Semi-Automatic Photograph Tagging by Combining Context with Content-Based Information

Hugo Feitosa de Figueirêdo[1,2], Cláudio de Souza Baptista[1], Marco Antônio Casanova[3],Tiago Eduardo da Silva[1], Anselmo Cardoso de Paiva[4]

[1]*University of Campina Grande, Computer Science Department, Av. Aprígio Veloso, 882, Bodocongó - Campina Grande, Paraíba, 58109-900, Brazil.*

[2] *Federal Institute of Education, Science and Technology of Paraíba – Campus Monteiro, Monteiro, Paraíba, Brazil*

[3] *Pontifical Catholic University of Rio de Janeiro, Av. Marques de São Vicente, 225 RDC*
CEP 22451-900 - Gávea - Rio de Janeiro, Rio de Janeiro, Brazil.

[4] *Federal University of Maranhão,Applied Computing Group NCA, Av. dos Portugueses, SN - São Luís, Maranhão, 58109-900, Brazil.*

hugoff@dsc.ufcg.edu.br, baptista@dsc.ufcg.edu.br, casanova@inf.puc-rio.br, paiva@deinf.ufma.br

Abstract. This paper proposes a semi-automatic technique for the annotation of people in photographs. The technique uses context and content information and is based on the weighted sum of estimators, which results in a list of the person's contacts more likely to be present in a photograph. Multivariable linear regression and slope functions are adopted to weight the estimators. The paper also describes the results of experiments performed with a collection of 4,050 photographs, with 365 different people, which indicate that the proposed technique outperforms techniques that adopt just context or just content, using as performance metrics the H-Hit rate of correct annotations.

Keywords: *Personal Photo Collections; Geo-Referenced Photos; People Photo Annotation; Metadata; Face Recognition; Context-Aware Multimedia.*

1. Introduction

Studies show that the annotation of information about the context in which a photograph was taken helps retrieve it. In particular, identifying people present in the scene is considered the most important information for a person to remember a photograph [11]. However, most digital photographs are not annotated with such information due to the low recall and precision of the automatic annotation techniques based on face recognition [13].

The process of photograph annotation can be performed in three different ways: (1) manually, where the user performs the photo annotations himself; (2) semi-automatically, where an application helps the user annotate the photograph; (3) automatically, where an application annotates photographs, perhaps with the help of external sources and content-based techniques. Techniques to annotate people in photographs may in turn be divided into three groups: (1) techniques that use face recognition; (2) techniques based on the use of patterns of contextual information related to the environment where the photograph was taken; (3) hybrid techniques, which improve the performance of face recognition algorithms by using patterns of contextual information.

Semi-automatic techniques to annotate people in photographs are reported [4, 11, 12]. In particular, improved techniques based on contextual information typically use estimators to generate a recommendation of the people most likely to be present in a certain photograph. The estimators are statistics based on contextual information of the previously captured photographs. Examples of estimators are: general recurrence of a contact in the collection of photographs; spatial recurrence of a contact in a certain area; and temporal recurrence of a contact in photographs taken within a time interval. Considering the hit rate of a person present in both the photograph and in the recommendation list, estimators vary according to precision and recall. However, in most cases, combining several estimators leads to better overall results.

Automatic face recognition techniques, despite their low precision and coverage, undoubtedly helps annotating people in photographs. Moreover, a semi-automatic annotation approach, which combines face recognition with contextual information, turns out to be a very attractive solution. Indeed, in applications following this approach, the user has to manually inspect for annotation only those people who were not detected by the face recognition algorithm; furthermore, the application reduces his work by presenting a recommendation list from which he selects people. In addition, the rate of correct recommendations can be gradually improved, if the algorithm incorporates a feedback module that learns from the user's selections.

In this paper, we propose a semi-automatic technique for the annotation of people in photographs. The technique combines contextual information and content analysis to generate a recommendation list of people possibly present in a photograph. It merges two estimators, where the first estimator uses linear regression and the second one uses the slope function for weighting.

The main contribution of the paper is a semi-automatic strategy for the annotation of people in photographs with the following characteristics: (i) it uses context plus content for annotation; (ii) it automatically computes estimator weights and proposes two weighting techniques to combine estimators; (ii) it incorporates of estimator filtering. By context we consider spatiotemporal metadata and list contacts and by content we mean face recognition. The paper also describes the results of experiments performed with a collection of 4,050 photographs, which indicate that the proposed technique outperforms techniques that adopt just context or just content, using as performance metrics the rate of correct annotations.

The rest of this paper is organized as follows. Section II describes related work. Section III overviews the proposed people photograph annotation method. Section IV concentrates on the technique used in annotation module, which combines context and content. Section V highlights an experimental evaluation of the proposed solution. Section VI discusses the results of the experiments carried out. Finally, Section VII contains the conclusions and suggestions for future work.

