A control operator for perceptual grouping based on the Gestalt vision's theory

Jimmy Nagau (1), Anne-Sophie Cappelle-Laizé (1), Christine Fernandez-Maloigne (1), Jean-Luc Henry (2)
(1) XLIM-SIC Laboratory, UMR CNRS 6172, Bd Marie et Pierre Curie - 86962 Futuroscope-Chasseneuil, France.
(2) LAMIA Laboratory, Campus de Fouillole, B.P. 592, 97157 Pointe à Pitre, French West indies.

Abstract

The fusion of regions from segmentation is often a required step in order to analyze objects in the scene. It allows to group under a same closure property the partitions of an image, which denote the same object. In this way, several studies have been made. Most of the algorithms are based on global information to evaluate the relevance of region combinations and fusion. In this paper, we propose a merging process based on Gestalt vision's theory. The particularity of our method is to preserve photographer's region of interest thanks to the definition of a new criterion based on Di Zenzo gradient. Thus, region fusion preserves photographer's relevant objects of the scene. The performance of the proposed method is evaluated using various natural images, including Berkeley image database and its accuracy is shown.

Introduction

A segmentation process creates a partition, formed of connected components representing the objects of an image. It exists numerous approaches (threshold, edge detection, region search...). Segmentation is often achieved using low-level attributes (color, texture, shape...). However, the segmentation result of natural images, which are often characterised by textural areas, is usually composed of a large number of regions; there is some over segmented. Consequently, it is necessary to proceed to a merging operation to allow content analysis and semantic interpretation. The fact remains that all these segmented methods are highly dependent on the pixel distribution of an image and the results, in the case of natural images, show divisions of regions due to the circumstances of the shooting. To solve this problem, many computer vision tools have emerged. They are based on *a priori* knowledge about the subject area. For example, the SIGMA system [9] or Shema [3] allow to define aerial images based on object classes. But all these approaches are dependent on the application domain and, therefore they do not provide a wide flexibility in the variety of information provided by the natural images. So, we are interested in the Gestalt vision's theory that can take into account information from a higher level and which is based on the spatial organization of objects and objects groups concepts. In this paper, we propose a new method for merging regions based upon Gestalt theory. The originality of this work is to integrate, next to the Gestalt properties, the notion of photographer's region of interest by defining on a new criterion. It takes into account the objects, which have interested the photographer during the shooting, by selecting the objects with clear outlines. Regions associated to the objects of interest are preserved while the fuzzy regions, located in the background, tend to merge together.

In the first part of the article, we briefly describe the basics of perceptual grouping in focusing on the Gestalt vision's theory. We present, in a second part, the new criterion to control the merging process and we discuss some of its properties. Finally in the last part, we evaluate the method and propose some results in various application fields.

Perceptual grouping: The Gestalt vision's theory

Gestalt vision theory is based on psychologist works and refers to the fact that the human visual system is highly influenced by the notion of group for image interpretation[13]. Region segmentation process is a necessary step for image interpretation and consists in separating image into homogeneous regions according to a chosen criterion. To allow automatic image analysis, regions have to be in relation with the objects of the scene, such as a human observer could identify them. Thus, it is quite natural to exploit human vision properties in segmentation process to obtain semantic results. In [12], perceptual grouping was exploited using Dempster-Shafer[2] formalism. In [8], Markov Random Fields [7] are used. The obtained results have proved the interest of using Gestalt properties. Some of them are illustrated on Figure 1.

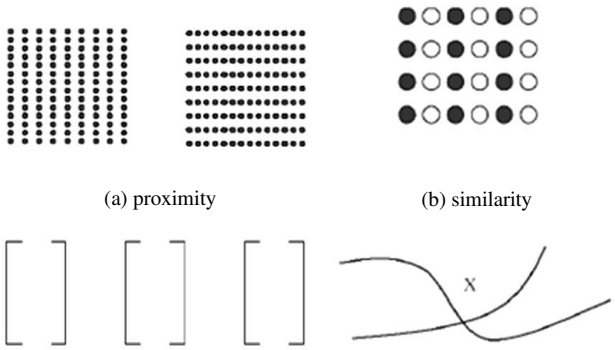

Figure 1. *Some Gestalt properties for perceptual grouping*

Considering that an initial segmentation process has divided a color image I into several regions R_i ($I = \cup R_i$), it is possible to compute some measures between two regions R_i and R_j of I, which reflect the Gestalt properties. In the following, we describe some of these measures.

