Using Three-Dimensional Features to Improve Terrain Classification *

Xiaoguang Wang, Frank Stolle, Howard Schultz, Edward M. Riseman, and Allen R. Hanson

Department of Computer Science
University of Massachusetts, Amherst, MA. 01003-4610
Email: {xwang, stolle, hschultz, riseman, hanson} @cs.umass.edu

Abstract

Texture has long been regarded as spatial distributions of gray-level variation, and texture analysis has generally been confined to the 2-D image domain. Introducing the concept of "3-D world texture", this paper considers texture as a function of 3-D structures and proposes a set of "3-D textural features". The proposed 3-D features appear to have a great potential in terrain classification. Experiments have been carried out to compare the 3-D features with a popular traditional 2-D feature set. The results show that the 3-D features significantly outperform the 2-D features in terms of classification accuracy.

1 Introduction

Texture analysis is an important area in computer vision and has been extensively studied (e.g. [1, 2, 3, 4]). While it is widely accepted that *texture* has no formal and precise definition, it is generally presumed to be a spatial distributions of gray-level variations, or regular structural "patterns", in the image. This presumption has dominated texture analysis for decades. Tuceryan and Jain [4] identified four basic approaches to texture analysis: statistical, geometrical, model-based, and signal processing methods. Although the methodologies vary substantially from one algorithm to another, they generally assume that texture is a part of the 2-D image and analysis is performed in image space. Typically, the gray-level patterns are characterized as textural features that are extracted in a localized image region.

The limitation of this traditional approach is that the 3-D properties of real world textures are not used directly. Consider the concepts *3-D world texture* and *2-D image texture* that we will use in this paper. World textures are reoccurring patterns caused by physical coarseness, roughness, and other characteristics on surfaces of objects in the real world, and generally have 3-D structures. For example, forests and grass cover possess different roughness because they have different 3-D structures. On the other hand, image textures are 2-D optical patterns in the image, reflecting perspective projection of object surfaces and world textures under particular sensor and illumination configurations. Attempting to characterize image textures by image features does not focus upon the underlying physical properties of the surfaces or objects creating the texture. All the feature extraction techniques in literature [2, 3, 4] are based on the concept of image texture, attempting to describe and discriminate 2-D gray-level patterns. We call them *2-D features.* In contrast, a feature is called a *3-D feature* if it measures some 3-D structural characteristics of a world texture. A significant amount of 3-D structural information relating to world texture can be obtained from multiple views of an object.

Consider the pair of oblique views of rural terrain of Ft. Hood, Texas, shown in Fig. 1(a). They are 2k×2k images taken with an aerial survey camera from an altitude of 2600m at tilts of 53° and 36°, respectively, off of vertical (nadir), with a camera separation of 3998m. (b) and (c) show the details of two small areas in the terrain after an epipolar resampling from the source images. Some significant differences occur between the image patches in (b) from the two views. Because the trees are off the ground plane, different parts of the trees are seen from different viewpoints. In addition, the crowns of the trees stretch out and occlude parts of the ground when seen from an oblique angle. However, the ground and road in (c) look quite similar in the two views, because they are almost flat, producing little perspective distortion. Generally speaking, a complicated (i.e. highly varying) 3-D structure such as vegetation tends to cause more variations in the different views than simple objects such as roads with simpler 3-D structures (i.e. less variation in depth). Hence, a measure of similarity in the image patches of the same location from different views reveals some information of the texture types at the location, and potentially can be used as a feature for texture classification. In summary, this information is directly a function of the 3-D variation in the scene.

We develop a set of 3-D textural features in Section 2. A set of experiments (Section 3 and 4) has been designed to compare the performance of the proposed 3-D features and a set of well-known 2-D features. In Section 5 we discuss future work.

2 Three Dimensional Features

In this section we propose a set of 3-D textural features that will be used in terrain classification. A de-

*This work was funded under DARPA/Army TEC contract DACA76-92-C-0041, DARPA/ARL contract DAAL02-91-K-0047, NSF grant CDA-8922572, and ARO DURIP grant DAAH04-95-1-0068.

