PERCEPTUAL ORGANIZATION APPROACH

BASED ON DEMPSTER-SHAFER THEORY

P. VASSEUR, C. PEGARD, E. MOUADDIB, L. DELAHOCHE

Perception in Robotics Group
GRACSY (Groupe de Recherche sur l’Analyse et la Commande des SYstèmes)
Université de Picardie Jules Verne
7, rue du Moulin Neuf
80000 AMIENS - FRANCE
Pascal.Vasseur@sc.u-picardie.fr
SUMMARY

In this paper, we deal with the perceptual organization which is a crucial problem in computer vision. Perceptual organization consists in structuring the image into perceptually significant groups. These structures will be used in higher level processes such as image understanding or object recognition. The grouping is based on application of the Gestaltic perceptual phenomena. Originally, the Gestaltic theory tended to explain psychological and physiological mechanisms of the biological vision. In computer vision, perceptual organization provides complexity reduction for high level treatments and also tackles the problem of noise which induces fragmentation.

However, its implementation is confronted to the combinatorial aspect of the process and the variability of the scene. Thus, we propose, in this paper, a perceptual grouping method based on the application of Gestaltic rules, by the Dempster-Shafer theory. The aim of this probabilistic approach is twofold. First, this method is applied in order to rectify segmentation mistakes by restoring the coherence of the primitives. Next, it allows to form groups of primitives which correspond to separate real objects in the scene.

The method, which is proposed in this paper, is completely bottom-up and is able to discern different objects in the scene without prior knowledge. Moreover, it uses no threshold and thus, the variability of the scene does not affect the final results. We show in this paper how we apply the Dempster-Shafer theory, usually used in data fusion, in order to obtain an optimal adequation between our problem and this tool. In order to demonstrate the robustness and the reliability of our algorithm, we finally show experimental results on both real indoor and outdoor scenes.
ABSTRACT
In this paper, we propose an application of the perceptual organization based on the Dempster-Shafer theory. This method is divided into two parts which respectively rectifies the segmentation mistakes by restoring the coherence of the segments and detects objects in the scene by forming groups of primitives.

We show how we apply the Dempster-Shafer theory, usually used in data fusion, in order to obtain an optimal adequation between the perceptual organization problem and this tool. We show that without any prior knowledge and any threshold, our bottom-up algorithm detects efficiently the different objects even in cluttered environment. Moreover, we demonstrate its robustness and flexibility on indoor and outdoor scenes without any modification of parameters.

Keywords: Dempster-Shafer Theory, Fusion, Grouping, Perceptual Organization, Preattentive Component.

1. INTRODUCTION
For many years, research on computer vision systems has been developed in order to design efficient automatic systems. However, image understanding is a particularly difficult and delicate task. Indeed, copying human vision abilities in artificial vision is very complex and a system does not exist yet which presents a real convincing solution\(^1\). Noise and computational complexity are the main obstacles to the achievement of such a system.

The noise is divided into two classes\(^2\). First, there are weak perturbations which have repercussions on the whole image. This noise can be compared with interferences. The second category is made of important noises as occlusions, which spoil small parts of the image. The problem of the noise is that it immediately induces defaults in the primitive extraction stage. Thus in our case of straight line segments, we obtain fragmented edges.

The computational complexity is essentially important during the recognition stage. Indeed, the matching is a combinatorial problem and the computational cost grows exponentially with the number of primitives. So, an exhaustive verification between the model and image primitives is then prohibitive. In order to solve this difficulty, a solution consists in grouping primitives into significant structures describing distinct objects.
The problem of organizing primitives into perceptual structures represents a crucial stage in computer vision and is a prerequisite in object recognition\(^{(3)}\).

In the next part, we outline the importance of perceptual organization through previous works in order to present our contributions. The third section is dedicated to an introduction to the evidence theory of Dempster-Shafer. We develop, respectively in the fourth an fifth parts, our applications to the segment fusion and to the perceptual grouping. Finally, we conclude with experimental results obtained on real indoor and outdoor scenes and we present our further works and perspectives for our method.

