

Advanced Vision Assignment 2 Report

Mindaugas Dabulskis, Marat Subkhankulov

26th February 2013

1 Introduction

This report describes the work carried out for the second assignment of the AV course. The aim of the assignment was to track three coloured balls through a set of video frames. The algorithm for ball detection and segmentation is discussed. The performance of the approach is evaluated using a gold standard.

2 Algorithm and implementation

In order to detect and track the balls we used the following algorithm:

1. Apply mask 1 - Average background subtraction
2. Apply mask 2 - Clothing mask based on intensity
3. apply mask 3 - Hand skin mask based on saturation
4. For each ball classify ball pixels based on a training sample
5. Clean up and track the centroids - store the centroid in a matrix

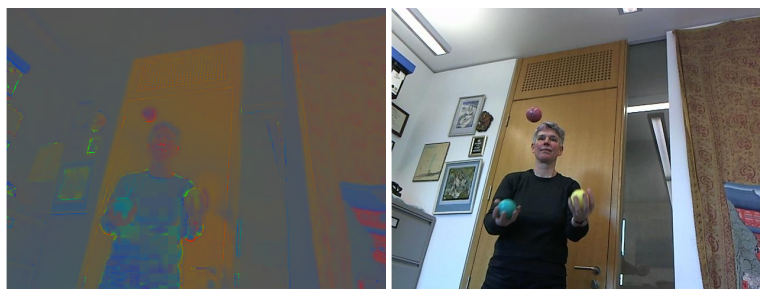


Figure 1: Training sample for each ball

2.1 Mask 1: Average background subtraction

A picture of the empty room had been given, however when subtraction of the background from the subject image did not eliminate the shadow of the door, which had similar values to the yellow ball even in normalized RGB. Thus we obtained a mean of all the frames and used the resulting image as the background; By converting the subject and the new background images to N-RGB, subtracting and thresholding out the low brightness regions we obtained a mask which eliminated the static background, the juggler's, the juggler's face.

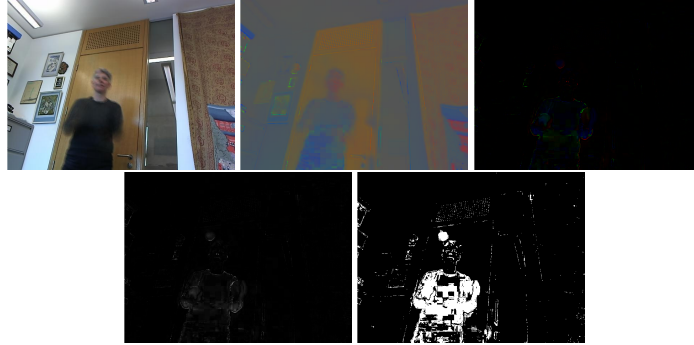


Figure 2: Training sample for each ball

2.2 Mask 2: Clothing mask

Mask 1 did not remove the clothing of the juggler due to the smoothing effect of averaging; This posed a problem as the clothes had similar chromaticity to the green ball. We took advantage of the clothes' dark colour to easily threshold out the clothing using a manually set intensity threshold.

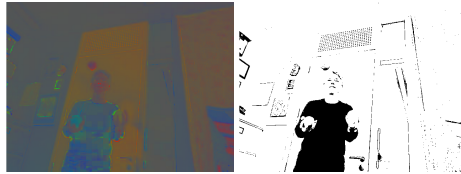


Figure 3: Training sample for each ball

2.3 Mask 3: Hands

It seemed reasonable to use N-RGB for segmenting out the balls as the lighting effects were largely removed, however the hands of the juggler were not removed by the previous masks and had similar chromaticity to the red ball. The hands were masked by taking advantage of the skin's unique saturation value.



Figure 4: Training sample for each ball

2.4 Classification of ball pixels

Having applied all of the masks - the background was largely removed. Classification could now be performed. Conversion to N-RGB eliminated diffuse and specular lighting effects on the balls, but necessitated thresholding on all three channels; Since the intensity of all the pixels was largely the same in N-RGB, it was clear that the majority of variation lies in the hue channel - using this single value simplified thresholding as only one channel could be used to successfully classify the ball pixels.

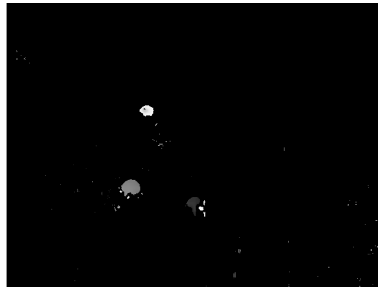


Figure 5: Training sample for each ball

Threshold values for hue were obtained from a training sample. The training sample obtained by sampling the pixel at the centroid of each ball given by the ground truth. The lower and upper thresholds were taken at two standard deviations from the median of the sample - this was done to ignore outliers in the sample.

