

Efficient Semantically-Aware Annotation of Images

Dimitris K. Iakovidis, and Christos V. Smailis

Dept of Informatics and Computer Technology
Technological Educational Institute of Lamia
Lamia, Greece

dimitris.iakovidis@ieee.org, xsmailis@teilam.gr

Abstract—The collection of knowledge about an imaging domain usually requires input from experts on that domain. In a two-dimensional image space this input can be provided through graphic annotation of image regions that represent objects of interest. This is a necessary process which is both time-consuming and cost inefficient, especially when sequences of images are involved. In this paper we present a novel software tool for efficient annotation of images and image sequences. This tool has a simple graphical user interface through which the average user can use alternatively two graphic annotation protocols and combine them with a versatile snake-based framework for semi-automatic image annotation to significant speedup the annotation process. Graphic annotations can be stored and retrieved directly from the web, whereas they can be associated with semantic identifiers available on any OWL ontology. The efficiency of the proposed annotation tool is demonstrated for the annotation of sequences of chest radiographs. The results obtained promise a wide applicability of the tool for a variety of intelligent imaging applications extensible to the semantic-web.

Keywords—*Images, sequences, annotation, active contour model, snake, ontologies, semantics*

I. INTRODUCTION

Supervised machine learning methods, commonly used for object recognition in digital images, are based on prior, ground-truth, knowledge acquired from domain experts. Such knowledge is provided through manual annotation of their content, which can be provided in a graphic and/or textual form. This is usually a time consuming process since it requires interaction of the domain expert with the software tool used for the annotation task, while the required effort can be thought as a function of the annotation detail and the annotator's skill. A software tool that would enable fast image annotation through a simple interface would definitely contribute to the reduction of both the annotation time and cost.

State of the art image annotation tools of generic applicability include LabelMe [1], ImageParsing.com [2], Peekaboom [3] and ESP [4]. LabelMe is a web-based annotation tool through which an increasing set of annotated images is being made available online. It is supported by a MATLAB toolbox for browsing and searching images. Limitations of the online version of LabelMe include inability to annotate images without publicizing them, slow response times if the user's internet connection is slow. These problems can be overcome by setting up LabelMe on a local server; however this is a quite complex procedure for the average user.

LabelMe has been recently extended for the annotation of video [5]. Imageparsing.com [2] is a commercial solution to image annotation also supported by a MATLAB toolbox, and accompanied with a high quality annotated dataset. Peekaboom and ESP [3]-[4] are interactive games aiming to collect image annotations by entertaining its users. The players of these games cooperate by exchanging textual and spatial information between each other that is likely to describe the content of an image. In [6] the Amazon Mechanical Turk has been proposed as an alternative to provide the users with an economic motivation to annotate images. Several other annotation tools have been proposed but most of them are application-specific. Such tools include DoctorEye [7], a multifunctional open platform for fast annotation and visualization of tumors in medical images, and a human annotation tool especially designed to cope with human articulation [8]. A review of image annotation methods mainly oriented in keyword/text-based annotation is presented in [9]. Another important issue discussed in that study is the need for association of image annotations with semantic identifiers that unambiguously characterize image objects. This can be effectively formalized through ontologies [10]. However, current annotation methods supported by ontologies are mainly domain-specific [9],[11]-[12]. More generic approaches to semantic annotation of images include M-OntoMat [13], and Photostuff [14]. However, these software tools are mainly oriented to the semantic rather than on the efficient graphic annotation of images.

To the best of our knowledge none of the above tools offer a framework for efficient annotation of multiple images. In this paper we present a versatile software tool, called Ratsnake (Rapid image annotation with snakes), implementing such a framework and we investigate its applicability for annotation of image sequences. Annotation efficiency is achieved through the combination of different annotation protocols and a properly modified snake model [15] enabling semi-automatic annotation of objects of interest. Graphic image annotation is complemented by semantics, formally represented in ontologies that can either be developed or retrieved from the semantic web.

