Using Local Features to Classify Objects having Printable Codes.

Filipe Ferreira*,+, Luis Davim+, Rui Rocha+, Vitor Santos* and Jorge Dias+

*University of Aveiro, Portugal

+Instituto de Sistemas e Robótica, University of Coimbra, Portugal

Abstract

The concept of tracking the component origin and manufacturing specification encoded in the form of Data-Matrices has been gaining ground. This has translated into a search for cost-effective alternatives to machine readable, printable codes such as barcodes. The advantage of Data-Matrices assumes special relevance in the case of miniature components and in the case of complex, multi-part assemblies in which a large amount of information must be encoded within a small foot-print. We propose the application of state of the art techniques in multi-scale, local image feature extraction (Corners and lines) to read the binary data in Data-Matrices. The columns of binary data are arranged in the form of a sequence of binary features each of which codes some information. Our aim is to increase the applicability of the codes by developing applications in object classification without supervision, achieving tolerance to incomplete information about the objects to be classified. The classification process is achieved by using techniques that are commonly utilized in bio-informatics to align protein sequence which recently have been applied to Place recognition problems in Mobile Robot Navigation.

1 Introduction

The need to tracking the origin of manufactured components and their specifications has been gaining ground. Whereas, barcodes were once utilised to code a small amount of information regarding the manufacturer, part and manufacturing batch numbers, the need for ever greater information regarding the history and origin of the components has resulted in their substitution by Data-Matrices. The advantage of Data-Matrices assumes special relevance in the case of miniature, albeit complex, components in which a large amount of information must be encoded within a small foot-print Fig. 1a, in middle. It has also been found to be very useful in the case of multi-part assemblies that use concatenated Data-Matrices to express very large amounts of information.

The use of two dimensions to store the information coupled with more robust algorithms for reading and error detection/correction means that a large amount of information can be stored in the data-matrix representation. Typically in a square symbol, between 10×10 to 144×144 bits can be stored. Using multi-scale techniques the principal 'sync' and 'handle'

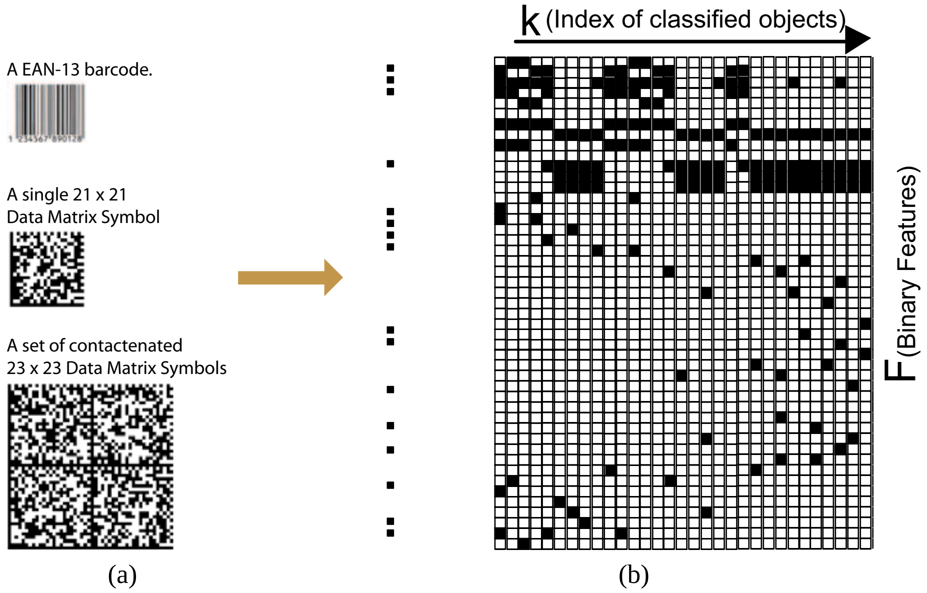

Figure 1: Rearranging the barcode/data-matrix symbols to create a Binary vector descriptor. The descriptors for a pilot set of classified object is represented as a Feature Incidence Matrix, shown at right

properties of a data-matrix are extracted in order to recover the orientation and to achieve the ability to correctly read the matrix. The black and white squares are then extracted, using fast corner extractors on the original and on the inverted images, in the form of sequences of binary features each of which codes some information together with the redundancy implicit in the coding process. Given these limiting features on a single data-matrix symbol, multiple symbols can be concatenated along either dimension to increase the amount of information that is stored, Fig. 1a, at bottom.

