in Fig. 9 from section 'Results'.

As our purpose is to combine more than two belief functions, note that those results can be generalized for a set of n belief functions by iteratively handle Eq. (4).

3. Gestalt measurements

This section will explain how the several grouping hypotheses are characterized, considering each Gestalt property. There is no real consensus, even among Gestalt psychologists [4], on how many properties are involved during perceptual grouping, and on how they should be implemented. Examples of properties are shown in Fig. 1. For instance, one tends to group circles from (b) according to a color-based similarity. In the same way, lines from (c) tend to be grouped in order to form close shapes, even if they are equally spaced. Note also the continuity property which states that patterns from (d) are viewed as two continuous lines intersecting rather than two adjacent cups.

We propose to characterize a grouping hypothesis with four different properties, which seem to be quite relevant: proximity, similarity, closure (compactness) and continuity/symmetry. Note that property of proximity is directly handled by the use of a RAG during hypothesis extraction.

Note also that the implementation of these properties is driven by computing-time considerations: we are looking for features that are fast to compute as perceptual grouping may latter takes place in indexing systems.

3.1. Similarity

Similarity 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.

Two descriptors are used for similarity feature evaluation: color and texture. 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.

Hence, we define the color similarity measure $M_1(G_{ij})$ in

$$M_1(G_{ij}) = \left(\sum_k (d_{i,k} - d_{j,k})^2\right)^{1/2}. \tag{11}$$

For texture, we use Gabor texture features [20] for analyzing regions. Four orientations are considered for Gabor filters. The region is sampled and the values of the four filters are computed for each zone. The resulting features for the region are defined as the average values for each filter.

As in Eq. (11), we use an euclidian distance for the texture similarity measure $M_2(G_{ij})$. $M_1(G_{ij})$ and $M_2(G_{ij})$ tend towards 0 when the sets of descriptors for the two regions tend to be identical.

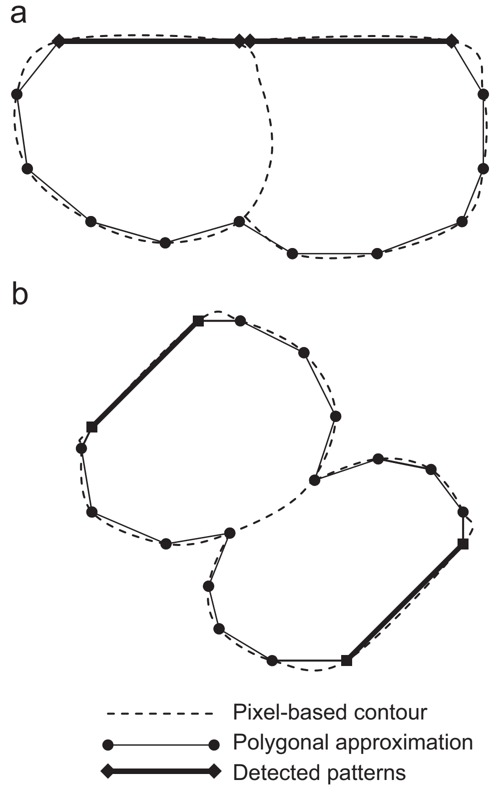

Fig. 7. Example of allowed pattern for continuity/symmetry properties. (a) Continuity-like pattern. (b) Symmetry-like pattern.

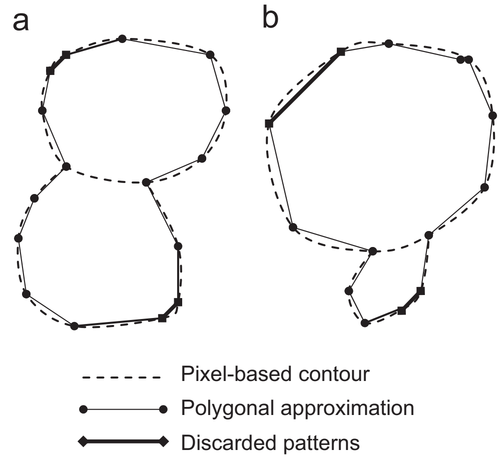

Fig. 8. Example of discarded pattern for continuity/symmetry properties. (a) Discarded segments are not representative of their region, (b) while discarded segments are representative of their region, they do not have a similar size.

3.2. Closure-compactness

According to Ref. [4], closure property tends to favor the perception of simple, closed and regular object. It has often been reduced in contour-based approaches only to closed objects, but it also refers to compactness. That is why we introduce:

$$M_2(G_{ij}) = \left|1 - \frac{area(R_i + R_j)}{area(ellipse(R_i + R_j))}\right|, \tag{12}$$

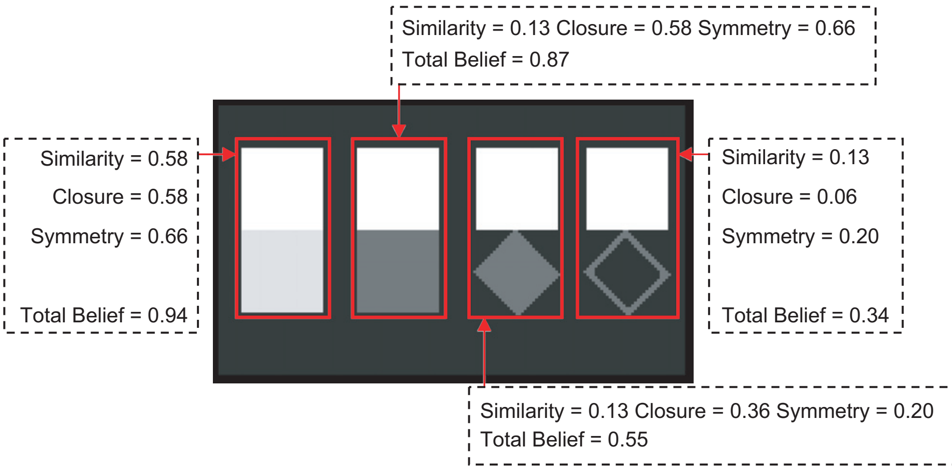

Fig. 9. Example of Dempster's combinations. Grouping hypotheses are surrounded in red.

