

# Quantitative description of 3D vascularity images: texture-based approach and its verification through cluster analysis

Artur Klepaczko · Marek Kociński ·  
Andrzej Materka

Received: 7 October 2009 / Accepted: 28 October 2010 / Published online: 4 December 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** This paper undertakes the problem of quantitative inspection of 3D vascular tree images. Through the use of cluster analysis, it confirms the correspondence between texture descriptors and various vessel system parameters, such as blood viscosity and the number of tree branches. Moreover, it is shown that unsupervised selection of significant texture parameters, especially in the synthetic data sets corresponding to noisy images, becomes feasible if the search for relevant attributes is guided by the clustering stability-based optimization criterion.

**Keywords** Vascularity image synthesis ·  
Cluster analysis · Unsupervised feature selection

## 1 Introduction

Image diagnosis of the vascular system provides fundamental information that facilitates detection of abnormal alternations present in its geometrical architecture. These may denote certain pathologies and thus the need for clinical or even surgical treatment. Thus, much effort has been put on developing novel imaging techniques which enhance visualization of arteries and veins. The 3D raster images can be acquired using susceptibility weighted imaging combined

with time-of-flight (SWI-TOF) sequences [2] under the magnetic resonance (MR) control or by the help of a confocal microscope. With the development of computational intelligence methods, image-based diagnostic procedure can overcome the restrictions of qualitative image analysis. It arises that a method for automatic recognition of the pathology type or its localization would provide significant support to a diagnostician. Quantitative description of various malformations or stenosis can aid in deciding on appropriate therapy.

In a general setting, vascular system can be viewed at three levels, depending on a vessel diameter size. The width of the largest, first-order vessels reaches almost 3 cm diameter of medium-sized veins and arteries is equal to approximately 1–5 mm. At the bottom-most level, there remain venules and arterioles, and capillaries, whose size does not exceed 10  $\mu\text{m}$  [3]. At each mentioned level, vascularity images involve specific processing techniques. Diverse approaches result from the relation between vessel size and image resolution (see Fig. 1).

At the top-most level, arteries and veins can be effectively tracked in MR images. As a result, geometrical architecture of a visualized portion of vascular system can be numerically described. This kind of analysis is restricted to vessels whose diameter exceeds image resolution. On the other extreme, the smallest vessels are too thin to be captured by any of the realistic state-of-the-art 3D imaging modality. Their quantification can be obtained indirectly, e.g. by measuring perfusion in the dynamic contrast susceptibility examination [4].

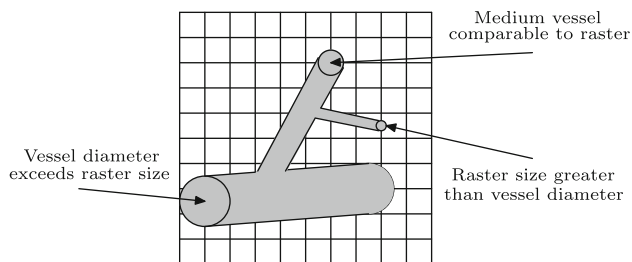
Similarly, segmentation of medium vessels, whose thickness is comparable to MR image resolution (0.5 mm voxel size for a 7-T magnet), remains a relatively difficult task. However, vasculature at this level still allows imaging which conveys important structural patterns inherent in the

---

This is the extended version of the paper: Klepaczko A et al (2009) Cluster analysis in application to quantitative inspection of 3D vascular tree images. In: Kurzynski M, Wozniak M (eds) Computer recognition systems 3, Advances in Intelligent and Soft Computing, Springer, Heidelberg, pp 87–94 [1].

---

A. Klepaczko (✉) · M. Kociński · A. Materka  
Institute of Electronics, Technical University of Lodz,  
ul. Wolczanska 211/215, 90-924 Lodz, Poland  
e-mail: aklepaczko@p.lodz.pl; arthur.klepaczko@gmail.com



**Fig. 1** Relation between image resolution and a vessel size

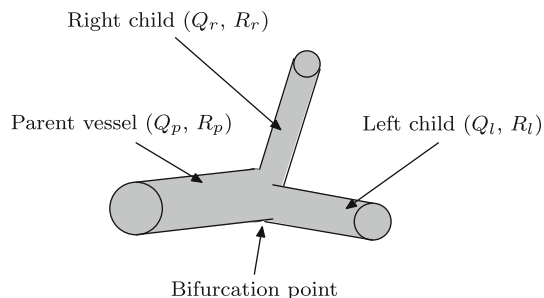
vessel system. Through computer simulation and experiments reported in this study, we argue for the usage of texture analysis in order to encapsulate quantitative characteristics of vascular tree images. It will be shown that there exists direct correspondence between physical blood parameters—determining a tree growth—and relevant texture features. Exploiting this dependence, it is possible to automatically classify vasculature images according to a series of criteria.

The experimental results presented so far in this area were derived with the use of supervised classification and feature selection techniques [5]. However, it can be argued that such an approach can artificially bias the constructed classification rule toward a predetermined assumptions inherent in a training data set. In this paper, the findings grounded on supervised learning methods will be further validated in the unsupervised manner.

## 2 Vascular tree image synthesis

Under regular clinical examination, image diagnosis is performed qualitatively. For this reason, data acquisition procedure not necessarily aids in quantitative inspection. This refers to both image resolution and noise. Good-quality scans, adequate for quantitative analysis, usually require time-consuming measurement sequence and also more computational power in the image reconstruction phase. In the qualitative diagnosis, low or medium resolution images often appear sufficient, what shortens acquisition time but simultaneously degrades efficiency of image processing methods.

The above-described circumstances disallow statistically credible verification of the proposed texture-based approach using real images. Therefore, in the experimental part of this study, a vessel tree algorithm is employed, and a series of synthetic vascularity images is generated. Naturally, the synthesized images will differ from the real ones, which contain noise and various imaging artifacts. However, several studies reported in the literature [6, 7] utilize vessel tree models and their raster representations to develop image processing methods capable of extracting



**Fig. 2** The adopted bifurcation model

quantitative information also from the real CT or MRI scans.

