Object Segmentation through Multiple Instance Learning

Iker Gondra¹, Tao Xu², David K. Y. Chiu², and Michael Cormier³

¹ Math, Stats, Computer Science Department, St. Francis Xavier University
Antigonish, Nova Scotia, Canada

igondra@stfx.ca

² School of Computer Science, University of Guelph
Guelph, Ontario, Canada

xut@uoguelph.ca, dchiu@cis.uoguelph.ca

³ School of Computer Science, University of Waterloo
Waterloo, Ontario, Canada
m4cormie@uwaterloo.ca

Abstract. An object of interest (OOI) in an image usually consists of visually coherent regions that, together, encompass the entire OOI. We use Multiple Instance Learning (MIL) to determine which regions in an over-segmented image are part of the OOI. In the learning stage, a set of over-segmented images containing, i.e., positive, and not containing, i.e., negative, an instance of the OOI is used as training data. The resulting learned prototypes represent the visual appearances of OOI regions. In the OOI segmentation stage, the new image is over-segmented and regions that match prototypes are merged. Our MIL method does not require prior knowledge about the number of regions in the OOI. We show that, with the coexistence of multiple prototypes corresponding to the regions of the OOI, the maxima of the formulation are good estimates of such regions. We present initial results over a set of images with a controlled, relatively simple OOI.

Keywords: Image segmentation, object recognition, multiple instance learning, diverse density, adaptive kernel.

1 Introduction

Homogeneity, which is largely related to local visual information extracted from an image and reflects how uniform a region is, plays a very important role in weak (traditional) image segmentation since the result of such segmentation is visually coherent regions. Algorithms that follow such traditional approach (e.g., [2], [7], [13]) perform well in narrow domains, where the variability of low-level visual content is limited. Unfortunately, in broader domains, homogeneous regions do not necessarily correspond to semantically meaningful units. In the case that there is a particular object of interest (OOI), i.e., we are not concerned with the quality of segmentation of other objects in the scene, a number of learning-based approaches have been proposed. Although these methods incorporate high-level knowledge, the major disadvantage is their lack of flexibility due to the

assumption that such knowledge in the form of, e.g., the full object model (e.g., [1], [16]), partial object model information such as its structure or shape of its probability density function (e.g., [4], [8]), hand-specified information (e.g., [20]), is provided in advance.

Multiple Instance Learning (MIL) [5] has been recognized useful for a wide spectrum of applications. In particular, in the context of images, applications of MIL include content-based image retrieval (e.g., [12], [22]), image classification [14], object detection ([6],[17],[20]), and our previous work on OOI segmentation ([3],[9],[10]). MIL generalizes standard supervised learning for problems with incomplete knowledge about the labels of training samples. In MIL, each training sample consists of a bag (set) of instances. A bag is labeled positive if at least one instance in the bag is positive, or negative otherwise. Different from standard supervised learning, we are only given the labels of the bags but not the labels of the instance(s). Given such a training set, the goal of MIL is to find the prototype(s) responsible for positive labeling. Or briefly, to find what all the positive bags “have in common”. Maximum diverse density (MDD) [15] is a popular method for solving the MIL problem. The intuition behind MDD is that, in the instance space, areas with high density of distinct (diverse) positive bags are more likely to embody a prototype. Rooted in the formulation, the performance of MDD depends heavily on prior knowledge about the number of prototypes that is often unknown in image analysis tasks. In this paper we extend our previous work on an adaptive kernel diverse density estimate (AKDDE)[21] by applying it to the problem of OOI segmentation and show that, upon the coexistence of multiple prototypes (corresponding to the regions in the OOI) the method generates correct estimates.

2 Preliminaries

2.1 Maximum Diverse Density

Maximum diverse density (MDD) [15] is based on the fact that a prototype should be close to at least one instance in each positive bag while far from all the negative ones. A maximum likelihood estimate can be formulated to find the most likely estimate(s) of the prototype(s). Assuming a unique prototype $\mathbf{t} \in \mathbb{R}^d$ accounting for the labels for all bags, $\mathbf{t}$ is located in a region that is not only dense in positive instances but also diverse in that it includes at least one instance from each positive bag.

