Appendix C

Object Image Databases

Several object image databases have been used throughout this thesis. In this appendix we want to show a relevant set of images of each database in order to have a general idea of which images have been used. Four image databases have been selected:

1. **Pharmaceutical products**. This color image database has been used for experiments which need the use of local color information. This image database is perfect for this kind of representations because some of the pharmaceutical products contain local regions with similar color tonalities. This means that a global approach fails and we appreciate the use of a local approach. We present two databases. The first one is only composed of 100 pharmaceutical products and nearly all of them are conflictive in terms of local similarities. The second one is composed of 400 pharmaceutical products but the level of conflictiviness is "acceptable".

2. **Corel Image Database**. This color image database has also been used in our experiments. Images of this database are divided into local regions from where we extract local color histograms. In this appendix we only present some of the images used for our experiments.

3. **AR Face Image Database**. This face image database is composed of 117 individuals (64 males and 53 females). The subjects were recorded twice at a 2-week interval and during each session 13 conditions with varying facial expressions, illumination and occlusion were captured. This image database is interesting since we can implicitly evaluate natural occlusions such as individuals wearing glasses and scarfs.

4. **MNIST Image Database**. This digit image database consists of 60000 training and 10000 test handwritten digits. This digit database has been extensively used in computer vision and there are nearly 30 algorithms which have evaluated their performances with this database. In our case, we are interested in it since consists of a huge number of training/testing instances. So that, results based on this huge number of instances are statistically reliable.

C.1 Pharmaceutical Products

Pharmaceutical products have been captured using a high resolution 3CCD color camera. We used a placement protocol to assure different object orientations with respect to the optical axis and different object locations in the field of view. A diffuse lighting system was used in order to minimize highlights and other specular reflections. We captured six instances of each pharmaceutical product but we only show one instance per product. We should note that the background of each image is the one shown in images. That is, a black background. Since illumination is controlled, we can directly work with these images without the use of any normalization technique.

We created a first pharmaceutical product database which contains 100 products. They contain a high level of similarity in terms of local appearance. They can be seen in figure (C.1) where we see that white color tonalities are abundant throughout all products.

A second pharmaceutical product database is created which contains 400 products. Now, the level of appearance similarity has been reduced since products do not contain the same level of white tonalities as the previous database. Figures (C.2,C.3,C.4,C.5) show the whole 400 pharmaceutical products that were used in our experiments. These pharmaceutical products show a high level of color variability.

We should note that these two pharmaceutical product object databases have been randomly chosen from a database of 7458 different products. This large database, composed of legal pharmaceutical products in Spain, has been created to develop an integrated color-, shape- and edge-based image indexing system.

C.2 Corel Database

948 color images were selected from the original corel image database [33]. These images have been selected in order to extract local patches of images with a specific meaning for humans. We wanted to divide local patches into ten natural classes of images. These classes are: Clouds, Grass, Ice, Leaves, Rocky Mountains, Sand, Sky, Snow Mountains, Trees and Water. Figure (C.6) shows some of the original images that were selected in order to extract local image patches which correspond to the above mentioned data classes. Labeling of each local patch has been performed by a human observer. This observer took into account the whole image scene in order to label a local region. So that, the local context of a patch was relevant in making such a decision. This is the reason that some of the local patches extracted from images contain similar colors (or even the same colors) but they belong to different data classes.

Images were divided into 100 local regions (10×10) of 3456 (48×72) RGB pixels. Then, each local region is represented using a color histogram and these histograms constitute the image database used in our experiments.

This image database has been created thanks to the strong collaboration with Bernt Schiele from the Perceptual Computing and Computer Vision (PCCV) group (Institute for Computational Science, Department of Computer Science, ETH Zurich).