2. PREVIOUS WORKS

Perceptual organization is a global term referring to several sets of methods which contribute to create order in visual information to be processed. This subject was firstly essentially treated by physiologists and psychologists in order to explain the biological vision\(^{(4)}\). The Gestaltic school brought one of the most important contributions to explanation of this phenomenon. Thus, numerous rules have been formulated in accordance with spatial and temporal relations between primitives, their sizes, natures and localizations.

With the emergence of computer vision systems and the associated requirements, many of these rules have constituted a particular topic of interest. Thus, many paradigms propose to apply these rules in order to organize the image data into perceptually significant groups and structures\(^{(5, 6)}\).

The benefit of this image organization is threefold. It allows to reduce the effect of the noise induced by segmentation by restoring the coherence of the primitives. Next, it implies that the number of primitives decreases in the image. This phenomenon is amplified by the organization into significant groups which represent hypotheses about the detected objects in the scene. These hypotheses, which constitute the third result, are then considered by the high level treatments in order to semantically interpret the image.

According to Sarkar\(^{(7)}\), three points characterize the perceptual organization in computer vision. First, it is essentially symbolic with attentive (top-down) and preattentive (bottom-up) components. These properties tend to solve the following dilemma : to interpret precisely any small part of an image, it is necessary to understand its larger
context, and it is impossible to understand the context with the only interpretation of small parts. On the one hand, the direct interpretation of a large patch is impossible because of its complexity and its variability; on the other hand, the context understanding implies to know which patches are significant groups in the image. Thus, the solution of this dilemma is reached by formulating hypotheses about groups during bottom-up process and next by verifying them thanks to the attentive component. In this paper, we only present the preattentive structure because we consider that the coherence verification and interpretation are ensured by the higher level process which is presented in (8).

The second characteristic underlines the hierarchic behavior of the perceptual organization. From local primitives, it is important that a global structure emerges. This hierarchy gives an explicit representation of the groups but implies an exhaustive search among the primitives of each level. Thus, the larger the number of levels, the stronger the description, the higher the computational complexity and vice versa. We are then confronted with a new dilemma between the description and the complexity. As we will see further down, we use in our application a description with two levels in order to bound the complexity to the detriment of an accurate representation.

The last characteristic, previously mentioned, concerns the reduction of the number of primitives which is the result of the hierarchical aspect and the provided groupings.

One of the pioneers in perceptual organization is Lowe (6). He applies several Gestaltic criterion in his SCERPO system. The significant groupings are performed according to the proximity, the colinearity and the parallelism of the primitives. From these geometric considerations, an accidentless apparition probability is computed in order to allow the association. We can distinguish the systems which employ the perceptual organization in order to form significant groups to be identified (9, 10) from those which extract primitives from the pixels of the image (11, 3). In (12), an excellent survey on the applications of the perceptual organization is proposed. A classificatory structure is also provided according to the dimension of the organization (2D, 2D ½ ...) and the kind of primitives (pixels, straight line segments, geometric shapes).
When the organization is top-down, the knowledge of the object allows to focus the research on particular points\(^6\). When it is bottom-up, the organization consists without prior knowledge in grouping the primitives\(^{13, 14}\).

The bottom-up processes are very difficult to manage essentially because of their computational complexity, but they offer a more important flexibility.

To solve the combinatorial difficulty, numerous solutions are possible. In general, they propose a compromise between the size of the space search and the complexity of the calculation. Thus, Lowe suggests to arrange the primitives according to their spatial distance. Then, the search is done only in the surroundings of each primitive\(^6\).

This method restricts the number of combinations to the detriment of a local aspect characterized by its myopia. On the contrary, Sarkar and Boyer examine the whole set of segments in order to induce, thanks to a voting method, the different groups\(^7\). In this case, the relations between edges form a graph in which cliques, connected components and cycles represent different types of perceptual structures. In this way, the relations are not privileged according to the distance but the complexity depends directly on the number of input primitives.

Two categories of paradigms allow to apply the Gestaltic principles in order to perform the perceptual organization. We distinguish the methods based on artificial intelligence\(^{13, 6}\) and the probabilistic approaches\(^{7, 11}\).

Artificial intelligence allows a fast implementation in developing simple rules such as “If ... then ...”. These rules use generally heuristic thresholds which give the validity of the associations. If these thresholds are static or manually modified, the method becomes very sensitive to the variability of the scene. Then, it is necessary to implement dynamically variable thresholds according to the context of the image. Probabilistic methods generally employ the Bayesian theory and avoid the defaults of the previous techniques often to the detriment of a complex modelization.