### 2.1 Vascularity model

The adopted approach to vascular tree growth follows the scheme proposed in [6]. It assumes cylindrical shape of a vessel with fixed radius along its entire length. Furthermore, construction of a tree model is based on the rules described below.

1. Tree growth is parameterized by:

- number of output branches,
- blood viscosity,
- input and output blood flow,
- input and output blood pressure,
- perfusion volume.

By varying one of the above arguments with the other kept unchanged or modified accordingly, it is possible to obtain a series of vascularity classes differing with respect to a chosen parameter.

2. In a single step of a tree construction one new branch is added. This action starts with random selection of a new vessel endpoint (within a predefined organ shape; in this study sphere was used). Then the algorithm searches for the nearest branch. At the center point of that branch, a temporary bifurcation is made. The parameters for all branches are adjusted so that the following laws hold:

- matter preservation law,

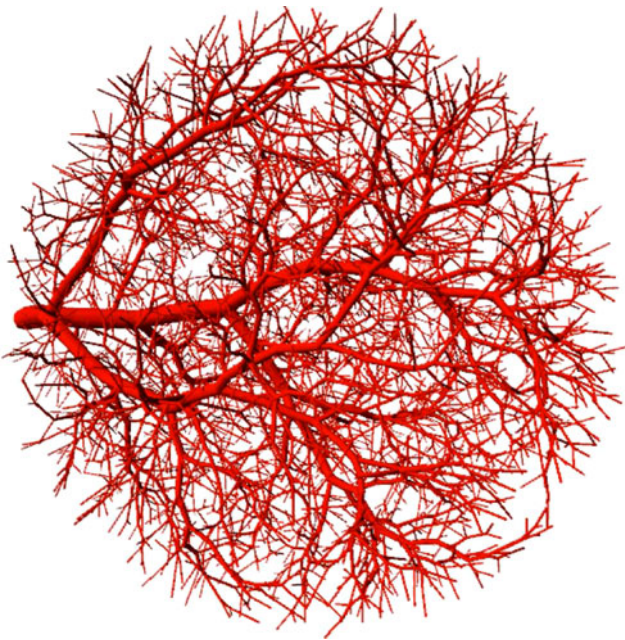
$$Q_p = Q_l + Q_r, \quad (1)$$

which couples flows ( $Q$ ) of parent and child vessel segments of a single bifurcation (cf. Fig. 2),

- Poiseuille's law,

$$\Delta P = \frac{8Q\eta l}{\pi R^4}, \quad (2)$$

where  $\Delta P$  denotes pressure drop on a vessel segment,  $\eta$  is the blood viscosity,  $R$  vessel radius, and  $l$  vessel length,



**Fig. 3** Visualization of a vascular tree vector model

- bifurcation law

$$R_p^\gamma = R_l^\gamma + R_r^\gamma, \tag{3}$$

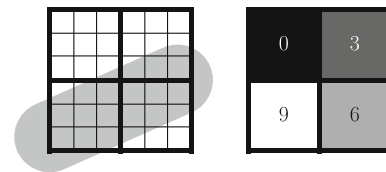
where  $\gamma$  is a bifurcation ratio—a factor that relates parent and child vessel radii.

3. The algorithm requires that a new bifurcation is optimized. For that purpose, a bifurcation point is moved so that the volume of affected vessels is minimized. This is performed through a *constrained constructive optimization* algorithm.
4. The output of the generation procedure is a vector description of a vascular tree. Such a description can be viewed as a list of (1) start and end points (expressed in 3D Cartesian coordinates) which define length and direction of a vessel axis and (2) vessel radius.

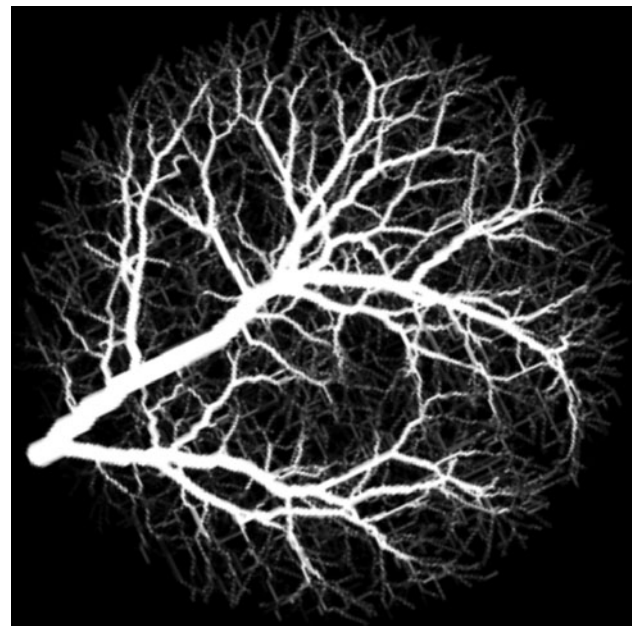
### 2.2 Imaging simulation

Vector description of a vascular tree is convenient for visualization purposes. Figure 3 illustrates example of such a tree model. However, when it comes to development of image processing methods, raster representation of this vector model is needed. In order to convert real-valued data into 3D discrete coordinates, the following imaging simulation procedure is involved.

Each voxel in the scene is assigned some intensity value which is proportional to a vessel volume present in it. In the two extreme cases, when a voxel is either totally filled with a vessel or it does not belong to vascular tree at all the



**Fig. 4** Subpixel analysis in calculation of pixel intensities



**Fig. 5** Example of a generated vascular tree raster image (MIP)

intensity becomes maximal or zero, respectively. To deal with the intermediate cases, it is assumed that a voxel can be divided into a number of subvoxels. In this research, we use 27 equal-sized subvoxels forming a  $3 \times 3 \times 3$  element cube. Then, if a voxel is occupied only partially, its intensity is adjusted to the value proportional to the number of subvoxels whose middle points belong to a vessel. This concept is visualized for the two-dimensional case in Fig. 4. An example of a raster tree image (maximum intensity projection) generated using the above-outlined procedure is depicted in Fig. 5.

