

# HIGHFOREST - FOREST PARAMETER ESTIMATION FROM HIGH RESOLUTION REMOTE SENSING DATA

Heikki Astola, Catherine Bounsaythip, Jussi Ahola, Tuomas Häme, Eija Parmes, Laura Sirro, Brita Veikkanen

Technical Research Centre of Finland, Information Technology, Information Systems,  
P.O. Box 1201, FIN-02044 VTT, Finland, <http://www.vtt.fi/tte/> -  
[heikki.astola@vtt.fi](mailto:heikki.astola@vtt.fi) (firstname.lastname)@vtt.fi

## Commission VII, WG VII/3

**KEY WORDS:** remote sensing, forestry, automation, estimation, inventory, high resolution

### ABSTRACT:

The aim of the study was to develop a tool for the estimation of forest variables using high-resolution satellite data. The tool included modular operative software. The image analysis methodology focused on the reduction of the known problems of the previous satellite image based methods, *i.e.* the saturation of the estimates at higher biomass levels and uncertainty in tree species estimation. Modern contextual image analysis methods were combined with the spectral information of the imagery. In the test application the tool used images from the Ikonos satellite with a ground resolution of one and four meters.

The developed Forestime software estimated the forest variables by segmenting the imagery to 'micro-stands', by computing stand-wise image feature vectors for the stands from the input satellite image, and by combining ground reference data with clusters from an unsupervised clustering stage. The estimates are produced as weighted sums of the input sample class probabilities. The target variables in the study were stem volume, average stem diameter, stem number and tree species proportions. The RMSE% for total stem volume was 37.4 % (% of mean), for average stem diameter 23.4 %, for stem number 87 %, for pine percentage 111 %, for spruce percentage 47 %, and for broad-leaved tree percentage 137 %.

## 1. INTRODUCTION

Forest owners and wood procurement organisations have a continuous need to acquire updated information on forests. The traditional stand-wise ground inventory is, however, expensive and its frequency is too slow. There is also a need to reduce stand sizes in order to delineate stands more flexibly when cuttings and silvicultural measures are planned, and in order to simulate future development of the forest more reliably. Satellite imagery with a ground resolution of twenty to thirty meters has been tested and is partly used even operatively for forest inventories. Although the estimates for stand characteristics are reasonable good using such data there are two major limitations that have prevented a wider operative use. The first limitation is saturation of the reflectance at higher levels of the growing stock volume. The saturation limit in Finnish conditions is at approximately 200 m<sup>3</sup>/ha. Another limitation is the difficulty to estimate tree species proportions since one pixel of a dimension of e.g. 30 meter includes often several tree species.

The spatial resolution of remotely sensed optical imagery has increased from 30 m (Landsat Thematic mapper) to 1 m panchromatic and 4 m multispectral (Ikonos, Quickbird) in the last 20 years. At VTT's remote sensing group various methods for forestry applications using optical and radar satellite imagery have been developed within the same time span, and are in operative use as well. One goal of the HighForest study was to combine some of the existing methods into a new operative software tool utilizing the new high resolution satellite data.

The suitable application areas are all operative forest inventories as well as wood procurement mapping. Strategic planning could also utilize the information whenever information on species is required. Significant cost savings can be achieved if estimates applicable in operative forestry can be computed using imagery with a resolution of one to few meters instead of using airborne data with a resolution of tens of centimeters. By making the method for forest mapping of large forest areas fully automatic, fresh information for wood procurement purposes will be rapidly available.

## 2. DATA AND PREPROCESSING

### 2.1 Satellite data

The acquisition of one Ikonos image was initiated mid June 2003. A geo-referenced and orthorectified Ikonos image of size 10 km x 13 km was acquired from the specified coordinates in Suonenjoki, Finland, dated 5.9.2003 (see *Figure 1*). Due to excess cloud cover in the eastern side of the image, it was extended three km towards west without additional charge. Fortunately the cloud cover did not extend over the study area located in the middle of the ordered image. The Ikonos images have four spectral channels: blue (450 - 520 nm), green (520 - 600 nm), red (630 - 690 nm) and NIR (760 - 900 nm), with a ground resolution of four meters, and one panchromatic channel (450 - 900 nm), with ground resolution of one meter. A smaller image covering the ground data area was extracted from the whole Suonenjoki image for testing (see *Figure 2*).