2. Related Work

For the annotation of people present in photographs, the most common mechanisms use face recognition algorithms or contextual information. However, as already pointed out, it is possible to combine both approaches. In this section, we cover some related work that uses content, context or both to perform the annotation of people in photographs.

Naaman et al. [11] use the idea that there are patterns for the presence of people in photographs. These patterns are used to estimate the probability that a certain person be present in a photograph. The following estimators have been proposed: popularity, co-occurrence, spatial and temporal recurrence. An application following this strategy suggests a list of people with a high chance of being in a certain photograph by calculating these estimators. The probability of a person being present in a photograph is calculated as the sum of all estimators, using the same weight for each of them. As, this work uses a simple sum of the estimators, it do not take in account the fact that some of them may be more relevant than others. This might reduces the accuracy.

PhotoMap [15] is a system that uses Bluetooth technology to identify and annotate friends nearer to the place where a certain photograph was taken. The system annotates people that can be in the photo, as they are near to the photo capture place. But some of the annotated people may be in the photo and others may not be. So, both precision and recall may be affected due to erroneous people photo annotation.

MediAssist [12] is a system that offers browsing, search and semi-automatic annotation of personal photographs by analyzing the content of the photograph and the context where it was taken. The main semi-automatic annotation performed by MediAssist identifies people in the photographs. To build the list of people with a higher chance of being in the photograph, MediAssist uses weighting techniques to combine the estimators. In the weighted combination, the weights were defined through a brute force approach, evaluating all possible values for the weights and adopting those which returned better results.

MMM2 [3] is a system which uses a client-server architecture to perform the semi-automatic annotation of people in photographs captured in mobile devices. To help the user in the annotation process, MMM2 uses face recognition, data and time of capture, geographic location

(CellID and GPS) of the mobile device at the moment of capture, and people nearby. This last geographic context information is obtained from a mapping between users and Bluetooth.

PhotoGeo [4] aims at helping in the annotation of people in photographs, the annotation of events in photographs and the storage and retrieval of personal digital photographs. PhotoGeo uses a simple summation of estimators to generate the recommended list for semi-automatic annotation of the people present in the photographs.

ACRONYM [10] uses analysis of factors and Pearson correlation to weight the values of the estimators. However, ACRONYM does not allow the addition of new estimators and does not filter the estimators which are more relevant for each user.

Gallagher and Chen [5] created a model for the relationship between context and people recognition in photographs. The context information was divided into three categories: pixel context, camera context and social context. Pixel context information is extracted from the image itself, such as clothing, other people, glasses and hats. Camera context information is extracted from the photograph metadata, such as flash, location, time and luminosity. Finally, social context information concerts people and their social networks.

Cooray and O'Connor [2] use face and body characteristics, combined with date and time, to perform semi-automatic annotation of people. They demonstrated that the semi-automatic annotation per event – automatically detected from the date and time of capture of the photograph – is more efficient that the techniques that use just the recognition in a collection of photographs that does not consider the event.

Lin, et al. [9] proposed a probabilistic model that combines the people, events and location domains for the annotation of information about these domains in photographs, using face recognition for the annotation of people and context information for event detection.

Hulsebosch and Ebben [6] used information about the location of people to improve the performance of face recognition. The location of people is discovered by means of Bluetooth and RFID.

Finally, we remark that, when a semi-automatic technique is used for annotation of people present in photographs, it is important to weight each estimator to achieve a higher hit rate in the annotations. However, most of the related work does not use any weight technique to combine the estimators. In addition, no work proposed the automatic calculation of the estimator weights, the filtering of estimators and the use of context plus content to annotate people present in photographs, which is the focus of this paper.

3. People Photo Annotation

Let F be a set of n photographs of a certain user, organized as albums, and P be a set of people who appear in the photographs.

For each photograph $f \in F$, let $p_f \subseteq P$ be the set of people who indeed occur in f. Given f, a people photograph annotation process creates a set $H_f \subseteq P$ of people suggested to be in f. A perfect system is such that $H_f = p_f$. Let $i \in P$. We say that a *hit occurs for* i iff $i \in H_f \cap p_f$, and that a *miss occurs for* i iff $i \in p_f - H_f$. We say that the system has a *hit* for f iff $H_f \cap p_f \neq \emptyset$, that is, at least one person from H_f is correctly associated with f; otherwise, we say that the system has a *miss*, that is, the system did not suggest anyone who actually is in f.

A Brief Summary of PhotoGeo

In this paper, we extend the people photo annotation method introduced in PhotoGeo [4]. PhotoGeo predicts the presence of a person in a given photograph using recurrence patterns of contextual information in other photograph of the same person. A ranking is generated with the people more likely to be present in a certain photograph in a given context. The main contextual information used to generate the ranking is the spatio-temporal information of the photographs.

