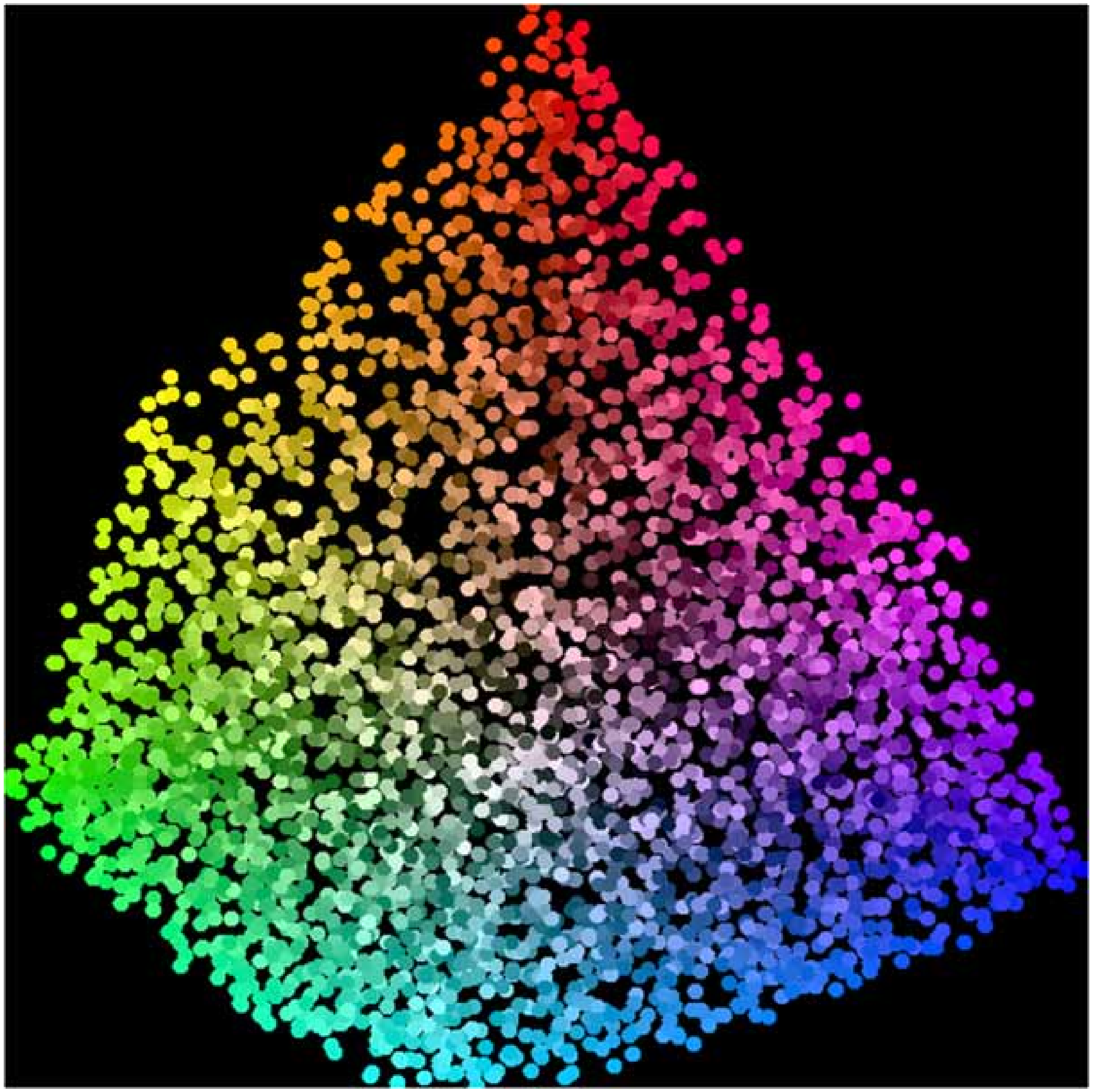
Arranging Arbitrary Data into Structured Layouts

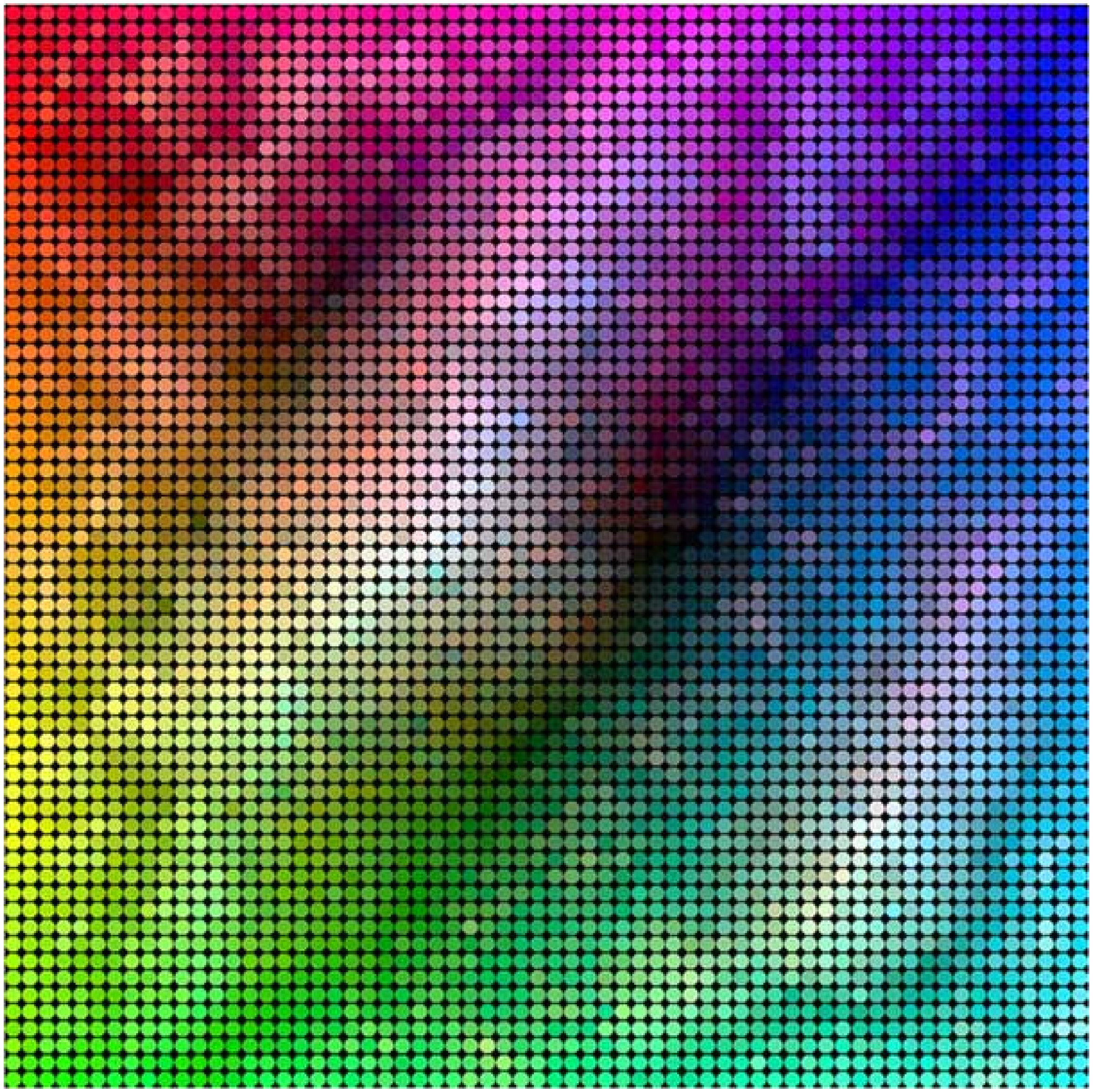
Minglun Gong (collaboration w/ Grant Strong)
Memorial University, Canada

Problem Formulation



Dimension Reduction

Map data onto low-dimension continuous space.



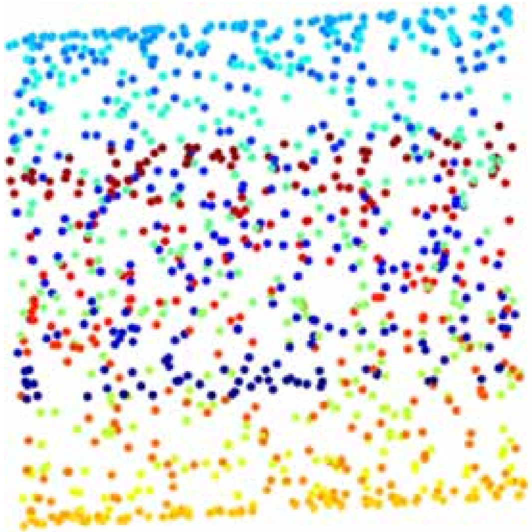
Occlusion Free Mapping

Map data onto low-dimension discrete positions without occlusion.

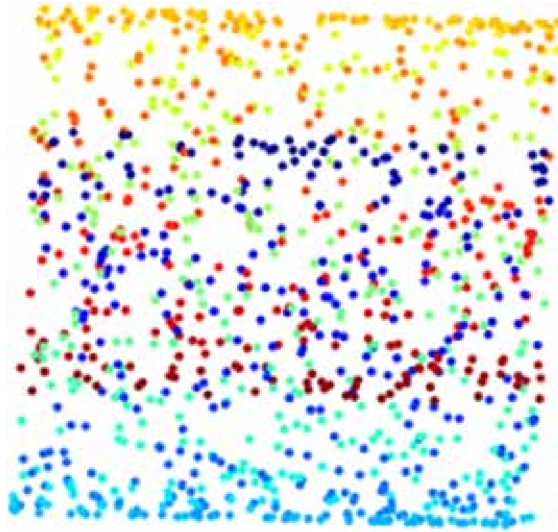
- **Find an occlusion-free mapping, where the proximity of data items in the structured layout correlates with their similarities**
- $\rho(M) = \frac{\sum_{\forall s,t \in \Omega}\left(\psi(M(s),M(t)) - \bar{\psi}\right)\left(\delta(s,t) - \bar{\delta}\right)}{\sigma_{\psi}\sigma_{\delta}|\Omega|^2}$
  - $M: \Omega \rightarrow \Gamma$: **A mapping function.**
  - $\delta(\cdot,\cdot)$: **dissimilarity between two items.**
  - $\psi(\cdot,\cdot)$: **distance between two cells in the layout.**

