Projected Texture for Object Classification

Avinash Sharma and Anoop Namboodiri

Center for Visual Information Technology,
International Institute of Information Technology,
Hyderabad, INDIA - 500 032

Abstract. Algorithms for classification of 3D objects either recover the depth information lost during imaging using multiple images, structured lighting, image cues, etc. or work directly the images for classification. While the latter class of algorithms are more efficient and robust in comparison, they are less accurate due to the lack of depth information. We propose the use of structured lighting patterns projected on the object, which gets deformed according to the shape of the object. Since our goal is object classification and not shape recovery, we characterize the deformations using simple texture measures, thus avoiding the error prone and computationally expensive step of depth recovery. Moreover, since the deformations encode depth variations of the object, the 3D shape information is implicitly used for classification. We show that the information thus derived can significantly improve the accuracy of object classification algorithms, and derive the theoretical limits on height variations that can be captured by a particular projector-camera setup. A 3D texture classification algorithm derived from the proposed approach achieves a ten-fold reduction in error rate on a dataset of 30 classes, when compared to state-of-the-art image based approaches. We also demonstrate the effectiveness of the approach for a hand geometry based authentication system, which achieves a four-fold reduction in the equal error rate on a dataset containing 149 users.

1 Introduction

Three dimensional object are characterized by their shape, which can be thought of as the variation in depth over the object, from a particular view point. These variations could be deterministic as in the case of rigid objects or stochastic for surfaces containing a 3D texture. The depth information are lost during the process of imaging and what remains is the intensity variations that are induced by the object shape and lighting, as well as focus variations. Algorithms that utilize 3D object shape for classification tries to recover the lost depth information from the intensity or focus variations or using additional cues from multiple images, structured lighting, etc. This process is computationally intensive and error prone. Once the depth information is estimated, one needs to characterize the object using shape descriptors for the purpose of classification.

Image-based classification algorithms tries to characterize the intensity variations of the image of the object for recognition. As we noted, the intensity

variations are affected by the illumination and pose of the object. The attempt of such algorithms is to derive descriptors that are invariant to the changes in lighting and pose. Although image based classification algorithms are more efficient and robust, their classification power is limited due to loss of information during the imaging process.

We propose the use of structured lighting patterns for the purpose of recognition without shape recovery. The depth variations of the object induces deformations in the projected patterns, and these deformations encode the shape information. We treat the deformed patterns as a texture; referred to as *projected texture*. The primary idea is to view the projected texture as a characteristic property of the object and use it directly for classification instead of trying to recover the shape explicitly. To achieve this we need to use an appropriate projection pattern and derive features that sufficiently characterize the deformations. The patterns required could be quite different depending on the nature of object shape and its variation across objects.

In this paper, we primarily concentrate on the problem of 3D texture classification. We propose a set of simple texture features that can capture the deformations in projected lines on 3D textured surfaces. Experiments indicate the superiority of the approach as compared to traditional image based classification algorithms. To demonstrate the flexibility of the idea, we also show the use of projected textures for hand geometry based person authentication.

Figure 1 shows the effect of projected texture on two similar texture classes: salt and sugar crystals. The projected texture based features are clearly different, while the image based features look similar. One should note that an approach using structured lighting has it limitations also as it requires some amount of control of the environment. However, it can be useful in a variety of applications such as industrial inspection, robot navigation, biometric authentication, supermarket billing, etc.
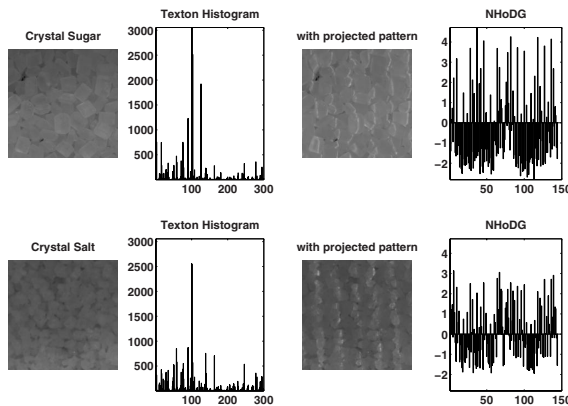
Fig. 1. Salt and Sugar crystal with and without projected texture and the corresponding feature representations