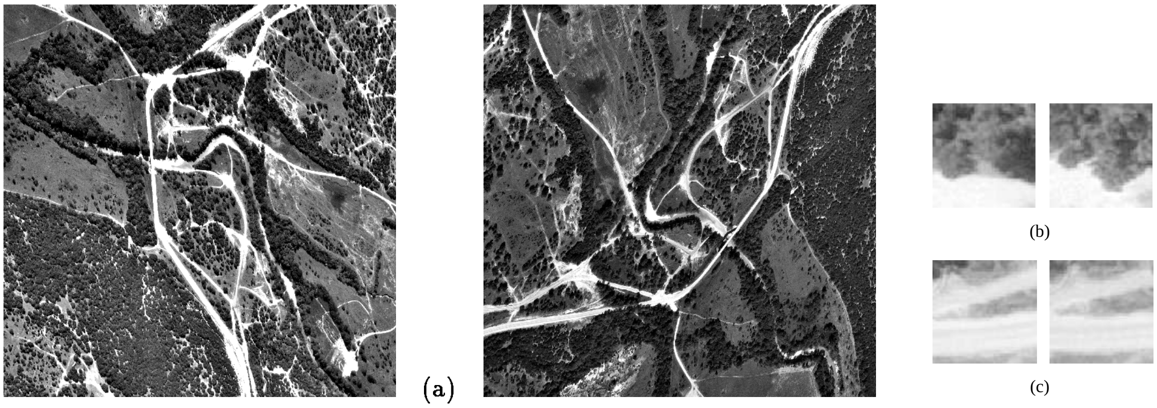

Figure 1: Sample source images
(a) two original source images (b)(c) epipolar resamplings of two small areas from the two original images

tailed description for the computation of these features can be found in [7].

2.1 Similarity function

The proposed 3-D features are derived from a *similarity function*, ρ, using a multi-view stereo terrain reconstruction algorithm [6]. Input to this algorithm are two aerial or satellite images of the same region. Using known intrinsic and extrinsic camera parameters, the image pair is resampled to satisfy the relationship $I^F(i,j) \cong I^G(i + D(i,j), j)$, where $D(i,j)$ is the disparity matrix. Under the assumption of a nearly Lambertian surface, a patch on the surface has the same apparent brightness pattern when seen from different views. The similarity function $\rho_{ij}(d)$ is computed between a window centered at (i,j) in I^F and a series of windows centered at $(i+d, j)$ in I^G, where d in the predefined search interval $[d_{\min}, d_{\max}]$ is incremented in sub-pixel steps. Using the weighted central moments

$$\mu_A(i,j) = \sum_{\eta=-n}^{n} \sum_{\xi=-m}^{m} \frac{I(i+\xi, j+\eta)A(\xi,\eta)}{(2m+1)(2n+1)}, \quad (1)$$

$$\sigma_A(i,j) = \quad (2)$$
$$\left(\sum_{\eta=-n}^{n} \sum_{\xi=-m}^{m} [I(i+\xi, j+\eta) - \mu_A(i,j)]^2 A^2(\xi,\eta) \right)^{1/2},$$

the similarity function is defined as

$$\rho_{ij}(d) = \quad (3)$$
$$\sum_{\eta=-n}^{n} \sum_{\xi=-m}^{m} \frac{I^F(i+\xi, j+\eta) I^G(i+d+\xi, j+\eta) A^2(\xi,\eta)}{\sigma_A^F(i,j)\sigma_A^G(i+d,j)}$$
$$- \frac{((2m+1)(2n+1))\mu_A^F(i,j)\mu_A^G(i+d,j)}{\sigma_A^F(i,j)\sigma_A^G(i+d,j)},$$

where A is a center-weighted Gaussian matrix, making the similarity function more reliable.

