

Defect Detection in Woven Fabrics by Analysis of Co-occurrence Texture Features as a Function of Gray-level Quantization and Window Size

P.S.H. Pallemulla, S.J. Sooriyaarachchi, C.R. de Silva and C.D. Gamage

Abstract: In this experimental research, the effects of gray-level quantization and tiling window size on 22 gray-level co-occurrence matrix features were investigated in the context of automated woven fabric defect detection. A dataset comprising 1426 128×128 images was used, in which defective and the defect-free images were split in a 50:50 ratio. Experiments were carried out with seven quantization levels ($L = 4, 8, 16, 32, 64, 128$ and 256) and four window sizes ($N = 8, 16, 32, 64$). The features were extracted from each image in the training set for each $\langle L, N \rangle$ combination and thereafter were ranked using the joint mutual information metric. Next, for each $\langle L, N \rangle$ combination, a k-nearest neighbour classifier was trained, first with only the highest-ranking feature and thereafter iteratively by adding features of lower ranks. It was observed that a minimum of nine features were needed to achieve an acceptable ($>90\%$) F1 score for any $\langle L, N \rangle$ combination, except when N is relatively large. The two features that contribute to improving the F1 score for any $\langle L, N \rangle$ combination were found to be Homogeneity I and Homogeneity II. It was also noted that using an 8×8 window on images with 128 gray levels resulted in a practically usable high F1 score (96.39%) with the least number of features (14).

Keywords: Fabric defect detection, GLCM, Gray-level quantization, Window size

1. Introduction

Almost 30% of fabric defects go unnoticed to the human eye during a typical fabric inspection process in the apparel manufacturing industry [19]. Undetected defects can result in large losses in the production as they affect the quality of the output and often result in the rejection of garments at the final quality control stages. Many researchers have attempted to solve this issue by using computer vision-based techniques. However, due to the diversity of properties in different fabric types used in the industry and their sensitivity to the algorithms used, most of these studies limit their scope to defect detection in a few types of fabric using few select algorithms.

Some more common types of defects in fabrics manifest as local anomalies in the fabric texture. Following this assumption, texture analysis has been widely used to separate defective areas of the fabric from the defect-free area. Among the many methods used, the gray-level co-occurrence matrix (GLCM) [1] has been frequently used to calculate textural features owing to its ability to discriminate between textures that are visually easily separable.

An important preprocessing step that is known to improve the textural features estimated from the GLCM is gray-level re-quantization. At present, studies on the effect of the gray-level

quantization on texture classification have been conducted for irregular textures in the geological [5] [7] and biomedical [6] [8] [9] fields, but not for near-regular textures such as those in textile samples. All these studies recommend using fewer gray levels in order to obtain better texture classification accuracy, whereas two of them [5][6] recommend gray-level quantization values that are dataset-specific. Therefore, it is clear that the current available literature is limited to irregular textures. As woven fabric has near-regular texture [18], we apply our method and conduct our analysis on such textures.

Eng. P.S.H. Pallemulla, AMIE(SL), B.Sc. Eng. Hons (KDU), Research Assistant, University of Moratuwa.
Email:sajithpallemulla7@gmail.com

 <http://orcid.org/0000-0002-2115-5904>

Eng. (Dr.) S.J. Sooriyaarachchi, AMIE(SL), MCS(SL), AMTILP, B.Sc. Eng. Hons (Peradeniya), M.Sc.

(Moratuwa), Ph.D. (Moratuwa), Senior Lecturer in Computer Science and Engineering, University of Moratuwa, Sri Lanka. Email:sulochanas@cse.mrt.ac.lk

 <http://orcid.org/0000-0003-0905-7196>

Eng. (Dr.) C.R. de Silva, C. Eng., MIE(SL), B.Sc. Eng. Hons (Moratuwa), M.Eng. (NTU), Ph.D. (NTU), Senior Lecturer in Computer Science and Engineering, University of Moratuwa, Sri Lanka. Email:chathura@cse.mrt.ac.lk

 <http://orcid.org/0000-0002-6760-6352>

Eng. (Dr.) C.D. Gamage, C. Eng., MIE(SL), B.Sc. Eng. Hons (Moratuwa), M.Eng. (AIT), Ph.D. (Monash), Senior Lecturer in Computer Science and Engineering, University of Moratuwa, Sri Lanka.

Email:chandag@cse.mrt.ac.lk.

 <http://orcid.org/0000-0002-8182-812X>



Motivated by the absence of a basis upon which an optimal gray-level quantization level can be reliably chosen for co-occurrence feature extraction in plain-weave fabrics, this study aims to investigate the effect of different values for gray level quantization on fabric defect detection. The study uses multiple window sizes for partitioning the image into smaller tiles, which is a common practice for GLCM in this domain. Moreover, the behaviors of 22 textural features that can be computed from the GLCM are also considered in the classification. The contributions of this paper are the highlighting of co-occurrence features that contribute most to the classification, the determination of the least number of features required to achieve appreciable detection accuracy and the establishment of an objective framework for the choice of the optimal quantization/window-size combination in future research on plain-weave fabric defect detection.

2. Materials and Methods

2.1. Dataset

Our dataset was created as follows. From the 195 512x512-pixel images in the publicly available Cotton Incorporated dataset [20], the images of unpatterned plain-weave fabrics containing Broken End, Broken Pick, Burl, Coarse End, Coarse Pick, Dirty Yarn, Foreign Yarn, Knot, Oil Spot and Slubs were selected. They were randomly cropped with a 128x128 window and annotated with the support of domain experts. As the defect-free images obtained thus outnumbered the defective images, the set of defect-free images was undersampled to obtain a perfectly balanced dataset comprising 1426 images.

2.2. Gray-level Quantization

The gray level depth (L) in an image has an effect on the textural features calculated from the GLCM. Many research encourage gray-level depth reduction through down sampling due to multiple reasons. First, a lower depth requires less computational time and cost and thereby facilitates real-time processing with limited resources. Additionally, when similar gray levels are merged for reduction, the effect of noise on the image is also averaged to a certain degree [5]. Similarly, the down-sampling process also mitigates undesirable effects of uneven illumination [6]. All these features contribute in reducing the extent of information included in the sample to match the extent

required in isolating defects from other natural variation in the fabric texture.

