

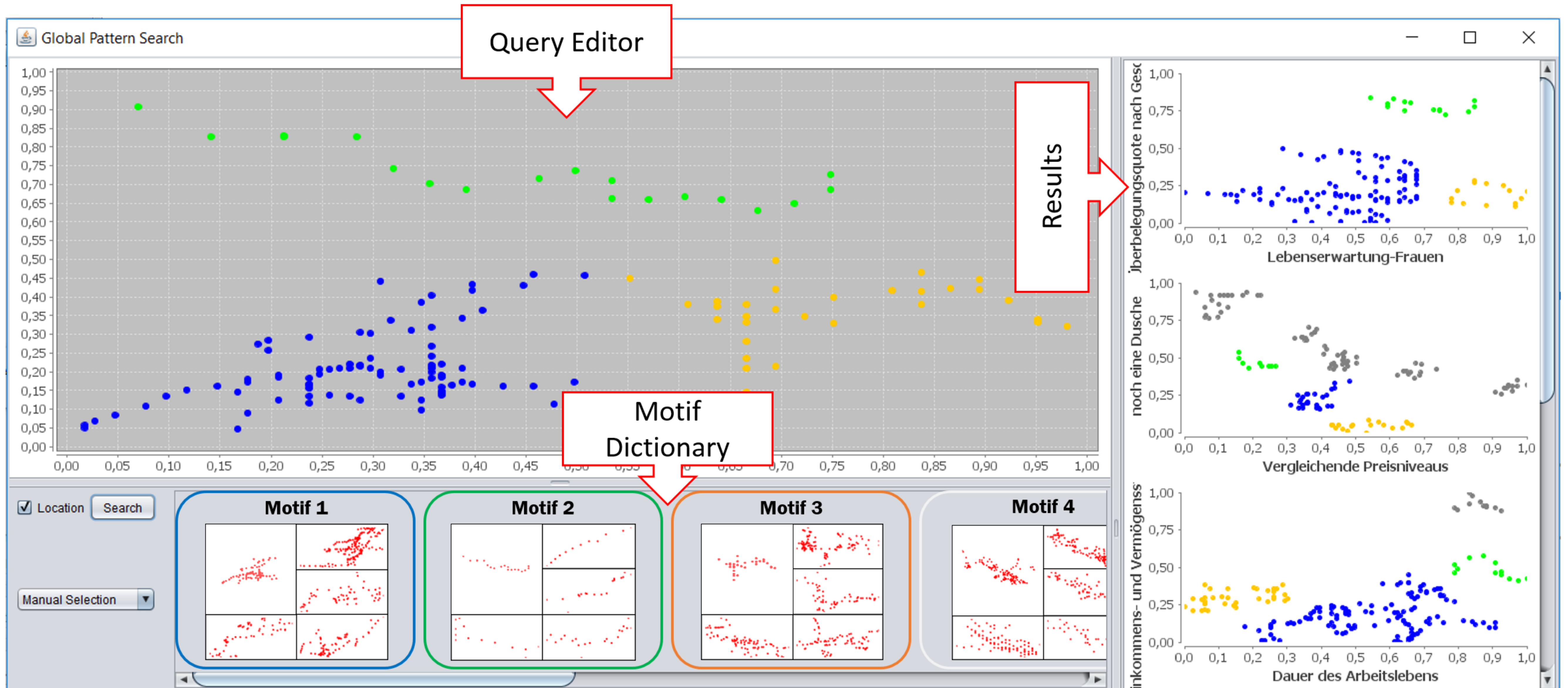
Query by Visual Words:

Visual Search for Scatter Plot Visualizations

Lin Shao¹, Timo Schleicher², Tobias Schreck¹

¹Graz University of Technology, Austria

²University of Konstanz, Germany

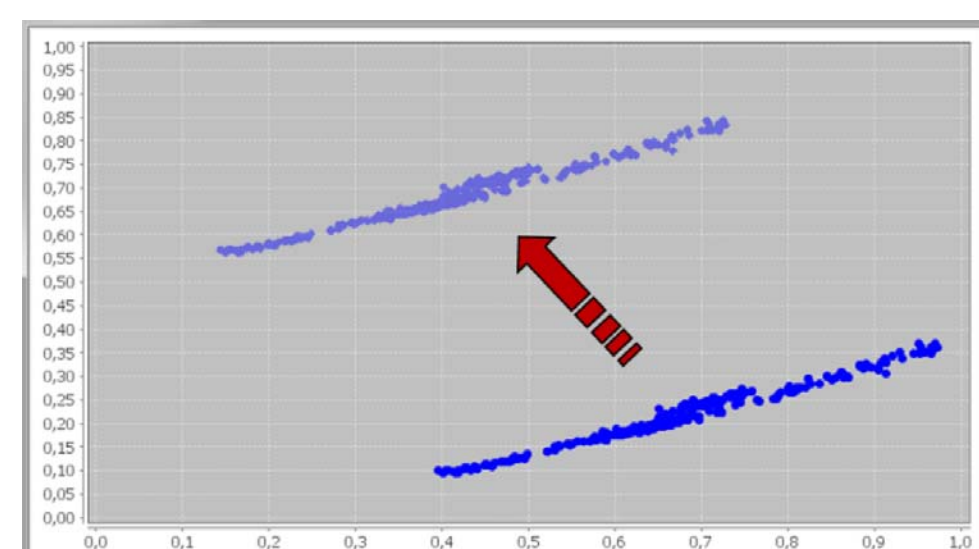


The Goal

- Transfer the Bag-of-Words concept from image retrieval to data visualization techniques.
- Decompose scatter plot views into regions of interest.
- Introduce a novel scatter plot search approach based on a dictionary of frequent local scatter plot patterns [1].