The information about geographic and temporal location is extracted from metadata using well-known formats, such as the EXIF and XMP. The contextual information extracted from a photograph is used by the estimators. This information is analyzed, together with the information of photographs already annotated. Then, the probability that each of the contacts be present in the photograph is estimated.

The results returned by the estimators are combined to generate a single value for each contact, representing the probability that the contact be present in the photograph. Next, a list with the H contacts more likely to be present in the photograph is computed. This list is presented to the user to help him with the identification of the people in the photograph.

Combining Content and Context for People Photo Annotation

For the semi-automatic annotation of people in photographs, we adopt a technique based on the weighted sum of estimators. The proposed algorithm uses context-based and content-based estimators to generate a list with the people who are more likely to be present in the photograph. People who were already annotated are removed from the list, and these annotations are used to improve the rate of correct annotations of the recommendation list.

Context-Based Estimators

The techniques proposed by Naaman et al. [11] and Lacerda et al. [7] use estimators which evaluate the probability that a person already identified by the system is present in a photograph. In this section, we briefly introduce the estimators used in the paper.

Let K be the set of people already identified by the system. The definition of the first three estimators use Eq. 1, for each $i \in K$:

$$R(i,X) = \sum_{f \in X} \frac{present(i,f)}{|X|}, \text{with } present(i,f) = \begin{cases} 1 \; if \; i \in p_f \\ 0 \; otherwise \end{cases} \qquad (1)$$

where X is a set of photographs that depends on the specific estimator, as explained below, and p_f is the set of people who indeed occur in photograph f, as defined before.

[GR] *Global Recurrence:* percentage of times that a contact was added to a user's photograph collection. To compute this estimator, we take X in Eq. 1 as the set of all annotated photographs.

[SR] *Spatial recurrence:* percentage of times that a contact was added within a distance *d* from the photograph to be annotated. In this work, we used *d* = 1,000 meters. Let $C \subseteq F$ be the set of photographs taken within *d* meters from the geographic location of the photograph. To compute this estimator, we take *X=C* in Eq. 1.

[TR] *Time recurrence:* percentage of times that a contact was annotated in an interval $[t_B,t]$, where *t* is the moment that the photograph to be annotated was captured. In this work, we used *t* = 12 hours. Let $T \subseteq F$ be the set of photographs taken in the interval $[t_B,t]$. To compute this estimator, we take *X=T* in Eq. 1.

[CO] *Co-occurrence:* percentage of times that a contact is present in a photo simultaneously with other contact that is already annotated for that photo. Considering two contacts i_1 and i_2, where $i_2 \in p_f$ (is already annotated as present in the photo f). The co-occurrence of i_1 and i_2 is expressed in Eq. 2:, for each $i \in P$:

$$CO(i_1, i_2) = \frac{\sum_{f=0}^{N_F} present(i_1, i_2, f)}{\sum_{f=0}^{N_F} present(i_1, f)},$$

$$\text{with } present(i_1, i_2, f) = \begin{cases} 1 \ if \ i_1, i_2 \in p_f \\ 0 \ otherwise \end{cases} \quad (2)$$

where p_f and $present(i, f)$ are as in Eq. 1.

As we must use just one co-occurrence estimator value for each $i \in P$, we consider the co-occurrence of i_1 as (Eq. 3):

$$CO(i_1) = Max_{i_2 \in p_f}\left(CO_{i_1,i_2}\right) \quad (3)$$

The performance of the estimators SR and TR respectively depends on the values of parameters *d* and *t*. In the experimental part of this paper (see Section EXPERIMENTS), we considered different values for *d* and *t*. In fact, the results obtained for the estimators SR and RT do not justify attributing a higher significance to geographic location than for time.

Content-Based Estimator

One of the techniques used for the annotation of people in photographs, as mentioned earlier, is based on an analysis of the content of the photograph. In this paper, we adopted a content-based strategy that uses face detection / face recognition algorithms.

The face detection step tries to identify regions in a photograph which represent faces. We adopted the classification algorithm proposed by Lienhart and Maydt [8] that uses Haar features. As input to the algorithm, we used a training base containing frontal face images. So, the classifier will not select faces which are not in a frontal position in a photograph. After the regions representing faces are identified in a photograph, they are inserted in the database.

The face recognition step tries to match the faces detected in the first step with those stored in a database, using the EigenFace algorithm proposed in [14]. Recall that *K* is the set of people already identified by the system. A score for each person $i \in K$ is calculated and stored in the database. The face recognition step returns a list H_f with the n best classified persons for a photograph f. If a person is scored in more than one region for the same photograph, the highest score will prevail. Finally, the user it then invited to confirm the persons returned in H_f.

The content-based estimator is performed adopting the score in the face recognition. The new estimator is the following:

[FR] *Face Recognition:* Higher contact rate between the values returned by the face recognition algorithm applied to the faces detected in the photograph.

This estimator is used only in the semi-automatic annotation technique based on both context and content. In the automatic annotation technique, this estimator is used separately, considering just one suggestion.

