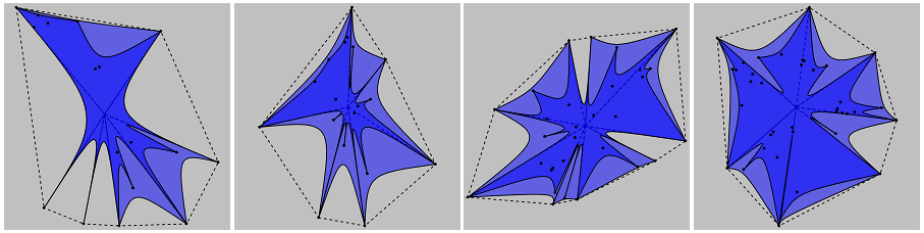
International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision '2008



Butterfly Plots for Visual Analysis of Large Point Cloud Data

Tobias Schreck[1], Michael Schüßler[2], Katja Worm[2], Frank Zeilfelder[1]

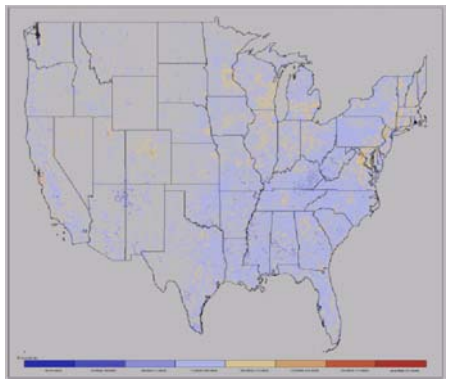
[1] Interactive Graphics Systems Group (GRIS)
Technische Universität Darmstadt, Germany

[2] Siemens ElectroCom Postautomation GmbH
Berlin, Germany

1. Introduction (1)

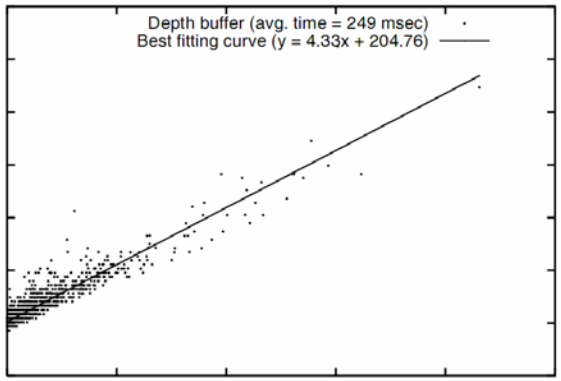
2D Point cloud data: One of the most basic, most used data types

- Geospatial data
- Business data
- Experimental data
- High-dimensional data (projection)

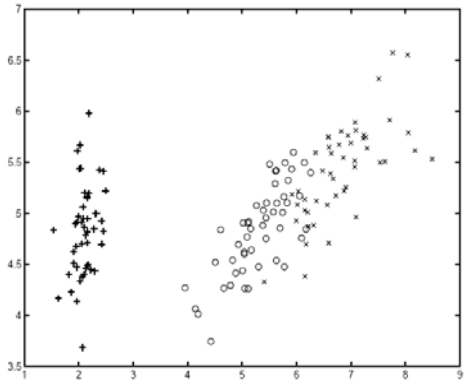


Geospatial data

(KPSS04)



Experimental scatter plot data



Projection-based data analysis

(DMS02)

1. Introduction (2)

Non-trivial point cloud data sets

- Large point sets
- Complex spatial distribution
- Class membership (e.g., given in classifier analysis)

Visual analysis tasks

- Class distribution properties (spatial patterns, frequencies)
- Spatial relationship between classes of points
- Application-dependent notions: separability, overlap, etc.

