On Improving Geotag Quality in Photo Collections

Yuri Almeida Lacerda[1,2], Hugo Feitosa de Figueirêdo[1,3], João Paulo Ribeiro da Silva[1], Daniel Farias Batista Leite[1], Cláudio de Souza Baptista[1]
[1]University of Campina Grande
Campina Grande,Paraíba, Brazil
[2] Federal Institute of Education, Science and Technology of Ceará - Campus Crato - Crato, Ceará, Brazil
[3] Federal Institute of Education, Science and Technology of Paraíba – Campus Monteiro - Monteiro, Paraíba, Brazil
yurilacerda@ifce.edu.br, hugo.figueiredo@ifpb.edu.br, joao.silva@ccc.ufcg.edu.br, daniel.leite@ccc.ufcg.edu.br,baptista@dsc.ufcg.edu.br

Anselmo Cardoso de Paiva[4]
[4] Federal University of Maranhão
Applied Computing Group NCA
São Luís, Maranhão, Brazil.
paiva@deinf.ufma.br

***Abstract*—Digital cameras equipped with GPS receivers allow storing geographic location into the photograph metadata. Geographic location constitutes a very important information to be used in systems that retrieve and organize photographs. However, cameras equipped with GPS may store either invalid or null geographic locations. This is usually due to a delay in obtaining the GPS signal. In this paper, we propose a new method for detection of inconsistencies of geographic locations in photograph metadata and for propagation of the geographic location annotation for these photographs. Besides, we also propose the incorporation of these methods in a Web-based system for correction and annotation of geotags. Another contribution of this work is the presentation of a case study to validate the system. The results prove that the proposed techniques increase precision (through the detection of inconsistencies) and recall (through the propagation of the geographic locations) in the retrieval of georeferenced photographs.**

***Keywords* - Geotag; Photo Metadata; GeotagPropagation; Metadata Inconsistency Detection.**

I. Introduction

The technological advances in the last years have enabled a wide use of electronic devices, such as: digital cameras, smartphones, and tablets. This popularization caused an exponential increase of the amount of multimedia files produced by people, such as videos, photographs and audio. This phenomenon can be easily verified in social networks, blogs and internet sites. The large number of multimedia files that has been generated by people has jeopardized information management. For example, imagine an user organizing manually a collection with thousands of photographs. Even manually organizing hundreds of photographs taken during a vacation trip is very time consuming and tedious task.

Several approaches have been proposed for automatic organization of photographs, with the objective of reducing the user's manual efforts, such as PhotoGeo [1], Naamanet al. [2], Cooper et al. [3] and Tsay et al. [4]. Such systems usually make use of the photo metadata to help in the organization process. Some examples of these metadata include date, time, geographic location of the camera at the moment of capture, camera manufacturer and model, tags and descriptive data.

Some studies argue that the place where the photograph was taken is one of the first things people remember when they want to retrieve that photograph [2]. This means that the geographic location of the camera at the time the picture was taken is very important for the process of photograph organization.

The integration of GPS chips into smartphones and digital cameras have allowed the storage of geographic location in the metadata of photographs automatically. Nevertheless, there can be some problems related to this information acquisition such as data imprecision, invalid data and indoor difficulties.

The low power of the GPS chips supplied with those devices, and the poor quality of the GPS signal in many places may generate imprecise geographic location data. Thus, the photographs may end up being indicated in places far away from the real point where the picture was taken.

Another problem is the fact that GPS receivers do not work well indoors, possibly generating invalid or imprecise data in this situation. Some smartphones uses the A-GPS [5] system to minimize this problem.

In other situations, the georeferencing is either absent or taken erroneously. For example, suppose that a given person with a camera equipped with GPS has captured some photographs. We know that it is necessary some instants until the chip receives the GPS signal. While the signal is not received, the camera will not make the georeference of the photographs, or it may use the last geographic information captured by the GPS, possibly generating incorrect information. However, if the user remains with the camera on and the GPS function activated, within a few instants the photographs will be taken with correct geographic information.

