The University of Amsterdam's Concept Detection System at ImageCLEF 2010

Koen E. A. van de Sande and Theo Gevers
Intelligent Systems Lab Amsterdam (ISLA), University of Amsterdam
Software available from: `http://www.colordescriptors.com`

Abstract

Our group within the University of Amsterdam participated in the large-scale visual concept detection task of ImageCLEF 2010. The submissions from our visual concept detection system have resulted in the best visual-only run in the per-concept evaluation. In the per-image evaluation, it achieves the highest score in terms of example-based F-measure across all types of runs.

1 Introduction

Our group within the University of Amsterdam participated in the large-scale visual concept detection task of ImageCLEF 2010. The Large-Scale Visual Concept Detection Task [2] evaluates visual concept detectors. The concepts used are from the personal photo album domain: *beach holidays*, *snow*, *plants*, *indoor*, *mountains*, *still-life*, *small group of people*, *portrait*. For more information on the dataset and concepts used, see the overview paper [2]. Our participation last year, in ImageCLEF 2009, focussed on increasing the robustness of the individual concept detectors based on the bag-of-words approach, and less on the per-image evaluation.

Last years experiments [3–8] emphasize in particular the role of visual sampling, the value of color invariant features, the influence of codebook construction, and the effectiveness of kernel-based learning parameters. This was successful, resulting in the top ranking for the large-scale visual concept detection task in terms of both EER and AUC. Both these measures do a per-concept evaluation. The per-image evaluation based on the ontology score suggested that the assignment of concept tags to images leaves room for improvement. Therefore, for this year, we focus on the per-image evaluation. The primary evaluation metric used in 2010 for the per-image evaluation is the average example-based F-measure.

2 Concept Detection System

Our concept detection system is an improved version of last years system [6]. For the ImageCLEF book [1], we have performed additional experiments [5] which give insight into the effect of different sampling methods, color descriptors and spatial pyramid levels within the bag-of-words model. Two of our runs this year correspond exactly to *Harris-Laplace and dense sampling every 6 pixels (multi-scale) with 4-SIFT* and *Harris-Laplace and dense sampling every pixel (single-scale) with 4-SIFT* from this book chapter [5]. These runs were also submitted to ImageCLEF@ICPR 2010. Please refer to the cited papers[1] for implementation details of the system.

To achieve better results in the per-image evaluation, where we need to perform a binary assignment of a tag to an image, we have modified the probabilistic output of the SVM. We have disabled Platts conversion method to probabilities, and instead use the distance to the decision boundary. The decision boundary lies at 0, positives are trained to lie at 1 and negatives are trained to lie at -1. In a cross-validation experiment, we have found a threshold of -0.3 to be good for most concepts: the default threshold of 0 would be too conservative when evaluating with an example-based F-measure where precision and recall are weighted equally. Optimizing the threshold on a per-concept basis instead of a single threshold was found to be less stable.

3 Submitted Runs

We have submitted five different runs. All runs use both Harris-Laplace and dense sampling with the SVM classifier. We do not use the EXIF metadata provided for the photos nor the provided text tags.

- **Harris-Laplace and dense sampling every 6 pixels (multi-scale) with 4-SIFT**: from ImageCLEF 2009 [6], ImageCLEF@ICPR 2010 and the ImageCLEF book [5].
- **Harris-Laplace and dense sampling every pixel (single-scale) with 4-SIFT**: from ImageCLEF@ICPR 2010 and the ImageCLEF book [5].
- **Harris-Laplace and dense sampling every 6 pixels (multi-scale) with 4-SIFT plus soft assignment and multiple kernel learning**: try to optimize the soft assignment parameters from [9] with multiple kernel learning.
- **mkl-bothdenseallharris-4sift-plus**: includes improved color descriptors which have not yet been published.
- **mkl-mixed-mixed**: includes improved color descriptors which have not yet been published.

[1] Papers available from `http://www.colordescriptors.com`

Table 1: Overall results of the our runs evaluated over all concepts in the Photo Annotation task with Average Precision.

