



Color object segmentation and tracking using flexible statistical model and level-set

Sami Bourouis^{1,2} · Ines Channoufi² · Roobaea Alroobaea¹ · Saeed Rubaiee³ · Murad Andejany³ · Nizar Bouguila⁴

Received: 6 April 2020 / Revised: 22 July 2020 / Accepted: 2 September 2020 /

Published online 8 October 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

This study presents an unsupervised novel algorithm for color image segmentation, object detection and tracking based on unsupervised learning step followed with a post processing step implemented with a variational active contour. Flexible learning method of a finite mixture of bounded generalized Gaussian distributions using the Minimum Message Length (MML) principle is developed to cope with the complexity of color images modeling. We deal here simultaneously with the issues of data-model fitting, determining automatically the optimal number of classes and selecting relevant features. Indeed, a feature selection step based on MML is implemented to eliminate uninformative features and therefore improving the algorithm's performance. For model's parameters estimation, the maximum likelihood (ML) was investigated and conducted via expectation maximization (EM) algorithm. The obtained object boundaries in the first step are tracked on each frame of a given sequence using a geometric level-set approach. The implementation has the advantage to help in improving the computational efficiency in high-dimensional spaces. We demonstrate the effectiveness of the developed segmentation method through several experiments. Obtained results reveal that our approach is able to achieve higher precision as compared to several other methods for color image segmentation and object tracking.

Keywords Color image segmentation · Object tracking · Mixture bounded model · Feature selection · Minimum message length · Level-set

1 Introduction

One of the most challenging problems in pattern recognition and image processing applications is object detection and tracking. Object detection and tracking is now widely used in video surveillance systems, smart vehicles, traffic monitoring and human computer interaction. To deal with this problem, we always proceed to involve more visual information

✉ Sami Bourouis
s.bourouis@tu.edu.sa

(characteristics) into the segmentation algorithm. Visual pixel's characteristics can be either informative or uninformative. In case of presence of uninformative information (like noise), the object detection process will be complex and extremely time consuming. Also, the presence of nonuniform illumination or self-shadow can generate false clusters and therefore all these matters may easily conduct to over-segmentation. On the other hand, taking into account all possible visual features (color, texture, shape, etc.) may decrease the algorithm's performance as cited in [17, 28, 29, 37, 41, 43, 54, 57, 58, 61]. For all these reasons, it is better to not consider irrelevant features in order to increase the detection accuracy. Recently, various distinguished scientific works involving computer vision techniques have been proposed to address the above-mentioned issues and particularly to deal with complicated real data sets. In particular, the class of approaches named "finite mixture models (FMM)" have been suggested as a potential alternative for complex data modelling and clustering. Finite mixture models represent one of the most crucial machine learning methods in the literature and are at the heart of several image processing and pattern recognition algorithms [18, 47, 52, 53]. Many probability distributions (mixtures) have been suggested in the state of the art. Based on the selected density function, the corresponding model enforces its structure to the treated image or video. It should be noted that there is no unique model is capable to cope with all possible data-shapes in a given dataset. Thus, it is important that the choice of any model (i.e density function) must consider the nature of the dataset. Even though conventional Gaussian mixture models have been broadly employed in different areas, recent publications have illustrated that other alternatives could offer superior performances in terms of data modeling and segmentation especially when dealing with non-Gaussian data. Moreover, the efficacy of such models depends on their ability in automatically determining the number of components (classes or objects) for a given image. Although finite unbounded mixtures (e.g. unbounded Gaussian mixture) have been successfully applied in data modelling given their advantage in terms of data approximation, other more flexible mixture variants notably the so-named "bounded Gaussian mixtures" were presented as an attractive choice for data analysis. In this paper, we focus on effective multidimensional (color-textured) data modeling and segmentation for which bounded mixture models have demonstrated efficiency in many applications. Indeed, bounded support mixtures offer an alternative for many other models such as Laplace mixture model (LMM), Gaussian mixture (GMM), generalized Gaussian mixture (GGMM) and bounded Gaussian mixture (BGMM) as special cases. Compared with classic unbounded Gaussian distribution, the non-Gaussian distributions such as bounded generalized Gaussian distribution (BGGMM) could offer more flexibility. In this present work, we are mainly motivated by the flexibility of this class of statistical models. Thus, the main target of this work is to propose an unsupervised learning method based on BGGMM and permits simultaneous segmenting color-textured images and selecting only relevant features which is very essential for multidimensional data modeling.

The rest of this manuscript is structured as follows. In Section 2 we introduce a state of the art related to the current context of object detection and tracking in image and video sequences. Section 3 describes the proposed segmentation approach and introduces the main formalism of our statistical feature selection mixture model. In Section 4, we introduce a post processing step with level-set approach for object tracking purpose. In Section 5, extensive experiments are conducted and discussed to show the robustness of the developed approach. Finally, Section 6 concludes the paper.

2 Related research work

Object detection and segmentation in images or video sequences are two closely and important steps for many applications including video surveillance, medical image analysis, face detection, smart vehicle, traffic monitoring, and pattern recognition applications. In the literature, these two tasks have been extensively studied and a lot of interesting papers have been proposed to address them. A possible categorization of existing approaches may be those based on machine learning or not-based. In particular, variational active contours are widely used in this area [7, 11, 22, 64]. Pattern classification techniques have also been broadly explored for object detection [3, 24, 31, 41, 43, 57, 58, 61]. For instance, deep learning-based approaches have been exposed recently as a good solution since they have the benefit to learn large scale data, making them a worthy tool for analyzing massive data. For object detection application, a single convolutional network can be used for example to predict both multiple bounding boxes and also their class probabilities simultaneously [58]. The coordinates of bounding boxes are predicted using fully connected layers on the convolutional feature extractor. Later, this algorithm has been enhanced by [57] where a real time object detection system named (YOLOv2) was developed to detect several object categories (about 9000). Another method called (Single Shot Detector: SSD) has been implemented in [43]. The advantage of this method is that it does not require features for bounding box however it permits encapsulating all stages in one network. Another recent work called (Feature Pyramid Network) was proposed in [41] which is based on a pyramid structure, a sliding windows step and a Fast R-CNN algorithm. Convolutional neural networks (CNN) was also applied in [61] and a deconvolution based CNN model and a corner based ROI estimator have been integrated to form a single CNN based detection model. Region proposals are combined with CNN in order to improve object detection and to overcome classic CNN drawbacks. The developed method is referred as “R-CNN” [31]. Later, this method was extended as “Fast R-CNN” [30] in which the training and testing speed was improved and also the detection accuracy was increased.

It is noteworthy that in object detection process, the more information is provided the better result can be obtained in term of accuracy. Taking into account that certain features are more relevant than others and therefore integrating only these informative features into such algorithm allows better detection of specific regions (or objects). The irrelevant characteristics can be only noise, thus not participating to effectively segment the desired objects. Thus, feature selection step is particularly important and plays a primary role in improving the accuracy of object detection algorithms and reducing the processing time especially when the data sets contain a lot of features. A comprehensive literature review show that the step of extracting and selecting relevant features has been investigated broadly for several related applications and it is addressed with different manners. In fact, feature selection can be used as a preprocessing step or integrated within the classifier. The diversity of methods and applications using the feature selection step is a good indicator of its importance, as we will show through some relevant published works.

Selecting informative features using Fisher information criterion is used for measuring the uncertainty of the classifier and for providing more effective real-time object tracking [66]. The latter has been extended in [65] and it was enhanced by taking into account prior knowledge. Another work, related to medical context, was developed for detecting and screening diseases in capsule endoscopy [60]. It involves the following steps: extracting and selecting relevant features then classifying images. In [63] a supervised learning approach was developed for discriminative regional feature integration. Support vector machines have

been also used in [20]. The proposed model is able to automatically exclude useless features from the feature pool. Recently, the authors in [35] proposed a robust feature selection mechanism to deal with image classification which identifies a set of mixed visual features from color spaces. The selection step is performed according to the least entropy of pixels frequency histogram distribution. The work [39] integrated cluster analysis and sparse structural learning into a joint framework for feature selection. They investigated a nonnegative spectral clustering and the hidden structure of features to learn an accurate clustering. The developed framework is formulated and optimized via an iterative algorithm. Another interesting work is developed in [38], where authors proposed a new scheme to identify and select the most useful and discriminative features based on the so-called nonnegative spectral clustering and redundancy analysis. The redundancy between the selected features is also exploited. The proposed formulation is optimized via an objective function through an iterative algorithm. In [33], a texture-based feature selection technique is proposed for segmenting fundus images. Texture features are derived from different statistical image descriptors. Another recent work shows the importance of feature selection via the implementation of a tensor based image segmentation algorithm [34]. Indeed, specific objects are distinguished and characterized on the basis of their characteristics (shape, colour, texture).

