Submitted to *Texture 2002 - 1 and 2 June 2002, Copenhagen (co-located with ECCV 2002)*

Image-based Skin Analysis

Oana G. Cula
CS Department

Kristin J. Dana
ECE Department

Rutgers University
Piscataway, NJ 08854
{*oanacula,kdana*}*@caip.rutgers.edu*

Abstract

Quantitative characterization of skin appearance is an important but difficult task. The skin surface is a detailed landscape, with features that depend on many variables such as body location (knuckle vs. torso), subject parameters (age/gender/health) and imaging parameters (lighting and camera). Computational modeling of skin texture has potential uses in many fields and applications including realistic rendering for computer graphics, robust face models for computer vision, computer assisted diagnosis for dermatology, topical drug efficacy testing for the pharmaceutical industry and quantitative product comparison for cosmetics. In this work, image-based representations of skin appearance are used in order to have descriptive capabilities without the need for prohibitively complex physics-based skin models. We present a method for representing and recognizing different areas of the skin surface that have visibly different texture properties. Our model takes into account the varied appearance of the skin with changes in illumination and viewing direction.

1 Introduction

Simple models of skin appearance are not sufficient to support the demands for high performance algorithms in computer vision and computer graphics. For example, in computer vision algorithms for face recognition, shape from shading and facial feature-tracking rely on accurately predicting appearance so that local matching can be done among images obtained with different imaging parameters. In computer graphics, image-based rendering is a popular technique to achieve realism, but this technique is often based on the assumption that varying the appearance of local texture can be accomplished by warping reference images to create new views. If fidelity of fine scale texture is needed, this rendering approach is not sufficient. The appearance of skin texture and the manner that appearance changes with illumination direction and viewing direction is difficult to model. Although much work has been done in modeling for facial animation [17, 13, 12, 2], accurately rendering skin surface detail has not been the primary emphasis and remains an open topic.

In addition to computer vision and graphics, accurate skin models may be useful in dermatology and several industrial fields. In dermatology, these skin models can be used to develop methods for computer-assisted diagnosis of skin disorders. In the pharmaceutical industry quantification is quite useful when applied to measuring healing progress. Such measurements can be used to evaluate and compare treatments and can serve as an early indicator of the success or failure of a particular treatment course. Consumer products and cosmetic industries can use computational skin representations to substantiate claims of appearance changes.

Skin texture can be considered a 3D texture, i.e. a texture in which the fine scale geometry affects overall appearance. In some cases there may be a variation of albedo in addition to fine scale geometric texture. Increasingly, recent work on texture representations deal with the complex changes of surface appearance with changes in illumination and viewing direction [4, 16, 18, 25, 7, 9, 8, 26, 20, 6, 5]. Terminology for texture that depends on imaging parameters was introduced in [10]. Specifically, the term bidirectional texture function (BTF) is used to describe image texture as a function of the four imaging angles (viewing and light source directions). The BTF is analogous to the bidirectional reflectance distribution function (BRDF). While BRDF is a term for the reflectance of a point, most real world surfaces exhibit a spatially varying BRDF and the term BTF is used for this situation. To illustrate the importance of accounting for variations with imaging parameters, consider Figure 5 (a), where each row shows a skin surface patch under several different viewing and illumination directions. In this paper we extend our approach described in [6] to skin analysis. This approach is useful because a single image can be used for fast non-iterative recognition. We present recognition results using clinical skin measurements from several locations on the hand.