In this work, we propose an automatic and semiautomatic photograph georeferencing system based on the detection of inconsistencies of geographic location data and correction propagation, with the objective of augmenting the georeference recall and precision in a photograph collection. The proposed system uses temporal segmentation and the geographic location of some georeferenced photographs to detect inconsistencies in the locations. The system also suggests new annotations for the photographs with mistakes or without annotation.

The remaining of the paper is structured as follows. In Section II, we highlight some studies related to the subject approached in this paper. Next, Section III focuses on the prototype architecture. Section IV addresses the solutions proposed in this paper, for detection of inconsistencies and suggestion of geographic location annotation. Section V presents the evaluation of the proposed solutions through experiments and analysis of the results. Finally, in Section VI, we present the conclusion and discuss further work to be undertaken.

II. Related Work

In this section, we discuss related work. Initially, we present studies dealing with the use of tags. Next, we focus on the use of content to make the images georeferenced and, finally, we present the studies on propagation of geotags.

Lee et al. [6] highlight the existence of a strong correlation between purely textual tags and the geographic location of the photographs in social networks. So, in the proposed approach, a computation of the similarity between tags and geographic location of the photographs was used to determine the relationship between the tags and geotags.

Hays and Efros [7] propose an algorithm called "im2gps", that estimates the geographic location of a photograph based on the geographic location of photographs with higher visual similarity. For such, they used a database containing more than 6 million georeferenced photographs.

Hollenstein and Purves[8] carried out a study demonstrating that the geotag must be according to the way people describe a place, that is, instead of georeference through latitude and longitude, tags like "Eiffel Tower", for example, should be added.

Many studies deal with the propagation of tags based on geographic location [9] and infer the geographic location based on the image content and tags [10].

Vandormael and Courdec[11] used the communication between devices to check the coherence of the geographic location of mobile devices at the moment the photographs are captured.

Ivanov et al. [9] proposed the propagation of geotags based on the combination of detection of duplicated objects and the user's confidence modeling. The idea is the propagation of geotags using other geotagged photographs.

CrEve [12] is a collaborative event annotation framework that uses photograph content found in social media sites. One of the addressed issues is inconsistency of photo metadata. However, the user must create an event using social media and this framework does not focus on personal photographs.

In this work, we propose a new automatic and semiautomatic scheme for georeferencing photographs that uses other already georeferenced photographs from the same personal collection. Furthermore, we propose a scheme for detection of inconsistencies in the photograph georeferencing that uses the spatiotemporal dimension of the collection. To the best of our knowledge, there is no work in the literature with a similar approach.

III. System Architecture

The proposed approach was integrated to PhotoGeo [1]. PhotoGeo is a digital multimedia library specialized in georeferenced photographs. It has a multilayer architecture and was developed in compliance to the MVC (Model-View-Controller) design pattern.

Figure 1 presents the PhotoGeo architecture, highlighting the data, business logic and view layers. In this work, we added the Photograph Georeferencing module, responsible for integrating the detection of inconsistencies and georeferencing propagation. This module is described in detail the next section.

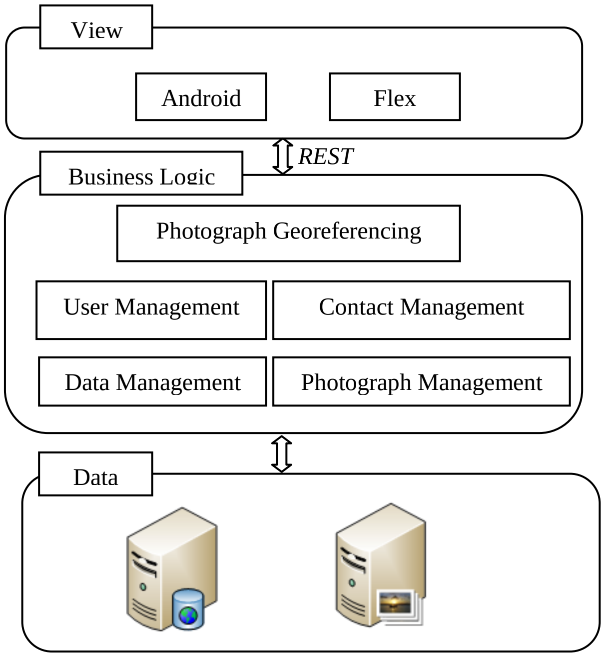

Figure 1. Prototype Architecture.

