Study of a New Approach for Labeling of Image Segments

Marinalva Dias Soares[1], Rafael Santos[2], Nandamudi Vijaykumar[2],
Luciano Vieira Dutra[3]

[1]Programa de Doutorado em Computação Aplicada – CAP
Instituto Nacional de Pesquisas Espaciais – INPE

[2]Laboratório Associado de Computação e Matemática Aplicada – LAC
Instituto Nacional de Pesquisas Espaciais – INPE

[3]Divisão de Processamento de Imagens – DPI
Instituto Nacional de Pesquisas Espaciais – INPE

{marinalva,dutra}@dpi.inpe.br, {rafael.santos,vijay}@lac.inpe.br

Abstract. *This paper presents the proposal for a new approach for labeling of image segments using citizen science. Citizen science has brought great results and findings in research projects in various industries, including discovery of new plants, flowers, birds and galaxies, for example. In this paper we present the concepts, conditions and expected results and consequences of the approach.*

Key words: *image segmentation, polygons labeling, citizen science.*

1. Introduction

In remote sensing, interpretation of image data (scenes) may be considered as associating meaning to the contents of the image [Pahl 2003]. This is possible by introducing knowledge into the process of interpretation, which requires the following two aspects: first, knowledge must be obtained from a human specialist and needs to be formulated and updated by the specialist; second, it should be possible to automatically represent and process that knowledge.

Humans can interpret scenes without much effort using knowledge, experience and visual evidence including the context of objects (or targets) on the image. Image processing systems for automatic scene identification often cannot use directly the information humans can. One reason is that image interpretation systems often apply image processing techniques in different levels, each being done with different techniques, algorithms and parameters. For example, one of the first steps in object identification on images is the conversion of pixels in regions for further processing, a process known as image segmentation.

Segmentation this is a process that partitions an image into regions, so that the elements belonging to each region are similar with respect to one or more properties [Jung 2007]. The segmentation process creates polygons corresponding to groups of pixels which are similar according to specific criteria. The segmentation process may create a huge number of polygons and these polygons may not be segmented properly due to the imperfection inherent of the segmentation algorithms.

Segmentation itself is just a step towards object identification – the regions obtained through segmentation must be labeled. Each of the polygons obtained with the segmentation step must receive a discrete label, usually with semantically high information about it. For an application using an urban scene, for example, one could use labels such as *roofs*, *trees*, *streets*, *pools*, etc.

Figures 1, 2 and 3 illustrate the process of image segmentation and labeling: a small region in an urban scene is shown in Figure 1; and its segmentation in Figure 2. In Figure 2 there are both oversegmented regions (perceptual objects divided into several regions) and undersegmented regions (regions which contains several different perceptual objects), which are practically unavoidable when using the majority of unsupervised image segmentation algorithms. In Figure 3 there are four regions which were labeled by an expert and painted with different colors to indicate its semantic class: yellow for ceramic-tiled roofs and green for trees.

Figure 1. Satellite Image.

Labels can be assigned by a human expert, with rules, samples or conditions dependent on the object identification task itself. Manual labeling, while not a particularly complex task, is a repetitive one which can lead to errors by the operators. On the other hand automatic labeling approaches, in order to be successful, must incorporate the knowledge humans use.

This a difficult task. Low-level vision tasks, such as segmentation, classification and recognition of objects, perform poorly [Neumann and Müller 2008] mostly because automatic systems do not have the ability to perceive context, similarity and other hints based on experience and common sense to perform such a task as humans can do. So, the automatic labeling of polygons (from a segmented image), is an error-prone process since algorithms cannot reproduce faithfully the knowledge and experience from the users.