Problem Formulation

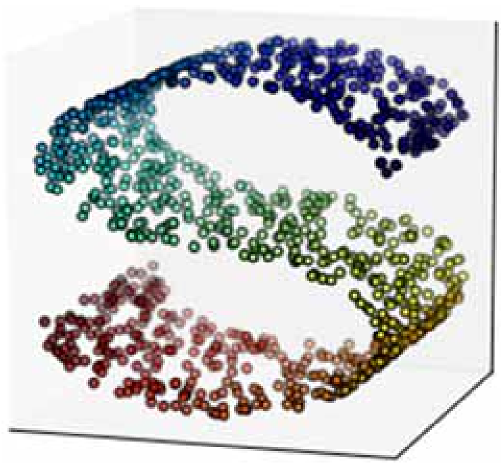
RELATED WORK

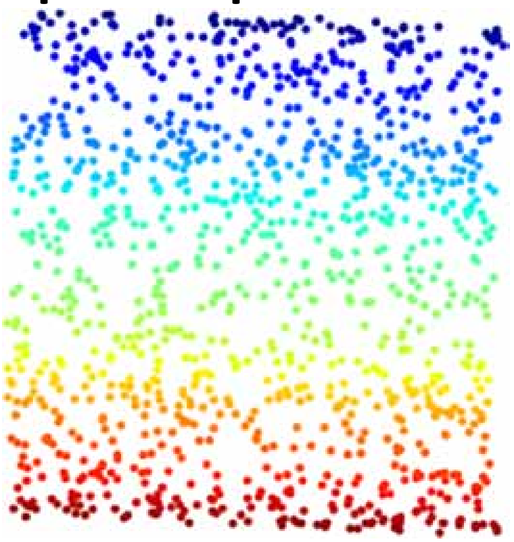


Principal component analysis

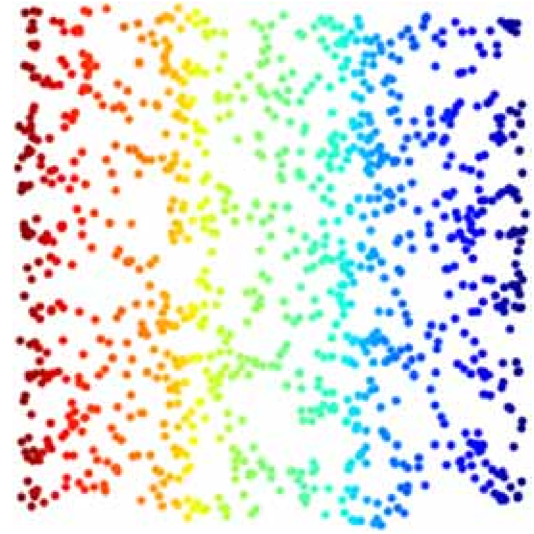


Multidimensional scaling





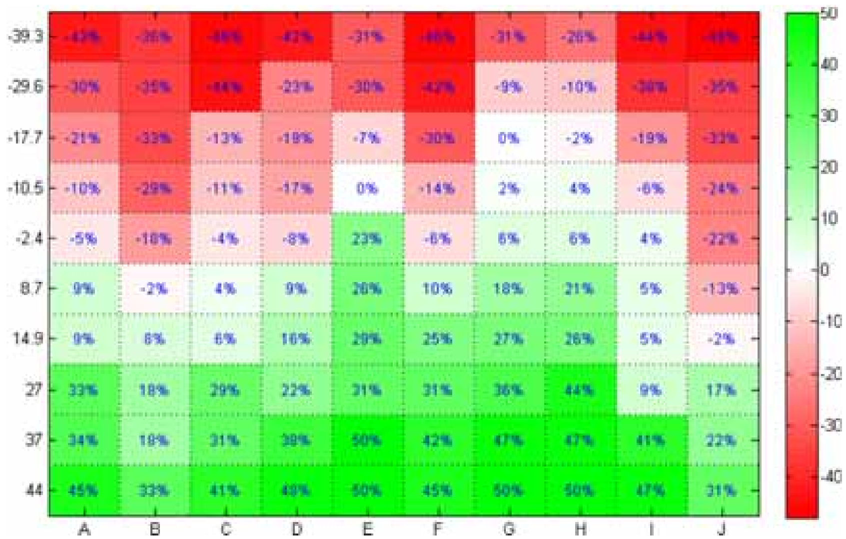
Locally-linear embedding



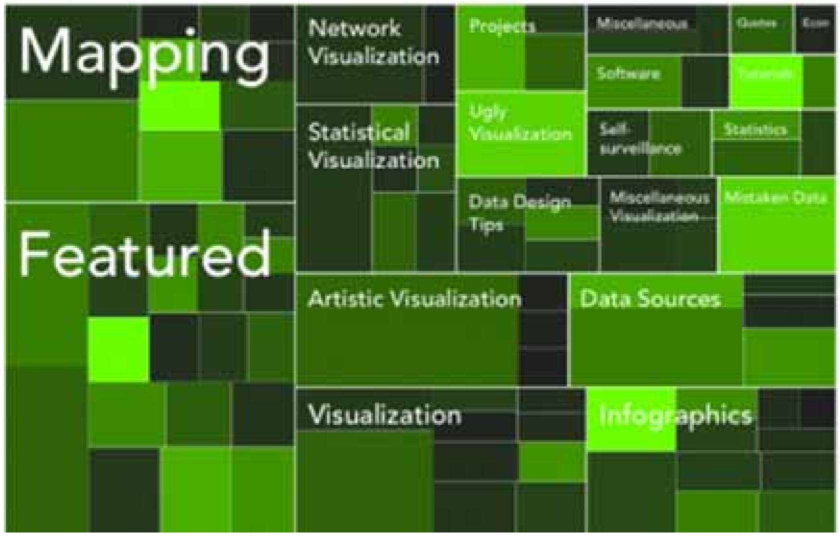
Isomap

Dimensionality Reduction

All map to continuous space, with possible occlusion.

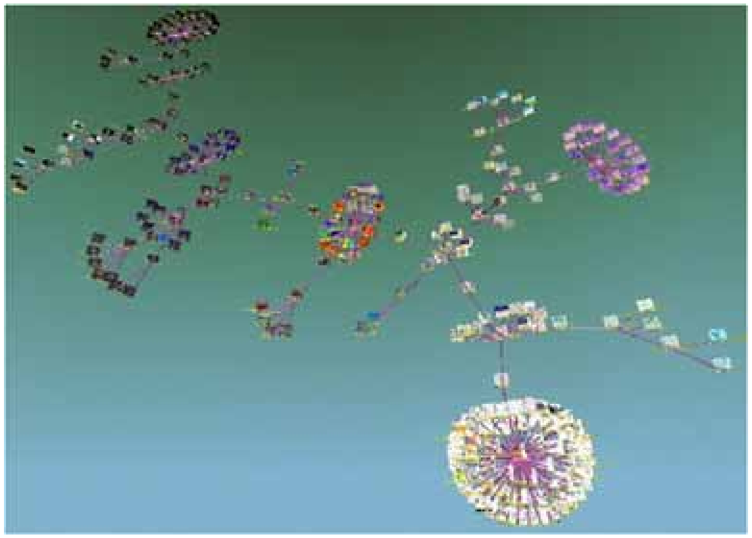


Heatmap



Treemapping

Tag cloud

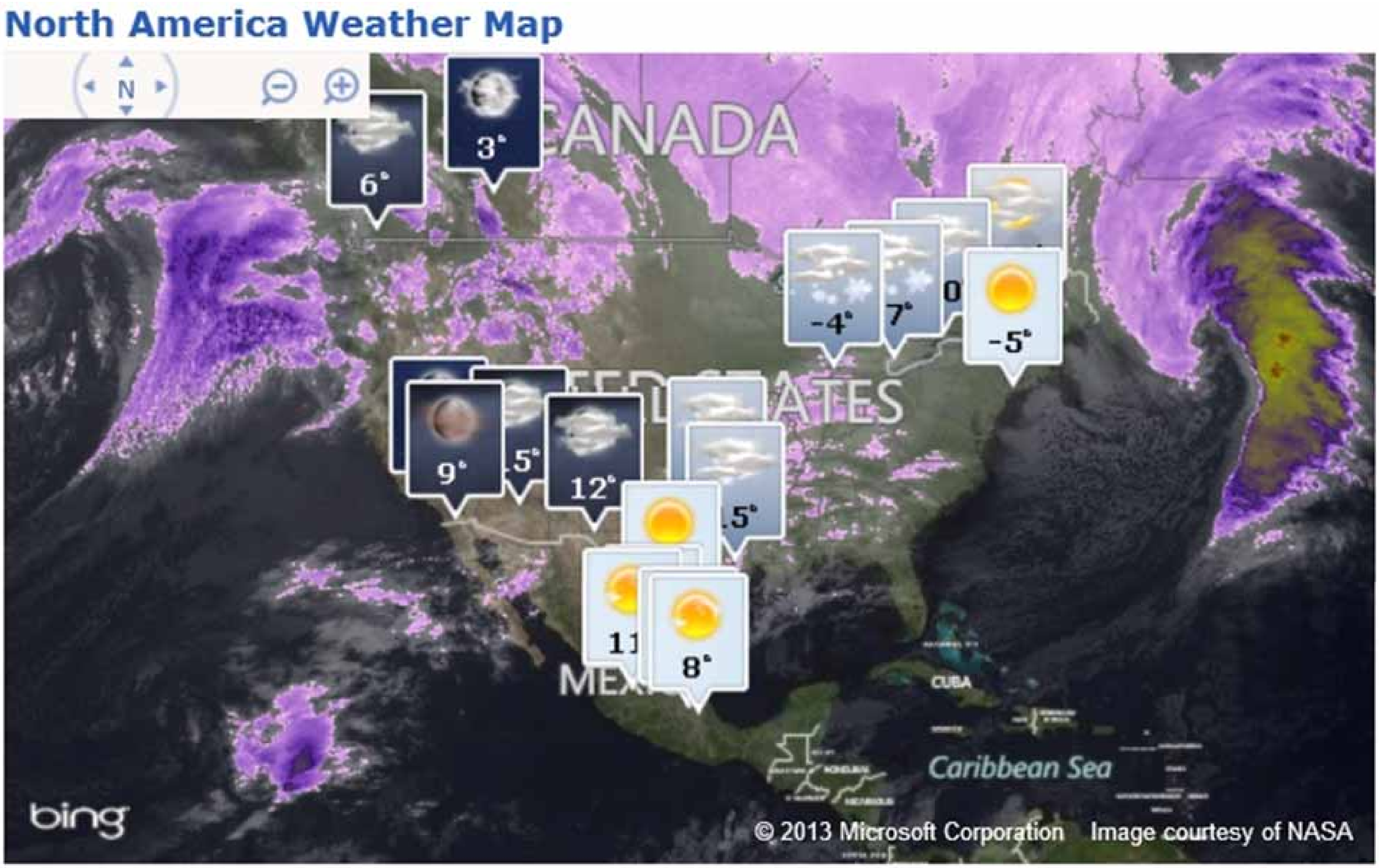


Pathfinder network

Information Visualization

Require effects to understand the visualization.

MOTIVATIONS



Interactive Map

Some information is occluded; requires user to interact with the map.

North America Weather Conditions

Calgary	3°	Mexico City	7°
Chicago	-4°	Monterrey	15°
Ciudad Juarez	12°	Montreal	-12°
Ciudad Nezahualcoyotl	10°	New York	-5°
Dallas	14°	Phoenix	15°
Detroit	-7°	Puebla	8°
Guadalajara	10°	Tijuana	9°
Houston	15°	Toronto	-10°
Leon	7°	Vancouver	6°
Los Angeles	12°	Zapopan	11°

Sorted List

Require search by name; Does not provide regional overview.

Structured Layout

Easy to grasp regional weather; No occlusion; Easy to refind a given city later.

Other Possible Applications

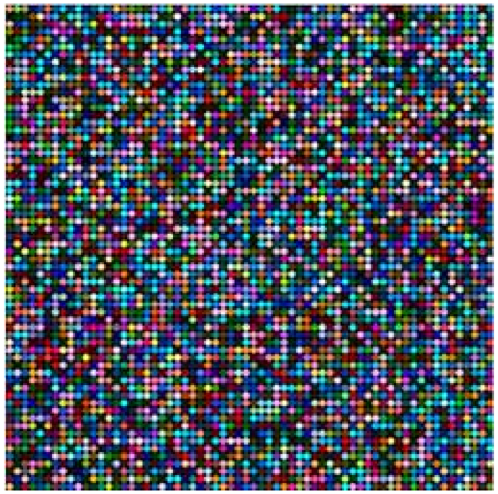
Apps in Windows 8

1st Attempt:

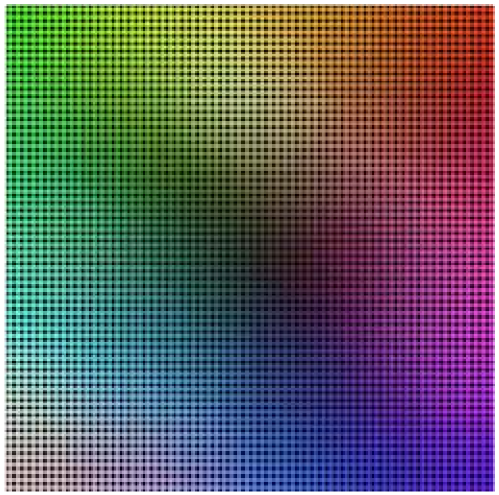
SELF-ORGANIZING MAP + KD-TREE

- Developed by Teuvo Kohonen in 1988.
- A special type of Artificial Neural Network
  - Maps high dimensional space to a low one
  - Topology preserving so visualization is meaningful
- Consist of interconnected units
  - Each unit keeps a K-dimensional weight vector
  - An unsupervised learning process is used to train the weight vectors

