Automatic Region of Interest Detection in Natural Images
ANUCHA TUNGKASTHAN¹, WICHIAN PREMCHAISWADI²
Graduate School of Information Technology in Business
Siam University
38 Petkasem Rd., Phasicharoen, Bangkok, 10160
THAILAND
¹aimdala@hotmail.com, ²wichian@siam.edu

Abstract: - Identifying the Region or Object of Interest in a natural scene is a complex task because the content of natural images consists of the multiple non-uniform sub-regions and the intensity inhomogeneities. In this paper, we present a novel Region of Interest (ROI) detection method to minimize the ROI in the images automatically. We applied the geometric active contours that forces the variational level set function to be close to object boundaries. In addition, the mean-shift algorithm was used to reduce the sensitivity of parameter change in variational level set equation. In order to achieve the experiment, varieties of natural images in different modalities were tested. We compared the efficiency of the proposed method with the method using the human segmentation of the images. In a less complex background, the precision and recall are 92.77% and 88.95%, respectively. In a complex background, the precision and recall are 88.93% and 89.10%, respectively. The experimental results show that our method is imitating human decision making for ROI detection and evaluation.

Key-Words: - Image Segmentation, Active Contour, Region of Interest, Object of Interest, Level Set, Mean Shift Algorithm

1 Introduction
Image segmentation is an extremely important task in image analysis, particularly those related to pattern classification, e.g., medical imaging, security surveillance, remote sensing. 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, mean shift and active contour method. Moreover, the definition of objects is largely subjective, pre-determined 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 is a 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) \( P(I) = aP_1(I) + (1-a)P_2(I) \). 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. In general, the 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 [4] 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 [5] 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. 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 subimage 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 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 has still 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.

2 Background

2.1 Mean Shift Segmentation

In image processing, mean shift algorithm is one of the most popular used in image segmentation. The mean shift technique builds upon a general machine learning approach called unsupervised clustering. In the mean shift algorithm, 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 \( \{ xi, \ i=1,...,n \} \) be an arbitrary set of \( n \) points in the \( d \)-dimensional Euclidean space \( R^d \), 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 [6], the difference between \( m(x)-x \) is called mean-shift, which can be calculated by equation (1).

\[
\sum \frac{x_i - x}{h_x} - x
\]

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 (2).

\[
\sum \frac{x_i - x}{h_x} - x
\]

Fig. 2 Results of the mean shift method for a color image of penguin and cartoon fish. Row 1 shows the original image. Row 2-4 shows the segmentation of change bandwidth hs.
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. $y_i$ is spatial part, representing the pixel locations with bandwidth parameter $h_s$, while $y_r$ is range part with bandwidth $hr$. 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 $hr$ and $h_s$ parameters. In (2), the results from the classification of components in images are much different when small changes were made to the value of parameters $h_s$ and $h_r$. 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. 2(a) and Fig. 2(b).

### 2.2 Active Contour

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

$$
E(C) = \alpha \int_{\Omega_1} |C(q)|^2 dq + \lambda \int_{\Omega_2} g(|\nabla I(C(q))|)|\nabla I(q)|^2 dq
$$

where $\alpha$ and $\lambda$ are real positive constants. $C(q) : [0,1] \rightarrow \mathbb{R}^2$ is a parameterized planar 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]. $g(|\nabla I|)$ 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.

### 2.3 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 $F$. The evolution equation of the level set function $\partial I(t,x,y) = 0$. If the contour $C(t)$ moves in the normal direction with a function speed $F$. The evolution equation of the level set function $\partial F/\partial t = 0$ can be written in the following partial differential equation general form

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

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 $F = \nabla \phi / |\nabla \phi|$ 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
Based on energy based active contour without re-initialization [9]. Let \( \phi \) be the evolution function periodically during the evolution of \( \partial^2 E/\partial \phi^2 = 0 \) and calculus of variations. The evolution equation of level set function is the following gradient flow

\[
\frac{\partial \phi}{\partial t} = \mu \Delta \phi - \text{div}(\nabla \phi | \nabla \phi|) + \lambda \delta(\phi) \text{div}(g \nabla \phi) + v g \delta(\phi)
\]

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 (8), 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 [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 140x93 pixel image of flower. These images were tested on Chunming Li et al.’s variational level set [9] with variation of parameter change in equation (8). We can observe that changing the parameter \( \sigma, \lambda, \mu, v \) and time step \( \tau \) cannot minimize the object of interest effectively.

Fig. 3 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.