In this paper, we propose a perceptual organization approach based on Gestaltic criteria associated with the evidence theory of Dempster-Shafer. Our algorithm allows to correct the segmentation mistakes or potential occultation over a wide range of images without any fixed threshold. Moreover, it enables us to form significant groups of primitives which can be easily interpreted as objects in high level treatments. The final result provides a two levels representation describing respectively the segments and the objects.
III. AN OVERVIEW OF THE DEMPSTER-SHAFER THEORY

The theory of evidence proposed by Dempster and Shafer belongs to the class of subjective probabilistic theories. This tool is generally applied in multi-sensor fusion\(^{(15)}\) or in image fusion\(^{(16)}\).

This theory is both an evidence theory, as it uses a degree of support based on evidence, and a probable reasoning theory as it gives an evidence combination\(^{(17)}\). This theory, called subjective theory, is based on an individual judgment and is more interested in credibility and plausibility than in probability. The Dempster-Shafer method also permits to represent the ignorance of a knowledge and then to function with partial knowledge.

The first step consists in developing a frame of discernment \(\Theta\) which is a finite set of mutually exclusive hypotheses. The frame of discernment may consist in the possible values of an attribute. For example, if we are trying to determine the color of an object, we may set \(\Theta\) to the set consisting of all possible colors.

Then we can define the \(m\) function which assigns an evidential weight to each subset \(A\) of \(\Theta\). This function is also called basic probability assignment. Thus if \(\Theta\) is a frame of discernment, then a function \(m : \mathcal{P}(\Theta) \rightarrow [0,1]\) is called basic probability assignment or weight function if :

\[
m(\emptyset) = 0 \tag{1}
\]

\[
\text{and } \sum_{A \subseteq \Theta} m(A) = 1 \tag{2}
\]

Then \(m\) is equal to 0 for the empty set and to 1 when the evidence is complete. This quantity measures the belief that one commits exactly to \(A\), not to the total belief that one commits to \(A\). To obtain the measure of the total belief committed to \(A\), one must add to \(m(A)\) the quantities \(m(B)\) for all proper subsets \(B\) of \(A\) :

\[
\text{Bel}(A) = \sum_{B \subseteq A} m(B) \tag{3}
\]

If \(\Theta\) is a frame of discernment, then a function \(\text{Bel} : \mathcal{P}(\Theta) \rightarrow [0,1]\) is a credibility function if and only if it satisfies the following conditions :
\[
\text{Bel}(\emptyset) = 0 \\
\text{Bel}(\Theta) = 1
\]

\(\forall n > 0\) and every collection \(A_1, A_2, ..., A_n\) of subsets of \(\Theta\),

\[
\text{Bel}(A_1 \cup \cdots \cup A_n) \geq \sum_{l \subseteq \{1, \ldots, n\}} (-1)^{|l|+1} \text{Bel}(\bigcap A_l) \\
\text{with } l \neq \emptyset
\]

In the same way, the plausibility function is defined as:

\[
\text{Pl}(A) = \sum_{B \cap A = \emptyset} m(B) = 1 - \text{Bel}(\overline{A})
\]

Thus, \(1 - \text{Pl}(A)\) is a measure of doubt in \(A\), noted \(\text{Dou}(A) = \text{Bel}(\overline{A})\). A subset \(A\) of a frame \(\Theta\) is called a focal element of a belief function \(\text{Bel}\) over \(\Theta\) if \(m(A) > 0\).

Assume \(\Theta\) is \(\{A, B, C\}\) and

\[
m(\{A\})=0.3 \\
m(\{A, B\})=0.2 \\
m(\{A, B, C\})=0.5
\]

then

\[
\text{Bel}(\{A\})=0.3 \quad \text{Pl}(\{A\})=1.0 \\
\text{Bel}(\{A, B\})=0.5 \quad \text{Pl}(\{A, B\})=1.0 \\
\text{Bel}(\{A, C\})=0.3 \quad \text{Pl}(\{A, C\})=1.0 \\
\text{Bel}(\{A, B, C\})=1.0 \quad \text{Pl}(\{A, B, C\})=1.0
\]

Belief functions are adapted to the representation of evidence because they admit a genuine rule of combination.

