EXTRACTION AND ANALYSIS OF LINEAR FEATURES IN MULTIDIMENSIONAL IMAGES BY HOMOLOGICAL METHODS

M. MROZEK, M. ŻELAWSKI, A. GRYGLEWSKI, S. HAN, AND A. KRAJNIAK

Abstract. We show that the problem of extracting linear features from a noisy image and counting the number of branching points may be successfully solved by homological methods applied directly to the image without the need of skeletonization and the analysis of the resulting graph. The method is based on the superimposition of a mask set over the original image and works even when the homology of the feature is trivial and in arbitrary dimension. We tested the method on computer-generated data, 2D images of blood vessels, 2D satellite images and 3D images of collagen fibers.

1. Introduction

Topology, a 100 years old theory oriented on the fundamental characteristics of shape, recently starts playing an important role in computer science. This happens mainly via homological methods, in part because such methods provide substantial reduction of the amount of analysed data [7]. Although the classical metric tools remain in the mainstream of the image analysis, the importance of topology in the field becomes significant too [10, 33, 35]. For instance, the simplest topological technique commonly used in image segmentation is the extraction of the connected components of a black and white image treated as a subset of $\mathbb{R}^d$. Such extraction may be performed in linear time and in the case of images stored as bitmaps the algorithm is extremely fast [29]. The process of extracting the connected components of the image may be viewed as constructing the zeroth homology group of the image. In this interpretation every connected component corresponds to a generator

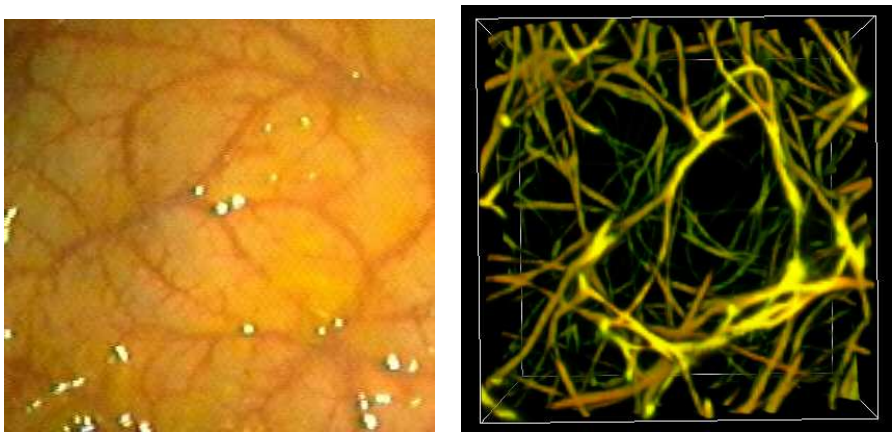
2000 *Mathematics Subject Classification.* Primary **68U10**; Secondary **55N99**.

Key words and phrases. linear feature, feature extraction, homological method, blood vessels image, collagen fibrils image, medical imaging, satellite image, branching point.

Supported by Polish MNSzW, Grant N201 037 31/3151 and N N201 419639.

of the zeroth homology group. Higher homology groups, which measure the presence of tunnels, cavities and higher dimensional holes in the image, also may be used in computer vision and image processing [1, 15, 25, 37, 36].

In this paper we develop the concept of feature extraction of multidimensional images by homological tools, introduced in [24]. Our aim is to show that higher homology groups may be successfully applied to image analysis even if there are no holes in the image itself and/or in the features to be extracted. This is because the superimposition of some pattern (mask) over the original image may form holes and the appearance (or lack of appearance) of holes may be used to analyse certain features of the image via the study of homology generators. Unfortunately, until recently there were no homology algorithms fast enough for such applications and very few homology algorithms constructing homology generators. This is in part because the classical homology algorithms are based on Smith diagonalization of integer matrices, which is super cubical. However, the recently developed reduction homology algorithms [23, 22] (with implementations available from [21], see also [5, 4]) offer speed comparable to the speed of the algorithms constructing connected components. Moreover, these algorithms offer not only homology groups but also homology generators. For another recent algorithm for homology groups and generators oriented on image analysis see [26].



a) b)

FIGURE 1. a) A sample colonoscope 2D picture of blood vessels in mucosa. b) A projection of a 3D image of collagen fibrils.

Although we are convinced that the method of the superimposition of masks over the image to facilitate the use of homological methods will find several applications in the future, in this paper we concentrate on a sample problem, which consists in the extraction of linear structures from the image under the presence of other features and noise. By a linear structure we understand a feature of a binary image consisting of thin threads which may branch. In images of varying origin such features are very common. For example, linear structures are formed by rivers and roads in satellite images of earth ([13, 32]) or blood vessels in medical imaging where the problem of extracting linear features is very important from the point of view of medical diagnostics ([16, 3, 9]).

Although the examples we have studied so far are only two and three dimensional, the techniques presented in this paper apply to images in any dimension. The image dimension is denoted throughout the paper by d. The sets of integers, positive integers and reals are denoted respectively by $\mathbb{Z}, \mathbb{Z}^+$ and $\mathbb{R}$.

The organization of the paper is as follows. We begin with Section 2, where we present our original motivation, the reconstruction of blood vessels and collagen fibrils and the model problem of the extraction of linear features from a noisy image. In Section 3 we present our model example, used in the following sections to explain our algorithm. In the next sections we introduce cubical sets, which form the bridge between the discrete concept of a raster image and the topology of subsets of R^d. Section 5 is devoted to a brief, intuitive and very informal presentation of the homology of cubical sets. The following section is the main section of the paper and consists of three subsections. The first subsection describes our extraction algorithm based on the homology generators of the image superimposed with masks. The following subsection contains sample results of the algorithm for computer-generated data, 2D endoscopic images of mucosa microcirculation and 3D images of collagen fibers. The last subsection contains a comparison of our algorithm with the common methods based on skeletonization and pruning. An algorithm counting branching points, based on another set of masks and homology is presented in Section 7. In the last section we present conclusions and some final remarks.

2. The motivation and the formulation of the problem

Our sample problem arises from two independent concrete issues: the analysis of 2D colonoscope images of blood vessels in colon mucosa (see Figure 1a) and the analysis of 3D reconstituted confocal microscopy images of type I collagen fibrils (see Figure 1b). The patterns

of blood vessels in colon mucosa have been tested in animals via histologic specimens and proved to differentiate pathology. However, to see the patterns animals had to be killed and specimens properly extracted [6, 17, 19, 34]. In theory, the patterns can be discerned by analysing colon mucosa images and measuring some characteristics, for instance the number of branching points. In practice, this is difficult because of the presence of noise and interference of other tissues. We found that the techniques presented in this paper provide a good method to get rid of noise, extract the vessels and count some valuable characteristics, in particular the number of branching points.

Counting of branching points is also important in the case of collagen fibrils. Type I collagen is the major matrix protein of skin, bone, tendon and other tissue. The normal form of type I collagen is a heterotrimer composed of two α1(I) and one α2(I) chains. In fetal tissues, tumors, and several heritable disorders, $\alpha 1(I)_3$ homotrimers have been found as well [14, 18, 27]. Structurally, the homotrimer and heterotrimer fibrils are similar, but they lead to different mechanical properties, e.g., resistance of tendons to tensile stress [20]. Recently, confocal microscopy images revealed that heterotrimers form entangled networks of flexible, thread-like fibers while homotrimers form spikes emanating from a common center [11]. The homotrimers clearly play a special role which still remains unclear. To understand how homotrimers cause pathologies, it is important to investigate morphological characteristics, e.g., the number of branching points of type I collagen fibrils. As in the case of blood vessels, this is complicated by the presence of noise. We found that the homological methods presented in this paper are very useful in understanding the structure of entangled 3D networks of collagen fibrils.

Common to the two problems is the separation of the noisy linear features (blood vessels and collagen fibrils) from the background and the knowledge of the number and location of branching points together with the length of branches. Such an analysis performed directly by human eye is difficult in 3D images even with the use of the best available visualization techniques. But also in 2 dimensions the analysis is not straightforward because of the presence of other features in the image both of natural (for example other tissues) and artifactual (for example unevenly distributed light) character. Therefore, an automated analysis is highly desirable. The typical technique applied to such problems begins with skeletonization, usually obtained by means of morphological openings. The process of skeletonization is usually complicated by the presence of noise and deficiencies of low level processing of the

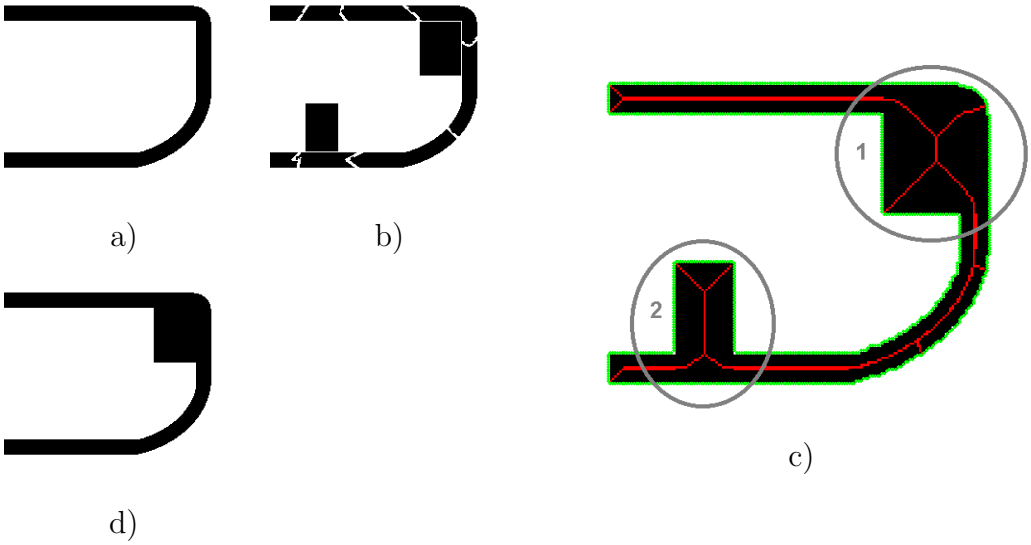


FIGURE 2. a) Sample image. b) Noisy version of the image. c) Contour (marked green) and skeleton (marked red) found by the method proposed in [2] in the image (b) after morphological closing. d) Result after trimming short branches.

