

# *Imagination: Accurate Image Annotation Using Link-analysis Techniques\**

Ilaria Bartolini and Paolo Ciaccia

DEIS, University of Bologna, Italy  
{ibartolini,pciaccia}@deis.unibo.it

**Abstract.** The advent of digital photography calls for effective techniques for managing growing amounts of color images. Even if the content-based approach represents a completely automatic solution to image retrieval, it partially fails due to the semantic gap existing between the user subjective notion of similarity and the one of a feature-based retrieval system. A possible way to fill such gap is to (semi-)automatically assign meaningful terms to images, so as to enable a high-level, concept-based, retrieval. In this paper we explore the opportunities offered by graph-based link analysis techniques in the development of a semi-automatic image captioning system. The approach we propose is appealing since predicted terms for an image: 1) are in variable number, depending on the image content, 2) represent correlated terms, and 3) can also represent abstract concepts. We present preliminary results on our prototype system and discuss possible extensions.

## 1 Introduction

The use of digital cameras and camera phones has become widespread in recent years. As a main consequence, individuals make frequent use of home computers with the aim of building sizeable personal digital photo collections. Photo sharing through Internet has also become a common practice at present time. There are many examples of on-line photo-sharing communities, such as flickr<sup>1</sup>, photo.net<sup>2</sup>, and airliners.net<sup>3</sup>, just to name a few. These archives of personal photo collections are growing at phenomenal rate, so that the need for effective and efficient techniques for managing color images becomes more and more pressing. Even if content-based image retrieval (CBIR) systems represent a completely automatic solution to image retrieval [13], low level features, such as color and texture, are hardly able to properly characterize the actual image content. This is due to the semantic gap existing between the user subjective notion of similarity and the one according to which a low level feature-based retrieval system evaluates two images to be similar. Just to give an intuitive example, let us consider the

---

\* This work is partially supported by a Telecom Italia grant.

<sup>1</sup> flickr: <http://www.flickr.com/>.

<sup>2</sup> photo.net: <http://photo.net/>.

<sup>3</sup> airliners.net: <http://www.airliners.net/>.

two images depicted in Figure 1. Even if they could be considered “similar” by a CBIR system, they indeed represent different animals (namely a horse and a bison). On the other hand, if the user is just looking for some “mammals on the grass”, the two images could be considered similar even at a semantic level.



**Fig. 1.** Two images with associated terms.

Note that, although user feedback [12, 3] and context-based techniques [1] can indeed be helpful in improving the precision of results, i.e., the percentage of returned images which are actually relevant to the query, they stay well below the optimal 100% precision value, in particular when the user is looking for images matching some high-level concept (e.g., landscape).

A possible way to fill the semantic gap is to (semi-)automatically assign meaningful terms to images, so as to indeed allow a high-level, concept-based, retrieval. For instance, assuming that the two images in Figure 1 are annotated as shown in the figure, it would be possible to discriminate among them if, say, one is looking for horses and, at the same time, to consider both relevant if one is looking for mammals on grass.

Several techniques [11, 6, 8, 10, 9] have been proposed in recent years and the first image annotation prototypes are now available on Internet (e.g., ALIPR.com<sup>4</sup> and Behold<sup>5</sup>). We can group state-of-the-art solutions into two main classes, namely *semantic propagation* and *statistical inference*. In both cases, the problem to be solved remains the same: *Given a training set of annotated color images, discover affinities between low-level image features and terms that describe the image content, with the aim of predicting “good” terms to annotate a new image.* With propagation models [9], a supervised learning technique that compares image similarity at a low-level and then annotates images by propagating terms over the most similar images is adopted. Working with statistical inference models [10, 6, 8, 11], an unsupervised learning approach tries to capture correspondences between low-level features and terms by estimating their joint probability distribution. Both approaches improve the annotation process and the retrieval on large image databases. However, among the predicted terms for unlabelled images, still too many irrelevant ones are present.

<sup>4</sup> ALIPR.com: <http://www.alipr.com/>.

<sup>5</sup> Behold: <http://go.beholdsearch.com/searchvis.jsp>.

