20^{th} *Computer Vision Winter Workshop*
Paul Wohlhart, Vincent Lepetit (eds.)
Seggau, Austria, February 9-11, 2015

Novel Concepts for Recognition and Representation of Structure in Spatio-Temporal Classes of Images

Ines Janusch, Walter G. Kropatsch
Institute of Computer Graphics and Algorithms
Pattern Recognition and Image Processing Group
Vienna University of Technology
Vienna, Austria
{ines, krw}@prip.tuwien.ac.at

Abstract. *This paper discusses open problems and future research regarding the recognition and representation of structures in sequences of either 2D images or 3D data. All presented concepts aim at improving the recognition of structure in data (especially by decreasing the influence of noise) and at extending the representational power of known descriptors (within the scope of this paper graphs and skeletons). For the recognition of structure critical points of a shape may be computed. We present an approach to derive such critical points based on a combination of skeletons and local features along a skeleton. We further consider classes of data (for example a temporal sequence of images of an object), instead of a single data sample only. This so called co-analysis reduces the sensitivity of analysis to noise in the data. Moreover, a representative for a whole class can be provided. Temporal sequences may not only be used as a class of data in a co-analysis process - focusing on the temporal aspect and changes of the data over time an analysis of these changes is needed. For this purpose we explore the possibility to analyse a shape over time and to derive a spatio-temporal representation. To extend the representational power of skeletons we further present an extension to skeletons using model fitting.*

1. Introduction

A single 2D image is defined in the spatial domain. By extending data from a single capturing to a sequence of such data temporal information is added and the data is defined in the spatio-temporal domain [8]. Instead of capturing a single 2D image or 3D data (e.g. a 3D point cloud) the data may be extended to an image sequence or a sequence of 3D data. Temporal information as motion or development over time are thereby added to the representation. Therefore, this paper focuses on novel concepts for the identification of structure from sequences of images or 3D data and on the representation of this structure.

Applications and spatio-temporal datasets for the concepts proposed in this paper can, for example, be found in biology or in medicine. For the latter, spatio-temporal data may describe recurring sequences which may be the motion of an organ or abnormal changes of an organ caused by an illness. In biology, temporal image sequences can, amongst others, be found in plant phenotyping where plants or their roots are imaged on successive days of growth [9]. Furthermore, phenotyping of animals is currently still based on the manual analysis of experts [4]. An analysis of a sequence of 3D scans of an animal may be a future alternative as it provides spatio-temporal data showing the animal as well as its movements.

For any analysis of the captured object this object first needs to be detected in the data and processed to compute a suitable representation. Well known representations are Reeb graphs (as described in [1] and skeletons as for example a medial axis or a more sophisticated 3D Curve Skeleton (as described in [2]). For the computation of these representations a binary segmentation of the input data into foreground regions, representing the object to be analysed, and background regions, showing the rest of the data that is not in focus, is needed. However, such a segmentation may introduce artefacts that falsify the representation. We encountered this problem in [10],

where we applied knowledge about the structure to be represented and post-processing methods as for example graph pruning in order to reduce segmentation artefacts kept in the representation. Werghi et al. applied a similar approach in [20]. They handle noise in the input data by knowledge about the structure to be represented and in this way detecting and discarding improper configurations in the representation.
In this paper we discuss general methods to improve representations of data based on a potentially flawed segmentation. In this context we discuss the use of co-analysis for classes of data as well as the application of co-analysis and co-representation for changing, respectively developing shapes. Mitra [15] provides a detailed survey on co-analysis and co-segmentation. Promising methods of co-analysis have for example been presented by Golovinskiy et al. in [6] and van Kaick et al. in [18].
Additionally to co-analysis we propose two novel skeleton based representations: A graph representation using skeletons together with local features and a model based representation that is derived using model fitting to an initial skeleton.

The rest of the paper is structured as follows: Section 2 proposes the use of local features for the computation of critical points while Section 3 bases this computation on a function according to time. The analysis of a whole class of data using so called co-analysis and the representation of such classes using a co-representation is discussed in Section 4. Section 5 introduces extensions to known skeletons that improve their representational power and Section 6 concludes the paper.