The received image was subjected to radiometric correction using VTT's in-house software (smac\_corr.exe) using the coefficients from Space Imaging (2004).



Figure 1 Ikonos true color image from Suonenjoki, acquisition date 5.9.2003. The test image area depicted with red square

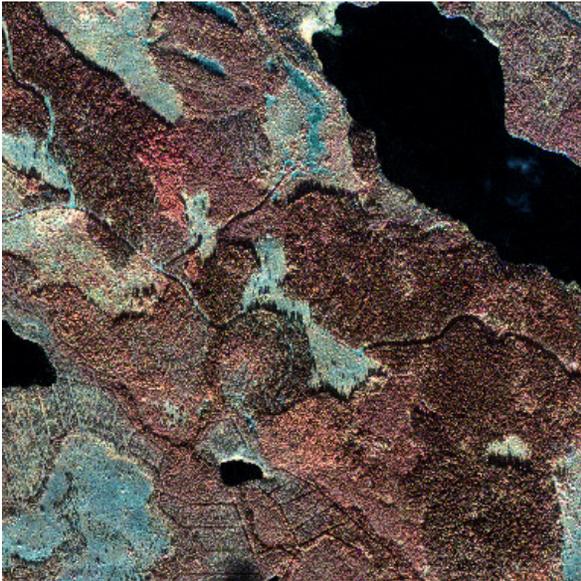


Figure 2 Ikonos False color image of Suonenjoki test area (R = NIR chn., G = green chn., B = blue chn., I = intensity layer = PAN chn.)

## 2.2 Ground data

The Suonenjoki Research Station of the Finnish Forest Research Institute has measured 327 sample plots for 7 variables for different tree species in summer 2001. The Suonenjoki ground data was thoroughly examined visually to exclude ground sample plots that distort or cause error to the estimation process.

A total of 74 data points were removed from the original ground data set (327 points) after a careful visual inspection. These data points were considered as erroneous or were too close to borders of very different ground segments. A set of 43 points of zero (39 points) or near zero (4 points) data were added on areas regarded as clear cuts or fields with zero total stem

volume. In the resulting ground data set there were thus 296 sample plots (see Figure 3).

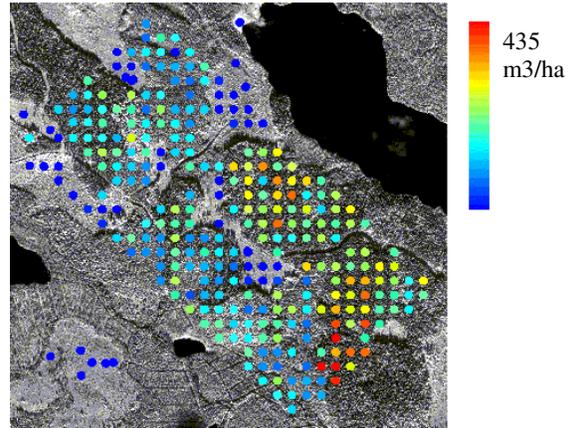


Figure 3 Ground data points after data exploration. Plot color proportional to stem volume (dark/blue  $V = 0$  m<sup>3</sup>/ha, bright/ red  $V = 435$  m<sup>3</sup>/ha)

## 3. METHODS

### FOREST VARIABLE ESTIMATION

#### 3.1 Feature extraction

The spectral features averaged from the radiometrically corrected Ikonos channels together with the contextual features calculated from the Ikonos PAN-chromatic channel form the input feature set to the Foresttime estimation process.

The test feature set contained five Haralick features (Haralick, et. al. 1973): contrast, entropy, inverse difference moment, homogeneity, and sum average. The associated grey-level co-occurrence matrix has been calculated for a 15 x 15 pixel window, and with distance relation of one pixel in both image directions. The occurrences in the four possible pixel separation combinations were summed together i.e. the direction information is lost. The co-occurrence matrix was calculated for a compressed image of 16 grey levels.