In this context we propose a new approach for image segments labeling, which
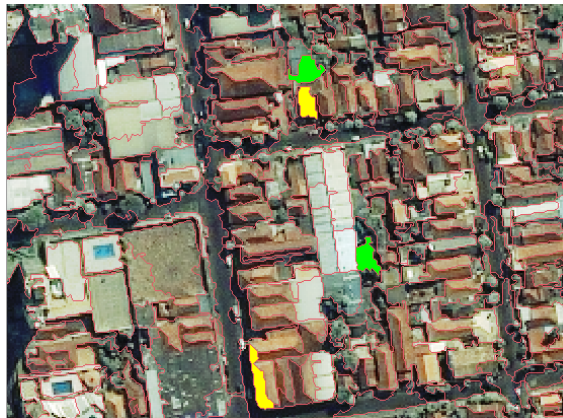
Figure 2. Segmentation of the image.

Figure 3. Manual labeling of the regions.

makes use of volunteer users. The use of communities or networks of citizens who act as participants or observers in some domain of science is often called *citizen science*, which can be defined as anything from citizens observing and reporting on natural events to a genuine revolution in science that democratizes the important social role of learning about the world around us, in order to advance its understanding [Working Group, Citizen science at the Cornell Lab of Ornithology 2007].

Citizen science-based approaches have been used in several different tasks that either could be performed by a lot of work by an human specialist or poorly by an automatic system, often with good results, and, more interestingly, with additional results that could not be obtained by traditional methods. Citizen science is more often than not based on volunteer users – the motivation of the participants (often unpaid volunteers) has also been studied [Droedge 2010, Raddick et al. 2010].

Examples of Internet-based projects that aims to promote public engagement with research, as well as with science in general include: SETI Search for Extraterrestrial Intelligence (http://setiathome.berkeley.edu/), Galaxy Zoo 2 (http://www.galaxyzoo.org/), Herbaria@home (http://herbariaunited.org/atHome), The ODP Open Dinosaur Project (http://opendino.wordpress.com/), Citizen Science Central (http://www.birds.cornell.edu/citscitoolkit).

The objective of this research paper is to present a framework to acquire and model user knowledge, with an application for labeling objects extracted from urban satellite image based on the approach of using citizen science. The process is composed by data acquisition, data preprocessing, data storage, and data analysis modules.

The paper is organized as follows: section 2 presents the methodology for achieve the objectives proposed here; section 3 describes the experiment and section 4 describes the expected results.

2. Methodology

This section presents, in general lines, the methodologies for data preprocessing, acquisition, analysis and application, including guidelines for the development of computational tools (such as a prototype for the proposed data acquisition). Figure 4 shows the processes considered for data preprocessing, acquisition and analysis required for the development of this work.
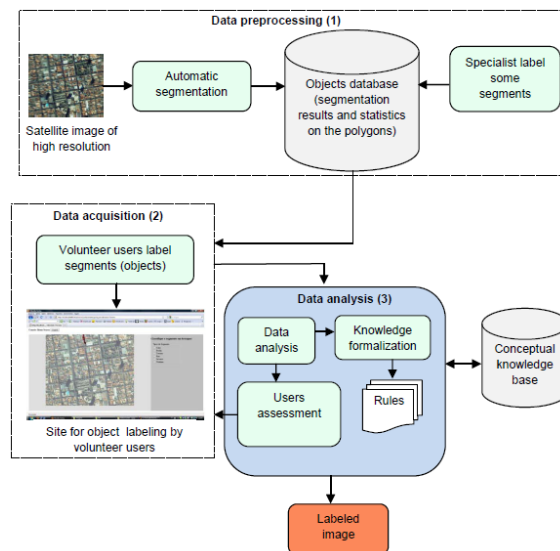

Figure 4. Data processing, acquisition and analysis tasks.

2.1. Data Preprocessing

The required preprocessing steps are:

- Selection of a high-resolution digital image from a satellite. It has been decided to chose an urban scene consisting of different types of objects (buildings, occluded streets, shadows, etc.) and with several instances of such objects.
- Segmentation of this image using a traditional region growing algorithm in Spring tool [Câmara et al. 1996]. Some parameters of the segmentation algorithm control the number and size of the regions it creates and the similarity tolerance used to create those regions; and it is very hard to determine the best parameters for a specific scene since the resulting segments are often evaluated subjectively. For the purposes stated in this work, we would rather have oversegmentation of the image (so an object may be broken into several small segments) than undersegmentation (which causes different targets on the image to be clumped into one single region),

because it may be easier for the user to identify parts of a stated object than be in doubt when several different objects are present in a polygon.