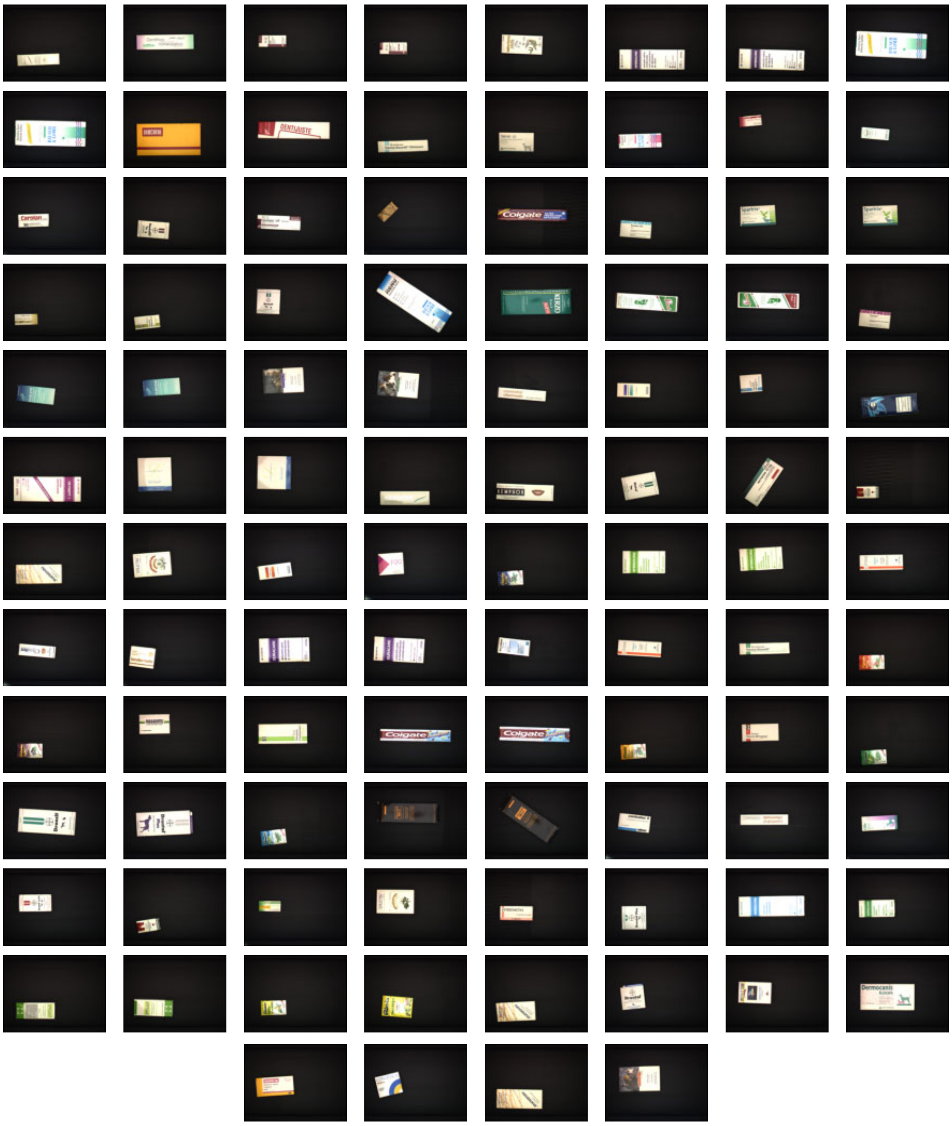

Figure C.1: A set of 100 pharmaceutical products used in the first experiments. As seen, this set of pharmaceutical products are very conflictive since they are very similar in color.