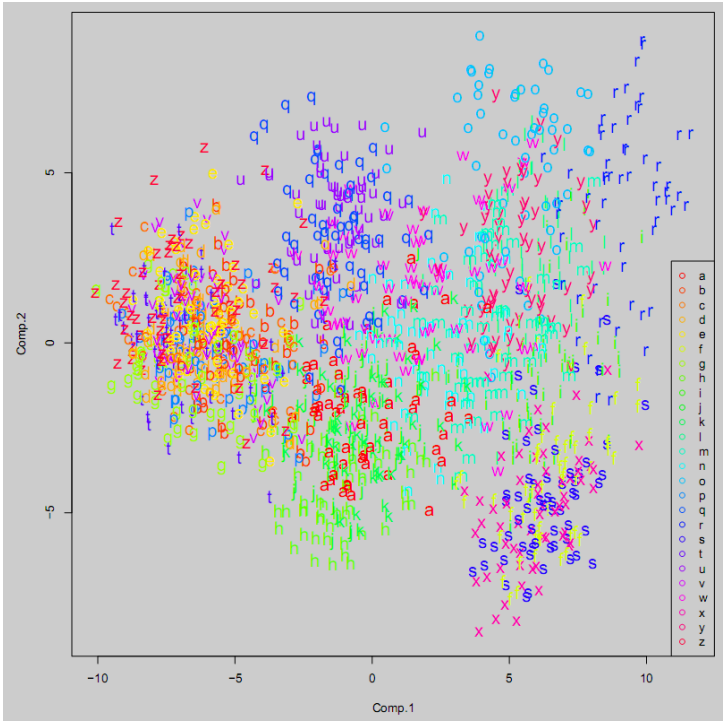
Visual variables for point cloud visualization

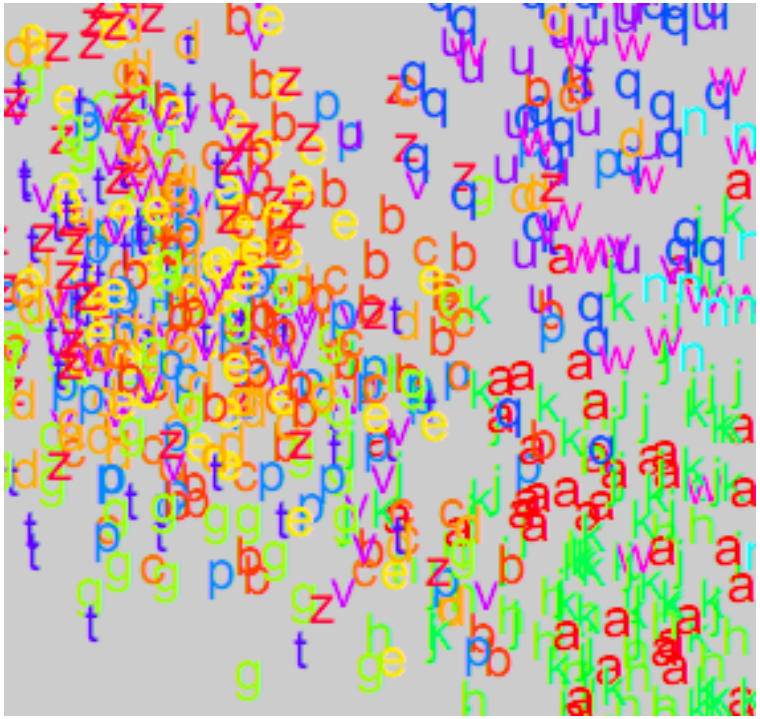
- Position
- Shape (label) and color
- Other

1. Introduction (3)

Scalability problem of standard approach

- Perceptual limitations
- Difficult to form mental model
- What does overlap mean for point clouds?





2. Related work (1)

Abstract points by shape

- Bivariate boxplot [RRT99]
  - Statistically sound notion of centralness
- Convex hulls [SP07]
  - Simple; sensitive to outliers

