Automatic classification of bottles in crates

Kjersti Aas Line Eikvil
Norwegian Computing Center, Oslo, Norway

Dag Bremnes Andreas Norbryhn
Tomra Systems A/S, Asker, Norway

ABSTRACT

This paper presents a statistical method for classification of bottles in crates for use in automatic return bottle machines. For the automatons to reimburse the correct deposit, a reliable recognition is important. The images are acquired by a laser range scanner coregistering the distance to the object and the strength of the reflected signal. The objective is to identify the crate and the bottles from a library with a number of legal types. The bottles with significantly different size are separated using quite simple methods, while a more sophisticated recognizer is required to distinguish the more similar bottle types. Good results have been obtained when testing the method developed on bottle types which are difficult to distinguish using simple methods.

1 INTRODUCTION

Tomra Systems AS is a Norwegian company producing different types of automatons for recirculation of bottles. Their assortment includes automatons for receiving single bottles as well as crates of bottles. To be able to reimburse the correct deposit, reliable recognition of the bottles is of great importance in these systems. Recognition of bottles in crates represents special problems, as the bottles can only be imaged from above.

The objective is to identify the crate and the bottles from a library with a number of legal types, where each crate may contain bottles of several different types. The bottles with different height or very different collar size are separated using quite simple methods, while a more sophisticated recognizer is required when distinguishing between more similar bottle types. The chosen method should be sufficiently fast to be implemented on an automatic return bottle machine.

This paper describes a method which is developed as a supplement to the simple methods with the aim of dealing with the difficult cases. A prototype is constructed for each bottle type. During recognition, a bottle of unknown type is matched against each prototype and assigned to the class giving the best match. The classification rule has a proper statistical basis; assuming that the bottle types are of equal prior probanbility it minimizes the expected error rate.

2 DATA

The images of the crates are acquired by a laser range scanner. The range scanner returns two different signals; the distance to the object and the strength of the reflected signal used to measure this distance. The result is an

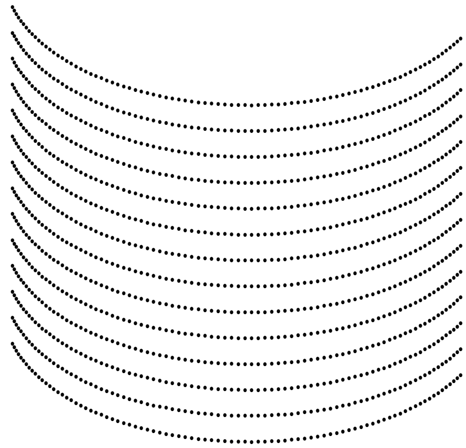

Figure 1: *The scan pattern.*

image with two components.

The scanner performs crescent shaped scans (figure 1) with a vertical laser beam which is moved across the crate forming a 2-dimensional image. Geometric correction is performed during the image acquisition to correct for the scanning pattern and reduce the amount of data. The principles for this correction are based on the fact that the most reliable range signals correspond to the strongest reflected intensity signals. When mapping pixels from the input image to a pixel in the output image, the pixel in the input image having the greatest strength is chosen.

The first step of the method is to specify a mathematical model defining the relation between the line and sample coordinates (u, v) in the output image and the line and sample coordinates (x, y) in the input image. Because direct knowledge of the geometric distortion produced by the scanner is available, the mathematical model may be derived analytically:

$$u = r_x * (1.0 + \cos(\theta)) \tag{1}$$

$$v = r_y * (1.0 - \sin(\theta)) + (\frac{\theta}{\pi} + y) * d \tag{2}$$

Here θ is the angle from the semi-major axis to sample x of an abritrary scanline, d is the distance between two scans, r_x is the x-radius of the scan ellipse and r_y is the y-radius of the scan ellipse.

Both the bands of the input image are identically transformed, and the resulting bands are 100×100 pixels with a resolution of 5mm. These images were the basis for the further processing.