A smooth function $\tilde{\rho}_{ij}$ is fit to the discrete similarity function $\rho_{ij}(d)$. At position $\hat{d}$, where $\tilde{\rho}_{ij}(\hat{d}) = \max\{\tilde{\rho}_{ij}(d)\}$, the window centered at $I^F(i,j)$ finds its best match in I^G. Thus, $\hat{d}$ is taken as the estimate for the disparity between $I^F(i,j)$ and the corresponding pixel in I^G, and is stored in matrix D. The algorithm utilizes an iterative low-to-high resolution method to improve stability. D is used as an estimate at the beginning of each iteration, and refined by warping I^G according to the estimate in D. Using the final disparity matrix $\hat{D}$, an orthographic version of the intensity image, as shown in Fig. 2(a), can be generated.

2.2 Generating 3-D features

The 3-D features are generated from the smoothed similarity function. First, the maximum value of the function is set to be the *match score* (MS) at pixel $I^F(i,j)$, that is, MS $= \tilde{\rho}_{ij}(\hat{d})$. MS tends to be high when the image patches being correlated are similar. Taking into account perspective distortion, the highest degree of similarity can be achieved when the image patches are smooth and there is no occlusion due to change of viewpoint. In the case of flat ground, sufficient texture usually exists and changes in viewpoint are unlikely to cause occlusion as the 3-D surface structures are very small; hence, MS naturally tends to be high. In the case of foliage areas, changes in viewpoint will often cause occlusion, and MS tends to be lower.

Curvature of similarity function (CSF) is a feature describing the distinctiveness of the match between window patches from two views. It is computed by fitting a parabola of the form ad^2+bd+c to the smoothed similarity function $\tilde{\rho}_{ij}(d), d \in [\hat{D}(i,j) - o, \hat{D}(i,j) + o]$, where o is a parameter that determines a range of statistical significance. In the experiments o was set to a value of 1.5 pixels. The curvature of the parabola, defined to be CSF $= 2a$, provides an estimate of the curvature at the peak of $\tilde{\rho}_{ij}(d)$. Typically CSF is a negative value. The absolute value of CSF tends to be high when distinctive features or texture with small but flat structures exist, i.e. there is a unique match. Texture with complicated 3-D structures such as foliage, or texture with little distinctive features such as plain ground, usually have lower values of $|\mathrm{CSF}|$.

The parabolic approximation fails if the peak of the parabola falls outside the search range or CSF has a positive value, and the curvature at this position is un-

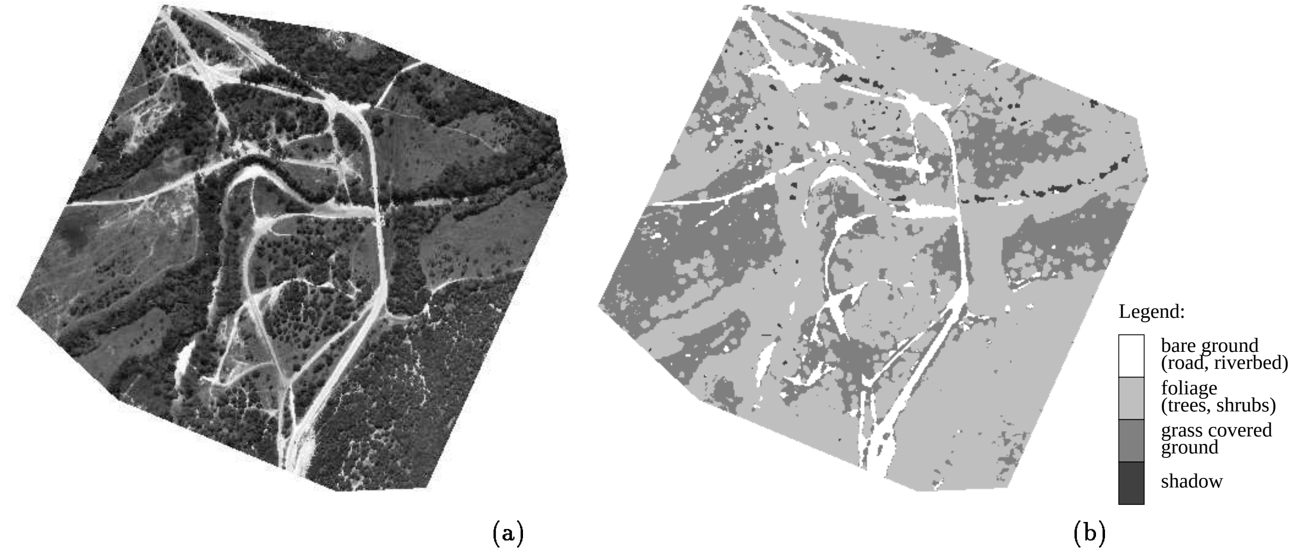

Figure 2: Orthographic intensity image and its classification
(a) the 2k×2k ortho-image from Ft. Hood Image Set, (b) classification using a combination of co-occurrence, 3-D, and intensity features (classifier trained by the four small image chips in Fig. 4)