In this study, the classification accuracy of co-occurrence textural features for seven possible choices for uniform gray-level quantization ($L = 4, 8, 16, 32, 64, 128$ and 256) are studied. The uniform quantization scheme has been used in several studies [5] [7] [9].

2.3. Window Size

In fabric defect detection, GLCMs are calculated for non-overlapping subregions of the image since most defects are extremely small compared to image size. The usual approach is to split the image into tiles using a $N \times N$ sliding window.

Theoretically, larger windows provide a more accurate classification of texture [7], but having a window smaller than the smallest defect ensures its detection [4]. Using a smaller window also means splitting the image into smaller tiles and hence the computation of more GLCMs, thereby increasing the computational cost further. Even though the largest window always provides the fastest speed of computation since it splits the image to the least number of tiles, it will fail to detect defects that are comparatively much smaller than its size.

To determine the window size that is sufficiently large so as to provide an accurate estimation of texture without ignoring the distortion at the defect, four window sizes ($N = 8, 16, 32, 64$) are studied.

2.4. Gray-level Co-occurrence Matrix

The GLCM is used to extract the second-order statistics of an image. The matrix represents the number of occurrences of gray-level pairs, i and j , that are d distance apart along a direction θ . Hence, the GLCM captures the frequency of spatial relationships among pixels in textured images. The GLCM is defined as [1]

$$C(i, j) = || \{ [x_1, y_1], [x_2, y_2] \} | x_2 - x_1 = d \cos \theta, \\ y_2 - y_1 = d \sin \theta, I(x_1, y_1) = i, I(x_2, y_2) = j || \dots (1)$$

where (x_1, y_1) and (x_2, y_2) are pixels, $I(\cdot)$ is the gray level of the pixel and $|| \cdot ||$ is the number of pixel pairs that satisfy the conditions.

Studies on textile images of multiple resolutions have revealed that a distance d of one pixel gives better classification accuracy [2]. For this reason and in order to allow relative comparisons to be made with respect to the parameters under consideration, a

distance of one pixel is used in this study. For the sake of making the textural features invariant to rotation, the GLCMs of all possible directions ($\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$) are obtained from a tile, and the features calculated from them are averaged [1] [4]. Prior to the calculation of features, the GLCMs derived from each tile are normalized.

2.5. Extracting the Feature Vector

In the interest of improving the detection accuracy, the textural feature vector extracted from each image is compared with that of a defect-free image of the same fabric. For referential clarity, images belonging to the latter group will be known as reference images and images in the dataset will be known as query images.

The procedure that is used to extract the textural feature vector from a query image and compare it with the fabric's reference image is detailed in the rest of this subsection. Apart from the gray scaling and gray-level re-quantization, no additional preprocessing was applied to the images, in order to preserve all micro-level features in their natural form.

An image that is re-quantized to L_1 gray levels and tiled with a window size of $N_1 \times N_1$ results in a set of m tiles, where m is given by

$$m = \frac{128 \times 128}{N_1 \times N_1} \quad \dots(2)$$

Four GLCMs, one for each θ , are derived from each tile. The 22 textural features as presented in [9] are calculated from each GLCM and the same features calculated from the four GLCMs of one tile are averaged to ensure rotational invariance. This results in each of the m tiles being described by a 22-element feature vector, given by the matrices $R_{m \times 22}$ and $Q_{m \times 22}$ for the reference image and query image, respectively, in Figure 1. The aforementioned set of 22 features includes GLCM features such as energy, entropy, correlation II and homogeneity II [2] [12] that are commonly used in fabric defect detection. The rest of the features are used for texture classification in general [4]-[8].

For the defect-free reference image of the fabric, it is assumed that the same feature across all m tiles is similar in value, because woven fabrics have near-regular texture. Therefore, to represent the reference image by one textural

feature vector ($r_{1 \times 22}$), each feature is averaged across all m tiles.

A query image may either be defect-free or defective. If a query image is defective, distortions in the texture of a tile, caused by the presence of a defect, lead to one or more textural features having abnormally high or low values. Therefore, both the maximum of each textural feature ($q_{max_{1 \times 22}}$) and the minimum ($q_{min_{1 \times 22}}$) are compared with $r_{1 \times 22}$ by calculating the absolute deviation between them. The maximum absolute deviation $t_{1 \times 22}$ is taken as the feature vector of the query image.

$$t_{1 \times 22} = \max(|r_{1 \times 22} - q_{max_{1 \times 22}}|, |r_{1 \times 22} - q_{min_{1 \times 22}}|) \quad \dots(3)$$

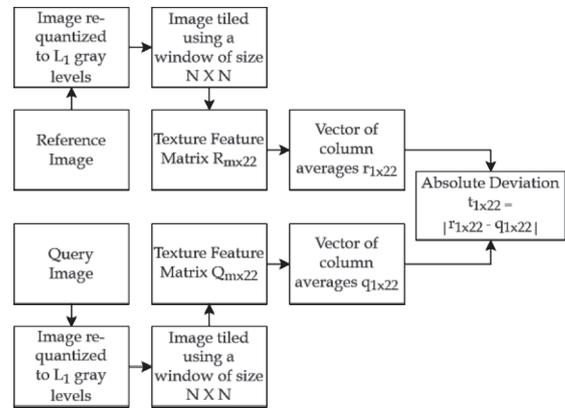


Figure 1 - Extracting and Comparing the Feature Vector of a Query Image ($q_{1 \times 22}$) and Reference Image ($r_{1 \times 22}$) to get the Relative Textural Feature Vector ($t_{1 \times 22}$).

According to a previous study [12], a comparison of the feature average across tiles compared with the feature maximum or minimum across tiles provides a feature set suitable for more accurate classification.

2.6. Feature Ranking

Some features in high-dimensional feature spaces may carry non-discriminative or correlated information from the problem domain's perspective. These features do not contribute to the classifier performance and, in some cases, could even influence it negatively by biasing the classifier on directions outside the problem domain. Feature ranking determines the features which are most discriminative and relevant to a given classifier model.

