Willkommen
Welcome
Bienvenue

Materials Science & Technology

Segmentation algorithms for porous (building) materials

An (incomplete and biased) overview

M. Griffa[1], B. Münch[1]

[1]Concrete and Construction Chemistry Laboratory, EMPA, Dübendorf, Switzerland

Outline

- Segmentation of porous (building) materials
- Overview of some methods
  - Global thresholding algorithms
  - Data clustering algorithms
- Conclusions
- Suggested bibliography

Outline

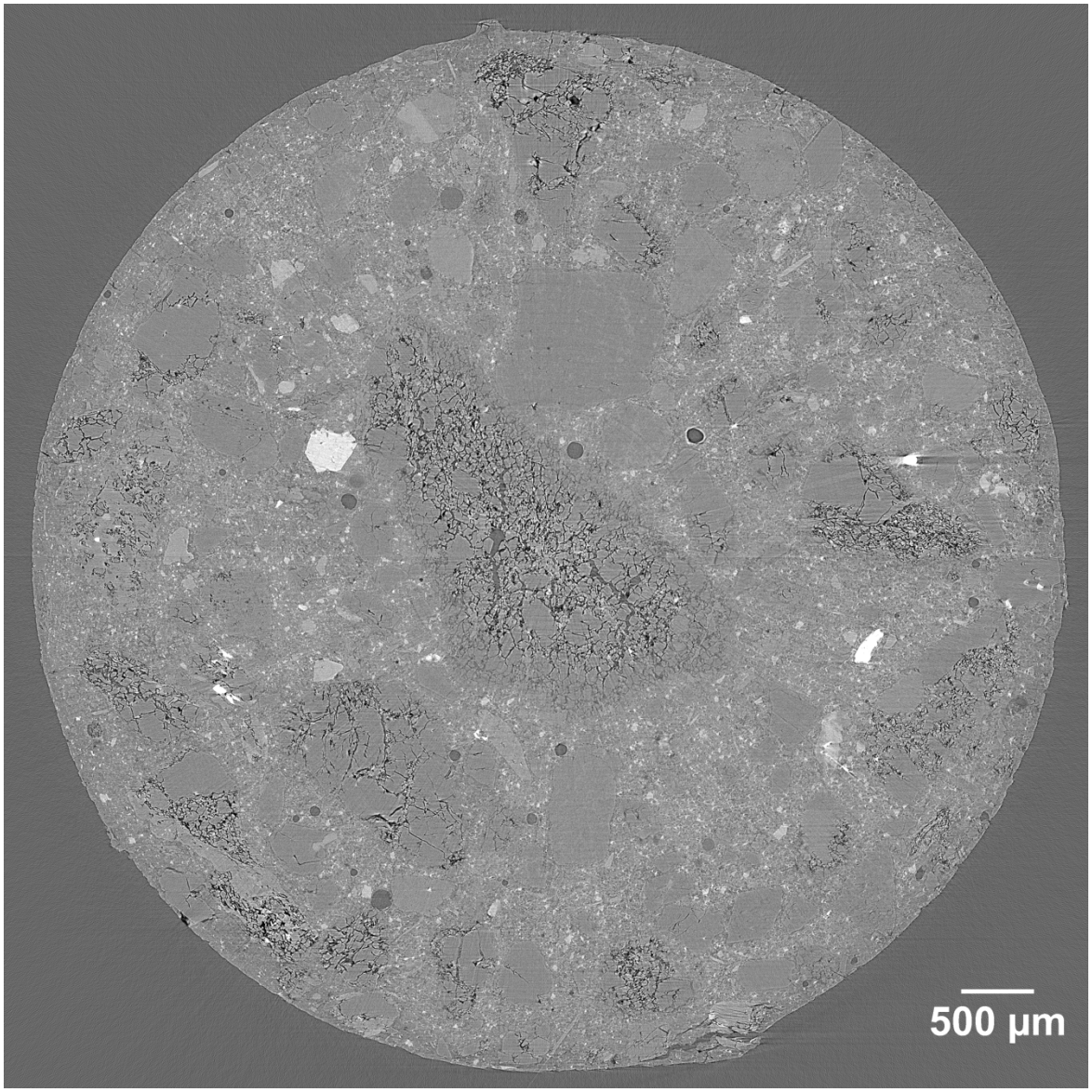
- Segmentation of porous (building) materials

- Overview of some methods
  - Global thresholding algorithms
  - Data clustering algorithms
- Conclusions
- Suggested bibliography

Digital porous (building) materials: an example

EMPA
Materials Science & Technology

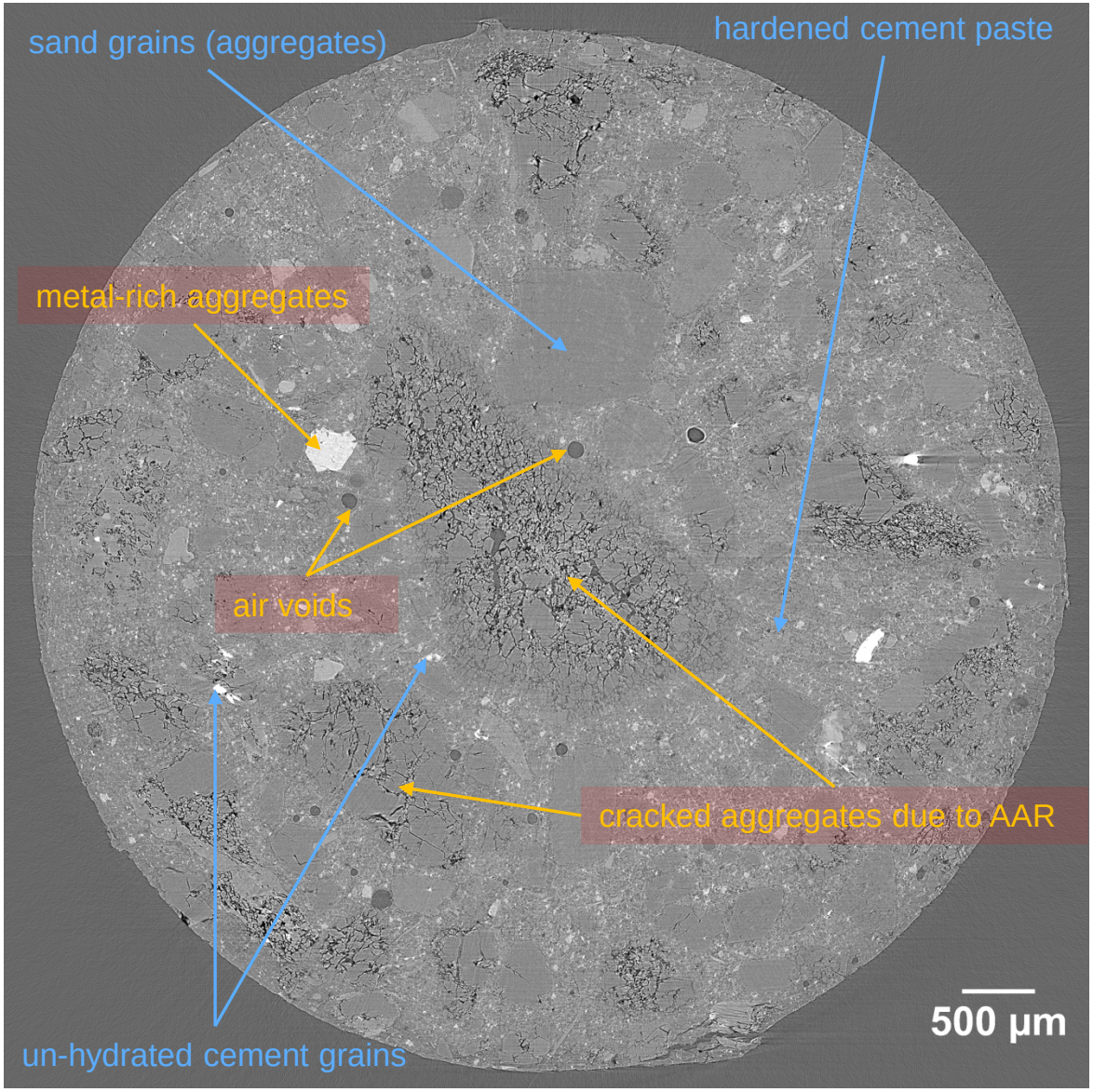
- single cross-sectional digital slice from a volume
- synchrotron radiation-based X-ray Tomographic Microscopy (XTM)
- measurements at the TOMCAT beamline, SLS, PSI.
- cylindrical concrete sample, 7 mm ø
- degraded by autogenous chemical reactions (Alkali-Aggregate Reactions, AAR)
- acceleration of the reactions with lab protocols
- the voxel (grey) values correspond to a superposition of the **X-ray linear attenuation coefficient**, μ, and of the **Laplacian** of the **X-ray index of refraction**, Δn
- wide Field-Of-View configuration (4020 x 4020 pixels)
- voxel lateral size = 1.85 μm.



