UG3 Introduction to Vision and Robotics
Vision Assignment

Clemens Wolff, Toms Bergmanis

March 7th, 2013

1 Introduction

This report describes the work done for the first assignment in the IVR course. It gives the aims and hypotheses that guided the work; describes the algorithms that were implemented and reports the results of experiments that were run. The goal of this assignment was to develop three algorithms: one that detects the robots in each image, one that correctly identifies the direction of the robot and one that links together detections of the robots in consecutive images. Several simplifying assumptions were made about the possible set-ups of the assignment. Firstly, it was assumed that colours of the robots may change only due to the light it is exposed to, thus giving a rise to an assumption that robots will each appear in various shades of red or blue or green. Secondly, it was assumed that camera will be set up in some reasonable angle with respect to the plane it is supposed to observe. Finally, due to colour dependant approach it was assumed that background of the image will be in colour other than any of the colours of the robots.

2 Methods

In the following sections the three algorithms are described in detail. *Further detection of the direction of the robot was done under the assumption that colour representing the robot is different and appears differently than that of the triangle indicating it's direction. Furthermore - it was assumed that the triangle will have lesser pixel colour value in the channel representing robot's colour (red or green, or blue) in RGB representation than the average pixel of the convex hull. These assumptions can be justified by the fact that area of the triangle indicating the direction of the robot constituted relatively smaller area of the convex hull than the rest of the convex hull which was expected to appear in some other colour than black.*

2.1 Detection of Robots

Input

I, a three channel image of dimensions $m \times n$ in the RGB colorspace.

Output

M, a $m \times n \times 3$ binary matrix where for each pixel P_{ij} of I, it holds that:
$M(i, j, 1) = 1 \leftrightarrow P_{ij}$ belongs to the red robot,

$M(i, j, 2) = 1 \leftrightarrow P_{ij}$ belongs to the green robot,
$M(i, j, 3) = 1 \leftrightarrow P_{ij}$ belongs to the blue robot.

Algorithm

1. Apply approximate RGB-normalisation to I, giving I_n:
   - For each pixel in I_n, calculate the sum S_{rgb} of the red, green, and blue values of that pixel.
   - If $S_{rgb} \neq 0$ (the pixel is not absolute black), set each of the pixel's red, green, and blue values to that value divided by S_{rgb}.
2. Calculate μ_r, μ_g, μ_b and $\sigma_r, \sigma_g, \sigma_b$, the means and standard deviations of the values in the three channels of I_n.
3. Assign each pixel P_{ij} in I to one of the robots or to the background:
   - Normalise P's red, green, and blue values, giving P_n.
   - Calculate the probabilities p_r, p_g, p_b that P_n was generated by the gaussian distributions $\mathcal{N}_r = (\mu_r, \sigma_r), \mathcal{N}_g = (\mu_g, \sigma_g), \mathcal{N}_b = (\mu_b, \sigma_b)$.
   - Calculate P's hue value h.
   - If h is whithin a certain range defined as red and p_r is sufficiently small, set $M(i, j, 1) = 1$ (similarly for ranges defined as green/blue and p_g/p_b. If none of these conditions are met, set $M(i, j, 1) = M(i, j, 2) = M(i, j, 3) = 0$.
4. Remove noise from each channel in M:
   - Set pixels to zero if they have fewer neighbours with value one than they have adjacent pixels with value zero.
   - Set zero-valued pixels to one if they have two one-valued horizontal or vertical neighbours.
5. Remove components that are distant from the main concentration of mass in each channel in M:
   - Compute the center of mass C of the channel.
   - Compute $c_1, c_2, \ldots$, the centers of mass of each connected component in the channel.
   - Compute the mean distance δ of the c_k to C.
   - Set $M(i, j) = 0$ for all the pixels (i, j) in the components k that satisfy $c_k > \tau\delta$ for some fixed threshold τ.
6. Exploit the fact that all robots have similar sizes by setting every channel in M to all-zeros if the number of pixels set in that channel is smaller than the number of pixels set in the most populated channel by some margin.

Figure 1 shows a visualisation of the output matrix M.

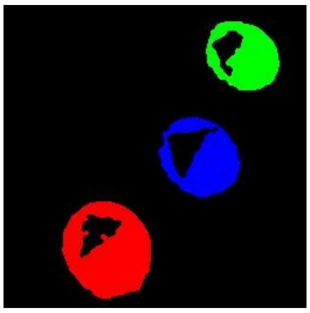


Figure 1: Result of Robot detection

2.2 Finding Robot Directions

Input

I, a three channel image of dimensions $m \times n$ in the RGB colorspace.

