Comparative Studies of 3-D Textural Features and Their Reliability in Terrain Classification *

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

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

Abstract

We study the reliability of a set of stereo-based 3-D textural features in classification in response to different training data. Two types of training data, "labeling-based" and "chip-based," are investigated. Experiments have been carried out to compare the 3-D features with a set of 2-D features based on co-occurrence analysis. Results show that the 3-D features consistently outperform the 2-D features in terms of classification accuracy under various configurations of training data.

1 Introduction

Texture analysis is an important area in computer vision and has been extensively studied (e.g. [7, 9, 14, 15]). 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 [7, 13, 15]) based on this presumption. The limitation of these algorithms is that the 3-D underlying structures cannot be used directly.

Recently, both Dana et al. [3] and Wang et al. [16] 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 angle samples to make a "complete" view of the textures. However, their system did not support texture classification analysis. Wang et al. [16] 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 3-D structural characteristics of the texture.

A problem in previous experiments with the 3-D features in classification is that only a very small amount of data is used to train the classifier. In [16], 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 set better represents the data than a smaller one. However, small training data sets are often preferred since they involve less human-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 study the reliability of a set of 3-D features (described in Section 2). Experiments are carried out under different types of training data (Section 3). Section 4 shows that the 3-D features consistently outperform a widely used 2-D feature set under various configurations of training data. We discuss future work in Section 5.

2 The Feature Space

2.1 The 3-D features

The 3-D features we propose [16] for texture classification are generated by a two-view stereo algorithm [12], which employs the well-known epipolar geometry. For an arbitrary 2-D point, P, in the first view, it is correlated along its epipolar line on the second view, under the assumption that the terrain is a nearly Lambertian surface. A *similarity function*, ρ, is defined to measure how P is correlated with the points on its epipolar line. The point Q with the best match on the epipolar line tends to be the true correspondence of P. The behavior of ρ at Q reveals some important information of the 3-D texture at P. Four types of 3-D features are defined: MS, NVMS, CSF, and NDC.

MS (match score) is the maximum value of the similarity function at the best match. It tends to be high

*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.

when the image patches being correlated are very similar. NVMS (neighborhood variation of match score) measures the homogeneity of MS. Flat surfaces tend to have a better homogeneity of MS than coarse surfaces. CSF (curvature of similarity function) reflects the distinctiveness of the similarity function at the best match. The absolute value of CSF tends to be high when distinctive features or structures exist in the texture. NDC (neighborhood density of well-defined curvature) is a feature that demonstrates the behavior of the similarity function when it is used for detail analysis. Complex 3-D structures usually show a low NDC. Detailed description of these features can be found in [16, 17]. In Fig. 2(c) we illustrate the 3-D features extracted from a portion of an area of Ft. Hood, TX. Fig. 1 shows the whole area and the location of the displayed portion **Y**.

2.2 The 2-D features for comparison

Co-occurrence features were introduced by Haralick et al. [8]. Previous theoretical and experimental studies [2, 4, 8, 11, 18] showed that co-occurrence features were more effective in terrain classification problems than many other well-known 2-D textural features.

Co-occurrence features describe the texture within a local 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, 11, 16, 18], we use the following feature types: *angular second-moment (ASM)*, *contrast (CON)*, and *entropy (ENT)*. Thus a total of twelve features are employed.

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 λ of a classification can be computed based on the comparison between $\Gamma_{T,F}$ and the hand produced ground truth classification Γ_0:

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

in which

$$\Delta[c_i, c_j] = \begin{cases} 1, & \text{if } c_i = c_j \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

λ 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 proposed 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 $\lambda(T, F_1)$ and $\lambda(T, F_2)$ 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:

$$\lambda_M(F_k) = M\{\lambda(T, F_k) | \text{ for all } T \subseteq X\}, \tag{3}$$

in which M is a measure to evaluate the λ's in all the experiments. Although $\lambda_M(F_k)$ 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 practicalities of preparing the data. 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. The other method, *chip-based*, is to pick a set of small image chips, in which each chip contains only one c_i.

In order to reduce interactive operations, the chip-based method is preferred. For example, consider the image in Fig. 1, 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. **X** in Fig. 1). 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 **X** tends to include these pixels at the same rate as in the whole image. The 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. 1, 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. Image chips {f1, g1, r1, s1} $\cup$ {f2, g2, r2, s2, f3, g3, r3, s3} (see Fig. 1(b)) are used to test the stability of the 3-D and 2-D features under different combinations

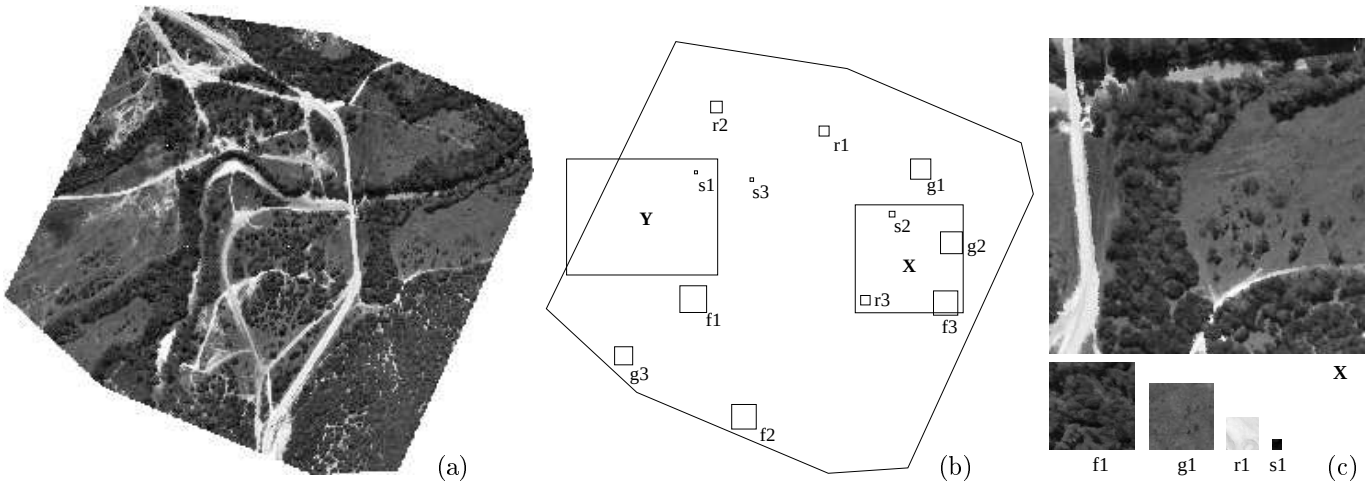

Figure 1: The orthographic intensity image and the locations of subimages used in the experiments
(a) the 2k × 2k ortho-image from Ft. Hood Image Set
(b) **Y**: the displayed subimage whose classification results are shown in Fig. 2
X: the hand-labeled subimage used as training data in the comparison experiment
{f1,g1,r1,s1}: the hand picked-up chips used as training data in the comparison experiment
{f2,g2,r2,s2,f3,g3,r3,s3}: additional chips used as training data in the stability experiment
(c) training data (upper: manually labeled subimage **X** (400 × 400); lower: f1 (99 × 99, foliage), g1 (75 × 75, grass covered ground), r1 (37 × 37, bare ground), s1 (11 × 11, shadow))

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 three feature sets are tested: Feature Set A contains the twelve co-occurrence features and the original image intensity as an additional feature; Feature Set B consists of the four 3-D features plus the intensity feature. Feature Set C 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) [6, 16, 18] and minimum Mahalanobis distance criterion. All the experiments are conducted on the 2k × 2k Ft. Hood image in Fig. 1(a). Using formula (1), 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 **X** and under the chip-based {f1, g1, r1, s1}. A portion (subimage **Y**) of the classification results are displayed in Fig. 2(d)(e). Table 1 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.

Some observations can be made from both the visual results and the quantitative analysis. First, the feature set containing only 2-D features (Fig. $2(\mathrm{d_A})(\mathrm{e_A})$) 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 is not reliable in that the brightness of ground covers might change due to different soil/grass types and lighting conditions. In the lower left part of Fig. 2(a), the ground region below the road is darker, and the classifier identifies it incorrectly when 3-D features are not in use. Third, when the four 3-D features are used (in Feature Sets B and C), the classification accuracy improves significantly (Fig. $2(\mathrm{d_B})(\mathrm{e_B})$ and $(\mathrm{d_C})(\mathrm{e_C})$). Under either training data set, the use of 3-D features increases the overall accuracy about 10 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(b)(c) shows that label-based training data (subimage **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. $2(\mathrm{d_B})(\mathrm{e_B})$ and $(\mathrm{d_C})(\mathrm{e_C})$ 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. $2(\mathrm{e_B})(\mathrm{e_C})$.