The parameter $S_{i,j}$ deals with color proximity between two regions R_i and R_j. It is an Euclidean distance between the average

values M_i and M_j of the pixels in regions R_i and R_j in the 3 dimensional CIE L*a*b* color space [4] which is calculated using the following formula:

$$S_{i,j} = \sqrt{\sum_{k=1}^{3} (M_{j,k} - M_{i,k})^2} \tag{1}$$

The spatial proximity between 2 regions is expressed by the parameter $CP_{i,j}$, with

$$CP_{i,j} = \frac{1}{n_j} \sum_{k=1}^{n_j} e^{-\alpha \frac{d(F_{1,i}, Pixel_{j,k}) + d(F_{2,i}, Pixel_{j,k})}{2A}} \tag{2}$$

where n_j is the number of pixels of region R_j, $d(.)$ the Euclidean distance, $Pixel_{j,k}$ is a pixel coordinate in the image space belonging to region R_j, $F_{1,i}, F_{2,i}$ are the ellipse's foci that encompasses the region R_i and the parameter A depends upon elliptic approximation of the region R_i with:

$$A = \frac{l}{4} + \frac{1}{2}\sqrt{L^2 + \left(\frac{l}{2}\right)^2} \tag{3}$$

where variables l and L are the dimensions of the rectangle encompassing a region (see Figure 2-left).

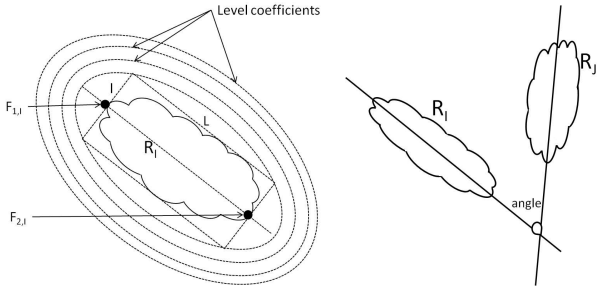

Figure 2. *Notations and relations between two regions R_i and R_j*

The parameter $CL_{i,j}$ (Eq. 4) measures the degree of mutual overlap. It is considered as closure parameter in Gestalt vision's theory:

$$CL_{i,j} = \frac{\min(Perimeter_i, Perimeter_j)}{Perimeter_{i,j}} \tag{4}$$

where $Perimeter_i$, $Perimeter_j$ are the respective perimeter of regions R_i and R_j, and $Perimeter_{i,j}$ their shared perimeter.

Finally, the parameter $CO_{i,j}$ (Eq. 5), that has been proposed in [10], from the original Gestalt parameters, measures the continuity between R_i and R_j.

$$CO_{i,j} = \min\left(\frac{1}{2} + Shape_{i,j} \times Direction_{i,j}\ ,\ 1\right) \tag{5}$$

with:

$$Shape_{i,j} = \frac{\min(Capacity_i, Capacity_j)}{\max(Capacity_i, Capacity_j)} \tag{6}$$

$$Capacity_i = \frac{Perimeter_i^2}{4\pi\ Surface_i} \tag{7}$$

$$Direction_{i,j} = \frac{||180 - angle| - 90|}{90} \tag{8}$$

This parameter is based on the general shape and direction of the treated regions and thus introduce the notions of symmetry and continuity. The $Direction_{i,j}$ is given by the value of the angle formed by the principle main direction of regions R_i and R_j as shown in Figure 2-right. In Eq. 5, the value $\frac{1}{2}$ was added because of presence of noise which can under-estimate the value of $CO_{i,j}$ [10].