- Creation of a generic database to hold the tables with data on the polygons, users and tasks of labeling.
- Creation of a hierarchical taxonomy for some classes (e.g. roofs and streets paved with different materials) presented to the users to identify. This taxonomy is dependent on the task itself and could be as detailed as possible, but when creating it one must consider that complex, multi-level hierarchical classes descriptions could lead to complex user interfaces, which could confuse or discourage users.
- Creation of a table in the database to collect the users' interaction with the data acquisition system. This database can be as simple as possible, but for the analysis there must be at least one table with the user identification (without personal data, but with information on the users' technical/educational background if possible) and the task record table (which records a decision by the user, i.e., the user's identification, the polygon he/she labeled, his/her choice for label and the timestamp for the inclusion of this data on the table).
- Partially populate the users' interaction table by preliminary labeling some of the polygons by a domain expert. These initially labeled polygons could be considered "ground truth", i.e. labels that are considered to be correct; and can be used to assess answers from the non-expert users, effectively allowing the evaluation of the users' ability. The domain expert will label only some polygons for each class in the labeling task, so he/she won't need to dedicate too much time for this task.

2.2. Data Acquisition

In this step data will be collected from the users that will be used to determine which labels will be used for the polygons and to model the users' knowledge and behavior.

For the data acquisition, was created a web-based interface (available at http://www.lac.inpe.br/UrbanZoo) that presents the tasks for the users and collect the users' decisions. A web site is a natural choice for presenting a task to the user: it will be designed without any special software or hardware requirements; it uses software (internet browsers) which are already well-used and well-known. The basic mechanism of interaction is based on what is used by the Galaxy Zoo project [Raddick et al. 2010], with some changes to allow collection of some metadata (e.g. timestamps of the users' interactions) and allowing users to skip decision tasks.

Interaction with the data acquisition should be as much non-intrusive as possible: the user will be presented with a polygon in the context of the image (see Figure 6) and a simple form with choices for the classes that can be used to label it (with the option to skip the labeling for that particular polygon). Any choice by the user will be stored and a new task will be presented. Apart from the login into the site (to identify the user) no further interaction will be required.

The presentation of the labeling tasks to the users will not be totally random: a *task selector* will be developed with the following features:

- While the user is still doing his/her first interactions with the site (i.e. after he/she has registered on the site and receiving the first tasks) tasks will be chosen in order to assess the users' ability to correctly label segments for which the label

Figure 5. Region highlighted to be labeled.

was already defined by the domain expert (details on this evaluation are shown in the subsection 2.3).

- When the user is proficient enough to label polygons of most of the labels, the polygons will be selected randomly.
- When enough data about this user's decision is collected, the task selector may either present specific labeling problems tailored for this users' proficiency (e.g. based on his/her ability to discriminate between two classes of polygons) mixed with random selections or more "ground truth" polygons to reassess the users' proficiency for a particular label.

2.3. Data Analysis

With enough data collected (i.e. decisions on labels done by the users with the segmented image and using the web interface) the analysis phase will can be initiated.

Several different analysis tasks and scenarios are considered. Among those are:

- The most trivial analysis, which is highly related to the target application (in this case image region labeling), is the evaluation of opinions of users' opinions on the labeling of the polygons obtained from the image segmentation. Simple rules could be used such as labeling polygons only when enough opinions have been collected about it, and use a simple weighted majority rule for labeling it. Weights could be derived from an user reliability metric (described further in this section). The labeled polygons won't be shown to the users to avoid influencing their future decisions, but could be evaluated by a domain specialist to assess quality of labeling for further use.
- One of the most important analysis, specially considering that the users may be mostly volunteers without training, is the evaluation of the reliability of the user through a reliability metric, which could be calculated using the number of "hits" of the user agains his/her "misses". For example, when the user is still under evaluation, he/she could be presented with some labeling tasks for which the expected classes are already known (being identified by the domain specialist beforehand). The evaluation will be taken into consideration for further analysis but won't be presented back to the user, to avoid influencing him/her. User reliability could be reassessed over time, periodically or when statistics indicate that labeling errors are becoming more frequent.