The use of Bidirectional Texture Function (BTF), which incorporates the variations in illumination and statistical texture variations, is popular in appearance based 3D texture models. Leung and Malik [1] extended the idea of modeling the appearance based on texture primitives (textons) using BTFs to define a set of 3D textons. Cula and Dana [2] modified the approach to by building dictionaries directly from filter output, making approach less sensitive to illumination and pose. Wang and Dana [3] extended the approach to incorporate geometric information computed from sampled BTF to make the representation suitable for tasks like texture prediction, synthesis, etc. Although the above algorithms work on 2D image features, their definitions are based on lighting variations in 3D. Varma and Zisserman[4] proposed image level features that are invariant of illumination and pose. They further extended the idea of textons by creating a dictionary from the most responsive filters for an image [5], as well as based on image patch exemplars [5]. Currently, these approaches are two of the best performing classifiers for texture images, and we use them as benchmarks for comparison. However, these approaches are computationally intensive for both training and testing. We show that a relatively simple texture measure that we propose is sufficient to achieve better performance, when combined with projected texture.
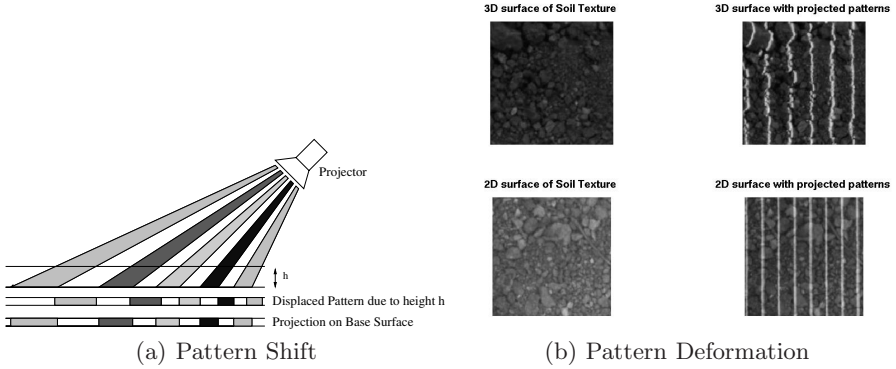
A different class of approaches use natural texture in the scene for recognition of objects or people [6,7,8] as well as for depth estimation [9,10]. The primary difference in our approach is that the texture we use is not an inherent property of the object, but superimposed on it during imaging. We demonstrate the flexibility of our approach with a second application in hand geometry based person authentication, where one is required to capture minor variations between similar samples (hands) belonging to different people. The performance is compared with popular image based features [11,12].

2   Projected Texture for Recognition

The primary idea of the approach, as described before, is to encode the depth variations of an object as deformations of a projected light pattern. There are primarily two categories of objects that we might want to characterize. The first class of objects, such as manufactured parts and human palm, are characterized by their exact 3D shape, while the second class of objects are characterized by the stochastic variations in depth such as 3D textured surfaces. In this paper, we primarily concentrate on classification of 3D textured surfaces, and the results of hand geometry based authentication is presented briefly.

The object is placed in the field of view of the camera and the projector, and a specific light pattern is projected on it. The projected pattern, or the original texture, falling on the surface containing the object, gets transformed according to the depth map of the object under illumination. These transformations can be primarily classified into two categories:

 – *Pattern Shift*: The position where a particular projected pattern is imaged by the camera depends on the absolute height from which the pattern in

3D surface of Soil Texture   3D surface with projected patterns

2D surface of Soil Texture   2D surface with projected patterns

Projector

h

Displaced Pattern due to height h

Projection on Base Surface

(a) Pattern Shift   (b) Pattern Deformation

Fig. 2. Pattern shift and deformation due to depth variations

reflected. Figure 2 illustrates this with a cross section of a projection setup. Note that the amount of shift depends on the height difference between the objects as well as the angle between the projector and camera axes.