The last issue that should be taken into account during vascularity image generation is noise introduced by any real-world image acquisition technique. In this research, in order to make the imaging procedure comparable—at least to some extent—to MR acquisition technique, the noise signal is modeled using the Rice distribution [4]. It is given by the following probability density function

$$f(x|v, \sigma) = \frac{x}{\sigma^2} \exp\left(\frac{-(x^2 + v^2)}{2\sigma^2}\right) I_0\left(\frac{xv}{\sigma^2}\right), \tag{4}$$

where  $x$  denotes a given voxel intensity,  $v$  and  $\sigma$  are the distribution parameters, while  $I_0$  is the modified Bessel

**Table 1** Summary of the experimental data sets (further explanation in the text)

Vascularity parameter	Range	Number of classes		Total number of classes	Number of instances per class
		Per noise level	Per vascularity parameter		
$N_{q_{inp}}$	3,000–5,000	5	5	25	32
$N_{q_{out}}$	3,000–5,000	5	5	25	32
$\eta$	1–10 cP	5	4	20	32

function of the first kind and order 0. Other imaging artifacts (such as chemical shift or intra-voxel dephasing in the case of MRI) are not handled in the current version of the algorithm.

### 2.3 Specification of synthesized data sets

Classification of synthetic vascularity images shall reveal the correspondence between vessels trees architecture and their textural description. Therefore, we formulate three groups of classification tasks corresponding to the following tree growth parameters:

1.  $N_{q_{inp}}$ : number of terminal branches at constant input flow,
2.  $N_{q_{out}}$ : number of terminal branches at constant output flow,
3.  $\eta$ : blood viscosity.

For groups 1 and 2, the number of branches varied between 3,000 and 5,000 with a 500-branch step (5 classes), whereas blood viscosity (measured in the poise units, P<sup>1</sup>) was equal to either 1, 3.6, 5 or 10 cP (4 classes). Within each group, the images were divided into further categories related to the quantization level of introduced noise signal. The Rice distribution parameter  $v$  [cf. (4)], primarily responsible for the noise level, took five optional values 0 (no noise), 1, 3, 5, 10. For each single class and noise level, 32 tree images were synthesized. Texture feature vectors calculated for these images were then combined into data sets corresponding to one quantized value of noise distribution parameter and one vascularity parameter (15 data sets altogether). Table 1 summarizes data sets constructed for the need of the experiments.

### 3 Texture analysis in 3D

Submitting an image to a classification procedure requires formulating its appropriate description tractable by a computer-implemented algorithm. Image characteristics

must reflect significant patterns of the visualized objects. Thus, it has been proposed to represent a vessel tree using texture parameters [8].

The literature describes several texture characterization models [9]. Choosing arbitrarily the proper one for a particular image type and application remains a difficult task. The common strategy is thus calculation of many texture parameters derived from various concepts. Such an approach based on the assumption that at least some of the computed features will reflect important regularities inherent in image. Within this study, three texture models are taken into consideration:

- Co-occurrence matrix (260 features),
- Run-length matrix (25 features), and
- Image gradient (5 features).

Thus, every vascular tree image was assigned a vector of 290 numerical attributes plus a class label. This obviously breeds the problem of feature space dimensionality reduction. It should increase classification accuracy as well as enable identification of attributes which are mostly correlated with the investigated tree growth parameters.

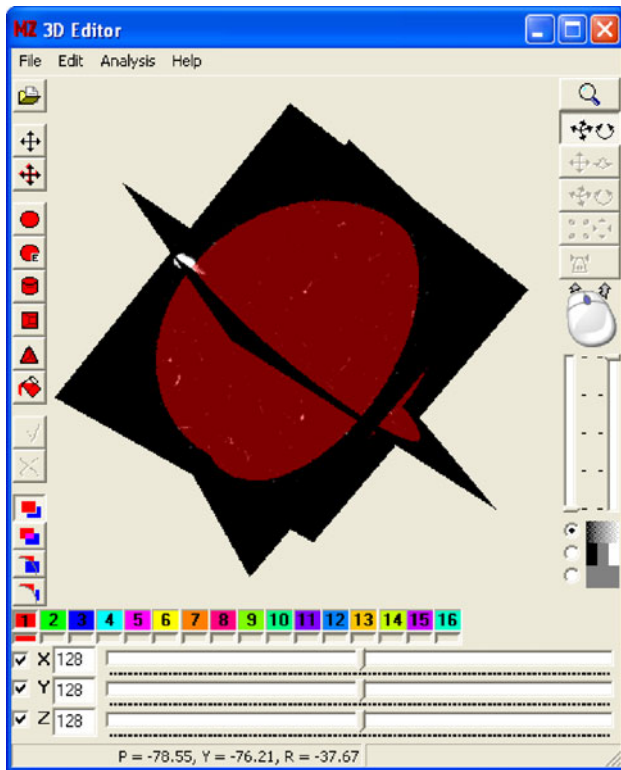
Texture features computation can be performed using one of the freely available software packages [10–12]. Among these, the *MaZda* suite (see Fig. 6), developed with the authors contribution, seems to offer the most versatile functionality [10], especially when 3D analysis is involved. It allows defining volumes of interest (VOI) for an image, calculating texture features within a VOI and classifying feature vectors using either supervised or unsupervised methods

### 4 Dimensionality reduction

As noted above, texture analysis leads to multidimensional data sets, which constitute—as it has been thoroughly described elsewhere (e.g. in [13])—a remarkable challenge for classification algorithms. In this case, dimensionality reduction needs to be involved. Since the goal of this study was to *objectively* confirm validity of the texture model applied to vascularity image classification, dimensionality reduction (as well as classification itself) should be approached in the unsupervised manner.