defined, indicating a bad match. (The CSF at these positions can be mended using a median filter.) The *neighborhood density of well-defined curvature* (NDC) contains information about the reliability of the match. NDC is low in an area that contains less distinguishable texture, or where significant occlusion is present.

The final measure is the *neighborhood variance of match score* (NVMS), computed by the second central moment of the two-dimensional match score in a local window. Clearly, NVMS is affected by the 3-D structure of the area under the window. For flat, textured ground such as roads or bare ground NVMS tends to be low; if occlusion is present it will be relatively high.

We illustrate the four types of newly defined 3-D features in Fig. 3, using a portion of the ortho-image shown in Fig. 2(a) (its position shown in Fig.4). It contains various types of terrain textures. Features NVMS and NDC are computed in a local window of size 17×17. It is visually apparent that the various types of texture are distinguished by the features.

3 Experimental Set-up

As an initial attempt to test the capability of newly introduced 3-D features, we employ a linear discriminant method (the Foley-Sammon transform (FST) [8]) for classification. FST is considered an effective algorithm in terms of linear discriminant ability [7, 9, 11]. Using FST, the original features are projected into a lower dimensional *algebraic feature* space, which returns optimal discriminability for the classes. After some preliminary tests, we used a fixed number $r = 3$ as the dimensionality of the algebraic feature space in the experiments described in Section 4. We further modified the FST by using Mahalanobis distance instead of Euclidean distance.

The experiments are designed to compare the proposed 3-D features with traditional 2-D features in their performance in terrain classification tasks. We chose the *co-occurrence features* [10] as the comparison 2-D feature set, which was claimed [11, 12, 13, 14] as a highly effective feature set for classification. Previous studies [7] suggest the use of features ASM (the angular second-moment), CON (the contrast), and ENT (the entropy). Based on our preliminary experimental results, we applied the following settings in our experiments: image window size 17×17, distance $d = 3$, four angles $\theta = 0°, 45°, 90°, 135°$, and the number of gray levels $G = 8$. Under these settings, a total of twelve features were employed (i.e. three types of features, each with four angles). The definition of ASM, CON and ENT, as well as the parameters, can be found in the literature [14].

The experiments have been carried out on a set of aerial images of Ft. Hood. Fig. 2(a) shows a 2k×2k image from an orthographic projection. Four object classes are considered: foliage (trees, shrubs), grass covered ground, bare ground (road, riverbed), and shadow. The 2k×2k ortho-image has been independently hand-labeled into the four classes to act as ground truth data in the experiments. A portion of the ground truth segmentation is shown in Fig. 3(b) (legend shown in Fig. 2). To reflect the reality that a large amount of training data is not easily available in practice, we use very small training data sets to design classifiers and observe their performance. One of the training data sets is shown in Fig. 4, randomly sampled from the ortho-image. It includes four small image chips, each containing a texture of a particular class. The sizes of the four image chips are: 99×99 (foliage), 75×75 (grass covered ground), 37×37 (bare ground), and 11×11 (shadow). They occupy only about 1% of the pixels in the entire ortho-image.

