IPL at ImageCLEF 2014: Scalable Concept Image Annotation

Spyridon Stathopoulos and Theodore Kalamboukis

Information Processing Laboratory,
Department of Informatics,
Athens University of Economics and Business,
76 Patission Str, 104.34, Athens, Greece
spstathop@aueb.com,tzk@aueb.gr
http://ipl.cs.aueb.gr/index_eng.html

Abstract. In this article we report on the experiments conducted by the IPL team within the context of the ImageCLEF 2014 challenge on Scalable Concept Image Annotation. Our approach encompasses, a CBIR phase following with a concept extraction procedure. The content based retrieval utilizes Latent Semantic Analysis on a set of multiple Compact Composite Features to retrieve the most similar images and in the sequel a number of concepts are extracted from the associated textual information, based on their posterior probabilities.

Key words: LSA, Image Annotation, Data Fusion

1 Introduction

The continuous increase of digital images over the web has led to the need for an efficient method of indexing and retrieving content, based on the semantic information presented in an image. A most common approach to achieve this, is by assigning metadata in the form of keywords to the image. Most methods rely on the use of a manually labeled training set of images. However, manually annotating images is a costly process and has obvious scalability problems.

In the Scalable Concept Image Annotation task of the ImageCLEF 2014 [1, 2], the goal is to develop a fully automatic procedure, that is able to annotate an image with a predefined set of labels without the use of any hand labeled training data. Our baseline algorithm is divided into two phases. Visual retrieval step: Given a test image, a sample of the K most visually similar images is retrieved. Annotation step: From the texts associated to the K retrieved images a set of candidate keywords is selected as labels. The final assigned keywords for the test image are determined by a probability score based on the co-occurrence of labels in the selected sample.

Section 4 presents the results obtained from our experiments using the development and test set made available by the task. We compare our results with those from the task's baseline system.

2 Image Representation and Retrieval

In order to retrieve a sample of the visually nearest images, several low-level visual descriptors were extracted form each image. Those descriptors are then combined using LSA to provide a latent semantic vector representation for each image. The low level features (CEDD,FCTH) were selected based on our previous research [3] and experience from our participation in CLEF, medical Image retrieval task [4]. Furthermore, the following descriptors were selected for our final submitted runs since their combination gave the best results on the development set:

1. Color and Edge Directivity Descriptor (CEDD)[5].
2. Fuzzy Color and Texture Histogram (FCTH)[6].
3. Opponent-SIFT [7], provided by the organizers with a codebook of 1,000 features used to create the histograms for the images.

The descriptors CEDD and FCTH were locally extracted from a 3x3 grid, resulting in a vector size of 1,296 and 1,728 features respectively. Content based retrieval was based on applying LSA to the feature matrix X=[CEDD; FCTH; OppSIFT] using the Matlab's routine eigs for matrix XX^T with k=50. It has been shown [3] that, this is an effective and efficient approach that overcomes the defficiencies of using the singular value decomposition analysis.

3 Image Annotation

Each image in the training data is associated with a set of keywords with a score assigned to each keyword. The keywords were extracted from the text surrounding the image within its webpage and the scores were calculated using:

- The term frequency (TF).
- The document object model (DOM) attributes.
- The word distance to the image.

More information on the data is provided by the organizers in [2]. Additionally, we have removed the stopwords from the keywords sets.

A concept is a construct, an idea, of something formed by mentally combining all its characteristics. In our perception, a concept c, corresponding to a keyword, w, is defined by the set $C = \{w, s_1(w), ..., s_n(w)\}$ where $s_i(w)$ are the synonyms of w extracted from WordNet [8]. For a test image, g, labels were selected based on the posterior probabilities $p(c|g)$, (probability to select concept c, given a test image g), [9], defined by:

$$p(c|g) = \sum_{j=1}^{K} p(c|j)p(j|g) \tag{1}$$

Probabilities $p(c|g)$ are approximated from the K nearest neighbors visually retrieved images from the training set when the test image g is submitted as

query. To obtain an estimate of $p(j|g)$ we use the method proposed by Platt [10] to extract probabilistic outputs from SVM. The basic idea is that the retrieval problem can be considered equivalent to classification. The hyperlane in the case of retrieval is defined from the query vector Q for and an appropriate constant θ ($h(x) = Qx + \theta$). The matching function (cos) for a test example is inversely proportional to the distance from the hyperplane defined by the classifier ($d(j, g) = 2-2cos(j, c)$). Thus following Platt's method we approximate the probability $p(j|g)$ by :

$$p(j|g) = \frac{1}{1 + e^{a \cdot d(j,g)}} \tag{2}$$

The conditional probability $p(c|j)$ is calculated by:

$$p(c|j) = \sum_{w \in C_j} \frac{\sqrt{s(w, j)}}{\sum_{w' \in W_j} \sqrt{s(w', j)}} \tag{3}$$

where C_j is the set of concepts of the ith retrieved image and W_j the set of keywords assigned to image j. Moreover, $s(w, j)$ are the scores provided by the organizers as part of the training set. Finally, the l concepts with the highest posterior probability $p(c|g)$ calculated from Equation (1), are selected as the concepts being present in the test image g. Different values of l were tested with the development set, with $l = 8$ giving the best results and thus this value was selected for our submitted runs.