<sup>1</sup> 1 P = 0.1 Pa. s, 1 cP = 1 mPa. s.





**Fig. 6** 3D editor window of the MaZda software used in the research. The picture presents three orthogonal cross-sections of a synthesized 3D vascularity image and a spherical VOI covering all vessels

Principally, there remain two alternative solutions to the problem—feature *transformation* (projection of original data space into the new space of feature aggregates) and feature *selection* (identifying the most discriminative attributes *in* the original feature space). In the following, it will be shown that feature space transformation exhibits poor effectiveness regarding the aforementioned research goal. It also disallows direct identification of texture features mostly correlated with the vascular tree growth parameters. Therefore, in Sect. 4.2, it will be postulated to use the strategy of feature selection.

#### 4.1 Feature space transformation

Two popular methods of *unsupervised* feature space transformation are principal component analysis (PCA) and random projection (RP). Both algorithms map input feature space onto the new space of feature aggregates, which has lower dimension than the original one and preserves its statistical properties—total data variance or distance relationships in the case of PCA or RP, respectively. Comparing to feature selection, the extraction strategy appears less computationally complex. On the other hand, feature

space transformation not necessarily improves separation between different classes.

To show this effect, performance of PCA and RP in application to vascular tree image classification was verified experimentally. First, prior to employing any feature transformation, each generated data set (cf. Table 1) was clustered using simple *k*-means method. Then data sets were transformed by both PCA and RP algorithms. PCA was set to preserve 98% of the total data variance. The projection matrix in the RP transformation was drawn with the *sparse* distribution function, defined as

$$r_{ij} = \sqrt{3} \begin{cases} -1 & \text{if } u < \frac{1}{6} \\ +1 & \text{if } u > \frac{5}{6} \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

where  $u \in (0,1)$  denotes a random number drawn according to the uniform distribution. After feature space transformation, data were submitted to a clustering procedure, where again *k*-means algorithm was used.

Due to intrinsic tendency of *k*-means to get stuck in a local minimum of the optimization criterion (sum of squared distances between feature vectors and their corresponding cluster centers), every call to clustering procedure involved ten executions with different initializations. Then the result with the best score was chosen for calculating clustering error. The latter was assessed using classes to clusters evaluation method [14]. The whole procedure (feature space transformation followed by clustering) was repeated ten times, and the obtained results were averaged (see Table 2). Note that several invocations are particularly needed for random projection because of its inherently undeterministic behavior. Thus, averaging clustering errors gives more statistically reliable estimates of RP performance.

It is worth noticing, that in the case of two vascularity parameters ( $N_{q_{inp}}$  and  $N_{q_{out}}$ ), RP transformation performs better than PCA. The latter, by exploiting information about data variance, leads to obtaining the so-called *most expressive features* [15]. Their discriminative power is not optimized during transformation. If the original data space contains many insignificant attributes, PCA can even enlarge clustering error. On the other hand, texture-based descriptors occurred remarkably significant for blood viscosity ( $\eta$ ). Here, it is random projection that degrades classification scores. However, generally good results of clustering viscosity classes motivates further effort in finding significant texture features also for the other two parameters. This can be accomplished through feature selection.

#### 4.2 Feature selection

One of the possible approaches to feature selection, the so-called *wrapper-approach* [16, 17], consists of four main steps:

**Table 2** Clustering errors and their corresponding standard deviations (in brackets) before and after feature transformation (%)

Vascularity parameter	Rice distribution parameter $\nu$				
	0	1	3	5	10
Original feature space					
$N_{q_{inp}}$	36.9 (4.0)	39.4 (3.5)	54.3 (10.9)	47.5 (8.3)	53.8 (13.0)
$N_{q_{out}}$	24.4 (3.0)	27.5 (3.0)	28.1 (6.1)	29.4 (4.1)	33.1 (9.3)
$\eta$	0.0 (0.0)	0.0 (0.0)	0.8 (1.1)	0.2 (0.3)	0.3 (0.6)
Principal component analysis					
$N_{q_{inp}}$	50.6 (4.5)	51.9 (2.7)	54.3 (3.5)	57.4 (4.8)	58.9 (5.6)
$N_{q_{out}}$	23.8 (9.5)	29.4 (6.7)	28.8 (10.1)	28.8 (12.6)	29.4 (13.0)
$\eta$	0.0 (0.0)	0.0 (0.0)	0.8 (1.1)	0.2 (0.4)	0.3 (0.6)
Random projection					
$N_{q_{inp}}$	31.3 (6.9)	38.1 (6.0)	40.0 (7.2)	49.4 (4.3)	51.3 (6.5)
$N_{q_{out}}$	21.8 (9.1)	24.4 (3.6)	25.9 (11.5)	20.1 (9.8)	30.6 (14.1)
$\eta$	6.5 (1.4)	7.4 (1.8)	9.7 (3.1)	2.9 (1.1)	8.4 (2.8)

1. Constructing feature subsets  $\Xi_l$ —candidates for selection—according to a chosen feature space exploration strategy.
2. Clustering vectors composed of attributes taken only from  $\Xi_l$  using a chosen clustering algorithm. Each investigated feature subset is assigned some partitioning  $\pi(\Xi_l)$ .
3. Evaluating correctness of  $\pi(\Xi_l)$  according to a criterion function  $J$ . The value  $J(\pi(\Xi_l))$  or simply  $J(\Xi_l)$  suggests in turn relevance of a subset  $\Xi_l$ .
4. After the feature space exploration terminates, selecting the best feature subset  $\hat{\Xi}$  which satisfies

$$\hat{\Xi} = \arg \max_{\Xi_l \subset \Omega} J(\Xi_l). \quad (6)$$

In (6),  $\Omega$  denotes a subset of all original features, whereas  $\Xi_l \in M_\Xi = M_\Omega \setminus \{\Omega, \emptyset\}$ , with  $M_\Omega$  being the collection of all subsets of  $\Omega$ , and  $l = 1, \dots, |M_\Xi|$ . The reason for choosing the above selection scheme is that features are evaluated in subsets (and not individually, as in the *filter approach*, cf. e.g. [16]). It is often observed that feature significance reveals only when it is accompanied by other discriminative attributes. Moreover, the wrapper approach is flexible—the same search strategy can be applied to various clustering algorithms and evaluation criteria.

