

Automatic image annotation with relevance feedback and latent semantic analysis

Donn Morrison, Stéphane Marchand-Maillet, and Eric Bruno

Centre Universitaire Informatique
Université de Genève, Genève, Switzerland
{donn.morrison, marchand, eric.bruno}@cui.unige.ch
<http://viper.unige.ch/>

Abstract. The goal of this paper is to study the image-concept relationship as it pertains to image annotation. We demonstrate how automatic annotation of images can be implemented on partially annotated databases by learning image-concept relationships from positive examples via inter-query learning. Latent semantic analysis (LSA), a method originally designed for text retrieval, is applied to an image/session matrix where relevance feedback examples are collected from a large number of artificial queries (sessions). Singular value decomposition (SVD) is exploited during LSA to propagate image annotations using only relevance feedback information. We will show how SVD can be used to filter a noisy image/session matrix and reconstruct missing values.

1 Introduction

Content-based image retrieval is burdened by a dichotomy between user-desired high-level concepts and the low-level descriptions that retrieval systems are capable of indexing. During an interactive query, *relevance feedback (RF)* attempts to narrow this gap by allowing a user to refine the query by selecting positive and negative examples. After each iteration, a model is updated to reflect the modified query and a new result set is returned to the user. Despite the advantages, however, a query session using relevance feedback is cumbersome and the user must possess a priori knowledge about the low-level structure of the desired image: the colours, shapes, and textures it contains.

1.1 Redefining the query

Keyword-based queries, popularised by internet search engines and text-based document retrieval, are more desirable from a user's perspective. For image retrieval, this changes the field substantially. Where users were previously confined to retrieval by low-level feature similarity (*query by example (QBE)*), they would now be able to specify their queries more naturally using high-level semantics. Where relevance feedback once dominated the narrowing of a query, the augmentation of query terms now stands to take its place.

However, this presents an obvious problem: keyword-based image retrieval can only be realised on accurately and completely annotated databases. Furthermore, indexed

keywords must describe the images and/or image regions in many ways. Polysemy and synonymy present problems as similar concepts can be described using different terms and different concepts can be described with similar terms. But ultimately, the prohibitive barrier is on the creation of annotated databases themselves. It is infeasible to manually annotate images given the scope and size of image collections nowadays.

This realisation changes the focus of content-based image retrieval from a query problem to that of a semantic extraction problem. This is precisely the goal of this paper: to experiment with a method of propagating image labels onto unannotated images via instances of relevance feedback made during user queries. In this way, an image collection can be incrementally annotated beginning with a small (but sufficiently uniform) subset of labelled images.

1.2 Traditional image annotation

Despite suffering problems relating to subjectivity, the manual annotation of image collections, that is having users hand-label the images, is the most accurate method for creating a database on which reliable keyword-based queries can be made. However, most image databases are too large and too complex to facilitate such a task. Some studies successfully harness social aspects of image labelling, such as the ESP Game [1], where users compete to label as many images as possible in a set amount of time, but for specialised databases with sensitive applications in medicine or security, this approach is impractical and manual annotation is beyond reach.

Another popular approach in the literature is the use of semi-automatic annotation methods [2]. Here users play a limited role, effectively managing, correcting, or judging annotations resulting from automatic algorithms. Automatic algorithms must make use of the features contained in the images, as mentioned above: colour, texture, shape, etc. Similarities between images mean that keywords can be shared, but as these approaches largely lack any semantic understanding of the actual content beyond the low-level features, users must provide this high-level information.

A need arises to extract the semantic knowledge users possess about images they view or interact with and apply it at the collection level so that relationships between images are formed. This has been acknowledged in the literature for some time [3–9], but we see it as an underdeveloped area that needs further activity in order to mature.

The rest of this paper is organised as follows. In Section 2 we review existing automatic annotation techniques in the literature. These include generative models, classification-, and cluster-based approaches. Next, in Section 3 we describe our experiments and the structure of the database used. Section 4 examines our initial results and finally Section 5 concludes with some avenues for improvement and future research.