One difficulty in building skin texture models is acquiring good data. Since changes in skin texture with viewing and illumination direction are to be modeled, a full BTF is ideal. However the BTF is the appearance of a surface patch as a function of four parameters, two angles for the viewing direction and two angles for the illumination direction. These measurements can be quite difficult to obtain because of mechanical requirements of the setup. When the sample is non-planar, non-rigid and not-fixed, as is the case for human skin, the measurements are even more difficult. Because of these issues, exhaustive measurement is impractical in a clinical setting. In our approach, three viewing directions are used, and for each viewing direction, ten lighting directions are measured, obtaining a total of 30 images per surface sample. The viewing direction is obtained with a tripod augmented with a manual articulated arm and illumination is controlled by a rotatable arc mounted on a tripod. The details of the measurement protocol are discussed in Section 4.2. These measurements are obtained as part of

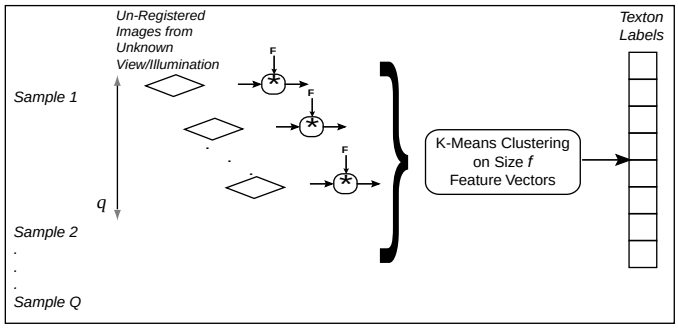


Figure 1: Creation of the image texton library. The unregistered texture images of different BTF's are filtered with the filter bank F. The filter responses for each pixel are grouped to form the feature vectors. The feature space is clustered via k-means to determine the collection of key features, i.e the image texton library.

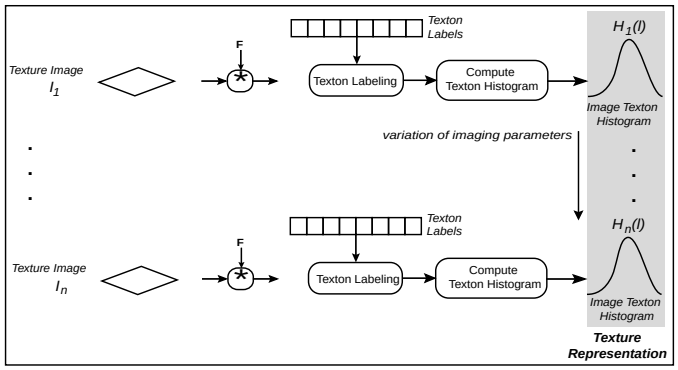


Figure 2: Our proposed texture representation. The texture image is filtered with filter bank F, and filter responses for each pixel are grouped to form feature vectors. The feature vectors are projected into the space spanned by the elements of the image texton library, then labeled by determining the closest texton. The distribution of labels is approximated by the texton histogram.

a collaboration with UMDNJ Department of Dermatology and the eventual outcome of this study will investigate disease and treatment progress as well as the appearance of normal skin on various body locations. In this paper, we present experiments using skin on the articulated segments of a finger, which can be divided into several classes. Recognition of the location the image was taken on the hand is obtained using an image-based approach. Often in the literature, texture recognition is applied to texture classes that are visually quite dissimilar, e.g. rocks vs. sand. The goal of this paper is to perform a classification of skin texture where the visual differences in appearance are more subtle.

2 Texture Representation

2.1 Image Textons

Based on the classic concept of textons [15], and also on modern generalizations [18], we follow the theory that there is a finite set of local structural features that can be found within a large collection of texture images from various samples. This reduced set of local structural representatives is called the image texton library.

A widely used computational approach for encoding the local structural attributes of textures is based on multichannel filtering [3, 14, 24, 18], this type of analysis being inspired by various evidences of similar processing in human vision system. In our approach, illustrated in Figure 1, to obtain a computational description of the local feature we employ a multi-resolution filter bank F, with size denoted by f, and consisting of Gaussians, derivatives of Gaussians and center surround derivatives on three scales. Each pixel of a texture image is characterized by a set of three multi-dimensional feature vectors obtained by grouping the corresponding filter responses over scale. For simplicity in our discussions we will refer to a single scale, but it is important to keep in mind that the processing is done in parallel for all three scales. Note that the resulting feature vector does not encode the variation of local appearance as the imaging conditions change. Instead we account for the change in appearance with viewing/illumination directions globally by populating the feature space with feature vectors from sampled BTFs.