Apart from the strategy of constructing subsets  $\Xi_l$ , performance of a selection procedure depends mainly on a feature evaluation function. However, choosing appropriate evaluation measure constitutes the major problem in unsupervised feature selection. Unlabeled data sets lack unambiguous clues which would facilitate direct identification of attributes most valuable for classification. Common solutions to this problem employ various—though conceptually similar—measures of clustering

*quality*, assuming that correctly classified data form high-quality clusters (i.e. well separated and internally compact) [18–20]. As it has already been shown by the authors in [21], this concept happens to fail since true data vector classes often exhibit opposite properties. Remaining in the context of texture classification, this situation takes place if analyzed images convey insufficient information about visualized structures—either because of low image resolution or the presence of noise. This leads to conclusion that some other criterion is needed to detect significant features when objectively different classes do not form high-quality clusters. Such a criterion can be defined using the notion of clustering stability.

#### 4.2.1 Clustering stability

Clustering stability assumes that good clusters occupy regions in the feature space which are consistent for more than one data samples constructed in the same classification task. Formally, it can be defined as [22]

$$\hat{\beta}(P, N, \mathcal{A}) = E_P(d_{\hat{\beta}}(\pi_1(S_1), \pi_2(S_2))), \quad (7)$$

where  $E_P$  denotes the expected value taken with respect to the (unknown) probability distribution  $P$ ,  $\pi_1$  and  $\pi_2$  are partitionings obtained with an algorithm  $\mathcal{A}$  for two data sets  $S_1$  and  $S_2$ , both generated from  $P$ ,  $N = |S_1| = |S_2|$ , and  $d_{\hat{\beta}}$  is some measure of distance between clustering results. It must be noticed that  $\hat{\beta}$ , being proportional to dissimilarity between different groupings, actually measures the *instability* of clustering. Thus, in order to infer the ratio of stability  $\beta$ , an inverse or negative value of  $d_{\hat{\beta}}$  should be taken.

Although conceptually simple, the notion of clustering stability constitutes a challenge when attempting to

formulate some measure of  $d_{\beta}^L$ . This is due to the fact that data vectors representing the same class in sets  $S_1$  and  $S_2$  may be assigned different labels by  $\mathcal{A}$ . Hence, there is no direct way for comparing independent cluster structures induced by  $\pi_1(S_1)$  and  $\pi_2(S_2)$ .

Ben-Hur et al. [23] approach the outlined problem by making the two sets  $S_1$  and  $S_2$  share some portion of feature vectors. These non-disjoint sets are then clustered separately and only their intersection  $S_c = S_1 \cap S_2$  is taken into account in calculation of the *Jaccard coefficient*

$$d_{\beta}^L(\pi_1(S_1), \pi_2(S_2)) = 1 - \frac{N^{11}}{N^{01} + N^{10} + N^{11}}, \tag{8}$$

where  $N^{11}, N^{01}, N^{10}$  denote numbers of data vectors  $x_i, x_j \in S_c, i \neq j$ , assigned to:

- the same cluster both in  $\pi_1(S_1)$  and  $\pi_2(S_2)$ ,
- the same cluster in  $\pi_1(S_1)$  but to different clusters in  $\pi_2(S_2)$ ,
- different clusters in  $\pi_1(S_1)$  but in the same cluster in  $\pi_2(S_2)$ .

The alternative measure is presented in [24]. First, the set  $S_1$  is clustered giving the partitioning  $\pi_1$ . Next,  $S_1$  together with its cluster labels are considered as a training sample and used to construct a classifier  $\mathcal{C}$  (in the supervised manner). Eventually,  $\mathcal{A}$  and  $\mathcal{C}$  are invoked on the set  $S_2$  to produce partitioning  $\pi_2$  and the vector of class labels  $\lambda_2$ . Estimation of (7) is performed through calculating the Hamming distance

$$d_{\beta}^H(\pi_1(S_1), \pi_2(S_2)) = \frac{2}{N(N-1)}[L^{10} + L^{01}], \tag{9}$$

where  $L^{10}$  denotes the number of data vector pairs  $x_i, x_j \in S_2$  which belong to the same class in accordance with  $\lambda_2$  but are differently clustered in  $\pi_2$ .  $L^{01}$  symbolizes the opposite condition.

Yet, another approach to stability estimation can be devised in terms of clusterings comparison measures, such as *Mallows distance*-based metric [25]. However, stability in principle refers to similarity between clusterings obtained for two at least partially different feature vector collections. On the contrary, Mallows distance presumes partitionings of the same data set. Hence, in order to adopt Mallows metric for stability assessment, we propose the following mechanism.

As in the case of Jaccard coefficient, let the intersection of samples  $S_1$  and  $S_2$  be a non-empty set  $S_c$ . For partitionings  $\pi(S_1), \pi(S_2)$  find the optimal clusters correspondence, such that the sum of distances between matched clusters is minimal. Then the contribution of a single data vector  $x_i \in S_c$  to the overall difference between cluster structures is given by [26]

$$d_i = \sum_{k=i}^K \sum_{j=1}^K p_{i,k} q_{i,j} L(k, j) (1 - I(\rho(j) = k)), \tag{10}$$

where  $p_{i,k}$  and  $q_{i,j}$  specify whether  $x_i$  is a member (when equal to one, zero otherwise) of a  $k$ th cluster in  $\pi(S_1)$  and a  $j$ th cluster in  $\pi(S_2)$ , respectively.  $L(k, j)$  denotes distance between the corresponding cluster centroids,  $I$  is the indicator function equaling one if the argument is true (zero otherwise) and  $\rho(j)$  gives an index of a cluster in partitioning  $\pi(S_1)$  that is matched to a  $j$ th cluster in  $\pi(S_2)$ . Then the total instability ratio can be defined as

