A Technical Report
On
Projected Texture for Object Recognition

Avinash Sharma
Center for Visual Information Technology
IIIT, Hyderabad 500032 INDIA
avinash_s@research.iiit.ac.in

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. These depth variations are lost during the process of imaging and what remains is the intensity variations that are induced by the shape and lighting, as well as focus variations. Algorithms that utilize 3D shape for classification tries to recover the lost 3D 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 as the 3D information is lost during the imaging process.

We propose the use of structured lighting patterns, which we refer to as *projected texture*, for the purpose of object recognition. The depth variations of the object induces deformations in the projected texture, and these deformations encode the shape information. The primary idea is to view the deformation pattern 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.

For the problem of 3D texture classification, where the depth variation in surface is stochastic, we proposed a set of simple texture features that can capture the deformations in projected lines on 3D textured surfaces.

we also proposed a solution to category level recognition of rigid objects, where the nature of these variations is deterministic. Different depth profiles can result for a particular shape depending on the

view. Here we propose a set of simple position and pose invariant features for characterizing the deformations based on the popular bag-of-words paradigm for object representation.

We also proposed a feature for position sensitive object recognition and demonstrated it on projected textures for hand geometry based person authentication.

Experiments indicate the superiority of the approach as compared to traditional image based algorithms.

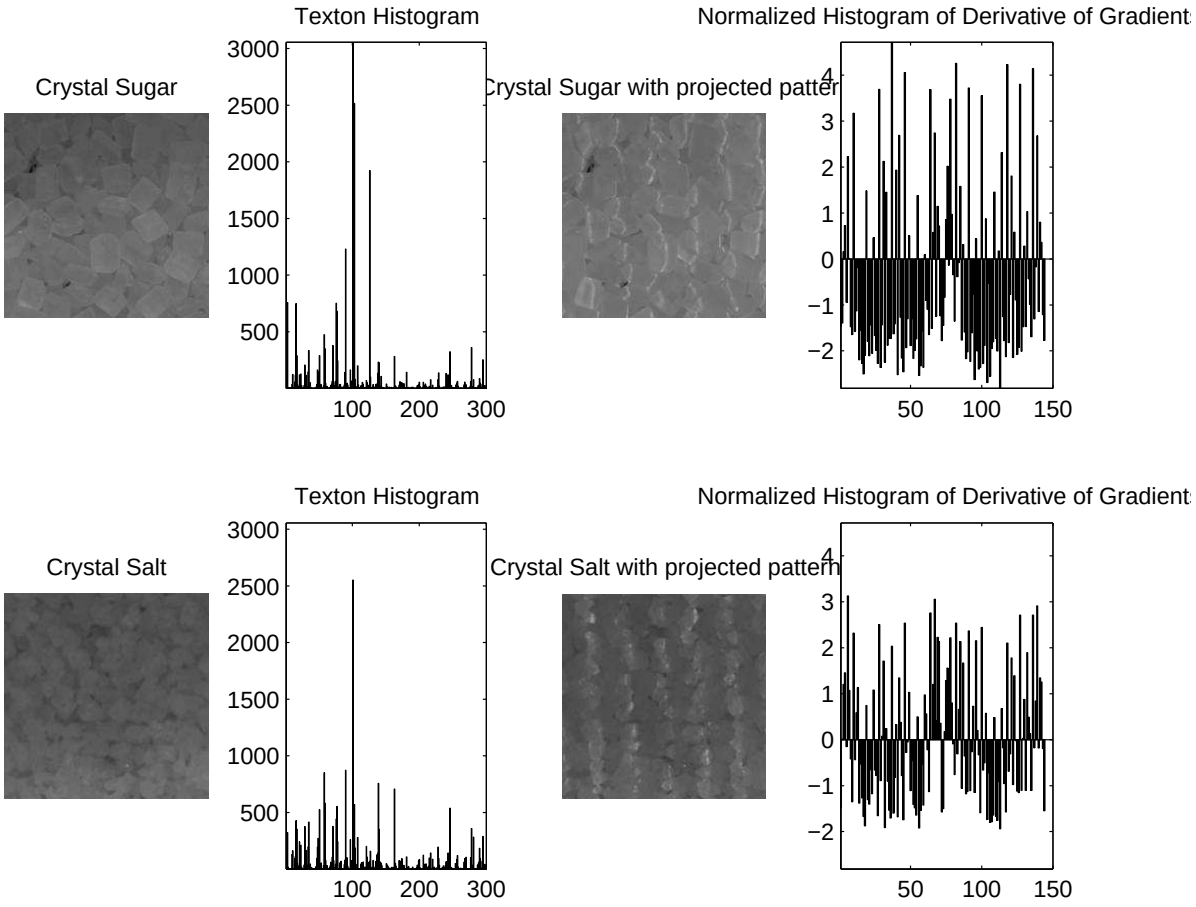


Figure 1: Salt and Sugar crystal with and without projected texture and the corresponding feature representations.

Figure 1 shows an example to two similar texture classes (salt and sugar crystals), under regular illumination and with projected texture, along with the computed feature vectors. One can clearly see the difference in feature representation with projected texture, while the image based feature sets 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 applied in a variety of applications such as industrial inspection, robot navigation, biometric authentication, etc. Figure 2 shows two similar views of different object, which are confusing in 2D appearance.

