Semantic, Automatic Image Annotation Based On Multi-Layered Active Contours and Decision Trees

Joanna Isabelle OLSZEWSKA
School of Computing and Engineering
University of Huddersfield
Queensgate, Huddersfield, HD1 3DH, UK

***Abstract*—In this paper, we propose a new approach for automatic image annotation (AIA) in order to automatically and efficiently assign linguistic concepts to visual data such as digital images, based on both numeric and semantic features. The presented method first computes multi-layered active contours. The first-layer active contour corresponds to the main object or foreground, while the next-layers active contours delineate the object's subparts. Then, visual features are extracted within the regions segmented by these active contours and are mapped into semantic notions. Next, decision trees are trained based on these attributes, and the image is semantically annotated using the resulting decision rules. Experiments carried out on several standards datasets have demonstrated the reliability and the computational effectiveness of our AIA system.**

***Keywords*—automatic image annotation; natural language tags; decision trees; semantic attributes; visual features; active contours; segmentation; image retrieval**

I. INTRODUCTION

With the increasing amount of available visual digital data, labeling [21] or searching [24] for an image remains a challenging task, not only because it necessitates a computationally efficient management of image storage and indexing processes, but also it requires the investigation of the semantic gap, i.e. the difference between the visual image representation and its linguistic description.

For this purpose, several image retrieval (IR) techniques have been developed in the literature. In the tag-based retrieval approach, images are retrieved on the basis of the textual information which has been beforehand manually associated to the images, whereas in the content-based image retrieval (CBIR) method, images are retrieved on the basis of low-level visual information automatically extracted from the images [1].

In this work, we focus on the most recent approach called Automatic Image Annotation (AIA), whose main steps are the automatic extraction of visual features from images and their automatic, semantic labeling. This latter step usually requires a training to learn the semantic concepts from image samples and to use these concepts to label new images. Thus, these images, which are automatically annotated with semantic labels, can be retrieved by users providing keywords such as in the tag-based retrieval approach rather than a query image as it is the case for CBIR. Hence, AIA combines the advantages of both tag-based and content-based image retrieval approaches.

Whereas most of the image annotation approaches are still manual [6] for both object delineation and labeling such as LabelMe [27], some automatic image annotation techniques have been recently developed [32].

AIA systems mainly use graph-based algorithms, e.g. Normalized cut (N-cut) [29] or region growing methods [5], as segmentation methods. In general, these approaches are appropriate for segmenting background objects, but not a main object itself, since they usually tend to oversegment the studied image. This results in the loss of the main object's entirety and in a mix of foreground's parts with the background ones. Hence, the features extracted from these resulting regions are not specific enough to characterize the main object. The active contour approach [10] does not present this drawback as it delineates the boundaries of the entire object. However, active contours have been used up to now only for semi-automatic graphic annotation processes [8], [17], [9], thus not providing fully automatic graphic nor semantic annotations, as these specific implementations found in the literature present weaknesses in presence of noise and do not offer any semantic computational framework.

In AIA, the semantic labeling of images usually implies the use of classifiers such as artificial neural network (ANN) [25], [11], or support vector machine (SVM) [4], [8], but these methods require computationally expensive training. Decision trees (DT) have been proven to be much faster and allow both categorical and numerical values [18]. The classification could thus rely on semantic rules and visual features.

In this paper, we propose a new fully automatic image annotation method based on efficiently implemented active contours and decision trees. Hence, our approach consists of the automatic recursive image segmentation in multiple layers using multi-feature active contours and the automatic semantic labeling of the image based on decision trees.

While being an unsupervised segmentation technique, the multi-layered multi-feature active contour approach does not

