Noname manuscript No.
(will be inserted by the editor)

Technical Report for the Label-TADD System

Ran Song · Tom Duckett

Received: date / Accepted: date

Abstract This paper reports an automatic system to detect the label region on a food pack. It is based on a list of modern computer vision techniques, including keypoint detection, keypoint matching, motion estimation, image registration, image erosion and dilation. In the proposed method, such computer vision techniques are specifically integrated, forming an effective object detection system which can handle various images of food packs. We also developed a friendly graphical user interface where the entire layout of the software is consistent with the standard of modern industrial software applications. To further demonstrate the advantages of the system, we incorporate text localisation and optical character recognition (OCR) into it and achieves desired results. Experiments using a variety of real food pack images show that the system can effectively and efficiently detect label positions and orientations and also improve the OCR results.

Keywords Food pack · Label detection · Image registration

1 Introduction

In food industry, more and more systems based on modern computer vision techniques have been witnessed in the emerging area of 'intelligent food quality control' where the principle is the reduction of human effort by replacing it with computational intelligence. In this report, we propose a system which is specifically designed to recognise, detect and segment the label region on a food pack. A label region on a food pack usually contains important information about the food product, such as time and date, price and weight, etc. Visually on a food pack, it is the region where most errors occur. More importantly, such errors as an incorrect expiry date could cause serious consequence since it is closely related to food safety. And the economic cost of these errors, if they cannot be detected and then stopped before entering the market, could be quite heavy.

Therefore, this work aims at detecting such errors in the packaging process where the food pack is still on the conveyor belt. Compared with the existing method which is mainly based on human effort, the proposed system relies on computational intelligence and thus is more reliable over long time operation and more efficient.

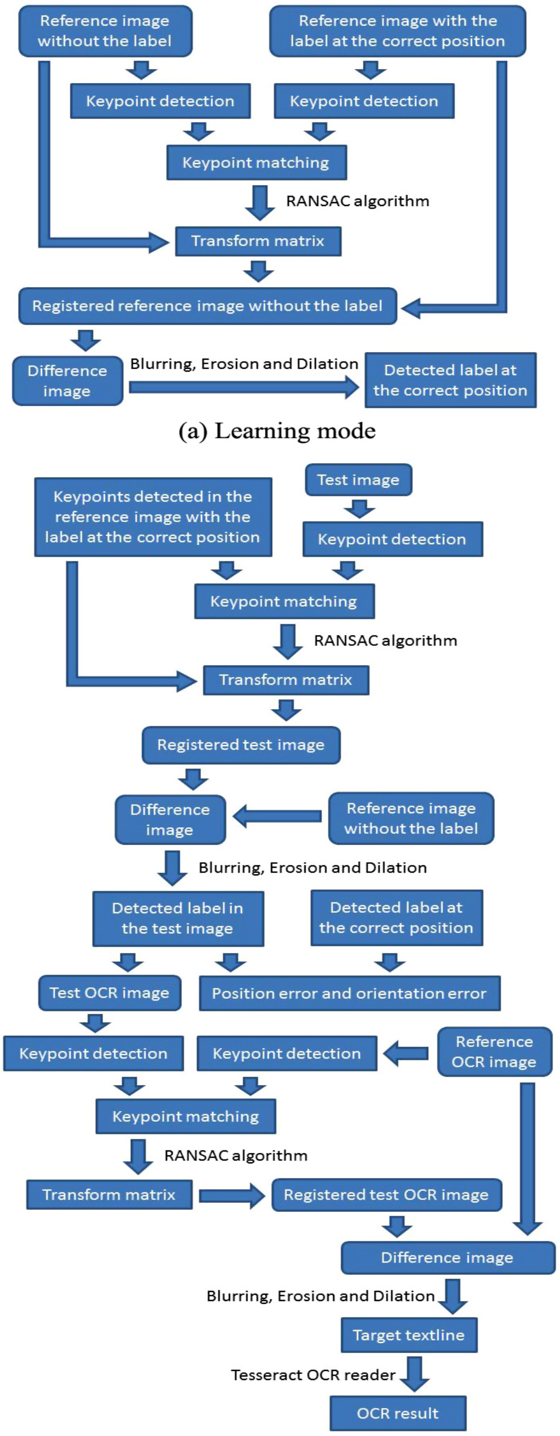
In this report, we shall focus more on the technical part of the system. And we found that a literature review to the related work is highly difficult because to the best of our knowledge, this is the first system for vision-based intelligent food packaging although quite a lot of vision-based methods have been presented for other types of food inspection tasks. This kind of works can be found in the review papers [4, 2, 9, 8, 5]