As in many approaches in texture literature [18, 1, 23, 19], we cluster the feature space to determine the set of prototypes among the population. Specifically, we invoke k-means algorithm, which is based on the first order statistics of data, and finds a predefined number of centers in the data space, while guaranteeing that the sum of squared distances between the initial data points and the centers is minimized. Empirical results suggest that the resulting collection of representatives in the space spanned by the local structural feature vectors, namely the image texton library, is generic enough to represent a large set of texture samples.

2.2 Texton Histograms

The histogram of image textons is used to encode the global distribution of the local structural attribute over the texture image. This histogram, denoted by $H(l)$, is a discrete function of the labels l induced by the image texton library and is computed as described in Figure 2. Note that in our approach, neither the image texton nor the texton histogram encode the change in local appearance of texture with the imaging conditions. These quantities are local to a single texture image. We represent the surface using a collection of image texton histograms, acquired as a function of viewing and illumination direction. This surface representation may be described by the term *bidirectional feature histogram*. It's worthwhile to explicitly note the difference between the bidirectional feature histogram and the BTF. While the BTF is the set of measured images as a function of viewing and illumination, the bidirectional feature histogram is a representation of the BTF suitable for use in classification.

In our work, each texture class is modeled using a large collection of texton histograms. The dimensionality of the histogram space is given by the cardinality of the image texton library, which should be inclusive enough to represent a large range of textured surfaces. Therefore the histogram space is high dimensional, and a compression of this representation to a lower-dimensional one is suitable, providing that the statistical properties of the bidirectional feature histograms are still preserved. To accomplish dimensionality reduction we employ principle component analysis (PCA), which finds an optimal new orthogonal basis in the space, while best describing the data. This approach follows [21], where a similar problem is treated, specifically an object is represented by set of images

taken from various poses, and PCA is used to obtain a compact lower-dimensional representation.

3 Recognition Method

The 3D texture recognition method consists of three main tasks: (1) creation of image texton library, (2) training, and (3) classification. The image texton library is described in Section 2.1 and illustrated in Figure 1. Each of the q texture images, from each of the Q texture samples, is filtered by the multi-channel filter bank F, and the filter responses corresponding to a pixel are grouped to form the feature vector. The feature space is populated with the feature vectors from all $Q \times q$ images. Feature grouping is performed via k-means algorithm, and the representatives among the population are found, forming the image texton library.

In the training stage a 3D texture model for each of the texture samples is constructed as illustrated in Figure 3. The subset of texture images used for training are arbitrarily sampled from the entire measured BTF. Each of the p texture images from each of the P texture samples to be classified, are filtered by the same filter bank F as the one involved for texton library construction. The resulting multidimensional feature vectors, obtained by grouping the filter responses for a certain pixel in the image, are labeled relative to the set the image textons, and for each texture image a texton histogram is computed. Therefore each texture sample is characterized by the set of p texton histograms, or bidirectional feature histogram. Furthermore, the set of bidirectional feature histograms corresponding to the set of texture samples are employed to compute the universal eigenspace, where each of the texture samples is modeled as a densely sampled manifold, parametrized by both viewing and illumination directions.

In the classification stage, illustrated in Figure 3, the subset of testing texture images is disjoint from the subset used for training. Again, each image is filtered by F, the resulting feature vectors are projected in the image texton space and labeled according to the texton library. The classification is based on a single novel texture image, and it is accomplished by projecting the corresponding texton histogram onto the universal eigenspace created during training, and by determining the closest point in the eigenspace. The texture sample corresponding to the manifold onto which the closest point lies is reported as the texture class of the testing texture image. To demonstrate the versatility of the image texton library we designed two classification experiments.