Given several belief functions over the same frame of discernment but based on distinct bodies of evidence, Dempster’s rule of combination enables us to compute their orthogonal sum, a new belief function based on the combined evidence.
So we obtain the following rule:

\[
\text{if } k = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \leq 1 \text{ then }
\]

\[
(m_1 \oplus m_2)(A) = \frac{\sum_{A = B \cap C} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B) m_2(C)}
\]

(6)

k can obviously serve as a measure of the extent of the conflict.

For example, if \( \Theta = \{D, D'\} \), and

\[
\begin{align*}
    m_1(\{D\}) &= 0.8 & m_1(\{D'\}) &= 0 & m_1(\{D, D'\}) &= 0.2 \\
    m_2(\{D\}) &= 0.9 & m_2(\{D'\}) &= 0 & m_2(\{D, D'\}) &= 0.1
\end{align*}
\]

we obtain the following table:

<table>
<thead>
<tr>
<th></th>
<th>( m_1 )</th>
<th>( m_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>{D}</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>{D'}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>{D, D'}</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

thus,

\[
\begin{align*}
    m_1 + m_2(\{D\}) &= 0.72 + 0.08 + 0.18 = 0.98 \\
    m_1 + m_2(\{D'\}) &= 0
\end{align*}
\]
In order to combine a collection of belief functions, one can form the pairwise orthogonal sums:

\[
\left( \left( \cdots \left( m_1 \oplus m_2 \right) \cdots \right) \oplus m_{n-1} \right) \oplus m_n
\]

Finally, several criteria may be applied in order to make a decision among the frame of discernment. Non-exhaustively, we can distinguish:

- maximum credibility
- maximum plausibility
- maximum credibility and plausibility
- \(A \text{ as } B, \text{Pl}(A) \geq \text{Bel}(B)\)
- \(A \text{ as } B\text{\textbackslash }A, \text{Bel}(A) \geq \text{Pl}(B)\).

4. APPLICATION TO THE SEGMENT FUSION

4.1. System Overview

In this part, we are presenting succinctly our algorithm and we are explaining each stage which is detailed further. Figure 1 shows the five main operations which allow to perform a complete extraction of the significant segments in the scene.

\[\text{figure 1}\]

The first stage consists in applying the Burns algorithm based on the orientation of the pixel gradient in order to extract the segments from the scene \((18)\). Then for each pair of segments, three criteria based on Gestaltic rules are computed. These values allow to estimate the basic probability assignments useful in order to realize the associations. When the segments must be merged, a link is created and once all the associations are provided, a relational graph is obtained.

4.2. Theory and System Development
After the primitive extraction stage, the merging of segments is essential. Indeed, for many reasons (noise, occlusions, gray level variation, ...), a real edge is often divided in several fragments in the image. Two main difficulties arise from this default of segmentation. First, the number of primitives is increased as well as the combinatorial complexity and secondly the matching is more difficult to perform. To solve these problems, we apply the Dempster-Shafer theory in order to significantly merge the segments.

As mentioned previously, the evidence theory is essentially dedicated to the multi-sensor and image fusion. However, in our case we only have one camera and we work on a single image of the scene. Thus, we decided to treat our problem as a fusion application in order to obtain an optimal use of the theory. So, we have made three virtual sensors which provide different information about the segments. Then the fusion of this information gives the decision concerning the association of the segments.

The three sensors respectively provide distance, angle and orientation difference information. Thus, the C1 sensor measures the shortest distance between the endpoints of the segments. C2 gives the angle formed by two segments in a range of \([0, \pi]\) and finally C3 measures the difference of the gradient orientation (each segment has a gradient orientation among 16 possible). Next, the whole set of segments pair is examined and the different associations are made according to the result of each measure.