Digital porous (building) materials: an example

EMPA
Materials Science & Technology

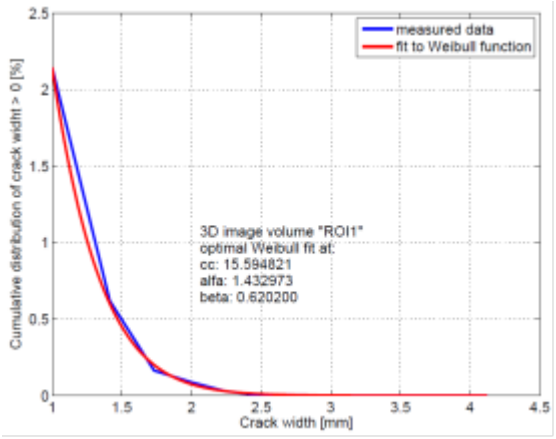
- single cross-sectional digital slice from a volume
- synchrotron radiation-based X-ray Tomographic Microscopy (XTM)
- measurements at the TOMCAT beamline, SLS, PSI.
- cylindrical concrete sample, 7 mm ø
- degraded by autogenous chemical reactions (Alkali-Aggregate Reactions, AAR)
- acceleration of the reactions with lab protocols
- the voxel (grey) values correspond to a superposition of the **X-ray linear attenuation coefficient**, μ, and of the **Laplacian** of the **X-ray index of refraction**, Δn
- wide Field-Of-View configuration (4020 x 4020 pixels)
- voxel lateral size = 1.85 μm.

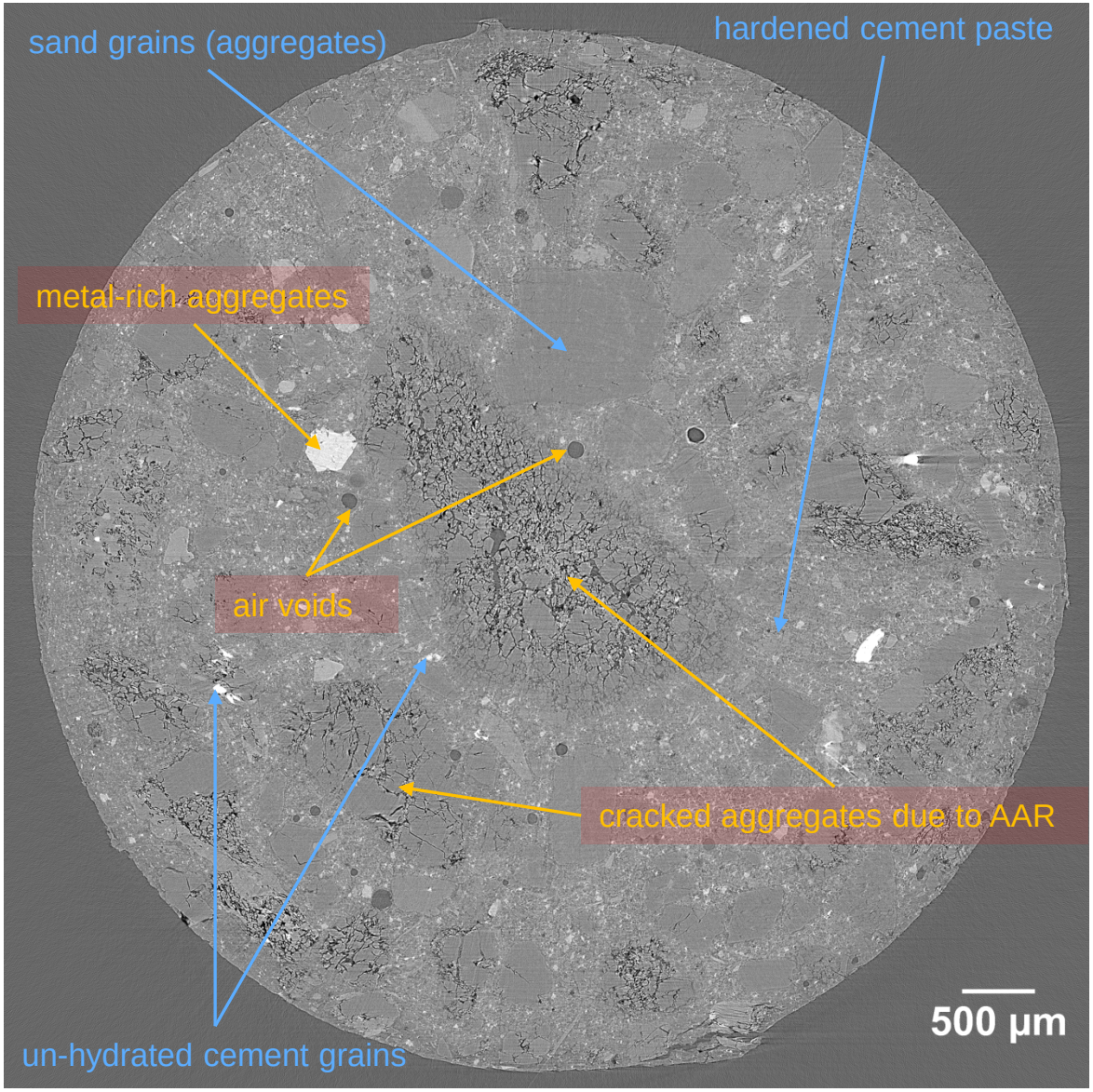


Segmentation of porous (building) materials

EMPA
Materials Science & Technology

- **segmentation** = distinction among pixels/voxels belonging to different physical objects, e.g., pore space, solid material phases, aggregates, cement paste, etc. …
- segmentation as the 1st step towards extracting quantitative information
- Example related with this XTM dataset:
  - ➢ find the voxels beloning to the pore space
  - ➢ separate pores from cracks
  - ➢ calculate the crack width distribution (by the Euclidean Distance Transform and the Canny edge detector)





Outline

- Segmentation of porous (building) materials
- Overview of some methods
  - Global thresholding algorithms
  - Data clustering algorithms
- Conclusions
- Suggested bibliography

Segmentation of porous (building) materials: types of algorithms

Materials Science & Technology

Elefanten Park – Zoo Zürich – Stadt Zürich

- a ``zoo´´ of algorithm types
- their classification is not unique

Segmentation of porous (building) materials: types of algorithms

Materials Science & Technology

Elefanten Park – Zoo Zürich – Stadt Zürich

- a ``zoo´´ of algorithm types
- their classification is not unique

Regarding images of porous materials:

Segmentation of porous (building) materials: types of algorithms

Materials Science & Technology

Elefanten Park – Zoo Zürich – Stadt Zürich

- a ``zoo´´ of algorithm types
- their classification is not unique

Regarding images of porous materials:

- segmentation based upon **ranges of pixel/voxel values** (**thresholding**)
  - ➢ global thresholding (thresholds defined for the whole image)
  - ➢ local thresholding (thresholds defined for each pixel/voxel's neighborhood)

Segmentation of porous (building) materials: types of algorithms

EMPA
Materials Science & Technology

Elefanten Park – Zoo Zürich – Stadt Zürich

- a ``zoo´´ of algorithm types
- their classification is not unique

Regarding images of porous materials:

- segmentation based upon **ranges of pixel/voxel values** (**thresholding**)
  - ➢ global thresholding (thresholds defined for the whole image)
  - ➢ local thresholding (thresholds defined for each pixel/voxel's neighborhood)
- segmentation based upon **proximity of pixel/voxel** in a **feature space** (**data clustering**)

Segmentation of porous (building) materials: types of algorithms

EMPA
Materials Science & Technology

Elefanten Park – Zoo Zürich – Stadt Zürich

- a ``zoo´´ of algorithm types
- their classification is not unique

Regarding images of porous materials:

- segmentation based upon **ranges of pixel/voxel values** (**thresholding**)
  - ➢ global thresholding (thresholds defined for the whole image)
  - ➢ local thresholding (thresholds defined for each pixel/voxel's neighborhood)
- segmentation based upon **proximity of pixel/voxel** in a **feature space** (**data clustering**)
- segmentation based upon spatial contiguity and similarity of values for pixels/voxels (**seeded region growing**)