2. Critical Points Based on Local Features

Graph based representations or skeletons rely on segmented input data. Thus, for the input data a binary segmentation - a separation between background (not of interest) and foreground (to be represented) - needs to be known. However, such a pre-processing of the data may introduce artefacts. Representations based on flawed segmented data can be improved using post-processing steps that detect and correct spurious parts of the representation. For graph representations a simple graph pruning may for example be applied. However, graph pruning may not remove all spurious branches (false negatives) or discard true branches (false positives).

$x_1 \geq c$	$x_2 < c$	$x_3 < c$
$x_8 \geq c$	c	$x_4 < c$
$x_7 \geq c$	$x_6 \geq c$	$x_5 \geq c$

(a) center pixel and neighbourhood

1	0	0
1	c	0
1	1	1

(b) comparison with neighbourhood

2^0	2^1	2^2
2^7	c	2^3
2^6	2^5	2^4

(c) neighbourhood pattern

$$\textbf{LBP} = 10001111 = 2^0 + 2^4 + 2^5 + 2^6 + 2^7 = \mathbf{241}$$

(d) LBP operator for center pixel c

Figure 1: Simple LBP computation.

A graph representation based on segmented data can only provide reliable results for a correct segmentation.
Instead of applying post-processing techniques to reduce artefacts introduced by the segmentation we propose to base the representation on the original unsegmented data. For a Reeb graph representation critical points may be computed on the original data instead of the segmented data. Local Binary Patterns (LBPs) [17] are considered as one method to derive such critical points on an unsegmented image.

LBPs were introduced as as a tool of texture classification and work (in their simplest version) as shown in Figure 1: The center pixel is compared to its neighbourhood. The relations of this comparison are stored as a bit pattern: In case a neighbouring pixel is larger or equal the center pixel its bit is set to 1 otherwise to 0. The neighbourhood pattern is encoded as the position of each neighbourhood pixel in a binary data item [16].
Critical points on a shape according to a Morse function build the nodes in a Reeb graph. Such critical points (in 2D) are minimum, maximum and saddle points. The configuration of the neighbourhood around a pixel encodes the local topology. The region may be a (local) maximum (the bit pattern contains only 0s), a (local) minimum (the bit pattern contains only 1s), a plateau (the bit pattern contains only 1s, but all pixels of the region have the same gray value), a slope (the bit pattern of the region contains one connected component of 1s and

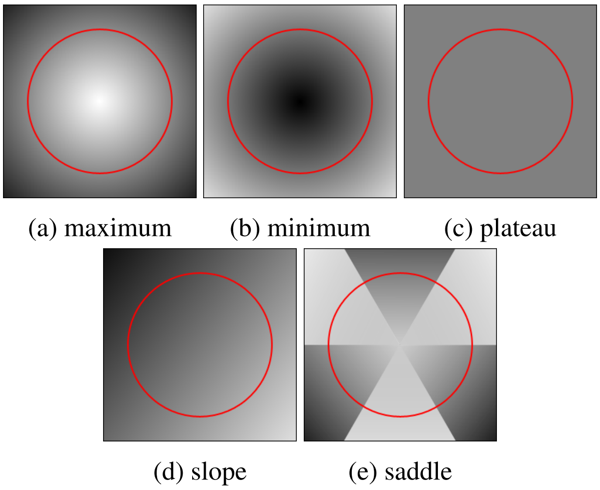


Figure 2: Neighbourhood configuration detected by LBPS. The red circle indicates the neighbourhood used in the LBP computation for the pixel at the position of the center of the circle.

one connected component of 0s) or a saddle point otherwise [7]. Figure 2 shows examples for all these region configurations that may be encoded by LBPs.