$$d_{\beta}^M(\pi_1(S_1), \pi_2(S_2)) = \sum_{x_i \in S_c} d_i. \tag{11}$$

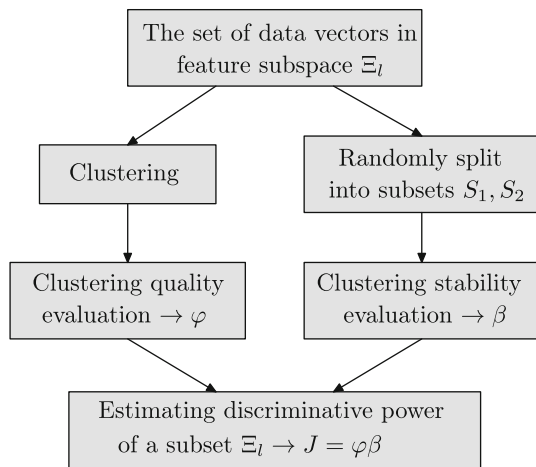
Naturally, the notion of stability can be encapsulated by a number of different methods. Among those, which are not included in this study, consensus-driven clustering (or meta-clustering in general) presents an attractive option. Similarity between clusterings can be evaluated using a measure of distance between the so-called *proximity matrices* (for a comprehensive description refer to [27]). Moreover, it is interesting to see that semi-supervised learning leads to more stable clustering results as they are forced to fit some a priori inferred pattern that is known to be correct. However, providing a detailed survey of all possible stability metrics would exceed the scope of this research. The primary goal was to demonstrate general need for reflecting clustering stability in the unsupervised feature selection problem. Thus, only the three above-described measures are included in the conducted experiments. In order to identify the optimal solution for (7) another in-depth study should be undertaken.

#### 4.2.2 Clustering stability-based feature selection

Following concepts introduced in the previous works of the authors [21], it is assumed that these are insignificant features which constitute one of the potential sources of clustering instability. Hence, it can be postulated to use clustering stability as a measure of feature relevance. However, the proposed approach avoids neglecting the importance of properties reflected in clustering quality measures. Therefore, in the experiments reported in Sect. 5, the combined quality/stability-based criterion was used to estimate features significance. It is defined as

$$Y(\Xi) = \varphi(\Xi)\beta(\Xi), \tag{12}$$

where  $\beta(\Xi)$  is a stability measure assessed by either (7), (8) or (11), while  $\varphi(\Xi)$  denotes a clustering quality function. Since this study focuses on the stability notion, quality measure was chosen arbitrarily—it was decided to employ the *average silhouette value* [28, 29]



**Fig. 7** Evaluating features relevance based on the clustering stability notion

$$\varphi^{SV} = \frac{1}{N} \sum_{i=1}^N \varphi_i^{SV}, \quad (13)$$

$$\varphi_i^{SV} = \frac{\min_{\vartheta_j} D(x_i, \vartheta_{j \neq k}) - A(x_i \in \vartheta_k)}{\max[\min_{\vartheta_j} D(x_i, \vartheta_{j \neq k}), A(x_i \in \vartheta_k)]}, \quad (14)$$

where

$$A(x_i \in \vartheta_k) = \frac{1}{N(k)} \sum_{x \in \vartheta_k} d(x_i, x)^2 \quad (15)$$

denotes average distance of a data vector  $x_i$  belonging to  $k$ th cluster from other data vectors included in  $\vartheta_k$ , while

$$D(x_i \in \vartheta_k, \vartheta_{j \neq k}) = \frac{1}{N(j)} \sum_{x \in \vartheta_j} d(x_i, x)^2 \quad (16)$$

is the average distance between  $x_i$  and data vectors belonging to  $j$ th cluster which is not the cluster of  $x_i$ . The  $d(\cdot, \cdot)$  operator stands for any desired distance metric (e.g. Euclidean measure),  $N$  is the total number of feature vectors, while  $N(j)$  denotes cardinality of a  $j$ th cluster. High scores on (13) indicate good quality of associated groupings.

The list below presents the subsequent steps involved in determining the value of  $\Upsilon$  (see also Fig. 7):

1. Find clustering of data vectors in a feature subspace  $\Xi_l$ .
2. Evaluate quality of clusters obtained in step 1 using (13)  $\rightarrow \varphi(\Xi_l)$ .
3. Randomly split data vectors into two groups  $S_1, S_2$ , either disjoint or non-disjoint, depending on the stability measure (given in (8), (9) or (11), respectively).
4. Estimate stability associated with  $\Xi_l$  basing on the two samples formed in step 3  $\rightarrow \beta(\Xi_l)$ .
5. Calculate  $\Upsilon$  according to (12).

## 5 Experiments

### 5.1 The experimental framework

In order to confirm the relevance of texture description applied to quantitative inspection of vascular tree images, and simultaneously evaluate the proposed approach to unsupervised feature selection, each data set mentioned in Table 1 was submitted to the following procedure:

1. Unsupervised feature selection using the average silhouette value only  $\rightarrow J = \varphi^{SV}$ .
2. Clustering data vectors in feature space determined in step 1.
3. Calculation of the error rate  $e_\varphi$ .
4. Unsupervised feature selection using the stability-based feature evaluation criterion defined in (8)  $\rightarrow J = \varphi^{SV} (1 - d_\beta^L)$ .
5. Clustering data vectors in feature space determined in step 4.
6. Calculation of the error rate  $e_\gamma^L$ .
7. Unsupervised feature selection using the stability-based feature evaluation criterion defined in (9)  $\rightarrow J = \varphi^{SV} (1 - d_\beta^H)$ .
8. Clustering data vectors in feature space determined in step 7.
9. Calculation of the error rate  $e_\gamma^H$ .
10. Unsupervised feature selection using the stability-based feature evaluation criterion defined in (11)  $\rightarrow J = \varphi^{SV} \left( \frac{1}{1 + d_\beta^M} \right)$ .
11. Clustering data vectors in feature space determined in step 10.
12. Calculation of the error rate  $e_\gamma^M$ .