Segmentation of porous (building) materials: types of algorithms

EMPA
Materials Science & Technology

Elefanten Park – Zoo Zürich – Stadt Zürich

- a ``zoo´´ of algorithm types
- their classification is not unique

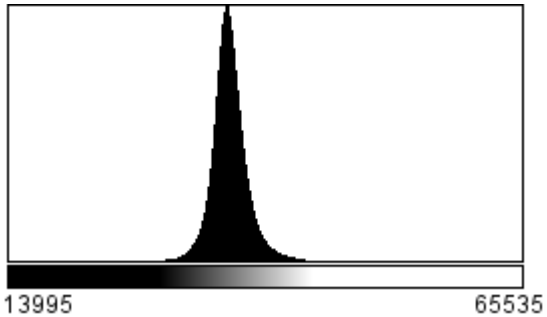
Regarding images of porous materials:

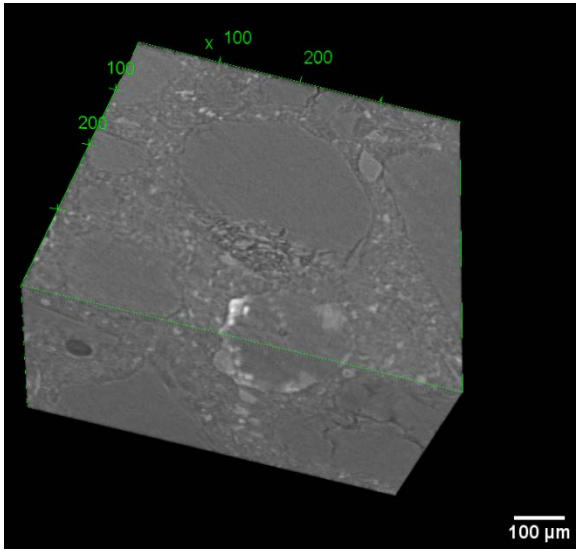
- segmentation based upon **ranges of pixel/voxel values** (**thresholding**)
  - ➢ global thresholding (thresholds defined for the whole image)
  - ➢ local thresholding (thresholds defined for each pixel/voxel's neighborhood)
- segmentation based upon **proximity of pixel/voxel** in a **feature space** (**data clustering**)
- segmentation based upon spatial contiguity and similarity of values for pixels/voxels (**seeded region growing**)
- segmentation based upon surface evolution (**level set + fast marching methods**)
- segmentation based upon a probabilistic analysis of the image (**Bayesian analysis**)

{1}: case study

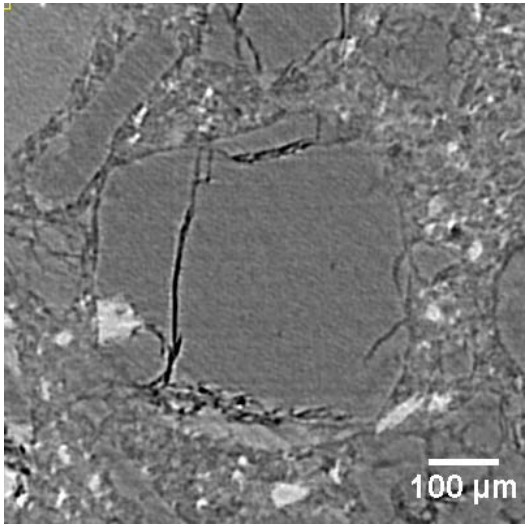
EMPA
Materials Science & Technology

- focus on a smaller ROI, 400 x 400 x 201 voxels
- unimodal voxel value histogram





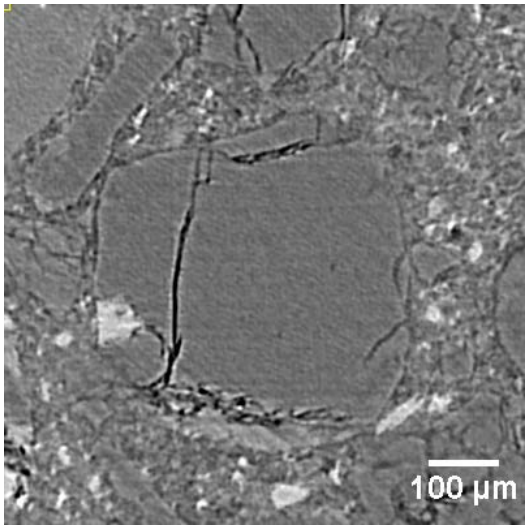
- I want to segment the pore space, all the rest is considered as solid materials
- complicated pore system: cracks + aggregate dissolution pores



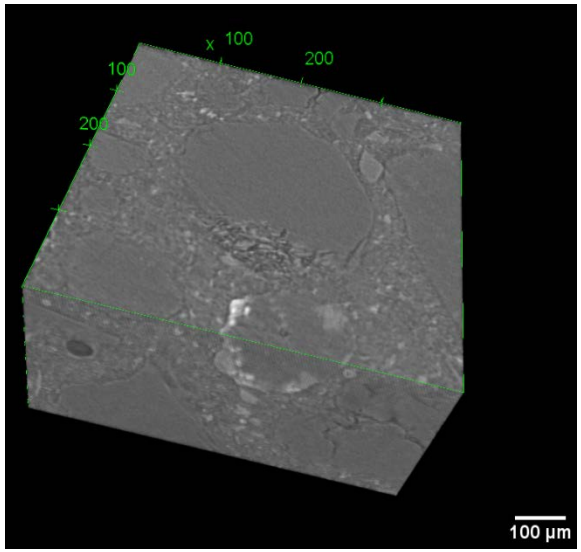
slice # 98 of 201

{2}: how to think of an image (mathematically)

- scalar/vector field $(f(\vec{x}), \vec{f}(\vec{x}) = (f_i(\vec{x})), \vec{x} \in \mathbb{R}^2 \; or \; \mathbb{R}^3)$



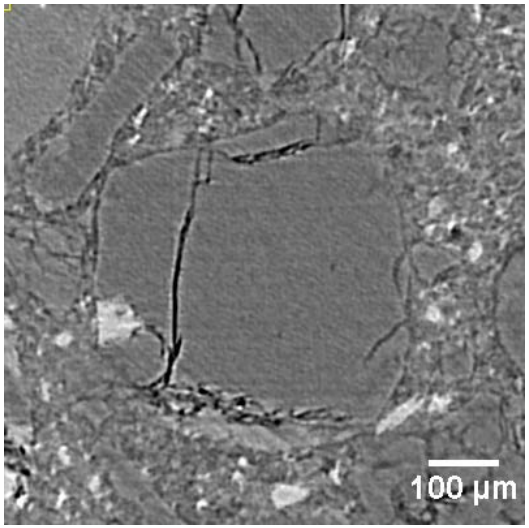
slice # 98 of 201



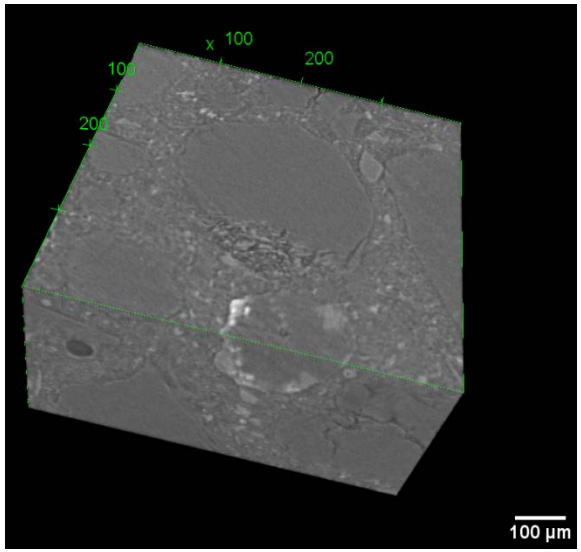
{2}: how to think of an image (mathematically)

EMPA
Materials Science & Technology

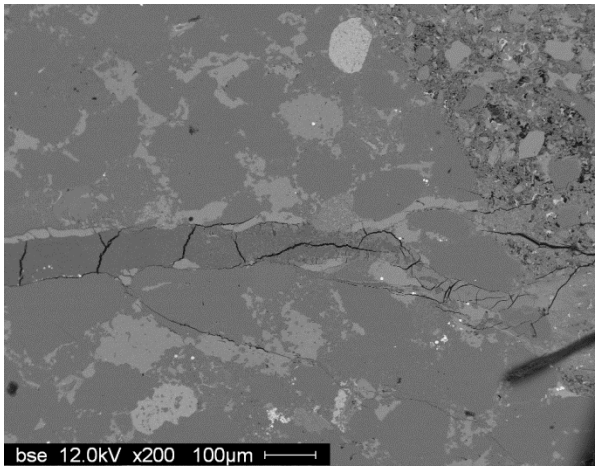
- scalar/vector field $(f(\vec{x}), \vec{f}(\vec{x}) = (f_i(\vec{x})), \vec{x} \in \mathbb{R}^2 \; or \; \mathbb{R}^3)$



