The Theory of Random Sets as Flexible Texture Descriptor for Biological and Medical Objects and Self-Similarity as Feature Descriptors for the Description of the Appearances of Cells and Motion

Petra Perner

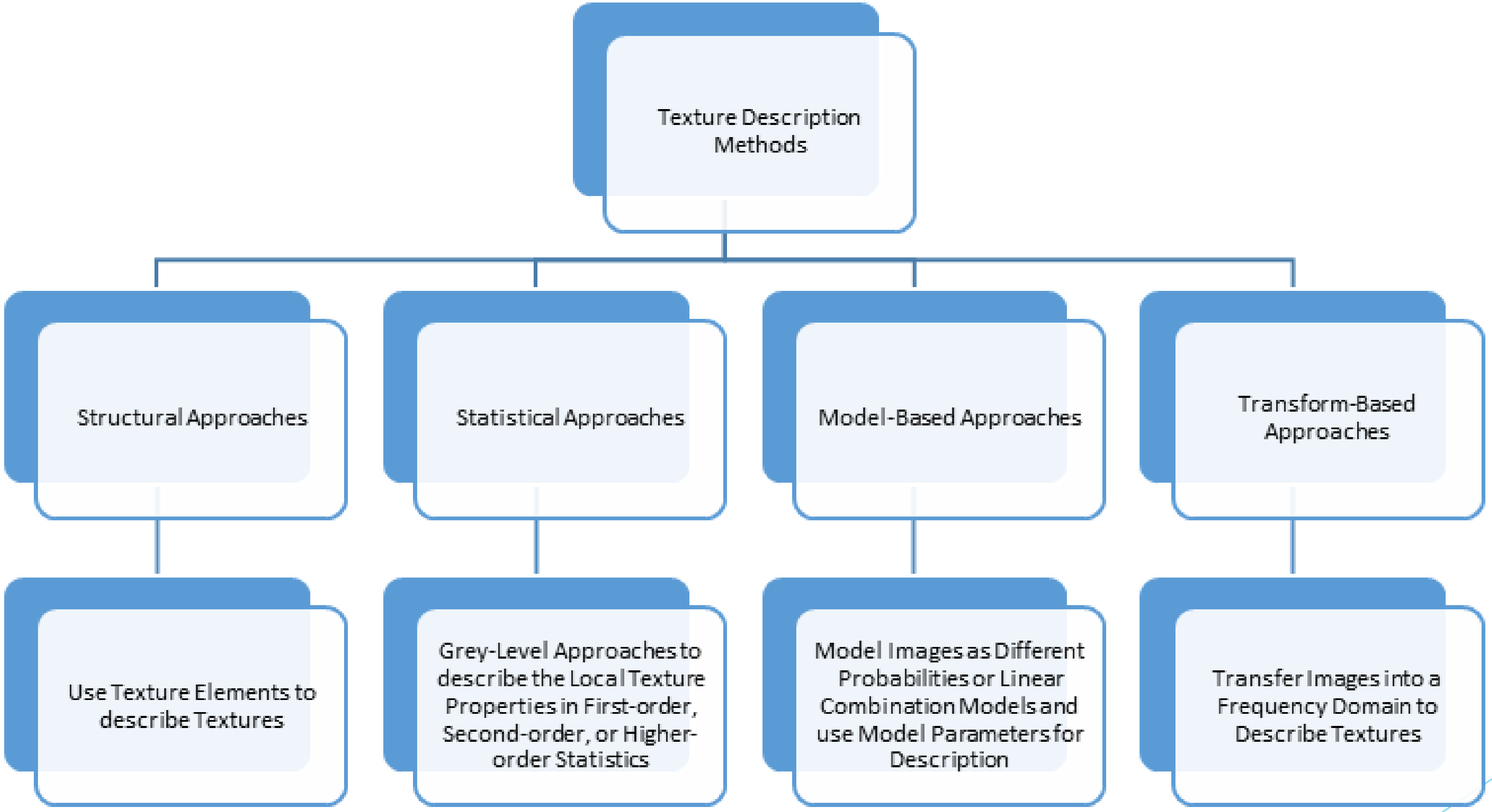
Institute of Computer Vision and Applied Computer Sciences, IBaI

Leipzig, Germany

Content

- Introduction
- Related Work
- Texture Descriptor based on Random Sets
- Material and Application
- Results
- Discussion
- Conclusions

Related Work



Related Work

- Often texture descriptors are compared on standard texture data sets
- Recently appeared work of texture description for real world problems such as description of objects in medical images, microscopic images for different purposes such as e.g. in system biology and for environmental applications, food inspection and so on.
- Texture became a valuable information about images.
- Researcher try to develop many new texture descriptor that take into account the variances of the texture, the spectral influences and so on.
- At lot of different methods exist and it is not easy to do a categorization of all these methods. Often that are variants of the above-described categories that have been evaluated on standard data sets.
- However nowadays, more work on real world applications appear.