In order to separate the feature ranking process from the classification, joint mutual information (JMI) is used to rank features in this study. The classifier-independent features selected thus are



generic. Additional advantages of using JMI are the lower chance of overfitting during classification and speed of computation [10] compared to classifier-dependent techniques. JMI is used from among other classifier-independent methods as it provides the best trade-off between accuracy and stability [11].

The mutual information I between two random variables, x and y is defined as

$$I(x; y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad \dots (4)$$

where $p(x, y)$ is their joint probability density function and $p(x)$ and $p(y)$ are their respective marginal probabilities. When the mutual information is conditioned by a third variable z , it can be expressed as

$$I(x; y|z) = \sum_k p(z_k) \sum_{i,j} p(x_i, y_j|z_k) \log \frac{p(x_i, y_j|z_k)}{p(x_i|z_k)p(y_j|z_k)} \quad \dots (5)$$

The JMI score finds the feature that adds the most new information to a set of already selected features ($f_j \in S$). The score is formulated for a feature under investigation (f_i) as

$$JMI(f_i) = \sum_{f_j \in S} (I(f_i; c) + I(f_i; c|f_j)) \quad \dots (6)$$

where c is the target class.

2.7. Detection

Past studies on texture analysis [7] [8] [9] similar to our work, have opted to use Fisher's linear discriminant analysis (FLDA) for classification. According to the earliest study among these [7], the reason for choosing FLDA over a machine learning technique was the scarcity of samples in their dataset. Due to the dataset used in this research having an adequate number of samples for a binary classification problem and due to the ubiquity and success of machine learning techniques at present, this study uses the k-nearest neighbour (k-NN) classifier for classification with a standard five-fold cross-validation (FFCV) to determine the optimal number of neighbours.

Since the classes in the dataset are balanced, the k-NN is an acceptable algorithm for the classification. The fact that it is a non-

parametric algorithm comes as an advantage. When the k-NN is provided with a sample to classify, it calculates the dissimilarity between said sample and each sample in the training set within the feature space. Here, the Manhattan distance is used as the dissimilarity measure. The Manhattan distance between two samples I_1 and I_2 is given by,

$$d(I_1, I_2) = \sum_i |x_i - y_i| \quad \dots (7)$$

where x_i and y_i are the feature vectors of I_1 and I_2 . Then, the nearest k ($k = \{2c + 1 \mid c \in \mathbb{Z}_0^+\}$) samples are considered and the test sample is assigned the majority class.

The F1 score is used to measure classifier performance. Accuracy has also been calculated for the purpose of comparing our results with those in recent studies. The definitions of these metrics are given below:

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \quad \dots (8) \\ \text{Recall} &= \frac{TP}{TP + FN} \quad \dots (9) \end{aligned}$$

$$F1 \text{ score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \dots (10)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + FN} \quad \dots (11)$$

where TP, FP, TN and FN are the true positives, false positives, true negatives and false negatives, respectively.

2.8. Assessment of Optimal Feature Set

The dataset is split into the training set which contains 1018 images and the test set containing 408 images. The reference images used in the feature extraction are not considered part of the dataset.

For notational clarity, let $\langle L, N \rangle$ denote a single combination of a gray level and the window size.

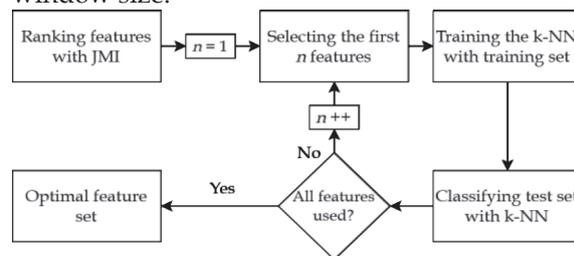


Figure 2 - Feature Ranking with JMI to Obtain the Optimal Feature Set i.e. the Set of Features that gives the Best Classification F1 Score

The experiments were performed as follows. The feature vectors for all $\langle L, N \rangle$ combinations were extracted using the method in Figure 1 for all images in the training set. For each $\langle L, N \rangle$, JMI, given by equation 6, was used to rank the 22 features in the feature vector. The ranked features were normalized to the range -1 to 1 . For each $\langle L, N \rangle$, a binary classification was carried out using a k-NN classifier with an increasing number of features, starting with the features of the highest rank and adding features of lower ranks until the full feature set is used for the classification as shown in Figure 2. The classifier performance is measured using the F1 score, given by equation 10. The set of features which produced the best F1 score is selected as the optimal feature set.

3. Results and Discussion

This section is split into two subsections. Subsection 3.1 discusses the results presented in Tables 1 and 2, and Figure 3, which shows the outcomes of the classification of the full dataset. Subsection 3.2 discusses the results presented in Figure 4, which shows the outcomes of the classification on subsets of the dataset that represent each defect.

3.1. Analysis of Results for the Full Dataset

From the results of the JMI ranking of each texture feature at each gray level quantization and window size in Table 1, it is clear that a subset of features, most prominently energy (P8), homogeneity I (P10), homogeneity II (P11) and maximum probability (P12) have considerably higher JMI scores than the rest of the features. The higher scores are mostly noticeable when the window size is smaller and fewer gray levels are used. The reason for this most likely stems from the texture characteristics of the fabric images.

Owing to the near-regularity of textures that occur in woven fabrics, the pixel intensities in their grayscale images will only take a few values that repeat throughout the image. This results in a GLCM in which a few elements are comparatively much larger than others. These elements correspond to the dominant intensity pairs occurring in the images. When the number of gray levels are reduced, pixels that contain similar intensities will be merged, thus making the GLCM fairly sparse and at the same time increasing the magnitude of the GLCM elements that correspond to the dominant intensity pairs at those gray levels. The occurrence of a defect will either introduce new

intensities or disturb the pattern of co-occurrence in its vicinity, hence resulting in GLCM elements of less magnitude in the tile(s) in which the defect occurs.