2 Automatic image annotation

2.1 Latent-space and generative models

Automatic image annotation can be approached with a variety of machine learning methods, from supervised classification to probabilistic to clustering. It is common to

borrow latent and generative models from text retrieval such as latent semantic analysis (LSA) [10] and its probabilistic cousin, PLSA [11]. These two latent-space models are compared in [12]. The authors pose the question of whether annotation by propagation is better than annotation by inference. LSA is shown to outperform PLSA. However, they explain that some of the reasons for this may be that LSA is better at annotating images from uniformly annotated databases.

In a later paper, the authors introduce an improved probabilistic latent model, called *PLSA-words*, which models a set of documents using dual cooperative PLSA models. The intention is to increase the relevance of the captions in the latent space. The process is divided into two stages: parameter learning, where the latent models are trained, and annotation inference, where annotations are projected onto unseen images using the generated models. In the first stage, the first PLSA model is trained on a set of captions and a new latent model is trained on the visual features of the corresponding images. In the second stage, the standard PLSA technique projects a latent variable onto the new image, and annotations of an aspect are assigned if the probability exceeds a threshold [13].

Extensions of PLSA have been described, for example *latent Dirichlet allocation (LDA)*, introduced in [14], which models documents as probabilistic mixtures of topics which are comprised of sets of words [15]. This model was applied to image annotation in a slightly modified version called *correspondence latent Dirichlet allocation (Corr-LDA)* [16]. In this study, the authors compare the algorithm with two standard hierarchical probabilistic mixture models. Three tasks are identified: modelling the joint distribution of the image and its caption, determining the conditional distribution of words in an image, and determining the conditional distribution of words in an image region. The Corr-LDA model first generates region descriptions from the image using an LDA model. Then corresponding caption words and image regions are selected, based on how the image region was selected.

2.2 Supervised learning approaches

In [17] a statistical model using multi-resolution hidden Markov models (MHMMs) is presented for automatic annotation of images. The authors use global features consisting of colour and texture features. Wavelets are used for representing the texture features and the LUV colour space is used for the colour features. Three benefits are given to the usage of the statistical model: simple retraining when new images are added; individual models can be retrained easily as they represent a single concept; likelihood for similarity based on global features. The images are first subsampled into multiple resolutions, each of which are used for different levels of feature extraction. For each concept, a statistical model is built based on the training images. Next, test images are introduced and the system searches a dictionary for the most keywords which have the greatest likelihood of describing the image. These terms are then associated with the image.

In [18], a method (confidence-based dynamic ensemble, or CDE) is proposed which utilises one-class SVMs at the base level to extract a confidence factor on the prediction of a two-class SVM on the classification of an image. This confidence factor is also used to dynamically alter the characteristics of the member classifiers. At the base level,

binary SVMs are used to make a distinction on one semantic label. The confidence factor is found by using the one-class SVM, and the prediction is mapped to a posterior probability. At the multi-class level, the confidence factors from the base classifiers are combined to predict a single class. At the bag-level, results from many classifiers are combined using different partitions of the training data in a standard bagging technique. The dynamic ensemble (DE) part of the system takes predictions with low confidence and tries to retrain specialist classifiers using reduced-class training data depending on which classes are most certain. The authors recognise the added complexity, but note that most automatic annotation is done off-line, and that the DE is only applied to low confidence instances, and that the DE only considers the relevant classes.

In [19], salient regions were used as the features for an automatic annotation framework. A statistical model was also used, dubbed the cross-media relevance model, to uncover correlation between visual features and keywords. The model finds the joint probability distribution between image regions and the keywords of the annotated training data. Salient regions are used in place of global image descriptors, and the Scale Invariant Feature Transform (SIFT) feature is also used. A hybrid approach joins the two methods and annotation accuracy is found to be better than LSI ($k=4$) and vector-space methods.

2.3 Unsupervised learning approaches

The combination of short-term and long-term learning is explored in [20] where K-means clustering is used to update a series of one-class support vector machines, each of which corresponds to an instance of relevance feedback during a query. Following the initial query, the nearest neighbour images (in features space and cluster proximity) are returned. After relevance feedback, a one-class SVM is trained on the relevant images. This new SVM becomes the new query. The process is repeated until the session terminates.