In our work, we focus on the treatments of naturals digital images. With such images, the regions of interest are the ones which were focused by the photographer during acquisition process, whereas the rest of the image is fuzzy and can be considered as background. The Figure 3 illustrates such images. Particular attention was driven on the flowers, which belong to first plan. The rest of image constitutes the background and is characterized by its fuzziness. When computing color gradient such as the Di Zenzo gradient on the images[15], magnitude of the gradient edge is higher on the region of interest whereas the background has weak gradient values. In the following, we propose to use such characteristic in a Gestalt fusion algorithm.

Figure 3. *Some natural images and magnitude of the Di Zenzo gradient (in blue are values near 0 and in red the high values)*

Merging process using Gestalt properties and contour information

Gestalt distance and Naive Fusion Algorithm (NFA)

From the original Gestalt properties used in [6], we define a distance Gestalt $DG_{i,j}$ between two regions R_i and R_j by the following equation:

$$DG_{i,j} = S_{i,j} \times CP_{i,j} \times CL_{i,j} \times CO_{i,j} \tag{9}$$

The distance $DG_{i,j}$ is in the interval $[0,1]$. A value close to 1 indicates that the regions R_i and R_j are visually homogeneous whereas a value near to 0 shows that the regions have very different characteristics. This Gestalt's parameter can be used in a simple process of regions merging as described by the Algorithm 1. From here, we call this algorithm *Naive Fusion Algorithm* (NFA).

Algorithm 1 Naive fusion algorithm (NFA)

do
  for each region $R_i \in I$
   find $R_j \in I$ such that $\max_{j,j\neq i} DG_{i,j} > 0$
    if R_j exists then
    merge R_i and R_j
while fusions are possible

New merging constraints

The NFA only uses the Gestalt's property to determine the regions to be merged. As illustrated with Figure 5, the number of regions widely decreases but a lot of details in the region of interest is also lost. When we observe an image, the human eye is sensitive to the reliefs contained in the image and the objects of interest in a scene are often reflected by their position in the focal depth of the lens when shooting. The objects in those regions have good edge sharpness. To increase the NFA's performance and to take into account the human behavior, we now propose to modify the initial Gestalt's criterion by using the notion of sharpness.

We here introduce some notations. If I is an image to segment, I_s is one initial region segmentation such that $I_s = \cup_i R_i$. Each region R_i is associated to its contour C_i. Let I_b be a binary image of I obtained by one contour segmentation process. We denote ζ the set of pixels of I_b non-equal to 0. The contour image I_b is then a new information we propose to exploit in a new fusion criterion as following.

For each region R_i of I_s we calculate a parameter DR_i (Eq. 10) reflecting the quality of the spatial organization of the contour C_i of R_i according to the high magnitude pixels of the contour segmentation I_b. The set of measures DR_i then reflects a similarity degree between the two segmentations: the region and the contour segmentations. The calculation of DR_i, is realized by dividing the edge C_i into a set of k intervals (Figure 4). We call C_i^j the set of pixels of C_i restricted to the interval j (for $j = 1 \ldots k$) and the coefficient DR_i which is defined by:

$$DR_i = \frac{1}{k}\sum_{j=1}^{k} e_j \quad \text{with} \tag{10}$$

$$e_j = \begin{cases} 1 & \text{when } C_i^j \cap \zeta \neq \emptyset \quad (case\ 1) \\ \frac{1/n+1/m}{2} & \text{otherwise} \quad (case\ 2) \end{cases}$$

with $n = \min_{p\in N^*}(C_i^{j-p} \cap \zeta \neq \emptyset)$ and $m = \min_{p\in N^*}(C_i^{j+p} \cap \zeta \neq \emptyset)$. Thus n and $m \in N^*$ represent the number of intervals to browse, respectively, clockwise and counterclockwise, before finding an intersection between the edge points of C_i and ζ from the position j.

