REAL TIME PHARMACEUTICAL PRODUCT RECOGNITION USING COLOR AND SHAPE INDEXING.

Albert Pujol, Jordi Vitria, Petia Radeva, Xavier Binefa, Robert Benavente, Ernest Valveny, Craig Von Land

Centre de Visio per Computador and Dept. Informàtica,

Universitat Autonoma de Barcelona

***Abstrac**t: In this paper, a real time application for recognition of pharmaceutical product boxes is presented. The system developed is able to recognize in less than a second a box from a set of 10000 pharmaceutical products using as cues the box size, an optimal description of a pose invariant representation of the shape and its color. To deal with the huge amount of product database as well as to achieve the real-time performance, the identification method has been designed in a hierarchical way. The inspected product box is classified in one of 50 size clusters, then, the shape and color descriptors are applied to identify the product inside a cluster in a way that the learning and recognition step can be done in an efficient manner. Additionally, principal component analysis is applied to reduce the dimensionality of the identification problem of the product boxes. The system is being currently validated and in use in the biggest pharmaceutical store serving most of the pharmacies in Cataluña, Spain.*

1 Introduction

Object indexing and recognition has been an active field of research in computer vision in the last years. Aspect-based methods have appeared as a powerful alternative to traditional 3D geometry based techniques (Lowe, 1987) (Huttenlocher *et al*. 1987) (Grimson, 1990) when geometrical models of the viewed objects can be difficult to obtain. Some problems remain difficult to be solved, like partial object occlusions and severe lighting changes, but important applications have been solved using these techniques. Turk and Pentland (Turk *et al.* 1991) used principal component analysis to describe face patterns with a lower-dimensional space than the image space. The appearance of a face is a combination of its shape, reflectance properties, pose in the scene and illumination conditions. All these factors can be compactly represented using a well-known compression technique: principal component analysis (PCA), also known as the Karhunen-Loéve transform. It uses the eigenvectors of an image set for representing each image in the set. PCA can be used for dimensionality reduction, yielding projection directions that maximize the total scatter across all images.

The problem of general object recognition and pose estimation was tackled by (Murase *et al.* 1993). They represent objects as manifolds in a low dimensional subspace formed by the *n* dominant eigenvectors of a set of training images representing all possible object views under all possible illumination conditions. Recognition is achieved by finding the manifold that is closest to the projection of an input image in the eigenspace formed by all objects. (Black *et al.* 1996) have addressed the problem of partial occlusion by using robust estimation techniques. (Buhman *et al.* 1990) use **local** appearance-based models, based on Gabor filter

responses which are used in a graph matching strategy for recognition, for minimizing object deformations. The responses of a set of local "feature" detectors have also been used in different forms by several authors (Martinez *et al.* 1997)(Schielle *et al.* 1996)(Rao *et al.* 1995).

Color distributions can be efficiently used as signatures for object recognition in the appearance-based framework. The earliest approach (Swain *et al.* 1991) showed the usefulness of color histograms for indexing large object databases independently of object's pose. Most of the recent approaches focus on illumination colour invariance (Gevers *et al.* 1997),(Funt *et al.* 1995) known as color constancy, but although these methods perform better than histogram indexing when color illumination changes, they use color information only where surface color varies and are very sensitive to noise.

Appearance-based methods have only been tested in large-scale image databases for the face recognition problem (Phillips *et al.* 998), resulting in recogntion rates around 85 per cent. Up to our knowledge, color based indexing has not been tested in very large image databases, being the largest one that referred in (Gevers *et al.* 1997).

The system presented in this paper integrates appearance based recognition and color indexing in a complete and operational prototype to recognize pharmaceutical product boxes. This problem constitutes a large scale test for these techniques, given that in our case (legal pharmaceutical products in Spain) this involves more than 10000 different products. Among these, we have identified 3500 "non-processable" products, corresponding to different causes (mirror-like box surfaces, plastic bags, etc.), which have not been considered in the test.