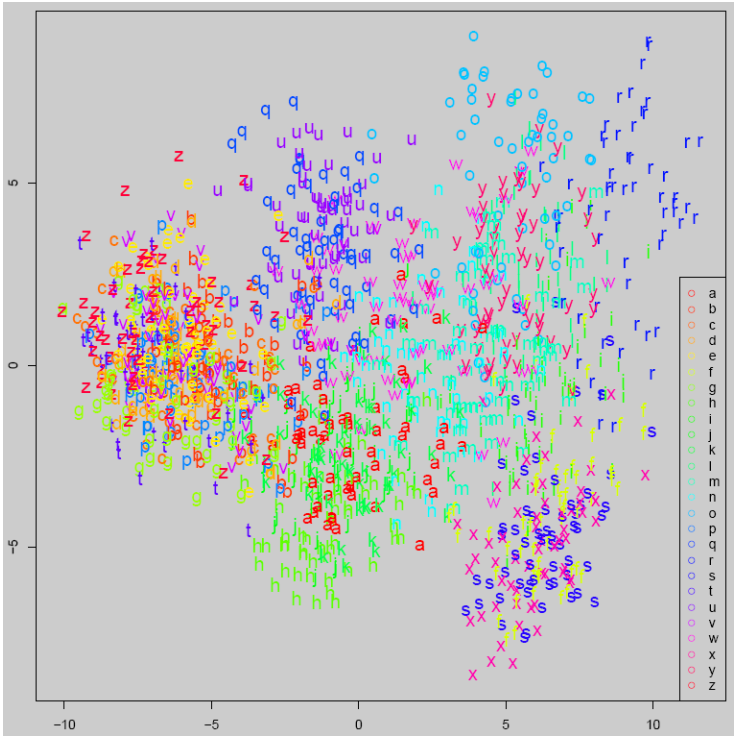
Introduction of distortion

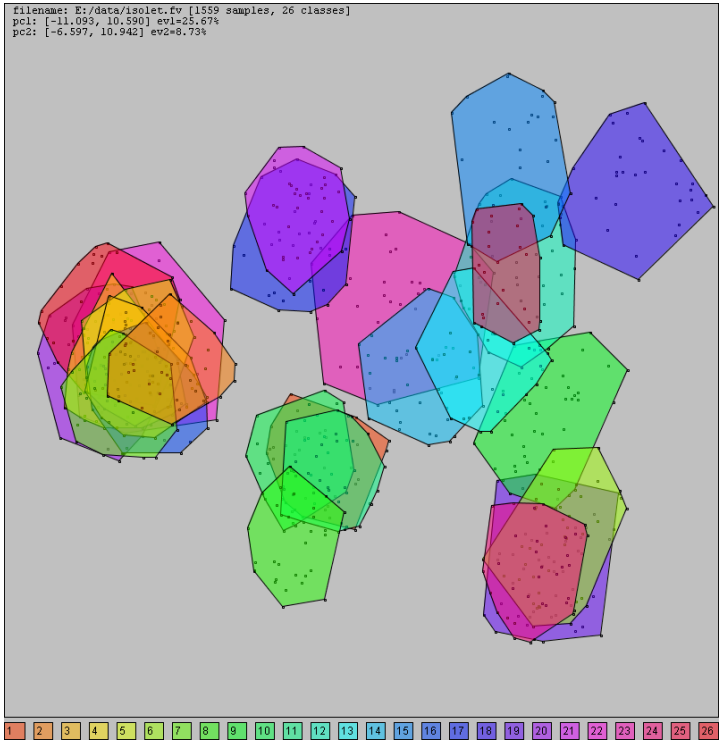
- Pixel placement [KPSS04]
  - Rearrangement of pixels; minimizing distortion (geospatial context)
- Geometric distortion [KNP05]
  - Mesh-based distortion of display area

2. Related work (2)

Convex hull-based shape formation [SP07**]**

- Outlier removal
- Rainbow colormap





3. Design (1)

Considerations for shape construction

- Enclose all points, no outlier removal
- Shape compactness
- Shape complexity
- User view: aim at predictability

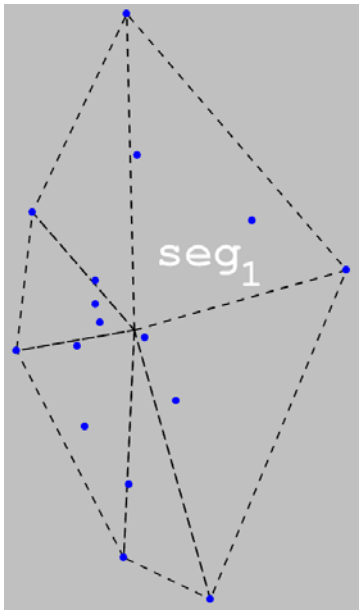
Starting point

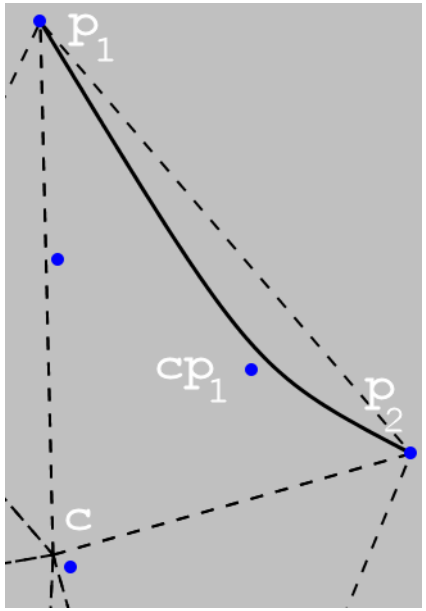
- Convex hull
- Refinement; adapt to the given point distribution