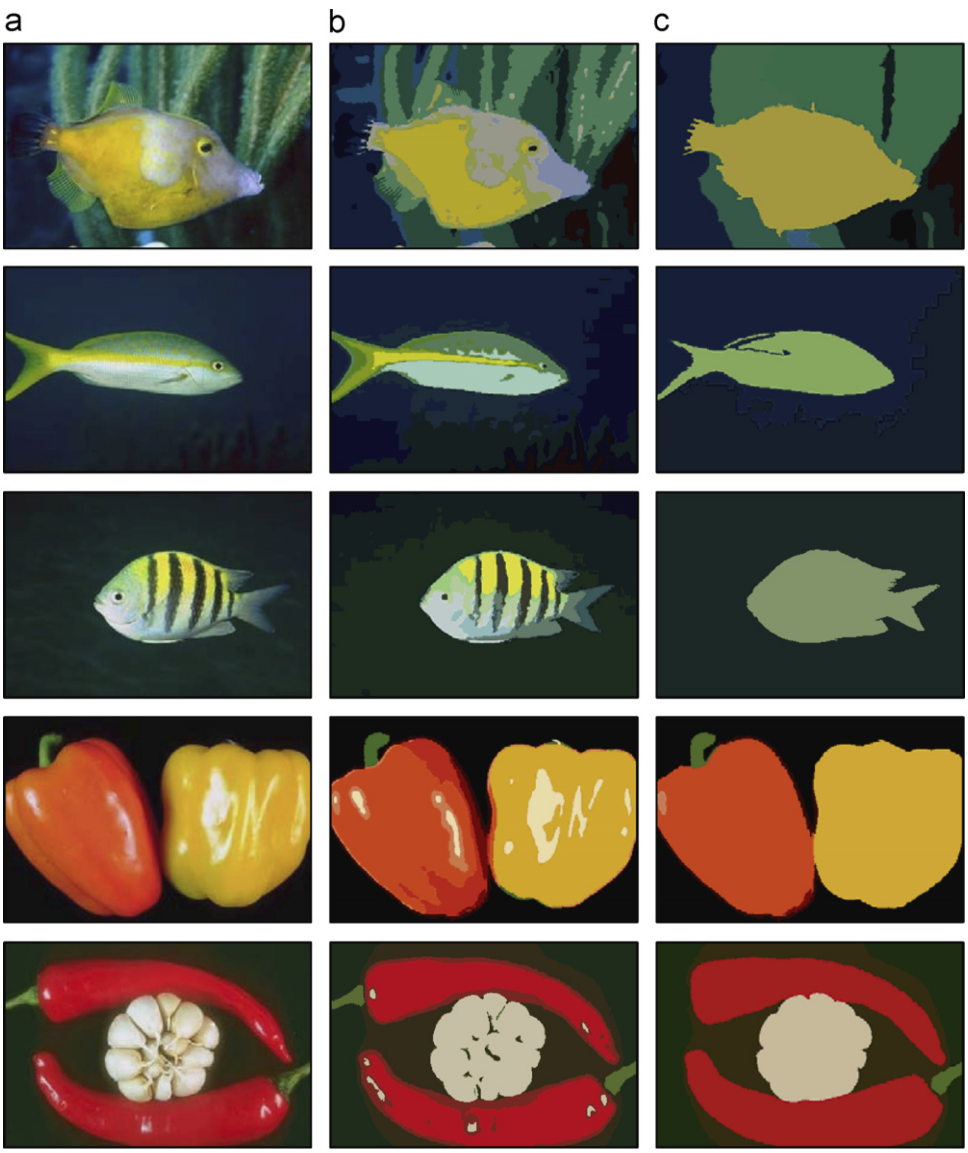

Fig. 10. Examples of results of our perceptual grouping (c) from segmented images (b). Original images are shown in (a). Parameter minBelief is set, respectively, to 50%, 40%, 35%, 62% and 55%.

