Statistical classification of raw textile defects

Vittorio Murino, Manuele Bicego, and Ivan A. Rossi
Dipartimento di Informatica, Università di Verona - Italy
{vittorio.murino@univr.it, bicego@sci.univr.it}

Abstract

In this paper, the problem of classification of defects occurring in a textile manufacture is addressed. A new classification scheme is devised in which different features, extracted from the gray level histogram, the shape, and co-occurrence matrices, are employed. These features are classified using a Support Vector Machines (SVM) based framework, and an accurate analysis of different multi-class classification schemes and SVM parameters has been carried out. The system has been tested using two textile databases showing very promising results.

1. Introduction

The analysis of raw textile images is an important feasible application of the pattern recognition research: the aim is to automatically detect and classify possible defects occurring in raw textile. Regarding the detection problem, several approaches appeared in the last years, based on statistical or spectral techniques (e.g., see [12, 4, 5]). On the other side, the classification problem has been poorly investigated in the past, and only few papers are present in the literature [17, 6, 1]. In particular, in [17] a multi-scale multi-channel wavelet approach has been proposed, based on the Minimum Classification Error training (MCE). In [6], a fractal-based defect classification has been proposed; another work [1] classifies first and second order statistic features using a competitive neural classifier. It is also important to note that commercial systems for textile quality control are limited to the detection of the defects, but, to the best of our knowledge, no classification tasks are actually carried out.

In this paper, a new approach is proposed, able to analyze raw textile images to detect and classify the defects in a precise set of typologies. The focus of this paper is only on the classification stage, so, the detection phase has been assumed already done. The proposed method assumes that the information needed for characterizing the different types of defects can be extracted from the shape, the grey-level profile, and the texture alteration. Such information is derived by computing from the image three kind of features, namely, those extracted from the gray-level histogram, those characterizing the shape of the defect [9], and especially those related to the co-occurrence matrix [8]. In particular, in the latter case, the typical approach lies in extracting several features from these matrices [8], which could indeed result correlated or poorly relevant for some kind of defects. In our approach, we directly consider the coefficients of the normalized co-occurrence matrix as features, representing a sort of statistical characterization of the different defects, like in the *Textons* theory [11]. The extracted features have been classified using a framework based on Support Vector Machines [3], a classification tool widely employed in recent years in several application domains.

The approach has been tested using several different textiles, with natural noisy contaminants, derived from two data sets, with different types of defects. We considered cases of both oriented or non-oriented textiles, obtaining really promising results.

The rest of the paper is organized as follows. Section 2 describes the feature extraction process, focusing on the motivations of the use of the several kinds of features. The actual classification is then reported in Section 3, addressing in particular different possible schemes based on SVM. In Section 4, experimental results on two different data sets are shown, and, in Section 5, final considerations are drawn.

2. Feature Extraction

The starting point of our algorithm is an image which contains the defect of the textile. The detection phase has been performed by using a simple method based on the second order statistics of the image, but, indeed, any other detection technique can be used. From this image, we have extracted features related to the histogram, to the defect shape, and to the co-occurrence matrix. The final feature pattern is then composed by all of these features. The choice of these features is due to the complexity and the subtle nature of the defects. Actually, some defects are quite easy to detect

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

and classify (see Fig. 1(a)), but others are very sneaky and hardly visible even to a human operator (Fig. 1(b)), so that a large number of features of different types and a robust classification method are justified to get quite accurate performances.