2 System overview

The Label-TADD software is based on the idea of teaching-by-showing to detect the label position with reference to the food pack. It first learns the information supplied by the reference image and then compares the information within the test image with that learned from the

Ran Song
School of Computer Science, University of Lincoln, Lincoln, UK
E-mail: rsong@lincoln.ac.uk

Tom Duckett
School of Computer Science, University of Lincoln, Lincoln, UK



(a) Learning mode

(b) Testing mode

Fig. 1 The workflow of the Label-TADD system is composed of two modes: learning and testing

reference image to find the difference which typically corresponds to the label.

However, we cannot compare the test image and the reference image directly since the packs usually appear at different positions with different degrees of rotation.

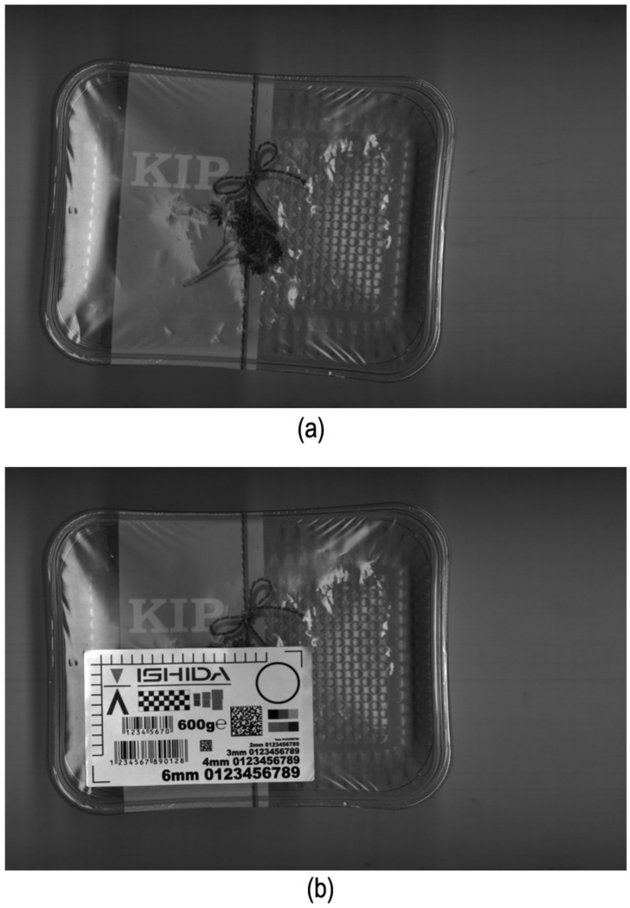


Fig. 2 Two reference images are needed to teach the system. (a) The first reference image (b) The second reference image. Please note that benefiting from the reliability of the system, here small rotations are allowed

This is caused by the nature of the image acquisition provided through the Ishida machine where a food pack is laid onto the conveyor belt and the line scan camera on top of it takes a photo of the surface of the food pack. Hence it is usual that the food pack has a small rotation and unknown amount of translation towards the reference position. Thus the core algorithm of the Label-TADD system is the so-called registration. Through the registration, two food packs can be laid at the same position with the same degree of rotation angle as long as the majority of their visual appearances are identical, which is just the very case for the same type of product.