Related Work

- Cheng et. al propose a texture method based on the co-occurrence matrix to detect colorectal polyps in colonoscopy images. They used support vector machines for classification and achieve a sensitivity of 86,2%.
- We have developed our own texture descriptor based on statistics that model the texture by a Poisson process after the image has been processed by a morphological operation.
- The remaining areas in the images can be described by first-order and second-order statistics as well as higher-order statistics if the number of remaining areas are large enough.
- The texture descriptor can be easily and fast computed and can handle different medical textures very well.

Related Work

- These medical textures are often not easy to describe as it is in case of the Brodatz texture data set.
- Our method has also explanation capability.
- A human can understand the differences in the texture by looking up the remaining images.
- If necessary, a symbolic description of the different textures can be found.
- Our texture descriptor has still some other properties that are of interest but here in this paper, we want to compare our texture descriptor to the co-occurrence matrix since it is from the category of statistic texture descriptors.
- The co-occurrence matrix is still the most used texture descriptor and we want to explore the differences between our texture descriptors and the co-occurrence matrix.

Texture Descriptor based on Co-occurence Matrix

A co-occurrence matrix $C_{(\Delta x, \Delta y)}$ with the offset $(\Delta x, \Delta y)$ is defined over an *nxm* Image *I*:

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, & I(p,q) = i \text{ and } I(p+\Delta x, q+\Delta y) = j \\ 1, & I(p,q) = j \text{ and } I(p-\Delta x, q-\Delta y) = i \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{(1)}$$

Features that can be calculated:

1. Angular Second Moment
2. Contrast
3. Correlation
4. Entropy
5. Sum of Variance, etc.

Texture Descriptor based on Random Sets
The Boolean Model

- The Boolean model allows to model and simulate a huge variety of textures e.g. for crystals, leaves, etc..
- The texture model X is obtained by taking various realizations of compact random sets, implanting them in Poisson points in Rn, and taking the supremum. The functional moment $Q(B)$ of X, after Booleanization, is calculated as:

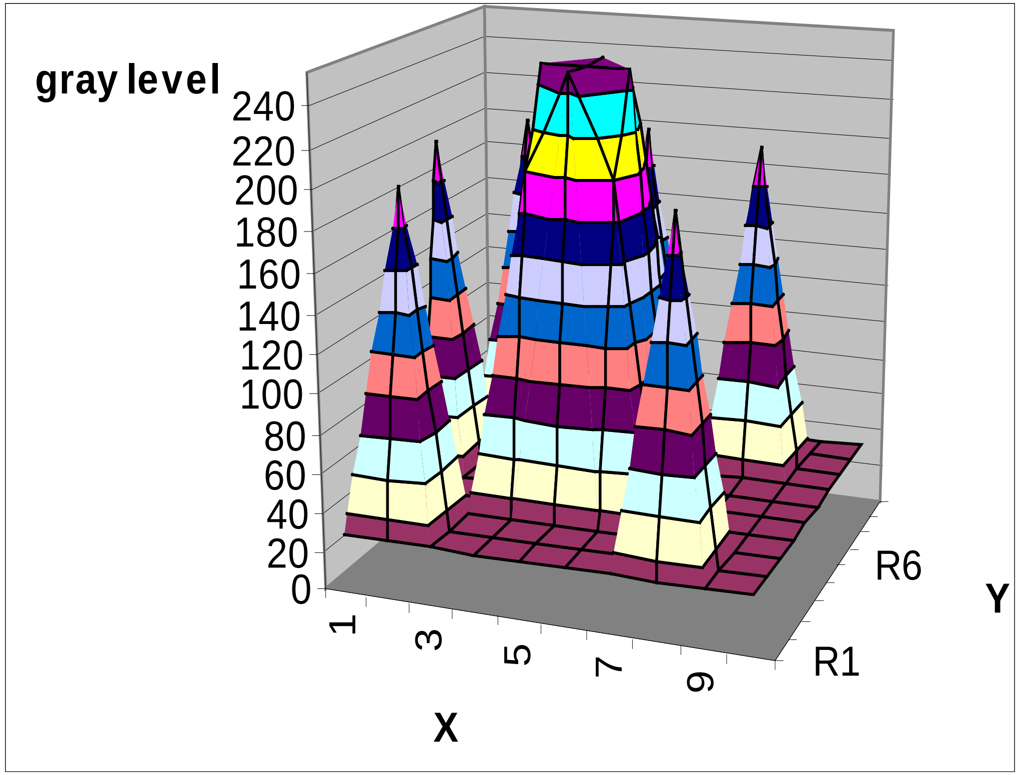
$$P(B \subset X^c) = Q(B) = \exp(-\theta \overline{Mes}(X' \oplus \check{B})) \qquad \forall B \in \kappa$$