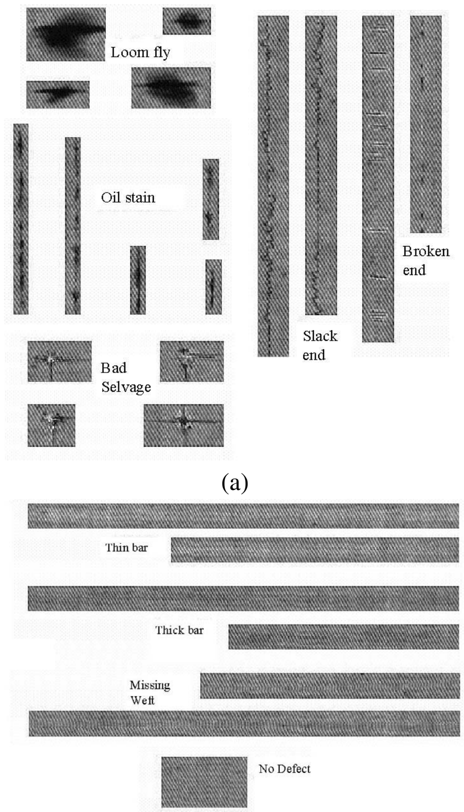

Figure 1. Some examples of textile defects: (a) easy defects; (b) difficult defects.

2.1. Gray level histogram

We have noted that histograms of defects of the same class show a similar general behavior; nevertheless, the standard features extracted from the histogram (like mean, standard deviation, entropy) are not able to properly characterize this similarity. In our approach, a study of the histogram function $H(x)$ has been performed: first, the profile was smoothed by applying a mean filter; second, the following features are computed:

$$\begin{aligned} c_1 &= \arg\max_{x\in[0,\ldots,255]} H(x) \\ c_2 &= \min_{x\in[0,\ldots,255] \text{ s.t. } H(x)\neq 0} H(x) \\ c_3 &= \max_{x\in[0,\ldots,255] \text{ s.t. } H(x)\neq 0} H(x) \\ c_4 &= (c_3 - c_2)/255 \\ c_5 &= c_4/c_1 \\ c_6 &= \frac{\sum_{x\in[0,\ldots,255] \text{ s.t. } H'(x)>0} H'(x)\times H(x)}{\sum_{y\in[0,\ldots,255]\nabla H'(y)<0} H'(y)\times H(y)} \end{aligned}$$

where $H'(x) = \frac{dH(x)}{dx}$ is the derivative of the histogram function. Some preliminary results (not presented in this paper) have shown that the proposed set of features is more descriptive with respect to the mean, standard deviation, skewness and other similar features that could be extracted from the histogram.

2.2. Shape Descriptors

An interesting analysis of features used for classifying defects in general textures has been proposed in [9], in which the Chain Code, the Pairwise Geometric Histogram (PGH), and the shape descriptors have been analyzed and compared. The results was that the best compromise between efficiency and descriptiveness is represented by shape descriptors. In the present work, the following features are used: area, centroid, major axis length, minor axis length, eccentricity, orientation, convex area, Euler number, equivalent diameter, and extension [10].

2.3. Co-occurrence matrices

This matrix represents one of the most used informative feature for texture analysis. Considering the textile case, when a defect is present, the resulting texture is surely modified and this behavior can be captured and characterized by the co-occurrence information. The gray level Co-occurrence Matrix (CM) can be defined as follows [8]. Given an image with Q gray levels, a CM is a matrix $Q \times Q$ where at position (i, j) there is the number of occurrences of dipoles with gray level i and j. The dipoles are defined with a vector $D = (\alpha, \theta)$, where α is the dipole orientation and θ is the length. The length parameter, together with the number of gray levels of the image (which has been fixed to 32 in our approach), is crucial for the effectiveness of the matrix, and strongly depends on the type of the texture under analysis [13]. Typically, several features are extracted from the CMs [8], computed for several vectors D, resulting in a large set of often correlated features. To solve this problem, an interesting approach has been proposed in [2], where a χ^2 test for determining the expressivity of a CM is used, allowing to choose the best vectors D.