For a region R_i, the coefficient DR_i depends on the number of shared pixels between the region segmentation and edge detection but also rests on the position of their relative spatial distribution. It returns a value close to 1 if the relief is fully demonstrated by the edge detector or if the contour points are distributed around a region on a regular basis. With this definition of DR_i criterion, we propose the *new Gestalt Fusion Algorithm* (new GFA) as described in algorithm 2 to achieve the fusion of regions in an image.

Algorithm 2 New Gestalt fusion algorithm (new GFA)

do
  for each region $R_i \in I$
   if $DR_i <$ *threshold*, then
   find $R_j \in I$ such that $\max_{j,j\neq i} DG_{i,j} > 0$
    if R_j exists then
    merge R_i and R_j
while fusions are possible

Results and discussion

Algorithm results

In the remainder of this work, the segmentation algorithm used to compute the initial region segmentation is performed using the global Mean Shift algorithm [1]. This version of Mean Shift is used in a discretized color space. This procedure allows a pixel to move directly to its neighbors in the color space. The Mean Shift has the particularity to require only one single parameter δ_r. It is a distortion parameter, used to set the quantization level[1]. We have chosen this algorithm because of its reduced number of parameters. Although, the choice of segmentation algorithm is dependent on the intended application and its choice should be adapted to the application context.

Dealing with color image, we chose to use Di Zenzo gradient [15] to compute edge in the original image. This multispectral gradient is calculated with the formula 11.

$$\| \vec{\nabla} I \|^2 = \frac{g_{11} + g_{22} + \sqrt{\Delta}}{2} \tag{11}$$

with:

$$\vec{v_1} = \begin{pmatrix} R_x \\ V_x \\ B_x \end{pmatrix}; \quad \vec{v_2} = \begin{pmatrix} R_y \\ V_y \\ B_y \end{pmatrix}; \quad g_{ij} = \vec{v_i} \cdot \vec{v_j}$$

$$\Delta = (g_{11} - g_{22})^2 + 4g_{12}^2 \quad \text{and} \quad \theta = \frac{1}{2}\arctan\left(\frac{2g_{12}}{g_{11} - g_{22}}\right).$$

The set of coefficients DR_i allows selecting regions. The selected regions have both edge sharpness and good Gestalt properties. Also, their are the best candidates for the merging. The figure5 shows a result obtained by our proposed algorithm. The figure 5-b shows the Mean-Shift initial segmentation results. As we can observe, these results are visually relevant however they cannot permit a semantic analysis of the segmented regions due to the too large number of regions (17 regions). This oversegmentation is typical when dealing with natural and textured images and justifies a merging process as a post-processing treatment.

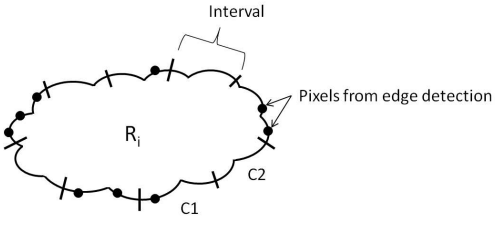

Figure 4. *A region and its cutting border*

The figure 5-c and 5-d are respectively the results of the NFA, which only used Gestalt properties and new GFA, which is our proposed one. Images 5-c and -d show more nuanced picture leaving better perception to analyze the semantic content of a scene. The segmentation results in figure 5-d are comprehensive (we can identify an animal in the middle of vegetation) even if we don't know anything about the original image. Using our method, we preserve regions of high relief (regions with high values of DiZenzo gradient). These regions are the one chosen by the photographer through the focusing of his camera lens.

