Content-Based Image Retrieval with LIRe and SURF on a Smartphone-Based Product Image Database

Kai Chen[1] and Jean Hennebert[2]

[1] University of Fribourg, DIVA-DIUF, Bd. de Pérolles 90, 1700 Fribourg, Switzerland
kai.chen@unifr.ch
[2] University of Applied Sciences, HES-SO//FR, Bd. de Pérolles 80, 1705 Fribourg, Switzerland
jean.hennebert@hefr.ch

Abstract. We present the evaluation of a product identification task using the LIRe system and SURF (Speeded-Up Robust Features) for content-based image retrieval (CBIR). The evaluation is performed on the Fribourg Product Image Database (FPID) that contains more than 3'000 pictures of consumer products taken using mobile phone cameras in realistic conditions. Using the evaluation protocol proposed with FPID, we explore the performance of different preprocessing and feature extraction. We observe that by using SURF, we can improve significantly the performance on this task. Image resizing and Lucene indexing are used in order to speed up CBIR task with SURF. We also show the benefit of using simple preprocessing of the images such as a proportional cropping of the images. The experiments demonstrate the effectiveness of the proposed method for the product identification task.

Keywords: product identification, CBIR, smartphone-based image database, FPID, benchmarking.

1 Introduction

There is now a growing interest for mobile applications allowing a consumer to automatically identify a product and access information such as prices comparisons, allergens or ecological informations. For usability reasons, the use case involves that the user takes a picture of the product of interest, from which an identification procedure derives the most probable product label. We have build such a product identification mobile application, namely GreenT.

In our work, we focus on consumer product identification using camera phone devices. Different approaches have attempted to identify the product using a detection and recognition procedure of the bar code. While overall efficient, such approaches suffer from two drawbacks. First, some products do not have bar codes such as luxury products or products with small packaging. Second, many mobile phones do not present autofocus capability resulting in a low-pass blurring effect when capturing a close shot of the bar code. Our approach is therefore to attempt recognizing the product from a product image using CBIR systems. As illustrated in Figure 1, such systems typically use a picture as query and find similar images from a reference database. Generally speaking, a CBIR system would proceed in five steps: image preprocessing, feature extraction, relevant images retrieval, post-processing, and closest product id identification.

J.F. Martínez-Trinidad et al. (Eds.): MCPR 2014, LNCS 8495, pp. 231–240, 2014.

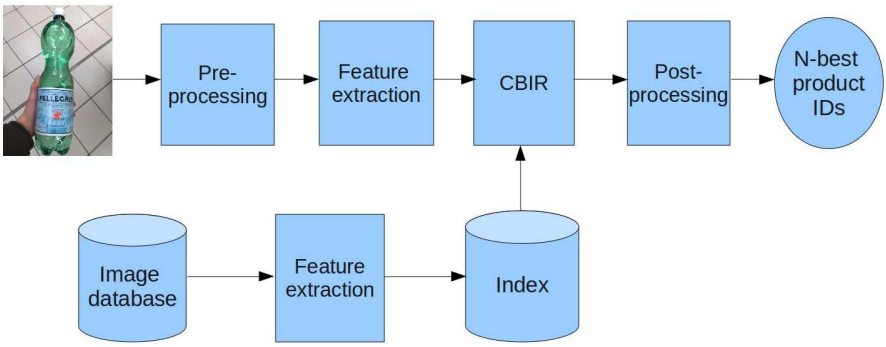


Fig. 1. General operations of a CBIR system using the FPID database

Many CBIR systems have been proposed and described in the literature, for example QBIC [4], GIFT [6], and FIRE [3]. In our work, we have chosen to use the open source java Lucene Image Retrieval (LIRe) library[1] [5]. Local feature SURF (Speeded-Up Robust Features) [1] is also employed for the product identification task. In this paper, we report on the evaluation of LIRe and SURF using the Fribourg Product Image Database (FPID) that has been released recently [2]. FPID is a smartphone-based image database. It contains more than 3'000 pictures of consumer products captured in supermarkets with various regular smart-phone cameras. Using the evaluation protocol proposed with FPID, we explore the performance of different preprocessing and feature extraction using LIRe and SURF. Compared with global feature methods, SURF takes more time. On the other hand, the size of image has a huge impact on the time taken for SURF approach. Therefore, image resizing and feature indexing approaches are used to speed up SURF in our system. The experimental results demonstrate the effectiveness of the proposed method for the product identification task.