The rest of this paper consists of four sections. Section II provides an overview of Ratsnake, whereas the annotation protocols it supports are described in Section III. The results obtained from the application of Ratsnake for the annotation of sequences of chest radiographs are apposed in the case study presented in Section IV. The conclusions of this work are summarized in the last section.

This work was realized in the framework of the project DebugIT which is co-funded by the EC FP7 grant agreement n° FP7-217139. The EC, DG INFSO, is not liable for any use that may be made of the information in this document. It reflects solely the views of the authors.

II. RATSSNAKE

Ratsnake is a cross-platform, publicly available¹, image annotation tool developed in Java. Its user interface was kept simple, in order to make image annotation easier, especially for the average and even the novice users, whereas it is installable with a single click. These considerations have been motivated by the fact that the experts in an imaging domain are not necessarily expert users.

Since image annotation mainly aims to generate ground truth knowledge about images in the scope of machine learning, additional considerations were taken towards the exploitation of knowledge that already exists on the semantic-web. Figure 1 provides an overview of Ratsnake, where it can be noticed that it is capable of retrieving and storing multiple images and annotations not only from and to local storage but also from and to the web. Each image may have multiple annotations, each of which can be associated with a semantic identifier from a single or multiple ontologies. The ontologies should be represented in Web Ontology Language (OWL) and can be either locally or remotely stored. The annotations can be graphic, stored either as binary masks or semi-structured data stored into text files. Ratsnake is fully compatible with LabelMe and therefore it can be used as a cross-platform alternative software to retrieve and edit image annotations from the large collection of LabelMe available online [1].

In order to speed up image annotation tasks, including the annotation of image sequences, we have considered novel approaches to both manual and semi-automatic graphic annotation. Ratsnake supports different annotation protocols that can be combined together so as to meet user needs and preferences.

III. ANNOTATION PROTOCOLS

A. Manual annotation

The polygon annotation protocol is adopted by most current image annotation tools. According to that protocol the user can create an annotation by setting consecutive landmarks around a region of interest (ROI). The landmarks can be interconnected either by linear or non-linear (spline) interpolation. Additional features available in Ratsnake include on-demand increment or decrement of the number of landmarks, and affine transformations of the shape of an annotation.

Ratsnake supports another simple, however practical annotation protocol that has not been previously considered in relevant tools. A grid is used to subdivide the image into small square regions according to the user's needs or preferences. The user can then annotate a ROI by selecting the appropriate grid cells. This can be performed in two ways: a) by selecting grid cells within the ROI; b) by selecting grid cells around the ROI in a freehand annotation style. In both cases, the smaller the resolution of the grid is, rougher the output annotation will be, and vice versa. However, a rough annotation is always performed faster than a detailed annotation.

In order to be able to exploit the advantages of these two annotation protocols, the user can use them alternatively during the annotation process. A significant speedup of the image

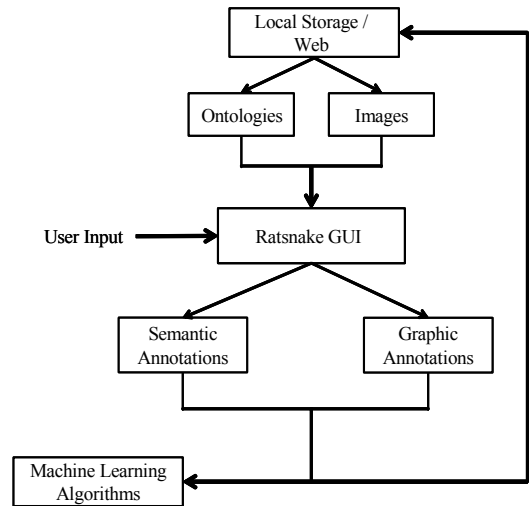


Figure 1. Overview of Ratsnake.

annotation process can be further achieved with a novel semi-automatic approach that is applicable both in the case of polygon and in the case of freehand annotations.