image, in particular the process of binarization. Immediate skeletonization does not make sense because of the presence of noise dividing the linear structure into many pieces. Instead, some kind of gluing should be applied first, for example wrapping, dilation or morphological closing [28, Chpt. 7]. However, after this step some artifacts may be glued to linear structures which do not belong to them. Sometimes such artifacts may be later removed by trimming the short branches of the skeleton and removing the features which do not intersect the trimmed skeleton. Unfortunately, this strategy, for example based on the algorithm presented in [2], often leads to improper results even in simple cases, as we can see in Figure 2d. After applying closing to the noisy image in Figure 2b in order to glue together the pieces of the linear structure, both rectangular artifacts are also attached to the linear structure. In Figure 2c the contour and skeleton of the resulting image obtained by applying the method described in [2] are presented. The analysis of the length of branches in order to remove artifacts is not a satisfactory strategy in this case. After removing only the shortest branches which do not branch themselves, the trimmed skeleton still intersects both artifacts. On the other hand, if we trim all the branches

leaving the skeleton consisting of one line only, we get rid only of the artifact in region 2, because the trimmed skeleton still intersects the artifact in region 1. Consequently, the standard method applied to this problem may always leave some undesired features in the extracted image. In the case of the noisy image in Figure 2b the best result we can hope for is presented in Figure 2d. Obviously, in this oversimplified example it is easy to detect the rectangular artifacts directly. However, in the more realistic problems this is a serious obstacle. We will discuss the problem related to skeletonization methods in more detail in Section 6.3.

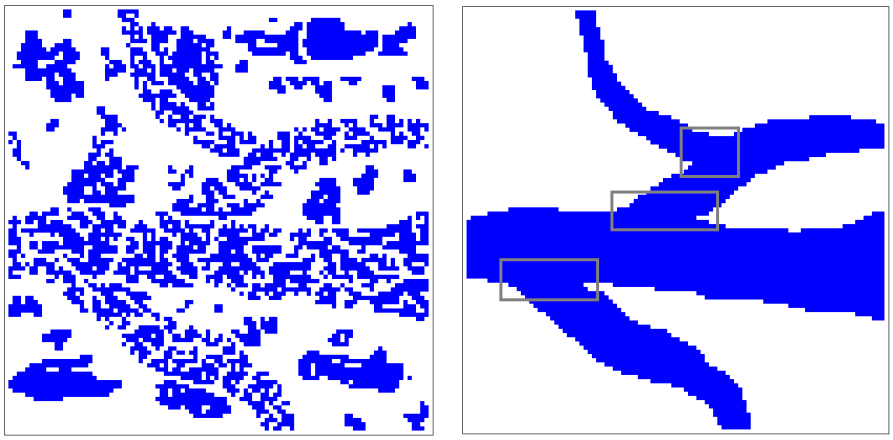
In many situations the understanding of the global structure of the image is needed to get rid of the artifacts. There are not many tools which can analyze the image globally. Among the few is the geometry of the graph resulting from the skeletonization (see [31] for an example of such an approach). The drawback is the need to construct such a graph. This is in contrast to our approach. We do not need neither a skeletonization nor a graph. Instead, we analyse directly the homology of the binarized image superimposed with masks. The approach succeeds, because the information captured by homology is global, despite the fact that the homology computation is a purely local process.

3. A model example

Consider the binary image in Figure 3a and the problem of extracting from this image the linear structure visible in Figure 3b consisting of four threads glued at three branching points.

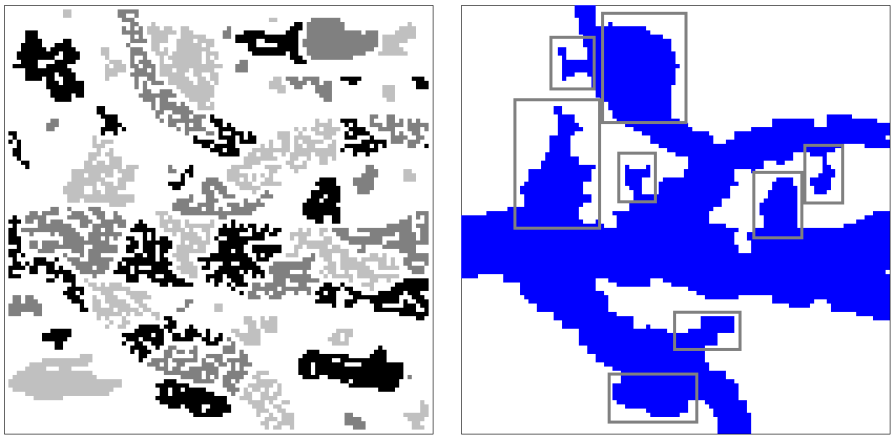
The segmentation by the size of the connected components of the image fails. This is because the noise together with the applied binarization techniques disconnect the branches into several components. Figure 4a shows the connected components of the original image in various shades of gray depending on the cardinality of the component. Clearly, there is no way to select a good threshold level. On the other hand, the image in Figure 4b resulting from the standard gluing technique based on morphological closing is not satisfactory either, because it attaches some nearby components, not belonging to the branches. The same happens when the path closing technique [12] is applied.

The human brain has no problem in extracting the desired feature. This is because the brain performs the analysis of the mutual location of the connected components searching for pieces which form long threads. Such analysis is global in nature, so it cannot be achieved by local tools, such as dilation and erosion. The standard procedure is to build the skeleton of the desired feature (see Figure 5), code it as a



a) b)

FIGURE 3. a) Sample binary image. b) The linear structure which needs to be extracted with the three branching points marked with rectangles.

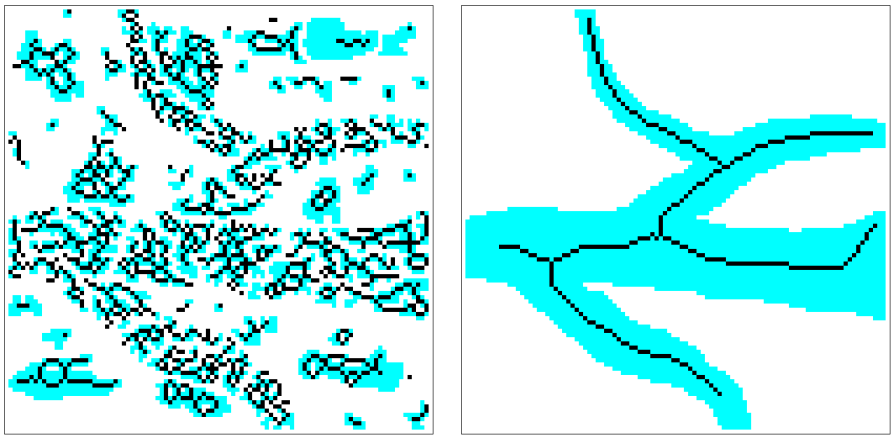


a) b)

FIGURE 4. a) Connected components of the original image in various shades of gray. b) The result of segmentation by component size processed by morphological closing. The undesired features attached to the image are marked with rectangles.

graph and perform the global analysis of the graph. In the approach proposed in this paper the extra levels of processing (construction of graph, graph analysis) are avoided by applying homological methods,

which are computationally local but provide global information. Moreover, unlike the standard case, they are straightforward to apply in higher dimensional images.



a) b)

FIGURE 5. a) Skeleton of the image in Figure 4a. b) Skeleton of the image in Figure 4b.

4. CUBICAL SETS.

We view a raster image as a finite array of pixels with a color associated to each pixel. Each pixel is indexed by its *coordinate vector* $q \in \mathbb{Z}^d$. Since in the case of a *binary raster image* only two colors are allowed, the image may be identified with a subset of all pixels sharing the same color, for instance black.

In order to tie some topology to a raster binary image, we identify every pixel of coordinates $q \in \mathbb{Z}^d$ with the *pixel cube*

$$Q_q = [q_1, q_1 + 1] \times [q_2, q_2 + 1] \times \cdots [q_d, q_d + 1]. \tag{1}$$

We denote the set of all pixel cubes in $\mathbb{R}^d$ by $\mathcal{K}$. This convention allows us to think about a raster binary image as a finite subfamily $\mathcal{X} \subset \mathcal{K}$. Given a raster pixel image $\mathcal{X}$, we consider the set

$$|\mathcal{X}| := \bigcup \{ Q \mid Q \in \mathcal{X} \} \subset \mathbb{R}^d,$$

i.e. union of all cubes corresponding to black pixels in the image. The union is a subset of $\mathbb{R}^d$, called a *cubical set*. A cubical set, as a subset of $\mathbb{R}^d$ has some induced topology. This is the topology we are interested in, often referred to as image topology. Conversly, given a cubical set X in $\mathbb{R}^d$, the family of all pixel cubes contained in a X, denoted by

$\mathcal{K}(X)$, may be thought of as a raster image associated to the cubical set X.

Two simple examples of cubical sets in R^d needed in the sequel are the cubical ball and cubical sphere centered at a point $a \in \mathbb{Z}^d$ and of radius $s \in \mathbb{Z}^+$ defined as follows

$$\begin{aligned} B^d_{a,s} &:= \bigcup \{\, Q_q \mid |q_i - a_i| \leq s \text{ for } i = 1, 2, \ldots d \,\}, \\ S^{d-1}_{a,s} &:= \bigcup \{\, Q_q \mid |q_i - a_i| = s \text{ for } i = 1, 2, \ldots d \,\}. \end{aligned}$$

Cubical sets, unlike the subsets of $\mathbb{Z}^d$, may have nontrivial topology. For instance, Figure 6 represents a 3-dimensional cubical set with a hole going from the top to the bottom.

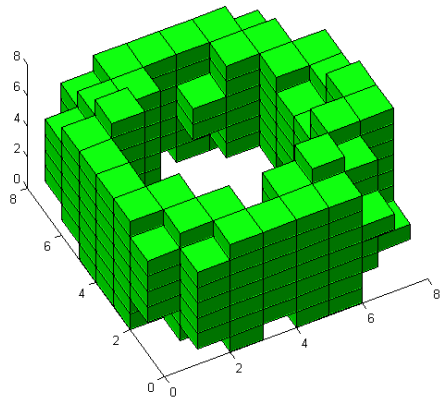


FIGURE 6. An example of a cubical set in $\mathbb{R}^3$.

