Classifying vessel shapes

using automated shape extraction and unsupervised classification

Martin Hinz[1], Caroline Heitz[2]

March 2018

1. Institute of Pre- and Protohistoric Archaeology, Kiel University; 2. Institut für Archäologische Wissenschaften, Bern University

motivation

previous approaches

Shape Extraction

Shape Analysis 1: PCA & hclust

Shape Analysis 2: t-sne & hdbscan

Case studies

motivation

shape vs. decoration for typology

levels of pottery design

criteria	shape	decoration
functional dependency	strong	weak
dependency on local ressources	stronger	weak
moment of determination	from the beginning	finished product
ad hoc changes	difficult	easy
more reflecting	habitus	personal signature
design decision	more collective	more individual
homogenity	more convergent	more divergent
degrees of freedom	lesser	more
category of style	more isochrestic	more emblemic
unit of analysis	ideal type	singularity

Phases of ceramic production

Design

- imagined and intended
- concept of a shape
- ideal type
- **social bounded options for creativity**

'style of action'

- learned, habitual movements, manipulations, routines
- at all stages of the *chaîne opératoire*
- instantiation
- **limited options for creativity**

'material style'

- physical traces and effects in the material
- Characteristics of the vessel
- finished object
- **material induced alterations of the original design possible**

resume

using vessel shapes to identify underlying large scaled ‘communities of practice’

- more related to covert communication structures, habitus and similar (economic) practises
- ‘better measurable’ from a practical quantitative perspective

previous approaches

quantitative approaches towards shape-based typology

using specific locations

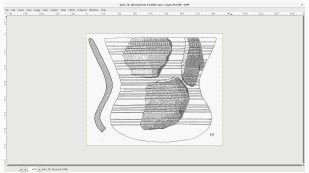
- Koch 1998

'Hollistic' approaches

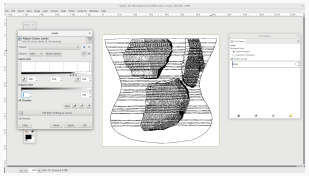
- Mom 2005
- Chapman et al., 2006
- Keogh et al., 2009

Shape Extraction

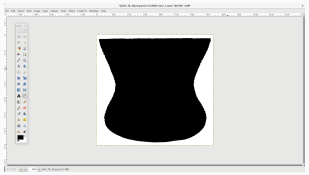
‘by hand’: workflow



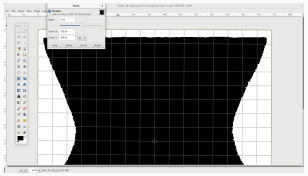
load



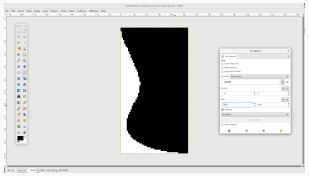
enhance contrast



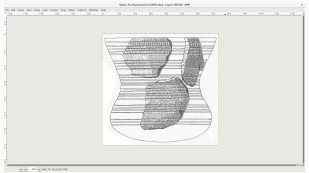
fill inside black



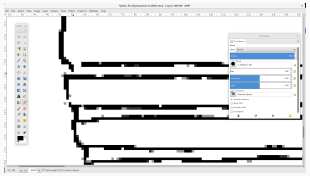
rotate



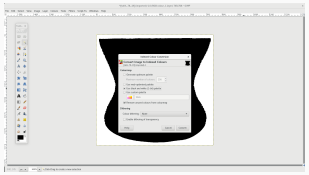
split in half



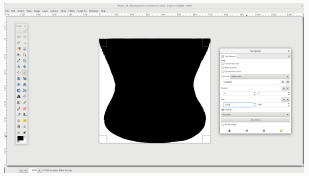
clean



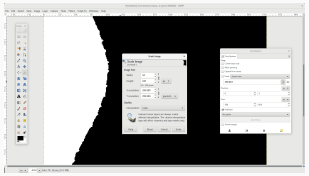
close gaps



turn bw (bitmap)

