INVESTIGATION OF THE MPEG-7 HOMOGENEUOS TEXTURE DESCRIPTOR FOR THE AUTOMATIC EXTRACTION OF TREES

B.-M. Straub

IPI, Institute for Photogrammetrie and GeoInformation, 30167 Hanover, Germany - bernd-m.straub@ipi.uni-hannover.de

Commission III, WG III/4

KEY WORDS: Urban, Vegetation, Automation, Recognition, Algorithms, Texture, Infrared, High Resolution.

ABSTRACT:

In this paper we describe our recent work on the automatic extraction of trees from high resolution aerial images. In order to be more independent of color information we have investigated textural properties of trees and buildings. The aim is to be able to differentiate between object classes based on textural information. Texture is a characteristic feature of trees, and if color information is not available it is an important cue to differentiate between trees and buildings. The Gabor filter bank of the standardized MPEG-7 Homogeneous Texture Descriptor (HTD) was used for the extraction of the textural properties. The qualification of the HTD for the extraction and classification of trees is evaluated. The evaluation is based on first experimental results, which are presented in the paper.

1. INTRODUCTION

Geographic information meets virtual reality (CROSSES, 2002). The aim of the CROSSES (Crowd Simulation System for Emergency Situations) project is to develop a realistic training system for emergency forces. An important aspect of such a system is the use real data, which give the training staff a good impression of the local situation. One of the tasks that we had to solve in this project was the extraction of trees and buildings from aerial imagery. The production of the data for the 3D city model should be done automatically, because the system is proposed to be installed in different cities, always with actual real data. We have developed algorithms for the automatic extraction of buildings (Gerke et al., 2001) and trees (Straub and Heipke, 2001). For CROSSES colour infrared (CIR) aerial images were acquired in summer 2000. The image flight was carried out with 80% overlap along and across the flight direction. The image scale is 1:5000, which leads to a GSD of 10 cm at a scanning resolution 20µm. Based on these images a digital surface model and a true orthoimage were automatically derived by the French company ISTAR (Gabet et al., 1994). The extraction of vegetation was planned from the beginning of the project, partly because reports had predicted an increasing request for vegetation in general, and trees in particular as a part of 3D city models (Fuchs et al, 1998). It seems, however, that CIR imagery is not readily available today, even though CIR imagery is well established for the extraction of vegetation information and does not handicap the extraction of man-made objects. Many customers of 3D city model data[1] prefer true color images due to the appearance of the orthophotos. In order to be more independent of the available colors and especially of the infrared channel, we have decided to investigate textural properties for the extraction of trees during the vegetation period, and potentially their classification into different types.

[1] We have learned that from discussions with other researchers and staff from companies working in the field of data production for 3D city models.

There is no doubt that the use of the textural information is helpful for the detection of objects from images. Human analysts discriminate between areas with vegetation and trees and areas with man-made objects by using textural features (Haralick and Shapiro, 1992), and many promising results are reported in the literature regarding the use of texture for the automatic object extraction. But, there is no commonly accepted way to select the texture operators and to link the different textural features (Shao and Förstner, 1994).

Today, the description of texture is a part of the ISO/IEC standard MPEG-7, different texture descriptors were investigated by the MPEG consortium (MPEG-7, 2002). The *Homogeneous Texture Descriptor* (HTD) (Man Ro et al., 2001) which is composed of a Gabor filter bank, a formal description of the extracted features as well as different similarity measures, is investigated in this paper for the extraction of trees from high resolution aerial imagery.

2. RELATED WORK

The extraction of trees from optical and/or height data was investigated by different research groups. The discrimination of vegetation and man-made objects using true-color images is discussed in (Niederöst, 2000). Niederöst proposes the use of an artificial channel denoted as degree of artificiality, which can be computed from the red and the green band of true color images.

(Brandtberg and Walter, 1998) have developed an approach for the extraction of trees from aerial images with a GSD of 10 cm based on the gray level curvature and length of edges in different scales. (Brunn and Weidner, 1997) proposed to use the variance of DSM surface normal to detect vegetation regions. Laser scanner data and a colour infrared image are used in combination by (Haala and Brenner, 1999) for the classification of an urban scene. A pixel based unsupervised classification algorithm is employed to perform the segmentation of the image.