This paper is organized as follows. We introduce the FPID smartphone-based image database and our CBIR protocol in Section 2. In Section 3 we present baseline CBIR performance as well as several improvements that we could obtain through the parameters of the feature extraction and preprocessing of the images. Section 4 presents conclusions and future works.

2 FPID and CBIR Evaluation Protocol

The evaluation of CBIR systems requires two elements. First, a database of reference images must be provided together with verified ground truth values for each images. Second, an evaluation protocol must be clearly defined, so that different teams can run their algorithms and compare their results.

For the work reported here, we used the "Fribourg Product Image Database" (FPID) [2]. This database has been recently released to the scientific community. Currently, FPID contains more than 3'000 pictures of retail products that can be found in

[1] http://www.semanticmetadata.net/lire/

Swiss and European supermarkets. The set of images covers about 350 products spread into 3 families: bottled water, coffee, and chocolate. Each product has at least one image in the database and the most popular products have about 30 images. The images have been captured using various mobile phones in different supermarkets without any control of the illumination. For identical products, the image features may therefore show a large variability. The ground truth information is the product label expressed as a character string, i.e., if two images have the same product label, then they are considered as relevant for a CBIR task. The product label is a character string uniquely identifying the product brand and model. The ground truth also includes the mobile phone brand/model, the shop name and its location that allows for some advanced error analysis. Some images taken from FPID are illustrated on Figure 2.

(a) 169, Nokia n95, Manor Fribourg

(b) 497, Samsung g600, Coop Fribourg

(c) 1052, Sonyericson w880i, Migros Fribourg

(d) 1216, Sonyericsson w880i, Migros Fribourg

(e) 23, nokia n95, Migros Fribourg

(f) 2041, Nokia N78, Manor Fribourg

Fig. 2. Example images from FPID with the image id, the device name and location of the acquisition

In this paper, we follow strictly the evaluation protocols for product identification that are proposed with FPID [2]. These protocols are based on a subset S of 1200 images including 100 different products with 12 images per product. From the set S, disjoint

sets T and Q for training and query are defined. The protocols are said to be *closed-set* as all query images belong to a product category that is represented in the training sets. In other words, the proposed protocols do not evaluate rejection performances of CBIR systems where a query image has zero corresponding relevant images.

As illustrated in Figure 3, different training sets T_k are defined, k representing the number of images per product in the set. All the training sets are balanced with, e.g., the training set T_4 containing exactly 4 images per products for a total of 400 images. Lists of images are provided on the web site of FPID[2] for $T_1, T_2, \ldots, T_{10}$. In a similar manner, a query set Q_2 including 200 images with 2 images per product is defined. Q_2 is of course disjoint to all the training sets T_n.