– *Pattern Deformation*: Any pattern that is projected on an uneven surface gets deformed in the captured image depending on the change in depth of the surface (see Figure 2). These deformations depend on the absolute angle between the projector axis and the normal to the surface at a point as well as its derivative.

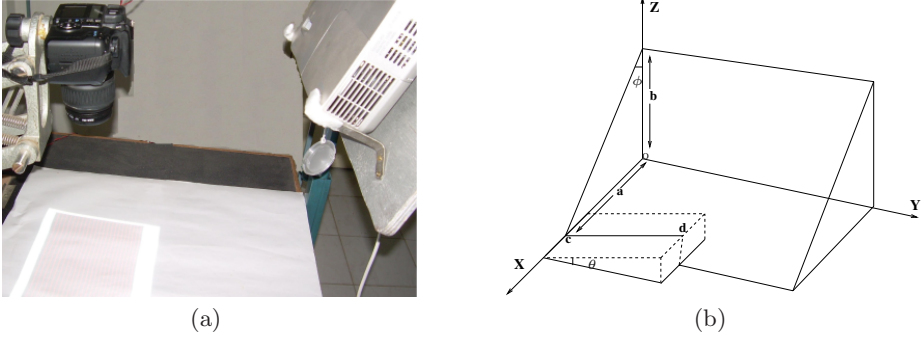## 2.1   Pattern Deformation and Projector Camera Configuration

We now take a closer look at the nature of depth variation in objects surface and how it affects projected patterns for a specific set of setup parameters. One of the important factor affecting deformation is the slope of the surface with respect to the projection axis. We first derive the relationship between the deformation in pattern to various parameters of physical setup and the height variations on object surface. Figure 3(a) shows the image capture setup and Figure 3(b) shows a schematic diagram with the object surface having slope θ to the Y-axis. We refer to this as the *object plane*. Figure 3(b), considers the projection of a single horizontal line pattern at an angle ϕ from Z-axis forming a plane that we will call the *light plane*. The light plane may be represented as $\frac{x}{a} + \frac{z}{b} = 1$, where $a = b \tan \phi$. The equation of the light plane and the object plane can hence be expressed as:

$$x \cot \phi + z - b = 0, \text{ and} \tag{1}$$

$$z - y \tan \theta = 0 \tag{2}$$

The line cd as shown in figure is the intersection of both of these planes in 3D, and can be expressed by cross product of the normals of both intersecting planes. The direction vector of cd is: $\boldsymbol{n_3} = [\cot \phi \ \ 0 \ \ 1]^T \times [0 \ \ \tan \theta \ \ -1]^T$ or,

$$\boldsymbol{n_3} = [-\tan \theta \ \ \ \cot \phi \ \ \ \tan \theta \cot \phi]^T \tag{3}$$

(a)   (b)

Fig. 3. The image capture setup and the pattern deformation geometry

One point common to both plane say p can be obtain by solving equation 1 and 2 as : $p = [\, b \tan \phi \quad 0 \quad 0\,]^T$. So equation of 3D line can be written as

$$r = [\, b \tan \phi - s \tan \theta \quad s \cot \phi \quad s \tan \theta \cot \phi\,]^T, \tag{4}$$

where s is line parameter, different value of s will give different points on line.