In consequence of the results of the associations, we define our frame of discernment with two values as follows: “Yes, the merging is good” and “No, the segments can not be merged”. According to the measures provided by the three sensors, we can establish the basic probability assignment for the subsets of the frame. We determine the correspondence between the measures and the basic probability assignment in function of personal experience (Figure 2, Figure 3 and Figure 4). Among these three functions of equivalence, the first one based on the shortest distance is dynamically variable according to the content of the image. Thus before the beginning of the algorithm, we analyze the set of segments extracted from the scene and we compute the mean length. This value is
used in order to determine the slope of the curves of correspondence. If the mean length is greater than 20 pixels, we fix the value to twenty. In this way, we avoid too large merging which may be rapidly unjustified and wrong.

\[(\text{figure 2})\]

\[(\text{figure 3})\]

\[(\text{figure 4})\]

Given the basic probability assignments $m_1$, $m_2$ and $m_3$ respectively associated to the sensors $C_1$, $C_2$ and $C_3$, it is possible to compute the combined belief thanks to the Dempster’s combination rule:

\[
m_1
\]

\[
\begin{array}{|c|c|c|}
\hline
& \text{YES} & \text{NO} \\
\hline
\text{YES} & \text{YES} & \emptyset \\
\hline
\text{NO} & \emptyset & \text{NO} \\
\hline
\end{array}
\]

\[
m_3
\]

\[
\begin{array}{|c|c|c|}
\hline
& \text{YES} & \text{NO} \\
\hline
\text{YES} & \text{YES} & \emptyset \\
\hline
\text{NO} & \emptyset & \text{NO} \\
\hline
\end{array}
\]

\[
m_1 \oplus m_2
\]

\[
\begin{array}{|c|c|c|}
\hline
& \text{YES} & \text{NO} \\
\hline
\text{YES} & \text{YES} & \emptyset \\
\hline
\text{NO} & \emptyset & \text{NO} \\
\hline
\end{array}
\]

From these tables, it is possible to deduce the measure of the conflict between the subsets of the frame of discernment with the following relation:

\[
k = m_1(\text{YES}).m_2(\text{NO}).m_3(\text{YES}) + m_1(\text{YES}).m_2(\text{YES}).m_3(\text{NO}) + m_1(\text{NO}).m_2(\text{YES}).m_3(\text{YES}) + m_1(\text{YES}).m_2(\text{NO}).m_3(\text{NO}) + m_1(\text{NO}).m_2(\text{NO}).m_3(\text{YES}) + m_1(\text{NO}).m_2(\text{YES}).m_3(\text{NO}).
\]
If \( k \) is smaller than 1, then the conflict is not total and the combined belief for each subset of the frame is given by:

\[
m_1 \oplus m_2 \oplus m_3(YES) = m_1(YES) \cdot m_2(YES) \cdot m_3(YES) / (1-k)
\]

\[
m_1 \oplus m_2 \oplus m_3(NO) = m_1(NO) \cdot m_2(NO) \cdot m_3(NO) / (1-k)
\]

Since we have only two subsets in our frame of discernment, we choose to take the final decision according to the highest degree of credibility. In order to actually illustrate the different rules that we have implemented, we present three simple examples. Each one symbolizes two segments which may be merged. Figure 5 represents two collinear segments which are close. In this case the three sensors provide a belief for YES which is higher than the belief for NO. So the combined belief for YES is the strongest and the two segments can be associated.

*(figure 5)*

Figure 6 proposes an example in which the segments are close but the angle is too large. In this case, the C2 sensor gives a complete belief for NO, and even if C1 provides a higher belief for YES, the final result is NO. Thus, the segments are not merged.

*(figure 6)*

Finally, the last example illustrates a conflict case. Indeed, the segments share a common endpoint, the angle is weak but the difference of orientation is maximum (Figure 7). This case is possible when two objects are only joined by a corner as the squares on chessboard. Then the C1 and C3 sensors provide complete beliefs respectively for YES and NO and the \( k \) coefficient is equal to 1. Thus, in this kind of situation, we privilege the NO decision in order to avoid the loss of information and the creation of false knowledge.

*(figure 7)*

Once the totality of associations is made, we have to detect the chains which link the different segments into significant edges. Then the endpoints of the new segments are evaluated in order to be used in the next stage.

4.3. Experimental Results
Our tests are performed on real images of 512x512 pixels, which include several different objects. No prior knowledge is available about these objects. Figure 8 represents a white parallelepiped on a black background. In this image, 55 segments are detected and several breakings appear in the edges. As a result of our algorithm, we finally obtain 11 segments and the object is completely rearranged with continuous segments.

