The Fribourg Product Image Database for Product Identification Tasks

Kai Chen[a,*], Jean Hennebert[b]

[a]DIVA Group, University Of Fribourg, Bd de Pérolles 90, Switzerland
[b]HES-SO, University of Applied Sciences, Bd de Pérolles 80 – CP 32, Switzerland

*Corresponding Author: kai.chen@unifr.ch

Abstract

We present in this paper a new database containing images of end-consumer products. The database currently contains more than 3'000 pictures of products taken exclusively using mobile phones. We focused the acquisition on 3 families of product: water bottles, chocolate and coffee. Nine mobile phones have been used and about 353 different products are available. Pictures are taken in real-life conditions, i.e. directly in the shops and without controlling the illumination, centering of the product or removing the background. Each image is provided with ground truth information including the product label, mobile phone brand and series as well as region of interest in the images. The database is made freely available for the scientific community and can be used for content-based image retrieval benchmark dataset or verification tasks.

Keywords: CBIR, image retrieval, image database, FPID, benchmarking, product identification.

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.

To be able to identify the object in the images becomes an interesting topic. Our project GreenT is based on the Content-based image retrieval (CBIR) technology. To achieve the product identification task, the CBIR technology is a suitable solution, it is efficient and scalable.

CBIR is any technology that helps to organize digital picture archives by their visual content. With the argument number of images around us, images become an important media of information transfer. People uses CBIR to retrieve the desired images from a large database by their feature such as color, texture or shape. It is nowadays an active field of research[1,2,13].

In our product identification project, we focused on three product categories: bottled water, chocolate, coffee. To be able to retrieve the information of the product, user takes an image on the product by holding it in the hand with his mobile phone, then sent it to our server, by comparing its visual content with the images in the database, we get the *N*-best relevant images, finally we filter the results and send the useful information about the product (e.g. comparison the price, allergens) to the user. Figure 1 gives the approach of GreenT project. In general, there are 5 steps: image preprocessing, feature extraction, retrieve the *N*-best relevant images, post-processing and show user the information on the relevant products.
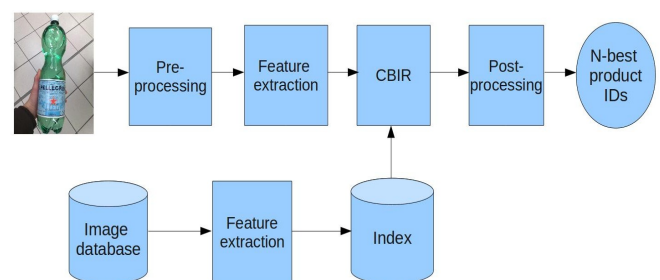

Fig. 1. Approach of GreenT product identification system.

To provide a image database for the image retrieval tasks and store the product information of each image, we have built a new image database, namely Fribourg Product Image Database (FPID). FPID could also be used as the benchmark set for any CBIR method evaluation. As we know, to be able to compare different image retrieval

algorithms, we need the benchmarking. However, few image sets are available. One widely in CBIR method evaluation is Corel dataset. However, Corel dataset has several limitations. First, it is commercial. The ground truth is not provided which is considered as the biggest limitation of Corel. That makes the CBIR evaluation difficult. Another signification problem is many semantically similar images in Corel are also very similar in term of their visual features, that makes CBIR task easy be achieved[7,11].

In this paper, we aim to introduce our new image database FPID. This paper is organized in the following way: A general overview of the existing benchmark databases is given in Section 2. In Section 3, we describe the FPID image database. Section 4 addresses the challenges of using FPID for CBIR. Section 5 concludes the paper.

2. Benchmark databases for CBIR

To be able to compare different image retrieval algorithms, we need the benchmarking. Several image databases have been constructed for this objective. In this section, we give an overview of the existing benchmark databases for CBIR evaluation. Benchmark databases are used by the scientific community to evaluate quantitatively the performance of CBIR systems. A benchmark database should contain not only images, but also the ground truth value for each image. To cover a wide range of different CBIR tasks, various domains images databases have been constructed[2]. Table 1 gives a summary of these benchmark databases.

