Classification of Textures under Various Lighting and Viewing Conditions

Funda Durupınar

Department of Computer Engineering, Bilkent University,
Ankara, 06800, Turkey

fundad@cs.bilkent.edu.tr

Abstract. In this paper, 3D texture classification problem is addressed. 3D textures are different from their 2D counterparts as they show significant change with respect to varying viewing and lighting conditions. In order to isolate the effect of these conditions, we make use of the Hue-Saturation-Value (HSV) information of the pixels in the images of different materials. We reduce the number of features by clustering the pixel information using the k-means algorithm. Classification is performed by clustering the test image and finding the similarity between the clusters of the test image and the training images of different materials under different conditions. In the experiments, different image features such as large-scale filter bank responses, or grayscale pixel intensities are compared with the HSV information.

Keywords: Texture recognition, texture classification, k-means clustering, HSV color space

1 Introduction

This paper is about the classification of surfaces that are made of different material types. Texture is a discriminative property of a material and can be learned from the images of the material photographed under different imaging conditions. Textured surfaces have a variety of color and reflectance properties, as well as local height variations. Such textures are called 3D textures due to the change in their appearance with respect to varying viewing and lighting conditions. Therefore, a single image of a texture is not sufficient to give enough information about the material type. This fact can be observed in Figure 1, where five images of the same texture, which are significantly different from each other, are presented.

In the past, researchers addressed texture recognition as a 2D problem, ignoring the viewpoint and illumination changes [7, 8]. Then, a significant progress has been made by Leung and Malik [6], who defined the concept of a 3D texton. The term "texton", which is the unit of the human texture perception, was first introduced by Julesz [4]. Then, Leung and Malik further extended this idea to three dimensions. Later, additional research on 3D textures has been done [1, 2, 3, 4, 11].

In most of the studies, filter response distributions are learnt from training images. The distributions are either represented by clusters or histograms. Only in a recent study [12] the necessity of filter banks when forming the model distributions is questioned and it is demonstrated that, instead of the response to large-scale filter banks, taking only local pixel neighborhood distributions performs better. It has been shown that textures with global structures far larger than the local neighborhoods can be classified by a distribution of local measurements.

The main purpose of this study is to classify a material given a single image of a texture and while doing this, being as time-efficient as possible. The learning strategy is based on the method presented by Leung and Malik [6], which uses 3D textons. In our study, we show that taking the hue-saturation-value (HSV) distributions of pixels show better classification accuracy compared to other methods. In taking the local neighborhoods, the dimension of the feature space increases with the increase in the scale of the neighborhood. In addition, applying large-scale filter banks (e.g. of size 49x49) increase the computational complexity. Compared to these methods, HSV values can be obtained more efficiently. In addition, the experimental results show that they perform more accurately in the classification.

The organization of the paper is as follows: Section 2 elaborates on the method and gives the algorithms adopted. Section 3 presents the experimental results. Finally, Section 4 gives conclusions.

Fig. 1. The same material (terrycloth) photographed under different viewing directions and illumination conditions.

2 Method

The method in this paper consists of two phases: training and classification. Training step constructs the texton vocabulary similar to [6] but using HSV values of pixels instead of large-scale filter bank responses. The main purpose of the classification step is to achieve time-efficiency along with a high accuracy.

2.1 Feature Selection and Training

When classifying 3D textures, the selection of features plays the most important role. Different materials can be represented by the pixel values of their images; however, this would be highly costly and unnecessary. For instance, an 80x80 image would present 6400 features. In order to reduce dimensionality, the repeating structure of textures can be exploited. For this purpose, we follow the approach in [6] and apply k-means clustering on the pixels. K-means clustering is a greedy algorithm that assigns data vectors to the nearest of the k centers and updates each cluster center as the mean of the data vectors belonging to that cluster until convergence.