However, to make a system really useful in real applications, rather than just a concept demonstration working under a laboratory environment, we need some secondary but non-trivial steps of processing for the input images. We combine all of these steps together with the core image registration techniques and show the workflow of both the learning and the testing modes of the proposed system in Figure 1. It can be seen that

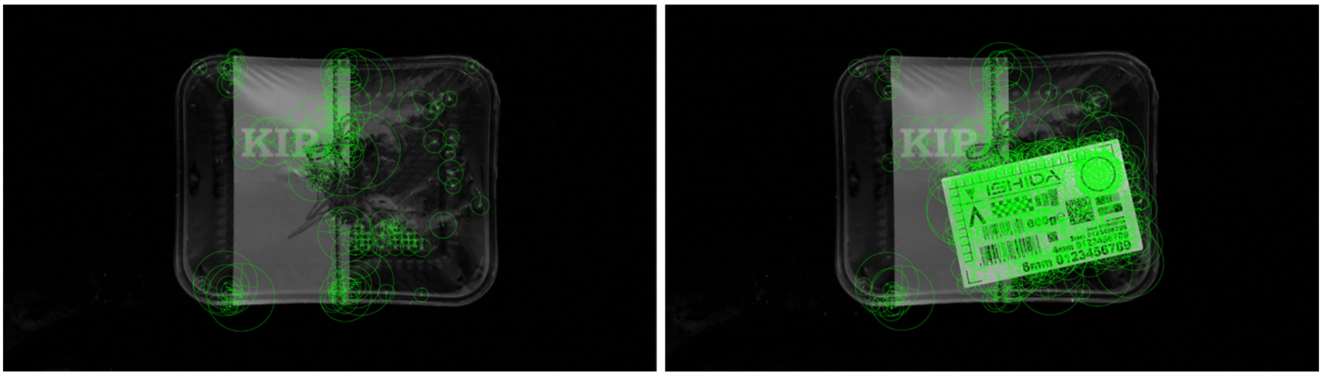


Fig. 3 SURF keypoint detection on images with and without a label. Each SURF keypoint is represented by a circle where the centre of the circle denote the location of the keypoint, the lenghth of the radium denotes the scale and we draw a line segment inside each circle to represent the orientation of the keypoint.

a couple of modern computer vision techniques are involved in the system, such as keypoint detection, keypoint matching, motion estimation, image registration, image erosion and dilation. In the next section, we shall give a detailed description about these techniques and how we apply them into the label detection task.

3 Technical details

The teaching-by-showing scheme in the system is realised by inputting two reference images as shown in Figure 2. The first one is an image of a food pack without the label and the second one is with the label and the label must be at the correct position on the surface of the food pack. Then, image registration technique is used to align the two reference images within the same coordinate system. It can be observed that in the testing mode illustrated by Figure 1 (b), image registration is also employed. Image registration is one of the most important topics in computer vision. A modern image registration method usually includes three stages: keypoint detection, keypoint matching and motion estimation.

3.1 Keypoint detection

In our system, we employ the speeded up robust features (SURF) algorithm [1] for keypoint detection. A SURF keypoint is defined by three arguments including location, scale and orientation. Figure 3 shows the results of SURF keypoint detection on two images. Please note that due to the requirement of real-time implementation, the efficiency of the system is vital. Considering the rolling speed of the real conveyor belt, we only have approximately 0.6 second to process one image. Thus the most well-known keypoint detection method, the scale invariant feature transform (SIFT) algorithm [6] is not quite fit for our specific application because of its relatively slow running speed.

3.2 Keypoint matching

To match corresponding keypoints detected from two different images, the Random Sample Consensus method (RANSAC) [3] is employed. The basic theory of the RANSAC algorithm is simple but ingenious. Firstly, it randomly samples four points to fix one motion defined by a transform matrix (details will be given in the next subsection). The underlay (or consensus) of this motion is defined as the points whose registration error to the corresponding transform matrix are less than some threshold. After many iterations of this random sampling approach, the final solution is the motion that has the largest underlay or the largest 'consensus' and the points within the underlay of this motion are the correct matching points which consist of the consensus.

3.3 Motion estimation

The motion between two images can be expressed mathematically by defining a transform matrix as below

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix}. \tag{1}$$

The transform matrix project a keypoint (x, y) in one image to the coordinate system of the other image through the following equations

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + 1}, \quad y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + 1}. \tag{2}$$

Fig. 4 Image registration where we deliberately give red and blue shading and some amount of transparency to the two images respectively to make them visually different.

