J. Vis. Commun. Image R. 43 (2017) 1–9

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Integrating machine learning with region-based active contour models in medical image segmentation ☆

Agus Pratondo [a,b,*], Chee-Kong Chui [c], Sim-Heng Ong [b]

[a] Dept. of Information Systems, Telkom University, Indonesia
[b] Dept. of Electrical and Computer Engineering, National University of Singapore, Singapore
[c] Dept. of Mechanical Engineering, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 30 August 2016
Revised 24 November 2016
Accepted 29 November 2016
Available online 2 December 2016

Keywords:
Machine learning
Active contour
Medical images
Segmentation

ABSTRACT

Region-based active contour models are effective in segmenting images with poorly defined boundaries but often fail when applied to images containing intensity inhomogeneity. The traditional models utilize pixel intensity and are very sensitive to parameter tuning. On the other hand, machine learning algorithms are highly effective in handling inhomogeneities but often result in noise from misclassified pixels. In addition, there is no objective function. We propose a framework which integrates machine learning with a region-based active contour model. Classification probability scores from machine learning algorithm, which are regularized using a non-linear function, are used to replace the pixel intensity values during energy minimization. In our experiments, we integrate the *k*-nearest neighbours and the support vector machine with the Chan-Vese method and compare the results obtained with the traditional methods of Chan-Vese and Li et al. The proposed framework gives better accuracy and less sensitive to parameter tuning.

1. Introduction

Image segmentation plays a significant role in computer vision and medical image analysis. Numerous segmentation methods have been proposed but none is universally applicable [1]. A number of modern approaches using energy minimization for image segmentation have been intensively studied, starting from the snake model introduced by Kass et al. [2]. A popular energy minimization approach is the level set method (LSM), which is widely used in medical image analysis. The basic idea of the LSM was first described in [3] and popularized by [4]. It was subsequently applied to image segmentation [5–7].

Generally, existing image segmentation models using level set methods can be grouped into two categories: edge-based models and region-based models [8–11]. Edge-based models utilize edge information while region-based models utilize a region descriptor to control the motion of the active contour [12]. Region-based models are not sensitive to objects with poorly defined boundaries but are sensitive to inhomogeneity of image intensities, i.e., the overlapping of the intensity ranges. Also, they are sensitive to parameter tuning [13,14] which are not desirable in practical use.

Another popular approach is the use of machine learning algorithms to classify each pixel based on training data. Many algorithms are described in the literature, e.g., the *k*-nearest neighbors (*k*-NN), support vector machine (SVM), extreme machine learning, etc [15,16]. These algorithms can handle complex patterns, but further post-processing such as morphological operations are often required to obtain the final solution without employing an objective function.

To overcome the limitation of those approaches, a number of studies includes classifier probability scores from overall pixel classification instead of pure intensity values [17–21]. Different from existing methods, our framework works in a simpler manner. The scores are written in a matrix in the range of $[0, 1]$ which are subsequently regularized by a non-linear function. Finally, the region-based active contour model proposed by Chan and Vese is applied to the matrix to find the optimal solution. The solution thus converges faster and is less sensitive to parameter tuning of the LSM.

2. Proposed framework

The proposed framework can be constructed from any classification algorithm and applied to any region-based model with an

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author at: Selaru lt. 3, School of Applied Science, Telkom University, Jl. Telekomunikasi No. 1, Terusan Buah Batu, Bandung, West Java 40257, Indonesia.
E-mail address: pratondo@gmail.com (A. Pratondo).

http://dx.doi.org/10.1016/j.jvcir.2016.11.019
1047-3203/

LSM. Here, we employ the k-NN and SVM to generate a matrix of classifier probability scores. The matrix is regularized and fed to the popular region-based active contour model proposed by Chan and Vese [12]. We present a review of these methods in this section.

2.1. Region-based active contour model

Chan and Vese introduced an energy functional $F(c_1, c_2, C)$ defined by

$$\begin{aligned} F(c_1, c_2, C) &= \mu \cdot \text{Length}(C) + \nu \cdot \text{Area}(inside(C)) \\ &+ \lambda_1 \int_{inside(C)} |u_0(x,y) - c_1|^2 dxdy \\ &+ \lambda_2 \int_{outside(C)} |u_0(x,y) - c_2|^2 dxdy \end{aligned} \tag{1}$$

where C is the evolving curve, c_1 and c_2 are, respectively, the values of u inside and outside of C, $\mu \geqslant 0$, $\nu \geqslant 0$, $\lambda_1, \lambda_2 > 0$ are constants, and u_0 is the input image. The minimization problem is expressed by

$$\inf_{c_1, c_2, C} F(c_1, c_2, C) \tag{2}$$

and can be solved by applying the level set method introduced by Osher and Sethian [4]. The curve C is implicitly represented by the zero level set of a Lipschitz function $\phi(x, y, t)$:

$$C(t) = \{(x,y) | \phi(x,y,t) = 0\} \tag{3}$$

where t is an artificial time variable. For $t = 0, \phi(x,y,0) = \phi_0(x,y)$ denotes the initial contour. After further derivation, Chan and Vese proposed a gradient flow

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= \delta_\epsilon(\phi) \left[\mu \ \text{div}\left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \nu - \lambda_1 (u_0 - c_1)^2 + \lambda_2 (u_0 - c_2)^2 \right] \\ &= 0 \text{ in } \Omega, \quad \frac{\delta_\epsilon(\phi)}{|\nabla \phi|} \frac{\partial \phi}{\vec{n}} = 0 \text{ on } \partial\Omega \end{aligned} \tag{4}$$

where δ_ϵ is the regularized Dirac function, Ω is a bounded open subset of $\mathbb{R}^2$ with $\partial\Omega$ its boundary, $\vec{n}$ denotes the exterior normal to the boundary, and $\frac{\partial \phi}{\vec{n}}$ denotes the normal derivative of ϕ at the boundary. Eq. (4) contains a number of parameters that should be tuned carefully [12].