The features used in the experiments consist of twelve 2-D co-occurrence features, four 3-D features (MS, CSF, NVMS, NDC), and the original intensity as an additional feature. For the purpose of comparison, four classifiers are designed, each using a different set of features, denoted as *Feature Set A, B, C,* and *D.*

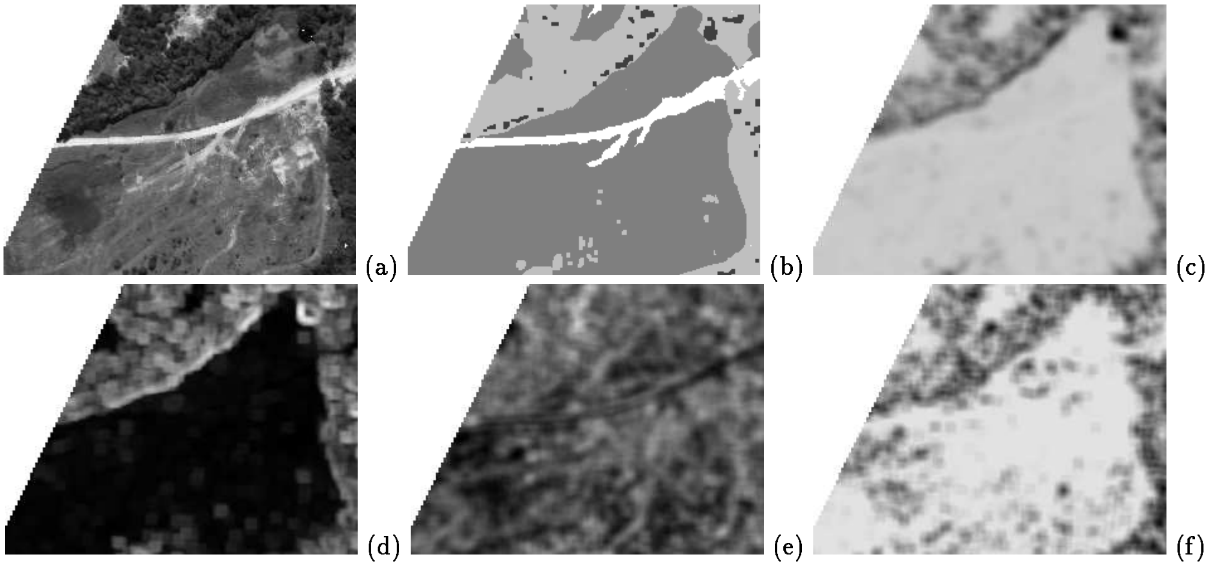

Figure 3: A portion of the 2k×2k ortho-image with a set of 3-D feature maps
(a) the intensity ortho-image, (b) ground truth hand classification, (c) match score, (d) neighborhood variation of match score, (e) curvature of similarity function, (f) neighborhood density of well-defined curvature

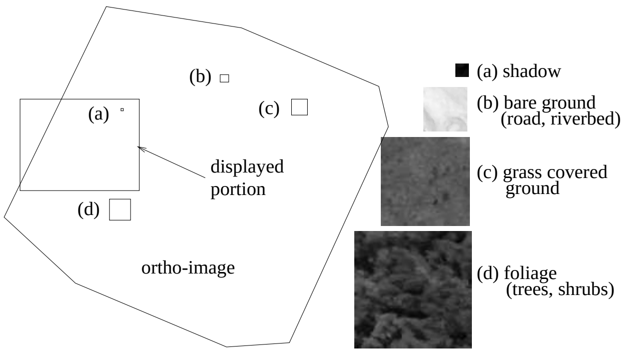

Figure 4: Positions of the training image chips and the displayed portion in Fig. 3 and 5

Set A contains the twelve co-occurrence features only. Set B adds the intensity feature into A. Set C consists of the four 3-D features plus the intensity feature. Set D includes all the twelve co-occurrence features, the four 3-D features and the intensity. All four classifiers employ the same classification algorithm based on Foley-Sammon transform and minimum Mahalanobis distance criterion.

4 Experimental Results

We test the four classifiers using Feature Set A, B, C, and D, to compare their classification accuracy. The entire ortho-image was classified based on a series of training data sets. In this section we only show the results based on the four image chips in Fig. 4. As shown in [7], the use of other training data sets yields similar experimental results.