4 Submitted Runs

Several initial experiments were performed using the development set. The most of notable ones were those which study the impact of the rate parameter a and the number of K for the top retrieved images. The corresponding results in Tables 2 and 3, show that these parameters can have an important impact on annotation performance. For the test set, a total of 10 runs were submitted:

- Run 1: K-NN with $K = 1000$ neighbors, retrieved by Early fusion and LSA on Opponent SIFT, CEDD, FCTH. Parameter $a = 6$. No synonym usage.
- Run 2: Same as Run 1, $a = 10$.
- Run 3: Same as Run 1, $a = 16$.
- Run 4: Same as Run 1, but with $K = 800$ and $a = 16$.
- Run 5: Same as Run 1, but with $K = 450$ and $a = 16$.
- Run 6: K-NN with $K = 1000$ neighbors, retrieved by Early fusion and LSA on Opponent SIFT, CEDD, FCTH. Parameter $a = 6$. Concepts include WordNet synonyms.
- Run 7: Same as Run 6, $a = 10$.
- Run 8: Same as Run 6, $a = 16$.
- Run 9: Same as Run 6, but with $K = 800$ and $a = 16$.
- Run 10: Same as Run 6, but with $K = 450$ and $a = 16$.

Table 4 present the corresponding results of our submitted runs. In Table 1 baseline results are presented using single descriptors for image representation. Tables 2, 3 present results with fusion of several descriptors and varying the values of the parameters a, K.

Table 1. Develop set baseline results

Run	**MF-Samples** %	**MF-Concepts** %	**MAP-Samples** %
oppsift	23.2	14.4	30.8
rgbsift	23.1	15.7	30.7
csift	23.0	12.7	30.3
sift	22.6	13.4	30.0
colorhist	21.4	10.2	28.5
gist2	20.8	10.9	27.9
getlf	20.2	8.8	27.1
random	7.2	5.0	15.8

Table 2. Develop set performance for different values of a. K-NN with $K = 20$ using OppSIFT, CEDD and FCTH.

a	**MF-Samples** %	**MF-Concepts** %	**MAP-Samples** %
2	25.4	18.9	34.7
6	25.6	18.7	35.0
10	25.6	18.9	35.2
14	25.6	18.8	35.2
16	25.6	18.9	35.3

Table 3. Develop set performance for different values of K. ($a = 16$). K-NN with using OppSIFT, CEDD and FCTH.

K	**MF-Samples** %	**MF-Concepts** %	**MAP-Samples** %
20	25.4	18.9	34.7
50	27.1	19.7	37.2
100	28.5	20.5	39.1
200	29.1	20.7	40.1
300	29.4	21.1	40.7
450	29.7	21.5	41.0
800	29.8	20.9	41.0
1000	29.8	20.6	41.0

Table 4. Test set IPL's submitted runs results

Run	MF-Samples %	MF-Concepts %	MAP-Samples %
Run 1	18.5	12.1	21.9
Run 2	18.6	12.4	22.1
Run 3	18.7	13.3	22.4
Run 4	18.9	13.3	22.5
Run 5	18.8	13.0	22.4
Run 6	17.3	12.0	21.3
Run 7	17.7	13.4	22.0
Run 8	18.4	15.7	23.4
Run 9	18.4	15.8	23.4
Run 10	18.3	15.5	23.4

5 Concluding Remarks

We have presented a baseline algorithm for image annotation based on the visual retrieval from the train set and extracted the labels of concepts from the top K-NN retrieved images. Our approach enhances the representation of an image, by fusing different low-level features using LSA. Results show that image representation play an important role in the annotation problem. Furthermore, the number of the retrieved images K, and the way these are compared with a test image g ($p(j|g)$), have an important impact on annotation.

References

1. Caputo, B., Müller, H., Martinez-Gomez, J., Villegas, M., Acar, B., Patricia, N., Marvasti, N., Üsküdarlı, S., Paredes, R., Cazorla, M., Garcia-Varea, I., Morell, V.: ImageCLEF 2014: Overview and analysis of the results. In: CLEF proceedings. Lecture Notes in Computer Science. Springer Berlin Heidelberg (2014)
2. Villegas, M., Paredes, R.: Overview of the ImageCLEF 2014 Scalable Concept Image Annotation Task. In: CLEF 2014 Evaluation Labs and Workshop, Online Working Notes. (2014)
3. Stathopoulos, S., Kalamboukis, T.: An svd-bypass latent semantic analysis for image retrieval. In Greenspan, H., Muller, H., Syeda-Mahmood, T., eds.: Medical Content-Based Retrieval for Clinical Decision Support. Volume 7723 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2013) 122–132
4. Müller, H., de Herrera, A.G.S., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S., Eggel, I.: Overview of the imageclef 2012 medical image retrieval and classification tasks. In Forner, P., Karlgren, J., Womser-Hacker, C., eds.: CLEF (Online Working Notes/Labs/Workshop). (2012)
5. Chatzichristofis, S.A., Boutalis, Y.S.: Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In: ICVS. (2008) 312–322
6. Chatzichristofis, S.A., Boutalis, Y.S.: Fcth: Fuzzy color and texture histogram - a low level feature for accurate image retrieval. In: WIAMIS. (2008) 191–196

7. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(9) (2010) 1582–1596
8. Fellbaum, C.: WordNet: An Electronic Lexical Database. Bradford Books (1998)
9. Villegas, M., Paredes, R.: A k-nn approach for scalable image annotation using general web data. In: Big Data Meets Computer Vision: First International Workshop on Large Scale Visual Recognition and Retrieval, held in conjunction with NIPS 2012. (2012) 1–5
10. Platt, J.C.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: ADVANCES IN LARGE MARGIN CLASSIFIERS, MIT Press (1999) 61–74