2.5 Clean up and tracking the centroids

Final clean up and ball identification was made using matlab **bwlabel** and **regionprops** functions. The function **bwlabel** takes our intermediate file, which we got after applying all the masks, and labels all the connected objects in the file. Next, this file is supplied to **regionprops** function, which measures and returns set of the properties for each connected object in the file. Three properties we used:

1. *Area* - number of actual pixels in the region,
2. *PixelIdxList* - vector containing the linear indices of the pixels in region,
3. *PixelList* - matrix specifying the locations of pixels in the region.

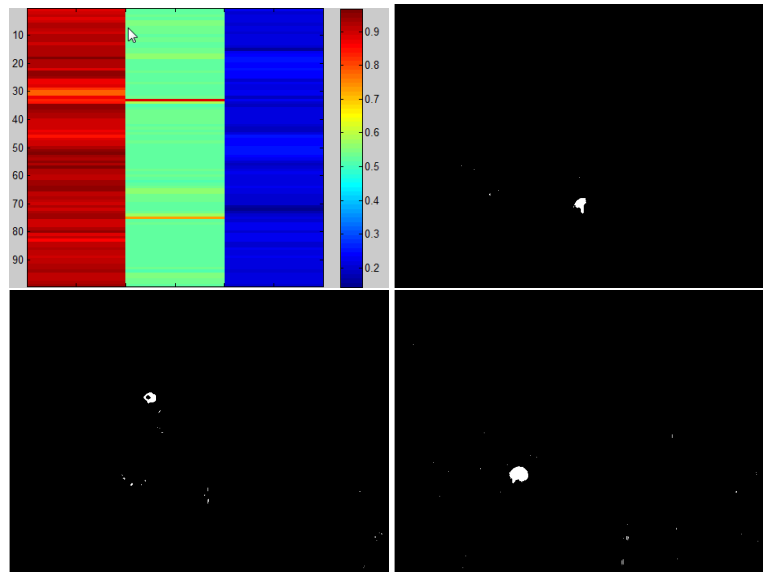


Figure 6: Training sample for each ball

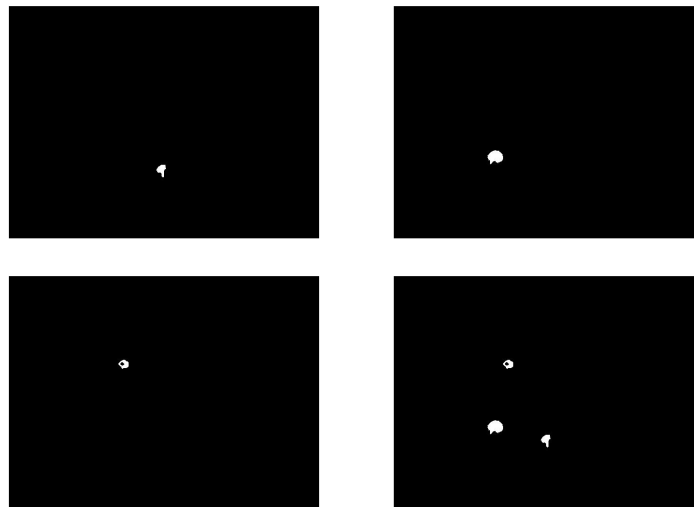


Figure 7: Cleaned images

The idea was:

1. To find maximum connected object in the file by the pixels area,
2. Set all other connected object areas pixel to 0 (delete these areas),
3. Find the middle point of the largest connected objects area.

Second point cleaned the image of all the noise, because the object with biggest area was the ball itself. Third point was needed for evaluation, to evaluate how much

our detected ball differs from ground truth.

3 Experimental results

In order to evaluate the algorithm, Euclidean distance between the centroids calculated by the algorithm and the ground truth was measured; A threshold of 10px was set to discriminate between successful and failed detections. In order to qualitatively assess the performance of the algorithm and screen for any outliers the ball trajectories were rendered. In addition we also wanted to find the most appropriate method for calculating the centroid of the detected balls. There were three approaches to calculating the centeroid of the ball, therefore three result sets are presented. These were:

1. Mean of the area pixels,
2. Median of the area pixels,
3. Maximum and minimum pixels in **y** and **x** coordinates and find the mean between them. This creates a bounding box of the area and finds it centre.