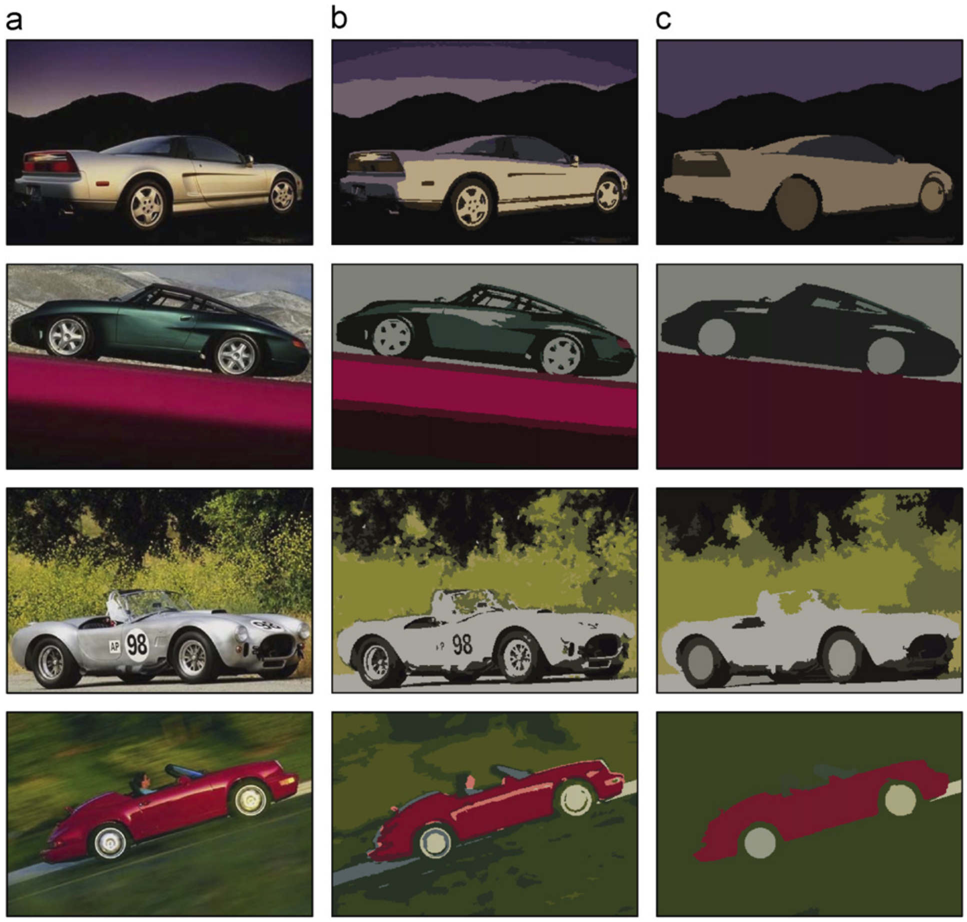

Fig. 11. Examples of results of our perceptual grouping (c) from segmented images (b). Original images are shown in (a). Parameter minBelief is set, respectively, to 43%, 68%, 70% and 62%.

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 $R_i + R_j$. We make use of an ellipse as it allows us to approximate a convex hull very quickly considering the computing time.

$M_2(G_{ij})$ tends towards 0 when $R_i + R_j$ tends to shape as an ellipse with the same second order moments. It therefore favors both groupings which shape like ellipses, and that involve one region strongly bounded by another one.

3.3. Continuity-symmetry

On the contrary to previous properties, continuity and symmetry far more rely on contours of regions than on regions only. We unify them into one single property, considering they both rely on the same kind of notion that is the orientation difference between primitive segments. The main difference is that continuity needs two segments to be close, while symmetry can handle more distant ones.

We use a polygonal approximation of regions contours, based on a recursive approximation [22]. Then, orientation θ_{s_m} is extracted for each segment s_m and a global measure is set:

$$M_3(G_{ij}) = \min_{(s_m,s_n)\in(\mathcal{S}_i\times\mathcal{S}_j)} (|\theta_{s_m} - \theta_{s_n}|\alpha_{s_m}\alpha_{s_n}\beta_{s_m s_n}), \tag{13}$$

where $\mathcal{S}_i$ stands for the set of segments issued from polygonal approximation of R_i.

α_{s_m} and $\beta_{s_m s_n}$ are two corrective parameters ($\alpha_{s_m} > 1$, $\beta_{s_m s_n} > 1$), which prevent the detection of pattern like in Fig. 8(d) and (c), respectively.

$$\alpha_{s_m} = \frac{\max_{s_k\in\mathcal{S}_i}(l_{s_k})}{l_{s_m}}, \quad \beta_{s_m s_n} = \frac{\max(l_{s_m}, l_{s_n})}{\min(l_{s_m}, l_{s_n})} \tag{14}$$

with l_{s_m} the length of segment s_m. Note that α_{s_n} is the analogous term of α_{s_m} for region R_j.

Examples of detected patterns for this continuity/symmetry property are shown in Fig. 7(a) and (b).

3.4. Normalization of measurements

In order to normalize raw measurements $M_k(G_{ij})$ into bpa m_k, we use the principle of non-accidentalness [6]. As explained before, it formulates that the significance of a grouping is proportional to the inverse of its prior probability of appearance. For instance, since three points are unlikely to be aligned in an image, if such an alignment is found, it represents a grouping of high significance. Following that idea, we argue that mean value $\overline{M_k}$ of a measure $M_k(G_{ij})$ cannot be regarded as salient. Considering also that raw measurements $M_k(G_{ij})$ tend