At this point, another question arises, i.e. what kind of pixel information to use. The first method that comes into mind is to use the gray-level intensities of the pixels. However, this approach is insufficient to capture the 3D information. Another method, which is proved to be successful in the literature is applying certain filters on the image and acquiring the responses of these filters. Filter responses characterize the illumination and viewing direction in an image. On the other hand, we try to eliminate the effects of illumination and the viewing direction in this study. Here, HSV color space becomes useful. In order to understand how to utilize this information, we need to understand what HSV is. The HSV stands for the hue, saturation and value components of the color, where hue is the dominant color, saturation is the amount of white light mixed with the hue and value is the intensity of a color. The hue and saturation components are related to the color perception of human eye. The representation of the HSV space can be seen in Figure 2. The HSV color space is fundamentally different and more intuitive than the widely known RGB (red, green, blue) color space, since it separates the intensity information from the color information. This separation is useful in 3D pattern recognition since the lighting and shading artifacts are isolated to the intensity channel [10].Thus, when selecting the features, just using the hue and saturation information and ignoring the intensity value will help to eliminate the illumination changes on the textures. Figures 3, 4 and 5 show the cluster centers for three different types of materials under different lighting conditions. In the figures, k is taken as 5, namely there are 5 clusters. It can be observed from these figures that, the distribution of the cluster centers are similar for the same type of material despite the illumination changes, and very different for different types of materials.

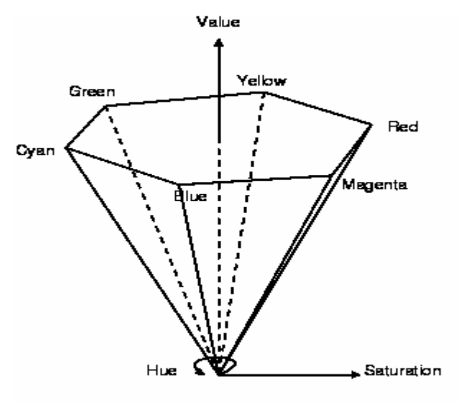
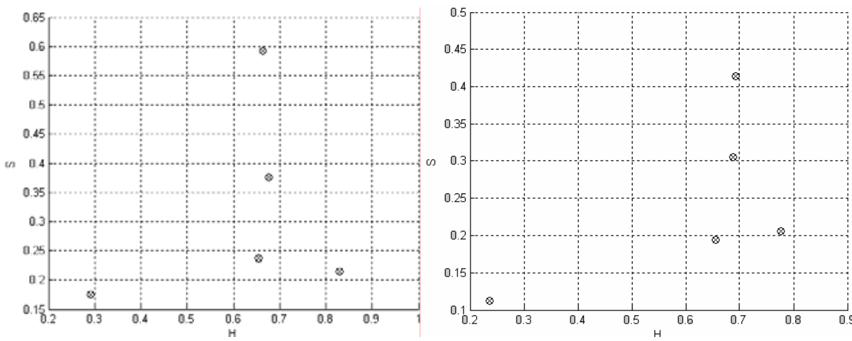


Fig. 2. The HSV color space

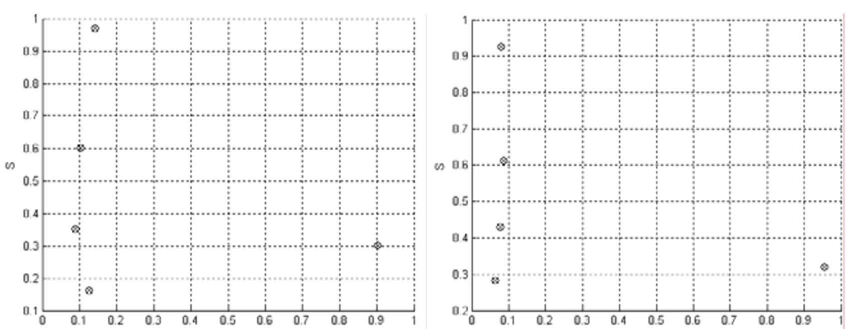
(a)



(b)

Fig 3. (a) The same material (felt) under different illumination conditions; (b) the distribution of cluster centers for the given images

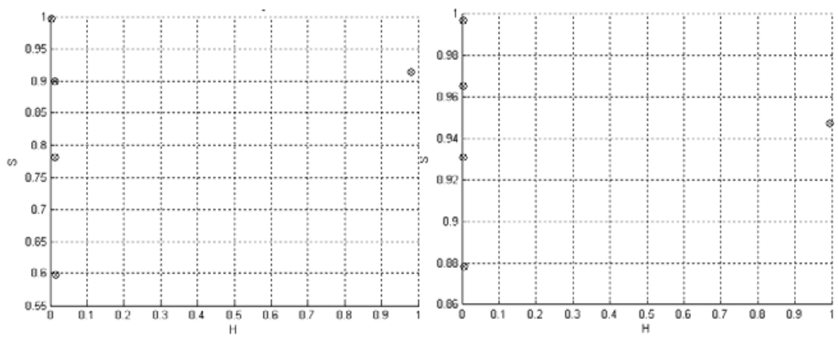
(a)