In addition to the Haralick features, a set of four Gabor features were calculated from the PAN-chromatic channel of the Ikonos image, using a bank of even-symmetric real-valued Gabor filter masks (Jain & F. Farrokhnia, 1991). In the method a set of Gabor filters, covering the image spatial-frequency domain nearly uniformly, are generated, and a filtered multi-channel image is produced from the input image.

The tree location tool is determines tree crown locations using the local maximum filtering (LM filtering) technique on the Ikonos PAN-chromatic channel (Wulder, et.al., 2002). The Foresttime averaging tool then produces a segmentwise local maximum density feature as its output. The tree species proportions tool registers the tree reflectance from the Ikonos image spectral channels at the locations given by the tree location tool. Near-infrared reflectance of the nearest pixel of each located stem is used to determine whether the tree is broad-leaved or a conifer. If near-infrared reflectance exceeds the given threshold the tree is labelled as broad-leaved,

otherwise coniferous. The Forestime averaging tool then produces a segmentwise feature proportional to the percentage of broad-leaved trees as its output.

### 3.2 Feature selection

To get an idea of best feature set for the parameter estimation models, the feature data was examined by calculating the mutual correlation of forest variables and the features, and by creating estimation models with stepwise linear regression of SPSS statistical software package. The models obtained with SPSS contained often five features or more. However, the marginal utility obtained by adding features above five was very small with the SPSS regression models, and hardly ever increased the accuracy of the forest variable estimates obtained with Forestime. Consequently the feature selection process concentrated on selecting features best explaining the target variable variance, and the feature sets used in the estimation contained six features maximum. In many cases the best features for separate target variables coincided, resulting in fairly similar input feature sets (the spectral channels augmented with the Haralick entropy are a good compromise to estimate most of the selected variables).

### 3.3 Segmentation

The basic component of the Ikonos image processing is the segmentation of the image into homogenous areas (segments), in the sense of tree species, stem number, tree height and other relevant forest parameters. The criteria for homogeneity is based on multispectral radiance or reflectance values and texture measures. The method allows segments small enough to include only a single tree.

The multispectral segmentation is based on the method developed at VTT (Parnes1992), derived from the segmentation approach presented by Narendra & Goldberg (1980). The method uses gradient images calculated from the original image spectral bands as inputs, and links the pixels belonging to the same image segment with the directed trees method by Narendra & Goldberg (1980).

### 3.4 Variable estimation

The segmentation is used in the next step to obtain stadwise averages from the selected features. The resulting average images are sampled with uniform random grid to produce a set of input data vectors. The data set is clustered by K-means algorithm into a pre-selected number of clusters. Each obtained cluster is then assigned with the ground reference vector of target variables as an average of the plot samples belonging to this cluster. The estimates are produced as weighted sums of the input sample class probabilities (Häme, et al. 2000, Häme et al. 2001). The estimation results are obtained as raster images, or in vector data format. The Forestime system principle is depicted in Figure 4.

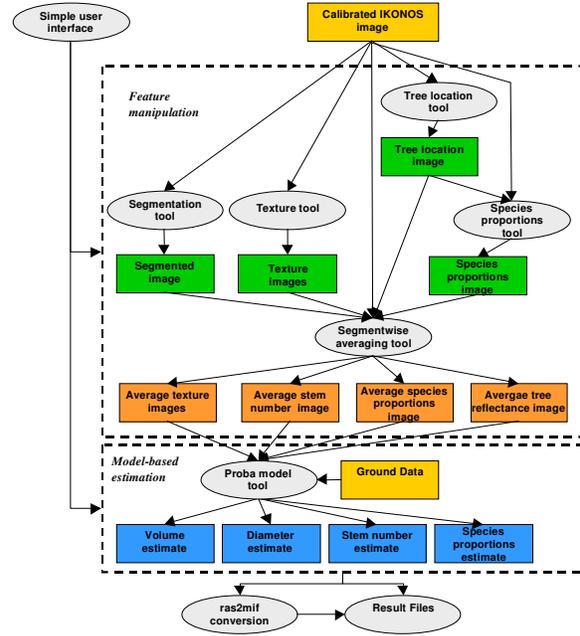


Figure 4 Forestime estimation process

## FORESTIME SOFTWARE

The first version of a modular software tool was developed (Forestime v. 1.0) in the study. The guidelines in the software design were modularity, development and utilization of in-house software, and making the method for forest mapping of large forest areas automatic. The modularity of the system has several benefits. It allows an easy addition or replacement of e.g. some feature extraction module with another. The output of the system might as well be the land cover classification instead of parameter estimates. The system is not restricted to any specific imagery, but can fairly easily be configured to use data from a different type of instrument, or even from several instruments. The software also includes an API (Application Program Interface) for integration with other systems like GIS databases.

The system is a multi-tier application (Figure 5). The Forestime is running in Windows operating system. The user interface for the System is provided by a designated GUI or external software via System's application program interface. The server communicates with analysis tools via command line calls. The Forestime software was realised as object oriented design, which makes the system easily configurable and expandable. In software development Rational Rose was used for UML modeling, and JBuilder for Java implementation. The version control system was MS Visual SourceSafe. All the analysis tools were written in C.

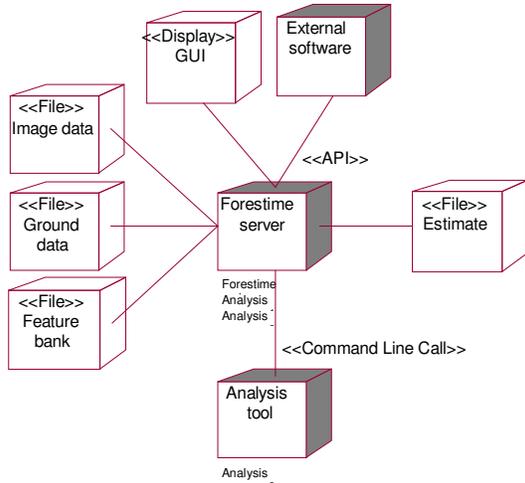


Figure 5 Deployment diagram of Forestime.

#### 4. RESULTS AND DISCUSSION

The Forestime system performance in forest variable estimation was tested with Ikonos image acquired on 5. Sep 2003 from Suonenjoki, and with FFRI sample plot data from the same area, measured 2001. The tests included producing estimates for total stem volume (V), average stem diameter weighted with basal area (DW), stem number (Nstems), and for the individual stem volumes for pine (Vma), spruce (Vku) and broad-leaved trees (Vlepu). The last three estimates were then used to calculate the tree species proportion estimates with an external program (ErMapper).

The root mean squared error (RMSE), the relative root mean squared error (RMSE%) wrt. mean of the variable, and the correlation of the variable estimate with ground data were then computed to obtain the estimation accuracy. Due to the small size of the ground data sets, a 5-fold cross-validation was used to the model validation. The shown figures of merit are the mean values of the values calculated for K=5 partial data sets.

Table 1 shows results for the estimation obtained with spectral features only (baseline set).

VARIABLE	RMSE	RMSE% [% of mean]	Corr. coeff
<b>Total stem volume</b>	56.6 m <sup>3</sup> /ha	38.9	0.81
<b>Mean diameter</b>	4.2 cm	24.0	0.86
<b>Stem number</b>	953 stems/ha	90.6	0.45
<b>Pine percentage</b>	26.9 %	132	0.64
<b>Spruce percentage</b>	30.5 %	60	0.69
<b>Broad-leaved percentage</b>	25.0 %	166	0.49

Table 1 Figures of merit for the baseline estimation

Table 2 shows the corresponding figures for estimation with feature set vol\_1 = spectral channels + Haralick entropy.