Some authors propose the use of texture for the detection of regions with trees in urban environments, for example (Zhang, 2001) who uses local directional variance with good results. The local variance was also used in (Straub et al., 2000) for the detection of vegetation areas in coarse scale. In (Baumgartner et al., 1997) the authors propose the use of Laws Filters (Haralick and Shapiro, 1992) for the detection of textured regions. These features are often used as an additional channel in the framework of a pixel per pixel classification, refer for example (Kunz and Vögtle, 1999), (Straub et al., 2000), (Zhang, 2001).

Summarizing one can say, that different texture parameters were investigated for the automatic detection of vegetation in aerial or satellite imagery. But, "The texture discrimination techniques are for the most part ad hoc" (Haralick and Shapiro, 1992, p.453), which is perhaps true until today. Standardization may overcome this problem. Thus we have investigated the qualification of the MPEG-7 *Homogeneous Texture Descriptor* (HTD) (Man Ro et al., Kim 2001) for the detection of vegetation in high resolution aerial images.

3. THE MPEG-7 HOMOGENEOUS TEXTURE DESCRIPTOR

MPEG-7 is an ISO/IEC standard developed by MPEG (Moving Picture Expert Group). The formal name of MPEG-7 is "Multimedia Content Description Interface". The standard provides a set of standardized tools to describe multimedia content, Geographic Information Systems and Remote Sensing are mentioned as possible application domains. Low level features of images like texture and color are described in the part "MPEG-7 Visual". Three texture descriptors are recommended, the HTD, the edge histogram descriptor (EHD), and the perceptual browsing descriptor (PDB). The HTD should allow to classify images with high precision (Wu et al., 2001). The detection of objects like "parking lots", or "vegetation patterns" is also directly mentioned in the standard (ISO/IEC, 2001). The MPEG-7 Homogeneous Texture Descriptor (HDT) is described in detail in (Man Ro et al., 2001). In this section we give a short summary of the used filter bank, the extracted feature vector and the proposed measures of similarity.

3.1 Extraction of Textural Features

The extraction of features is done with a Gabor filter bank. In radial direction the feature channels are spaced with octave scale, center frequencies and octave bandwidths are given in Table 1. Thirty feature channels C_i are defined for the features which are extracted with Gabor filters in six orientations f = 0°, 30°, 60°, 90°, 120°, 150° with an angular bandwidth of 30°. This frequency layout is motivated by the human visual system. It was confirmed by psychophysical experiments, that the brain decomposes the incoming signal into sub bands in frequency and orientation (Branden Lambrecht, 1996).

Radial Index, r	*0*	*1*	*2*	*3*	*4*
Center frequency *w*	¾	3/8	3/16	3/32	3/64
Octave Bandwidth	½	¼	1/8	1/16	1/32

Table 1: Parameters of Gabor Filter Bank in radial Direction

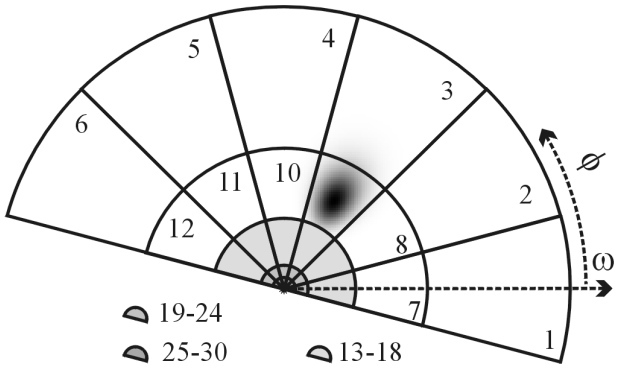

Figure 1: Frequency layout of the Gabor filter bank with ID's of feature channels C, depicted is the Gabor filter for feature channel C_9.

3.2 The Feature Vector

The mean value f_{DC} and standard deviation f_{SD} of the original image, as well as the energies e_i and their standard deviations d_i of the Gabor filtered image constitute the feature vector *TD*, as follows:

$$TD = [f_{DC}, f_{SD}, e_1, e_2, e_3, \ldots, e_{30}, d_1, d_2, d_3, \ldots, d_{30}]$$

All 62 elements together are called the *enhancement layer*, the reduced feature vector without d_i values is called *base layer*. The computation of the feature vector can be done in advance, and then the feature vector can be stored together with the image. The quantization of the TD values to 1 byte leads to a total length of the texture descriptor of 62 bytes for the enhancement layer, respectively 32 bytes for the base layer. If one uses tiles with a size of 128*128 pixel, and stores only the base layer of the feature vector, the amount of storage for the feature layer is then 1/512 of the uncompressed size of one image channel (1/264 for the enhancement layer).