The data layer comprises two databases: an object-relational database (PostgreSQL) with spatial support (PostGIS); and the user photograph collections.

The business logic layer comprises the following main modules: Data Management, User Management, Contact Management, Photograph Management and Photograph Georeferencing. These modules will be detailed next.

Data Management is responsible for accessing and mapping of data in objects. In this module, JPA (Java Persistence API) is used to make the data persistence.

User Management and Contact Management modules are responsible for managing of system users and contacts, respectively.

The Photograph Management module is in charge of inserting, removing and retrieving photographs, besides extracting metadata. Moreover, in this module the access permissions for the photographs are checked.

The communication between the view layer and the business logic layer is made through REST (Representational State Transfer). Figure 2 shows the prototype interface used to present inconsistencies, and suggestions for the correction and georeferencing of photographs.

IV. Detection and Correction of Inconsistencies

In this section, we present the solution proposed for propagation and detection of inconsistencies in geotags.

A. Geotags Propagation

In this subsection, we present the solution proposed for geotag propagation from georeferenced photographs to non georeferenced ones, in the same personal collection. First, we perform a temporal segmentation on the set of photographs, using the $t_{\max}$ segmentation time input parameter, in minutes.

Assuming that F is a set of n photographs, the temporal segmentation is responsible for separating the photographs into k non intersecting clusters g, in such a way that:

$$\left(\bigcup_{j=1}^{k} g_j\right) = F \tag{1}$$

This segmentation produces clusters whose photographs have a maximum temporal difference of t_{max} minutes between two temporal consecutive photographs. That is, considering that f_i, and f_{i+1} are two consecutive photographs in the cluster g_k, and that t_i and t_{i+1} are their timestamps, then $t_{i+1} - t_i \leq t_{max}$. The photograph collection will be segmented based on Equation (2) and each photograph will belong to exactly one cluster.

$$\begin{gathered} f_1 \in g_1 \\ f_{i+1} \in g_s \text{if } (t_{i+1} - t_i) \leq t_{max} \\ f_{i+1} \in g_{s+1} \text{if } (t_{i+1} - t_i) > t_{max} \end{gathered} \tag{2}$$

For every cluster g_k and considering fg_i as the geographic location of the photograph f_i, the iteration is made in order to find the non georeferenced photographs ($fg_i = \emptyset$). For each photograph with no geotag, f_r, we search for a photograph that is temporally closer, f_s, inside the same cluster, and that is georeferenced ($fg_s \neq \emptyset$). In these cases, photographs with GPS state in interoperability mode are not considered. This mode indicates that the geotag of the photograph may be imprecise. The geotag will be propagated from the photograph f_s to f_r. The propagation can be automatic, without user interaction, or semiautomatic, when the user may accept or reject the suggestion.

When the propagation occurs, the same procedure is done, recursively, for other non georeferenced photographs. So, the new geotag of f_r, which was propagated fromf_s, may be propagated to other photographs.

B. Geotag Inconsistence Detection

In this subsection, we present the proposed algorithm for detection of geotag inconsistences in photographs. This algorithm iterates on a subset of photographs, F, in a personal collection of a certain user, locating the georeferenced photographs. For each georeferenced photograph, f_i, we retrieve the photograph f_j, that has the smallest timestamp difference with respect to f_i. Right after that, the *maximum tolerable spatial distance* (*mtsd*) between the photographs is computed, through Equation (3). This computation is made by computing the difference from f_i timestamp (t_i) to $f_i$timestamp (t_j), and multiplying it by the *mean shift speed* (*mss*).

$$mtsd = \left|t_i - t_j\right| \times mss \ (3) \tag{3}$$

The *mss* is the speed of the camera. However, there can be adverse circumstances, for example, in the case of a photograph being captured from inside an airplane which travels at 900 km/h. In this case, the photographs will have a considerable spatial distance, due to the shift speed. The same problem could happen when pictures are taken from inside a train, car or any other kind of vehicle which travels at high speed. For these situations, it is possible to capture speed information from the GPS chip. The *mss* must be supplied as one of the input parameters of the algorithm.