As the energy (P8) is simply the sum of the squares of the GLCM elements, the energy of GLCMs of defect-free tiles will be comparatively higher than those of defective tiles. Because the squaring of the elements further emphasizes the difference, it should theoretically be able to distinguish between defect-free tiles and tiles with defects that introduce new intensities to the image. This is presumably the reason why energy has a higher score. The maximum probability (P12) extracts the number of occurrences of the most frequently-occurring pixel pair i.e. the maximum of the GLCM and therefore behaves similar to the energy. The two homogeneity features (P10 and P11) not only depend on the magnitude of the GLCM elements, but also on their row and column indices. In other words, the homogeneity features will depend on the difference of the intensities of pixel pairs as well as the number of occurrences. Therefore, P10 and P11 should be able to detect tiles with defects that change the contrast of the image in their vicinity.

From the results of the k-NN classification on the test set with an increasing number of features in Figure 3, it is clear that using a very large window, relative to the image size, will invariably result in lower F1 scores. This result is contrary to the statement in [5] that larger windows, in theory, provide more accurate classifications. The reason smaller windows produce better results in this study is the small size of most defects that commonly occur in fabrics. Using larger windows will fail to capture the texture distortion in the locality of a defect which is significantly smaller than the defect-free area. The claim that the window size should be smaller than the defective area [4] cannot be verified here, as the smallest window size used in this study is 8×8 and there are much smaller defects in the dataset. However, it is believed that using overly small windows will have a negative effect on the classification since even subtle differences in the texture that are not considered defects will affect the textural features being calculated. These subtle differences might arise as a result of the changeable nature of the yarn structure or stray yarn fibres sticking out of the fabric.



It should also be noted in Figure 3 that, even if a greater number of gray levels is used, having a sufficiently large textural feature set will eventually result in acceptably accurate classifications (with F1 scores > 90%) given that

the window size used for tiling is considerably small. This finding is in contrast with the result in [9] which claims such a non-variation is only present if the textural features

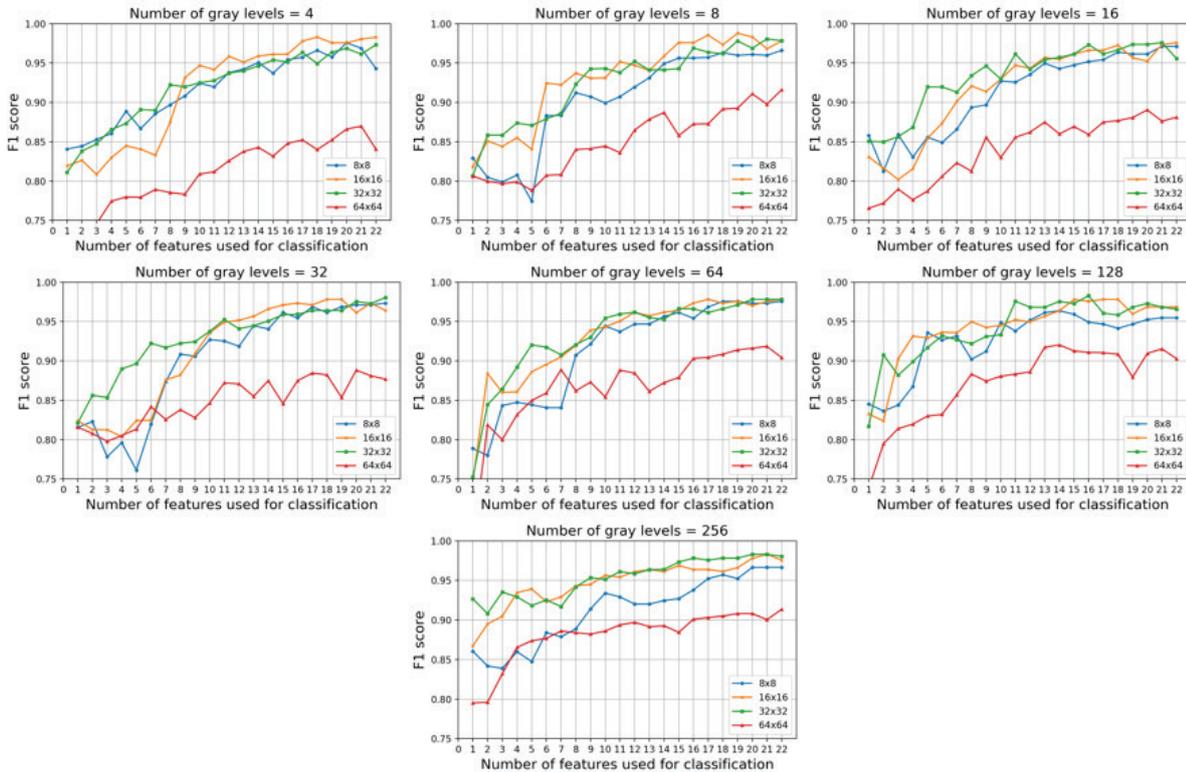


Figure 3 - F1 Scores Obtained for the Test Set after Classification by k-NN with an Increasing Number of Features Added in the Order of their Ranks for each < L, N > Combination

Table 1 - JMI Rankings for each Texture Feature (P1 to P22) for each Gray Level (L = 4 to L = 256) at Window Sizes (N = 8x8, 16x16, 32x32 and 64x64)

