Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2008, Article ID 489202, 11 pages
doi:10.1155/2008/489202

Research Article

A Robust Approach to Segment Desired Object Based on Salient Colors

Jérôme Da Rugna and Hubert Konik

Laboratoire LIGIV, Université Jean Monnet, Bâtiment E, 18 Rue Benoît Lauras, 42000 Saint-Etienne, France

Correspondence should be addressed to Jérôme Da Rugna, jerome.darugna@univ-st-etienne.fr

Received 13 September 2007; Revised 29 October 2007; Accepted 22 November 2007

Recommended by Alain Tremeau

This paper presents a clustering-based color segmentation method where the desired object is focused on. As classical methods suffer from a lack of robustness, salient colors appearing in the object are used to intuitively tune the algorithm. These salient colors are extracted according to a psychovisual scheme and a peak-finding step. Results on various test sequences, covering a representative set of outdoor real videos, show the improvement when compared to a simple implementation of the same K-means oriented segmentation algorithm with ad hoc parameter setting strategy and with the well-known mean-shift algorithm.

1. INTRODUCTION

Digital videos are nowadays widespread on the World Wide Web or mobile phones but, whereas text documents are self-describing, their utility suffers as they do not give any explicit description of their content. The MPEG-7 standard gives however the true-content-based representation of any video that allows manipulation and adaptation [15] but the challenge is still to develop a system that is able to segment automatically and accurately any videos.

Indeed, more precisely, in the field of new multimedia services, and more specially around the digital content creation, distribution, and services, the technology for creating clickable videos allowing the viewers to click on objects in the video and purchase products or obtain some complementary information is a real challenge. This technology supposes firstly an automatic extraction from the image of each object of interest.

Several segmentation approaches have been proposed using principally inherent motion [6, 25] or more complex information [23] in a tracking objective [24]. Moreover, the well-known semantic gap problem can be narrowed down using object ontology to define high-level concepts or using machine learning methods to associate low-level features with query concepts [12]. Only homogeneity of pixels within a region plays a role. Similarity identification is calculated over simple continuous pixel neighborhood similarity without guiding the result through a postsegmentation step based on human vision [27]. In our work, the deal is not to discuss about the tracking problem, but only to discuss on how to improve the segmentation step using some a priori information on considered objects. Furthermore, the parameters have to be few and with a clear interpretation. Besides, only the segmentation step, that is to say the low-level one, has to be analyzed. No posttreatment will be possible in order to improve the results as in [8], where color saliency is introduced, defined from average border contrast, or in [14] where a probabilistic model for the nonpurposive grouping problem is performed. In this study, we can assume that the following object will appear similar enough along the sequence. On the other hand, the lighting conditions can change during the sequence because of shadows or point of view changing for example.

The segmentation, when talking about image processing and computer vision, is one of its fundamental problems. In several approaches, the task of segmentation is divided into two parts. First part concentrates on low-level processing which can be rather implemented in computers. The second part is then provided either from a high-level processing through a more semantic processing (machine learning) or simply from a human user who will correct in order to produce the final segmentation result [12].

(a) User selection (b) Reference object

Figure 1: Selection step of the reference object. The user selects a frame in the sequence where the desired object is representative enough. He locates by hand the object to create a mask and then initiates the process.

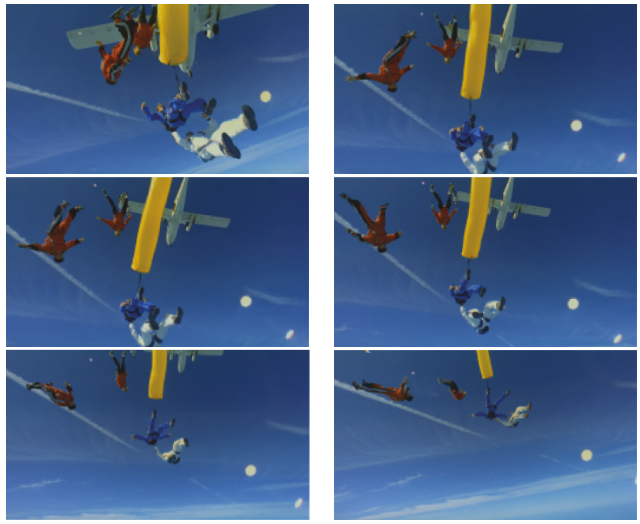


Figure 2: Some images extracted of the "sky diver" sequence. During this short cut, that lasted for about 3 seconds, the reference object, that is the blue sky diver, changes in size and in shape as well as the lighting conditions.

Primarily classified into four types: thresholding, boundary-based, region-based, and hybrid techniques [13], published low-level techniques are innumerable. Unfortunately, segmentation is still nowadays a very challenging task as no method that is effective for each color image has been developed so far. Our approach is then not to develop another method but to improve first naively, and then saliency oriented, this step in adding some features on the desired object, previously provided by the user, as illustrated in Figure 1.

This paper then discusses the robustness of segmenting general images, that is, images of any sort of scene under any illumination, where only one shot of the desired object is taken as a reference [20]. More precisely, even if the rest of the image is rawly segmented, the more robust the segmentation of the object is, the better the results are. Some illumination changes or shades can perturb the segmentation step too. Lets cite the example of the blue sky diver filmed during his drop (Figure 2). When the white sky diver is too close or when he becomes smaller and smaller, the robustness may be defective. The end goal, the tracking of the desired object during the sequence, will be improved if the segmentation result is not too sensitive and changing. Partitioning the image into a set of meaningful regions is in fact prerequisite before any analysis can be applied. The object tracking is then generally based on the visual features extracted from these regions.

Among all recent image segmentation techniques, instead of implementing all of them [3, 10, 18, 20, 27], we have focused our work on two more significant methods and classically used in the concerned context: a mean-shift-based method, called MS [7], and the K-means clustering method [4], called KM. As previously noticed, our goal deals with how to improve the results and the robustness of these methods in using some color features extracted from the desired objects. Two important properties for color features detection are repeatability, meaning that the colors should be invariant of the varying viewing conditions, and distinctiveness, meaning that they should have high discriminative power. First of all, the use of MPEG-7 dominant color descriptor (DCD) will be implemented, and to avoid an overfitting behavior, we introduce a new approach based on a perceptive saliency model [9].