The original LBP operator was defined for the spatial domain only. Similar to the work of Laptev who extended the Harris and Förstner interest point operator to space-time interest points in [12] and [11] the LBP description of local structures was extended in time to describe local features in the spatio-temporal domain [21]. The so called Volume Local Binary Pattern (VLBP) represents dynamic textures as volumes of (X, Y, T), where X and Y are the spatial coordinates, T, as a temporal coordinate, represents points in time. A sequence of dynamic textures over time is therefore represented by a VLBP.

Reeb graphs are derived on binary segmented 2D or 3D data using an analysis based on a Morse function as for example a height function. In order to analyse unsegmented data local descriptors as for example LBPs may be used as Morse function, provided that the descriptors satisfy the necessary conditions, analog to the conditions of Morse functions [3].
Despite the idea to avoid segmentation as a preprocessing step, this approach works on a segmented image as a first input. However, the critical points are computed on the unsegmented data, the segmented image is only needed to guide the computation of the critical points as follows:
On the initial segmentation the medial axis is computed for the foreground region. The medial axis is formed by the centers of maximal circles that cover the shape completely. Therefore, the medial axis implicitly provides a measure of width, as for each point along the medial axis the radius of the inscribed maximal circle (the distance to the boundary) is known [13]. Along this skeleton LBPs are computed for each skeleton pixel. The LBP kernel size is thereby determined by the radius associated with the individual skeleton pixels. Minima, maxima and saddle points that are encountered in this way may then be used as critical points (nodes) in a graph, connections of these nodes can be derived from the skeleton.
In case the position of the skeleton, respectively the critical points, can be estimated (for example in video data based on the position in a previous frame) the segmentation as well as the initial skeleton do not need to be recomputed. Rather this known approximation can be reused to guide the computation of the critical points (in the next frame).

3. Time as Morse Function

Analysing data over time adds one dimension to the original data domain. Edelsbrunner et al. introduced time varying Reeb graphs in [5]. They present an algorithm to maintain Reeb graphs through time and to store the graph's evolution.
For 2D images of shapes that change over time we can augment the spatial information of the pixels with temporal information by storing as a third coordinate the point in time the according pixel was first

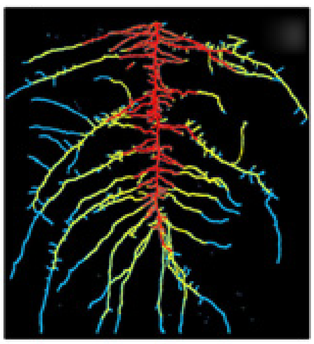
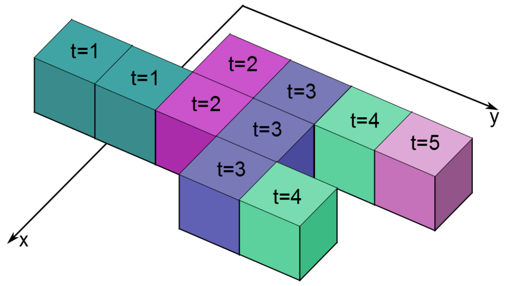
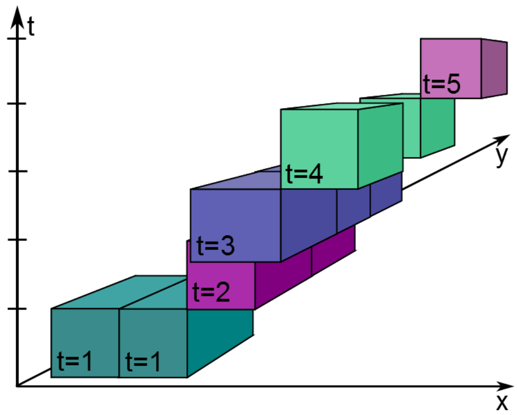


Figure 3: Spatial information of a growing root augmented with temporal information: the image shows a segmented lupine root, the colors indicate measurement time. Image courtesy of Leitner et al. [14]

encountered. Shapes that grow are imaged on several points of time through this development process. After an alignment of the acquired data according to the last acquisition (the most mature one), parts of the growing shape are labeled according to the time they were first encountered. This representation aims specifically at the representation of growth, temporal deformations are not represented. Figure 3 acquired by Leitner et al. [14] shows an example for such a dataset: individual parts of the root are labeled according to the time these parts were first observed. For such a configuration a height function along the temporal axis (time function) may be used to extract level sets. These level sets represent the evolution of the shape over time. Figure 4 illustrates the proposed approach: Figure 4a shows the augmented spatial information, Figure 4b shows level sets of this data according to time.
A Reeb graph can be built, as the time is used equally to a height function as Morse function. In order to build the Reeb graph the individual components are connected by tracing through the spatial information from one time step to the next.



