A Logo-Based Approach for Recognising Multiple Products on a Shelf

Trisha Mittal(✉), B. Laasya, and J. Dinesh Babu

Information Technology, IIIT Bangalore, Bangalore, India
{Trisha.Mittal,B.Laasya}@iiitb.org,
jdinesh@iiitb.ac.in

Abstract. This paper addresses detecting, localizing and recognizing various grocery products in retail store images. Our object recognition algorithm achieves this goal using just one image per product for training, assuming that the category of the products (like cereals, rice, etc.) is known. This algorithm uses logo detection as a precursor to product recognition. So, the first step involves detecting and classifying products, at a broader level, based on their brands. The second step is the finer classification step for recognizing and localizing the exact product label, which involves using colour information. This hierarchical approach limits the confusion in classifying similar looking products and outperforms product recognition that was implemented without logo detection. The algorithm was tested on 80 annotated grocery shelf images containing 238 different products that fall under 3 categories. This facilitates smarter inventory management in retail stores on a large scale and on a day to day basis for the visually impaired people.

Keywords: Object detection and recognition · Grocery products · Local feature · Template matching

1 Introduction

Multiple instance recognition is a relevant problem in both object recognition and image retrieval literature. A general object recognition algorithm should be able to recognize multiple instances of varied known objects. In a real world scenario, this would be highly challenging due to diverse backgrounds, occlusion, illumination and many other factors. Confining this general problem to the retail scenario makes it more addressable, nevertheless the challenges persist. One major challenge would be to deal with large number of products that keep getting added to the dataset. With an ever growing dataset, getting large number of training images for each product is impractical. Developing a recognition system that uses one image per product for training, addresses this problem. Even if there were just one image per product for training, it would be unrealistic to assume that the testing and training images belong to the same distribution.

To this end, we designed a scalable algorithm that performs well even in cross dataset settings, as shown in Fig. 1. We exploit the fact that logos are designed to be unique and distinct so that they can be easily associated with their respective brands.

Y. Bi et al. (eds.), *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016*,
Lecture Notes in Networks and Systems 16, DOI 10.1007/978-3-319-56991-8_2

Fig. 1. Our product recognition algorithm recognises the products on the left column in the shelf image on the right side

This association between logos and products plays an integral role in our algorithm for recognizing grocery products on a shelf.

The algorithm mainly comprises two steps. The first step involves detecting and classifying products based on their brands. This is achieved by matching the extracted SIFT [1] keypoints from the image of the grocery shelf with the logos of all products in the training dataset and picking the top k logos that have the most number of matches. Then, template match, which is modified to be scale invariant, is used for localization of the logos. The second step is the finer classification step for recognizing and localizing the exact product label, which involves using colour information.

Grocery image interpretation can prove beneficial to store owners for inventory management, companies to monitor its brands in large retail store and help the visually impaired.

The paper has been structured as follows. Section 2 presents the related work. Section 3 is Methodology, where we explain our algorithm in detail. In Sect. 4 we delve into the details about the dataset that we used and tested our algorithm on. Also, we present our accuracy results calculated using mean Average Precision (mAP). Finally, we conclude the paper with some details about the Future Work in Sect. 5.

2 Literature Review

Substantial work has been done in recognition and classification of objects. Most of this work falls in either using global or local features. Global features like colour histograms [2] and Principal Component Analysis (PCA) [3] were used in some of the earlier works in this field. More recently, local features like SIFT [1] and SURF [8] have gained a lot of popularity and are widely used. These local features have had an upper hand over global features as they overcome problems like occlusion, scale-variance and orientation.

But, these local features cannot be directly used for multiple object recognition as the keypoints spanning over two or more instances will be considered for homography and will lead to a skewed bounding box. Skoczylas [11] solve this problem for detecting multiple identical logos by using a sliding window method. Hence, generalising single instance detection to multiple instance detection has its own set of challenges. One of the methods [5] estimates the number of instances of a particular object in challenging situations by using the traditional shape matching algorithms [6, 7] and Hough-like collaborative voting. In place of the geometric constraints used in the previous work, which need to be obtained manually, Collet et al. [9] use local descriptors from several training images to learn a 3-dimensional model of an object. Given a test image, they group local keypoints belonging to the same object based on their 2D position and use RANSAC to recognize them. However, this approach might not be able to distinguish between two similar instances lying in very close proximity. And also both the above approaches would be impractical for the retail scenario as they require training images to be collected for a large number of products.