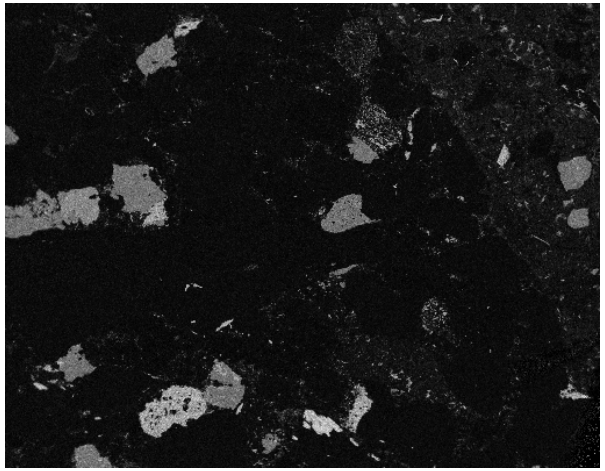
slice # 98 of 201



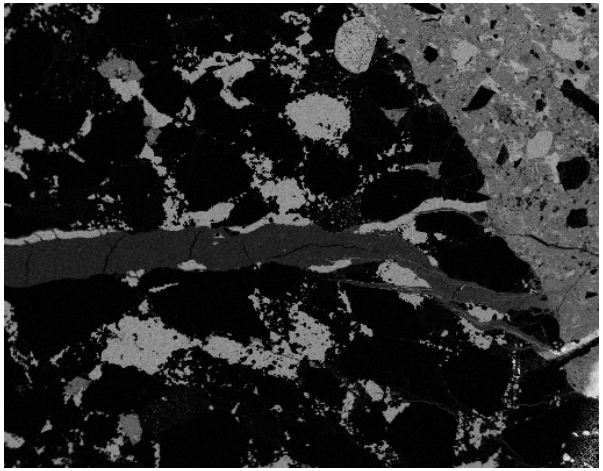
AAR-damage concrete sample



back-scattering SEM micrograph



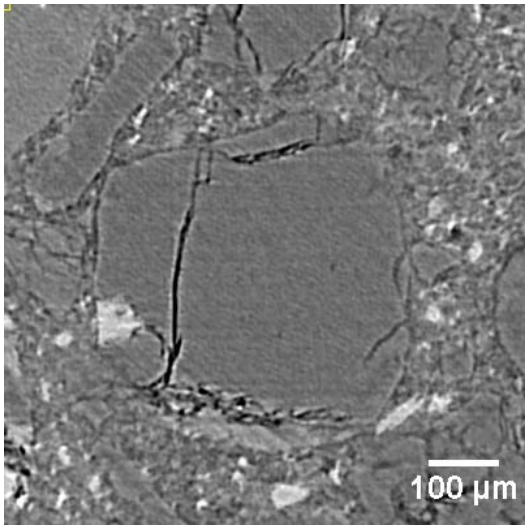
K-edge *Al* EDX micrograph



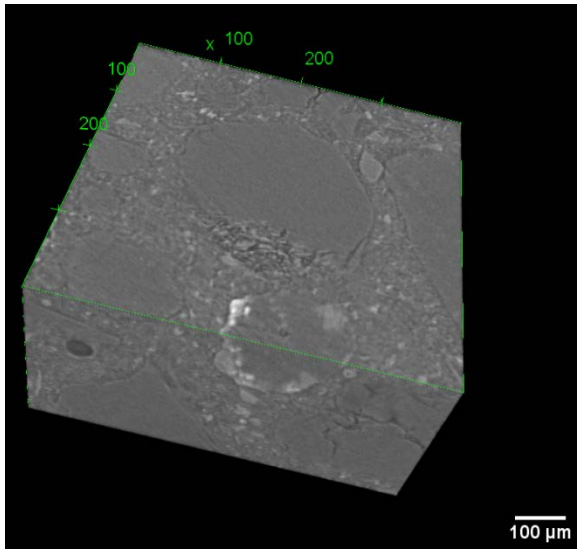
K-edge *Ca* EDX micrograph

{2}: how to think of an image (mathematically)

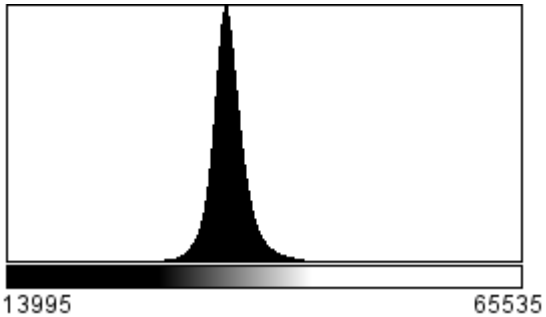
- scalar/vector field ($f(\vec{x}), \vec{f}(\vec{x}) = (f_i(\vec{x})), \vec{x} \in \mathbb{R}^2 \; or \; \mathbb{R}^3$)



slice # 98 of 201



- statistical *ensemble* of realizations of one or more scalar/vector random variables



Outline

- Segmentation of porous (building) materials
- Overview of some methods
  - Global thresholding algorithms
  - Data clustering algorithms
- Conclusions
- Suggested bibliography

Global thresholding

- different objects = different material phases
- **Hp**.: different material phases ⇔ different pixel/voxel value ranges (classes)
- select *n* thresholds based upon the pixel/voxel value histogram

Global thresholding

- different objects = different material phases
- **Hp**.: different material phases ⇔ different pixel/voxel value ranges (classes)
- select *n* thresholds based upon the pixel/voxel value histogram
  - manual selection is useful for qualitative analysis in 2D but is not reliable
  - several categories of automatic selection algorithms based upon
    - the geometrical features of the histogram (peaks, valleys and curvature)
    - probabilistic analysis of the image
      - ✓ *a-priori* models of each class' probability distribution (statistical inference)
      - ✓ maximizazion of entropies of the random variable (the pixel/voxel value)
      - ✓ analysis of the moments of the random variable

Global thresholding

EMPA
Materials Science & Technology

- different objects = different material phases
- **Hp**.: different material phases ⇔ different pixel/voxel value ranges (classes)
- select *n* thresholds based upon the pixel/voxel value histogram
  - ➢ manual selection is useful for qualitative analysis in 2D but is not reliable
  - ➢ several categories of automatic selection algorithms based upon
    - o the geometrical features of the histogram (peaks, valleys and curvature)
    - o probabilistic analysis of the image
      - ✓ *a-priori* models of each class' probability distribution (statistical inference)
      - ✓ maximizazion of entropies of the random variable (the pixel/voxel value)
      - ✓ analysis of the moments of the random variable

Global thresholding: Otsu's algorithm

EMPA
Materials Science & Technology

- ✓ analysis of the moments of the random variable
- ✓ find the pixel/voxel value that minimizes the ***intra-class* variances**, i.e., that maximizes the ***inter-class* variance**
- ✓ for *n* = 1 (background = pore space, foreground = solid phases):

$$\boldsymbol{t_o \equiv arg\,max_{v_j}\left\{P_{Backg}(v_j)\cdot\left(\mu_{Backg}(v_j)-\mu_{Tot}\right)^2 + P_{Foreg}(v_j)\cdot\left(\mu_{Foreg}(v_j)-\mu_{Tot}\right)^2\right\}}$$

$P_{Backg}(v_j) \equiv$ probability that a pixel/voxel belongs to the background class $= \sum_{i=1}^{j} P(v_i) = \frac{1}{N}\cdot\sum_{i=1}^{j} n_i$

$P_{Foreg}(v_j) \equiv$ probability that a pixel/voxel belongs to the foreground class $= 1 - P_{Backg}(v_j)$

$\mu_{Backg}(v_j) \equiv$ average pixel/voxel value within the background class $= \sum_{i=1}^{j} v_i \cdot P(v_i | Backg)$

$\mu_{Foreg}(v_j) \equiv$ average pixel/voxel value within the foreground class $= \sum_{i=j}^{M} v_i \cdot P(v_i | Foreg)$

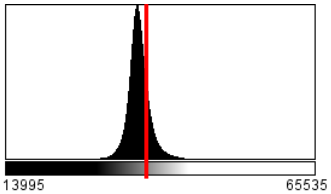
$\mu_{Tot} \equiv$ average pixel/voxel value within the whole image $= \sum_{i=i}^{M} v_i \cdot P(v_i)$

Global thresholding

Shortcomings:

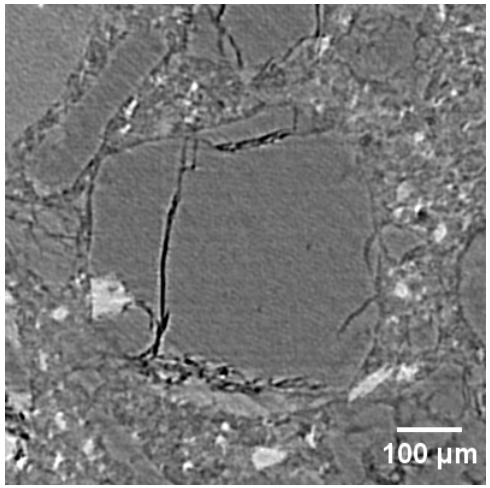
- images with low contrast or large noise levels ⇒ low performance