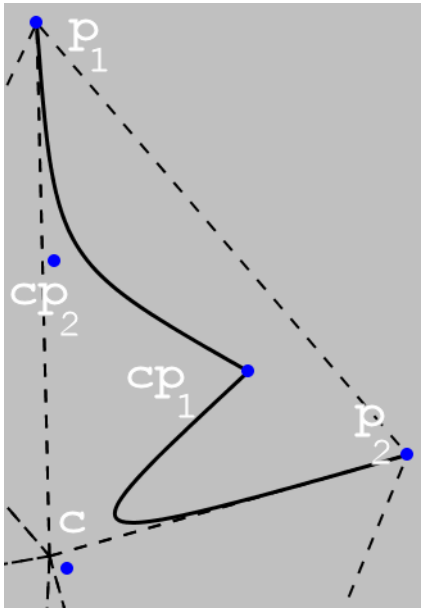
3. Design (3)

Approach: refine convex hull boundary lines by curves controlled by interior points

- Smooth boundary: splines
- Recursive refinement
- Controlled by nearest neighbor points (eventually: centroid)

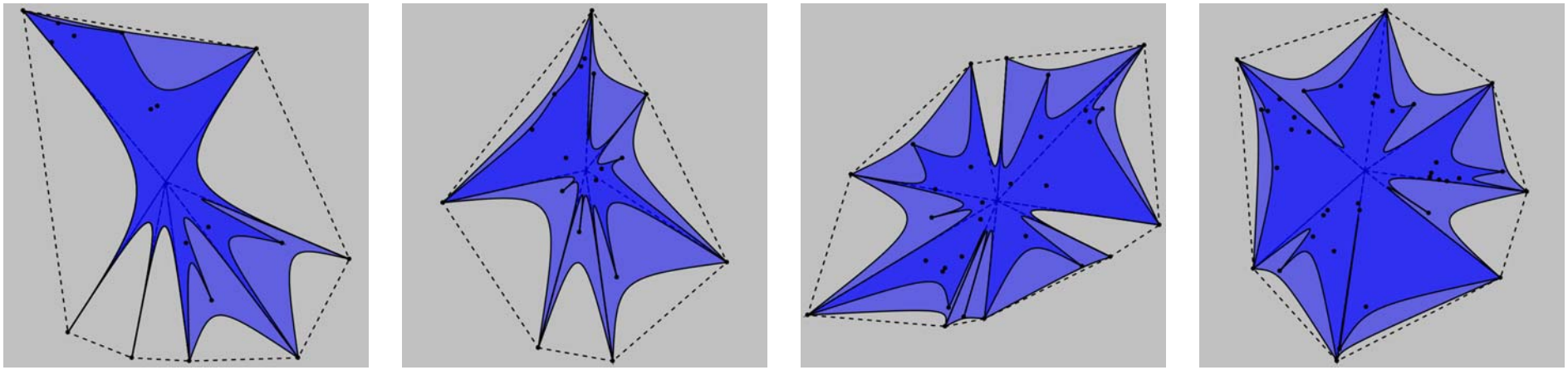






3. Design (4)

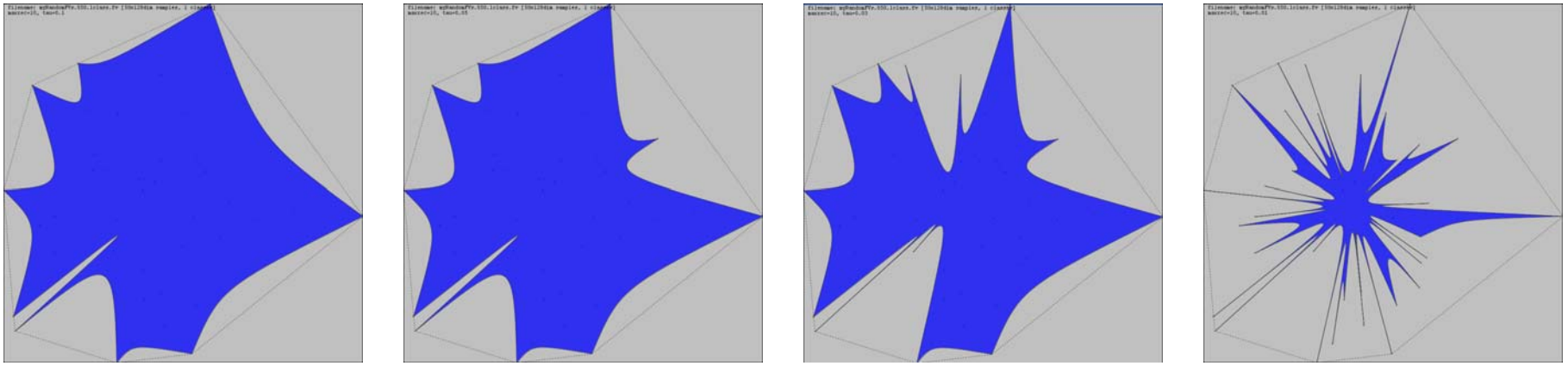
Main parameter: recursion depth



- Smooth boundary; edges are known to locate data points
- Deeper recursion depth decreases area but increases complexity
- Area reduction achieved is data dependent
  here: 20%-55% (1x refinement) | 46%-74% (2x refinement)

3. Design (5)

Variation: adaptive refinement