3.1 Prior work in 3D texture recognition

Prior methods for texture recognition based on representations derived from image sets are discussed in [25, 8, 18]. In [8] the texture representations are the conditional histograms of the response to a small multiscale filter bank. PCA is performed on the histogram of filter outputs and recognition is done using the SLAM library [22, 21] with the histogram vectors forming the appearance-based feature vectors. The preliminary recognition results in this experiment were encouraging and motivated the work presented in this paper. In [25] the individual texture images are represented using multiband correlation functions that consider both within and between color band correlations. PCA is used to reduce dimensionality of the representation and this color information is used to aid in recognition. However, the use of texture color in [25] greatly assists the recognition since

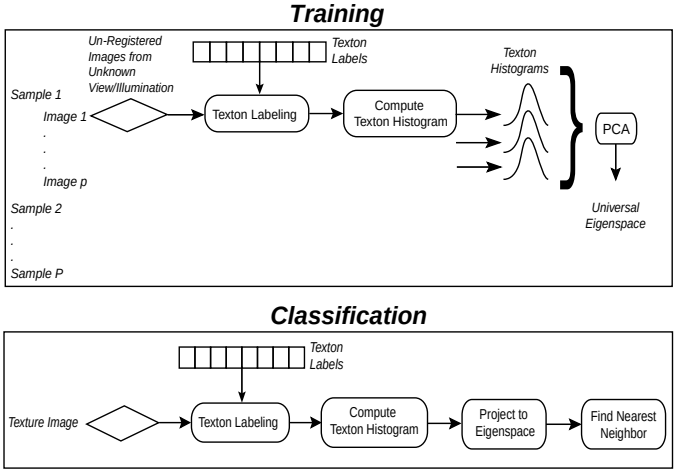


Figure 3: Training and classification stages of image texton recognition method. During training PCA is performed on the histograms of image textons. Recognition is accomplished by finding the closest neighbor to the novel point in the texton histogram eigenspace.

many of the samples in the test dataset are well separated in color space. In [18] a 3D texton is created by using a multiscale filter bank applied to an image set for a particular sample. The filter responses as a function of viewing and illumination directions are clustered to form appearance features that are termed 3D textons. In this paper we refer to this method as the *3D texton method*.

The classical standard framework for texture representation consists of a primitive and a statistical distribution of this primitive over space. So the pertinent question is: how does one extend a standard texture representation to account for changes in image texture with imaging parameters? To accomplish this extension, either the primitive or the statistical distribution should be a function of the imaging parameters. Using this framework, the comparison of our approach with the 3D texton method is straightforward. The 3D texton method uses a primitive that is a function imaging parameters, while our method uses a statistical distribution that is a function of imaging parameters. In our approach the histogram of features representing the texture appearance is a function of viewing and illumination and therefore is called a bidirectional histogram. We accomplish a similar goal to that of the 3D texton method, namely precise recognition of textured surfaces. However our approach uses less a priori information and a fundamentally simpler computational representation.

4 Experimental Results

In developing the recognition method, several experiments have been conducted with the CUReT database [11], a collection of BTF/BRDF measurements for 61 real-world textured surfaces, each imaged under more than 200 different combinations of viewing and illumination directions. In this section we summarize these results to show the overall performance of the algorithm on a large dataset. For the task of skin analysis, we describe a new clinical measuring project that is now underway in collaboration with UMDNJ Department of Dermatology. We apply the image texton method to preliminary images from this study and the results are encouraging.