- unimodal pixel/voxel value histograms ⇒ low performance

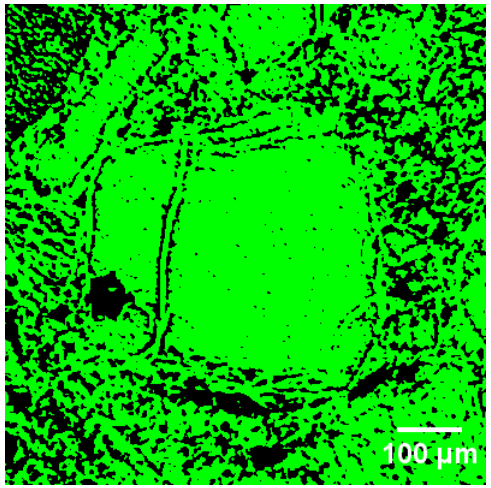


segmentation by the 3D Otsu algorithm

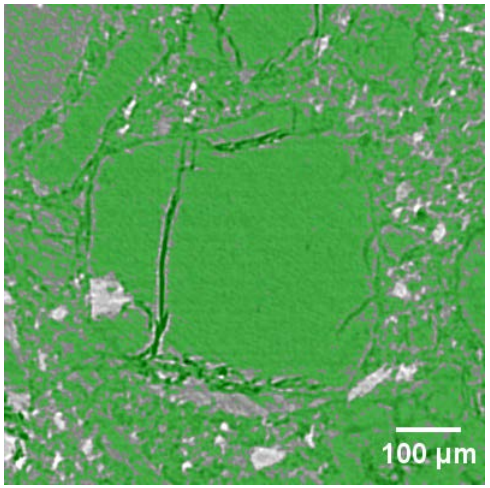
green pixels = porosity
black pixels = solid phases



slice #98, original volume



slice #98, segmented porosity

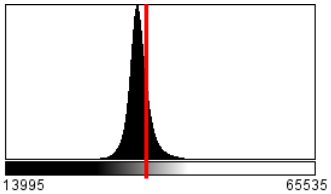


slice #98, overlay

Global thresholding

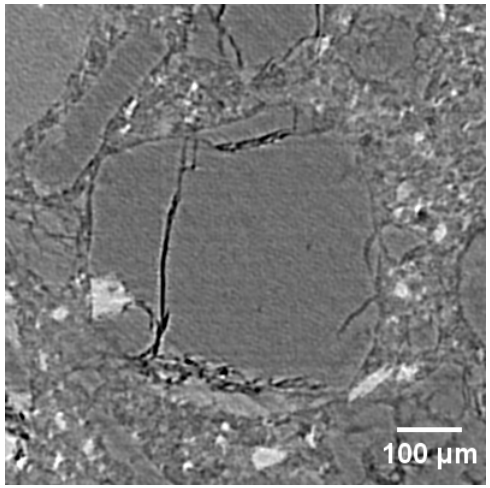
Advantages:

- computationally not very demanding
- used as the 1st of many steps in segmentation
  - ➢ help in choosing where to put the seed points in **seeded region growing** (**SRG**)-based segmentation

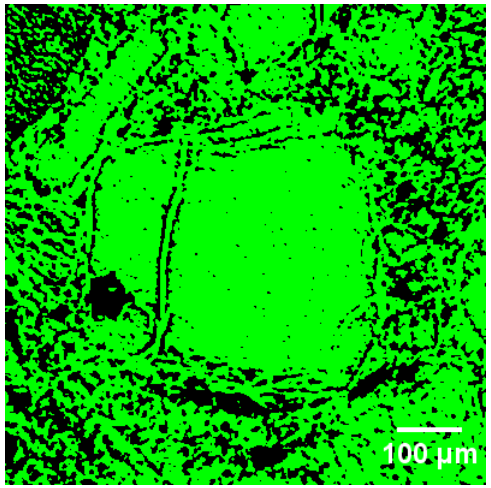


segmentation by the 3D Otsu algorithm

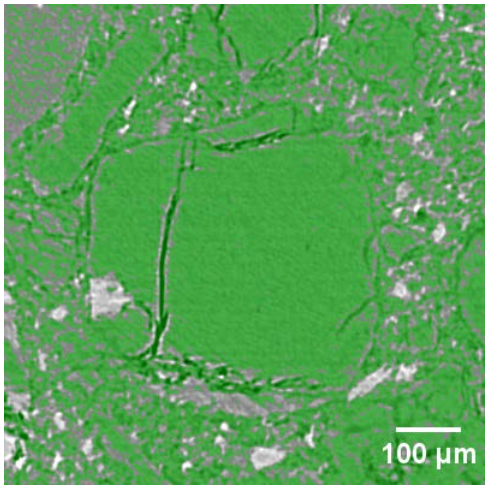
green pixels = porosity
black pixels = solid phases



slice #98, original volume



slice #98, segmented porosity



slice #98, overlay

Outline

- Segmentation of porous (building) materials
- Overview of some methods
  - Global thresholding algorithms
  - Data clustering algorithms
- Conclusions
- Suggested bibliography

Image segmentation by clustering

- dataset = set of points $\vec{x}_i, i = 1, \ldots, M$, in a Euclidean vector space R^n (the **feature space**)

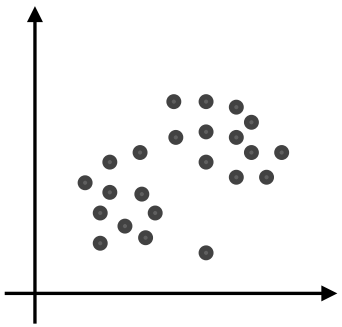


Image segmentation by clustering

- dataset = set of points $\vec{x}_i, i = 1, \ldots, M$, in a Euclidean vector space R^n (the **feature space**)
  - image as an ensamble of *M* points $\vec{x}_i$ (pixels or voxels), where *n* = 1 (grey value image) or *n* = 3 (RGB image) or whatever (not only the pixel/voxel value as feature)

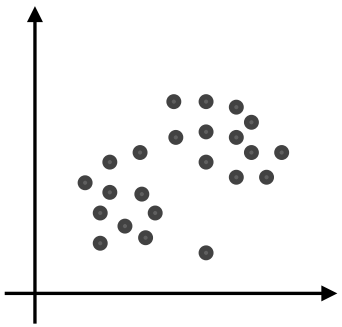
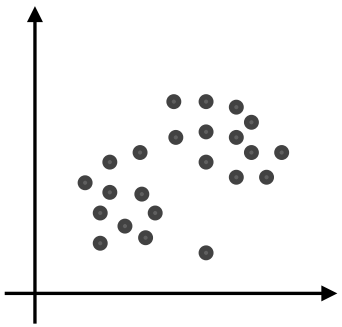


Image segmentation by clustering

- dataset = set of points $\vec{x}_i, i = 1, \ldots, M$, in a Euclidean vector space R^n (the **feature space**)
  - image as an ensamble of *M* points $\vec{x}_i$ (pixels or voxels), where *n* = 1 (grey value image) or *n* = 3 (RGB image) or whatever (not only the pixel/voxel value as feature)



- **clustering** ⇔ partitioning the set of points in *K* subsets (clusters), finding *K* **reference points**, $\vec{y}_l, l = 1, \ldots K,$ representatives of the subsets

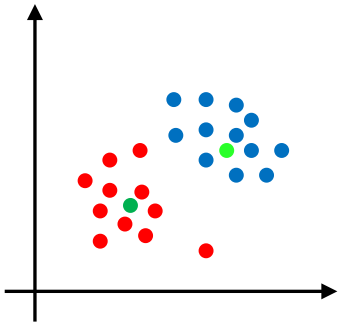
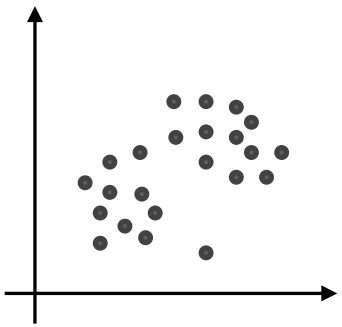


Image segmentation by clustering

- dataset = set of points $\vec{x}_i, i = 1, \dots, M$, in a Euclidean vector space R^n (the **feature space**)
  - image as an ensamble of *M* points $\vec{x}_i$ (pixels or voxels), where *n* = 1 (grey value image) or *n* = 3 (RGB image) or whatever (not only the pixel/voxel value as feature)