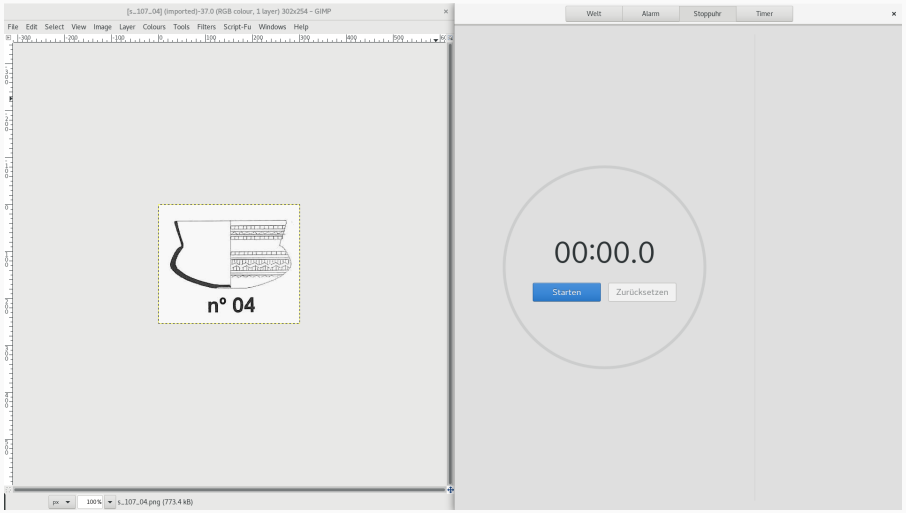


crop

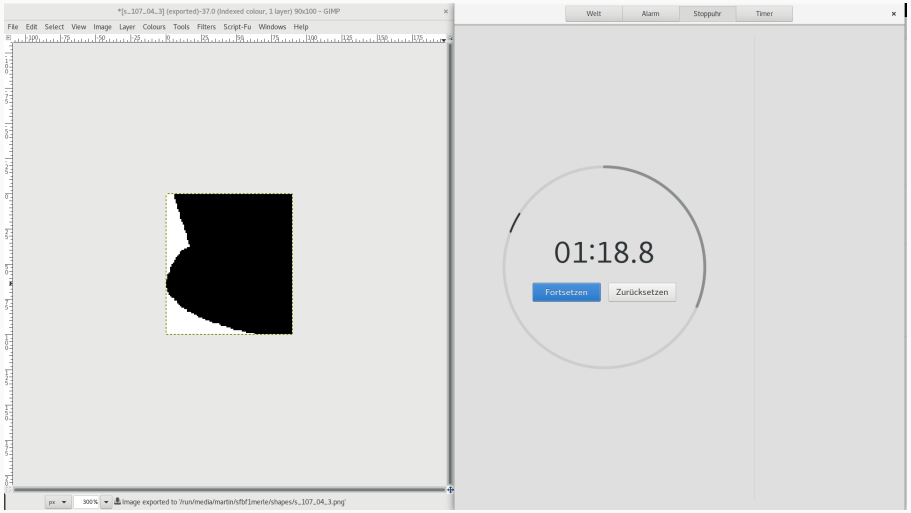


scale to resolution (100px)

‘by hand’: speed



↓



Adaptive Contour Motivation

- 1:30 min for 800 images -> 20 hours
  - automation makes sense!
- scanned vessel drawings not trivial
  - might have holes (dashed lines)
  - might have additional image components beside the vessels
- no simple image segmentation with background color and floodfill possible

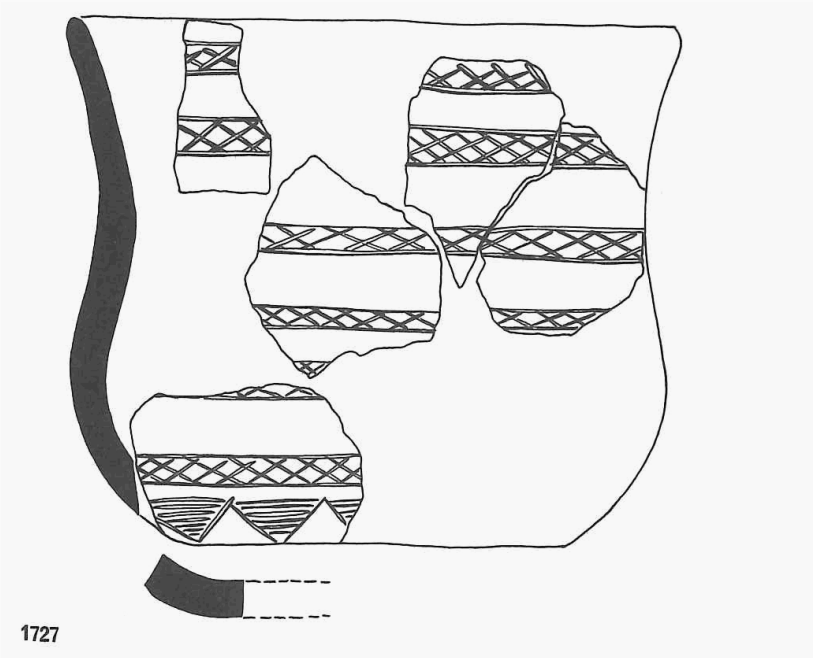


Figure 1: Bell Beaker (Harrison 1977)