Formally, let $\mathcal{B}_i^+ = \{\mathbf{b}_{i,1}^+, \mathbf{b}_{i,2}^+, \dots\}$ be the i^{th} positive bag with $\mathbf{b}_{i,j}^+ \in \mathbb{R}^d$ as the j^{th} instance, and similarly defined for negative bags. We can find $\mathbf{t}$ by examining a random vector $\mathbf{x} \in \mathbb{R}^d$ that maximizes probability of being prototype $\mathbf{t}$, conditioned on observations $\mathcal{B} = \{\mathcal{B}^+, \mathcal{B}^-\} = \{\mathcal{B}_1^+, \mathcal{B}_2^+, \dots, \mathcal{B}_n^+\}, \{\mathcal{B}_1^-, \mathcal{B}_2^-, \dots, \mathcal{B}_m^-\}\}$, i.e., $Pr(\mathbf{x} = \mathbf{t} | \mathcal{B}_1^+, \mathcal{B}_2^+, \dots, \mathcal{B}_n^+, \mathcal{B}_1^-, \mathcal{B}_2^-, \dots, \mathcal{B}_m^-)$. Applying Bayes’ rule and assuming equal prior probabilities $Pr(\mathbf{x} = \mathbf{t})$ everywhere in the instance space, the hypothesis is the one that maximizes the following likelihood

$$h_{ML} = \arg \max_{\mathbf{x}} Pr(\mathcal{B}_1^+, \mathcal{B}_2^+, \dots, \mathcal{B}_n^+, \mathcal{B}_1^-, \mathcal{B}_2^-, \dots, \mathcal{B}_m^- | \mathbf{x} = \mathbf{t}).$$

With an additional assumption that bags are conditionally independent of each other given $\mathbf{t}$, it becomes

$$h_{ML} = \arg \max_{\mathbf{x}} \prod_{i=1}^n Pr(\mathcal{B}_i^+ | \mathbf{x} = \mathbf{t}) \prod_{i=1}^m Pr(\mathcal{B}_i^- | \mathbf{x} = \mathbf{t}).$$

Again with the assumption of a uniformly distributed probability of $\mathbf{t}$ over the entire instance space, applying Bayes' rule once more, it becomes

$$h_{ML} = \arg \max_{\mathbf{x}} \prod_{i=1}^n Pr(\mathbf{x} = \mathbf{t} | \mathcal{B}_i^+) \prod_{i=1}^m Pr(\mathbf{x} = \mathbf{t} | \mathcal{B}_i^-). \quad (1)$$

This is the definition of MDD. Three feasible probabilistic models were suggested to estimate $Pr(\mathbf{x} = \mathbf{t} | \mathcal{B}_i)$ [15].

As a maximum likelihood estimate, the performance of MDD depends heavily on prior knowledge about the number of prototypes, which may not be known. In the absence of this knowledge, MDD may end up with false estimates. As indicated in [15], applying the single prototype formulation given in Eq.1 to the problem with two (or more) distinct prototypes results in an estimate close to neither one of them but somewhere in between. This was also verified in the experiments in [11,12].

2.2 Adaptive Kernel Diverse Density Estimate

We revise the definition of diverse density [15] (Section 2.1) as *the probability density function of diverse positive bags*. Let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ where $\mathbf{x}_i \in \mathbb{R}^d$ be a set of independent and identically distributed samples drawn from an unknown probability density function (*pdf*) $f(\mathbf{x})$, then its density estimate is given by

$$\hat{f}(\mathbf{x}) = \frac{c}{nh^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right), \quad (2)$$