- **clustering** ⇔ partitioning the set of points in *K* subsets (clusters), finding *K* **reference points**, $\vec{y}_l, l = 1, \dots K,$ representatives of the subsets
  - each point $\vec{x}_{j,l}$,$j = 1, \dots, M_l$, belonging to the *l*-th cluster, is closer to the *l*-th reference point than to any other reference point, according to a certain distance metric function

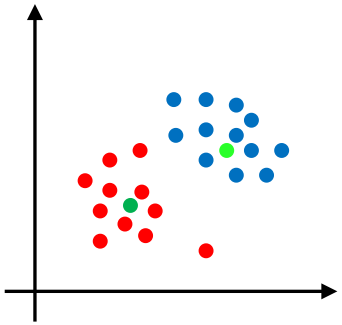
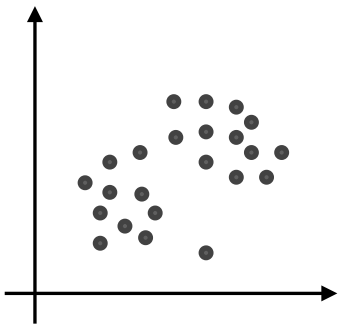


Image segmentation by clustering

- dataset = set of points $\vec{x}_i, i = 1, \dots, M$, in a Euclidean vector space R^n (the **feature space**)
  - image as an ensamble of *M* points $\vec{x}_i$ (pixels or voxels), where *n* = 1 (grey value image) or *n* = 3 (RGB image) or whatever (not only the pixel/voxel value as feature)



- **clustering** ⇔ partitioning the set of points in *K* subsets (clusters), finding *K* **reference points**, $\vec{y}_l, l = 1, \dots K,$ representatives of the subsets
  - each point $\vec{x}_{j,l}$,$j = 1, \dots, M_l$, belonging to the *l*-th cluster, is closer to the *l*-th reference point than to any other reference point, according to a certain distance metric function
  - starting from initial guesses, iteratively find *K* reference points $\vec{y}_l, l = 1, \dots K$, and corresponding clusters, that decrease a global (i.e., for all the clusters) distance metric function till finding its minimum

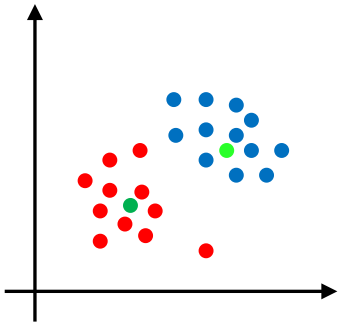
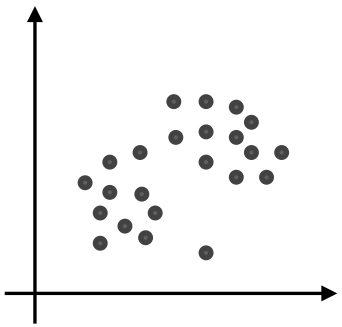


Image segmentation by clustering

- dataset = set of points $\vec{x}_i, i = 1, \dots, M$, in a Euclidean vector space R^n (the **feature space**)
  - image as an ensamble of *M* points $\vec{x}_i$ (pixels or voxels), where *n* = 1 (grey value image) or *n* = 3 (RGB image) or whatever (not only the pixel/voxel value as feature)



- **clustering** ⇔ partitioning the set of points in *K* subsets (clusters), finding *K* **reference points**, $\vec{y}_l, l = 1, \dots K,$ representatives of the subsets
  - each point $\vec{x}_{j,l}$,$j = 1, \dots, M_l$, belonging to the *l*-th cluster, is closer to the *l*-th reference point than to any other reference point, according to a certain distance metric function
  - starting from initial guesses, iteratively find *K* reference points $\vec{y}_l, l = 1, \dots K$, and corresponding clusters, that decrease a global (i.e., for all the clusters) distance metric function till finding its minimum
  - segmented image ⇔ pixel/voxels belonging to the *l*-th cluster are assigned the value of the respective *l*-th reference point

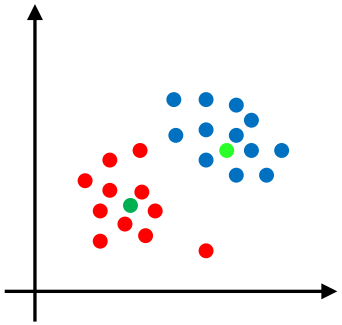


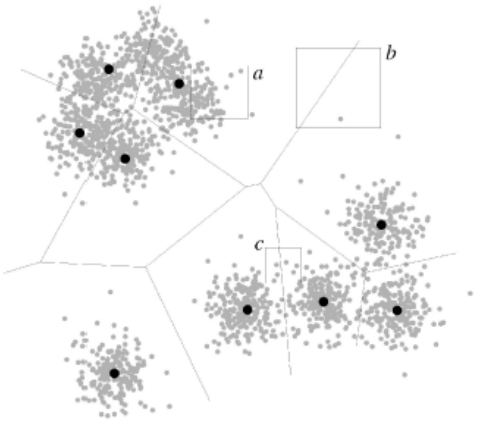
Image segmentation by *K*-means clustering

- ***K*-means clustering**: a class of data clustering algorithms

Image segmentation by *K*-means clustering

- ***K*-means clustering**: a class of data clustering algorithms

  (1) at each iteration, create the clusters by the **Voronoi partitioning** of the feature space based upon the current reference points



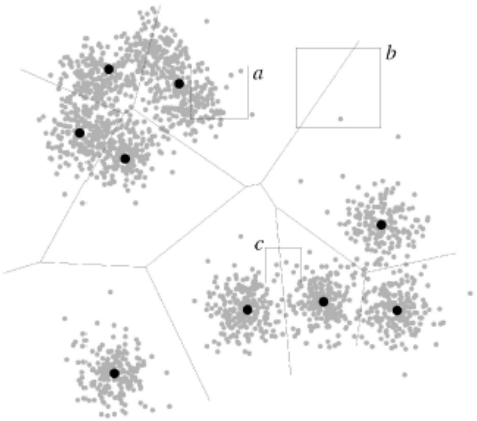
Adapted from Fig. 3 of T. Kanungo et al., IEEE Trans. Patt. Anal. Mach. Intell. 24 (7), 881 – 892 (2002)

Image segmentation by *K*-means clustering

- ***K*-means clustering**: a class of data clustering algorithms

  (1) at each iteration, create the clusters by the **Voronoi partitioning** of the feature space based upon the current reference points

  (2) for each cluster, calculate the **centroid** (mean/average value)



Adapted from Fig. 3 of T. Kanungo et al., IEEE Trans. Patt. Anal. Mach. Intell. 24 (7), 881 – 892 (2002)

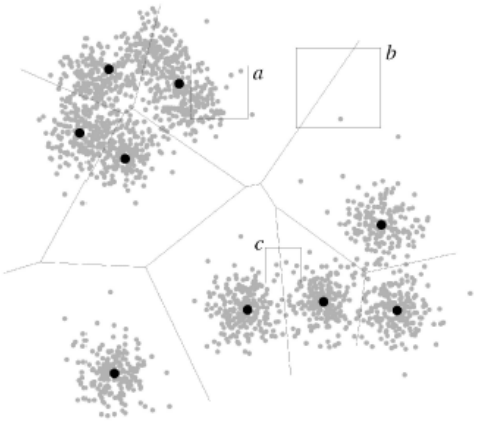
Image segmentation by *K*-means clustering

- ***K*-means clustering**: a class of data clustering algorithms

(1) at each iteration, create the clusters by the **Voronoi partitioning** of the feature space based upon the current reference points

(2) for each cluster, calculate the **centroid** (mean/average value)

(3) move the reference point of each cluster to the centroid; go back to (1)



Adapted from Fig. 3 of T. Kanungo et al., IEEE Trans. Patt. Anal. Mach. Intell. 24 (7), 881 – 892 (2002)

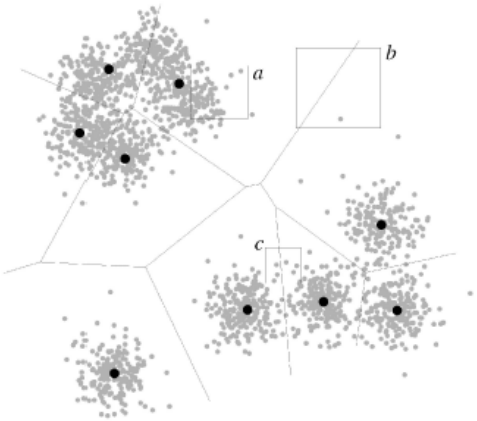
Image segmentation by *K*-means clustering

- ***K*-means clustering**: a class of data clustering algorithms

  (1) at each iteration, create the clusters by the **Voronoi partitioning** of the feature space based upon the current reference points