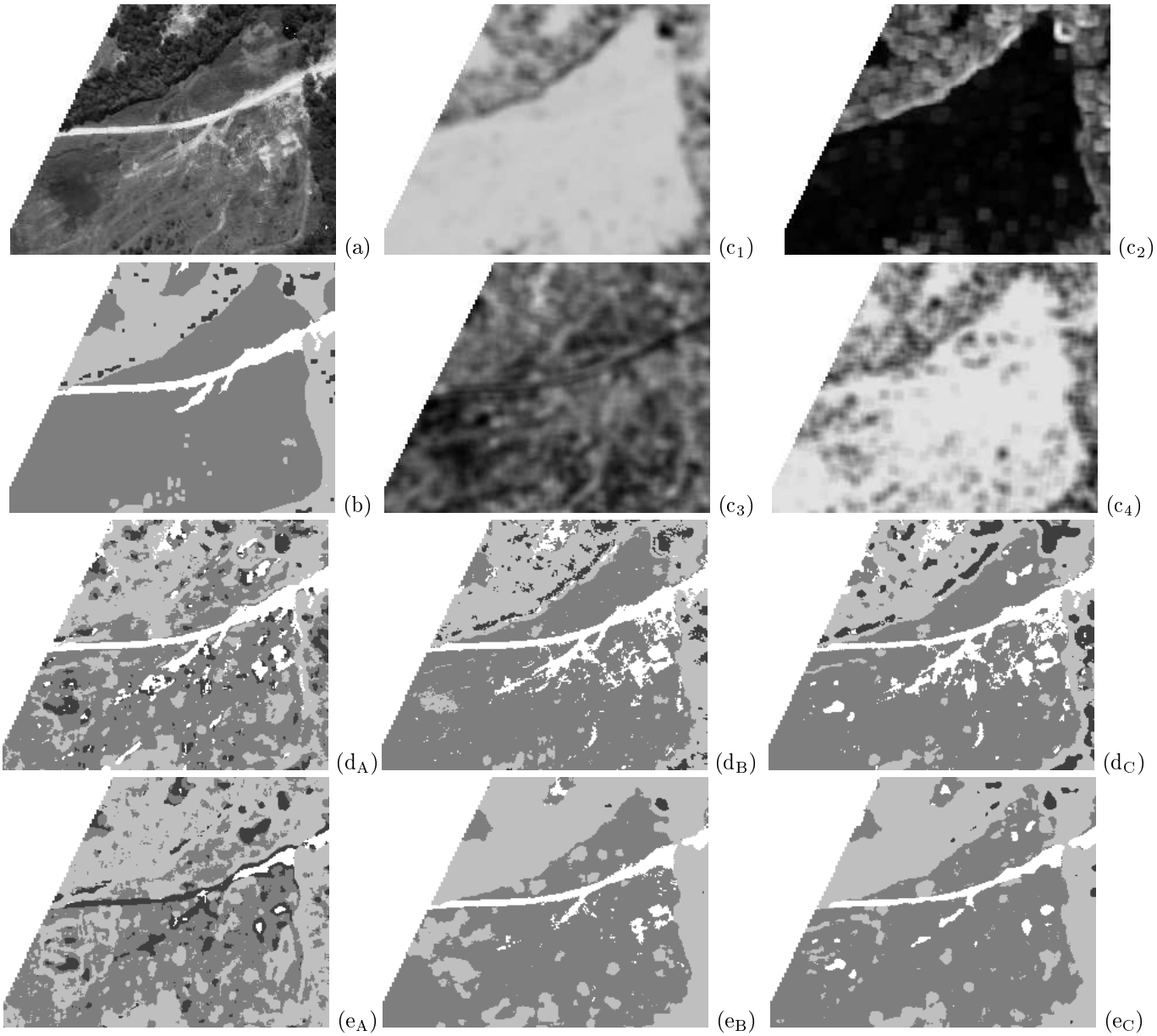

Figure 2: A portion (subimage **Y**, 560 × 430) of Ft. Hood area: its 3-D features and classification results
(a) the intensity ortho-image
(b) the ground truth hand classification
(c) 3-D features: (c_1) = MS, (c_2) = NVMS, (c_3) = CSF, (c_4) = NDC
(d) classification results using subimage **X** as training data
(e) classification results using four image chips {f1,g1,r1,s1} as training data

Feature sets used in (d)(e):
A = {twelve co-occurrence features and one intensity feature} for (d_A) and (e_A),
B = {four 3-D features and one intensity feature} for (d_B) and (e_B),
C = {twelve co-occurrence features, four 3-D features, and one intensity feature} for (d_C) and (e_C).