User intervention is relied upon in a clustering approach in [2] where the authors employ similarity methods with semi-supervised fuzzy clustering for semi-automatic image annotation. The premise is that user intervention may be required to help accurately define clusters by specifying pairwise relationships (must-link, cannot-link) between instances which are at the borders of automatically defined clusters. This method relies on the assumption that users can easily specify the pairwise constraints on images and that after specifying such constraints the clusters in question are better separated. The centroids of strong clusters are found by isolating and purging spurious small clusters. The method is compared with two other clustering methods: CA and K-means. The authors conclude that as long as the user intervention is minimal and simple, important information can be gleaned from this step.

The work in [21] and [22] builds and improves on a semantic ensemble which uses an ontology such as WordNet for determining a hierarchical representation of the semantic information for automatic annotation. One process specialises in propagating generic semantic information while the other process works to introduce specific knowledge about the document.

[23] describe a system for pruning irrelevant terms from automatically annotated images using word similarity in WordNet. Images are segmented into blobs using the

normalised cuts algorithm [24] which are associated with keywords based on the distance between the image and the centroids of the blobs. The similarity of the keywords is measured using a hybrid approach comprising three WordNet noun distance measures. Scores are combined using Dempster-Shafer Evidence Combination. Words which are viewed as outliers to others in the annotation are pruned.

2.4 Exploiting user interaction

A different modality can be introduced to harnesses the knowledge generated by users or groups of users interacting with an image database, whether it be browsing or performing longer queries (including but not limited to relevance feedback). By observing these interactions and the associations made between relevant and non-relevant images during a query, semantic themes can start to become apparent. These themes need not be named entities such as words describing objects or concepts, but can simply be relationships between images indicating some level of semantic similarity.

This type of learning is dubbed *inter-query learning* due to the feature space spanning multiple (or all) query sessions. The converse is the traditional *intra-query learning*: the utilisation of relevance feedback examples during the current query only (after the session has ended the weights are discarded). Inter-query learning takes an approach similar to collaborative filtering; interaction (in the form of queries with relevance feedback) is required to increase density in the feature space. It is in this way that a collection can be incrementally annotated. The more interaction and querying, the more accurate the annotations become.

The Viper group produced one of the first studies which looked at inter-query learning [3]. The authors analysed the logs of queries using the *GIFT (GNU Image Finding Tool)* demonstration system over a long period of time and used this information to update the *tf-idf* feature weightings. Images were paired based on two rules: images sharing similar features and also marked relevant have a high weight while images sharing similar features but marked both relevant and irrelevant should have a low weight (indicating a semantic disagreement). Two factors were introduced to manage the relevance feedback information. The first being a measure of the difference between the positively and negatively rated marks for each feature and the second re-weighting the positively and negatively marked features differently such that the ratio is scaled non-linearly.

Later, in [4], the authors focus more formally on annotation. A general framework is described which annotates the images in a collection using relevance feedback instances. As a user browses an image database using a CBIR system, providing relevance feedback as the query progresses, the system automatically annotates images using the relationships described by the user.

Taking a direction toward the fusion of the two modalities, [5] combine inter-query learning with traditional low-level image features to build semantic similarities between images for use in later retrieval sessions. The similarity model between the request and target images are refined during a standard relevance feedback process for the current session. This refinement and fusion is facilitated by a *barycenter*. The paper also discusses the problems with asymmetrical learning, where the irrelevant images are marked irrelevant by the user for a variety of reasons, whereas relevant images are

marked relevant only because they relate semantically to the query. Therefore, the authors reduce the relevance of irrelevant images during the fusion of feedback stages. Similarly, in [6], a statistical correlation model is built to create semantic relationships between images based on the co-occurrence frequency that images are rated relevant to a query. These relationships are fused with low-level features (256 colour histogram, colour moments, 64 colour coherence, Tamura coarseness and directionality) to propagate the annotations onto unseen images.

In [7], inter-query learning is used to improve the accuracy of a retrieval system and latent semantic indexing (LSI) is used in a way such that the interactions are the documents and the images correspond to the term vocabulary of the system. The authors perform a validation experiment on image databases consisting of both texture and segmentation data from the MIT and UCI repositories. Random queries were created and two sessions of relevance feedback were conducted to generate the historical information to be processed by LSI. From experiments on different levels of data, they conclude that LSI is robust to a lack of data quality but is highly dependent on the sparsity of interaction data.