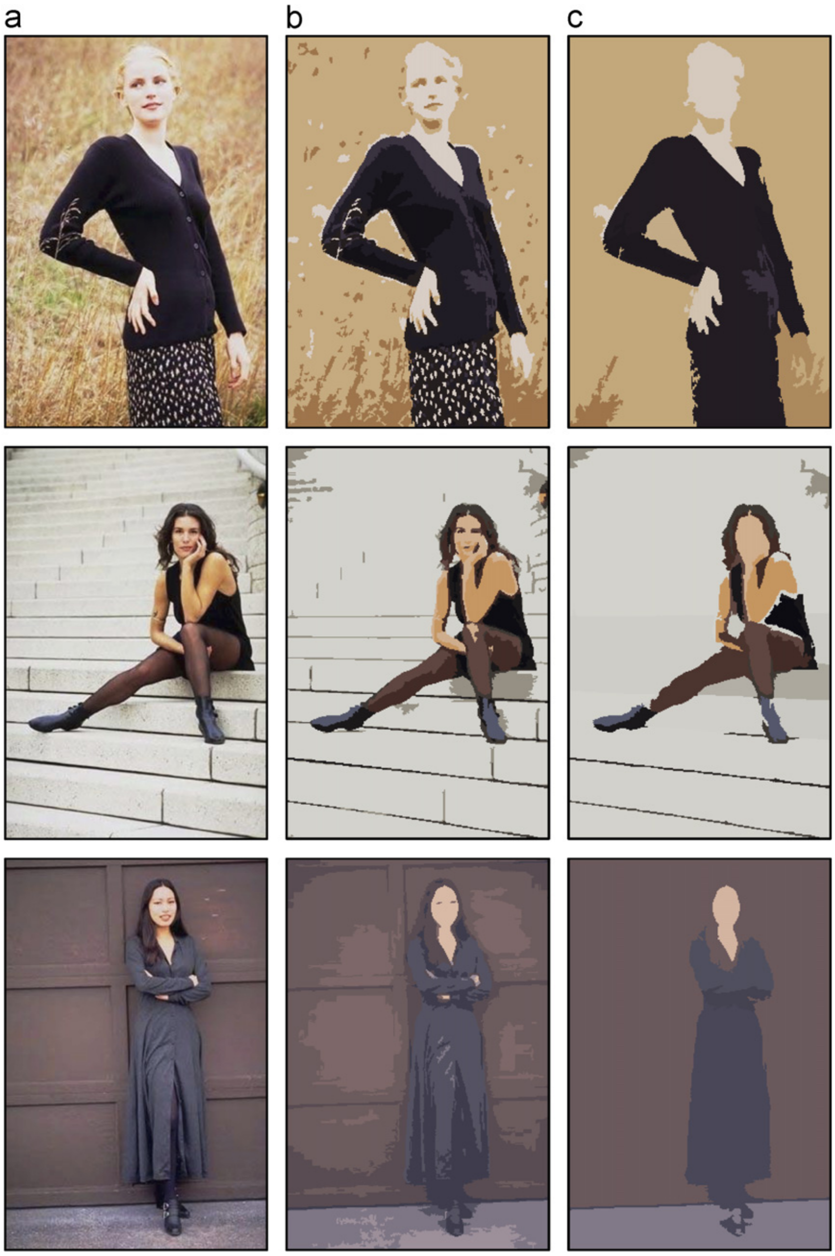

Fig. 12. Examples of results of our perceptual grouping (c) from segmented images (b). Original images are shown in (a). Parameter minBelief is set, respectively, to 60%, 50% and 71%.

to 0 when corresponding Gestalt property tends to be active, we set

$$m_k(G_{ij}) = \frac{2}{N}\left[1 - \frac{M_k(G_{ij})}{\overline{M_k}}\right]$$
$$\text{if } M_k(G_{ij}) < \overline{M_k} \text{ or } m_k(G_{ij}) = 0 \quad \text{else.} \tag{15}$$

N is the number of Gestalt properties used. The normalization factor $2/N$ ensures that one single property cannot give a combined belief of 1 (total belief) in one hypothesis but rather needs other properties to cooperate with it in order to make a grouping occur.

4. Results

4.1. Results on artificial images

Fig. 9 shows how our system behaves on artificial images. For each of the four grouping hypotheses (surrounded in red), three partial beliefs are exhibited, corresponding to the Gestalt properties of similarity, closure and symmetry. Finally, a total belief for each grouping is computed, based on Dempster's rule. Recall that value of one for belief corresponds to total truth. First of all, we can see that our implementation of each Gestalt property seems quite efficient: symmetries between squares are