Adaptive Contour: what is it

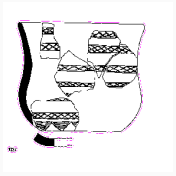
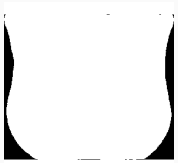
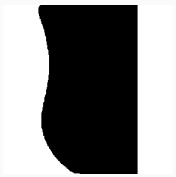
Active contour model, *also called* ***snakes***, *is a framework in computer vision for delineating an object outline from a possibly noisy 2D image. The snakes model is popular in computer vision, and snakes are greatly used in applications like object tracking,* ***shape recognition***, *segmentation, edge detection and stereo matching. Wikipedia*

source: https://github.com/pmneila/morphsnakes

Active contour with Bell Beakers

See shapAAR vignette

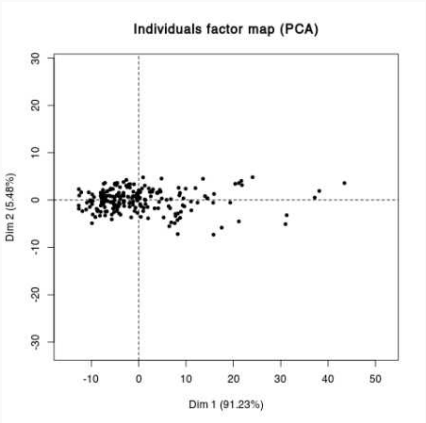
https://github.com/ISAAKiel/shapAAR/blob/master/vignettes/object-extraction.md

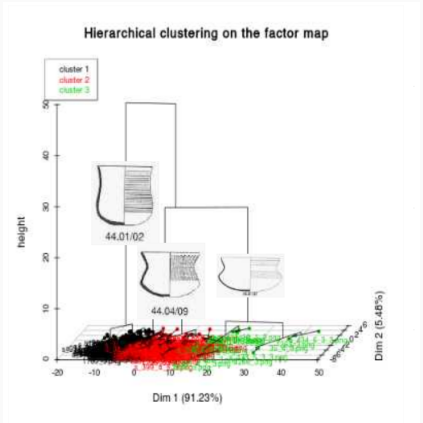
	Image preparation	bw, blur
	Image segmentation	active_contour
	Selecting the biggest object	EBImage
	Rectify and crop	get Bounding box, rotate upright, crop
	half and side mean	split in middle, mean left-right

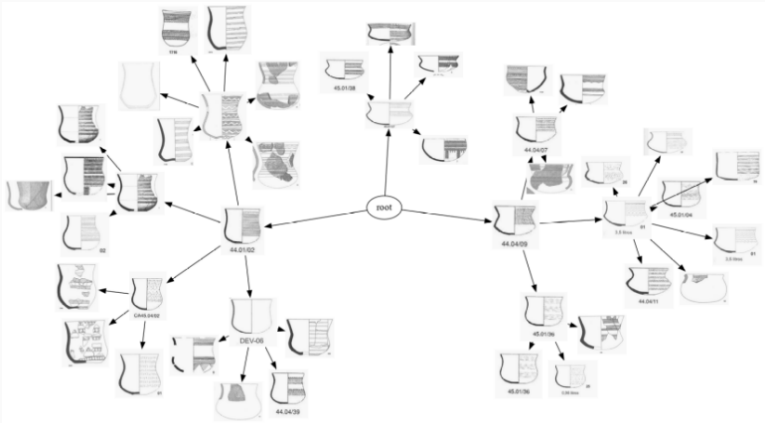
Shape Analysis 1: PCA & hclust

approach/workflow

- using profile distances + additional nominal variables as input
- conduct PCA
- using hclust (euclidean/Ward) to cluster
- package `FactoMinR`: HCPC with automatical cut (number of clusters according to higher relative loss of inertia)
- for individual clusters, repeat with cluster members as new dataset (usually 3 levels)







problems

- non-metric (non-euclidean) variables not considered correctly
- all vessels had to belong to one cluster (hand made ceramics)
- global dissimilarity resulted in suboptimal separation