Lastly, we propose different objective criteria to compare the results. Since the development of common and reasonable ones for evaluating and comparing the segmentation results performance is yet problematic [16], besides the ground-truth where the desired objects are given by some experts, our results will be compared with a classical measure introduced in [2], which integrates color and spatial distribution of the regions without requiring any user-set parameter or threshold value.

This paper is organized as follows. Section 2 reconsiders the lack of unsupervised segmentation algorithms and discusses their use considering the desired objects features. Section 3 gives an overview of our constraining algorithm introducing representative colors, while presenting some experimental illustrations in comparison with the other techniques reviewed. Finally, Section 4 concludes this paper.

2. USING THE DESIRED OBJECT TO ORIENT THE SEGMENTATION ALGORITHM

As our objective is to supervise the segmentation method, we have focused our work on a simple method where the parameters tuning seems to be logical. The clustering approach [4] permits to adapt the partition of color space in regards to the desired object. The principal idea is that adaptive histograms can represent more efficiently the distributions with much less bins. In [19], the authors proposed a clustering-based color model where the color space of the object is partitioned adaptively but with an empirical setting. In order to be more robust, the desire to automatically determine the number of bins is given as a conclusion. Before introducing a clustering-based approach, lets first introduce the objective evaluation used in this study in order to measure the improvement done.

The ill-defined nature of the segmentation problem makes actually the evaluation of any algorithm difficult. Unnikrishnan et al. [22] list three characteristics crucial for a segmentation algorithm to possess: correctness, that is the ability to produce a segmentation which agrees with ground-truth, stability with respect to parameter choice, and stability with respect to image choice. From now on, the assessment introduced in this study will rely on a heterogeneous ground-truth coupled to two objective criteria measuring the quality and the robustness of the results.

2.1. Ground-truth

Simulations have been performed to evaluate the performance of the proposed algorithm. The experiments have been carried out on different outdoor sequences, chosen for their diversity and illumination variations. The first one consists in the DCI-StEM mini movie that provides a full 2 k HD noncompressed video. The second one is the classical "coastguard" sequence, where a little boat guided by a man in red crosses a bigger one. Each frame is of size 352×288. The third (of size 1440×1080) and fourth ones (of size 1280×720) present, respectively, a skier passing near the boundary of a forest implying shadows and divers in a sunny sky with local changes of illumination conditions. These sequences are parts of the Microsoft WMV high definition content showcase, available at the company's website ("adrenaline rush" and "to the limit" sequences). The first three sequences are presented in Figure 3 while the fourth one has previously been shown in the introduction part. The temporal resolution of the test sequences is 25 images per second. Each frame has been segmented by hand with all desired objects by some experts.

2.2. Objective evaluation criteria

In the field of data clustering, different measures for evaluation have been developed; Borsotti et al. [2] proposed an empirical function $B(I)$ design for the evaluation of the segmentation results and checked for different clustering techniques:

$$B(I) = \frac{\sqrt{R}}{10000 \times (N\cdot M)} \times \sum_{i=1}^{R}\left(\frac{e_i^2}{1+\log A_i} + \left(\frac{\Psi(A_i)}{A_i}\right)^2\right), \tag{1}$$

where I is the segmented image of size $N \times M$, R is the number of regions of the segmented image, $A_{1\leq i\leq R}$ is the number of pixels of the ith region, e_i is the color error of the region i, and $\Psi(A_i)$ is the number of regions of area A_i. e_i is calculated as the sum of the distances to the region color average. In this formula, the first term is a normalization factor, the second penalizes oversegmentation, and the third term penalizes results with nonhomogeneous regions, that is to say undersegmentation.

Moreover, segmentation is only a part of a larger tracking system and the larger system will be improved if the segmentation does not misclassify objects pixels as the background. The ground-truth segmentation is available and we could evaluate the percentage of misclassified pixels (object/background) for each frame. While the entire object is important and not particularly the distribution of the regions inside it, without using an overlapping area matrix [16], the discrepancy measure is then based on a number of missegmented pixels, called OBC as object-background confusion. Let $Y = \bigcup_{j=1}^{NY_j}$ be a segmentation of the object X and $\overline{X}$ the complementary part, that is the part of the image not covered by the object X. Then the OBC coefficient is defined by

$$\text{OBC} = \frac{\sum_{j=1}^{N}(\text{Card}(Y_j \cap X) \times \delta_j)}{\text{Card}(X)} \tag{2}$$

with

$$\delta_j = \begin{cases} 1 & \text{if } \dfrac{\text{Card}(Y_j \cap \overline{X})}{\text{Card}(Y_j)} \geq t, \\ 0 & \text{else,} \end{cases} \tag{3}$$

where $\text{Card}(A)$ is the number of pixels of the region A. t is a threshold, set to 5%, that enables a region to have a small part of pixels mixed in the background without being considered as mixed.

The lower these measures are, the better the segmentation results are. The robustness of the tracking step will then depend on small values for these criteria and also on low variances favorable to a good stability.

2.3. Object oriented K-means algorithm

This classical clustering process is based on an iterative algorithm: each pixel is first allocated to initial cluster K_i with the closest cluster center using a specific distance and the main idea is to change the position of cluster centers as long as at least one of them is modified by the iteration step. Generally, dominant colors in the images create dense clusters in the color space in a natural way. Nevertheless, the results depend on the position of the initial clusters center. To avoid inherent problem of random initialization, we use an efficient partitioning of the image color space to specify initial cluster centers [28]. The authors propose a scheme based on a coarse division of the RGB color space. The initial clusters correspond to the centroids of the most representative color bins.

Considering the complexity and the color quantity of outdoor real scenes, the K-means method suffers from a lack of adaptability. Our aim is to follow an object in a video sequence with the knowledge of it. The matter of this study is to focus only on the color information without considering neither the motion nor texture or geometry information [23].

The initial step is then now to extract dominant colors that will constrain the segmentation algorithm. Considering one object, to extract the representative or dominant colors is a complex problem. First of all, we may discuss about the following question: what are these colors? Subjectively, it is commonly known that dominant colors are absolutely not unique and very relative to the person who defined them. In this paper, we will discuss about representative colors extraction only in one aim: to use these colors to refine the K-means segmentation algorithm.

MPEG-7 defined multimedia content description and specially color descriptors. The MPEG-7 committee has approved several color descriptors including the DCD [21].