Right after that, we compute the *spatial distance* (*d*) between the photographs f_i and f_j, using their geotags. In the case d is greater than $mtsd$, the algorithm points out that there is a georeference inconsistence between the photographs, because it is unlikely that, moving from ft_j, with a mean speed *mss*, the user reaches the location of fg_i.

$$\begin{gathered} d = distance(f_i, f_j); \\ \text{if}\begin{cases} d > mtsd, inconsistence \\ d \leq mtsd,\ consistence \end{cases} \end{gathered} \tag{4}$$

I. Evaluation

In this section, we present the experiments carried out to validate the solutions proposed in this work for geotag propagation and inconsistency detection, respectively. Besides, we also present the methodologies used to perform the experiments.

A. Geotag Propagation

In order to validate the geotag propagation scheme, we performed experiments to compare the geotags propagated by the algorithms, both in automatic and semiautomatic mode, with the real location of the non georeferenced photographs.

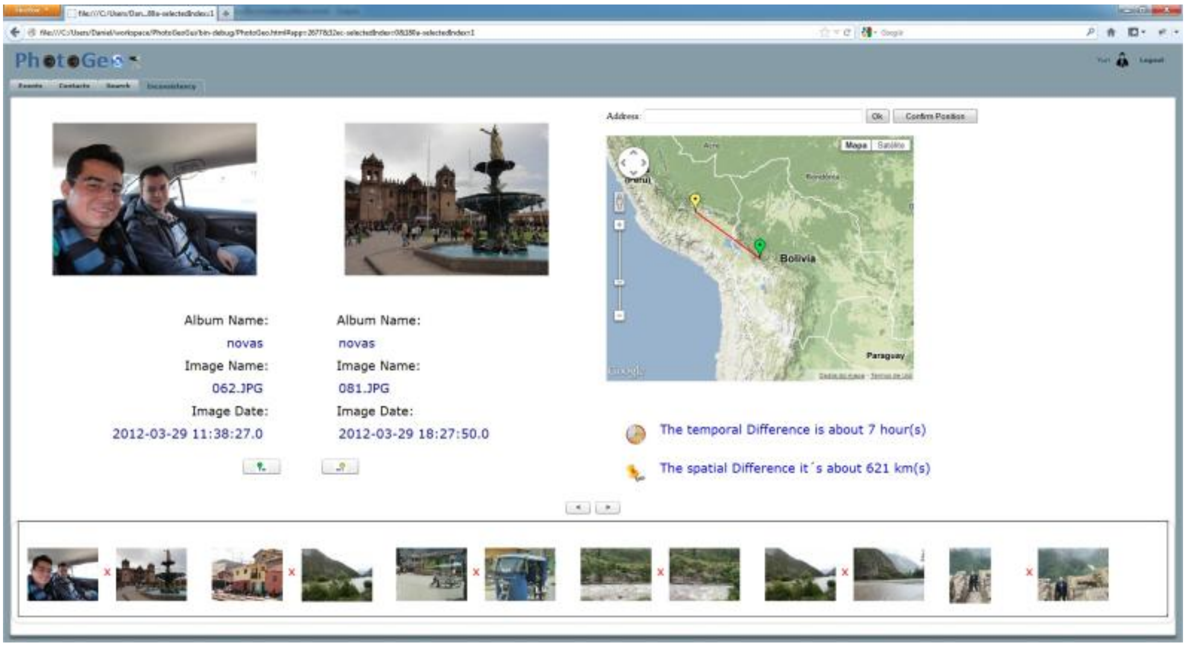

Figure 2. Photograph georeferencing interface.