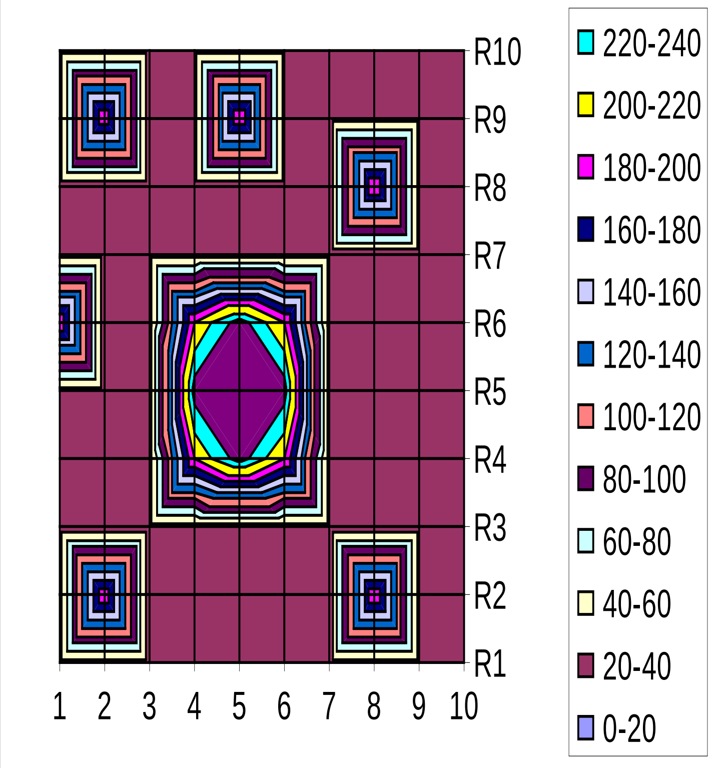
- where κ is the set of the compact random set of R_n, θ the density of the process and $\overline{Mes}(X' \oplus \check{X})$ is an average measure that characterizes the geometric properties of the remaining set of objects after dilation.

Texture Descriptor based on Random Sets The Boolean Model

- Formula is the fundamental formula of the model. It completely characterizes the texture model. $Q(B)$does not depend on the location of B, i.e., it is stationary. One can also provide that it is ergodic so that we can peak the measure for a specific portion of the space without referring to the particular portion of the space.
- Formula shows us that the texture model depends on two parameters:
- • the density θ of the process and
- • a measure $\overline{Mes(X' \oplus \check{B})}$that characterizes the objects. In the one-dimensional space it is the average length of the lines and in the two-dimensional space $\overline{Mes(X' \oplus \check{B})}$ is the average measure of the area *A* and the perimeter of the objects under the assumption of convex shapes.