Figure 8: Image in which the distance between ground truth (+) and found mass centre (.) is bigger than 10 pixels

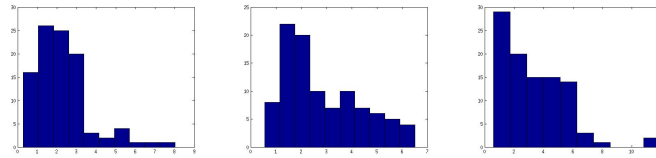


Figure 9: Cleaned images

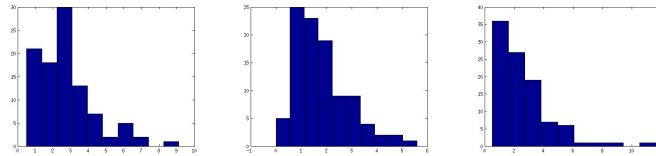


Figure 10: Cleaned images

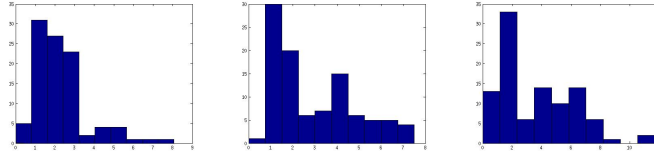


Figure 11: Cleaned images

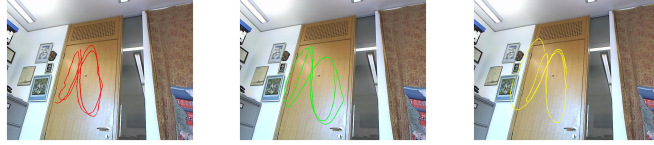


Figure 12: Cleaned images

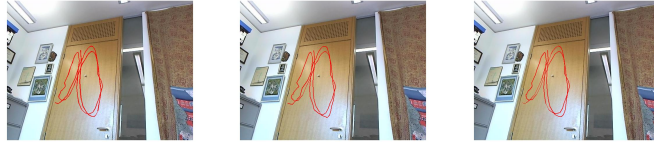


Figure 13: Cleaned images

Table 1: *Mean*

| | Red | Green | Yellow |
|--|----------|----------|----------|
| ball found in images | 98/98 | 98/98 | 98/98 |
| mass centre is less than 10 pixels from ground truth | 98/98 | 98/98 | 96/98 |
| maximum error | 8.0093 | 6.5060 | 11.9423 |
| minimum error | 0.2799 | 0.5622 | 0.5995 |
| one ball error | 224.0428 | 276.9718 | 328.6480 |
| total error | 829.6625 | | |

Table 2: *Bound*

| | Red | Green | Yellow |
|--|-----------|----------|----------|
| ball found in images | 98/98 | 98/98 | 98/98 |
| mass centre is less than 10 pixels from ground truth | 98/98 | 98/98 | 97/98 |
| maximum error | 9.1241 | 5.5902 | 11.6297 |
| minimum error | 0.5000 20 | 0 | 0.5000 |
| one ball error | 264.7848 | 189.4427 | 258.1291 |
| total error | 712.3566 | | |

4 Discussion

Qualitative assessment showed 100

From statistics we can see that bounding box approach has the smallest total error

Table 3: *Median*

| | Red | Green | Yellow |
|--|----------|----------|----------|
| ball found in images | 98/98 | 98/98 | 98/98 |
| mass centre is less than 10 pixels from ground truth | 98/98 | 98/98 | 96/98 |
| maximum error | 8.0623 | 7.5000 | 11.7047 |
| minimum error | 0 | 0 | 0 |
| one ball error | 227.6602 | 298.2528 | 363.9370 |
| total error | 889.8500 | | |

and has only one image in which the centre deviates above threshold, compared to other approaches, which have have two such cases. Using the centre of the bounding box performs well in cases where the ball can be seen through the grasping fingers, so the left, right, top and bottom edges are visible and an accurate centre can be found; However approach is highly sensitive to outliers as any random point can offset it enourmously.

If we look into individual ball error, we notice that with bounding box approach red ball has the largest error compared with other approaches. This is probably caused by the similarity of the hands and the red ball, which causes misclassifications. That is why in mean and median approach the red balls error is smallest, as the outliers are largely ignored; Because of this we decided not to use the bounding box approach.

4.1 References

References should be numbered in order of appearance, for example [1], [2], and [3]. You *can* use `bibtex` to prepare references, or do it by hand if there are very few.

References

- [1] Smith, J. O. and Abel, J. S., “Bark and ERB Bilinear Transforms”, IEEE Trans. Speech and Audio Proc., 7(6):697–708, 1999.
- [2] Lee, K.-F., Automatic Speech Recognition: The Development of the SPHINX SYSTEM, Kluwer Academic Publishers, Boston, 1989.
- [3] Rudnick, A. I., Polifroni, Thayer, E. H., and Brennan, R. A. ”Interactive problem solving with speech”, J. Acoust. Soc. Amer., Vol. 84, 1988, p S213(A).