In addition to the approaches mentioned above, other interesting algorithms were applied with success to tackle the issue of unsupervised feature selection and object detection notably those based on model-based approaches [2, 51]. Model-based approaches are extensively applied for data classification which is a critical step in many applications of practical importance such as pattern recognition applications [4, 13, 15, 19], medical image analysis [14] and intrusion detection systems [1]. For instance, in [2] authors combines both the unbounded generalized Gaussian mixture model and feature selection for image segmentation. Although these models (i.e unbounded Gaussian and generalized Gaussian) are capable to offer good results, their limitation is that the underlying distribution is not bounded and their support range is $(-\infty, +\infty)$ which can be an obstacle to achieve high performances. In fact, in real applications, it is crucial to select the appropriate model for the data such as the pixel's values which belong to $[0, 255]$ and not to $]-\infty, +\infty[$. Indeed, most source's supports are bounded and it is therefore important to consider this assumption as useful for the challenges of data modeling. Motivated by the promising results obtained with the bounded mixture models for some applications like in speech modelling [42] and image denoising [18], we propose in this work to develop a novel unsupervised flexible bounded mixture model to address the aforementioned challenging problems. In particular, we propose to investigate the flexibility of bounded generalized Gaussian mixture models to fit different data-shapes which have (bounded models) the advantage to model observed images using different supports. We are also motivated by selecting and weighting only most relevant data-features. Thus, by estimating automatically and simultaneously model's parameters, feature weights, and the number of clusters, we can achieve more competitive results. For tracking purpose, we tackle this problem by following the developed mixture model with a post-processing step implemented with a variational active contour.

3 Proposed method for image segmentation and feature selection

The proposed method for simultaneously segmenting color-texture images and selecting relevant visual features is established in an unsupervised manner. The problem of feature selection is reformulated as a problem of model selection as proposed in [15, 36]. Indeed,

instead of identifying some characteristics, we assess a saliency measure for each characteristic named “feature saliency”. This process is performed through the Maximization Estimation (EM) algorithm. For such situation, one must avoid the case that all saliencies have the maximum value. This process is performed by integrating the principle of the minimum message length (MML) [62] in our mixture model algorithm. The MML criterion encourages the saliency of irrelevant features to reach the value of zero, which will allow us to reduce the number of total features. Thus, by integrating the process of estimating the saliency of each characteristic in the proposed algorithm, we will have a method capable of simultaneously selecting the relevant characteristics and properly segmenting the input images by determining the optimal number of components. In this section, we start by offering a useful nomenclature in Table 1 in order to simplify for the reader the understanding of the notation used in this manuscript. Then, we present our developed flexible statistical mixture model.

3.1 The model

If we have an input image X defined by d -dimensional vector and having N pixels, then we can describe this vector by a mixture model with M components as follows:

$$p(X|\Theta) = \sum_{j=1}^M \pi_j \prod_{l=1}^d f(x_l|\theta_{jl}) \tag{1}$$

where $f(x_l|\theta_{jl})$ represents the probability density of the feature l in the component j . For the case of a mixture of bounded generalized Gaussian distributions [18] with M components, the complete likelihood is expressed as:

$$p(X|\Theta) = \prod_{i=1}^N \sum_{j=1}^M \pi_j \frac{f_{ggd}(X_i|\theta_j)H(X_i|\Omega_j)}{\int_{\partial_j} f_{ggd}(X_i|\theta_j)dy} \tag{2}$$

where $H(X_i|\Omega_j) = \begin{cases} 1 & \text{if } X_i \in \partial_j \\ 0 & \text{Otherwise} \end{cases}$ is an indicator function able to define ∂_j indicates the bounded support region in \mathfrak{R} for each Ω_j and the density function f_{ggd} is the generalized

Table 1 Nomenclature related to the proposed method

GMM	Gaussian Mixture model
GGMM	Generalized Gaussian Mixture Model
BGGMM	Bounded Generalized Gaussian Mixture Model
BGGMM+FS	BGGMM with Feature Selection
N	Number of pixels
M	Number of components
π_j	Mixing proportions
θ_j	Model’s parameters for the j th component
λ	Shape parameter
Θ	Model’s parameters for the mixture
μ	Mean parameter
σ	Variance parameter

Gaussian distribution. This probability density is defined as:

$$f_{ggd}(X_i|\theta_j) = \mathcal{A}(\lambda_j) \exp\left(-\left[\frac{\Gamma(3/\lambda_j)}{\Gamma(1/\lambda_j)}\right]^{\lambda_j/2} \left|\frac{X_i - \mu_j}{\sigma_j}\right|^{\lambda_j}\right) \tag{3}$$

where: $\mathcal{A}(\lambda_j) = \frac{\lambda_j \sqrt{\frac{\Gamma(3/\lambda_j)}{\Gamma(1/\lambda_j)}}}{2\sigma_j \Gamma(1/\lambda_j)}$ and $\Gamma(\cdot)$ is the Gamma function given by: $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, $x > 0$.

Finally, in order to characterize the developed mixture model, we propose the following complete set of parameters:

$$\Theta = \{\mu_1, \dots, \mu_M, \sigma_1, \dots, \sigma_M, \lambda_1, \dots, \lambda_M, \pi_1, \dots, \pi_M\}.$$

Thus, each component (as illustrated in (2)) models the data with different supports Ω_j . The objective of the segmentation consists in assigning each pixel x_i to one region (or component) of the given image. This assignment is defined through a latent variable $z_i = (z_{i1}, z_{i2}, \dots, z_{iM})$ such as $z_{ij} \in \{0, 1\}$ and $\sum_{j=1}^M z_{ij} = 1$. Therefore, the conditional distribution of x_i given the label z_i is defined as follows:

$$p(x_i|z_i, \Theta) = \prod_{j=1}^M \left(\prod_{l=1}^d p(x_{il}|\theta_{jl}) \right)^{z_{ij}} \tag{4}$$

where $\theta_{jl} = (\lambda_{jl}, \mu_{jl}, \sigma_{jl})$, and $p(x_{il}|\theta_{jl})$ is the bounded generalized Gaussian distribution (BGGD). In this case, these parameters are the missing information. Given the set of parameters Θ , the missing information represented by the parameters z_{ij} can be identified by applying the Bayes theorem as follow:

$$p(z_{ij}|x_i, \Theta) = p(x_i, z_{ij}|\Theta) / p(x_i|\Theta) \propto p_j p(x_i|z_{ij}, \Theta) \tag{5}$$

where $p_j = p(z_{ij})$. Therefore, for M components in the mixture, the primary segmentation step consists in estimating the optimal parameters $\Theta^* = (p_j^*, \theta_{jl}^*)$, with $j = 1 \dots M$ and $l = 1 \dots d$. On the other hand, when dealing with images having various features like color and texture, it is clear that their importance and contribution is not the same for discriminating pixels. The distribution of some of these features can be independent of the different regions. We can found for example noise which can make the modeling task more complex and can lead to false classification.

Thus, we propose to extend the BGGMM model by taking into account the importance of each feature separately. The feature irrelevance can be defined as follows: features are considered irrelevant if they have a common density, $p(x_{il}|\varphi_l)$, in all model's components. Let ϕ_l a binary variable such as:

$$\phi_l = \begin{cases} 0 & \text{if then } l\text{th feature is irrelevant} \\ 1 & \text{Otherwise} \end{cases}$$

Thus, each x_{il} has the following distribution [15, 36]:

$$p(x_{il}|\theta_{jl}^*, \varphi_l^*, \phi_l) \simeq (p(x_{il}|\theta_{jl}))^{\phi_l} (p(x_{il}|\varphi_l))^{1-\phi_l} \tag{6}$$

where $\Theta^* = (p_j^*, \theta_{jl}^*, \varphi_l^*)$. The superscript star indicates the unknown real distribution of the feature l ; $p(x_{il}|\theta_{jl})$ and $p(x_{il}|\varphi_l)$ are both univariate bounded generalized Gaussian distributions. Based on this equation, we notice that the underlying mixture model could lead to false positives, that is to say non-informative features can be considered relevant [15]. To deal with this issue, we propose to generalize the definition of the features pertinence by taking into account the following component $p(\cdot|\varphi_l)$ as irrelevant one and being

a mixture of BGGD independent of the region label (z_i). Our choice is encouraged by the flexibility of the proposed mixture to found close arbitrary distribution for the irrelevant characteristics [10].