This method of exchanging RF instances and images for the documents and term vocabulary was also used in a later study where the authors use long-term learning in the PicSOM retrieval system [8]. PicSOM is based on multiple parallel tree-structured *self-organising maps* (SOMs) and uses MPEG7 content descriptors for features. The authors claim that by the use of SOMs the system automatically picks the most relevant features. They note that the relevance feedback information provided by the users is similar to hidden annotations. Using Corel images with a ground truth set of 6 classes, MPEG7 features scalable colour, dominant colour, colour structure, colour layout, edge histogram, homogeneous texture, and region shape, the authors reported a significant increase in performance.

In [9] long term user interaction with a relevance feedback system is used to make better semantic judgements on unlabelled images for the purpose of image annotation. Relationships between images which are created during relevance feedback can denote similar or dissimilar concepts. The authors also try to improve the learning of semantic features by “a moving of the feature vectors” around a group of concept points, without specifically computing the concept points. The idea is to cluster the vectors around the concept centres.

In this section we have reviewed the many approaches to automatic image annotation. The traditional approaches, such as those which attempt to label images using generative probabilistic models or that rely on document similarity to label images clearly have not succeeded in narrowing the semantic gap. Until the time when these automatic methods are able to recognise and infer semantic concepts, the knowledge required to fill this gap must come from the user.

3 Experimental framework

For our experiments, a subset of the Corel database was chosen due to its ubiquitous use in the image retrieval community. Ten categories were arbitrarily chosen to reduce the complexity of the data set for the initial experiment: *architecture*, *beach*, *bird*, *cloud*,

flower, insect, leopard, lizard, mushroom, sunrise/sunset. For each category there are 20 images, bringing the total number of images in the collection to 200.

We group categories into concepts to build an artificial data set for inter-query learning. Concepts are used to describe semantically similar categories. For example, the concept “landscape” can describe images depicting beaches and sunrise/sunset and “animals” can describe images depicting birds and lizards. We use concepts to model more general themes from relevance feedback examples. Table 1 shows the concept-category relationships used in the experiments.

Table 1. Concept-category relationships

Concept	Category
landscape	beach
	sunrise/sunset
sky	beach
	cloud
	sunrise/sunset
animals	birds
	insect
	leopard
	lizard
plants	flower
	mushroom
man-made	architecture

A pool of 100 artificial relevance feedback sessions was created by setting all images under a concept as relevant to that query. In essence, the matrix created is a groundtruth matrix where all concepts are related to the categories through artificial sessions. This data simulates query sessions where users would have a concept image in mind (for example, images depicting animals), and would construct the query by selecting a number of positive and negative examples.

Thus, a matrix A is created where the columns represent the RF sessions and the rows represent the images [7, 8]. Figure 1 shows this matrix. It is highly redundant because of the large number of RF sessions generated and the low number of concepts.

In this experiment, only the positive examples ($A(i, j) == +1$) are used in order to simplify the propagation stage (we will ignore irrelevant concepts for the moment). Next, to simulate missing relevance feedback data, the values of A are randomly dropped (set to 0) to form a new noisy matrix, A_n . Singular value decomposition (SVD) is applied to this matrix to yield:

$$A_n = U \Sigma V^T. \tag{1}$$

The diagonal matrix Σ contains the singular values. We retain only $k = 5$ concepts from Σ to filter out unimportant concepts and reconstruct A as A_r .