where c is a normalization constant, and $h > 0$ is the bandwidth (of the estimating window) that controls the smoothness of the estimate function. If $K(\cdot)$ is a *pdf*, i.e., non-negative everywhere and integrating to one, so is $\hat{f}(\mathbf{x})$. In such a case, $\hat{f}(\mathbf{x})$ is a *kernel density estimate* (KDE) [18]. If h is not fixed but allowed to vary adaptive to the distribution of data, the method is called *adaptive kernel density estimate* (AKDE) [19]. Formally, an AKDE using k nearest neighbors can be written as

$$\hat{f}(\mathbf{x}) = \frac{c}{n} \sum_{i=1}^n \frac{1}{h_i^d} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_i}\right), \quad (3)$$

where h_i is dynamically decided by the distance from $\mathbf{x}_i$ to the k^{th} nearest neighbor.

Based on the intuition that a prototype of MIL falls into the region that is dense of diverse positive bags, we first redefine diverse density as the *probability*

density function of diverse positive bags, then the AKDE of diverse positive bags is written as

$$\hat{f}_{B^+}(\mathbf{x}) = \frac{c^+}{n} \sum_{i=1}^n \frac{1}{h_i^d} K\left(\frac{\mathbf{x} - \mathbf{b}_{i,nrst}^+}{h_i}\right), \quad (4)$$

where $\mathbf{b}_{i,nrst}^+$ denotes the nearest instance in $\mathcal{B}_i$ to $\mathbf{x}$, c^+ is the normalization constant, and h_i is the maximum instance-bag distance from $\mathbf{b}_{i,nrst}^+$ to all $n-1$ positive bags excluding $\mathcal{B}_i^+$ itself, i.e.,

$$h_i = \max_{\mathcal{B}_j^+ \in \mathcal{B}^+, i \neq j} \{dist(\mathbf{b}_{i,nrst}^+, \mathcal{B}_j^+)\}, \quad (5)$$

where the (Euclidean) instance-bag distance is defined as

$$dist(\mathbf{x}, \mathcal{B}_i^+) = \min_{\mathbf{b}_{i,j}^+ \in \mathcal{B}_i^+} \{\|\mathbf{x} - \mathbf{b}_{i,j}^+\|_2\}. \quad (6)$$

In short, Eq.4 together with Eq.5 and Eq.6 states that the *adaptive kernel diverse density estimate* (AKDDE) [21] of positive bags at $\mathbf{x}$ is locally decided by the bound that contains the nearest instance from each positive bag. In the case that each positive bag contains multiple coexisting and heterogeneous positive instances, MDD's single prototype formulation (used when the number of prototypes is unknown) may end up with false estimates. However, AKDDE can still generate correct estimates.

Because of the homogeneity of negative bags (i.e., all instances are negative), it is safe to treat all negative bags as one, i.e., $\mathcal{B}_c^- = \mathcal{B}_1^- \cup \mathcal{B}_2^-, \dots, \cup \mathcal{B}_m^-$. For the combined negative bags $\mathcal{B}_c^-$, the concept of “diversity” is not necessary. Hence, the *pdf* can be estimated in the conventional way by Eq.3. Moreover, because the nearest negative instance matters most on the decision of an estimate, the bandwidth for the density estimate can be simply chosen as the distance from $\mathbf{x}$ to the nearest negative instance in $\mathcal{B}_c^-$. Summarizing all of the above, we have the KDE of $\mathcal{B}_c^-$ as

$$\hat{f}_{B^-}(\mathbf{x}) = \frac{c^-}{|\mathcal{B}_c^-| h_{\mathcal{B}_c^-}^d} \sum_{\mathbf{b}_{c,i}^- \in \mathcal{B}_c^-} K\left(\frac{\mathbf{x} - \mathbf{b}_{c,i}^-}{h_{\mathcal{B}_c^-}}\right), \quad (7)$$

where

$$h_{\mathcal{B}_c^-} = dist(\mathbf{x}, \mathcal{B}_c^-) = \min_{\mathbf{b}_{c,i}^- \in \mathcal{B}_c^-} \{\|\mathbf{x} - \mathbf{b}_{c,i}^-\|_2\} \quad (8)$$

where c^- in Eq.7 is the normalization constant, and $|\cdot|$ represents the cardinality operator. Clearly, if $K(\cdot)$ is differentiable on its domain, so is $\hat{f}_{B^-}(\mathbf{x})$. The objective function is defined as

$$J(\mathbf{x}) = \hat{f}_{B^+}(\mathbf{x}) - \gamma \hat{f}_{B^-}(\mathbf{x}). \quad (9)$$