Texture Descriptor based on Random Sets

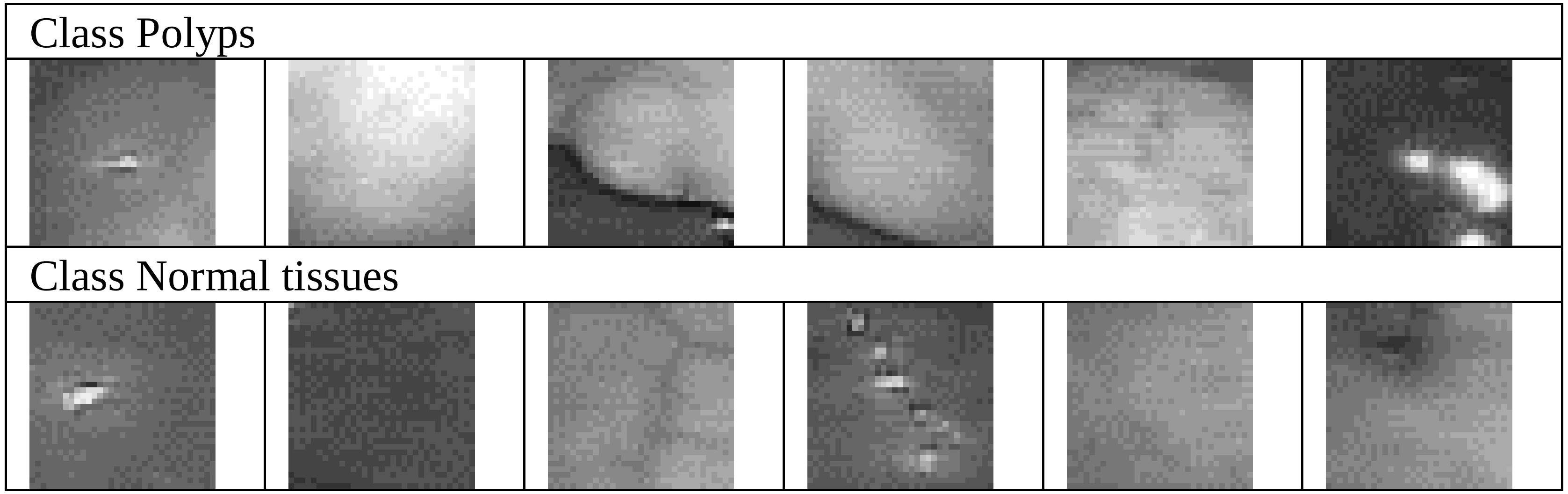




Texture Descriptor based on Random Sets

Description	Name	Type	Formula
Area in class image *t*	***Area_t***	num	$Area_t = \begin{cases} Area_t = Area_t + 1 & \text{if } f(x,y,t) = 1 \\ Area_t = Area_t & \text{if } f(x,y,t) = 0 \end{cases}$
Density in class image *t*	***Dens_t***	num	$Dens_t = \begin{cases} Dens_t = Dens_t + \frac{1}{A} & \text{if } f(x,y,t) = 1 \\ Dens_t = Dens_t & \text{if } f(x,y,t) = 0 \end{cases}$ with $A = \sum_{t=1}^{S} Area_t$
Number of objects	***Count_t***	num	$n(t)$
Mean area of objects in class image *t*	***AreaMean_t***	num	$\overline{A(t)} = \frac{1}{n(t)} \sum_{i=1}^{n(t)} A_i(t)$
Standard deviation of the area of the objects in class image *t*	***AreaStdDev_t***	num	$S(t) = \sqrt{\frac{1}{n(t)} \sum_{i=1}^{n(t)} (A_i(t) - \overline{A(t)})^2}$
The contour length of a single object is $u = l + \sqrt{2} \cdot m$ with *l* being the number of contour pixels having odd chain coding numbers and *m* being the number of contour pixels having even chain coding numbers.			
Mean contour length of objects in class image *t*	***ContMean_t***	num	$\overline{u}(t) = \frac{1}{n(t)} \sum_{i=1}^{n(t)} u_i(t)$
Standard deviation of the contour length of objects in class image *t*	***ContStdDev_t***	num	$S(t) = \sqrt{\frac{1}{n(t)} \sum_{i=1}^{n(t)} (u_i(t) - \overline{u(t)})^2}$