  (2) for each cluster, calculate the **centroid** (mean/average value)

  (3) move the reference point of each cluster to the centroid; go back to (1)



Adapted from Fig. 3 of T. Kanungo et al., IEEE Trans. Patt. Anal. Mach. Intell. 24 (7), 881 – 892 (2002)

- ***K*-means clustering**: ``the devil is in the details´´

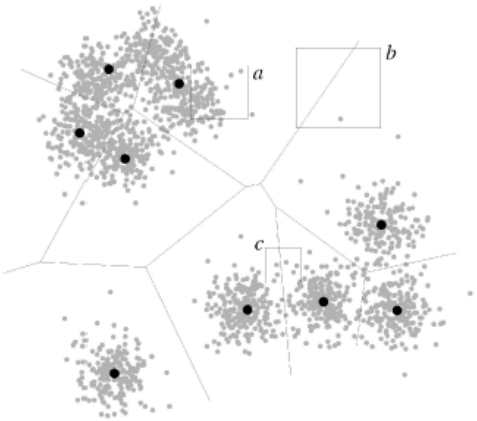
Image segmentation by *K*-means clustering

- ***K*-means clustering**: a class of data clustering algorithms

  (1) at each iteration, create the clusters by the **Voronoi partitioning** of the feature space based upon the current reference points

  (2) for each cluster, calculate the **centroid** (mean/average value)

  (3) move the reference point of each cluster to the centroid; go back to (1)



Adapted from Fig. 3 of T. Kanungo et al., IEEE Trans. Patt. Anal. Mach. Intell. 24 (7), 881 – 892 (2002)

- ***K*-means clustering**: ``the devil is in the details´´

  ➢ different algorithms have different computational efficiency and overload

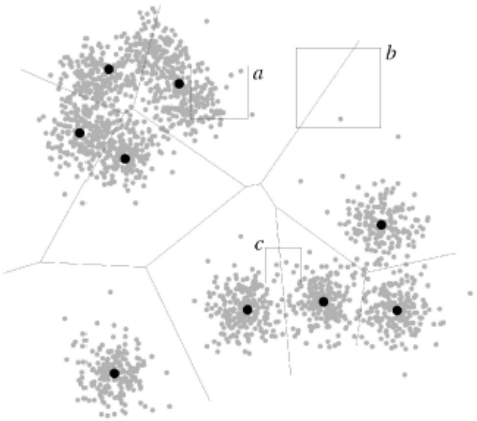
Image segmentation by *K*-means clustering

- ***K*-means clustering**: a class of data clustering algorithms

(1) at each iteration, create the clusters by the **Voronoi partitioning** of the feature space based upon the current reference points

(2) for each cluster, calculate the **centroid** (mean/average value)

(3) move the reference point of each cluster to the centroid; go back to (1)



Adapted from Fig. 3 of T. Kanungo et al., IEEE Trans. Patt. Anal. Mach. Intell. 24 (7), 881 – 892 (2002)

- ***K*-means clustering**: ``the devil is in the details´´

  - different algorithms have different computational efficiency and overload
  - belong to the class of **Centroidal Voronoi Tessellations** algorithms

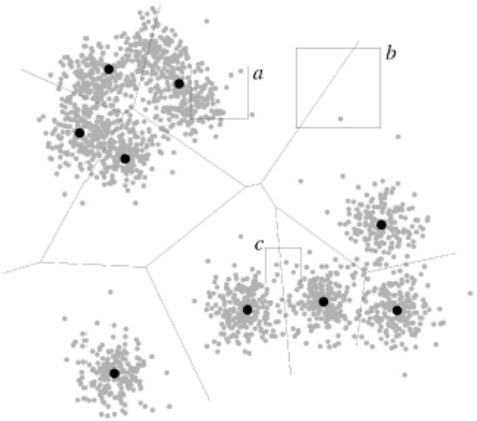
Image segmentation by *K*-means clustering

- ***K*-means clustering**: a class of data clustering algorithms

(1) at each iteration, create the clusters by the **Voronoi partitioning** of the feature space based upon the current reference points

(2) for each cluster, calculate the **centroid** (mean/average value)

(3) move the reference point of each cluster to the centroid; go back to (1)



Adapted from Fig. 3 of T. Kanungo et al., IEEE Trans. Patt. Anal. Mach. Intell. 24 (7), 881 – 892 (2002)

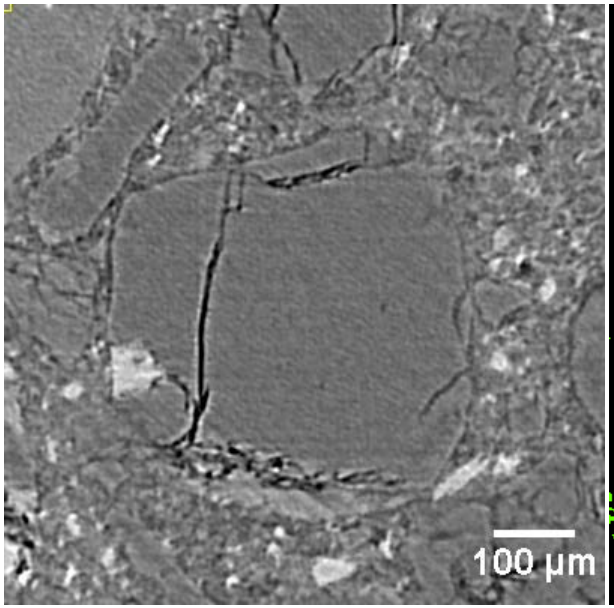
- ***K*-means clustering**: ``the devil is in the details´´

  - different algorithms have different computational efficiency and overload
  - belong to the class of **Centroidal Voronoi Tessellations** algorithms
  - no parameters involved, except for *K* (number of clusters) and total number of iterations

Pore space segmentation by *2*-means clustering

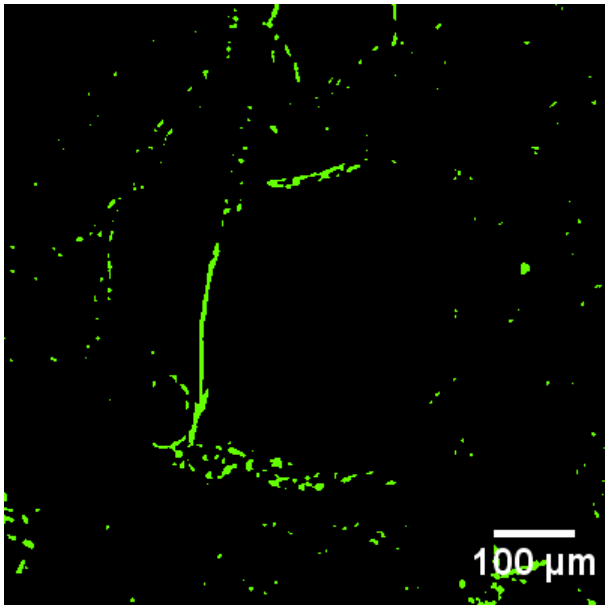
- assumed K=2 clusters (pores and solid phases)
- *2*-means clustering of the entire 3D ROI (volume)
- feature space dimensionality n = 1 (only voxel grey values used)

filtered, original ROI



slice # 98 of 201

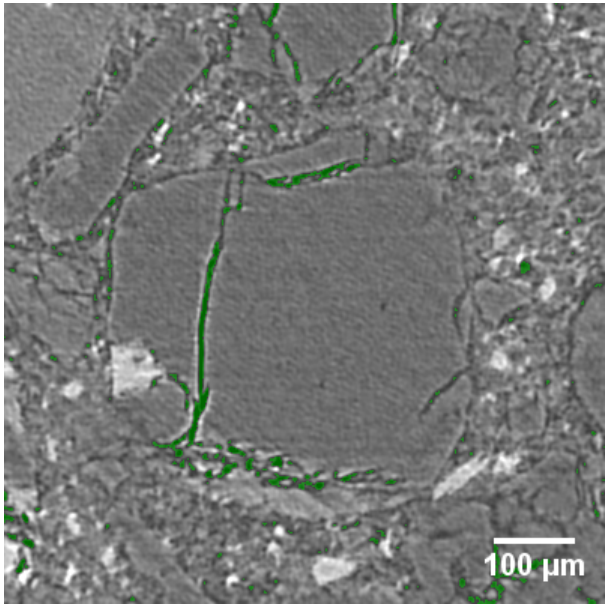
after 2-means clustering (binary image)



slice # 98 of 201

green = pore space voxels
black = solid phase voxels

overlay of the binary image on top of the original



slice # 98 of 201

blue = voxels classified as belonging to the pore space

Outline

- Segmentation of porous (building) materials
- Overview of some methods
  - Global thresholding algorithms
  - Data clustering algorithms
- Conclusions
- Suggested bibliography

