Identifying the regions or objects of interest in a natural scene is a very difficult task because the content of natural images consists of the multiple non-uniform sub-regions. In this paper, we present a novel Region of Interest (ROI) detection method to automatically and efficiently minimize the ROI in the images. We applied the geometric active contours that forces the variational level set function to be close to the object boundaries in image segmentation. In addition, the mean-shift algorithm was used to reduce the sensitivity of parameter change in variational level set formulation. Varieties of natural images in different modalities were tested in order to achieve the purpose of the experiment. When image background was not complex, the precision and recall are 93.69% and 89.74%, respectively. For an image with a complex background, the precision and recall are 88.59% and 90.92%, respectively. The results confirmed that the proposed methods could efficiently solve the segmentation problems and experimental results show that it is closer to human behavior and decision making for detection and evaluation of Region of Interest.

Keywords: Active Contour Model, Variational Level Set Function, Mean Shift Segmentation, Object Detection, Region of Interest Detection

1. Introduction

In most image analysis operation, pattern classifiers require individual objects to be separated from the image, so the description of objects can be transformed into a suitable form for computer processing. Image segmentation is a fundamental task, responsible for separating operation. Typically, the level of image segmentation depends on the problem being solved. That is, segmentation should stop when the region of interest (ROI) or object of interest (OOI) in the application have been isolated [1]. Due to this property of problem dependence, automatically identify the regions or object of interest in the natural scene that are complex in
content and can be any shape is one of the most difficult task in image analysis. The reason is because it depends on the efficiency of image segmentation which normally relies on sensitivity of parameter change, for example the popular method such as mean shift algorithm \[6\] and active contour model \[3\]. Moreover, the definition of objects is largely subjective, predetermined objects cannot serve as a universal template for matching in all applications \[2\]. Consider a given vector valued image \(I(x,y) : \Omega \rightarrow \mathbb{R}^d\) is a multispectral image that is defined as a function on a two-dimensional spatial domain \(\Omega \subset \mathbb{R}^2\), and \(d \geq 1\) is the dimension of the vector \(I(x,y)\), in particular, \(d = 1\) for gray level images, while \(d = 3\) for color images. In image processing, the statistical distribution of image intensity with the multispectral image is represented by multiple sub-classes or modes, and sub-region in the image has a uniform property \[3\]. In fact, non-uniform sub-regions and intensity inhomogeneities often occur in the natural images. The example of the classic problem on a white zebra image with non-uniform sub-regions that consists of white and black stripes. The statistical distribution of image intensity within the zebra would be the graph of the probability mixture density function (PDF) 
\[
P(I) = \alpha P_1(I) + (1-\alpha)P_2(I).
\] Since the statistical distribution of image intensity within the zebra has at least two sub-classes including black and white stripes. In this case, a good segmentation method should be able to segment a mixture of sub-classes as the class representing the zebra, not separating the black and white stripes. Moreover, the method should be able to identify the objects or regions of interest that is zebra object in the image together.