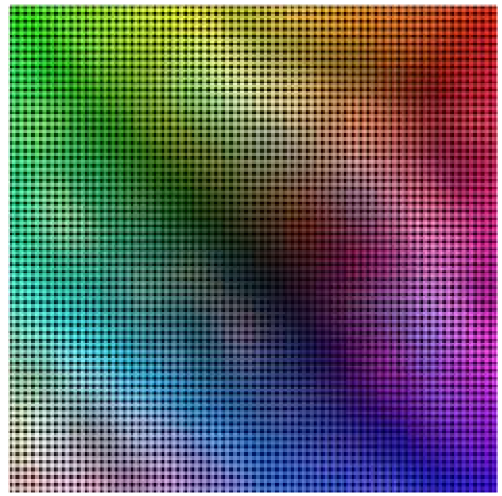
Self-Organizing Map



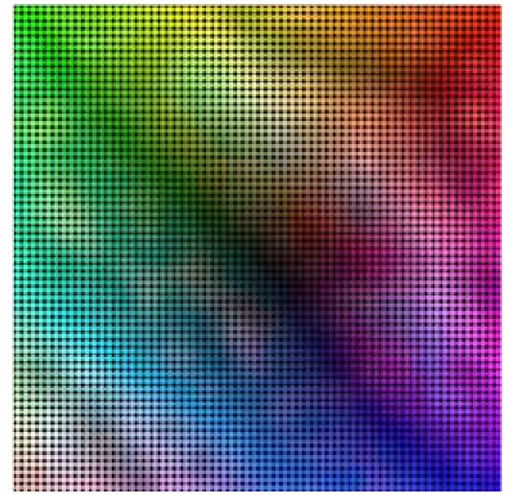
Initial Weights
$I = 0$



Early Weights
$I = 5$



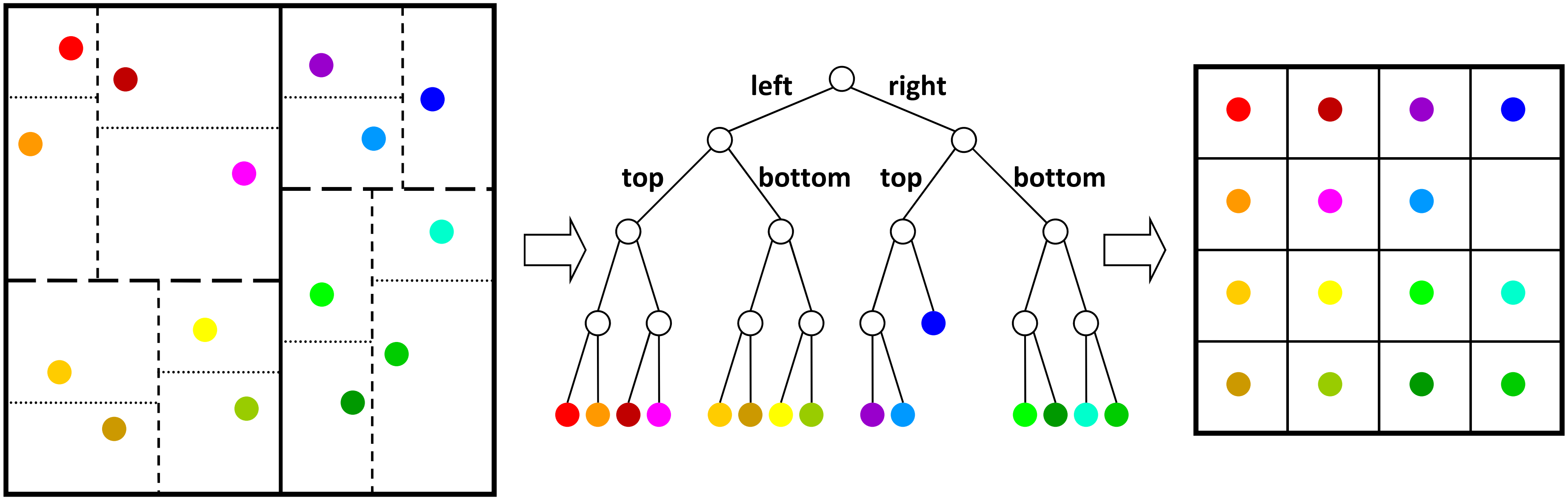
Later Weights
$I = 10$

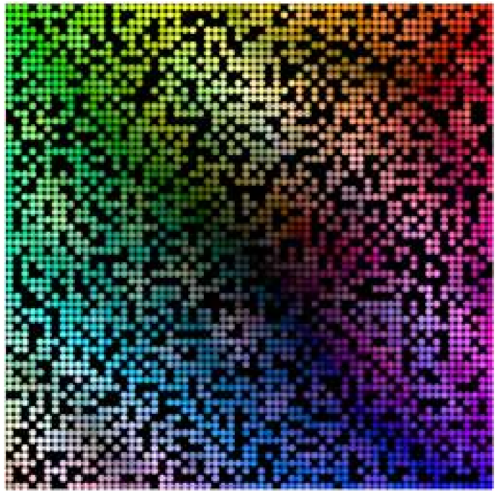


Final Weights
$I = 15$

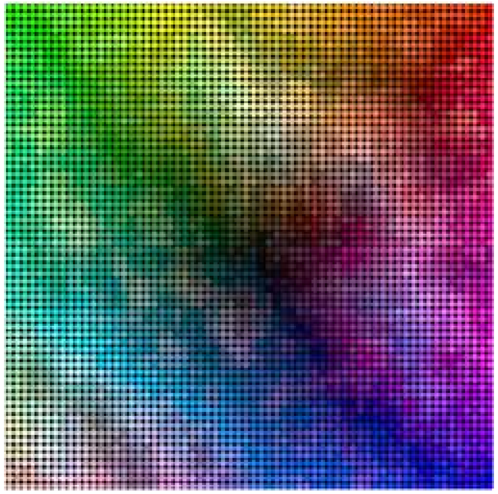
Convergence of SOM

Start with random vectors; gradually converges as the neighborhood size reduces.





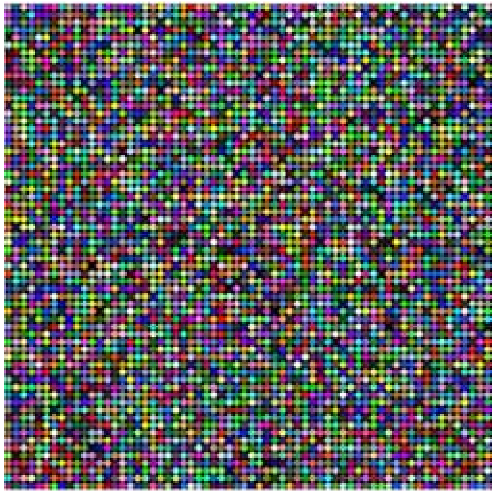
Unaligned Items



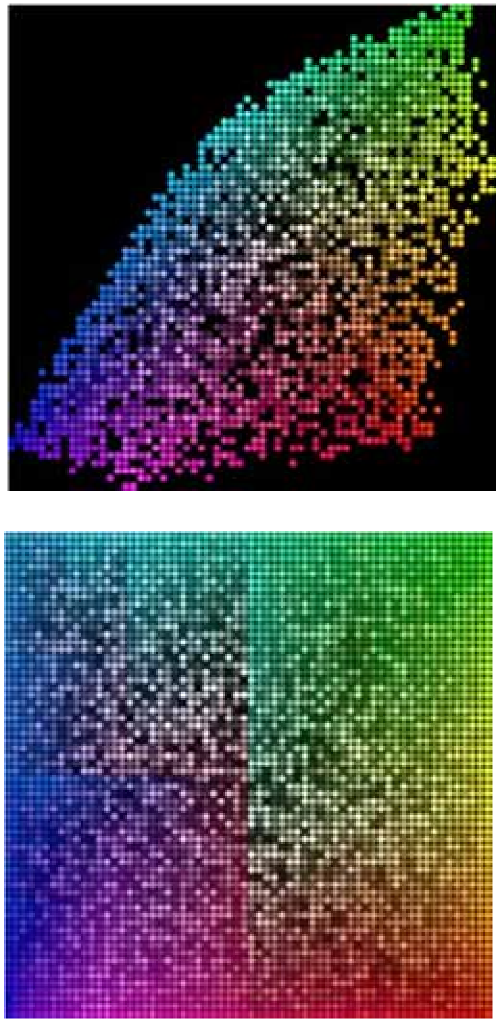
Aligned Items $C = 0.794$

Self-Organizing Map

Occlusion can be removed using a K-D Tree.

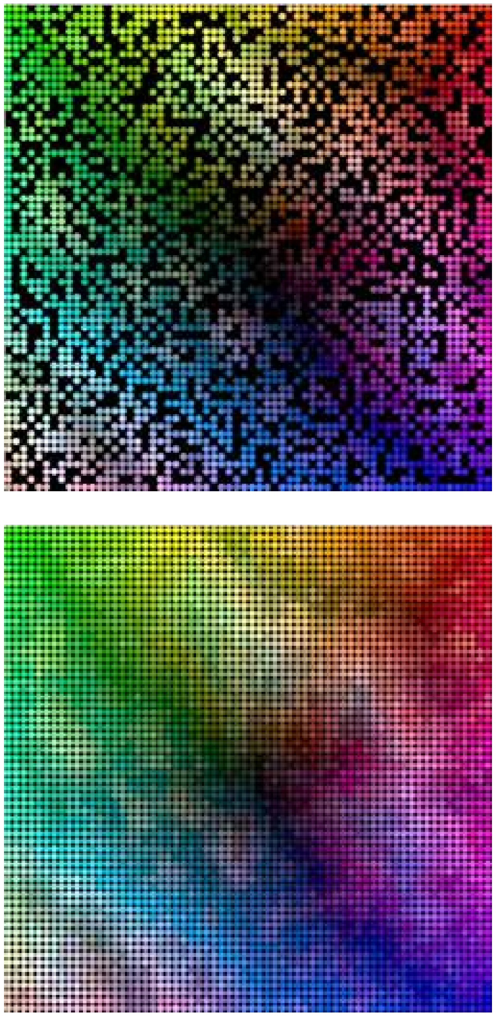


Initial Items



Sammon:

$T = 5.9s, C = 0.814$



SOM:

$T = 9.4s, C = 0.794$

SOM vs. Sammon's Mapping

Effects of KD-tree is more noticeable in Sammon's mapping

= Eiffel tower Sky Lawn +

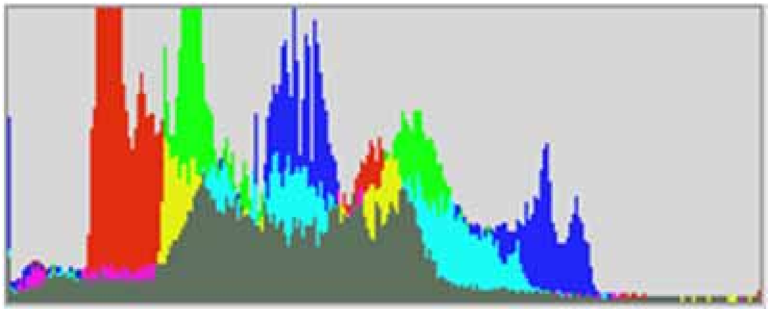
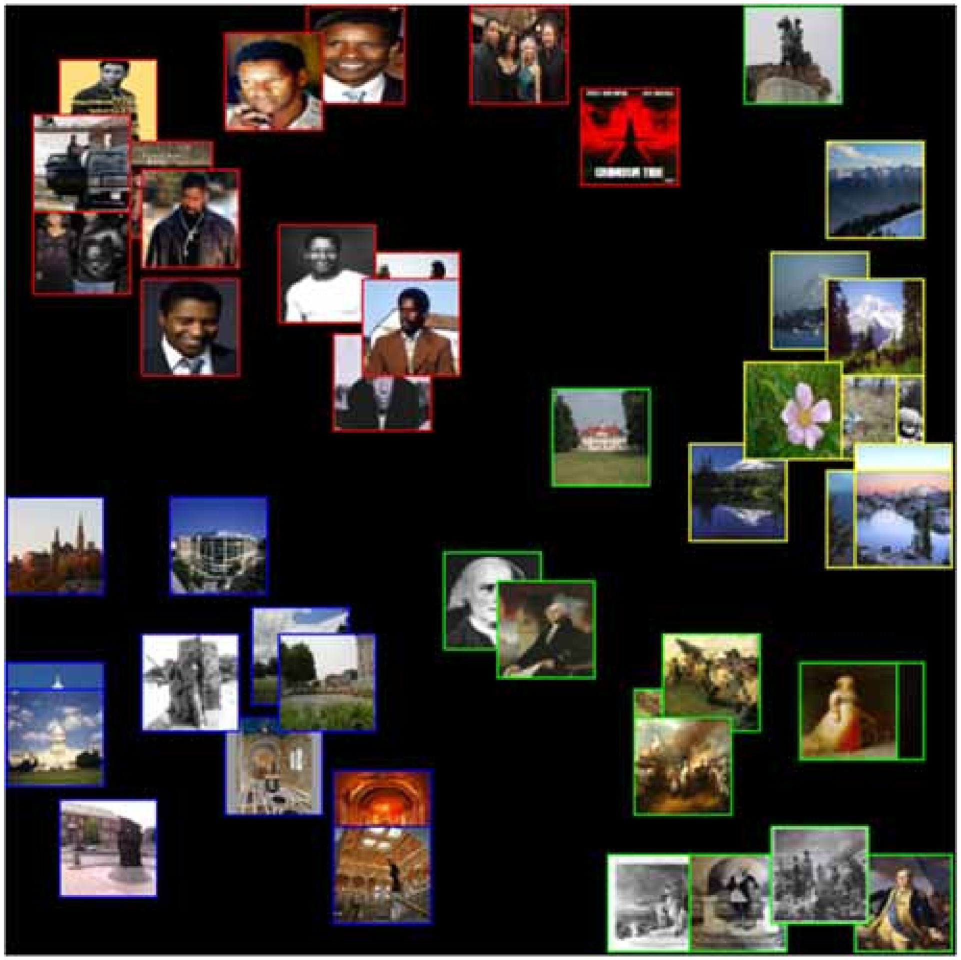


Image = Semantic features + Visual features

Complex Data: Images

Images carry both visual and semantic information.



SOM: $T = 0.21s, C = 0.635$

KD-tree: $T = 0.141s, C = 0.482$

SOM with Images

Trained over hybrid vectors representing both semantic and visual information.

	DW	GW	WDC	WS
Denzel Washington	0	0.024	0	0.517
George Washington	0.024	0	0.67	0.914
Washington DC	0	0.67	0	0.687
Washington State	0.517	0.914	0.687	0

Limitations of SOM + KD-tree

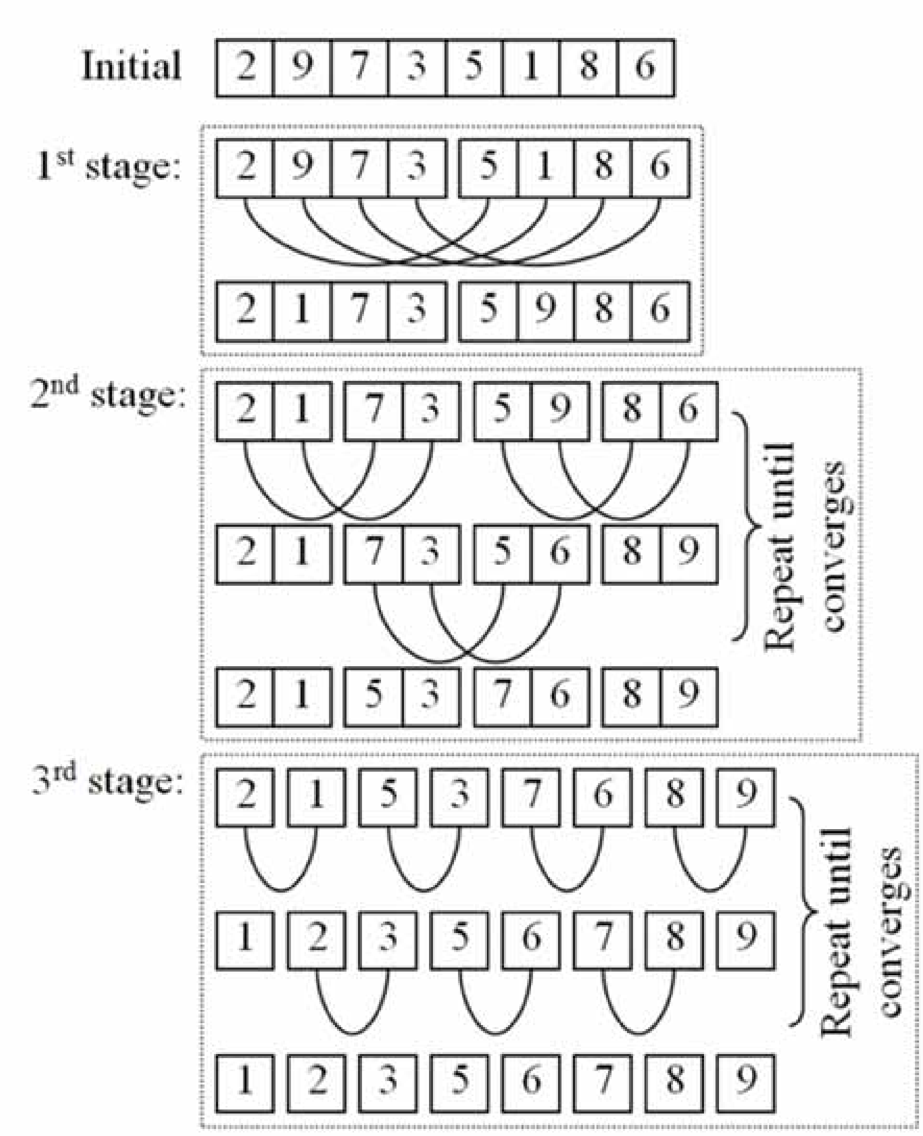
Representing data with vectors may introduce errors.

