Image matching for individual recognition with SIFT, RANSAC and MCL

Dávid Papp, Ferenc Mogyorósi, Gábor Szűcs

Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Magyar Tudósok krt. 2., H-1117, Budapest, Hungary,

pappd@tmit.bme.hu, mogyi006@gmail.com, szucs@tmit.bme.hu

Abstract. Monitoring the whale individuals in the ocean is a current problem among conservationists. Biologists often use photos of whale caudal for this problem as it is the most discriminant pattern for distinguishing an individual whale from another, but it often requires laborious visual analysis. There was a challenge announced in the SeaCLEF of LifeCLEF campaign for automatic whale individual recognition based on visual contents. We elaborated a solution to compare the photos of individuals by SIFT features (as simple image representation) with spatial consistency refinement method RANSAC (based on a rotation and scale transformation model). After determining the similarity of every pair, to discover the wrong similarity values and correct them, a clustering method was applied on the similarity graph of the dataset.

Keywords: whale recognition, similarity, SIFT, SeaCLEF, unsupervised learning

1   Introduction

Analysis of images usually requires very time-consuming and expensive input by human observers, and this is true for sea images as well, although the statistics of data collection would be very useful for exploratory applications, in particular for fisheries and biological areas. This demands effective methods for automatic content analysis to enable proactive provision of analytical information; and in order to solve this problem a challenge is announced in SeaCLEF 13 of the LifeCLEF 3 campaign of ImageCLEF.

In this challenge there were four subtasks: (1) Automated Fish Identification and Species Recognition on Coral Reef Videos, (2) Automated Frame-level Salmon Identification in Videos for Monitoring Water Turbine, (3) Marine Animal Species Recognition using Weakly-Labelled Images and Relevance Ranking, and (4) Whale Individual Recognition. We enrolled in the last subtask, where the goal was to find the images that correspond to the same individual whale.

2   Related works

The aim is to monitor the organisms at the individual level rather than at the species level in the task of discovering whale individuals in a large collection of pictures collected by nature observers. The process is initiated without any knowledge on the number of individuals and without any training samples of these individuals; thus, the problem is entirely unsupervised. The task was identical to the challenge in the last year, so there are some previous works dealing with this problem.

In the paper 10 GMM (Gaussian Mixture Model) 12 was used to determine visual codebook with parametric probability density function represented as a weighted sum of Gaussian component densities. According to the codebook the next step was to create a descriptor that specifies the distribution of the visual codewords in any image, called high-level descriptor. To represent an image with high-level descriptor, the GMM based Fisher vector 11 was calculated. These vectors were the final representations (image descriptor) of the images. Finally a kernel matrix was built using RBF (Radial Basis Function) kernel function, an entry of kernel matrix $K(i,j)$ describes how similar the ith Fisher vector to the jth Fisher vector (i.e. the ith image to the jth image), and the retrieval of predicted pairs was based on these similarity values.

In another paper 4 a scalable fine-grained matching system was developed to discover small rigid visual patterns in highly clutter background. The work was based on RANSAC algorithm 1, as a solution to perform epipolar geometry estimation. It consists in generating transformation hypotheses using a minimal number of low-level visual correspondences and then evaluating each hypothesis based on the number of inliers among all features under that hypothesis.

3   Method for Whale Individual Recognition

The aim was to find the images that correspond to the same individual whale; and the basic idea was to create the mathematical representation of each image based on visual content, then calculate the pairwise similarity values (these ideas were different from our previous works 101415).

We elaborated more alternative ways in the continuation of the process. In one of the alternatives the matching decisions among the images are binary, and in the other there are continuous (presenting the reliability that they are matching pairs).

We used the segmentation propagation technique introduced in 56 for separating the background (the water) from the whale's caudal fin, and we used the segmented images for further calculations. We used the default configuration of the segmentation technique proposed in 56. As it can be seen in Figure 1, the algorithm was not able to perfectly separate the fins in all cases. In the first column of the figure, there are a few

examples for a perfect segmentation; the second column includes acceptably good segmentations, where a portion of water still present in the masked version of the images; finally, some wrong segmentations can be seen in the last column, but this occurred rarely.

Fig. 1 Some examples for the segmentation of whales

3.1 Extracted features and SIFT matching

Lots of different feature types can be detected in an image, e.g. corners, edges, ridges, as "interesting" part of an image, furthermore many possible feature extraction methods are available for images. We used the Harris-Laplace corner detector 29 for feature (i.e. keypoint) detection and we chose the SIFT (Scale-Invariant Feature Transform) 78 algorithm to create descriptor vectors.