Let define K , the total number of components of this mixture defined by $(\varphi_{11}, \dots, \varphi_{Kl})$. Let $w_{ilk} \in \{0, 1\}$, where $\sum_{k=1}^K w_{ilk} = 1$, the label associated to an irrelevant feature, where $w_{ilk} = 1$ if x_{il} is generated by the component k of this mixture and 0 otherwise. Let Φ be the set of all ϕ_l . From this, it is simple to demonstrate that the (4) can be redefined as:

$$p(x_i|z_i, \Phi, \{w_{il}\}, \Theta) = \prod_{j=1}^M \left(\prod_{l=1}^d (p(x_{il}|\theta_{jl}))^{\phi_l} \left(\prod_{k=1}^K (p(x_{il}|\varphi_{kl}))^{w_{ilk}} \right)^{1-\phi_l} \right)^{z_{ij}} \tag{7}$$

where z_i , Φ and $\{w_{il}\}$ are the missing information. Thus, it is practical to marginalize the complete likelihood function $p(x_i|z_i, \Phi, \{w_{il}\}, \Theta)$ with respect to the variables z_i , Φ and $\{w_{il}\}$. This is used to define the likelihood of the observations $p(x_i, \Theta)$. To reach this objective, we must first define the following prior distributions:

$$\begin{aligned} p(z_i) &= \prod_{j=1}^M p_j^{z_{ij}}, \\ p(\Phi) &= \prod_{l=1}^d \rho_{l1}^{\phi_l} \rho_{l2}^{1-\phi_l}, \\ p(w_{il}|\phi_l) &= \prod_{k=1}^K ((\pi_{kl})^{w_{ilk}})^{1-\phi_l} \end{aligned} \tag{8}$$

where $\rho_{l1} = p(\phi_l = 1)$ determines whether the feature l is relevant or not. On the other hand, $\rho_{l2} = p(\phi_l = 0)$ measures the irrelevance of l . In this case $\rho_{l1} + \rho_{l2} = 1$. Here π_{kl} is the prior probability that x_{il} is coming from the component k given that l is considered as irrelevant feature (i.e. $\phi_l = 0$).

To estimate the set of all model’s parameters, we need to optimize an objective functional which leads to perform the segmentation of color images by considering the mechanism of visual feature selection. These parameters are denoted by $\Theta = (p, \theta_{jl}, \varphi_{kl}, \pi_l)$ where $p = (p_1, \dots, p_M)$, $\theta_{jl} = (\lambda_{jl}, \mu_{jl}, \sigma_{jl})$ and $\pi_l = (\pi_{l1}, \dots, \pi_{lK})$. Now $p(x_i|\Theta)$ is estimated by successively marginalizing the likelihood function with respect to the parameters $p(x_i|z_i, \Phi, \{w_{il}\}, \Theta)$ and we can deduce the following final model for image segmentation with feature selection mechanism:

$$p(x_i|\Theta) = \sum_{j=1}^M p_j \prod_{l=1}^d \left(\rho_{l1} p(x_{il}|\theta_{jl}) + \rho_{l2} \sum_{k=1}^K \pi_{kl} p(x_{il}|\varphi_{kl}) \right) \tag{9}$$

3.2 Model’s parameters estimation

Several techniques have been proposed to deal with the complex problem of parameters estimation for statistical mixture models [47]. In this work, we opt for the maximum likelihood estimation method [48, 56]. In practice, the expectation maximization method [27] is applied in order to estimate the parameters of the model. The log-likelihood is given by:

$$\log p(X|\Theta) = \sum_{i=1}^N \log(p(x_i|\Theta)) \tag{10}$$

The posterior probability is estimated as follows:

$$p(j|x_i) = \frac{p_j \prod_{l=1}^D [\beta_j(x_{il})]}{\sum_{j=1}^M p_j \prod_{l=1}^D [\beta_j(x_{il})]}, \tag{11}$$

where $\beta_j(x_{il}) = \rho_{l1} p(x_{il}|\theta_{jl}) + \rho_{l2} p(x_{il}|\varphi_l)$, and $p(x_{il}|\varphi_l) = \sum_{k=1}^K \pi_{kl} p(x_{il}|\varphi_{kl})$.

We derive now the equation to compute the relevance of the respective features.

$$\frac{1}{\hat{\rho}_{l1}} = 1 + \frac{\max \left(\sum_{i=1}^N \sum_{j=1}^M p(j|x_i) \frac{\rho_{l2} p(x_{il}|\varphi_l)}{\beta_j(x_{il})} - \frac{3K}{2}, 0 \right)}{\max \left(\sum_{i=1}^N \sum_{j=1}^M p(j|x_i) \frac{\rho_{l1} p(x_{il}|\theta_{jl})}{\beta_j(x_{il})} - \frac{3M}{2}, 0 \right)} \tag{12}$$

The weights p_j and π_{kl} are updated as follows:

$$p_j = \frac{\max \left(\sum_{i=1}^N p(j|x_i) - \frac{3D}{2}, 0 \right)}{\sum_{j=1}^M \max \left(\sum_{i=1}^N z_{ij} - \frac{3D}{2}, 0 \right)} \tag{13}$$

$$\pi_{kl} = \frac{\max \left(\sum_{i=1}^N \sum_{j=1}^M p(j|x_i) \frac{\rho_{l2} \pi_{kl} p(x_{il}|\varphi_{kl})}{\beta_j(x_{il})} - \frac{3}{2}, 0 \right)}{\sum_{k=1}^K \max \left(\sum_{i=1}^N \sum_{j=1}^M p(j|x_i) \frac{\rho_{l2} \pi_{kl} p(x_{il}|\varphi_{kl})}{\beta_j(x_{il})} - \frac{3}{2}, 0 \right)} \tag{14}$$

For $j = 1, \dots, M$ and $k = 1, \dots, K$ the model’s parameters are estimated as follow: First, the mean is updated according to the following equation:

$$\hat{\mu}_{jl}^\theta = \frac{\sum_{i=1}^N p(j|x_i) \frac{\rho_{l1} p(x_{il}|\theta_{jl}) |x_{il} - \mu_{jl}^\theta|^{\lambda_{jl}^\theta - 2}}{\beta_j(x_{il})} x_{il}}{\sum_{i=1}^N p(j|x_i) \frac{\rho_{l1} p(x_{il}|\theta_{jl}) |x_{il} - \mu_{jl}^\theta|^{\lambda_{jl}^\theta - 2}}{\beta_j(x_{il})}} \tag{15}$$

Then, the standard deviation parameter is computed as:

$$\hat{\sigma}_{jl}^\theta = \sqrt{\frac{\sum_{i=1}^N \frac{p(j|x_i) \rho_{l1} p(x_{il}|\theta_{jl}) \lambda_{jl}^\theta A(\lambda_{jl}^\theta) |x_{il} - \mu_{jl}^\theta|^{\lambda_{jl}^\theta}}{\beta_j(x_{il})}}{\sum_{i=1}^N \frac{p(j|x_i) \rho_{l1} p(x_{il}|\theta_{jl})}{\beta_j(x_{il})}}} \tag{16}$$

Finally, the shape parameters $\hat{\lambda}_j^\theta$ and $\hat{\lambda}_k^\varphi$ are estimated using the Newton-Raphson method as follow:

$$\hat{\lambda}_{ol}^* \simeq \hat{\lambda}_{ol}^* - \left[\frac{\partial^2 MML(M, K)}{\partial \hat{\lambda}_{ol}^{*2}} \right]^{-1} \left[\frac{\partial^2 MML(M, K)}{\partial \hat{\lambda}_{ol}^*} \right] \tag{17}$$

3.3 Optimal model selection

In our case, the problems of both relevant features and optimal model selection are solved with the minimum message length (MML) principal [62] which is able to identify the best statistical learning model with less complexity. It is noted that the weights p_j , ρ_{l1} and ϕ_{kl} for unwanted components are forced to zero. We determine the message length (MessLength) criterion as follows:

$$MessLength = -\log p(\Theta) + \frac{1}{2} \log(|I(\Theta)|) + \frac{c}{2} \left(1 + \log \frac{1}{12} \right) - \log p(X|\Theta) \tag{18}$$

where $p(\Theta)$, $p(X|\Theta)$, and $I(\Theta)$ denote the prior distribution, the likelihood and the Fisher information matrix, respectively. The constant c denotes the total number of parameters. In order to facilitate the calculation of MML, we assume the independence of the different groups of parameters. This assumption allows the factorization of $p(\Theta)$ and $|I(\Theta)|$. The Fisher information $|I(\Theta)|$ is approximated from the complete likelihood which assumes labeled observations [25]. Given that p , ρ_l and π_l are defined on the simplexes $(p_1, \dots, p_M) : \sum_{j=1}^{M-1} p_j < 1$, $(\rho_{l1}, \rho_{l2}) : \rho_{l1} < 1$, and $(\pi_{l1}, \dots, \pi_{lK}) : \sum_{k=1}^{K-1} \pi_{lk} < 1$, respectively, a natural choice is the Dirichlet distribution for conjugate prior. The hyper-parameters of these distributions are set to 0.5. The latter are defined as:

$$\begin{aligned}
 p(p) &\propto \frac{1}{\prod_{j=1}^M p_j^{1/2}} \\
 p(\rho_l) &\propto \frac{1}{\rho_{l1}^{1/2} \rho_{l2}^{1/2}} \\
 p(\pi_l) &\propto \frac{1}{\prod_{k=1}^K \pi_{kl}^{1/2}}
 \end{aligned} \tag{19}$$

The Fisher information of the θ_{jl} is approximated on the basis of the second derivatives of the minus log-likelihood of the l 'th feature. Indeed, by discarding the first order terms and substituting the prior and Fisher information in (18), the minimum message length objective to be minimized becomes:

$$\begin{aligned}
 MessLength &= -\log p(X|\Theta) + \frac{c}{2} \log N + \frac{3d}{2} \sum_{j=1}^M \log p_j \\
 &+ \frac{3}{2} \sum_{l=1}^d \sum_{k=1}^K \log \pi_{kl} + \frac{c}{2} \left(1 + \log \frac{1}{12} \right) \\
 &+ \frac{3M}{2} \sum_{l=1}^d \log \rho_{l1} + \frac{3K}{2} \sum_{l=1}^d \log \rho_{l2}
 \end{aligned}$$

3.4 Proposed algorithm/framework

The developed framework and the proposed algorithm are both summarized in Fig. 1 and Algorithm 1.