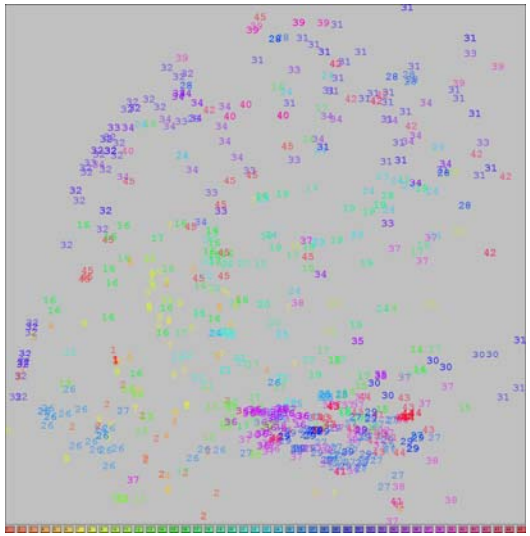
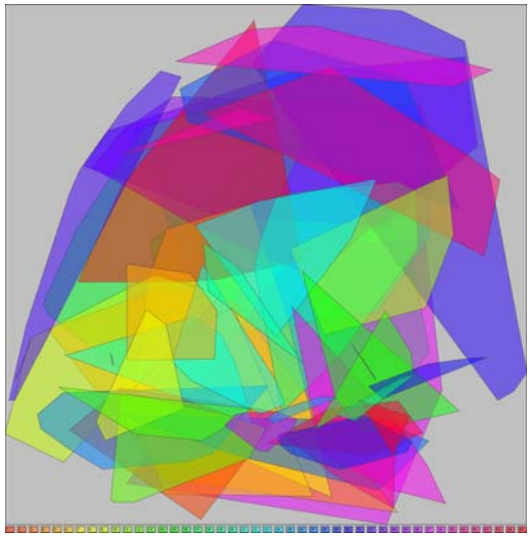
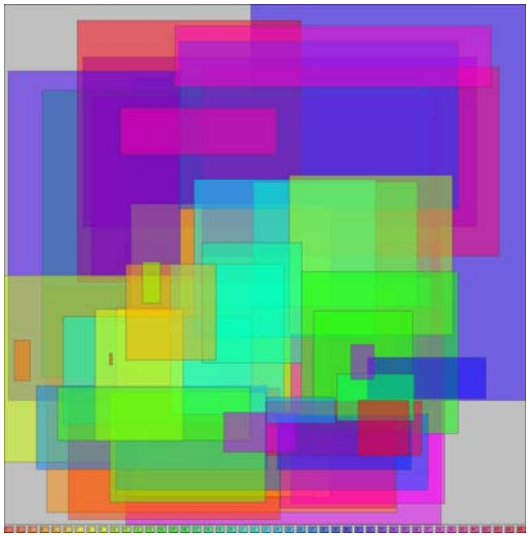


t=1.00 *t*=0.05 *t*=0.03 *t*=0.01

- Always do first level refinement (lines→curves)
- Execute recursion only if at least *t%* of shape area are recovered
- *t* implements the tradeoff between compactness and copplexity

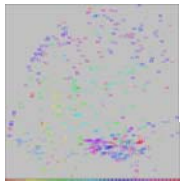
4. Application (1)

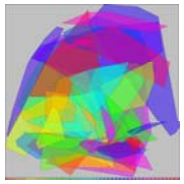
1. **Comparison with standard hull-based visualization**
   - ESB data set with 1000 points in 45 classes
   - CPX features; PCA-based projection
   - Visualization: point cloud | convex hull | MBR | Butterfly

