Advanced Vision Assignment 2 Report

Mindaugas Dabulskis, Marat Subkhankulov

26th February 2013

1 Introduction

This report describes the work done 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 segment out the balls from the subject image (the current frame being considered), we removed background regions, and thresholded using values obtained from training data.

The following regions had to be removed from the image:

1. Static background
2. Juggler's clothing
3. Juggler's hands
4. Juggler's face
5. Shadow caused by the juggler on the door

To remove the above regions the following techniques were applied:

1. Mask 1: Average background subtractionand value thresholding
2. Mask 2: Clothing mask based on intensity
3. Mask 3: Hand skin mask based on saturation
4. Classification based on a training sample

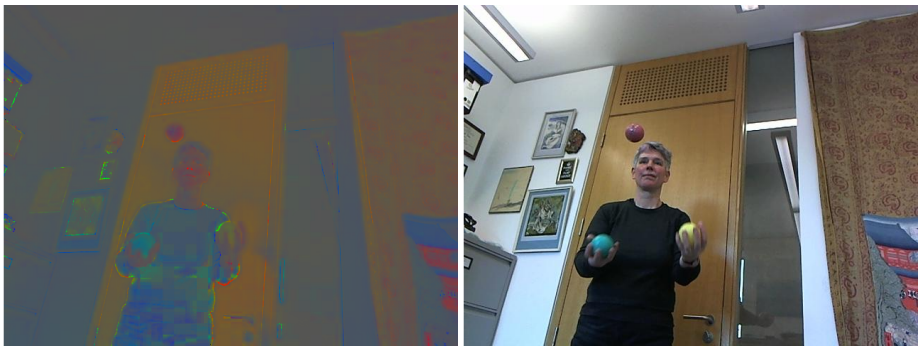

Figure 1: Training sample for each ball

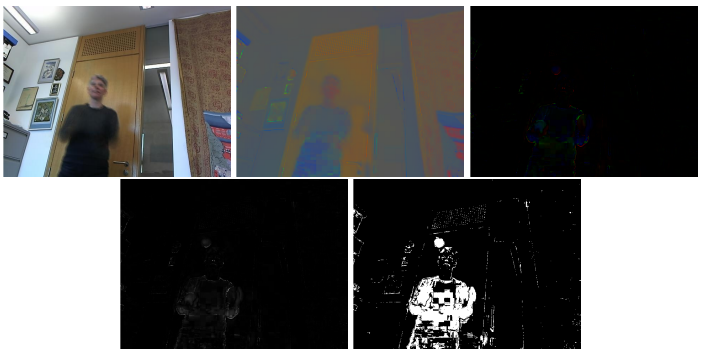

Figure 2: Training sample for each ball

2.1 Mask 1: Average background

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:

1. Static background
2. Shadow on the door
3. Juggler's face

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 theshold out the clothing using a manual threshold.

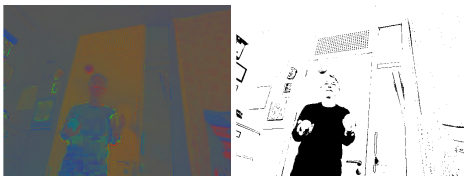

Figure 3: Training sample for each ball

Figure 4: 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 lare largely removed, however the hands of the juggler were not removed by the previous masks and had similar chromaticity to the red ball. We removed the hands by taking advantage of the skin's unique saturation value.

2.4 Segmentation

Having applied all of the masks the background was largely removed. Thresholding 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 after normalization, it was clear that the majority of variation lies in the hue of the spheres - using this single value simplified thresholding as only one channel could be used to successfully classify the ball pixels.

Threshold values for hue were obtained from a training sample. The training sample obtained by sampling the pixel as the centroid of each ball given by the ground truth. The lower and upper thresholds were 2 standard deviations from the median of the sample - this was done to ignore outliers.

3 Discussion and Conclusion

3.1 Formatting: tables

An example of a table is shown as Table 1. Somewhat different styles are allowed according to the type and purpose of the table.

