

Automatic Whale Matching System using Feature Descriptor

S. M. Jaisakthi¹, P. Mirunalini², Rutuja Jadhav¹, and Vatsala¹

¹ School of Computer Science and Engineering,
VIT University, Vellore, India

{jaisakthi.murugaiyan, jadhav.jitendra2014,
vatsala2014}@vit.ac.in

² Department of Computer Science and Engineering, SSN College of Engineering,
Kalavakkam, Chennai, India
miruna@ssn.edu.in

Abstract. Whales play an important role in ocean ecosystem by maintaining a stable food chain. Whales continued to be endangered animals though they don't have direct predators. To ensure the survival of this endangered species, marine biologists are tracking them to know their status and health of the species at all times. Manual recognition of whale is most tricky and hence automated system helps biologist to develop conservation strategies for different species of whale. This can be primarily achieved through individual whale recognition and tracking their behavior by analyzing the data collected. In this paper we have proposed a method for finding the matching pairs of whales by analyzing the caudal fin images of whales in the data set. For finding the matching pairs we have segmented the caudal fins from the background images using GrabCut and FloodFill algorithm. From the segmented images key features are extracted using Scale Invariant Feature Transform (SIFT) and key features are matched using the FLANN (Fast Library for Approximate Nearest Neighbors) matcher. Finally the matched pairs are ranked using the confidence values.

Keywords: Whale, Matching, Segmentation, Recognition, GrabCut, FloodFill, SIFT features

1 Introduction

To effectively manage and conserve animals, it is important to understand and predict the movement of animals. Recent study shows that the movement and behaviors of highly migratory marine species is challenging because of logistical and technological limitations. However automatic recognition and tracking tools significantly helps the scientist to study the movements and behavior of the animals. Automatic recognition of individual whales plays a crucial role for tracking target population, biological samples, acoustic recordings, behavior patterns, migration, sexual maturity and health assessments. To track and monitor the population, whales are photographed during aerial surveys and then manually matched to an online photo-identification catalog. The development of the photo-identification has been proven to be a useful tool for learning

many species. Since the process is done manually it is impossible to get accurate count of all individuals in a given large collection of observations [9]. Automating the photo-identification process could speed-up and provide better result in large data collection of data. This can be done by comparing the features extracted between the whales. Generally whales have four fins : two pectoral fins, a caudal fin and a dorsal fin. The caudal fin is meant for propulsion of animals and the pectoral fins act as rudders and stabilizers. Caudal fins is the powerful fins with different shapes. The different shapes in caudal fins, patterns of coloration and patterns of marks designed by nature over the caudal fins can play a valuable contribution in recognition and tracking. So, analysis of caudal fins plays an important role in studying whales. Scientists compare and match the image taken, with images in database and locate known whales during their migration.

2 Tasks and Dataset

The report in this research work is a challenge in SeaClef 2017 which is a part of LifeCLEF 2017 [6]. This task aims in automatically matching whale image pairs, over a large set of images, of same individual whales through the analysis of their caudal fins. The caudal fins are considered as the most discriminant pattern for distinguishing an individual whale from another. The data set shared for the task contains 2005 images of humpback whales caudals.

Individual whales can be recognized by their pattern of markings and/or the scars that appear along the years [7]. It is very difficult to match the whales in whole data set because number of individual whales is high, features to discriminate the whales are very similar and similar water background.

3 Proposed Methodology

To automatically detect the matching pairs of whale from the given data set we have first preprocessed the images to remove the noise like clouds, sea water and landscapes. From the noise removed images we have extracted the scale invariant key features and these extracted features are matched to obtain the matching pairs of whales. The steps involved in automatic whale matching system are :

- Preprocessing using Grabcut Segmentation and Flood fill algorithm
- Key Feature Extraction using SIFT Algorithm
- Feature Matching using FLANN Matcher
- Ranking the Matched Pairs

The overall architecture of the proposed methodology is depicted in Figure 1.

3.1 Preprocessing

The data set contains images of caudal fins with some background details such as sea water, clouds, land cover, rock etc. The presence of these background details may reduce the performance of the system considerably. So we have preprocessed the images in the data set to segment the caudal fins alone from the background using GrabCut and FloodFill algorithm.