Output

$\Lambda = \{(c_r^m, c_r^t), (c_g^m, c_g^t), (c_b^m, c_b^t)\}$, a set where c_r^m is the center of mass of the red robot and c_r^t is the point towards which the robot is facing (similarly for the green and blue robots).

Algorithm

1. Get a matrix of robot masks M using the algorithm in Section 2.1. Let M_i be the ith channel of M i.e. the set of points $\{M(a, b, i) | 1 \leq a \leq m, 1 \leq b \leq n\}$. Apply the remainder of the algorithm to each channel ξ in M.

2. Calculate the convex hull H of the points in the channel and create the set of pixels of I that are inside H: $P = \{p_{ij} | p_{ij} \in \xi \wedge M(i, j, \xi) = 1\}$

3. Calculate μ, the average rgb-value over P. Generate $\Pi = \{p | p \in P \wedge rgbvalue(p) < \mu\}$, the set of pixels in P that have a below-average rgb value.

4. The black triangles on the robots are the pixels in Π. Get rid of them by setting the relevant indices in M to zero.
   Recompute the convex hull of M.

5. Repeat the previous step and remember the pixels in Π.
   This reduces noise in M by giving a tighter estimate on the robot's pixels when the triangles were under-detected by the algorithm in Section 2.1.
   Figure 2 shows the result of this step - a notable improvement in clarity of the triangles compared to Figure 1.

6. Update Λ: c_ξ^m is the center of mass of M_ξ, c_ξ^t is the center of mass of Π.
   A line from c_ξ^m to c_ξ^t indicates the direction of the robot.

Figure 3 shows a visualisation of the output set Λ.

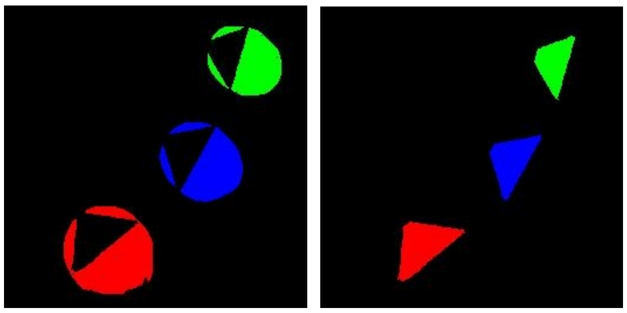

(a) After Removing Triangles (b) Removed Triangles

Figure 2: Triangle Detection via Local Thresholding

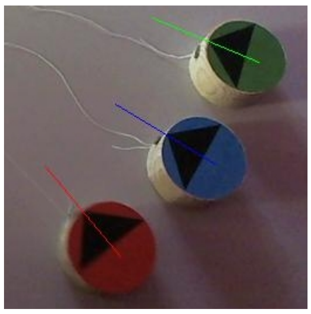


Figure 3: Detected Directions

2.3 Tracking Robots Over a Sequences of Frames

Input

$\Upsilon = \{I_1, I_2, \cdots\}$, a sequence where each of the I_i is a three channel image of dimensions $m \times n$ in the RGB colorspace.

Output

Ω, a visualisation of the robot positions over Υ.

Algorithm

1. Use a median-filter to generate a background Ω from Υ.
   For each $1 \leq i \leq m, 1 \leq j \leq n$:
   - Create $\omega_{ij} = \{I_k(i,j) | I_k \in \Upsilon\}$, the set of the colors of the pixels at location (i,j) of all the images in Υ.
   - Set $\Omega(i,j) = median(\omega_{ij})$.
2. For each $I_i \in \Upsilon$:
   - Use the algorithm in Section 2.2 to get the set Λ. Let $\lambda = \{c | (c, \_) \in \Lambda\}$.
   - Overlay Ω with a line from each element in λ_{i-1} to the corresponding element in λ_i, thus linking the centroids from image I_{i-1} to the centroids in image I_i.

Figure 4 shows a visualisation of the resulting track.