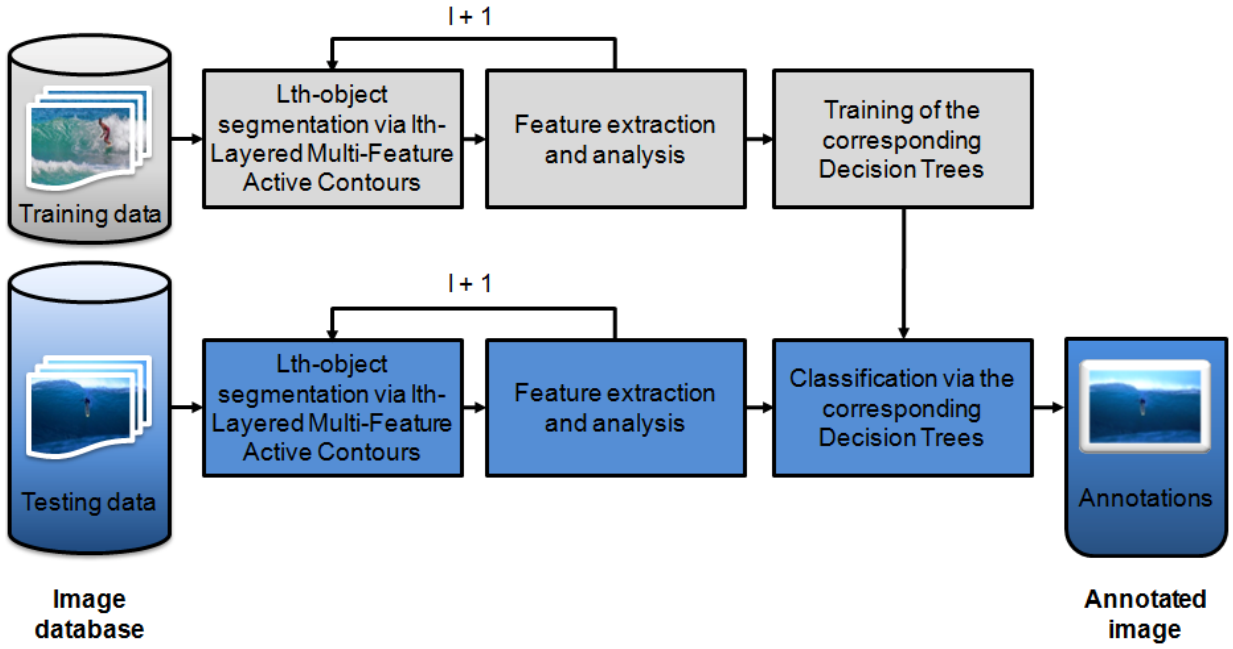

Fig. 1. Architecture of our Automatic Image Annotation process.

use any prior knowledge about the foreground unlike top-down segmentation methods [2] and reaches a semantically coherent segmentation of the objects more accurately than the bottom-up segmentation techniques [28] and faster than the combined ones [13], [14] or [12].

On the other hand, our segmentation method also provides the background region. However, in this work, we only exploit the information about the main object and its subparts, in order to process the training of the corresponding decision trees and the automatic labeling of the dataset images in a more computational efficiently way than background-based systems like [8].

AIA approaches consider usually that the main object is in the center part of the image [11] or constitutes the largest region of the image [25]. Because of these constraining assumptions on the position or importance of the main object, these systems cannot classify nor annotate an image properly if the object appears in another part of the image. This is not the case for the adopted multi-feature active contour approach which allows the detection of any object in any part of the image [19], [22].

Moreover, when compared to [7], our multi-layered multi-feature active contour method provides not only semantically coherent objects but also a semantically meaningful sub-object decomposition without any training.

The contributions of this paper are as follows:

- the use of active contours into a computationally efficient, full AIA system;
- the introduction of multi-layered active contours based on the robust and effective multi-feature active contours, in order (i) to precisely and automatically segment the image into background and semantically meaningful foreground regions and (ii) to extract coherent and semantically meaningful sub-regions of the extracted main object;
- the proposed architecture of the novel automatic image annotation process involving decision trees relying on hierarchic, semantic attributes derived from multi-stage visual features, which ones have been extracted from the image regions segmented by the corresponding multi-layered multi-feature active contours.

The paper is structured as follows. In Section II, we present our Automatic Image Annotation (AIA) approach based on the unsupervised, semantic labeling of an image under investigation, given visual features of objects extracted from the segmented image by means of multi-layer active contours and given trained decision trees. The resulting annotation system has been successfully tested on a challenging database containing real-world images with very close semantic classes as reported and discussed in Section III. Conclusions are drawn up in Section IV.

II. Our Proposed Annotation System

In this section, we describe our AIA system illustrated in Fig. 1, which performs both the automatic visual segmentation of the image and its automatic semantic annotation. The main steps of the process are the multi-layered partition of the image in terms of background, foreground and foreground´s semantically meaningful sub-regions (Section II.A), the extraction of the corresponding metric features from these delineated regions as well as the definition of the semantic attributes based on the visual features (Section II.B), and the

(a) (b) (c) (d)

Fig. 2. Feature extraction with (a) N-cut method; (b) edge map; (c) first-layer multi-feature active contour; (d) second-layer multi-feature active contour. Best viewed in color.