B. Semi-automatic image annotation

A novel semi-automatic approach, based on snakes [15]-[16], is proposed for the enhancement of the efficiency of image annotation. Snakes are time-varying parametric curves of the form $v(s,t) = (x(s,t), y(s,t))^T$ where x and y represent coordinate functions of $s \in [0, L]$ and time t in the image plane. Given an image I of $N \times M$ pixels size with values in $\Omega \subseteq \mathbb{R}$, the energy functional that dictates the shape of the snake is given by $E(v) = S(v) + P(v)$ where

$$S(v) = \frac{1}{2} \int_0^L \left(a \left| \frac{\partial v}{\partial s} \right|^2 + \beta \left| \frac{\partial^2 v}{\partial s^2} \right|^2 \right) ds \quad \text{and} \quad P(v) = \int_0^L P_I(v) ds \quad (1)$$

represent an *internal* and an *external energy* forcing the contour to move. In Eq.(1), a and β are weight parameters controlling the continuity (or tension) and the curvature (or rigidity) of the contour respectively. Typically, the snake algorithm considers a scalar *potential function* estimated as $P_I(v) = -\gamma \cdot \|\nabla I\|$ or $P_I(v) = -\gamma \cdot \|\nabla(G_\sigma * I)\|$, with the latter to be a more robust alternative in the presence of noise. In these equations γ is a weight parameter, G_σ is a 2D-Gaussian function and σ is its standard deviation. According to [16] the user may guide the evolution of the snake by adding constraining terms to $P_I(v)$.

State of the art snake models based on the one described include the gradient vector field (GVF) [17] and the boundary vector field (BVF) [18] models. These models use different potential functions leading to enhanced image segmentation results. In this work we propose a modification of the snake's potential function so that semi-automatic image annotation is feasible. Semi-automatic annotation requires a scalar function $P_I(v)$ such that it a) constrains the evolution of the contour around the manually defined initial boundary, and b) forces the contour to move towards the image edges that are more likely to characterize the boundary of the target object.

In order to achieve that, we consider a binary image B of size $N \times M$ pixels, generated as the projection of (the interpolated) contour $v_0 = v(s, 0)$ on the image plane for all $s \in [0, L]$ i.e.

$$B(x, y) = \begin{cases} 1, & \text{for each } v_0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

This image is subsequently morphologically eroded by d pixels, resulting in a new image B_d whose non-zero pixels form a band, within which the evolution of the snake can be constrained by multiplication as $P_i^s(v) = B_d(v_0) \cdot P_i(v)$, instead of the additive approach considered in the original snake model.

We assume that during the annotation process the user intuitively tries to approximate the boundaries of the target object as close as possible. Therefore, the pixels that are closer to the initial contour ($t=0$) are more likely to be closer to the boundaries of the target object. This consideration has been incorporated into our model, by diffusing B_d so as its intensity becomes lower towards the boundaries of the constraining band. To achieve this the Euclidean distance transform (EDT) is applied on B_d [20]. EDT produces a 2D distance map $T(B_d)$ assigning to each binary pixel a value equal to the distance from the pixel to the nearest edge. Thus, instead of the simple multiplicative approach, and in order to amplify mid-tone edges that are also likely to belong to the boundaries of the target object, we propose

$$P_i^{cs}(v) = (1 - e^{-\eta T(B_d(v_0))}) \cdot \Pi_f(v) \quad (3)$$

where η controls the curvature of the exponent and

$$\Pi_f(v) = -\sum_{i=1}^{\Delta} \delta_i \cdot f_i(I) \quad (4)$$

where $f_i : \Omega \rightarrow \Omega$ a user-defined pre-processing function of I , such that the force driving the snake towards the boundaries of the target object is increased, and δ_i is a weight parameter that controls the degree to which $f_i(I)$ contributes to the external energy. The value of d is set according to the user's annotation profile for a specific application domain. For example, if the user's annotations tend to be closer to the subject, smaller values of these parameters would be more suitable. Equation (4) is used to indicate that Ratsnake can accommodate several (Δ) terms in the potential function each of which may have a different influence in the evolution of the snake. The functions f_i can be easily implemented as plug-in modules of Ratsnake in simple Java. The generality of this approach allows any of the recent models such as GVF or BVF to be incorporated.