Figure 2: *The upper parts of the different bottles; coke (left), pepsi (middle) and hero (right).*

Figure 3: *Bottle types used in this study*

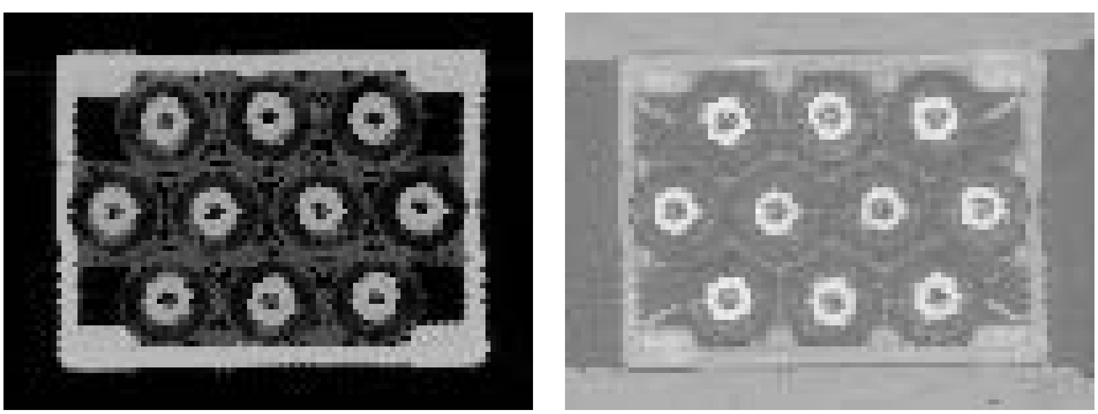

Figure 4: *Two-component image of a crate containing coke-bottles. Range component (left) and strength component (right).*

The methods described in this paper were tested on bottles of three different types. All the bottles were made of plastic, their true heights were between 33 and 34 cm and their caps were missing. The three classes were *coke*, *pepsi* and *hero*. The *hero*-bottles are made of a thinner plastic than the others. Figure 3 shows samples of each type, while the two-component image of a crate containing *coke*-bottles is shown in figure 4. The range values corresponding to the slanting part of a transparent bottle shoulder are known to be unreliable. Hence, only the upper part of each bottle is used in the classification. Figure 2 shows a photograph of the tops of three different types of bottles, taken from the side.

3 METHODS

The recognizer consists of two main parts where the bottles are first located and then classified. Accurate information about the location of the bottles is important to obtain a successful classification. Prior to the recognition, the images must also be calibrated. This calibration is necessary to make the images scanner independent. The following sections treat the problems of calibration, bottle localization, classification and training of the classifier.

3.1 Calibration of images

It is desirable to be able to use the same set of prototypes for all images, independent of the scanner producing the images. Then, for the classification methods to work, both the range and strength values of the images have to be calibrated. If the images were not calibrated, we would need a prototype for each scanner, and the images to be classified would have to originate from the same scanner as the images from which the prototypes were computed.

3.1.1 Calibration for range

We denote the original and the calibrated range values as r_0 and r, respectively. The range values can then be calibrated according to the following linear formula:

$$r(x, y) = a * r_0(x, y) + b \tag{3}$$

The parameters a and b are adjusted to correctly calibrate heights corresponding to the part of the bottle which is used in the classification.

In addition to the linear calibration, a correction of the range values due to the elliptic orbit of the scanner is necessary. The heights measured at the middle of a scanline are almost correct, while the heights measured near the ends of a scanline will be larger than they should be. A scanline from a flat area is used to compute a correction table, where each element of the correction table specifies the difference between the measured height and the true height. These values should be subtracted from the measured values before the linear calibration is performed, i.e:

$$r(x, y) = a * (r_0(x, y) - \delta(x)) + b \tag{4}$$

Here, $\delta(x)$ is the correction value for pixel number x at a scanline. $\delta(x)$ is found from the correction table containing correction factors for the current scanner.

3.1.2 Calibration for strength

The strength calibration is based on two calibration marks in the machine. Let s_{low} and s_{high} be the strength of the lowest and highest of these, respectively. Then a strength threshold, t_s is given by:

$$t_s = \frac{s_{low} + s_{high}}{2} - C_s \tag{5}$$

where C_s is a scanner-dependent constant. The strength threshold is subtracted from all strength values. Moreover, to avoid negative strength values, a constant is added to all the thresholded values:

$$s(x, y) = s_0(x, y) - t_s + 127 \tag{6}$$

3.2 Determination of bottle centres

Prior to the classification of the bottles, the crate is identified. Since the bottles in a crate are usually placed in a rectangular grid, we have some idea of where to look for them. Moreover, information of the types of bottles normally appearing in this particular type of crate can be utilized. If a bottle of an unexpected type appears, this would probably result in a wrong location as well as a misclassification. However, this is not crucial, since an unexpected bottle type does not usually give a reimbursement of deposit.

Our method for finding the bottle centers attempts to segment the collars of the bottles, and determine their centroids. For this, we need an approach to segmentation which is suited for real time applications. To reduce the amount of data and simplify the segmentation, only the range data are used for the estimation of these centres.