In order to express 2D projection of 3D line onto image plane of camera, we need to take two points on 3D line such that they are in FOV of camera. Let Q_1 and Q_2 be two such points corresponding value of s as $s = l_1$ and $s = l_2$ respectively.

$$Q_1 = [\, b \tan \phi - l_1 \tan \theta \quad l_1 \cot \phi \quad l_1 \tan \theta \cot \phi\,]^T \tag{5}$$

$$Q_2 = [\, b \tan \phi - l_2 \tan \theta \quad l_2 \cot \phi \quad l_2 \tan \theta \cot \phi\,]^T \tag{6}$$

For simplifying the things let us assume camera to be pinhole camera with camera matrix $P = K[R|t]$. Let $K = I$ i.e., the internal parameter matrix is unity matrix and R an t be

$$R = \begin{bmatrix} R_1 & R_2 & R_3 \\ R_4 & R_5 & R_6 \\ R_7 & R_8 & R_9 \end{bmatrix}, t = \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}^T$$

The image of these points in camera plane be $q_1 = PQ_1$ and $q_2 = PQ_2$. q_1 can be represented in matrix form in terms of R_1 to R_9, l_1 and ϕ, θ as:

$$q_1 = \begin{bmatrix} R_1(b\tan\phi - l_1\tan\theta) + R_2 l_1 \cot\phi + R_3 l_1 \tan\theta \cot\phi + t_1 \\ R_4(b\tan\phi - l_1\tan\theta) + R_5 l_1 \cot\phi + R_6 l_1 \tan\theta \cot\phi + t_2 \\ R_7(b\tan\phi - l_1\tan\theta) + R_8 l_1 \cot\phi + R_9 l_1 \tan\theta \cot\phi + t_3 \end{bmatrix} \tag{7}$$

Similarly q_2 can be represented in terms of R_1 to R_9, l_2 and ϕ, θ. Let us write q_1 and q_2 as:

$$q_1 = \begin{bmatrix} X_1 & Y_1 & Z_1 \end{bmatrix}^T q_2 = \begin{bmatrix} X_2 & Y_2 & Z_2 \end{bmatrix}^T \tag{8}$$

In the homogeneous coordinate system q_1 and q_2 can be represented as:

$$\overline{q_1} = \left[\begin{array}{cc} \frac{X_1}{Z_1} & \frac{Y_1}{Z_1} \end{array} \right]^T \overline{q_2} = \left[\begin{array}{cc} \frac{X_2}{Z_2} & \frac{Y_2}{Z_2} \end{array} \right]^T \tag{9}$$

Thus equation of line in 2D image plane is $\boldsymbol{L} : \overline{q_1} \times \overline{q_2} = 0$. i.e.,

$$\boldsymbol{L} \ : \ X(Z_1Y_2 - Z_2Y_1) - Y(Z_1X_2 - Z_2X_1) - X_1Y_2 + X_2Y_1 = 0 \tag{10}$$

$$m = (Z_1Y_2 - Z_2Y_1)/(Z_1X_2 - Z_2X_1) \tag{11}$$

From equation of line it can inferred that slope m of this line will depend upon b, ϕ and θ thus slope of height variation directly affects orientation of projection of 3D line onto image plane subject to setup specific setup parameters as shown before.

Hence, we can compute the projection angle given the minimum angle in deformation that can be detected by the camera and the slope variation of the surface. One other factor is the shadow effect if slope is in opposite direction of illumination. In that case response of any transform will be zero or low. Internal reflection of the surface is an important factor which depends on physical property of object surface. Thus all these factor combine to form a deformation pattern which we have used to recognize the surface.

2.2   Design of Projected Pattern

The choice of an appropriate projection pattern is important due to a variety of factors:

1. For the deformation to be visible at any point in the captured image, the gradient of the projected pattern should not be zero in the direction of gradient of the object depth.
2. One should be able to capture the deformations of the projected pattern using the texture measure employed for this purpose.
3. The density of the projected pattern or its spatial frequency should be related to the frequency of height variations to be captured. Hence, analyzing the geometry of an object with a high level of detail will require a finer pattern, whereas in the case of an object with smooth structural variations, a sparse one will serve the purpose.
4. Factors such as the color, and reflectance of the object surface should be considered in selecting the color, intensity and contrast of the projected pattern.

For the purpose of 3D texture recognition, we use a set of parallel lines with regular spacing, where the spacing is determined based on the scale of the textures to be recognized. For hand geometry based authentication, we have selected a repetitive star pattern that has gradients in four different directions. The width of the lines and the density of patterns in the texture were selected experimentally so that it captures the height variations between the palms at the angle of projection selected.