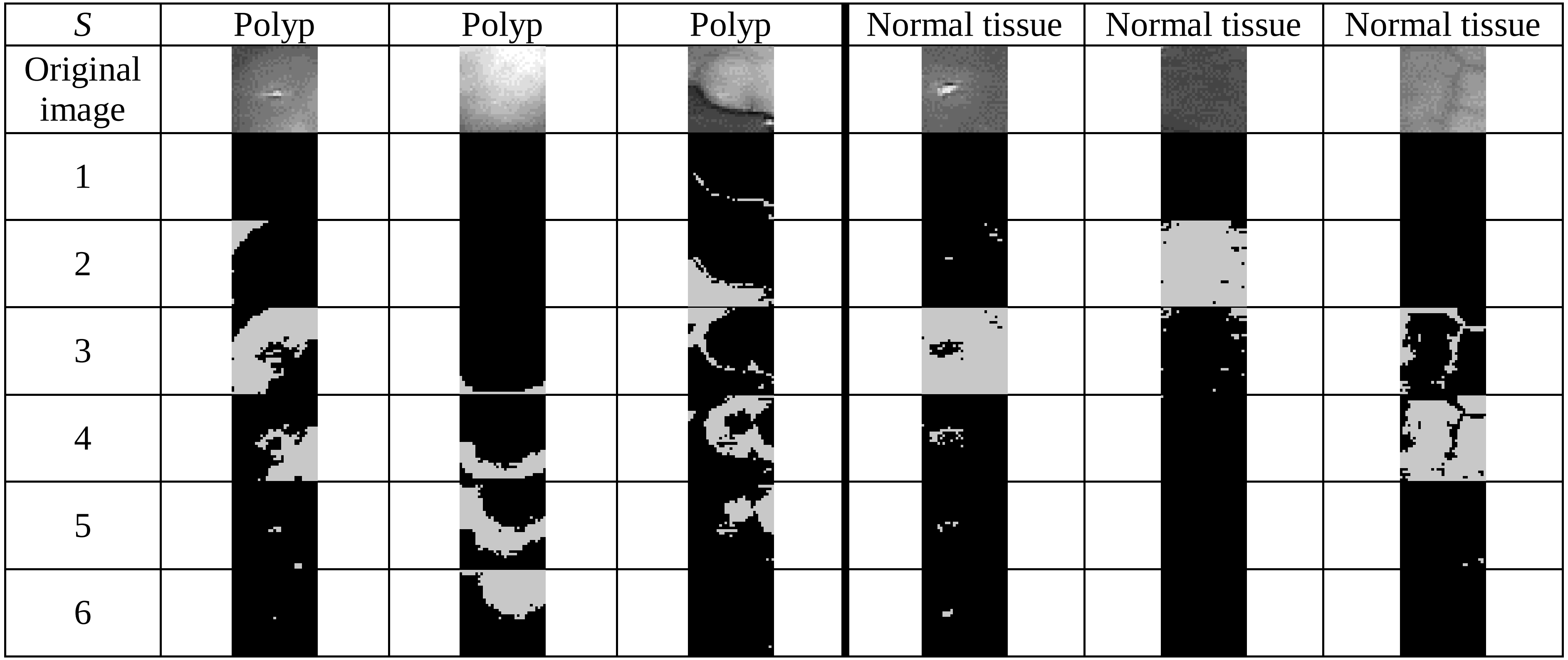
Material and Application



Material and Application

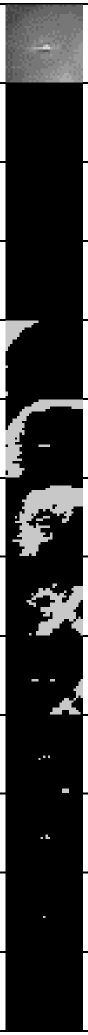
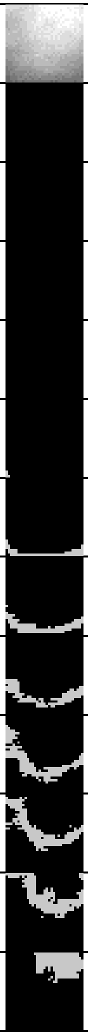
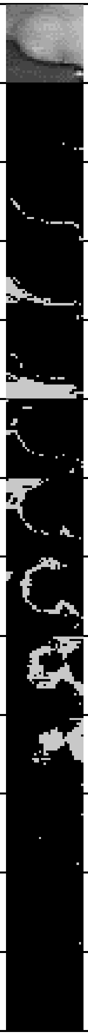
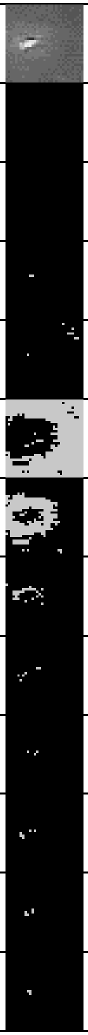
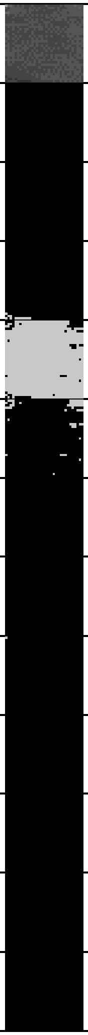
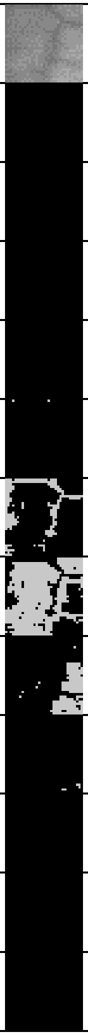
- We studied the performance of the two texture descriptors based on a data set of 344 images.
- These images come from an endoscopic video system used for colon examination.
- The data set contains 283 normal tissue images and 61 polyp images in the form of sub-images of a size 33x33 that are derived from 37 original colonoscopic images.
- The polyps in the 37 original colonoscopy images were identified and selected by a “well-trained” medical expert.
- A polyp is split into as many as possible sub-images.
- The 283 normal images consist of dark regions, reflections etc. of the 37 original colonoscopy images.
- This presents a two class problem; one must decide if the image shows a polyp or not.
- The texture descriptions were calculated from these images.
- The resulting data set was used to train a decision tree based on the C4.5 algorithm. Cross-validation was used to estimate the error rate.