(figure 8)

Figure 9 includes two distinct objects, respectively polyhedral and spherical. The extractor provides 121 segments in the scene, and we keep, after the application of our method, only 25 of them. As in the previous case, the polyhedron is totally rearranged. However, the most interesting point concerns the approximation of the ring and we can even consider its future recognition in spite of the slight warping.

(figure 9)

An indoor scene is presented in Figure 10 in which several objects with different sizes are include with a natural lighting. In this scene, we can count 656 segments which constitute fragments of objects such as a cupboard, a pyramid, a parallelepiped or a chair. In consequence of our merging, 68 segments are formed and correctly represent the previous mentioned objects. Figure 11 presents the analysis of the segments extracted from the image thanks to the Burns algorithm. The mean length is equal to 13.02 and represents the maximum distance allowed in order to be taken into account. Moreover, the uniform distribution of the orientation and gradient of the segments favor a good merging and avoid accidental association.

(figure 10)

(figure 11)

Thus, the results obtained on real images are very satisfactory and the recovering of the edges of the objects may be useful for the high level treatments. Moreover, our method is efficient over a large range of scenes without any threshold.

To make our method totally valid and to demonstrate its robustness, we apply the algorithm to aerial images used as references in perceptual organization. The aim in this kind of scene is to find relevant information such as roads,
ways and buildings. Figure 12 presents the different results provided on this kind of scene. The initial number of extracted segments is equal to 2438. After merging and suppression of the segments shorter than 10 pixels, 245 segments are considered as significant and representative of the scene.

(figure 12)

4.4. Conclusion

We have presented a method of segment merging based on perceptual organization and Dempster-Shafer theory. We demonstrate its robustness over a large range of images snapped in different conditions and moreover without modification of any parameter. The main goal of the treatments applied during the merging is to recover the segments which have been divided in consequence of the extraction of the primitives. The Gestaltic rules provide criteria that the evidence theory enables to exploit and to combine efficiently.

5. APPLICATION TO THE PERCEPTUAL GROUPING

5.1. Method Description

In the previous section, our aim was to correct the defaults caused by the segmentation. Thus, the number of primitives is appreciably reduced and their perceptual significance is increased. The goal is now to group the segments which are likely to represent the distinct objects in the scene. Unlike the algorithm described in (13) and based on artificial intelligence rules, we do not build a hierarchical representation used as the spine of the method. Indeed, the final result is obtained with only a single analysis of each pair of segments. The drawback of this reduction is a less accurate description of the objects.

Figure 13 describes the global scheme of our algorithm. We favor the criterion of distance which appears as the most important. Then, two segments sufficiently close are automatically grouped. No additional verification is performed in this case because we consider that it is too difficult to deny this result with a bottom-up process. In this way, in order to avoid too many special cases, we use a dynamic threshold similar to the mean length of the previous section.

(figure 13)
If the distance criterion is not verified, then the association is not automatically discarded. In this case, we consider the parallelism and the coherence of these segments. Indeed, the first verification consists in measuring the distance between the centroids of both segments. If this test is negative, then the segments association is immediately canceled. Otherwise, the Dempster-Shafer is applied in order to evaluate the grouping cohesion. This value is estimated by the fusion of the criterion based on similarity, symmetry and parallelism of the segments. As in the previous stage, if the coefficient of conflict is under 1 and if the belief in the association is higher than the belief in its contrary, then a link is created between the two segments. Finally the different chains of segments are covered in order to form the different perceptual significant groups.

5.2. Algorithm Implementation

In this part, we describe the different tests and the basic probability assignments associated with each sensor that we used in our algorithm:

- Distance test: this test is delicate and essential because this criterion has priority but may generate easily irreparable mistakes. It is completely impossible to fix an absolute threshold because of the variability of the images. Thus, as in the first stage, we establish a statistical analysis in which we evaluate the mean length of the segments. Then, if the mean length is higher than 20 pixels, we bound the limit to 10 pixels. Otherwise, the threshold is equal to the half mean length.

- Gap test: this verification allows to limit the search of parallel segments in order to avoid inconsistent associations. The principle consists in measuring the distance between the centroids of the segments. If this length is less than the shortest segment, then we inspect more precisely the potential association. The limit of separation may seem to be a default of the algorithm. However, it seems inconsistent to group two short segments placed at the extremities of the image even if they are similar and parallel. So we have introduced this verification in order to give a likelihood characteristic to the method.