2nd Attempt:

SELF-SORTING MAP

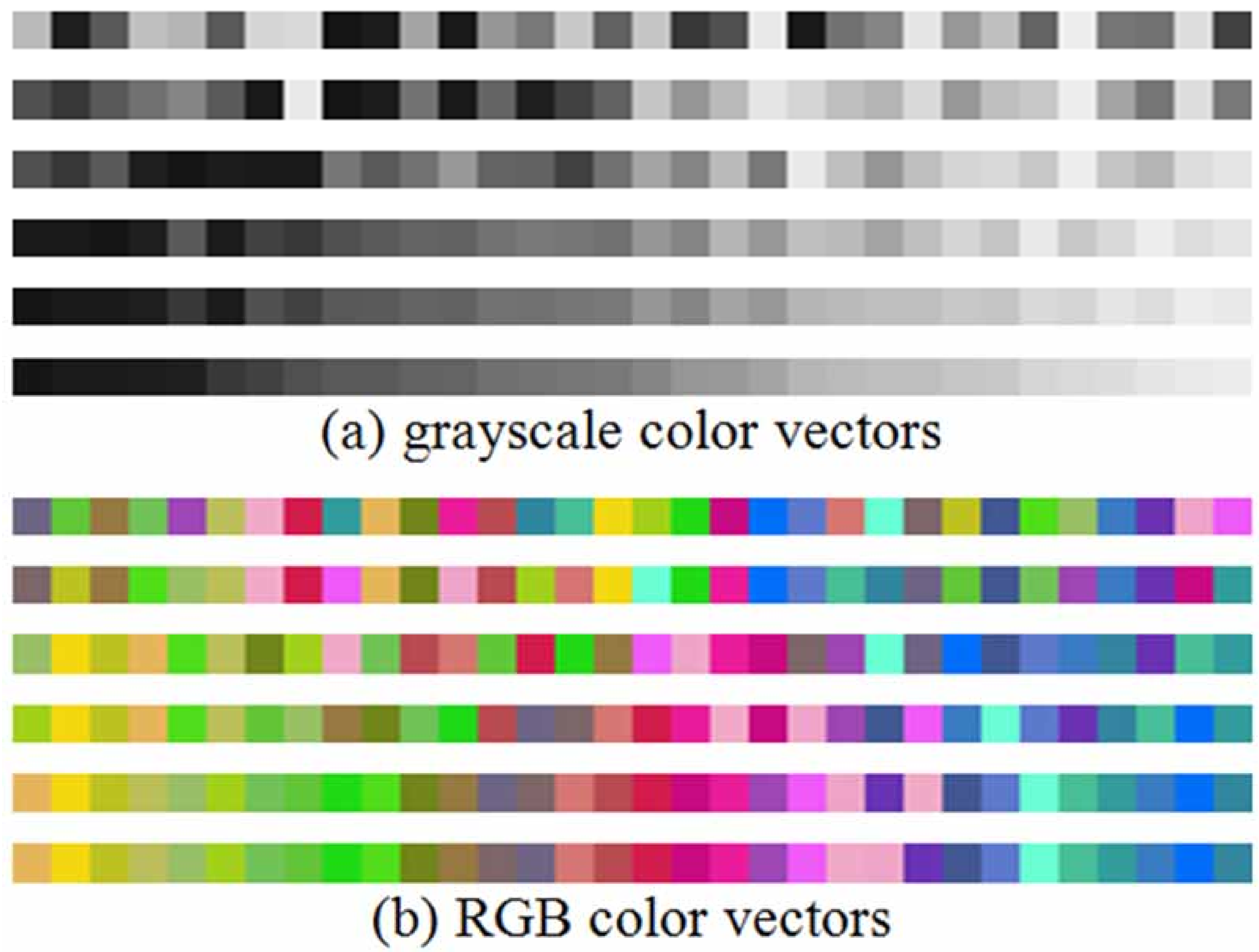
- **Select a target for each area of the map.**
- **Move items similar to the target into the corresponding area.**
- **Incorporate the ideas of sorting, k-mean clustering, and self-organizing map.**

Basic Idea for SSM



Sort 1D Numbers into 1D Array

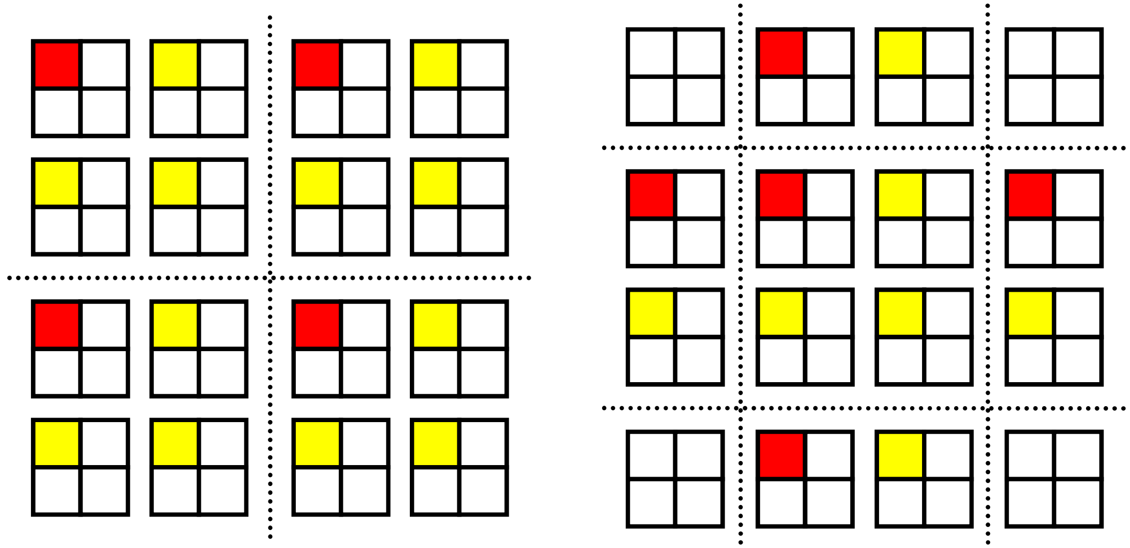
Odd-even sorting with shell sort spacing.



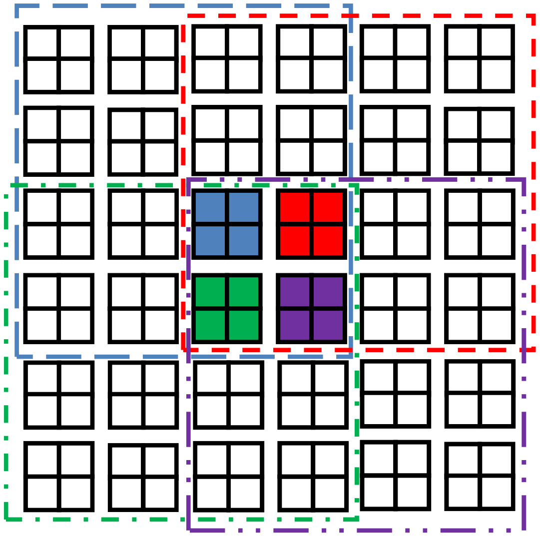
Organize 3D Vectors into 1D Array

Use a similarity measure, instead of an ordering function.

Cell Grouping

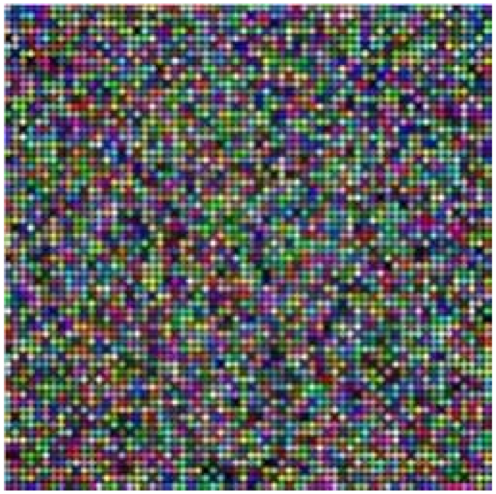


Neighborhood

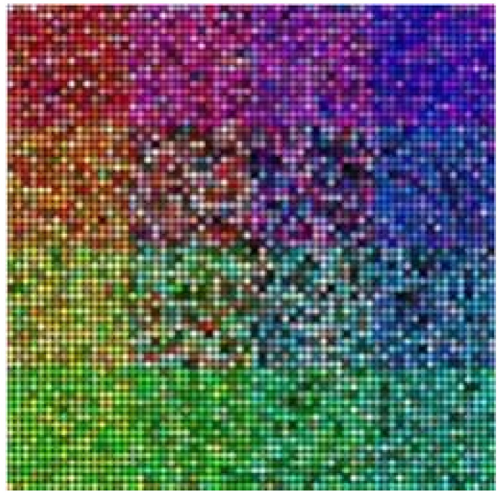


Organize Data into 2D Grid

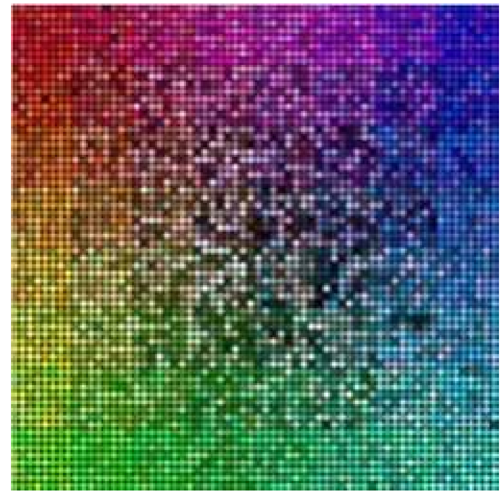
Alternate between target selection and alignment at each block level.



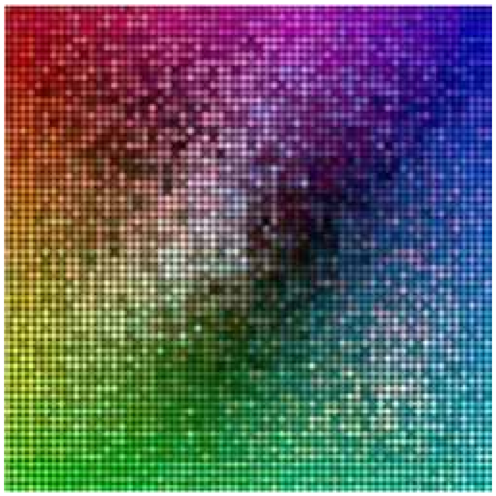
1st stage: $C = 0.009$



2nd stage: $C = 0.604$

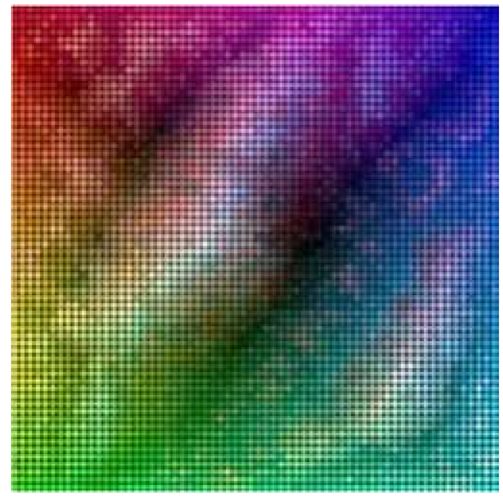


3rd stage: $C = 0.789$



4th stage: $C = 0.825$

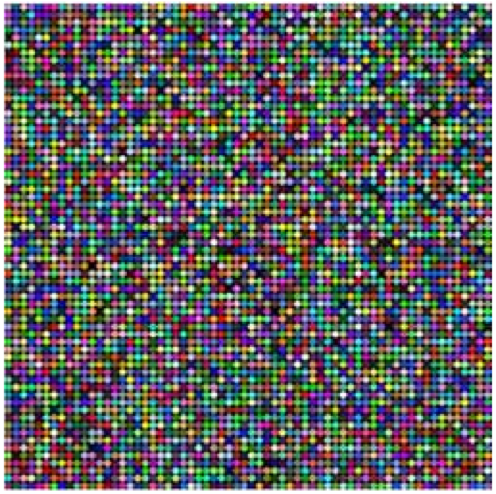
5th stage: $C = 0.830$



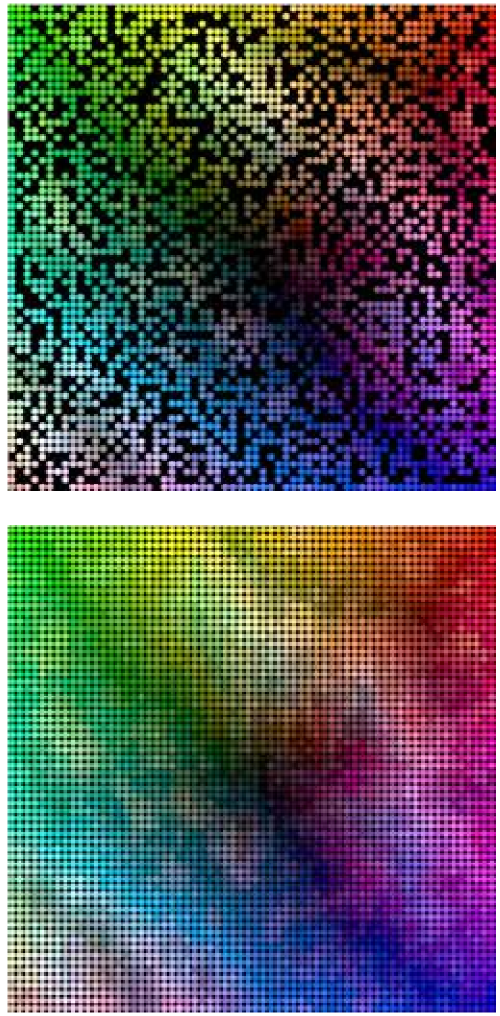
Final: $C = 0.831$

Convergence of SSM

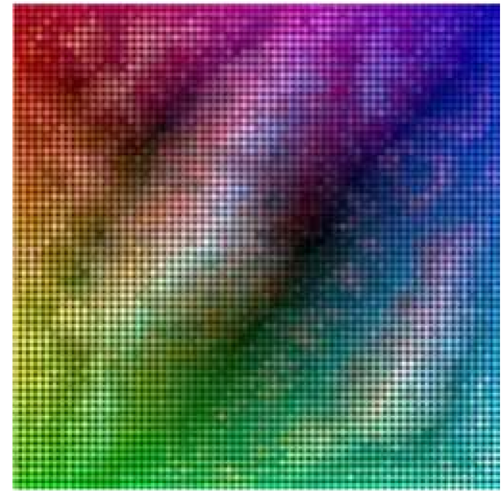
Organize 3D Vectors into 2D Array.