Shape Analysis 2: t-sne & hdbscan

t-Distributed Stochastic Neighbor Embedding

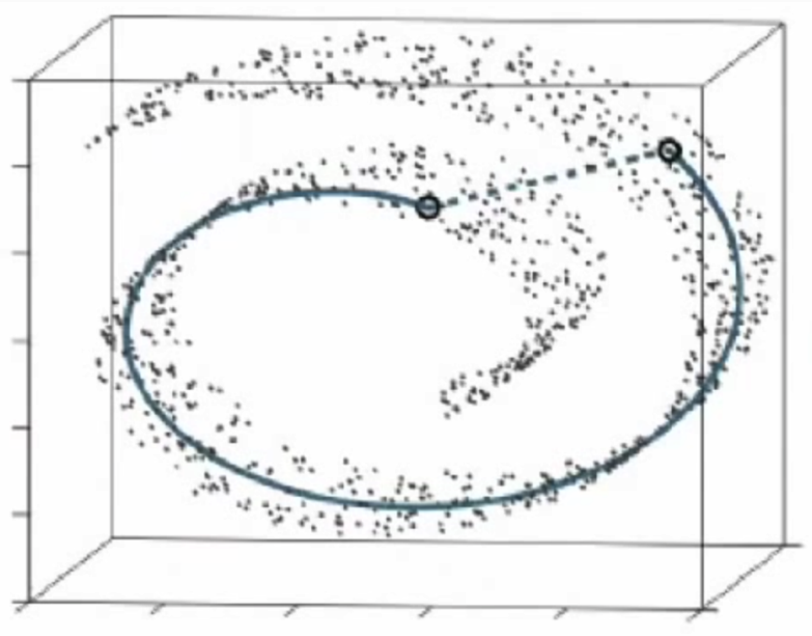
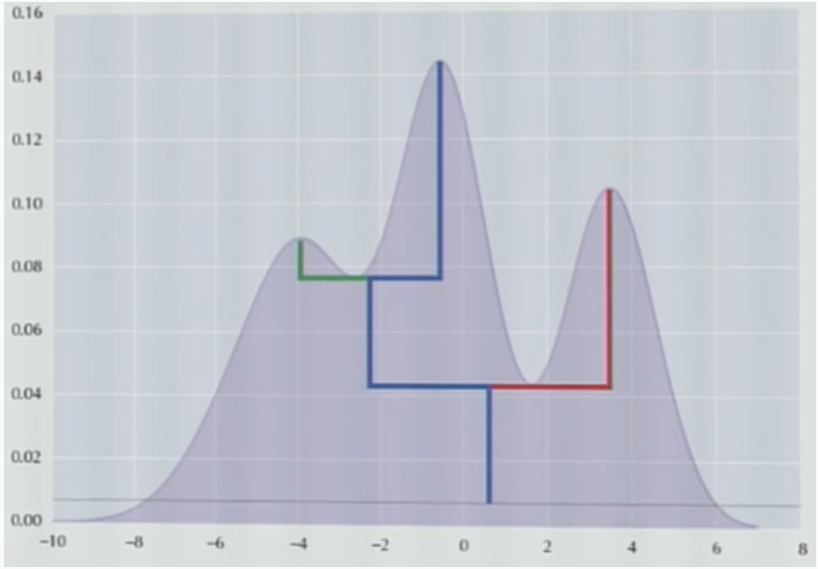


Figure 2: van der Maaten 2008; 2009; 2012; 2014;
https://lvdmaaten.github.io/tsne/
https://www.youtube.com/watch?v=RJVL80Gg3lA

Challenge

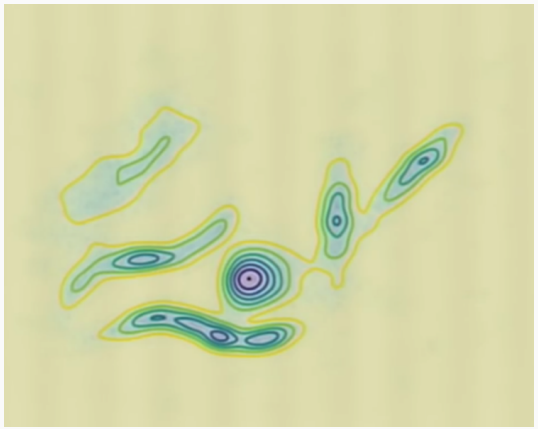
- high dimensional non-linear data distribution
- consider not the global, but the local neighbourhood (contrasting PCA)