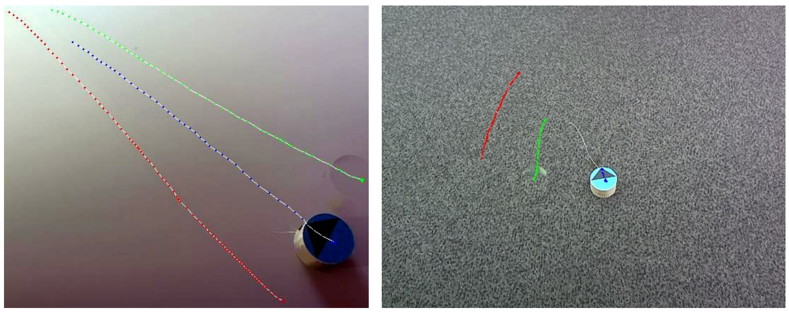

(a) Dataset 1 (b) Dataset 2

Figure 4: Output of Tracing Algorithm

3 Results

This section evaluates and visualises the performance of the three algorithms presented in Sections 2.1, 2.2, and 2.3. Table 1 describes the properties of the datasets used for this evaluation.

#	Background	Robot Size	Robot Color	Illumination
1	uniform, gray	large	saturated, dark	uniform, red hue
2	noisy, gray	small	faded, blue robot is cyan	histograms are bell-shaped
3	patterned, brown	large	saturated	daylight only
4	patterned, brown	large	saturated	daylight and artificial light

Table 1: Properties of evaluation datasets

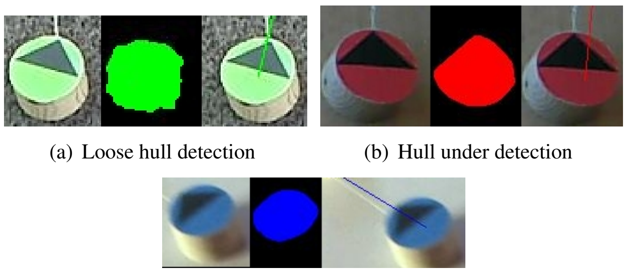
3.1 Detection of Robots

The algorithm described in Section 2.1, worked perfectly on datasets 1 and 2. Evaluation on the third dataset led to the worst performance over all datasets, with $\sim$60% of the occurences of the blue robot being undetected and $\sim$40% of the occurences of the green robot being under-detected (leading to bad direction detection). The performance on the fourth dataset was interesting: the red robot was under- detected in $\sim$45% of the cases (with the blue and green robots being found just fine) - while in the other datasets the red robot was usually detected with the highest confidence. Over all four datasets, about 10% of the robot instances were badly detected.
The fact that the color-detection algorithm works well on both datasets 1 and 2 leads to the conjecture that it is invariant under texture changes in the scene background and variations in robot-color saturation. The bad performance on dataset 3 can be explained by interference from the color of the scene background and by changes in scene illumination. The fact that the algorithm offers almost top-level performance on dataset 4 (captured on the same background as dataset 3) implies that the change in scene illumination is probably the largest influence on the algorithm's performance.
This is in keeping with the intuition that daylight has more inherent variation than arti-

ficial light, thus introducing a higher degree of variability into the characteristics of the captured images.

3.2 Detection of Directions

Performance of robot direction detection, understandably, is heavily dependent on the performance of the detection of the robots. If the robots are well isolated by Section 2.1's algorithm, robot orientations are perfectly detected.
The algorithm is invariant under loose detection - false positive cases where some addition non-robot region is misleadingly detected as a robot. This is due to the algorithm's ability to filter-out noisy detections.
In case of under-detection - false negative cases where some parts of the image representing the robot where not detected - the algorithm breaks: the predicted direction is skewed towards the opposite side of the under-detection. This is due to the algorithm not putting strict circular or elipsoidal constraints on the shape of the robots: under detection is thus able to move the center of mass of either the triangle or the robot-convex-hull. The error introduced by this is proportional to the area of the under-detected region.


(a) Loose hull detection (b) Hull under detection

(c) Good hull detection

Figure 5: Direction detections for convex hulls of different qualities

3.3 Tracking of the robots

Section 2.3's algorithm to track robots over consecutive frames is trivial - a mere visualisation of half of the results of the robot-direction- detection algorithm presented in Section 2.2. The tracking algorithm's performance is therefore directly related to the performance of the robot-direction-detection algorithm and the same observations as in Section 3.2 apply: generally speaking, the algorithm performed well.
Figure 4 begets one additional observation related to the evaluation of the robot-tracking algorithm: both datasets considered in this report exhibit the property that one of the objects of interest does not move much for most of the frames. This entails that generating a background from data employing a simple frame-difference base approach (such as the median- filter used in Section 2.3) to perform background subtraction is bound to fail as one of the objects of interest will be considered a part of the background due to being mostly stationary. This is unfortunate since pre-processing the datasets with background subtraction would increase the accuracy of Section 2.1's algorithm by reducing noise and increasing resolution in the image.