Table 1. Overview of existing benchmark image databases for CBIR evaluation.

database	images	queries	avg rel	type	objects
WANG	1'000	1'000	99.0	color	various
UW	1'109	1'019	59.3	color	various
IRMA	12'677	1'733	520.2	gray	radiography
ZuBuD	1'005	115	5.0	color	building
UCID	1'338	262	3.5	color	various
FPID	3'159	200	8.0	color	products

The WANG database is the subset of 1'000 images of the Corel stock photo database[3,14]. The images are divided into 10 classes and each class contains 100 images. Figure 2 illustrates one example image from the WANG database with the class label from the 10 classes (e.g. Africa people and village, Beach, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains and glaciers, Food). The image size is 256 x 384 or 384 x 256 pixels. In WANG database,

the test images set and training images set are the same set, i.e. the whole set. To evaluate a CBIR system, user have to pick an image from the database and the rest 99 images in the same class are considered as the relevant images. Since each image has 99 relevant images in the database, it is easy to get a high precision value, however to get a high recall value is difficult.
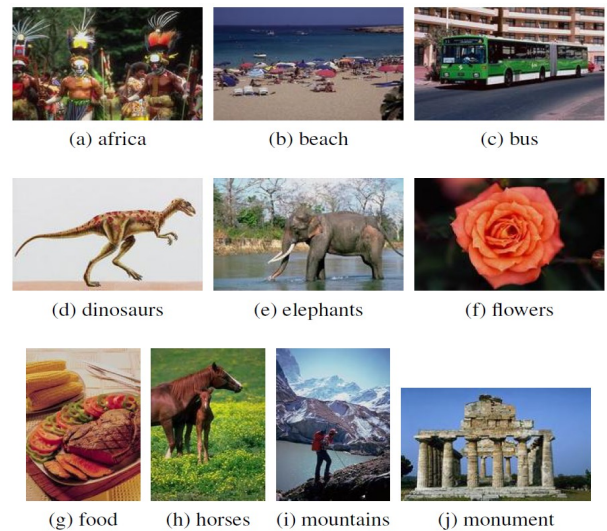

(a) africa   (b) beach   (c) bus

(d) dinosaurs   (e) elephants   (f) flowers

(g) food   (h) horses   (i) mountains   (j) monument

Fig. 2. Example images of WANG database

(a) Water Buildings Boats Sailboats Green Trees Sky   (b) Buildings Leafless Trees Clear Sky   (c) Water Buildings Boats Leafless Trees Sky Snow Mountains

(d) Clear Sky Water Mountains Rockes   (e) Clear Sky Water Mountains Rockes Greens   (f) Clear Sky Water Mountains Rockes

(g) buildings sky car street lantern   (h) buildings sky street trees grass   (i) buildings car people sky trees

Fig. 3. Example images with annotation from UW database

The UW database is created at the University of Washington which consists of 1'109 images. These images are divides into 22 classes. These images are partly

annotated using keywords. The remaining images have been annotated by the RWTH group of Aachen University. Figure 3 consists of some example images with their annotation from the UM database. The complete annotation consists of 6'383 words with a vocabulary of 352 unique words. The maximum number of keywords per image is 22 and minimum number of keywords per image is 1. On average, each image has about 6 words of annotation. If two images contain common keywords in the annotation, then they are considered as relevant. Therefore, like WANG database it is relatively easy to retrieve similar images for a query. In other words, precision is high, but the recall is low, because for a given image, it has many relevant images in the database, i.e. these images have common keywords in their annotation. For example, in Figure 3, all the images contains the keyword "sky", therefore they are considered as relevant to each other, but they are not visually similar.

The IRMA database consists of anonymous radiography, which have been arbitrary selected from routine at the Department of Diagnostic Radiology, Aachen University of Technology (RWTH). The image represents different ages, genders, view positions and pathologies. The images were downscaled to 512 x 512 bounding box. maintaining the original aspect ratio. They were classified according to the label of IRMA. Based on this label, 193 categories were defined. For 12'677 images, these categories are provided. The remaining 1'733 images without label are used as test data for the ImageCLEFmed 2009 competition. All images are gray value images. Figure 4 illustrates some example images with their annotation from IRMA.
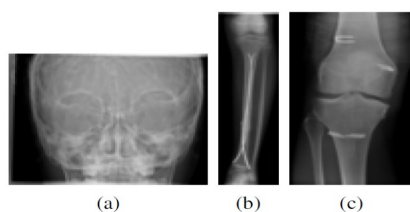