$$A_r = U \Sigma' V^T \tag{2}$$

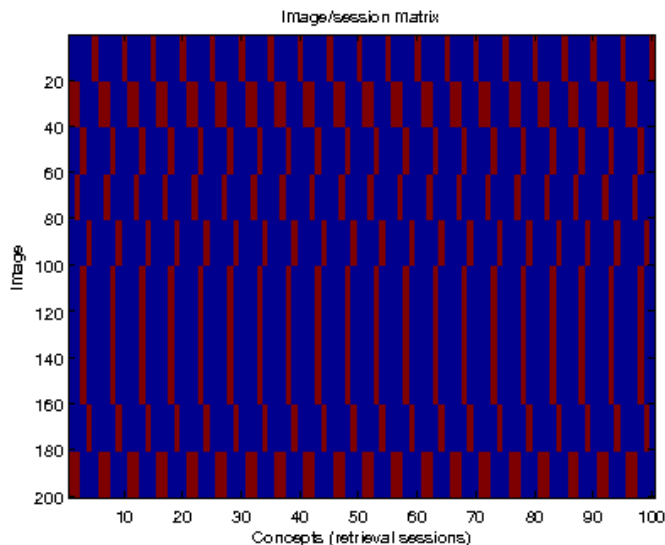


Fig. 1. Complete artificially generated relevance feedback matrix A

With A_n reconstructed as A_r , we now apply a thresholding measure to allow diffusion of relevance feedback examples into cells with missing data. As a result of SVD, cells previously zero will now be non-zero. These values are normalised into the same space as A_n and then thresholded at 0.7, giving:

$$A_r(i, j) = \begin{cases} 1 & \text{where } A_n(i, j) > 0.7 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The result is the reconstructed matrix with values that should minimise the difference from A_n :

$$D_{nr} = \sum_{i,j} |A_n(i, j) - A_r(i, j)| \quad (4)$$

Based on this reconstructed data, we can propagate image labels from similar concepts. Note that this process does not differentiate between terms at the keyword level. Different keywords in the annotations can describe different parts of the image, some of which may not be present in images under a similar concept. In order to alleviate this, an extension using regions must be employed.

We want to annotate the unlabelled images in the database to allow for keyword-based queries. Image similarities are specified by the user by way of relevance feedback. This alone could be sufficient for labelling, but normally the feature space is very sparse, and some diffusion is needed to propagate image labels throughout the collection.

We start by simulating a uniform partially annotated database. We discard all annotations except for 25% in each concept. Thus, for the concept “sky,” we have 15 anno-

tated images out of 60. For each concept in the reconstructed matrix A_r , we determine the most common keywords by image.

Consider the reconstructed matrix A_r . Recall that the rows correspond to images, I_i and the columns correspond to concepts, C_j . Using A_r , the concepts C_j for image I_i are summed and a vote is cast determining to which concept I_i belongs. There is a case when A_n will be too sparse that reconstruction will yield no concepts for an image I_i . In this case the image can only be labelled when further query sessions have been added to A , or in the case of our artificial situation, when the number of cells deleted from A is reduced.

4 Results and discussion

4.1 SVD and missing value recovery

Figure 2 shows the effect that incremental random deletions have on the distance from A_n and A_r . Starting with a fully annotated set and incrementally removing data has a non-linear effect on the distance measure. Note that as the deletions approach the maximum, the distance begins to decrease slightly due to the instability of the SVD on a very sparse matrix.

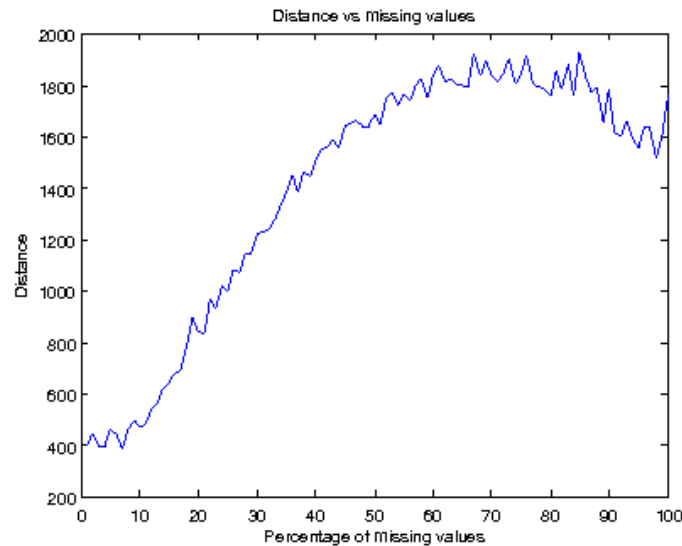


Fig. 2. Distance between A_n and A_r versus percentage of deleted cells

Figure 3 shows A_n , which results from the random cell deletion on A at 80%. In our experiments, the percentage of cell deletion was varied to see how SVD handles incremental missing values. Finally, Figures 4 and 5 show the reconstructed image/session

matrix A_r , before and after thresholding, respectively. It can be seen that one category of images (Figure 5, images 61-80) suffers more corruption after reconstruction than the rest, with almost no associated concepts. This is because the category in question, “cloud”, belongs to only one concept (“sky”), while the other members of that concept belong to two concepts (“landscape” and “sky”). This causes the “cloud” category to have less influence, and thus, SVD tries to map the “cloud” concept onto these images.