Window size = 8X8																						
L	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
L = 4	0.202	0.105	0.196	0.289	0.186	0.161	0.107	0.308	0.113	0.339	0.352	0.322	0.179	0.144	0.183	0.254	0.131	0.197	0.271	0.163	0.256	0.259
L = 8	0.164	0.069	0.125	0.218	0.118	0.092	0.063	0.292	0.078	0.280	0.280	0.352	0.104	0.105	0.118	0.223	0.036	0.165	0.187	0.201	0.190	0.190
L = 16	0.131	0.031	0.107	0.219	0.088	0.065	0.059	0.249	0.056	0.226	0.247	0.347	0.053	0.063	0.091	0.130	0.000	0.183	0.124	0.203	0.145	0.145
L = 32	0.106	0.013	0.066	0.138	0.025	0.039	0.019	0.256	0.000	0.218	0.215	0.303	0.018	0.032	0.062	0.101	0.009	0.193	0.076	0.200	0.127	0.128
L = 64	0.117	0.022	0.146	0.140	0.040	0.096	0.026	0.214	0.000	0.233	0.210	0.227	0.048	0.098	0.111	0.157	0.000	0.190	0.067	0.273	0.135	0.141
L = 128	0.198	0.158	0.290	0.109	0.136	0.147	0.171	0.287	0.000	0.279	0.300	0.136	0.168	0.154	0.162	0.136	0.000	0.053	0.263	0.350	0.327	0.329
L = 256	0.154	0.204	0.284	0.120	0.205	0.208	0.209	0.266	0.000	0.281	0.317	0.104	0.247	0.168	0.215	0.152	0.217	0.198	0.288	0.351	0.338	0.338
Window size = 16 X16																						
L	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
L = 4	0.218	0.089	0.219	0.244	0.205	0.180	0.073	0.290	0.093	0.363	0.353	0.333	0.095	0.103	0.177	0.260	0.198	0.134	0.262	0.118	0.268	0.271
L = 8	0.197	0.060	0.199	0.217	0.161	0.165	0.060	0.253	0.020	0.322	0.349	0.336	0.068	0.057	0.157	0.207	0.130	0.132	0.234	0.209	0.204	0.204
L = 16	0.127	0.062	0.213	0.185	0.114	0.109	0.042	0.220	0.064	0.265	0.241	0.328	0.053	0.048	0.137	0.142	0.194	0.171	0.205	0.220	0.204	0.203
L = 32	0.142	0.038	0.164	0.163	0.124	0.089	0.034	0.195	0.050	0.270	0.258	0.327	0.028	0.028	0.113	0.131	0.066	0.180	0.148	0.199	0.176	0.176
L = 64	0.152	0.022	0.231	0.123	0.087	0.113	0.029	0.215	0.086	0.229	0.208	0.229	0.044	0.091	0.181	0.122	0.013	0.197	0.028	0.203	0.163	0.163
L = 128	0.157	0.107	0.320	0.136	0.162	0.205	0.134	0.297	0.223	0.301	0.255	0.213	0.197	0.123	0.154	0.157	0.000	0.057	0.097	0.274	0.281	0.278
L = 256	0.123	0.167	0.382	0.118	0.235	0.276	0.162	0.317	0.248	0.344	0.318	0.176	0.225	0.111	0.214	0.213	0.219	0.202	0.106	0.339	0.328	0.325
Window size = 32 X32																						
L	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
L = 4	0.172	0.110	0.251	0.254	0.159	0.199	0.120	0.274	0.036	0.321	0.322	0.276	0.104	0.117	0.184	0.195	0.183	0.147	0.194	0.104	0.274	0.273
L = 8	0.161	0.071	0.248	0.203	0.168	0.167	0.072	0.259	0.032	0.327	0.317	0.328	0.066	0.086	0.148	0.132	0.110	0.129	0.193	0.213	0.273	0.273
L = 16	0.132	0.046	0.230	0.184	0.141	0.130	0.078	0.235	0.082	0.286	0.281	0.268	0.050	0.079	0.105	0.144	0.092	0.092	0.128	0.168	0.216	0.216
L = 32	0.091	0.055	0.222	0.129	0.099	0.077	0.074	0.199	0.058	0.282	0.282	0.270	0.026	0.071	0.095	0.100	0.078	0.100	0.163	0.211	0.189	0.189
L = 64	0.120	0.025	0.258	0.169	0.122	0.088	0.033	0.164	0.025	0.230	0.214	0.254	0.041	0.071	0.165	0.139	0.099	0.130	0.096	0.122	0.167	0.167
L = 128	0.143	0.087	0.359	0.130	0.172	0.125	0.122	0.243	0.184	0.250	0.290	0.237	0.198	0.110	0.180	0.126	0.028	0.078	0.089	0.251	0.205	0.206
L = 256	0.149	0.190	0.427	0.135	0.217	0.184	0.212	0.304	0.279	0.344	0.325	0.225	0.254	0.102	0.157	0.112	0.161	0.179	0.155	0.297	0.298	0.301
Window size = 64 X64																						
L	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
L = 4	0.107	0.000	0.207	0.095	0.117	0.104	0.019	0.194	0.000	0.219	0.212	0.149	0.022	0.031	0.103	0.057	0.123	0.033	0.085	0.051	0.152	0.152
L = 8	0.107	0.000	0.167	0.083	0.120	0.088	0.014	0.163	0.013	0.239	0.212	0.155	0.027	0.048	0.099	0.079	0.085	0.038	0.108	0.107	0.180	0.180
L = 16	0.085	0.017	0.186	0.070	0.153	0.074	0.000	0.128	0.019	0.236	0.239	0.113	0.041	0.037	0.058	0.079	0.047	0.018	0.042	0.063	0.155	0.155
L = 32	0.110	0.000	0.202	0.096	0.119	0.044	0.005	0.168	0.000	0.256	0.241	0.143	0.048	0.006	0.086	0.062	0.054	0.068	0.072	0.110	0.156	0.156
L = 64	0.056	0.009	0.227	0.056	0.112	0.043	0.000	0.098	0.024	0.133	0.145	0.084	0.024	0.034	0.083	0.032	0.042	0.046	0.047	0.000	0.078	0.078
L = 128	0.041	0.002	0.217	0.057	0.076	0.036	0.028	0.049	0.087	0.110	0.089	0.061	0.080	0.046	0.052	0.031	0.011	0.014	0.049	0.036	0.095	0.095
L = 256	0.045	0.035	0.268	0.066	0.055	0.010	0.040	0.050	0.142	0.118	0.135	0.042	0.119	0.045	0.064	0.019	0.039	0.063	0.068	0.033	0.057	0.057

used in the classification had not been averaged over multiple orientations of the GLCM. The reason for the difference might be that the textures analysed in [9] were natural, irregular textures whereas the textures analysed in this study are near-regular and

thus respond differently to the various orientations of the pixel pairs compared when deriving the GLCM.