C. Annotation of image sequences

The different annotation protocols supported by Ratsnake can be combined together for efficient annotation of similar objects in image sequences. This is a common task in visual monitoring of phenomena and activities such as the detection of biological structures in microscopy images, and infection monitoring from chest radiographs [21]. Image annotation in this kind of applications can facilitate the generation of ground truth image sets for algorithm training or evaluation purposes.

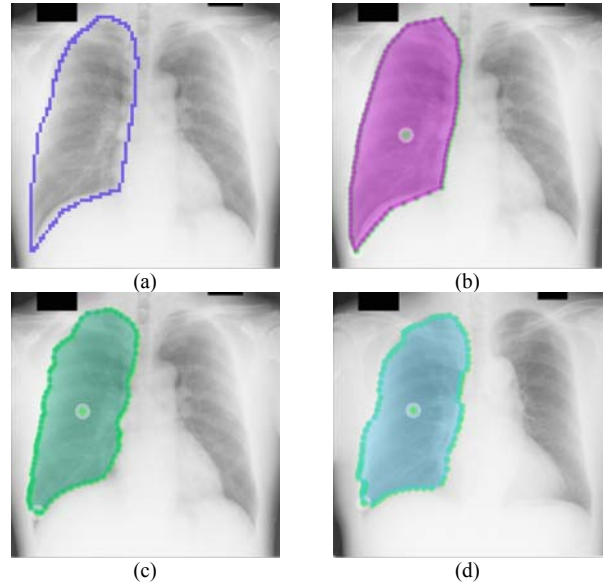


Figure 2. Example annotation of two images in a sequence using the proposed semi-automatic approach. The chest radiographs have been obtained from the publicly available JSRT dataset [22]. (a) Quick freehand user annotation of a lung field. (b) Automatically derived polygon annotation. (c) Fine annotation of the lung field with the proposed snake approach (accuracy 90.2%). (d) Fine annotation of the lung field in another image of the sequence after steps 5-6 described in Section II.C (accuracy 87%).

The annotation process for a sequence of images involves the following steps:

- Step 1: Load image sequence; Set $count = 0$;
- Step 2: Annotate object of interest in image $\#count$;
- Step 3: Copy the annotation; Set $count = count + 1$;
- Step 4: Paste the annotation on the relevant object(s) on image $\#count$;
- Step 5: Modify the annotation;

Repeat steps 3 to 5 until all images are annotated.

The fifth step of this process may involve only a single click in case the semi-automatic approach is used for the modification of the transferred annotation. In any case, the modification of an initial annotation on each image of a sequence is expected to be generally faster than the annotation of each image without having an initial annotation. An example application of the proposed approach on two chest radiographs in a sequence, according to the described process, is illustrated in Fig. 2.

IV. CASE STUDY

In order to demonstrate the efficiency of Ratsnake for the annotation of image sequences we have considered an indicative case study motivated by our need to obtain hundreds of ground-truth annotations of sequences of chest radiographs. The annotations were necessary for training and evaluation of machine learning algorithms that detect the boundaries of the lung fields in chest radiographs [21]. The evaluation of Ratsnake on such an annotation task was performed on a database of 247 chest radiographs provided by the Japanese Society of Radiological Technology (JSRT) [22]. We selected this database because it is accompanied with ground truth

annotations [23] that can be used to assess the quality of the annotations obtained by the proposed semi-automatic approach.