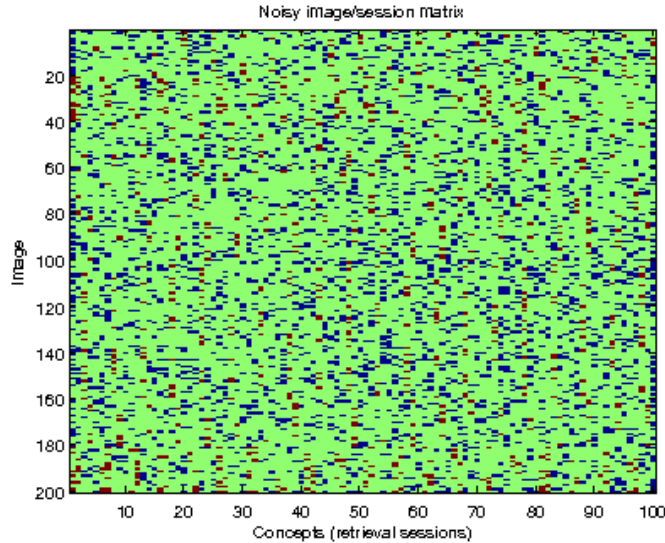


Fig. 3. Noisy image/session matrix A_n with 80% percent cell deletion

4.2 Annotation propagation

Figure 6 shows some example images which have been labelled using our propagation method. It is apparent that the method blankets images under each concept with words that have the highest distribution. Although this allows some relevant keywords to be propagated onto unlabelled images, the final annotations leave much room for improvement.

An ontology such as WordNet could be used to find parent concepts shared between labels to apply a more general annotation, as was done in [22]. For example, instead of the top-middle image in Figure 6 being labelled “bird tree insect lizard,” when it is clearly a leopard, the common parent between “lizard,” “bird,” and “insect”, which is “animal,” could be used. This would more naturally model the concepts we have chosen.

This method has a quantization effect on the propagation of image labels due to the use of broad concepts. In the example of architecture in Figure 6, the labels more closely model the image than some of the other examples because the “man-made” concept contains only the “architecture” category. For the “animals” concept, the images

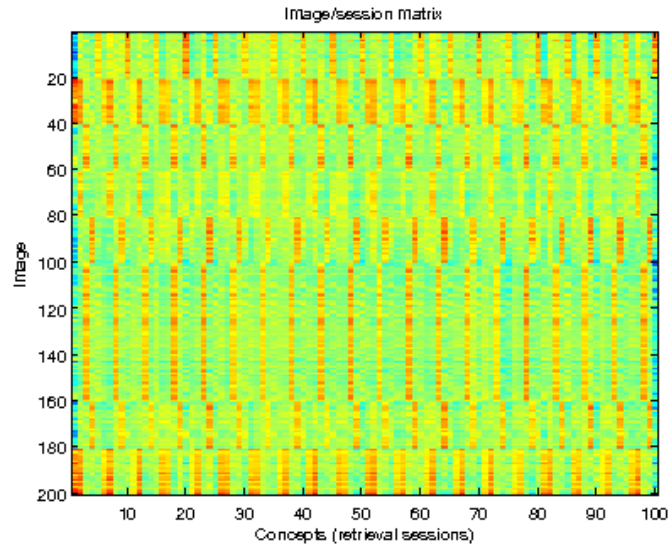


Fig. 4. Reconstructed image/session matrix A_r .