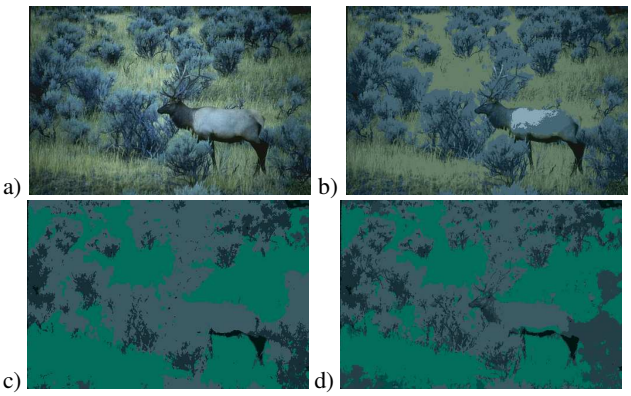

Figure 5. *Photograph of a deer in a plain, (a) original image, (b) discretized global Mean Shift, (c)result of the merger with only the Gestalt distance (NFA) and (d) our proposal (new GFA).*

As shown in the pseudo-algorithm (algorithm 2), the fusion result depends on some parameters: the number of intervals k and the threshold value. The parameter k influences the strength of the fusion process. As an example, image 6 shows different results of segmentation based on the value of k.

In the first case (6-c) for an interval value k large ($k = 1000000$), we are looking for regions with a large number of contour points. The algorithm so will not prevent the submission of the regions of mountainous to a potential merging because the low number of edge point contained. In the second case (6-d), with a k parameter sufficiently small ($k = 5$), the system is more flexible and edge map pixel distribution around the mountainous regions is sufficient to block its merging with the background.

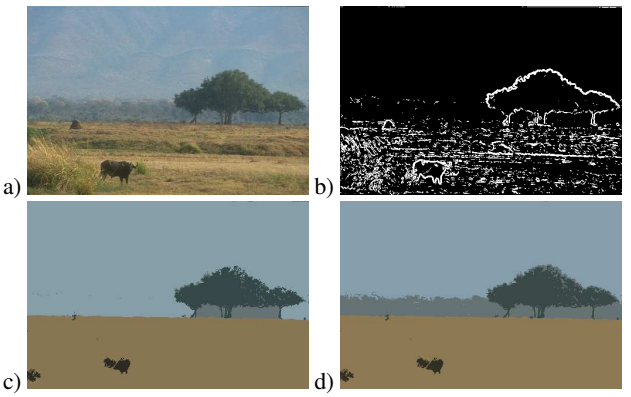

Figure 6. *Photograph of a Savannah (a) original image, (b) edge map, (c)result of the new GFA with $k = 1\,000\,000$ and (d) new GFA with $k = 5$*

On figures 5 and 6, we observe that the new GFA provides a better representation of objects in the scene. The algorithm avoids over-segmentation by reducing drastically the number of regions in the image while preserving areas preferred by the photographer: only the fuzzy outline objects were merged. We observe an average reduction about 90% of the region number and the labeling shows the morphological conservation of focused regions. This method remains heavily dependent of the used edge detector. Indeed, this one must be adapted to process data and must minimize detection of false edges. However, the color edge detector of Di Zenzo exactly meets these criteria. To process the images, a threshold of 0.5 was used for the distribution evaluation of edge points around a region. The parameter k (value of 5 for testing), flexible enough show more details in the scene.

The proposed method is applying on plant image database. Goal is to extract a set of regions to obtain morphological, structural and color characteristics of plants in order to compare them with known patterns. The database is a set of color images of natural scenes in various environments which we cannot have any information. As previously, discretized global Mean Shift and respectively the Di Zenzo operator are used to provide respectively the initial segmentation and respectively the edge map. The result of the Mean Shift and the proposed merging algorithm are shown Figures 7 and 8. Using global Mean-Shift in discretized color space, which requires only one parameter, brings speed and reduced color diversity. As observed in Figures. 7 and 8, morphological parts of the plants (leaves, petals, stems…) are better retrieved using new GFA. This example shows the efficiency of the proposed method on natural, color images.