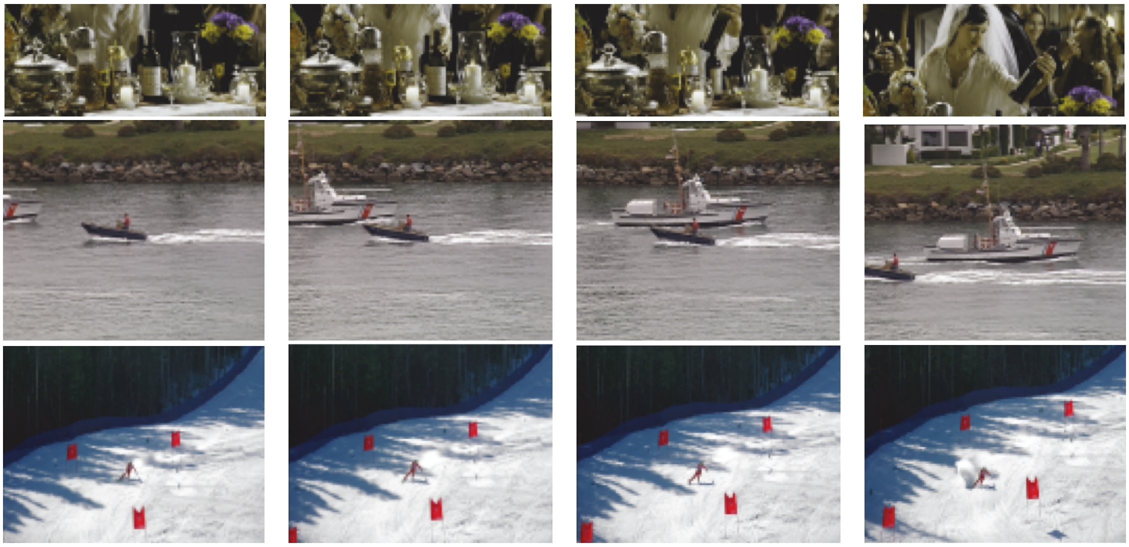


Figure 3: Some frames extracted of sequence 1, sequence 2, and sequence 3, where the reference objects are, respectively, the bottle of wine, the little boat, and the skier.

Input: A 3D Color Histogram H
Output: Significant peaks of the Histogram
Peaks $\leftarrow$ Local maxima of H
Peaks $\leftarrow$ Local maxima of Peaks
$T_\alpha \leftarrow \alpha \cdot \max(\text{Peaks})$
Peaks $\{p \in \text{Peaks};\ H(p) \geq T_\alpha\}$
foreach $(p_1, p_2) \in \text{Peaks} \times \text{Peaks}$
  if $\|p_1, p_2\| \leq \beta$
    if $H(p_1) < H(p_2)$
      Peaks $\leftarrow$ Peaks $\setminus \{p_1\}$
    else
      Peaks $\leftarrow$ Peaks $\setminus \{p_2\}$

Algorithm 1: Peak-finding algorithm.

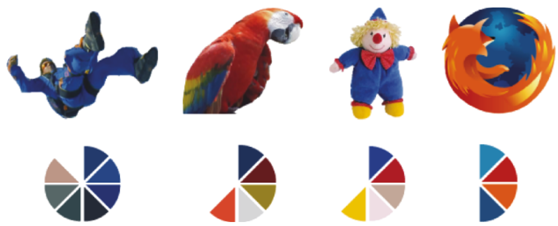


Figure 4: Some dominant colors extraction examples. The same parameters of the peak-finding algorithm are used. The variation of the number of colors depends on the method that focuses only on the histogram properties and not on a desired number of colors.

While classical techniques are low-cost, fast, and coarse privileged [11, 28], our objective is to take care of very small regions and local variations of color images. In this context, the peak-finding algorithm (see Algorithm 1) introduced in [5] by Cheng and Sun is used to identify the most significant peaks of the histogram in the RGB color space. α is a threshold used to exclude not enough representative peaks and β represents the minimum distance allowed between two peaks. The authors set α to 0.05 and β to 15.

Figure 4 illustrates some dominant colors extracted on some colorful objects.

Then, the adapted method, named ooKM for object oriented K-means, is the initial method where the dominant colors, extracted from the desired object as previously described, are added to the list of initial cluster centers. More precisely, the clusters are issued from two families: those which are obtained considering the entire image and those obtained with the initial object. We expect that object clusters, after the iterations during the K-means classification, will be attractive enough to continue in the final result.

2.4. First results

Table 1 presents the comparative results using dominant colors versus the original KM algorithm. In order to be on a level playing field between the two methods, a number of regions quasiequivalent for each method is as much as possible retained.

As regards Table 1, the values of Borsotti and OBC criteria are lower for ooKM method. But the difference is not significant enough to conclude to a superiority of this constrained approach. To explain this slight improvement, it is necessary to focus on the behavior of each method along the sequence. Figures 5 and 6 give the evolution of Borsotti and OBC criteria along the sequence 4 while using the dominant colors selected on the object first taken on frame 16 as a reference and on frame 36, respectively. Even if the results are noticeably improved around these frames, this fact is not present on the entire sequence. We are confronted to an

Table 1: Comparative results KM versus ooKM obtained with oriented approaches with test sequences. Average values and standard deviations are given. The Borsotti and #N values are computed only on the object ground-truth mask.

Criteria	Borsotti		OBC		#N	
Method	KM	ooKM	KM	ooKM	KM	ooKM
Sequence 1	3.14 ± 2.63	2.87 ± 2.42	0.06 ± 0.16	0.04 ± 0.01	9 ± 3.3	9 ± 3.1
Sequence 2	0.06 ± 0.02	0.06 ± 0.02	0.22 ± 0.22	0.21 ± 0.17	8 ± 0.2	8 ± 0.5
Sequence 3	0.50 ± 0.10	0.33 ± 0.09	0.46 ± 0.04	0.28 ± 0.11	4 ± 1.4	6 ± 0.7
Sequence 4	0.97 ± 0.37	0.96 ± 0.47	0.43 ± 0.19	0.37 ± 0.23	12 ± 2.1	10 ± 1.1

overfitting problem where the learned colors are too precise: they cannot be generalized to the complete sequence.

