MRIM-LIG at ImageCLEF 2010 Visual Concept Detection and Annotation task

Rami Al Batal, Philippe Mulhem
Rami.Albatal@imag.fr, Philippe.Mulhem@imag.fr

Laboratoire Informatique de Grenoble (LIG), Grenoble University, CNRS, LIG

Abstract. This paper focuses on one of the Image CLEF Photo tasks at which the MRIM research group of the LIG participated: the Visual Concept Detection and Annotation. For this task, we applied a simple state of the art technique based on bag of visual words. We extracted SIFT-like features that integrate colors (rgSIFT) proposed by van de Sande[10]. We used then a Kmeans clustering in a way to group these features according to 4000 clusters. We generated then for each image of the training set a 4000 dimensions histogram by summing all the occurrences of each cluster, using the nearest neighbour centroid for each extracted feature. For the recognition we extracted the rgSIFT features from the test set, before generating the 4000 dimensional histograms. We applied then SVMs with RBF kernels using a probabilistic estimation of recognition. The results obtained by our run are presented.

1 Introduction

We describe here the experiments that have been conducted by the MRIM group at the LIG in Grenoble for the ImageCLEF 2010 campaign. We participated at the Visual Concept Detection and Annotation. We present our approach and the results obtained.

The paper is organized as follows. First we describe our image representation in section 2. In this section, we focus on the features that were extracted to represent the images, before describing the histogram representation and the learning process used. We present in section 3 the official results obtained avccording to the two sets of measures proposed. Then, we conclude in section 4.

2 Image representation

This year, we only worked on applying a simple state of the art technique based on bag of visual words for the annotation of images. This approach is inspired by the work on text categorization in [4]. In the context of visual media, this approach has been originaly proposed by Sivic et Zisserman in [9] for the retrieval of video documents, before been applied on still images initially by Csurka and his colleagues in [1] for image classification and then in numerous works ([10], [3], [5], [2]) for image annotation.

2.1 Visual feature extracted

We focus now on the features extracted from the images. Scale Invariant Feature Transforms, namely SIFT[6], have been successsul for the classifiaction and the annotation of images. The images considered are consumers photographs in which color may play a great role, we considered then SIFT-like features that integrate colors. Among such existing features, after experiments we used for the CLEF task the rgSIFT feature proposed by van de Sande in [10]. The rgSIFT features include color information around the salient points in the images. The set of features extracted from the trainig set S_{train} is named S_{feat_train}.

As usual in bag of words approaches, we need to group several features in clusters, in a way to identify visually similar features. To do that, we applied on a subset of S_{feat_train} a Kmeans clustering in a way to group these features according to N_c clusters. Recent studies demontrated that large numbers for N_c (vocabulary size), namely several thousands, perform better for image classification and retrieval ([7], [8],[10]). That is why, after some tests, we chose to use N_c=4000.

2.2 Learning of concepts

For learning step of our approach, we generated for each image of S_{train} a 4000 dimensions histogram by summing all the occurrences of each cluster, using the nearest neighbour centroid for each extracted feature.

Then, a learning of each concept model is achieved using Support Vector Machines (SVMs). The one against all (OAA) approach was experimented: all the positive sample and negative samples are used to learn each concept. In the SVMs, we use the common Radial Basis Function kernel defined by equation (1).

$$K(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} \tag{1}$$

For the definition of the values of the parameter $sigma$ we learned the models for each label using half of S_{train} for testing, namely S_{train_train}, and half of S_train for validation, namely S_{train_valid}. There two subsets form a partition of S_{train}, and they were selected randomly. For each concept the same subsets S_{train_train} and S_{train_valid} were used.

2.3 Annotation of images

For the generation of the results, we extracted the same rgSIFT features for the test set S_{test}, before generating the 4000 dimensional histograms (one per image of S_test).

We applied then the recognition based on the SVM models defined during the laerning step, using a probabilistic estimation of recognition. We submitted only one result.

3 Submitted run and results

We submitted one run based on the characteristics described above. The run has the following identifier: LIG__1277153756343__clefResults.txt_binary.txt .

We focus first on the Mean Average Precision (MAP) result obtained by our approach. We obtained the rank 30 on 45 submissions, with a MAP value of 0.225 . This value is 0.1 lower than the median value for theses runs, 0.237. We can mention however that for one concept, $Visual\_Art$, the MAP that we obtained is the second result, 0.374, after the IJS result at 0.385 . In this case we think that the color aspect integrated in the rgSIFT features is the reason for this result.

For the hierarchical recognition measure based on example-based F-measure, we obtained the rank 27 on 45. The value obtained is 0.477, and the median value is 0.530; we achieve then quite poor results according to this measure. For the hierarchical recognition measure based on the Ontology Score incorporating the Flickr Context Similarity, we achieved the 20th rank, with a value of 0.530. Our result is above the median value of 0.515 for the 45 visual only runs considered.

4 Conclusion

To summarize our work for the ImageCLEF 2010 Visual Concept Detection and Annotation task, we proposed a simple state of art method. Our work here demonstrates that such state of the art techniques are a basis for further extensions.

In the future, we will integrate grouping of regions of interest to increase our results.

Acknowledgment

This work was partly supported by: a) the French National Agency of Research (ANR-06-MDCA-002), b) the Région Rhones Alpes (projet LIMA).

References

1. Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.
2. L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In IEEE, editor, *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, volume 2, pages 524–531 vol. 2, Washington, DC, USA, June 2005. IEEE Computer Society.
3. Yu G. Jiang, Chong W. Ngo, and Jun Yang. Towards optimal bag-of-features for object categorization and semantic video retrieval. In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 494–501. ACM, 2007.

4. Thorsten Joachims. Text categorization with support vector machines: learning with many relevant features. In Claire Nédellec and Céline Rouveirol, editors, *Proceedings of ECML-98, 10th European Conference on Machine Learning*, number 1398, pages 137–142, Chemnitz, DE, 1998. Springer Verlag, Heidelberg, DE.
5. Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2169–2178. IEEE Computer Society, October 2006.
6. David G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, 1999.
7. David Nister and Henrik Stewenius. Scalable recognition with a vocabulary tree. In *In CVPR*, volume 2, pages 2161–2168, 2006.
8. James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, 2007.
9. J. Sivic and A. Zisserman. Video google: a text retrieval approach to object matching in videos. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1470–1477. IEEE, April 2003.
10. Koen E. A. van de Sande, Theo Gevers, and Cees G. M. Snoek. Evaluation of color descriptors for object and scene recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, June 2008.