4.1 Results with CUReT Database

In order to analyze the performance of our texture representation and recognition algorithm, we designed various experiments which employed a large number of texture images corresponding to 40 of the 61 texture samples from CUReT database. For each texture sample we considered a highly sampled BTF, consisting of $N = 156$ texture images. To create the image texton library, we used $Q = 20$ texture samples, each characterized by $q = 64$ arbitrary texture images. To form feature vectors, each image is filtered with a filter bank F of size $f = 45$ filters, grouped on three scales. For each scale, there are 12 oriented Gaussian filters with 6 orientations and 2 phases, 2 center-surround derivatives of Gaussian filters, and one low-pass Gaussian filter.

We conducted two series of recognition experiments. First set of experiments uses for classification the same 20 texture samples as the ones used for creating the texton library. For this case we analyzed the performance of the method as the size of the training set for each class is varied. We obtain an excellent global recognition rate of more than 96%, for the case of using $p = 56$ texture images for training, and the remaining of 100 texture images for testing, for each of the classes. To demonstrate the generality of the texton library, the second set of recognition experiments classifies a different set of 20 texture samples, samples which have not been used to construct the texton library. Again, for the case of using $p = 56$ texture images for training, and the remaining of 100 texture images for testing, for each of the classes, we obtain a classification rate well over 96%. This result confirms the assumption that different textured surfaces have similar local structural properties, therefore an image texton library obtained from a limited collection of texture samples might also convey the texture characteristics of other surfaces.

4.2 Skin Imaging Protocol

Our skin measurement device is comprised of a Sony DFW-V500 IEEE-1394 digital camera equipped with InfiniMini video lens with variable focus from 750mm to 150mm, and video magnification 2.9-15X, a rheostat controlled high intensity quartz halogen illuminator, a personal computer running Windows 2000 with IEEE-1394 digital camera Windows driver (available freeware at URL http://www-2.cs.cmu.edu/ iwan/1394/).

Various camera poses are obtained by using a tripod and a manually articulated arm. The light source is positioned by using a rotatable positioning arc, which ensures constant distance between surface of interest and light source. The lamp is positioned distant enough from the surface of interest such that an approximative collimation of light is ensured. In order to monitor the light intensity across various imaging sessions, a white diffuse reflectance target providing an extremely flat spectral response in UV, VIS, and IR, is imaged during each imaging session.

4.3 Recognition of Skin Images

In our experiments we employed skin texture images corresponding to three distinct regions of a finger: bottom segment on palm side, fingertip, and bottom segment on the back of the hand. Images have been obtained from two subjects: for subject 1 both the index and middle fingers of left hand have been imaged, for subject 2 the index finger of left hand has been measured. Each imaging session consists of measuring the skin texture of a certain region, finger and subject, as the imaging parameters are varied. Therefore we had 9 imaging sessions, and for each session 30 images were acquired, corresponding to 3 camera poses, and 10 light source positions for each camera pose. As a result the entire dataset employed during the experiments contains 270 skin texture images. Figure 5 illustrates few examples of texture images in this dataset.

For constructing the image texton library, we consider a set of skin texture images from all three classes, however only from index finger of subject 1. This reduced subset of images is used because we assume that the representative features for a texture surface are generic. This assumption is even stronger for skin textures, given the local structural similarities between various skin texture classes.

Each texture image is filtered by employing a filter bank, consisting of $f = 18$ oriented Gaussian filters with six orientations corresponding to three distinct scales. The filter outputs corresponding to a certain scale are grouped to form six-dimensional feature vectors. The resulting three sets of feature vectors are used each to populate a feature space, where clustering via k-means is performed to determine the representatives among the population. We empirically chose to employ in our experiments a texton library consisting of 50 textons for each scale.