It can be seen from Figure 5 that around the frame where the object is extracted the difference between the KM results and the ooKM ones is larger. In fact, the clusters are preserved on the object implying better Borsotti results. On the contrary, when the dominant colors are used for segmenting frames where the lighting conditions have noticeably varied, the clusters are mixed with the background ones and the results are similar considering the two approaches. The difference likewise exists with the OBC criteria but the results seem to be less influenced.

Objectively, we can assume that the results will be improved if we select more dominant colors in order to entirely cover the object color distribution. Nevertheless, the curves presented in Figures 7 and 8 illustrate this point of view: it is possible to parameter the KM algorithm (by notably defining more seeds) to perform best results for both criteria.

These curves show the evolution of the Borsotti and OBC criteria on increasing the number of regions. The behavior is logically an improvement of these both criteria even if sometimes they rise again. The dot, representing the ooKM algorithm, seems to be a good deal between criteria results and number of regions. Indeed, our aim is to fit as best as possible the data, without creating a large amount of regions. This is first because erroneous image segmentation, that is over-segmentation, is a source of errors and difficulties in further tracking step; second because, as we have previously said, no posttreatment leading to a fusion step between adjacent regions will be used.

As a first conclusion, the naive idea to constrain the K-means clustering using dominant colors as complementary clusters is neither sufficient nor better enough compared to the KM algorithm alone.

3. OBJECT SALIENT COLORS METHODOLOGY

Extracting the dominant colors of the object in order to improve the K-means clustering has lead to a certain deadlock even in increasing the number of clusters. The aim is now to implement a saliency-based mechanism to focus the attention on a well selection of the retained colors as original clusters.

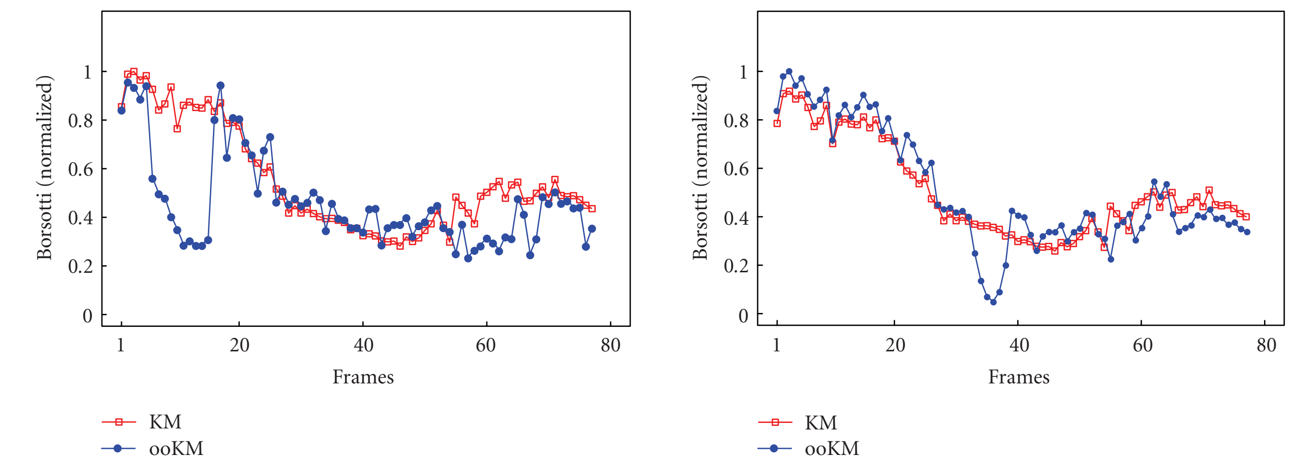
3.1. Itti model and dominant colors extraction

Itti et al. [1, 9] have proposed a model mapping the saliency of objects in the visual environment. The aim of this map is to simulate the human visual attention during the bottom-up phase using 3 kinds of features: intensity, colors, and orientations (at 0, 45, 90, and 135 degrees). Several spatial scales, computed using a Gaussian pyramid, allow to simulate human visual receptive fields: center-surround reception is implemented as the difference between two levels of the pyramid. Six-feature maps are designed 2–5, 2–6, 3–6, 3–7, 4–7, and 4–8; 2, 3, 4, 5, 6, 7, and 8 corresponding to the pyramid levels. This process, applies, respectively, to color, intensity, and orientations, and permits to compute 42 maps separated in 7 groups: intensity contrast, red/green and blue/yellow double opponent channels, and 4 encoding orientation contrasts (at 0, 45, 90, and 135 degrees). After a normalization step, all these feature maps are summed to obtain a saliency map where maxima represent the focus of attention during the bottom-up phase [17].

Figure 9 presents some salient maps obtained on different images. The maxima of intensity correspond to the focusing zones: in the second image we can estimate for example that the skier, for which a zoom is proposed, and bottom flags are clearly attracting attention.

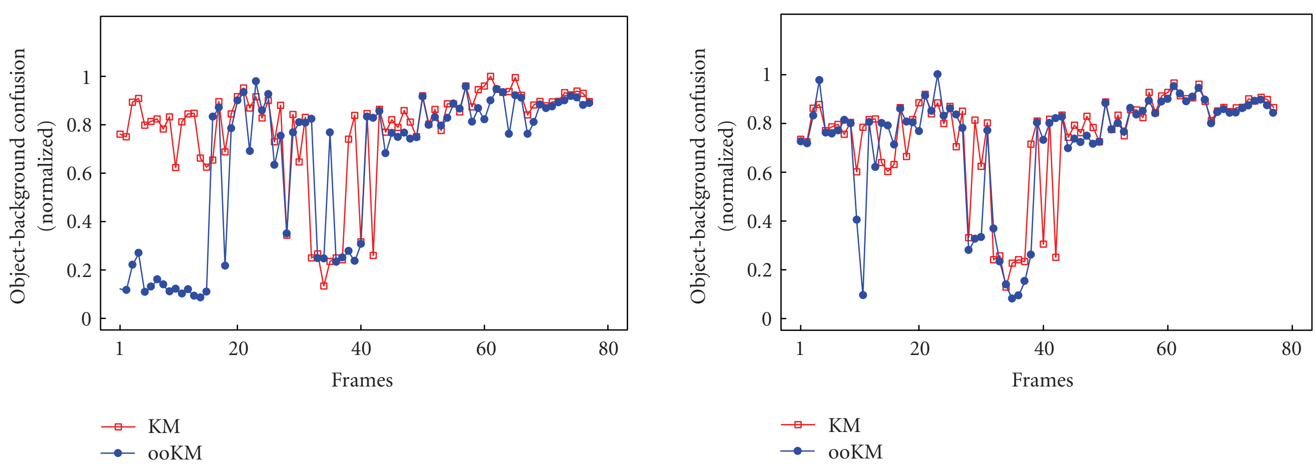
To avoid the overfitting problem issued from classical colors extraction, the basic idea is to search the representative colors not on the whole object but in two zones of it: the high-focusing one and the low-focusing one. From the visual attention point of view, they represent the low and the high frequencies. We may note here that the salient map is computed on the reference object and not in the complete image. As literature fixed the focus threshold at 0.3, we consider that any pixel whose salient value is higher than this threshold is the high-focusing pixel group. Reciprocally, we set a threshold of 0.05 to create the low-focusing pixel group.