Fig. 4. Example images with annotation from the IRMA database the annotations are: (a) facial cranium musculosceletal; (b) lower leg musculosceletal; (c) knee musculosceletal

The "Zurich Buildings Database for Image Based recognition" (ZuBuD) database has been created at the Computer Vision Lab of Swiss Federal Institute of Technology in Zurich[9,10]. This database contains 1'005 color images of 201 buildings, 5 images per buildings and 115 images are chosen as test images. Each query image contains one of the buildings from the main part of the database, however the imaging conditions do not match

exactly. These query images have been acquired at different view and under varying illumination condition. For a given image, only images containing exactly the same building are considered as relevant. Figure 5 consists one query with 5 relevant database images in different view. This database can be used to identify buildings in pictures, i.e. by comparing the image taken with a mobile phone camera with the images in the database, we get the most relevant images set, then we can provide certain information about the building to the user.

Fig. 5. One query image and 5 images of the same building with different view from ZuBuD-database.

The Uncompressed Color Image Database (UCID database) is a benchmark dataset for image retrieval where all images were captured and are available in uncompressed form[7,8]. UCID is built for evaluating the impact of compression algorithms. Currently it consists of 1'338 uncompressed TIFF (Tagged Image File Format) images on a variety of topics including natural scenes and man-made objects, both indoors and outdoors. All images were taken with a Minolta Dimage 5 digital color camera which, in contrast to many other models, also allows images to be captured in uncompressed form. The dataset is used for the evaluation of image retrieval techniques that operate directly in the compressed domain and to investigate the effect of image compression has on the performance of CBIR methods. A subset of 264 images as query images are manual assigned. Most query images will have more than one corresponding model image. Figure 6 illustrates a query image with assigned matches. For each image a list of relevant images is associated, these relevant images are considered as the ground truth. The lists of the relevant images are given by a text file which is available online[1]. The number of relevant images per image is between 1 to 16. UCID is similar to the UW database. The two databases

---

1   http://vision.cs.aston.ac.uk/datasets/UCID/ucid.html

consist of vacation images. The problem is that there are high visual similar images in the database but they are not considered as relevant. Therefore it is difficulty to get high precision, but since the number of relevant images is small, it is relatively easy to get the high recall value[2].

Fig. 6. One query image with its relevant images from the UCID database

We have built a new image database that we called the "Fribourg Product Image Database" (FPID). Currently FPID consists of 3'159 color images. These images are from 353 products from 3 categories: bottled water, coffee and chocolate. Each product has at least one image in the database, the most popular products have 34 images. These images have been taken using different mobile phones with different size of pixel settings and from different supermarkets in Switzerland. These images are saved in JPEG format and are resized to 450 x 600 or 800 x 600 which depends on the original image's ratio of width and height. The ground truth information is the product id, i.e. if two images have the same product id, then they are considered as relevant. The information of the image also includes image id, mobile phone name, shop name and location. Figure 7 illustrates some images from FPID.

Comparing with the existing image databases, the ZuBuD database is the most similar one to FPID. The images are taken by camera on the mobile devices under varying illumination, so the imaging conditions do not match exactly, however the images in ZuBuD have been acquired at different viewpoint. Each image of the two databases contains one significant object. In FPID, each image contains a product, on the other hand, in ZuBuD, each image contains a building. The two databases have been used as the database for identification system. ZuBuD database can be used to evaluate the task of buildings identification, i.e. by comparing the visual content of the image, users get a list of similar buildings with the