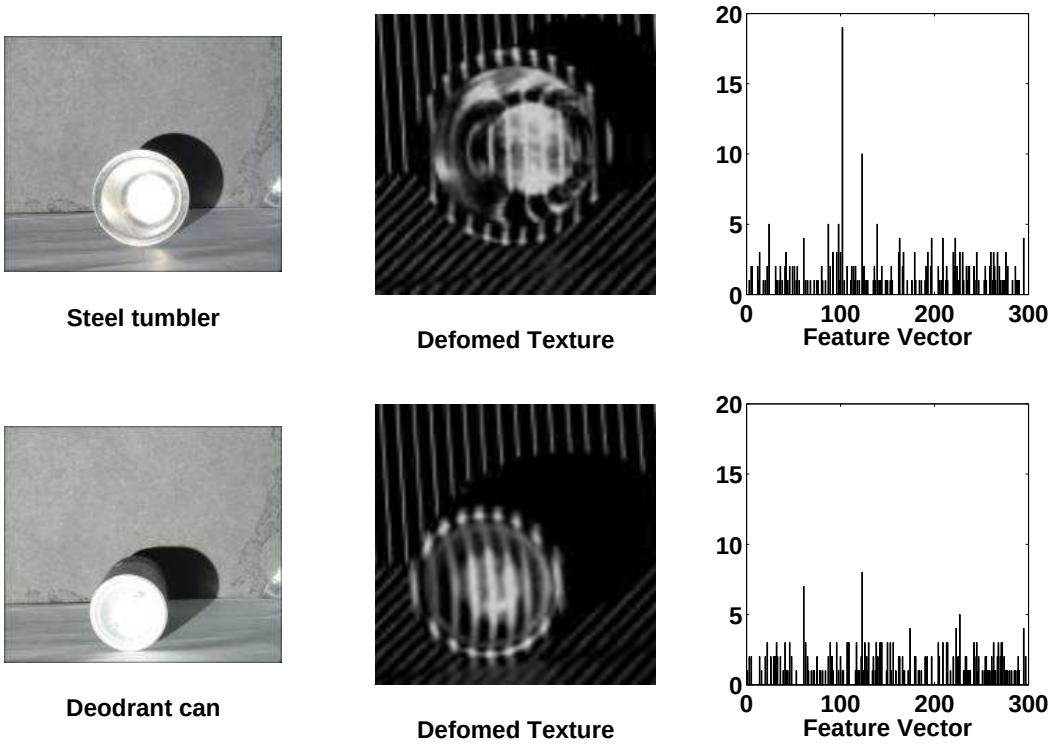


Figure 2: Variations in deformation on similar looking objects.

2 Literature Survey

As we noted before the class of object recognition algorithms work either at image level or by inferring shape information explicitly. In the case to 3D textures, Cula and Dana [2] models the appearance based on their reflectance function and use it for classification. Leung and Malik [11] proposed the use of a set of appearance primitives (3D textons) to model the 3D textures. Wang and Dana [16] infers geometric information of 3D textures from Bidirectional Texture Function [1]. Although the above algorithms work on 2D image features, their definitions are based on lighting variations in 3D. Varma and Zisserman[14] proposed image level features that are invariant of illumination and pose. They further extended the idea of a texton dictionary to achieve highly accurate classification of 3D textures in [15]. Currently, this is one of the best performing classifiers for 3D textures in the image domain, and hence we use it as a benchmark for classification accuracy. However, the approach is 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 projection of structured patterns.

Natural texture in the scene has been used both for recognition of objects such as biometric traits [10, 3, 8] as well as for depth estimation [6, 12]. The primary difference in our approach is that the texture is not an inherent property of the object, but superimposed on it during imaging. Moreover, it is the resultant deformations that characterize the object shape, and not the texture itself. We also provide an example of application of projected texture for recognition of people based on their hand geometry, and

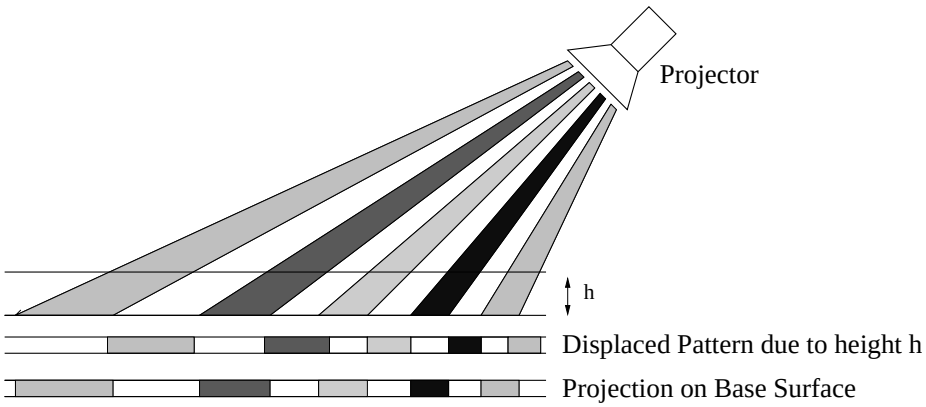


Figure 3: Pattern Shift due to different heights.

the performance is compared with popular 2D approaches such as Jain *et al*[9] and Faundez-Zanuy *et al*[4].

In area of Object category recognition algorithms that uses the shape information explicitly, try to recover the depth information from images using the above cues [17], correspondence across multiple images [7], structured lighting [13], etc. Moreover, for the purpose of recognition, one needs to map the depth profile thus recovered into an underlying 3D shape, which is then characterized using shape based features. The whole process is both error prone and computationally intensive.

On the other hand image-based classification algorithms usually compromise the 3D information for high computation efficiency and robustness. They characterize the objects directly based on the edges and texture in the image. The challenge here is to come up with a representation that is invariant to illumination and pose, thus providing robust classification. Such approaches have become popular both due to its simplicity and robustness to object and pose variations within a category [5]. However, the classification power of such approaches is limited as some of the 3D information is lost during the imaging process.

3 Projected Texture for Recognition

The primary idea of the approach as described before is to encode the depth variations in an object as deformations of a projected texture. 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 under controlled pose and illumination and a specific texture 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 reflected. Figure **??** illustrates this with a cross