2 System description

The system consists of three major components: the image acquisition component, the object-localization component, and the object indexing component. This section describes each of these components and their integration.

2.1 Image Acquisition

For the capture of the image, a high resolution 3CCD color camera is employed. The camera is coupled with a standard frame grabber with a resolution of 768 *x* 494 pixels. The optics of the camera consists of a 25mm lens. This parameter allows for the acquisition of images corresponding to objects ranging from 2cm *x* 2cm to 20cm *x* 25cm. Due to the fact that products are transported by continuos moving conveyor belt, only one field of the captured image can be considered. Hence, the useful pixel resolution is 384 *x* 247. Objects are placed on the conveyor belt by a human operator that assures that the most representative face of each box is presented to the camera[1]..

In order to create a learning set of images, six images of every product were acquired, following a placement protocol to assure different object orientations with respect to the optical axis and different object locations in the field of view.

---

[1] The most representative face of the product corresponds to the one where commercial information -name, company, etc.- is printed

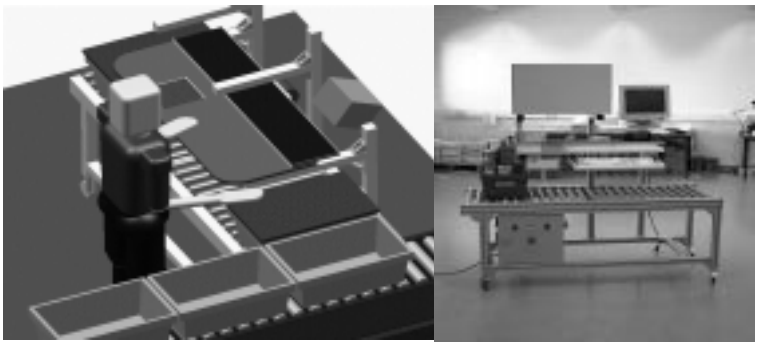

Figure 1. The ergonomic design of the system and the installed vision system in the pharmaceutical store

Lighting deserves special attention. In order to minimize lighting effects on the acquisition process, we have designed a lighting diffuser dome. In spite of the fact that this lighting architecture minimizes object aspect changes due its relative position, it does not eliminate them. This fact was decisive for deciding the acquisition of six images for each product.

2.2 Object localization and normalization

The aim of the segmentation process is to obtain the orientation, position and extent of the product box. We have handled this problem through the straight lines Hough transform. It has been chosen in order to avoid the problem that appears with boxes, that due to their colors, present a contour with holes (figure 2).

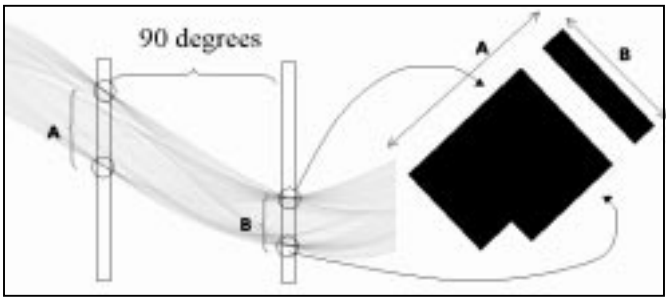

Figure 2. Representation of the Hough transform (a) of a box (b) with one hole and gap, in this image the relation between the desired measures and the maxims of the accumulation space can be easily seen.

Each contour point of the box accumulates a sinusoidal in the Hough space, where each point of the sinusoidal is a vote for one of the possible straight lines (defined in a polar representation, angle and distance to the origin) that passes through that point.

The developed system recognizes only rectangular boxes. This means that the four lines that describe the boundary of the box will be disposed in the accumulation space so that, the two highest maxims (the most voted lines), corresponding to the longer box side, will appear in the same column. The maxims corresponding to the shorter side will appear in a column shifted 90 degrees respect to the previous one.