Initial Items



SOM:
$T = 9.4s, C = 0.794$



SSM:
$T = 0.82s, C = 0.831$

SSM vs. SOM

SSM is faster and better correlated in the resulting structured layout.

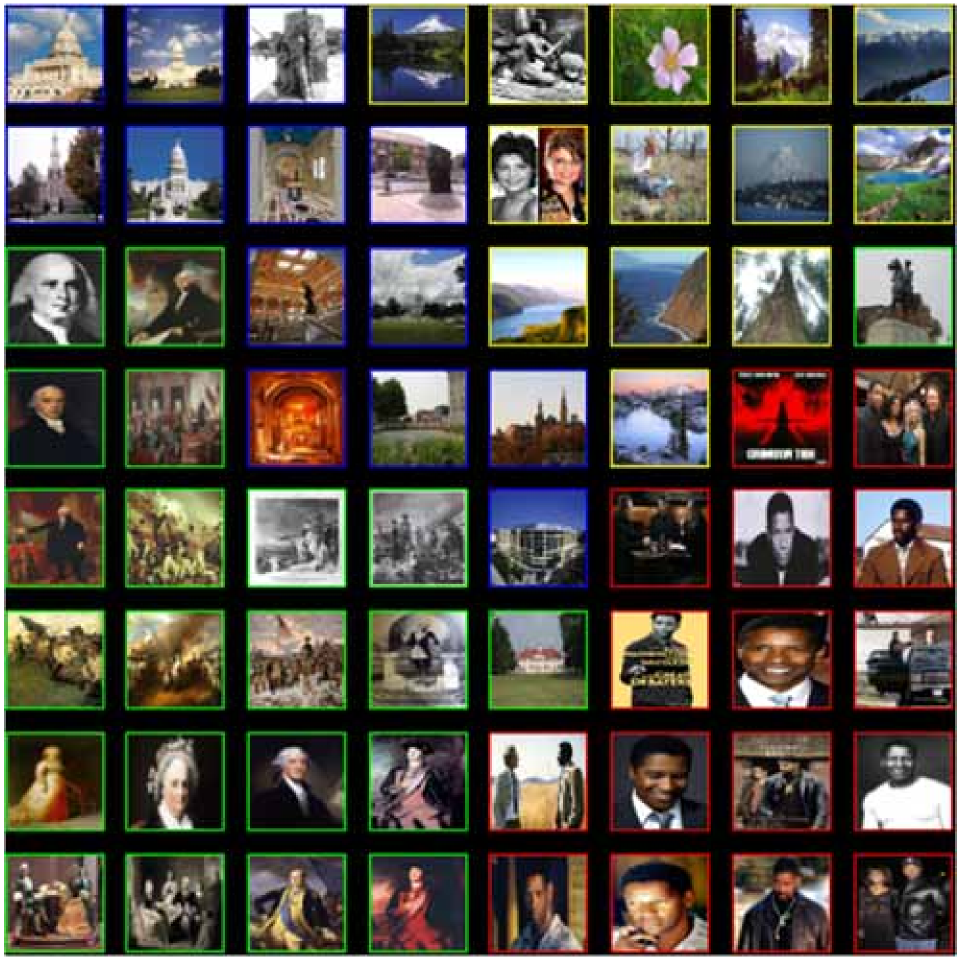
$C = 0.506$

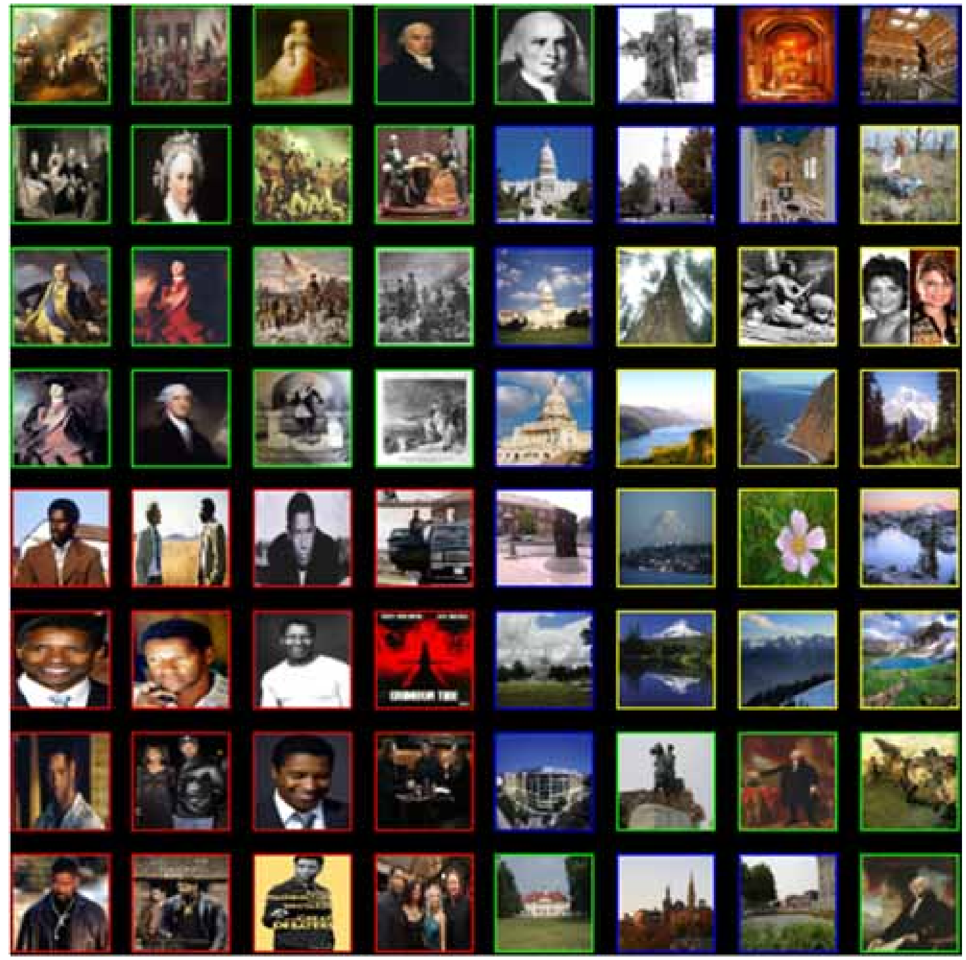
$C = 0.581$

$C = 0.541$

User Intervention by Boundary Condition

Users can define the desired properties along the borders.





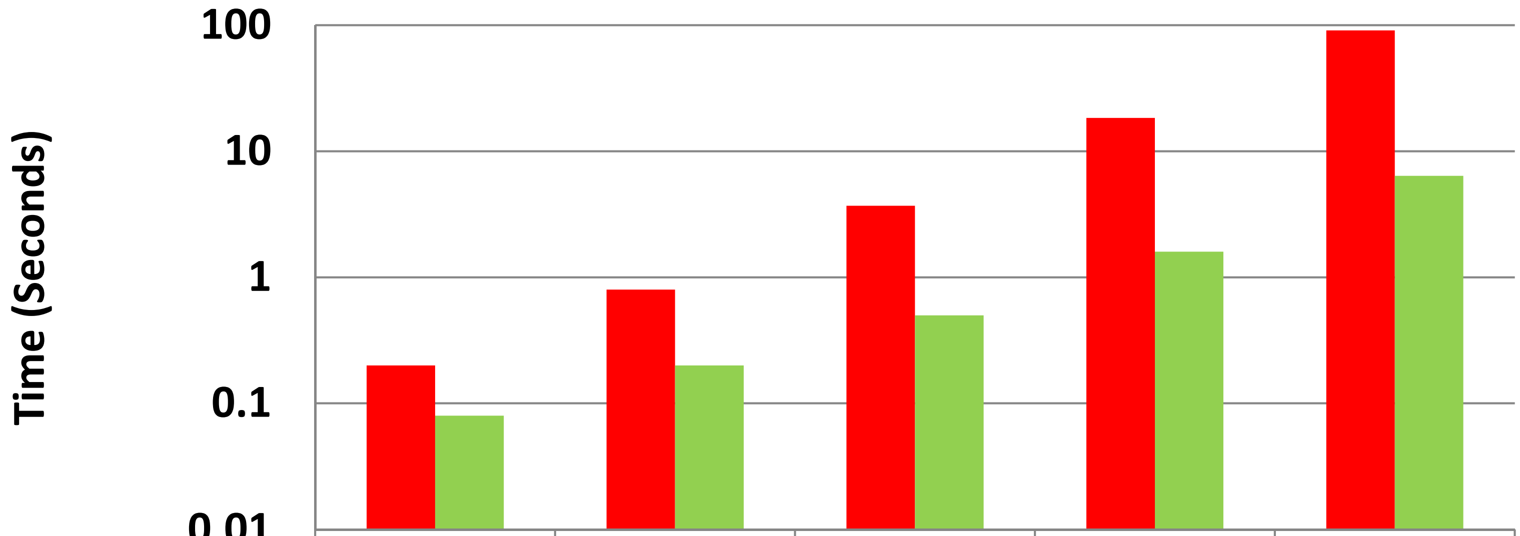
Means: $T = 0.148s, C = 0.679$

Centroids: $T = 0.141s, C = 0.482$

SSM with Images

Mean result produces slight better result over SOM; Centroids does not work as well.

CPU vs. GPU Runtimes



	4096	16384	65536	262144	1048576
Serial CPU	0.2	0.8	3.7	18.4	91
Parallel GPU	0.08	0.2	0.5	1.6	6.4

Number of random color vectors

GPU Implementation

GPU implementation achieves significant speedup; Million+ items arranged in 6.4 seconds.

3rd Attempt:

Max Correlation Map

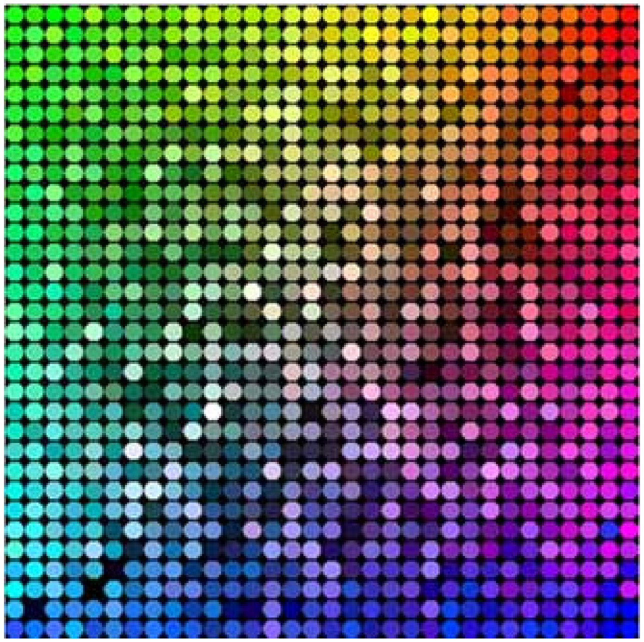
- **Directly optimize**

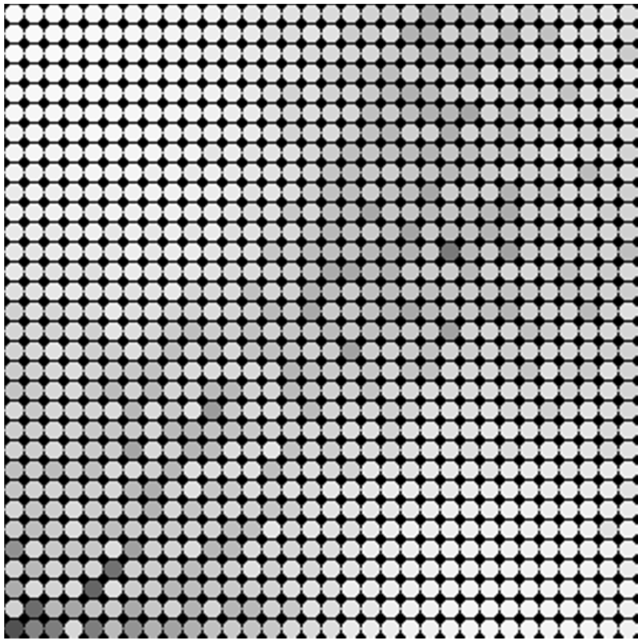
$$\underset{M}{argmax} \frac{\sum_{\forall s,t \in \Omega} \left(\psi\left(M(s),M(t)\right)-\bar{\psi}\right)\left(\delta(s,t)-\bar{\delta}\right)}{\sigma_{\psi}\sigma_{\delta}|\Omega|^2}$$

  - **Start with a random mapping.**
  - **Randomly select two cells and swap the data if such a swapping increases the correlation.**
- **Avoid local optimum.**