Given a cubical sets X and a point $x \in X$, by $\mathrm{cc}(X, x)$ we denote the connected component of X which contains x. We extend this notation to a set $A \subset X$ by

$$\mathrm{cc}(X, A) := \bigcup_{x \in A} \mathrm{cc}(X, x).$$

The standard operation of dilation and erosion in the context of a cubical set X are defined as

$$\begin{aligned} \mathrm{dil}(X) : &= \bigcup \{\, Q \in \mathcal{K} \mid Q \cap X \neq \emptyset \,\} \\ \mathrm{ero}(X) : &= \bigcup \{\, Q \in \mathcal{K}(X) \mid P \in \mathcal{K} \text{ and } P \cap Q \neq \emptyset \text{ implies } P \subset X \,\} \end{aligned}$$

and the 1-step closing operation is a dilation followed by erosion

$$\mathrm{clo}(X) := \mathrm{ero}(\mathrm{dil}(X)).$$

5. Overview of cubical homology.

Homology groups constitute a relatively simple means of studying the topology of cubical sets by algebraic methods. We briefly present an intuitive approach to homology groups of cubical sets. For a formal treatment of this subject we refer the reader to [15].

The concept of homology requires not only cubes of the form (1), later on referred to as *full cubes*, but also faces of cube Q of the form

$$P = [q_1, r_1] \times [q_2, r_2] \times \cdots [q_d, r_d], \tag{2}$$

where $r_i = q_i$ or $r_i = q_i + 1$. We call such cubes *elementary cubes*. The *dimension* of P is the number

$$\dim P := \operatorname{card} \{\, i \mid q_i \neq r_i \,\}.$$

The faces of dimension zero are called *vertices* and the faces of dimension one are called *edges*. Given a cubical set X, by an n-chain in X we mean a linear combination of n-dimensional elementary cubes contained in X of the form $\sum_{i=1}^{k} \alpha_i P_i$. The coefficients α_i may be taken from an arbitrary ring R, although the typical choices are the field of integers modulo two $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$ (the simplest choice) and the ring of integers $\mathbb{Z}$ (the most general choice). In particular every elementary cube P in X may be viewed as an n-chain $1 \cdot P$, where n is the dimension of P. The *support* a chain c, denoted $|c|$, is the minimal cubical set containing all the elementary cubes appearing in the chain with a non-zero coefficient.

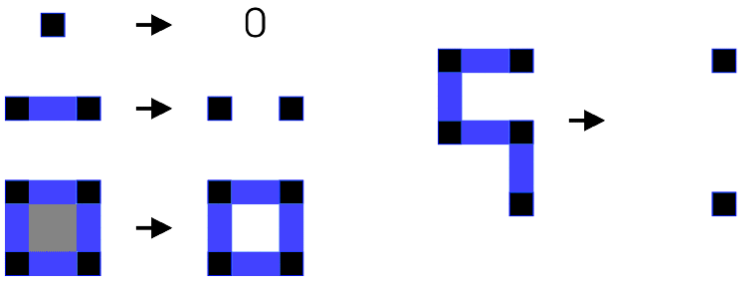


Figure 7. a) Boundaries of a vertex, an edge and a two dimensional elementary cube. b) A 1-chain and its boundary. Coefficients are not visualised. Vertices and edges are fattened for better visualization.

All n-chains in X form a group called the group of n-dimensional chains of X, denoted $C_n(X)$. There is a well defined homomorphism

of groups $\partial_n^X : C_n(X) \to C_{n-1}(X)$ which sends every n-chain $1 \cdot P$ to a certain linear combination of its $(n-1)$-dimensional faces representing its boundary (see Figure 7a).

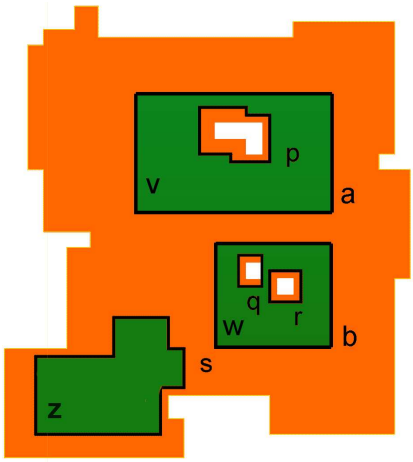


FIGURE 8. A cubical set X (orange and green, individual cubes are not marked), 1-cycles p, q, r, s (black) and 2-chains v, w, z (green).

The boundary operator is defined in such a way, that in the case of a 1-chain in Figure 7b the vertices shared by two edges cancel out.

Chains whose boundary is zero, i.e. chains for which all faces in the boundary cancel out, are called *cycles*. They form a subgroup of $C_n(X)$, denoted $Z_n(X)$. Cycles are important, because they may be used to detect holes. Figure 8 shows a cubical set X with some 1-chains and 2-chains. The supports of 1-chains a and p surround the same whole in X. This is reflected algebraically by the fact that the difference $p - a$ is the boundary of the 2-chain v. The support of the chain s does not surround a hole. Again, this is seen algebraically, because s is the boundary of the 2-chain z. The n-chains which are boundaries of $(n+1)$-chains form a subgroup of $Z_n(X)$ denoted $B_n(X)$. The nth homology group of X is the quotient group

$$H_n(X) := Z_n(X)/B_n(X).$$

The elements of this group are equivalence classes, called *homology classes*. Roughly speaking, each hole in the space X corresponds to a generator of a homology group of X. In the case of cubical sets contained in $\mathbb{R}^d$ for d not exceeding three or in the case of coefficients in $\mathbb{Z}_2$ the structure of this group is very simple: it is a free group generated by a finite number of homology classes. In this case the number

of generators in $H_n(X)$ is called the nth Betti number of X and denoted $\beta_n(X)$. In the general case the definition of Betti number is slightly more complicated, because the group may contain cyclic elements. These elements detect torsions in the set and must be factored out before one can define the Betti number. However in this paper we will not consider such a situation.

In the case of the cubical set X in Figure 8 the first homology group is generated by the homology classes of the 1-chains p, q and r, so $\beta_1(X) = 3$.

6. Extracting linear features

6.1. **The algorithm.** After the brief introduction to homology theory we proceed with the analysis of the image in Figure 3a. As we mentioned in Section 3, we need a global tool to extract the linear feature. The tool we choose is based on the detection of one dimensional holes in a modified image. First, the morphological closing X' of the original image X is constructed in order to glue all the nearby connected components. Next, the generators of the relative homology $H(X', X)$ are computed and their supports are added to X. Now a mask consisting of a collection of parallel hyperplanes (lines in the case of a two-dimensional image) is added to detect holes formed by the branching points and the one-dimensional homology generators of the resulting set are computed (see Figure 9a). The mask itself has no holes, but the superimposition of the mask over the original image forms holes. These holes are surrounded by generators having non-empty intersection with the hyperplanes in the mask. Thus, the parts of the generators disjoint from the mask must extend along the threads and branching points in the image. Nonetheless, they are unlikely to intersect the other features of the image which may have been glued to the set in the closing process. In the next step the support of the selected generators is added to the original image (see Figure 9b). Finally, we select the connected components of the resulting image which have non-empty intersection with the selected generators (see Figure 9c). At this stage the linear features are extracted. The standard techniques consisting in removing the remaining small components and applying smoothing based on the closing process may be still applied to obtain the final image (see Figure 9d). This procedure is summarized in Algorithm 1 presented in Table 1.

Notice that this algorithm works in every dimension.

Some artifacts may be perceived as very short branches (see Figure 4b). The size of branches detected by the algorithm is controlled

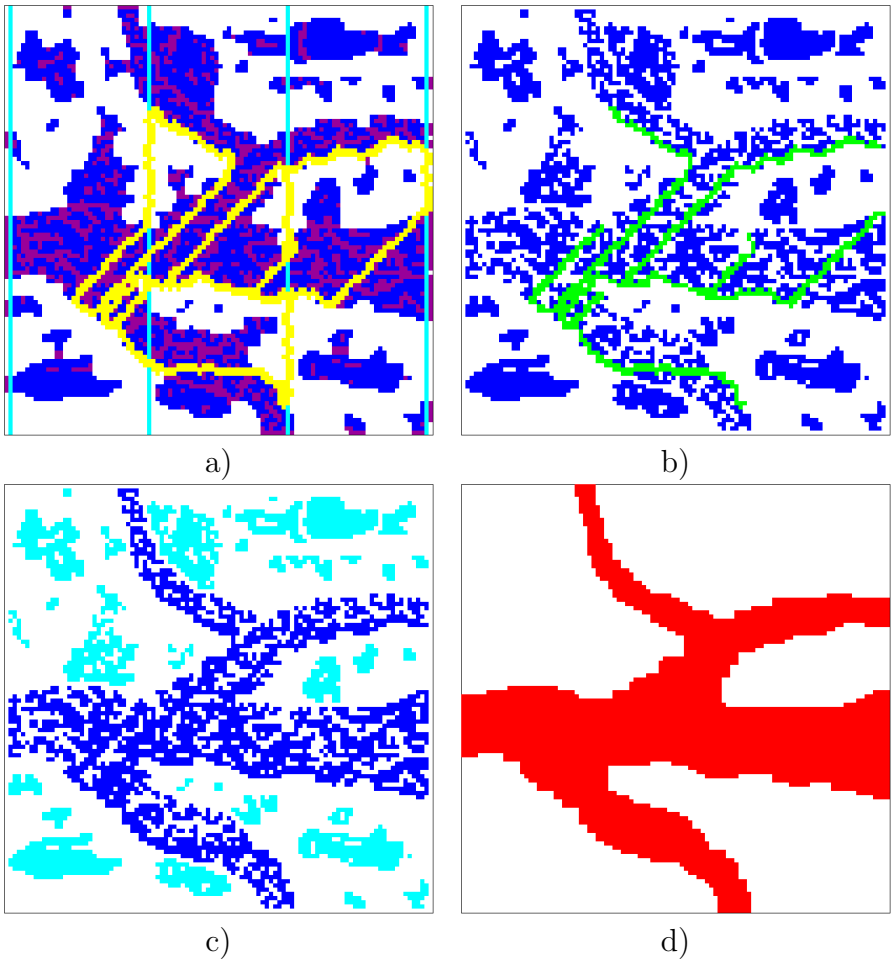


FIGURE 9. a) The image in Figure 3a after applying one cycle of closing process (fill-in marked in violet), adding masks (marked in cyan) and computing 1-homology generators (marked yellow). b) The union of the original image and the trace of generators on the closed image (marked green). c) Components of the union which have nonempty intersection with the trace (marked blue) and the remaining components (marked cyan). d) The extracted feature after discarding small components and smoothing via the closing process.

by the parameter m. The larger m, the larger are the branches removed by the algorithm. Sometimes a short branch coming from an artifact may be left by the algorithm even if m is large. This situation may occur when a hyperplane of the mask passes close to the artifact.