Here γ is a regulation parameter that weights the impact of negative samples so as to control the degree of deviation of the estimate away from the maximum

of $\hat{f}_{\mathcal{B}^+}(\mathbf{x})$. Obviously, an estimate $\mathbf{t}$ in the instance space having the maximum difference between $\hat{f}_{\mathcal{B}^+}(\mathbf{x})$ and $\hat{f}_{\mathcal{B}^-}(\mathbf{x})$ is a good estimate of a prototype, i.e., $\mathbf{t} = \arg \max_{\mathbf{x}} J(\mathbf{x})$. To ease the mathematical analysis of modes as well as the computation of gradient, we use the Gaussian kernel with equal variance σ in all dimensions for both Eq.4 and Eq.7, i.e.,

$$K(\mathbf{x}) = \left(\frac{1}{2\pi\sigma} \right)^{d/2} e^{-\left(\frac{(\mathbf{x}-\mu)^T(\mathbf{x}-\mu)}{2\sigma^2} \right)}. \quad (10)$$

3 OOI Segmentation

3.1 Learning Stage

Let an image $\mathcal{I}$ be defined as a non-empty set of pixels. Given a set of training images $\mathcal{I} = \{\mathcal{I}^+, \mathcal{I}^-\} = \{\{\mathcal{I}_1^+, \mathcal{I}_2^+, \dots, \mathcal{I}_n^+\}, \{\mathcal{I}_1^-, \mathcal{I}_2^-, \dots, \mathcal{I}_m^-\}\}$, where the positive or negative labeling indicates whether or not the image contains an instance of the OOI, each $\mathcal{I}_i$ (regardless of the labeling) is segmented into a set of visually homogeneous regions $\{\mathcal{R}_{i,1}, \mathcal{R}_{i,2}, \dots\}$ where $\mathcal{R}_{i,j} \subset \mathcal{I}_i$ and $\mathbf{b}_{i,j} \in \mathbb{R}^d$ is the visual feature vector representation of $\mathcal{R}_{i,j}$. Then, the training set $\mathcal{B} = \{\mathcal{B}^+, \mathcal{B}^-\} = \{\mathcal{B}_1^+, \mathcal{B}_2^+, \dots, \mathcal{B}_n^+\}, \{\mathcal{B}_1^-, \mathcal{B}_2^-, \dots, \mathcal{B}_m^-\}$, where $\mathcal{B}_i^+ = \{\mathbf{b}_{i,1}^+, \mathbf{b}_{i,2}^+, \dots\}$, and similarly defined for negative bags, is formed in preparation for AKDDE. After the optimization of the AKDDE objective function (Eq.9) takes place on this MIL formulation, we obtain a set of prototypes $\mathcal{P} = \{\mathbf{t}_1, \mathbf{t}_2, \dots\}$ where $\mathbf{t}_i \in \mathbb{R}^d$ represents the visual appearance of pixels in a common instance, i.e., OOI fragment, among the bags, i.e., images. Algorithm 1 summarizes the learning stage.

3.2 Segmentation Stage

Given a new image $\mathcal{I}$, it is segmented into a set of visually homogeneous regions $\{\mathcal{R}_1, \mathcal{R}_2, \dots\}$ where $\mathcal{R}_i \subset \mathcal{I}$ and $\mathbf{b}_i \in \mathbb{R}^d$ is the visual feature vector representation of $\mathcal{R}_i$. Then, for each $\mathbf{b}_i$, $k^* = \arg \min_{k \in \{1, 2, \dots, |\mathcal{P}|\}} (\|\mathbf{b}_i - \mathbf{t}_k\|^2)$. If $\|\mathbf{b}_i - \mathbf{t}_{k^*}\|^2 < \delta$ then $\mathcal{R}_i$ is assumed to be an instance of the k^* th part of the OOI. The set of such regions (if any) are then merged. Algorithm 2 summarizes the segmentation stage.