Figure 10 shows an example of the salient colors retained on the blue sky diver object. Colors that are attractive and those that are on the contrary rather dark are automatically selected. We used the peak-finding algorithm previously presented during the dominant colors extraction process. We present in Figure 11 extraction of some salient colors from objects previously used in Figure 4. Compared to the classical dominant color extraction, this method generates colors representing main zones and small zones of the object where the classical one is more concentrated only on the main zones.



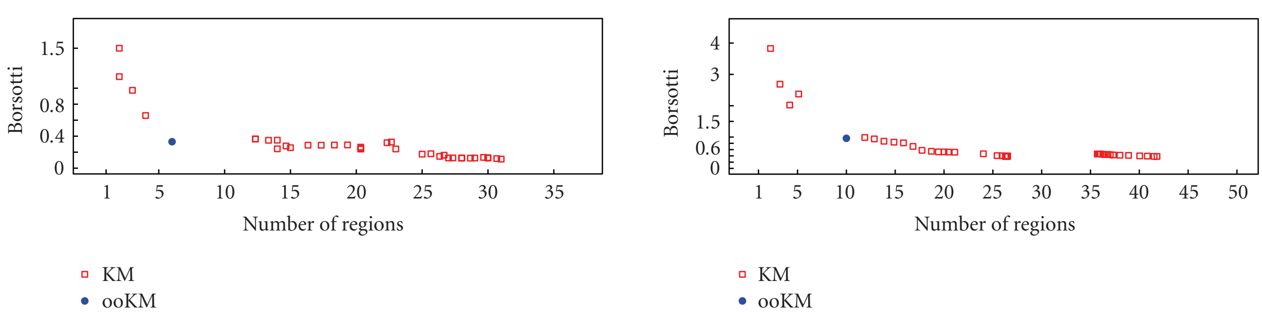
(a) (b)

Figure 5: Illustrations of the "overfitting" problem. The reference is, respectively, selected on frames 16 and 36. The figure shows the Borsotti criteria for KM and ooKM methods.



(a) (b)

Figure 6: Illustrations of the "overfitting" problem. The reference is, respectively, selected on frames 16 and 36. The figure shows the OBC criteria for KM and ooKM methods.



(a) (b)

Figure 7: Illustrations of the difficulty to reach the best deal between Borsotti optimization and number of regions in the object (sequence 3 and sequence 4). KM results are obtained by setting the number of germs from 4 to 50. The final number of regions depends on the number of clusters but there is not a strict equivalence.

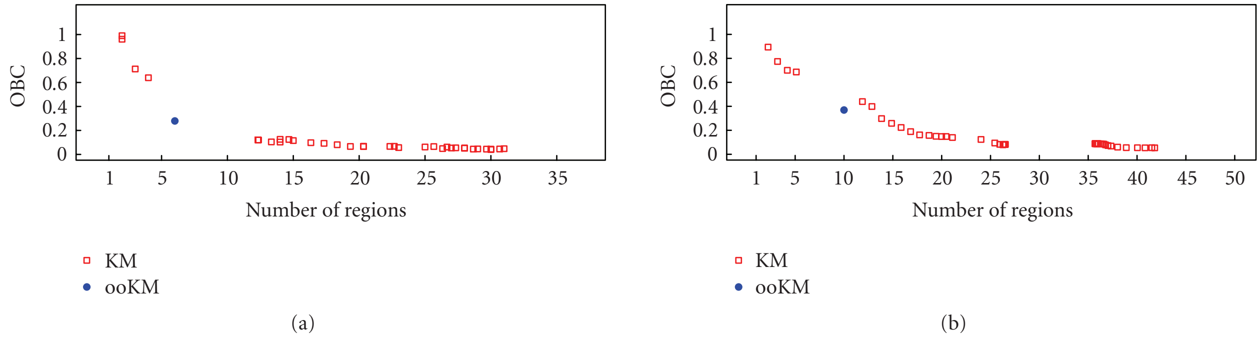


Figure 8: Illustrations of the difficulty to reach the best deal between OBC optimization and number of regions in the object (sequence 3 and sequence 4). KM results are obtained by setting the number of germs from 4 to 50.

Figure 9: Examples of salient maps. The two first maps are computed on the complete images. The last map is obtained by computing saliency only on the red skier object.

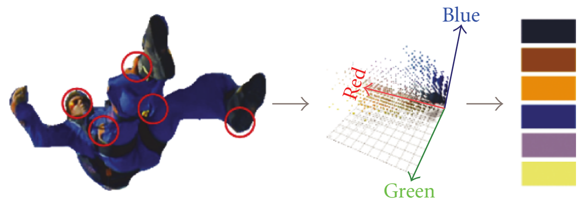


Figure 10: Principle of colors extraction based on saliency. After the thresholding in three classes of the saliency map, peaks are extracted on the color histogram with the previous algorithm to generate the final colors.

Input: n frames F_i and one object O
Output: Object-oriented segmentation of the n frames
map ← Salient-map of O
$ObjSeeds$ ← Colors extraction computed on map
foreach frame F_i
$ImgSeeds$ ← Extraction based on F_i color partitioning
$Seeds$ ← $ImgSeeds \cup ObjSeeds$
K-means segmentation of F_i using $Seeds$

Algorithm 2: soKM algorithm.

As in the ooKM methodology, the soKM method (salient-oriented KM) consists in combining the extracted colors through the saliency-map with the basic cluster seeds. Algorithm 2 resumes the overall steps of this methodology.