Algorithm 1.

1. Input:
   - raster binary image represented as a cubical set X,
   - integer m
2. $X' := \mathrm{clo}(X)$
3. $\mathcal{A} :=$ homology generators of $H_1(X', X)$
4. $G := \bigcup_{a \in \mathcal{A}} |a|$
5. $Y := X \cup G$
6. $M :=$ union of parallel hyperplanes
   separated by distance m
7. $\mathcal{C} :=$ homology generators of $H_1(Y \cup M)$
8. $D := \emptyset$
9. for each $c \in \mathcal{C}$ if $|c| \cap M \setminus X \neq \emptyset$ then $D := D \cup (|c| \setminus M)$
10. $\mathcal{W} := \mathrm{cc}(D \cup X, D)$
11. Output: $\bigcup \mathcal{W}$

TABLE 1. The algorithm extracting the linear structure

However, the chances of this situation are small and may be further minimized by running the algorithm a few times with shifted location of the hyperplanes and taking the intersection of the outputs. The proper value of the parameter m for a concrete input image can be estimated and chosen automatically on the basis of the sizes of connected components in the image.

The quality of the output may be dependent on the direction of the hyperplanes. In the case of the lack of any natural choice of the direction, one can run the algorithm for several choices of the direction and take the union of the outputs.

6.2. **Sample results.** In order to validate Algorithm 1, several computer-generated linear structures in 2D images were produced (see Figure 10a,b). These linear structures were broken randomly and random artifacts were added (see Figure 10c,d where computer-generated artifacts are marked by red color). The processed noisy images are shown in Figure 10e,f. Red artifacts removed by the algorithm are marked by blue color.

Other 2D examples are presented in Figure 11. Red artifacts, which were not removed, are marked by yellow color (see Figure 11f where a

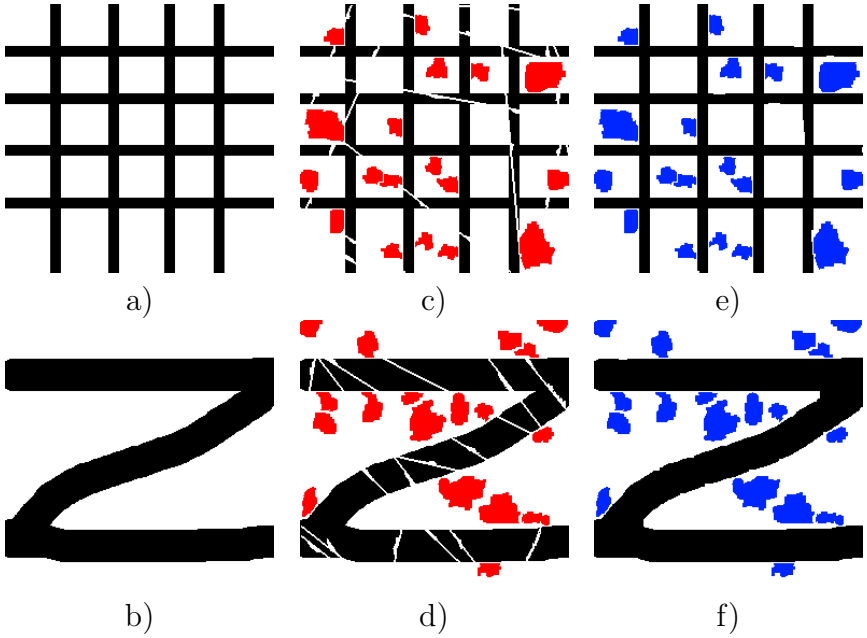


FIGURE 10. a,b) Sample computer-generated 2D images. c,d) The images with computer-generated random noise. e,f) The noisy images (c,d) processed by Algorithm 1 and smoothed (colors are explained in the text).

longer artifact was classified as a broken linear branch because it was too similar to linear structures and was localized too close to them).

The algorithm was also tested on 3D images generated in the same way (see Figure 12). The meaning of colors is the same as in the 2D examples.

Other 3D examples consisting of bending fibers are presented in Figure 13 and Figure 14. A yellow artifact in Figure 14d were classified improperly because of the same reasons as previously mentioned (similarity to linear structures and their closeness).

Figure 15 shows the results of the presented algorithm for real endoscopic images of colon mucosa. Unequal endoscopic illumination and some colon mucosa features cause that in original input images many vessels are broken or partially visible. Thereofore, color images (Figure 15a,b) were binarized (Figure 15c,d) by a special strategy which is based on the observation that vessels are narrow and vessel pixels are always darker than background pixels in the nearest neighborhood of the vessel (details will be presented in a future paper). Next, the

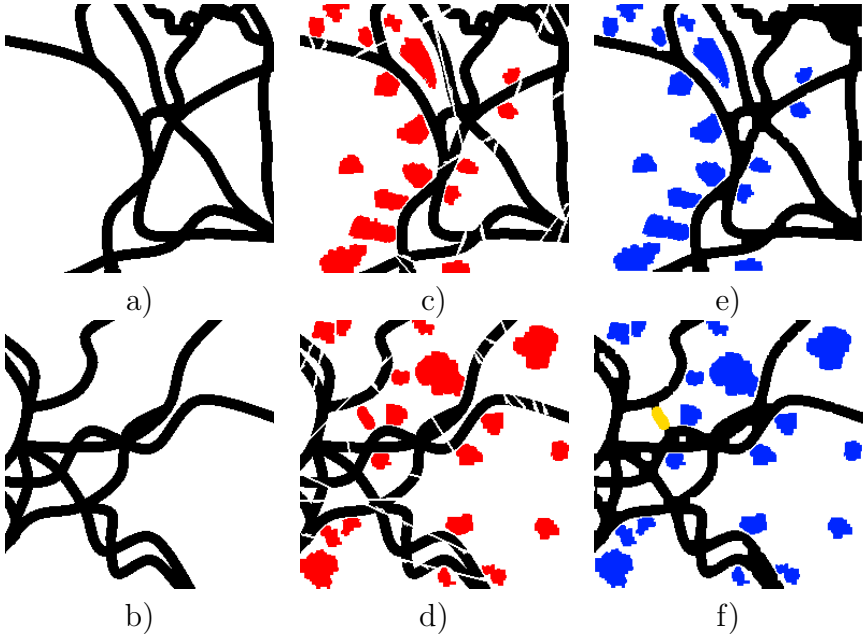


FIGURE 11. a,b) Next sample computer-generated 2D images. c,d) The images with computer-generated random noise. e,f) The noisy images (c,d) processed by Algorithm 1 and smoothed (colors are explained in the text).

binarized images were processed by Algorithm 1 in order to extract and reconstruct the microcirculation architecture (see Figure 15e,f).

The presented algorithm can also be applied to the extraction of other linear structures like rivers and roads in satellite and aerial images of earth. Figure 17 shows the river extracted from a sample noisy binarization of a satellite image. In this case the noisy image has many artifacts which are glued to the linear structure and cannot be removed in any step. The improved version of Algorithm 1 applies morphological opening before the first step in order to disconnect such artifacts. Therefore, in the next steps disconnected artifacts are classified properly and removed. Obviously opening may also disconnect some parts of the linear structure, but they will be glued again by the algorithm during the reconstruction process.

6.3. **Comparison with other methods.** The algorithm of extracting linear structures presented in this paper was compared to methods based on skeletonization and pruning (see [2, 30]). As we mentioned earlier, immediate skeletonization does not make sense because of the

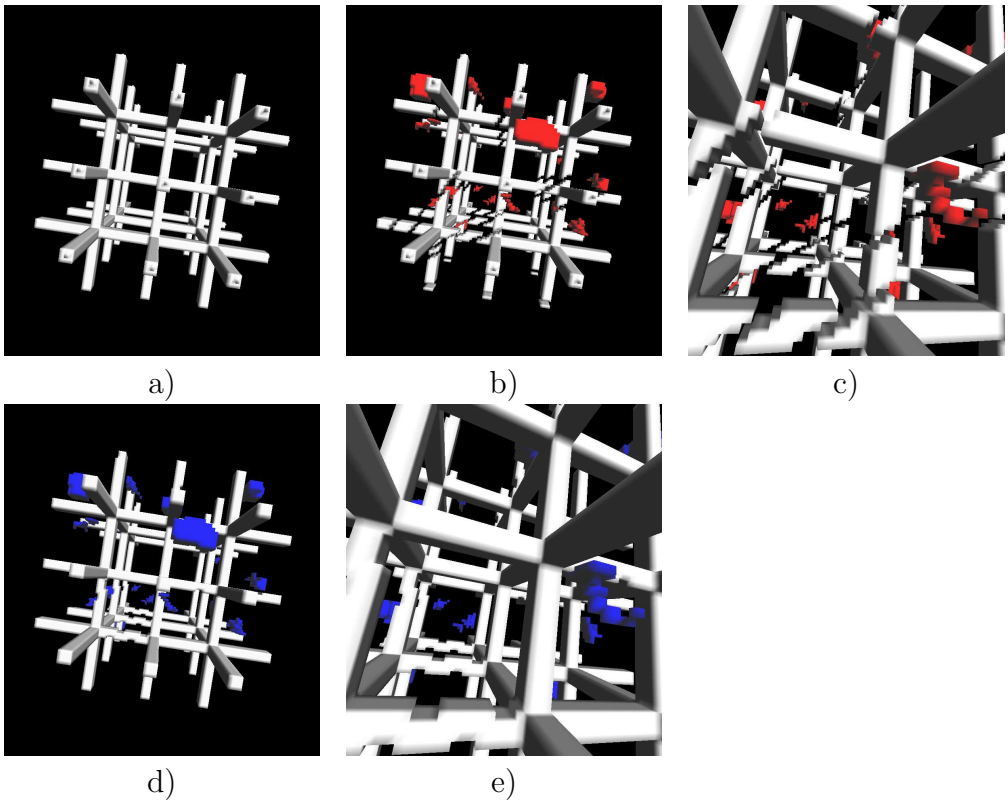


FIGURE 12. a) Sample computer-generated 3D image. b,c) The image with computer-generated random noise and its enlargement. d,e) The noisy image (b) processed by Algorithm 1 and smoothed and its enlargement (colors are explained in the text).