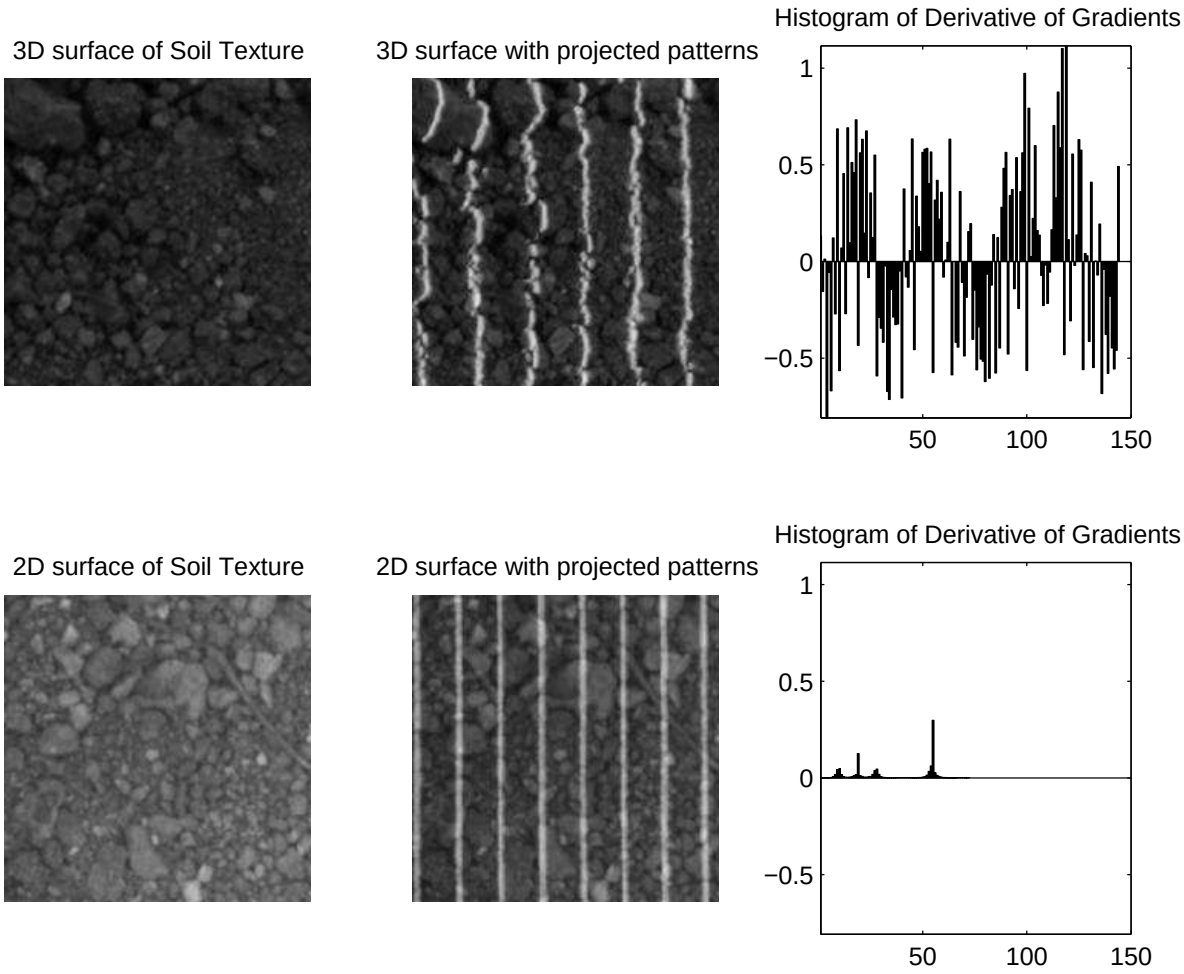


Figure 4: Deformation due to depth variation.

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 axis and the plane of the palm.

- *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 **??**). 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.

3.1 Pattern Deformation and Projector Camera Configuration

We now 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 derive the relationship between the deformation in pattern to various parameters of physical setup and the height variations on object surface. Figure 5(a) shows the image capture setup and Figure 5(b) shows a part of the object surface with slope θ to the Y-axis. We refer to this as the *object plane*. As illustrated in Figure 5(b), we consider the projection of a single

horizontal line pattern at an angle ϕ from Z-axis forming a plane which we will call the *light plane*, with equation : $\frac{x}{a} + \frac{z}{b} = 1$, where $a = b\tan\phi$. The equation of the light plane and the object plane can hence be expressed as:

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

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

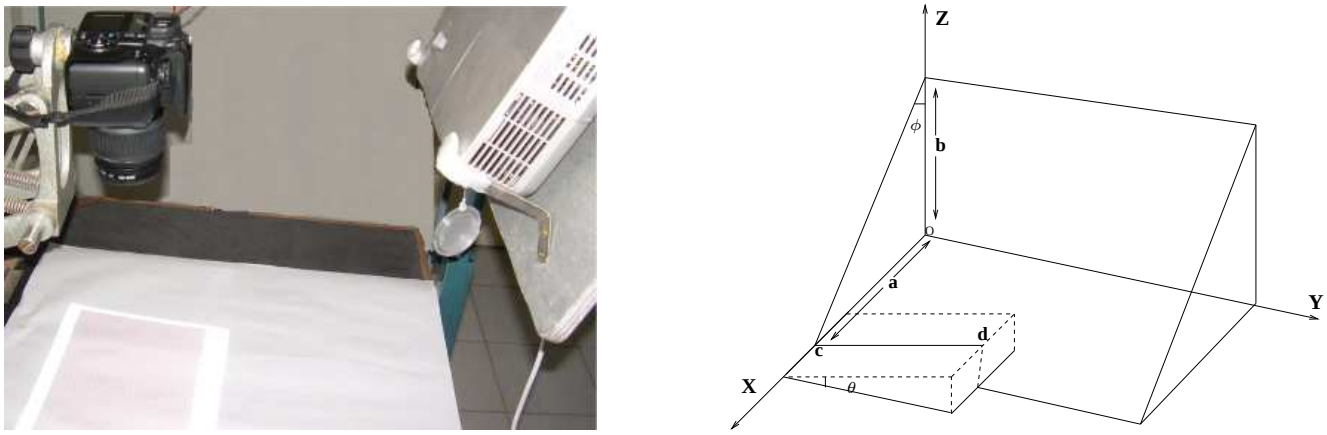
The line cd as shown in figure is the intersection of both of these planes in 3D coordinate system and it can be expressed by cross product of the normals of both intersecting planes. Thus the 3D direction vector of the line cd will be: $\vec{n_3} = [\ \cot\phi \ \ 0 \ \ 1]^T \times [\ 0 \ \ \tan\theta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

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

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

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

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



(a) (b)

Figure 5: The image capture setup and the pattern deformation geometry.

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:

$$\overline{q_1} = \begin{bmatrix} \frac{X_1}{Z_1} & \frac{Y_1}{Z_1} \end{bmatrix}^T \overline{q_2} = \begin{bmatrix} \frac{X_2}{Z_2} & \frac{Y_2}{Z_2} \end{bmatrix}^T \tag{9}$$

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

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

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

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.

3.2 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 in the captured at any point in the image, the gradient of the texture at that point 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 correspond 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 texture, 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 texture so that one can identify the deformations in the captured image.

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 patten that has gradients in eight 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.