There are several segmentation algorithms available. The region growing approaches[1] attempts to merge pixels having similar characteristics. The edge detection and boundary following approaches searches for discontinuities with respect to some characteristics[2] and try to assemble the pixels found into meaningful boundaries. As bottles are circular, the Hough transform[3] is one method for linking edge pixels method which may work well. This approach is, however, computationally expensive. We found thresholding,[1] an old technique for segmenting digital images, to be the most suitable approach. The simplest of all thresholding methods is to segment the image by determining a threshold T and classify all pixels having a range value less than T as background and all pixels having a range value greater than or equal to T as bottle collars. However, it is difficult to find a global threshold which will work for all types of bottles, and another problem is that the collars may be confused with other regions of the image.

The first problem was solved by defining a search area for each bottle and compute the threshold separately for each search area. In this way we could treat one bottle at the time, and the threshold would be adapted to the current bottle type. As we were dealing with circular bottles, we chose a circular search area centred at the preliminary bottle position, c_0. To ensure that the collar of the bottle was completely contained in the search area, the radius of the circular disk was set to $r + \Delta$, where r is the maximum radii of any bottle collar, and Δ is the maximum error of the initial center position. The search area, Φ, can then be expressed as:

$$\Phi = \{\mathrm{c} : |\mathrm{c} - \mathrm{c0}| \leq \Delta + r\} \tag{7}$$

We adapted the threshold to the current area. First restrictions were put on the ratio, R, of the size of the collar region to the background, assuming that R should be almost equal for every search area. The meaning of this is that the pixels belonging to the bottle always should represent about the same portion of all the pixels in the search area. To achieve this, the lowest 80 % of the pixels in the search area were pre-classified as background, and the highest 10 % of the pixels were pre-classified as foreground. The percentages were determined by trial and error. To allow for some variance, the remaining 10 % of the pixels were thresholded, where the threshold was computed by Otsu's method.[4] This method finds the threshold T which maximizes the between-class variance. To remove noise and fill in bays and gaps, the search area was filtered with a max operator after thresholding.

The problem of confusing collar regions and other regions of the image was solved by inspecting the regions segmented as possible bottle collars. To be classified as collars the regions had to satisfy conditions restricting the size, the ratio of width to height, and the third order moments. Hence, regions that satisfied the conditions, were assumed to correspond to the collar of a bottle, and the centroids of the regions were computed and chosen as the new bottle positions. That is, the x- and y-coordinates of the centre were computed as:

$$x_c = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{and} \qquad y_c = \frac{1}{N}\sum_{i=1}^{N} y_i$$

where $(x_i, y_i), i = 1, ..., N$ are the pixels of a collar region.

3.3 Classification

The fact that the bottles are viewed only from above, places considerable constraints on the types of features that may be extracted. Magee et al.[5] present a method for classification of bottles viewed from above. However, in this study the classification was mainly based on the characteristics of the bottle caps which reflect strong signals. A different approach was necessary in our case where the caps are usually removed from the bottles. The reflected signals around the top of the bottle are then weaker. Moreover, information from the logos that are usually printed on the caps, can not be utilized.

A prototype was constructed for each bottle type. The prototype is a statistical description of the image area corresponding to the bottle. The prototype consists of a fixed number of measurements corresponding to the number of pixels of which the area is composed. In what follows a Bayesian classification rule[6] is to be derived.

Each measurement, Y_n is a two dimensional vector and we assume it to be bivariate Gaussian distributed with density:

$$f(y_n; \mu_n, \Sigma_n) = (2\pi)^{-1}|\Sigma_n|^{-1/2}\exp\{-\frac{1}{2}(y_n - \mu_n)'\Sigma_n^{-1}(y_n - \mu_n)\}$$

μ_n and Σ_n are unknown, but maximum likelihood estimates can be obtained from a training set.

The locations of the measurements which form the basis of the classification are given relative to the centre of the bottle. The measurements may be configurated in many ways and the issue of selecting the prototype that is most appropriate in this case is treated in section 3.3.1.