presence of noisy cuts separating the linear structure into many pieces. Before skeletonizing morphological closing or another gluing process such as wrapping or dilation should be applied to the input image in order to connect (reconstruct) partially visible broken vessels. After skeletonizing short branches of the obtained skeleton may be for example trimmed by iterative deleting simple points from the set of end points (a simple point is a point whose deletion does not alter the topology of the skeleton and an end point is a point with only one neighbour in the skeleton). Finally, connected components of the input image which do not intersect the trimmed skeleton may be classified as artifacts and removed. Unfortunately, the analysis of the length of skeleton branches often leads to improper extraction (see Figure 2). Another result of such strategy based on the method described in [2] (skeleton

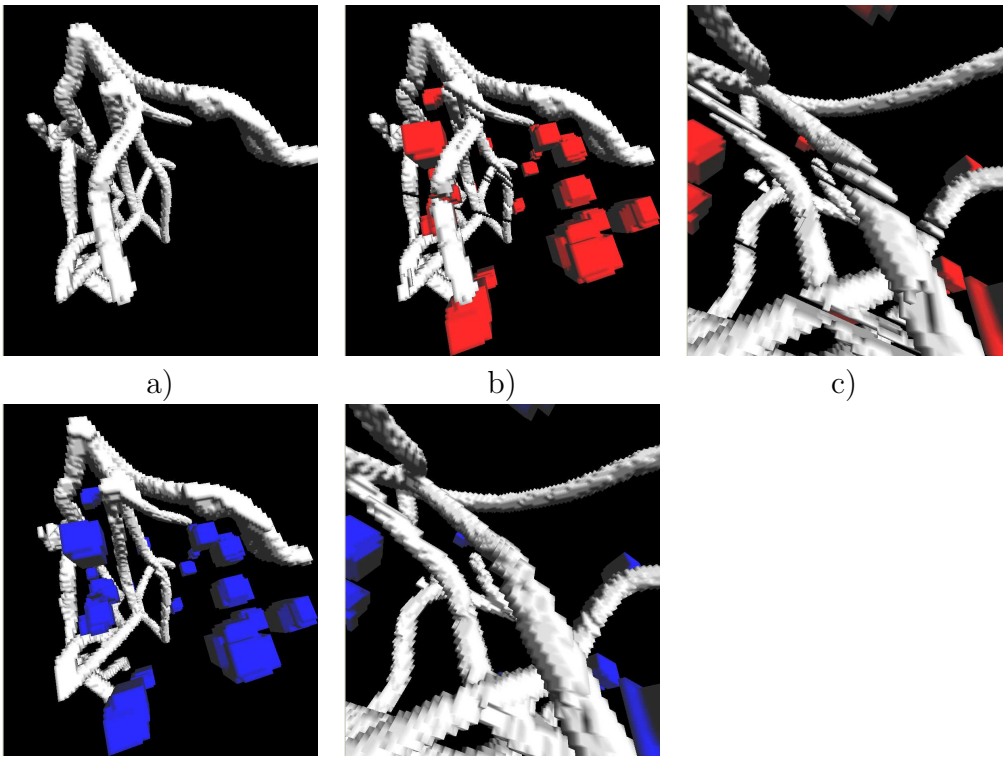


FIGURE 13. a) Next sample computer-generated 3D image. b,c) The image with computer-generated random noise and its enlargement. d,e) The noisy image (b) processed by Algorithm 1 and smoothed and its enlargement (colors are explained in the text).

pruning by the contour partitioning with the Discrete Curve Evolution technique) is presented in the more realistic problem in Figure 16a. Observe that some glued artifacts do not generate any branches and change significantly the skeleton. For example, the main line of the skeleton is modified because of the presence of the artifact in region 1. In consequence, this artifact intersects the main skeleton line (see Figure 16b) and is classified as a part of the vessels. Moreover, some glued artifacts generate longer or similar skeleton branches than branches connected with proper linear structures (compare the branch in region 2 with the branches in regions 3, 4, 5 and the branch in region 6 with the branches in region 4). Shorter branches may be connected with the components of the main vessel and should not be trimmed. Deleting simple points from the whole skeleton trims skeleton parts connected

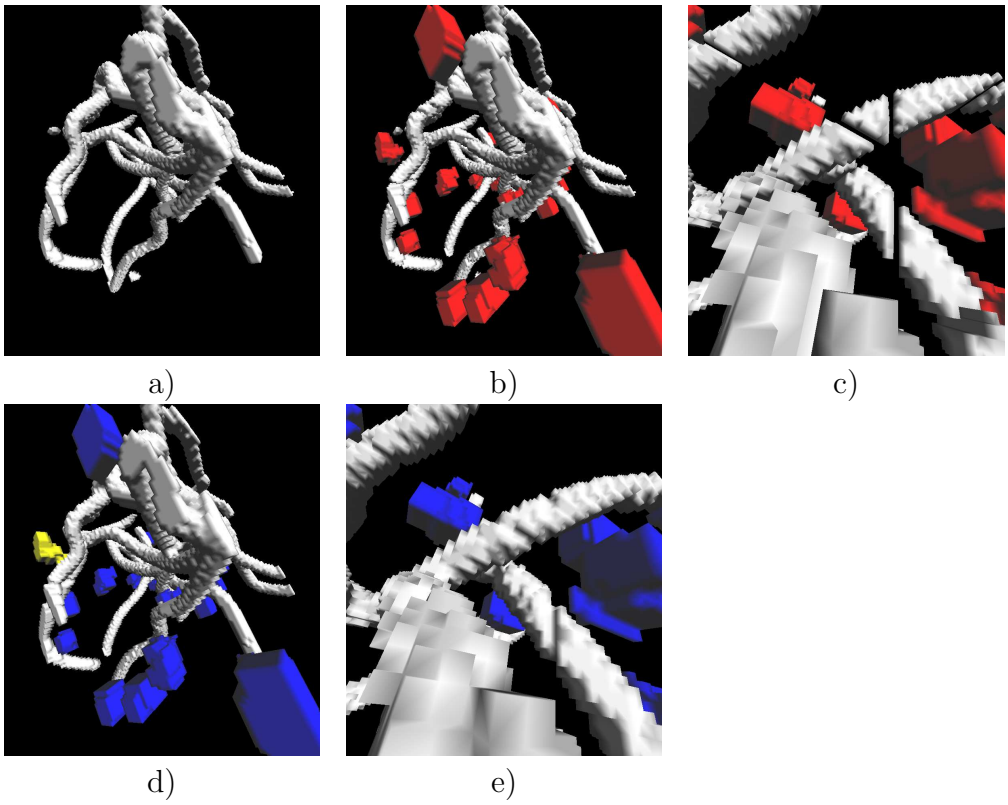


FIGURE 14. a) Next sample computer-generated 3D image. b,c) The image with computer-generated random noise and its enlargement. d,e) The noisy image (b) processed by Algorithm 1 and smoothed and its enlargement (colors are explained in the text).

with components of vessels (see regions 3, 4, 5) and these components will be later removed as artifacts. It is not obvious how to improve this strategy, because it is difficult or impossible to decide, how long we should repeat the procedure of trimming simple points and which branches should be trimmed. If we trim only the shortest branches, we do not get rid of the artifact in region 2. Selecting proper branches to trim is not easy, because we are not always sure where the main line of the skeleton leads and consequently vessel components may be removed instead of artifacts. For example, if we decide to trim a shorter branch which is a natural choice, we finally remove the vessel component in region 3 (marked black in Figure 16b) instead of the artifact in region 2 because the branch in region 3 is shorter than the branch in region 2. In region 4 every choice of a branch to be trimmed is bad because both

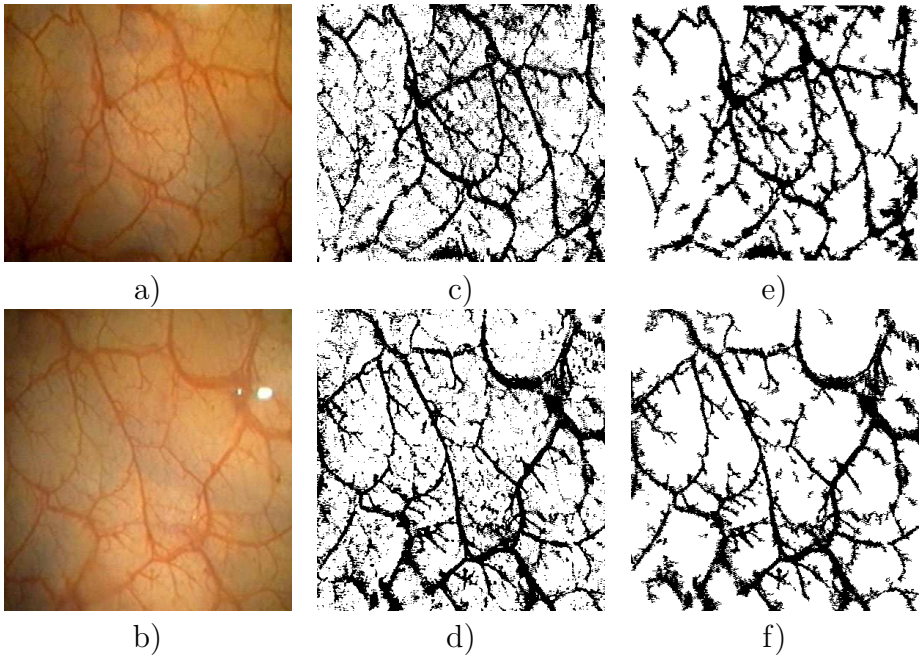


FIGURE 15. a,b) Sample endoscopic images of colon mucosa. c,d) The binarized endoscopic images. e,f) The binarized images (c,d) processed by Algorithm 1 and smoothed.

branches are connected with vessel components. This approach cannot properly identify artifacts without branches like the artifact in region 1 and all possible final results are unsatisfactory (see Figure 16c).

