TSB-funded Project 'TADD' - Trainable vision-based anomaly detection and diagnosis Technical Report for December 2013

Ran Song and Tom Duckett

Agri-Food Technology Research Group, Lincoln Centre for Autonomous Systems Research, School of Computer Science, University of Lincoln, UK

Abstract

In this report, we proposed an automatic method in order to detect the region of the label within a tray image. This method is based on some classic computer vision techniques, including feature extraction, image registration, image erosion and dilation. Such computer vision techniques are specifically integrated, forming an effective object detection system which can handle various images of food trays with different positions and orientations. Preliminary experiments show that the proposed method is efficient, reliable and easy-to-use (in particular for non-expert) while certainly more tests need to be done to comprehensively evaluate it.

1. Introduction

The automatic detection of the label region on a food tray is very important in modern food industry. On the one hand, we need to make sure that the label, usually containing such information as barcode, dates, price and weight is presented at the right position and orientation. On the other hand, the detection of the label region is typically a necessary preprocessing step for the Optical Character Recognition (OCR) so that the image information of the printed texts within the label region can be converted into machine-encoded/computer-readable text and then the system can judge whether the printed dates, price and weight are semantically right or wrong. Importantly, the accuracy of the label region detection largely affects the reliability of the following OCR.

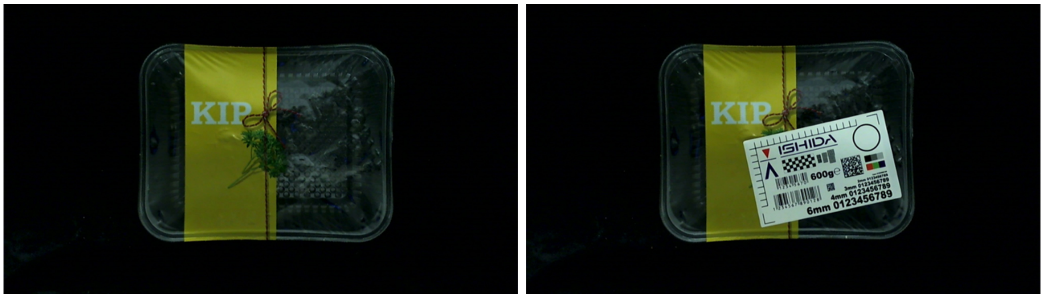

Figure 1: The two input images required by our method

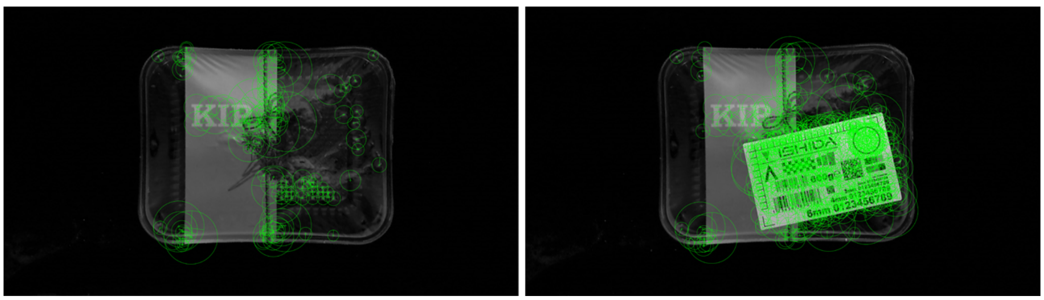

Figure 2: The results of feature extraction using SURF where the centre of each circle denotes the location of each feature, the size of the circle denotes the scale of the feature and the line segment within the circle denotes the orientation of the feature.

However, detecting the label region from a complicated image of food tray is not a simple task, which usually consists of some advanced computer vision and/or pattern recognition techniques due to the complex and various content of the input images [2]. If we further consider whether the detection is fast enough and the system is easy to operate (both are vital for industrial applications), the task will be more challenging. In this report, we propose a novel method to solve the problem.

2. Method

The proposed method is basically a teaching-by-showing approach. We first input a food tray image without the label to teach the system. Then we input a food tray image with the label and expect that the system can automatically detect the position and the orientation of the label based on the information it is taught.

Our method contains three major steps:

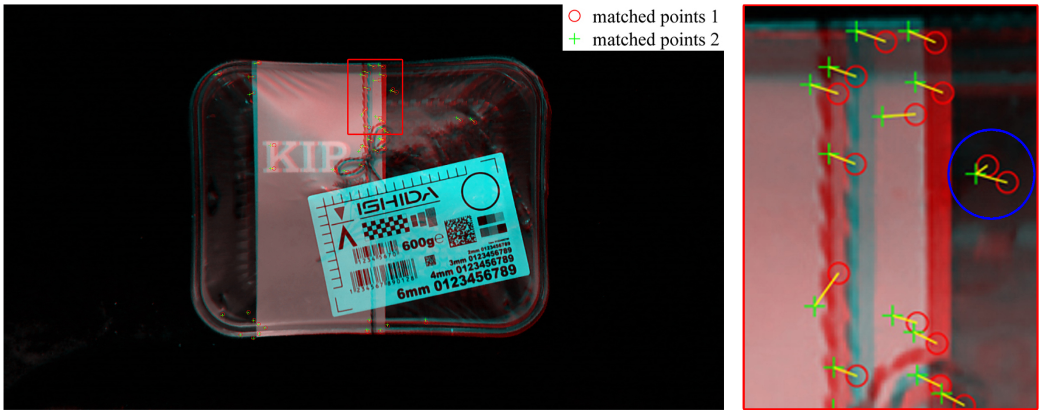

Figure 3: The raw matches produced by SURF contains some incorrect matches. To visualise the results, we deliberately colour the two images in red and blue respectively. In the blue circle marked in the enlarged subfigure on the right hand side, there are incorrect matches since two points cannot correspond to the same point.

1. Feature extraction using the Speeded Up Robust Features (SURF) algorithm [1];
2. Feature matching and image registration using the Random Sample Consensus method (RANSAC) [3] and the Levenberg-Mardquardt algorithm;
3. Label region localisation and repairing.

In the following, we describe each step in details.

2.1. Feature extraction via SURF

The proposed method requires two input images. As shown in Fig. 1, one is a base image of a food tray without the label and the other is a test image of a food tray with the label. Then, we employ the SURF method to perform feature extraction. Please refer to our report of November 2013 for the details of the implementation of the SURF method. Here, we just show the results of our implementation in Fig. 2.

2.2. Feature matching and image registration

Fig. 3 shows the result of the raw feature matching produced by SURF where we use the yellow lines to illustrate the matches. It can be seen that some features points are not correctly matched. A correct match means that the two matched points should correspond to the same point in the real

Figure 4: By using RANSAC, we remove incorrect raw matches or 'outliers', leaving only the 'inliers'.

world although they are from different images. Incorrect matches can lead to a poor registration between the two images. To address this problem, we employ the RANSAC method to remove incorrect matches. The details of the RANSAC method have been given in our report in November. The outcome of applying it is a refinement of the matches, only preserving the correct matches as shown in Fig. 4.

The image registration via Levenberg-Mardquardt algorithm has also been mentioned in our last report. Basically, we utilise this technique to align the base image and the test image in a uniform coordinate system. It is worth noting that in this specific task, the registration is actually dependent on the matched feature points in the non-label regions since they are the common regions contained by both images. Thus one assumption of the method is that the non-label regions of the tray contain enough number of features. In practice, such an assumption can hold because (i) the films of food trays usually contain some printed patterns or (ii) there is some food in the tray which actually forms some visual patterns even if the film is completely transparent.

The outcome of the registration is shown in Fig. 5. It can be seen that each pair of corresponding feature points have already been moved to the

Figure 5: After registration, the feature points from different images are fully overlapped.

same position after the registration. This offers us a good basis for the next step.

2.3. Label localisation and repairing

To localise the label region, we first normalise the intensity values of the registered base image and the test image for the later image subtraction. To make the subtraction not very sensitive to the ever-present image noise (mostly caused by the sensor), we apply an average filtering to the subtraction image where the size of the convolution kernel is set to 10×10. This leads to Fig. 6 where the subtraction image looks a little bit blurred because of the effect of filtering. To achieve a better visualisation, we also normalise the subtraction image to interval $[0, 1]$. Thus in Fig. 6, the warmer the colour, the larger the difference between the base image and the test image.

To better localise the label, an intuitive idea inspired by Fig. 6 is to set a threshold and then detect the pixels with values greater than the threshold as the label region. Unfortunately, this idea is rather problematic. This is because the values are not consistent within the label region. As shown in Fig. 6, the top right corner has a strong response while quite a few regions inside the label are in blue. A thresholding scheme is thus not robust enough to detect a complete rectangular region as the label region. Instead, we