(a) (b) (c) (d)

Fig. 3. Image segmentation with multi-layered multi-feature active contours: (a) main-object segmentation with the first-layer multi-feature active contour; (b) main-object extraction; (c) sub-object segmentation with the second-layer multi-feature active contour; (d) sub-object extraction. Best viewed in color.

labeling of the image followed by the final online annotation of the image using offline-trained decision trees (Section II.C).

A. *Multi-Layered Multi-Feature Active Contours*

Active contours [10] are deformable two-dimensional closed curves that evolve in the image plane from a given initial position to the foreground boundaries characterizing thus the shape and the position of the object of interest.

In this work, we have chosen multi-feature active contours [19] to segment images in order to provide visual information to the system. Multi-feature active contours are particularly suitable for image annotations, since they can precisely segment images in semantically meaningful parts. Indeed, they could extract main object(s) entirely and very efficiently as illustrated in Figs. 2(c)-(d). It is worth to note that this is not the case of most of the state-of-art segmentation methods such as N-cut [29], [3] or edge detection [31], which usually suffer from over-segmentation and do not necessarily grasp the objects of interest as shown in Figs. 2 (a)-(b), respectively.

Other major advantages of multi-feature active contours [19] are as follows:

- they use both region-based and edge-based image representation;
- they combine the positive properties of bottom-up and top-down approaches, whereas do not require any prior knowledge, in order to not constrain the contour evolution, leading to the accurate delineation of main objects with highly varying shape and appearance;
- they are robust towards noise, clutter, and complex backgrounds.

Multi-feature active contour representation consists in a parametric plane curve $\mathcal{C}(s) : [0,1] \rightarrow \mathbb{R}^2$ modeled by a B-Spline formalism, while its evolution is guided by internal forces (α: elasticity, β: rigidity) described by the curve´s mechanical properties and the external force Ξ resulting from multiple characteristics of the image under study, computed by the dynamic equation as follows:

$$\mathcal{C}_t(s,t) = \alpha\, \mathcal{C}_{ss}(s,t) - \beta\, \mathcal{C}_{ssss}(s,t) + \Xi. \qquad (1)$$

The external force Ξ based on the Multi-Feature Vector Flow (MFVF) [19] has a large capture range as well as a bidirectional convergence and owns additional capacities related to the properties of the extracted features. Equation (1) sets the general framework of the multi-feature active contours, allowing the use of an extensible number of different features describing the shape and appearance of the objects of interest [19].

Multi-layered multi-feature active contours segment an image I into several parts or equivalently in $l + i$ layers, namely, the background ($i = 0$), the foreground ($i = 1$) and the foreground sub-regions ($i = 2$). The segmentation is recursively performed by applying *ith*-times multi-feature active contours. In the first step, the multi-layered multi-feature active contours divide the image into background/foreground such as illustrated in Fig. 3 (a). The background corresponds to the layer l, while the main object or foreground F_{l+1} lies in the layer $l + 1$ shown in Fig. 3 (b).

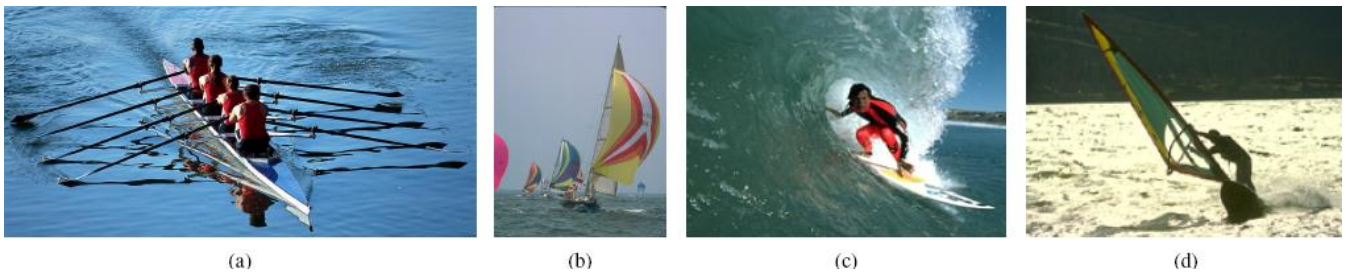

Fig. 4. Examples of four semantic classes of our dataset: (a) rowing; (b) sailing; (c) surfing; (d) windsurfing.