In Figure 18 one can see that the applied skeletonization method cannot detect all parts of linear structures even in the oversimplified regular shape without any noise where interior lines are not skeletonized (Figure 18a). Isolated components of linear structures without skeleton are classified as noise and the reconstruction is not effective. This also is a serious limitation in the applications we are interested in.

The strong advantage of Algorithm 1 is that it applies to images in any dimension (without any change) as opposed to skeletonization and pruning methods which work in 3 dimensions ([30]) or only in 2 dimensions and their extension to 3D space is not obvious ([2]).

7. Counting the branching points

Among the main features characterizing a linear structure is the number of branching points. A branching point is not a hole, so it is

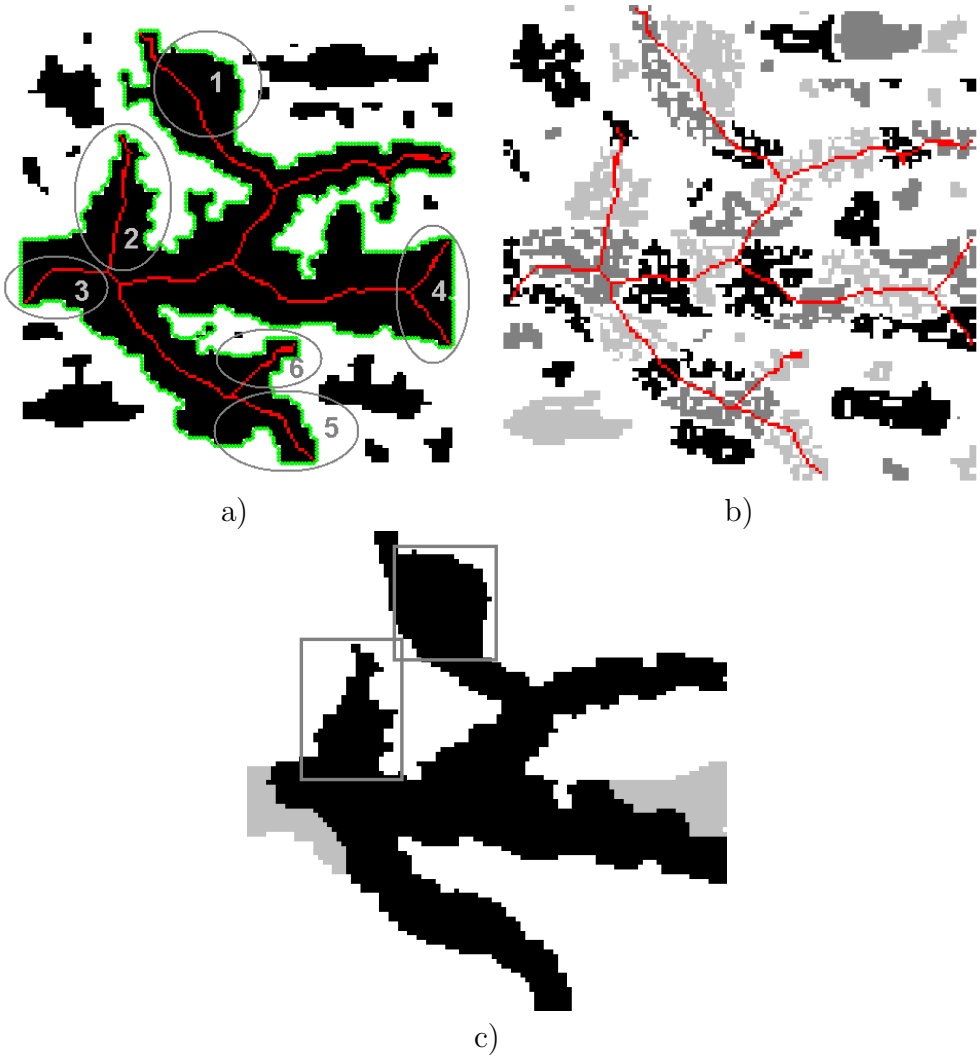
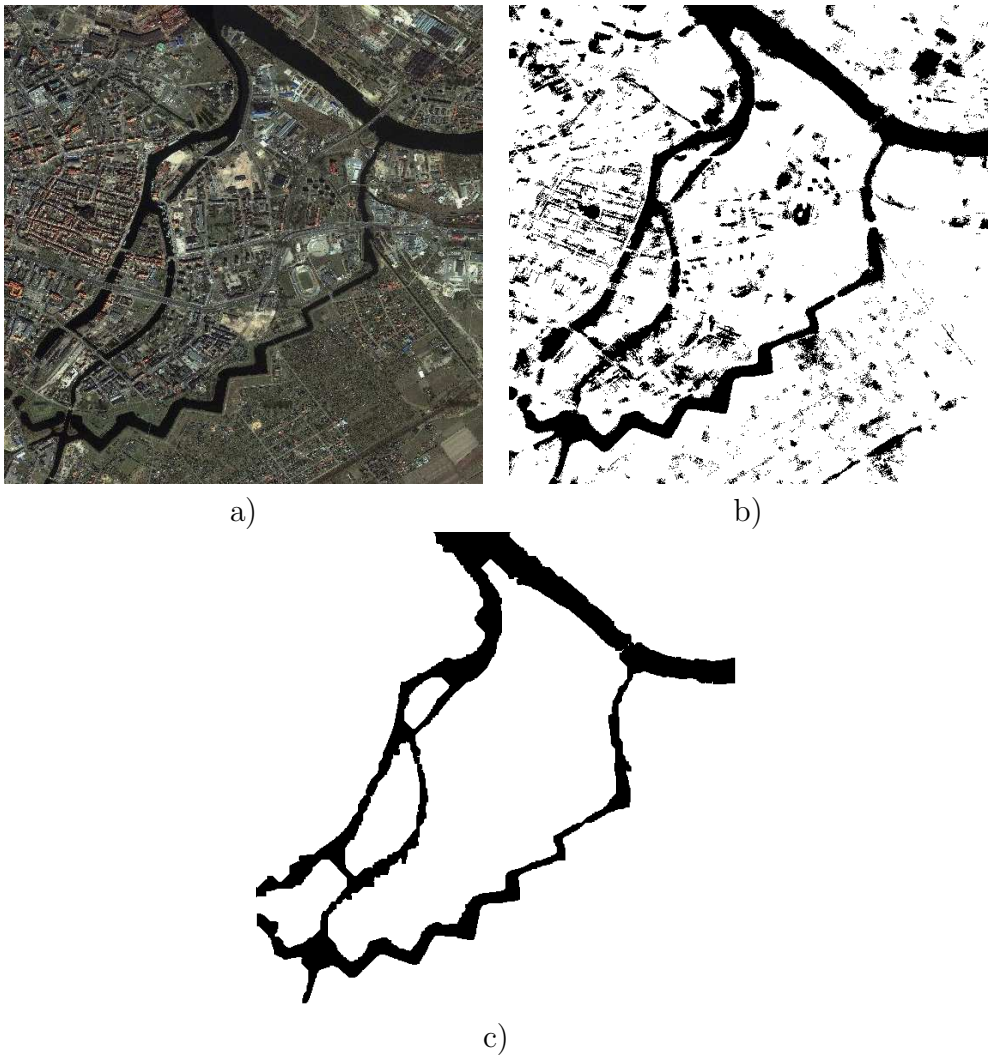


FIGURE 16. a) Contour (marked green) and skeleton (marked red) found by the method proposed in [2] in the image in Figure 3a after closing. b) The skeleton in a noisy input image with distinguished connected components. c) Sample result after trimming branches. The undesired features are marked with rectangles. Missing vessel components are marked by light gray.

not automatically reflected in the topology of a linear structure. However, intersecting the image around a branching point with a cubical ball $B^d_{a,s}$ and adding a cubical sphere $S^{d-1}_{a,s}$ to the intersection, we may differentiate the branching point from a single thread by counting the



a) b)

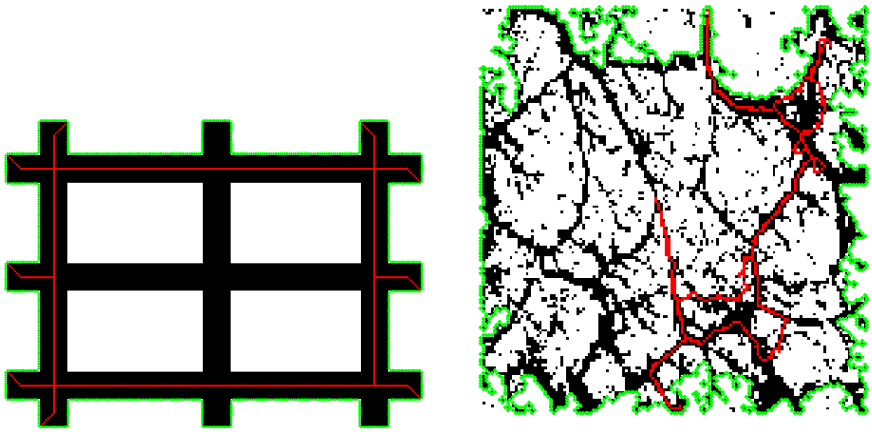
c)

FIGURE 17. a) Satellite image of a river (source: geoserwer.pl). b) Sample binarization. c) The binarized image processed by Algorithm 1 and smoothed.

first Betti number of the union. In the case when only a single thread crosses the sphere, we see two holes: the first one is $(d-1)$-dimensional, generated by the sphere itself and the other one is 1-dimensional and comes from the thread. However, if the thread branches, we see additional 1-dimensional holes (see Figure 19).