4 Characterization of Pattern Deformation

An effective method for characterization of the deformations of the projected texture is critical for its ability to discriminate between different object. We propose a set of texture features the captures the statistics of deformation in the case of 3D textures.

4.1 3D Texture Surface with Stochastic Depth Variation

As we noted before the projection pattern used for 3D texture classification was a set of parallel lines. The feature set that we propose (NHoDG), captures the deformations in the lines and computes the overall statistics.

4.1.1 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 between histograms, equally, and hence employ the Euclidean distance for comparison of histograms. The Normalized histograms, or $NHoDG$ is a simple but extremely effective feature for discriminating between different texture classes. Figure 6 illustrates the computation of the NHoDG feature from a simple image with 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.

4.2 Category Recognition of Rigid Objects

The primary concerns in developing a representation for object category is that the description should be invariant to both shape and pose of the object. Note that the use of projected patterns allows us to avoid object texture, and concentrate only on its shape. Approaches such as 'bag of words' computed from interest points have been successfully employed for image based object category recognition [5].

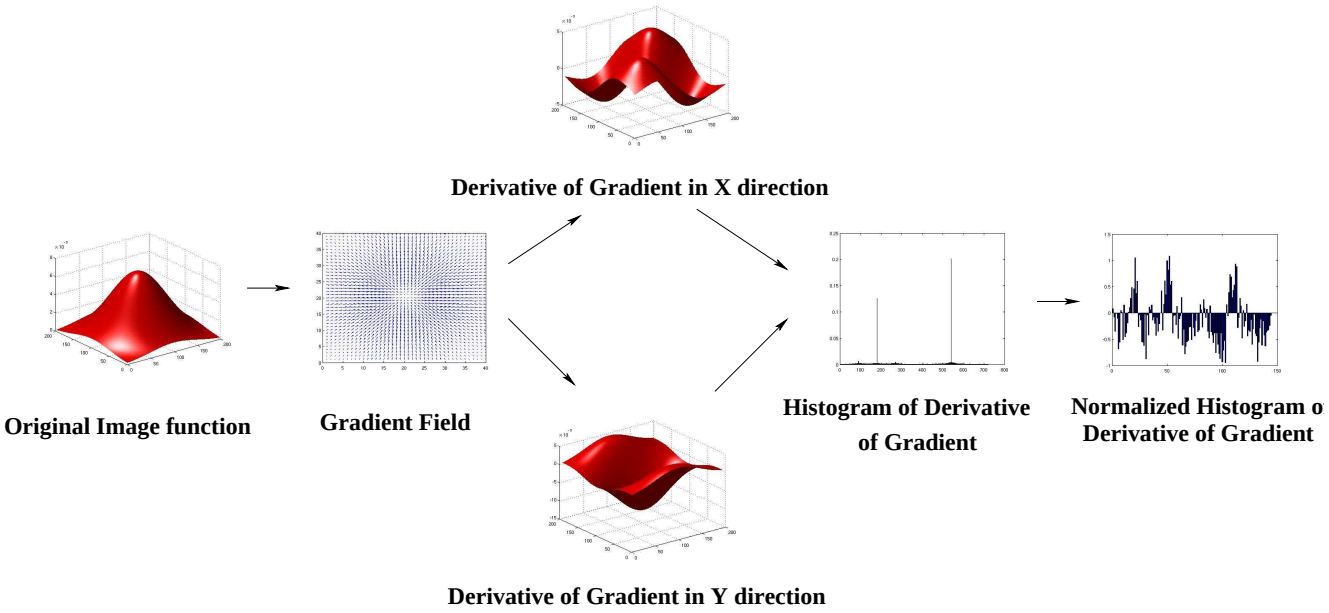


Figure 6: Computation of NHoDG feature vector.

4.2.1 FFT Feature with Bag-of-word Approach

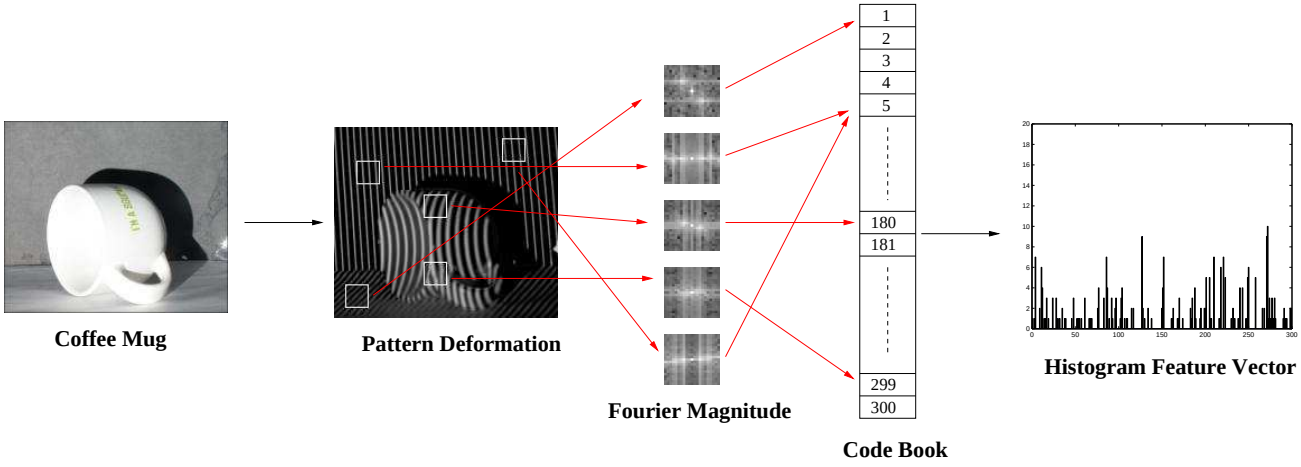


Figure 7: Computation of feature vector.

Our approach is similar in spirit to achieve pose invariance. We learn the class of local deformations that are possible for each category of objects by creating a codebook of such deformations from a training set. Each object is then represented as a histogram of local deformations based on the codebook. Figure 7 illustrates the computation of the feature vector from a scene with projected texture.

There are two primary concerns to be addressed while developing a parts based shape representation:

The location of points from which the local shape descriptor is computed is important to achieve position invariance. In image based algorithms, the patches are localized by using an interest operator that is computed from object texture or edges. However, in our case the primary objective is to avoid

using texture information and concentrate on the shape information provided by the projected texture. Hence we choose to use a set of overlapping windows that covers the whole scene for computation of local deformations. Our representation based on the codebook allows us to concentrate on the object deformation for recognition.

