The goal of this section is to obtain a groundtruth annotation for fish recognition. More specifically, we want to know which fish images belong to the same species, along with the species names. This allows us to train and evaluate our fish recognition methods in the F4K project. In order to support the manual labelling of images, we propose to use an automatic clustering method to groups and retrieve similar images, which allows us to label large dataset in an efficient manner.

This section is organised as follows: We start with a discussion of the fish clustering method that is used to support the labelling task. We then explain how we combine the clustering method with the annotation interfaces to support manual annotation work. Two types of annotation interfaces are defined: i) the interface for normal annotators to label fish images; and ii) the interface for marine biologists to verify the obtained labels. We close this section with a discussion of the quality of the labels obtained.

I changed the subsection titles here, the original one was “First method to determine the similarity between fish”. It’s confusing: why is it a “first” method? as there doesn’t seem to have a “second” method ....

Fish Clustering

Measuring similarity between fish images The fish clustering method starts with the assumption that the segmentation of the fish is correctly performed. In the case that we do not have groundtruth segmentation data, we can use visual inspection to manual remove failures in the segmentation from the dataset, this migt however be very time consuming.Do you mean manually remove the wrong segmentations? What’s the relation of this to the similarity measure?

It is confusing here whether we are talking about similarity measures/fish representations or clustering method.I thought we are talking about similarity measures/fish representations... In order to compare fish, we need a method which is also able to compare unseen fish species with already know species in the dataset. The similarity method needs to be invariant against a lot of variantions because of the uncontrolled nature of the video recording. The last requirement on the method is that it must be able to deal with objects which are quite similar, as opposed to most methods in image retrieval where the classes with are quite dissimilar (car, building, people, etc). To compute the similarity between images, we use a method that is very similar to the method described in goldberger$_unsupervised_2006$.

The feature used for fish recognition are the color of the fish, the texture of the fish and the fish contour. In the case of color, we transform all the pixel values of the segmented fish to HSV (Hue, Saturtion, Value), for the Hue channel two values are used, namely the sine and cosine of the Hue channel. This removes the big difference between the different red values because of the cylindrical color definition. Using all the 4 dimensional pixel values $X = \{x_1, \dots, x_N\}$ in the segmented image, we fit a Gaussian Mixture Model (GMM) $f(x)$, using the Expectation Maximalization algorithm described in zivkovic$_recursive_2004, whichusesMinimumDescriptionLenghttoautomaticallydeterminethenumberofGuassiandensit$ is determined by using a Monte-Carol simulation to approximate the Kullback-Liebler divergence:

equation D(f$_1\|f_2) = \int f_1 \log f_1 f_2 \approx 1N \sum_{t=1}^{N} \log f_1(x_t) f_2(x_t)$