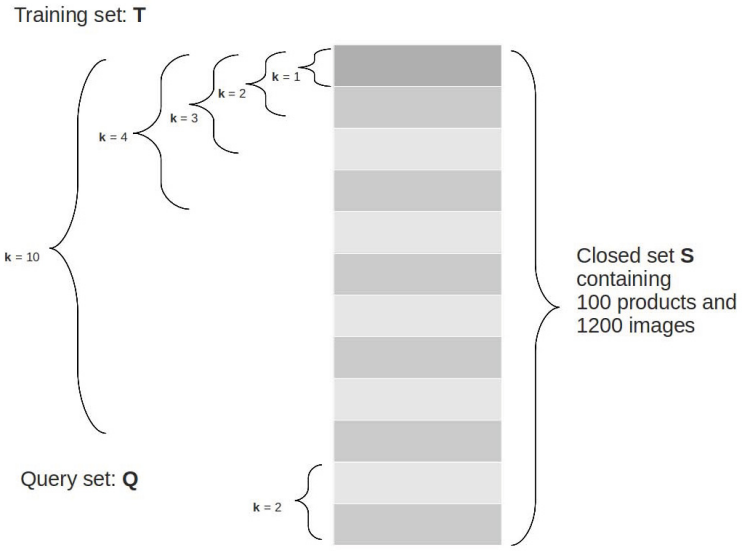


Fig. 3. FPID evaluation protocols

With FPID, it is proposed to report system performance using the recognition rate $RR_I(n)$ considering the n-best retrieved set of images. The rate $RR_I(n)$ is computed as the ratio of positive matches divided by the total number of queries. A match is considered positive when, for a given query image, there exists at least one relevant image in the retrieved n-best set of images. If there is no relevant image in the retrieved n-best set of images then this is a miss. Increasing the value of n in the n-best retrieved set of images will make the task easier. When n is equal to 1, the rate $RR_I(1)$ is actually equal to the *precision at 1* or $P@1$, frequently measured when benchmarking CBIR systems. For the experiments reported in this work, we used $n \in \{1, 2, 5, 10, 15, 20\}$. In a similar manner, we also measure the recognition rate $RR_P(n)$ which is the recognition rate considering the set of n-best retrieved product categories. In this case we retrieve, for

[2] `http://diuf.unifr.ch/diva/FPID/`

a given query image, the set of n-best images in which we keep only one representing image per product, the other one being discarded.

3 System Description and Results

Our CBIR system is based on the open-source library LIRe and SURF. The LIRe library offers different feature extraction possibilities that we have explored in this work. One of the difficulties of CBIR systems is indeed to select the most suitable feature extraction technique regarding the specificities of a given task. We have also explored the impact of the n value on the RR_I and RR_P performances as described in Section 2. We applied some meaningful preprocessing of the images and method of choosing most suitable feature that have lead to improvements over our baseline results. The SURF is a robust algorithm for local, similarity invariant representation and comparison. It outperformed all global features available in LIRe, but on the other hand it suffers the speed. We observed that the size of image has a direct impact on time taken. Image resizing and feature indexing approaches are taken to speed up the task. All our experiments are achieved in DALCO High Performance Linux Cluster with 64GB per Compute node in the University of Fribourg.

Global Features Comparison. Several global feature extraction are available in LIRe. Figure 4 shows the performances for the different features and the evolution of RR_I as a function of n in the set of n-best retrieved images. In LIRe, image is presented in feature vectors. For a given query image q, a training set T and feature f, we compute the distance $d_{q,i,f}$, such that $d_{q,i,f}$ is the Euclidean distance between feature vector of q and i^{th} images in T. Then we sort images in T by ascending order of this distance value. The first n images in T are considered as n-best relevant images to q.

As expected, the performances increase when n is getting bigger. We can also observe that the image feature "MPEG-7 edge histogram" gives the best performance. The second best feature is the "MPEG-7 color layout". Such results are actually meaningful if we consider that product packagings have purposely different shapes and colors that form their marketing identity. The texture features such as "Tamura"or "Gabor" seem to characterize less efficiently the products. Overall, the RR_I performances are not so much satisfying with, for example, 72% measured with $n = 10$ for "MPEG-7 edge histogram". We can also observe that the performances are very low when n tends to 1, which is a clear indication of the difficulty of the task with a probable explanation to find in the large variability of image characteristics.

Suitable Feature Selection. We observe that for different kind of products, same feature has different performance. In order to find the most suitable feature, for a given query image q and i^{th} images in T, the distance is re-defined as $d_{q,i} = \min_{f \in F} d_{q,i,f}$, where F is the feature set available in LIRe. After sorting the distances of ascending order, first n^{th} images are considered as the relevant images to q. The results shown in Figure 5a indicate us this method improve the performance slightly.

Products vs. Image Recognition. We show on Figure 5b the comparison of RR_I and RR_P rates for the test (T_{10}, Q_2) using "Combine global features" approach which is