3 The Proposed Automatic Region of Interest Detection

Based on the problem of active contour model and mean shift algorithm presented in section 2, this section proposed 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 one boundary simultaneously, and multiple initial contours can be placed. However, the main drawbacks of traditional level set methods are re-initialized the function \( \phi \) 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 \( \phi \) at \( x \) to remove the re-initialization step [9].

The total energy functional in (3), the energy can be represented by variational level set formulation and the total energy functional is defined by

\[
E(\phi) = \mu E_m(\phi) + E_{ext}(\phi)
\]

where \( E_m(\phi) = \int_\Omega \frac{1}{2} |\nabla \phi|^2 dxdy \) is a metric term to characterize how to close a function \( \phi \) whose task is to remove the re-initialization step [9]. \( \mu > 0 \) is parameter controlling the effect of penalizing the deviation of \( \phi \) from a signed distance function. While the external energy \( E_{ext}(\phi) \) drives the zero level set toward the object boundaries. Initialized \( E_{ext}(\phi) = E_\phi(\phi) \), 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 (3) for a function \( \phi(x,y) \) can be written as

\[
E_{ext}(\phi) = \int_\Omega \lambda g \delta(\phi) |\nabla \phi| dxdy + \int_\Omega g H(-\phi) dxdy
\]

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 \( \phi \) is computed by the first term, and \( \lambda > 0 \) is constant.

We shall return to the variational level set denoted by \( \partial E/\partial \phi \) the Gateaux derivative of the functional \( E \) in (3). The gradient flow of evolution that minimizes the functional \( E \) can be represented by

\[
\frac{\partial \phi}{\partial t} = \frac{\partial E}{\partial \phi}
\]
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 \( \{x_i, i=1,...,n\} \) be an arbitrary set of n points within a data set can be obtained by
   2.1 Calculate the mean shift vector \( \mathbf{m}(x) \) by (2)
   2.2 Move kernel density window \( G(x) \) by \( \mathbf{m}(x) \) and re-compute the weighted mean
   2.3 Repeat the step 2.2 and 2.3 until the centralized x is not moved, gradient move closer to zero or less than the specific value (threshold). Assume that \( y_j \) is the initial value, the new movement of the centralized x is calculated as (9).

\[
y_{j+1} = m(y_j) + y_j - \frac{\sum_{k=1}^{n} \left( \mathbf{y}_{ij} - s_k \right) \left( g \left( \left\| \mathbf{y}_{ij} - F(s_k) \right\| \right) \right)}{\sum_{k=1}^{n} \left( g \left( \left\| \mathbf{y}_{ij} - F(s_k) \right\| \right) \right)}
\]

(9)

3) Image that passes the mean shift process is converted to gray images.
4) Create the initial contour, supposing \( \phi_0 \) 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 \( \phi_0 \) is defined as [9].

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

(10)

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 (6) from an arbitrary region \( \Omega_0 \) in the image domain \( \Omega \). The following parameters were used to test in (6), including \( \sigma, v, \lambda, \mu, \tau \), which took 360 iterations of curve evolution. An example is shown in Fig. 4 to illustrate how the progressive ROI detection proceeds.

Fig. 4 The progressive ROI detection proceeds of the proposed method with color image (a) an original image (b) mean shift segmentation (c) initialized level set function (d) 100 iterations of curve evolution (e) 360 iterations of curve evolution (f) final ROI results (white color).

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 defied by equation (11) and (12).

\[
\begin{align*}
\text{Precision} & = \frac{A \cap B}{A} \\
\text{Recall} & = \frac{B \cap A}{B}
\end{align*}
\]

(11)

(12)

Based on the precision and recall measurement, the relation of precision and recall illustrate on Fig 5. 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.5(c) shows a precision of 100% at recall 50% and Fig. 5(d) shows a precision of 50% at recall 100%.
In order to achieve the experiment, a variety of natural images in different modalities from Correl 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 \(\sigma = 1.5, v = 3.0\) and \(\sigma\)...
$v = 1.5$ were used for fitting the boundaries of objects with complex and less complex of image content, respectively. We used 20 images with a complex background and a less complex background in the experiment. The experimental results are shown in table 1 and table 2. Fig. 6 and Fig. 7 illustrate the results from application of proposed method in natural images that are typical images with less complex and complex of image content, respectively. The boundaries of object of interest were efficiently and accurately extracted by the proposed method.

## 5 Conclusion

In this paper, the active contour model based on variational level set method and its application in region of interest segmentation is investigated. 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. For parameter constancy, the mean-shift algorithm was used to reduce the sensitivity of parameter change of level set formulation. The results show that our method has achieved high accuracy when it is tested on 20 natural images, many with non-uniform sub-region and intensity inhomogeneities of foreground or background distractions.

References: