Reliability of 3-D Textural Features in Terrain Classification *

Xiaoguang Wang and Allen R. Hanson

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

Abstract

3-D textural features are textural features that reveal some 3-D characteristics of the texture. In this paper we study the reliability and stability of a set of 3-D features proposed earlier [22, 23]. We emphasize the performance of the 3-D features in response to two types of training data, "labeling-based" and "chip-based." Experiments have been carried out to compare the 3-D features with a set of 2-D co-occurrence features under different types of training data. The results show that the 3-D features significantly and consistently outperform the 2-D features in terms of classification accuracy.

Keywords: texture analysis, terrain classification, 3-D textural feature, textural feature reliability, training data

*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 DAAG55-97-1-0026.

1 Introduction

Texture classification analysis is an important area in computer vision and has been extensively studied (e.g. [8, 10, 19, 20]). Traditionally, texture has always been presumed as a kind of spatial distributions of gray-level variations, or regular structural "patterns", in the image. A very large number of texture analysis algorithms have been proposed (see surveys in [8, 18, 20]) based on this presumption. The limitation of these algorithms is that the 3-D structures existing in the textures on object surfaces cannot be used directly.

Recently, both Dana et al. [3] and Wang et al. [22, 23] have used the concept of *3-D texture*. 3-D textures are recurring patterns caused by physical coarseness, roughness, and other characteristics on object surfaces in the real world. An important property of a 3-D texture is that its 2-D appearance varies when viewed from different positions, if there exist 3-D structures in the texture. Based on this observation, Dana et al. [3] established a texture database, in which images of object surfaces are taken from various angles to make a complete view of the textures. However, their system did not support texture classification analysis. Wang et al. [22, 23] proposed a set of *3-D features* for terrain classification. In contrast to traditional 2-D image features, which are extracted from a single image, these 3-D features are obtained by a multi-view analysis of the texture, and reflect some 3-D structural characteristics of the texture.

Some problems, however, exist in Wang et al.'s experiments. For example, they used a very small amount of data to train the classifier. In the reported result, these data accounted for only about 1% of the test data. In a large terrain that contains various types of ground covers (forests, grass, bare ground, etc.), such a small amount of training data might not reflect the reality of the entire test data. Generally, in classification problems, a classification result is dependent not only on the classification algorithm and the selected features but also on the data that are chosen to train the classifier. A classification result is reliable only if the

training data are sufficiently representative.

Training data selection is a practical issue in classification problems. Generally speaking, a larger training data set has a better representativity than a smaller one. However, small training data sets are often preferred since they involve less man-machine interaction. Good textural features are those that perform well not only under a large training data set but also in the case of small training data sets. This property is called "reliability" of the textural features. Using a reliable textural feature set, the interactive operations in the training stage can be minimized because only a small training data set is necessary.

In this paper we review the set of 3-D features (Section 2) and test its reliability under two types of training data, "chip-based" and "labeling-based" (Section 3). The experimental results, given in Section 4, show that the set of 3-D features outperforms a well-known 2-D feature set consistently under a wide collection of training data. We discuss future work in Section 5.

2 The Feature Space

In this section we review the 3-D features proposed by Wang et al. [22, 23]. They are MS (match score), CSF (curvature of similarity function), NVMS (neighborhood variation of match score), and NDC (neighborhood density of well-defined curvature). The 2-D features we use for comparison are ASM (angular second-moment), CON (contrast), and ENT (entropy), chosen from the co-occurrence feature set proposed by Haralick et al. [9].

2.1 The 3-D feature set

The 3-D features are derived from a two-view stereo algorithm [15], which has been used for terrain modeling [16, 17]. These features have been used for terrain classification recently [14, 22, 23].

The algorithm employs the epipolar geometry depicted in Fig. 1(a). For an arbitrary 2-D point, P, in the left image, it is correlated along its epipolar line on the right image, under the assumption that the terrain is a nearly Lambertian surface. Fig. 1(b) shows the *similarity function*, ρ, which indicates how P is correlated with the points on its epipolar line. The point Q with the best match – the highest correlation on the epipolar line – tends to be the true correspondence of P. The shape of the similarity function at Q reveals some important information of the 3-D texture at P, and thus is used to define the 3-D features.

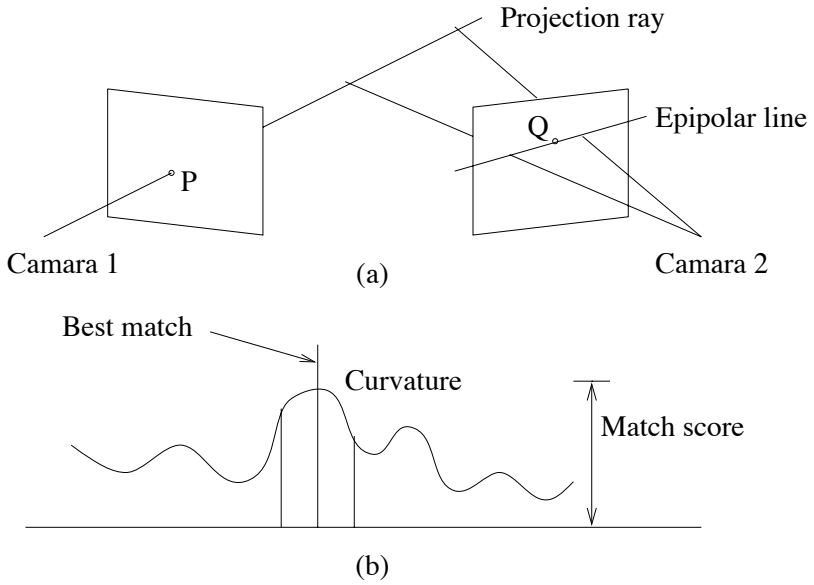

Figure 1: Epipolar geometry and correlation function

(a) epipolar geometry of an image pair

(b) similarity function of a 2-D point in the left image

The maximum value of similarity function at the best match is set to be the *match score (MS)* feature:

$$\mathrm{MS}(P) = \rho(Q). \tag{1}$$

MS tends to be high when the image patches being correlated are very similar. This happens

in smooth areas, such as flat ground, where there is no occlusion due to change of viewpoint between the two images.

The second feature, *neighborhood variance of match score (NVMS)*, is defined in a local window $W(P)$, centered at P, on the MS image:

$$\mathrm{NVMS}(P) = \sqrt{\frac{1}{N} \sum_{P' \in W(P)} [\mathrm{MS}(P') - E(W(P))]^2}, \tag{2}$$

in which N is the total number of pixels in the window and $E(W(P))$ is the average value of MS in the window. NVMS measures the homogeneity of MS in a piece of texture. In forested areas, for example, the occlusions caused by view point changes occur in a random way, and the values of MS may have a larger variance than grass areas (for example).

The distinctiveness of the similarity function at the best match is described by the feature *curvature of similarity function (CSF)*, which is computed by fitting a parabola of the form $ax^2 + bx + c$ to the function ρ at the best match. The curvature of the parabola

$$\mathrm{CSF}(P) = 2a, \tag{3}$$

provides an estimate of the curvature at the peak of the similarity function. The absolute value of CSF tends to be high when distinctive features or structures exist in the texture, i.e. there is a unique match.

At some points, MS and CSF might not have a well-defined value because there may not be a well-defined or unique local maximum over the certain search range in ρ. We fix the MS and CSF at these points by setting the value of MS to the global maximum in the search range and by filtering the CSF map using a median filter. A feature called *neighborhood density of well-defined curvature (NDC)* is defined to measure how densely these points occur, because missing a well-defined local maximum generally indicates a bad match and a (perhaps)

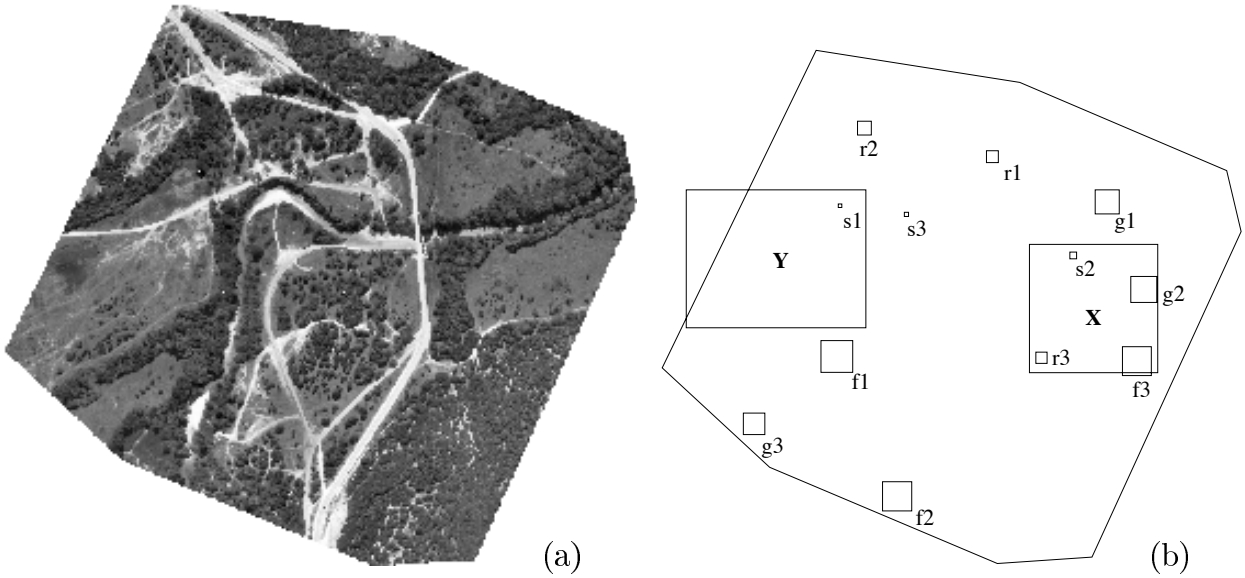

Figure 2: The orthographic intensity image and the locations of subimages used in the experiments

(a) the 2k × 2k ortho-image from Ft. Hood Image Set

(b) **Y**: the displayed subimage whose classification results are shown in Fig. 5 and Fig. 6

X: the hand-labeled subimage – training data in the comparison experiment

{f1,g1,r1,s1}: hand picked-up chips – training data in the comparison experiment

{f2,g2,r2,s2,f3,g3,r3,s3}: additional chips – training data in the stability experiment

complicated 3-D structure on the ground. In a local window $W(P)$ centered at P,

$$\mathrm{NDC}(P) = \frac{1}{N} \sum_{P' \in W(P)} \delta[CSF(P')], \tag{4}$$

where N is the total number of pixels in $W(P)$ and

$$\delta[CSF(P')] = \begin{cases} 1, & \text{if CSF is well-defined at } P' \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

In Fig. 3 we illustrate the 3-D features extracted from a portion of an area of Ft. Hood, TX. Fig. 2 shows the whole area and the location of the displayed portion $\mathbf{Y}$.

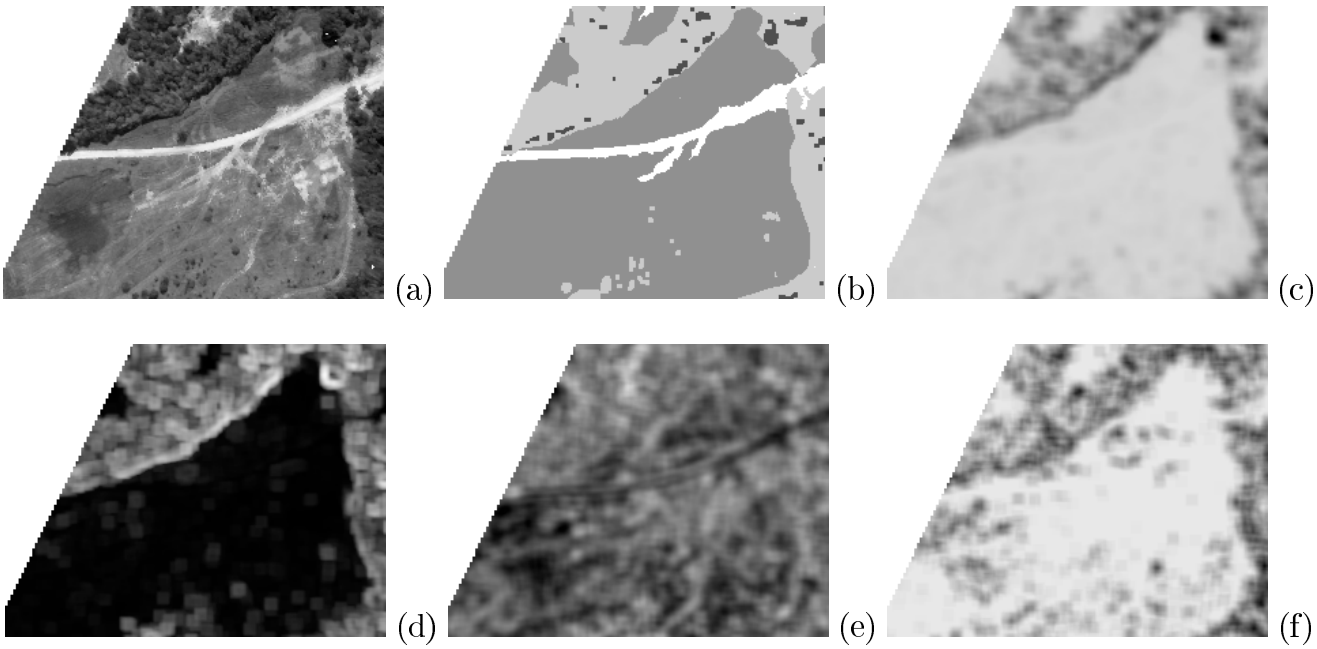

Figure 3: Subimage $\mathbf{Y}$ (560×430) and its 3-D feature maps

(a) the intensity ortho-image

(b) ground truth hand classification

(c) match score (MS)

(d) neighborhood variation of match score (NVMS)

(e) correlation curvature (CSF)

(f) neighborhood density of well-defined curvature (NDC)

2.2 The 2-D features for comparison

Co-occurrence features were introduced by Haralick et al. [9]. Previous theoretical and experimental studies [2, 5, 9, 13, 24] showed that co-occurrence features were more effective in terrain classification problems than many other well-known 2-D textural features, such as

the Fourier power spectrum, the gray-level run-length, the gray-level difference, the power spectrum, Markov Random Field parameters, multi-channel filtering features, and fractal based features, among others. Most recently the co-occurrence features have been studied by Valkealahti and Oja [21].

Co-occurrence features describe the texture within a local image window based on gray-tone spatial dependencies. Each feature type is associated with four separate features describing the gray-scale dependencies in four directions. As suggested by other researchers [2, 13, 23, 24] and by our preliminary experimental results, we use the following feature types: *angular second-moment (ASM)*, *contrast (CON)*, and *entropy (ENT)*. Thus a total of twelve features were employed. The definition of these features can be found in the literature [9, 13, 22].

3 The Training Data Space

In classification problems, training data selection is an important, yet often ignored, practical issue. Training data is the data, with known classification, used in the training stage to train the classifier in order to determine the parameters of the classifier (which are used in the later classification stage). The quality of a classification result is determined by the parameters of the classifier, hence by the training data.

A classification of an object x (e.g. a pixel in an image) is an assignment $\Gamma(x) = c_i$, where $x \in X$ (the object set) and $c_i \in \mathcal{C}$ (the class set). For the same classifier, when different feature sets and training data sets are used, it will output a different assignment. Let $\Gamma_{T,F}$ denote the classification assignment produced by the classifier using F as the feature set and $T \subseteq X$ as the training data set. The accuracy ρ of a classification can be computed based on the comparison between $\Gamma_{T,F}$ and the ground truth classification Γ_0:

$$\rho(T, F) = \frac{1}{|X|} \sum_{x \in X} \Delta[\Gamma_{T,F}(x), \Gamma_0(x)], \tag{6}$$

in which

$$\Delta[c_i, c_j] = \begin{cases} 1, & \text{if } c_i = c_j \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

ρ is a function of both the features set F and the training data set T.

As stated in Section 2, our goal is to compare the performance of the 2-D co-occurrence features and the new 3-D features. However, any experiment for comparing two feature sets, F_1 and F_2, must be conducted under a fixed training data set T; that is, any conclusion drawn from $\rho(T, F_1)$ and $\rho(T, F_2)$ is dependent on T. If T is not a good representation of the entire object set X, the result of the experiment is not reliable. Theoretically, a thorough understanding of the performance of two feature sets F_1 and F_2 needs a complete examination of all the subsets of X as training data. Measures can be defined to evaluate the results in these experiments:

$$\rho_M(F_k) = M\{\rho(T, F_k) | \text{ for all } T \subseteq X\}, \tag{8}$$

in which M is a measure to evaluate the ρ's in all the experiments. Although $\rho_M(F_k)$ thoroughly represents the performance of the feature set F_k on the object set X, its computation is practically impossible for it involves $O(2^{|X|})$ experiments.

In practice, there are only limited ways to choose T because of the limitation in interactive operations. In the training stage, a human operator must define a training data set T by choosing some portions of the image dataset and labeling them with the correct categories (i.e. specify $\Gamma_0(x)$ for each $x \in T$). There are two methods commonly used for choosing such a T. One method, *labeling-based*, is to select a subimage which contains pixels that cover all the $c_i \in \mathcal{C}$, and then label each pixel into a c_i by hand (e.g. Fig. 4(a)). The other method, *chip-based*, is to pick a set of small image chips, in which each chip contains only one c_i (e.g. Fig. 4(b)).

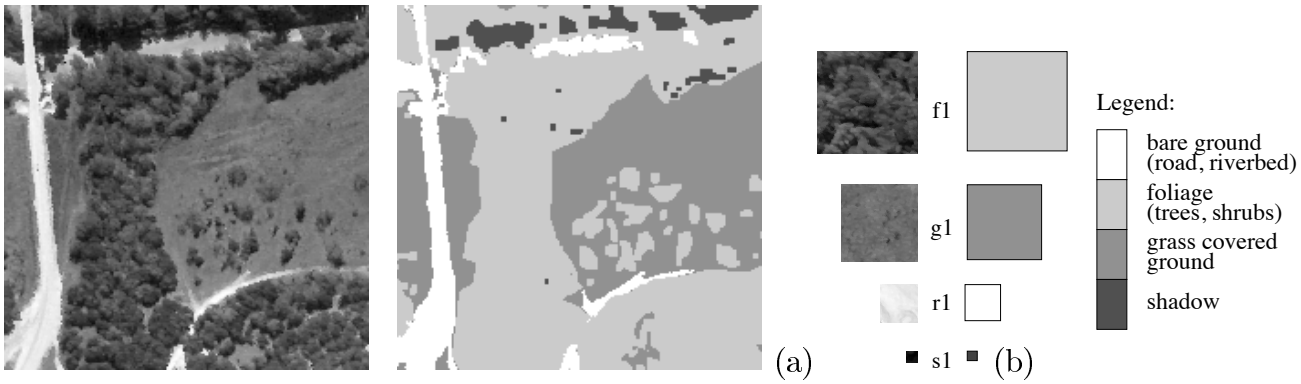

Figure 4: Training data (positions shown in Fig. 2(b))

(a) manually labeled subimage $\mathbf{X}$ (400×400)

(b) f1 (99×99, foliage), g1 (75×75, grass covered ground), r1 (37×37, road/riverbed), s1 (11×11, shadow))

In order to reduce interactive operations, the chip-based method is preferred. For example, consider the image in Fig. 2, which is to be classified into four classes: foliage(trees, shrubs), grass covered ground, bare ground (road, riverbed), and shadow. The operator only need to pick up minimally four small chips (e.g. the set {f1, g1, r1, s1}) as training data and label each with four class labels. On the other hand, if the labeling-based method is used, the operator has to label the whole subimage manually pixel by pixel (e.g. $\mathbf{X}$ in Fig. 2). The labeling-based training data, however, have the advantage that their distribution in the feature space tends to resemble the distribution over the whole set X. For instance, sometimes ambiguities exist around the boundary of two classes: it is hard to decide whether a pixel is a "foliage pixel" or a "shadow pixel". A subimage like $\mathbf{X}$ tends to include these pixels at the same rate as in the whole image. Th human operator is "forced" to make a decision on these pixels, and the decision is fed into the classifier as a part of the training data. In the case of chip-based method, the operator can escape from this hardship by avoiding those boundary pixels and choosing only those "easy" pixels with obvious texture types. As a result, the

distribution of training data T in the feature space might be significantly different from that of X.

To test the reliability of the 3-D and 2-D textural features, we will test their performances (Section 4.1) under both labeling-based and chip-based training data sets. Subimage **X** in Fig. 2, randomly selected and manually labeled, is used as the labeling-based training data. Image chips {f1, g1, r1, s1} are used as chip-based training data. Because the chip-based method is more desirable in real applications, we increase the number of image chips by including {f2, g2, r2, s2, f3, g3, r3, s3} (see Fig. 2(b)) in Section 4.2 to test the stability of the 3-D and 2-D features under different combinations of training data.

4 Experiments

Experiments have been carried out to compare the proposed 3-D features and the 2-D co-occurrence features under different training data. For each configuration of training data, the performance of the following four feature sets are tested: Feature Set A contains the twelve co-occurrence features only; Feature Set B includes the original image intensity as an additional feature into Set A; Feature Set C consists of the four 3-D features plus the intensity feature. Feature Set D includes all the features – the twelve co-occurrence features, the four 3-D features and the intensity. All the experiments employ the same classification algorithm based on the Foley-Sammon Transform (FST) [7, 12, 22, 24] and minimum Mahalanobis distance criterion. All the experiments are conducted on the 2k $\times$ 2k Ft. Hood image in Fig. 2(a). Using formula (6), the classification result in each case is compared with a hand produced Γ_0 that was manually generated by a human operator.

4.1 The comparison experiment

In the comparison experiment, we compare the performance of the feature sets under the labeling-based training data $\mathbf{X}$ and under the chip-based {f1, g1, r1, s1}. A portion (subimage $\mathbf{Y}$) of the classification results are displayed in Fig. 5 and Fig. 6, respectively. Table 1 and Table 2 shows the quantitative analysis of these results, each entry (i, j) of the table being the number of pixels classified into c_j when the ground truth was c_i.

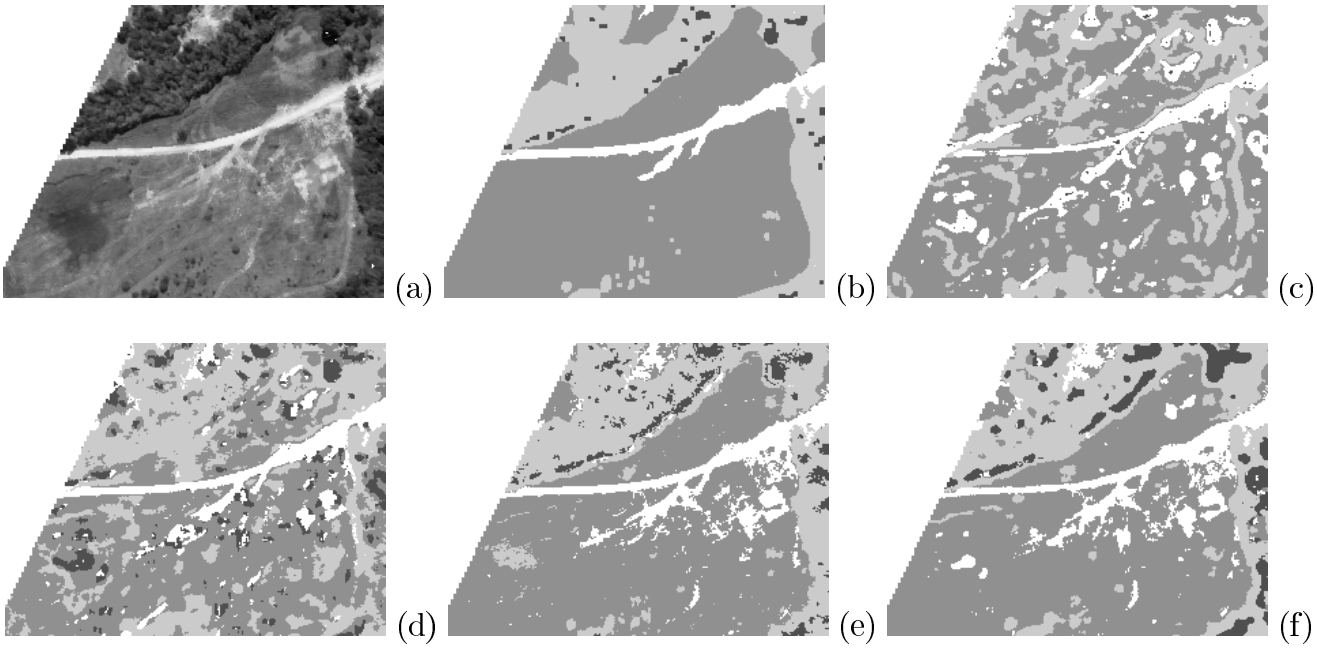

Figure 5: Classification results on subimage $\mathbf{Y}$ (560×430) using subimage $\mathbf{X}$ as training data (see Fig. 4 for legend)

(a) the intensity ortho-image

(b) ground truth hand classification

(c) Feature Set A: twelve co-occurrence features

(d) Feature Set B: twelve co-occurrence features and one intensity feature

(e) Feature Set C: four 3-D features and one intensity feature

(f) Feature Set D: twelve co-occurrence features, four 3-D features, and one intensity feature

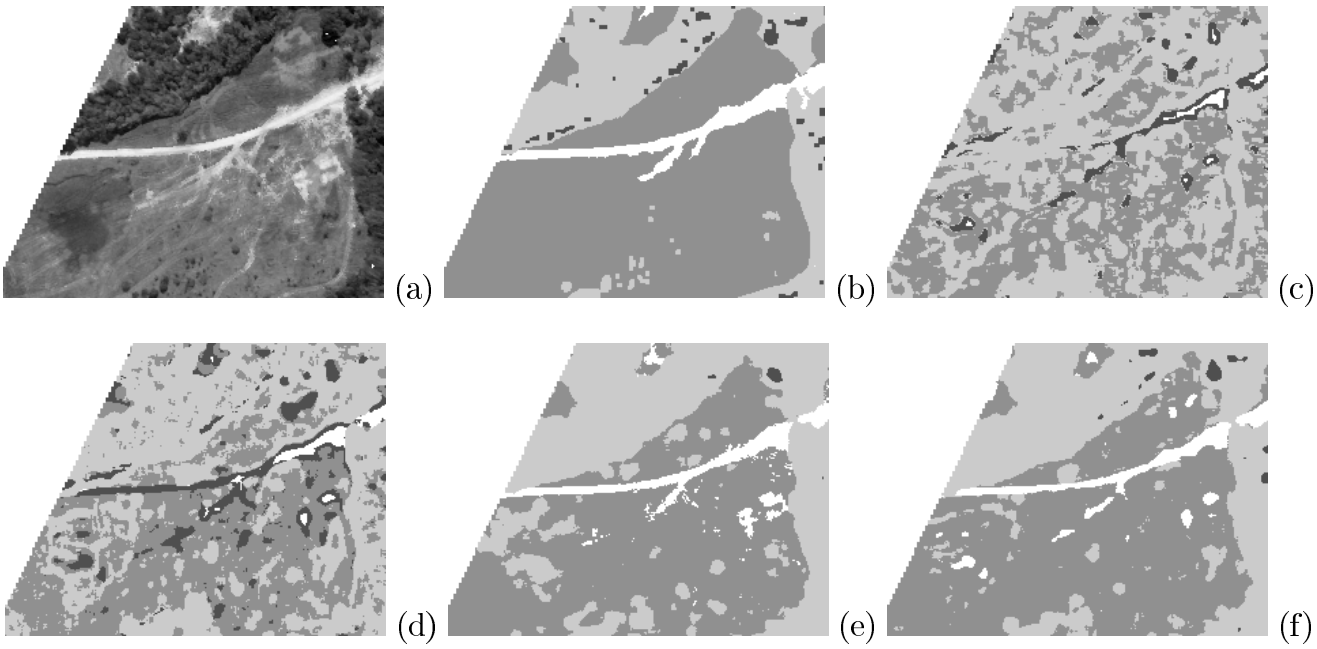

Figure 6: Classification results on subimage $\mathbf{Y}$ (560×430) using image chips {f1,g1,r1,s1} as training data (see Fig. 4 for legend)

(a) the intensity ortho-image

(b) ground truth hand classification

(c) Feature Set A: twelve co-occurrence features

(d) Feature Set B: twelve co-occurrence features and one intensity feature

(e) Feature Set C: four 3-D features and one intensity feature

(f) Feature Set D: twelve co-occurrence features, four 3-D features, and one intensity feature

Some observations can be made from both the visual results and the quantitative analysis. First, the feature set containing only the 2-D co-occurrence features (Fig. 5(c) and Fig. 6(c)) produces the worst classification result using either training data set. In particular, the ground and foliage labels are confused in many places, because the 2-D features are not able to reflect the 3-D structures in foliage areas, while the existence of grass and vehicle tracks intermingled with ground patches makes the ground area have a mottled 2-D textural appearance. Second,

the intensity feature provides some discriminability among the classes (Fig. 5(d) and Fig. 6(d)), especially the road and shadow classes which have extreme intensities. However, the intensity feature is not reliable in that the brightness of ground covers might change due to different soil/grass types. (In the lower left part of Fig. 5(a) and Fig. 6(a), the ground region below the road turns darker, and classification deteriorates in this area when the intensity feature is added into Feature Set A.) Third, when the four 3-D features are used (in Feature Sets C and D), the classification accuracy improves significantly (Fig. 5(e)(f) and Fig. 6(e)(f)). Under either training data set, the use of 3-D features increases the overall accuracy about 10-15 percentage points. In summary, the involvement of 3-D features consistently improves classification under the two types of training data.

Some differences in the classification can be observed between the two types of training data. Table 1 and Table 2 shows that label-based training data (subimage $\mathbf{X}$) leads to a better classification result in the categories of shadow and bare ground than the chip-based method does (using {f1,g1,r1,s1}), although the overall accuracy remains the same. This phenomenon can be explained by the difference of the feature space distribution of the two types of training data. As stated in Section 3, the distribution of chip-based training data in the feature space may be different from that of the test data, because the "boundary pixels" are usually not included. Therefore, the classifier may have a different "understanding" of these boundary pixels than the human operator does. Fig. 5(e)(f) and Fig. 6(e)(f) shows that the major differences between the two types of training data happen at those boundary shadow and bare ground pixels. Since these pixels are ambiguous pixels in nature, a mis-classification does not deteriorate the visual effect too much in Fig. 5(e)(f) and Fig. 6(e)(f).

4.2 The stability experiment

Since chip-based method for choosing training data is more desirable in real applications, a stability experiment has been carried out to test the stability of the 2-D and 3-D features using various combinations of image chips as training data. For each texture class, we included two more image chips randomly sampled from the ortho-image (Fig. 2), with sizes similar to those used in the comparison experiment. Twelve combinations (subsets of training chips) are tested in the experiments, and the results are shown in Table 3. Each Training Set 1, 2, and 3 contains all the training chips for the bare ground, grass covered ground, and shadow classes, but only one (different) chip for the class of foliage. Each Training Set 4, 5, and 6 contains one chip for the bare ground class and all the chips for other classes. Similarly, Training Set 7, 8, and 9 each contains one chip for grass covered ground and all the chips for the others, and Training Set 10, 11, and 12 each has one for shadow and all for the others.

For each combination of training data, the four feature sets, A, B, C, and D, are tested. From Table 3 we can see that, under any combination, the classifier using the 3-D features always performs better than Feature Set A and B, where no 3-D feature was involved. On average, Feature Set C and D outperforms Feature Set A and B by nearly 10-15 percentage points.

From the standard deviation we can see that the set with co-occurrence features plus the intensity is most sensitive to different training sets. Hence the quality of its classification is most unreliable. The sets with 3-D features have a good stability against various training data. This fact indicates that the proposed 3-D features represent some consistent physical characteristics of textures with 3-D structures, and can be considered as candidate features in real applications where a large amount of training data are not easily available.

5 Discussions

In order to evaluate the quality of a set of textural features, the performance of the features under different configurations of training data must be investigated. In this paper we have experimentally compared a set of proposed 3-D features with the set of 2-D co-occurrence features, which has been claimed to be one of the best traditional textural feature sets. Two most frequently used training data selection methods, labeling-based and chip-based, have been investigated. Experimental results have shown that the proposed 3-D features significantly and consistently outperform the co-occurrence features in response to the two types of training data. It is also shown that the 3-D features have a good stability over different pieces of training data, suggesting that they can be used in circumstances where only a small amount of training data is available.

A further inspection of Table 1 and Table 2 shows that a large portion of mis-classifications made by the 3-D features involves mistaking grass pixels for foliage pixels. Because of this confusion, the edges of forests almost always include some grass covered ground, as can be seen in Fig. 5(e)(f) and Fig. 6(e)(f). This is because the 3-D features are based on multi-view analysis, and various parts of grass covered ground may be occluded by the foliage when the viewpoint changes. As a result, a forest tends to have a larger area than it actually has when this juxtaposition of the two classes occurs in the image. This problem could be solved by using the knowledge of the height of the trees, and thus the 3-D feature based classification could be further improved.

Future studies include an investigation of the distributions of the terrain data in the space of the 3-D features. All the experiments conducted in this paper are based on the FST classifier, which performs the best when each class has a Gaussian distribution in the feature space. Since the distributions in the feature space are unknown, how the 3-D features would perform in other kinds of classifiers is an open question. We are currently seeking to test the

reliability and stability of the 3-D features using a decision tree classifier [1, 4, 14]. Classifiers based on neural networks (e.g [11]) are good choices as well.

References

[1] C. Brodley and P. Utgoff, "Multivariate Decision Trees," *Machine Learning*, 19, pp. 45-77, 1995.

[2] 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.

[3] K. Dana, S. Nayar, B. van Ginneken, and J. Koenderink, "Reflectance and Texture of Real-World Surfaces," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Puerto Rico, pp. 151-157, June 1997.

[4] B. Draper, C. Brodley, and P. Utgoff, "Goal-Directed Classification using Linear Machine Decision Trees," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 9, pp. 888-893, 1994.

[5] J. du Buf, M. Kardan, and M. Spann, "Texture Feature Performance for Image Segmentation," *Pattern Recognition*, Vol. 23, pp. 291-309, 1990.

[6] D. Dunn, W. Higgins, and J. Wakeley, "Texture Segmentation Using 2-D Gabor Elementary Functions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 2, pp. 130-149, 1994.

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

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

[9] 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.

[10] R. Haralick, "Statistical and Structural Approaches to Texture," *Proc. IEEE*, Vol. 67, No. 5, pp. 786-804, May 1979.

[11] A. Jain and K. Karu, "Learning Texture Discrimination Masks," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 2, pp. 195-205, 1996.

[12] 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.

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

[14] J. Piater, E. Riseman, and P. Utgoff, "Interactively Training Pixel Classifiers," to appear in *Proc. FLAIRS'98*, AAAI Press, 1998.

[15] 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, April 1995.

[16] H. Schultz, C. Jaynes, M. Marengoni, A. Schwickerath, F. Stolle, X. Wang, A. Hanson, and E. Riseman, "3D Reconstruction of Topographic Objects at the University of Massachusetts," *Joint ISPRS Commission III/IV Workshop: 3D Reconstruction and Mod-*

elling of Topographic Objects, E. Baltsavias et al. (Ed.), pp. 77-87, Stuttgart, Germany, September 1997.

[17] H. Schultz, F. Stolle, X. Wang, E. Riseman, and A. Hanson, "Recent Advances in 3D Reconstruction techniques Using Aerial Images," *Image Understanding Workshop*, pp. 977-982, New Orleans, LA, 1997.