In our approach, we propose an alternative use of the co-occurrence matrices, that computes the best dipole length and avoids the feature computation. More in detail, first, we perform the χ^2 test for CMs computed in the four principal directions $0^o, 45^o, 90^o, 135^o$ on many samples of the original textiles without defects, in order to determine the best dipole length for any direction used. Second, we compute and normalize the CM obtaining a sort of probability matrix. Finally, features are not computed from the CM, but the CM coefficients are directly used as features for the Support Vector Machine architecture. This permits to remove the feature computation phase, which is the major drawback of the technique, letting to the SVM the goal of finding the real useful data in different CMs. Moreover, these normalized matrices describe the statistical behavior of the defect in the four principal directions, independently from the image dimension, and using the most suitable dipole for that kind of textile. These matrices are in some way similar to the *textons*, a concept introduced by Julesz [11], representing internal geometrical texture primitives used to differentiate between different textures. It has been shown [15] that the CMs are able to characterize the oriented microstructures present in the texture, so as to characterize in some way the *textons* of the image.

3. Classification

The features extracted from the previous stage are organized in a feature vector, and used to train the Support Vector Machines (SVM). SVMs have been successfully employed in a wide range of applications in the recent years, and have been chosen due to their high generalization capability, and to their major ability to deal with high dimensionality space, such as that resulting from our features. Due to the limited space, SVMs are not fully described here, and a good introduction can be found in [3].

The basic SVM scheme relies to binary classification and in order to deal with multi-class problems a generalized scheme should be introduced. In this paper we have compared three schemes:

- *SVM 1-vs-R (One versus Rest)*: this is the standard scheme, which consists in building one SVM for each class C_i, using as negative examples all the patterns of the other classes. An unknown pattern is assigned to the class whose SVM responds positively.
- *SVM 1-vs-1 (One versus One) Binary Decision Tree, BDT*: this method, also called of "tennis tournament" has been introduced in [14], and builds a binary decision tree where each node represents a SVM. The recognition is performed following the rules of a tennis tournament. Each class is regarded as a player, and in each match the system classifies the test defect according to the decision of the SVM trained on the pair of players involved in the match. The winner identities, proposed by each SVM, will be propagated to the upper level of the tree, playing again. The process continues until the root is reached, and the winner is the class to be assigned to unknown pattern.
- *SVM 1-vs-1 Max Win*: this method, proposed in [7], trains one SVM for each pair of classes. Given an unknown pattern, all SVM are evaluated, counting for each class the number of wins. The pattern is assigned to the class with the maximum number of wins.

We have performed several testing (not reported here) using all these three methods, and the best scheme resulted the SVM 1-vs-1 - BDT: this scheme is also the most efficient in terms of time required for training and classification. Another session of preliminary testing has shown that the best kernel is the Exponential Radial Basis Function, with $\sigma = 15$.

4. Experimental Results

The proposed approach has been tested using two databases: the first is the PARVIS database[1], which contains 2 kinds of textiles. In this case, the orientation of the textile is known. Some of the defects have been synthetically reproduced, in order to augment the population of the data in the testing phase, using tools of photo-realistic graphics. The second is the TILDA database [16], realized by the *Technische Universitt Hamburg* in 1995, which consists in 4 different textiles. In this case, the orientation of the textile is not known: this is the hardest case, typically disregarded in literature, but important to test the robustness of the approach. In both cases, we have included also the class of "No Defect", in order to validate and to refine the defect detection scheme. We have computed the Leave One Out (LOO) error, reporting the results for each class of defect. The defects nomenclature used derives from the Italian Textile Industry and from literature. Results for the PARVIS database are reported in the Table 1. The total numbers of elements is 1117. Results are really satisfactory, a nearly perfect accuracy is reached.

Results for the four textiles of the TILDA database are shown in Table 2. The number of elements is 1333. Since the orientation of the textile is not known, instead of using four gray level CMs, we used only one isotropic matrix, where the dipole length θ is the best over the four main directions. Also in this case the results are satisfactory, even if in this non oriented case the method loses part of its effectiveness. However, in actual applications, the raw textile is typically analyzed assuming a nearly constant orientation, and, indeed, in the PARVIS database, in which the textile orientation is known, results are much better.

[1] Database provided by PARVIS srl - Milano.

Defect	Population	Accuracy
No Defect	81	100%
Loom fly	42	100%
Thin bar	41	97.56%
Broken end	41	97.56%
Thick bar	41	100%
Double weft	41	100%
Slack end	41	100%
Missing draw	41	100%
Wrong draw	46	100%
Bad selvage	31	93.54%
Oil stain	46	100%
Missing Weft	20	100%
Average classification accuracy: 99.11%		

Table 1. Classification accuracies for the PARVIS database.