Generation of the Recommendation List

In order to generate the recommendation list H_f, the people photo annotation method may adopt various estimators to evaluate the probability $Pr_{i,f}$ that a person $i \in P$ is in p_f. Weights are used to combine the estimators. The candidates with higher probability are included in the set H_f.

The final value of $Pr_{i,f}$ is given by Equation (4):

$$\Pr_{i,f} = \sum_{e \in E} \left[estimator(e, i, f) * weight(e) \right] \quad (4)$$

where E is the set of estimators, *estimator* is a function that returns the value of estimator e for person i be in photograph f, *weight* is a function that returns the weight of estimator e. In the next subsection, we describe the weighting of estimators.

Weighting of the Estimators

Weighting is the act of increasing or decreasing the importance of certain criteria in the calculation of a certain value. The problem of weighting is known as an optimization problem, i.e., finding the combination that returns the extreme value of a function (maximum or minimum).

Estimators may be combined either by filling or by weighting, with the objective of obtaining the best result. In the combination by filling, estimators are put in a list in such a way that the most precise are put first and, then, the rest of the list is filled with a combination of the less precise, but with higher recall.

Some weighting techniques use heuristics to determine which estimators have higher weights in the combination. However, these techniques have no mathematical foundation to prove that the adopted weightings are the best possible.

For this reason, we propose in this paper the use of two techniques for weighting estimators. The first technique adopts multivariable linear regression, using the method of minimum squares, to obtain the weights. The second technique uses the slope function, which is the variation rate in the regression line between two points.

To calculate weights using both techniques, we generate a file with the following user information: contact identification, global recurrence, temporal recurrence, co-occurrence and a binary value to inform whether the contact is present in the photograph (1) or not (0).

The method M5 was used for variable selection . It removes the estimators that do not contribute to the improvement of the error estimation given by the minimum squares method.

We introduce Equations (5), (6) and (7) to calculate the weights using the slope function:

$$Weight(e) = \frac{\sum_{f=0}^{N_F}\left\{\sum_{i=0}^{N_f}\left[estimator(e,i,f) - \overline{estimator(e)}\right].\left[present(i,f) - \overline{present(i)}\right]\right\}}{\sum_{f=0}^{N_F}\left\{\sum_{i=0}^{N_f}\left[estimator(e,i,f) - \overline{estimator(e)}\right]^2\right\}} \quad (5)$$

$$\overline{estimator(e)} = \sum_{f=0}^{N_F}\left[\sum_{i=0}^{N_f}\frac{estimator(e,i,f)}{N_f}\right] \quad (6)$$

$$\overline{present(i)} = \sum_{f=0}^{N_F}\frac{present(i,f)}{N_F} \quad (7)$$

where functions *estimator* and *weight* are as in Eq. 4, *present* is a function that returns a binary value that represents the presence of person i in photograph f.

O slope é utilizado com um peso para majorar a significância dos estimadores que com pequenos valores (valores próximos de 0) indicam a presença da pessoa i na fotografia f.

The weighting of the estimators must be periodically recomputed (however, we have not experimentally verified how often the weighting must be recomputed).

4. Experiments

Datasets

To evaluate the techniques for combining estimators, we used four different collections of personal photographs, with a total of 4,050 pictures with 365 different people (see Table 1). These people were annotated 7,912 times. All photos have in their metadata the date and time and geographic location and were manually annotated with the people present by the owners of the collections.

Each collection was partitioned into two sub-collections, each containing 50% of the original collection. The first sub-collection was used for training, i.e., weighting the estimators, whereas the second one was used to evaluate the combination techniques.

Table 1. Statistics of the collection of photographs used in evaluation.

Collection	Photographs	Different People annotated	People annotation
1	1,644	110	2,851
2	648	51	698
3	898	186	2,936

4	860	18	1,427
Total	4,050	365	7,912

We considered two situations for evaluation. In the first one, the user wants to annotate all people that he knows to be present in his photograph collection. In the second situation, the user wants to annotate only the 20 most popular contacts of his photograph collection. For each situation, eight experiments were performed: four experiments using each estimator separately; one using the sum of each estimator; one using the weighted sum obtained by linear regression; one using weighted sum by slope; and one using weighted sum obtained by the Pearson correlation proposed in ACRONYM [10] for combination of estimators. Hence, 16 experiments were performed for each collection, giving a total of 64 experiments

We used the H-Hit rate as metrics for the evaluation of the recommended list generated by each technique, since it is used in the evaluation of several other related works [1, 11, 12].

Weighting Evaluation

Tables 2, 3 and 4 present the weights obtained for the estimators of the four collections of photographs used in the experiment obtained by the linear regression, the slope and the correlation weighting techniques, respectively.