In the second step, the foreground is segmented again using the multi-layered multi-feature active contours as depicted in Fig. 3 (c). It results in s sub-regions or sub-parts of the main object $F_{l+2,1}$, ... $F_{l+2,s}$, with Fl+i = Fj=1,...,s Fl+i,j (i, j 2 N). Figure 3 (d) shows $F_{l+2,1}$ which is semantically meaningful. This process of image partition leads to the delineation of coherent objects, allowing efficient foreground labeling and automatic image annotation as described in the next sections.

B. Feature Extraction and Analysis

Each region $F_{l+i,j}$ segmented by the active contours at the layer *i* could be characterized by metric features such as their mean color values in the *RGB* color space. However, humans use semantic concepts to identify and describe colors [16]. Hence, we adopt both numeric features directly extracted from the image such as Region_Average_Color and semantic features like Region_Color_Name mapped from the visual features as described in [20].

Unlike [30] or [8], texture features are not considered in our work in order to allow our automatic system to annotate low-resolution and noisy images as well.

Geometric properties of the delineated regions could be described with notions such as Region_Center_of_Gravity, Region_Shape = {*oval*, *rectangular*, *triangular*}, and Region_Area. Indeed, linguistic concepts have been proven to complement well visual information in the process of scene understanding [23].

C. Decision Trees (DT)

Decision trees (DTs) [26], [15] are a form of multiple variable analysis based on multi-level decisions which split data into a hierarchy of branches that produce the characteristic inverted tree shape. Each segment or branch is called a node. Each node could be of two types, namely, internal node and terminal node also called leaf. Each internal node corresponds to a decision governed by an attribute dividing the data samples the most effectively. Each leaf represents the outcome of the data samples that follow the path from the root (top node) of the tree to the corresponding leaf. The leaves have mutually exclusive assignment rules, and thus, they can be expressed with unique $if - then$ rules, called decision rules, which are interpretable semantically.

In fact, each data sample is represented by a vector of attributes and its associated values. The discovery of the decision rules to create the branches underneath the root node is based on the extraction of the relationship between the input attributes of the samples and the outcomes. The standard DT process implies that each sample has only one possible outcome, i.e. belongs to a single class.

A DT is trained using a set of labeled samples. During the training phase, a DT is built by recursively dividing the training samples into non-overlapping sets. Every time the samples are divided, the attribute used for the division is discarded. The procedure continues until all samples of a same class reach the tree's maximum depth when no attribute remains to separate them.

The classification of new samples is done by performing a sequence of tests. Hence, during the testing phase, the DT is traversed from the root to a leaf node using the attribute values of each new sample. The decision of the sample is the outcome of the leaf node where the sample reaches.

Compared to other machine learning methods such as support vector machine (SVM) such as in [8] or Artificial Neural Networks (ANN) used in [25], DT is naturally interpretable in human language, is fast, and its learning requires only a small numbers of samples. Moreover, DT is robust for incomplete and noisy data and handles both semantic and numeric values.

In this work, we used several decision trees to achieve the goal of automatically annotating images. Our approach consists in using semantic decision rules to classify the images into classes based on their semantic attributes, which were defined using trained decision trees involving both semantic and visual features. Hence, decision trees are first induced in order to define keywords based on numeric visual features and semantic features introduced in Section II.B. Next, higher semantic level decision trees are built to classify the images into the classes based on these natural-language keywords.

As an example, we use a dataset with ‘water sport’ images with *M* and *N*, the width and the height of the image under

Fig. 5. Examples of automatically segmented images from the dataset. Best viewed in color.