2.3   Characterization of Pattern Deformation

An effective method for characterization of the deformations of the projected pattern is critical for its ability to discriminate between different objects. We propose a set of texture features the capture the statistics of deformation in the case of 3D textures. As we noted before, the projection pattern used for 3D texture classification was a set of parallel lines. Hence the feature set that we propose should capture the deformations in the lines and compute the overall statistics.

Normalized Histogram of Derivative of Gradients (NHoDG). Gradient directions in images are the directions of maximal intensity variation. In our scenario, the gradient directions can indicate the direction of the projected lines. As the lines get deformed with surface height variations, we compute the differential of the gradient directions in both x and y axes to measure the rate at which the surface height varies. The derivatives of gradients are computed at each pixel in the image, and the texture is characterized by a Histogram of the Derivatives of Gradients (HoDG). The gradient derivative histogram is a good indicator of the nature of surface undulations in a 3D texture. For classification, we treat the histogram as a feature vector to compare two 3D textures. As the distance computation involves comparing corresponding bins from different images, we normalize the counts in each bin of the histogram across all the samples in the training set. This normalization allows us to treat the distance between corresponding bins in the histograms, equally (employ the Euclidean distance). The NHoDG is a simple but extremely effective feature for discriminating between different texture classes. Figure 4 illustrates the computation of the NHoDG feature from a simple image with a bell shaped intensity variation.

We compare the effectiveness of this feature set under structured illumination in the experimental section using a dataset of 30 3D textures.
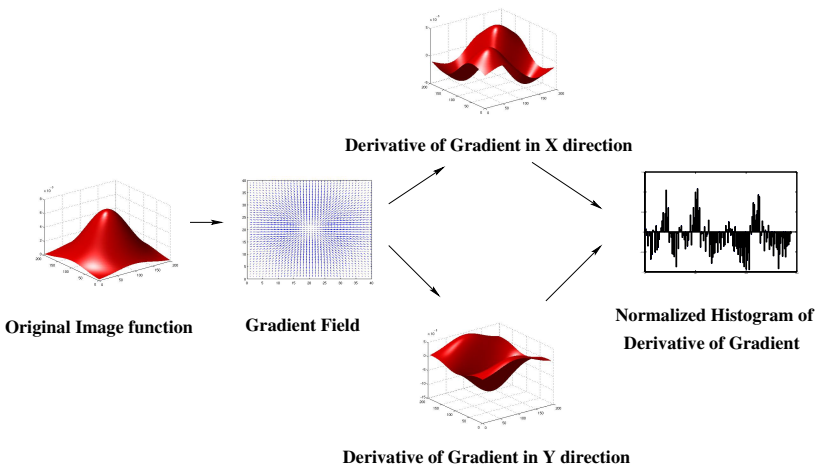

Derivative of Gradient in X direction

Original Image function   **Gradient Field**   **Normalized Histogram of Derivative of Gradient**

Derivative of Gradient in Y direction

Fig. 4. Computation of the proposed NHoDG feature vector

Characterizing Deterministic Surfaces. In the case of hand geometry based authentication, we need to characterize the exact shape of the object, and not the statistics of height variations. Hence we divide the hand image into a set of non-overlapping sub-windows, and compute the local textural characteristics of each window using a filter bank of 24 Gabor filters with 8 orientations and 3 scales (or frequencies). In our experiments we have used a grid of 64 sub-windows (8×8), and the mean response of each filter forms a 1536 dimensional feature vector, that is used to represent each sample.