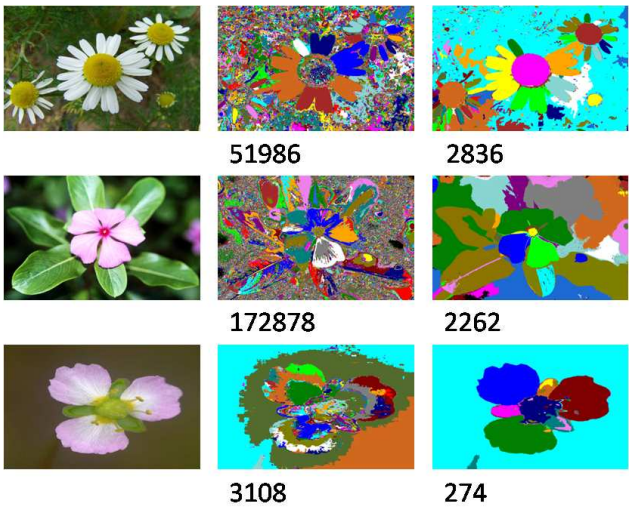

Figure 7. *Some results on images of plants taken in the wild. The labeling of regions from the segmentation by a discretized Mean Shift (center) is followed by a labeling of regions from the new GFA. The number of regions is precised under the images.*

An other application concerns image and text analysis of medieval epigraphy. The database is composed of set of images of stone carving. Objective is to study engraving characteristics and text drawing and meaning. In this application, images are grey images. So, only luminance is used for the calculation parameter $S_{i,j}$ (Eq. 1). Illustrations are given Figures 9 and 10. These ancient documents have rough surfaces and present some alter-

Figure 8. *Images of plants (left), the labeled regions using Mean Shift (center) and labellings regions after the new GFA (right)*

ations. Treatments on Figure 9 (bottom-left) shows the NFA is not able to provide efficient segmentation. To many regions have merged and regions with text have disappeared. On the contrary, the new GFA reduces the number of regions and preserves the region of interest. On Figure 10, segmentation using new GFA is followed by binarization.

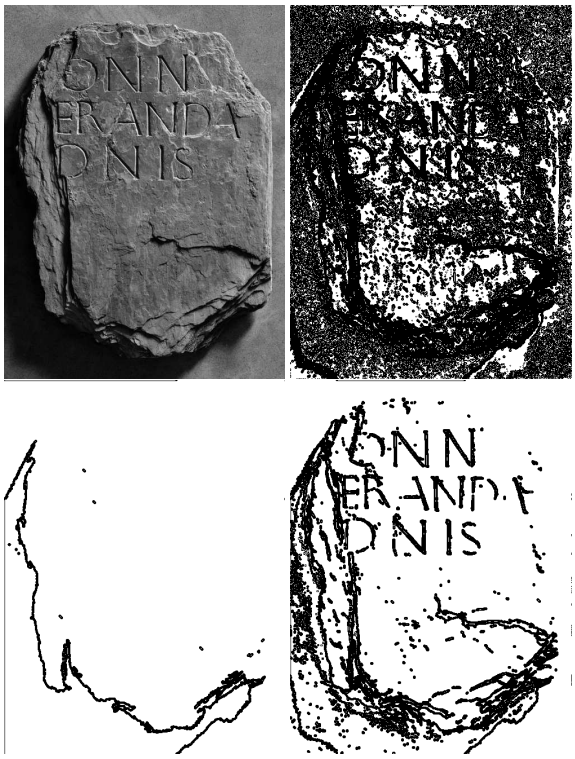

Figure 9. *Top: original epigraphy (left) and Mean-Shift segmentation (right). Bottom: NFA result (left) and new GFA result (right)*

Algorithm evaluation

In order to measure the performance of the method, we use the images of the Berkeley database [5], which includes 300 natural images and hand-labeled segmentations. This collection offers