In this paper we explore the opportunities offered by graph-based link analysis techniques in the development of an effective semi-automatic image captioning system - namely *Imagination* (IMAGe (semI-)automatic anNotATION). In our approach each image is characterized as a set of *regions* from which low-level features are extracted. The training set is built by associating a variable number of terms to each image. In this way, not only terms related to a particular region of the image, but even abstract concepts associated to the whole image (e.g., “landscape” and “pasture”) are possible.

We turn the annotation problem into a set of *graph-based* problems. First, we try to discover *affinities* between terms and an unlabelled image, which is done using a *Random Walk with Restart* (RWR) algorithm on a graph that models current annotations as well as regions’ similarities. Then, since the RWR step might predict unrelated, or even contradictory, terms, we compute pairwise *term correlations*. Again, this relies on the analysis of links in a (second-order) graph. Finally, we combine the results of the two steps to derive a set of terms which are both *semantically correlated* each other and affine to the new image. This final step amounts to solve an instance of the Maximum Weight Clique Problem (MWCP) on a small graph. Doing this way, the number of terms we predict for each new image is variable, and dependent on the actual image content.

The paper is organized as follows: In Section 2 we define the problem. Section 3 shows how to compute affinities between an image and the terms of the training set and Section 4 analyzes correlations of terms. In Section 5 we show how we derive the most correlated affine terms and in Section 6 we provide some preliminary results obtained from *Imagination*. Section 7 concludes and discusses possible extensions.

## 2 Problem Definition

Before presenting our image annotation technique, we need to precisely define the problem. We are given a dataset of  $N$  manually annotated images that constitute the image *training set*  $\mathcal{I}$ . Each image  $I_i \in \mathcal{I}$  is characterized as a set of *regions*  $R_j$ , for each of which a  $D$ -dimensional feature vector is automatically extracted. For instance, features could represent the color and the texture of  $R_j$  [2]. Moreover, each image  $I_i \in \mathcal{I}$  is manually annotated with  $m_i$  *terms*  $\{T_{i_1}, \dots, T_{i_{m_i}}\}$ . Thus, each image  $I_i$  is represented as  $I_i = (\{R_{i_1}, \dots, R_{i_{n_i}}\}, \{T_{i_1}, \dots, T_{i_{m_i}}\})$ .

**Problem 1** *Given an unlabelled (or query) image  $I_q$ , with regions  $\{R_{q_1}, \dots, R_{q_{n_q}}\}$ , exploit the knowledge of images in  $\mathcal{I}$  to predict a “good” set of terms  $\{T_{q_1}, \dots, T_{q_{m_q}}\}$  able to effectively characterize the content of  $I_q$ .*

We turn the annotation problem, an instance of which is depicted in Figure 2, into a *graph-based* problem that is split into three main steps:

1. **Affinities of terms and query image:** Starting from the training images  $\mathcal{I}$ , we build a graph  $G_{MMG}$  and “navigate” it so as to establish possible *affinities* between the query image  $I_q$  and the terms associated to images in  $\mathcal{I}$ .

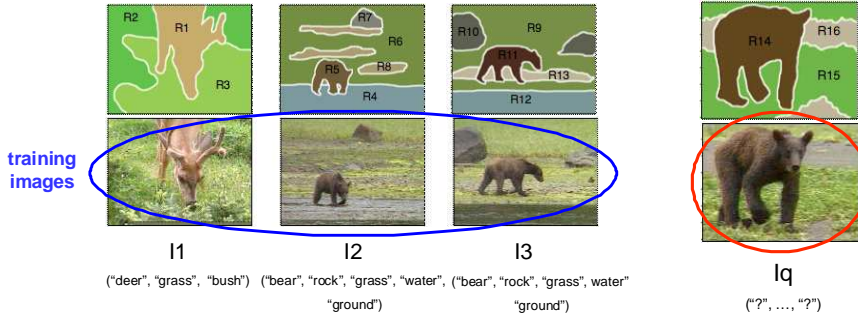


Fig. 2. Visual example of the image annotation problem.

2. **Correlation of terms:** Starting from  $G_{MMG}$ , we derive a *second-order* graph  $G_T^2$  from which to compute the *similarity* among terms.
3. **Correlated affine terms:** In this step we combine the results of the first two steps and derive the set of most *semantically correlated* terms to label the query image  $I_q$ .

