Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-23 (2016) 470–475

Hand gesture recognition from multibeam sonar imagery ⋆

Franka Gustin * **Ivor Rendulic** * **Nikola Miskovic** * **Zoran Vukic** *

* *University of Zagreb, Faculty of Electrical Engineering and Computing (e-mail: gustin.franka@gmail.com, ivor.rendulic@fer.hr, nikola.miskovic@fer.hr, zoran.vukic@fer.hr).*

Abstract: Divers perform demanding tasks in a complex and hazardous underwater environment, which prevents them from carrying special devices that may allow them to communicate with their robotic diving buddies. In this world of natural human–robot interaction in the underwater environment, envisioned by the FP7 Cognitive Robotics project CADDY, hand detection and gesture interpretation is a prerequisite. While hand gesture recognition is most often performed with cameras (mono and stereo), their use in the underwater environment is compromised due to water turbidity and lack of sunlight at greater depths. This paper deals with this lack of performance by introducing the concept of using high resolution multibeam sonars (often referred to as acoustic cameras) for diver hand gesture recognition. In order to ensure reliable communication between the diver and the robot, it is of great importance that the classification precision is as high as possible. This paper presents results of hand gesture recognition which is performed by using two approaches: convex hull method and the support vector machine (SVM). A novel approach that fuses the two methods is introduced as a way of increasing the precision of classification. The results obtained on more than 1000 real sonar samples show that the precision using the convex hull method is around 92%, and using the SVM around 94%, while fusing the two approaches provides around 99% classification precision.

Keywords: Multibeam sonar, gesture recognition.

1. INTRODUCTION

Exploring the underwater environment requires use of autonomous marine (surface and underwater) robots mostly due to the fact that oceans and seas present a hazardous environment for people. Nowadays, unmanned, autonomous underwater vehicles are commonly used in applications such as seabed mapping, detection of hazardous oil spills Vasilijevic et al. (2015), inspection of ships and underwater infrastructure Ridao et al. (2010), mine countermeasures Miskovic et al. (2011), etc. Only recently have some advances in unmanned underwater robotic manipulation been made, mainly in applications related to offshore industry Ahmadzadeh et al. (2013), and search and rescue scenarios (e.g. black-box recovery) Prats et al. (2012). There is a large number of tasks underwater that still have to be executed by experienced divers.

In order to aid divers during their activities in the underwater environment, the FP7 CADDY project envisions the development of an underwater robotic diving buddy, that will be capable of performing three basic functionalities: "buddy guide" that leads the diver through the underwater environment; "buddy observer" that monitors the diver during the dive and tries to perform early detection of any catastrophic behaviour; and "buddy slave" that helps the diver perform tedious tasks such as making a mosaic of a designated area, illuminating an area, etc. The prerequisite for the diver–robot interaction scenario, is to keep the natural way of communication between two divers – using the diving sign language. The buddy vehicle can interpret those signs either by using mono camera, stereo camera or a multibeam sonar mounted on the vehicle. The biggest downside to using cameras (mono or stereo) lies in the fact that visibility underwater is often compromised due to increased turbidity. On the other hand, multibeam sonars are not influenced by turbidity, at the expense of lower resolution.

Using mono and stereo cameras on dry land to determine hand gestures has already been done in applications related to easier human-computer interaction, deaf language interpretation, etc. Using these methods in the underwater environment introduces additional problems related to ego-motion compensation and segmentation of the hand relative to the dynamic background Armesto et al. (2007). However, using high resolution multibeam sonars for the detection of diver hands has not yet been done before.

In this paper, we present the use of high resolution multibeam sonar imagery for the detection of diver hands in the underwater environment. First, we deal with the issue of detecting the diver hand from the sonar imagery. Second, we use already existing image processing algorithms based on the convex hull method and the support vector machines for detection of the gesture (from a set of predefined gestures). To increase recognition and classification precision of the system, we present a method that fuses the results obtained by the convex hull method and the support vector machine based approach. The proposed approach results in an increased robustness of diver symbol interpretation.

⋆ This work has been done within the scope of CADDY, a collaborative project funded by the European Community's Seventh Framework Programme FP7 - Challenge 2: Cognitive Systems and Robotics - under grant agreement 611373.

2405-8963
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.10.450

The paper is organized as follows. Section II describes the hand detection algorithm which is based on the cascade of boosted Haar classifiers. In section III, gesture recognition algorithms are described, while Section IV describes the experimental setup and multibeam sonar used. Section V describes the results of hand gesture recognition, together with the results of the proposed method for increasing the robustness of the recognition process. The paper is concluded with Section VI.

2. HAND DETECTION

Hand gesture recognition is an area of active research, with many possible applications Liang and Ouhyoung (1998), Malima et al. (2006), Brethes et al. (2004). Most approaches for hand detection are based on skin color segmentation Kang et al. (2008), Rasim and Alexander (2013), which is not applicable on sonar images, due to the fact that sonar is not providing color information. In this paper, hand recognition is implemented using the cascade of boosted Haar classifiers. The main goal is to detect different hand gestures and extract them from the original image.