Figure 10. *Image of text on a graphite (left) and the segmentation (right) by an mean shift and the merging from the new GFA. We made a binarization on pixels intensity denoting engraving areas*

a wide variety of possible human segmentation's and thus reflects a general trend. We evaluate the segmentation results obtained using Mean-Shift algorithm, NFA and the new GFA. To analyse the results, we used the Probabilistic Rand Index (*PRI*). This operator gives the number of pixels correctly classified according to the ground truth. The Rand Index between test and ground truth segmentation, respectively denoted S and G, is given by the sum of the number of pairs of pixels that have the same label in S and G and those that have different labels in both segmentations, divided by the total number of pairs of pixels. Variants of the Rand Index have been proposed [11, 14] for dealing with the case of multiple ground-truth segmentations. In the Berkeley database, we use human segmentations with the largest number of regions. It is within these levels of segmentation that humans revealed the most detail and the probabilistic Rand Index is defined by:

$$PRI(S,\{G_k\}) = \frac{1}{T}\sum_{i<j}[c_{ij}p_{ij} + (1-c_{ij})(1-p_{ij})] \qquad (12)$$

where c_{ij} is the event indicating that the pixels i and j in image space have the same label and p_{ij} represents this probability. Table 1 represents the PRI average values (and variance) obtained on all test images from the database. Moreover, we indicate the average number of regions in the segmented images. As observed, the all algorithms have essentially the same values of *PRI*. As expected, the PRI values for new GFA is better than the NFA's PRI values. It indicates that the proposed method increases the basic approach of Gestalt fusion algorithm. Moreover, new GFA gives results between the result of the segmentation operator that can produce over-segmentation and NFA which often produces under-segmentation. The new algorithm offers a better images partition in applications where we want to retrieve objects in a scene even if the colors are poorly nuanced.

PRI values and number of regions on Berkeley database

	PRI *average (variance)*	region number *average (variance)*
Human	0.87	36.6
Mean-Shift	0.771 (0.012)	31.9 (16.2)
NFA	0.6909 (0.0332)	17.47 (10.36)
new GFA	0.7115 (0.0232)	19.41 (11.46)

The results figure 11 show the influences of the new algorithm on some poorly nuanced images. One can see, thanks to use of the edge map information, a better conservation of all the retailers while the NFA lose these regions.

Conclusion

Natural image segmentation often produces over-segmentation that prevents a semantic analysis of a scene. There are ways to overcome this phenomenon through fusion methods. In particular, using criteria based on the Gestalt vision's theory in the goal to keep a semantic approach of problem. However, these approaches can lead to the inverse of the under-segmentation. So we study how to regularize a post segmentation process by controlling the fusion process. We proposed, in this paper, a new criterion for guiding the regions merging when analyzing images of natural scenes. We use new information coming from edge map, that is to say the edge of objects of interest in the first plan of the scene that have been focused by the photographer. Such an approach can significantly improve the segmentation results to go to a semantic natural scenes analysis. In future works it would be interesting to study the optimal value of the number of interval k to be used on regions of an image to reduce parameters to have just the edge map to manage. This parameter k has an influence on DR value depending on the size of regions contour. A region of low dimension for example, with a large k value cause a search for precision on the distribution of edge map pixels around this small region and supports their submission to a potential merger.

Acknowledgments

The authors would like to thanks Professor Cecile Treffort and Vincent Debiais from CESCM Laboratory (University of Poitiers, France) for the collaboration and discussions about Epigraphy. This research was support by Poitou-Charentes Region under the *Contrat de projets État-Région 2007-2013*.

References

[1] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17 no. 18:790–799, 1995.

[2] A. P. Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society, Series B*, 30:205–247, 1968.

[3] B. Draper, A. Colins, and J. Brolio. The schema system. *IEEE transactions on pattern Analysis and Machine Intelligence*, pages 1349–1380, 2000.

[4] A. Elif, D. Kocalar Erturk, and A. Khokhar Ashfaq. Quantized cielab space and encoded spatial structure for scalable indexing of large color image archives (2000) icassp. *IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 4:1995–1998.

[5] C. Fowlkes, D. Martin, and J. Malik. The berkeley segmentation dataset and benchmark (bsdb). *www.cs.berkeley.edu/projects/vision/grouping/segbench/*.