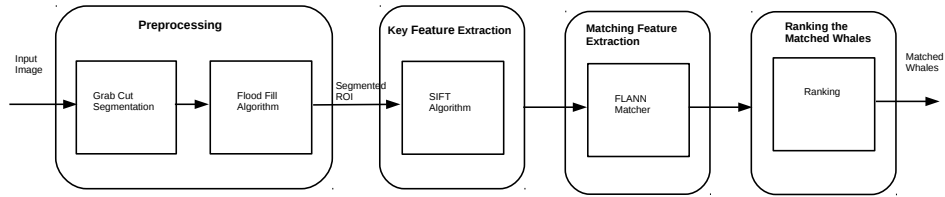


Fig. 1. Architecture Diagram of the Proposed Whale matching System

ROI segmentation using Grabcut method GrabCut algorithm [1] efficiently segments the foreground object from the background images using graph cut method. This algorithm takes foreground information in the form of rectangle and the information which lies outside the rectangle is considered as background. Using this foreground and background information Gaussian Mixture Models (GMM) are modeled. Each pixel in the foreground is assigned the most probable Gaussian component in the foreground GMMs. Similarly each pixel in the background is assigned the most probable Gaussian component in the background GMMs. From this pixels a graph is built with pixel being represented as nodes. Two more additional nodes Source node and Sink node are added where every foreground pixel is connected to Source node and every background pixel is connected to Sink node. The weights for edges between source and end node is defined based on the probability of a pixel being foreground/background. The weights between the pixels are defined by the edge information or pixel similarity. Using the mincut algorithm graph is segmented. After the cut, all the pixels connected to Source node become foreground and those connected to Sink node become background. The same process is repeated until the foreground image is segmented.

In this work, we have used GrabCut segmentation to segment the caudal fins from the background. In order to automatically obtain the foreground information in the form of rectangle we have divided the image into grid like structure by dividing rows and columns of the image into 20 equal parts. Sample grid image is shown in Figure 2. From this grid like structure we have eliminated two topmost row strip and one bottom most row strip. Similarly one left most and right most column strip is eliminated. After eliminating some of the rows and columns the grids in the middle is formed into single rectangle and this rectangle is given as input foreground rectangle. The result of the GrabCut segmentation for two sample images are depicted in Figure 3.

For the first image foreground is perfectly segmented while in the second image foreground is not segmented properly. These images need further segmentation which is done using FloodFill algorithm.

Improving Segmentation Results using FloodFill Algorithm GrabCut method fails to give good segmentation results for some images. The segmentation process is further improved using floodFill algorithm which clusters same colored pixels. To obtain a seed point automatically for clustering, blue colored pixels are searched along the column stripes and row stripes. Once the seed points are selected FloodFill algorithm groups

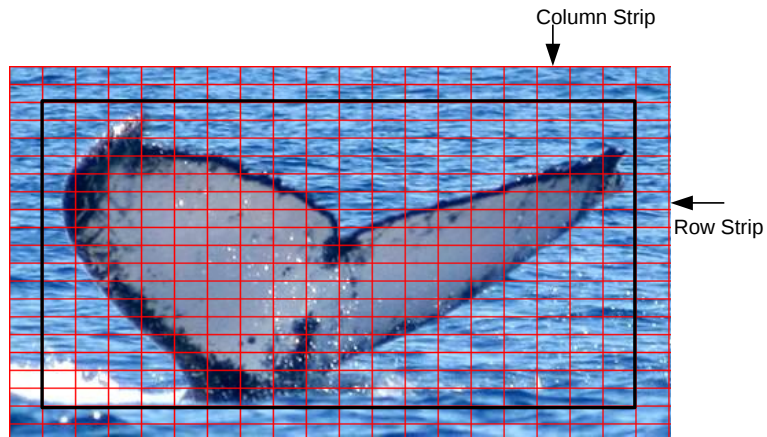


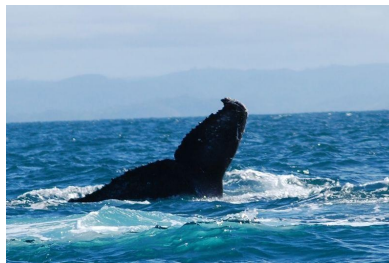
Fig. 2. Extraction of Foreground Rectangle Input for GrabCut Algorithm



Input Image : 18.jpg



Result of Grabcut Segmentation



Input Image : 995.jpg



Result of Grabcut Segmentation

Fig. 3. Result of Grabcut Segmentation

pixels with same color and these grouped pixels are then segmented out. This algorithm ensures a better segmentation of caudal fins by removing the noise present after GrabCut segmentation. The result of applying GrabCut combined with FloodFill algorithm is shown in Figure 4.