4 Experiments

The dataset consists of JPEG-format photos of an arrangement of Lego bricks, which we call the OOI. The OOI consists of four visually coherent regions: red, blue, green and yellow bricks. The dataset consists of two parts. The first part is the training set, which contains 50 “positive” images depicting the OOI in a variety of settings and 57 “negative” images which do not contain the OOI. The second part is the test set, consisting of 74 images, also placing the OOI in a variety of settings. Examples of these images are given in Table 1.

Algorithm 1. Learning Stage

Input: $\mathcal{I} = \{\mathcal{I}^+, \mathcal{I}^-\}$, γ , δ
Output: $\mathcal{P} = \{\mathbf{t}_1, \mathbf{t}_2, \dots\}$

for each $\mathcal{I}_i^+ \in \mathcal{I}^+$ **do**
 $\mathcal{B}_i^+ \leftarrow \emptyset$
Oversegment to generate $\{\mathcal{R}_{i,1}^+, \mathcal{R}_{i,2}^+, \dots\}$
for each $\mathcal{R}_{i,j}^+$ **do**
Extract $\mathbf{b}_{i,j}^+$
 $\mathcal{B}_i^+ \leftarrow \mathcal{B}_i^+ \cup \{\mathbf{b}_{i,j}^+\}$
 $\mathcal{B}_c^- \leftarrow \emptyset$
for each $\mathcal{I}_i^- \in \mathcal{I}^-$ **do**
Oversegment to generate $\{\mathcal{R}_{i,1}^-, \mathcal{R}_{i,2}^-, \dots\}$
for each $\mathcal{R}_{i,j}^-$ **do**
Extract $\mathbf{b}_{i,j}^-$
 $\mathcal{B}_c^- \leftarrow \mathcal{B}_c^- \cup \{\mathbf{b}_{i,j}^-\}$
 $\mathcal{P} \leftarrow \emptyset$
Form objective function $J(\mathbf{x})$ (Eq. 9)
for each local maximum $\mathbf{t}_i$ of $J(\mathbf{x})$ **do**
if $J(\mathbf{t}_i) > \delta$ **then**
 $\mathcal{P} \leftarrow \mathcal{P} \cup \{\mathbf{t}_i\}$

Algorithm 2. Segmentation Stage

Inputs: $\mathcal{I}$, $\mathcal{P}$, δ
Output: OOI segmentation $\mathcal{S}$

$\mathcal{S} \leftarrow \emptyset$
Oversegment $\mathcal{I}$ to generate $\{\mathcal{R}_1, \mathcal{R}_2, \dots\}$
for each $\mathcal{R}_i$ **do**
Extract $\mathbf{b}_i$
 $k^* = \arg \min_{k \in \{1,2,\dots,|\mathcal{P}|\}} (\|\mathbf{b}_i - \mathbf{t}_k\|^2)$
if $\|\mathbf{b}_i - \mathbf{t}_{k^*}\|^2 < \delta$ **then**
 $\mathcal{S} \leftarrow \mathcal{S} \cup \mathcal{R}_i$

Table 1. Sample training images: positive (leftmost two), negative. OOI in positive images consists of four visually coherent regions: red, blue, green, yellow bricks.