3.3 Measurement of Similarity

The similarity *d(R, J)* between two images *R* and *J*, can be measured with the Euclidian Distance in feature space. Once the feature vector *TD* is computed, the following similarity measurements can be performed. In the following the TD_R is the feature vector of the reference image *R* (in the domain of image retrieval the term *query image* is more usual), index *j* of TD_j assigns the feature vector of another image *J*, and the index *k* marks the *k*-th element of the feature vector.

$$d(R, J) = \text{distance}(TD_R, TD_J) \tag{1}$$

$$d(R, J) = \sum_k \left| \frac{w(k)\left[TD_R(k) - TD_J(k)\right]}{a(k)} \right| \tag{2}$$

The weighting factor *w(k)* of the *k*-th *TD* value and the normalization values $a(k)$ depend on the used images. In (Man Ro et al., 2001) it is proposed to use values from a reference data base.

The similarity measurement *d(R, J)* depends of the intensity, the scale, and the rotation of the texture. Since this dependency is undesirable in many applications, three matching procedures are proposed: the Intensity-invariant-, the Scale-invariant-, and the Rotation-invariant matching.

3.3.1 Intensity-Invariant Matching

If one is only interested in the textural features, f_{DC} has to be ignored for the computation of the similarity measure.

$$w(0) = 0 \tag{3}$$

3.3.2 Scale-invariant Matching

The querying image *R* is zoomed with *N* different zoom factors, leading to *n* different scaled versions of the querying image.

$$d(R, J, n) = \text{distance}\left(TD_{R,n}(k), TD_J(k)\right) \tag{4}$$

The similarity measure *d(R, J)* is then the minimum of the *n* obtained distances.

$$d(R, J) = \text{minimum}\{d(R, J, n); n = 1, 2, ..., N\} \tag{5}$$

3.3.3 Rotation-Invariant Matching

Here the feature vector TD_R of the reference image R is shifted in the angular direction by $\phi = 30°$:

$$d(R, J, mf) = \text{distance}\left(TD_{R,mf}(k), TD_J(k)\right) \tag{6}$$

The distance used for the rotation invariant descriptor is then calculated as:

$$d(R, J) = \text{minimum}\{d(R, J, mf); m = 1, 2, ...6, f = 30°\} \tag{7}$$

4. EXPERIMENTS

An investigation of the HTD for the extraction of vegetation like bushes or trees is presented in this section. The qualification of the HTD for the coarse segmentation of aerial imagery was shown in (Manjunath et al., 2000), (Newsam et al., 2002).

As mentioned above we are mainly interested in the differentiation between roofs and trees. In our experiment we have selected quadratic regions from an aerial image, which show roofs or trees. These data will be used for the further investigation. The performance of the Intensity-Invariant Matching (refer 3.3.1) was tested. The Scale- and Rotation Invariant Matching Methods were not investigated. Scale-Invariant Matching is not so interesting in the given context, because the scale is usually known for aerial images. The Rotation-Invariant Matching procedure is not applicable for the distinction between roofs and trees, because tree textures are more or less isotrop and roof not.

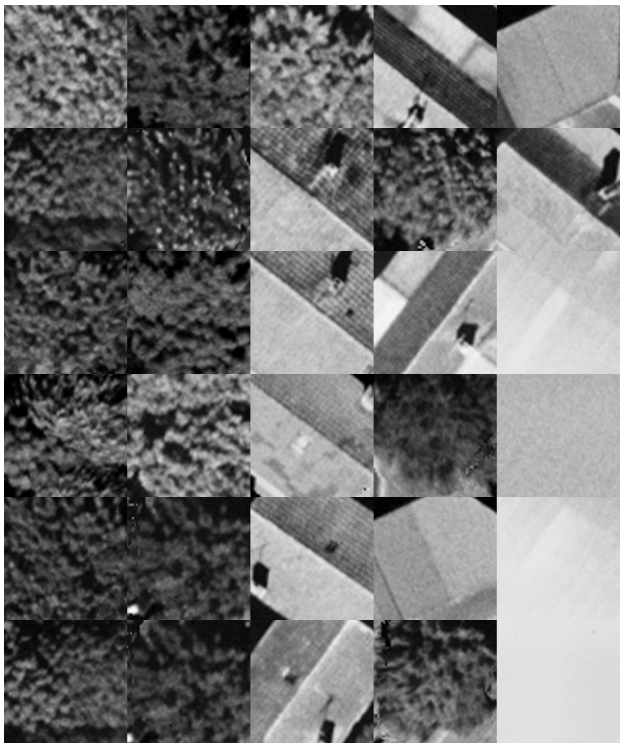

Figure 2: Example textures used for the investigation

A closer investigation of the *TD* is presented in the second part, focusing on object specific properties. These properties can be used in a more object specific approach for the extraction of trees in an urban environment.

4.1 Description of the Used Image Data

The quadratic image tiles, which we have used for a first test of the qualification of the HTD are taken from the green channel of an aerial CIR image with a GSD (Ground Sampling Distance) of 10 cm. The regions with trees and different kinds of roofs were selected by hand. A subset of 64*64 pixel is taken from the image, and enlarged to a size of 128*128 pixel using bilinear interpolation. The size of the subset was selected such, that the tiles cover a homogeneous textured region. The size of 128*128 pixel for a tile was also used in the performance tests of the MPEP-7 consortium. The resulting image tiles show tree- or roof textures with a simulated GSD of 5 cm. A part of these tiles are depicted in Figure 2, 16 examples with typical tree texture, and 14 examples with roof textures. Two different types of roofs can be clearly distinguished, one type with a preferred direction of the texture, and another type without that property. The tiles in Figure 2 are ordered by the similarity to the tile in the upper left corner, refer section 4.2 for details.

1	7	13	19	25
2	8	14	20	26
3	9	15	21	27
4	10	16	22	28
5	11	17	23	29
6	12	18	24	30

Table 2: ID's of the textures depicted in Figure 2

4.2 Intensity Invariant Matching

The test of the Intensity Invariant Matching method should give us a first idea of the performance of the algorithm. The feature

vectors TD_i are calculated for every tile (refer 4.1), the first image is marked as reference image *R*. The weighting parameters *w(k)* and $a(k)$ in Equitation (2) are computed from the feature vectors TD_i as mean value and standard deviation. The weight of a sub band, used for the computation of the similarity measure *d*, increases with its energy and decreases with its noise.

$$w(k) = \frac{1}{n} \sum_{i=1-n} TD_i(k) \tag{8}$$

$$a(k) = \frac{1}{n-1} \sum_{i=1-n} \left(TD_i(k) - w(k)\right)^2 \tag{9}$$

The distances *d(R,J)* between the image *R* and all images *J*, with *j=2,3, ..30*, are plotted in Figure 3, refer Figure 2 together with Table 2 for a qualitative inspection. The distances *d(R,J)* are scaled with the largest distance value *d(R,30)*, which occurred in the test data.

Figure 2 indicates the tiles *J*, with index *j=2,3,...,30* ordered by the distances to the reference image *R*, with index *1*. The visual inspection of this data shows that, with some exceptions, the tree textures are well separated from the roof textures using the intensity invariant matching method. Thus the HTD seems to be qualified for the differentiation between buildings and trees.

The numerical values for the similarity are represented in Figure 3, the largest distance *d(R,30)=4.5* is used to scale the distances *d(R,J)*. As the textural energy in image 30 is very poor, this should give an idea of the range of distance values, which may occur in praxis. In our example a threshold value, placed in the centre of the range, i.e. $d_{THRESHOLD}=0.5$, will lead to a retrieval of the first ten tiles, a success rate of 63%. A threshold value $d_{THRESHOLD}=0.65$, would lead to a success rate of 82%. An order of magnitude comparable with the results of the performance test using Brodatz Textures (Brodatz, 1966) in (Man Ro et al., 2001). Nevertheless, the computation of $d_{THRESHOLD}$ from the values in Figure 3 is not obvious.

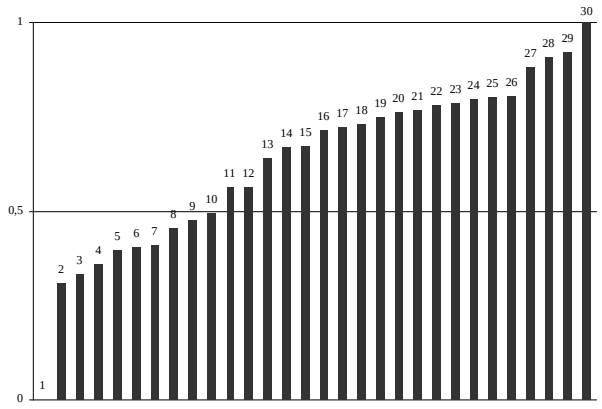

Figure 3: Scaled Distances for the Intensity Invariant Matching

A further investigation should include, besides an enlarged data set, tree-textures which are clearly different, for example textures from conifers. Nevertheless, the first results are well promising, taking into account that there was no room for a fine tuning of the parameters of the filter bank, and the threshold. The intensity invariant matching method seems to be suited for the texture based separation of trees and roofs.