Results

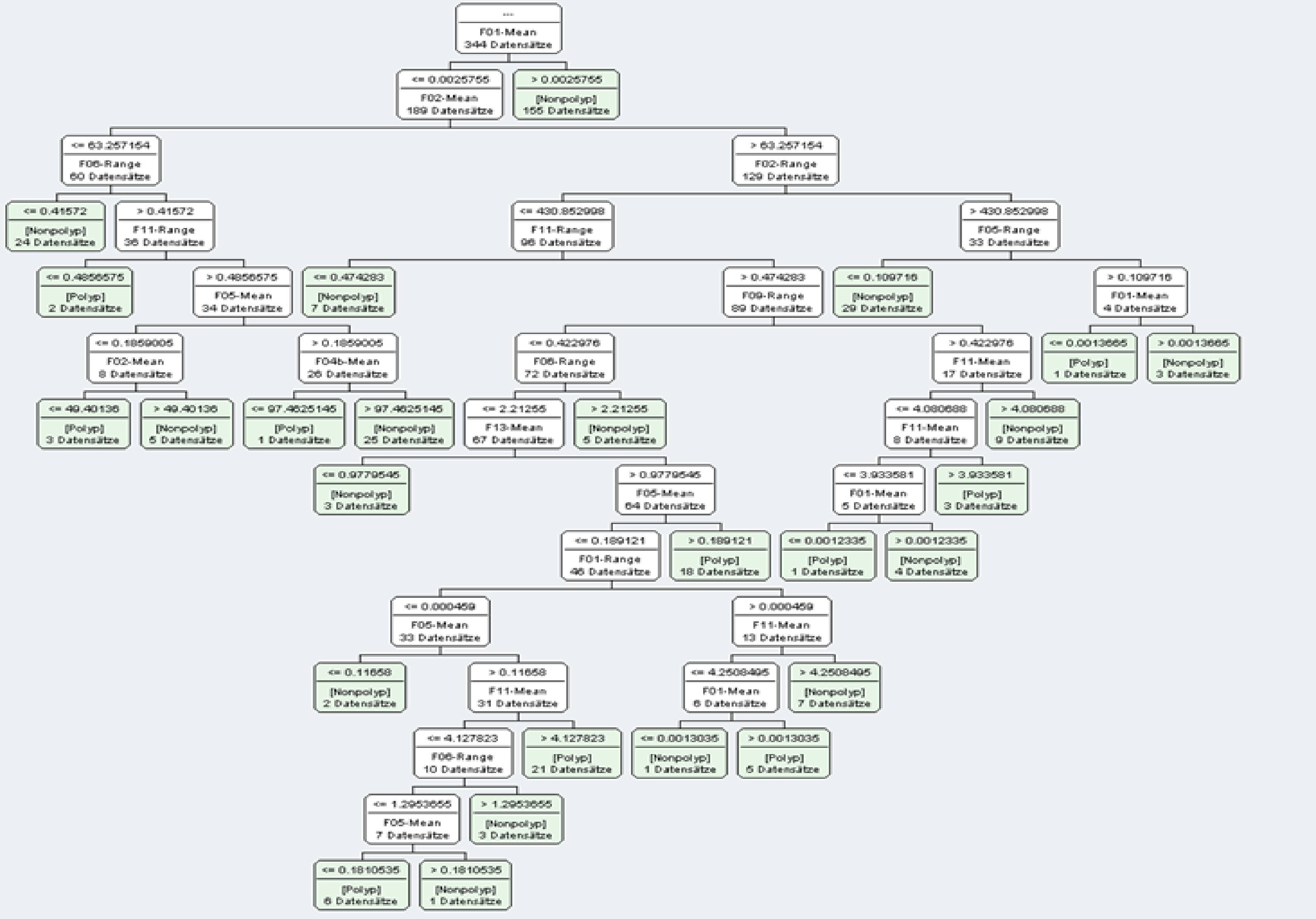


The images f(x,y,t) with S=6

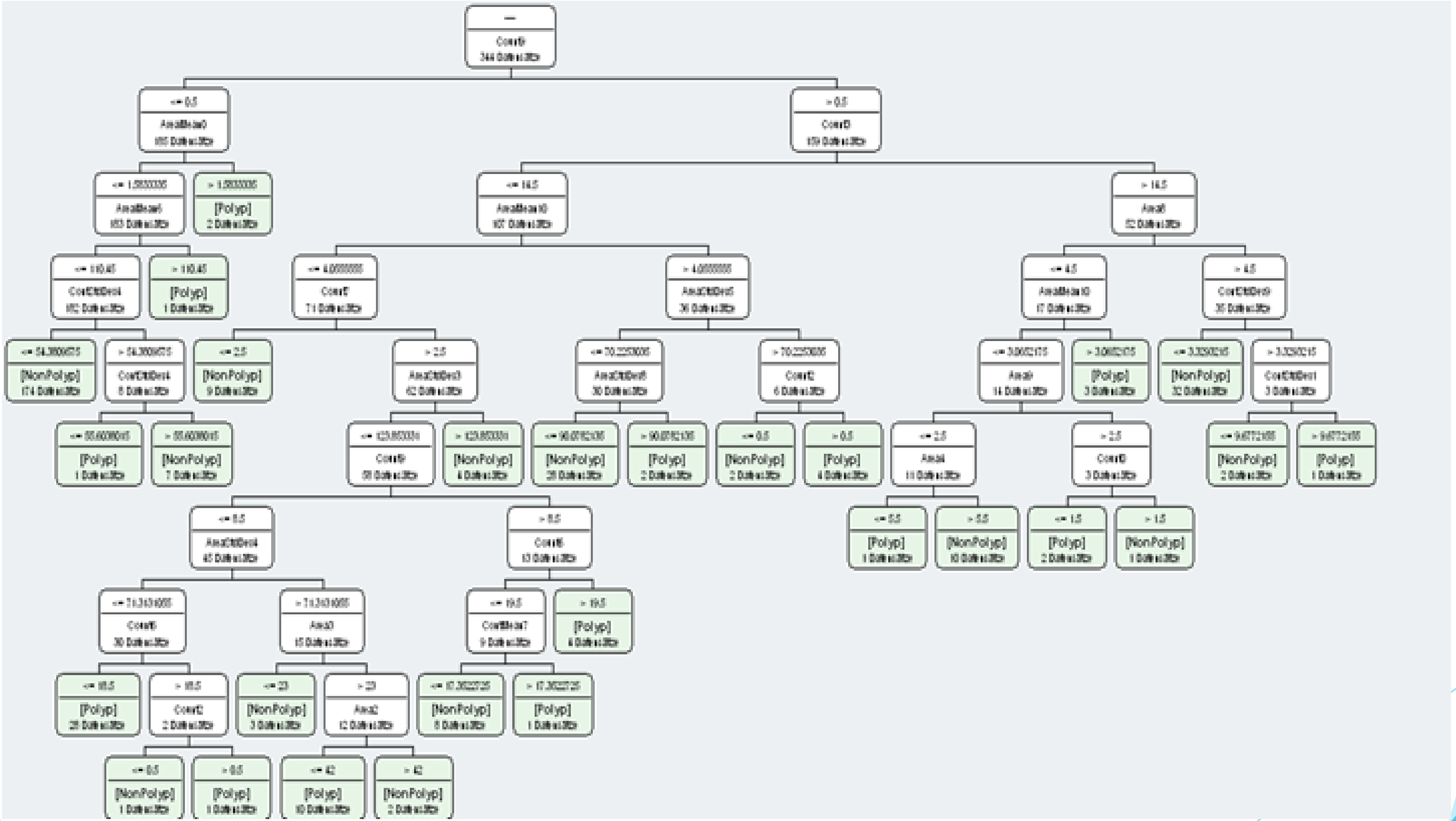
Results for the Images $f(x,y,t)$ with $S=12$

S	Polyp1	Polyp6	Polyp20	Normal tissue	Normal tissue	Normal tissue
Original						
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