Basic Idea for MCM

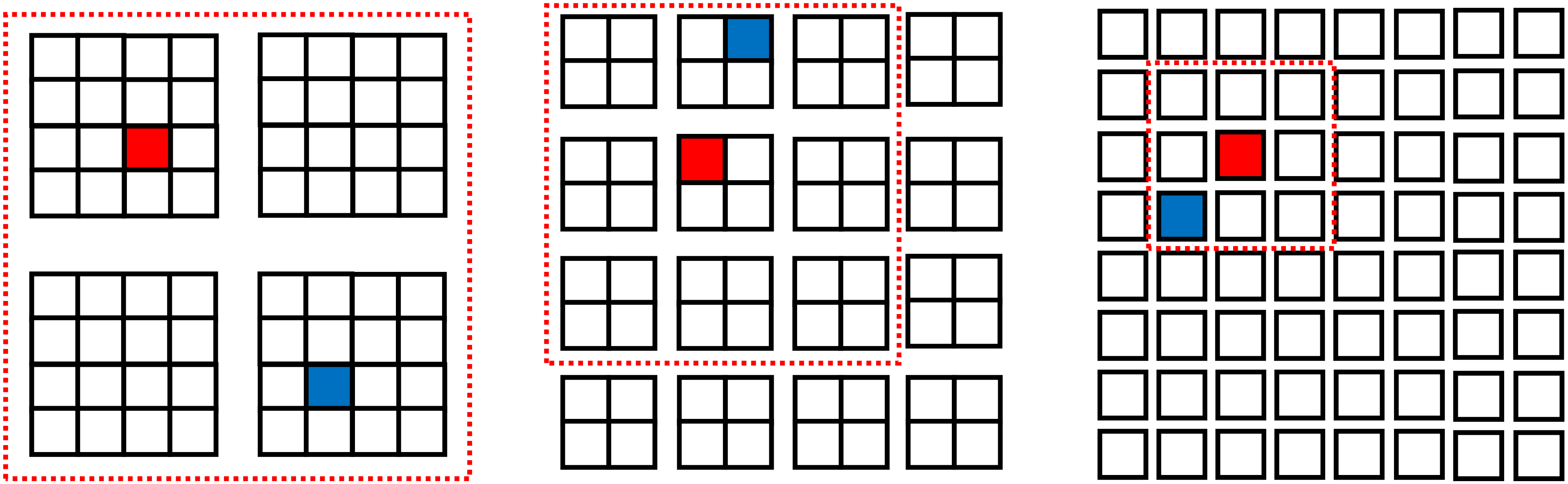
$$\boldsymbol{\rho(M) = \frac{\sum_{\forall u,v \in \Gamma} \Psi(u,v) \cdot \Delta\left(M^{-1}(u), M^{-1}(v)\right)}{|\Gamma|^2} = \frac{\sum_{u \in \Gamma} \rho_u(M)}{|\Gamma|}}$$





Cell Correlation

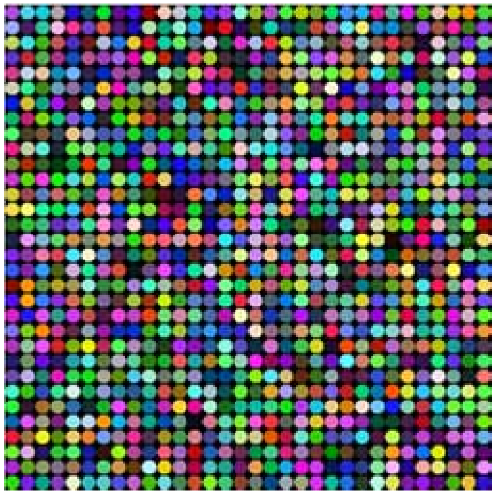
Total correlation is the sum of cell correlations.



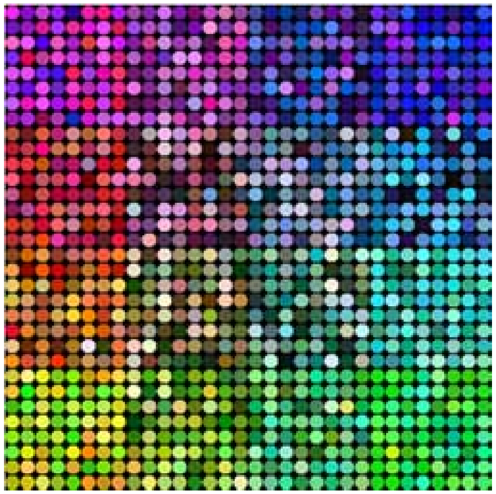
Swap Neighborhoods

Coarse-to-Fine Swap-based Search

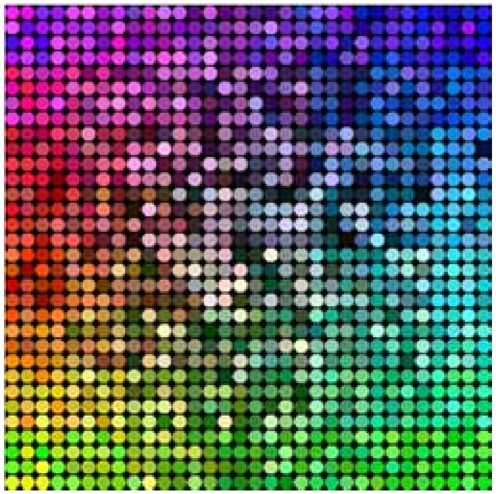
Avoid local optimum.



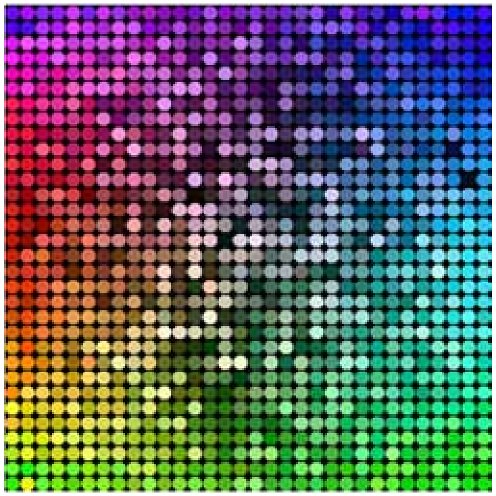
1st stage



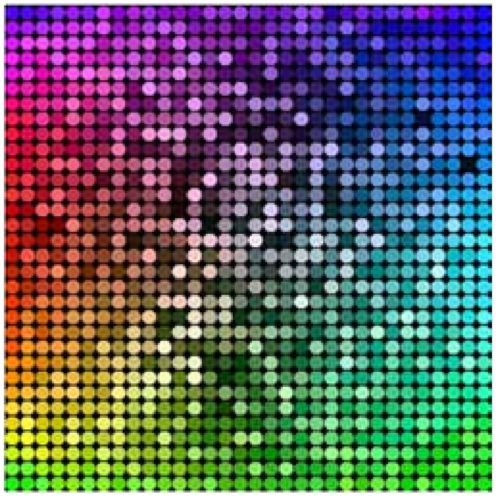
2nd stage



3rd stage



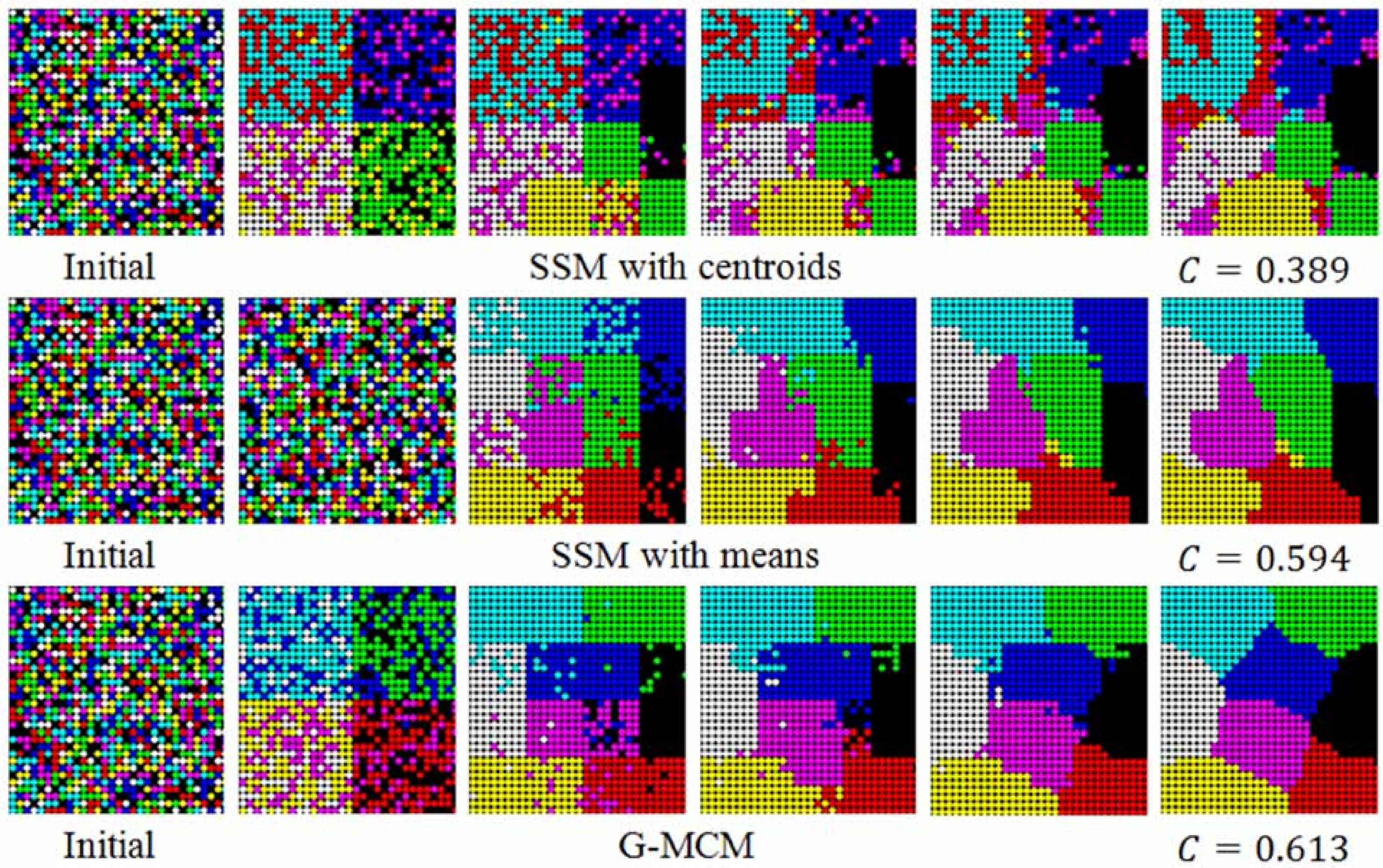
4th stage



5th stage

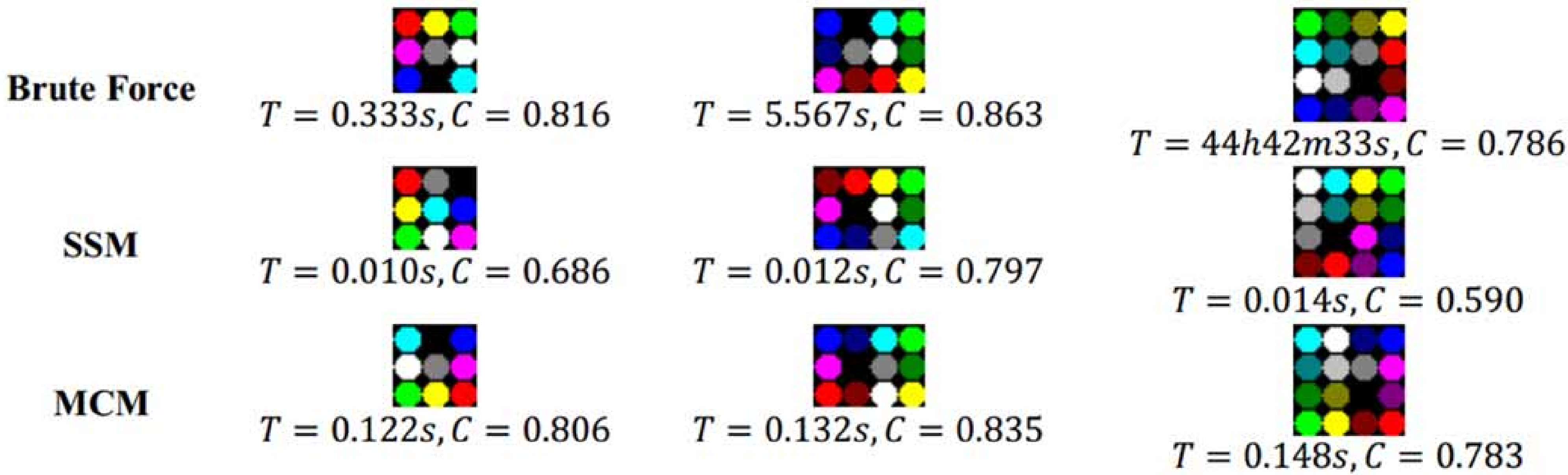
Convergence of MCM

The organization is globally smoother than SSM, but locally more noisy.



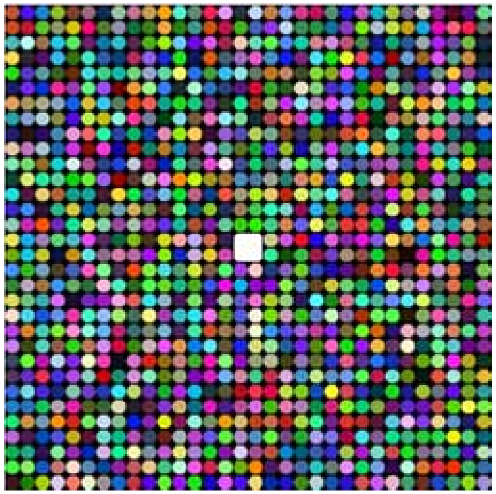
MCM vs. SSM

Better handles data clusters (binary color vectors).