(b)

Fig 4. (a) The same material (pebbles) under different illumination conditions; (b) the distribution of cluster centers for the given images

(a)



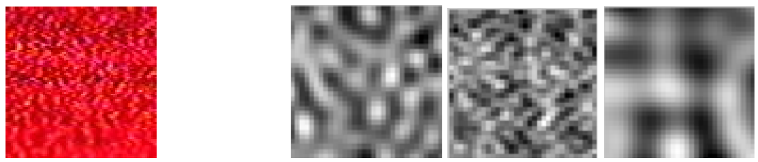
(b)

Fig 5. (a) The same material (velvet) under different illumination conditions; (b) the distribution of cluster centers for the given images

The training algorithm is as follows:

```
1. for each material set M
2.      for each image I ∈ M
3.          convert I from RGB to HSV
4.          concatenate the H and S values at each pixel
            to form a    vector of size 2  per pixel. Let
            the result be HSImg
5.          apply  k-means  clustering  algorithm  to  HSImg
            and  assign  the  clusters  to  {C1,….,Ck},  where
            each Ci is of size 1x2
```

This method is simpler than taking the filter responses or taking local pixel neighborhood distributions and requires less storage and computational cost. In order to compare the accuracy results, a large-scale (49x49) filter bank is applied to the images and the responses are taken as the features. For this purpose, the images are first converted to grayscale. The filter bank used is the Schmid set [11]. It contains 13 rotationally invariant, isotropic filters. Applying a filter on an image gives information about the local geometric and photometric properties of the surface. Thus, after representing each pixel by the 13 filter responses at that pixel, clustering the pixels gives the dominant features in the image under all lighting and viewing conditions. Figure 6 shows the filter responses of a velvet surface.



(a) (b)

Fig. 6. (a) Input texture (velvet); (b) filter responses of the input texture

2.2 Classification

Most of the classification methods in the literature are based on the construction of a model for each material to be classified. The model construction task involves finding the texton distribution of each image, which is performed by finding the most similar texton label for each pixel in the image. Thus, texton histograms are constructed for each image.

However, a simpler approach is followed in this paper. First of all, the image to be classified is clustered similarly as in the training procedure. Then, each cluster center of the test image is checked against the cluster labels and classified accordingly. The testing algorithm is as follows:

```
Most_Similar_Material(I_T ∈ Test set)

  min = +∞
for each cluster S_L ∈ Training set
     S_T = k-means(I_T)
     dist = Minimum_Distance_Between_Clusters(S_T,S_L)
     if dist < min
          min = dist
          materialIndex = Index_Of(S_L)
return materialIndex
```

The problem is to find the minimum distance between two data sets. For instance, we are given two sets $S_1 = \{ v_1, v_2,\ldots,v_n\}$ and $S_2 = \{w_1, w_2,\ldots,w_n\}$ whose elements are composed of vectors of a specific size. The difference between these two sets is taken as the Euclidean distance as:

$$\Delta S_1S_2= (\| v_1-w_1\|^2 +\| v_2-w_2\|^2 +\ldots\| v_n-w_n\|^2)^{0.5}.$$

We should find such an ordering of these sets that the difference is the minimum. For this purpose, we perform a greedy $O(n^2)$ time algorithm, which first takes a vector v_i from set S_1, finds the closest vector w_j to it in S_2, deletes w_j from S_2, and performs the same operations on the remaining vectors in S_1 and S_2.The pseudocode for the algorithm is given as:

```
Minimum_Distance_Between_Clusters(S1,S2)

   1. Given two cluster sets S1 = { v1, v2,….,vn} and S2 =
      {w1, w2,…,wn},
   2. totalDist = 0
   3. for i=1 to n
   4.          Let wj be the closest vector to vi
   5.          d = || vi-wj ||^2
   6.          delete wj from S2
   7.          totalDist = totalDist + d
   8. return (totalDist)^0.5
```

3 Experiments

All the images in this paper are obtained from the Columbia-Utrecht dataset. The dataset consists of 61 materials, each with 205 images, photographed under differing lighting and viewing conditions. We have selected 10 materials, each with 20 images for training and 10 random images for testing. Testing is performed on the same 10 materials; each with 10 new images of different viewing and lighting conditions from the training set images. Thus, a total of 300 images have been used. In addition, we have done our experiments on the 80 x 80 regions cropped from these images.