Grocery Products Recognition is a specific case of multiple object detection. Among the few available works that deal with this, is the ShelfScanner [10], which aims at detecting products on a shelf in real time. This is achieved by using enhanced SURF descriptors extracted from images to train a multiclass naïve Bayes classifier. This work achieves good performance on real world data but requires high quality training data. Whereas George and Floerkemeier [4] achieve runtime efficiency by using one single image per product label as training data. They adapt a hierarchical approach that first recognizes the category (like biscuits, chocolates, etc.) using discriminative random forests. Then an intra-class classification to simultaneously recognize and localize the product is done using deformable dense pixel matching and genetic algorithm.

3 Methodology

Two approaches to solve the grocery products recognition problem have been discussed in considerable detail in this section. First approach that we propose is to use logo detection as a precursor to the product recognition step. We compare this with the second approach that involves only the product recognition step. The second approach serves as the baseline. George and Floerkemeier [4] try to solve the same grocery products recognition problem. But our approach differs from theirs mainly in the usage of logo detection to aid the product recognition step.

3.1 Using Logo Detection

Our product recognition algorithm involves detecting and localizing logos of the products belonging to a particular category present on a shelf. In order to achieve this, we first aim at shortlisting the probable logos within a given category. Initially, the SIFT keypoints and descriptors of all the logos in the training set are extracted and stored in a database. Then, we extract SIFT keypoints and descriptors obtained from the grocery shelf image, which are matched with the already extracted keypoints of the logos that were stored in the database. The matching is done using a simple FLANN matcher. A number is associated with every logo in the database, which represents the number of keypoints from the shelf that match with its keypoints. These numbers are normalized and sorted. Then, the top k logos were shortlisted based on the number of keypoints that got matched with the shelf. In Fig. 2, top few logos corresponding to the shelf have been shown in a sorted order. Here, every keypoint of the shelf is matched with the closest keypoint among the logos. This makes the algorithm more scalable as it avoids going through each logo and checking whether it is present on the shelf.

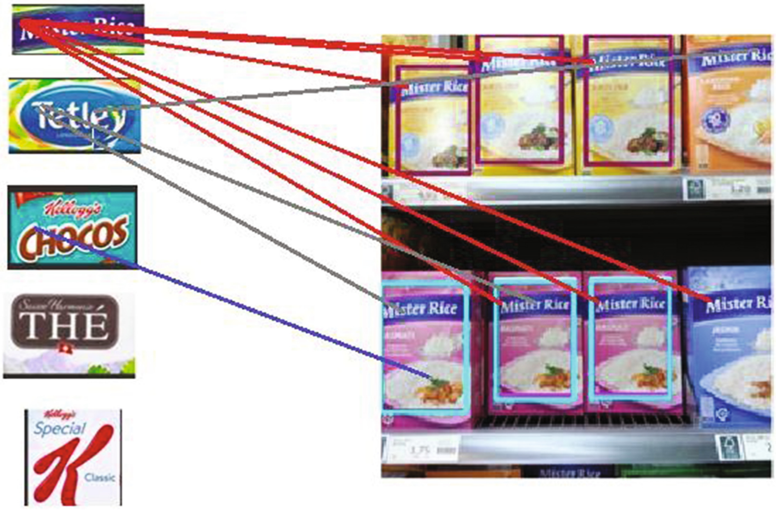

Fig. 2. A pictorial illustration of choosing the top k logos in the logo detection step. The red lines depict the matches between the keypoints in the shelf and the first logo. Similarly, the grey and the blue lines are matches with the next two logos

Once the top-k logos are decided, we perform a scale invariant template match. The template here is the logo which is matched with the shelf image varied over different scales. The template match will give the best matches for a given template in the form of rectangular bounding boxes around the matches found. These bounding boxes have to be further filtered to reduce the number of false positives. This can be done by adding a verification step that involves computing the SIFT keypoints and