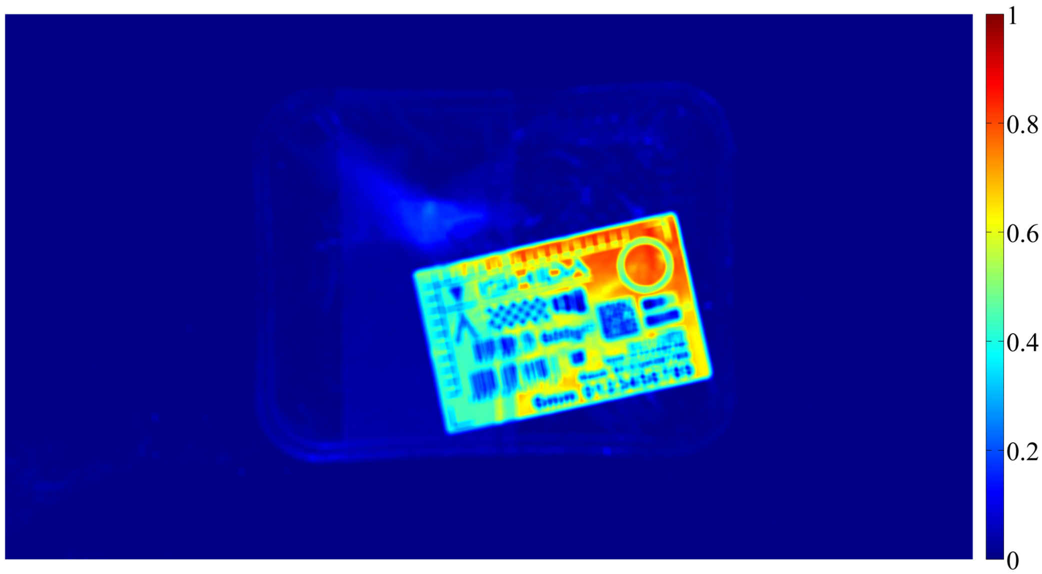

Figure 6: The result of image subtraction using the registered base image and the original test image.

develop a method to more reliably detect the repaired rectangular lable region using image erosion and dilation.

First, we convert the subtraction image to a binary image by thresholding. The threshold is set as the double of the mean value of the subtraction image. Second, a flood-fill operation is performed on the binary image to fill the holes. A hole is a set of background pixels which cannot be reached by filling in the background from the edge of the image. We then perform image erosion by creating a disk-shaped morphological structuring element. The radius of the disk is fixed to 70. Next, we implement image dilation using the same morphological structuring element. To pursue an adaptive scheme, we iteratively adjust the threshold used for the binarisation of the subtraction image. In each iteration, we measure the number of the connected components in the image after the operations of erosion and dilation. If it is not equal to 1, the threshold is added by 0.1 and the aforementioned steps will be repeated. Such an adaptive scheme guarantees that one complete label region can be detected. The final step is to draw the minimum bounding box around the detected label region by computing its convex hull. Fig. 7 shows the final result of our label detection method.

Figure 7: The final result of label region detection

3. More results

Figs. 8–12 show more results where we can see that the proposed method is highly reliable and the positions and the orientations of the trays are various. In general, the running time is 1-2 seconds using an Matlab implementation. We expect that the speed could be much faster if we can develop a C++ implementation. Certainly, before doing so, more tests need to be done to demonstrate that the current version of this method is really reliable and to make sure that the method is flawless.

4. Future work

Besides the redevelopment of the proposed method in C++ in the near future, more functionalities can be added. For example, we can teach the system what is the right position of the label by inputting a food tray image with a label at the right place. Then, by combining our previous development on background segmentation, the system should be able to recognise whether the label position is right for any input food tray images. Another issue is to obtain more datasets to test the reliability of the system. In particular, we need to establish our own ground truth data to evaluate the accuracy of the system.

Figure 8: Top left: the base image; Top right: the test image; Bottom: label detection

References

[1] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up robust features, in: Proc. ECCV, Springer, 2006, pp. 404–417.

[2] L. Fang, C. Xie, 1-d barcode localization in complex background, in: Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on, IEEE, pp. 1–3.

[3] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981) 381–395.

Figure 9: Top left: the base image; Top right: the test image; Bottom: label detection

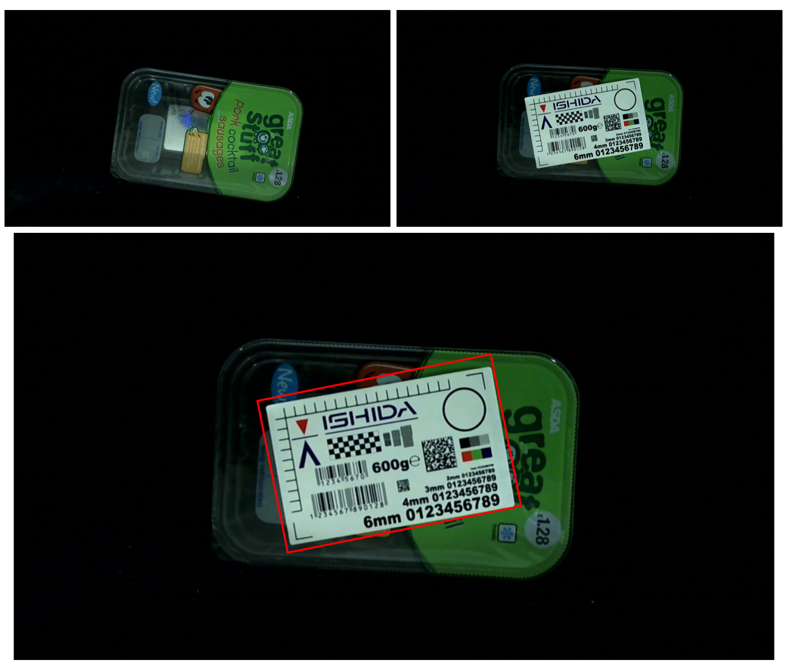

Figure 10: Top left: the base image; Top right: the test image; Bottom: label detection

Figure 11: Top left: the base image; Top right: the test image; Bottom: label detection

Figure 12: Top left: the base image; Top right: the test image; Bottom: label detection