4 Discussion

The algorithms in Sections 2.1, 2.2, and 2.3 operate under a limited number of reasonable simplifying assumptions and performed well across a range of datasets captured under very different conditions.
The robot color detection algorithm of Section 2.1 could be improved by adding more stringent conditions on the shape of the robots e.g. fitting a circle or ellipse to them rather than a convex hull. This should improve the subsequent performance of the direction-detection algorithm by reducing the amount of under-detections.
If assumptions about the size, scale, and shape of the robots can be made, Hugh transforms or other shape-based techniques could be used to detect the robots. This would be more robust than the current color/pixel based approach but not invariant under scale or differences in camera positions.
Another way to improve the detection of the robots could be augmenting the current approach with the utilisation of second order spatial image statistics to pre-process the images. This would lead to a rough estimate of the robot positions with few false negatives, thus improving the currently utilised algorithms due to a reduced search space (implying a denser signal). A reduced search space also allows for computationally expensive but high-precision techniques to be used. One example of such a technique is a local color-based search starting from a small seed of pixels that are hypothesised to be part of a robot.
Direction detection could be improved by making it less dependent on the accuracy of the robot detection. One way to achieve this would be to base the direction calculation on properties of the detected triangles (e.g. direction = tangent line to the longest side) rather than on the triangles' centers of mass. This would increase robustness to under-detections.
Cross-frame tracking could be improved by abusing spatio-temporal proximity. One way to do this would be put a cap on how far the centroid of the robot can move in a certain frame interval, thus smoothing out the odd completely wrong prediction.

5 Code

```
function res = main(path, image_type, start_offset, time_step)
    if nargin < 2
        image_type = 'jpg';
    end
    if nargin < 3
        start_offset = 0;
    end
    if nargin < 4
        time_step = 0.33;
    end

    files = dir(sprintf('%s/*.%s', path, image_type));
    filenames = {files.name};
    [~, num_files] = size(filenames);

    [m n ~] = size(imread(sprintf('%s/%s', path, filenames{1})));
    track_mask = zeros(m, n);
```

```
  rc = zeros(num_files - start_offset, 2);
  gc = zeros(num_files - start_offset, 2);
  bc = zeros(num_files - start_offset, 2);

  reds = cell(num_files - start_offset, 1);
  greens = cell(num_files - start_offset, 1);
  blues = cell(num_files - start_offset, 1);

  for i = 1 + start_offset : num_files
      image = imread(sprintf('%s/%s', path, filenames{i}));
      [directions, centroids] = analyse_image(image);

      r = uint16(centroids(1,:));
      if r(1) > 0 && r(2) > 0
          track_mask = overlay_cross(track_mask, 1, r(1), r(2));
          rc(i,:) = [r(1), r(2)];
      end

      g = uint16(centroids(2,:));
      if g(1) > 0 && g(2) > 0
          track_mask = overlay_cross(track_mask, 2, g(1), g(2));
          gc(i,:) = [g(1), g(2)];
      end

b = uint16(centroids(3,:));
      if b(1) > 0 && b(2) > 0
          track_mask = overlay_cross(track_mask, 3, b(1), b(2));
          bc(i,:) = [b(1), b(2)];
      end

      reds{i} = image(:,:,1);
      greens{i} = image(:,:,2);
      blues{i} = image(:,:,3);
      imshow(overlay_mask(image, directions));

      pause(time_step);
  end

  rc = rc(any(rc,2),:);
  gc = gc(any(gc,2),:);
  bc = bc(any(bc,2),:);

  red_median = median(cat(3, reds{:}), 3);
  green_median = median(cat(3, greens{:}), 3);
  blue_median = median(cat(3, blues{:}), 3);

  bg = cat(3, red_median, green_median, blue_median);

  bg = overlay_polygon(bg, rc, [255, 255, 255]);
```

```
    bg = overlay_polygon(bg, gc, [255, 255, 255]);
    bg = overlay_polygon(bg, bc, [255, 255, 255]);
    bg = overlay_mask(bg, track_mask);
    imshow(bg)
end
```