3.2. Results

Regarding the previous conclusion using dominant colors, lets compare now the results obtained with this saliency-based approach. First of all, the global results will be presented, second the problem of overfitting will be reconsidered, and finally the improvement according to the classical mean-shift method will be shown.

Table 2 gives the average criterion on the four sequences with ooKM versus soKM methods. For both criteria, the soKM method is more efficient than ooKM, with a noticeable improvement of the stability. Indeed, if we consider the sequence 4, where the difference between the criteria values is the less important, the standard deviation is divided by 3 for OBC and Borsotti criteria. And the lower the deviation is, the more stable the segmentation is expected to be.

Figures 12 and 13 illustrate obtained results initialized with the object contained in frame 16: the overfitting problem is not present for the soKM method. Using saliency map allows to initiate germs able to generalize the extracted colors; in this point, classical dominant color method fails.

The improvement in injecting clusters based on salient colors instead of dominant colors during the K-means algorithm has been noticed in Table 2. Compare our results with the MS method [7] used recently in color image segmentation [22, 26]. While this quite general method is used without similar prior information considered, we consider its large using in the literature as a necessary benchmark reference.

Table 2: Comparative results ooKM versus soKM obtained with oriented approaches with test sequences. Average values and standard deviations are given.

Criteria	Borsotti		OBC		#N	
Method	ooKM	soKM	ooKM	soKM	ooKM	soKM
Sequence 1	2.87 ± 2.42	0.56 ± 0.12	0.04 ± 0.01	0.01 ± 0.01	9 ± 3.1	9 ± 2.2
Sequence 2	0.06 ± 0.02	0.04 ± 0.01	0.21 ± 0.17	0.12 ± 0.09	8 ± 0.5	8 ± 0.9
Sequence 3	0.33 ± 0.09	0.28 ± 0.06	0.28 ± 0.11	0.15 ± 0.06	6 ± 0.7	6 ± 0.8
Sequence 4	0.96 ± 0.47	0.82 ± 0.14	0.37 ± 0.23	0.26 ± 0.08	10 ± 1.1	10 ± 0.8

Table 3: Comparative results MS versus ooKM obtained with oriented approaches with test sequences. Average values and standard deviations are given.

Criteria	Borsotti		OBC		#N	
Method	MS	soKM	MS	soKM	MS	soKM
Sequence 1	2.58 ± 0.34	0.56 ± 0.12	0.03 ± 0.01	0.01 ± 0.01	11 ± 1.2	9 ± 2.2
Sequence 2	0.14 ± 0.01	0.04 ± 0.01	0.27 ± 0.14	0.12 ± 0.09	7 ± 0.5	8 ± 0.9
Sequence 3	0.99 ± 0.18	0.28 ± 0.06	0.66 ± 0.08	0.15 ± 0.06	7 ± 0.8	6 ± 0.8
Sequence 4	1.36 ± 0.68	0.82 ± 0.14	0.27 ± 0.26	0.26 ± 0.08	10 ± 1.4	10 ± 0.8

Figure 11: Some salient colors extraction examples. These colors differ from the dominant colors in values as well as in number. As expected, some retained colors are not present in majority but seem to fit visual attractive colors.

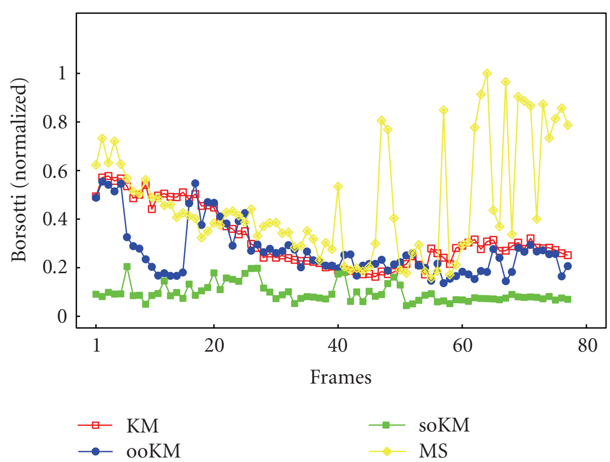


Figure 12: Results of Borsotti criterion on *sequence 4* with all segmentation methods. The blue sky diver is taken from frame 16: instead of ooKM method, the soKM one does not suffer from overfitting. MS method is not stable at the end of the sequence, where object is really small and near, in colors, to the background, that is the sky. In overall sequence, soKM gets best results in value and in variation.

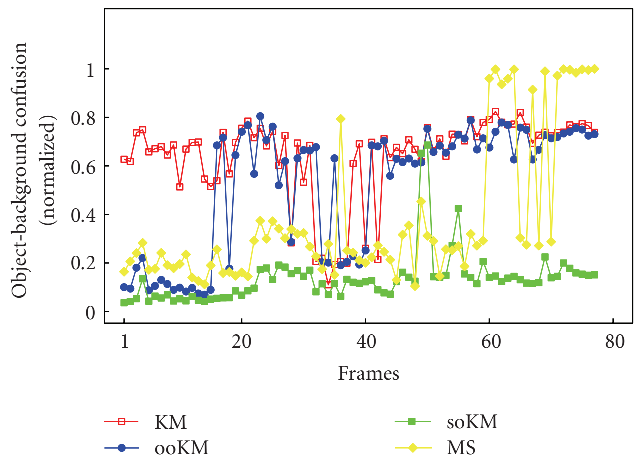


Figure 13: Results of OBC criterion on *sequence 4* with all segmentation methods. MS and soKM are comparable at the beginning of the sequence but only the soKM method is efficient at the end of it.

The results given in Table 3 confirm the efficiency of our soKM model. In fact, with similar number of regions, the soKM algorithm always leads to better results as the MS one for both criteria. Nevertheless, the MS algorithm is applied on each frame without taking into account any color information of the object.

Figures 14, 15, 16, and 17 present the stability of our method among the 4 selected entire sequences. In these graphics, the nearer the data from $(0, 0)$ are, the more efficient the method is expected to be. Thus, we first retrieve the previous results: soKM is the most stable and remains stable on all sequences.