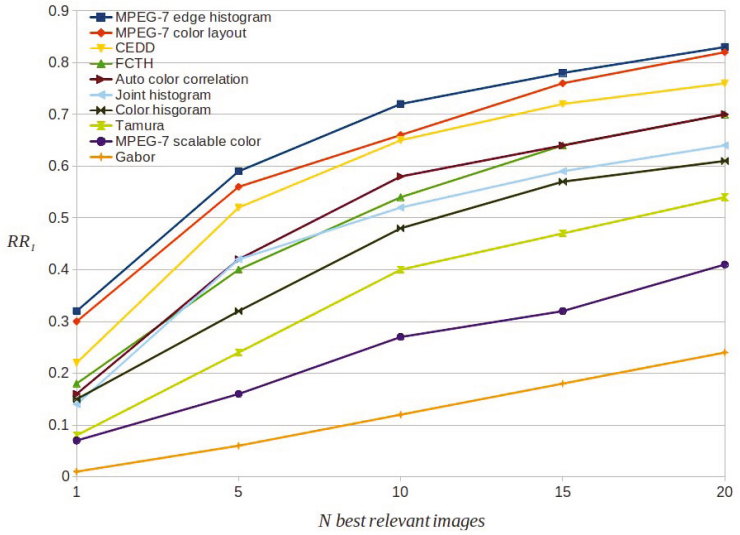
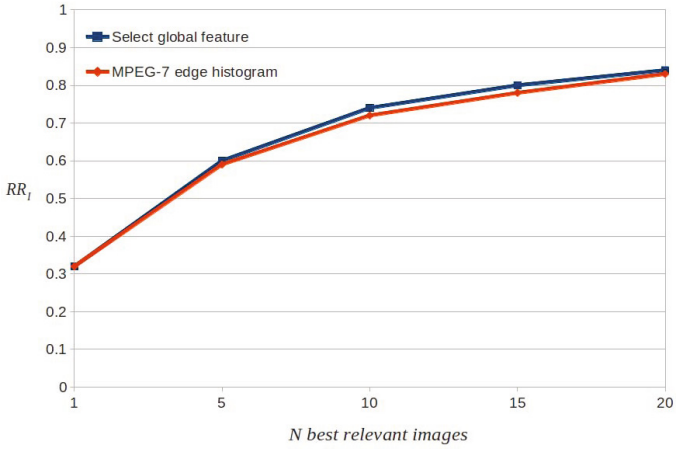


Fig. 4. RR_I evolution using (T_{10}, Q_2) for different features

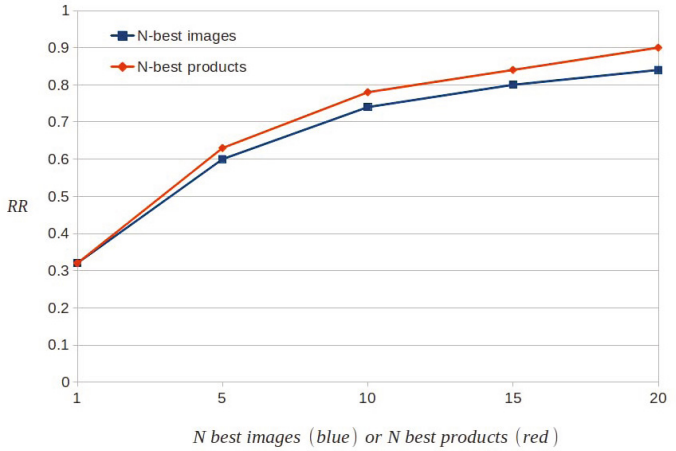
described in previous section. As expected, RR_P rates are systematically higher as the task is easier. From the results we can observe an increase of 6% between $RR_I(20)$ and $RR_P(20)$. Similar results have been observed for all features.

Image Cropping. In our task, we can reasonably assume that most of the information to identify the products is located at the center of the image. The outer parts are, for most of the FPID images, showing the floor of the supermarket. We therefore attempted to remove these outer parts by implementing a proportional cropping. As illustrated in Figure 6a, we define the cropped area as a rectangle where the top left coordinate is defined by $x = (p \times width)/2$ and $y = (p \times height)/2$ where p is expressed as a percentage of width and height that is removed. The bottom right coordinate is computed in a similar manner. Figure 6b shows the evolution of $RR_P(10)$ as a function of p using our "Combine global feature" method. Interestingly, we see a significant gain of performance up to 86%, 85%, 85% recognition rate when 60%, 30%, 10% of the outer parts of the images are removed.

Local Features. SURF is a scale and rotation invariant detector and descriptor. The SURF algorithm contains three steps: (1) interest points detection. (2) building the descriptor associated with each interest points. (3) descriptor matching. The first two steps are rely on scale-pace representation, and on first and second order differential operators. All the three steps are speed-up by using integral image and box filters. For details of SURF, we refer to [1]. Table 1 gives the results on (T_{10}, Q_2) with cropping factor $p = 0.2$. The results indicate that SURF improve the performance drastically, e.g. $RR_P(3)$ increase from 52% to 94%.