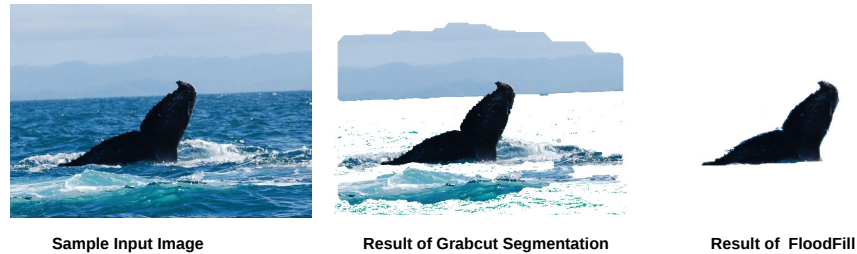


Fig. 4. Result of GrabCut Segmentation combined with Floodfill Algorithm

3.2 Key Feature Extraction using SIFT algorithm

After obtaining the ROI, whale recognition has been done by extracting the key features from the ROI and comparing them with the features of all other images. The key features of the whale image are obtained using SIFT [2] feature extraction algorithm since, SIFT features are scale and rotational invariants and also invariant to change in illumination. The descriptors obtained rely on local pixel information which makes them robust to minor occlusions. The following steps are involved in generating the key features of the image.

- Scale-Space Construction : In this step several octaves (images of same size) of the original images is generated. The size of image in each octave is reduced to half the size of previous octave's image. The images within an octave is blurred out progressively.
- Scale-space extrema detection: In this step, potential interest points that are invariant to scale and orientation are identified by estimating Difference-of-Gaussian (DoG) over all scales and image locations.
- Key point Localization: To find the key points first maxima/minima is located in DoG images and the subpixel maxima/minima is determined. This is achieved by comparing neighbouring pixels in the current scale, the scale "above" and the scale "below".
- Orientation assignment: One or more orientations are assigned to each key point location based on local image gradient directions.
- Key point descriptor: The local image gradients are measured at the selected scale in the region around each key point. These are transformed into a representation

called descriptors which encompasses position, scale, orientation and local image structure.

The descriptors are called as key features which are extracted from the ROI of the data set. To find the matching pairs, the extracted features of an image are compared with the features obtained from all other images in the data set. After feature extraction, feature matching is done using FLANN matcher. The sample image after key point extraction using SIFT algorithm is depicted in the Figure 5.

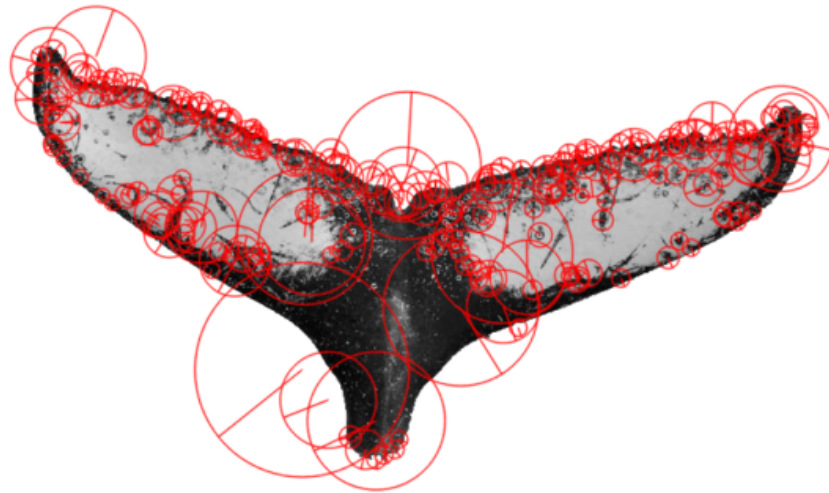


Fig. 5. Extracted Key Points for image 14.jpg

3.3 Feature Matching using FLANN Matcher

The extracted key features from all the images are matched using FLANN matcher [8]. This is an optimized algorithm, does a fast local approximate nearest neighbor calculation between two sets of feature vectors of high dimensional features. FLANN uses the Hierarchical K-means tree for generic feature matching and uses a priority-queue (Best-Bin-First) to find approximate nearest neighbors from Hierarchical K-Means tree. It compares each image with their neighbors based on the ratio of their distance, smaller the ratio, higher the distance with respect to the neighbors. Figure 6 depicts key point matching for image 14.jpg and 9.jpg.