data-matrix have been used to mark small items such as integrated circuits and printed circuit boards. In such applications the code can typically hold upto fifty characters within a 3×3 square and can be read using a video camera in situations of atleast 20 percent contrast ratio.

While the main use of the data from barcodes and Data-Matrices is to correctly and reliably access databases containing information regarding the historical information of a component, there are some applications that could use the information that is stored in the pattern on the label itself. Our aim is to increase the applicability of the codes by developing applications in object classification without supervision, achieving tolerance to incomplete information about the objects to be classified. Most airline tickets are printed by a type of barcode label printer, and another similar machine is also used to make luggage tags that help the airline keep track of luggage.

While service industries, retailers and storage facilities already utilise barcodes and Data-Matrices to adequately label spare parts and keep track of stocks and storage locations, the increased requirement for recycling products and parts has opened up a need to incorporate and handle much more information.

Corresponding Author: Filipe Ferreira, email:cfferreira@isr.uc.pt

Proof of concept projects and research is still ongoing into the informational databases and product labelling that would lower the costs of recycling just as in other industries they have increased efficiencies [Saar & Thomas, 2002]. Getting as much information as possible is pertinent in the case of products that contain as toxic substances [Saar & Thomas, 2004], or might have come in contact with such substances over the course of their useful lives. Since the specification of the European Waste Electrical and Electronic Equipment (WEEE) directive, Industry associations such as SPECTARIS, the German Industrial Association for Optical, Medical and Mechatronic Technologies promote the use of technologies that integrate manufacturing information with usage history in order to make recycling easier. The information that can be coded in a data-matrix or an RFID includes quantity and type of the old device, quantity of the re-used old device, quantity of the raw materials used, and notification of the components and working materials used. Similarly, the US Department of Defence has mandated that components used in DOD project must include Data-Matrices in order to enhance the traceability of these components [Agapakis & Stuebler, 2006].

The classification technique that is proposed has been previously utilized in measuring the similarity between documents, based on the word-similarity. It has also been recently applied to Place recognition problems in Mobile Robot Navigation[Ferreira et al., 2006].

2 Local Features from images

Local image image feature detectors are ubiquitous in applications of vision to robotics. Among the various local features, corners and edges are among the most common features used.

In Schmid and Mohr[Schmid & Mohr, 1995], it is mentioned that the only image invariants are the curvature of the isophote line and the flow line. This same article goes on to suggest that this property is of little practical value given that the associated calculations are difficult, and that the noise in real images and their limited resolution negates the remaining usefulness of the result. So a lot of research goes into the extraction of 'partially invariant' features in images. Schmid constructs a 'Local-jet' and extracts points of interest according to a Heitger and Rosenthaler detector. In Schmid and Mohr [Schmid & Mohr, 1997], the same authors utilize a Harris detector, a corner detector, to choose interest points in an image.

In Mikolajczyk and Schmid [Mikolajczyk & Schmid, 2001], a comparison is made of various methods of extracting points of interest and the efficacy of different region descriptors. Invariance to scale variations is achieved differently.

Image pyramids have been utilized for quite some time in order to take advantage of memory savings and faster processing as information redundancy in an image increases. Lowe [Lowe, 1999], utilizes a sampled Gaussian kernel to smooth an image repeatedly. Each smoothing by convoluting with a Gaussian filter is followed by the reduction of the resolution of the image. The Difference of Gaussian(DoG) image as an approximation of the laplacian, is used to identify the interesting points.