As we can observe in Table 2, the global recurrence estimator (GR) was discarded for all the collections using the linear regression technique to weight the estimators. The weights of collections 1 and 2 obtained the lowest standard deviations (0.1), and collection 3 obtained the highest (0.23). These deviations demonstrate a considerable difference between the collections used for evaluation, where collection 3 is highlighted due to the high ratio between the number of people annotations and the number of photographs, with the co-occurrence estimator (CO) having the highest weight (0.61). The deviations of the weights per estimator were not as significant as per collection, with the spatial recurrence estimator (SR) having the highest standard deviation (0.11).

Table 2. Weights of the estimators computed by the linear regression technique.

Collection	GR	SR	TR	CO	Mean	Deviation
1	X	0.31	0.49	0.47	0.34	0.10
2	X	0.37	0.29	0.49	0.31	0.10
3	X	0.17	0.30	0.61	0.30	0.23
4	X	0.14	0.38	0.44	0.25	0.16
Mean	X	0.25	0.37	0.5		
Deviation	X	0.11	0.09	0.08		

In Table 3, we can see that the global recurrence estimator (GR) has the smallest weight for all the collections and has the smallest mean (0.15).

Table 3. Weights of the estimators computed by the slope technique.

Collection	GR	SR	TR	CO	Mean	Deviation
1	0.17	0.33	0.41	0.39	0.33	0.11
2	0.17	0.30	0.30	0.32	0.27	0.07
3	0.12	0.27	0.30	0.54	0.31	0.17
4	0.14	0.19	0.22	0.25	0.20	0.05
Mean	0.15	0.27	0.31	0.38		
Deviation	0.02	0.06	0.08	0.12		

In Table 4, we can see that the global recurrence estimator (GR), despite having the smallest weight in all the collections, was not discarded because the correlation and slope weighting techniques do not eliminate estimators that do not alter results. Contrasting with the linear regression technique, the standard deviations of the weights per collection and per estimator did not obtain high values and, once again, the deviation of the weights of collection 3 had the highest values. The deviations of the weights in the collections and in the estimators for the slope weighting technique were higher than those obtained through correlation. So, the slope technique results in more spread weights than the correlation technique.

Table 4. Weights of the estimators computed by the correlation technique.

Collection	GR	SR	TR	CO	Mean	Deviation
1	0.41	0.54	0.59	0.54	0.52	0.08
2	0.35	0.52	0.52	0.51	0.48	0.08
3	0.33	0.46	0.47	0.62	0.47	0.12
4	0.32	0.35	0.41	0.42	0.38	0.05
Mean	0.35	0.47	0.5	0.52		
Deviation	0.04	0.09	0.08	0.08		

The rest of this section contains figures where the X axis represents the value of *H*, i.e., the size of the recommendation list, and the Y axis represents the rate of correct recommendations.

In Figure 1, we show the results of the experiment for the annotation of all the contacts of each collection. Collection 3 had the worst results among the collections for all estimators and values of *H*. Analyzing the estimators separately, i.e., without combinations, SR was the best estimator for collections 2 and 3, where TR, GR and CO are the next best estimators, in this order. On the other hand, for collections 1 and 4, the TR estimator had the best results, followed by SR, GR and CO, in this order. So, we may conclude that each collection has a distinct behavior and may have different weights for each estimator.

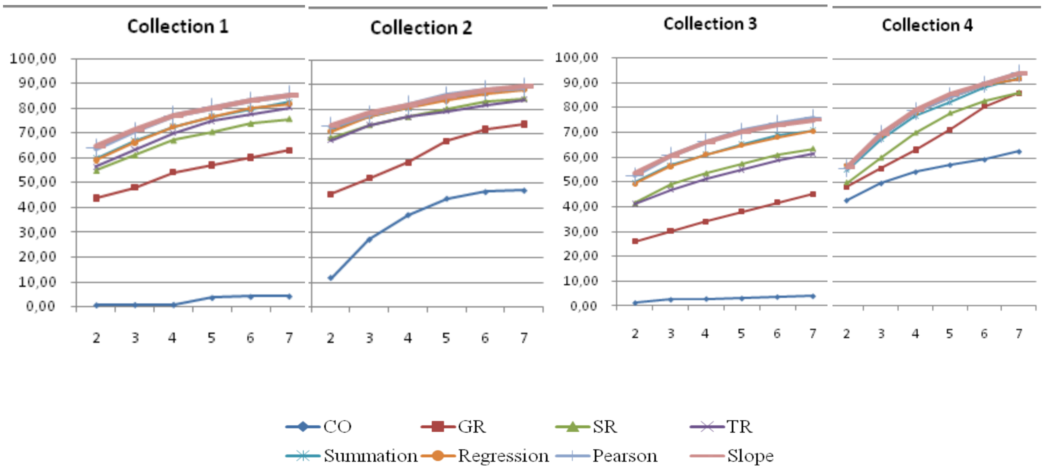


Figure 1. Results for the classification of all contacts of each collection.