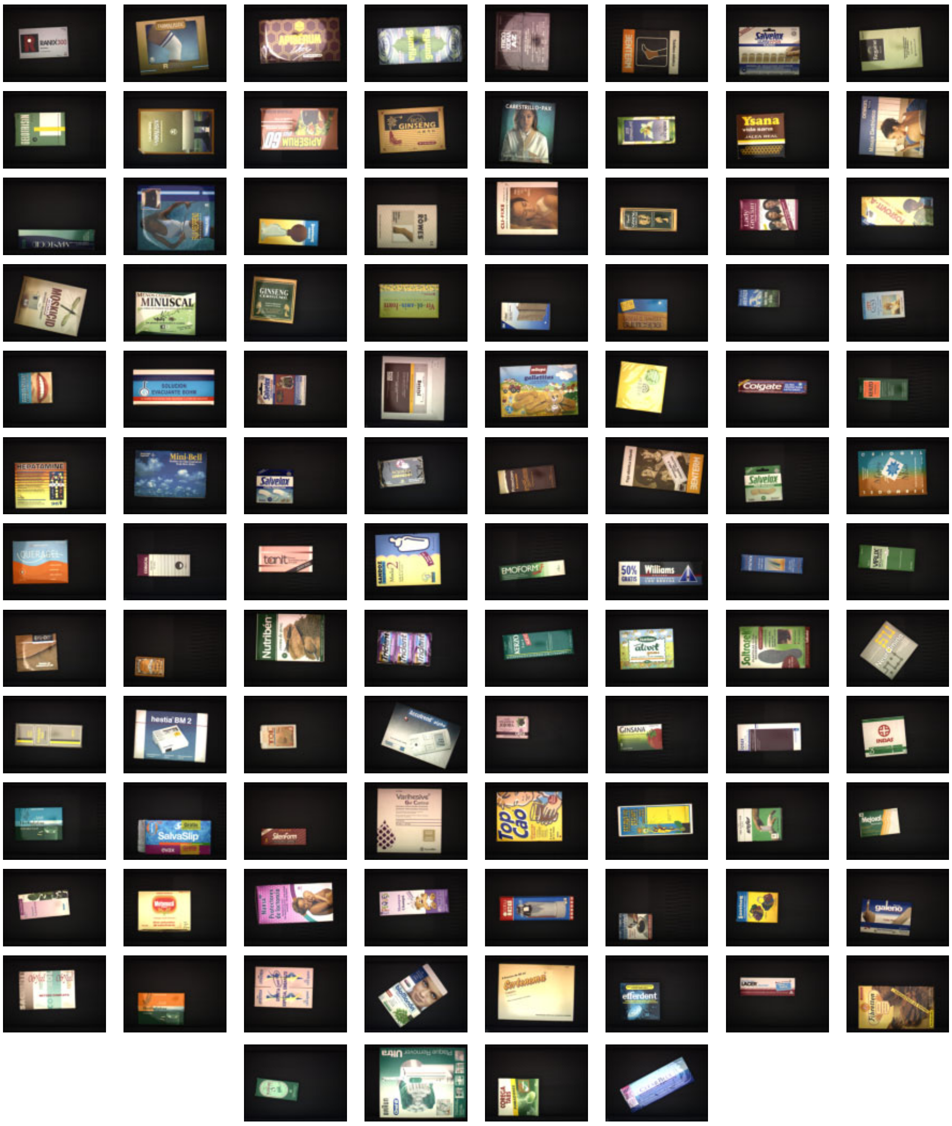

Figure C.2: A first set of 100 pharmaceutical products.