Corresponding Author: Filipe Ferreira, email:cfferreira@isr.uc.pt

3 Classification using Bernoulli Mixture Model

While classification of a few binary properties might be easy, using many, correlated features is very difficult. In applications such as Mobile-robot-localisation, Image retrieval and robot localisation methods typically make use of a large number of features. In the application to Mobile-robot-localization, we have employed up to 16000 binary features to allow the robot to recover its position within the environment. Having to reduce the dimension of data from a sensor or multiple sensors in order to make the procedures more tractable or robust is therefore a common problem. A common solution to the problem is to perform data fusion according to some model to create composite features from different features(from the same or from different sensors) after registering the data of one sensor with another. This approach seeks to impose correlations that are suspected to exist in the data as a result of a particular environment.

On the other hand a data-driven approach will attempt to extract these correlations. Methods that reduce the dimension of features with continuous values are common in many perception fields including face recognition, speech recognition etc. Among these approaches, Mixture Models are a common solution to modelling data that is thought to a follow non-parametric distribution. Sajama and Orlitsky in [Sajama & Orlitsky, 2005] demonstrate the use Mixture models composed of Gaussian, Bernoulli and exponential distributions as a solution to the classification problem. To a greater or lesser extent these clustering or classification methods seek to identify features that are more correlated with members of their own group than with members from another group. Mclachlan and Peel [McLachlan & Peel, 2000] provide a good reference to the general topic of Finite Mixture Models.

Articles such as [Kaban & Girolami, 2000] and [Wang & Kaban, 2005] provide a healthily different viewpoint and go some way to demonstrated the usefulness of binary features. In [Wang & Kaban, 2005] the context in which a word is used in a sentence is converted into multiple binary features. Similarly [Juan & Vidal, 2004] and [García-Hernández et al., 2004] seek to model some training data as a sample of sets of binary features taken from a population of binary features, each distributed according to a mixture of Bernoulli distributions. The application of binary features to the classification of images and text has motivated us to apply the approach to classifying other types of features, such as barcode/data-matrix features.

Mixture models assume that there exists a finite number of distributions which, when mixed together in a particular proportion, result in a distribution that best describes the data.

3.1 Specifying the Classification Task

Supposing we classify a pilot set of objects into two or more subsets according to certain criteria that are reflected in the binary digit code that could come from a barcode or from a data-matrix.

The binary features from each of the objects in the pilot set are represented within a Feature Incidence Matrix (FIM), $\mathcal{V}$. Each row i, of the FIM corresponds to a feature Y_i and each column j, to an object, V_j, from the reference sequence (each entry in the FIM might be represented as $Y_{i,j}$ where the first subscript indicates the feature and the second subscript, the object). $Y_{i,j}$ takes value 1 if feature Y_i appears (is present in the code) in object V_j, 0

Corresponding Author: Filipe Ferreira, email:cfferreira@isr.uc.pt

otherwise, see Fig. 1.

$$\mathcal{V} = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \dots & Y_{1,K} \\ Y_{2,1} & Y_{2,2} & \dots & Y_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{N,1} & Y_{N,2} & \dots & Y_{N,K} \end{bmatrix} \tag{1}$$

A simple metric for verifying the similarity between a new object to be classified, and another in the pilot set, could be the number of binary features in each object that are unchanged. This metric would assume that the individual features on each object are independent. Unfortunately, given that features arise in groups and persist/disappear as a result of the behaviour of the object and its exposure to the environment, the assumption of independence between the features does not reflect reality.

Inferences made under this assumption would be biased toward certain objects in the FIM and in practice, some of the features are highly correlated while others are less. In such circumstances we need to employ methods that deal with the correlation between the features. As stated earlier, this thesis describes our use of Mixtures of Bernoulli Distributions to model the binary FIM.