4.1 Qualitative analysis

Fig. 5 shows the classification results on the image portion of Fig. 3(a). The first observation for these results is that the classifier in (c) using only the 2-D co-occurrence features produces the worst classification result. In particular, the ground and foliage labels are misclassified for each other in many places. An explanation of this phenomenon is that the ground category is affected by the existence of grass and vehicle tracks intermingled with ground patches so that it has a mottled 2-D textural appearance. Thus, the 2-D features find the ground regions in some cases with rough texture producing a variation of the 2-D intensity to be similar to foliage.

In contrast, since the presence of grass and vehicle tracks does not change 3-D smoothness of ground surface noticeably, they have little impact on the values of the 3-D features. For example, two image patches with the same vehicle tracks will still have a high match score because the variations are distinctive as opposed to patches of foliage texture which are far less distinct. Therefore, classifiers using 3-D features in (e) and (f) have a greater ability to discriminate the smooth ground surface from the rough forest surface.

Intensity, as a feature, provides some discriminability among the classes, especially the road and shadow classes which have extreme intensities. Thus, there is some improvement in (d) when the intensity feature is added to the co-occurrence features. However, because ground cover may have variations in brightness due to some variations in surface slope and soil/grass type, the intensity feature is not fully reliable. For example, in the left part of (a), the ground regions below the road turn darker, and classification deteriorates in these areas in (d). However, once again these regions receive satisfactory classification in (f) when 3-D features are added, since the darker ground remains smooth (i.e. has little change in 3-D variation), allowing 3-D features to perform effectively.

The classification result on the entire 2k×2k ortho-image using Feature Set D is shown in Fig. 2(b).

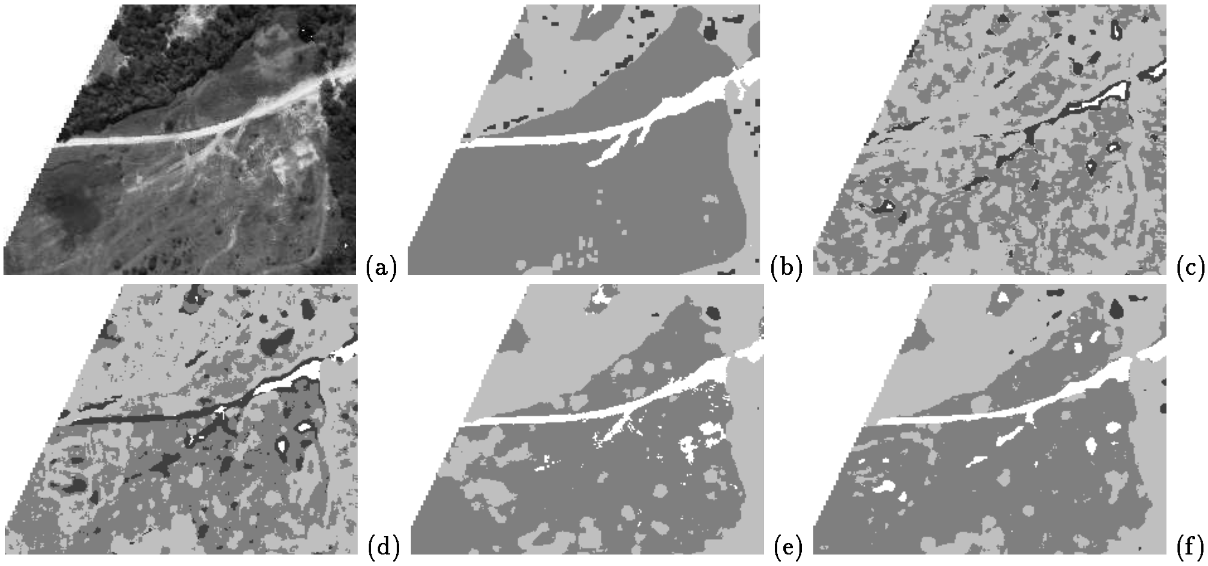

Figure 5: Classification results using various feature sets (see Fig. 2 for legend)
(a)(b) the intensity ortho-image and ground truth hand classification
(c) Feature Set A: twelve co-ocurrence features
(d) Feature Set B: twelve co-ocurrence features and one intensity feature
(e) Feature Set C: four 3-D features and one intensity feature
(f) Feature Set D: twelve coocurrence features, four 3-D features, and one intensity feature

4.2 Quantitative analysis

A quantitative analysis of the classification accuracy is shown in Table 1. The statistics of all the four experiments using different feature sets are given in the table. For each experiment, the entry (i, j) of the table shows the number of pixels that belong to Class i while being classified into Class j. The elements on the diagonals are the numbers of pixels that are correctly classified. The overall accuracy of each classification experiment is also given in the table (obtained by summing the diagonal elements and dividing by the total number of testing pixels).