The first step in the gesture recognition framework (shown in figure 8) is hand detection. Haar cascade method is proposed by Viola and Jones for fast and robust object detection using boosted classifiers based on Haar-like features Viola and Jones (2001). The algorithm is based on several key concepts:

Haar-like features Haar-like features are simple rectangular features that can be calculated as the difference between the sum of pixel intensities in rectangular areas at any position and scale within the original image. Features can indicate certain characteristics in the image, such as edges or changes in texture. On the detection window with resolution 24×24, there are over 160,000 rectangular features.

Integral image Rectangular features can be rapidly calculated using intermediate image representation called the integral image. The integral image at location x, y is defined as the sum of the pixels above and left of x, y:

$$ii(x,y) = \sum_{x' \leq x, y' \leq y} i(x',y'), \tag{1}$$

where $i(x,y)$ is the original image and $ii(x,y)$ is the integral image. Integral image can be computed in only one pass over the original image using the following expressions:

$$s(x,y) = s(x,y-1) + i(x,y) \tag{2}$$

$$ii(x,y) = ii(x-1,y) + s(x,y) \tag{3}$$

where $s(x,y)$ is the cumulative row sum, $s(x,-1) = 0$ and $ii(-1,y) = 0$. Integral image enables to calculate any rectangular sum in only four array references.

AdaBoost Each Haar feature can represent one weak classifier (classifier with high false detection). Although features can be computed very efficiently, computing the whole set can be time consuming process. The task of AdaBoost learning algorithm is to choose small set of features which best separate positive and negative examples, and combine them to form an effective classifier.

Cascade Cascade of classifiers is used to achieve better performance while radically reducing computation time. Classifiers are trained using the Adaboost algorithm and arranged into a cascade by complexity. The overall detection process is similar to a decision tree. The first classifier evaluates all the sub-windows, while each following only those examples that weren't eliminated by his predecessor. If the sub-window is eliminated in some level of the cascade, it is no longer considered in the following stages. Large number of negative sub-windows is rejected in the lower levels of the cascade when simpler classifiers are used, while the more complex processing takes place only on a small number of remaining sub-windows.

Classifier is trained using previously described Viola and Jones algorithm. Learning process is based on positive and negative images. Positive images are images that contain the object of interest, i.e. hand gestures in different forms (Figure 1). Negative or background images do not contain any object of interest, in this case they are images of diver or underwater surroundings (Figure 2). The number of images used for training process are: 1406 positive images and 3357 negative images. Images with different hand gestures are obtained in laboratory environment, while diver and background images are conducted on field testings. After the training stage, the cascade classifier of totally 21 stages is generated. In the detection stage classifier is applied on the input image using sliding window of fixed size (24×24). In order to detect objects of different sizes, the image size is reduced at each image scale. Parameter specifying reduction size is called the scale factor. In our tests, the hand was always at approximately the same distance from the sonar (around 1m). The reason is that not only the hand gets smaller with increasing distance, but the details are rapidly lost. That is why the gestures were performed only at a close range. Examples of detection are shown in Figure 3. In case of positive detection, object is marked and extracted from the original image.

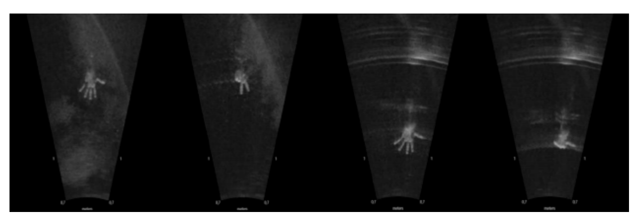

Fig. 1. Positive training images

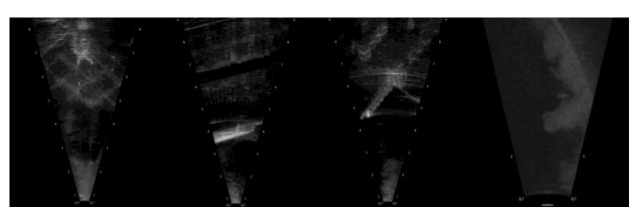

Fig. 2. Negative training images

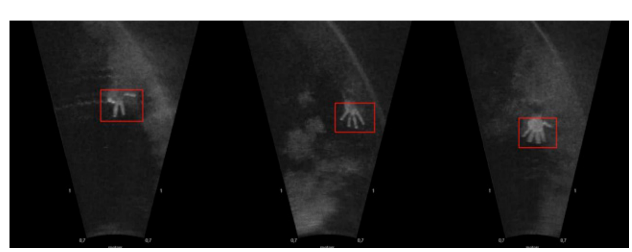

Fig. 3. Examples of hand detection