3.2 Formulation of the Bernoulli Mixture-Model

Mixture models assume that there exist a finite number of parametric distributions which, when mixed together in a particular proportion, result in a distribution that best describes the data we wish to characterize. In this case we model any code that is observed V_{obs} as a vector of binary features $\{0, 1\}^N$ which is obtained from a particular mixture of Bernoulli distributions, as in (2).

We present a brief legend of the terms we will utilize below. $\mathcal{V}$ represents the complete set of objects as collected during the classification fo the pilot set, V_k is a single object with an index k within the pilot set, $\mathcal{V}$, described by multiple features, V_{obs} is a single object that must be classified and F is a A single [named] Feature. Z is a Hidden or incomplete data in a Mixture Model, α_i represents the mixture component coefficient or component Prior probability and Θ_i is a single component of the mixture model with the named features $\prod$ represents the product operator while $\sum_{k=1}^{K}$, represents the sum operator with the index k varying from 1 to K.

$$P(V_{obs}|\Theta) = \sum_{i=1}^{M} \alpha_i P(V_{obs}|\Theta_i) \tag{2}$$

where Θ denotes the parameters of the distribution of the objects that compose our Mixture Model. These parameters include the M component vectors, the Θ_is, and the proportions in which these are mixed, the α_is. Each α_i represents the prior probabilities of the component i in the mixture model, subject to the constraint $\sum_i \alpha_i = 1$. The term $P(V_{obs}|\Theta_i)$, can be determined using (3) where each Θ_i is a multivariate vector of Bernoulli probabilities each of whose N components indicate the probability of success for a particular feature.

$$P(V_{obs}|\Theta_i) = \quad \prod_{j=1}^{N} {}_j\Theta_i^{{}^jV_{obs}} (1 - {}_j\Theta_i)^{(1 - {}_jV_{obs})} \tag{3}$$

Corresponding Author: Filipe Ferreira, email:cfferreira@isr.uc.pt

To obtain the mixture parameters that explain a particular FIM $\mathcal{V}$, consisting of K observations it is assumed that the objects are independent and the likelihood of the mixture satisfying the FIM is expressed thus (4).

$$P(\mathcal{V}|\Theta) = \prod_{k=1}^{K} P(V_k|\Theta) = \mathcal{L}(\Theta|\mathcal{V}) \tag{4}$$

The optimisation task to find the mixture that best explains this $\mathcal{V}$ can be expressed as in (5), i.e. find the value of Θ that best satisfies the distribution of features in $\mathcal{V}$.

$$\Theta^* = argmax_{\Theta}\mathcal{L}(\Theta|\mathcal{V}) \tag{5}$$

The preferred method of solving the Mixture Model problem is the Expectation Maximisation algorithm. McLachlan ([McLachlan & Peel, 2000], page 19) states '...it will be seen that conceptualization of the mixture model ...(hidden data + component distributions)... is most useful in that it allows the Maximum likelihood estimation of the mixture distribution to be computed via a straightforward application of the EM algorithm.'. The EM method applied to the Mixture problem assumes that the data is only partially available. It becomes fully known through the use of a vector of coefficients denoted henceforth as Z, called the 'missing data' or the 'hidden data' or still the 'unobserved data'. If we introduce this Z to expression (2) we can now express the likelihood of the observations given the entire data as in (6) and further simplify it to (7).

$$\mathcal{L}(\Theta|\mathcal{V}, Z) = \sum_{k=1}^{K} z_k log(\sum_{i=1}^{M} \alpha_i P(V_{obs}|\Theta_i)) \tag{6}$$

$$\mathcal{L}(\Theta|\mathcal{V}, Z) = \sum_{k=1}^{K}\sum_{i=1}^{M} z_{ki}(log(\alpha_i) + log(P(V_{obs}|\Theta_i))) \tag{7}$$

The EM algorithm proceeds in two stages: the *Expectation* stage attempts to reach the best value for the missing data Z, by keeping the parameters of the Mixture model constant(8), while the subsequent *Maximization* stage attempts to optimise the components and mixing parameters themselves by using the values of the 'missing data' obtained in the expectation step just performed (9), (10). The method then alternates between the two steps until some termination criteria is satisfied.