4.3 Discussion of Typical Object Properties

Once the feature vector is computed, a closer look onto possible object specific properties, which are reflected in the feature vector, is obvious.

The texture of a tree does not have a major orientation. Thus the energies for all directions of one sub-band should have similar values. This observation is always valid for deciduous trees, for conifers only if they are quite close to the center of the image. Against textures of trees, roof textures have one or two main directions. This property should lead to one or two peaks in every sub band, and the angular index should be the same for these peaks.

The TD mean values of the tree and building textures confirm this assumption, refer Figure 4. In the three middle sub bands the energy values are relatively homogeneous for trees, and for buildings peaks reflect their main orientation.

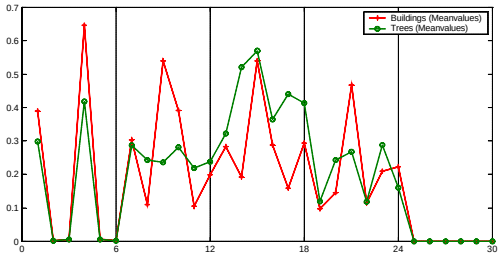

Figure 4: Mean values of energies for buildings and trees

The standard deviation $s_{(i, SB, r)}$ for the sub bands with constant *w* -values can be used to measure this property.

$$m_{i,r} = \frac{1}{6} \sum_{m} TD_i(6r+m), \quad \text{with } r = [0, 1, 2, 3, 4],\ m = [0, 1, 2, 3, 4, 5] \tag{10}$$

$$s_{i,SB,r} = \frac{1}{5} \sum_{m} \left(m_{i,r} - TD_i(6r+m)\right)^2 \tag{11}$$

The mean value s_i of the standard deviations $s_{i, SB, r}$ can be used to differentiate between trees and buildings. From the example data set, which is depicted in Figure 2 one gets s_{TREE}=0.1 and $s_{BUILDING}$=0,18.

5. SUMMARY AND OUTLOOK

In this paper we have investigated the performance of the MPEG-7 Homogeneous Texture Descriptor. The main focus of this paper is on the investigation and discussion of the HTD's qualification for the detection and possibly reconstruction of trees from high resolution imagery. The results are well promising, and it seems that an integration of the HTD in our system for the extraction of trees in urban environments will

overcome some restrictions regarding the pre-conditions with regard to the image data.

At the moment it is not clear if the HTD leads to really better results than approaches using simple texture measurements like the local variance or Laws energy approach. But the standardization of the HTD provides the chance to use a common base for the extraction and representation of textural features. In future work the performance will be compared with the both approaches mentioned above.

The performance of the HTD will also be investigated closer using larger data sets. The enlarged data set should be selected such, that that different types of, clearly different, trees can be investigated.

6. REFERENCES

Baumgartner, Albert, Eckstein, W., Mayer, H., Heipke, C., and Ebner, H., 1997. Context Supported Road Extraction. *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Vol. II. ed. E.P. Baltsavias, O.Henricson A. Gruen, Birkhäuser. Basel, Boston, Berlin. pp. 299-308.

Branden Lambrecht, Christian J. van den, 1996. A Working Spatial Model of the Human Visual System for Image Restoration and Quality Assessment Applications. *International Conference on ASSP*. Vol. 4. IEEE. New York. pp. 2291-2294.

Brandtberg, Tomas, and Walter, Fredrik, 1998. Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple scale analysis. *Machine Vision and Applications*, Vol. (1998) No. 11, pp. 64-73.

Brodatz, P., 1966. *A Photographic Album for Artists and Designers*. New York, Dover.

Brunn, Ansgar, and Weidner, Uwe, 1997. Extracting buildings from digital surface-models. *IntArchPhRS*. Vol. 32. No. Part 4-4W2. ISPRS. pp. 27-34.

CROSSES, 2002. *Website of the CROSSES Project*. http://crosses.matrasi-tls.fr/ (21.3.2002).