The description of the local deformations should be sufficient to distinguish between various local surface shapes within the class of objects. The feature vector used exploits the periodic nature of projected patterns. Since Fourier representation is an effective descriptor for periodic signals, and since we are interested in the nature of deformation and not its exact location, we compute magnitude or the absolute value of the Fourier coefficients (AFC) corresponding to each of the window patch as our feature vector. To make comparisons in a Euclidean space for effective, we use a logarithmic representation of these coefficients (LAFC). We show that this simple Fourier magnitude based representation of the patches can effectively achieve the discriminative power that we seek.

The feature extraction process proceeds as follows: The images in the training set are divided into a set of overlapping windows of size 20×20 (decided experimentally). Each window is then represented using the magnitude of Fourier representation in logarithmic scale. This results in a 200 dimensional feature vector (due to symmetry of Fourier representation) for each window. A K-means clustering of windows in this feature space allows us to identify the dominant pattern deformations, which forms a codebook for the classification problem (see figure 8). During the testing phase, the feature representations of the windows in an image is computed as above, and each window is mapped to the nearest codebook vector. A histogram of the codes present in an image forms the representation of the object contained in it. As shown in figure 6 the patches that are part of the background maps to one location in codebook. Thus codebook can isolate the words that captures maximum information for defining an object category. Detailed comparison of recognition results of new features and existing one are presented in next section.

We note that the representation is independent of the position, while the classification algorithm achieves pose invariance due to the generalization from different poses in the training set.

4.3 Recognition of Aligned Deterministic shapes

In this class of objects the position of object is assumed to be fixed with respect to projector-camera setup. Here we try to recognize two aligned object having subtle variation with projected texture. We demonstrated this class of recognition problem with hand geometry based person authentication.

4.3.1 Simple Gabor based feature

Here we need to characterize the exact shape of the object, and not the statistics of height variations. We can also use the position of depth variation as a clue for recognition. 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 16 sub-windows (4×4), and the mean response of each filter forms a 384 dimensional feature vector, that is used to represent each sample.

Figure 8: 100 Words from codebook

5 Experimental Results and Analysis

We have conducted exhaustive experiments to validate our approach. Our contribution includes the use of deformed projection patterns as texture as well as proposing the new feature set for each of the class of objects. We have done exhaustive

5.1 3D Texture Surface with Stochastic Depth Variation

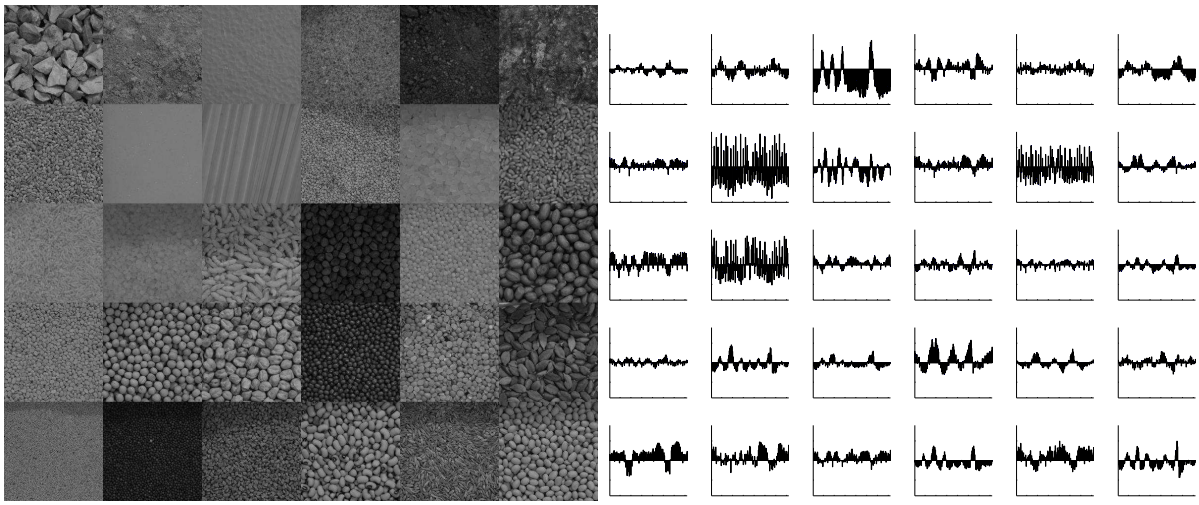
Experimental setup consists of a planar surface to place the object samples, an LCD projector fixed at an angle to the object surface, and a camera located directly above the object with its axis perpendicular to the object plane (see Figure 5(a)). We considered a set of 30 3D textures which has considerable depth variation. Details of each texture is given in Table 1 Total 3600 images were collected, with 120 samples for each of the 30 classes. The 120 samples consist of 24 different object samples, each taken with different projected patterns and illumination conditions. The projected patterns are parallel lines having uniform spacing of $5, 10, 15$ and 20 pixels between them. We call these patterns as W5,W10,W15 and W20 for rest of our experimental validation part.

Our data has large scale variation across the textures, while having several surfaces with similar depth variation profiles, making the recognition task, very challenging. However, we have not varied the

Class Id	Texture Name	Class Id	Texture Name	Class Id	Texture Name
01	Pebble	11	Crystal Sugar	21	Chick Peas
02	Concrete	12	Wheat grain	22	Green gram
03	Thermocol	13	Rice grain	23	Red gram /Pigeon Pea
04	Sand	14	Crystal Salt	24	Cardamom
05	Soil	15	Puffed Rice	25	Poppy seeds
06	Stone	16	Black Gram	26	Mustard seeds
07	Barley	17	Sago	27	Fenugreek seeds
08	Sponge	18	Ground Nut	28	Soybean seeds
09	Ribbed Paper	19	Split Gram beans	29	Fennel/Aniseed
10	Sesame Seed	20	Green Peas	30	White beans

Table 1: List of 3D texture surfaces used in our experiments. We have used to set of grains and pulses to create surfaces with similar scale of depth variations, which makes the classification problem, challenging.

pose of imaging as the application under consideration require controlled illumination conditions. The illumination variations are also limited due to this fact. Images of the 30 different classes along with their NHoDG feature representations are shown in Figure 9.