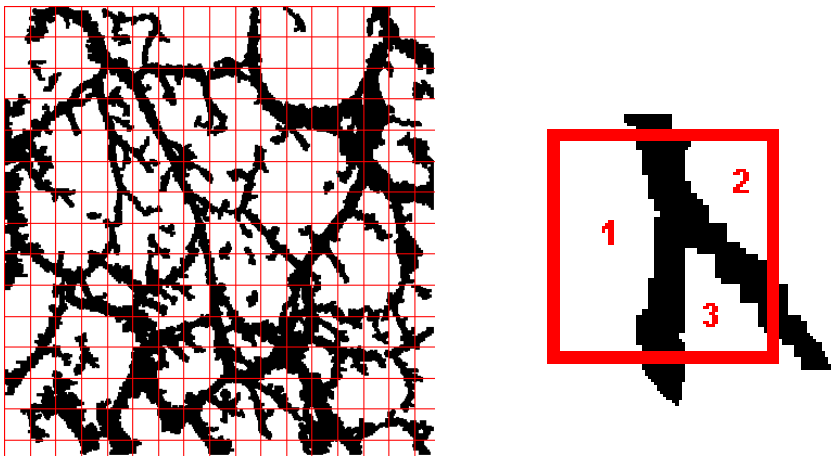
Therefore, the number of extra 1-dimensional holes given by

$$b_d(X,a,s) := \max\{0, \beta_1(X \cap B^d_{a,s} \cup S^{d-1}_{a,s}) - \beta_1(S^{d-1}_{a,s}) - 1\} \tag{3}$$



a) b)

FIGURE 18. a,b) Contour (marked green) and skeleton (marked red) found by the method proposed in [2] in: a) a simple image with regular linear structures, b) the image in Figure 15d after closing.



a) b)

FIGURE 19. a) Sample vessels covered by estimating boxes (red). b) An estimating box with 3 holes numbered 1,2,3.

may be used as an estimate of the number of branching points inside the ball $B^d_{a,s}$. This leads to Algorithm 2 presented in Table 2.

Algorithm 2.

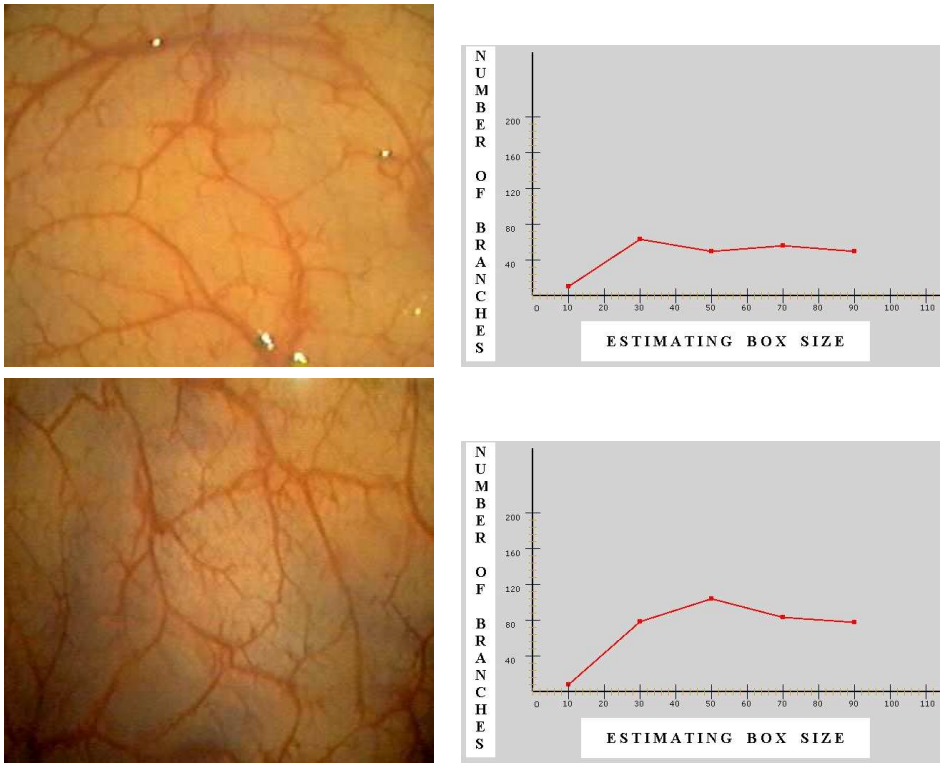
1. Input: image X, an integer s
2. $A := \{\, q \in X \mid s|q[i] \text{ for } i = 1, 2, \ldots d \,\}$
3. $t := 0$
4. for every $a \in A$ do $t := t + b_d(X, a, s)$
5. Output: t

TABLE 2. The algorithm estimating the number of branching points

The parameter s in this algorithm controls the size of the branches which should be captured. The extreme values of s cannot give reasonable results. If s is too small, of the order of magnitude of the thickness of threads, then obviously no branching points may be captured. The same happens when s is too large, of the order of magnitude of the size of the image. However, for the intermediate values of s, the algorithm counts only these branching points whose lengths of branches are at least s. When the length of the threads between the branching points does not vary significantly, then the outcome of Algorithm 2 is approximately constant for the intermediate values of s. This is the case for the images of blood vessels presented in the left column of Figure 20. This feature may be used to estimate the length of the threads between the branching points. The situation is different in the case of collagen fibers (see Figure 21). In this case the number of detected branching points decreases with s, which indicates that the length of branches varies.

Of course, Algorithm 2 gives only a rough measure of the number of branching points and its accuracy depends on several factors. First of all, we need to assume that the algorithm is applied to a linear feature. If not, then the 1-dimensional holes in the set will distort the results. If this is the case, then one can modify formula 3 by subtracting the number of holes contained entirely inside the cubical sphere. The number of such holes may be obtained as $\beta_1(X \cap B_{a,s})$. Moreover, the threads cannot be broken by noise. As we showed in Section 6 this may be achieved by Algorithm 1.

There may happen that two independent threads are captured in the same box. This will also result in an extra hole, although not coming



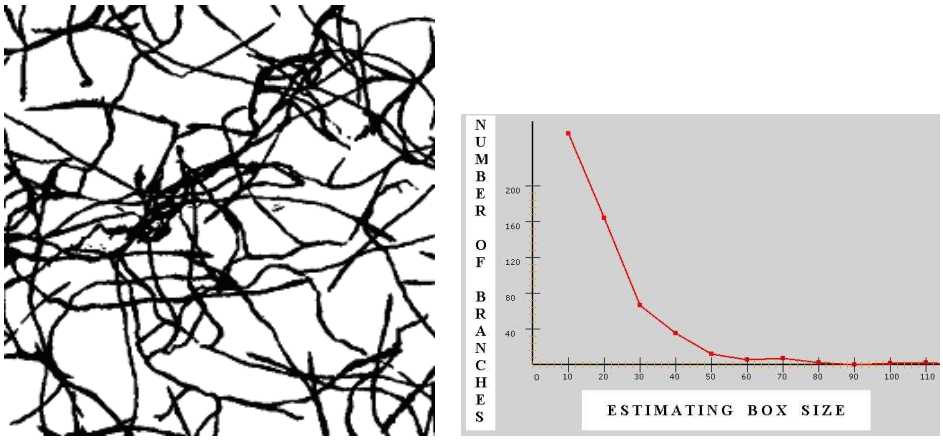
a) b)

FIGURE 20. a) Sample endoscopic images of blood vessels b) The respective number of branches as a function of different estimating box sizes.

from a branching point. One can discard such a case by looking at the generators: at least one generator will intersect the cubical sphere $S_{a,s}$ in at least to components. However, such a situation is relatively rare in the case of a linear feature.

Some branching points may fail to fall inside a box and will not be counted. To avoid such a situation one can do the computations a few times, each time shifting the grid by a fraction of the grid size and average the results.

Finally let us mention an important advantage of the presented method: it is straightforward to adapt Algorithm 2 so that it not only counts the number of branching points, but also gives their locations. This is particularly easy, because the algorithms operate on the original image, whereas the algorithms operating on graphs require special data structures which bind the graph with the original image.



a) b)

FIGURE 21. a) A projection of the 3D image of collagen fibrils b) The respective number of branching points in dependency on different estimating box sizes.

8. Conclusions and final remarks.

We presented the method of extracting linear features from noisy images in arbitrary dimension and counting the number of branching points. The algorithms work directly on the binarized image and do not require any intermediate data structures such as graphs. We overcame the problem of noise which makes the linear structure visible only via a global look by applying homological methods and masks. The algorithms have been implemented and tested on computer-generated data and concrete problems and proved to be useful in medical diagnostics. Details will be presented in separate papers, in specialistic journals.

The techniques presented in this paper apply directly only to binary images. Since we concentrate on homological methods, we do not discuss here the thresholding techniques used in the examples to obtain binary images. They will be presented in the future papers.

Let us mention that apart from thresholding, a more sophisticated method is possible to analyse gray scale and color images by topological methods, namely the gray scale or color image may be decomposed into a filtration of black and white images and investigated via studying the homology of the inclusion maps and/or persistent homology (cf. [8]). This is left for future investigation.

References

[1] M. Allili, K. Mischaikow, A. Tannenbaum, Cubical Homology and the Topological Classification of 2d and 3d Imagery, IEEE International Conference on Image Processing 2 (2001) 173-176.

[2] X. Bai, L.J. Latecki, W. Liu, Skeleton Pruning by Contour Partitioning with Discrete Curve Evolution, IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) 29 (3) (2007) 449–462.

[3] K. Bühler, P. Felkel, A. La Cruz, Geometric Methods for Vessel Visualization and Quantification - a Survey, in: G. Brunnett, B. Hamann, H. Müller (Eds.), Geometric Modelling for Scientific Visualization, Kluwer Academic Publishers, 2004, pp. 399–420.

[4] *Computational Homology Project*, http://chomp.rutgers.edu.