VARIABLE	RMSE	RMSE% [% of mean]	Corr. coeff
<b>Total stem volume</b>	54.4 m <sup>3</sup> /ha	37.4	0.82
<b>Mean diameter</b>	4.1 cm	23.4	0.87
<b>Stem number</b>	917 stems/ha	87.1	0.50
<b>Pine percentage</b>	23.4 %	111	0.74
<b>Spruce percentage</b>	25.1 %	47	0.79
<b>Broad-leaved percentage</b>	21.8 %	137	0.53

Table 2 Figures of merit for feature set vol\_1

Examples of scatterplots for parameter estimates vs. ground reference data are presented in Figure 6, Figure 7 and Figure 8 (stem volume (V), average stem diameter (DW), and spruce percentage (Vku) respectively). The estimate images for the same variables are presented in Figure 9, Figure 10 and Figure 11 respectively.

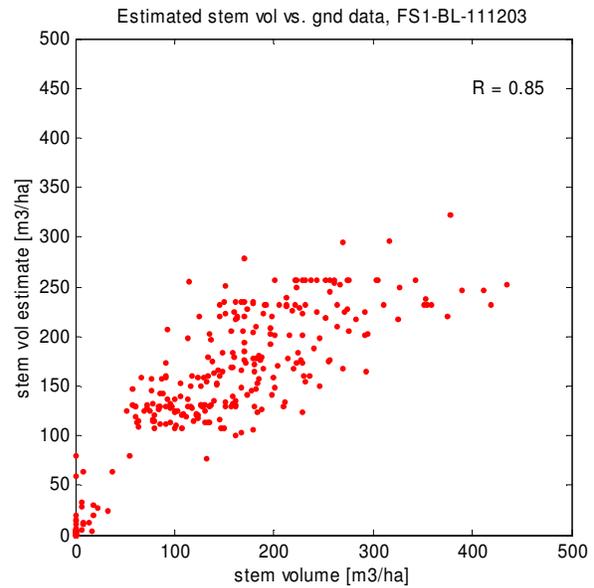


Figure 6 Stem volume(V) estimate plotted against ground truth data (N = 296), baseline feature set

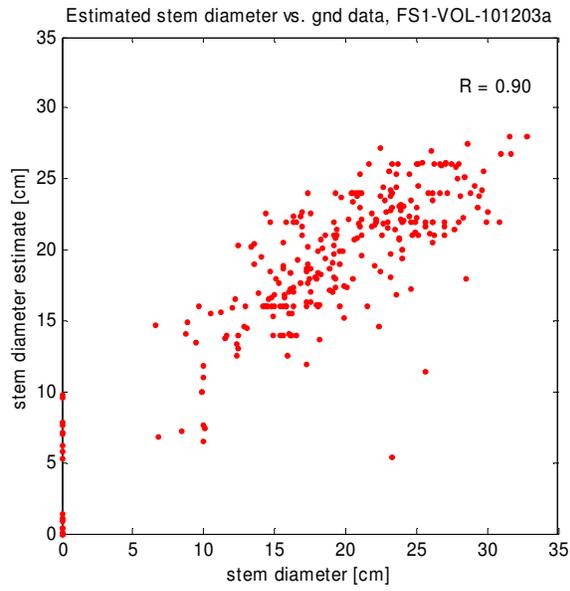


Figure 7 Average stem diameter (DW) estimate plotted against ground truth data (N = 296)

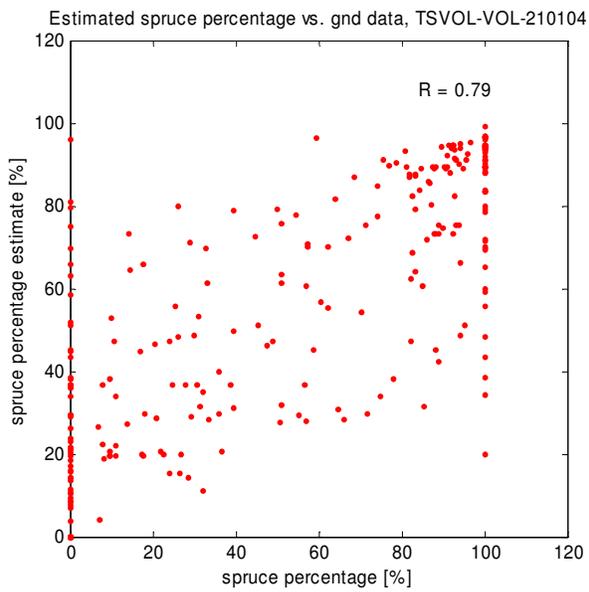


Figure 8 Estimated percentage of spruce plotted against ground truth data (N = 296)