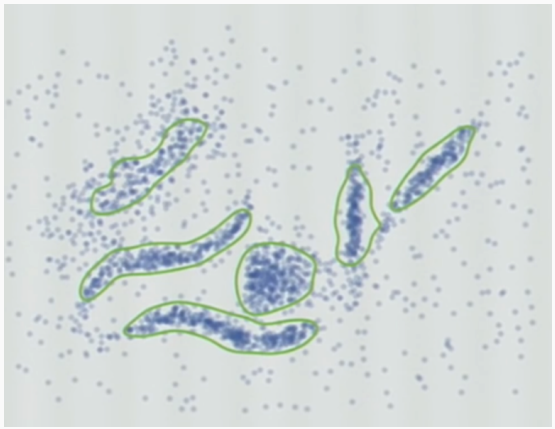
Hierarchical Density-Based Spatial Clustering of Applications with Noise

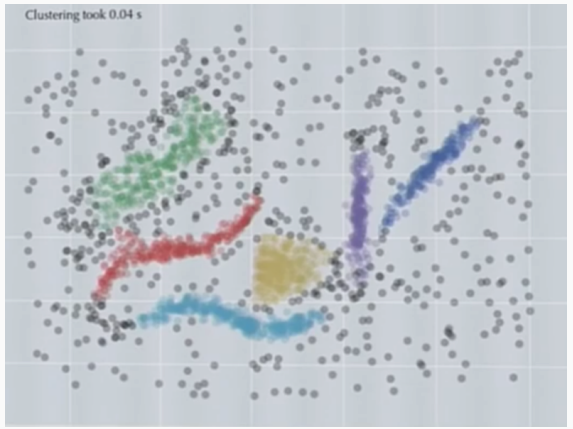


Benefits:

- separating non-circular clusters
- 'identification' and exclusion of noise (hand made ceramics!)

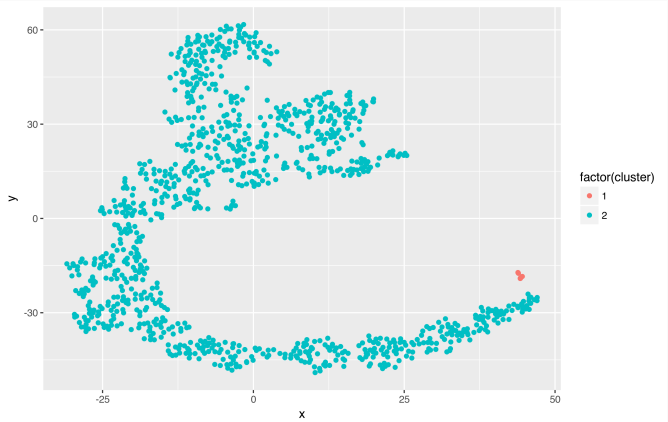


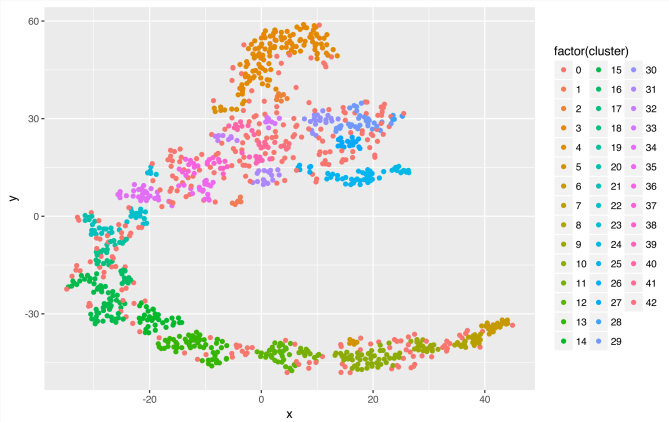




approach/workflow

- using profile distances + additional nominal variables as input
- conduct t-sne
- using hdbscan to cluster
- package `Rtsne`: conduct t-sne
  - `Rtsne(as.matrix(shapes),perplexity=30,dims=2,theta = 0)`
- package `dbscan`: conduct hdbscan on t-sne result
  - `hdbscan(this_res_sne, minPts = 5)`
- for individual clusters, repeat with cluster members as new dataset (usually only 2 levels)





Case studies

Bell Beakers of the Iberian Peninsula

Neolithic Swiss Ceramic

Thank you

 **Martin Hinz**

martin.hinz@ufg.uni-kiel.de

 **Caroline Heitz**

caroline.heitz@iaw.unibe.ch