[6] K. Idrissi, G. Lavoué, J. Richard, and A. Baskurt. Object of interest-based visual navigation, retrieval, and semantic content identification system. *Computer Vision and Image Understanding*, pages 271–294, 2004.

[7] R. Kindermann and J. Laurie Snell. Markov random fields and their applications. *American Mathematical Society*, 1980.

[8] A. Mabmann, S. Posch, and G. Sagerer. Using markov random fields for perceptual grouping. *in proc of International Conference on Image Processing*, 2:207–210, 1997.

[9] T. Matsuyama and V. Hang. Sigma : A framework for image understanding integration of bottom-up and top-down analysis. *Plenium, New-York*, 1990.

[10] J. Nagau. *Identification de plantes dans des images numériques par champs de gradient avec une application aux plantes médicinales de la région caraïbe.* PhD thesis, Université des Antilles et de la Guyane. Campus de Fouillole., 2010.

[11] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *PAMI*, 2007.

[12] P. Vasseur, C. Pégard, and M. Mouaddib. Perceptual organization approach by dempster-schafer theory. *Pattern Recognition*, pages 1449–1462, 1999.

[13] M. Wertheimer. Principles of perceptual organization. *readings in Perception*, pages 115–135, 1958.

[14] A. Yang, J. Wright, Y. Ma, and S. Sastry. Unsupervised segmentation of natural images via lossy data compression. *CVIU*, 2008.

[15] S. Di Zenzo. A note on the gradient of multi-image. *Computer Vision, Graphics and Image Processing (CVGIP)*, 33:116–125, 1986.

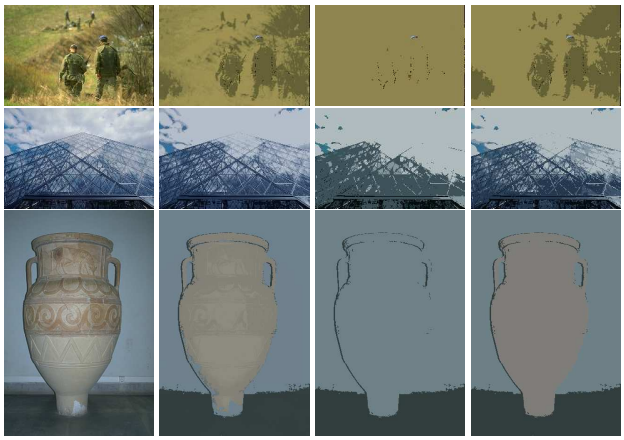

Figure 11. *Some results on images on the Berkeley database for the low contrast images. The original image (left), the Mean-shift (center left), the Gestalt fusion (center right) and our method (right)*

Author Biography

Jimmy Nagau received his Ph.D. April 15, 2010. He worked on the automatic recognition of plants under the direction of Professor Jackie Desachy from LAMIA, a French West Indies laboratory. Since October 2010, he make a post-doctorate in XLIM-SIC Laboratory, in Poitiers.

Anne-Sophie Capelle-Laizé received her PhD in Image Processing from Poitiers University, France (2003). Since 2006, she is associated Professor at the University of Poitiers, France. Her work focuses on color image processing, segmentation, data fusion and imprecise data fusion.

Christine Fernandez-Maloigne is Professor at the University of Poitiers, France, since 1996, where she created a new research pole for color imaging. She is currently director of XLIM-SIC (Signal Image and Communications) laboratory, selected as one the 15 Laboratories of Excellence in the area of digital Sciences in France in 2011. She was General Chair and organizer of the first IS&T European Conference about Colour in Graphics Image and Processing (CGIV) held in Poitiers in 2002. She is also French representative of the CIE Division 8 where she manages a Technical committee about image and video compression quality assessment.

Jean-Luc Henry is assistant professor at the University of West indies and Guyana since 1997. Its fields of research relate to pattern recognition, image processing and recognition of printed characters. Its current work refers to the identification of medicinal herbs from digital images