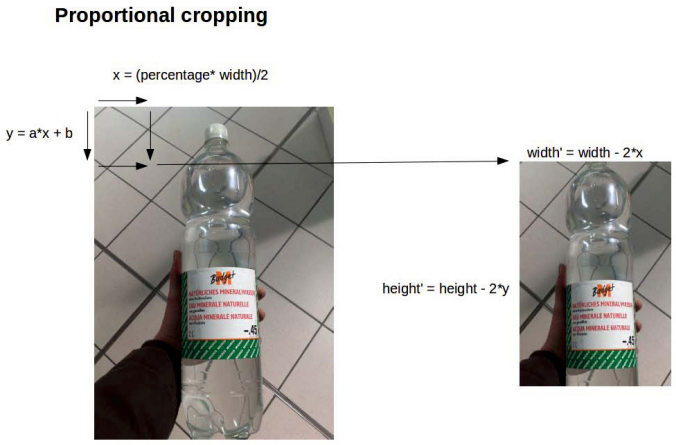
(a)



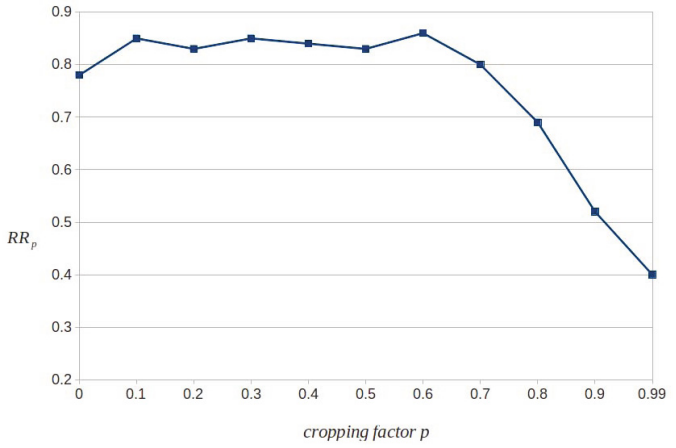
(b)

Fig. 5. 5a: RR_I evolution using (T_{10}, Q_2) for "combine global features" and "MPEG-7 edge histogram". 5b: $RR_I(10)$ and $RR_P(10)$ rates evolution for (T_{10}, Q_2) using as approach "combine global features"

However, compare to other global features methods, SURF suffers from the time of interest points extraction and matching. For a given query image and training set T_{10}, the average time used for global features is less than 2 seconds, on the other hand, by applying SURF method, it takes about 240 seconds. In order to reduce the time taken for SURF method, we proportionally reduce the image size. For a given image I, such that $width_{I'} = width_I \times w$, where $width_{I'}$ is the resized image width and w is the resizing factor. Figure 7a illustrates the impact of image size on the time taken. The impact of image size on performance is given in Figure 7b. We observe that reducing image



(a)



(b)

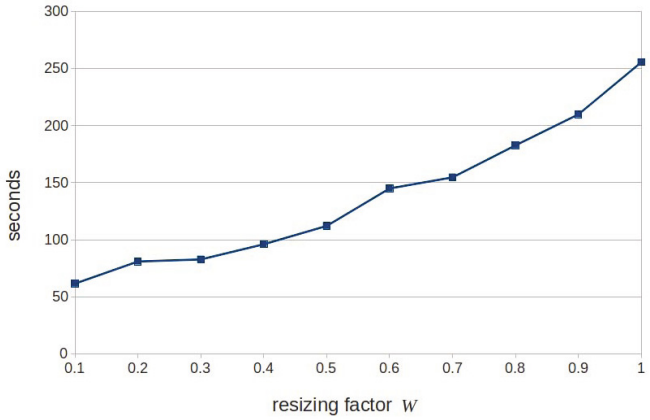
Fig. 6. 6a: Proportional image cropping. 6b: $RR_P(10)$ evolution as a function of p, the proportional cropping factor, for (T_{10}, Q_2) using "combine global features" approach.

size by choosing $w = 0.7$, we still have a high accuracy $RR_P(3) = 92\%$ compare to $RR_P(3) = 94\%$ with the original image size, but the time has been reduced from 240 seconds to 158 seconds. Inspired by LIRe, we use Apach Lucene[3] as feature indexing tool to speed-up SURF. For each image in T, we first extract the interest points with SURF, then these points are saved into a index by using Lucene. With this approach, we reduce the time from 158 to 80 seconds.