information, e.g. name of the building, location, address. FPID can be used for the product identification task. By comparing the visual content of the image, we return a list of similar products with their information. The two databases can be used as benchmarking for image retrieval algorithms performance comparison.
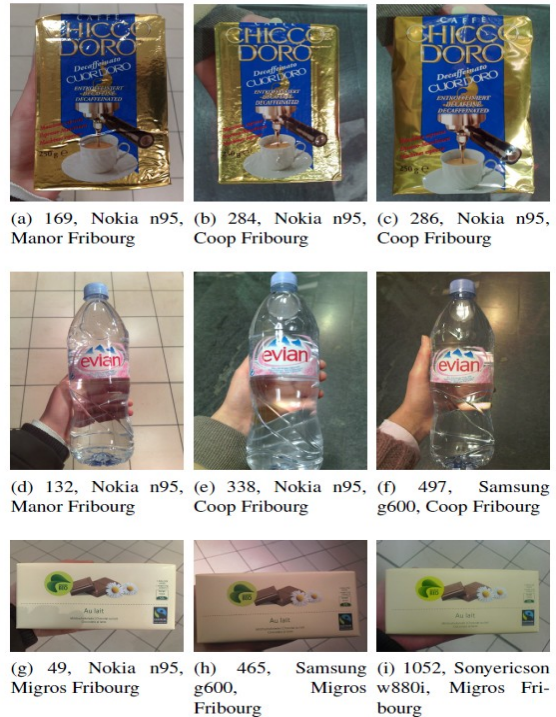

Fig. 7. Some images from FPID of products "155, caffé chicco d'oro decaffeinato cuor d'oro machines espresso 250g", "evian eau minérale naturelle 1l" and "bio migros au lait fairtrade 100g". The caption of each image contains "id", "device name", "location".

3. The Fribourg Product Image Database (FPID)

In this section, we present the detail about FPID and how we generate the closed set for CBIR methods evaluation.

3.1 Data acquisition

All the images of FPID are captured by 9 different mobile phones in different supermarkets. These images are taken in real-life conditions without any identifications to the user excepted the she/he should take off the product from the shelf in the supermarket, held it by hand and take a picture of it directly in the shop. The information of the

image is manually assigned and saved in the database. The information is found in the label of image which is in the format:

shop.*shop name*-device.*brand.series*-location.*country name in abbreviation.city name-current image number in the same set[2]*.e.g. shop.manor-device.nokia.n95-location.ch.fribourg-set2.

The product id of the image is considered as the ground truth. In Figure 8 we show image image121.jpg. Each image is associated with a xml file which contains the information being used for CBIR method evaluation, the information contains: id, image name, image path, product information (i.e. id, name), shop name, mobile phone information (i.e. name, series), location (i.e. country, city) and relevant images. We shows a part of xml file image121.xml in Figure 9.

Fig. 8. Image121 of product: s. pellegrino acqua minerale naturale frizzante 50cl. This image is taken by Nokia N95 at supermarket manor in Fribourg.

3.2 Statistics

In FPID, the number of images per products is from 1 to 34. In Figure 10, we show the histogram of the occurrence of products with at least N images. The variety mobile phones affect the image retrieval performance, we present the challenge in the next section. We have used 9 mobile phones for the data acquisition. Figure 11 lists the mobile phones we have used with the number of images taken by them.
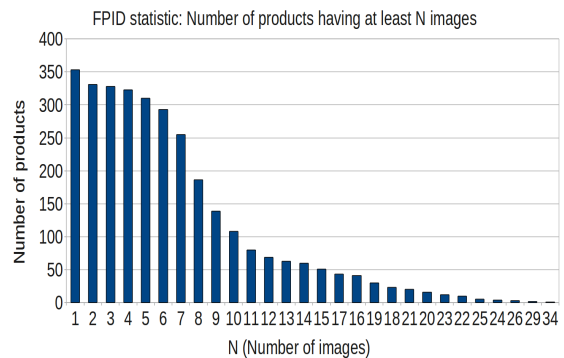
```
<image>
  <id>121</id>
  <filename>image121.jpg</filename>
  <completepath>
/home/fpid_data/data/image121.jpg
  </completepath>
  <label>
shop.manor-device.nokia.n95-location.ch.fribourg-set2
  </label>
  <product>
    <id>67</id>
    <name>
s. pellegrino acqua minerale naturale frizzante 50cl
    </name>
  </product>
  <shop>
    <name>manor</name>
  </shop>
  <phone>
    <name>nokia</name>
    <series>n95</series>
  </phone>
  <location>
    <country>ch</country>
    <city>fribourg</city>
  </location>
  <relevant>
    <image>
      <id>343</id>
      <filename>image343.jpg</filename>
      <label>
shop.coop-device.nokia.n95-location.ch.fribourg-set3
      </label>
    </image>
    ...
```

Fig. 9. Image 121.xml

Fig. 10. FPID statistic: number of products have at least N images.
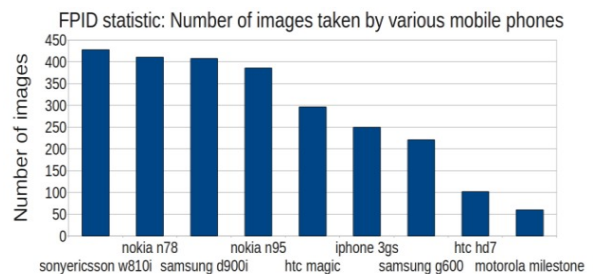

Fig. 11. FPID statistic: mobile phones.