(a) (b)

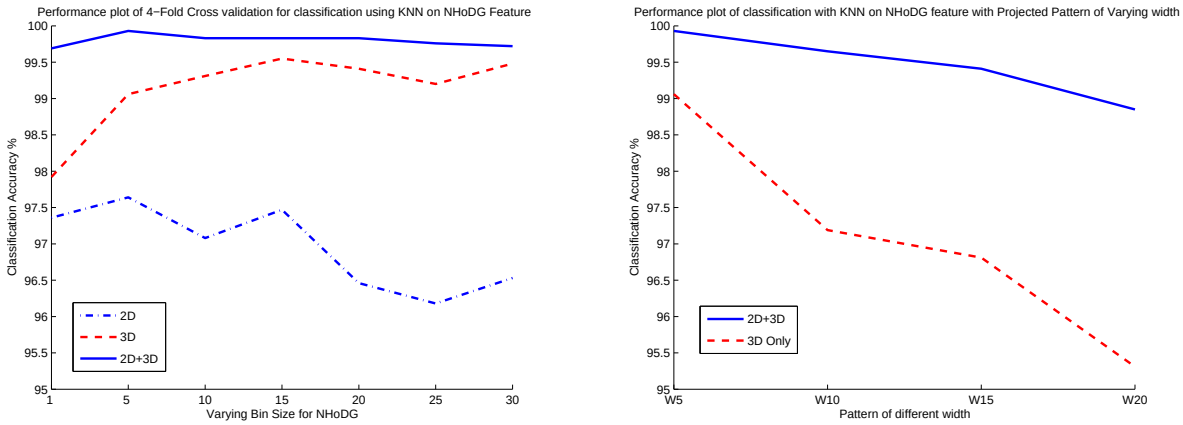
Figure 9: Examples of textures from the 30 classes and their NHoDG representations.

We have run our experiments with and without projection patterns, as well as using the proposed and traditional 2D features.

As the texton dictionary [15] is one of the best performing 2D image feature sets, we have used it for comparison with our approach. We have included two more filters with higher scale (MR12 now instead of MR8) so as to improve the results of texton dictionary on our dataset with higher scale variation. A maximum 50 iteration were used for k-means clustering (details can be found in [15]).

Figure 10 show 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 select a bin size of 5 degrees and pattern separation of 5 pixes for the rest of the experiments.



(a) h (b) h

Figure 10: Classification performance with varying histogram bin sizes and pattern separations.

Table 2 gives a detailed view of the results using MR12 approach and the proposed NHoDG feature set on our dataset as well as the Curet dataset. However, note that the performance on the Curet dataset is 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 approach achieves an error rate of 1.18%, i.e., 34 misclassifications on our dataset of 2880 samples. In comparison, the projection based approach with NHoDG features achieves an error rate of 0.07% when combined with 2D images, which corresponds to just 2 samples being misclassified. Figure 11 shows one of the misclassified samples, and the corresponding NHoDG and Texton 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.

Texton(Unnormalized)	Curet(MR8)	2D	3.15
	Ours(MR12)	2D	1.18
		3D	0.76
		2D + 3D	0.31
NHoDG(Bin Resolution 5)	Curet	2D	12.93
	Ours	2D	2.36
		3D	1.15
		2D + 3D	0.07

Table 2: Error rates of classification using Texton and NHoDG features on our dataset and the Curet dataset (in %).

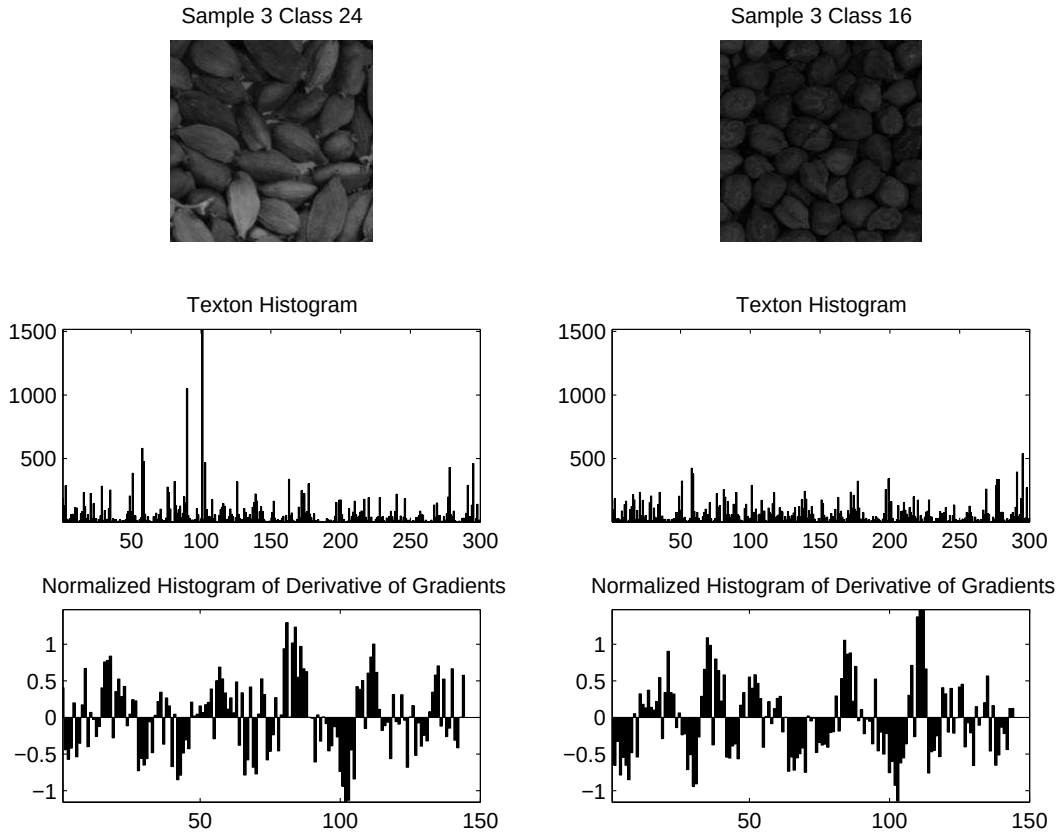


Figure 11: One of the two misclassifications in the dataset using NHoDG feature set.