If we assume that the prototype consists of N independent measurements, the joint density of these is given by:

$$f(y_1, ..., y_n) = (2\pi)^{-N}\prod_{n=1}^{N}|\Sigma_n|^{-1/2}\exp\{-\frac{1}{2}\sum_{n=1}^{N}(y_n - \mu_n)'(\Sigma_n)^{-1}(y_n - \mu_n)\} \tag{8}$$

An unknown bottle is assigned to the bottle type which maximizes the joint probability in (8). Due to the exponential form of the Gaussian density, working with the natural logarithm of (8) is more convenient. Hence, the following classification rule is the result:

Assign the unknown bottle y to the bottle type i if

$$\sum_{n=1}^{N}\left[\ln|\Sigma_n^i| + (y_n-\mu_n^i)'(\Sigma_n^i)^{-1}(y_n-\mu_n^i)\right] < \sum_{n=1}^{N}\left[\ln|\Sigma_n^j| + (y_n-\mu_n^j)'(\Sigma_n^j)^{-1}(y_n-\mu_n^j)\right] \tag{9}$$

for all j different from i.

This rule has a proper statistical basis; assuming that the bottle types are of equal prior probability, it minimizes the expected error rate. (The term $(2\pi)^{-N}$ and the factors of $1/2$ were eliminated as they do not affect the final decision process.)

3.3.1 Choice of appropriate prototype

An ideal prototype would cover the areas of a bottle where the different types of bottles differ. Without any knowledge of the location of these areas, the best prototype will be the one covering the whole bottle. However, that means that the number of points to be matched for every bottle will be huge, which in turn is not very desirable from a computational point of view. Hence, it is necessary to find a configuration which consists of a smaller number of points but still gives a good discrimination between bottle types.

To determine an appropriate prototype, we investigated the differences between the bottle types. This was accomplished in the following way. For each pair of bottle types, we computed the generalized Mahalanobis distance[7] at each point of a prototype covering the whole of the bottle. The generalized Mahalanobis distance between two multinormal distributions with mean vectors and covariance matrices equal to $\mu_1, \mu_2, \Sigma_1, \Sigma_2$, respectively, is defined as

$$\omega = (\delta^2 + \gamma^2)^{1/2} \tag{10}$$

where

$$\delta^2 = (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2) \qquad \gamma^2 = 4\log\frac{|\Sigma|}{|\Sigma_1|^{1/2}|\Sigma_2|^{1/2}} \qquad \Sigma = \frac{1}{2}(\Sigma_1+\Sigma_2) \tag{11}$$

If the Mahalanobis distance is computed for each point, the distribution of the Mahalanobis distances over the bottle shape may be visualized as an image. The results of these computations are shown in figure 5. The Mahalanobis distance between each point is represented as gray values. Large distances correspond to dark pixels and small distances to light pixels.

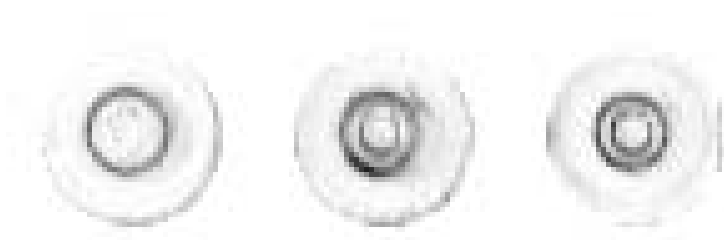

Figure 5: *Mahalanobis distances between prototypes visualized as images; coke and pepsi (left), coke and hero (middle) and pepsi and hero (right)*

From figure 5 it can be seen that *coke* has a wider collar than *pepsi* which in turn has a wider collar than *hero*. Moreover, the diameter of the opening of *hero* is somewhat smaller than those of *coke* and *pepsi*. The contribution

	Results scanner 3			Results scanner 2			Results scanner 1		
	coke	pepsi	thin	coke	pepsi	thin	coke	pepsi	thin
coke	**87.0**	13.0	0.0	**93.0**	7.0	0.0	**90.0**	9.0	1.0
pepsi	3.0	**97.0**	0.0	4.0	**95.0**	1.0	1.0	**97.0**	2.0
thin	1.1	0.0	**98.9**	0.0	0.0	**100.0**	0.0	0.0	100.0

Table 1: *Confusion matrices with the individual classification rates for each scanner.*

to the discrimination from the shoulder of the bottles is minor. It appears that a combination of two doughnuts would form a suitable prototype, one to detect the difference in opening diameter and the other to detect the difference in collar width. However, the positioning of such doughnuts have to be very accurate not to miss the difference between the bottle types. This in turn is not very easily obtained due to the coarse image resolution. Hence, we were forced to choose a prototype that covered a larger area, but to reduce the size, it was shaped in a star-like fashion.

3.4 Parameter estimation

The classification rule described in section 3.3 requires estimation of the mean vector and covariance matrix for each point of each prototype. During a learning procedure, the well known maximum likelihood estimates[6] for these parameters were computed. In what follows we omit to express the dependency on the bottle type.