---

2   Images being captured on the same product with same mobile in the same shop are considered as in the same set.

3.3 Closed set for CBIR methods evaluation

A closed set of FPID has been generated for CBIR evaluation. It includes two disjoint sets: query set Q_n and training set T_n, n indicates the number of images per products. To reduce potential bias due to product having a high occurrence in the database, we propose that each category has balanced number of images in the training set. To construct the training and query set, we first create a product set P where each product in P contains at least N images in the database. Then for each product we randomly choose n images to construct the training set T_n, $|T|=|P|\times n$. In the rest of images, we randomly choose n' images of each product to construct the query set Q_n', $|Q_n|'=|P|\times n'$. The ideas behind the training and query set generation are: separate the training and query images and balance the number of relevant images for each each training image. The closed set images are generated once and saved as lists for the later image retrieval experiments. Figure 12 describes the approach of closed set generation. We chose $N=12$ $|P|=100$. The images comparing the training and query sets are generated once and saved as lists in the files for the later experiments. The query set Q_2 is actually the same for all the experiments, however users are free to choose the training set T_n The closed set we provide includes: $T_{1,}T_{2,}...,T_{10}$ and Q_2. They are available at http://diuf.unifr.ch/diva/FPID/. If you use them
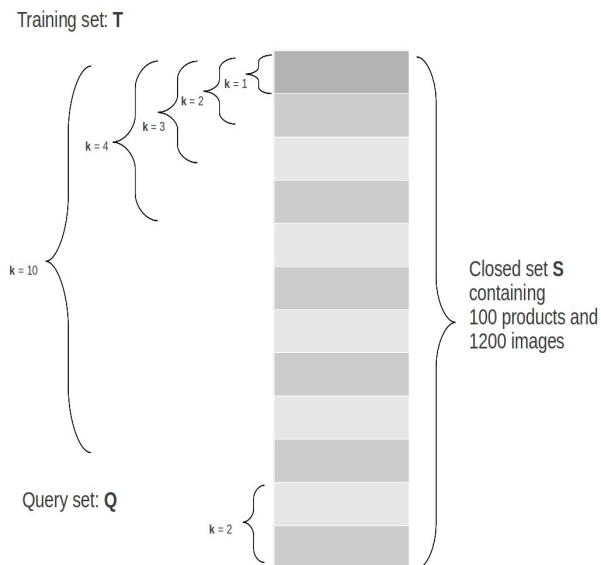

Fig. 12. Closed set generation for CBIR methods evaluation.

for CBIR evaluation please cite our paper .

4. Challenges

Mobile device and supermarket illumination influence. The variety of mobile phones and environment of supermarkets are the main influence on image retrieval. Figure 13 shows the example images in FPID of product "caffé chicco d'oro decaffeinato cuor d'oro machines espresso 250g", these images are taken in different shops with different mobile phones. We can found significant visual difference between these images.

(a) 169, Nokia n95, Manor Fribourg (b) 284, Nokia n95, Coop Fribourg (c) 286, Nokia n95, Coop Fribourg (d) 608, Samsung g600, Coop Fribourg (e) 639, Sonyericsson w810i, Coop Fribourg (f) 890, Sonyericsson w880i, Manor Fribourg (g) 1111, Sonyericsson w880i, Coop Fribourg (h) 1345, Sonyericsson w880i, Coop Fribourg (i) 1650, Sonyericsson w880i, Manor Vezia (j) 2010, Nokia n78, Manor Fribourg (k) 2429, Samsung d900i, Manor Fribourg (l) 2431, Samsung d900i, Manor Fribourg

Fig. 13. "caffé chicco d'oro decaffeinato cuor d'oro machines espresso 250g" taken by different mobile phones in different shops. The caption of each image contains "image id", "device name", "location".