$$z_{ki} = \frac{\alpha_i P(V_k|\Theta_i)}{\sum_{j=1}^{M} \alpha_j P_j(V_k|\Theta_j)} \tag{8}$$

$$\alpha_i = \frac{\sum_{k=1}^{K} z_{ki}}{K} \tag{9}$$

$$\Theta_i = \frac{\sum_{k=1}^{K} z_{ki} V_k}{\sum_{k=1}^{K} z_{ki}} \tag{10}$$

Corresponding Author: Filipe Ferreira, email:cfferreira@isr.uc.pt

This termination criteria is usually a lack of change in the mean error when the Mixture Parameters are applied to the original data. In the case of such applications, where the parameters of the Mixture models are required for the purpose of classification, the process is usually stopped quite early, when the reduction in the Mean Error is not significant.

Mixture models used for classification make use of both, the Mixture parameters and the posterior probabilities over the components, the Z are used to evaluate the likelihood in the space of the objects in the reference sequence as in (11) where $P(V_k)$ represent the prior probabilities on each index k.

$$P(k|V_{obs}) = \frac{\sum_{j=1}^{M} P(V_k) z_{ki} \alpha_j P(V_{obs}|\Theta_j)}{\sum_{k=1}^{K} \sum_{j=1}^{M} P(V_k) z_{kj} \alpha_j P(V_{obs}|\Theta_j)} \tag{11}$$

The *Maximum Likelihood Estimation* approach is used to obtain the matching object, the index k^*, in $\mathcal{V}$ that best describes the object to be matched, V_{obs}.

$$P(k = k^*|V_{obs}) = \max_k P(k|V_{obs}) \tag{12}$$

4 Simulation results

To demonstrate the classification procedure, a matrix of binary data corresponding to a simulated set of barcode/data-matrix objects is utilised. As was explained earlier a row in the Feature Incidence Matrix denotes a single feature. Each column denotes a particular object. The objects index itself advances, from the left to the right. The presence of a feature is indicated by a black square and the absence by a white square. As can be seen in Fig. 2, the FIM is composed of what can be easily recognised as two, distinct objects each of which is a complement of the other. Some features in some objects are flipped in order to simulate small variations between the types of objects to be classified. If we model this FIM as a mixture model we would still expect to get two main classes, as denoted by the main components as long as the variations in the features were not significant.

After running the EM algorithm for mixture models, with four components respectively, we end up with the components as shown in Table 1 and the mixture coefficients as shown in Table 2, respectively. As can be seen, most of the layout of the FIM is explained by components Θ_1 and Θ_3 (the original components in the noiseless data) and the distribution of these components is quite similar to the corresponding components for the noiseless FIM.

Thus it can be seen that the objects in the pilot set are neatly separated into classes based on the components of the Mixture Model. By using these components we can easily classify any new object. Additionally, it is possible to evaluate the quality of the classification depending on how close the binary descriptor of the object to be classified is to each of the components.

5 Conclusions

The Bernoulli mixture method has already proved its ability to classify objects that are represented using multiple binary features. As applied to the Mobile-robot-localization, it has allowed for the recovery of the position of the robot under difficult conditions.

Corresponding Author: Filipe Ferreira, email:cfferreira@isr.uc.pt

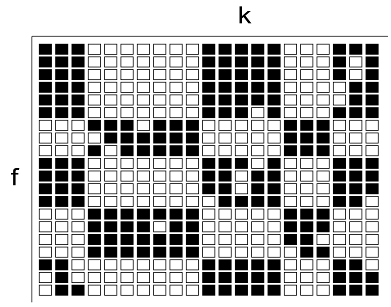

Figure 2: A simulated FIM to demonstrate the working of the Bernoulli Mixture Model

Table 1: Components for noisy FIM shown in Fig. 2.