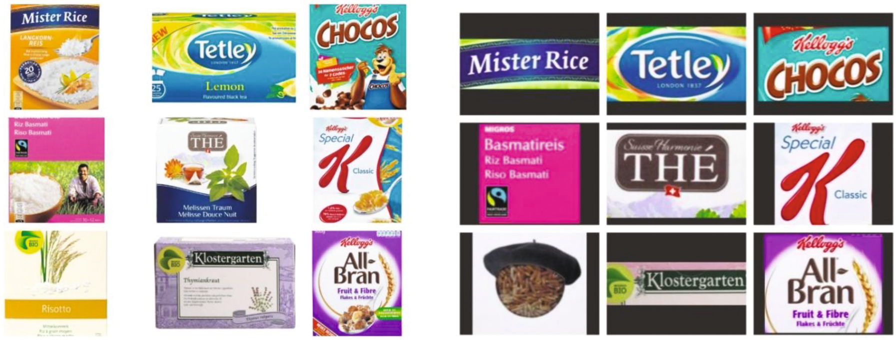

Fig. 3. (a) Some products from the grocery products dataset of three categories (b) Logos of some products

descriptors within each bounding box and matching it with the template using the Fast Library for Approximate Nearest Neighbour (FLANN) [12–14] matcher. FLANN automatically chooses the best algorithm and optimum parameters for the nearest neighbour search. These matches are then used to compute the homography and check for the presence of the correct logo.

After detecting the logos, a certain region around each logo is extracted for the product recognition step. To narrow down to a particular product containing the detected logo, it is important to consider the colour information. This colour information was captured through three indices obtained by performing a template match in Red, Green and Blue (RGB) space separately. These three indices are summed up and used to determine the best match among the products with the same logo. We also experimented with Lab colour space (without the luminance component) in place of the RGB space. This approach has been formulated as a flowchart in Fig. 4.

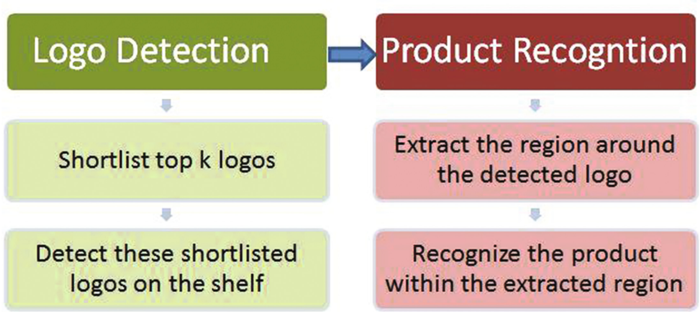

Fig. 4. Flowchart depicting the Methodology

3.2 Without Using Logo Detection

As opposed to the above approach, an implementation of grocery products recognition was attempted without the detection of logos. As mentioned before, SIFT keypoints and descriptors of all the products as well as that of the grocery shelf image are extracted. The top k products are shortlisted and a verification step is performed as discussed earlier.

4 Experimental Setup and Results

4.1 Dataset

The algorithm was tested on a subset of the Grocery Products Dataset [4]. One of the important features of this dataset is that the training images were taken under ideal conditions and the testing images were taken in real life scenarios using a mobile phone. Three categories were picked from this cross-dataset, which comprise 238 products and 80 images. The categories were picked such that they mainly comprise of non-deformable packaged products. The split within the chosen categories is shown in Table 1. Also the number of grocery shelf images that were used for testing our algorithm is listed in the table for each category.

Table 1. Statistics about various categories chosen for testing the algorithm

Category	No. of Logos	No. of Products	No. of Shelves
Cereal	58	138	25
Rice	21	31	22
Tea	9	69	32

This dataset provides training data consisting of one image per product label. These products are further grouped together based on their logos and a set of distinct logos are extracted manually for each category, which serve as input for the algorithm. Some of the products and logos from the three categories chosen have been shown in Fig. 3.

4.2 Results

The performance of our proposed approach was evaluated on the three categories using mean Average Precision (mAP) as the measure on two different variants of the system: (i) the logo detection step, and (ii) the full system.