Many algorithms have been proposed for automatic ROI detection in the images. Two main categories of these are bottom-up and top-down methods \[4\]. Bottom-up method is ideally selecting the small areas that are different from the surroundings. It first computes a segmentation of image by using several attributes in an image. A winner-take-all strategy then is used to determine that region’s important value. For example, Osberger and Maeder \[5\] used a weighted combination of a region’s low-level features such as size, intensity, contrast, location, and shape to determine that region of core areas in the images. Srivani and Damon \[6\] presented an algorithm that followed the previous work of Osberger and Maeder, and used various additional factors including color, foreground/background, edge strength and category (new high-level factor) to determine perceived interest. The main contribution of the approach was the use of a Bayesian framework based on precise likelihood functions to estimate
the perceived interest of object in images. However, data set training process is still needed for an approach. Levente et al. [1] exploited the local blur/focus information in the image for automatically classifying image areas relative to each other. The main novelty lies in the use of localized blind deconvolution for automatic estimation of focused areas on ordinary images. The advantage of this approach is that it does not require a priori knowledge about the image or the shooting condition. Based on the literature, the main drawback of the bottom-up methods depends on the efficiency of segmentation technique that is very sensitive with its parameters. Top-down methods, on the other hand, assume that people pay more attention to areas corresponding to semantic objects, which is additional to low-level features such as face detection and text detection [4]. Various approaches to face detection are discussed in [7], [8], [9], [10]. In contrast, some methods do not depend on image segmentation, for instance, Khanh Vu et al. [2] proposed an efficient core-area detection algorithm for image retrieval based on ROI. A core area was optimal in the scene that it was maximized to capture the characteristics of noise-free query (NFQ) to the greatest extent. The result of core area detection was a large square sub image with noise no greater than the maximum tolerable level. However, the effectiveness of core area identifying was determining a set of good seed pixels. In this paper, we present a novel unsupervised detection method to automatically and efficiently minimize the ROI in the images. We focus on applying an edge-based active contour model that draws upon edge information in local regions based on variational level set formulation. Since the traditional variational level set function still has parameterized sensitivity, therefore the mean-shift algorithm is used to reduce the sensitivity of parameter changes for variational level set method of active contour model that occurs in variation and complexity of intensity inhomogeneities in natural images. Based on this proposed method, the difficulties of non-uniform sub-region and intensity inhomogeneities in natural image segmentation are overcome.

The remainder of this paper is organized as follows. In section 2, some of efficient existing image segmentation methods and their limitations are presented. A description of our proposed method is given in Section 3. Experimental results of our proposed methods and some implementation details are given in section 4 and finally, we will conclude our paper on section 5.
2. Background and Knowledge

Our approach is based on the active contour model with adaptive edge information to obtain a robust segmentation model. The mathematical implementation of the proposed active contour models was accomplished using variational level set formulation. In addition, the mean-shift algorithm was used to reduce the sensitivity of parameter change of level set formulation. Before introducing our framework of automatic ROI detection, we will brief some background of these methods and their limitations.

2.1 Active Contour Model

Active contour is one of the most popular models that typically applied to image segmentation. An advantage of using active contours as image segmentation methods is that they partition an image into sub-region with continuous boundaries. There are two major classes of active contour such as edge-based models and region-based model [7]. In this paper, edge-based model is selected. It uses local edge information to attract the active contour toward the object boundaries. For an image \( I(x, y) \), let \( C \) be a closed curves (Contour) on the image \( \Omega \), which separates \( \Omega \) into two regions: \( \Omega_1 = \text{inside}(C) \) and \( \Omega_2 = \text{outside}(C) \). A modifies energy functional of the classical energy based snakes formulation [8]

where the rigidity coefficient of second order smoothness component is set to zero \( \beta = 0 \) given by

\[
E(C) = \alpha \int_{\Omega_1} |C(q)|^2 \, dq + \lambda \int_{\Omega_1} \gamma(|\nabla I(C(q))|) \, dq \\
= \int_{\Omega_1} E_{in}(C(q)) \, dq + \int_{\Omega_2} E_{ext}(C(q)) \, dq
\]

where \( \alpha \) and \( \lambda \) are real positive constants. \( C(q) : [0,1] \rightarrow \mathbb{R}^2 \) is a parameterized planer curve and \( I: [0,a] \times [0,b] \rightarrow \mathbb{R} \) be a given image in which we want to detect the objects boundaries[8]. \( \gamma(|\nabla I|) \) is strictly decreasing function where \( |\nabla I| \) is acting as edge indicator. The second term (external energy) is responsible for attracting the contour toward the object boundaries, while the first term controls the smoothness of contours to be detected (internal energy). The classic snakes provide an accurate location of the edges only if the initial contour is given sufficiently near the edges because they make use of only the local information along the contour but an estimating a proper position of initial contours without prior knowledge is a difficult problem. Also classic snakes cannot detect more than one boundary simultaneously because the snakes maintain the same topology during the evolution stage. That is, snakes cannot split to multiple boundaries or merge from multiple initial contours [1]. In contrast, Level set theory has given a solution for this problem. In this paper we represented the energy based snakes formulation by a level set.
formulation, and then the energy minimization problem can be converted to solve a level set evolution in the next subsection.