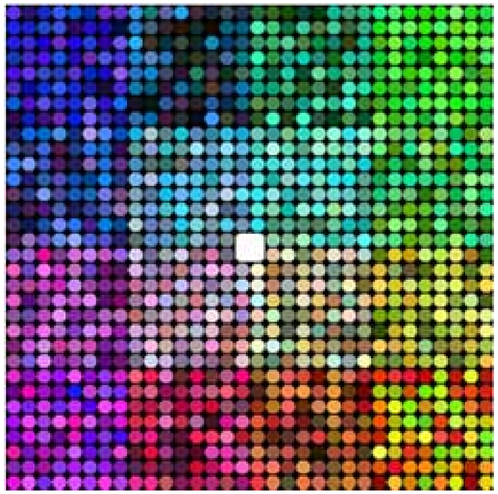


Evaluation using Ground Truth

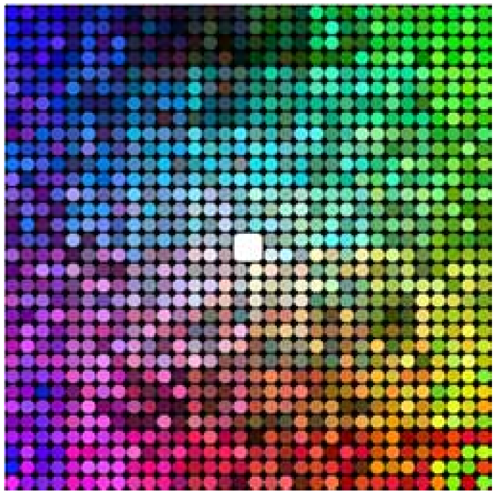
Closer approximation to the best found by enumeration.



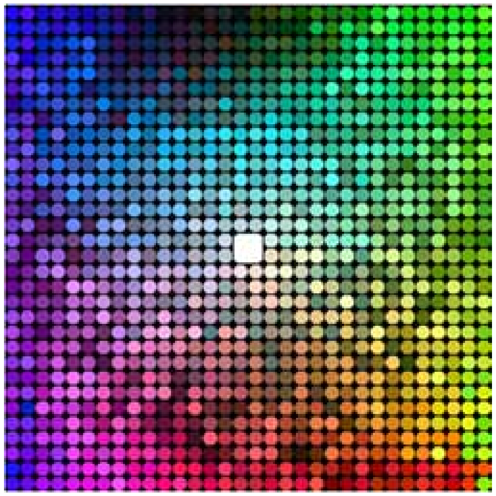
1st stage



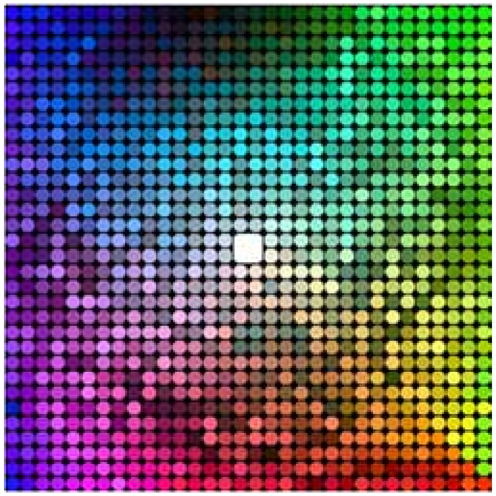
2nd stage



3rd stage



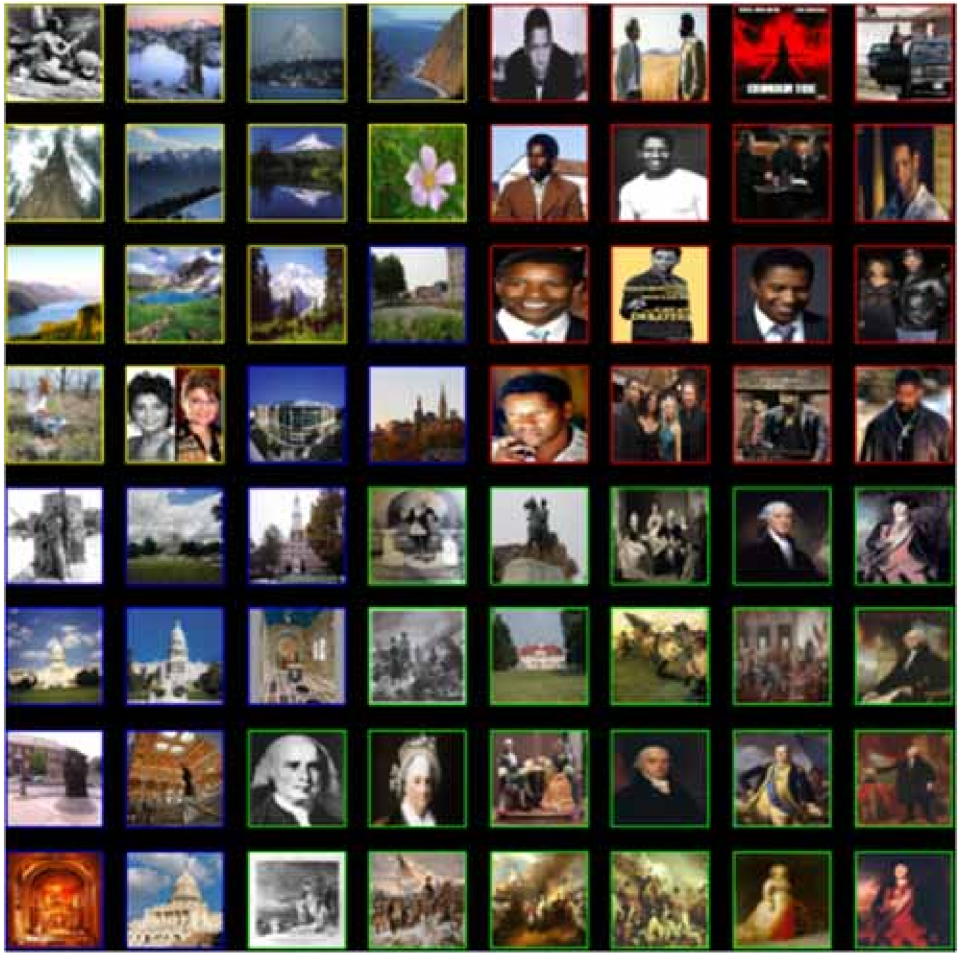
4th stage



5th stage

User Intervention by Fix Items

Users can specify one or more fixed items to influence the layout.

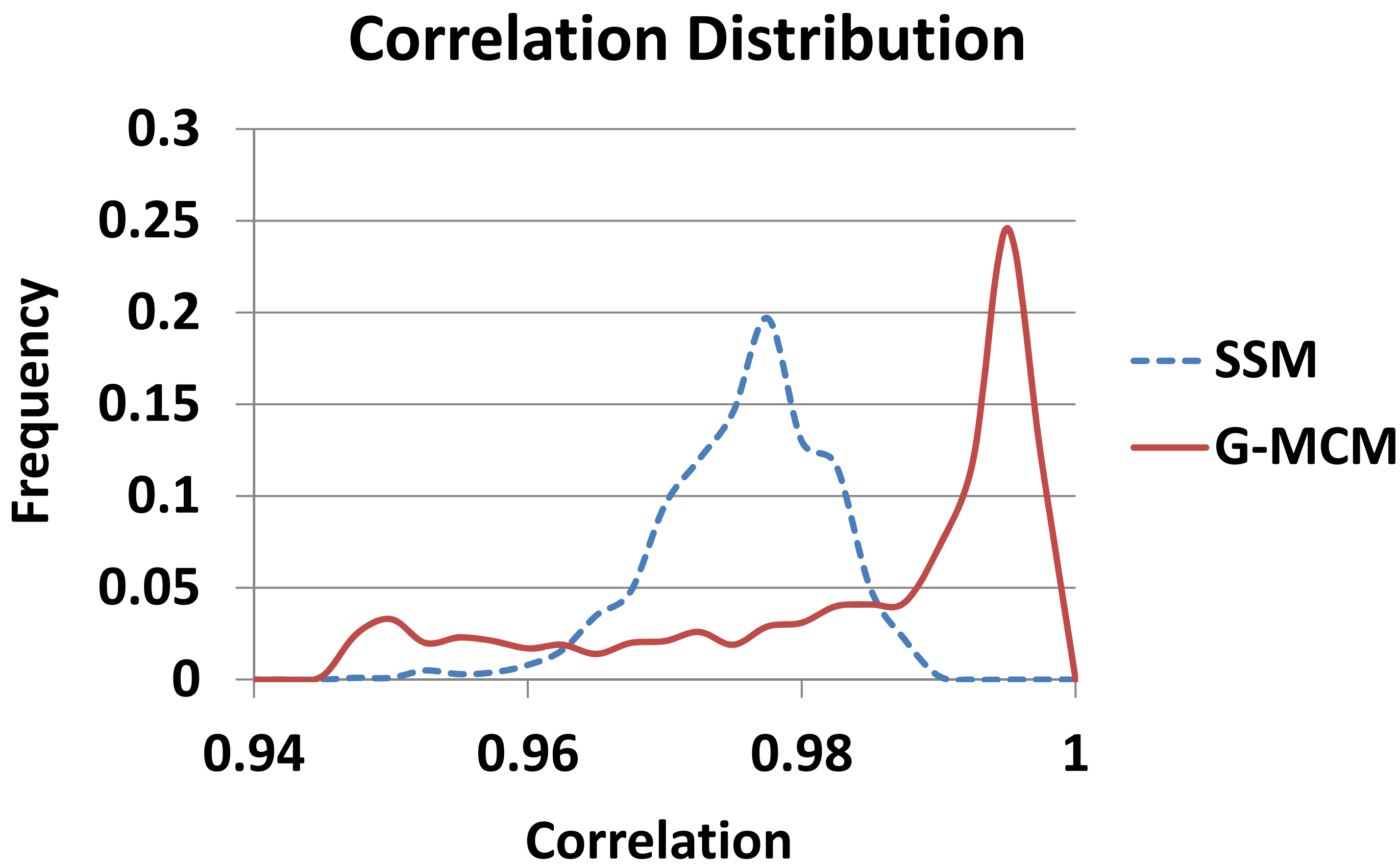


$T = 0.104s, C = 0.685$

$T = 0.108s, C = 0.640$

MCM with Images

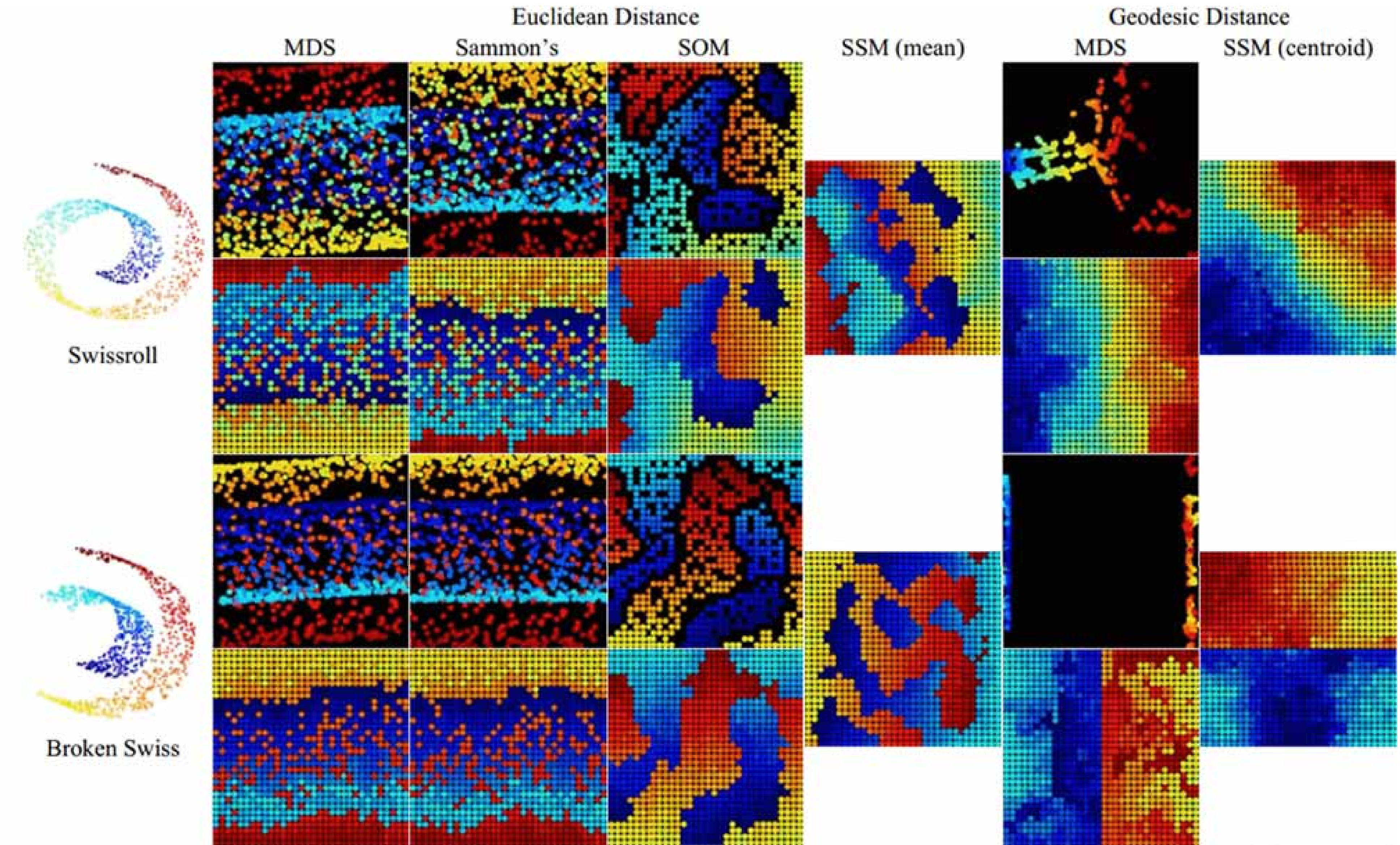
Higher correlation score than SSM; allow user-specified fixed items.