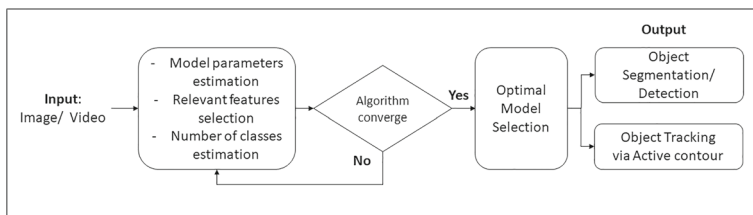


Fig. 1 Flowchart of the proposed method for the object detection/segmentation

Algorithm 1 BGGMM Model learning using EM+FS.

Input : Image designed with variable X
Output: Segmented image

- 1 **Initialization algorithm** :
- 2 Model's parameters $\Theta = (p, \theta_{jl}, \varphi_{kl}, \pi_l)$ are initialized using the K-means algorithm.
- 3 Pick an maximum value for M such as $M = M_{max}$.
- 4 **foreach** $1 \leq m \leq M_{max}$ **do**
- 5 **repeat**
- 6 {E-step} : Evaluate the posterior probability using (11).
- 7 {M-step} : Re-estimate the following parameters:
- 8 Feature relevance using (12).
- 9 Weights using (13) and (14).
- 10 Means μ_{jl} using (15).
- 11 Standard deviation σ_{jl} using (16).
- 12 Shape λ_{jl} using (17).
- 13 **until** *Algorithm converge*
- 14 Calculate the associated MML criterion.
- 15 Delete the j th component having the smallest π_j ($M = M - 1$)
- 16 **end**
- 17 Find the best matching model M^* such that: $M^* = \arg \min_M MML(M)$. Then, select the statistics for this best model.
- 18 Then input image is segented according to the selected model.

For the convergence, the log-likelihood should be evaluated through checking the convergence criterion between two successive iterations t and $(t+1)$. If the following condition is not satisfied $|\log(X|\Theta_{t+1}) - \log(X|\Theta_t)| < \epsilon$ then, the process is re-iterated from E-step where ϵ is a predefined threshold.

4 Object tracking with BGGMM and a variational-based approach

The purpose of this step is to detect speedily and accurately the contour of the object of interest in a sequence of images. Thus, we propose to apply a variational active contour which is controlled by an effective speed function (level set function). This function is derived from both local and global information. It is noteworthy that variational-based approaches are widely explored previously [7, 12, 64]. Unlike classical edge detection algorithms, variational models constitute a suitable framework able to combine heterogeneous information (e.g. local and global) and offer an effective geometrical representation for a image analysis. One of the fashionable developed model is called “level-set” [59]. The use of level-set makes it possible to avoid any possible parameterization and changes in topology are easily treated. In addition to these benefits, it has been shown also that this approach has flexibility properties especially in shape modeling and object tracking. These advantages make level-set a good alternative given its flexibility in shape segmentation and tracking. Whereas detailed proofs regarding the level-set principle are not given here, the reader can refer to [59].

The key idea behind this approach is to handle and update the displacement of the 2D curve into the motion of 3D surface. The shape of the 2D object (named as the front $\Gamma(t)$) is represented by the zero-Levelset function ϕ . ϕ is evolved by resolving the subsequent PDE

equation:

$$\frac{\partial \phi}{\partial t} = F \cdot |\nabla \phi| \tag{20}$$

where F is the speed function building on the local geometric curvature k . The symbol ∇ is used for the gradient operator.

In the literature, several level-set based speed functions (called also evolution equation) have been developed. We can categorize them into three main classes: edge-based information, region-based information, and prior-based information. The main difficulties facing when using level-set are the dependency to an accurate initialization step (i.e adequate initial active contour) and the choice of a robust speed function that guarantee the convergence of the deformable model to the optimal solution. In this work, we deal with these two subproblems by considering as initial active contour the one obtained with our developed statistical BGGMM+FS. The prior segmentation with BGGMM+FS will provide an initial contour (C) for the region of interest (ROI) for the variational model. Then, the obtained object boundaries in the first step are tracked on each frame of a given sequence (X) by using a robust level-set model proposed by Chan and Vese [16]. The proposed scenario for object tracking is depicted in Fig. 2 where both BGGMM+FS and level-set approaches are used in a cooperative scheme to detect accurately moving objects. Indeed, after a certain number p of frames, a step of boundary-detected verification is carried out with the statistical model BGGMM+FS. This will allow us to correct introduced errors by the variational model. As a result, parameters C_1 and C_2 (from the level-set equation (21)) are updated.

It is also noted that the advantage of applying the variational model is to speed up the online tracking process and to maintain high precision since, thanks to the accurate initialization-step, only a small number of iterations is needed to detect the boundaries of the object of interest. The used variational model is formulated by minimizing an energy functional which is a particular case of the Mumford-Shah formulation [50]. This function is defined as:

$$E = \lambda_1 \int_{inside(C)} |X - C_1|^2 dX + \lambda_2 \int_{outside(C)} |X - C_1|^2 dX \tag{21}$$

where C_1 and C_2 are two constants. C_1 is the average intensity inside the delimited region by the initial contour and C_1 is the average intensity outside the region. The variational level set is then reformulated as:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\mu \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \nu \right] + \delta(\phi) \left[-\lambda_1 (X - C_1)^2 + \lambda_2 (X - C_2)^2 \right] \tag{22}$$

where $\delta(x)$ is the derivative of the Heaviside function (i.e the Dirac mass function), $\mu \geq 0$, $\nu \geq 0$, $\lambda_1 > 0$, $\lambda_2 > 0$ are prefixed parameters. ν amplifies the propagation speed; λ_1 and λ_2 derive the image force inside and outside the contour, μ controls the smoothness of the level set model.

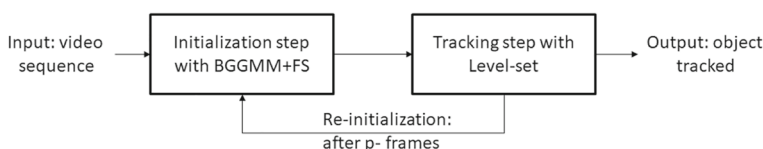


Fig. 2 Tracking step based on BGGMM+FS and Level-set approaches

5 Experiments and results

Our purpose here is to evaluate the effectiveness of the proposed method “bounded generalized Gaussian mixture model with feature selection” that we refer to as (BGGMM+FS). Other results for the problem of tracking obtained using BGGMM+FS and level set (LS) are also presented. We propose to compare the obtained results with respect to those offered by other models and methods such as GMM, GGMM and BGMM. The experiments were carried out on several real-world color-textured images. To deal with the convergence issue, we use two criteria: a threshold value (set to 0.01) that assesses the parameters’s difference between two successive iterations and also a maximum number of iterations.

5.1 Experiment 1: color-texture image segmentation

Images and their ground truth used in this section are offered by the well known Berkeley benchmark [45]. Indeed, each pixel (i, j) is modelled by a vector of several features $x(i, j)$. This vector includes both color and texture characteristics. We opt here for 19 features given as: 3 color characteristics calculated from from the RGB color space and the remaining 16 features describe the texture content of the image. They are obtained from the color correlogram matrix (CC) [32].

The entry of this matrix account for the probability that a pixel x_2 at distance d and orientation θ from a pixel x_1 . In our experiments, we calculated the correlogram matrix for 4 different orientations such as $\theta = [0, \pi/4, \pi/2, 3\pi/4]$. Texture features are mainly evaluated as follow [2]:

– Energy (EN)

$$EN(d, \phi) = \sum_{c_i, c_j} (C^{d, \phi}(c_i; c_j))^2$$

– Entropy (ET)

$$ET(d, \phi) = \sum_{c_i, c_j} -C^{d, \phi}(c_i; c_j) \log(C^{d, \phi}(c_i; c_j))$$

– Inverse-Difference-Moment (IDM)

$$IDM(d, \phi) = \sum_{c_i, c_j} \frac{1}{1 + \|c_i - c_j\|^2} C^{d, \phi}(c_i; c_j)$$

– Correlation (C)

$$C(d, \phi) = \sum_{c_i, c_j} \frac{(c_i - M_x)(c_j - M_y)^T}{|\sum_x| |\sum_y|} C^{d, \phi}(c_i; c_j)$$

where

$$M_x(d, \phi) = \sum_{c_i} c_i \sum_{c_j} C^{d, \phi}(c_i; c_j),$$

$$M_y(d, \phi) = \sum_{c_j} c_j \sum_{c_i} C^{d, \phi}(c_i; c_j),$$

$$\sum_x(d, \phi) = \sum_{c_i} (c_i - M_x)^T (c_i - M_x) \sum_{c_j} C^{d, \phi}(c_i; c_j),$$

$$\sum_y(d, \phi) = \sum_{c_j} (c_j - M_y)^T (c_j - M_y) \sum_{c_i} C^{d, \phi}(c_i; c_j),$$

$|\sum_x|$ and $|\sum_y|$ are the determinant of the matrices \sum_x and \sum_y .