![Diagram](image)

Fig. 1 Shows an example of the topological changes observed in the evolution of the level set function and propagation of corresponding contours: (a) the topology change on level set function $\partial(t,x,y)$, (b) shows the initially separated contours merge as topology of level set function varies.

### 2.2 Variational Level Set Formulation of Active Contour

The level set method was first introduced by Osher and Sethian [15]. It is a powerful numerical and theoretical tool for propagating interfaces. The basic idea is to start with a closed curve in two dimensions or a surface in three dimensions and allow the curve to move perpendicular to itself at a prescribed speed. In two dimensions, the level set method amounts to representing a closed curve $C(t)$ using an auxiliary function, called the level set function $\phi$. $C(t)$ is represented as the zero level set $C(t) = \{ (x,y) | \phi(t,x,y) = 0 \}$. If the contour $C(t)$ moves in the normal direction with a function speed $F$. The evolution equation of the level set function $\phi(t,x,y)$ can be written in the following partial differential equation general form

$$\frac{\partial \phi}{\partial t} + F |\nabla \phi| = 0$$

(2)

where $|\cdot|$ is the Euclidean norm and $t$ is time. The function $F$ is called the speed function to push or pull the contours. A particular case is the motion by mean curvature, when $F = \text{div}(\nabla \phi |\nabla \phi|)$ is the curvature of the level-curve of $\phi$ passing through $(x,y)$. For image segmentation, the function $F$ depends on the image data and the level set function $\phi$. The advantage of using the evolution of level set function is that contours can split or merge as the topology of the level set function changes. Therefore, level set method can detect more than one boundary simultaneously, and multiple initial contours can be placed as show Fig. 1. However, the main
drawbacks of traditional level set methods are re-initialized the function $O$ to be a signed distance function periodically during the evolution [10][11][12]. In this paper, the variational level set formulation of active contour without re-initialization is used to achieve the zero level curves toward the object boundaries in an image. Normally, the level set function is an approximate signed distance function during the evolution in a neighborhood around the zero level set. The key role in its formulation is how to close a function $O$ into a signed distance function 

$$\Omega \subset \mathbb{R}^2$$

where signed distance function must satisfy a desirable property of $|\nabla \phi| = 1$ [8][9]. Based on energy based active contour in (1), the energy can be represented by variational level set formulation and the total energy functional is defined by

$$E(\phi) = \mu E_{\text{int}}(\phi) + E_{\text{ext}}(\phi)$$

(3)

where $E_{\text{int}}(\phi) = \int_{\Omega} \frac{1}{2} \left( |\nabla \phi| - 1 \right)^2 dxdy$ is a metric term to characterize how to close a function $O$ whose task is to remove the re-initialization step [9]. $\mu > 0$ is parameter controlling the effect of penalizing the deviation of $O$ from a signed distance function. While the external energy $E_{\text{ext}}(\phi)$ drives the zero level set toward the object boundaries. Initialized $E_{\text{ext}}(\phi) = E_g(\phi)$, let $g = 1/(1 + |\nabla K_\sigma * I|^2)$ be the edge indicator function in image $I$ and $K_\sigma$ is Gaussian kernel with standard deviation $\sigma$. The external energy in (1) for a function $O(x,y)$ can be written as

$$E_{\text{ext}}(\phi) = \int_{\Omega} g \delta(\phi) |\nabla \phi| dxdy + v \int_{\Omega} gH(-\phi) dxdy$$

(4)

![Fig. 2](image_url)

Fig. 2 shows the results of variational level set method for a grayscale image of flower with parameters $\sigma$ and $v$. The regions shown in row 1 are the results obtained from traditional level set method while the regions shown in row 2 are the results obtained from the proposed method.

where $\delta$ and $H$ are univariate Dirac and Heaviside functions [9][13], respectively. The second term is introduced to speed up curve evolution, which is positive or negative constant $v$ depending on the relative position of initial closed curve are placed outside or inside of the object of interest. While, the length of the zero level curve of $O$ is computed by the first term, and $\lambda > 0$ is constant.