Query Editor

- The query editor can be used to edit a query scatter plot as the basis of the search.
- Users can select motifs from the dictionary as building elements of the query.
- Selected motifs can be freely positioned within the editor to formulate the final query scatter plot.



Motif Dictionary

Based on [1], we build the dictionary by

1. Segmenting local scatter plot patterns.
 2. Visual feature extraction of local patterns.
 3. Clustering the set of scatter plot segments.
- The Dictionary consists of clusters of similar scatter plot segments (so called visual words).
 - Dictionary entries are visualized by overview glyphs (motif prototype + associated patterns).
 - Users can quickly explore the local pattern space and choose interesting motifs for search.

Similarity Search

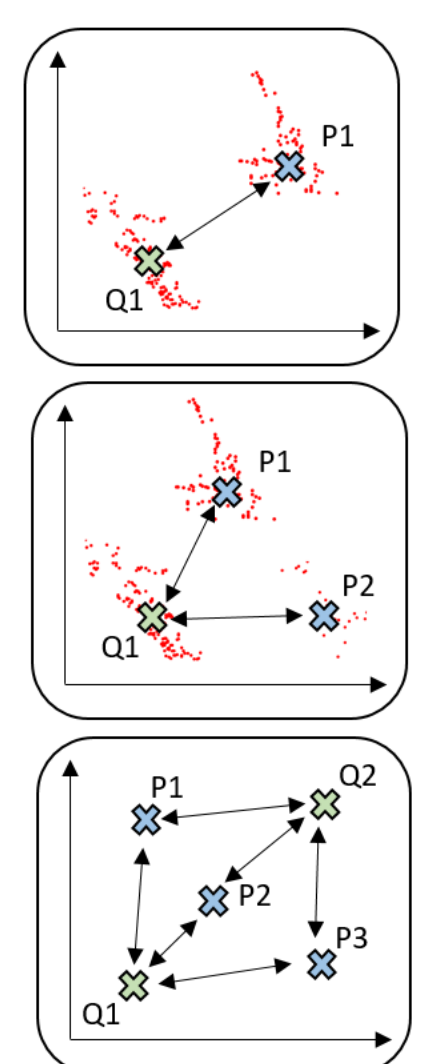
We use the motif dictionary, as well as the defined spatial positions of each motif to compute a similarity score between queries and target scatter plots. The overall similarity score is iteratively computed as a sum over the distances between each motif of the query (possibly several) and the matching motifs in the target. There are different cases:

1:1: A query motif occurs only once and matches to exactly one target motif.

1:N: A query motif occurs only once and matches to a set of target motifs.

N:M: A query motif occurs several times and is matched to a single or set of target motifs.

The similarity between a pair of motifs is computed based on the spatial distance between the motifs, optionally also considering the feature-based distance between the motif instances.



Related Work

- Local pattern analysis [1].
- Quality metrics for scatter plot visualizations [2,3].
- Sketch-based search techniques [4,5].

Future Work

- Sorting of dictionary glyph representation.
- Rescaling of query motifs.
- Comparison of different similarity functions (e.g., image features).

[1] L. Shao, T. Schleicher, M. Behrlich, T. Schreck, I. Sipiran and D. A. Keim: Guiding the exploration of scatter plot data using motif-based interest measures. In Proc. Int. Symp. on Big Data Visual Analytics (2015), pp. 1–8.
 [2] L. Wilkinson, A. Anand and R. Grossman: Graph-theoretic scagnostics. In In Proceedings of the IEEE Symposium on Information Visualization (2005), pp. 157–164.
 [3] A. Tatu, G. Albuquerque, M. Eisemann, P. Bak, H. Theisel, M. Magnor and D. A. Keim: Automated analytical methods to support visual exploration of high-dimensional data. Visualization and Computer Graphics, IEEE Transactions on 17, 5 (2011).
 [4] L. Shao, M. Behrlich, T. Schreck, T. van Landesberger, M. Scherer, S. Bremm and D. A. Keim: Guided Sketching for Visual Search and Exploration in Large Scatter Plot Spaces. In Proc. Int. Workshop on Visual Analytics (2014), EG.
 [5] B. Lee, G. Smith, N. H. Richie, A. Karlson and S. Carpendale: Sketchinsight: Natural data exploration on interactive whiteboards leveraging pen and touch interaction. In 2015 IEEE Pacific Visualization Symposium (PacificVis) (2015), pp. 199–206.