Since as shown in Equation (1), the transform matrix contains eight unknown parameters, we need at least four pairs of matching points to construct eight equations according to Equation (2) for the estimation of each parameter.

The aforementioned registration error between two images is defined as below

$$D = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i' - x_i'')^2 + (y_i' - y_i'')^2} \qquad (3)$$

where the point (x_i'', y_i'') is the matching point of (x_i, y_i) in the other image. Therefore, the outcome of RANSAC is a transform matrix which can minimise the registration error function based on the set of matching points.

The result of the registration is shown in Figure 4 where the two images are mostly overlapped and in particular, the matching points are almost fully overlapped with each other.

3.4 Blurring, erosion and dilation

Erosion and dilation are two basic operators in the area of mathematical morphology. They are typically applied to binary images where white regions usually represent the regions of interest. It can be imagined that small white regions will disappear due to the effect of erosion and large white regions will be largely unchanged since the dilation can compensate the shrinkage effect caused by the erosion. Also, some small holes (black regions with in a large white region) can also be filled through erosion and dilation.

In our system, the idea is to analyse the difference between the two registered images to find the label position because ideally in the resultant difference image, the pixels within the label region should have high values and those out of the label region should have zero

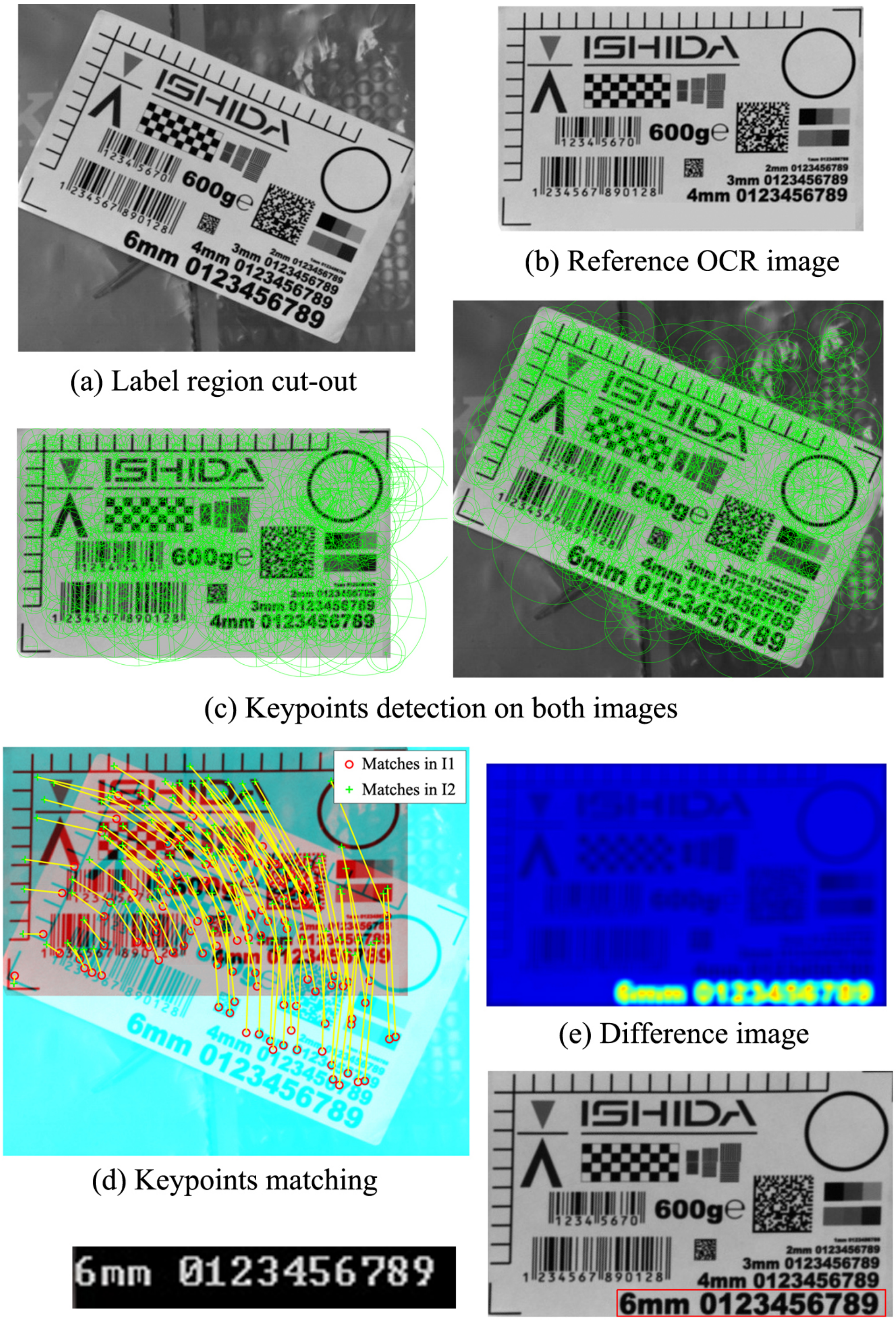
Fig. 5 The heat map of the difference image of the two registered images