During the first set of experiments, the training and testing image sets for each class are disjoint, corresponding to different imaging conditions or being obtained from different surfaces belonging to the same class. For each of the classes we considered all available data, that is, each texture class is characterized by 90 images. We varied the size of the training set for each of the classes, from 45 to 60, and, consequently the cardinality of the test set was varied from 45 to 30. For a fixed dimensionality of the universal eigenspace, i.e. 30, the profiles of individual recognition rates for each class, as well as of global recognition rate indexed by the size of the training set are illustrated in Figure 4 (a). As the training set for each class is enlarged, the recognition rate improves, attaining the value 100% for the case of 60 texture images for training and rest of 30 for testing. To emphasize the strength of this result consider that the classification is based on either: a single texture image captured under different imaging conditions than the training set; or a single texture image captured under the same imaging conditions, but from a different skin surface. The variation of recognition rate as a function of the dimensionality of the universal eigenspace, when the size of the training set is fixed to 60, is depicted in Figure 4 (b). As expected, the performance improves as the dimensionality of the universal eigenspace is increased.

In the second experiment training and testing images correspond to the same surface captured under the same imaging conditions, however the image regions are spatially disjoint. We divide each skin texture image into two non-overlapping subimages, denoted as lower half subimage, and upper half subimage. As a result we obtain for each class a set of 60 texture subimages, two for each of the 30 combinations of imaging parameters. For this experiment we considered data obtained from index finger of subject 1. The training set is constructed by alternatively choosing lower half and upper half subimages, which correspond to all 30 imaging conditions. The testing set is the complement of training set relative to the set of 60 subimages for each class. The recognition rate indexed by the dimensionality of the universal eigenspace is plotted in Figure 4 (c). For

the case of a 30-dimensional eigenspace, the global recognition rate is about 95%, when for class 1 is attained a recognition rate of 100%, class 3 is classified with an error smaller than 4%, and for class 2 the recognition rate is about 87%. Class 2 is the most problematic to be classified, due in part to the non-planarity of the fingertip.

These excellent preliminary classification results confirm that our proposed texture representation captures well the characteristics of skin texture, allowing good discrimination between different classes which correspond to different regions of the human body. The results are even more encouraging when one considers that visual differences in appearance of various skin areas are rather subtle.

5 Conclusion and Future Work

In summary, we have measured skin appearance and applied image-based texture representation for skin classification. The techniques used enable classification of textured surfaces of unknown viewing and illumination directions. The observed skin appearance changes significantly with changes in viewing and illumination direction because the surface microgeometry introduces local occlusion, shadowing, and foreshortening. Features apparent in one view seemingly disappear in another image, while new features reappear. This multiview approach (that is many images characterize the surface) provides a more comprehensive surface representation than any single image texture representation.

Future work includes fine-tuning filter characteristics to detect skin features. As in prior work, the image texton library should be generic when trained on a sufficiently large set. We showed this to be the case here, but improvements for skin data can be made. Methods to explore compact representations have been explored in prior work [5] and those methods can be used to guide measurements, to ensure that the current measurement protocol is sound.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0092491 and Grant No. 0085864.

References

[1] S. Aksoy and R. M. Haralick. Graph-theoretic clustering for image grouping and retrieval. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1:63–68, 1999.

[2] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. *Proceedings of SIGGRAPH*, pages 187–194, 1999.

[3] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, January 1990.

[4] M.J. Chantler. Why illuminant direction is fundamental to texture analysis. *IEE Proceedings Vision, Image and Signal Processing*, 142(4):199–206, 1995.

[5] O. G. Cula and K. J. Dana. Compact representation of bidirectional texture functions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, I:1041–1047, December 2001.

[6] O. G. Cula and K. J. Dana. Recognition methods for 3D textured surfaces. *Proceedings of SPIE Conference on Human Vision and Electronic Imaging VI*, 4299:209–220, January 2001.

[7] K. J. Dana and S. K. Nayar. Histogram model for 3D textures. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 618–624, June 1998.

[8] K. J. Dana and S. K. Nayar. 3D textured surface modeling. *IEEE Workshop on the Integration of Appearance and Geometric Methods in Object Recognition*, pages 46–56, June 1999.

[9] K. J. Dana and S. K. Nayar. Correlation model for 3D texture. *International Conference on Computer Vision*, pages 1061–1067, September 1999.