### 3 Affinities of terms and query image

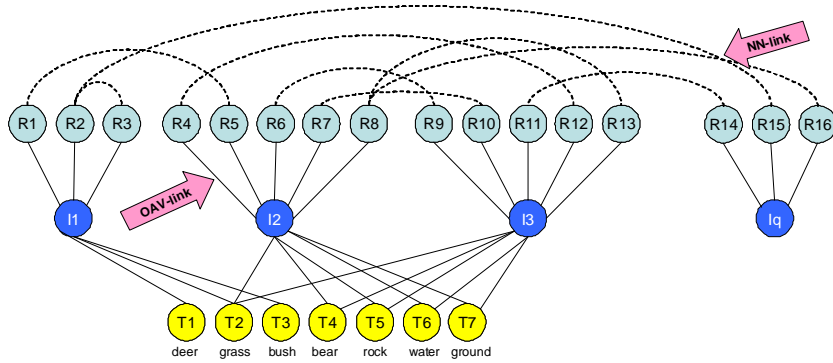
As for the implementation of the 1st step, we follow the Mixed Media Graph approach [11].

#### 3.1 Graph Construction

The Mixed Media Graph (MMG)  $G_{MMG} = (V, E)$  is a 3-level undirected graph, where each node represents an image (identifier), a region, or a term, in the training set. More precisely, if  $\mathcal{T}$  is the set of terms and  $\mathcal{R}$  is the set of regions, then  $V = \mathcal{I} \cup \mathcal{T} \cup \mathcal{R}$ . Edges in  $E$  are of two types. An *object-attribute-value* (OAV) edge connects an image node with either a region or a term node. Therefore for each image  $I_i \in \mathcal{I}$ , there are edges  $(I_i, R_j)$  for all regions  $R_j$  in  $I_i$ , and similarly for terms. *Nearest neighbor* (NN) edges connect a region to its  $k$  ( $k \geq 1$ ) nearest neighbors regions in  $\mathcal{R}$ , where the similarity between two regions is computed based on the regions' feature vectors. The graph  $G_{MMG}$  can be extended, so as to account for a new unlabelled image  $I_q$ , into the graph  $G_q = (V_q, E_q)$  by adding nodes for  $I_q$  and its regions, and NN edges for the regions of  $I_q$ . Figure 3 shows the  $G_q$  graph for the example in Figure 2.

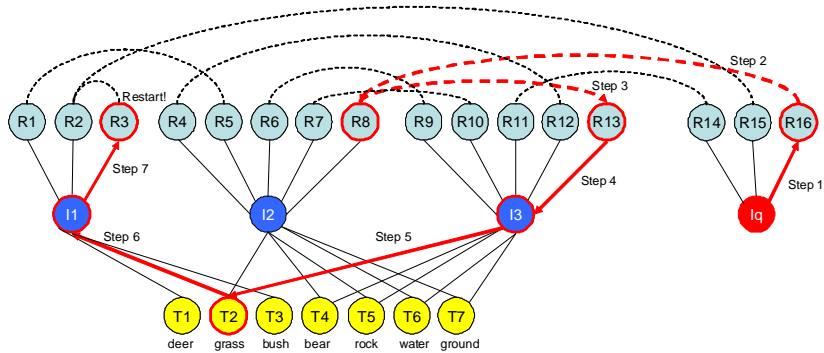
#### 3.2 Graph Navigation

As we turn the annotation problem into a graph problem, methods for determining how related a node  $X$  is to a "start" node  $S$  are needed to establish the affinity between the query image  $I_q$  and the terms in  $G_{MMG}$ . For this task we



**Fig. 3.** The  $G_q$  graph for the example depicted in Figure 2, assuming  $k = 1$

find appropriate to adopt the *random walk with restart* (RWR) technique [11]. The basic idea of RWR is to consider a random walker that starts from node  $S$  and at each step chooses to follow an edge, randomly chosen from the available ones. Further, at each step, with probability  $p$  the walker can go back to node  $S$  (i.e., it *restarts*). The *steady state probability* that the random walker will find itself at node  $X$ , denoted  $u_S(X)$ , can be interpreted as a measure of affinity between  $X$  and  $S$ . In our case it is  $S = I_q$  and relevant steady state probabilities are only those of term nodes (i.e.,  $X \in \mathcal{T}$ ). Intuitively, if  $u_{I_q}(T_j)$  is high, this is an evidence that  $T_j$  is a good candidate for annotating  $I_q$ . Details on how the steady state probabilities can be efficiently computed even for large graphs can be found in [14].