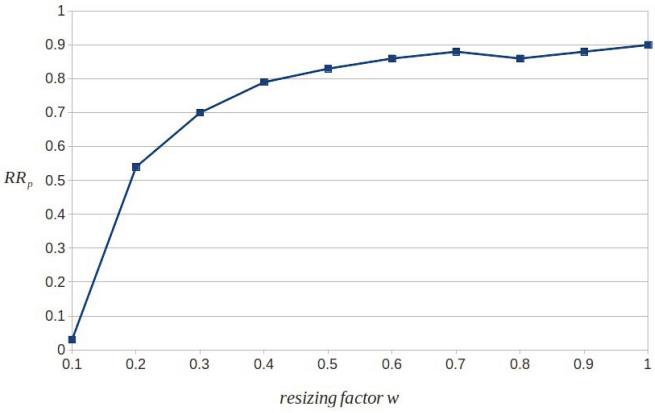
[3] `http://lucene.apache.org/core/`

Table 1. Comparison ”combine global features” method with SURF

N-best products	1	2	3
Combine global features	0.32	0.43	0.52
SURF	0.72	0.90	0.94



(a)



(b)

Fig. 7. 7a: Impact of image size on time taken for SURF. 7b: Impact of image size on performance for SURF.

4 Conclusion

We presented a product identification system based on CBIR. Experiments are made on LIRe (an open-source CBIR system part of Lucene) and SURF. The performances of

the system have been evaluated using the protocols proposed with the FPID database. We have also found that the global features based on color layouts and edge histograms are the most performing one on our product identification task. Considering that product packagings have purposely different shapes and colors for marketing identity, such results are probably meaningful. Several improvements have been proposed over the baseline use of the LIRe system, including proportional image cropping and global features combination. By using SURF, product recognition performance increased to 94% (with $n = 3$ in the n-best retrieval products), to be compared with the 52% obtained using baseline configuration of LIRe with "MPEG-7 edge histogram". By resizing the images and using Lucene for indexing, for a given image, we reduce the time from 240 to 80 seconds, with 92% accuracy of product recognition (with $n = 3$ in the n-best retrieval products) by resizing factor $w = 0.7$. Future works will probably to increase the speed of SURF in our system, since for our product recognition system, nearly 1 minute per query is too long for a mobile application. Another interesting approach would be to combine CBIR with camera-based OCR as there are frequently texts on the labels of the products.

Acknowledgements. This work was partly supported by the grant Green-T RCSO ISNet from the University of Applied Sciences HES-SO//Wallis, by the HES-SO//Fribourg and by the University of Fribourg.

References

1. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding (CVIU) 110(3), 346–359 (2008)
2. Chen, K., Hennebert, J.: The Fribourg Product Image Database for Product Identification Tasks. In: Chen, K., Hennebert, J. (eds.) IEEE/IIAE International Conference on Intelligent Systems and Image Processing (ICISIP), pp. 162–169 (2013)
3. Deselaers, T., Keysers, D., Ney, H.: FIRE – flexible image retrieval engine: ImageCLEF 2004 evaluation. In: Peters, C., Clough, P., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B. (eds.) CLEF 2004. LNCS, vol. 3491, pp. 688–698. Springer, Heidelberg (2005)
4. Faloutsos, C., Equitz, W., Flickner, M., Niblack, W., Petkovic, D., Barber, R.: Efficient and Effective Querying by Image Content. Journal of Intelligent Information Systems 3, 231–262 (1994)
5. Lux, M.: Content based image retrieval with LIRe. In: Proceedings of the 19th ACM International Conference on Multimedia, pp. 735–738 (2011)
6. Squire, D.M., Müller, W., Müller, H., Raki, J.: Content-Based Query of Image Databases, Inspirations From Text Retrieval: Inverted Files, Frequency-Based Weights and Relevance Feedback. Pattern Recognition Letters, 143–149 (1999)