Quantitative performances are obtained based on 500 images provided from the Berkeley segmentation database (BSD) [6]. All images are provided with their manual segmentations for validation purpose. Performance are evaluated on the basis of the following metrics: accuracy, sensitivity, specificity, recall, MCC (Matthews correlation coefficient), F1-measure and the boundary displacement error (BDE) [26]. These measures are often applied by the image segmentation research community to assess the segmentation output. Figure 4 shows the segmentation output for some images selected from the Berkeley database.

The convergence test is based on the stabilization of the parameters and the log-likelihood function. We have not noticed a problem related to the convergence of the EM algorithm although we are not sure obviously that we converge to a global maximum which is a common problem with EM. Indeed, in Fig. 3, which represents the log-likelihood function as a function of the number of iterations, we show that the log-likelihood does not change much after a certain number of iterations. Thus, after 20 iterations, the log likelihood stabilizes and then the learning algorithm converges (Fig. 4).

A comparative study is also given in Tables 2 and 3. It represents the average performance for the Berkeley benchmark database. Accordingly, some interesting conclusions can be deduced: first, BGGMM+FS is able to offer very encouraging results. It outperforms other conventional Gaussian-based models and other methods from the literature (Table 3). Furthermore, both BGGMM+FS and GGMM+FS are capable to offer better accuracy. This is due to the importance of considering only relevant features if we want to enhance the expected results in term of segmentation accuracy. If we look at both Tables 2 and 3, the obtained values for the accuracy metrics are 92.10% for GMM, 93.26% for GMM+FS, 93.87% for GGMM, 95.03% for GGMM+FS, 94.98% for BGGMM, and 96.68% for

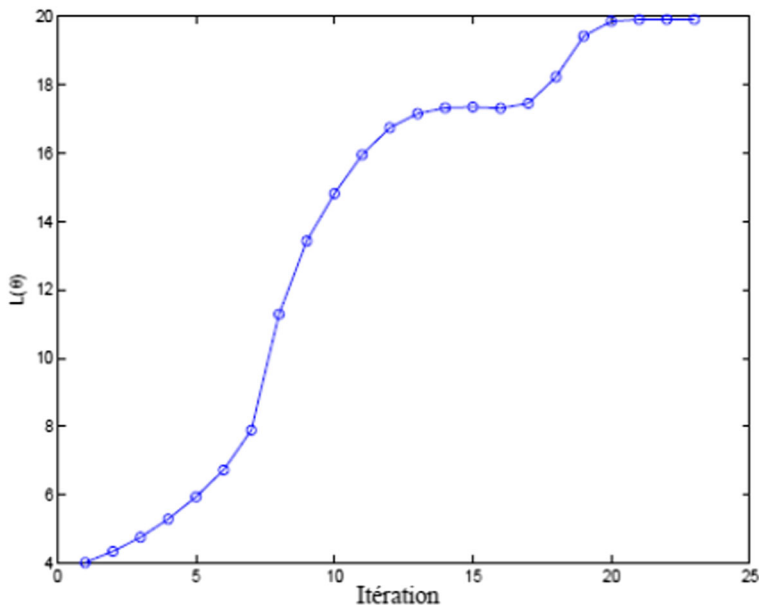


Fig. 3 The log-likelihood function as a function of the number of iterations shows that the log-likelihood stabilizes after 20 iterations and then the learning algorithm converges

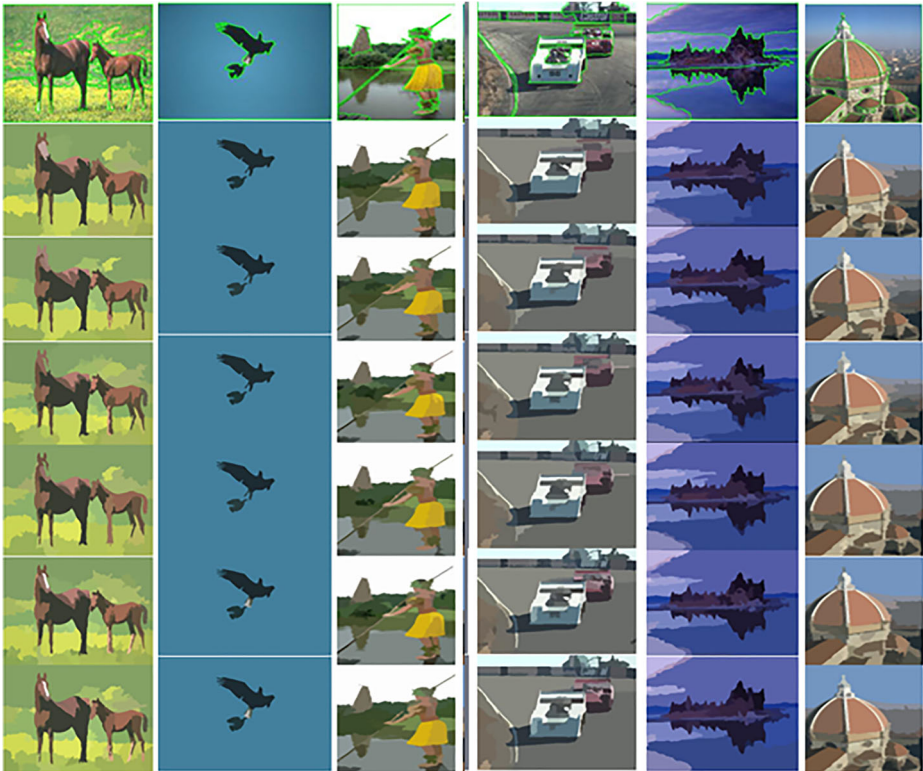


Fig. 4 Image Segmentation Results; First row: Ground Truth, second row: GMM, third row: GMM+FS, fourth row: GGMM, fifth row: GGMM+FS, sixth row: BGGMM, and seventh row: BGGMM+FS

Table 2 Average performance metrics for the Berkeley dataset generated by different Gaussian-based models

Method	GMM	GMM+FS	GGMM	GGMM+FS	BGGMM	BGGMM+FS
Accuracy(%)	92.10	93.26	93.87	95.03	94.98	96.68
Precision(%)	63.4	64.8	65.1	66.9	66.4	73.2
Sensitivity(%)	69.4	75.1	76.8	79.9	81.5	89.4
Specificity(%)	91.2	92.7	92.4	93.9	95.3	96.9
Recall(%)	69.4	75.1	76.8	79.9	81.5	89.4
F1(%)	59.1	62.5	60.8	62.9	63.1	64.3
MCC(%)	52.9	55.7	54.1	58.7	57.4	62.7
BDE	0.324	0.276	0.291	0.218	0.223	0.186

Table 3 Comparative study between different algorithms from the state-of-the-art on the Berkeley Segmentation Database

Method	PRif [49]	UCM [5]	MS [5]	WS [5]	gPb [44]	GMM	BGGMM	BGGMM+FS
Precision(%)	63	65	60	58	68	63	66	73
Recall(%)	62	69	67	68	71	69	81	89
F-measure(%)	63	67	63	63	70	59	63	64

BGGMM+FS. We notice that if we combine a feature selection step within the mixture model, then we can achieve better performance than if we did not consider this step.

5.2 Experiment 2: object detection

In this experiment, we focus on extracting a region of interest from an image. To this end, a series of experiences are performed on the “Microsoft Common Objects in COntext (MSCOCO)” dataset [40]. COCO dataset contains more than 200K images and 91 common object categories with 82 of them having more than 5000 labeled instances. Some samples are given in Fig. 5. This dataset is composed of more than 100K images for training, 5000 images for validation and about 40K for testing. We perform our experiments on the training subset designed with (*train_2017*). The algorithm is performed on several training sets chosen randomly where images are selected from the training subset. A comparative study was also carried out for this dataset. Indeed, we have compared the performance of BGGMM+FS against some relevant methods from the state-of-the-art on the basis of the metric mean Average Precision(mAP). This metric is often applied for object detection problems. Obtained results for different methods are provided in Table 4. According to this table, we can conclude that our method is very competitive and outperforms some other methods. This is justified by the generative nature of the developed approach that allows more flexibility and interpretability of the results. It is noteworthy that the proposed model has the advantage to take into account the nature of the data which is compactly supported. Moreover, the integration of the feature selection process in the generative model enables us to accurately distinguish and identify different objects of interest.