Our idea was to determine a similarity metric between two images based on their SIFT descriptors, therefore we searched for matching descriptor pairs on the images. Let us consider image I and image J, in order to avoid time consuming process caused by the large amount of possible pairs an approximate nearest neighbors search (in a k dimensional tree) 7 was used for investigation of matching. Our solution searched for the two nearest neighbors (the nearest and the second nearest one) to apply a ratio test, so the nearest neighbor of a descriptor was rejected as a matching pair if that was not significantly closer than the second nearest. In our experiments a ratio (r) was used for this purpose, and the condition of acceptance for a matched feature pair can be seen in Equation 1.

$$p_{ij} = \begin{cases} match & | \ r \times d_{ij} < d_{il} \\ reject & | \ otherwise \end{cases} ; \quad i = 1 \dots n ; \ j, l = 1 \dots m ; \ j \neq l \quad (1)$$

where p_{ij} denotes a pair of SIFT descriptors, d_{ij} and d_{il} are Euclidean distances between the appropriate descriptors (l is the descriptor second closest to i), and n, m denote the number of descriptors on images I and J respectively. In the examination of an image pair many matched feature pairs can be found, therefore a possible way to decide whether an image pair is a match is setting a minimum required number (t) of matched features between them. This gives us a binary decision for each pair of images, and we used this during the creation of 'BME_DCLab_whalerun_3' that we discuss later on.

We define a new metric to measure the similarity between two images (I and J), and we denote it by s_{IJ}.

$$s_{IJ} = \frac{N}{\max{(d_{ij})}} \qquad (2)$$

where N denotes the number of matched feature pairs. In this expression minimizing the $\max{(d_{ij})}$ automatically minimizes all the other distances between the matched feature pairs. On the other hand, N should be as high as possible, because the more matches of SIFT descriptors are discovered the more likely for the image pair to be a match. Therefore, the aim was to maximize s_{IJ}, so higher score represents more confidence. According to this we sorted the image pairs based on the s_{IJ} values in a decreasing order, and this was our 'BME_DCLab_whalerun_1' solution. Note that we used $r = 1.7$ as ratio during the calculation of the s_{IJ} similarity values.

Our solution examined every pair of images to reveal cliques, i.e. images that are mutually the closest to each other according to the similarity values. After that the previous run file was modified by moving the cliques to the beginning of the sorted list, and we submitted this slightly different run file under the name of 'BME_DCLab_whalerun_2'.

3.2   Homography and clustering

In the feature matching lot of false matches can occur due to case where the features belong to different parts of the images in the comparison. To decrease the number of these false positive matches, we applied a rotation-and-scale transformation model, which transforms the keypoints of one image to the keypoints of the other; and RANSAC 1 was used for this. The transformation was applied to the corner points of one image and then it was tested on convexity and area. A positive pair can be seen in Figure 1, there are lots of matched features (linked with green line), and the white line is the border of the transformed rectangle.
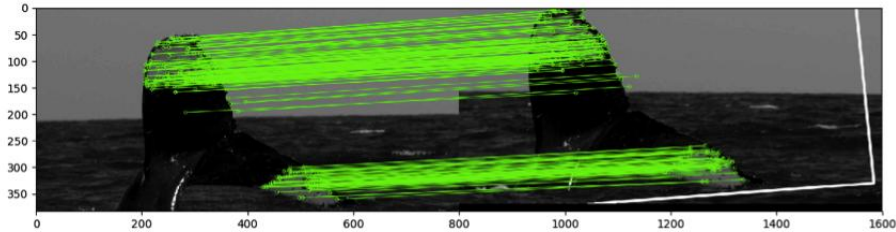
Fig. 2 A positive pair with many matched features

Now let us consider the binary decisions for each pair of images with $t = 30$ (minimum number of SIFT matches). This gives us an addition threshold, therefore we lowered the distance ratio to $r = 1.4$ for balancing the requirement of acceptance. We were able to detect the possible false matches by using rotation-and-scale transformation model (i.e. some 1 turned to 0). Regarding the options of RANSAC, the value of maximal difference in pixels was 5, and the values of other parameters were the default ones.