value. However, due to the effect of sensing noise, reflection and unexpected distortion (e.g. the surface of the food pack is not flat), the difference image of the two registered images could contain not just the label but also quite a few unexpected patterns even if we first use a Gaussian filter for raw denosing. Figure 5 use a heat map to show this effect where non-zero regions do not only occur within the label region and also there are some low-value pixels within the label region. Therefore, in the binarised difference image, we use image erosion and dilation to remove the small white regions corresponding to the unexpected patterns and only keep the largest one corresponding to the label.

In our system, the Gaussian blurring is operated in a 40×40 neighbourhood for each pixel.

3.5 OCR

Optical character recognition (OCR) is the mechanical or electronic conversion of images of typewritten or printed text into machine-encoded text. It is widely used as a form of data entry from printed paper data



(a) Label region cut-out

(b) Reference OCR image

(c) Keypoints detection on both images

(d) Keypoints matching

(e) Difference image

(f) Textline localisation and OCR reading

Fig. 6 The steps of operation towards textline localisation and OCR reading

records, whether passport documents, invoices, bank statement, receipts, business card, mail, or other documents. It is a common method of digitizing printed texts so that it can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as machine translation, text-to-speech, key data and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision. In our system, we cut out the label region and input it as an intensity image into the Tesseract OCR reader [7]. Since the Tesseract OCR reader works better for a single textline than for multi-