Let the training set consist of measurements from M bottles. The maximum likelihood estimates are then given by:

$$\mu_n = \frac{1}{M} \sum_{i=1}^{M} y_{ni} \qquad n = 1, ..., N \tag{12}$$

$$\Sigma_n = \frac{1}{M} \sum_{i=1}^{M} (y_{ni} - \mu_n)(y_{ni} - \mu_n)^T \qquad n = 1, ..., N \tag{13}$$

where $y_{n1}, ..., y_{nM}$ are the training samples used to estimate point n of the prototype.

4 EXPERIMENTAL RESULTS

The set of images were obtained from three different laser range scanners. As a total we had 866 different images of bottles. The training was performed on images from two of the scanners, while classifying images from the third scanner in a leave-one-out approach. The individual results for the three scanners are shown in the confusion matrices in table 1. As a total, for all the three scanners, 95.2% of the bottles were classified to the correct bottle type.

Of the three classes above, the *coke* and *pepsi* bottles usually give a reimbursement of deposit, while there is no reimbursement for the *hero* bottles. Classifying the bottles into the two classes "reimbursement of deposit" and "no reimbursement", a correct rate of 99.4% was obtained.

5 DISCUSSION AND CONCLUSION

In this paper we have presented a method for locating and identifying bottles in crates based on images from a laser range scanner. The bottles were of equal size, but their shapes were slightly different. An unknown bottle was isolated by a thresholding procedure and its center was determined as the centroid of the top region of the bottle. A Bayesian classifier was then used to determine the type of the bottle. The method shows good performance, producing a recognition rate of 95.2% for our test set. However, the datasets have been limited. It is therefore necessary to perform tests on a larger set of data, and this will be the natural next step in the process.

Currently, we have looked at images from three different scanners. However, the recognition software must be prepared to handle images from a wide range of scanners with slightly different properties. To study the effect of these differences, images representing the entire spectre of scanners should be used for testing. Also, in our study only a few difficult bottle types were selected for the test. The current method is able to distinguish quite well between these types. Still, we do not know enough about the performance on bottles not represented in the set. To get a thorough evaluation of the methods, they should therefore be tested on a larger set of bottle types.

Since the datasets used have been limited, we have avoided some problems that may occur in practice. We have for instance not yet looked into the problem of recognizing bottles with the capsule still on. Neither have we studied thoroughly the effect of deformed or missing stickers, and it may also be that the colour and shape of the crate may influence the images of the bottles. All of this should be tested systematically.

For the algorithms to be interesting in a practical environment, speed is of great importance. The throughput rate of the system as it is today is satisfactory. However a further increase in the effectivity of the method is desirable as this means that the number of candidate bottle types can be increased.

One way of achieving an increase in the effectivity of the system is to reduce the amount of data. We therefore investigated the contribution from the strength band by examining the results of the system prior to adding this information. Using range data only, the correct classification rate was 88.5%. Hence, the strength band proved to be critical to avoid misclassifications.

The classification rule (9) assumes that corresponding range and strength pixels are dependent. This assumption is not necessarily correct. Moreover, it increases the computational complexity of the system. Experiments showed that the results were only slightly poorer when independency was assumed.

From a computational point of view, the number of points, N, to be matched for every bottle should be as small as possible. Presently we are working on different ways of reducing N and the preliminary results are promising.

6 ACKNOWLEDGEMENTS

Support for this research was provided by the Norwegian Research Council (NFR). The authors would also like to thank Rangnar Bang Huseby at NR for valuable input.

7 REFERENCES

[1] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Addison-Wesley, New York, 1992

[2] D. Marr and E. Hildreth, “Theory of Edge Detection” Proc. P. Soc. Lond. B, Vol 207, pp 187-217, 1980

[3] P. V. C. Hough, “Method and means for recognising complex patterns” US Patent 3069654, 1962

[4] N. Otsu, “A Threshold Selection Method from Gray Level Histograms”, IEEE Transactions on Systems, Man and Cybernetics, Vol.SMC-9, No.1, 1979

[5] M. Magee, R. Weniger and D. Wenzel, “Multidimensional Pattern Classification of Bottles using Diffuse and Specular Illumination”, Pattern Recognition, Vol 26, No 11, pp 1639-1654, 1993

[6] R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, John Wiley & Sons, New York, 1973

[7] N. L. Hjort “Notes on the Theory of Statistical Symbol Recognition”, Norwegian Computing Center, Report no.778, 1986