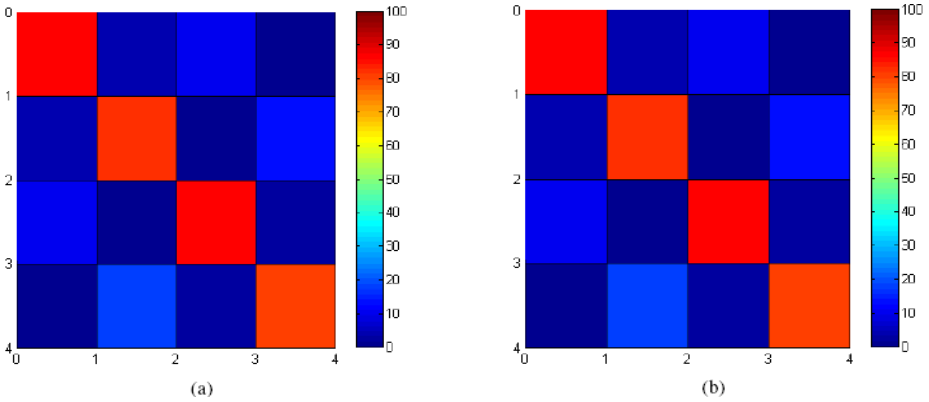

Fig. 6. Fig. 6 est viewed in color.

that should be automatically annotated. More information about this dataset are provided in Section III.

For this purpose, we consider at first the definition of keywords such as ‘board’, ‘boat’, ‘paddle’, and ‘sail’ based on the extracted visual features from the regions delineated by multi-layer active contours. The corresponding induced decision rules are as follows:

$$\begin{cases} if\ \text{Region\_Shape} = oval \\ \quad then\ \text{outcome} = board \\ if\ (\text{Region\_Shape} = rectangular \\ \quad and\ \text{Region\_Color\_Name} = black) \\ \quad then\ \text{outcome} = paddle \\ if\ \left(\text{Region\_Shape} = rectangular \right. \\ \quad \left. and\ \text{Region\_Area} > \dfrac{M \times N}{80}\right) \\ \quad then\ \text{outcome} = boat \\ if\ \text{Region\_Shape} = triangular \\ \quad then\ \text{outcome} = sail \end{cases} \quad (2)$$

investigation, respectively.

Next, we classify the ‘water sport’ images of the dataset into four classes (Fig. 4), namely, ‘rowing’, ‘sailing’, ‘surfing’, and ‘windsurfing’, which are semantically closely related, by inducing a decision tree whose leaf nodes can be expressed with unique *if* − *then* semantic rules as follows:

$$\begin{cases} if\ \text{paddle} = yes \\ \quad then\ \text{outcome} = rowing \\ if\ (\text{paddle} = no\ and\ \text{board} = yes \\ \quad and\ \text{sail} = yes) \\ \quad then\ \text{outcome} = windsurfing \\ if\ (\text{paddle} = no\ and\ \text{board} = yes \\ \quad and\ \text{sail} = no) \\ \quad then\ \text{outcome} = surfing \\ if\ (\text{paddle} = no\ and\ \text{board} = no \\ \quad and\ \text{boat} = yes) \\ \quad then\ \text{outcome} = sailing. \end{cases} \quad (3)$$

Fig. 7. Examples of semantically annotated images with our automatic system.

Some samples of automatically annotated images with our approach are presented in Fig. 7. More results are discussed in Section III.

III. Results and Evaluation

In order to test our segmentation and labeling approach for the automatic image annotation application, we have built a database called 'water sport image dataset' based on two standards datasets, namely, Berkeley Image dataset and Vitterbi USC-SIPI image database that we have merged and enhanced with Google-retrieved images in order to obtain a broad domain of images suitable for public applications involving image annotation. Berkeley Image dataset contains images in jpeg format with a resolution of 321x481, while Vitterbi USC-SIPI image database is a collection of digitized images in tiff format with an average size of 512x512.

Hence, the 'water sport image dataset' groups together 3148 images of 4 types of common outdoor water sports, namely, 'rowing', 'sailing', 'surfing', and 'windsurfing', with a resolution ranging from 320x433 pixels to 1280x650 pixels and in different image format such as tiff, jpeg, or png. Each category contains from 600 to 800 images. Some samples of our database are displayed in Fig. 4.

This dataset presents challenges of scale, pose and light variations as well as shadow effect and noise due to the water element. In this way, the images of our database have different size and resolution as well as large inter-class similarities, e.g. both windsurfing and surfing sports involve the use of a board, and intra-class variations, e.g. the water color could vary from light blue (Fig. 4 (a)) to dark blue (Figs. 4 (b)-(c)) or even be white (Fig. 4 (d)). Hence, the difficulty of the image segmentation, classification, and annotation in this dataset is very high.