**Fig. 4.** RWR navigation example ( $k = 1$ ).

In Figure 4 an example of RWR navigation is shown. In particular, the  $G_q$  graph displayed in Figure 3 is navigated (with  $k = 1$ ) by starting from the query image  $I_q$  and crossing nodes  $R_{16}$ ,  $R_8$ ,  $R_{13}$ ,  $I_3$ ,  $T_2$ ,  $I_1$ ,  $R_3$ , respectively.

### 3.3 Limits of MMG

Even if MMG with RWR is usually able to find some relevant terms for annotating a query image, it suffers some limits. First of all, the predicted terms are those that have been crossed most frequently during the graph navigation. It can be argued that using only frequency to evaluate the relevance of each term for annotating a new image is rather imprecise. For instance, when using MMG, querying *Imagination* with an image representing a “horse” often returned as result the term “cow”. Indeed, one should bear in mind that the MMG + RWR method heavily relies on the NN edges involving the regions of  $I_q$ , thus on low-level similarities. If a region  $R_{q_i}$  of  $I_q$  is (highly) similar to a region  $R_j$  of an image  $I$ , which however has some terms unrelated to  $I_q$ , this might easily lead to have such terms highly scored by RWR (consider the example in Figure 1).

Another shortcoming of MMG regards the number of terms,  $PT$ , with the highest steady state probabilities that are to be used for annotation. There are two alternatives here. If one insists to take only the best  $PT$  terms, then each image will be annotated with a same number of terms, thus independently of the actual image content. Note that setting  $PT$  to a high value might easily lead to wrong annotations, whereas a low value might easily miss relevant terms. The same problem would occur should the predicted terms be all those whose steady state probability exceeds a given threshold value.

## 4 Analyzing Correlations of Terms

The approach we take to overcome MMG limitations is to perform a link analysis on a sub-graph of  $G_{MMG}$  so as to find highly-correlated terms. In turn, this is evidence that such terms are also semantically related, thus good candidates to annotate a new image.

### 4.1 Link Analysis

From the graph  $G_{MMG} = (V, E)$ , we derive the sub-graph  $G_T = (V_T, E_T)$ , where  $V_T = \mathcal{I} \cup \mathcal{T}$ , i.e.,  $G_T$  is derived from  $G_{MMG}$  by deleting region nodes. With the aim of estimating the similarity between couples of terms, we derive from  $G_T$  a *second-order* (bipartite) graph  $G_T^2 = (V_T^2, E_T^2)$ . A node in  $V_T^2$  is either a pair of images  $(I_i, I_j)$ ,  $I_i, I_j \in \mathcal{I}$ , or a pair of terms  $(T_r, T_s)$ ,  $T_r, T_s \in \mathcal{T}$ . An edge between nodes  $(I_i, I_j)$  and  $(T_r, T_s)$  is added to  $E_T^2$  iff the two edges  $(I_i, T_r)$  and  $(I_j, T_s)$  (equivalently,  $(I_i, T_s)$  and  $(I_j, T_r)$ ) are both in  $E_T$ . This is to say that each image  $I_i$  and  $I_j$  contains (at least) one of the two terms, and the two images, when taken together, contain both terms. Notice that when  $I_i = I_j$ , then terms  $T_r$  and  $T_s$  appear together in image  $I_i$ . An intuitive example of  $G_T$  and of the derived  $G_T^2$  graph are depicted in Figures 5 (a) and (b), respectively.

Given the second-order graph  $G_T^2$ , the problem of estimating the correlation of two terms transforms into the problem of assigning a score to nodes corresponding to pairs of terms. For this one can use any link-based algorithm, such