  The user reliability metric could be calculated for each different label. Users with a record of reliable decisions for a particular label could be assigned more tasks related to that label (e.g. disambiguation tasks). Decisions by a particular user and label that are considered not reliable would imply in a smaller weight when

deciding which label ought to be applied to which polygon.

This reliability metric is essential to guide the task selector algorithm, which will determine which tasks each user should receive next.

- A measure of reliability could be calculated for each polygon when enough opinions were acquired about it, to determine whether it is an "easy" or "difficult" labeling task. Further data could be collected differently for easy or difficult polygons, eventually difficult enough tasks could be sent to the domain expert for disambiguation. A simple measure could be the entropy of the users' classes for that polygon – if most users decide for a class this measure will be low, if several different opinions arise, the polygon could be tagged as difficult.
- Polygons for which labeling task were often skipped by users could also be labeled as difficult for further evaluation by the domain expert.
- With enough labels collected, one or more simple agents could be developed and applied to use the label on some "well-known" polygons (polygons which were unanimously or almost-unanimously labeled by enough users) to label polygons with similar spectral/shape/statistical features. The performance (reliability) of these agents could be measured using its decisions compared against the users' decisions for further fine-tuning.
- Although this is not expected to happen, the original labeling of some polygons by the expert could be verified against a large majority of the users' decisions. Differences on the chosen labels could indicate either a mistake on the decision from the expert or a difficulty on labeling a particular polygon, with the expert possessing information that is not available or known by the user base.
- Eventually, with enough data collected from an user, it will be possible to model the users' abilities (related to the task being performed) to assess whether he/she is performing better with time.

All those analysis tasks may be implemented using several different classes of algorithms, ranging from using basic statistics to clustering and classification techniques to data mining techniques [Santos 2009]. In particular, it is expected that algorithms and approaches used to mine and explore recommendation systems [Anand and Mobasher 2007, Mobasher et al. 2004] and user activity modeling [Frías-Martínez and Karamcheti 2003] could be successfully applied to the data collected in this setup.

3. Experiment

The research presented here seeks to obtain a more comprehensive understanding of what the volunteer users say about complex and imprecise objects resulting from segmentation process. For this, the experiment presented here uses citizen science to identify objects extracted from a satellite image of Sao Jose dos Campos city.

The image size is 900x900 pixels and was segmented in such a way to create an oversegmented image with 2430 polygons. To enable the data collection process described in the previous section, a website (http://www.lac.inpe.br/UrbanZoo) was developed in which volunteer users are presented with polygons and a list of options for labeling those polygons.

Based on the chosen image, we established the following hierarchy of classes (labels) for identification of targets of interest: *roofs* (generic), *ceramic roofs*, *tin roofs*,

cement asbestos roofs, *trees*, *streets*, *swimming pools*, *shadows*, *open fields* (any kind of vegetation other than trees), *bare soil*, *water* and *mixed targets* (for when the polygon is composed of different classes of objects). These classes are shown to the user along with the options *none of the above* (in case the user knows the polygon class, but its class is not shown in list of options) and *unknown* (when the user does not know which is the correct class for the polygon). Users label the polygons by accessing the UrbanZoo site and selecting one of the classes for a particular polygon. This is repeated until the user decides to finish his/her interaction with the site. Each interaction is stored in a database for further analysis.

4. Expected Results

The experiment is ongoing and, with the data that are being collected by the approach proposed in this paper, it is intended to demonstrate the feasibility of completing the classes or labels based on the decisions of users, demonstrating an application of citizen science to the problem of object identification in remote sensing images.