[10] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real world surfaces. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 151–157, June 1997.

[11] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real world surfaces. *ACM Transactions on Graphics*, 18(1):1–34, January 1999.

[12] D. DeCarlo, D. Metaxas, and M. Stone. An anthropometric face model using variational techniques. *Proceedings of SIGGRAPH*, pages 67–74, 1998.

[13] B. Guenter, C. Grimm, D. Wood, H. Malvar, and F. Pighin. Making faces. *Proceedings of SIGGRAPH*, pages 55–66, 1998.

[14] A.K. Jain, S. Prabhakar, and L. Hong. A multichannel approach to fingerprint classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):348–369, 1999.

[15] B. Julesz. Textons, the elements of texture perception and their interactions. *Nature*, (290):91–97, 1981.

[16] J. J. Koenderink, A. J. van Doorn, K. J. Dana, and S. K. Nayar. Bidirectional reflection distribution function of thoroughly pitted surfaces. *International Journal of Computer Vision*, 31(2-3):129–144, 1999.

[17] Y. Lee, D. Terzopoulos, and K. Walters. Realistic modeling for facial animation. *Proceedings of SIGGRAPH*, pages 55–62, 1995.

[18] 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, 2001.

[19] W. Y. Ma and B. S. Manjunath. Texture features and learning similarity. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 425–430, 1996.

[20] G. McGunnigle and M. J. Chantler. Rough surface classification using first order statistics from photometric stereo. *Pattern Recognition Letters*, 21:593–604, 2000.

[21] H. Murase and S. K. Nayar. Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, pages 5–24, 1995.

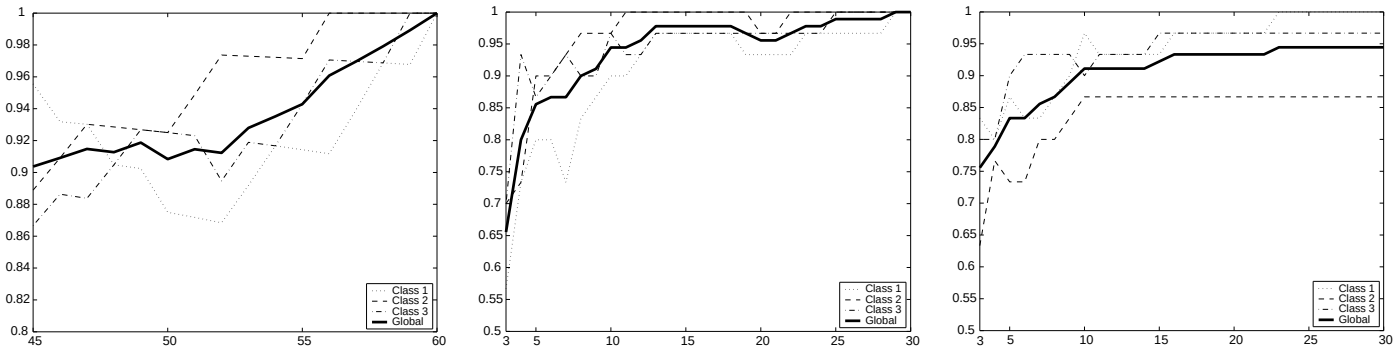
[22] S. A. Nene, S. K. Nayar, and H. Murase. Slam: A software library for appearance matching. *Technical Report CUCS-019-94*, Proceedings of ARPA Image Understanding Workshop, November 1994.

[23] J. Puzicha, T. Hoffman, and J. Buchmann. Histogram clustering for unsupervized image segmentation. *International Conference on Computer Vision*, 2:602–608, 1999.

[24] T. Randen and J. H. Husoy. Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):291–310, 1999.