(a) spatial information augmented with temporal information



(b) level sets with respect to time

Figure 4: Augmenting the spatial information of a growing structure with temporal information adds one dimension. A height function along this dimension provides a Morse function with respect to time.

4. Co-Analysis and Co-Representation

For the recognition and representation of structure methods based not only on a single object, but on a class of similar objects may be used. This so called co-analysis focuses on a common structure of all objects in the class and on relations between parts of the object and thereby reduces influence of noise in the capturing of a single image. Co-analysis may further reduce the time needed for a training phase, as objects of a class are simultaneously analysed in co-analysis [19].
To perform co-analysis an alignment of the individual data samples is needed first. Van Kaick et al. for example in [18] assume the shapes to be upright-oriented and partitioned into meaningful parts. Golovinskiy et al. in [6] base the shape alignment on the alignment of axis according to a principal component analysis. We propose to align 2D or 3D image data samples using standard representations as for example skeletons or Reeb graphs. Other than the alignment according to an axis, skeletons and graphs allow for an alignment of data samples of articulated or varying objects.
Spatio-temporal data may create classes of images as for example in biological datasets a growing organism may be imaged over time. Therefore, the data consists of sets of related data samples that have certain features in common. An analysis of a collection of data samples is called co-analysis [15] . The aim of this procedure is to label the same entities with the same labels although they may appear in variations. Considering for example drinking glasses: in co-analysis a collection of different glasses is used. All glasses have a body that may hold liquids, some of them may have a stem. Independent on the actual design of the stem (long and thin or short and decorated) this part of the glass should always be detected as stem.
Further knowledge based on co-analysis of a class of data can be used to verify decisions for a single data sample.
Considering the mentioned biological applications, co-analysis of data samples could for example be beneficial in the area of plant phenotyping. Here classes of images are formed as images of several different plants (therefore potentially different

(a) day 12 (b) day 16 (c) day 20

Figure 5: Application for co-analysis: root development. Small branches (indicated by red arrows) may only represent noise. An image of a later day can verify a decision - in this case true branches (indicated by the green tickmark).

phenotypes) exits for the same genotype. Moreover, the plants are imaged on successive days of growth. The temporal stack of images of a single plant can be seen as a class of images. Analysis decisions are not taken for a single image but considering the whole class. Artefacts may therefore be detected and reduced as decisions for a single image are verified considering the remaining images of the class.

After an initial segmentation and alignment of the images, representational decisions, for example whether a branch is a spurious branch or not, can be verified using an image acquired at a later time. Figure 5 shows such a temporal sequence of root images. In Figure 5a and 5b two small branches that may be identified as noise in a single image are indicated by red arrows. The later images in Figure 5b and 5c identify these small branches as true branches (indicated by the green tickmark). In this case the small branches would be kept in the final representations.

Co-analysis may further be used in the development of representations of a whole class of data (co-representation) instead of a single data sample. The reduction of a class of data to its characteristic properties provides a general representation of the whole class of data. Such a co-representation can for example be given by a graph that represents the properties valid for the whole class of data. For shape representation geometric graphs may be used. Especially when analysing and representing the content of an image, a node may be assigned to a pixel, therefore geometry is implicitly represented. In order to use a graph as a co-representation of a class of data, we may represent each data sample using a geometric graph. However, for the general graph derived as a co-representation the graphs representing single data samples may be analysed for common topological structures which are represented in the final co-representation graph. Geometric properties are in this case disregarded in the co-representation.
However, for a developing shape the representation of the latest acquisition may be taken as the co-representation for the whole sequence. In this way geometric properties can be kept in the representation and further data samples can be mapped to the co-representation.