All the experiments have been performed on a commercial computer with a processor Intel(R) Core(TM)2 Duo CPU T9300 2.50 GHz, 2 Gb RAM and using MatLab (Mathworks, Inc.) software.

To assess the accuracy of our AIA system, we adopt the standard criterion as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \qquad (4)$$

with *TP*, true positive, *TN*, true negative, *FP*, false positive, and *FN*, false negative.

In the first carried-out experiment, we aim to assess the importance of the precise and semantically meaningful segmentation of the image on the resulting semantic annotation of the image. Thus, the images are segmented using different approaches as presented in Fig. 2. We can observe that if the image is segmented using N-cut or edge detector techniques (Figs. 2 (a)-(b)), it results in semantically incoherent foreground objects. These resulting, meaningless visual information prevent the labeling system to process properly. Hence, a bad segmentation of the image leads to the misclassification of this image, and thus to its incorrect annotation. In the opposite case, when applying our multi-layer multi-feature active contour approach (Figs. 2 (c)-(d)), the segmentation is accurate and provides semantically meaningful foregrounds such as illustrated in Figs. 5 (a)-(d). In this case, the image labeling is performed well, leading to the hierarchical categorization of the images, with a computational time in the range of few ms (Figs. 7 (a)-(d)).

In the second experiment, we assess the influence of the number of layers on the classification accuracy. Images in all the dataset are first segmented by first-layer active contours, and in a second batch, by second-layer multi-layer active contours. The resulting confusion matrices are presented in Fig. 6 (a) and Fig. 6 (b), respectively. It results that more layers have the active contours, better is the image classification, thus the precision of the image annotation. Indeed, at each layer, visual and semantic information are gathered in smaller and more meaningful regions. Thus, the extracted features could be more precisely mapped into

linguistic notions, based on which decisions are made, leading to more reliable annotations.

The mean average accuracy reached by our AIA system is 95%. Compared to other approaches, little are automatic in the literature. We can note that [15] achieves 73% of accuracy, however it uses very distant categories. On the other hand, the technique presented in [11] is 84% accurate, but it involves constraining assumptions, e.g. only the center of the image is studied, and thus it is not processing data with foregrounds not in the middle of images. Hence, performance of our fully automatic image annotation method are better than those of the state-of-the-art ones, while our dataset is challenging as it contains closely-related classes and foregrounds not in the center of the images (Figs. 5 (b),(d)) and distracted by noise and/or shadows caused by the water element (Figs. 5 (a)-(d)).

IV. Conclusions

In this paper, we propose (a) original multi-layer active contours segmenting the image into semantically meaningful objects and sub-objects and (b) new unsupervised semantic labeling technique based on trained decision trees relying on both numeric and linguistic concepts. Thus, the novel fully automatic image annotation method based on (a) and (b) is performed by using semantic knowledge and visual content analysis together and is efficient in terms of precision, while being compatible with online applications.

References

[1] T. Alqaisi, D. Gledhill, and J. I. Olszewska, "Embedded double matching of local descriptors for a fast automatic recognition of real-world objects", in *Proceedings of the IEEE International Conference on Image Processing*, October 2012, pp. 2385-2388.

[2] E. Borenstein and S. Ullman, "Combined top-down/bottom-up segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2109-2125, December 2008.

[3] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation", *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109-131, November 2006.

[4] J. Cai, Z.-J. Zha, Y. Zhao, and Z. Wang, "Evaluation of histogram based interest point detector in web image classification and search", in *Proceedings of the IEEE International Conference on Multimedia and Expo*, July 2010, pp. 613-618.

[5] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of colour-texture regions in images and video", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800-810, August 2001.

[6] I. Endres, A. Farhadi, D. Hoiem, and D. A. Forsyth, "The benefits and challenges of collecting richer object annotations", in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, June 2010, pp. 1-8.

[7] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions", in *Proceedings of the IEEE International Conference on Computer Vision*, September 2009, pp. 1-8.

[8] Y.-F. Huang and H.-Y. Lu, "Automatic image annotation using multi-object identification", in *Proceedings of the IEEE Pacific-Rim Symposium on Image and Video Technology*, November 2010, pp. 386-392.

[9] D. K. Iakovidis and C. V. Smailis, "Efficient semantically-aware annotation of images", in *Proceedings of the IEEE International Conference on Imaging Systems and Techniques*, May 2011, pp. 146-149.