Nonetheless there still may be some wrong predicted values, and some of them can be corrected if we consider the similarity graph of the images with 1 and 0 weight values. The images of a common whale individual determine a cluster, and in the cluster high weight values (close to 1) and among clusters low weight values (close to 0) are expected. For this unsupervised learning, i.e. clustering we used Markov Cluster Algorithm (MCL) 16, which is a random walks based clustering algorithm. By creating the possible clusters, we further refined the initial decisions, because it was able to discover new connections or to reject existing ones. Note that, we set the "max_loop" parameter of MCL to 30, to increase the chance of detecting hidden relationships between the data points (other parameters were used at their default value). At this point we had a binary score (1 for matches 0 for non matches) and the s_{IJ} similarity value for each image pair. Thus we created an ordered list of image pairs based on the s_{IJ} metrics with the addition that a pair with 0 binary score can never precedes one with 1 binary score; this concludes the 'BME_DCLab_whalerun_3'.

3.3   Official evaluation of Whale Individual Recognition

The metric used for evaluating the submitted run files was the Average Precision (i.e. the precision averaged across all good matches of the ground truth). The following table provides the AP values of our submitted runs. Because of limited number of runs we could not measure contributive ratios of the clustering algorithm and other solution parts separately. However, the combination of solution parts in the

'BME_DCLab_whalerun_3' eventuated the best average precision as can be seen in the Table 1.

Table 1 Results of Whale Individual Recognition

Run name	Average Precision
BME_DCLab_whalerun_3	0.51
BME_DCLab_whalerun_1	0.30
BME_DCLab_whalerun_2	0.30

4   Conclusion

We elaborated three different methods (two of them are only slightly differ) for whale individual recognition. We used SIFT descriptors to represent the visual content of the images, and we searched for matching descriptor vectors between each pair of images. We defined a new metric to measure the similarity of the pairs by a confidence value, and then refined these scores to get even more precise predictions. For this purpose, we used RANSAC and MCL algorithms. The most complex run file that we submitted was able to achieve 0.51 Average Precision in the official evaluation.

References

1. Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381-395.
2. Harris, C, Stephens, M.: A combined corner and edge detector. In C. J. Taylor, editors, *Proceedings of the Alvey Vision Conference*, pages 23.1-23.6. Alvey Vision Club, September 1988. doi:10.5244/C.2.23.
3. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W. P., Lombardo, J. C., Planqué, R., Palazzo, S., Müller, H.: LifeCLEF 2017 Lab Overview: multimedia species identification challenges, *Proceedings of CLEF 2017* (2017)
4. Joly, A., Lombardo, J. C., Champ, J., & Saloma, A. (2016). Unsupervised individual whales identification: spot the difference in the ocean. In Working Notes of CLEF 2016-Conference and Labs of the Evaluation forum (pp. 469-480).
5. Kuettel, D., Guillaumin M., Ferrari V.: ImageNet Auto-annotation with Segmentation Propagation, Technical Report, *International Journal of Computer Vision*, 2013.
6. Kuettel, D., Guillaumin, M., Ferrari, V.: Segmentation Propagation in ImageNet *European*

Conference on Computer Vision (ECCV), Firenze, Italy, October 2012.

7. Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision*, 60, 2, pp. 91-110, (2004)

8. Lowe, D.: Object recognition from local scale-invariant features. In: *ICCV* (1999)

9. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors*,* International *Journal on Computer Vision* 60(1), 2004, pp. 63-86.

10. Papp, D., Lovas, D., Szűcs, G. (2016). Object Detection, Classification, Tracking and individual Recognition for Sea Images and Videos, Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum, Évora, Portugal, 5-8 September, 2016, CEUR Workshop Proceedings, ISSN 1613-0073, Vol-1609, pp. 525-533.

11. Perronnin, F., Dance, C.: Fisher kernel on visual vocabularies for image categorization, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, (2007)

12. Reynolds D. A.: Gaussian Mixture Models, *Encyclopedia of Biometric Recognition*, Springer, February, pp. 659-663 (2009)

13. SeaCLEF 2017 http://www.imageclef.org/lifeclef/2017/sea (LifeCLEF 2017), *CLEF working notes 2017*

14. Szűcs, G., Papp, D., Lovas, D., SVM classification of moving objects tracked by Kalman filter and Hungarian method, In: L Cappellato, N Ferro, G J F Jones, E S Juan (eds.) *Working Notes of CLEF 2015 Conference*, Toulouse, France, September 8-11, 2015, Paper 40. 10 p. Vol. 1391.

15. Szűcs, G., Papp, D., Lovas, D., Viewpoints Combined Classification Method in Image-based Plant Identification Task In: Cappellato L., Ferro N., Halvey M., Kraaij W. (eds) *Working Notes for CLEF 2014 Conference*. Sheffield, Great Britain, September 15-18, 2014., pp. 763-770. Vol.1180.

16. Van Dongen, S. (2000). A cluster algorithm for graphs. Report-Information systems, (10), 1-40.