Defect	Population	Accuracy
No Defect	376	99.46%
Slack end	110	88.18%
Broken end	169	89.35%
Hole	89	86.51%
Oil stain	256	94.92%
Rip	106	87.73%
Kink	86	87.21%
Unrelated corpus	49	97.96%
Missing weft	64	85.93%
Bad selvage	28	89.29%
Average classification accuracy: 92.87%		

Table 2. Classification accuracies for the TILDA database.

5. Conclusions

In this paper a novel approach to textile defects classification has been proposed. The method uses features extracted from the histogram, shape descriptors and the co-occurrence matrices so as to efficiently classify a large number of defects. These features are classified using a Support Vector Machines based method. Classification accuracies obtained on two different dataset are really satisfactory, and make the proposed approach an effective and promising visual inspection approach for the classification of textile defects.

6. Acknowledgments

The authors gratefully acknowledge Dr. Giovanni De Toni of PARVIS srl (www.parvis.it) to have provided one data set of textile defects.

References

[1] M. Abdulhady, H. Abbas, and S. Nassar. Fabric fault classification using neural trees. In *Proc. of IEEE Int. Conf. on System, Man and Cybernetics*, volume 6, 2002.

[2] A. Bodnarova, J. Williams, M. Bennamount, and K. Kubik. Optimal textural features for flaw detection in textile materials. In *Proc. of IEEE Conf. on Speech and Image Technologies for Computing and Telecommunications*, volume 1, pages 307–310, 1997.

[3] C. Burges. A tutorial on support vector machine for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[4] J. Campbell, C. Frawley, D. Stanford, F. Murtagh, and A. Raftery. Linear flaw detection in woven textiles using model-based clustering. *Int. J. of Imaging Systems and Technology*, 10:339–346, 1999.

[5] F. Cohen, Z. Fan, and S. Attali. Automated inpsection of textile fabrics using textural models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(8), 1991.

[6] A. Conci and C. Proena. A system for real-time fabric inspection and industrial decision. In *Proc. of ACM Int. Conf. on Software Engineering & Knowledge Engineering*, pages 707–714, 2002.

[7] J. Friedman. Another approach to polychotomous classification. Technical report, Stanford University, 1996.

[8] R. Haralick, K. Shanmugan, and I. Dunstein. Textural features for image classification. *IEEE Trans. on Systems, Man and Cybernetics*, SMC-3:610–612, 1973.

[9] J. Iivarinen and A. Visa. An adaptive texture and shape based defect classification. In *Proc. of Int. Conf. on Pattern Recognition*, volume 1, pages 117–122, 1998.

[10] R. Jain, R. Kasturi, and B. Schunck. *Machine Vision*. McGraw-Hill, 1995.

[11] B. Julesz. A theory of preattentive texture discrimination based on first order statistics of textons. *Biological Cyber-Netics*, 41:131–138, 1981.

[12] A. Kumar and G. Pang. Defect detection in textured materials using gabor filters. *IEEE Trans. on Industry Applications*, 38(2):425–440, 2002.

[13] S. Majumdar and D. Jayas. Classification of bulk samples of cereal grains using machine vision. *Journal of Agric. Eng. Res.*, 73:35–47, 1999.

[14] M. Pontil and A. Verri. Support vector machines for 3-D object recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20:637–646, 1998.

[15] L. Soh and C. Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurances matrices. *IEEE Trans. on Geoscience and Remote Sensing*, 37(2):780–795, 1999.

[16] Technische Universitt Hamburg-Harburg. Ein referenzdatensatz zur evaluierung von sichtprfungsverfahren fr textiloberflchen. Interner Bericht 96, 1996.

[17] X. Yang, G. Pang, and N. Yung. Fabric defect classification using wavelet frames and minimum classification error training. *IAS Industry Application Conference*, 1:290–296, 2002.