Run name	Type	AP
Harris-Laplace and dense sampling every 6 pixels (multi-scale) with 4-SIFT	Visual	0.3963
Harris-Laplace and dense sampling every pixel (single-scale) with 4-SIFT	Visual	0.4026
Harris-Laplace and dense sampling every 6 pixels (multi-scale) with 4-SIFT (soft/MKL)	Visual	0.3939
mkl-bothdenseallharris-4sift-plus	Visual	0.4069
mkl-mixed-mixed	Visual	**0.4073**

Table 2: Results using the per-image evaluation measures for our runs in the Large-Scale Visual Concept Detection Task. Measures are the average example-based F-measure and Ontology Score with Flickr Context Similarity.

Run name	F-measure	OS with FCS
Harris-Laplace and dense sampling every 6 pixels (multi-scale) with 4-SIFT	0.6754	0.5953
Harris-Laplace and dense sampling every pixel (single-scale) with 4-SIFT	0.6785	0.6005
Harris-Laplace and dense sampling every 6 pixels (multi-scale) with 4-SIFT (soft/MKL)	0.6512	0.5222
mkl-bothdenseallharris-4sift-plus	0.6782	0.5857
mkl-mixed-mixed	**0.6801**	0.5908

4 Evaluation Per Concept

In table 1, the overall scores for the evaluation of concept detectors are shown. The features with sampling at every pixel instead of every 6 pixels perform better (0.4026 versus 0.3963), which is similar to the result obtained in ImageCLEF@ICPR 2010. Optimizing the parameters of soft assignment using Multiple Kernel Learning did not have the desired effect. A possible explanation is that the slack parameter for MKL was set to 1, whereas the normal SVM runs optimize this parameter and tend to select 10 as a good slack setting. The two final runs perform better than the two 'baseline' runs from the ImageCLEF@ICPR 2010. However, the color descriptors present in these two runs have not yet been documented.

Compared to other ImageCLEF participants, our runs are the best visual-only submissions. However, combinations of text and visual methods do get a higher overall AP. For concepts like *Birthday* and *Party*, an attached tag with the words party or birthday implies presence of that concept, whereas the visual presence might be more ambiguous. The best visual+text method scores 0.4553, compared to 0.4073 for our best visual-only run and 0.2338 for the best text-only run.

5 Evaluation Per Image

For the per-image evaluation, overall results are shown in table 2. Our emphasis on optimizing the threshold for tag assignment has resulted in the best overall run in terms of example-based F-measure, *i.e.* this visual-only run outperforms visual+text methods.

6 Conclusion

The submissions from our visual concept detection system in the ImageCLEF 2010 large-scale visual concept detection task have resulted in the best visual-only run in the per-concept evaluation. In the per-image evaluation, it achieves the highest score in terms of example-based F-measure across all types of runs.

References

[1] H. Mueller, P. Clough, T. Deselaers, and B. Caputo. *ImageCLEF*, volume 32 of *Lecture Notes in Computer Science: The Information Retrieval Series*. Springer, 2010.

[2] S. Nowak and M. Huiskes. New strategies for image annotation: Overview of the photo annotation task at imageclef 2010. In *Working Notes of CLEF 2010*, 2010.

[3] C. G. M. Snoek, K. E. A. van de Sande, O. de Rooij, B. Huurnink, J. R. R. Uijlings, M. van Liempt, M. Bugalho, I. Trancoso, F. Yan, M. A. Tahir, K. Mikolajczyk, J. Kittler, M. de Rijke, J. M. Geusebroek, T. Gevers, M. Worring, D. C. Koelma, and A. W. M. Smeulders. The MediaMill TRECVID 2009 semantic video search engine. In *Proceedings of the TRECVID Workshop*, 2009.

[4] J. R. R. Uijlings, A. W. M. Smeulders, and R. J. H. Scha. Real-time bag-of-words, approximately. In *ACM International Conference on Image and Video Retrieval*, 2009.

[5] K. E. A. van de Sande and T. Gevers. *University of Amsterdam at the Visual Concept Detection and Annotation Tasks*, chapter 18, pages 343–358. Volume 32 of *The Information Retrieval Series: ImageCLEF* [1], 2010.

[6] K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. The university of amsterdam's concept detection system at imageclef 2009. In *Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers*, Lecture Notes in Computer Science. Springer, 2010.

[7] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Color descriptors for object category recognition. In *European Conference on Color in Graphics, Imaging and Vision*, pages 378–381, 2008.

[8] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010.

[9] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J.-M. Geusebroek. Visual word ambiguity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1271–1283, 2010.