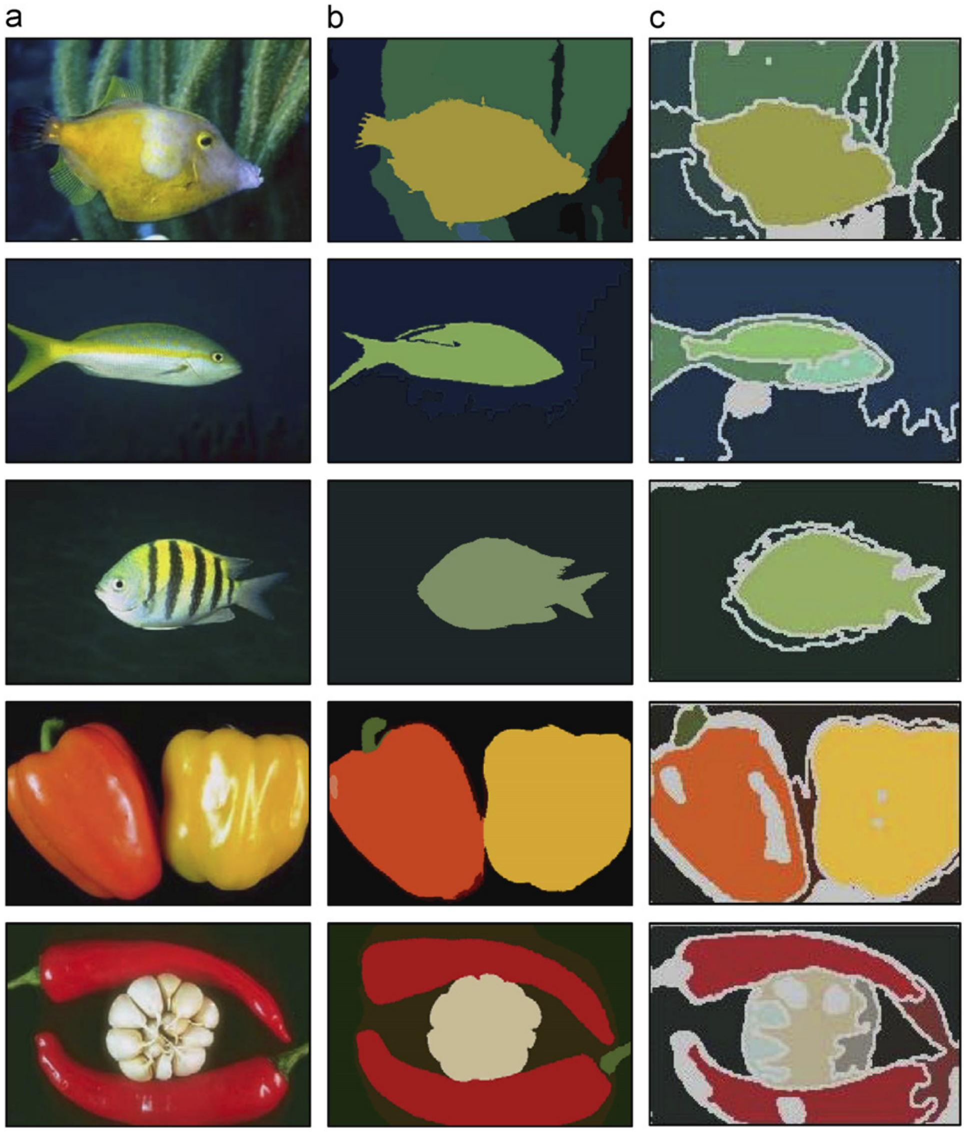

Fig. 13. Evolution of the results when image is under segmented. The first row shows the "normal" segmentation, the second one the under segmentation. The first column is the segmented image. Parameter minBelief is set, respectively, to 80% (b), 70% (c), 60% (d) and 50% (e).

detected, and so as to the color-based similarity. Note also that the closure belief tends to favor compact groupings.

Secondly, we can see that properties tend to reinforce their beliefs. Hence, the pattern on the left has the strongest total belief, since all Gestalt properties have been triggered. On the contrary, when some properties disagree on the grouping, our system is able to find a total belief, though, taking into account the conflicting viewpoints (two patterns on center).

4.2. Results on natural images

Given a total belief value for each hypothesis (i.e. edge), RAG is then iteratively reduced. At that moment, we use a High Confidence First (HCF) algorithm, which iteratively merges regions linked by the edge with the strongest associated belief. This ensures the reduction process to converge, but does not prevent it from falling into local extrema. Further works will be directed in this way. RAG is iteratively reduced until there is no more edge whose associated belief is greater than a value (denoted minBelief). This one is linked to the granularity of the grouping. The more it will be, the less the regions will be grouped. Note that the setting of this value is hence quite intuitive for a user.

We have tested our perceptual grouping on previously segmented images. We used a subset of Corel© database, segmented by a mean-shift color-based algorithm described by Comaniciu and Meer [3]. Examples of results are shown in Fig. 10 where three images are displayed: original image in (a), segmented image in (b) and perceptually grouped image in (c). Results show that perceptual grouping is able to significantly reduce noise issued from segmentation step. Besides, perceptual grouping tends to make semantic objects emerge (fishes in

Fig. 14. Comparisons on original images (a) of perceptual grouping (b) vs Blobworld (c). Note that gray regions in Blobworld correspond to discarded pixels. Parameter minBelief is set, respectively, to 50%, 40%, 35%, 62% and 55%.

top images, and different vegetables on last two images). Note for example that artifacts due to illumination on last two images are removed thanks to closure property. Note also that artifacts on left of yellow bell pepper (fourth image), issued from segmentation are removed during perceptual grouping.

Other results of perceptual grouping are displayed in Figs. 11 and 12. As content is far more complex, perceptual grouping is not able to fully extract semantic objects. However, groupings make salient structures emerge which may be useful for subsequent treatments of object recognition for instance. First of all, perceptual grouping leads to an extraction of subject from background. In addition, several regions of interest are correctly extracted such as skin, clothes for images of women and wheels or body for cars. Note that girl's hair in third image is wrongly merged with a long thin region from background, due to segmentation. It is therefore merged with the whole background during perceptual grouping thanks to a combined similarity–closure property. Note also that in the second image of cars, the ground is fully recovered due to the use of both similarity and symmetry properties. The same holds for the grouping of the car's body in the first image.

Finally, we conduct some experiments to evaluate how the proposed perceptual grouping handles a poorly segmented image like under segmentation. Based on the Comaniciu segmentation method, we have modified parameters to obtain under segmentation and the perceptual grouping has been performed. Fig. 13 presents the merging process for accurate segmentation (first row of images) and for an under segmentation example (second row). When the segmentation is correct, the merging process follows regular evolution function of the parameter minBelief. When a under segmentation is performed, the merging process is not activated in the first steps (high values for minBelief). For lower values of minBelief, our method leads to equivalent merging results as the accurate segmentation (Fig. 13(c) and (d)).

Computation time depends on the initial segmentation and more precisely on the number of regions. On the subset of images used (500 images), it is less than 1 s on a 3 GHz Pentium