Finally, Figure 18 gives some visual results and illustrates how the object influences the obtained segmentation. We have extracted in Figure 18(a) two sky divers: a red one and a

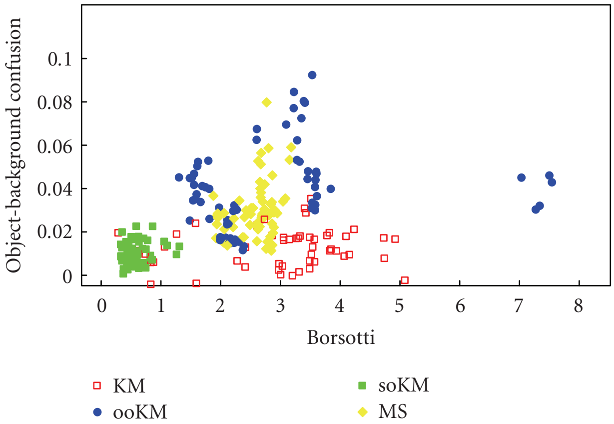


Figure 14: Results of Borsotti versus OBC on sequence 1 with all segmentation methods. This figure illustrates the stability of soKM method compared to the 3 other methods. We also retrieve the good behavior for the OBC criterion for method KM, nevertheless penalized by a high Borsotti value.

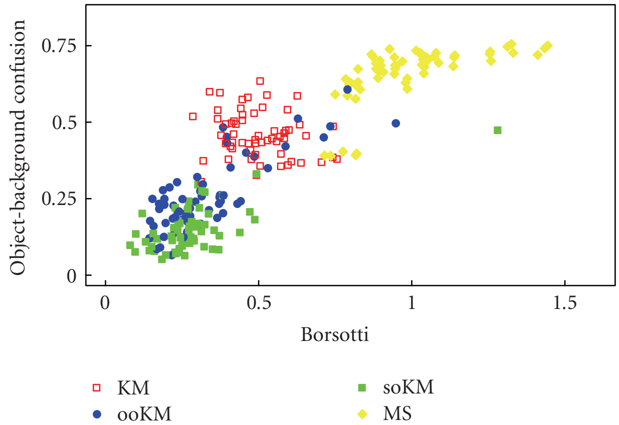


Figure 16: Results of Borsotti versus OBC on sequence 3 with all segmentation methods. ooKM and soKM reach quite same efficiency except for some frames, these ones corresponding to the "skier in shadow" event. MS seems again penalized by the few colors contained in each frame.

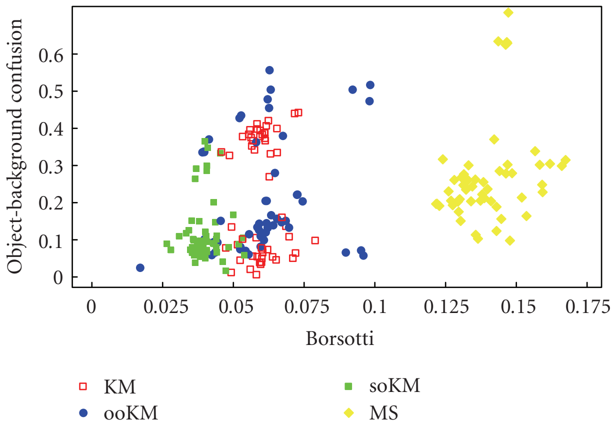


Figure 15: Results of Borsotti versus OBC on sequence 2 with all segmentation methods. MS method suffers from the poor quality of the sequence 2: KM and oriented KM methods seem more efficient considering these few colors and low-resolution frames. For the 4 methods the same behavior is present: on some frames, the OBC values are strongly increased without the same behavior on the Borsotti criteria. These frames correspond to the two boats crossing.

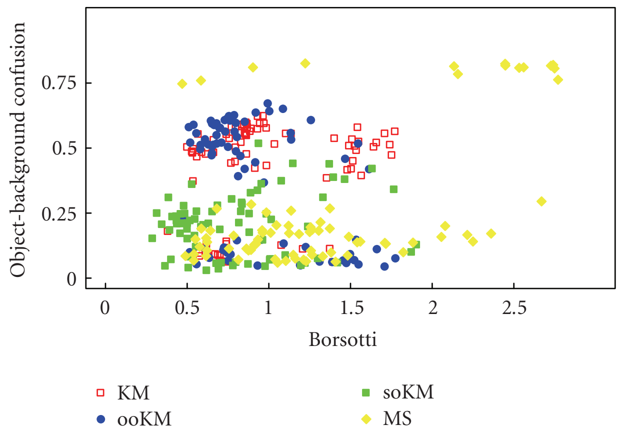


Figure 17: Results of Borsotti versus OBC on sequence 4 with all segmentation methods. We retrieve previous results: MS and soKM are quite comparable, but MS is no more efficient on some frames (the end of the sequence).

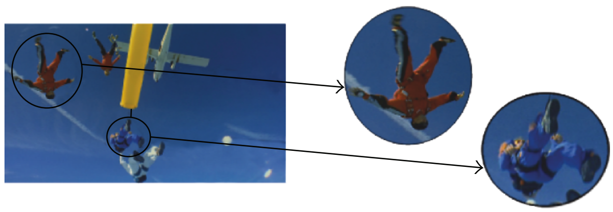
blue one. KM method gives on the red sky diver very poor results: the red color was not fitted correctly by a germ. The MS segmentation seems visually correct on the two sky divers, which was relatively expected for this method. However, the best segmentations are obtained using the soKM method in Figures 18(d) and 18(e). These examples also show how much soKM is object oriented: the other object is absolutely bad segmented.

4. CONCLUSION

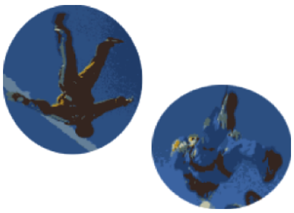
In this paper, we have presented a new strategy to tune the K-means algorithm for adaptive video segmentation. This method is only the first low-level step of a more general scheme of objects tracking in a context of content-enhancement called video clicking. In order to automatically follow a desired object chosen by the user, each step of the image processing must be optimized. Our response consists then in using available a priori knowledge on it to constrain the segmentation.

In addition to the first insufficient use of dominant colors, we have introduced a saliency-based improvement of K-means algorithm, where salient colors are coupled to primary clusters. The assessment used in this study on heterogeneous sequences (lighting conditions, view-point and geometry changes, etc.) has demonstrated a better efficiency of this model. Its generalization ability implies a noticeably better behavior both in quality and in robustness.