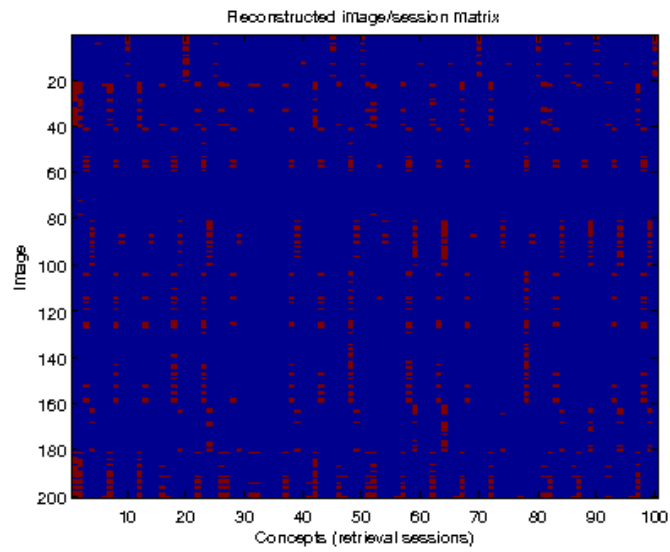


Fig. 5. Thresholded image/session matrix A_r , showing reconstruction by SVD

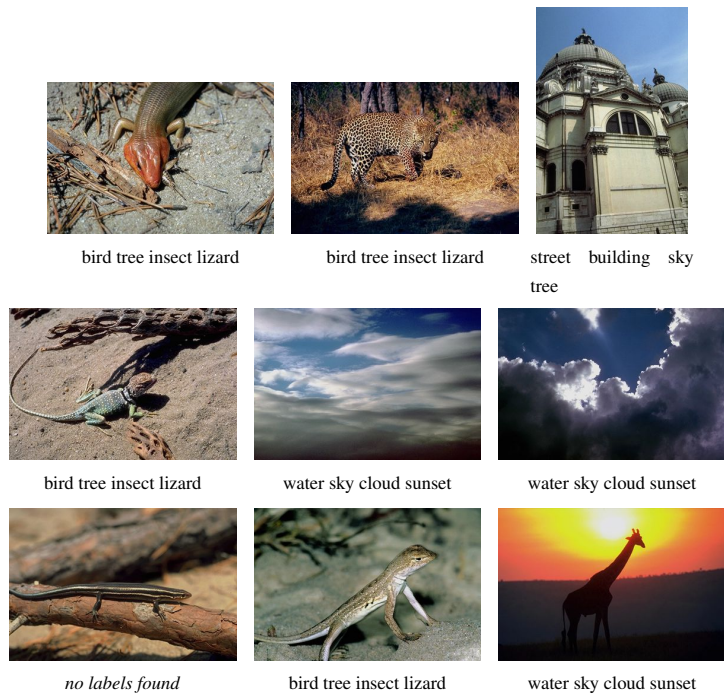


Fig. 6. Automatically labelled images with cell deletions at 80% and SVD threshold of 0.7

are labelled with a high proportion of “lizards,” most likely due to the distribution of keywords in the annotations.

In the case of the bottom left-most image in Figure 6, no labels were propagated because the reconstructed matrix was too sparse to determine a concept. As mentioned above, this can be alleviated by adding more relevance feedback examples to the original matrix A which sufficiently cover the concept and involve the image.

5 Conclusion

In this paper we reviewed several state of the art image annotation methods in the literature. The main goal was to determine whether the image-concept relationship, derived from inter-query learning, could be used to annotate images in a partially labelled corpus. We demonstrated how the SVD algorithm responds to incremental cell deletions in the image/session matrix and showed how it can be used with thresholding to reconstruct a matrix with missing values. This reconstructed matrix can be used to annotate unlabelled images in the collection. A set of artificial relevance feedback examples were used as a source of semantic knowledge, alleviating the reliance on low-level features. We have shown that labels can be propagated based on these semantic relationships.

We would like to address some of the drawbacks introduced with this method of label propagation. The first is that the categories are blurred due to the broad (and narrow) scope of the chosen concepts. This can be improved by diversifying the initial image-session matrix A so that there are more concepts or so that there are more than one concept per session. Second, because artificial image-concept data was used, the initial results may be inflated. We hope to acquire a large amount of natural data through long-term use of a retrieval system. This would enable the extraction of more diverse concepts.