In Figure 2, we present the results for annotations of all the contacts of all collections together. The change in size of the recommendation lists (*H*) affects the rate of correct recommendations in approximately 5% between the sizes 2 and 3. However, this change decreases to approximately 1% when *H* varies from 6 to 7. So, the change in this rate of correct recommendations decreases noticeably above 5 recommendations.

The rate of correct recommendations for the simple sum of estimators obtained a mean of 74% for $H = 5$, while the weighted sum using linear regression obtained a rate of 75%, with cases where the sum had better results. Therefore, the weighting using the linear regression technique had a performance approximately equal to that of the non-weighted combination, but using one less estimator.

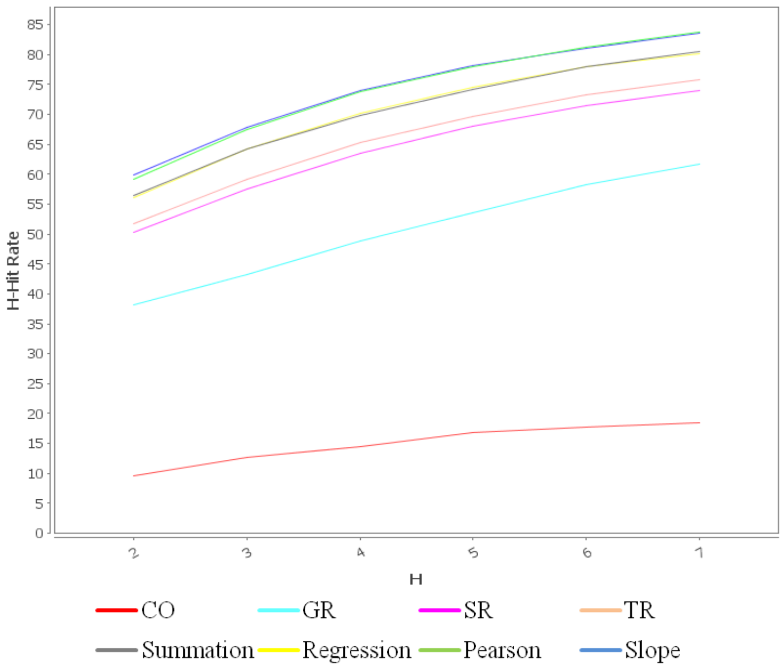


Figure 2. Results for the classification of all the contacts of all collections.

On the other hand, the weighted sum using the slope technique had the best results for all collections and for all values of *H*, except for collection 3 with *H* values of 6 and 7, where the correlation technique had the best results. The rate of correct recommendations using the slope technique increased about 4% with respect to the simple combination and weighted combination using linear regression. For *H*=5, this rate was approximately 79%, achieving 85% for collections 2 and 4. The worst rate (70%) for this value of *H* using the slope technique was for collection 3. However, in this collection, it was observed the biggest increase of the rate of correct recommendations compared to the simple combination and the weighted combination using linear regression, with an increase of about 5%.

To evaluate how a smaller number of contacts affect the system, we performed experiments using just the 20 most popular contacts of each collection.

In Figure 3, we show the results of the experiments for the annotation of the 20 most popular contacts of each collection. Collection 4 had no changes because it has annotated only with 18 different contacts. Collection 3, on the other hand, had the worst results in this experiment, having a mean increase of 15%. Collection 2 had the best results, obtaining a rate of correct recommendations of 91%, for *H*=5. The best rate obtained in all experiments was for collection 4, for H=7, of approximately 95%.

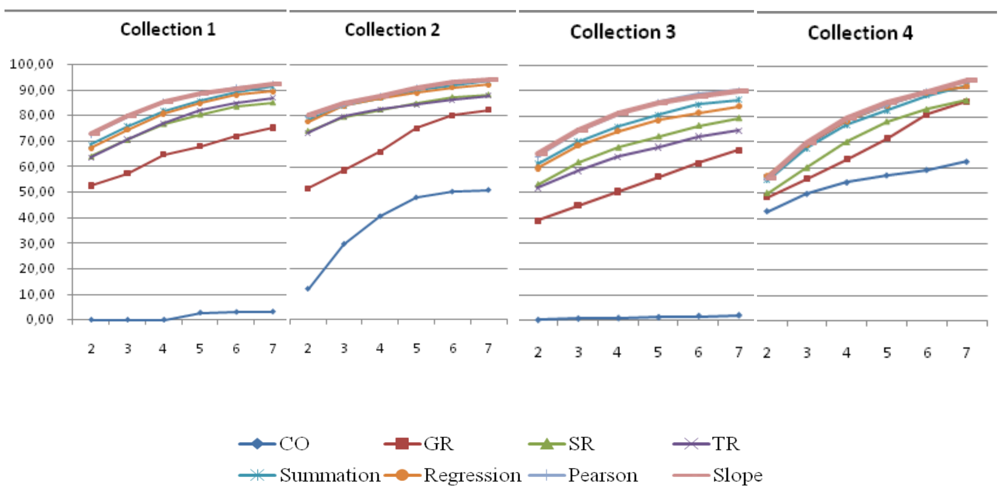


Figure 3. Results for classification of the 20 most popular contacts of each collection.