To carry out the experiments, we used the precision and recall metrics. Precision is defined as the ratio between the correct propagation and all the photographs with geotag propagated. The propagation is considered correct when it is within a *maximum tolerable distance* (*mtd*), in meters, from the real geographic location of the photograph (informed by the user). On the other hand, recall is computed as the ratio between the correct propagation to the all non-georeferenced photographs in the collection.

To perform the experiments, we used a collection containing 4,153 photographs, from which 503 (12.11%) are not georeferenced. In order to automate the experiments, we obtained manually from the users the correct location of each photograph using a map.

Figure 3 presents the results for the computation of precision in semiautomatic mode, varying the parameters *mtd* from 10 to 300, with a step of 10. Each line in that graphic represents an experiment for a different t_{max}. The parameter t_{max} varied from 5 to 60 minutes, at every five minutes. Figure 4 presents precision for the automatic mode, with the same variations of the parameters *mtd* and t_{max}.

We notice that precision increases when the segmentation time t_{max} falls, because the longest the temporal distance between the photographs in the cluster is, more distant will be the propagated geotag, that is, it becomes more imprecise. With respect to the *mtd* parameter, precision is directly proportional, because the longest the tolerable distance is, more propagations will be considered to be correct.

In both cases, the best value of the t_{max} parameter was five minutes. So, we chose the value to analyze the best precision for both modes. Thus, in semiautomatic mode, we achieved precisions varying from 94.89% to 97.08%. On the other hand, in automatic mode, we achieved precisions between 92.05% and 96.97%. We notice, then, in both cases, high precision was achieved, but semiautomatic mode gave better results.

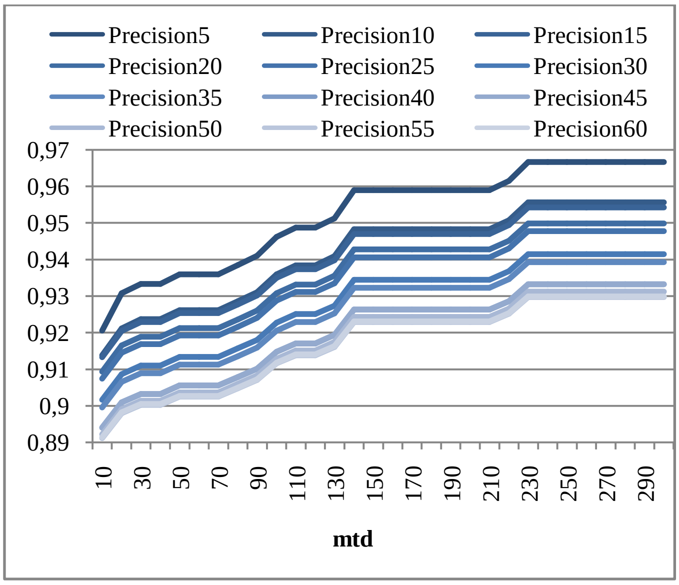

Figure 3 - Precision for the propagation experiment in semiautomatic mode.

For recall, using 5 minutes for t_{max}, we did not achieve good results (71.3% - 74.9% in semiautomatic mode, and 71.1%-72.7% in automatic mode). In Figure 5 we illustrate the result for recall in semiautomatic mode and in Figure 6 for automatic mode. Each line in the graphics represents a variation of the t_{max} parameter. It can be noticed that recall is inversely proportional to t_{max}. This happens because the longer is the temporal distance between photographs in a cluster, the higher will be the number of geotag propagations in this cluster. So, the best value for t_{max} to maximize recall was 60 minutes.