The training set is provided as input to Algorithm 1. The initial over segmentation of each image is obtained by partitioning the image into a grid of equally sized (4x4) blocks. The blocks are processed to produce feature vectors, describing average color (CIELUV colorspace) and texture (Haralick contrast

and correlation measures). The same method is used to generate initial over-segmentation of each image in the test set, which is provided as input to Algorithm 2. The value for the threshold parameter δ in Algorithm 1 is chosen so that only the top 20% of maxima of the objective function are part of the resulting set $\mathcal{P}$ of prototypes. Afterwards, the same value of δ is used in Algorithm 2. We compare the segmentation performances of both traditional MDD and AKDDE. For MDD, the same initial over-segmentation method and feature vector representation is used but, in the training phase, the MDD objective function (Eq.1) is used instead. The same threshold selection method is used for MDD.

4.1 Qualitative Results

Table 2 shows distance maps of randomly selected images from the test set. A distance map labels each (4x4) block of pixels with the Euclidean distance between the block's feature vector and a region prototype (generated by either MDD or AKDDE). It can be observed that, qualitatively, AKDDE prototypes tend to better distinguish OOI regions from background regions than do MDD prototypes. These results are consistent with what we expected would occur. That is, notice that, in the case that each positive bag (i.e., image) contains multiple coexisting but heterogeneous positive instances, i.e., regions, MDD's single prototype formulation, which is the default in the absence of knowledge about the number of prototypes, ends up with false estimates. However, AKDDE still generates correct estimates. This is because, different from MDD which may result in an estimate close to neither one of the multiple distinct prototypes but somewhere in between, the use of a localized estimating window allows AKDDE to generate estimates that are close to the prototypes.

4.2 Quantitative Results

We quantify the quality of the OOI segmentation (in a test image $\mathcal{I}$) $\mathcal{S}_{machine}$ generated by Algorithm 2 by comparing it with an ideal, i.e., human-generated or groundtruth $\mathcal{S}_{human}$, segmentation with the measure

$$Q = \frac{TP}{|\mathcal{S}_{human}|} \times \frac{TN}{|\mathcal{I} - \mathcal{S}_{human}|}$$

where $|\cdot|$ represents the cardinality operator and TP is the number of true positives, i.e., pixels in both $\mathcal{S}_{machine}$ and $\mathcal{S}_{human}$, $TP = |\mathcal{S}_{machine} \cap \mathcal{S}_{human}|$, and TN is the number of true negatives, i.e., pixels that are both not in $\mathcal{S}_{machine}$ and not in $\mathcal{S}_{human}$, $TN = |(\mathcal{I} - \mathcal{S}_{machine}) \cap (\mathcal{I} - \mathcal{S}_{human})|$.

Let p be the proportion of times that, for a particular $\mathcal{I}$, the Q value obtained with the AKDDE-based algorithm is larger than the corresponding Q value obtained with the MDD-based algorithm. The AKDDE-based algorithm

Table 2. Sample test images with distance maps. The lighter the color of a block is, the closer its feature vector is to prototype. AKDDE’s top two prototypes coincide with two OOI regions: yellow and blue Lego bricks respectively. MDD’s top two prototypes cannot discriminate between OOI and background, they are not OOI region prototypes.

Image	MDD Prototypes	AKDDE Prototypes

performs better (larger Q value) than the MDD-based algorithm on 50 of the 74 test images. Thus, $\hat{p} = \frac{50}{74}$. The hypothesis test

$$H_0 : p = 0.5 \quad H_1 : p > 0.5$$

with test statistic

$$z = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{74}}} = \frac{\frac{50}{74} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{74}}} = 3.02$$

allows us to reject H_0 with a p-value of 0.0013. Thus, the AKDDE-based algorithm will perform better in most cases (the mean and median improvements are both 0.063, though the standard deviation is 0.107).

5 Conclusions

We applied our work on AKKDE to the OOI segmentation problem. The method was used to determine which regions in an over-segmented image are part of an OOI. The method does not require prior knowledge on number of regions in the OOI. Experimental results with a controlled, relatively simple OOI are encouraging. Our short-term future work will be on testing the segmentation framework over images containing different (higher visual variability) OOI instances.