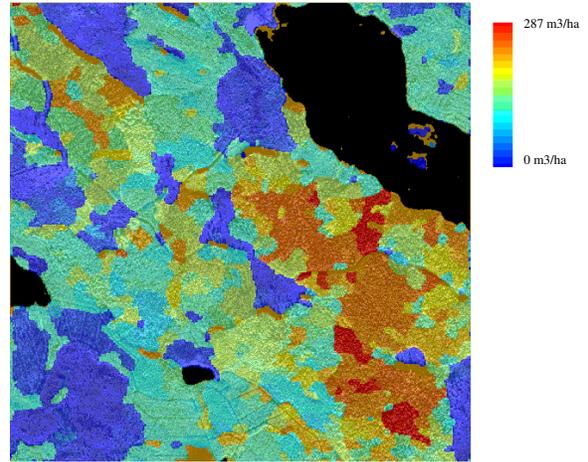


Figure 9 Stem volume estimate (input features B,G,R,NIR,Har1)

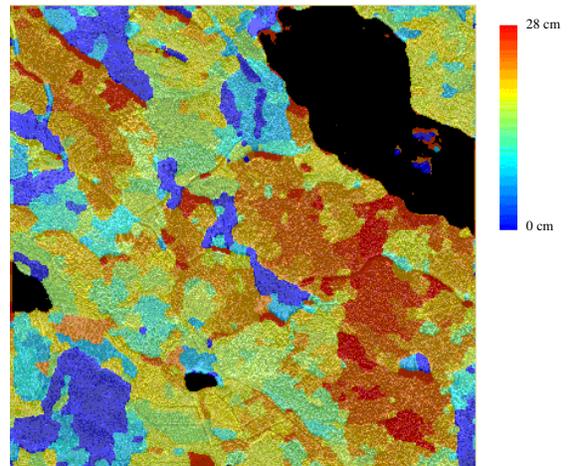


Figure 10 Stem diameter estimate (input features B,G,R,NIR,Har1)

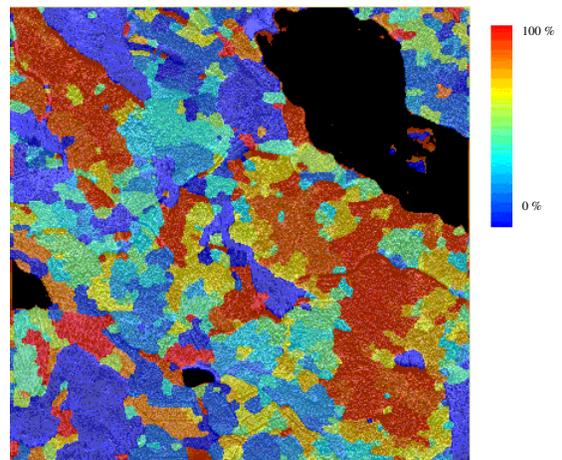


Figure 11 Percentage of spruce [%] (input features B,G,R,NIR,Har1)

At the current state Forestime lacks some key features and functionality is still missing (e.g. the necessary user interaction in the estimation process modeling). However, with a reasonable effort Forestime can be developed into a fully-featured operative forest variable estimation tool.

The results from the forest variable estimation tests show that the relative errors stem volume and stem diameter estimates are relatively small, while stem number and tree species proportions estimates give higher errors. The upper limit for stem volume estimates with the Forestime is around 300 m<sup>3</sup>/ha, while with previous estimates the limit has been somewhere between 200 m<sup>3</sup>/ha and 250 m<sup>3</sup>/ha. The large errors in the tree species proportions are partly due to the fact that the proportion of the broad-leaved trees is quite small in the ground data set.