Table 1 shows that Feature Set A provides the lowest classification accuracy. The additional intensity feature in Set B produces a minor improvement. The use of the 3-D features and intensity of Set C improves the classification dramatically over Set A – about 20 percentage points. Set D of all features only shows marginal improvement over Set C. In other words, if the four 3-D features and the intensity are used, then the participation of the 2-D co-occurrence features will hardly improve the classification accuracy.

A further observation from Table 1 is that major false classifications using Feature Set C and D fall into two cases: (1) ground truth shadow pixels being classified into foliage, or (2) ground truth grass pixels being classified into foliage. Case (1) is easy to understand. Shadow is usually an inherent component of foliage texture. Therefore, shadow in small pieces tends to be labeled as foliage by the classifier. In Fig. 2 we can see that large and continuously shadowed areas (e.g. in the upper right part) have been correctly classified. The mis-classification in case (2) is related to the 3-D structures of trees. In Fig. 1(a) we have seen that the crown of a tree is stretched out if seen from an oblique view. If a tree is observed from two different viewpoints, the projected crown will stretch out in two different directions, and will occlude different sections of ground. Suppose that a piece of ground appears in one source image, but is occluded by the stretch-out of tree crowns in the other. The 3-D features of this piece of ground is close to that of foliage (e.g. MS is low). The overall effect is that the tree seems to cover more ground area than it really does, and ground pixels near a foliage area tend to be mis-classified. This problem became worse in the test images, due to the existence of a lot of sparse trees. In Fig. 5(e)(f) we can see the dilations of the sparse tree areas. One method to solve this problem is to use morphological algorithms to erode the classified foliage areas, or areas of complicated 3-D structural textures. If some a priori knowledge, such as height of trees, can be obtained (e.g. via context-sensitive analysis of nearby tree regions), this problem could be solved in a knowledge-based manner. These are topics subject to future studies.

5 Conclusions and Discussions

We have provided a new perspective for understanding the nature of texture. Based on utilizing features for 2-D image texture and 3-D world texture, we extend our analysis to 3-D structural patterns on object surfaces in the real world. Methodologically, we have proposed a set of 3-D features that takes multiple views of 3-D objects into account for terrain texture classification in aerial images. Experimental results have shown that the set of new 3-D textural features significantly outperform the set of co-occurrence features, which is one of the best traditional 2-D feature sets. Additional experiments [7] show that the 3-D fea-

Table 1: Contingency analysis of classification results in the fundamental experiments (unit: 1000 pixels)

ground truth	(total)	classification result using Feature Set A				classification result using Feature Set B			
		shadow	grs. cvd. ground	foliage	bare ground	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	21.2	7.1	11.9	1.7	23.7	7.0	11.1	0.0
grs. cvd. ground	(683.0)	17.7	259.9	405.2	0.3	40.8	332.6	308.1	1.6
foliage	(1018.6)	10.8	121.2	886.6	0.0	11.7	55.0	950.7	1.2
bare ground	(193.4)	77.5	19.3	50.5	46.1	52.2	12.9	31.2	97.2
total	(1936.8)	127.2	407.5	1354.1	48.1	128.2	407.5	1301.1	100.0
		correctly classified pixel total: 1213.8				correctly classified pixel total: 1404.1			
		overall classification accuracy: 62.67%				overall classification accuracy: 72.50%			

ground truth	(total)	classification result using Feature Set C				classification result using Feature Set D			
		shadow	grs. cvd. ground	foliage	bare ground	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	3.0	0.0	38.8	0.0	13.8	0.0	27.8	0.2
grs. cvd. ground	(683.0)	0.0	439.5	231.4	12.2	0.0	468.6	202.7	11.8
foliage	(1018.6)	0.4	20.0	995.3	2.9	2.0	18.0	995.7	2.8
bare ground	(193.4)	0.0	17.3	25.1	150.9	0.0	33.9	21.4	138.1
total	(1936.8)	3.3	476.8	1290.6	166.0	15.8	520.6	1247.5	152.9
		correctly classified pixel total: 1588.7				correctly classified pixel total: 1616.1			
		overall classification accuracy: 82.03%				overall classification accuracy: 83.44%			