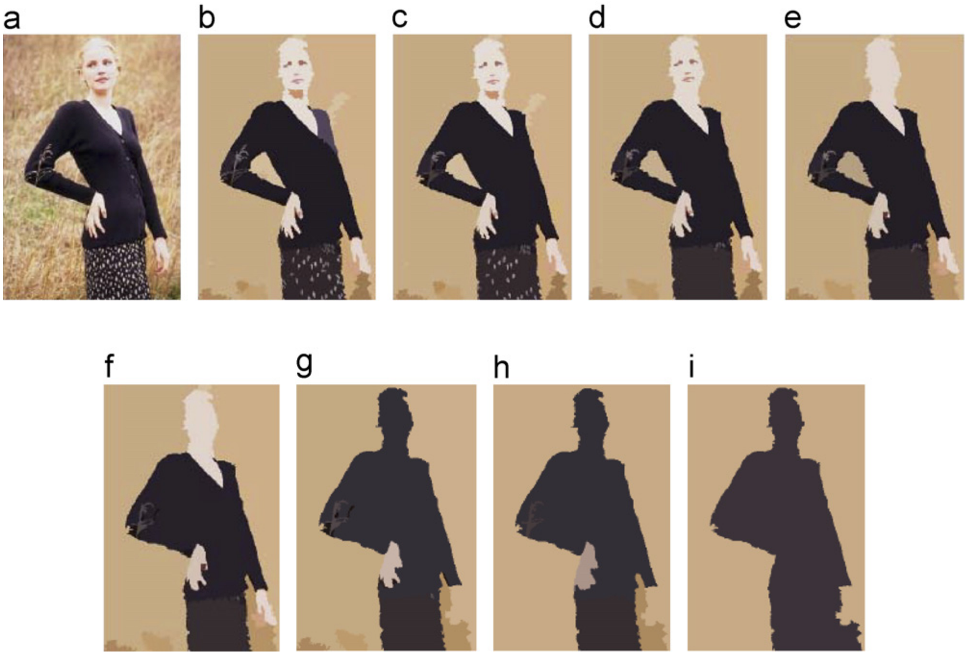

Fig. 15. Comparisons on original image (a) of perceptual grouping (b) vs works from Idrissi et al. (c) and from Luo and Guo (d). Parameter minBelief is set, respectively, to 60%, 50% and 70% for the perceptual groupings.

4 for both segmentation and perceptual grouping of one image. Images' size is 256*384 pixels. Segmentation step typically produces between 200 and 900 regions depending on its content, and perceptual grouping is able to produce a description that handles between 4 and 50 regions.

4.3. Comparisons to other systems

It is very difficult to evaluate segmentation results. As a matter of fact, one should keep in mind that segmentation is not a goal on its own, but a preliminary step for computer vision. Martin et al. [23] propose a framework to quantify segmentation results, based on a comparison with hand-labeled regions in color images. However, such hand-labeling is highly subjective and may become relevant or not depending on the overall tasks in which segmentation happens to be. For instance, one given segmented image could be though of as under- or over-segmented, depending on the level of detail needed. Hence, if we are interested in extracting subject from images, result in first image from Fig. 10 could be though of as a correct segmentation as it tends to isolate the fish's body from background. On the contrary, the same result becomes undersegmented if we are focusing on identification tasks, comparing extracted fish with models from a database. That is why we did not use such a framework to evaluate our results. Instead, we made comparison with other systems on a subset of Corel© database.

In order to show the relevancy of perceptual grouping over other methods based on segmentation only (that is: similarity and proximity properties), we have made comparisons with a well-known segmentation framework which has proven to be efficient: Blobworld system [1]. It models image with a mixture of Gaussians in a multidimensional space, involving color, texture and position for each pixel and uses expectation-maximization principle in order to estimate the model's parameters. Examples of results are shown in Fig. 14(c). Note that gray regions in Blobworld results correspond to discarded ones. We can see that our system performs better, in so far that it is able to group regions that have different color or texture descriptors but present a kind of unity though. Hence, the fishes are almost recovered in the first two images because perceptual grouping does not rely solely on low-level descriptors like color or texture. On the contrary, Blobworld's description of first image keep regions near fish's eye and tail separated both from the background and the fish. Besides, the use of continuity or closure property can also help to remove artifacts due to reflections for instance in the last two images.

Further comparisons have been made with two other existing systems, which specifically deal with perceptual grouping. Results are shown in Fig. 15, where original images are displayed in (a) and results from our perceptual grouping in (b). Then, column (c) shows results from Idrissi et al. [18], who implement proximity, similarity and closure properties on region-based color quantified images. Each grouping hypothesis is assigned with a score, based on color-similarity, weighted by a corrective parameter representing closure. While interesting results are shown with this method, the over-emphasis put on color similarity may lead to inconsistencies as in the second image, where background is merged with woman's body. Another example of such inconsistency is shown in third image, where girl's arm is wrongly grouped with her knee. Besides, results from this method are generally over-segmented compared with ours. This comes from the fact that such