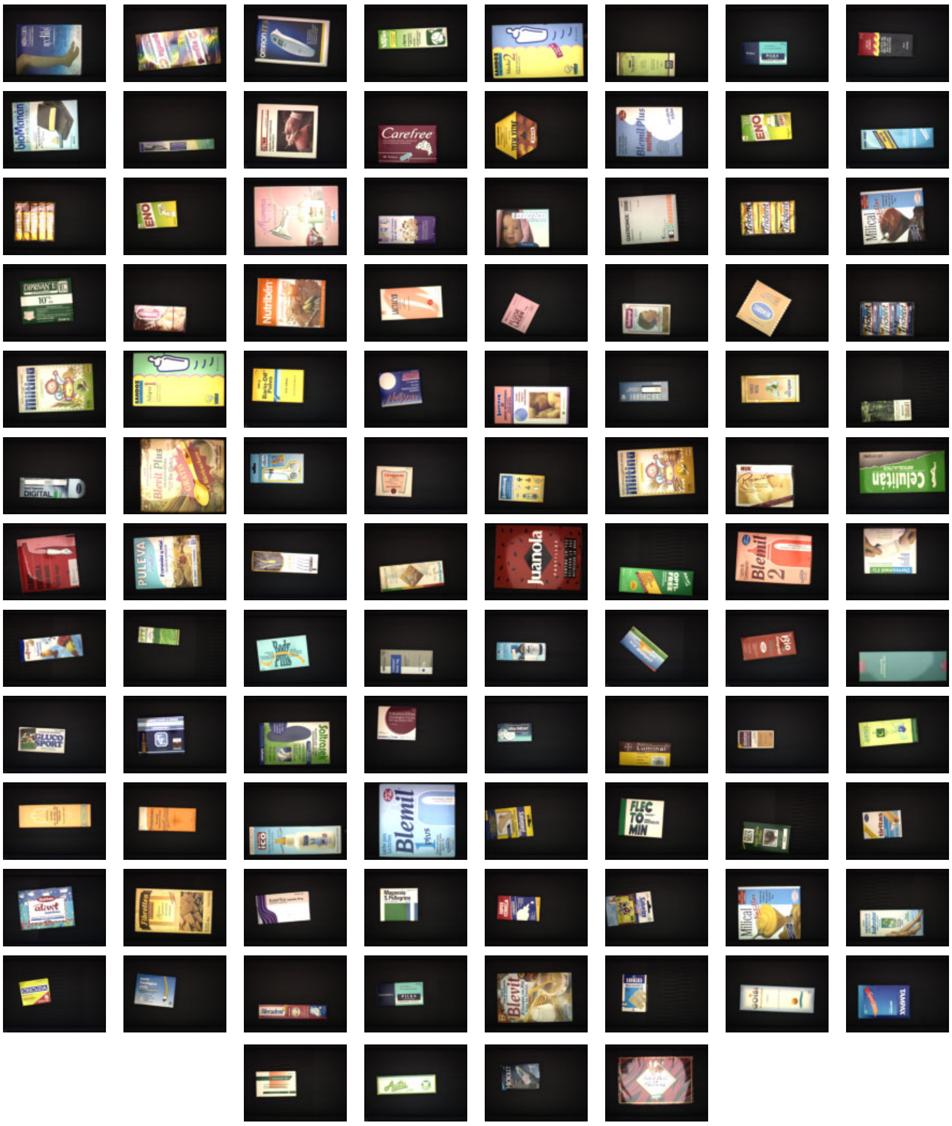

Figure C.3: A second set of 100 pharmaceutical products.