Acknowledgments. This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

1. Bhanu, B., Lee, S., Das, S.: Adaptive image segmentation using genetic and hybrid search methods. *IEEE Transactions on Aerospace and Electronic Systems* 31(4), 1268–1291 (1995)
2. Carson, C., Belongie, S., Greenspan, H., Malik, J.: Blobworld: color and texture-based image segmentation using EM and its applications to image querying and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(8), 1026–1038 (2002)
3. Cormier, M., Gondra, I.: Supervised object segmentation using visual and spatial features. In: *Proceedings of International Conference on Image Processing, Computer Vision and Pattern Recognition*, pp. 557–563 (2011)
4. Crandall, D., Huttenlocher, D.: Weakly supervised learning of part-based spatial models for visual object recognition. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006, Part I*. LNCS, vol. 3951, pp. 16–29. Springer, Heidelberg (2006)
5. Dietterich, T.G., Lathrop, R.H., Lozano Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence* 89(1-2), 31–71 (1997)
6. Dollár, P., Babenko, B., Belongie, S., Perona, P., Tu, Z.: Multiple component learning for object detection. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part II*. LNCS, vol. 5303, pp. 211–224. Springer, Heidelberg (2008)

7. Fan, J., Yau, D.K.Y., Elmagarmid, A.K., Aref, W.G.: Automatic image segmentation by integrating color-edge extraction and seeded region growing. *IEEE Transactions on Image Processing* 10(10), 1454–1466 (2001)
8. Fergus, R., Perona, P., Zisserman, A.: Weakly supervised scale-invariant learning of models for visual recognition. *International Journal of Computer Vision* 71, 273–303 (2007)
9. Gondra, I., Alam, F.I.: Learning spatial relations for object-specific segmentation using bayesian network model. *Signal, Image and Video Processing* (in press)
10. Gondra, I., Alam, F.I.: Learning-based object segmentation using regional spatial templates and visual features. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) *ICCVG 2012*. LNCS, vol. 7594, pp. 397–406. Springer, Heidelberg (2012)
11. Gondra, I., Xu, T.: Adaptive mean shift-based image segmentation using multiple instance learning. In: *Proceedings of the Third IEEE International Conference on Digital Information Management*, pp. 716–721 (2008)
12. Gondra, I., Xu, T.: Image region re-weighting via multiple instance learning. *Signal, Image and Video Processing* 4(4), 409–417 (2010)
13. Guy, G., Medioni, G.: Inferring global perceptual contours from local features. *International Journal of Computer Vision* 20(1-2), 113–133 (1996)
14. Maron, O., Lakshmi Ratan, A.: Multiple-instance learning for natural scene classification. In: *Proceedings of the 15th International Conference on Machine Learning*, vol. 15, pp. 341–349 (1998)
15. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. In: Jordan, M.I., Kearns, M.J., Solla, S.A. (eds.) *Advances in Neural Information Processing Systems*, vol. 10, pp. 570–576. MIT Press (1998)
16. Peng, J., Bhanu, B.: Closed-loop object recognition using reinforcement learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(2), 139–154 (1998)
17. Qi, Z., Xu, Y., Wang, L., Song, Y.: Online multiple instance boosting for object detection. *Neurocomputing* 74(10), 1769–1775 (2011)
18. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27, 832–837 (1956)
19. Terrell, D.G., Scott, D.W.: Variable kernel density estimation. *Annals of Statistics* 20, 1236–1265 (1992)
20. Viola, P., Platt, J.C., Zhang, C.: Multiple instance boosting for object detection. In: *Proceedings of Neural Information Processing Systems*, vol. 18, pp. 1417–1424 (2005)
21. Xu, T., Gondra, I., Chiu, D.: Adaptive kernel diverse density estimate for multiple instance learning. In: Perner, P. (ed.) *MLDM 2011*. LNCS, vol. 6871, pp. 185–198. Springer, Heidelberg (2011)
22. Yang, C., Lozano-Pérez, T.: Image database retrieval with multiple-instance learning techniques. In: *Proceedings of IEEE International Conference on Data Engineering*, pp. 233–243 (2000)