Feature Sets: A = {twelve co-occurrence features}; B = {twelve co-occurrence features and intensity}; C = {four 3-D features and intensity}; D = {twelve co-occurrence features, four 3-D features, and intensity}

tures are reliable to various training data, suggesting that they can be used in the circumstance that only a small amount of training data is available.

The motivation of this paper is to reveal the performance of the proposed 3-D feature set in comparison with the traditional 2-D features. To focus, we only pay attention to the raw results of pixel classification, and leave many issues that a real system must face as future topics, such as efficient use of the features and improvement of classification accuracy. These issues include designing sophisticated classifiers to deal with non-linear separability of feature space. Jain and Karu [5]'s recent work on learning schemes is a good candidates in this regard.

We are also investigating expansion of the number of object classes and the effect of this on classification accuracy. More terrain types (such as rocky ground and water surface) will be classified by using the 3-D features or a combination of 3-D and 2-D features.

References

[1] H. Tamura, S. Mori, and T. Yamawaki, "Textural Features Corresponding to Visual Perception," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 8, pp. 460-473, 1978.

[2] R. Haralick, "Statistical and Structural Approaches to Texture," *Pro. IEEE*, Vol. 67, pp. 786-804, 1979.

[3] L. Van Gool, P. Dewaele, and A. Oosterlinck, "Texture Analysis Anno 1983," *Computer Vision, Graphics, and Image Processing*, Vol. 29, pp. 336-357, 1985.

[4] M. Tuceryan and A. Jain, "Texture Analysis," In *The Handbook of Pattern Recognition and Computer Vision*, C. Chen, L. Pau, and P. Wang, eds. World Scientific Publishing Co., pp. 235-276, 1993.

[5] A. Jain and K. Karu, "Learning Texture Discrimination Masks," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 18, pp. 195-205, 1996.

[6] H. Schultz, "Terrain Reconstruction from Widely Separated Images," *Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II*, SPIE Proceedings Vol. 2486, pp. 113-123, Orlando, FL, 1995.

[7] X. Wang, F. Stolle, H. Schultz, E. Riseman, and A. Hanson, "A New Approach to Terrain Classification Using Three-Dimensional Features," Technical Report #97-021, Dept. of Computer Science, Univ. of Massachusetts at Amherst, April 1997.

[8] D. Foley and J. Sammon, Jr., "An Optimal Set of Discriminant Vectors," *IEEE Trans. on Computers*, Vol. 24, No. 3, pp. 281-289, 1975.

[9] K. Liu, Y. Cheng, and J. Yang, "Algebraic Feature Extraction for Image Recognition Based on an Optimal Discriminant Criterion," *Pattern Recognition*, Vol. 26, No. 6, pp. 903-911, 1993.

[10] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 3, No. 6, pp. 610-621, 1973.

[11] J. Weszka, C. Dyer, and A. Rosenfeld, "A Comparative Study of Texture Measures for Terrain Classification," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 6, No. 4, pp. 269-285, 1976.

[12] R. Conners and C. Harlow, "A Theoretical Comparison of Texture Algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 2, No. 3, pp. 204-222, 1980.

[13] J. du Buf, M. Kardan, and M. Spann, "Texture Feature Performance for Image Segmentation," *Pattern Recognition*, Vol. 23, No. 3/4, pp. 291-309, 1990.

[14] P. Ohanian and R. Dubes, "Performance Evaluation for Four Classes of Textural Features," *Pattern Recognition*, Vol. 25, No. 8, pp. 819-833, 1992.