Again, as in the experiments with feature transformation, error rates were assessed using classes to clusters evaluation technique. Feature space exploration was performed using *hybrid genetic algorithm* [30], *k*-means clustering algorithm, and 1-NN classifier [for clustering stability estimation in (9)]. It must be noted that in the general case of unsupervised classification task, along with data set partitioning, the number of clusters  $K$  needs to be determined. However, in order to simplify the problem, it was assumed that the correct value of  $K$  is known a priori. Such a simplification does not undermine the conclusions derived from the experiments. This research focuses on comparison of two approaches to estimation of features significance (based on clustering quality and stability concepts) whose efficiency in relation to each other is tested for exactly the same conditions and shall remain unchanged whether value of  $K$  is available or not.



**Table 3** Clustering error after feature selection (%)

Evaluation criterion	Rice distribution parameter $\sigma$				
	0	1	3	5	10
<b>Constant input flow</b>					
$\varphi$	0.0	26.3	34.4	46.3	53.1
$\varphi(1 - d_{\beta}^J)$	2.5	3.1	9.4	39.4	40.0
$\varphi(1 - d_{\beta}^H)$	0.0	3.1	6.9	31.3	45.0
$\varphi\left(\frac{1}{1+d_{\beta}^M}\right)$	0.0	1.9	8.8	34.3	47.5
<b>Constant output flow</b>					
$\varphi$	0.0	0.6	10.0	18.5	26.3
$\varphi(1 - d_{\beta}^J)$	0.0	0.0	6.9	5.6	26.3
$\varphi(1 - d_{\beta}^H)$	0.0	0.0	0.6	4.4	26.3
$\varphi\left(\frac{1}{1+d_{\beta}^M}\right)$	0.0	1.8	5.6	6.9	28.1
<b>Viscosity</b>					
$\varphi$	0.0	0.0	0.8	0.0	0.0
$\varphi(1 - d_{\beta}^J)$	0.0	0.0	0.8	0.0	0.0
$\varphi(1 - d_{\beta}^H)$	0.0	0.0	0.8	0.0	0.0
$\varphi\left(\frac{1}{1+d_{\beta}^M}\right)$	0.0	0.0	0.8	0.0	0.0

5.2 Discussion of the results

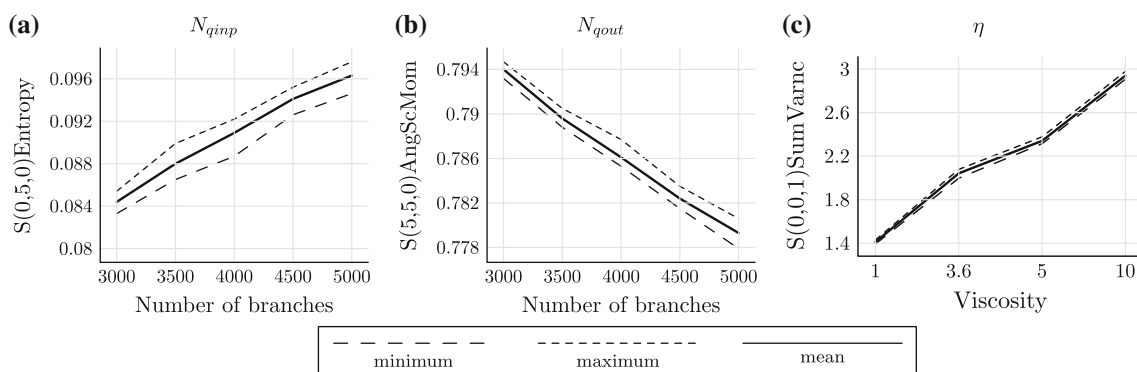
The analysis of the experimental results (see Table 3) reveals the following observations. First of all, texture description of vessel tree images indeed encapsulates intrinsic vascularity patterns. Small errors of unsupervised classification achieved for the less noisy or noise-free images (independently from a criterion employed for feature selection procedure) objectively confirms the correspondence between texture features and tree growth parameters. As it is depicted in Fig. 8, linear relationship can be observed between selected most significant features and each of the tested vascularity classes.

Second, it has been demonstrated that using solely clustering quality measure as a feature selection criterion may occur insufficient to find good attribute subspace ensuring proper class separation. In the reported experiments, this took place when images were corrupted by noise. Objectively different clusters are then neither compact nor isolated. Taking into account the notion of clustering stability helps resolving the problem. In that case, the classification error remains small until the impact of noise becomes dominant.

Eventually, it is worth noticing that that effectiveness of the stability-based selection depends on the measure used to estimate consistency between two clustering results. According to Table 3, the Hamming distance defined in (9) allows obtaining better results than the Jaccard coefficient and the Mallows distance-based metric used in this research. Recalling arguments raised by Lange et al. [24], the common part of data samples  $S_1$  and  $S_2$  affects structures of both partitionings obtained for them. This in turn makes clustering results apparently stable even for extremely noisy data. More reliable stability estimation can be achieved for two disjoint samples.

6 Conclusion

In this paper, we have confirmed applicability of the texture model in quantitative description of computer-simulated vascular tree images. A series of experiments with synthesized vascularity images was performed to show correspondence between selected texture features and tree growth parameters such as number of branches or blood viscosity. In these experiments, different vascularity classes were successfully clustered by k-means algorithm. This demonstrates that texture has great potential application in detecting structural abnormalities in the vessel system. It is worth noticing that preliminary studies on real data sets containing rat brain vascularity images from a confocal



**Fig. 8** Relation between texture and tree growth parameters for **a** number of branches at constant input flow, **b** number of branches at constant output flow, and **c** blood viscosity. Depicted results refer to images corrupted by Rician noise with parameter  $\nu = 1$

microscope have been described in [5]. The most important aspect of this study is the validation of that approach using relatively large sample of images and also objectivism of the analysis achieved by employing unsupervised learning methods—both for classification and feature selection.