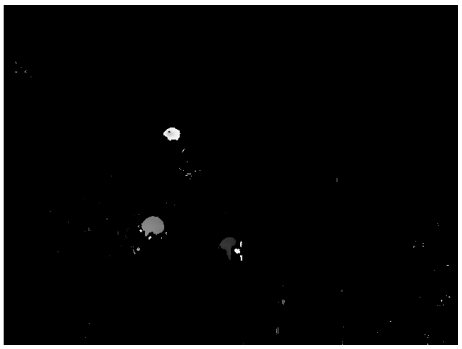

Figure 5: Training sample for each ball

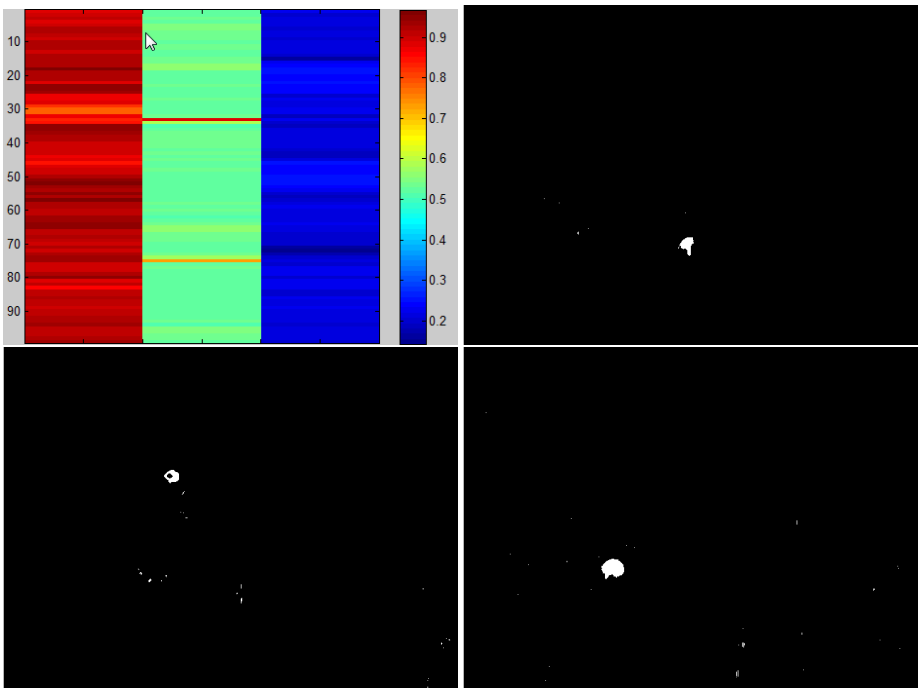

Figure 6: Training sample for each ball

Table 1: *This is an example of a table.*

ratio	decibels
1/1	0
2/1	≈ 6
3.16	10
10/1	20
1/10	-20
100/1	40
1000/1	60

To include text without formatting, use this (scriptsize uses a significantly smaller font, intermediate sizes are footnotesize and small):

```
I\O     1     2     3     4     5     6     7     8     9    10
  1  71.2   8.8   1.2   0.0   2.5   3.8   7.5   0.0   5.0   0.0
  2   0.0  87.5   1.2   0.0   2.5   2.5   0.0   5.0   0.0   1.2
  3   0.0   0.0  67.5   5.0   1.2  11.2   3.8   7.5   3.8   0.0
  4   0.0   0.0   1.2  62.5   3.8  22.5   0.0   6.2   2.5   1.2
  5   0.0   2.5   0.0   0.0  76.2   0.0   1.2   6.2   0.0  13.8
  6   5.0   1.2   6.2  21.2   5.0  47.5   1.2   5.0   1.2   6.2
  7  17.5   6.2   3.8   0.0   5.0   0.0  57.5   0.0  10.0   0.0
  8   0.0   0.0   2.5   1.2   8.8   0.0   0.0  73.8   2.5  11.2
  9  11.2   0.0   2.5   8.8   2.5   3.8   5.0   2.5  61.3   2.5
 10   1.2   0.0   0.0   2.5  20.0   0.0   0.0  12.5   0.0  63.7
```

If you want to use both columns, put it in a figure*: (figure* uses both columns, figure just 1): it is likely to float away to an unexpecte place, though.

```
I\O     1     2     3     4     5     6     7     8     9    10
  1  71.2   8.8   1.2   0.0   2.5   3.8   7.5   0.0   5.0   0.0
  2   0.0  87.5   1.2   0.0   2.5   2.5   0.0   5.0   0.0   1.2
  3   0.0   0.0  67.5   5.0   1.2  11.2   3.8   7.5   3.8   0.0
  4   0.0   0.0   1.2  62.5   3.8  22.5   0.0   6.2   2.5   1.2
  5   0.0   2.5   0.0   0.0  76.2   0.0   1.2   6.2   0.0  13.8
  6   5.0   1.2   6.2  21.2   5.0  47.5   1.2   5.0   1.2   6.2
  7  17.5   6.2   3.8   0.0   5.0   0.0  57.5   0.0  10.0   0.0
  8   0.0   0.0   2.5   1.2   8.8   0.0   0.0  73.8   2.5  11.2
  9  11.2   0.0   2.5   8.8   2.5   3.8   5.0   2.5  61.3   2.5
 10   1.2   0.0   0.0   2.5  20.0   0.0   0.0  12.5   0.0  63.7
```

Figure 7: Confusion Matrix

3.2 Maths, if needed

$$x(t) = s(f_\omega(t)) \tag{1}$$

where $f_\omega(t)$ is a special warping function

$$f_\omega(t) = \frac{1}{2\pi j} \oint_C \frac{\nu^{-1k} d\nu}{(1 - \beta\nu^{-1})(\nu^{-1} - \beta)} \tag{2}$$

A residue theorem states that

$$\oint_C F(z)dz = 2\pi j \sum_k Res[F(z), p_k] \tag{3}$$

Applying (3) to (1), it is straightforward to see that

$$1 + 1 = \pi \tag{4}$$

And here is an included image (png and pdf formats are allowed).

3.3 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] Rudnicky, 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).