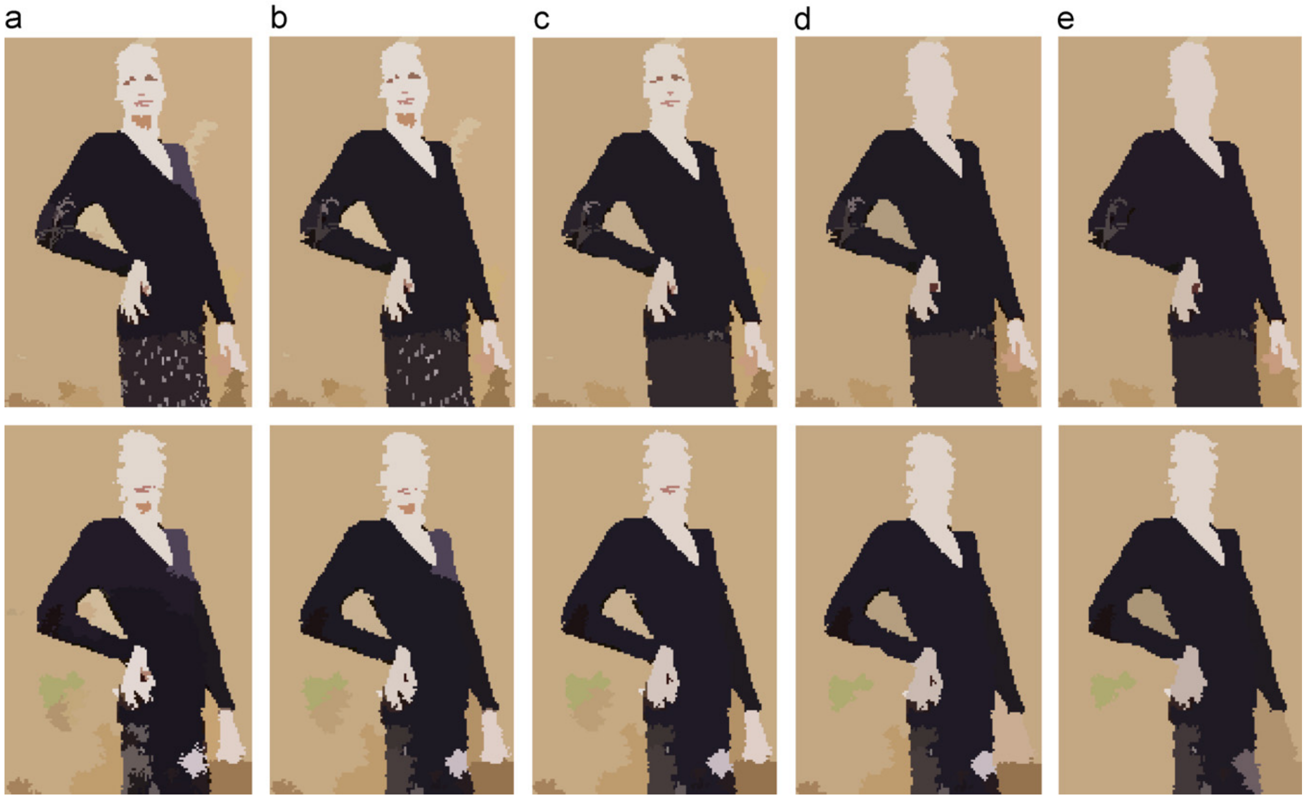

Fig. 16. Evolution of perceptual grouping with respect to the parameter minBelief (c–i). Original image is shown in (a) and segmented image on (b). Parameter minBelief is set, respectively, to 80% (c), 70% (d), 60% (e), 50% (f), 40% (g), 20% (h) and 5% (i).

groupings are often early stopped, to avoid too much inconsistency. On the contrary, as our system needs several laws to interact locally for a grouping, it can handle more groupings while preventing inconsistencies from being too active. This leads to a strong increase in robustness.

The second system to be compared with was the one proposed by Luo and Guo [19], which is based on MRF model from region-based segmented images. Column (d) in Fig. 15 shows some examples of results. Quality of grouping is quite similar to our (column (b)). For instance, if girl's hair in second image is missed by our method, the whole face is recovered in one single region, contrary to (d). However, one must note that the system proposed by Luo relies on empirically derived weights of importance for each Gestalt law, which is a strong limitation. As a matter of fact, such settings may soon become hazardous, as it is not intuitive to a user. On the contrary, our algorithm involves only one parameter minBelief (see Section 4.2), which is related to the granularity of the segmentation and thus is far more easy to set intuitively. Moreover, its range is not too large, since it goes from 40% to 70%, in the image subset used. In Fig. 16, the minBelief influence is presented for a given subset taking a large range of values. For this image, we note that the best perceptual grouping corresponds to 60% (Fig. 16(e)) where we can distinguish all the relevant details of the woman. Below this value, the grouping leads to more global information: separation of the object and the image background (5%) (Fig. 16(i)).

5. Conclusion and future work

We have presented a new framework for perceptual grouping in a pre-attentive context. We rely on a region-based segmented image. This allows us to use contour or region primitives when they are the most suitable for a given Gestalt property. Each grouping hypothesis is characterized by several belief values standing for different Gestalt properties. Then, Dempster–Shafer theory is used for combining belief values and for deriving a global significance for each hypothesis. The interaction model represents a gain for robustness during grouping, as it prevents one single law from triggering alone a grouping. Besides, our system relies on dynamic normalization for the computation of bpa and is thus easy to control, by a unique parameter representing the granularity of the process. Finally, it is independent of segmentation step, as it can handle any kind of region-based segmented image. But, even if we consider Gestalt and Dempster–Shafer theory for intermediate level, they could be used to improve other levels of treatments. For example, during the segmentation process, the similarity principle could be applied to merge pixels with close colors or to define texture patterns according proximity principles. In the same way, our proposition is in accordance with high-level aspects. As the perceptual grouping produces more complex objects, it is important to focus research works on the relevance of these objects. We addressed this subject in another paper [25] in which the perceptual grouping is done incrementally on a Region Adjacency Graph and all these merging steps are stored into a hierarchical structure. The aim is to exploit the various granularities of objects for efficient indexing and object-based retrieval. For instance, perceptual description could be used as an image's signature for indexing. Hence, similarities between images could be derived with graph-based similarities. In doing so, we hope handling more meaningful descriptors, integrating a strong spatial dimension, than those currently used in content-based indexing. Another prospect could be to use principles of perceptual grouping as a new similarity measure between two images. Finally, this model approach combined with a domain description like ontology [24] could be extended to a more semantic process of interpretation.

Further works will be directed on the graph reduction process. As a matter of fact, we use at that moment a greedy algorithm that ensures convergence. However, it can still fall into local extrema. Hence, a graph-based algorithm might be more efficient. For instance, the min-cut algorithm could be tested. Another improvement concerns the choice of the minBelief parameter which can be adapted to the type of the images subset or has to be selected to obtain the required granularity.

Moreover, in order to solve the problem of occlusions or geometry, some Gestalt laws seem to be well fitted to give information about occlusion (T junction) or perspective (Y junction). These features could help to obtain new information about the 3D scene represented by the photo and specifically relative disposition of objects. Also, contour-based features could enrich our perceptual grouping, especially when the images contain line drawings or synthetic objects. The use of these features for the scene description is actually under investigation.

Finally, one should remember that segmentation is not a goal on its own, but rather an essential first-step for vision. Patterns extracted by perceptual grouping have to be verified and interpreted by attentive processes. That is why we are currently working on the use of structural shape models for object recognition. Promising results have already been obtained for non-deformable objects.