Once all the points of the boundary of the box have been accumulated, the box orientation, size and position are obtained using the position of the maxims. Thus, the Φ position

(column) of the global maximum gives us the box main orientation and the distance between the two maxims of this column give us the size of the shortest side of the box. The distance between the most distant maxims in the column of the shortest side (column shifted 90 degrees respect to the previous one), gives the size of the longest box axis.

Once the box has been located, it is warped to a normalized box size of 32x32 pixels.

The normalization method is unable to distinguish between two boxes rotated 180 degrees. To avoid this problem a 180 degrees rotational invariant transformation of the normalized image has been applied. The transformation applied works as follow:

Given the point *(x,y)*, we will call *g(x,y)* to the position of this point when a 180 degrees rotation is applied. It is,

G(x,y)=(Tx-x,Ty-y)

Where *Tx,Ty* is the size of the image (in our case *Tx*=32 and *Ty*=32 pixels). Then the original image *I(x,y)* is transformed to its invariant representation *F[I(x,y)]*, where,

$$F[I(x,y)] = \begin{cases} |I(x,y) - I(g(x,y))| & if\ \ y < 16 \\ \dfrac{(I(x,y) + I(g(x,y))}{2} & if\ \ y \geq 16 \end{cases}$$

It can be seen easily that this transformation is invariant to 180 degrees rotations, it is,

$$F[I(x,y)] = F[I(g(x,y))]$$

Since,

$$F[I(g(x,y))] = \left\{ \begin{array}{c} |I(g(x,y)) - I(g(g(x,y)))| = |I(g(x,y)) - I(x,y)| = |I(x,y) - I(g(x,y))| \\ \dfrac{I(g(x,y)) + I(g(g(x,y)))}{2} = \dfrac{I(g(x,y)) + I(x,y)}{2} = \dfrac{I(x,y) + I(g(x,y))}{2} \end{array} \right\} = F[I(x,y)]$$

A small amount of the considered boxes are squared, in this cases, instead of 180 degrees rotations, we must take into account its rotations to 0, 90, 180 and 270 degrees. In these cases the image of the box and its 90 degrees rotation are considered as two independent descriptors of the same product. This duplication of descriptors does not represent an important decrease of the efficiency of the system due to the reduced number of products with exactly squared boxes per size cluster.

Figure 3. Pharmaceutical objects and their orientation invariant image

2.3 Object indexing

Object indexes are automatically learned from examples. As we have already commented, six images are acquired for each product, so that we have 6 or 12 invariant representations for each object. We then create a set of three indexes for each representation corresponding to three kinds of image features: color, appearance and edges. Indexes are computed using principal component analysis of the feature vectors. The feature vectors and them principal components are computed as follow:

2.3.1 Feature vectors

As we have seen before, in addition to the box size, we have used its color, appearance and edges as product descriptors. This is because although the color is a good object descriptor, it is not good enough for our problem due to the huge number of products. In some cases the color distribution is similar for two boxes and we need to know where each color is located to differentiate between them (appearance). On the other hand, using only appearance and color descriptors have the drawback of giving much more weight to the similarity between uniform regions, than the dissimilarity between small drawings of the box. This problem is handled through the use of the third descriptor. The sizes of these descriptors have been empirically chosen so that the complexity of the next steps of the process is minimized and the responses of the classifiers are preserved.

The appearance descriptor of the image is the 180 degrees rotational invariant transformation explained in the previous section. Therefore, the 32*x*32 pixels normalized image is directly used as appearance feature vector. The color descriptor is a 6*x*6*x*6 log-RGB histogram of the warped image (before 180º rotational invariant normalization), that is, once we have computed the RGB histogram, the logarithm of each bin has been computed. Finally, to compute the edge descriptor, an edge image of the warped image is constructed using the Sobel operator. Then the 180º rotational invariant transformation is applied on these contours image, the horizontal and vertical projections of this image are computed, and these projections are then used as contour descriptors.