Legend used in (b)(d)(e): four gray levels from lighter to darker: bare ground (road, riverbed), foliage(trees, shrubs), grass covered ground, shadow.

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. 1), 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 2. 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 three feature sets, A, B, and C, are tested. From Table 2 we can see that, under any combination, the classifier using the 3-D features always performs better than Feature Set A, where no 3-D feature was involved. On average, Feature Set B and C outperforms Feature Set A and B by about 10 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 Discussion

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(b)(c) 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. 2(d_B)(e_B) and (d_C)(e_C). 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. Classifiers based on decision trees [1] and neural networks (e.g [10]) are good choices for this purpose.

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

[5] 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.

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

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

[8] 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.

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

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

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

[12] 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.

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

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

(a) using Feature Set A = {twelve co-occurrence features and one intensity feature}:

		large training data set: subimage **X**				small training data set: {s1,r1,g1,f1}			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	28.5	8.5	2.3	2.5	23.7	7.0	11.1	0.0
grs. cvd. ground	(683.0)	37.9	381.9	199.5	63.7	40.8	332.6	308.1	1.6
foliage	(1018.6)	29.2	95.5	850.2	43.7	11.7	55.0	950.7	1.2
bare ground	(193.4)	10.6	10.1	3.2	169.5	52.2	12.9	31.2	97.2
total	(1936.8)	106.2	496.0	1055.2	279.3	128.2	407.5	1301.1	100.0
		correctly classified pixel total: 1430.2				correctly classified pixel total: 1404.1			
		overall classification accuracy: 73.84%				overall classification accuracy: 72.50%			

(b) using Feature Set B = {four 3-D features and one intensity feature}:

		large training data set: subimage **X**				small training data set: {s1,r1,g1,f1}			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	32.1	0.0	9.7	0.0	3.0	0.0	38.8	0.0
grs. cvd. ground	(683.0)	0.5	546.7	77.0	58.8	0.0	439.5	231.4	12.2
foliage	(1018.6)	55.7	75.5	861.7	25.7	0.4	20.0	995.3	2.9
bare ground	(193.4)	0.0	10.1	1.6	181.7	0.0	17.3	25.1	150.9
total	(1936.8)	88.4	632.4	949.9	266.1	3.3	476.8	1290.6	166.0
		correctly classified pixel total: 1622.2				correctly classified pixel total: 1588.7			
		overall classification accuracy: 83.76%				overall classification accuracy: 82.03%			

(c) using Feature Set C = {twelve co-occurrence features, four 3-D features, and one intensity feature}:

		large training data set: subimage **X**				small training data set: {s1,r1,g1,f1}			
ground truth	(total)	shadow	grs. cvd. ground	foliage	bare ground	shadow	grs. cvd. ground	foliage	bare ground
shadow	(41.8)	37.1	0.4	3.7	0.6	13.8	0.0	27.8	0.2
grs. cvd. ground	(683.0)	0.4	517.2	91.2	74.3	0.0	468.6	202.7	11.8
foliage	(1018.6)	50.5	79.0	849.1	40.0	2.0	18.0	995.7	2.8
bare ground	(193.4)	0.1	12.0	2.2	179.1	0.0	33.9	21.4	138.1
total	(1936.8)	88.1	608.5	946.2	294.0	15.8	520.6	1247.5	152.9
		correctly classified pixel total: 1582.5				correctly classified pixel total: 1616.1			
		overall classification accuracy: 81.71%				overall classification accuracy: 83.44%			

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

	training data sets													stand.
	1	2	3	4	5	6	7	8	9	10	11	12	mean	dev.
Feature Set A	65.6	66.3	64.7	65.4	65.3	66.0	65.5	74.6	66.6	65.7	76.5	68.7	67.59	13.81
Feature Set B	77.6	75.8	78.8	77.5	74.3	75.8	70.9	78.0	73.9	80.1	81.1	81.7	77.12	9.25
Feature Set C	83.0	80.8	79.5	80.6	79.0	82.4	79.8	80.1	78.6	83.3	83.0	83.1	81.09	2.82

Feature Set A, B, and C: same as in Table 1; Training Set 1-12: see Section 4.2

[14] 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.

[15] 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.

[16] 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.

[17] X. Wang and A. Hanson, "Reliability of 3-D Textural Features in Terrain Classification," Technical Report #98-002, Dept. of Computer Science, Univ. of Massachusetts at Amherst, January 1998.

[18] 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.