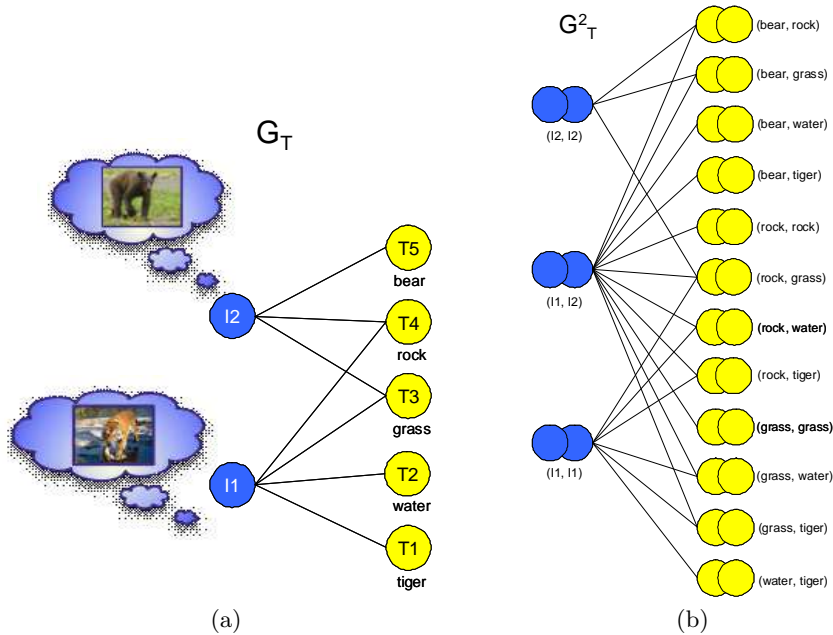


Fig. 5. Example of  $G_T$  graph (a) and the derived second-order graph  $G_T^2$  (b).

as those adopted for ranking Web pages [7]. We denote with  $corr(T_r, T_s)$  the (correlation) score computed by such an algorithm for the node in  $V_T^2$  corresponding to the pair of terms  $(T_r, T_s)$ . Note that this step can be performed off-line, since it is independent of the query image.<sup>6</sup>

## 5 Putting it All Together

In this last step we combine the results of the previous phases. As to the output of the MMG + RWR step, we always take the set of  $PT$  terms with the highest steady state probabilities,  $\mathcal{T}_{MMG} = \{T_1, \dots, T_{PT}\}$ . This will be possibly reduced considering terms correlations,  $corr(T_r, T_s)$ , so as to obtain a set of terms to annotate the query image  $I_q$  that: 1) are affine to  $I_q$ , and, at the same time, 2) are all tightly correlated each other.

We solve the problem by modelling it as an instance of the Maximum Weight Clique Problem (MWCP) [5]:

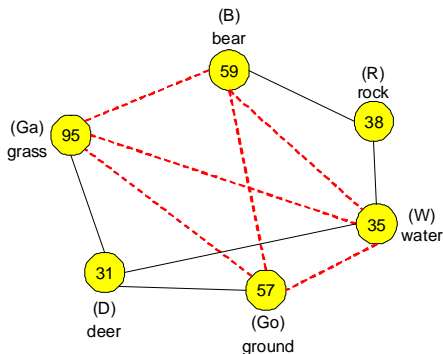
**Definition 1 (MWCP)** Let  $G = (V, E, w)$  be an undirected and weighted graph, where the  $j$ -th component of the weight vector  $w$  is the weight of the  $j$ -th node in  $V$ . A clique  $G' = (V', E')$  is a complete sub-graph of  $G$ , i.e.,  $V' \subseteq V$ , and

<sup>6</sup> We are currently studying how correlations can be efficiently updated in front of insertions in the training set.

there is an edge in  $E'$  between every pair of nodes in  $V'$ . The weight of clique  $G'$  is the sum of weights of the nodes in  $V'$ ,  $W(G') = \sum_{j \in V'} w_j$ . The Maximum Weight Clique Problem (MWCP) is to find the clique,  $G_{max}$ , with the maximum weight.

The correspondence with our problem is almost immediate. The set of nodes in the graph consists of the terms in  $\mathcal{T}_{MMG}$  (i.e.,  $V = \mathcal{T}_{MMG}$ ), and each node  $T_j$  is weighted by its steady state probability  $u_{I_q}(T_j)$  (i.e.,  $w_j = u_{I_q}(T_j)$ ). As to edges, we only add to  $E$  those between nodes (terms) whose correlation exceeds a given threshold value  $c$ , i.e.,  $(T_r, T_s) \in E$  iff  $corr(T_r, T_s) > c$ . Doing this way, solving the MWCP amounts to find the subset  $\mathcal{T}_{OPT}$  of optimal terms in  $\mathcal{T}_{MMG}$  such that: 1) all terms in  $\mathcal{T}_{OPT}$  are highly correlated, and 2) there is no other set of terms satisfying the same condition whose global affinity is higher.<sup>7</sup>