We plan to extend the model to utilise low-level image features such as colour, texture, and shape to further improve incremental annotation accuracy. Because keywords are propagated blindly with respect to actual image content, a more accurate approach would specify annotations that correspond to image regions so that irrelevant labels are not assigned. The use of WordNet will also be investigated to find common concepts within the vocabulary.

Acknowledgements

This study was funded by the Swiss NCCR Interactive Multi-modal Information Management (IM2).

References

1. von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM Press (2004) 319–326

2. Grira, N., Crucianu, M., Boujemaa, N.: Active semi-supervised fuzzy clustering for image database categorization. In: *MIR '05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, New York, NY, USA, ACM Press (2005) 9–16
3. Mueller, H., Mueller, W., Squire, D.M., Marchand-Maillet, S., Pun, T.: Long-term learning from user behavior in content-based image retrieval. Technical report, Université de Genève (2000)
4. Wenyin, L., Dumais, S., Sun, Y., Zhang, H., Czerwinski, M., Field, B.: Semi-automatic image annotation. (2001)
5. Fournier, J., Cord, M.: Long-term similarity learning in content-based image retrieval (2002)
6. Li, M., Chen, Z., Zhang, H.: Statistical correlation analysis in image retrieval (2002)
7. Heisterkamp, D.: Building a latent-semantic index of an image database from patterns of relevance feedback. (2002)
8. Koskela, M., Laaksonen, J.: Using long-term learning to improve efficiency of content-based image retrieval (2003)
9. Cord, M., Gosselin, P.H.: Image retrieval using long-term semantic learning. In: *IEEE International Conference on Image Processing*. (2006)
10. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by latent semantic analysis. *Journal of the American Society of Information Science* **41**(6) (1990) 391–407
11. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. *IEEE Trans. on PAMI* **25** (2000)
12. Monay, F., Gatica-Perez, D.: On image auto-annotation with latent space models. In: *Proc. ACM Int. Conf. on Multimedia (ACM MM)*, Berkeley, 2003. (2003)
13. Monay, F., Gatica-Perez, D.: PLSA-based image auto-annotation: constraining the latent space. In: *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, New York, NY, USA, ACM Press (2004) 348–351
14. Blei, D.M., Ng, A.Y., Jordan, M.I.
15. Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D., Jordan, M.: Matching words and pictures. *Machine Learning Research* **3** (2003) 1107–1135
16. Blei, D.M., Jordan, M.I.: Modeling annotated data. In: *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, New York, NY, USA, ACM Press (2003) 127–134
17. Wang, J.Z., Li, J.: Learning-based linguistic indexing of pictures with 2-d mhmm. In: *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, New York, NY, USA, ACM Press (2002) 436–445
18. Goh, K.S., Chang, E.Y., Li, B.: Using one-class and two-class svms for multiclass image annotation. *IEEE Transactions on Knowledge and Data Engineering* **17**(10) (2005) 1333–1346
19. Tang, J., Hare, J.S., Lewis, P.H.: Image auto-annotation using a statistical model with salient regions. In: *In Proceedings of IEEE International Conference on Multimedia & Expo (ICME)*, Hilton Toronto, Toronto, Ontario, Canada. (2006)
20. Gondra, I., Heisterkamp, D.R.: Incremental semantic clustering summarizing inter-query learning in content-based image retrieval via. In: *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC04)*. (2004)
21. Kosinov, S., Marchand-Maillet, S.: Multimedia autoannotation via hierarchical semantic ensembles. In: *Proceedings of the Int. Workshop on Learning for Adaptable Visual Systems (LAVS 2004)*, Cambridge, UK (2004)
22. Kosinov, S., Marchand-Maillet, S.: Hierarchical ensemble learning for multimedia categorization and autoannotation. In: *Proceedings of the 2004 IEEE Signal Processing Society Workshop (MLSP 2004)*, São Luís, Brazil (2004) 645–654

23. Jin, Y., Khan, L., Wang, L., Awad, M.: Image annotations by combining multiple evidence & wordnet. In: MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia, New York, NY, USA, ACM Press (2005) 706–715
24. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(8) (2000) 888–905