Figure 4 shows the results for the annotation of the 20 most popular contacts of each collection of the whole collection. The rate of correct recommendations, compared with the experiment that annotated all the contacts, increased 9% in average. The difference of rates among estimators decreased due to the higher increase of the isolated estimators without combination with respect to the combined estimators.

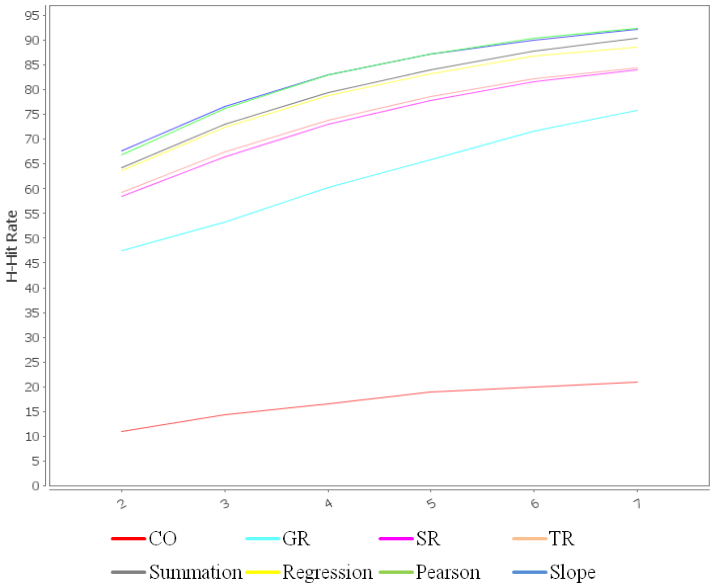


Figure 4. Results for classification of the 20 most popular contacts of each collection.

Annotating only the 20 most popular contacts, the simple combination had results approximately equal to those of weighted combination using linear regression, being the non-weighted slightly better.

For the combination using weighted sum obtained by the slope technique, annotating just the 20 most popular contacts obtained once more the best results for all users and all values of *H* that were analyzed.

The ranking of the contacts for the recommendation list can be based on an estimator. However, the combinations using the simple sum and weighted sum obtained results better than those of isolated estimators.

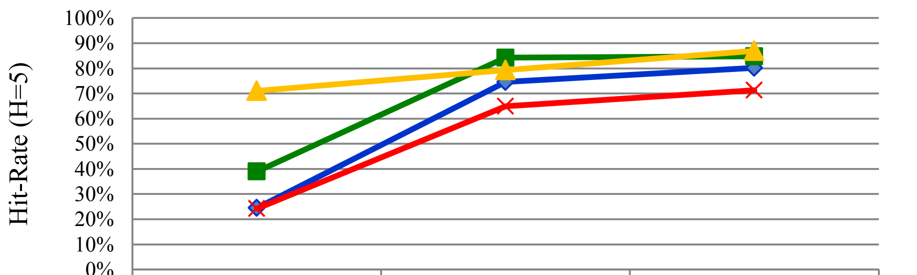
The results of the combination of estimators using the weighted sum obtained by the linear regression technique were equivalent to those of the simple sum. However, the weighted sum had the advantage of eliminating the use of one estimator - the global recurrence - decreasing the processing time. Hence, this weighting technique is interesting for evaluating which estimator must compose the value that will be used in the contact ranking for the recommendation list. This technique seems to be useful when there are many estimators to be used.

The combination of estimators using a weighted sum obtained through the slope technique achieved the best results with the H-Hit rate metric. Nevertheless, this method does not perform the selection of estimators, and so, it is not recommended to be used with a large number of estimators.

Context and Content Evaluation

In this section, we present the results of the experiments with the people annotation algorithms using context information only (EC), content analysis only (RF) and context plus content (RF+EC).

In Figure 5, we present the results of the experiments. As it can be seen, the algorithm that analyzes content and context information to recommend the presence of people in photographs has a higher rate of correct recommendations, as compared with the algorithms that analyze just context information or just content information.



	Content	Context	Content+Context
A	25%	75%	80%
B	39%	84%	85%
C	71%	79%	87%
D	24%	65%	71%

Figure 5. Comparative between people photo annotation methods.

5. Conclusions

In this paper, we presented a people photo annotation method that uses a technique for combination of context and content using weighted sum of estimators. Thus, we proposed the use of two techniques for weighting estimators. The first one uses linear regression to select and obtain the weights of the estimators, being useful to filter estimators, because it selects the best ones for each user. The second technique uses slope to weight each estimator.