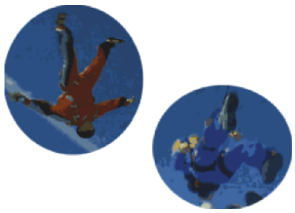
Currently, one static reference of the object is employed over the whole sequence. It is desirable to update and learn salient colors to adjust the model to sudden variations, which is our future work.



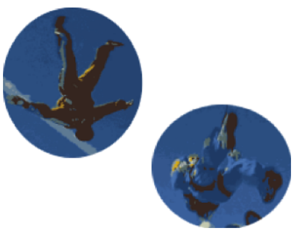
(a) Extracted frame from sequence 4



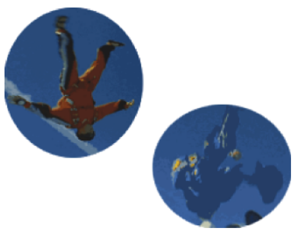
(b) KM segmentation. The red sky diver segmentation is not good enough: many details have been lost. Details seem respected in other segmentation but a part of the blue sky diver is combined with the sky



(c) MS segmentation. Inversely at the red sky diver, the blue one is badly segmented as many details do not remain and a part of the object is combined with the sky



(d) soKM, blue sky diver oriented. Blue sky diver is visually correctly segmented and correctly separated from the sky. The red one is segmented similar to KM method



(e) soKM, red sky diver oriented. Like the blue one oriented results are efficient on the red sky diver. The blue sky diver is badly segmented even far away from the KM method

FIGURE 18: Some segmentation examples on a frame of sequence 4. Two objects are considered: the red and the blue sky divers, in order to well illustrate the constraining approach according to the desired object.

ACKNOWLEDGMENT

This research was supported by the Région Rhône-Alpes, project LIMA, cluster ISLE.

REFERENCES

[1] J. Bonaiuto and L. Itti, "The use of attention and spatial information for rapid facial recognition in video," *Image and Vision Computing*, vol. 24, no. 6, pp. 557–563, 2006.

[2] M. Borsotti, P. Campadelli, and R. Schettini, "Quantitative evaluation of color image segmentation results," *Pattern Recognition Letters*, vol. 19, no. 8, pp. 741–747, 1998.

[3] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy *c*-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognition*, vol. 40, no. 3, pp. 825–838, 2007.

[4] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern Recognition*, vol. 34, no. 12, pp. 2259–2281, 2001.

[5] H.-D. Cheng and Y. Sun, "A hierarchical approach to color image segmentation using homogeneity," *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2071–2082, 2000.

[6] A. Colombari, A. Fusiello, and V. Murino, "Segmentation and tracking of multiple video objects," *Pattern Recognition*, vol. 40, no. 4, pp. 1307–1317, 2007.

[7] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[8] G. Heidemann, "Region saliency as a measure for colour segmentation stability," *Image and Vision Computing*, vol. 26, no. 2, pp. 211–227, 2008.

[9] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research*, vol. 40, no. 10-12, pp. 1489–1506, 2000.

[10] B.-K. Jeon, Y.-B. Jung, and K.-S. Hong, "Image segmentation by unsupervised sparse clustering," *Pattern Recognition Letters*, vol. 27, no. 14, pp. 1650–1664, 2006.

[11] J. Jiang, Y. Weng, and P. Li, "Dominant colour extraction in DCT domain," *Image and Vision Computing*, vol. 24, no. 12, pp. 1269–1277, 2006.

[12] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262–282, 2007.

[13] R. Lukac and K. N. Plataniotis, *Color Image Processing: Methods and Applications*, CRC Press, Boca Raton, Fla, USA, 2007.

[14] J. Luo and C.-E. Guo, "Perceptual grouping of segmented regions in color images," *Pattern Recognition*, vol. 36, no. 12, pp. 2781–2792, 2003.

[15] J. M. Martinez, R. Koenen, and F. Pereira, "MPEG-7: the generic multimedia content description standard, part 1," *IEEE Multimedia*, vol. 9, no. 2, pp. 78–87, 2002.

[16] A. Ortiz and G. Oliver, "On the use of the overlapping area matrix for image segmentation evaluation: a survey and new performance measures," *Pattern Recognition Letters*, vol. 27, no. 16, pp. 1916–1926, 2006.

[17] N. Ouerhani, R. von Wartburg, H. Hugli, and R. Muri, "Empirical validation of the saliency-based model of visual attention," *Computer Vision and Image Analysis*, vol. 3, no. 1, pp. 13–24, 2004.

[18] M. Ozden and E. Polat, "A color image segmentation approach for content-based image retrieval," *Pattern Recognition*, vol. 40, no. 4, pp. 1318–1325, 2007.

[19] L. Peihua, "A clustering-based color model and integral images for fast object tracking," *Signal Processing: Image Communication*, vol. 21, no. 8, pp. 676–687, 2006.

[20] L. Pi, C. Shen, F. Li, and J. Fan, "A variational formulation for segmenting desired objects in color images," *Image and Vision Computing*, vol. 25, no. 9, pp. 1414–1421, 2007.

[21] P. Salembier and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface.*, John Wiley & Sons, New York, NY, USA, 2002.

[22] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 929–944, 2007.

[23] H. Veeraraghavan, P. Schrater, and N. Papanikolopoulos, "Robust target detection and tracking through integration of motion, color, and geometry," *Computer Vision and Image Understanding*, vol. 103, no. 2, pp. 121–138, 2006.

[24] H. Wang and D. Suter, "A consensus-based method for tracking: modelling background scenario and foreground appearance," *Pattern Recognition*, vol. 40, no. 3, pp. 1091–1105, 2007.

[25] Y. Wang, K.-F. Loe, T. Tan, and J.-K. Wu, "Spatiotemporal video segmentation based on graphical models," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 937–947, 2005.

[26] Y. Wang, J. Yang, and N. Peng, "Unsupervised color-texture segmentation based on soft criterion with adaptive mean-shift clustering," *Pattern Recognition Letters*, vol. 27, no. 5, pp. 386–392, 2006.

[27] A. v. Wangenheim, R. F. Bertoldi, D. D. Abdala, and M. M. Richter, "Color image segmentation guided by a color gradient network," *Pattern Recognition Letters*, vol. 28, no. 13, pp. 1795–1803, 2007.

[28] N. Yang, C. M. Kuo, W. Chang, and T. Li, "A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval," *Journal of Visual Communication and Image Representation*, vol. 19, no. 2, pp. 92–105, 2008.