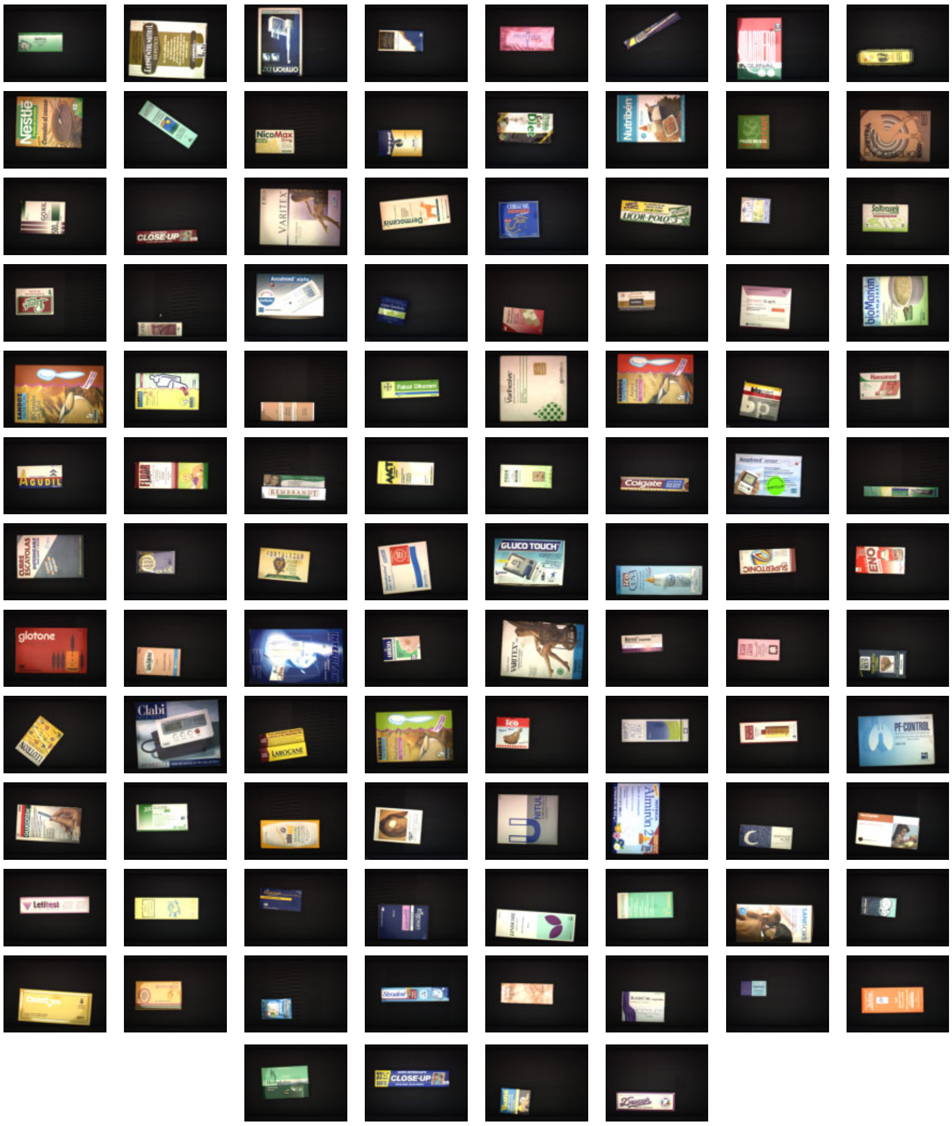

Figure C.4: A third set of 100 pharmaceutical products.