Results for COO Features



Decision Tree for Random Sets Features



Results

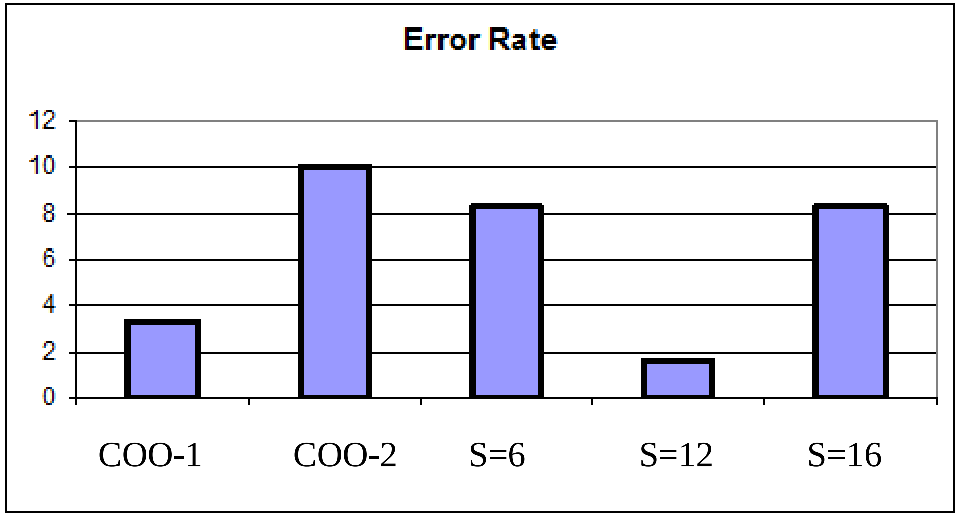


Fig. 5. Error rate (in percent) for Test 1

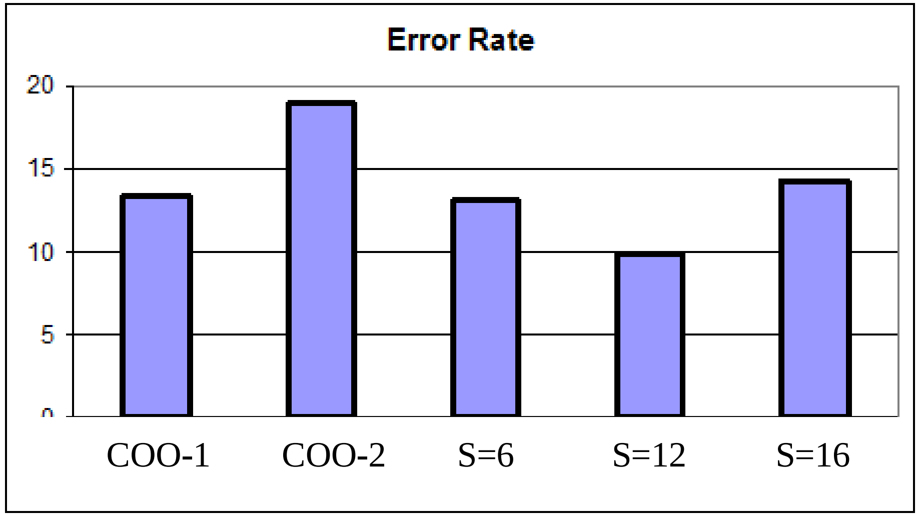


Fig. 6. Error rate (in percent) for Test 2

Runtime	COO-1	COO-2	Texture Descriptor based on Random Sets
	91.03s	83.22s	13.75s

Discussion

- In this application the texture descriptor based on random sets outperformed the COO texture descriptor. The accuracy is 3.49 % higher than that of COO texture descriptor in case of COO-1 and 9.01% higher in case of COO-2.
- Decision trees are sensitive to unbalanced class distribution. Therefore, the error rate in the second experiment rises since the ratio of the two classes is 1/5 in the data set. Nonetheless, the tendency of the error rate of the three descriptors is the same.
- Advantage of the texture descriptor based on random sets over COO texture descriptor is the reduced time required for computing the features.
- In addition, we can understand the semantics behind the numerical texture description. The texture features based on random sets have a semantic meaning and give an expert an understanding about texture (see Table 1).
- The choice of the number of slices S emerges to S=12 in all the applications we have done. The number S=12 provides a feature set of 84 features. It might be that this is a compromise between a rich description of texture and the large feature set problem (curse-dimensionality).
- The decision tree induction method performs feature selection during the tree building process. Therefore, the method can also be seen as a feature selector. The number of features selected for COO texture descriptor is always lower than the number selected for the texture descriptor based on random sets. The texture descriptor based on random sets may provide a more richer description of texture. Features from almost all slices are included in the decision.

Conclusions

- To study human image cognition is more than ever an important topic since the number of vision-based materials has been increased over the years.
- We have studied the human image cognition based on texture for medical images.
- Texture seems to be a powerful tool to describe the appearances of objects.
- Therefore, very flexible and powerful texture descriptors are of importance that allow to recognize the texture and to understand what makes up the texture.
- We give in our paper the methodology how to study the human image cognition by automatically calculating texture descriptors from a set of images, using decision tree induction in order to learn the classifier, and recognizing the performance of the texture-based object recognition by performance measures such as accuracy, run-time, explanation capability.
- Many texture descriptors are known from the literature. The most used texture descriptor is the texture descriptor based on the co-occurrence matrix.
- We proposed a texture descriptor based on random sets and in this paper compared both texture descriptors based on polyp images that were derived from colon examination. We learnt a classifier model based on decision trees. Then we compared both texture descriptors.
- We have found that the texture descriptor based on random sets outperform COO texture descriptor based on the error rate, tree properties and the runtime. COO texture descriptor uses fewer features from the set of calculated texture features than the texture descriptor based on random sets. However, this might only demonstrate that COO texture descriptor has limited description power since the error rate is much higher than that for the texture descriptor based on random sets.
- In addition, the texture descriptor based on random sets has semantic meanings. An expert can understand the properties of a texture when looking into the slices produced during the calculation of the texture features. The medical texture object are often not large objects. That limits the statistics we can use. Higher-order statistics make no sense since the number of objects gets less. Further work will study the behavior of our texture descriptor when the objects are large.

Acknowledgement

- This work has been sponsored under the grant title “Study of the Cognitive Aspects of Human Vision” CogVision under the grant number IS 2012-4.

Prof. Dr. Petra Perner

Institute of Computer Vision and

Applied Computer Sciences, IBaI

Kohlenstr. 2, Leipzig, Germany

www.ibai-institut.de

Thank you for your Attention!