Sequences of chest radiographs were generated as random subsets of 5% of the available dataset using images of 256×256 pixels. Two experts were asked to annotate the sequences a) using Ratsnake, and b) using LabelMe [1]. The latter was selected as a representative state of the art tool for comparisons. In all the experiments the settings of the snake used were $\alpha = 0.5$, $\beta = 1.4$, $\gamma = 0.7$ and $\sigma = 3$. The annotation times were measured on a conventional laptop with Intel Core™ 2 Duo 1.83 Ghz 2MB L2 cache processor and 3GB RAM. The average results per image are presented in Fig. 3. This figure compares the average times obtained a) with Ratsnake using the approach described in Section II.C, b) with Ratsnake using the approach described in Section II.B, i.e. without using the previous annotations, and c) with LabelMe set up on a local server. It can be observed that the speedup of the annotation using Ratsnake and the proposed annotation approach is significant reaching $15/6=2.5$.

The average accuracy (overlap with ground truth) obtained by the proposed snake algorithm for these annotations is $87.7 \pm 8.3\%$. This accuracy is comparable with most of the results obtained in [23]. However, it should be noted that Ratsnake is intended for rapid construction of ground-truth datasets and not for computer-aided detection of the lung fields for diagnostic purposes. The expert performing the annotation uses the result of Ratsnake to get closer to the ground truth so that the latter is obtained with only a few modifications of the shape of the annotation.

CONCLUSIONS

We presented Ratsnake, a software tool for efficient, semantically-aware annotation of images and image sequences featuring novel annotation approaches. Its efficiency has been validated on a case study involving the annotation of sequences of chest radiographs.

Information on Ratsnake, demos and other relevant deliverables of the EU project DebugIT are available at <http://ivibis.ctr.teilam.gr/debugit/downloadsEN.htm>.

REFERENCES

- [1] A. Torralba, B.C. Russell, J. Yuen, "LabelMe: Online Image Annotation and Applications," *Proc. IEEE*, vol.98, no.8, pp.1467-1484, 2010.
- [2] B. Yao, X. Yang, and S.-C. Zhu, "Introduction to a Large Scale General Purpose Ground Truth Dataset: Methodology, Annotation Tool, and Benchmarks." *EMMVCVPR*, 2007.
- [3] L. von Ahn, R. Liu, and M. Blum, "Peekaboom: A Game for Locating Objects in Images," In *ACM CHI*, 2006.
- [4] L. von Ahn and L. Dabbish, "Labeling Images with a Computer Game," In *SIGCHI*, 2004.
- [5] J. Yuen, B. C. Russell, C. Liu, and A. Torralba, "LabelMe video: Building a Video Database with Human Annotations," *IEEE International Conference on Computer Vision (ICCV)*, 2009.
- [6] A. Sorokin, D. Forsyth, "Utility Data Annotation with Amazon Mechanical Turk," in *Proc. Computer Vision and Pattern Recognition Workshops* 2008, pp.1-8, 23-28 June 2008.
- [7] E. Skounakis, V. Sakkalis, K. Marias, K. Banitsas, and N. Graf, "DoctorEye: A Multifunctional Open Platform for Fast Annotation and

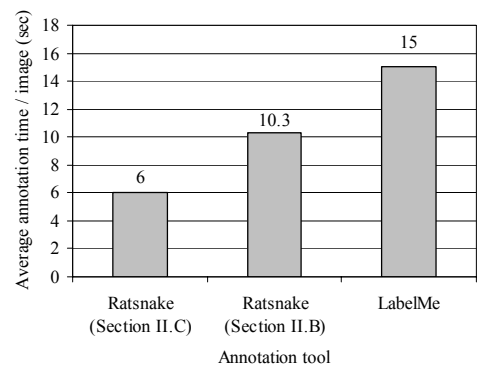


Figure 3. Average time per image (in seconds) required with Ratsnake and LabelMe for the annotation of the image sequences used in the experiments.