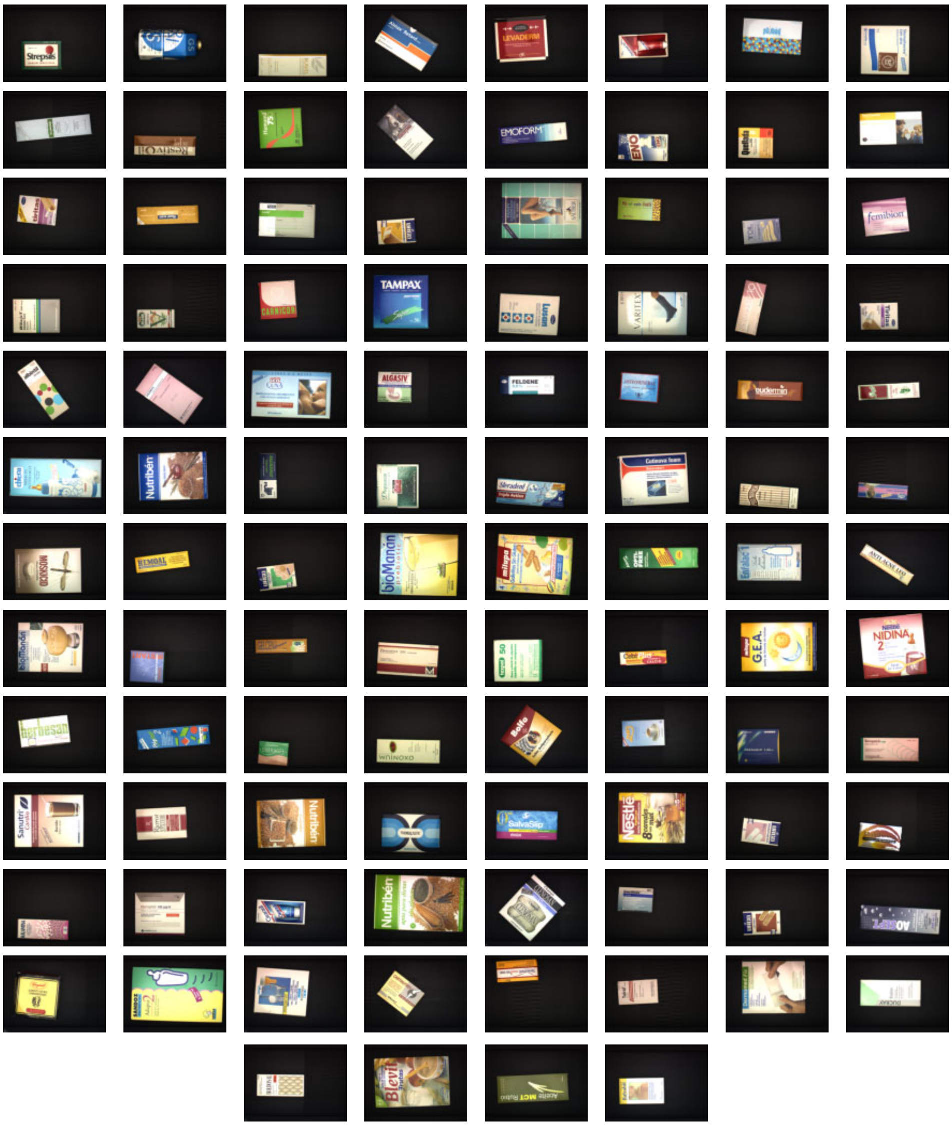

Figure C.5: A fourth set of 100 pharmaceutical products.

Figure C.6: Corel image database.

C.3 AR Face Database

This face image database is composed of 117 individuals (64 males and 53 females). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions. The original images are 768×576 pixels in size with 24-bit color resolution.

13 different facial images were taken. The conditions are: (1) neutral, (2) smile, (3) anger, (4) scream, (5) left light on, (6) right light on, (7) both lights on, (8) sun glasses, (9) sun glasses / left light, (10) sun glasses / right light, (11) scarf, (12) scarf / left light, (13) scarf / right light.

Face recognition is a two step process consisting of face detection and recognition. Usually, the face has to be located in the image and registered against an internal model. The result of this stage is a normalized representation of the face, which the recognition algorithm can be applied to. Then, in order to evaluate the face recognition accuracy it is a necessary step to provide the correct locations of the left and right eyes of a face. So that, we have applied a pose normalization to align all face images. The original AR face images have been reduced to face images of 40×48 pixels. In order to avoid external influences of background, we have defined an elliptical region that removes possible pixel artifacts. So that, the size of each face image is of 1505 pixels. We should note that the elliptical region considered has been calculated in order to reject all those pixels that do not have a relevant statistical influence in a face.

Figures (C.7,C.8,C.9,C.10,C.11,C.12,C.13,C.14) show all the individuals of the AR face database according to the 13 facial conditions exposed before.

C.4 MNIST Digit Database

The MNIST database of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image. It is a good and well-known database for people who want to try learning techniques and pattern recognition methods on real-world data while spending minimal efforts on preprocessing and formatting.

The original black and white (bilevel) images from NIST were size normalized to fit in a 20×20 pixel box while preserving their aspect ratio. The resulting images contain grey levels as a result of the anti-aliasing technique used by the normalization algorithm. the images were centered in a 28×28 image by computing the center of mass of the pixels, and translating the image so as to position this point at the center of the 28×28 field.

Figures (C.15,C.16,C.17,C.18,C.19) contain 30 instances of each digit. We see that some digit instances are really difficult to be recognized and modeled.