We shall return to the variational level set denoted by $\frac{\partial E}{\partial \phi}$ the Gateaux derivative of the functional $E$ in (1). The gradient flow of evolution that minimizes the functional $E$ can be represented by
Based on the function $\mathcal{O}$ that minimizes functional (3) satisfies the Euler-Lagrange equation $\frac{\partial E}{\partial \phi} = 0$ and calculus of variations \cite{12}. The evolution equation of level set function is the following gradient flow

$$\frac{\partial \phi}{\partial t} = \mu \Delta \phi - \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda \delta(\phi) \text{div} \left( g \frac{\nabla \phi}{|\nabla \phi|} \right) + \nu g \delta(\phi)$$

(6)

where $\Delta$ is the Laplacian operator. Importantly, the function of driving the zero level curves toward the object boundaries can be computed by the last two terms in the right hand side of (6), which is external energy.

Although, a variational level set formulation of active contours give us the good segmentation in real images such as microscope cell, ultrasound and so on that are presented in \cite{9}, but it cannot deal with the natural images that contain object and background complexity. The first row in Fig. 2 shows the results on 140x90 pixel image of flower. These images were tested on Chunming Li et al.’s variational level set \cite{9} with variation of parameter change in equation (6). We can observe that changing the parameter $\sigma, \lambda, \mu, \nu$ and time step $\tau$ cannot minimize the object of interest effectively.

### 2.3 Mean Shift Segmentation

Mean shift algorithm is one of the most popular algorithms used in image segmentation. The mean shift technique builds upon a general machine learning approach called unsupervised clustering. The mean shift algorithm is a non-parametric and iterative technique that has been used for finding modes of an estimated probability density. The first step is to calculate the density of points in a specific area compare with the interested point $x$. The second step is to calculate the average density in order to find out the location for the movement. The movement occurs in such a way that it moves to the area, which contains higher point density. It stops moving if the movement location obtained from the calculation is unchanged or the calculated value is less than the expected value. The process will stop calculating and the point that moves to the same location is classified as the same group. Given $\{x_i, i=1,\ldots,n\}$ be an arbitrary set of $n$ points in the $d$-dimensional Euclidean space $\mathbb{R}^d$, and let $K(x)$ be the kernel function where the choice of the kernel function $K$ is flexible, chosen as density gradient kernel $G(x)$ proposed in \cite{6}, the difference between $m(x)-x$ is called mean-shift, which can be calculated by equation (7).
where \( \| \) is Euclidean distance. In multispectral image, an image is typically represented as two domains such as spatial (two-dimensional) and range (gray or color) domain. When the location and range vectors are concentrated in joint spatial-range domain of dimension \( d = p + 2 \), their different natures have to be compensated by proper normalization. We can calculate a new mean shift [6] by equation (8).

\[
m(y_{i,j}) = \frac{\sum x_i \cdot g \left( \frac{y_{i,j} - s_{i,j}}{h_s} \right) \left| \left| \frac{y_{i,j} - F(s_{i,j})}{h_s} \right| \right|^2}{\sum g \left( \frac{y_{i,j} - s_{i,j}}{h_s} \right) \left| \left| \frac{y_{i,j} - F(s_{i,j})}{h_s} \right| \right|^2} - y_{i,j}
\]

(8)

where \( s \) and \( r \) represent data in spatial and range domain respectively. We assume that the spatial domain and range domain are independent. \( F(s) \) is the filtered image. \( y_s \) is spatial part, representing the pixel locations with bandwidth parameter \( h_s \), while \( y_r \) is range part with bandwidth \( h_r \). Further details may be found in [6]. Even through, the classification of components in images using mean shift yield good results but it slightly increases sensitivity to \( h_r \) and \( h_s \) parameters. In (8), the results from the classification of components in images are much different when small changes were made to the value of parameters \( h_r \) and \( h_s \). This resents the major problem for classifying natural images that have diversity in a sub-region and intensity inhomogeneities in each image. Moreover, mean shift algorithm not able to segment the objects correctly when the objects consist of mixture of sub-classes. They are illustrated in Fig. 3(a) and Fig. 3(b).