To show the mobile phone influence on CBIR. We pick image 169 as query and by using different features to retrieve the first 20th most relevant images in FPID. We have chosen LIRe (Lucene Image REtrieval)[12]. Table 2 gives the experimental results.

Table 2. Given query image 169, by choosing different features, we get different relevant images. Here we return the top 20th most relevant images. We only link the id of the images which have the same product id as image 169.

Feature	Relevant images id (in descending order of retrieval scores from left to right)
All MPEG-7 Descriptors	286 2429 2010 2431
Color layout	284 286
Edge histogram	2429 2431
Color and edge directivity, CEDD	2431
RGB histogram	286

We found that images taken by different mobile phones in different shops affect the CBIR performance. For example, if we compare the color spatial distribution of the images, by calculate the L2 distance of the "Color layout" feature[5,12]. then image 284, 286 are found in the retrieved images list. It is not surprising, since in Figure 13 we can easy discover that the color distribution in images 169, 284, 286 are more similar than the others. It appears that different mobile phone cameras have different color distribution in the images. On the other hands, we have found that the edge distribution of the image is related to the shop. If we choose "Edge histogram"[6]. as feature for image retrieval, then we get relevant images 2429, 2431 and images 169, 2429 and 2431 are taken in the same shop (i.e. Manor Fribourg). This is apparently related to the background. Consequently, to be able to increase the recognition accuracy, we suggest to augment the number of images taken by the cameras of various brands of mobile phones. On the other hand, the background info (e.g. floor, hand, sleeve) has a negative influence on CBIR. Removing the background of the images is an important step before the CBIR in our case.

Illumination. The luminosity in the shops may cause shadows in different position on the images. These shadows can affect the image recognition performance. We experimented this with image21 used as the query image. It belongs to product "m budget migros eau minérale naturelle non gazéifiée 2l". Figure 14 and Table 3 give the relevant images retrieved by using the "ALL MPEG 7 Descriptors" features[4]. We found that image 22 and 23 are more similar to image 21 than the other images, because image 22, 23 were acquired in the same shop with the same mobile phone. We can see a shadow of the shelf on the left corner of the images. That's why the scores of image 22, 23 are the same, on the other hand, the other images have lower scores. Consequently, removing shadows before image retrieval is recommended.

Other challenges have been discovered our experiments. They are: package similarity, e.g. The products of migros budget, coop quarantine have the similar look, some images are given in Figure 15 , product package update. Suppliers may change the package of some products for certain reasons, e.g. promotion, luck draw. In Figure 16 , you find the bottled water of "evian" with two different packages.
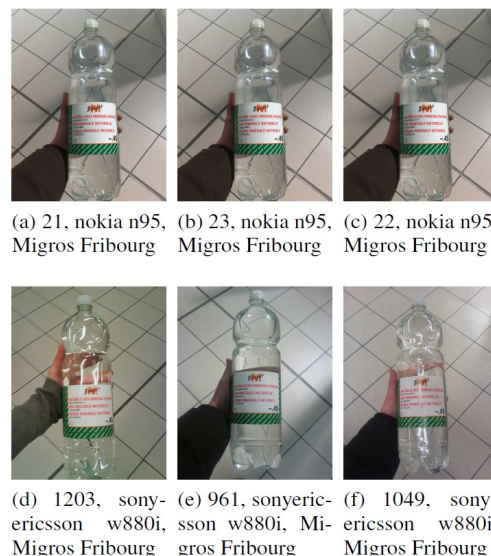

(a) 21, nokia n95, Migros Fribourg (b) 23, nokia n95, Migros Fribourg (c) 22, nokia n95, Migros Fribourg

(d) 1203, sony-ericsson w880i, Migros Fribourg (e) 961, sonyeric-sson w880i, Migros Fribourg (f) 1049, sony-ericsson w880i, Migros Fribourg

Fig. 14. Impact of similar packaging. Query image: 425, by comparing feature "All MEPG-7 Descriptors". We have Relevant image 1216 with score 0.245 at position third and irrelevant image 1239 at position first.