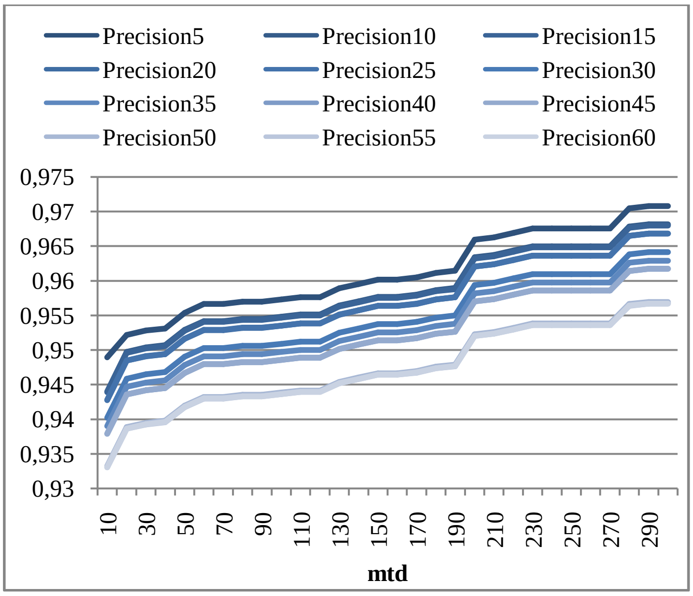

Figure 4 - Precision for the propagation experiment in automatic mode.

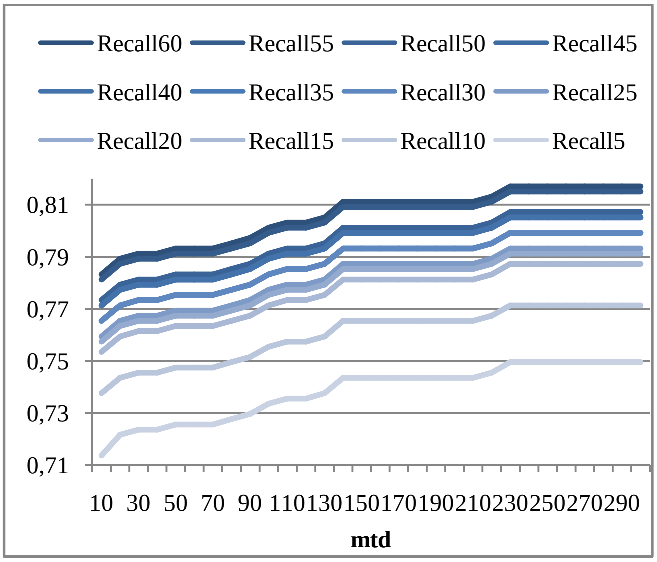

Figure 5. Recall for geotag propagation experiment in semiautomatic mode.

Considering a t_{max} of 60 minutes, recall varied from 78.33% to 81.71% in semiautomatic mode, and from 72.14 to 73.97 in automatic mode.

Figure 7 and Figure 8 illustrate the comparison between precision and recall for the automatic and semiautomatic modes, for t_{max} = 5 minutes (Figure 7) and t_{max} = 60 minutes (Figure 8).

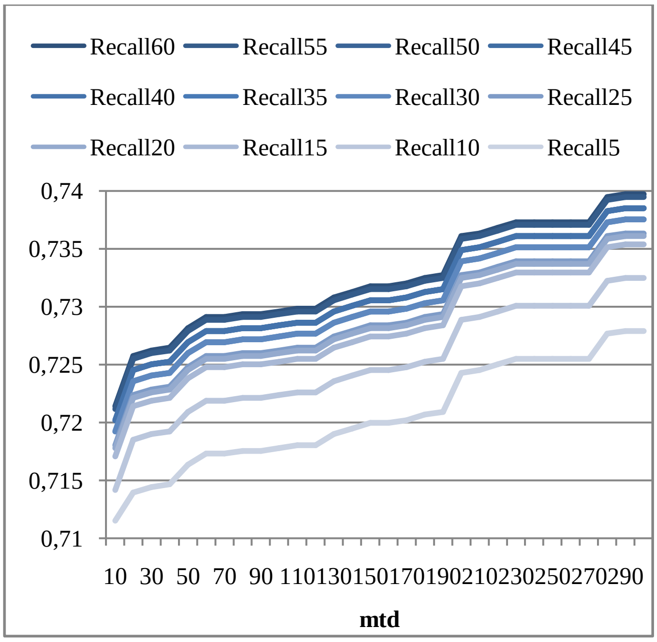

Figure 6. Recall for geotag propagation experiment in automatic mode.