- Visualization of Tumors in Medical Images," in *Proc. IEEE EMBC 2009*, pp.3759-3762, 2009.
- [8] C.J. Taylor. "Reconstruction of Articulated Objects from Point Correspondences in a Single Image", *Computer Vision and Image Understanding*, vol. 80, no. 3 pp 349-363, 2000.
- [9] A. Hanbury, "A Survey of Methods for Image Annotation," *Journal of Visual Languages and Computing*, vol. 19, pp. 617-627, 2008.
- [10] D.K. Iakovidis, D. Schober, M. Boeker, and S. Schulz, "An Ontology of Image Representations for Medical Image Mining," in *Proc. IEEE ITAB*, 2009.
- [11] D.L. Rubin, P. Mongkolwat, V. Kleper, K.S. Supekar, and D.S. Channin. "Medical Imaging on the Semantic Web: Annotation and Image Markup," *AAAI Spring Symposium Series, Semantic Scientific Knowledge Integration*, Stanford, 2007.
- [12] A.T.G. Schreiber, B. Dubbeldam, J. Wielemaker, B. Wielinga, Ontology-based photo annotation, *IEEE Intelligent Systems*, vol. 16, no. 3, pp. 66-74, 2001.
- [13] K. Petridis, D. Anastasopoulos, C. Saathoff, N. Timmermann, I. Kompatsiaris and S. Staab: "M-OntoMat-Annotizer: Image Annotation. Linking Ontologies and Multimedia Low-Level Features", *KES 2006*, Oct. 2006
- [14] C. Halaschek-Wiener, J. Golbeck, A. Schain, M. Grove, B. Parsia, J. Hendler, "Photostuff-An Image Annotation Tool for the Semantic Web," *ISWC*, 2005.
- [15] J. Liang, T. McInerney, D. Terzopoulos, "United Snakes," *Med. Image Anal.*, vol. 10, no. 2, pp. 215-233, 2006.
- [16] M. Kass, A. Witkin, and D. Terzopoulos, "Snake: Active Contour Models," *Int. J. Com. Vis.*, vol.1, no.4, pp.321-331, 1988.
- [17] C. Xu and J.L. Prince, "Snakes, Shapes, and Gradient Vector Flow", *IEEE T. Image Proc.*, vol. 7, no. 3, pp. 359-369, 1998.
- [18] K.W. Sum, and P.Y.S. Cheung, "Boundary Vector Field for Parametric Active Contour," *Patt. Rec.*, vol. 40, no. 6, pp. 1635-1645, 2007.
- [19] D.J. Williams, and M. Shah, "A Fast Algorithm for Active Contours," *CVGIP: Image Und.*, vol. 55, no. 1, pp. 14-26, 1990.
- [20] H. Brey, J. Gil, D. Kirkpatrick, M. Werman, "Linear Time Euclidean Distance Transform Algorithms," *IEEE T. Patt. Anal. and Mach. Intell.*, vol. 17, pp. 529-533, 1995.
- [21] D.K. Iakovidis, M.A. Savelonas, G. Papamichalis, "Robust Model-based Detection of the Lung Field Boundaries in Portable Chest Radiographs Supported by Selective Thresholding," *Meas. Sci. Technol.*, vol. 20, pp. 104019, 2009.
- [22] J. Shiraishi, et al., "Development of a Digital Image Database for Chest Radiographs with and without a Lung Nodule: Receiver Operating Characteristic Analysis of Radiologists' Detection of Pulmonary Nodules," *Am. J. Roentgenol.*, vol. 174, pp. 71-74, 2000.
- [23] B. van Ginneken, M.B. Stegmann, M. Loog, "Segmentation of Anatomical Structures in Chest Radiographs using Supervised Methods: a Comparative Study on a Public Database", *Med. Image Anal.*, 2006, vol. 10, pp. 19-40.