Fuchs, Claudia, Gülch, Eberhard, and Förstner, Wolfgang, 1998. OEEPE Survey on 3D-City Models. *OEEPS Publication*. No. 35. Bundesamt für Kartographie und Geodäsie. Frankfurt. pp. 9-123.

Gabet, L., Giraudon, G., and Renouard, L., 1994. Construction automatique de modèles num'eriques de terrain haute résolution en milieu urbain. *Société Francaise de Photogrammétrie et Télédétection*, No. 135, pp. 9-25.

Gerke, Markus, Heipke, Christian, and Straub, Bernd.-M., 2001. Building Extraction From Aerial Imagery Using a Generic Scene Model and Invariant Geometric Moments. *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*. IEEE. Rome. pp. 85-89.

Haala, Norbert, and Brenner, Claus, 1999. Extraction of Buildings and Trees in Urban Environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. (54) No. 2-3, pp. 130-137.

Haralick, R., and Shapiro, L.G., 1992. *Computer and Robot Vision (I)*. Vol. **1**. Addison Wesley Publishing Company, p. 672.

ISO/IEC JTC1/SC29/WG11, 2001. *Overview of the MPEG-7 Standard (version 6.0)*. http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm (21.3.2001).

Kunz, Dietmar, and Vögtle, Thomas, 1999. Improved land use classification by means of a digital topographic database and integrated knowledge processing. *International IGARSS '99 Proceedings*. Vol. 2. IEEE. Hamburg, Germany. pp. on CD.

MPEG-7, 2002. *The MPEG homepage*. http://mpeg.telecomitalialab.com/, (11.3.2002).

Man Ro, Yong, Kim, Munchurl, Kang, Ho Kyung, Manjunath, B.S., and Kim, Jinwoong, 2001. MPEG-7 Homogeneous Texture Descriptor. *ETRI Journal*, Vol. (23) No. 2, pp. 41-51.

Manjunath, B.S., Wu, P., Newsam, S., and Shin, H.D., 2000. A texture descriptor for browsing and similarity retrieval. *Journal of Signal Processing: Image Communication*, Vol. (16) No. 1-2, pp. 33-43.

Newsam, Shawn, Tesic, Jelena, El-Saban, Motaz, and Manjunath, B.S., 2002. *MPEG-7 Homogeneous Texture Descriptor Demo*. http://nayana.ece.ucsb.edu/M7TextureDemo/Demo/client/M7TextureDemo.html, (12.3.2002).

Niederöst, Markus, 2000. Reliable Reconstruction of Buildings for Digital Map Revision. *Geoinformation for All*. Vol. XXXIII. No. Part B3. IntArchPhRS. Amsterdam. pp. 635-642.

Shao, Juliang, and Förstner, Wolfgang, 1994. Gabor Wavelets for Texture Edge Extraction. *ISPRS Commission III Symposium on Spatial Information from Digital Photogrammetry and Computer Vision*. Vol. XXX-3, IntArchPhRS. Munich, Germany. pp. 8.

Straub, Bernd-M., and Heipke, , 2001. Automatic Extraction of Trees for 3D-City Models from Images and Height Data. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*. Vol. 3. ed. A. Gruen, L. van Gool E. Baltsavias, A.A.Balkema Publishers. Lisse/Abingdon/Exton(PA)/Tokio. pp. 267-277.

Straub, Bernd.-M., Wiedemann, Christian, and Heipke, Christian, 2000. Towards the automatic interpretation of images for GIS update. *GeoInformation for All*. Vol. 33. No. B2. IntArchPhRS. Amsterdam. pp. 521-532.

Wu, Peng, Man Ro, Yong, Won, Chee Sun, and Choi, Yanglin, 2001. Texture Descriptors in MPEG-7. *Computer Analysis of Images and Patterns*. ed. W. Skarbek, Springer. Berlin Heidelberg. pp. 21-28.

Zhang, Yun, 2001. Texture-Integrated Classification of Urban Treed Areas in High-Resolution Color-Infrared Imagery. *Photogrammetric Engineering & Remote Sensing*, Vol. (67) No. 12, pp. 1359-1365.

7. ACKNOWLEDGEMENT

This work was developed within the IST Program CROSSES financed by the European Commission under the project number IST-1999-10510.
Thanks to Eberhard Gülch for his questioning of customers, und the discussions regarding the image data.
Thanks to Christian Heipke for critical reading and patient explanations.