3 Experimental Results and Analysis

The image capture setup consists of a planar surface to place the object samples, an LCD projector fixed at an angle to the object surface. The camera is located directly above the object with its axis perpendicular to the object plane (see Figure 3(a)). We considered a set of 30 3D textures with considerable variations in depth profile. The texture surfaces included *pebbles, concrete, thermocol, sand, soil, gravel, sponge, ribbed paper, crystal sugar and salt*, and a variety of *grains* and *pulses*. The materials were chosen to have texture classes with similar scale and nature of depth variations, which makes the problem challenging. However, the overall scales varied considerably from pebbles to sand. A total $14,400$ images were collected, with 480 samples for each of the 30 classes. The 480 samples consisted of 24 different object samples, each taken with 5 different projected patterns (including no projection pattern case) under 4 different illumination conditions. The projected patterns are parallel lines having uniform spacing of $5, 10, 15, 20$ pixels between them. We refer to these patterns as W5, W10, W15 and W20 in the rest of this section. The overall object surface was nearly planar, which is normal to the camera axis. Sample images of the 30 different classes along with their feature space representations are shown in Figure 5. We refer to this dataset as *Projected Texture Dataset* or *PTD* from now on.

(a)   (b)
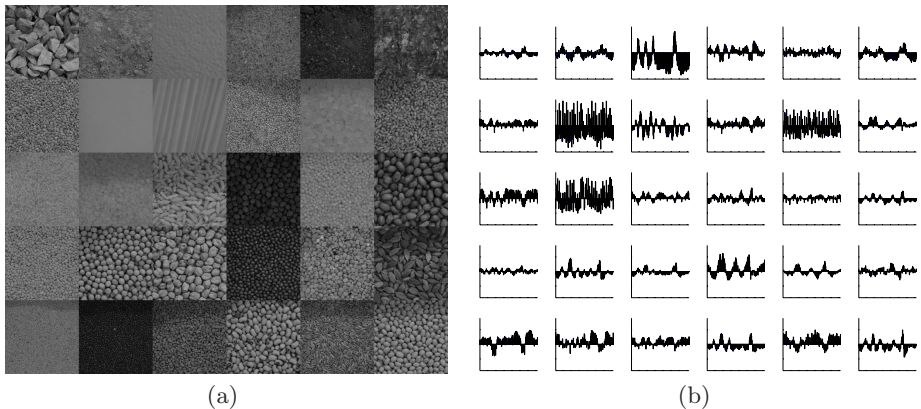
Fig. 5. Examples of the 30 texture classes and their NHoDG representations

For evaluation of the hand geometry based person authentication algorithm, we collected a dataset of 1341 images from 149 users, each user providing 9 samples each. For comparison, we collected two sets of images from each user, with projected texture as well as with uniform illumination.

3.1 Texture Classification

We have conducted exhaustive experiments to validate our approach. Our contribution includes the use of deformed projection patterns as texture as well as proposing an appropriate feature set to capture the deformations. We conducted experiments with and without projection patterns, using the proposed and traditional 2D features.

As mentioned before, we have used maximum filter response (MR) and image patch exemplar based features [5], as benchmarks for comparing our approach. We have included four additional filter responses with two higher scales (MR12 now instead of MR8) so as to improve the results of the MR approach, as our dataset contained higher scale variation. Patch-based texture representation with three different patch sizes were also used for comparison.

Table 1. Error rates of classification using NHoDG, MR, and Image Patch features on the PTD and Curet datasets (in %)

Dataset	Projection	NHoDG	MR	Image Patch		
				3x3	5x5	7x7
Curet	Without	12.93	**3.15**	4.67	4.38	3.81
PTD	Without	2.36	**1.18**	3.51	1.53	1.46
	With	1.15	**0.76**	1.60	1.18	0.90
	Combined	**0.07**	0.31	1.70	0.66	0.62

Table 1 gives the results using the proposed NHoDG feature set as well as the MR and patch based features. Results are presented on PTD as well as the Curet datasets. However, note that the results on the Curet dataset are without projected patterns. All the results are based on a 4-fold cross validation, where the dataset is divided into non-overlapping training and testing sets, which is repeated 4 times and the average results are reported. We note that the 2D image based approaches achieves an error rate of 1.18%, i.e., 34 misclassifications on our dataset of 2880 samples (with no projection pattern). Clearly the MR12 feature set performs better in pure image based classification. However, while combining the image information with the projected texture feature, the NHoDG feature achieves an error rate of 0.07%, which corresponds to just 2 samples being misclassified. We had experimented with the patch based approach also, which performed worse than the MR filter approach. The best accuracies for 3x3, 5x5 and 7x7 patches were 1.70%, 0.66%, and 0.62%, as opposed to 0.31% of MR12 filter.