After analysis, we expect to obtain the following:

- Label the yet-unlabeled polygons on a segmented image, using proper measures of reliability.
- Complement knowledge that can be partially extracted from a domain specialist and eventually use this knowledge in similar labeling tasks.
- Identify and possibly relabel polygons for which a label may had been incorrectly given by the domain specialist
- Identify polygons for which there isn't a consensus about its class – this may indicate an ambiguous or hard-to-label polygon, a problem on the image segmentation step or even a badly defined list of labels for the labeling task.
- Model the knowledge of the users, assessing his/her performance in general and related to specific classes (users may perform differently depending on the types of image regions presented to him/her) for further analysis and usage.
- Demonstrate the feasibility of using citizen science to achieve the specific goal (image segment labeling) which may be less expensive and more robust than the traditional, expert-only labeling.

References

Anand, S. S. and Mobasher, B. (2007). Contextual recommendation. In Berendt, B., Hotho, A., Mladenič, D., and Semeraro, G., editors, *From Web to Social Web: Discovering and Deploying User and Content Profiles – Workshop on Web Mining, WebMine 2006, Berlin, Germany, September 18, 2006, Revised Selected and Invited Papers (LNAI 4737)*, pages 142–160.

Câmara, G., Souza, R. C. M., Freitas, U. M., and Garrido, J. (1996). SPRING: Integrating remote sensing and GIS by object-oriented data modelling. *Computers & Graphics*, 20(3):395–403.

Droedge, S. (2010). Just because you paid them does not mean their data are better. http://www.birds.cornell.edu/citscitoolkit/conference/proceeding-pdfs/Droege

Frías-Martínez, E. and Karamcheti, V. (2003). A customizable behavior model for temporal prediction of web user sequences. In Zaïane, O. R., Srivastava, J., Spiliopoulou, M., and Masand, B., editors, *WEBKDD 2002 – Mining Web Data for Discovering Usage Patterns and Profiles, 4th International Workshop, Edmonton, Canada, July 23, 2002, Revised Papers (LNAI 2703)*, pages 66–85.

Jung, C. (2007). Unsupervised multiscale segmentation of color images. *Pattern Recognition Letters*, 28(4):523–533.

Mobasher, B., Jin, X., and Zhou, Y. (2004). Semantically enhanced collaborative filtering on the web. In Berendt, B., Hotho, A., Mladenic, D., van Someren, M., Spiliopoulou, M., and Stumme, G., editors, *Web Mining: From Web to Semantic Web – First European Web Mining Forum, EWMF 2003, Cavtat-Dubrovnik, Croatia, September 22, 2003, Invited and Selected Revised Papers (LNAI 3209)*, pages 57–76.

Neumann, B. and Müller, R. (2008). On Scene Interpretation with Description Logics. *Image and Vision Computing*, 26(1):82–101.

Pahl, M. (2003). *Arquitetura de um sistema baseado em conhecimento para a interpretação de dados de sensoriamento remoto de múltiplos sensores*. PhD, Hannover University. http://urlib.net/sid.inpe.br/mtc-m17@80/2008/03.07.18.31. Translated by Hermann Johann Heinrich Kux and Raul Queiroz Feitosa.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., Szalay, A. S., and Vandenberg, J. (2010). Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review*, 9(1):010103.

Santos, R. (2009). Conceitos de Mineração de Dados na Web. In Teixeira, M. M., Teixeira, C. A. C., Trinta, F. A. M., and Farias, P. P. M., editors, *XV Simpósio Brasileiro de Sistemas Multimídia e Web, VI Simpósio Brasileiro de Sistemas Colaborativos – Anais*, pages 81–124.

Working Group, Citizen science at the Cornell Lab of Ornithology (2007). What is this, ´CITIZEN SCIENCE´? http://www.birds.cornell.edu/citscitoolkit/about/SYNTHESIS0Citizen