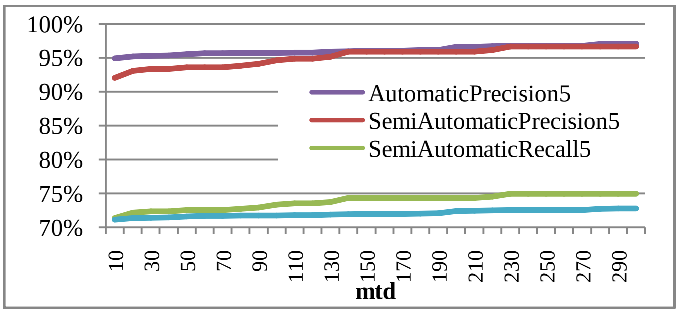

Figure 7. Comparison of recall and precision of the geotag propagation experiment automatic vs. semiautomatic, for t_{max} = 5 minutes.

B. Detection of Inconsistencies in Geotags

To validate the geotag inconsistency detection scheme, we ran an experiment to compare the inconsistencies detected by the scheme and current inconsistencies in a database.

In order to carry the experiments out, we used a real collection with 1,040 photographs captured with a camera withan integrated GPS chip. From those photographs, 944 (90.77%) are georeferenced, among which 112 (11.86%) have the GPS status in interoperability mode. The user informed, through an application, which photographs had inconsistencies in their geotags, so that they could be compared to the inconsistencies pointed by the system, and possibly has allow the automation of the experiment.

Figure 9 presents a graphic with the precision and recall metrics. The parameter *mss* varied from 1 to 120 km/h, in steps of 1 km/h. Through the graphic, we notice that precision is directly proportional to the parameter *mss*, i.e., the higher the mean shift speed, the longer will be the maximum tolerable spatial distance between the photographs, with fewer

 ISBN: 978-1-61208-251-6

inconsistencies detected. However, with an increasingly number of correct indications. For this reason, recall is expected to be inversely proportional to *mss*, since the number of inconsistencies pointed by the algorithm becomes smaller.

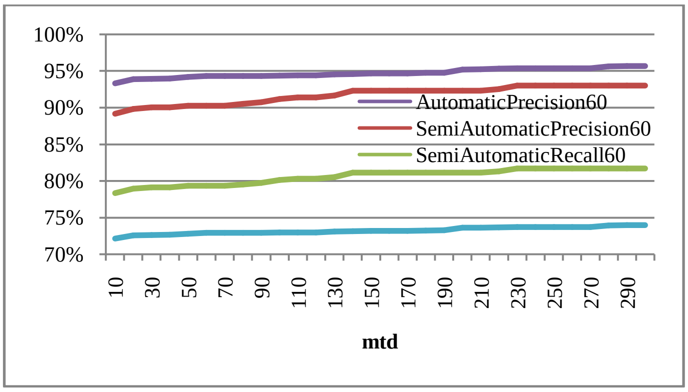

Figure 8. Comparison of recall and precision of the geotag propagation experiment automatic vs. semiautomatic, for t_{max} = 60 minutes.

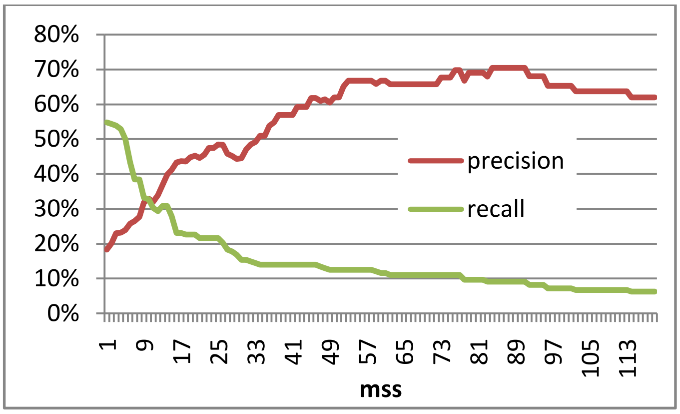

Figure 9. Graphic of precision and recall for the detection of inconsistencies.