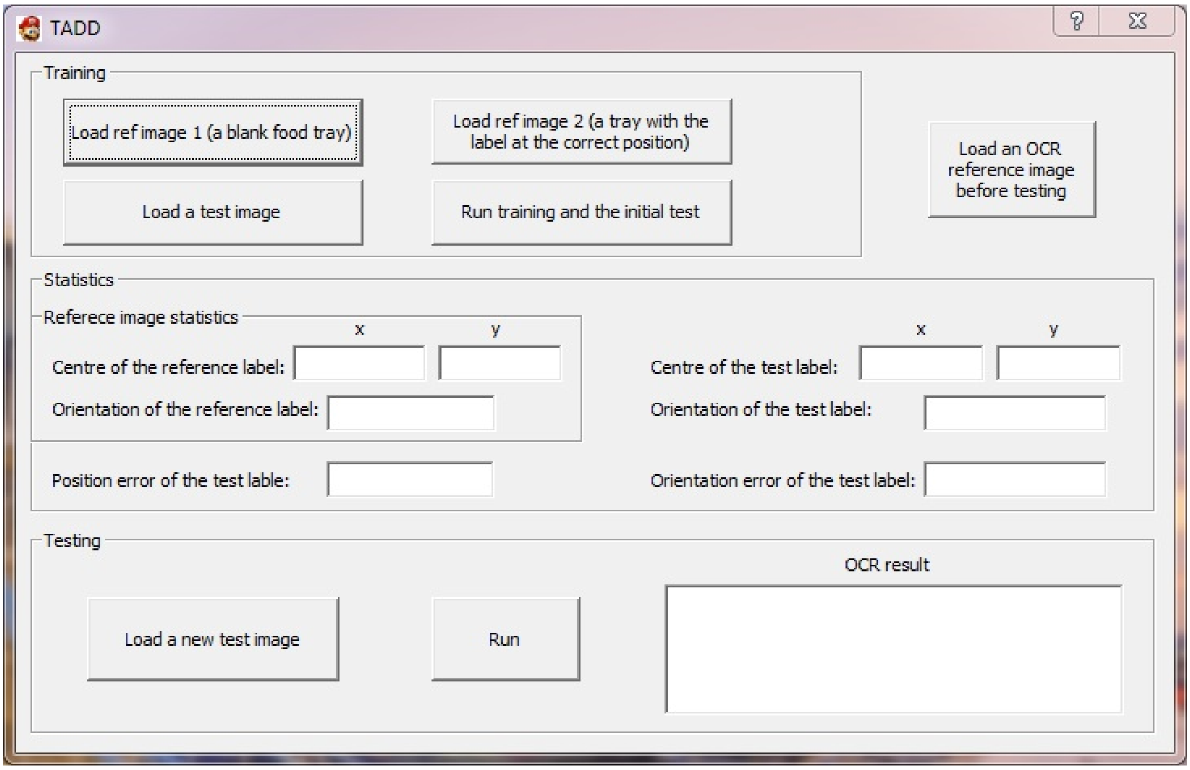


Fig. 7 The GUI for the Label-TADD system

ple textlines, we utilise a reference OCR image to detect the target textline within the label. The step-by-step operation of our text localisation and OCR component is shown in Figure 6.

3.6 GUI design

We also designed a friendly graphical user interface (GUI) for the system although a command line version connected to the conveyor belt is also available. In computing, graphical user interface (GUI) is a type of user interface that allows users to interact with electronic devices through graphical icons and visual indicators such as secondary notation, as opposed to text-based interfaces, typed command labels or text navigation. GUIs were introduced in reaction to the perceived steep learning curve of command-line interfaces (CLI), which require commands to be typed on the keyboard. In this work, we design a GUI for the system as shown in Figure 7

4 Experimental results

We test a wide range of images of real food packs using the proposed label detection system. Some results are shown in Figure 8. In Figure 8, please note the two images at the top right corner, it demonstrates that the label region detection is adjustable by using different learning images. It can be seen that the systems can offer a very reliable label detection no matter what content the food pack has, whether the food pack suffers from some rotation or translation, whether it is distorted and in some cases, whether there exists reflection. The entire system also work quite efficiently by processing two images per second on average in the testing mode, which makes it fully competent for a real-time application by receiving and handling images directly captured by the line-scan camera mounted on top of the conveyor belt.

Figure 9 shows that the software can also compute the position and the orientation errors of the label in the current food pack and output the OCR results on its interface.

5 Conclusions

We present a system for automatically detecting the label region on a food pack based on a teaching-by-showing scheme. The entire system is developed using C++ and can be implemented in real time rate. We tested it by using a variety of images and the results were highly accurate and reliable.