The mAP is computed as the average precision of all the test images within a category.

$$precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Where, TP is the number of true positives and FP is the number of false positives

Table 2. Accuracy for logo detection step

Category	Accuracy (mAP)
Cereal	80.58
Rice	90.22
Tea	71.86

Table 3. Accuracy for our grocery products recognition system

Category	Baseline accuracy (mAP)	Accuracy (mAP)
Cereal	49.43	60.21
Rice	25.083	61.15
Tea	21.34	30.70

The results for both the steps have been tabulated in Tables 2 and 3.

The accuracy and the run-time efficiency increase considerably when logo detection is incorporated as a step. This is because when we do logo detection first, the number of probable matches are reduced.

So, as opposed to matching a large number of templates against a test image, only a small subset of those templates are matched against the test image now. All the results mentioned above are obtained using the RGB colour space in the later part of the algorithm. When experiments were conducted using the Lab colour space without the luminance component, the accuracy of the full system decreased by 5% on an average.

All the results were performed on Intel Core i5 processor with clock speed of 2.2 GHz and 4 GB RAM. The product recognition algorithm was implemented in python, using the OpenCV 2.4.11 library. Our approach (using logo detection) takes 60 s on an average per shelf and outperforms the baseline by nearly 10 s.

5 Conclusion

In this paper, we show how the novelty of logo recognition helps to improve product recognition accuracy and the run-time on the Grocery Products Dataset. Also, the fact that we use only one image per product as training under cross-dataset settings makes it more realistic and adds to the novelty. Given the minimalist requirements of the dataset, it is convenient to alter and expand the dataset as and when required.

Future work could be to make the product recognition affine transformation invariant, so that it will be able to recognise products with deformable packaging, for example, bakery products.

References

1. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
2. Swain, M.J., Ballard, D.H.: Indexing via color histograms. In: Active Perception and Robot Vision, pp. 261–273. Springer, Heidelberg (1992)
3. De la Torre, F., Black, M.J.: Robust principal component analysis for computer vision. In: Eighth IEEE International Conference on Computer Vision, 2001, ICCV 2001, Proceedings, vol. 1, pp. 362–369. IEEE (2001)
4. George, M., Floerkemeier, C.: Recognizing products: a per-exemplar-label image classification approach. In: Computer Vision–ECCV 2014, pp. 440–455. Springer (2014)
5. Guo, G., Jiang, T., Wang, Y., Gao, W.: Finding multiple object instances with occlusion. In: 2010 20th International Conference on Pattern Recognition (ICPR), pp. 3878–3881. IEEE (2010)
6. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. **24**(4), 509–522 (2002)
7. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Comput. Vis. Image Underst. **89**(2), 114–141 (2003)
8. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Computer Vision–ECCV 2006, pp. 404–417. Springer, Heidelberg (2006)
9. Collet, A., Berenson, D., Srinivasa, S.S., Ferguson, D.: Object recognition and full pose registration from a single image for robotic manipulation. In: IEEE International Conference on Robotics and Automation, 2009, ICRA 2009, pp. 48–55. IEEE (2009)
10. Winlock, T.: ShelfScanner: toward real-time detection of groceries for the visually impaired (2010)
11. Skoczylas, M.: Detection of positions and recognition of brand logos visible on images captured using mobile devices. In: 2014 International Conference and Exposition on Electrical and Power Engineering (EPE), pp. 863–868. IEEE (2014)
12. Muja, M., Lowe, D.G.: Scalable nearest neighbor algorithms for high dimensional data. IEEE Trans. Pattern Anal. Mach. Intell. **36**(11), 2227–2240 (2014)
13. Muja, M., Lowe, D.G.: Fast matching of binary features. In: 2012 Ninth Conference on Computer and Robot Vision (CRV), pp. 404–410. IEEE (2012)
14. Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. In: VISAPP, vol. 1(2) (2009)

http://www.springer.com/978-3-319-56990-1

Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016
Volume 2
Bi, Y.; Kapoor, S.; Bhatia, R. (Eds.)
2018, XIII, 1070 p. 472 illus., Softcover
ISBN: 978-3-319-56990-1