Figure 6 shows the variation in classification performance as the histogram bin sizes and the pattern separation are varied. We note that the performance is consistently good, and we selected a bin resolution of 5 degrees and pattern
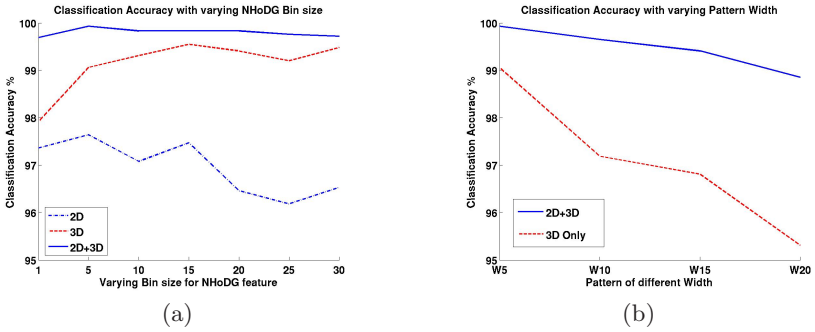
Fig. 6. Performance with varying histogram bin sizes and pattern separations
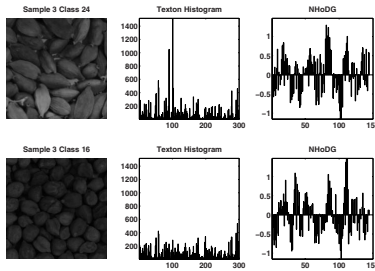

Fig. 7. One of the two misclassifications in the dataset along with the MR and NHoDG feature representation

separation of 5 pixels for the above experiments. Figure 7 shows one of the misclassified samples, and the corresponding NHoDG and MR features. We also note that the proposed feature set is primarily intended for use with projection and does not perform well on datasets such as Curet, without projected patterns.

3.2   Hand Geometry Based Authentication

We compare the performance of three different feature sets in this experiment: i) *Feat-1*: A set of 17 features based of finger lengths, widths and heights, proposed by Jain *et al.* [11], ii) *Feat-2*: A set of 10 features computed from palm contours proposed by Faundez *et al.* [12], and iii) *Feat-3*: The proposed projected texture based features.

Figure 8(a) shows the difference in deformation of the projected pattern based on the 3D shape of the palm. An effective method to compare the utility of a matching algorithm for authentication is the ROC curve, which plots the trade off between genuine acceptance and false acceptance of users in an authentication system. The ROC curves in Figure 8(b) clearly indicate the superiority of the proposed feature set. As the purpose of this experiment is to compare the feature sets, we have provided the ROC curve based on the Euclidean distance

| Original Image | Deformed Patterns | Zoomed View |
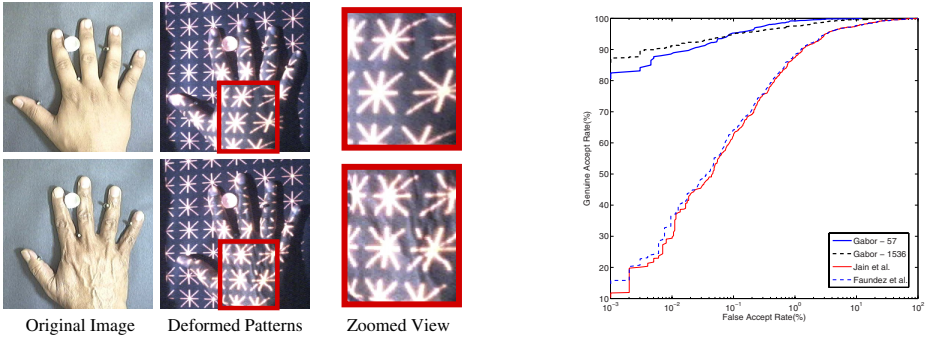
Fig. 8. Deformation in projected texture due to hand geometry, and the ROC curve for the different algorithms