The parameters were adjusted, and the highest precision achieved was 70.37%, for *mss* of 77 km/h. The maximum recall was of 55.81%, for *mss* = 1 km/h.

Considering a system for detecting photographs with inconsistencies, allowing the user to correct the geotags, high precision is more interesting for the algorithms, because it avoids the user from getting bored with checking too many inconsistencies erroneously reported. So, for this objective, the best value for *mss* is 77 km/h.

II. CONCLUSION

In this paper, we presented a new method for detection of inconsistencies in geographic locations of photographs, and for the propagation of geotags to non georeferenced photographs.

Both approaches presented had good results. The geotag propagation achieved precision of 97.08% and recall of 73.97% in semiautomatic mode, and precision of 96.76% and recall of 81.71% in semiautomatic mode. On the other hand, the detection of inconsistencies achieved precision of up to 70.37%, proving to be a good alternative for georeference inconsistence correction schemes.

As future work, we will apply new photograph clustering methods and machine learning techniques to estimate the best input parameters for the proposed method. Besides, we will also analyze the behavior of the proposed ideas for other photograph collections.

ACKNOWLEDGMENT

The authors would like to thank the CNPQ - Brazilian Research Council for funding this research.

REFERENCES

[1] H. F. Figueirêdo, Y. A. Lacerda, A. C. Paiva, M. A. Casanova, and C. Souza Baptista, “PhotoGeo: a photo digital library with spatial-temporal support and self-annotation,” *Multimedia Tools and Applications*, vol. 59, no. 1, pp. 279–305, Feb. 2012.

[2] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina, “Automatic organization for digital photographs with geographic coordinates,” in *Proc. of the 2004 joint ACM/IEEE conference on Digital libraries - JCDL '04*, 2004, p. 53.

[3] M. L. Cooper, “Clustering geo-tagged photo collections using dynamic programming,” in *19th ACM international conference on Multimedia - MM '11*, 2011, pp. 1025–1028.

[4] K.-E. Tsay, Y.-L. Wu, M.-K. Hor, and C.-Y. Tang, “Personal Photo Organizer Based on Automated Annotation Framework,” in *IIH-MSP '09*, 2009, pp. 507–510.

[5] G. M. Djuknic and R. E. Richton, “Geolocation and assisted GPS,” *Computer*, vol. 34, no. 2, pp. 123–125, Mar. 2001.

[6] S. S. Lee, D. Won, and D. McLeod, “Tag-geotag correlation in social networks,” in *2008 ACM workshop on Search in social media - SSM '08*, 2008, p. 59.

[7] J. Hays and A. A. Efros, “IM2GPS: estimating geographic information from a single image,” in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[8] L. Hollenstein and R. Purves, “Exploring place through user-generated content: Using Flickr to describe city cores,” *Journal of Spatial Information Science*, vol. 1, no. 1, pp. 21–48, Jul. 2010.

[9] I. Ivanov, P. Vajda, J.-S. Lee, L. Goldmann, and T. Ebrahimi, “Geotag propagation in social networks based on user trust model,” *Multimedia Tools and Applications*, vol. 56, no. 1, pp. 155–177, Jul. 2012.

[10] D. Joshi, A. Gallagher, J. Yu, and J. Luo, “Inferring photographic location using geotagged web images,” *Multimedia Tools and Applications*, pp. 131–153, 2012.

[11] P.-J. Vandormael and P. Couderc, “An Algorithm to Ensure Spatial Consistency in Collaborative Photo Collections,” in *SEUS '09*, 2009, pp. 332 – 342.

[12] C. Zigkolis, S. Papadopoulos, G. Filippou, Y. Kompatsiaris, and A. Vakali, “Collaborative event annotation in tagged photo collections,” *Multimedia Tools and Applications*, Jun. 2012.

 ISBN: 978-1-61208-251-6