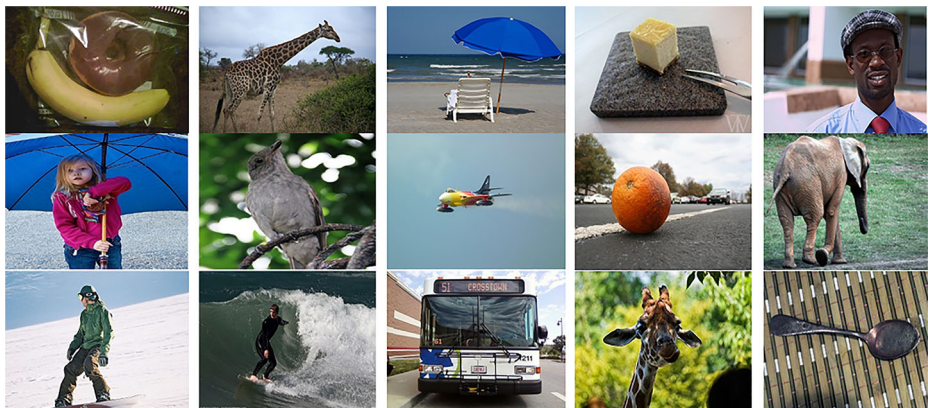
**Fig. 5** Sample images from the COCO dataset

Table 4 Comparative study using the mean average precision (mAP) scores for different object detection methods based on COCO dataset

Method	Detection result (mAP) (%)
SSD300 [43]	41.2
SSD512 [43]	46.4
FPN [41]	33.9
DeNet [61]	33.8
Proposed method	44.3



Fig. 6 Result 1 for Object tracking: First row presents the initial detection of the region of interest ROI (from left to right: first original frame, classification of the frame with BGGMM+FS, and the detected ROI). Second row presents the tracking of the ROI using the C-V level-set model for different frames

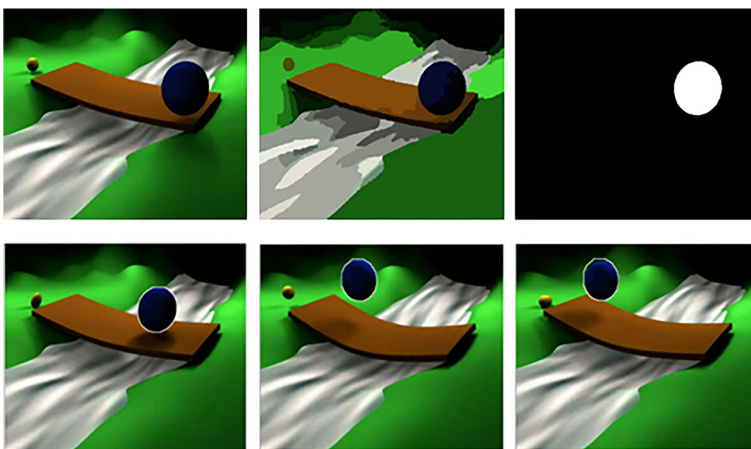


Fig. 7 Result 2 for Object tracking: First row presents the initial detection of the region of interest ROI (from left to right: first original frame, classification of the frame with BGGMM+FS, and the detected ROI). Second row presents the tracking of the ROI using the C-V level-set model for different frames

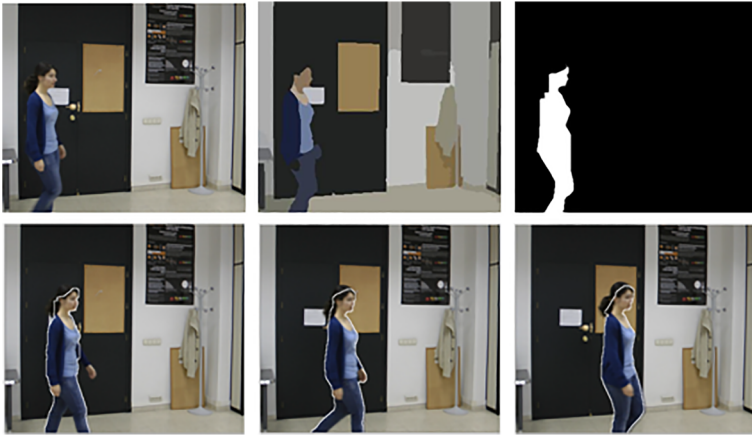


Fig. 8 Result 3 for Object tracking: First row presents the initial detection of the region of interest ROI (from left to right: first original frame, classification of the frame with BGGMM+FS, and the detected ROI). Second row presents the tracking of the ROI using the C-V level-set model for different frames

5.3 Experiment 3: object tracking

In this experiment, we focus on identifying a region of interest (ROI) in a sequence of images. To this end, for each sequence, the first frame is segmented using the BGGMM+FS model, then, the obtained segmentation will be considered as the initial contour for the level-set and will be evolved according to the level-set function in order to detect the boundaries of the same ROI in other frames. Subsequently, the output of the current result will be applied as an accurate initialization step (initial contour) for the following frame in the sequence and so that. Some obtained results for object tracking are depicted in Figs. 6, 7, 8, 9 and 10. The First row of each result presents the initial detection of the region of interest ROI and

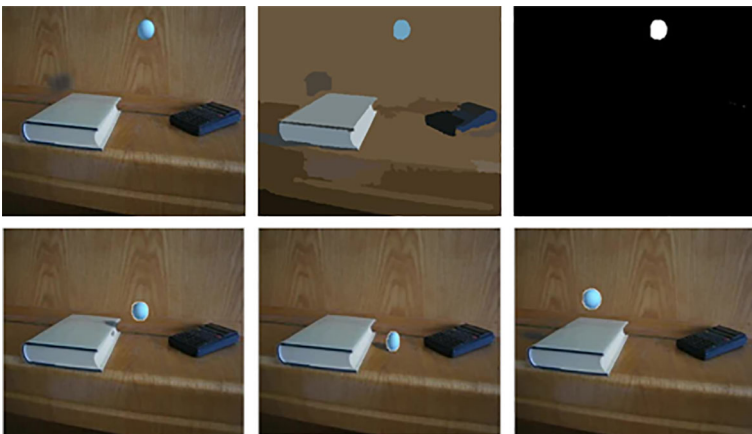


Fig. 9 Result 4 for Object tracking: First row presents the initial detection of the region of interest ROI (from left to right: first original frame, classification of the frame with BGGMM+FS, and the detected ROI). Second row presents the tracking of the ROI using the C-V level-set model for different frames

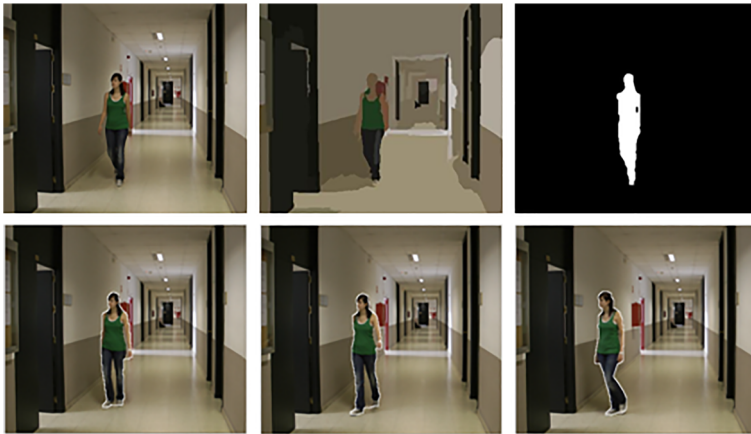


Fig. 10 Result 5 for Object tracking: First row presents the initial detection of the region of interest ROI (from left to right: first original frame, classification of the frame with BGGMM+FS, and the detected ROI). Second row presents the tracking of the ROI using the C-V level-set model for different frames

the second row presents the tracking of the ROI using the C-V level-set model for different frames.

Qualitative results are obtained on the basis of the LASIESTA dataset which contains many real image sequences with their corresponding ground truth [21]. Quantitative measures are performed using the accuracy and the boundary displacement error (BDE) metrics. The accuracy measures the proportion of correctly labelled pixels over all available pixels, and the boundary displacement error (BDE) measures the displacement error [26]. Obtained values are depicted in Tables 5 and 6. A comparative study with other approaches is also provided in Table 7. According to these results, we can conclude that the application of a variational active contour for object tracking with an accurate initialization step provided

Table 5 Quantitative results for five different sequences produced by BGGMM+FS with C-V level-set

Sequence 1	Frame 204	Frame 223	Frame 240	Average
Accuracy(%)	91.66	91.33	90.73	91.24
BDE	0.2709	0.2785	0.2819	0.2771
Sequence 2	Frame 204	Frame 223	Frame 240	Average
Accuracy(%)	96.69	96.41	95.96	96.35
BDE	0.1860	0.1891	0.1936	0.1895
Sequence 3	Frame 204	Frame 223	Frame 240	Average
Accuracy(%)	94.07	94.22	94.75	94.34
BDE	0.2229	0.2219	0.2201	0.2216
Sequence 4	Frame 15	Frame 20	Frame 33	Average
Accuracy(%)	95.49	95.21	97.53	96.07
BDE	0.1877	0.1993	0.1836	0.1902
Sequence 5	Frame 240	Frame 245	Frame 252	Average
Accuracy(%)	97.06	96.83	97.12	97.0
BDE	0.1954	0.2031	0.1940	0.1975

Table 6 Average metrics for different video sequences segmentation models

	GMM	GMM+FS	GGMM	GGMM+FS	BGGMM	BGGMM+FS
Accuracy(%)	94.1	94.4	94.9	95.6	96.2	96.6
BDE	0.354	0.320	0.317	0.269	0.246	0.207

by the BGGMM+FS model enables us to maintain an accurate track of the object of interest (OOI) even if the topology of the object changes over time. Indeed, high performances are obtained and the average accuracy value is more than 93% for all sequences. Furthermore, we can see that the obtained boundary displacement error values are very encouraging thanks to the use of the level-set formalism and the BGGMM+FS model. These results show the merit of using a variational approach initialized by the BGGMM+FS model for a robust object tracking.