2.3.2 Dimensionality reduction

Principal component analysis (Jollife *et al.* 1986) is a dimension reduction method which its first goal is to minimize the dimension of *n*-dimensional vectors to *m*-dimensional vectors (where $m<n$). PCA can be seen as a linear transformation that extracts a lower dimensional space that preserves the major linear correlation in the data and discards the minor ones. Vector projections can be used as representatives of original vectors for recognition purposes.

Having a data vector x ($x \in R^n$), PCA projects it onto the m dimensional linear subspace spanned by the leading eigenvectors of the data covariance matrix:

$$\sum = E\left[(x-\mu)(x-\mu)^T\right]$$

Where $\mu = E[x]$ and E denotes an expectation with respect to x. The leading m eigenvectors $\{e_i | i = 1,\cdots,m\}$ of a positive semidefinite matrix are the m eigenvectors which correspond to the m largest eigenvalues. The indices are assigned such that the corresponding eigenvalues in decreasing order are given by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$.

Let be $f : R^n \rightarrow R^m$ an encoding function from a vector $x \in R^n$ to a vector $z = f(x) \in R^m$, where $m<n$. Let be $g : R^m \rightarrow R^n$ a decoding function from z to $x' = g(f(x)) \in R^n$, a reconstruction of x. PCA encodes x as:

$$z = f(x) = V(x-\mu) = (e_1^T(x-\mu), \cdots, e_n^T(x-\mu))$$

Where V is a m x n matrix whose rows, e_i are the leading m orthonormal eigenvectors of Σ and z is the m dimensional encoding. The components of z are called the principal components. PCA reconstructs/decodes x' from z

As[2]:

$$x' = (z) = V^T z + \mu$$

The mean squared error in reconstructing the original data is:

$$\varepsilon = E\left[(x - g(f(x)))^T (x - g(f(x)))\right]$$

Using PCA we obtain the least error in terms of reconstructing the original vector. PCA builds a global linear model of the data: an m dimensional hyperplane spanned by the leading eigenvectors of the data covariance matrix.

PCA involves the computation of eigenvalues and eigenvectors of the data covariance matrix. When dealing with large high dimensional data sets, this computation can be difficult due to space requirements and complexity (Oja, 1983). Due to this fact, we have considered an alternative representation, based on splitting the data in several groups and computing an encoding function for each group.

This strategy has two advantages: the learning step can be performed more efficiently, and encoding functions are more adapted to objects, since they are computed for classes with fewer members.

In our case, the natural way of splitting data is based on size clustering. Boxes are distributed in size space in a non-uniform way, describing different clusters. In order to obtain a hard partition of this space, we have used a version of the watershed algorithm with markers (Vicent *et al.* 1991).

Markers have been selected as the 50 local maxima of the image that have the higher dynamics (Grimaud, 1992). This characteristic, defined in the framework of mathematical morphology, is a good alternative to classical hard partition methods such as k-means algorithm. Using this method we get 50 different clusters, from which we compute 50 different groups of descriptors.

---

[2] Note that due to the orthonormality of the eigenvectors, $V^T = V^{-1}$.

3 Recognition system

Each one of the 6 views of the objects in the database is stored as an array (p_i, d_i, c_i, a_i, e_i), containing the following information: p_i product identifier, d_i cluster identifier, c_i color index, a_i appearance index and e_i edges index. The cluster is determined by object size, and feature indexes are computed using the encoding functions learned from the set of images corresponding to that cluster.

The goal of the computer vision station is to assess that the products served in a request are correct. To do that, the list of descriptors of each request is sent to the computer before its arrival. When a basket with products arrives to the station, the operator introduces the identification number of the request. Thus, the computer is able to identify the descriptors of the requested products. The system verifies each one of the products that goes through the conveyor belt assuring that it has been requested.

