

UAIC at ImageCLEF 2009 Photo Annotation Task

Adrian Iftene, Loredana Vamanu, Cosmina Croitoru

UAIC: Faculty of Computer Science, “Alexandru Ioan Cuza” University, Romania
{adiftene, loredana.vamanu, cosmina.croitoru}@info.uaic.ro

Abstract. The present article describes the system used for the our first participation in the imageCLEF 2009 Photo Annotation task. For the image classification we used four components: (1) first uses face recognition, (2) second one use training data, (3) third one uses associated exif file and (4) the fourth uses default values calculated according to the degree of occurrence in the training set data. The UAIC team’s debut in the ImageCLEF competition has enriched us with the experience of developing the first system for the Photo Annotation task, at the same time setting the scene for subsequent ImageCLEF participations.

Keywords: ImageCLEF, Visual Concept Detection, Image Annotation

1 Introduction

Following the tradition from 2008, ImageCLEF 2009 offered again a visual concept detection and annotation task. In 2009, the organizers were focused on the extension of the task concerning the amount of data available and the amount of concepts annotated. In comparison with 2008 when were over 1800 images for training and 1000 images for testing available, in 2009 the training and test set consisted from thousand images from Flickr image database.

All images had multiple annotations with references to holistic visual concepts and were annotated at an image-based level. Few examples of categories for the visual concepts are: **Abstract Categories** (*Landscape, Family&Friends, Partylife, etc.*), **Seasons** (*Sumer, winter, etc.*), **Persons** (*no, single, big groups*), **Quality** (*blurred, underexposed, etc.*), **Representation** (*portrait, macro image, canvas, etc.*).

The visual concepts were organized in a small ontology with 53 concepts. Participants may use the hierarchical order of the concepts and the relations between concepts for solving the annotation task.

The structure of the rest of the paper is as follows: Section 2 describes our system and it main components; Section 3 describes our results. In the last Section we summarize our participation in the track and give some conclusions about the experience.

2 The UAIC system

The system built by us for this task has four main components. First component try to identify in every image people faces and after that accordingly with the number of these faces make the classification. Second one use for classification clusters built from training data and calculates for every image the minimum of distances between image and clusters. Third one uses for classification details extracted from associated exif file. If none from these components cannot performs the image classification, it is making by fourth module that uses default values calculated according to the degree of occurrence in the training set data.

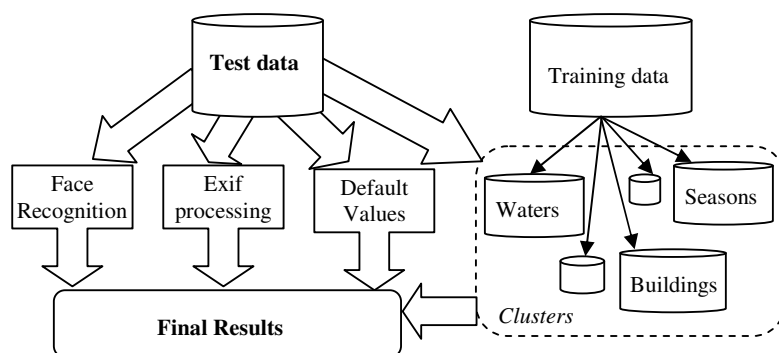


Figure 1: UAIC System

2.1 Face Recognition

Some of categories implied the presence of people, such as *Abstract Categories* (through concepts: Family Friends, PartyLife, Beach Holidays, CityLife), *Activity* through the concept Sports, of course *Persons* and *Representation* having one of the concepts Portrait. Finding a program that recognized faces in a photo and that could distinguish if it is a portrait or not was maybe the easiest solution that could help, in correlation with other concepts discovered with other means, to annotate the obvious concepts but also the more abstract ones.

We discovered a JAVA library, Faint¹ (developed by Malte Mathiszig, University of Oldenburg), that did exactly what we need it: recognizes if there are any faces in the photo given. Also, we can receive how many and how much percent of that picture is the face. In this way we were able to decide if there is a big group, a small one, if the photo was a portrait.