3. Proposed Detection of Region of Interest Detection

The main problem in active contours based approach and mean shift algorithm are presented in previous section. This section presents the efficient method for an automatic ROI detection in natural images for an image analysis. The concept of the proposed method is
to find out the minimization of region of interest obtained from the segmentation of the images using the mean shift algorithm. It is used to gather similar images into a group of features without the necessity of classifying different components in images for completeness and to reduce the sensitivity for adapting parameter values. After that, the closed contour with pixels that passes the mean shift process is constructed. It is used to specify an area or important objects within the images. The weighted combination of sub-images (button-up properties) is not considered. Moreover, the proposed method does not require prior knowledge or supervised learning. Each iteration requires the following steps.

1) Reduce the size of image to 140x90 pixels
2) Compute a segmentation of an image by using mean shift algorithm. The mean shift procedure for certain x where \( \{x_i, i=1,\ldots,n\} \) be an arbitrary set of n points within a data set can be obtained by

\[
\vec{m}(x) = \frac{\sum_{i=1}^{n} x_i g \left( \frac{y_{i,j} - s_j}{h_j} \right) \left( \frac{y_{i,j} - F(s_j)}{h_j} \right)}{\sum_{i=1}^{n} g \left( \frac{y_{i,j} - s_j}{h_j} \right) \left( \frac{y_{i,j} - F(s_j)}{h_j} \right)}
\]

3) Image that passes the mean shift process is converted to gray images.
4) Create the initial contour, supposing \( \varphi_0 \) as the initial function, let \( \Omega_0 \) be a subset in the image domain \( \Omega \), and \( \partial \Omega_0 \) be all the points on the boundaries of \( \Omega_0 \), which can be efficiently identified by some simple morphological operations. Then, the initial function \( \varphi_0 \) is defined as

\[
\varphi_0(x,y) = \begin{cases} 
-\rho & (x,y) \in \Omega_0 - \partial \Omega_0 \\
0 & (x,y) \in \Omega_0 \\
\rho & \Omega_0 - \Omega_0 
\end{cases}
\]

where \( \rho > 0 \) is a constant. In this paper we specific \( \rho = 4 \) and \( \Omega_0 \) containing two separate regions.

5) The initial level set functions are computed by equation (4) from an arbitrary region \( \Omega_0 \) in the image domain \( \Omega \). The following parameters were used to test in (4), including \( \sigma, \nu, \lambda, \mu, \tau \), which took 360 iterations of curve evolution. An example is shown in Fig. 4 to illustrate how the progressive ROI detection proceeds.
4. Experimental Results

In the past years, many algorithms have been proposed for automatic ROI detection in natural images. Unfortunately, these methods were often evaluated only on specific and small data sets that are not publicly available [14]. The lack of benchmarks makes experiments non-repeatable and quantitative evaluation difficult. Therefore this paper compared the efficiency of the proposed method with the method using the human segmentation of the images. We evaluate classification performance by applying precision and recall measurement. Let A and B are set of area of the ROI identified by the system and human segmentation, respectively. Precision is the fraction of retrieved area (the set A) which has been which is relevant. Recall is the fraction of relevant area (the set B) which has been which is retrieved. The precision and recall are defined by equation (11) and (12).

\[
\text{Precision} = \frac{A \cap B}{A} \quad (11)
\]

\[
\text{Recall} = \frac{B \cap A}{B} \quad (12)
\]

Based on the precision and recall measurement, the relation of precision and recall illustrate on Fig 6. The detail of each case can be described as the following. Fig. 5(a) shows a precision of 50% at recall 50% and Fig. 5(b) shows a precision of 0% at recall 0% (the system cannot identify the OOI). Fig.6(c) shows a precision of...
100% at recall 50% and Fig. 5(d) shows a precision of 50% at recall 100%

Fig. 5 Several Cases for Precision and Recall

In order to achieve the experiment, a variety of natural images in different modalities from Correll photo image database (image data set for image retrieval system) were tested using the proposed method. These images contained varieties of complexity of backgrounds and objects which were difficult to specify the region of interest. In this paper, the experiment was conducted by considering two different levels of complexity of content in the images, which were less complex and complex which separate by background complexity. The first step was done by applying Mean shift filtering with uniform kernel \((h_r, h_s) = (6.5, 7)\) for testing with all images. To identify ROI within the image, the proposed variational level set formulation of active contour based on edge information without re-initialization was used to achieve the closed curves toward the object boundaries in an image.