To give an example, Figure 6 shows a sample graph  $G$  in which  $PT = 6$ . Numbers within each node represent unnormalized steady state probabilities (normalizing would not change the net effect). Solving MWCP, the optimal terms (maximum weight clique) turn to be  $\mathcal{T}_{OPT} = \{grass, bear, ground, water\}$ , as it can be seen from Table 1 in which we report all cliques in  $G$  together with their weights. Notice that, without taking into account terms correlations, the affinity of *rock* is higher than that of *water*. However, *rock* is loosely correlated with almost all other terms in  $\mathcal{T}_{MMG}$ , thus it does not enter into the solution.



**Fig. 6.** Dashed edges define the clique with the maximum weight.

## 6 Preliminary Results

We have implemented all above-described algorithms within our prototype system *Imagination*. In particular, *Imagination* runs on top of the Windsurf system, which provides functionalities for image segmentation and support for  $k$ -NN

<sup>7</sup> Although the MWCP problem is NP-hard, the graphs we deal with are rather small (e.g., tens of nodes), so the computational overhead is negligible.



1	Ga (95)	B (59)	Go (57)	R (38)	W (35)	D (31)
2	Ga,B (154) Ga,Go (152) Ga,W (130) Ga,D (126)	B,Go (116) B,R (97) B,W (94)	Go,W (92) Go,D (88)	R,W (73)	W,D (66)	
3	Ga,B,Go (211) Ga,B,W (189) Ga,Go,W (187) Ga,Go,D (183) Ga,W,D (161)	B,Go,W (151) B,R,W (132)	Go,W,D (123)			
4	<b>Ga,B,Go,W (246)</b> Ga,Go,W,D (218)					

**Table 1.** All the (weighted) cliques in the graph  $G$  of Figure 6. For the sake of conciseness, terms are represented in abbreviated form. In particular, the correspondence is Ga for *grass*, B for *bear*, Go for *ground*, R for *rock*, W for *water*, and D for *deer*.

queries [2]. Each image is automatically segmented into a set of homogeneous regions which convey information about color and texture features. Each region corresponds to a cluster of pixels and is represented through a 37-dimensional feature vector. With respect to regions comparison (thus, to define the NN edges of  $G_{MMG}$ ) the Bhattacharyya metric is used [4]. The dataset we used was extracted from the IMSI collection.<sup>8</sup>

We trained *Imagination* by manually annotating about 50 images with one, two, or three terms. The query workload consists of other about 50 randomly chosen images. Each query image was assigned a set of terms by a set of volunteers so as to obtain a ground truth to evaluate the effectiveness of our system. We computed the annotation accuracy in terms of *precision* (i.e., the percentage of relevant terms predicted) and *recall* (i.e., the percentage of relevant predicted terms with respect to those assigned by our volunteers), averaged over the 50 query images. Table 2 summarizes the parameters used by *Imagination*, together with their default values which we used in our preliminary experiments.

parameter	default value
Average number of regions per image	4.4
Number of NN edges per region	$k = 5$
Maximum number of terms per image	$PT = 6$
RWR restart probability	$p = 0.8$
Correlation threshold	$c = 0.3$

**Table 2.** Parameters used by *Imagination* together with their default values.

<sup>8</sup> IMSI MasterPhotos 50,000: <http://www.imsisoft.com/>.

## 6.1 Effectiveness

Figure 7 shows the annotation accuracy in term of precision and recall. In particular, we compare our results with those obtained when using only MMG (i.e., without considering term correlations). As we can observe from the figure, *Imag-*

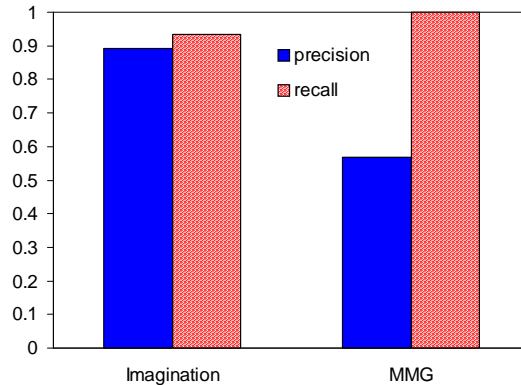


Fig. 7. Precision and recall levels of *Imagination* and *MMG* ( $PT = 6$ ,  $c = 0.3$ ).