Unfortunately, if the light isn't normal for example if there is a picture taken on a black night or if there is fog or a shaken photo the result will not be accurate. After

¹ Faint: <http://faint.sourceforge.net/>

our estimations, it works well for 80% of these cases, in comparison with day pictures.

2.2 Clustering using Training Data

We also used a similarity processing for finding some concepts. For this we have selected the most representing pictures for some concepts from test data and we used JAI² (Java Advanced Imaging API) for manipulating images easily and a small program that calculates a rate of similarity between the chosen photos and the photo wanted to be annotated.

It was hard to find the most representing photos for concepts as every concept can be seen so different in different seasons, different time of day, etc; but the hardest part was to decide the acceptance rate. Using the training data, we ran for some images that we expected to be annotated with one or more of the concepts that were illustrated by the pictures in our small clusters and we notated the rates and we also made the same thing with pictures that shouldn't be annotated with one of the concepts but were very similar with the pictures chosen to compare to and we also notated the rates.

In the end, we made kind of a compromise average rate and this was our limit rate. Of course that this algorithm could be improved, it can be calculated a rate for every cluster and maybe this way the program would be more accurate. The concepts that we tried to annotate in this way were CityLife, Clouds, Snow, Desert, Sea and Snow.

2.3 Exif Processing

We processed the exif information for every picture and according to the parameters of the camera with which the picture was taken we were able to annotate concepts for example related to illumination, but also correlating with other concepts found, some more abstract concepts like City Life or Landscape-Nature could also be found this way. Because the information taken from exif can or can not be accurate, and some of our limits can be subjective the concepts discovered annotated this way that were not clearly were set to 0.5.

2.4 Default values

There are five categories that contain disjoint concepts, which implies that only one concept from this kind of category can be found in a photo. Taking this into consideration, if by any other method a concept from a disjoint category was not discovered than a default value will be inserted. The default concept is selected according to the degree of occurrence in the training set data, the statistic with the occurrence values for every concept were delivered at the same time with the test data.

² JAI: <http://java.sun.com/javase/technologies/desktop/media/jai/>

There are two cases: first one if there was no way in order to find one of the concepts from the category then the neutral (the “no” concept (for example: *No_Visual_Time*)) will be annotated and the other case is: if one of the concepts could be distinguished but with no certainty then the concept that has appeared the most in the test data will be the one annotated, as it is the most probably the one that appears in the photo.

For example for the Seasons category, if there was nothing in the picture that indicated a season then the *No_Visual_Season* concept will be annotated, otherwise it will be annotated with Summer as it is by far the most frequent in the test data.

3 Results

The time necessary to run the application for the test data was 24 hours. It took so long because of the similarities process as this supposed to compare a photo with 7 up to 20 photos for every category that was chosen for this process. The computer on which it ran had 2 GB memory, an Intel E8200 Dual Core processor with 2600 MHz and a Western Digital 320GB HDD with 10 000 rotations/min HDD.

For the annotation process it was taken into consideration the relation between categories and the hierarchical order.

We submitted only one run with the following official evaluation (Paramita et al., 2009):

Table 1: UAIC run in Photo Annotation Task

Run ID	Average EER	Average AUC
UAIC_34_2_1244812428616_changed	0.4797	0.105589

Our detailed results regarding EER on classes are presented in Figure 2. We can see how the results are almost the same on different classes (around 0.5) with an average of 0.4797.

The better values were obtained for classes where default value rule was applied: 17 (*No_Visual_Place*) with precision 0.529716, 33 (*Canvas*) with precision 0.522709, 8 (*Sports*) with precision 0.518878, and 19 (*Flowers*) with precision 0.517818.

The lower values were obtained for classes where clusters from training data were used: 13 (*Winter*) with precision 0.335529, 12 (*Autumn*) with precision 0.354693, 29 (*Night*) with precision 0.394617.