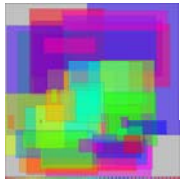




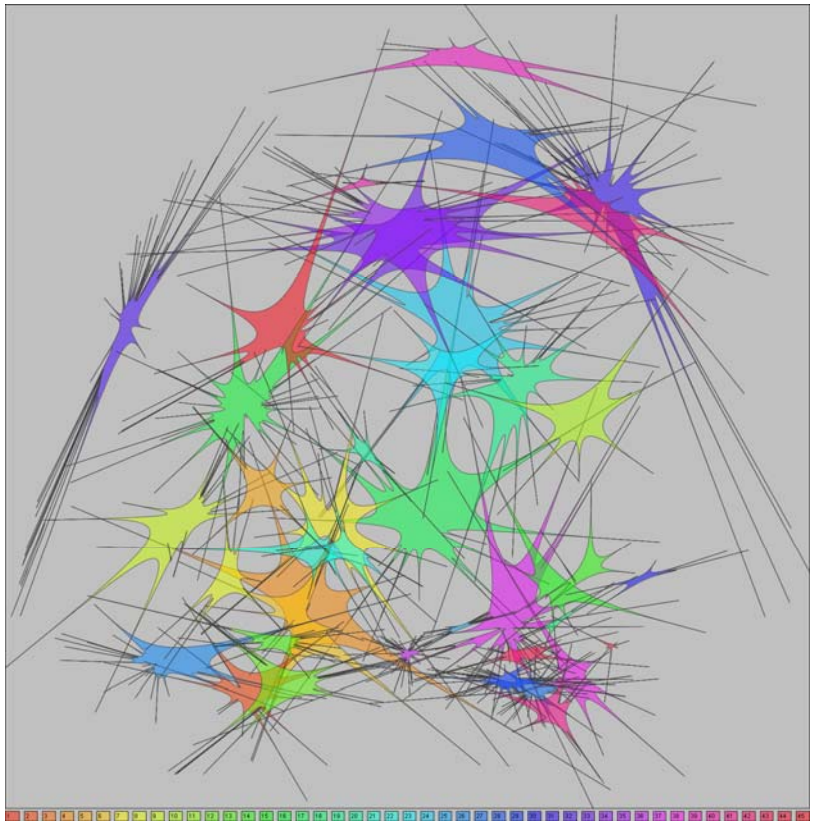
4. Application (1)

1. **Comparison with standard hull-based visualization**
   - ESB data set with 1000 points in 45 classes
   - CPX features; PCA-based projection





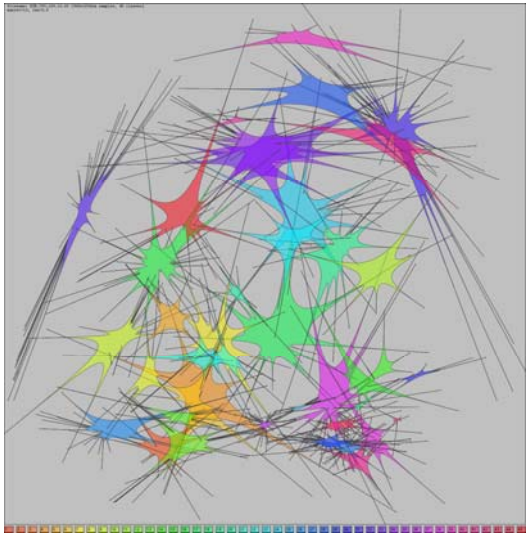
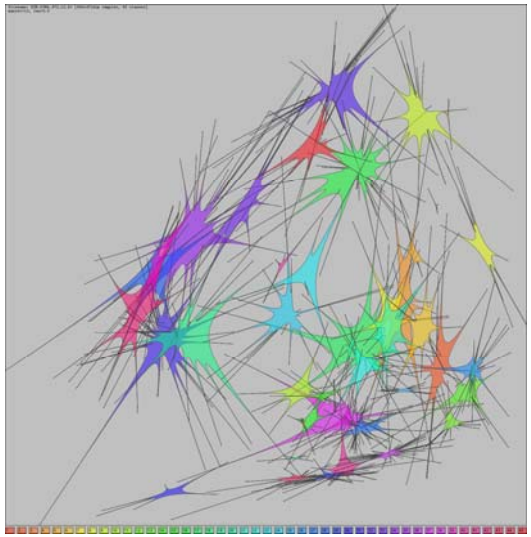
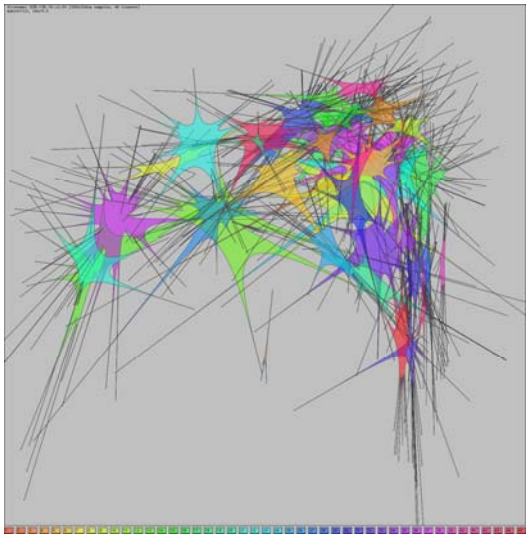