The method obtained better results than the context-based and content-based methods when the metric H-Hit rate was used. Furthermore, the experiments showed that the weighting technique obtained by the slope method achieved better results when the metric H-Hit rate was used, obtaining a rate of correct recommendations of 85% for the collection with the best results. When annotating just the 20 most important contacts of each collection, the rate of correct suggestions for a list of 5 suggestions raised to 87%, for all the collections, and to 91%, for the collection with better results.

As a future work, we plan to use face recognition in a step before the generation of the recommendation list, so that the semi-automatic annotation is done only in the cases not covered by the automatic annotation. Besides, experiments will be carried out to verify the ideal frequency of the application of weighting for a certain user's collection.

Acknowledgements

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

References

[1] Choi, J.Y., De Neve, W., Plataniotis, K.N. and Ro, Y.M. 2011. Collaborative Face Recognition for Improved Face Annotation in Personal Photo Collections Shared on Online Social Networks. *IEEE Transactions on Multimedia*. 13, 1 (Feb. 2011), 14–28.

[2] Cooray, S.H. and O'Connor, N.E. 2009. Enhancing Person Annotation for Personal Photo Management Applications. *2009 20th International Workshop on Database and Expert Systems Application* (2009), 251–255.

[3] Davis, M., Smith, M., Stentiford, F., Bamidele, A., Canny, J., Good, N., King, S. and Janakiraman, R. 2006. Using Context and Similarity for Face and Location Identification. *Proceedings of the IS&T/SPIE 18th Annual Symposium on Electronic Imaging Science and Technology Internet Imaging VII* (San Jose, California, 2006).

[4] Figueirêdo, H.F., Lacerda, Y.A., Paiva, A.C., Casanova, M.A. and Souza Baptista, C. 2011. PhotoGeo: a photo digital library with spatial-temporal support and self-annotation. *Multimedia Tools and Applications*. (Feb. 2011).

[5] Gallagher, A.C. and Chen, T. 2009. Using Context to Recognize People in Consumer Images. *IPSJ Transactions on Computer Vision and Applications*. 1, (2009), 115–126.

[6] Hulsebosch, R.J. and Ebben, P.W.G. 2008. Enhancing Face Recognition with Location Information. *2008 Third International Conference on Availability, Reliability and Security* (Mar. 2008), 397–403.

[7] Lacerda, Y.A., de Figueirêdo, H.F., de Souza Baptista, C. and de Paiva, A.C. 2008. Expandindo e utilizando informações de contexto para a sugestao de anotações de fotografias digitais. *Proceedings of the 14th Brazilian Symposium on Multimedia and the Web - WebMedia '08* (New York, New York, USA, 2008), 162–169.

[8] Lienhart, R. and Maydt, J. 2002. An extended set of Haar-like features for rapid object detection. *Proceedings. International Conference on Image Processing* (2002), I–900–I–903.

[9] Lin, D., Kapoor, A., Hua, G. and Baker, S. 2010. Joint People , Event , and Location Recognition in Personal Photo Collections using Cross-Domain Context. *Proceedings of The 11th European conference on Computer vision: Part I* (Heraklion, Crete, Greece, 2010), 243–256.

[10] Monaghan, F.P. 2008. Context-Aware Photograph Annotation on the Social Semantic Web. *Framework*. December (2008).

[11] Naaman, M., Yeh, R.B., Garcia-Molina, H. and Paepcke, A. 2005. Leveraging context to resolve identity in photo albums. *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries - JCDL '05* (New York, New York, USA, 2005), 178.

[12] O'Hare, N. and Smeaton, A.F. 2009. Context-Aware Person Identification in Personal Photo Collections. *IEEE Transactions on Multimedia*. 11, 2 (Feb. 2009), 220–228.

[13] Pham, T.-T., Maillot, N.E., Lim, J.-H. and Chevallet, J.-P. 2007. Latent semantic fusion model for image retrieval and annotation. *Proceedings of the sixteenth ACM conference on*

Conference on information and knowledge management - CIKM '07 (New York, New York, USA, 2007), 439–444.

[14] Turk, M. a. and Pentland, A.P. 1991. Face recognition using eigenfaces. *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1991), 586–591.

[15] Viana, W., Miron, A.D., Moisuc, B., Gensel, J., Villanova-Oliver, M. and Martin, H. 2011. Towards the semantic and context-aware management of mobile multimedia. *Multimedia Tools and Applications*. 53, 2 (Mar. 2011), 391–429.

*Highlights (for review)

- We propose a semi-automatic technique for the annotation of people in photographs;
- The technique uses context and content information and is based on the weighted sum of estimators;
- Multivariable linear regression and slope functions are adopted to weight the estimators;
- The technique obtained better results than the context-based and content-based methods when the metric H-Hit rate was used.