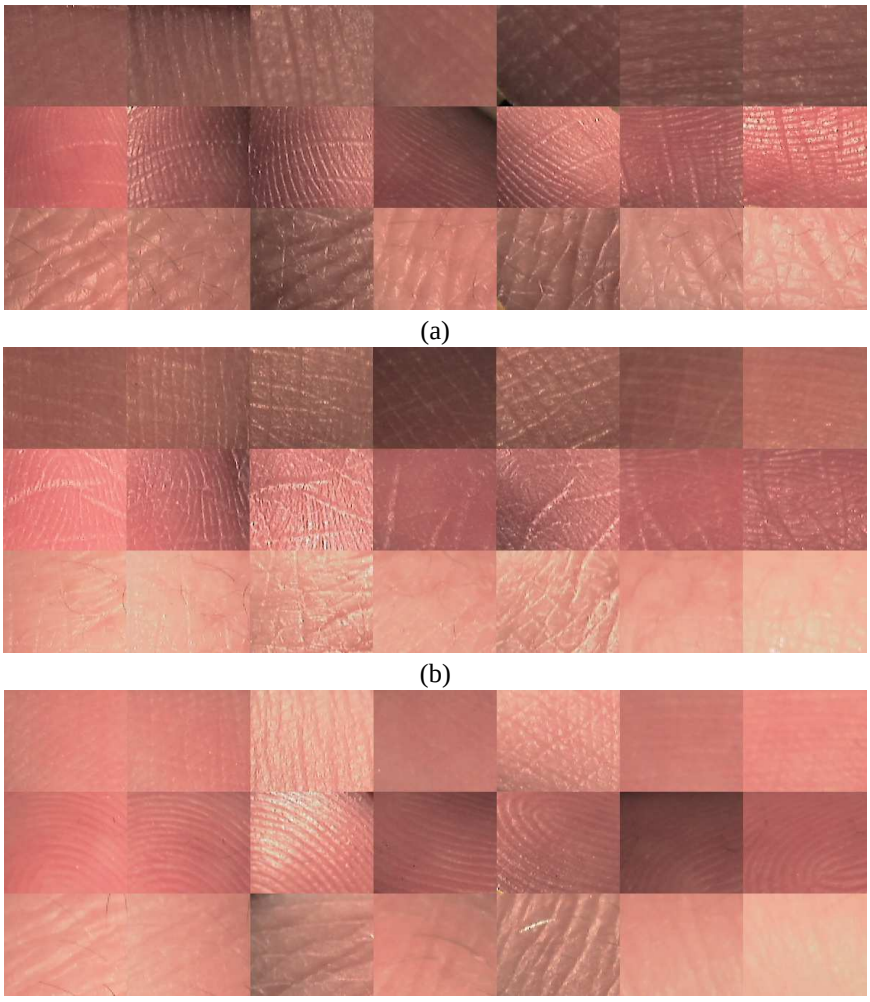
[25] P. Suen and G. Healey. Analyzing the bidirectional texture function. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 753–758, June 1998.

[26] P. Suen and G. Healey. The analysis and recognition of real-world textures in three dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):491–503, May 2000.



(a) (b) (c)

Figure 4: Recognition rate as a function of the size of the training set (a) (when dimensionality of the universal eigenspace is fixed to 30), and as a function of the dimensionality of the universal eigenspace (b) (when the training set of each class has cardinality 60), both corresponding to first set of recognition experiments reported in Section 4.3. (c) Profile of recognition rate as a function of the dimensionality of the universal eigenspace, corresponding to second recognition experiment, described in Section 4.3.



(a)

(b)

(c)

Figure 5: Examples of skin texture images for each class, and for each of the three fingers imaged during our experiments. In each of the pictures first row depicts skin texture corresponding to class 1 (bottom segment, palm side), second row presents texture images from class 2 (fingertip), and third row consists of texture images from class 3 (bottom segment, back of palm). In (a) images are obtained from index finger of subject 1, in (b) from middle finger of subject 1, and in (c) from index finger of subject 2.