The maximum F1 scores that were obtained for each $\langle L, N \rangle$ combination and the features

Table 2 - Maximum F1 Scores Obtained for each $\langle L, N \rangle$ Combination and the Number of Features used in the Classification. The Rankings of each of the 22 Features are also shown. The Numbers in the Shaded Cells are the Ranks of the Features that were used for the Classification that Resulted in the Maximum F1

Window size	Gray Levels	Max F1 score	No. of features used	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	
N = 8x8	L = 4	97.57%	20	10	22	12	5	13	17	21	4	20	2	1	3	15	18	14	9	19	11	6	16	8	7	
	L = 8	96.62%	22	12	20	13	6	14	18	21	2	19	3	4	1	17	16	15	5	22	11	10	7	9	8	
	L = 16	97.12%	21	10	21	13	5	15	16	18	2	19	4	3	1	20	17	14	11	22	7	12	6	8	9	
	L = 32	97.35%	22	10	20	13	7	17	15	18	2	22	3	4	1	19	16	14	11	21	6	12	5	9	8	
	L = 64	97.57%	19	12	20	8	10	18	15	19	4	21	2	5	3	17	14	13	7	22	6	16	1	11	9	
	L = 128	96.39%	14	9	13	5	19	16	15	10	6	21	7	4	18	11	14	12	17	22	20	8	1	3	2	
L = 256	96.67%	21	18	15	6	20	14	13	12	8	22	7	4	21	9	17	11	19	10	16	5	1	2	3		
N = 16x16	L = 4	98.30%	18	11	21	10	9	12	14	22	4	20	1	2	3	19	18	15	8	13	16	7	17	6	5	
	L = 8	98.79%	19	12	20	11	6	14	13	19	4	22	3	1	2	18	21	15	8	17	16	5	7	9	10	
	L = 16	97.58%	22	15	19	6	11	16	17	22	4	18	2	3	1	20	21	14	13	10	12	7	5	8	9	
	L = 32	97.82%	18	12	19	9	10	14	16	20	5	18	2	3	1	21	22	15	13	17	6	11	4	7	8	
	L = 64	97.82%	17	11	21	1	12	16	14	19	4	17	2	5	3	18	15	8	13	22	7	20	6	9	10	
	L = 128	97.82%	17	14	19	1	16	12	10	17	3	8	2	7	9	11	18	15	13	22	21	20	6	4	5	
L = 256	98.30%	21	19	17	1	20	10	8	18	7	9	2	6	16	11	21	13	14	12	15	22	3	4	5		
N = 32x32	L = 4	97.32%	22	14	19	8	7	15	9	17	4	22	2	1	3	21	18	12	10	13	16	11	20	5	6	
	L = 8	98.07%	21	13	20	7	9	11	12	19	6	22	2	3	1	21	18	14	15	17	16	10	8	5	4	
	L = 16	97.60%	21	12	22	5	8	11	13	20	4	18	1	2	3	21	19	15	10	16	17	14	9	6	7	
	L = 32	98.07%	22	15	21	4	10	13	17	18	6	20	2	1	3	22	19	14	11	16	12	9	5	7	8	
	L = 64	97.83%	22	14	22	1	5	13	17	20	9	21	3	4	2	19	18	8	10	15	11	16	12	7	6	
	L = 128	98.29%	16	13	20	1	14	12	16	17	5	10	4	2	6	9	18	11	15	22	21	19	3	8	7	
L = 256	98.30%	20	19	13	1	20	11	14	12	4	8	2	3	10	9	22	17	21	16	15	18	7	6	5		
N = 64x64	L = 4	86.98%	21	10	21	3	13	9	11	20	4	22	1	2	7	19	18	12	15	8	17	14	16	6	5	
	L = 8	91.63%	22	11	22	5	15	8	13	20	6	21	1	2	7	19	17	12	16	14	18	9	10	3	4	
	L = 16	89.05%	20	9	21	3	12	6	11	22	7	19	2	1	8	17	18	14	10	15	20	16	13	5	4	
	L = 32	88.83%	20	10	21	3	11	8	18	20	4	22	1	2	7	17	19	12	15	16	14	13	9	5	6	
	L = 64	91.85%	21	11	20	1	10	4	14	22	5	19	3	2	6	18	16	7	17	15	13	12	21	8	9	
	L = 128	92.04%	14	15	22	1	10	8	16	19	12	6	2	5	9	7	14	11	18	21	20	13	17	3	4	
L = 256	91.36%	22	15	19	1	7	12	22	17	13	2	5	3	16	4	14	8	21	18	9	6	20	11	10		
No. of times used in classification:				26	27	24	23	21	23	22	24	24	26	25	27	27	27	28	28	28	25	27	27	24	25	27
No. of times ranked among top 9:				2	0	22	11	6	2	0	26	5	28	28	23	5	0	4	5	1	6	9	18	26	25	27