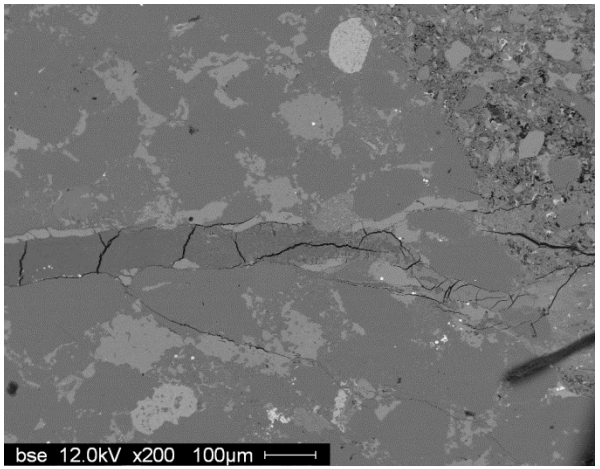
A few conclusions

EMPA
Materials Science & Technology

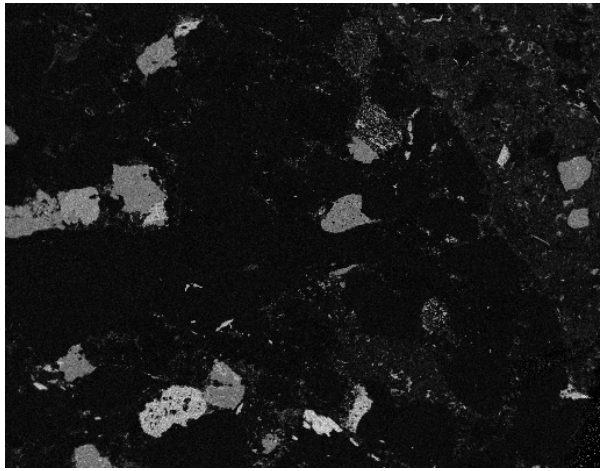
- segmentation by 2-means clustering worked very well on my 3D images
  - ➢ not perfect, but much more efficient than other algorithms, e.g., Otsu-based thresholding
  - ➢ it may not work as well on your images: the best segmentation algorithm type is image-dependent !
- only 1D feature space for the data clustering !

  Better results achievable by choosing additional feature variables for the pore space voxels, e.g., spatial coordinates and/or local morphological properties of the image itself

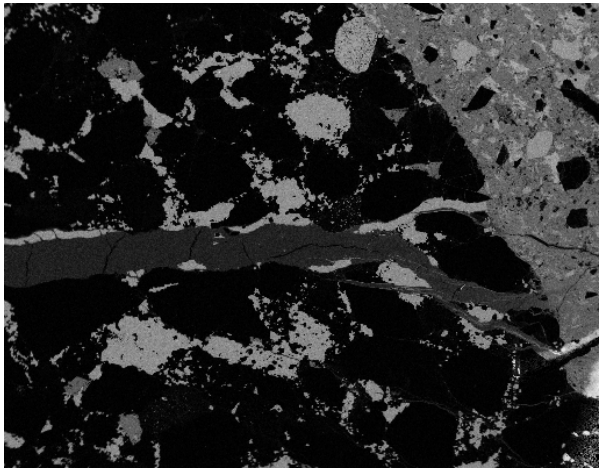
AAR-damage concrete sample



back-scattering SEM micrograph



K-edge *Al* EDX micrograph



K-edge *Ca* EDX micrograph

Outline

- Segmentation of porous (building) materials
- Overview of some methods
  - Global thresholding algorithms
  - Data clustering algorithms
- Conclusions
- Suggested bibliography

Suggested bibliography

EMPA
Materials Science & Technology

General reviews

- A.P. Sheppard, R.M. Sok, H. Averdunk, *Techniques for image enhancement and segmentation of tomographic images of porous materials*, Phys. A 339, 145-151 (2004).
- A. Kaestner, E. Lehmann, M. Stampanoni, *Imaging and image processing in porous media research*, Adv. Wat. Res. 31, 1174-1187 (2008).
- P. Iassonov, T. Gebrenegus, M. Tuller, *Segmentation of X-ray computed tomography images of porous materials: A crucial step for characterization and quantitative analysis of pore structures*, Wat. Res. Res. 45, W09415/1-12 (2009).

Global thresholding

- N. Otsu, *A Threshold Selection Method from Gray-Level Histograms*, IEEE Trans. Sys. Man. Cyb. 9 (1), 62 – 66 (1979)

Data clustering and image segmentation

- V. Farber, *Clustering and the Continuous k-Means Algorithm*, Los Alamos Science 22, 138 – 144 (1994).
- A.K. Jain, M.N. Murty, P.J. Flynn, *Data Clustering: A Review*, ACM Comp. Surveys 31 (3), 264 – 323 (1999).
- Q. Du, V. Farber, M. Gunzburger, *Centroidal Voronoi Tessellations: Applications and Algorithms*, SIAM Rev. 41 (4), 637 – 676 (1999).
- T. Kanugo *et al.*, *An Efficient k-Means Clustering Algorithm: Analysis and Implementation*, IEEE Trans. Patt. Anal. Mach. Intell. 24 (7), 881 – 892 (2002).

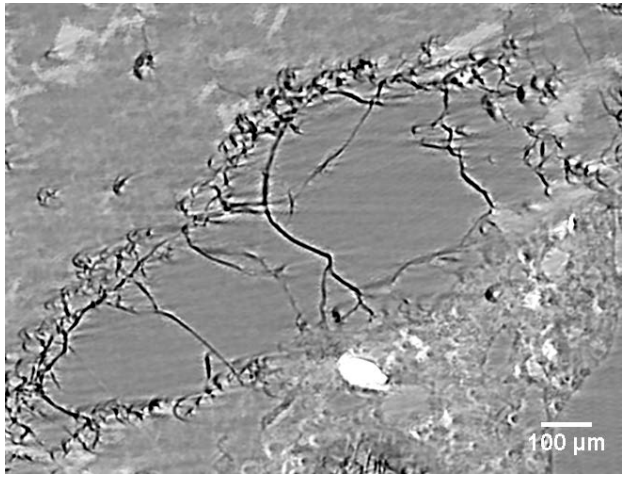
EMPA
Materials Science & Technology

Additional Materials

Pore space segmentation by *2*-means clustering

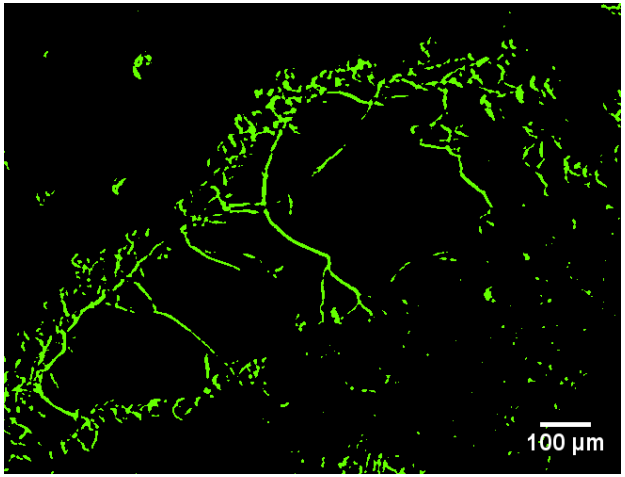
- another 3D ROI, 656 x 500 x 400 voxels
- same processing as for the previous ROI

filtered, original ROI



slice # 179 of 400

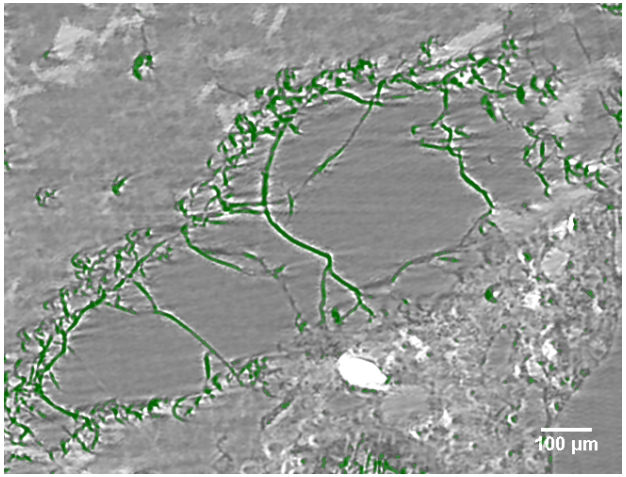
after 2-means clustering (binary image)



slice # 179 of 400

green = pore space voxels
black = solid phase voxels

overlay of the binary image on top of the original



slice # 179 of 400

blue = voxels classified as belonging to the pore space