5. Extensions to Skeletons

Skeletons (medial axes) given as a thinned version of a shape with equal distance to the boundary, are widely used shape descriptors. In order to extend the definition and representational power of such a common skeleton, we propose a combination of skeletons and model fitting.

The contemplated approach works as follows:

1. The medial axis is computed for the shape first. On the obtained skeleton a constrained distance transform is performed - the geodesic distance along the skeleton is computed in this way.

2. To allow for the fitting of simple models, the axis needs to be straightened first. Therefore, the medial axis is decomposed into single curve segments at branching points. For each pixel along the skeleton the distance to the starting point of the segment (and further to a starting point of the whole skeleton) as well as the distance to the boundary (the radius of the maximal inscribed circle at this position) are known.

3. Simple models as a parabola, an ellipse, a cylinder or similar (higher order e.g. super-quadrics) models can be fitted to the transformed data.

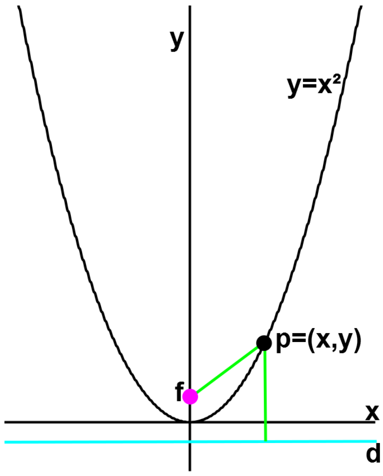


Figure 6: Example parabola, together with focus f and directrix d.

4. Fitted models may further be back-projected to the original domain.

We describe the process of model fitting in more detail using a parabola:

A parabola is defined as the locus of points equidistant from one point (the focus) and one line (the directrix). Figure 6 shows an example.
The straightened skeleton serves as the axis of symmetry of the parabola to be found. In order to fit this parabola to the data points two positions along the axis of symmetry need to be determined:

a. position f along the axis: position of the focus

b. position d along the axis: position of the directrix

For any point $p = (x, y)$ of the parabola the following equation holds:

$$\sqrt{x^2 + (f - y)^2} = |d - y| \qquad (1)$$

Reformulating equation 1 yields the dependency of x from y as follows:

$$x(y)^2 = (d - f)(d + f - 2y) \qquad (2)$$

We determine f and d by minimizing the sum of squared errors between the actual measured values x_i^2 (x_i corresponds to the radius stored with each skeleton pixel) and the value $x(y_i)^2$ given by a model parabola as formulated in equation 2, thus determine the parameters of an optimal fitted parabola.

Skeletons enhanced in this way may amongst others be used to:

- detect artefacts in an image segmentation;
- segment an object into meaningful parts based on fitted models;
- represent a particular shape or its properties using the parameters of the fitted model.

Applications for the above mentioned enhanced skeletons can for example be found in the analysis of biological data: Roots due to their elongated shape and narrow root tip can be approximated by a parabola. The parabola model may be used to improve the segmentation of an image into root region an background. Such a segmentation tends to introduce artefacts due to for example root hairs that falsify the segmented regions. Figure 7 illustrates an example of a parabola fitted to a root branch, according to the described approach. Additionally, to an improvement of the segmentation, the parabola parameters themselves may be used to describe the root and to model its growth.
Another application for enhanced skeletons is presented by the segmentation of 3D objects into rigid

