JKU-Satellite at MediaEval 2015: An intuitive approach to locate single pictures within a Session

Michael Mayr
Johannes Kepler
Universität Linz

Thomas Hintersteiner
Johannes Kepler
Universität Linz

Michael Weingartner
Johannes Kepler
Universität Linz

Florian Schröckeneder
Johannes Kepler
Universität Linz

Peter Knees
Johannes Kepler
Universität Linz

Markus Schedl
Johannes Kepler
Universität Linzl

ABSTRACT

In this paper we describe our solution to the placing task of the MediaEval 2015 workshop. In particular we provide a solution for Sub-Task 1.2, which is the Mobility-based placing task [1]. We offer a intuitive solution for run 1, a different approach for run 2 and combine these two solutions in run 3.

Run 1 just uses known coordinates and timestamps to predict a missing location. The second run uses only given image features. In a third run we combine these two approaches.

1. INTRODUCTION

The goal of the Mobility-based placing Task is to predict coordinates of pictures. What we have is a series of pictures, taken by one photographer in one city. Some of these pictures lack geographical information. We then have to estimate these missing coordinates by using the locations of the other pictures and their image features.[1]

2. OUR APPROACH

The following part describes our solution in detail. For the first run only coordinates and timestamps are used, for the second run only given image features are used. For the third run both resources are used.

In order to get an idea whether our predictions are anywhere near the real coordinates, we deleted randomly 20% of the coordinates from the training set and checked the predictions against the actual coordinates.

2.1 Run 1

The most intuitive solution we could come up with was implemented in run 1. For this approach the dataset is split up in single sessions of pictures. A session is a set of pictures as described in the Introduction.

MediaEval 2015 Workshop, Sept. 14-15, 2015, Wurzen, Germany

In this sessions we look for pictures with missing coordinates. The normal case is a picture with missing coordinates between two pictures where the coordinates are known. In this case a vector between the two known coordinates is computed. With the timestamps of the pictures we can estimate how far from the last known location to the next known location the picture with the missing coordinates has been taken.

In other words we just compute the direction and the time between two known locations. We then just place the missing location on this direction, according to the time it took to get there. For example, picture A and picture C are 100 meters apart and picture C was taken 10 minutes after A. Between A and C there is another picture B. B has no coordinates, but it has a time stamp. B was taken 3 minutes after A. With a simple division and a multiplication we estimate the location of picture B to be 30 meters from A in direction B.

This approach also works if more than one picture is missing a location. In that case the direction between the last two known locations is computed.

If the location of the first picture is missing, the first known point is set to the next picture with coordinates and if the last picture is missing we set the last known point to the last picture with coordinates. This produces the best solution for this 2 cases, because there is no direction, which can be read out of the data.

2.2 Run 2

For this run, only the provided image features are used. After experimenting with single features and combining them we decided to use the CEDD feature, because it provided the best results on the training set.

The experimentations where done as described at the beginning of the chapter. As stated we deleted 20% of the coordinates,this 20% remained the same during the different runs of the experiments in order to see which feature works best. We do not know why the chosen feature worked best and if there exists a better combination of features. In the following few lines we describe the algorithm used for the experiments. For run 2 we take this algorithm and apply it to the test data with the chosen feature.

The algorithm itself is relatively simple. First the feature vector of a picture with missing coordinates is read. This vector is compared to every other feature vector of the whole data set. This comparison is done by computing the Euclidean distance between these vectors. The picture with the shortest distance, the most similar is then taken and its coordinates are predicted.

We know that this approach by itself does not lead to really useful predictions. However it should improve the predictions which are already near the actual coordinates. This approach is designed to just further improve the results of run 1.

2.3 Run 3

In our final run, we combine run 1 and run 2, with a few alterations. We start by splitting the pictures into single sessions, as we do in run 1. The next step is to get the first picture without a location. Now we get all pictures that belong to the same node. A node in this case means a area as described in the data set. After that, we keep the 25 pictures with the shortest Euclidean distances and rank them from shortest to longest distance. After we retrieved these similar pictures, we look at all pictures of the current session and compute an average speed with the timestamps. Now we take the next picture of the session that has a location. This can be a picture before the one with the missing location or afterwards.

What we have now is a picture with a location, an average speed and 25 pictures with similar features to that one with the missing coordinates. Now we can compute a radius from the last known location and the average speed and get a rough idea where the missing location should be. The next step is simply to see if one of the 25 similar pictures was taken is inside this radius. If this is the case, the coordinates of this picture become the predicted ones. We take the first (most similar) picture that was taken inside the radius.

The idea behind this is that many people take pictures of the same things, so if we already narrow the location down to a certain radius chances are that another person took a picture at the same place that looks similar.

If we do not find a similar picture, that was taken inside the reachable radius, we use the unaltered approach from run 1 to predict the coordinates of that picture.

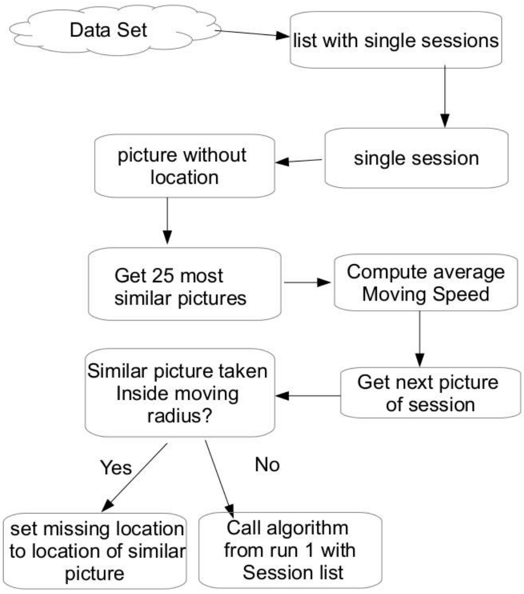

Figure 1: Process of Run 3

3. RESULTS

Our best run was run 1. With this approach our estimations of about a quarter (23.18%) of all locations where within 0.1 km of the actual location. And 76.94% where within a kilometre. In a range of 10 km away from the actual coordinates we got 96.92%.

With this result we can say that our method for run 1 works quite well for placing the pictures in a certain range, but is in many cases not enough to estimate exact coordinates. One of the weaknesses is a location missing at the beginning or the end of a session. But due to the fact that the pictures are taken within 10 minutes, the error distance is within the other estimation errors.

Another problem is that people who go from some point to somewhere else do not move at a constant speed, they might stop, slowdown or speed up along the way. Nevertheless we can get a rough idea of the location, where the picture might have been taken.

Not surprisingly, the predicted coordinates of run 2 where way off the actual coordinates. Because the coordinates of the other pictures of a session where not taken into account, only the image features where considered, under 1% of the locations where within 10 km from the actual location. This approach is just intended to improve the results of run 1 and of course on its own leads to a low success rate.

Run 3 had similar results than run 1, 23.06% of all locations where within 0.1 km, 76.28 % where within a kilometer and 96.82% where in a range of 10km. The predictions where a bit further away from the actual coordinates than run 1. This comes as bit of a surprise, because in our experiments with the development set, we got a slight improvement of results compared to run one. We guess it just depends on how many pictures in the whole data set are taken at the same area as the pictures with the missing coordinates.

Due to limited resources and time we could not experiment with combinations of all available features. One way to further improve our results from run 3 would certainly be more experimentation in that direction.

4. REFERENCES

[1] J. Choi, C. Hauff, O. V. Laere, and B. Thomee. The Placing Task at MediaEval 2015. In Working Notes Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, September 2015.