The ground data validity is essential: the removal of erroneous data points, and points residing too close to the border of very different segments improved the estimate accuracy drastically. The plot data obtained using a regular sampling grid is not optimal for a system using segmentwise averaged feature vectors as its inputs. The usage of this kind of ground data requires additional manipulation and probably also removal of relatively many sample data points. The best option would be accurate standwise data with the average stand size close to average Forestime segment size. In the future the availability of accurate reference data obtained by airborne imaging may prove useful. One has to remember also that the amount of the ground data together with the range of the target variables affects directly to the accuracy of the estimates.

## 5. CONCLUSIONS

In HighForest study VTT implemented the prototype version of a modular software tool for forest variable estimation (Forestime v. 1.0). Even if the functionality, speed and the resource consumption of the implemented software are still very limited, the operational tests have shown the feasibility of the system. The software was tested as a stand-alone system, as well as an external application program from commercial software through the API. The operation of Forestime was as expected and the objective for the software implementation was well achieved.

Summarizing the results of forest variable estimation with Forestime v 1.0, it can be stated that the objective to reduce the reflectance saturation effect was achieved partly, as well as the objective for accurate tree species estimation. The best results were achieved using the spectral channels and the Haralick entropy as input features. These are the recommended inputs for the use of the system as its present form. The estimation speed can still be improved, if only the spectral channels are used.

In the produced forest variable estimations, the target data variance in the clusters is relatively large, which leads to averaged estimates. One main issue of the future studies will be the possibility to reduce the variance using different kind of clustering, e.g. by taking the target variable information into the process, and its effect on estimate accuracy. The future development of the system should also focus on taking full advantage of the contextual features in segmentation, and on introducing means for easy ground reference data exploration. The development of feature bank concept for the wide operational use of the system is also regarded as one key development areas of Forestime system.

## ACKNOWLEDGEMENTS

The project was conducted in co-operation of four Finnish parties: the Technical Research Centre of Finland (VTT), the University of Helsinki (HY), Stora Enso Forest Consulting (SEFC), and Gisnet Solutions Finland Oy. The HighForest project had two parallel studies: the satellite image study (VTT) and the aerial image study (HY). The Suonenjoki Research Station of the Finnish Forest Research Institute (FFRI) provided the sample plot ground data (measured in summer 2001) for the system verification. Sixty percent of the funding was granted by the National Technology Agency of Finland (Tekes).

## REFERENCES

Haralick., R.M.; Shanmugam, K.; Dinstein, I. 1973: Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics, 3:610-621

Jain, A.K. and Farrokhnia, F. 1991: Unsupervised Texture Segmentation Using Gabor Filters, Pattern recognition, Vol. 24, No. 12, pp. 1167-1186

Narendra, P.M., Goldberg, M. 1980. Image segmentation with directed trees. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, Nr. 2, March 1980.

Häme, T.; Stenberg, P.; Rauste, Y. A methodology to estimate forest variables at sub-pixel level. - Conference on Remote Sensing and Forest Monitoring. Rogow, PL, 1 - 3 June 1999. European Communities (2000), s. 451 - 461.

Häme, T.; Stenberg, P.; Andersson, K.; Rauste, Y.; Kennedy, P.; Folving, S.; Sarkeala, J.. AVHRR-based forest proportion map of the Pan-European area. - Remote Sensing of Environment. (2001) 77, s. 76 - 91.

Parnes, E.; Lohi, A.; Häme, T.; Holm, M.; Väättäinen, S. 2000. Pienkuviollaisen tiedon tuottaminen metsistä kaukokartoituksen avulla. Loppuraportti, VTT Automaatio, 2000.

Parnes, E. 1992. Segmentation of Spot and Landsat satellite imagery. The Photogrammetric Journal of Finland, Vol. 13, Nr. 1, 1992. pp. 52 – 58.

Wulder, M.; Niemann, K.O.; Goodenough, D.G. 2002: error reduction methods for local maximum filtering of high spatial resolution imagery for locating trees, Canadian Journal of Remote Sensing, Vol. 28, No. 5, pp. 621-628

### References from websites:

Space Imaging; IKONOS Relative Spectral Response and Radiometric Cal Coefficients ©2004 Space Imaging, Inc., <http://www.spaceimaging.com/products/ikonos/spectral.htm>