4. Application (2)

2. Projection comparison application

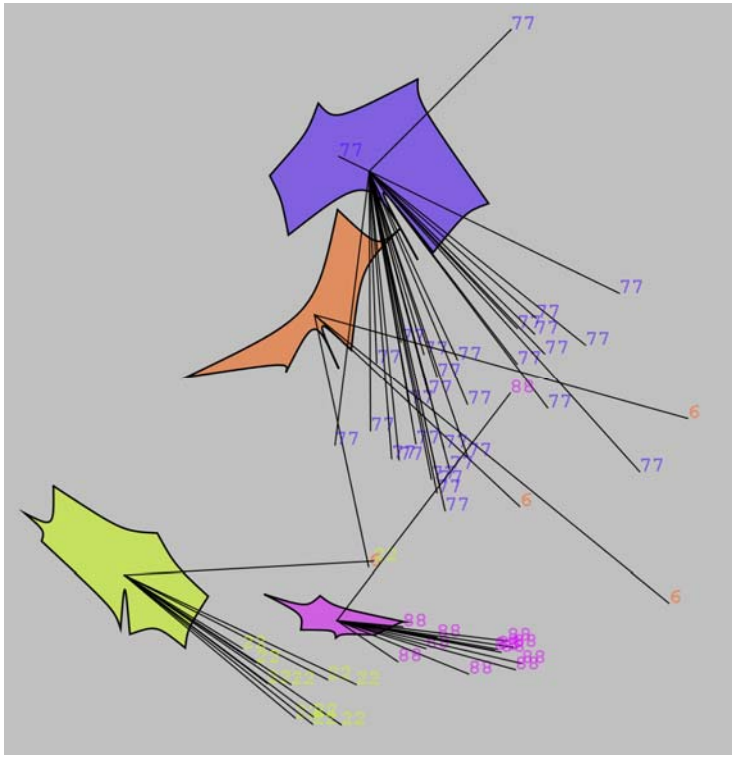
- ESB data set
- PCA-based projection of 3 competing feature representation
- Use case: Visual feature benchmarking

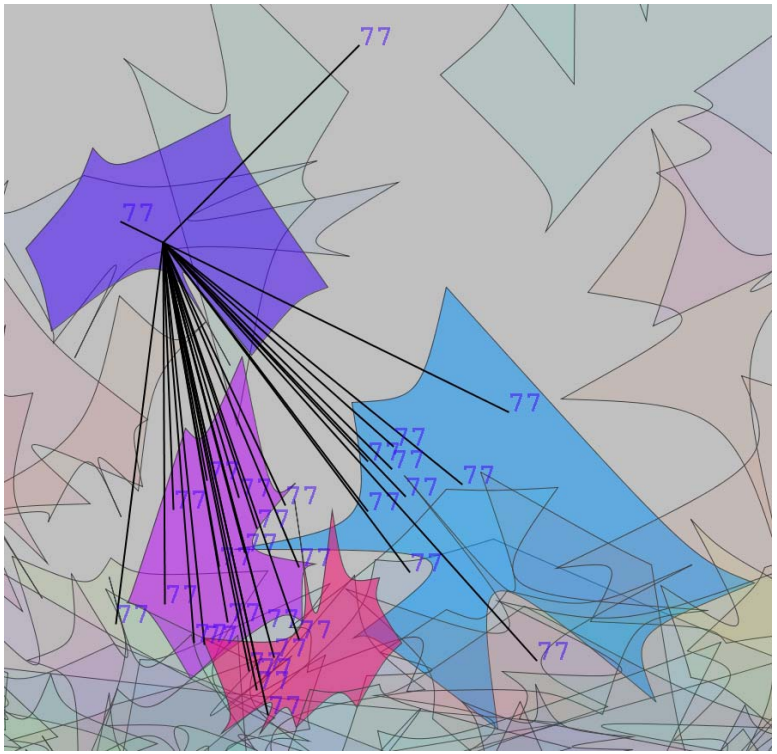




4. Application (3)

3. Missclassification application

- Problem: recognition of stamp images in postal sorting application
- Data set: 109 training classes; missclassified samples