6 Discussion

In this paper, a flexible and robust learning model followed with a post processing step is proposed. The later is implemented with a variational active contour for both image/video sequences segmentation and object tracking. Our main purpose is to improve these tasks by investigating the flexibility of bounded models such as the bounded generalized Gaussian mixture model and also by taking into account a feature selection mechanism. For tracking purpose, we tackle this problem by considering an active contours via the well known level set approach. Our method complexity is about $O(NM)$, where N represents pixel's number for the treated image and M is used to designate the number of components. Thus, the first main contribution of the current work is to tackle the segmentation problem by implementing a flexible bounded statistical model given that unbounded models are obviously not the appropriate approximation for data modelling and segmentation. The second main advantage of the proposed work is the consideration of a feature selection mechanism which is able to remove irrelevant features (i.e. noise) which make the detection of the real regions more and more difficult. Obtained results have proven these assumptions and more accurate results are obtained compared to the state of the art. However, it is noteworthy that using the conventional EM-algorithm has some problems related to its dependency on initialization and convergence to local maxima. To overcome this shortcoming, we plan to replace it with the enhanced ECM-algorithm (Expectation/ Conditional Maximization) in order to overcome the problems related to the use of the EM algorithm [9]. Thus, the complicated M-step

Table 7 Comparative study with some methods from the state of the art

Method	Accuracy
Babu et al. [8]	94%
Oussalah et al. [55]	87%
Mavska et al. [46]	86%
Dzyubachyk et al. [23]	85%
Our approach	96.6%

of EM will be replaced with several computationally simpler CM-steps. Moreover, the feature extraction step can be improved by considering other type of visual and spatio-temporal features. We plan also to consider other datasets for validation purpose.

7 Conclusion

We have developed, in the current work, an effective framework for both color images and sequence of images segmentation and tracking. The main goal was to investigate the flexibility of the proposed bounded model combined with a feature selection mechanism (BGGMM+FS) for image segmentation and object detection-tracking. The choice of the model is motivated by its high flexibility for multidimensional data modelling and its ability to integrate a feature selection mechanism. The developed statistical model is also followed with a post-processing step implemented with a variational active contour for tracking a particular object of interest in a sequence of color images. The learning model is implemented on the basis of the expectation-maximization algorithm taking into account the minimum message length (MML) criterion. The validation process is carried out through extensive series of experiments and the final results show high accuracy for both segmentation and tracking. It is noted that the BGGMM+FS offer better capabilities than the conventional and classic generative models.

Future works could be devoted to improve the feature selection mechanism by taken into account other visual (and spatio-temporal) local and/or global features for both images and video sequences. Another future work could be developing a unified framework that integrates the statistical model and the variational model into the same formalism. Moreover, the learning approach may be improved if it follows a Bayesian approximation instead of frequentist strategy in order to overcome the convergence to local maxima. Finally, it is possible to develop an online algorithm based on the proposed mixture model if one want to track specific object in real time.

Acknowledgments Taif University Researchers Supporting Project number (TURSP-2020/26), Taif University, Taif, Saudi Arabia.

References

1. Alhakami W, ALharbi A, Bourouis S, Alroobaea R, Bouguila N (2019) Network anomaly intrusion detection using a nonparametric bayesian approach and feature selection. *IEEE Access* 7:52181–52190
2. Allili MS, Ziou D, Bouguila N, Boutemedjet S (2010) Unsupervised feature selection and learning for image segmentation. In: 2010 Canadian conference on computer and robot vision (CRV). IEEE, pp 285–292
3. Alroobaea R, Alsufyani A, Ansari MA, Rubaiee S, Algarni S (2018) Supervised machine learning of kfcg algorithm and mbtc features for efficient classification of image database and cbir systems. *Int J Appl Eng Res* 13(9):6795–6804
4. Alroobaea R, Rubaiee S, Bourouis S, Bouguila N, Alsufyani A (2020) Bayesian inference framework for bounded generalized gaussian-based mixture model and its application to biomedical images classification. *Int J Imaging Syst Technol* 30(1):18–30
5. Arbelaez P (2006) Boundary extraction in natural images using ultrametric contour maps. In: IEEE conference on computer vision and pattern recognition, CVPR, p 182
6. Arbelaez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. *IEEE Trans Pattern Anal Mach Intell* 33(5):898–916
7. Ayed IB, Mitiche A, Belhadj Z (2005) Multiregion level-set partitioning of synthetic aperture radar images. *IEEE Trans Pattern Anal Mach Intell* 27(5):793–800





8. Babu G, Aneesh R, Nayar GR (2017) A novel method based on chan vese segmentation for salient structure detection. In: 2017 IEEE international conference on circuits and systems (ICCS). IEEE, pp 414–418
9. Bouguila N, Ziou D (2005) On fitting finite dirichlet mixture using ecm and mml. In: Wang P, Singh M, Apté C, Perner P (eds) Pattern recognition and data mining, third international conference on advances in pattern recognition, ICAPR 2005, Bath, UK, August 22–25, 2005, proceedings, Part I, vol 3686. Springer, pp 172–182
10. Bouguila N, Ziou D, Monga E (2006) Practical bayesian estimation of a finite beta mixture through gibbs sampling and its applications. *Stat Comput* 16(2):215–225
11. Bourouis S, Hamrouni K (2010) 3d segmentation of MRI brain using level set and unsupervised classification. *Int J Image Graph* 10(1):135–154
12. Bourouis S, Hamrouni K, Betrouni N (2008) Automatic MRI brain segmentation with combined atlas-based classification and level-set approach. In: 5th International conference, ICIAR 2008 image analysis and recognition, Póvoa de Varzim, Portugal, June 25–27, 2008. Proceedings, pp 770–778
13. Bourouis S, Al Mashrgy M, Bouguila N (2014) Bayesian learning of finite generalized inverted dirichlet mixtures: application to object classification and forgery detection. *Exp Syst Appl* 41(5):2329–2336
14. Bourouis S, Zaguia A, Bouguila N, Alroobaea R (2019) Deriving probabilistic SVM kernels from flexible statistical mixture models and its application to retinal images classification. *IEEE Access* 7:1107–1117
15. Boutemedjet S, Bouguila N, Ziou D (2007) Feature selection for non gaussian mixture models. In: 2007 IEEE workshop on machine learning for signal processing. IEEE, pp 69–74
16. Chan TF, Vese LA (2001) Active contours without edges. *IEEE Trans Image Process* 10(2):266–277
17. Channoufi I, Bourouis S, Bouguila N, Hamrouni K (2018) Color image segmentation with bounded generalized gaussian mixture model and feature selection. In: 4th International conference on advanced technologies for signal and image processing, ATSSIP 2018, Sousse, Tunisia, March 21–24, 2018, pp 1–6
18. Channoufi I, Bourouis S, Bouguila N, Hamrouni K (2018) Image and video denoising by combining unsupervised bounded generalized gaussian mixture modeling and spatial information. *Multimed Tools Appl* 77(19):25591–25606
19. Channoufi I, Najar F, Bourouis S, Azam M, Halibas AS, Alroobaea R, Al-Badi A (2020) Flexible statistical learning model for unsupervised image modeling and segmentation. Springer International Publishing, Berlin, pp 325–348
20. Cong Y, Wang S, Liu J, Cao J, Yang Y, Luo J (2015) Deep sparse feature selection for computer aided endoscopy diagnosis. *Pattern Recognit* 48(3):907–917
21. Cuevas C, Yáñez EM, García N (2016) Labeled dataset for integral evaluation of moving object detection algorithms: Lasiesta. *Comput Vis Image Underst* 152:103–117
22. Darolti C, Mertins A, Bodensteiner C, Hofmann UG (2008) Local region descriptors for active contours evolution. *IEEE Trans Image Process* 17(12):2275–2288
23. Dzyubachyk O, Van Cappellen WA, Essers J, Niessen WJ, Meijering E (2010) Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. *IEEE Trans Med Imaging* 29(3):852–867
24. Falco ID, Pietro GD, Cioppa AD, Sannino G, Scafuri U, Tarantino E (2018) Preliminary steps towards efficient classification in large medical datasets: structure optimization for deep learning networks through parallelized differential evolution. In: 11th International joint conference on biomedical engineering systems and technologies (BIOSTEC), pp 633–640
25. Figueiredo MAT, Jain AK (2002) Unsupervised learning of finite mixture models. *IEEE Trans Pattern Anal Mach Intell* 24(3):381–396
26. Freixenet J, Muñoz X, Raba D, Martí J, Cuff X (2002) Yet another survey on image segmentation: region and boundary information integration. In: Heyden A, Sparr G, Nielsen M, Johansen P (eds) Computer vision - ECCV 2002, 7th European conference on computer vision, copenhagen, Denmark, May 28–31, 2002, proceedings, Part III, vol 2352. Springer, pp 408–422
27. Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning, vol 1. Springer series in statistics. New York
28. Gao Z, Wang D, Wan S, Zhang H, Wang Y (2019) Cognitive-inspired class-statistic matching with triple-constrain for camera free 3d object retrieval. *Future Gener Comput Syst* 94:641–653
29. Gao Z, Xue H, Wan S (2020) Multiple discrimination and pairwise CNN for view-based 3d object retrieval. *Neural Netw* 125:290–302
30. Girshick RB (2015) Fast R-CNN. In: 2015 IEEE international conference on computer vision, ICCV 2015, Santiago, Chile, December 7–13, 2015, pp 1440–1448
31. Girshick RB, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR, pp 580–587