*ination* is able to guarantee an improvement in average precision of about 32.66% with respect to *MMG*, even if maintaining an average recall level that is comparable to that of *MMG*. Although it happened that for some queries predicted terms also included irrelevant ones, the precision of *Imagination* was better than that of *MMG* alone on each query image, thus validating the effectiveness of correlation analysis.

In Figure 8 an example of *Imagination* in action is reported. In this case, the optimal terms that *Imagination* returns are *sheep* and *grass*, which are indeed the only appropriate ones among the  $PT = 6$  predicted by *MMG*. Finally, Table 3 gives some examples of annotation given by *Imagination*. For the first and third images, *Imagination* annotates them correctly, whereas for the middle image *Imagination* predicts the term “grass” instead of the term “tree”.

## 6.2 Influence of parameters

We conclude the experimental section by discussing the role of parameters shown in Table 2. With respect to  $k$  (number of NN links per region) and  $p$  (RWR restart probability), which influence the *MMG+RWR* step, we used the values suggested in [11]. In particular, in [11] the authors prove that the effectiveness of RWR is almost insensitive to the  $k$  value, as long as  $k \in [3, 10]$ . As for the RWR restart probability  $p$ , it is demonstrated in [11] that good results are obtained for  $p \in [0.8, 0.9]$ .

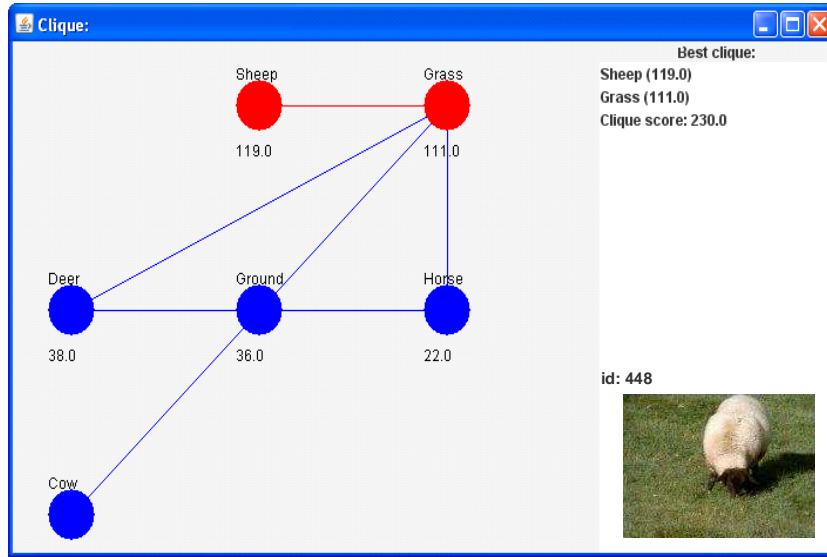


Fig. 8. The maximum weight clique for the image on the right.




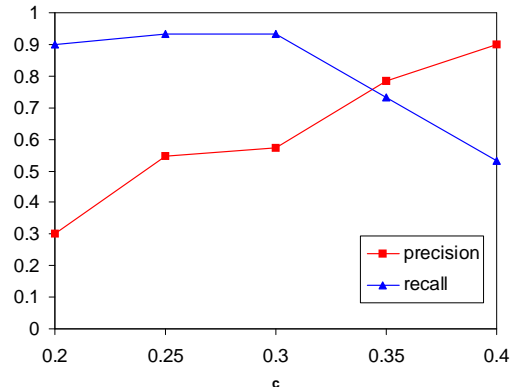
Query image			
Ground truth	<i>bear, grass, water, ground</i>	<i>deer, water, grass, rock</i>	<i>horse, sky</i>
Imagination	<i>grass, bear, ground, water</i>	<i>deer, tree, water, rock</i>	<i>horse, sky</i>

Table 3. Sample terms predicted by *Imagination*.