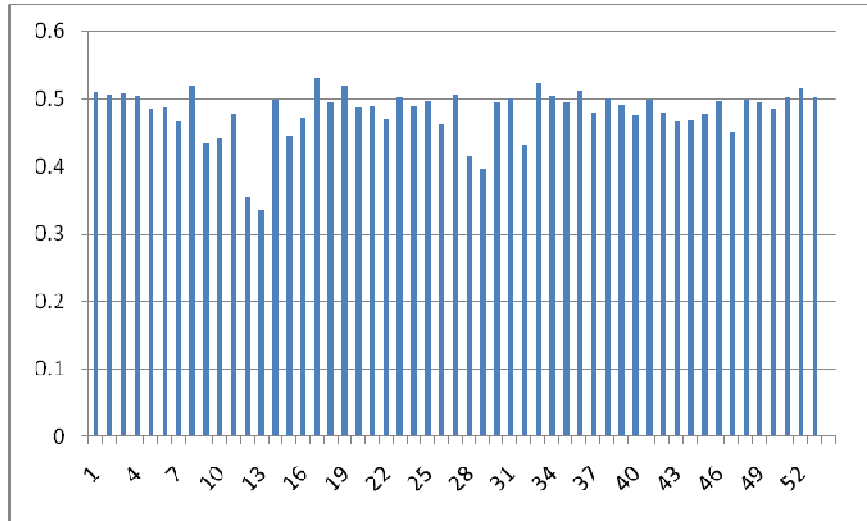


Figure 2: EER - Detailed Results on Classes

Similar, our detailed results regarding AUC on classes are presented in Figure 3.

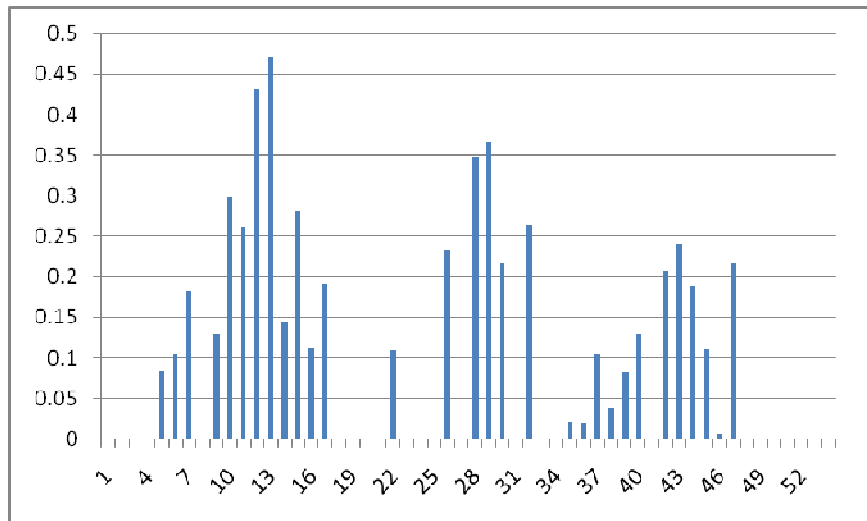


Figure 3: AUC - Detailed Results on Classes

4 Conclusions

The system built by us for this task has four main components. First component try to identify in every image people faces and after that accordingly with the number of these faces make the classification. Second one use for classification clusters built from training data and calculates for every image the minimum of distances between image and clusters. Third one uses for classification details extracted from associated exif file. If none from these components cannot performs the image classification, it is making by fourth module that uses default values calculated according to the degree of occurrence in the training set data.

From our run evaluation we conclude that some of the applied rules are better than others. Thus, the second rule from rules presented above that uses clusters from training data for classification is the worst rule, and the fourth rule that uses the default values calculated according to the degree of occurrence in the training set data is the best rule.

For the future, we will perform a more detailed analysis of the results in order to identify exactly the confidence of our rules and we will try to apply rules in descending order of these values.

Acknowledgements

The authors would like to thank the following members of the UAIC team: all students from group 1A, second year, for their help and support at different stages of system development.

References

Paramita, M., Sanderson, M. and Clough, P. Diversity in photo retrieval: overview of the ImageCLEFPhoto task 2009. CLEF working notes 2009, Corfu, Greece, 2009.