[5] *Computer Assisted Proofs in Dynamics*, http://capd.ii.uj.edu.pl.

[6] R.S. Da Costa, B.C. Wilson, N.E. Marcon, Optical Techniques for the Endoscopic Detection of Dysplastic Colonic Lesions, Current Opinion in Gastroenterology 21 (1) (2005) 70–79.

[7] T.K. Dey, H. Edelsbrunner, S. Guha, Computational Topology, in: B. Chazelle, J.E. Goodman, R. Pollack (Eds.), Advances in Discrete and Computational Geometry, Contemporary Mathematics 223, American Mathematical Society, Providence, 1999, pp. 109–143.

[8] H. Edelsbrunner, D. Letscher, A. Zomorodian, Topological Persistence and Simplification, Discrete Comput. Geom. 28 (2002) 511–533.

[9] P. Felkel, R. Wegenkittl, A. Kanitsar, Vessel Tracking in Peripheral CTA Datasets an Overview, Proceedings of the 17th Spring conference on Computer graphics (2001) 232–239.

[10] P. Frosini, C. Landi, Size Theory as a Topological Tool for Computer Vision, Pattern Recognition And Image Analysis 9 (4) (1999) 596–603.

[11] S. Han, D.J. McBride, W. Losert, S. Leikin, Segregation of Type I Collagen Homo- and Heterotrimers in Fibrils, Journal of Molecular Biology 383 (1) (2008) 122–132.

[12] H. Heijmans, M. Buckley, H. Talbot, Path Openings and Closings, Journal of Mathematical Imaging and Vision 22 (2-3) (2005) 107–119.

[13] S. Hinz, A. Baumgartner, Automatic Extraction of Urban Road Networks from Multi-view Aerial Imagery, ISPRS Journal of Photogrammetry and Remote Sensing 58 (1-2) (2003) 83–98.

[14] S.A. Jimenez, R.I. Bashey, M. Benditt, R. Yankowski, Identification of Collagen α1(I) Trimer in Embryonic Chick Tendons and Calvaria, Biochem. Biophys. Res. Commun. 78 (4) (1977) 1354–1361.

[15] T. Kaczynski, K. Mischaikow, M. Mrozek, Computational Homology, Applied Mathematical Sciences 157, Springer-Verlag, 2004.

[16] C. Kirbas, F. Quek, A Review of Vessel Extraction Techniques and Algorithms, ACM Comput. Surv. 36 (2) (2004) 81–121.

[17] M.A. Konerding, E. Fait, A. Gaumann, 3D Microvascular Architecture of Precancerous Lesions and Invasive Carcinomas of the Colon, Brit. J. Cancer 84 (10) (2001) 1354–1362.

[18] V. Mann, E.E. Hobson, B. Li, T.L. Stewart, S.F. Grant, S.P. Robins, R.M. Aspden, S.H. Ralston, A COL1A1 Sp1 Binding Site Polymorphism Predisposes to Osteoporotic Fracture by Affecting Bone Density and Quality, J. Clin. Invest. 107 (7) (2001) 899–907.

[19] W.J McLaren, P. Anikijenko, S.G. Thomas, P.M. Delaney, R.G King, In Vivo Detection of Morphological and Microvascular Changes of the Colon in Association with Colitis Using Fiberoptic Confocal Imaging (FOCI), Digestive Diseases and Sciences 47 (11) (2002) 2424–2433.

[20] K. Misof, W.J. Landis, K. Klaushofer, P. Fratzl, Collagen from the Osteogenesis Imperfecta Mouse Model (Oim) Shows Reduced Resistance Against Tensile Stress, J. Clin. Invest. 100 (1) (1997) 40–45.

[21] M. Mrozek, *Homology Software*, http://www.ii.uj.edu.pl/~mrozek/software/homology.html, 2006.

[22] M. Mrozek, B. Batko, Coreduction Homology Algorithm, Discrete and Computational Geometry 41 (2009) 96–118.

[23] M. Mrozek, P. Pilarczyk, N. Żelazna, Homology Algorithm Based on Acyclic Subspace, Computers and Mathematics with Applications 55 (11) (2008) 2395–2412.

[24] M. Mrozek, M. Żelawski, A. Krajniak, A. Gryglewski, S. Han, Homological Methods in Feature Extraction of Multidimensional Images, Proceedings of the 2009 2nd International Congress on Image and Signal Processing, IEEE (2009) 1061–1066.

[25] M. Niethammer, A.N. Stein, W.D. Kalies, P. Pilarczyk, K. Mischaikow, A. Tannenbaum, Analysis of Blood Vessel Topology by Cubical Homology, Proceedings of the International Conference on Image Processing 2 (2002) 969–972.

[26] S. Peltier, S. Alayrangues, L. Fuchs, J.O. Lachaud, Computation of Homology Groups and Generators, Computers and Graphics 30 (2006) 62–69.

[27] B.J. Pfeiffer, C.L. Franklin, F.H. Hsieh, R.A. Bank, C.L. Phillips, Alpha 2(I) Collagen Deficient Oim Mice Have Altered Biomechanical Integrity, Collagen Content, and Collagen Crosslinking of Their Thoracic Aorta, Matrix Biol. 24 (7) (2005) 451–458.

[28] J.C. Russ, The Image Processing Handbook, 2nd ed., CRC Press, 1995.

[29] M. Sonka, V. Hlavac, R. Boyle, Image Processing, Analysis and Machine Vision, Brooks/Cole Publishing Company, 1999.

[30] S. Svensson, G. Sanniti di Baja, Simplifying Curve Skeletons in Volume Images, Computer Vision and Image Understanding 90 (3) (2003) 242–257.

[31] A. Szymczak, A. Stillman, A. Tannenbaum, K. Mischaikow, Coronary Vessel Trees from 3D Imagery: A Topological Approach, Med Image Anal. 10 (4) (2006) 548–559.

[32] F. Tupin, B. Houshmand, M. Datcu, Road Detection in Dense Urban Areas Using SAR Imagery and the Usefulness of Multiple Views, IEEE Transactions on Geosciense and Remote Sensing 40 (11) (2002) 2405–2414.

[33] A. Verri, C. Uras, Metric-topological Approach to Shape Representation and Recognition, Image Vision Comput. 14 (1996) 189–207.

[34] R.K. Wali, H.K. Roy, Y.L. Kim, Y. Liu, J.L. Koetsier, D.P. Kunte, M.J. Goldberg, V. Turzhitsky, V. Backman, Increased Microvascular Blood Content Is an Early Event in Colon Carcinogenesis, Gut 54 (2005) 654–660.

[35] D. Ziou, M. Allili, Generating Cubical Complexes from Image Data and Computation of the Euler Number, Pattern Recognition 35 (12) (2002) 2833–2839.
[36] M. Żelawski, Detecting Pathologies with Homology Algorithms in Magnetic Resonance Images of Brain, Machine Graphics & Vision 18 (3) (2009) 253–266.
[37] M. Żelawski, Pattern Recognition Based on Homology Theory, Machine Graphics & Vision 14 (3) (2005) 309–324.

About the author - MARIAN MROZEK received PhD degree in Mathematics from the Jagiellonian University in 1982 and then he worked as and adjunct professor at this university. From 1991 to 1993 he worked as a visiting researcher and then in the academic year 1996/97 as a visiting associate professor at Georgia Institute of Technology in Atlanta, USA. Since 1999 he is a professor at the Department of Computer Science of the Jagiellonian University in Kraków, Poland. His research interest is in the field of Computational Topology and Dynamics and their applications in sciences.

About the author - MARCIN ŻELAWSKI is an Assistant Professor at the Institute of Computer Science, Jagiellonian University, Kraków, Poland. He obtained M.Sc. degrees in computer science and psychology at the Jagiellonian University, in 1995 and 1998 respectively. In 2002 he received his PhD degree in computer science from the Jagiellonian University. His research interests include computational homology, pattern recognition, computer assisted proofs, computational complexity of algorithms and artificial intelligence.

About the author - ANDRZEJ GRYGLEWSKI is a Senior Assistants Professor in the I-st Department and Clinic of General and Gastroenterological Surgery, University Hospital in Kraków, Poland. From 2002 to 2007 he was a Senior Assistants Professor in the Department of Anatomy Collegium Medicum, Jagiellonian University in Kraków, Poland. In 1982 he obtained Surgery Specialization (I degree), in 1984 - Ph.D. and in 1991 - Final Surgery Specialization (II degree).

About the author - SEJIN HAN earned her BS in Physics at Yonsei University (Korea) in 1999, an MS (2006) and PhD (2009) in Physics at University of Maryland (USA). Presently she is a post-doctoral fellow in Mechanical engineering at University of Hong Kong. Her research interests include Tissue engineering and Stem cell mechanosense.

About the author - ANDRZEJ KRAJNIAK was born in Katowice, Poland, on 16 September 1984. He received the M.Sc. in 2008 from Jagiellonian University in Kraków, Poland and since then he is Ph.D. student

under professor Marian Mrozek supervision. His main research interest is image recognition.

Marian Mrozek, Institute of Computer Science, Jagiellonian University, ul. prof. Stanisława Łojasiewicza 6, 30-348 Kraków, Poland and Chair of Computational Mathematics, WSB-NLU, ul. Zielona 27, 33-300 Nowy Sącz, Poland
E-mail address: Marian.Mrozek@ii.uj.edu.pl

Marcin Żelawski, Institute of Computer Science, Jagiellonian University, ul. prof. Stanisława Łojasiewicza 6, 30-348 Kraków, Poland
E-mail address: Marcin.Zelawski@ii.uj.edu.pl

Andrzej Gryglewski, Department of General and Gastroenterological Surgery, University Hospital, ul. Kopernika 40, 31-501 Kraków, Poland
E-mail address: msgrygle@cyf-kr.edu.pl

Sejin Han, University of Hong Kong, Mechanical Engineering Department, Hong Kong, SAR of China
E-mail address: sejinhan@hku.hk

Andrzej Krajniak, Institute of Computer Science, Jagiellonian University, ul. prof. Stanisława Łojasiewicza 6, 30-348 Kraków, Poland
E-mail address: krajniak@ii.uj.edu.pl