The verification process measures the target box and chooses its size cluster. The size cluster is used to recover the eigenvectors that will be used to project the color, appearance and edge descriptors of the target image to its eigenspaces. These projections are used to recover the closest model descriptor to the target image. If the similarity measure between the recovered model and the target is greater than a threshold value, the target product is identified and the request list actualized.

4 Results and Discusion

The big amount of products (more than 6000) makes it difficult to develop a statistically significant test population. That is why, by the moment, it has been only tested under laboratory conditions in small populations of boxes. The results of these previous experiments have been good enough to develop the system, which is actually under online test working.

5 Acknowledgement

This work was supported by CICYT grants TAP98-0631, TIC98-1100, TAP97-463 and 2FD97-0220, and by contract CERF.

6 References

BLACK M., JEPSON A. 1996. *Eigentracking: Robust matching and tracking of articulated objects using a view-based representation*. Proceedings of ECCV, pp.329-342.

BUHMANN J., LADES M., VON DER MARLSBURG C.1990 *Size and distortion invariant object recognition by hierarchical grpah matching*. Proc. IEEE IJCNN, San Diego, pp. 411-416.

FUNT B., FINLAYSON G. 1995 *Color Constant Color Indexing*. IEEE Trans. PAMI, 17, 5, pp.522-528.

GEVERS T., SMEULDERS A. 1997 *Color Based Object Recognition*. In Image Analysis and Processing, Alberto del Bimbo (Ed), LNCS 1310, pp. 319-326.

GRIMSON W. 1990 *Object Recognition by Computer*, MIT Press.

HUTTENLOCHER D.P., ULLMAN S. 1987 *Recognizing Solid Objets by Alignment.* Proc. IEEE International Conference on Computer Vision, pp 102-111.

JOLLIFFE I.T. 1986 *Principal Component Analysis*. Springer Verlag, New York.

LOWE D.G. 1987 *Three-Dimensional Object Recognition From Single Two Dimensional Images.* Artificial Intelligence, 31:355-395.

MARTINEZ A., VITRIÀ J. 1997. *Dimensionality Reduction for Face Recognition,* in Advances in Visual Form Analysis, pp. 405-414, World Scientific, Singapore.

MOGHADDAM B. AND PENTLAND A. 1996 Probabilistic Visual Learning for Object Recognition, in Nayar S. and Poggio T. (eds.), *Early Visual Learning*, Oxford University Press.

MURASE H., NAYAR S.K. 1993 *Learning and recognition of 3D objects from appearance.* Proc. IEEE Qualitative Vision Workshop, New York, pp. 39-49.

OJA E. 1983 *Subspace methods of pattern recognition.* Res. Studies Press, Hertfordshire.

PHILLIPS P.J., MOON H., RIZVI S., RAUSS P. 1998 *The FERET evaluation. In Face Recognition, from thery to applications*, Springer, berlin,

RAO R., BALLARD, D. 1995 *Object Indexing using an Iconic Sparse Distributed Memory.* In Proc. of the ICCV, pp.24-31.

RADEVA P., VITRIÀ J., BINEFA X. 1999 *EigenHistograms: using low dimensional models of color distribution for real time object recognition*. CAIP'99.

SCHIELLE B., CROWLEY J.L. 1996. *Object recognition using multidimensional receptive fields histograms.* In Proc. of ECCV, pp.610-619.

SWAIN M., BALLARD D. 1991 *Color Indexing*. Intern. J. of Computer Vision, 7, 1, pp. 11-32.

TURK M.A. AND PENTLAND A. 1991.*Eigenfaces for recognition. Journal of Cognitive Neuroscience*, 3 (1), 71-86.

ULLMAN S. 1996 *High-Level Vision*, MIT Press.

VINCENT L. AND SOILLE S. 1991, *Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13,no.6,pp.583-598,.

M.GRIMAUD M. 1992 *A new measure of contrast: the dynamics*. SPIE Vol. 1769, Image Algebra and Morphological Image Processing III, pp. 292-305.