We conducted all experiments by using the same parameters \(\lambda = 2.5, \mu = 0.04\), and time step \(\tau = 5.0\). In particular, the parameters \(\sigma = 1.5, \nu = 3.0\) and \(\sigma = 0.5, \nu = 1.5\) were used for fitting the boundaries of objects with complex and less complex of image content, respectively. We used 30 images with a complex background and a less complex background in the experiment. The experimental results using our approach base on precision and recall measurement are shown in table 1. In a less complex background, the precision and recall are 93.69% and 89.74%, respectively. In a complex background, the precision and recall are 88.59% and 90.92%, respectively. Fig. 6 and Fig. 7 illustrate the results from application of proposed method on a 140 x 90 pixel that are typical images with less complex and complex of image content, respectively. The original images are show in the first column. The second column shows the segmentation of images using the mean shift algorithm. In the last two columns are show the final ROI results (white color) of the proposed method and method using the human segmentation of the images, respectively. The boundaries of object of interest were efficiently and accurately extracted by the proposed method.
Fig. 6  Result of the proposed method with a less complex background of color images
Fig. 7 Result of the proposed method with a complex background of color images
Table 1. Segmentation Result using our approach in natural images with less complex and complex background based on precision and recall measurement

<table>
<thead>
<tr>
<th>Images</th>
<th>Less Complex</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma = 0.5, v = 1.5$</td>
<td>$\sigma = 1.5, v = 3.0$</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>zebra</td>
<td>91.24%</td>
<td>95.20%</td>
</tr>
<tr>
<td>rhinoceros</td>
<td>97.18%</td>
<td>94.66%</td>
</tr>
<tr>
<td>worm</td>
<td>97.04%</td>
<td>91.20%</td>
</tr>
<tr>
<td>bird</td>
<td>96.46%</td>
<td>92.85%</td>
</tr>
<tr>
<td>penguin</td>
<td>93.29%</td>
<td>89.20%</td>
</tr>
<tr>
<td>hawk1</td>
<td>97.96%</td>
<td>87.98%</td>
</tr>
<tr>
<td>hawk2</td>
<td>97.96%</td>
<td>87.98%</td>
</tr>
<tr>
<td>elephant</td>
<td>96.84%</td>
<td>77.49%</td>
</tr>
<tr>
<td>wolf</td>
<td>94.21%</td>
<td>95.10%</td>
</tr>
<tr>
<td>horse</td>
<td>86.21%</td>
<td>87.79%</td>
</tr>
<tr>
<td>bear</td>
<td>98.75%</td>
<td>88.69%</td>
</tr>
<tr>
<td>turtle</td>
<td>88.61%</td>
<td>80.83%</td>
</tr>
<tr>
<td>bug</td>
<td>83.35%</td>
<td>88.34%</td>
</tr>
<tr>
<td>butterfly</td>
<td>91.05%</td>
<td>95.61%</td>
</tr>
<tr>
<td>flower</td>
<td>95.24%</td>
<td>93.22%</td>
</tr>
<tr>
<td>Average</td>
<td>93.69%</td>
<td>89.74%</td>
</tr>
</tbody>
</table>

5. Conclusion

In this work, we present the proposed method for unsupervised segmentation of natural images. The proposed unsupervised detection of Region Of Interest uses the framework of active contours. We applied an edge-based active contour model that drew upon edge information in local regions. The mathematical implementation of the proposed active contour model was accomplished using a variational level set formulation. By presenting contours as a level of a topological function, multiple contours can be merged into one contour, or can split a contour into multiple contours providing a good flexibility in the use of active contour. The mean-shift algorithm was used to reduce the sensitivity of parameter change of level set formulation. The results show that the proposed method can specify the region or object of interest in natural images more accurately and efficiently.

In future work, based on this study, we plan to work on the complete set of solutions toward autonomous ROI extraction based on the experimental results by combining parameters $\sigma$, $v$ and bandwidth $h$ selection technique with high level top-down information.

References