[18] J. Strand and T. Taxt, "Local Frequency Features for Texture Classification," *Pattern Recognition*, Vol. 27, No. 10, pp. 1397-1406, 1994.

[19] 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.

[20] 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.

[21] K. Valkealahti and E. Oja, "Reduced Multidimensional Co-Occurrence Histograms in Texture Classification," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 90-94, 1998.

[22] 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.

[23] X. Wang, F. Stolle, H. Schultz, E. Riseman, and A. Hanson, "Using Three-Dimensional Features to Improve Terrain Classification," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Puerto Rico, pp. 915-920, June 1997.

[24] 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.

Table 1: Contingency analysis of classification results using subimage **X** as training data (unit: 1000 pixels)

		classification using Feature Set A			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	1.8	7.6	0.4	32.1
grass covered ground	(683.0)	0.9	391.5	219.7	71.0
foliage	(1018.6)	0.2	274.3	688.8	55.3
bare ground	(193.4)	3.5	20.6	9.7	159.6
total	(1936.8)	6.4	693.8	918.5	318.0
		correctly classified pixel total: 1241.7 (64.11%)			
		classification using Feature Set B			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	28.5	8.5	2.3	2.5
grass covered ground	(683.0)	37.9	381.9	199.5	63.7
foliage	(1018.6)	29.2	95.5	850.2	43.7
bare ground	(193.4)	10.6	10.1	3.2	169.5
total	(1936.8)	106.2	496.0	1055.2	279.3
		correctly classified pixel total: 1430.2 (73.84%)			
		classification using Feature Set C			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	32.1	0.0	9.7	0.0
grs. cvd. ground	(683.0)	0.5	546.7	77.0	58.8
foliage	(1018.6)	55.7	75.5	861.7	25.7
bare ground	(193.4)	0.0	10.1	1.6	181.7
total	(1936.8)	88.4	632.4	949.9	266.1
		correctly classified pixel total: 1622.2 (83.76%)			
		classification using Feature Set D			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	37.1	0.4	3.7	0.6
grs. cvd. ground	(683.0)	0.4	517.2	91.2	74.3
foliage	(1018.6)	50.5	79.0	849.1	40.0
bare ground	(193.4)	0.1	12.0	2.2	179.1
total	(1936.8)	88.1	608.5	946.2	294.0
		correctly classified pixel total: 1582.5 (81.71%)			

Table 2: Contingency analysis of classification results using image chips {f1,g1,r1,s1} as training data (unit: 1000 pixels)

		classification using Feature Set A			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	21.2	7.1	11.9	1.7
grass covered ground	(683.0)	17.7	259.9	405.2	0.3
foliage	(1018.6)	10.8	121.2	886.6	0.0
bare ground	(193.4)	77.5	19.3	50.5	46.1
total	(1936.8)	127.2	407.5	1354.1	48.1
		correctly classified pixel total: 1213.8 (62.67%)			
		classification using Feature Set B			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	23.7	7.0	11.1	0.0
grass covered ground	(683.0)	40.8	332.6	308.1	1.6
foliage	(1018.6)	11.7	55.0	950.7	1.2
bare ground	(193.4)	52.2	12.9	31.2	97.2
total	(1936.8)	128.2	407.5	1301.1	100.0
		correctly classified pixel total: 1404.1 (72.50%)			
		classification using Feature Set C			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	3.0	0.0	38.8	0.0
grs. cvd. ground	(683.0)	0.0	439.5	231.4	12.2
foliage	(1018.6)	0.4	20.0	995.3	2.9
bare ground	(193.4)	0.0	17.3	25.1	150.9
total	(1936.8)	3.3	476.8	1290.6	166.0
		correctly classified pixel total: 1588.7 (82.03%)			
		classification using Feature Set D			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	13.8	0.0	27.8	0.2
grs. cvd. ground	(683.0)	0.0	468.6	202.7	11.8
foliage	(1018.6)	2.0	18.0	995.7	2.8
bare ground	(193.4)	0.0	33.9	21.4	138.1
total	(1936.8)	15.8	520.6	1247.5	152.9
		correctly classified pixel total: 1616.1 (83.44%)			

Table 3: Classification accuracy in percentage on the 2k×2k ortho-image using 12 different training data sets

	Feature Set A	Feature Set B	Feature Set C	Feature Set D
Training Set 1	63.3	65.6	77.6	83.0
Training Set 2	63.2	66.3	75.8	80.8
Training Set 3	63.6	64.7	78.8	79.5
Training Set 4	63.6	65.4	77.5	80.6
Training Set 5	63.8	65.3	74.3	79.0
Training Set 6	63.4	66.0	75.8	82.4
Training Set 7	63.6	65.5	70.9	79.8
Training Set 8	63.9	74.6	78.0	80.1
Training Set 9	63.1	66.6	73.9	78.6
Training Set 10	63.5	65.7	80.1	83.3
Training Set 11	64.7	76.5	81.1	83.0
Training Set 12	64.2	68.7	81.7	83.1
mean	63.67	67.59	77.12	81.09
standard deviation	0.19	13.81	9.25	2.82

Feature Set A: twelve co-occurrence features

Feature Set B: twelve co-occurrence features and one intensity feature

Feature Set C: four 3-D features and one intensity feature

Feature Set D: twelve co-occurrence features, four 3-D features, and one intensity feature

Training Set 1-12: see Section 4.2