- Dempster-Fusion measures fusion: As in the stage of segments merging, we first develop the frame of discernment which is always made of two subsets: “YES, the segments must be grouped” and “No, the segments
must not be grouped”. We have also developed three new virtual sensors which respectively measure the similarity, the symmetry and the parallelism of the two segments. Figure 14, Figure 15 and Figure 16 represent the functions of correspondence that we have established between the measures and the basic probability assignments of each subset of the frame. The similarity is equal to the ratio of the lengths of the segments. The shortest is always equal to the numerator. The complete belief for YES is provided when the ratio is equal to 1 and for NO if the ratio is equal or less than 0.5. The second sensor measures the parallelism and the tolerance is fixed to plus or minus $\pi/8$ as in the first stage of fusion. Finally the third sensor evaluates the symmetry of the segments. In this way, we measure the angle between the longest segment and the line which passes through the centroid of both segments. A tolerance of plus or minus $\pi/8$ is also taken into account. The measure is performed with the longest segment in order to give the maximum likelihood to this value.

(figure 14)

(figure 15)

(figure 16)

Then, we compute the coefficient of conflict as in the previous stage and the belief for each subset of the frame of discernment as follows:

$$k = m_1(YES).m_2(NON).m_3(YES) + m_1(YES).m_2(YES).m_3(NON) + m_1(NON).m_2(YES).m_3(YES) + m_1(YES).m_2(NON).m_3(NON) + m_1(NON).m_2(NON).m_3(YES) + m_1(NON).m_2(YES).m_3(NON)$$

$$m_1 \oplus m_2 \oplus m_3(YES) = m_1(YES).m_2(YES).m_3(YES) / (1-k)$$

$$m_1 \oplus m_2 \oplus m_3(NON) = m_1(NON).m_2(NON).m_3(NON) / (1-k)$$

The final decision is always taken according to the greatest value between both beliefs. When each link is created, the different groups of primitives are formed in order to be used in higher level process.

5.3. Experimental Results

To demonstrate the efficiency of the method, we apply the perceptual organization to the results of the segment merging previously described. For each image we represent the significant groups with different colors. However, if
there are too many groups, many colors may be repeated. Figure 17 illustrates the very simple case of a single object in the scene. The method detects this object and we can validate it for simple cases.

(figure 17)

Figure 18 corresponds to the scene with a parallelepiped and a ring. We can see that the polyhedral object and the ring are detected separately. It is interesting to note that the spherical object is completely recovered even if it is not our main goal. This situation gives proof of the robustness of the method and of its flexibility.

(figure 18)

Since the validation for simple scenes is made, we are interested in cluttered environment as in Figure 19. This image contains many objects with different shapes and several different sizes. It is worth noting that we change nothing in the parameters of the algorithm. We can affirm that the results are very interesting and satisfactory. Indeed, pyramid, box, desk and rack are separated in the final results.

(figure 19)

In order to provide a comparison with the results obtained by our method based on artificial intelligence\(^\text{(13)}\), we apply the Dempster-Shafer on the same image called “Desk” (Figure 20). The method based on inference rules detects two objects in the image which are significant. The new algorithm based on the Dempster-Shafer theory finds these objects but also detects the wall and other details. The great advantage of the probabilistic method concerns the time of computation. Indeed, we can estimate that this time is divided by 20.

(figure 20)

Finally, in order to complete the evaluation of our method, we compare our results with the results obtained by Sarkar and Boyer on the aerial image that we call “Buildings” (Figure 21) (This image is available on the http://marathon.csee.usf.edu/~sarkar/PO_comparison.html site). In both images, the main buildings are distinctly detected and the different other groups are nearly similar. However, our method is less efficient in the recovering of the road but it is worth noting that our initial aim was to detect polyhedral objects in indoor scenes. Moreover, even for these outdoor scenes, the adjustment of the different parameters is always the same as for the indoor scenes.
Thus we can affirm that our method is efficient in the detection of significant groups in either kinds of scene. Indeed, it presents efficient and robust characteristics which allow to perform higher level process of good quality.