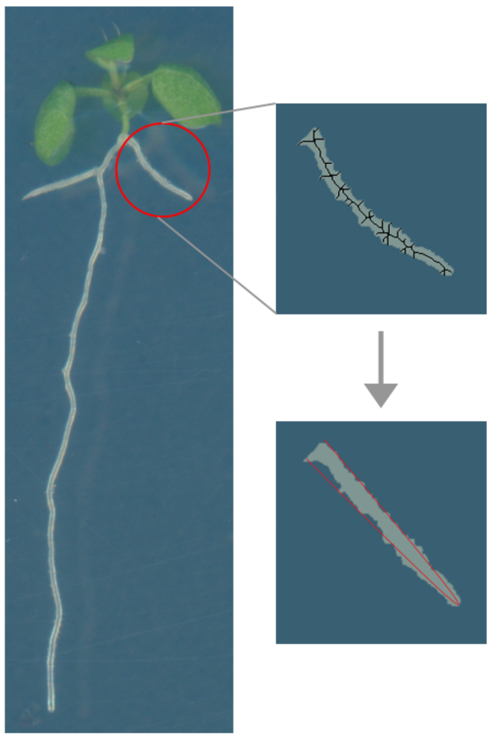


Figure 7: Root segment straightened according to a medial axis and parabola model fitted to the root tip (illustrative model).

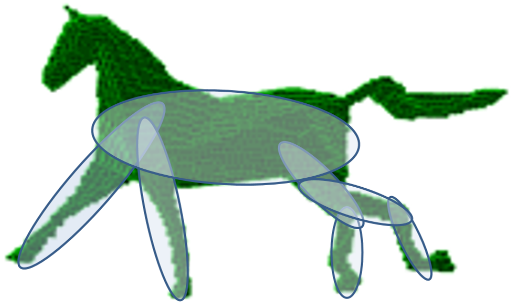


Figure 8: Representation of rigid parts of a horse using fitted ellipses (illustrative model).

parts. Instead of basing this segmentation on a simple medial axis alone, a 3D model of for example a horse can be segmented into rigid parts, by fitting ellipsoids to the individual parts. Elongated shapes as for example the torso of a horse may be better represented by ellipsoids than by spheres which are fitted for the medial axis representation. Moreover, the focal points of the fitted ellipsoids can be overlapped in the individual rigid parts, thereby representing connections between parts (in the horse model case these connections are the joints). Figure 8 illustrates this example.

6. Conclusion

The presented approaches are novel concepts and extensions to the current state-of-the-art. Common approaches for the extraction and representation of structure (for example skeletons and Reeb graphs) are sensitive to noise and depend on the quality of the binary segmented input image. The concepts introduced in this paper aim at decreasing this sensitivity towards noise.

An initial segmentation and an initial skeleton representation may be improved by considering local features along the skeleton or by model fitting using straight segments of the skeleton as axis of symmetry of a model. Both approaches in the end provide compact representations of the input data. A potentially noise flawed segmentation can be used as an initial input as the contemplated approaches can cope with a rough segmentation - while a representation is found, the segmentation may simultaneously be improved.

We further extend the known co-analysis to a more general approach by aligning several data samples according to skeletons or graphs instead of an axis. Therefore, the limitation of the alignment to non-articulated objects only is revoked. Based on graphs or skeletons objects in varying poses may then be used in the co-analysis by aligning their rigid parts.

For well aligned data samples over time we propose to add the temporal information as an additional dimension. A Reeb graph representation may in this case be built by using a height function along the time axis and by tracing back the evolution of the connected components.

These separate ideas may be joined together to improve representations of structures in spatio-temporal data: Robust skeleton representations may be used as an alignment of several data samples respectively their contained structure. Such aligned data samples in turn build the input for a Reeb graph representation over time, representing temporal changes in the structure. As well as for co-analysis which may provide a co-representation - a representation of a whole class of objects. For comparison of data samples, new samples may then be mapped to the co-representation and compared with the class. Furthermore, an initial representation may be improved by fitting a model to it. This model may even be back projected to the input data thereby also improving the segmentation.

Future work includes the implementation of these presented ideas and evaluation on (for example the mentioned biological roots and horses) data sets. An investigation of further approaches to overcome the discussed open problems in the recognition and representation of structures in spatio-temporal data are as well subject to future work.

Acknowledgements

We thank the anonymous reviewers for their constructive comments.