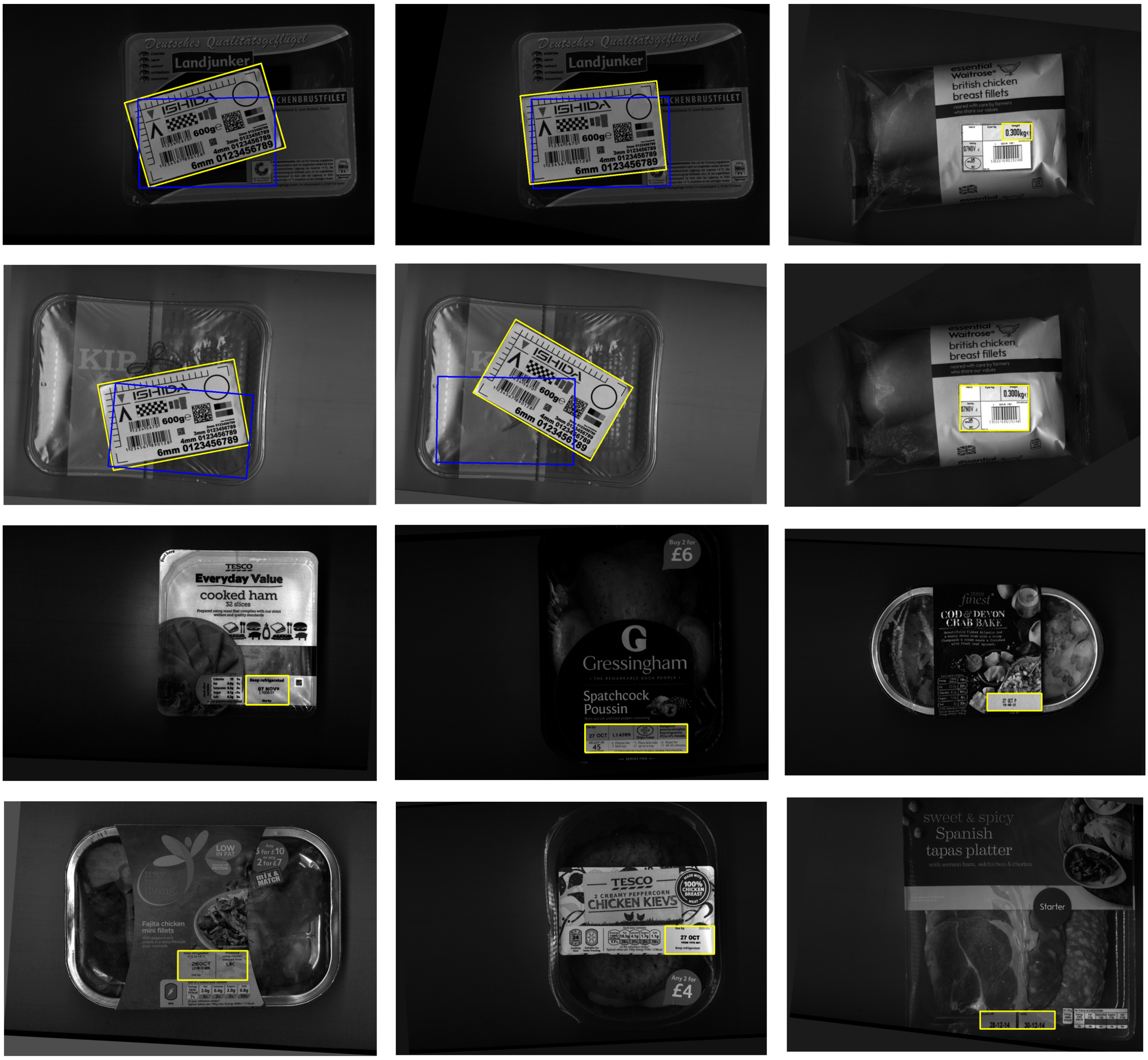


Fig. 8 Some label detection results where we use a yellow rectangle to denote the detected label position and a blue one (optional) to denote where it should be.

Acknowledgements The work is funded by the Technology Strategy Board, Ishida Europe and other industrial partners through the TADD (Trainable Vision-Based Anomaly Detection and Diagnosis) project. This support is gratefully acknowledged.

References

1. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Proc. ECCV, pp. 404–417. Springer (2006)
2. Brosnan, T., Sun, D.W.: Improving quality inspection of food products by computer vision—-a review. Journal of Food Engineering **61**(1), 3–16 (2004)
3. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
4. Gunasekaran, S.: Computer vision technology for food quality assurance. Trends in Food Science & Technology **7**(8), 245–256 (1996)
5. Jackman, P., Sun, D.W.: Recent advances in image processing using image texture features for food quality assessment. Trends in Food Science & Technology **29**(1), 35–43 (2013)
6. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004)
7. Smith, R.: An overview of the tesseract ocr engine. In: Proc. Ninth Int. Conference on Document Analysis and