We have also introduced a novel measure of features relevance suited to unlabeled data. The motivation for its development emerges from the fact that state-of-the-art feature selection algorithms promote features which ensure high-quality clusters, i.e. compact and mutually dispersed. Although noiseless vascularity images do possess such properties, the noisy ones cause these quality-based methods to fail. The proposed criterion is based on the notion of clustering stability. It thus facilitates unsupervised feature selection when true clusters are not necessarily well separated but retain their structure independently from random data samples.

**Acknowledgments** This work was supported by the Polish Ministry of Science and Higher Education grant no. 1205/DFG/ 2007/02.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Klepaczko A, Kociński M, Materka (2009) Cluster analysis in application to quantitative inspection of 3D vascular tree images. In: Kurzynski M, Wozniak M (eds) *Computer recognition systems 3, Advances in Intelligent and Soft Computing*, Springer, Heidelberg, pp 87–94
- Deistung A, Ditttrich E, Sedlacik J, Rauscher A, Reichenbach JR (2009) ToF-SWI: simultaneous time of flight and fully flow compensated susceptibility weighted imaging. *J Magn Reson Im* 29(6):1478–1484. doi:10.1002/jmri.21673
- Keener J, Sneyd J (2008) *Mathematical physiology*, 2nd edn. Springer, Berlin
- Tofts P (2003) *Quantitative MRI of the brain: measuring changes caused by disease*. Wiley, Chichester
- Kocinski M, Materka A, Lundervold A, Chekenya M (2008) Classification of vascular tree images on numerical descriptors in 3D. In: Tkacz E, Komorowski D, Kostka P, Budzianowski Z (eds) *Proceedings of 9th international conference SYMBIOSIS 2008*, Kamien Slaski
- Karch R, Neumann F, Neumann M, Schreiner W (1999) A three-dimensional model for arterial tree representation, generated by constrained constructive optimization. *Comput Biol Med* 29: 19–38
- Kretowski M, Rolland Y, Bezy-Wendling J, Coatrieux JL (2003) Fast algorithm for 3-D vascular tree modeling. *Comput Meth Prog Bio* 70:129–136
- Kociński M, Materka A, Lundervold A (2007) On the effect of vascular tree parameters on 3D texture of its image. In: *ISMRM-ESMRMB conference*, Berlin
- Materka A (2006) What is the texture? In: Hajek M, Dezortova M, Materka A, Lerski R (eds) *Texture analysis for magnetic resonance imaging*, Med4Publishing, Prague, pp 11–44
- Szczypinski P, Strzelecki M, Materka A, Klepaczko A (2009) Mazda—a software package for image texture analysis. *Comput Meth Prog Biol* 94:66–76
- KeyRes (2009) KeyRes technologies—information and communication technologies. [http://www.keyres-technologies.com/medical\\_imaging/index.html](http://www.keyres-technologies.com/medical_imaging/index.html). Accessed Sep 2009
- LS2W (2009) The LS2W Software Suite. <http://www.maths.bris.ac.uk/~wavethresh/LS2W/>. Accessed Sep 2009
- Duda R, Hart P, Stork D (2001) *Pattern classification*. Wiley, Chichester
- Witten IH, Frank E (2005) *Data mining. Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco
- Swets D, Weng J (1996) Using discriminant eigenfeatures for image retrieval. *IEEE Trans Pattern Anal Mach Intell* 18(8): 831–836
- Blum A, Langley P (1997) Selection of relevant features and examples in machine learning. *Artif Intell* 97(1–2):245–271
- Kohavi R, John G (1997) Wrappers for feature subset selection. *Artif Intell* 97(1–2):273–324
- Kim Y, Street W, Menczer F (2002) Evolutionary model selection in unsupervised learning. *Intell Data Anal* 6:531–556
- Morita M, Sabourin R, Bortolozzi F, Suen C (2003) Unsupervised feature selection using multi-objective genetic algorithms for handwritten word recognition. In: *7th IEEE conference on document analysis and recognition*, pp 666–670
- Dy J, Brodley C (2004) Feature selection for unsupervised learning. *J Mach Learn Res* 5:845–889
- Klepaczko A, Materka A (2009) Clustering stability-based feature selection for unsupervised texture classification. *Mach Graph Vis* 18(2):125–141
- Luxburg Uv, Ben-David S (2005) Towards a statistical theory for clustering. Technical report, Max Planck Institute for Biological Cybernetics, London, presented at the PASCAL workshop on clustering
- Ben-Hur A, Elisseeff A, Guyon I (2002) A stability based method for discovering structure in clustered data. In: *Pacific symposium on biocomputing*
- Lange T, Braun M, Roth V, Buhmann J (2004) Stability-based validation of clustering solutions. *Neural Comput* 16:1399–1323
- Fowlkes E, Mallows C (1983) A method for comparing two hierarchical clusterings. *J Am Stat Assoc* 78(383):553–569
- Zhou D, Li J, Zha H (2005) A new Mallows distance based metric for comparing clusterings. In: *Proceedings of 22nd international Conference on machine learning*, pp 1028–1035
- Pedrycz W, Hirota K (2008) A consensus-driven fuzzy clustering. *Pattern Recogn Lett* 29:1333–1343
- Struyf A, Hubert M, Rousseeuw P (1997) Integrating robust clustering techniques in S-Plus. *Comput Stat Data Anal* 26:17–37
- Handl J, Knowles J (2006) Feature subset selection in unsupervised learning via multiobjective optimization. *Int J Comput Intell Res* 2(3):217–238
- Oh I, Lee J, Moon B (2004) Hybrid genetic algorithms for feature selection. *IEEE Trans Pattern Anal Mach Intell* 26(11): 1424–1437