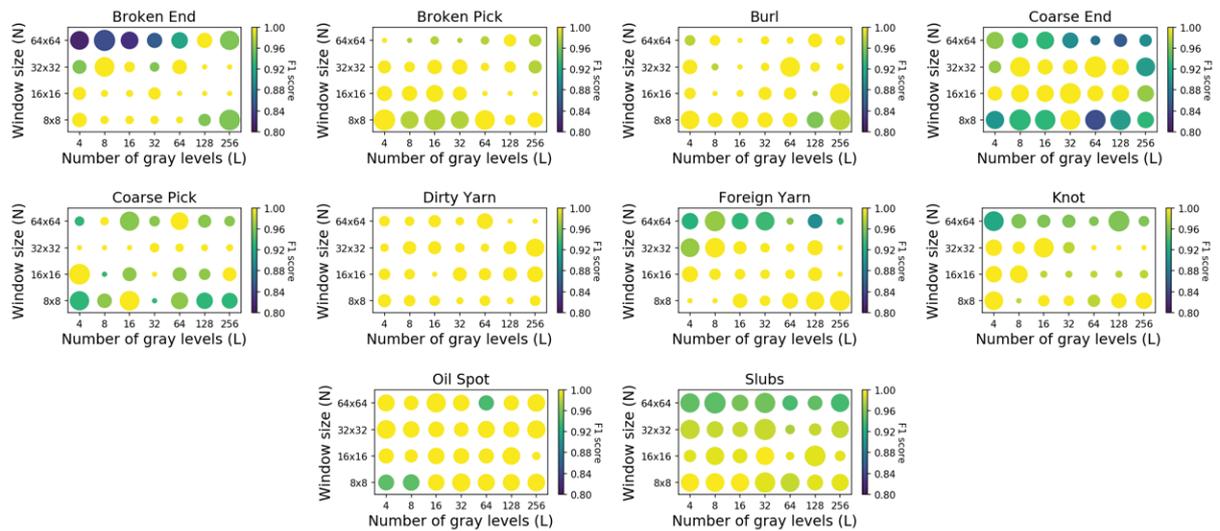


Figure 4 - The Number of Features Required for Detecting each Defect Type with the Maximum F1 Score at each $\langle L, N \rangle$ Combination. The Number of Features is Indicated by the Size of the Circular Markers and the F1 Score is Indicated by the Colour of the Circular Markers

used for the corresponding classification, along with their ranks, are given in Table 2. It can be seen that the highest F1 score (98.79%) is achieved when using a 16x16 window on images with 8 gray levels. However, this F1 score is not significantly different from the

maximum F1 scores achieved at other $\langle L, N \rangle$ combinations (which have F1 scores >96%) excluding those obtained with the largest window size. It should be noted, though, that to achieve this F1 score for some $\langle L, N \rangle$ combinations, almost all features had



to be used. The combination that uses the least number of features (14) to produce an acceptable accuracy is 128 gray levels and a 8×8 window. The F1 score of the classification in this case is 96.39%, which is only 2.4% less than the maximum.

It can also be seen from Figure 3 that using 9 features in the classification will increase the F1 score above 90% for any $\langle L, N \rangle$ combination, excluding the $N = 64 \times 64$ case. Therefore, it can be inferred that at least 9 feature computations are required to reliably

Table 4 - Comparison of Our Approach with Similar Studies

Research paper	Defect types	Method	Classifier	Performance		Dataset size
				Accuracy	F1 Score	
Our approach	10	GLCM-based feature extraction	k-NN	0.9877	0.9879	1426
[13]	4	Computation of gray-level statistics	Naïve Bayes	0.9962	-	128
[14]	NM	Extraction of visual saliency map features	Support vector machines	0.98	-	1140
[15]	89	Yarn tracking with UNet convolutional neural network	Minimum Covariance Determinant (MCD)	0.96	-	1431
[16]	5	Haar-like feature extraction	Cascade classifier	-	0.96	NM
[17]	7	Thresholding-based segmentation	Classified based on detected contour	0.95	-	NM

achieve a F1 score greater than 90%, given that the window size is relatively small. From Table 2, judging by the features that are consistently ranked among the top 9 features for a majority of $\langle L, N \rangle$ combinations, it can be inferred that the features P10, P11, P21, P8, P22, P12, P3, P20 and P4 are most likely to improve the classification F1 score for any $\langle L, N \rangle$ combination. It is also clear that the two features P10 and P11 are crucial in raising classification accuracy, as they have been ranked among the top 9 features for all $\langle L, N \rangle$ combinations. A set of 9 features is much less in number than the 17 features that another study [9] has proposed. The reduction in the number of features might also be a consequence of the near-regularity of woven fabric texture compared to the natural, irregular textures of ultrasound images used in the aforementioned study. The performance comparison of several commonly-used classifiers with the k-NN are given in Table 3 showing the superior performance of the k-NN for our feature set, which is the reason for its choice in this analysis. All classifiers had been optimized by performing a grid search and FFCV.

3.2. Analysis of Results for Each Defect

This subsection compares the results of the classification on each defect type as shown in Figure 4 with inferences made in the previous subsection.

It is immediately apparent that the Dirty Yarn defect is classified with a very high F1 score for all window sizes and all quantization levels. Judging by the sample given in Figure 5, the defect is distinct enough to not merge with the surrounding defect-free region during the gray-level down sampling and spans over the fabric in a tapering manner such that the textural information it produces can be captured by both large and small windows. These properties are not prominent in the other defects shown in the same figure.

Table 3 - Performance of Different Classifiers on the Optimal Feature Set Extracted at L= 8 and N=16X16

Classifier	Accuracy	F1 score
k-NN	98.77%	98.79%
Support Vector Machine	96.81%	96.84%
Multi-layer Perceptron	93.87%	93.86%
Random Forest	84.56%	84.21%
Naïve Bayes	73.28%	65.83%

Even though the 64×64 window generally produces lower classification F1 scores, it can be seen that the Dirty Yarn defect is perfectly classified for all quantization levels, whereas Oil Spot and Burl are perfectly classified for six of the seven quantization levels when using this window size. Observation of samples of these defects as presented in Figure 5 shows that these defects are all “global” defects i.e. ones that occupy a considerable area of the image. The issue of large windows

not significantly capturing the texture distortion created by small defects does not apply for such defects.

Regarding the number of features required for optimal F1 score, it can be inferred that Oil Spot, Slubs and Coarse End require many features to quantify the texture information they produce at any $\langle L, N \rangle$ combination, whereas all other defects can be classified accurately with fewer features.

It is also clear from Figure 4 why the overall classification F1 score was highest at $N = 16$ and $L = 8$, as all defects except for Broken Pick and Coarse End, have been classified with near-perfect F1 scores. However, it should be noted that, even though the least number of features for an appreciable F1 score for the overall classification is obtained at $N = 8 \times 8$ and $L = 128$, the least number of features required to achieve maximum F1 score for each defect varies considerably.

A comparison of the best performance achieved in this study for the full dataset is compared with the results of similar studies in Table 4.

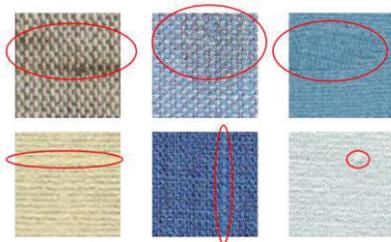


Figure 5 - From Left to Right: Dirty Yarn, Oil Spot, Burl, Slub, Coarse End and Knot Defects from the Dataset with the Defective area Circled

4. Conclusions

The aim of this research was to analyze the effect of the window size and the number of gray levels on varying sets of GLCM features.