In order to see how different feature sets perform, we have performed several experiments on the data. The experiments are done in Matlab 6.5. Clustering and RGB to HSV conversion are done by the built-in functions of the statistics and image processing toolboxes.

The recognition results for different types of training are given in Table 1 and Table 2, where the number of clusters is taken as 5 and 10 respectively. As the results indicate, the best results are obtained when the features are taken as the HSV information of the pixels. Filter responses and RGB information perform similarly, which is surprising since the RGB features normally fail to determine the color and intensity variations and come up with clusters that put neighboring pixels with similar color but different shading to different clusters [10].

Success rate drastically reduces when the grayscale image is directly clustered. Moreover, taking local neighborhoods of pixels does not work well either. This shows us that the information that each pixel carries is an important issue. In the method where color information is used, this is the hue and saturation components of each pixel. Similarly, the method where filter responses are used, this information is the local height variations, illumination changes etc. In fact, using filter responses has proved to be a successful method in the literature. This study shows us that color information, especially HSV, can also be used for classification.

The number of clusters also determines the accuracy of recognition. For instance, the accuracy increases from 88% to 92% when HSV information is used. However, after a certain number, the increase in clusters does not improve the performance, which is related to the number of the training instances.

Table 1. Recognition rates for feature sets with 5 clusters

Training method	% Correct classification on training data	% Correct classification on test data
Filter responses	100	72
3x3 local neighborhood	72	60
RGB information	96	80
Grayscale information	96	56
HSV information	96	88

Table 2. Recognition rates for different feature sets with 10 clusters

Training method	% Correct classification on training data	% Correct classification on test data
Filter responses	96	78
3x3 local neighborhood	80	56
RGB information	96	76
Grayscale information	76	48
HSV information	100	92

5 Conclusion

In this paper, a method for the classification and recognition of three dimensional textures is proposed. The method uses the hue and saturation information for each pixel, and reduces the number of features by means of the k-means clustering algorithm. Classification includes clustering the test image and finding the similarity between the clusters of the test image and the training images of different materials under different conditions. The experimental results show that using the HSV color space for feature selection performs better than using other information.

The results about the HSV space comply with our initial hypothesis. However, as future work, it should be further investigated why using the RGB information performs the same with the filter responses and why the local neighborhood distributions perform poorly.

6 References

1. Cula, O. G. and Dana, K. J. 2004, 3D texture recognition using bidirectional feature histograms'. *International Journal of Computer Vision* 59(1).
2. Dana, K. J. and Nayar, S. 1998, histogram model for 3d textures. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 618-624.

3. Dana, K. J. and Nayar, S. 1999, correlation model for 3D texture. In: Proceedings of the International Conference on Computer Vision. pp. 1061-1067.
4. Dana, K. J., van Ginneken, B., Nayar, S. K. and Koenderink, J. J. 1999, reflectance and texture of real world surfaces'. *ACM Transactions on Graphics* 18(1), 1-34.
5. Julesz, B. 1981. Textons, the elements of texture perception, and their interactions. *Nature*, 290(5802):91–97.
6. Leung, T. and Malik, J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *IJCV*, 43(1):29–44, 2001.
7. Malik, J. and Perona, P. 1990. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7(5):923–932.
8. Rubner, Y. and Tomasi, C. 1999. Texture-based image retrieval without segmentation. In *Proceedings IEEE 7th International Conference on Computer Vision*, Vol. 2. Corfu, Greece, pp. 1018–1024.
9. Schmid, C. 2001. Constructing models for content-based image retrieval. In: *Proceedings of CVPR*.
10. Sural S., Gang, Q. and Pramanik, S. 2002. Segmentation and histogram generation using the HSV color space for image retrieval. *IEEE International Conference on Image Processing*,.Vol. 2., pp. 589-592.
11. Varma, M. and Zisserman, A. 2002. Classifying images of materials: achieving viewpoint and illumination independence. In: *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark,* Vol. 3. pp. 255-271, Springer-Verlag.
12. Varma, M. and Zisserman, A. 2003. Texture Classification: Are Filter Banks Necessary? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.