Concerning the  $PT$  parameter (number of terms with the highest affinities with the query), in our experiments we observed that values of  $PT > 6$  (for  $c = 0.3$ ) often resulted in lowering the precision. A possible explanation, consistent with the results we observed, is as follows. When  $PT$  grows, more terms are candidate to be used for annotation. If the correlation threshold  $c$  is not too high, the chance that the maximum weight clique is composed of many nodes (terms) with not-so-high affinity with the query grows as well. In turn, this suggests that  $PT$  and  $c$  are tightly related. To justify this claim, we considered a high  $PT$  value,  $PT = 10$ , and changed the correlation threshold. Figure 9 shows precision and recall curves we obtained. As it can be observed from the figure, the precision at  $PT = 10$ ,  $c = 0.3$  is quite lower than that observed in Figure 7, in which  $PT = 6$ . On the other hand, if  $c$  is increased, *Imagination* is still able to reach a remarkable 90% precision level. However, in this case the recall



**Fig. 9.** Precision and recall levels varying the correlation threshold  $c$  ( $PT = 10$ ).

drops to about 50%. Summarizing, if  $PT$  grows,  $c$  should grow as well to stay at a given precision level. On the other hand, not all  $(PT, c)$  combinations are equally good if one also considers recall.

## 7 Conclusions

In this paper we have presented *Imagination*, an effective approach for semi-automatic image annotation based on link-analysis techniques. Our approach is able to predict terms that are highly correlated each other, which improves the accuracy of the annotation. At present, we are working on a more accurate, large-scale, evaluation. Further, we plan to extend our term analysis by means of ontologies, so as to exploit, besides term correlations, also their semantic relationships (e.g., “the horse is a mammal”). This will likely lead to further improve the precision of our approach.

## References

1. I. Bartolini. Context-Based Image Similarity Queries. *Adaptive Multimedia Retrieval: User, Context, and Feedback, AMR 2005, Revised Selected Papers (Springer Lecture Notes in Computer Science)*, 3877:222–235, 2006.
2. I. Bartolini, P. Ciaccia, and M. Patella. A Sound Algorithm for Region-Based Image Retrieval Using an Index. In *Proceedings of the 4th International Workshop on Query Processing and Multimedia Issue in Distributed Systems (QPMIDS 2000)*, pages 930–934, Greenwich, London, UK, Sept. 2000.
3. I. Bartolini, P. Ciaccia, and F. Waas. FeedbackBypass: A New Approach to Interactive Similarity Query Processing. In *Proceedings of the 27th International Conference on Very Large Data Bases (VLDB 2001)*, pages 201–210, Rome, Italy, Sept. 2001.

4. M. Basseville. Distance Measures for Signal Processing and Pattern Recognition. *European Journal of Signal Processing*, 18(4):349–369, 1989.
5. I. Bomze, M. Budinich, P. Pardalos, and M. Pelillo. *The Maximum Clique Problem*, volume 4. Kluwer Academic Publishers, Boston, MA, 1999.
6. P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth. Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. In *Computer Vision - ECCV 2002, 7th European Conference on Computer Vision*, pages 97–1123, Copenhagen, Denmark, May 2002.
7. D. Fogaras and B. Rácz. Scaling Link-based Similarity Search. In *Proceedings of the 14th International Conference on World Wide Web (WWW 2005)*, pages 641–650, Chiba, Japan, May 2005.
8. J. Jeon, V. Lavrenko, and R. Manmatha. Automatic Image Annotation and Retrieval Using Cross-media Relevance Models. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 119–126, Toronto, Canada, Aug. 2003.
9. O. Maron and A. L. Ratan. Multiple-instance Learning for Natural Scene Classification. In *Proceedings of the 15th International Conference on Machine Learning (ICML 1998)*, pages 341–349, San Francisco, CA, USA, July 1998.
10. Y. Mori, H. Takahashi, and R. Oka. Image-to-word Transformation Based on Dividing and Vector Quantizing Images with Words. In *Proceedings of the 1st International Workshop on Multimedia Intelligent Storage and Retrieval Management (MISRM 1999)*, 1999.
11. J.-Y. Pan, H.-J. Yang, C. Faloutsos, and P. Duygulu. Automatic Multimedia Cross-modal Correlation Discovery. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 653–658, Seattle, USA, Aug. 2004.
12. Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.
13. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.
14. H. Tong, C. Faloutsos, and J.-Y. Pan. Fast Random Walk with Restart and Its Applications. In *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 2006)*, pages 613–622, Hong Kong, China, Dec. 2006.