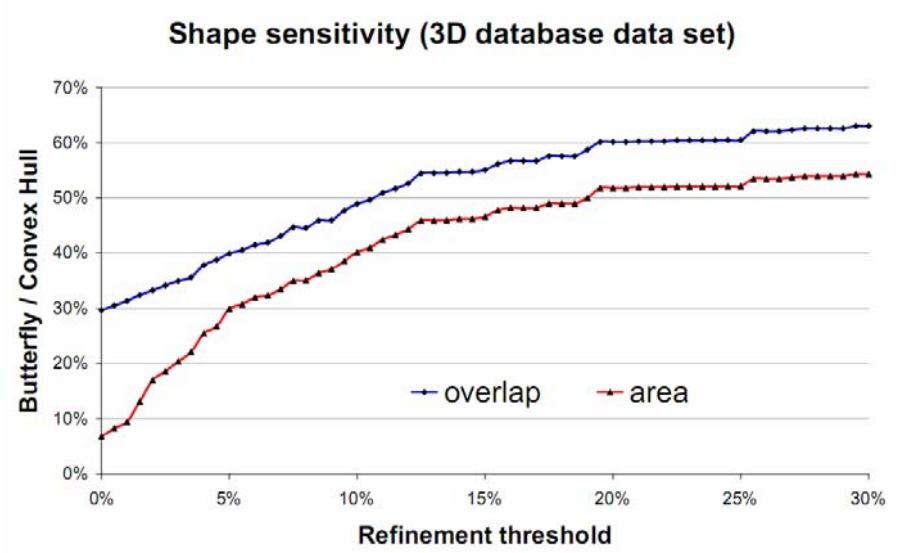


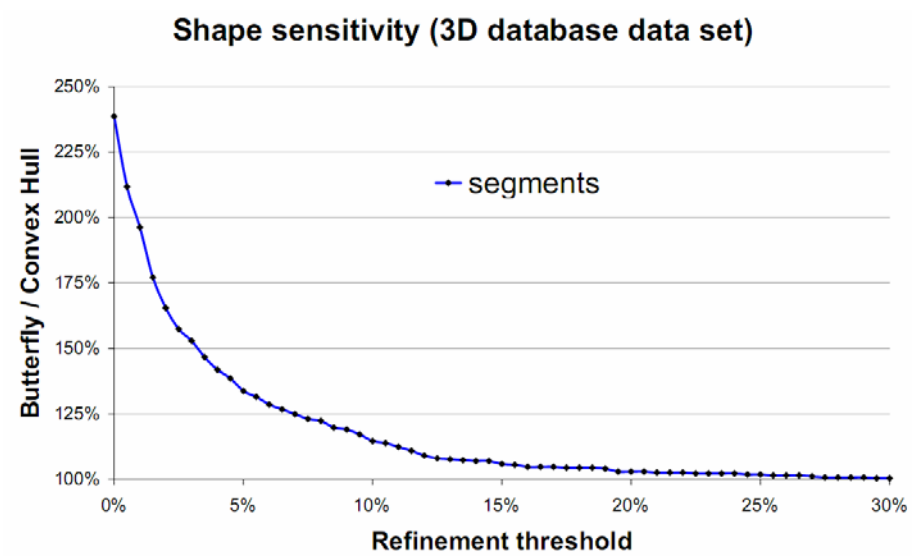


5. Evaluation (1)

Parameterization of the Butterfly plots

- Area and boundary complexity tradeoff
- Here: experimental sensitiv analysis
- Approaches: manual setting (slider) | elbow criterion





6. Conclusion

This work

- Construction of compact hull (Butterfly plots) for visualization of large point cloud data
- Flexible approach, automatic or interactive parameter specification
- Applications in data analysis, cluster analysis

Future work

- Research other shape-based point visualizations (distance fields; density-based metaphors; etc.)
- Optimize color and rendering order based on actual shapes
- Hybrid shape aggregation
- Preprocessing for very large point clouds (10.000s of classes…)