5.2 Category Recognition of Rigid Objects

The pre-processing stage includes removal of object texture by subtracting a uniformly illuminated image of the object from the image with projection and Gaussian smoothing to reduce the imaging noise. We have collected dataset with total 5 object categories: i) *Coffee Cup*, ii) *Steel Tumbler*, iii) *Plastic Mug*, iv) *Deodorant Spray*, and v) *Alarm Clock*. The categories were chosen to introduce challenging similarities between object categories, and 5 objects of each category were chosen to have large intra-class variations (see Figure 12). For each object, 9 different images were collected with views around 45 degrees apart, making the dataset an challenging one. All images were captured under 8 different texture patterns with varying widths as well as under uniform illumination for comparison. For the purpose of classification, we have used two different classifiers: Multi Layer Perceptron (MLP), which has good generalization capabilities, and a simple Nearest Neighbor (1NN). All results reported are the mean error rates based on 4 Fold cross validation. The number of hidden nodes in the MLP was set to 20 for all experiments.

For the purpose of comparison, we conducted similar experiments with feature proposed by [5] on our dataset without the projected patterns. Note that the comparison is made only to show the effect of the additional information introduced by the projected patterns into the classification process and is not a testimony of the classification algorithm itself. In fact, the algorithms are remarkably similar, and the primary difference is in the local patch representation. Table 5.2 presents the mean error for both of the

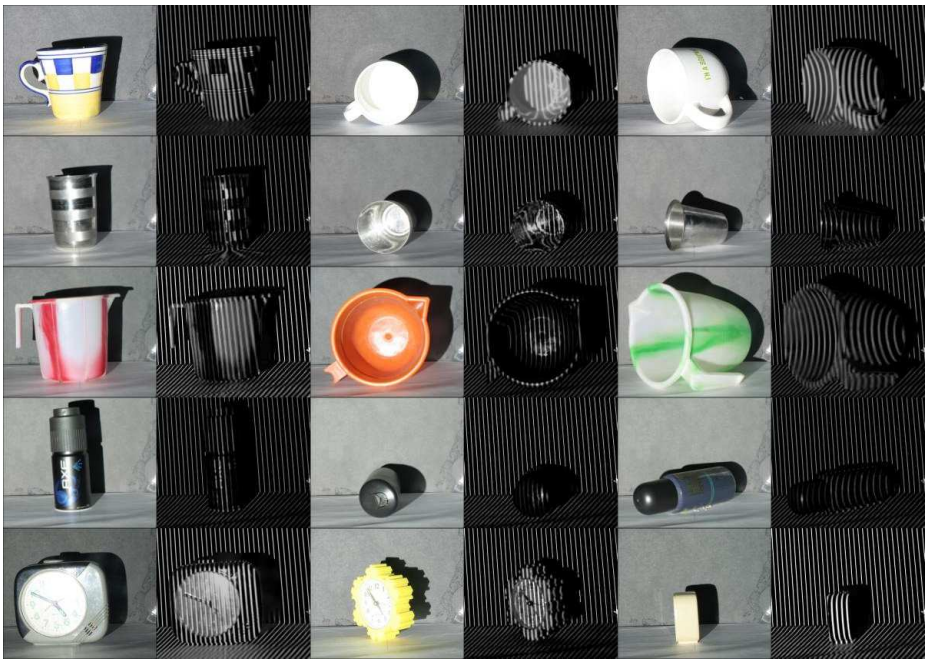


Figure 12: Dataset

approaches. It clearly shows superiority of our approach over the state-of-the-art. Table 5.2 is showing confusing matrix corresponding to results. The error rate is only 1.33% or three misclassification, which includes two between classes 2 and 4, and one between classes 3 and 5 (see Figure 13).

	LAFC	SIFT
MLP	1.33	21.33
1-NN	5.73	20.09

Table 3: Recognition Error Rates

We also conducted experiments with different codebook sizes and pattern variations. Figure 14(a) shows the graph of accuracy vs size of code book. Figure 14(b) explains variation in performance with change in width of projected pattern.

5.3 Recognition of Aligned Deterministic shapes

We have presented the experimental details of hand geometry based person authentication to varify our approach for recognizing aligned deterministic objects. We have our results as well as the performance of the state of the art method on our dataset.

	1	2	3	4	5
1	45	0	0	0	0
2	0	45	0	0	0
3	0	0	45	0	0
4	0	2	0	43	0
5	0	0	1	0	45

Table 4: Confusion Matrix of the proposed approach.

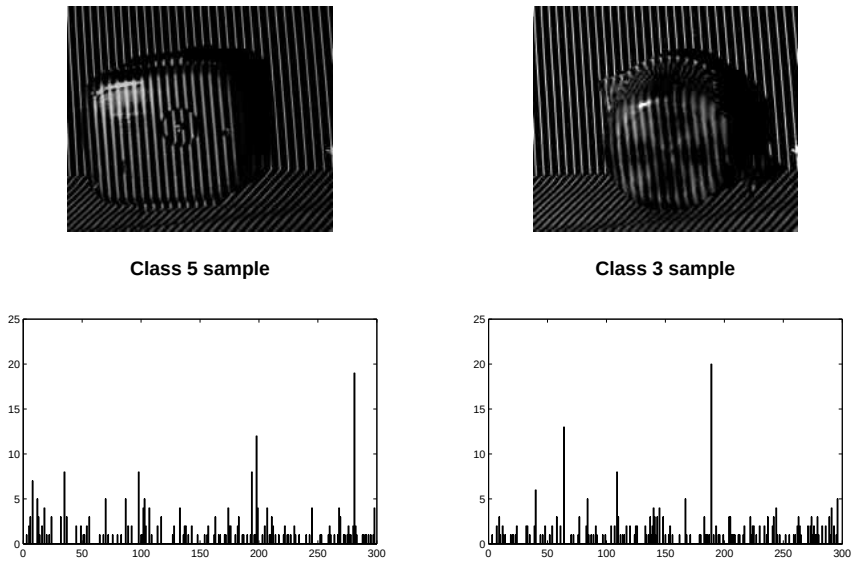
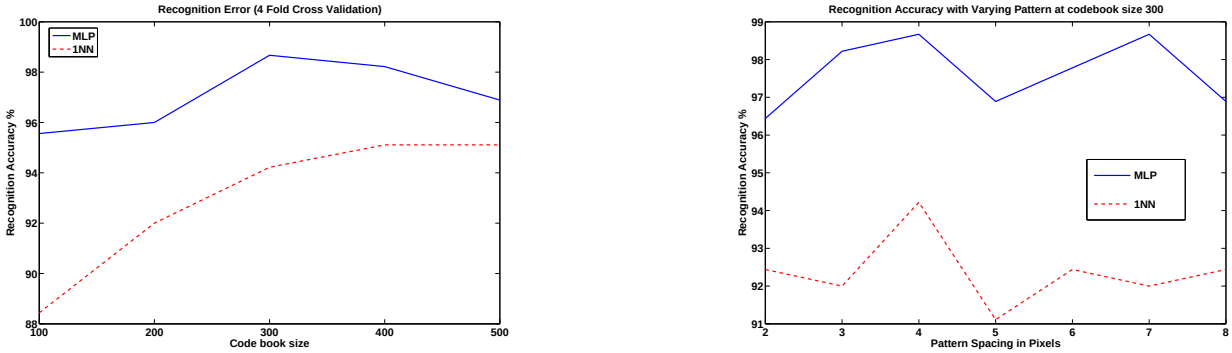