The results of the experimental study show that a set of 9 features is sufficient to reach an F1 score exceeding 90%. Out of these 9 features, Homogeneity I and Homogeneity II have contributed to improving the classification accuracy regardless of the size of the window or its number of gray levels. The combination that produced an acceptable F1 score (96.39%) using the least number (14) of features was 128 gray levels and a 8×8 window. Another outcome of the study is the

result that the use of a greater number of gray levels will still produce reasonably high F1 scores if a greater number of features are used in the classification. Furthermore, the fact that large windows impact the classification accuracy negatively has been confirmed. This work has shown that the requirements for classifying near-regular textures with the GLCM differ from those for classifying irregular textures, mainly in that fewer features can be used to achieve acceptable classification accuracies.

Acknowledgement

This research was supported by the Accelerating Higher Education Expansion and Development (AHEAD) Operation of the Ministry of Higher Education funded by the World Bank.

References

1. Haralick, R. M., Shanmugam, K., and Dinstein, I., "Textural Features for Image Classification," *IEEE Trans. Syst., Man, and Cybern.*, Vol. SMC-3, No. 6, pp. 610-621, 1973.
2. Bodnarova, A., Williams, J. A., Bennamoun, M., and Kubik, K. K., "Optimal Textural Features for Flaw Detection in Textile Materials," *IEEE Region 10 Ann. Conf. Speech and Image Technol. for Comput. and Telecommun.*, 1997, pp. 307-310.
3. Cuenca, S. A., and Camara, A., "New Texture Descriptor for High-Speed Web Inspection Applications," in *Proc. 2003 Int. Conf. Image Process.*, 2003, pp. 537-540.
4. Baraldi, A., and Parmiggiani, F., "An Investigation of the Textural Characteristics Associated with Gray Level Cooccurrence Matrix Statistical Parameters," in *IEEE Trans. Geosci. Remote Sens.*, Vol. 33, No. 2, pp. 293-304, 1995.
5. Soh, L., and Tsatsoulis, C., "Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-occurrence Matrices," *IEEE Trans. Geosci. Remote Sens.*, Vol. 37, No. 2, pp. 780-795, March 1999.
6. Karthikeyan, S., and Rengarajan, N., "Performance Analysis of Gray Level Cooccurrence Matrix Texture Features for Glaucoma Diagnosis," *Amer. J. Appl. Sci.*, Vol. 11, No. 2, pp. 248-257, 2013.
7. Clausi, D., "An Analysis of Co-occurrence Texture Statistics as a Function of Grey Level



- Quantization," *Can. J. Remote Sens.*, Vol. 28, No. 1, pp. 45-62, 2002.
8. Patel, M. B., Rodriguez, J. J., and Gmitro, A. F., "Effect of Gray-Level Re-quantization on Co-occurrence Based Texture Analysis," 2008 15th *IEEE Int. Conf. Image Process.*, 2008, pp. 585-588.
 9. Gomez, W., Pereira, W. C. A., and Infantosi, A. F. C., "Analysis of Co-Occurrence Texture Statistics as a Function of Gray-Level Quantization for Classifying Breast Ultrasound," in *IEEE Trans. Med. Imag.*, Vol. 31, No. 10, pp. 1889-1899, Oct. 2012.
 10. Suto, J., Oniga, S., and Sitar, P. P., "Comparison of Wrapper and Filter Feature Selection Algorithms on Human Activity Recognition," 2016 6th *Int. Conf. Comput. Commun. and Control*, 2016, pp. 124-129.
 11. Brown, G., Pocock, A., Zhao, M.-J., and Luján, M., "Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection," *J. Mach. Learn. Res.*, Vol. 13, pp. 27-66, 2012.
 12. Raheja, J. L., Ajay, B., and Chaudhary, A., "Real Time Fabric Defect Detection System on an Embedded DSP Platform," *Optik*, Vol. 124, No. 21, pp. 5280-5284, 2013.
 13. Habib, M. T., Shuvo, S. B., Uddin, M. S., and Ahmed, F., "Automated Textile Defect Classification by Bayesian Classifier Based on Statistical Features," in 2016 *Int. Workshop Comput. Intelligence (IWCI)*, 2016.
 14. Zhang, H., Hu, J., and He, Z., "Fabric Defect Detection Based on Visual Saliency Map and SVM," in 2017 2nd *IEEE Conference on Computational Intell. and Appl.*, 2017.
 15. Weninger, L., Kopaczka, M., and Merhof, D., "Defect Detection in Plain Weave Fabrics by Yarn Tracking and Fully Convolutional Networks," in 2018 *IEEE Int. Instrum. and Meas. Technol. Conf.*, 2018.
 16. Wang, Y., Liqing, L., Wan, X., and Wang, J., "Woven Fabric Defect Detection Based on the Cascade Classifier," in 2019 12th *Int. Congr. Image and Signal Process., Biomed. Eng. and Inform.*, 2019.
 17. Vladimir, G., Evgen, I., and Aung, N. L., "Automatic Detection and Classification of Weaving Defects Base on Digital Image Processing," in 2019 *IEEE Conf. of Russian Young Researchers in Elect. and Electron. Eng.*, 2019.
 18. Zhou, J., Pan, R., and Gao, W., "Automated Woven Fabric Texture Periodicity Extraction by Spectral Analysis and Patch-DMF," *J. Textile Inst.*, Vol. 112, pp. 991-998, 2021.
 19. Sabeenian, R. S., and Paramasivam, M. E., "Defect Detection and Identification in Textile Fabrics Using Multi Resolution Combined Statistical and Spatial Frequency Method," 2nd *Int. Adv. Comput. Conf.*, Patiala, 2010, pp. 162-166.
 20. *Standard Fabric Defect Glossary*, Cotton Incorporated, Jan. 2019. [Online]. Available: <https://www.cottoninc.com/quality-products/textile-resources/fabric-defect-glossary>.