References

[1] S. Biasotti, D. Giorgi, M. Spagnuolo, and B. Falcidieno. Reeb graphs for shape analysis and applications. *Theoretical Computer Science*, 392(1-3):5–22, Feb. 2008. 1

[2] G. S. di Baja, L. Serino, and C. Arcelli. 3d curve skeleton computation and use for discrete shape analysis. In *Innovations for Shape Analysis*, pages 117–136. Springer, 2013. 1

[3] H. Doraiswamy and V. Natarajan. Efficient algorithms for computing Reeb graphs. *Computational Geometry*, 42(67):606–616, Aug. 2009. 3

[4] T. Druml, A. Gabdulkhakova, N. Artner, G. Brem, and W. Kropatsch. The use of image data in the assessment of equine conformation – limitations and

solutions. In R. B. Fisher, J. Hammal, B. Boom, and C. Spampinato, editors, *ICPR 2014 workshop on Visual observation and analysis of Vertebrate And Insect Behavior*, page in press, 2014. 1

[5] H. Edelsbrunner, J. Harer, A. Mascarenhas, V. Pascucci, and J. Snoeyink. Time-varying reeb graphs for continuous space–time data. *Computational Geometry*, 41(3):149–166, 2008. 3

[6] A. Golovinskiy and T. Funkhouser. Consistent segmentation of 3d models. *Computers & Graphics*, 33(3):262–269, 2009. 2, 4

[7] R. Gonzalez-Diaz, W. Kropatsch, M. Cerman, and J. Lamar. Characterizing configurations of critical points through lbp. In *SYNASC 2014 Workshop on Computational Topology in Image Context*, 2014. 3

[8] B. Jähne. *Spatio-temporal image processing: theory and scientific applications*, volume 751. Springer, 1993. 1

[9] I. Janusch, W. G. Kropatsch, and W. Busch. Reeb graph based examination of root development. In *Proceedings of the 19th Computer Vision Winter Workshop*, pages 43–50, Feb 2014. 1

[10] I. Janusch, W. G. Kropatsch, W. Busch, and D. Ristova. Representing roots on the basis of reeb graphs in plant phenotyping. In *ECCV 2014 Workshop on Computer Vision Problems in Plant Phenotyping*, page in press, 2014. 1

[11] I. Laptev. On space-time interest points. *International Journal of Computer Vision*, 64(2-3):107–123, 2005. 3

[12] I. Laptev and T. Lindeberg. Space-time interest points. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV03)*, pages 432–439, 2003. 3

[13] D. T. Lee. Medial axis transformation of a planar shape. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-4(4):363–369, July 1982. 3

[14] D. Leitner, B. Felderer, P. Vontobel, and A. Schnepf. Recovering root system traits using image analysis exemplified by two-dimensional neutron radiography images of lupine. *Plant Physiology*, 164(1):24–35, 2014. 3, 4

[15] N. J. Mitra, M. Wand, H. Zhang, D. Cohen-Or, V. Kim, and Q.-X. Huang. Structure-aware shape processing. In *ACM SIGGRAPH 2014 Courses, SIGGRAPH '14*, pages 13:1–13:21, 2014. 2, 4

[16] T. Ojala and M. Pietikäinen. Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32(3):477–486, 1999. 2

[17] M. Pietikäinen. Image analysis with local binary patterns. In *Image Analysis*, pages 115–118. Springer, 2005. 2

[18] O. van Kaick, K. Xu, H. Zhang, Y. Wang, S. Sun, A. Shamir, and D. Cohen-Or. Co-hierarchical analysis of shape structures. *ACM Transactions on Graphics (TOG)*, 32(4):69, 2013. 2, 4

[19] Y. Wang, S. Asafi, O. van Kaick, H. Zhang, D. Cohen-Or, and B. Chen. Active co-analysis of a set of shapes. *ACM Transactions on Graphics (TOG)*, 31(6):165, 2012. 4

[20] N. Werghi, Y. Xiao, and J. Siebert. A functional-based segmentation of human body scans in arbitrary postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(1):153–165, 2006. 2

[21] G. Zhao and M. Pietikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):915–928, 2007. 3