[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models", *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321-331, January 1988.

[11] S. Kim, S. Park, and M. Kim, "Image classification into object/non-object classes", in *Proceedings of the International Conference on Image and Video Retrieval*, July 2004, pp. 393-400.

[12] P. Kohli, L. Ladicky, and P. Torr, "Robust higher order potentials for enforcing label consistency", *International Journal of Computer Vision*, vol. 82, no. 3, pp. 302-324, May 2009.

[13] I. Kokkinos and P. Maragos, "An expectation maximization approach to the synergy between image segmentation and object categorization", in *Proceedings of the IEEE International Conference on Computer Vision*, October 2005, pp. I.617-I.624.

[14] A. Levin and Y. Weiss, "Learning to combine bottom-up and top-down segmentation", *International Journal of Computer Vision*, vol. 81, no. 1, pp. 105-118, January 2009.

[15] Y. Liu, D. Zhang, and G. Lu, "Region-based image retrieval with high-level semantics using decision tree learning", *Pattern Recognition*, vol. 41, no. 8, pp. 2554-2570, August 2008.

[16] A. Mojsilovic, J. Gomes Boykov, and B. Rogowitz, "Semantic-friendly indexing and querying of images based on the extraction of the objective semantic cues", *International Journal of Computer Vision*, vol. 56, no. 1-2, pp. 79-107, January 2004.

[17] G. S. Muralidhar, A. C. Bovik, J. D. Giese, M. P. Sampat, G. J. Whitman, T. Miner Haygood, T. W. Stephens, and M. K. Markey, "Snakules: A model-based active contour algorithm for the annotation of spicules on mammography", *IEEE Transactions on Medical Imaging*, vol. 29, no. 10, pp. 1768-1780, October 2010.

[18] S. Nowozin, C. Rother, S. Bagon, T. Sharp, B. Yao, and P. Kohli, "Decision tree fields", in *Proceedings of the IEEE International Conference on Computer Vision*, November 2011, pp. 1668-1675.

[19] J. I. Olszewska, Unified Framework for Multi-Feature Active Contours, PhD Thesis, UCL, 2009.

[20] J. I. Olszewska, "Spatio-temporal visual ontology", in *Proceedings of the EPSRC/BMVA Workshop on Vision and Language*, September 2011.

[21] J. I. Olszewska, "A new approach for automatic object labeling", in *Proceedings of the EPSRC/BMVA Workshop on Vision and Language*, December 2012.

[22] J. I. Olszewska, "Multi-target parametric active contours to support ontological domain representation", in *Proceedings of the RFIA Conference*, January 2012, pp. 779-784.

[23] J. I. Olszewska and T. L. McCluskey, "Ontology-coupled active contours for dynamic video scene understanding", in *Proceedings of the IEEE International Conference on Intelligent Engineering Systems*, June 2011, pp. 369-374.

[24] J. I. Olszewska and D. Wilson, "Hausdorff-distance enhanced matching of scale invariant feature transform descriptors in context of image querying", in *Proceedings of the IEEE International Conference on Intelligent Engineering Systems*, June 2012, pp. 91-96.

[25] S. B. Park, J.W. Lee, and S. K. Kim, "Content-based image classification using a neural network", *Pattern Recognition Letters*, vol. 25, no. 3, pp. 287-300, February 2004.

[26] J. R. Quinlan, "Induction of decision trees", *Machine Learning*, vol. 1, no. 1, pp. 81-106, March 1986.

[27] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation", *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 157-173, May 2008.

[28] J. Shao, D. He, and Q. Yang, "Multi-semantic scene classification based on region of interest", in *Proceedings of the IEEE International Conference on Computational Intelligence for Modelling, Control and Automation*, December 2008, pp. 732-737.

[29] J. Shi and J. Malik, "Normalized cuts and image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888-905, August 2000.

[30] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modelling texture, layout, and context", *International Journal of Computer Vision*, vol. 81, no. 1, pp. 2-23, January 2009.

[31] M. Tabb and N. Ahuja, "Multiscale image segmentation by integrated edge and region detection", *IEEE Transactions on Image Processing*, vol. 6, no. 5, pp. 642-655, May 1997.

[32] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques", *Pattern Recognition*, vol. 45, no. 1, pp. 346-362, January 2012.