(figure 21)

6. CONCLUSION

Computer vision systems are confronted with two main difficulties such as the noise and the computational complexity of the matching. Perceptual organization sets necessary laws which manage primitive merging and grouping. However in order to efficiently form perceptually significant groups, it is essential to use tools which enable to take into account uncertainty, inaccuracy, loss of information (occlusions) and image variability.

In this paper, we have presented an approach of the perceptual organization which applies Gestaltic rules based on the theory of evidence of Dempster-Shafer. We apply it at two different levels of treatment: Segments merging and perceptual grouping. Experimental results obtained with images different in nature and complexity, give proof of the quality, efficiency and robustness of the method. Besides, the latter is very simply implemented. An interesting furtherwork is to apply this method with different primitives such as ribbons or curves, and to manage several knowledge about a scene in order to extract a lot of information like as is done by Sarkar and Boyer with a bayesian network$^{(19)}$.

VII. REFERENCES


LIST OF CAPTIONS OF THE FIGURES

Figure 1: Algorithm description
Figure 2: Equivalence functions between the sensor C1 and our frame of discernment
Figure 3: Equivalence functions between the sensor C2 and our frame of discernment
Figure 4: Equivalence functions between the sensor C3 and our frame of discernment
Figure 5: Fusion example
Figure 6: No fusion example
Figure 7: Conflict example
Figure 8: Parallelepiped
Figure 9: Parallelepiped and ring
Figure 10: Pyramid
Figure 11: Segments analysis of “Pyramid”
Figure 12: Buildings
Figure 13: Perceptual Grouping Algorithm
Figure 14: Equivalence functions between the sensor C1 and our frame of discernment
Figure 15: Equivalence functions between the sensor C2 and our frame of discernment
Figure 16: Equivalence functions between the sensor C3 and our frame of discernment
Figure 17: Dempster-Shafer on “Parallelepiped”
Figure 18: Dempster-Shafer on “Parallelepiped and ring”
Figure 19: Dempster-Shafer on “Pyramid”
Figure 20: Comparison on “Desk”
Figure 21: Comparison on “Buildings”
Segments Extraction by Burns Algorithm

For each segments pair, Application of Gestaltic Criterion

Decision of the Fusion by Dempster-Shafer

Links Creation between associated Segments

Significant Segments Determination
Distance : 15.2595
m1(YES) = 0.6948 ; m1(NO) = 0.3052
Angle = 174.9584°
m2(YES) = 0.7759 ; m2(NO) = 0.2241
Orientation difference = 1
m3(YES) = 0.8750 ; m3(NO) = 0.1250
Conflict k = 0.5198
m1 ⊕ m2 ⊕ m3(YES) = 0.9822 m1 ⊕ m2 ⊕ m3(NO) = 0.0178
Decision : YES
Distance : 11.5030

m1(YES) = 0.7699 ; m1(NO) = 0.2301

Angle = 146.8032°

m2(YES) = 0.0000 ; m2(NO) = 1.0000

Orientation difference = 3

m3(YES) = 0.6250 ; m3(NO) = 0.3750

Conflict k = 0.9137

m1 ⊕ m2 ⊕ m3 (YES) = 0.0000 m1 ⊕ m2 ⊕ m3 (NO) = 1.0000

Decision : NON
Distance : 0.0000  
m1(YES) = 1.0000  ; m1(NO) = 0.0000  
Angle = 177.8117°  
m2(YES) = 0.9026  ; m2(NO) = 0.0974  
Orientation difference = 8  
m3(YES) = 0.0000  ; m3(NO) = 1.0000  
Conflict k = 1.0000  
No decision because the conflict is complete
Image of luminance  Extracted segments  Final result
Image of luminance

Extracted segments

Final result
Segments length  
Segments orientation  
Gradient orientation
Pair of segments extraction
Distance verification
Gap test
YES
Parallelism
YES
Similarity
Symmetry
Data Fusion by Dempster-Shafer
Association verification
NO
YES
Link creation
Image of luminance  Dempster-Shafer  Artificial Intelligence
<table>
<thead>
<tr>
<th>Dempster-Shafer</th>
<th>Sarkar-Boyer</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Dempster-Shafer" /></td>
<td><img src="image2.png" alt="Sarkar-Boyer" /></td>
</tr>
</tbody>
</table>