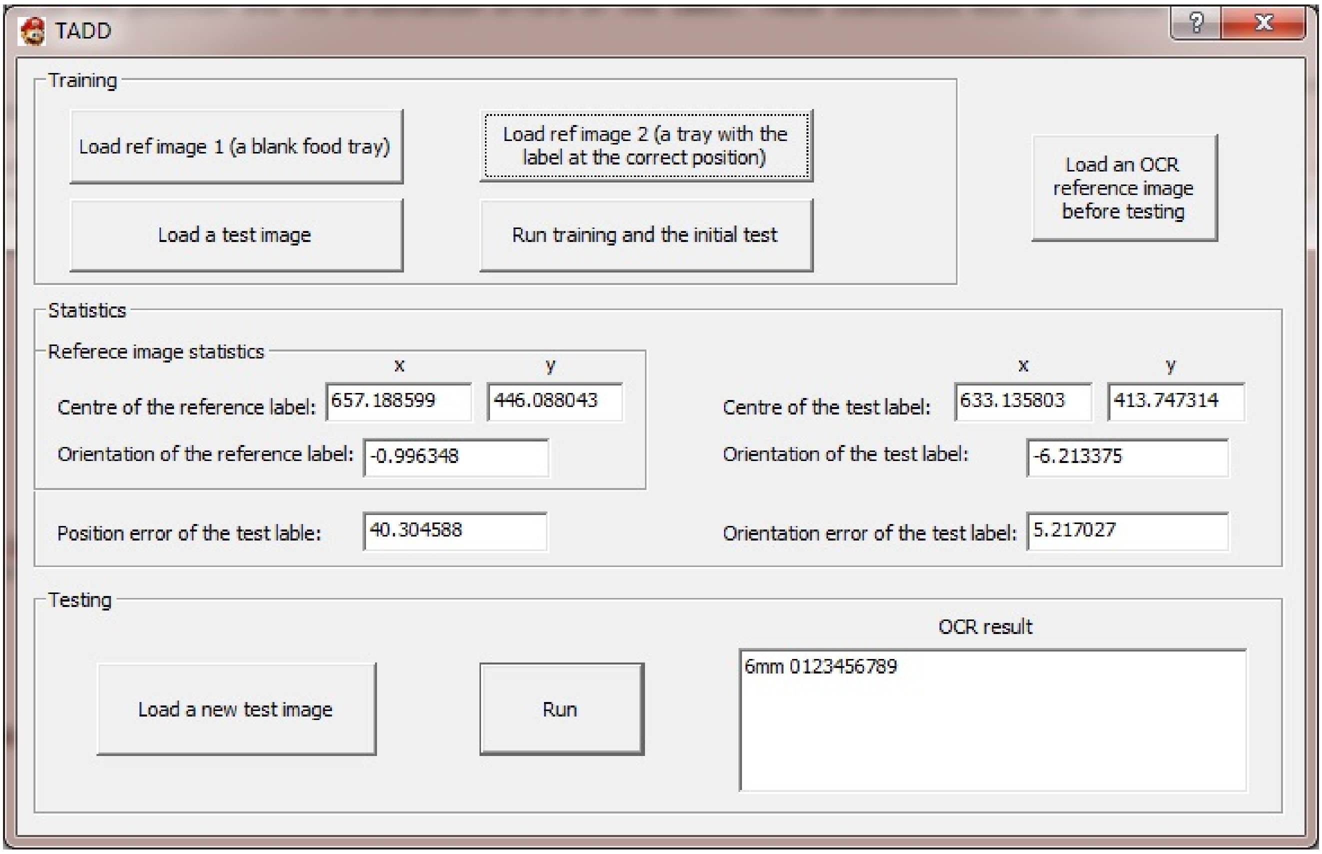


Fig. 9 The GUI shows a list of information related to the detected label.

Recognition (ICDAR), pp. 629–633 (2007)
8. Wu, D., Sun, D.W.: Colour measurements by computer vision for food quality control–a review. Trends in Food Science & Technology **29**(1), 5–20 (2013)
9. Zheng, C., Sun, D.W., Zheng, L.: Recent developments and applications of image features for food quality evaluation and inspection–a review. Trends in Food Science & Technology **17**(12), 642–655 (2006)