Table 3. Given query image 21, by using "ALL MPEG 7 Descriptors" as the image retrieve feature and return the 100 most relevant images as the result. In this table, the information: id image, phone, location, shop name, the score (between 1-0) of retrieval and the image's position in the retrieval list are displayed. The score is the normalized L2 distance between the features.

id	phone-location-shop	score	rank in the retrieved list
23	nokia n95-ch-fribourg-migros	0.81	2
22	nokia n95-ch-fribourg-migros	0.81	3
1203	sonyericsson w880i-ch-fribourg-migros	0.12	31
961	sonyericsson w810i-ch-fribourg-migros	0.05	61
1049	sonyericsson w810i-ch-fribourg-migros	0.02	79

(a) 425, m budget espresso migros en grains 1kg, Sam-sung g600, Migros Fribourg (b) 1216, m budget espresso migros en grains 1kg, Sonyeric-sson w880i, Migros Fribourg, 0.245, 3 (c) 1239, m budget caf en grains 1kg, Sonyer-icsson w880i, Migros Fribourg, 0.279, 1

Fig. 15. Impact of similar packaging. Query image: 425, by comparing feature "All MEPG-7 Descriptors". We have Relevant image 1216 with score 0.245 at position third and irrelevant image 1239 at position first.

Fig. 16. Impact of productions on the packing. Two different label of the bottled water of evian.

5. Conclusion

In this paper, we present the existing benchmark image databases for CBIR methods evaluation. We introduce a new image database FPID, the images are captured directly in the supermarkets of Switzerland without controlling the illumination, centering of the product or removing the background. Each image is provided with pre-defined ground truth. The database is made for the mobile product identification application, namely GreenT. Another objective of this database is to provide a benchmark dataset for CBIR methods performance comparison. We have also addressed some challenges of using the images from FPID for image retrieval. The FPID images are available from http://diuf.unifr.ch/diva/FPID/.

References

(1) R. Data, D. Joshi, J. Li, and Z. Wang : "Image retrieval: Ideas, influences, and trends of the new age", ACM Computing Surveys, Vol. 40, No. 2, pp. 5:1--5:60, 2008

(2) T. Deselaers : "Features for image retrieval", Master's thesis, Human Language Technology and Pattern Recognition Group, RWTH Aachen University, Aachen, Germany, 2003

(3) J. Li, and J. Z. Wang : "Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, pp. 1075--1088, 2003

(4) M. Lux : "Content based image retrieval with LIRe", MM '11 Proceedings of the 19th ACM international conference on Multimedia, pp. 735--738, 2011

(5) B. S. Manjunath, J. rainer Ohm, V. V. Vasudevan, and A. Yamada : "Color and Texture Descriptors", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, pp. 703--715, 1998

(6) D. K. Park, Y. S. Jeon, and C. S. Won : "Efficient use of local edge histogram descriptor", Proceedings of the 2000 ACM workshops on Multimedia, pp. 51--54, 2000

(7) G. Schaefer : "CVPIC Colour/Shape Histograms for Compressed Domain Image Retrieval", 26th DAGM Symposium, pp. 424--431, 2004

(8) G. Schaefer, and M. Sitch : "UCID - An Uncompressed Colour Image Database", Storage and Retrieval Methods and Applications for Multimedia, Vol. 5307, pp. 472--480, 2004

(9) H. Shao, T. Svoboda, and L. V. Gool : "ZuBuD --- Zürich buildings database for image based recognition", Computer lecture notes on image retrieval and video retrieval, LNCS 2728, pp. 71--80, 2003

(10) H. Shao, T. Svoboda, T. Tuytelaars and, L. V. Gool : "HPAT indexing for fast object/scene recognition based on local appearance", Proceedings of the 2nd internal conference on Image and video retrieval, pp. 71--80, 2003

(11) N. V. Shirahatti, and K. Barnard : "Evaluating Image Retrieval", Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01, pp. 955--961, 2005

(12) T. Sikora : "The MPEG-7 Visual Standard for Content Description – An Overview", IEEE Trans. Circuits and Systems for Video Technology, Vol. 11, No. 6, pp. 696—702, 2001

(13) A. W. M. Smeulders, S. Member, M. Worring, S. Santini, A. Gupta, and R. Jain : "Content-based image retrieval at the end of the early years", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 22, No. 12, pp. 1349--1380, 2000

(14) J. Z. Wang, J. Li, and Wiederhold : "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture LIbraries", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, pp. 947--963, 2001