Θ_1	Θ_2	Θ_3	Θ_4
1	1	0	1
0.90	1	0	1
0.90	1	0	1
0.90	1	0	1
0.90	1	0	1
0.90	1	0	1
0	0	0.91	0
0	0	0.91	0
0	0	0.91	0
0.90	1	0	1
0.90	1	0	1
0.90	1	0	1
1.00	0	0	0
0	0	1	0
0	0	0.91	0
0	0	0.91	0
0	0	0.91	0
0.90	0.50	0	0.44
0.80	1	0	1
0.90	1	0	1

Table 2: Coefficients for noisy FIM shown in Fig. 2.

α_1	α_2	α_3	α_4
0.44	0.06	0.48	0.02

Corresponding Author: Filipe Ferreira, email:cfferreira@isr.uc.pt

The Bernoulli Mixture model manages to capture the correlation between the features on the pilot set of objects that were classified. These correlations can be extracted and utilized to classify new objects, even when the there is some variation in the values of features within the group to be classified. Classification improves when the classes are substantially different from each other and when the proportion of features that change is smaller than the features that define one or the more classes.

We now wish to test the method using real data obtained by reading codes off Data-Matrices created to specifically represent values of interest at the time of disassembly of a component or object. We expect methods such as these to accompany the development of recycling technologies for electronic products and other toxic technologies.

6 Acknowledgments

The first author was partially supported by the EU-BACS FP6-IST-27140 project. The second author is supported by the EU-BACS FP6-IST-45048 project.

References

[Agapakis & Stuebler, 2006] Agapakis, J. & Stuebler, A. (2006). Online: http://www.assemblymag.com.

[Ferreira et al., 2006] Ferreira, F., Santos, V. & Dias, J. (2006). In IEEE Conference on Multisensor Fusion and Integration, Heidelberg, Germany.

[García-Hernández et al., 2004] García-Hernández, J., Alabau, V., Juan, A. & Vidal, E. (2004). In Proc. of the LEARNING04, ISBN 84-688-8453-7, (et al., A. R. F.-V., ed.),, Elche (Spain).

[Juan & Vidal, 2004] Juan, A. & Vidal, E. (2004). In International Conference on Pattern Recognition (ICPR'04) vol. 3, pp. 367–370,.

[Kaban & Girolami, 2000] Kaban, A. & Girolami, M. (2000). In Proceedings of the 15th International Conference on Pattern Recognition (ICPR'00) vol. 2,.

[Lowe, 1999] Lowe, D. G. (1999). In Proc. of the International Conference on Computer Vision, Corfu pp. 1150–1157,.

[McLachlan & Peel, 2000] McLachlan, G. & Peel, D. (2000). Finite Mixture Models. John Wiley and Sons.

[Mikolajczyk & Schmid, 2001] Mikolajczyk, K. & Schmid, C. (2001). In International Conference on Computer Vision, Vancouver.

[Saar & Thomas, 2002] Saar, S. & Thomas, V. (2002). Journal of Industrial Ecology, MIT Press 6, 133–146.

[Saar & Thomas, 2004] Saar, S. & Thomas, V. M. (2004). Resources, Conservation and Recycling 41, 15–22.

[Sajama & Orlitsky, 2005] Sajama & Orlitsky, A. (2005). In Proceedings of the 22 nd International Conference on Machine Learning, Bonn, Germany.

Corresponding Author: Filipe Ferreira, email:cfferreira@isr.uc.pt

[Schmid & Mohr, 1995] Schmid, C. & Mohr, R. (1995). Technical report INRIA.

[Schmid & Mohr, 1997] Schmid, C. & Mohr, R. (1997). IEEE Transactions on Pattern Analysis and Machine Intelligence 19, 530–535.

[Wang & Kaban, 2005] Wang, X. & Kaban, A. (2005). In Sixth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL ’05).

Corresponding Author: Filipe Ferreira, email:cfferreira@isr.uc.pt