Acknowledgment

We want to thank Emilie Naud for her participation to final experimentations and reviews of this article.

References

[1] C. Carson, S. Belongie, H. Greenspan, J. Malik, Blobworld: image segmentation using expectation-maximization and its application to image querying, IEEE Trans. Pattern Anal. Mach. Intell. 24 (8) (2002) 1026–1038.

[2] J. Wang, J. Li, G. Wiederhold, Simplicity: semantics-sensitive integrated matching for picture libraries, IEEE Trans. Pattern Anal. Mach. Intell. 23 (9) (2001) 947–963.

[3] D. Comaniciu, P. Meer, Robust analysis of feature spaces: color image segmentation, in: Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR'97), 1997, pp. 750–757.

[4] K. Koffka, Principles of Gestalt Psychology, Harcourt, New York, 1935.

[5] G. Kanizsa, Organization in Vision, Praeger, New York, USA, 1979.

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

[7] S. Sarkar, K. Boyer, Perceptual organization in computer vision: a review and a proposal for a classificatory structure, IEEE Trans. Syst. Man Cybern. 23 (2) (1993) 388–399.

[8] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.

[9] S. Sarkar, K. Boyer, Integration, inference, and management of spatial information using Bayesian networks: perceptual organization, IEEE Trans. Pattern Anal. Mach. Intell. 15 (3) (1993) 256–274.

[10] S. Sclaroff, L. Liu, Deformable shape detection and description via model-based region grouping, IEEE Trans. Pattern Anal. Mach. Intell. 23 (5) (2001) 475–489.

[11] D. Forsyth, M. Fleck, Automatic detection of human nudes, Int. J. Comput. Vision 32 (1) (1999) 63–77.

[12] A. Desolneux, J. Morel, A grouping principle and four applications, IEEE Trans. Pattern Anal. Mach. Intell. 25 (4) (2003) 508–513.

[13] S. Sarkar, K. Boyer, A computational structure for preattentive perceptual organization: graphical enumeration and voting method, IEEE Trans. Syst. Man Cybern. 24 (2) (1994) 246–267.

[14] A. Ackermann, A. Massmann, S. Posh, G. Sagerer, D. Schlüter, Perceptual grouping of contour segments using Markov random fields, Pattern Recognition Image Anal. 7 (1) (1997) 11–17.

[15] H. Kang, E. Walker, Multilevel grouping: combining bottom-up and top-down reasoning for object recognition, in: Proceedings of International Conference on Pattern Recognition (ICPR'94), 1994, pp. 559–562.

[16] R. Mohan, R. Nevatia, Perceptual organization for scene segmentation and description, IEEE Trans. Pattern Anal. Mach. Intell. 14 (6) (1992) 616–635.

[17] V. Murino, C. Regazzoni, G. Foresti, Grouping as a searching process for minimum-energy configurations of labelled random fields, Comput. Vision Image Understanding 64 (1) (1996) 157–174.

[18] K. Idrissi, G. Lavoue, J. Ricard, A. Baskurt, Object of interest-based visual navigation, retrieval, and semantic content identification system, Comput. Vision Image Understanding 94 (1–3) (2004) 271–294.

[19] J. Luo, C. Guo, Perceptual grouping of segmented regions in color images, Pattern Recognition 36 (12) (2003) 2781–2792.

[20] A.K. Jain, F. Farrokhnia, Unsupervised texture segmentation using Gabor filters, Pattern Recognition 24 (12) (1991) 1167–1186.

[21] P. Vasseur, C. Pegard, E. Mouaddib, L. Delahoche, Perceptual organization approach based on Dempster–Shafer theory, Pattern Recognition 32 (8) (1999) 1449–1462.

[22] T. Pavlidis, S. Horowitz, Segmentation of plane curves, IEEE Trans. Comput. 23 (8) (1974) 860–870.

[23] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of International Conference of Computer Vision (ICCV), vol. 2, 2001, pp. 416–423.

[24] N. Zlatoff, B. Tellez, A. Baskurt, Image understanding using domain knowledge, in: Proceedings of RIAO Conference (Recherche d'information Assistée par Ordinateur), 2004, pp. 277–290.

[25] N. Zlatoff, G. Ryder, B. Tellez, A. Baskurt, Content-based image retrieval: on the way to object features, in: Proceedings of International Conference on Pattern Recognition (ICPR'06), 2006, pp. 153–156.

About the Author—NICOLAS ZLATOFF graduated as an engineer from the Ecole Central of Lyon in 2002. He received a Ph.D. in Image Indexing in 2006. His researches concerned image indexing and retrieval.

About the Author—BRUNO TELLEZ graduated from the University of Bordeaux and the Engineering University, Institut National des Sciences Appliquées de Lyon (M.S. degree in 1996). He received a Ph.D. degree in 1999 also from INSA of Lyon. He is an Assistant Professor in the University of Lyon.
His present research interests within the LIRIS laboratory include image analysis, retrieval and interpretation.

About the Author—ATILLA BASKURT was born in Ankara, Turkey, in 1960. He received a B.S. degree in 1984, an M.S. degree in 1985 and a Ph.D. in 1989, all in Electrical Engineering from Institut National des Sciences Appliquées (INSA), Lyon, France.
He is currently a Professor in the Department of Telecommunication of INSA, Lyon, France. Since 2003, he is the Chair of this department.
He works in animation at the IMAGINE team of LIRIS laboratory and belongs also to the M2DisCo team. These teams perform image and 3D data analysis and segmentation for image compression, image retrieval, shape detection and identification. His technical research and experience include digital image processing, 2D–3D data analysis, compression, retrieval and watermarking, especially for multimedia applications.
Since 2006, he is also “Chargé de Mission” on Information and Communication Technologies at the French Research Ministry.