2.2. Classifier probability scores

Classification algorithms are used to generate a matrix of classifier probability scores from the image u_0. In this paper, two classification algorithms, where each of the outputs has a different range, are investigated, namely, k-NN and SVM.

2.2.1. k-NN

k-NN provides scores in the range $[0, 1]$ which can be implemented easily using the fuzzy k-NN rule. This rule is derived from the fuzzy set [22] and the k-NN classifier in machine learning [23]. Given a reference set $X_R = \{x_i\}_{i=1}^{m_R}$, and a set of l-dimensional vectors $W = \{w_i\}_{i=1}^{m_R}, w_i = (w_{i,1}, w_{i,2}, \ldots, w_{i,l})$, where l and m_R are the number of classes and the number of elements in the reference set X_R, respectively, the following property holds on the fuzzy k-NN rule:

$$\sum_{j=1}^{l} w_{i,j} = 1, 0 \leqslant w_{i,j} \leqslant 1. \tag{5}$$

For $1 \leqslant i \leqslant m_R$ and $1 \leqslant j \leqslant l$, the value of $w_{i,j}$ is the membership value of the i-th object to class j. For a particular x to be classified, the set K of indices corresponding to the classes of k-nearest neighbours of x in X_R is obtained. Instead of applying the majority vote in the original k-NN, the fuzzy k-NN rule generates a fuzzy decision-vector computed by

$$v = \frac{1}{k} \sum_{s \in K} w_s. \tag{6}$$

The maximum $v_j, 1 \leqslant j \leqslant l$ where $v = (v_1, v_2, \ldots, v_l)$ is used to define the object class in the original k-NN.

2.2.2. Support vector machine

An SVM is a supervised learning method that classifies data using the best separation hyperplane which separates the data of a class from those of another, and gives the largest margin between these two classes [24,25]. The classification is performed using a sign function $\text{class}(x) = \text{sgn}(h(x))$ where $h(x)$ is the separating hyperplane for the two classes. For linearly separable data in dimension d, the hyperplane is expressed by

$$h(x) = \mathbf{w}_0^T \mathbf{x} + b_0 \tag{7}$$

where $\mathbf{w}_0 \in \mathscr{R}^d$ is the optimal weight vector, $\mathbf{x} \in \mathscr{R}^d$ is the data, and b_0 is the optimal bias. Since it may be difficult to separate the data in the original input space, a transformation of the data into a higher dimensional space through function φ is introduced. Then $h(x)$ can be expressed as

$$h(x) = \mathbf{w}_0^T \varphi(\mathbf{x}) + b_0. \tag{8}$$

Finding an explicit φ is often difficult; instead, kernel [26,27] $K(\mathbf{x}, \mathbf{x}_i)$ is used to compute directly the dot product expressed by

$$h(x) = \sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b_0 \tag{9}$$

where a_i is the estimated SVM parameter, and $y_i \in \{+1, -1\}$ is the desired class for the corresponding $\mathbf{x}_i$. The value of $h(x)$ is the SVM evaluation score and the sign is the predicted class [28].

Evaluation scores from classification algorithms generally fall in the range $[0, 1]$ or $(-\infty, +\infty)$. The scores of the k-NN are of the first type while those of the SVM are of the second. Methods are available to convert the second type to a prior probability score [29].

2.3. Regularization for classifier probability score

Originally, classifiers generate binary results by applying a hard limiter function to the probability scores. Let $s \in [0, 1]$ be a probability score and ρ a regularization function that maps s to a real value in $[0, 1]$. The traditional classifier generates binary results by

$$\rho(s) = \begin{cases} 1 & \text{if } s \geqslant \frac{1}{2}, \quad \text{(a)} \\ 0 & \text{if } s < \frac{1}{2}. \quad \text{(b)} \end{cases} \tag{10}$$

Instead of refining these binary scores using machine learning algorithms, we retain the probability scores which are processed further by applying any region-based active contour model. This aims to find an optimal solution where the function $\rho(s)$ can be simply expressed by

$$\rho(s) = s. \tag{11}$$

The plot of (10) is shown in Fig. (1) as ρ_1 and that of (11) as ρ_2. The former is binary while the latter is linear.

Based on our preliminary results, a non-linear function ρ approximately lying under ρ_2 for $s > 0.5$ and above ρ_2 for $s < 0.5$ leads to better results. It is worth noting the properties of a good ρ:

(a) the domain, s, as well as the range, $\rho(s)$, lie in $[0, 1]$,
(b) it is monotonically increasing,
(c) the following equations hold