3.4 Ranking the Matched Pairs

Matching Whales To find the best match, each image in the data set is matched with all other images in the data set using Flann matcher. While matching the whales, there may

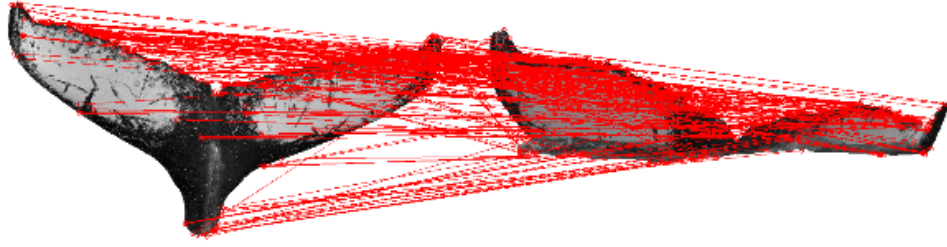


Fig. 6. Matching Key Points for images 14.jpg and 9.jpg

be duplicate pairs. In order to remove the duplicate pairs we have constructed matrix of size $N \times N$, where N represents the number of images in the data set. The value v in the location (i,j) represents the number of good matches that matches image i and image j . The duplicate matching of the images are eliminated by considering only the values in lower triangular matrix and discarding the upper triangular matrix. Higher the number of good matches represents high possibility of matching pairs.

Confidence Value Calculation The confidence value for the matching pairs should be in the interval of 0 to 1. To estimate the confidence value we have used min-max normalization method which is given by

$$v' = \frac{v - \min}{\max - \min}(new_max - new_min) + new_min \quad (1)$$

Min-max normalization transforms the obtained good matches v of an image to a new interval from new_min to new_max using min and max values of good matches.

Ranking the Whales Pairs The matched pairs are identified by fixing the threshold t for the confidence value. Generally the threshold is fixed as 95%. The pairs which is having confidence value more than the threshold are filtered out and then ranking is performed by sorting the confidence values.

4 Resources

We have used the data set released by the SeaClef 2017 for whale matching and for implementation we have used Opencv in c++ environment.

5 Results and Discussion

Using the proposed method we could able to obtain 0.01% average precision only. The accuracy can be further improved by using other matcher such as Brutforce and knn-matcher. The results can be further enhanced by obtaining best estimates of the matches using feature correspondence using homography technique.

6 Conclusion

In this paper we have proposed an automatic method to match the whales in a large data set which is the picture of caudal fins. Our method first segments the whale image from the background using grabcut segmentation. From the segmented foreground images key points are obtained using SIFT descriptors. subsequently, using FLANN matcher key features of an image is matched with the key features of all other images in data set and good matches are obtained. Using the obtained good matches confidence value is calculated using min-max normalization and the estimated confidence value is ranked.

References

1. Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. "Grabcut: Interactive foreground extraction using iterated graph cuts." *ACM transactions on graphics (TOG)*. Vol. 23. No. 3. ACM, 2004.
2. David G. Lowe. "Distinctive Image Features from Scale-Invariant Key-points", *Int. J. Comput. Vision* 60, 2 (November 2004), 91-110. DOI:<http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
3. <https://www.challenge.gov/challenge/right-whale-recognition-challenge/>
4. Malmer, Tomas. "Image segmentation using GrabCut." *IEEE Transactions on Signal Processing* 5.1 (2010): 1-7.
5. Tatiraju, Suman, and Avi Mehta. "Image Segmentation using k-means clustering, EM and Normalized Cuts." University Of California Irvine (2008).
6. Joly, Alexis and Goëau, Hervé and Glotin, Hervé and Spampinato, Concetto and Bonnet, Pierre and Vellinga , Willem-Pier and Lombardo, Jean-Christophe and Planqué, Robert and Palazzo, Simone and Müller, Henning, 2017,LifeCLEF 2017 Lab Overview: multimedia species identification challenges,Proceedings of CLEF 2017
7. Alexis Joly, Jean-Christophe Lombardo, Julien Champ, Anjara Saloma: Unsupervised Individual Whales Identification: Spot the Difference in the Ocean. CLEF (Working Notes) 2016: 469-480
8. Marius Muja and David G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration", in *International Conference on Computer Vision Theory and Applications (VISAPP'09)*, 2009
9. Papp, David, Daniel Lovas, and Gabor Szucs. "Object Detection, Classification, Tracking and individual Recognition for Sea Images and Videos." CLEF (Working Notes). 2016.
10. <http://us.whales.org/whales-and-dolphins/whales>