Figure 13: Misclassification example

5.3.1 Hand Geometry Based Authentication

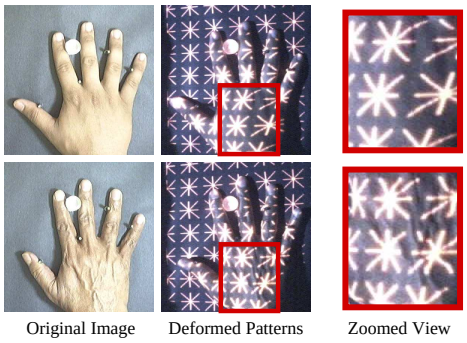
For evaluation of the hand geometry based person authentication algorithm, we collected a dataset of 1380 images from 138 users, each user providing 10 samples each. For comparison, we collected two sets of images from each user, with projected texture as well as with uniform illumination. We compare 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* [9], ii) *Feat-2*: A set of 10 features computed from palm contours proposed by Faundez *et al* [4], and iii) *Feat-3*: The proposed projected texture based features.

Figure 15(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 15(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 between the samples of the same user as well as different users.



(a) (b)

Figure 14: Performance with variation in Codebook size and Pattern width.



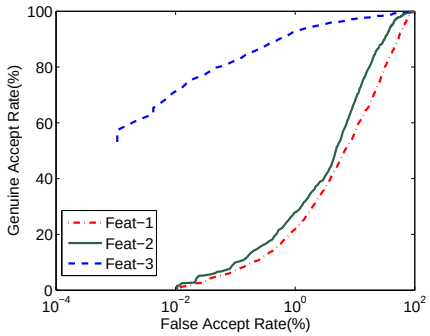


Figure 15: Deformation in projected texture due to hand geometry, and the ROC curve for the different algorithms.

The equal error rate (EER), or the point at which false reject rate equals false acceptance rate, for *Feat-1* and *Feat-2* were 25.0% and 18.65% respectively. In contrast, the proposed feature set achieved and EER of 4.01%. Clearly the projected patterns induce a large amount of discriminating information into the computed features.

6 Conclusion and Future Work

A novel technique for recognition of 3D objects using projected texture is proposed. The results were demonstrated in the case of three different object classes, one for 3D texture classification, second for object category recognition for rigid objects, and third for hand geometry based person authentication as position sensitive feature. The overall approach is computationally efficient as we need not to find correspondences or reconstruct the 3D object model. Moreover, the computational requirements are comparable to the simpler 2D image based recognition approaches. The proposed approach is also robust to noise and occlusion issues, which are difficult to handle for approaches that rely on recovery of 3D models.

However, a several issues still remain unaddressed in applying the recognition approach to generic

objects. The method is sensitive to the relative positioning of the camera, the projector, and the object for deterministic surfaces. Object reflectance and transparency might be another interesting area to explore. Temporal variations in dynamic texture deformations could also give us a cue towards designing optimal classifiers for recognition.

References

[1] O. G. Cula and K. J. Dana. Compact representation of bidirectional texture functions. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 1041–1047, 2001.

[2] O. G. Cula and K. J. Dana. 3d texture recognition using bidirectional feature histograms. *International Journal of Computer Vision*, 59(1):33–60, 2004.

[3] J. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993.

[4] M. Faundez-Zanuy, D. A. Elizondo, M. Ferrer-Ballester, and C. M. Travieso-González. Authentication of individuals using hand geometry biometrics: A neural network approach. *Neural Processing Letters*, 26(3):201–216, 2007.

[5] R. Fergus. *Visual Object Category Recognition*. PhD thesis, University of Oxford, Dec. 2005.

[6] D. Forsyth. Shape from texture and integrability. *Proceedings of IEEE International Conference on Computer Vision*, 2:447–452, July 2001.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[8] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti. Filterbank-based fingerprint matching. *IEEE Transactions on Image Processing*, 9(5):846–859, 2000.

[9] A. K. Jain, A. Ross, and S. Pankanti. A prototype hand geometry-based verification system. In *Proceedings of the AVBPA'99*, pages 166–171, Washington D.C., Mar. 1999.

[10] A. Kumar and D. Zhang. Personal recognition using hand shape and texture. *IEEE Transactions on Image Processing*, 15(8):2454–2461, Aug. 2006.

[11] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, June 2001.

[12] A. Loh and R. Hartley. Shape from non-homogeneous, non-stationary, anisotropic, perspective texture. *Proceedings of the British Machine Vision Conference*, pages 69–78, 2005.

[13] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Proc. of CVPR*, pages 195–202, 2003.

[14] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *Proceedings of the 7th European Conference on Computer Vision*, volume 3, pages 255–271, Copenhagen, Denmark, May 2002. Springer-Verlag.

[15] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis*, 62(1-2):61–81, April-May 2005.

[16] J. Wang and K. J. Dana. Hybrid textons: modeling surfaces with reflectance and geometry. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 372–378, 2004.

[17] R. White and D. Forsyth. Combining cues: Shape from shading and texture. In *Proc. of CVPR*, volume 2, pages 1809–1816, 2006.