32. Huang J, Kumar SR, Mitra M, Zhu WJ, Zabih R (1997) Image indexing using color correlograms. In: 1997 IEEE computer society conference on computer vision and pattern recognition, 1997. Proceedings. IEEE, pp 762–768
33. Ilyasova N, Paringer R, Kupriyanov A, Kirsh D (2017) Intelligent feature selection technique for segmentation of fundus images. In: 2017 Seventh international conference on innovative computing technology (INTECH), pp 138–143
34. Jackowski K, Cyganek B (2017) A learning-based colour image segmentation with extended and compact structural tensor feature representation. *Pattern Anal Appl* 20(2):401–414
35. Junfeng L, Jinwen M (2016) Effective selection of mixed color features for image segmentation. In: 2016 IEEE 13th international conference on signal processing (ICSP), pp 794–798
36. Law MH, Figueiredo MA, Jain AK (2004) Simultaneous feature selection and clustering using mixture models. *IEEE Trans Pattern Anal Mach Intell* 26(9):1154–1166
37. Li Y, Guo L (2008) TCM-KNN scheme for network anomaly detection using feature-based optimizations. In: Proceedings of the 2008 ACM symposium on applied computing (SAC), Fortaleza, Ceara, Brazil, March 16–20, 2008, pp 2103–2109
38. Li Z, Tang J (2015) Unsupervised feature selection via nonnegative spectral analysis and redundancy control. *IEEE Trans Image Process* 24(12):5343–5355
39. Li Z, Liu J, Yang Y, Zhou X, Lu H (2014) Clustering-guided sparse structural learning for unsupervised feature selection. *IEEE Trans Knowl Data Eng* 26(9):2138–2150
40. Lin T, Maire M, Belongie SJ, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft COCO: common objects in context. In: 13th European conference ECCV, pp 740–755
41. Lin T, Dollár P, Girshick RB, He K, Hariharan B, Belongie SJ (2017) Feature pyramid networks for object detection. In: IEEE Conference on computer vision and pattern recognition, CVPR, pp 936–944
42. Lindblom J, Samuelsson J (2003) Bounded support gaussian mixture modeling of speech spectra. *IEEE Trans Speech Audio Process* 11(1):88–99
43. Liu W, Anguelov D, Erhan D, Szegedy C, Reed SE, Fu C, Berg AC (2016) SSD: single shot multibox detector. In: Computer vision—ECCV 2016—14th European conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I, pp 21–37
44. Maire M, Arbelaez P, Fowlkes CC, Malik J (2008) Using contours to detect and localize junctions in natural images. In: 2008 IEEE Computer society conference on computer vision and pattern recognition CVPR
45. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Eighth IEEE international conference on computer vision, 2001. ICCV 2001. Proceedings, vol 2. IEEE, pp 416–423
46. Maška M, Matula P, Daněk O, Kozubek M (2010) A fast level set-like algorithm for region-based active contours. In: International symposium on visual computing. Springer, pp 387–396
47. McLachlan G, Peel D (2000) Finite mixture models. Wiley, New York
48. Meignen S, Meignen H (2006) On the modeling of small sample distributions with generalized gaussian density in a maximum likelihood framework. *IEEE Trans Image Process* 15(6):1647–1652
49. Mignotte M (2010) A label field fusion bayesian model and its penalized maximum rand estimator for image segmentation. *IEEE Trans Image Process* 19(6):1610–1624
50. Mumford D, Shah J (1989) Optimal approximations by piecewise smooth functions and associated variational problems. *Commun Pure Appl Math* 42(5):577–685
51. Najaf F, Bourouis S, Bouguila N, Belghith S (2017) A comparison between different gaussian-based mixture models. In: 14th IEEE/ACS International conference on computer systems and applications, AICCSA 2017, Hammamet, Tunisia, October 30–Nov. 3, 2017, pp 704–708
52. Najaf F, Bourouis S, Zaguia A, Bouguila N, Belghith S (2018) Unsupervised human action categorization using a riemannian averaged fixed-point learning of multivariate GGMM. In: Image analysis and recognition - 15th international conference, ICIAR, pp 408–415
53. Najaf F, Bourouis S, Bouguila N, Belghith S (2019) Unsupervised learning of finite full covariance multivariate generalized gaussian mixture models for human activity recognition. *Multimed Tools Appl* 78(13):18669–18691
54. Najaf F, Bourouis S, Bouguila N, Belghith S (2020) A new hybrid discriminative/generative model using the full-covariance multivariate generalized gaussian mixture models. *Soft Comput* 24(14):10611–10628
55. Oussalah M, Shabash M (2012) Object tracking using level set and mpeg 7 color features. In: 2012 3rd International conference on image processing theory, tools and applications (IPTA). IEEE, pp 105–110
56. Pi M (2006) Improve maximum likelihood estimation for subband ggd parameters. *Pattern Recognit Lett* 27(14):1710–1713
57. Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. In: 2017 IEEE Conference on computer vision and pattern recognition, CVPR 2017, Honolulu, HI, USA, July 21–26, 2017, pp 6517–6525

58. Redmon J, Divvala SK, Girshick RB, Farhadi A (2016) You only look once: unified, real-time object detection. In: IEEE Conference on computer vision and pattern recognition, CVPR, pp 779–788
59. Sethian J (1999) Level set methods and fast marching methods: evolving interfaces in geometry, fluid mechanics, computer vision, and materials science, 2nd edn. Cambridge University Press, Cambridge
60. Szczypiński P, Klepaczko A, Pazurek M, Daniel P (2014) Texture and color based image segmentation and pathology detection in capsule endoscopy videos. *Computer Methods Progr Biomed* 113(1):396–411
61. Tychsen-Smith L, Petersson L (2017) Denet: scalable real-time object detection with directed sparse sampling. In: 2017 IEEE International conference on computer vision (ICCV). IEEE, pp 428–436
62. Wallace CS (2005) Statistical and inductive inference by minimum message length. Springer Science & Business Media
63. Wang J, Jiang H, Yuan Z, Cheng M, Hu X, Zheng N (2017) Salient object detection: a discriminative regional feature integration approach. *Int J Comput Vis* 123(2):251–268
64. Zhang K, Zhang L, Song H, Zhou W (2010) Active contours with selective local or global segmentation: a new formulation and level set method. *Image Vis Comput* 28(4):668–676
65. Zhang K, Zhang L, Yang MH (2013) Real-time object tracking via online discriminative feature selection. *IEEE Trans Image Process* 22(12):4664–4677
66. Zhang K, Zhang L, Yang MH, Hu Q (2013) Robust object tracking via active feature selection. *IEEE Trans Circ Syst Video Technol* 23(11):1957–1967

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Sami Bourouis^{1,2}  · Ines Channoufi² · Roobaea Alroobaea¹  · Saeed Rubaiee³  · Murad Andejany³ · Nizar Bouguila⁴ 

Ines Channoufi
ines.channoufi@esprit.tn

Roobaea Alroobaea
r.robai@tu.edu.sa

Saeed Rubaiee
salrubaiee@uj.edu.sa

Murad Andejany
mbazzar@uj.edu.sa

Nizar Bouguila
nizar.bouguila@concordia.ca

¹ Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, P.O. Box 11099, Taif 21944, Saudi Arabia

² LR-SITI Laboratoire Signal Image et Technologies de l'Information, Université de Tunis El Manar, Tunis, 1002, Tunisia

³ Department of Industrial and Systems Engineering, Faculty of Engineering, University of Jeddah, Jeddah, Saudi Arabia

⁴ The Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC H3G 1T7, Canada