between the samples of the same user as well as different users. The equal error rate (EER), or the point at which false reject rate equals false acceptance rate, for *Feat-1* and *Feat-2* were 4.06% and 4.03% respectively. In contrast, the proposed feature set achieved and EER of 1.91%. In addition to the equal error rate, we note that the genuine acceptance rate continues to be above 80%, even at false acceptance rates of 0.001% for the proposed features, while the performance of the 2D image based features degrade considerably at this point. We also conducted an experiment in feature selection to choose a subset of the 1536 features that would help us in reducing the computations required. We note that even with just 57 features out of 1536, the ROC curve is similar to that of the complete feature set. Moreover, the equal error rate improves to 0.84% with the reduced feature set. This is possible as the selection process avoids those sub-windows, where the intra-class variations in pose are high. Clearly the projected patterns induce a large amount of discriminating information into the computed features.

4   Conclusions and Future Work

A novel technique for recognition of 3D objects using projected texture is proposed. The results were demonstrated in the case of two different object classes, one for 3D texture classification, and the second for hand geometry based person authentication. The approach is robust to occlusions and noise as we need not find any correspondences or recover the depth map of the object. Moreover, the computational requirements are comparable to the simpler 2D image based recognition approaches, while being far more accurate.

We are currently working on extending the approach for arbitrary pose object recognition, and the initial results are promising. Future work in this direction could be to handle objects of high reflectance and transparency. Automatic adaptation of the projected pattern to a particular application could also be interesting. Temporally varying projection patterns, giving rise to dynamic deformations, could also give us a cue towards designing optimal classifiers for recognition.

References

1. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision 43(1), 29–44 (2001)
2. Cula, O.G., Dana, K.J.: 3d texture recognition using bidirectional feature histograms. International Journal of Computer Vision 59(1), 33–60 (2004)
3. Wang, J., Dana, K.J.: Hybrid textons: modeling surfaces with reflectance and geometry. In: Proc. of CVPR 2004, vol. 1, pp. 372–378 (2004)
4. Varma, M., Zisserman, A.: Classifying images of materials: Achieving viewpoint and illumination independence. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2352, pp. 255–271. Springer, Heidelberg (2002)
5. Varma, M.: Statistical Approaches To Texture Classification. PhD thesis, University of Oxford (October 2004)
6. Kumar, A., Zhang, D.: Personal recognition using hand shape and texture. IEEE Transactions on Image Processing 15(8), 2454–2461 (2006)
7. Daugman, J.: High confidence visual recognition of persons by a test of statistical independence. IEEE Transactions on PAMI 15(11), 1148–1161 (1993)
8. Jain, A.K., Prabhakar, S., Hong, L., Pankanti, S.: Filterbank-based fingerprint matching. IEEE Transactions on Image Processing 9(5), 846–859 (2000)
9. Forsyth, D.: Shape from texture and integrability. In: Proc. of ICCV 2001, vol. 2, pp. 447–452 (July 2001)
10. Loh, A., Hartley, R.: Shape from non-homogeneous, non-stationary, anisotropic, perspective texture. In: Proc. of the BMVC 2005, pp. 69–78 (2005)
11. Jain, A.K., Ross, A., Pankanti, S.: A prototype hand geometry-based verification system. In: Proc. of the AVBPA 1999, Washington D.C., pp. 166–171 (1999)
12. Faundez-Zanuy, M., Elizondo, D.A., Ferrer-Ballester, M., Travieso-González, C.M.: Authentication of individuals using hand geometry biometrics: A neural network approach. Neural Processing Letters 26(3), 201–216 (2007)