Effect of Randomness

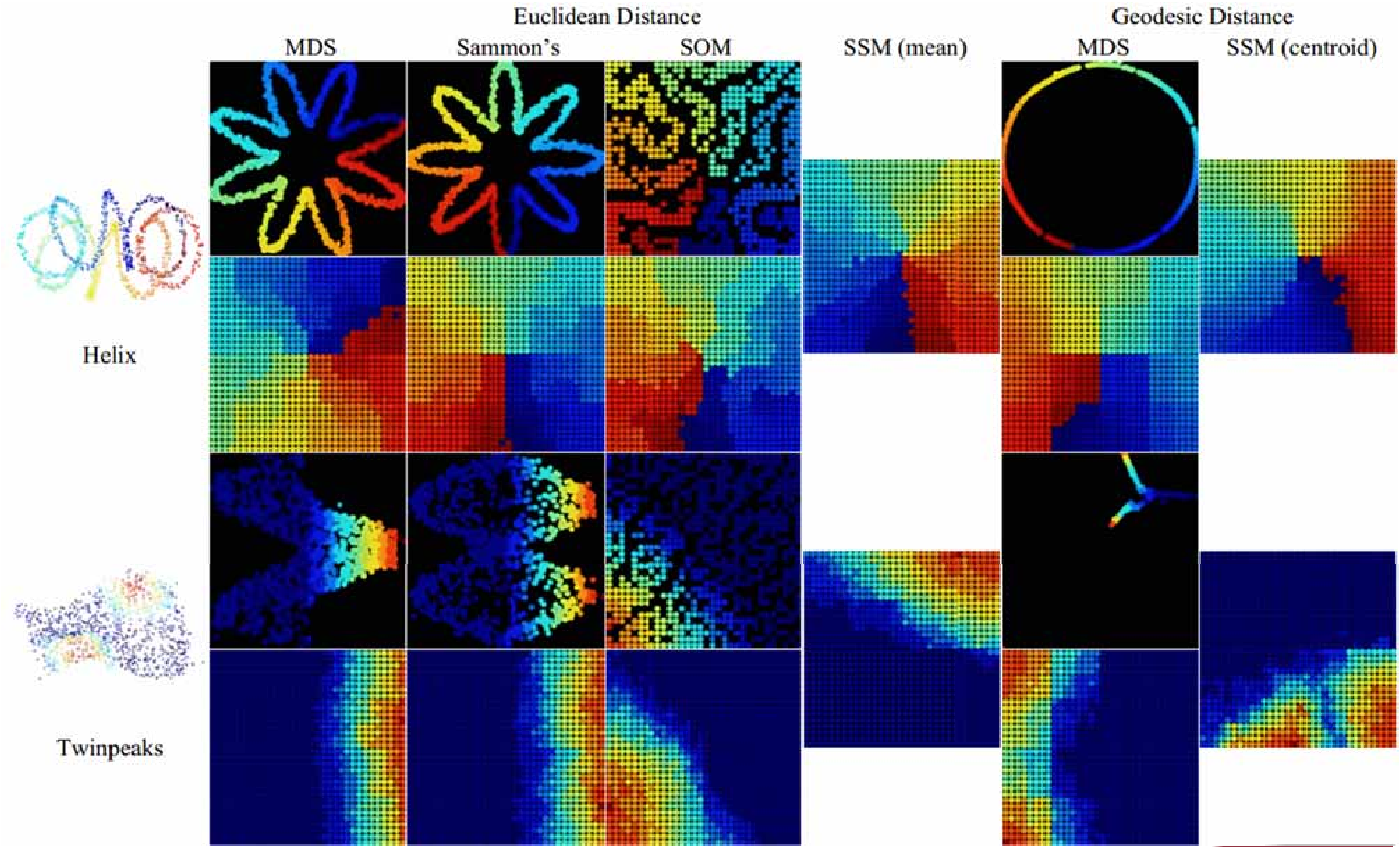
MCM has overall smaller variation.

RESULTS AND APPLICATIONS



Artificial Data

SSM works on both Euclidean and Geodesic distances.

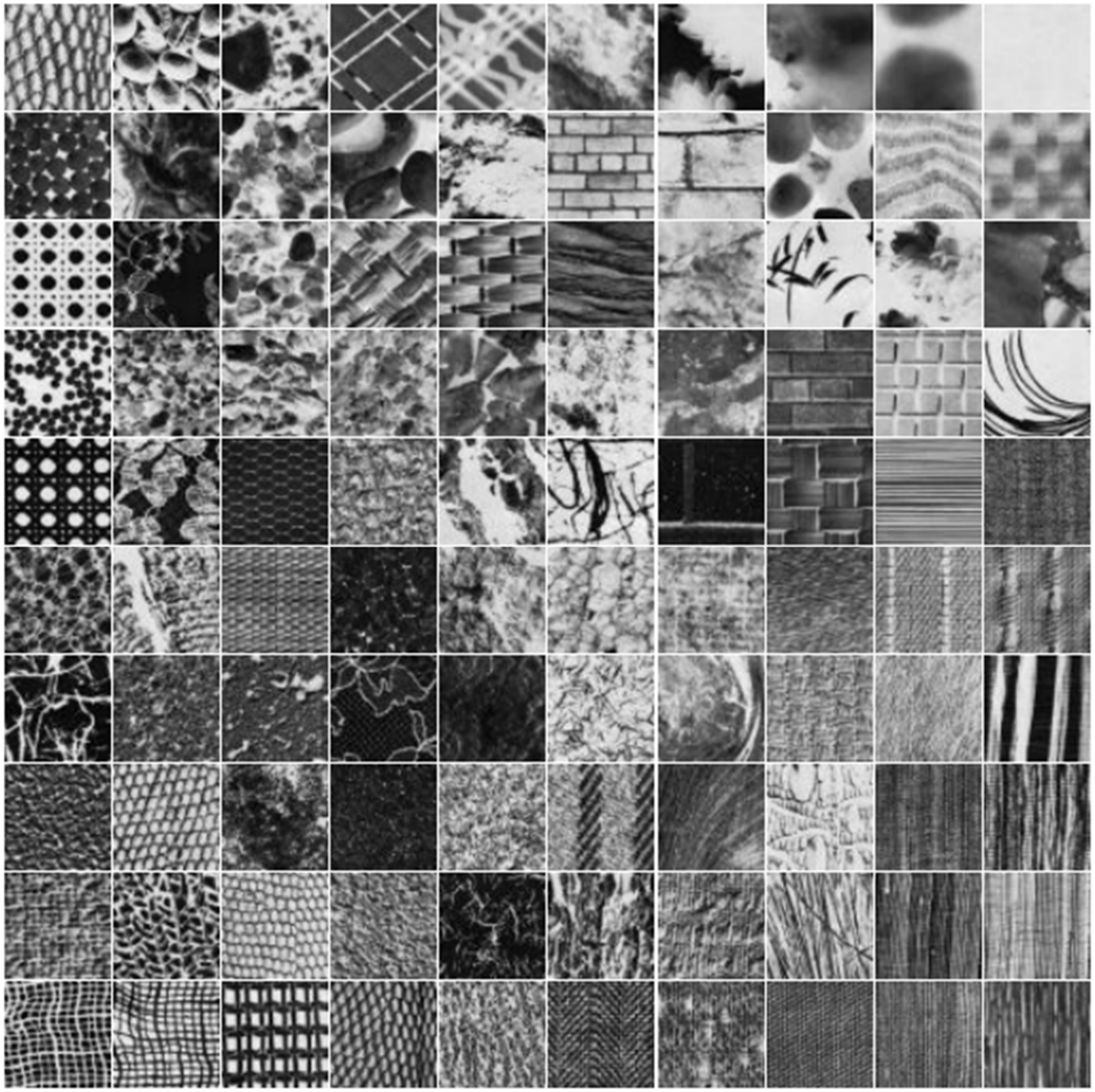


Artificial Data (Cont'd)

SSM (mean) produces good result for helix; SSM (centroid) splits twinpeaks well.

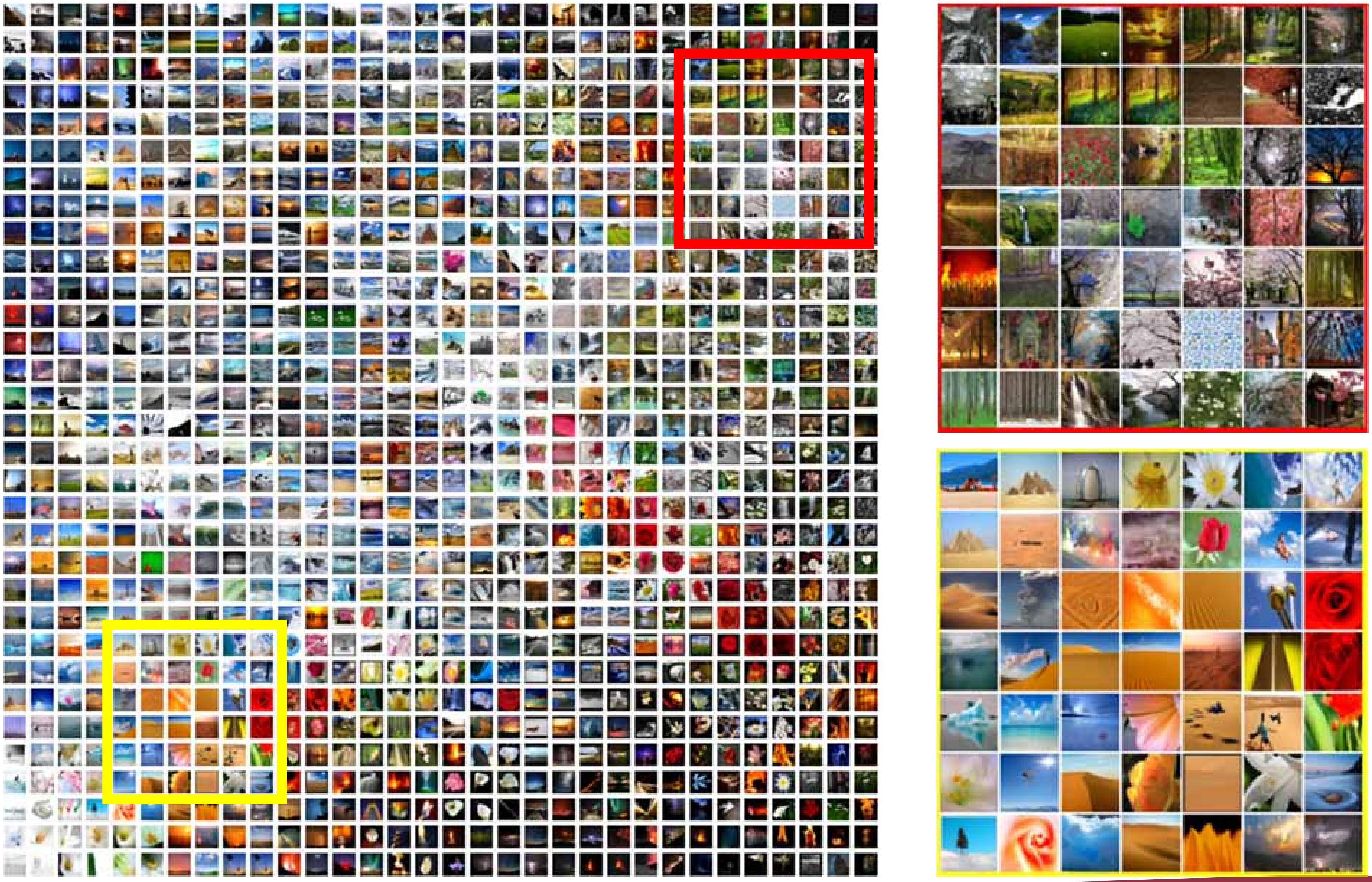
64 Eiffel Tower images

Organized based on 64-dimensional histogram vectors.



100 Texture Images

Organized using GIST descriptors.



1024 Images from Flickr

Organized using a combination of gist and histogram vectors.

	TD	TGD	CF	FQC	DW	OT	IM	TBC	RC	GH	MV	MW	BF	GW	JM	CWP	AR	VF	FB	WC	USC	WM	WH	VC	GU	GWU	HU	WDC	Sea	MR	Ore	Mon	CR	ONP	CRi	WS
Training Day	1	0.71	0.7	0.58	0.62	0.68	0.67	0.62	0.67	0.64	0.54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.32	0.1	0.09	0	0	0	0	0	0	0
The Great Debaters	0.71	1	0.62	0.68	0.75	0.74	0.71	0.66	0.52	0.47	0.52	0	0	0	0	0	0	0	0	0	0	0	0	0	0.31	0.22	0.45	0.19	0	0	0	0	0	0	0	0
Cry Freedom	0.7	0.62	1	0.64	0.68	0.64	0.59	0	0.5	0.62	0.5	0	0	0	0	0	0	0	0	0	0.36	0	0	0	0	0	0	0.15	0	0	0	0	0	0	0	0
For Queen and Country	0.58	0.68	0.64	1	0.67	0.73	0.63	0	0.48	0	0.54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Denzel Washington	0.62	0.75	0.68	0.67	1	0.67	0.67	0.71	0.55	0	0.51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Out of Time	0.68	0.74	0.64	0.73	0.67	1	0.72	0.67	0.48	0	0.54	0	0.38	0	0	0	0	0	0	0	0	0	0.29	0	0.35	0	0	0	0.22	0	0	0	0	0	0	0
Inside Man	0.67	0.71	0.59	0.63	0.67	0.72	1	0.75	0.58	0.44	0.49	0	0	0	0	0	0.23	0	0	0	0	0	0.21	0	0	0	0	0	0.13	0	0	0	0	0	0	0
The Bone Collector	0.62	0.66	0	0	0.71	0.67	0.75	1	0.45	0.46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Russell Crowe	0.67	0.52	0.5	0.48	0.55	0.48	0.58	0.45	1	0.64	0	0	0.17	0	0	0	0	0	0	0	0.22	0.35	0.06	0	0.13	0.02	0	0.21	0.09	0	0.11	0.07	0	0	0	0
Gene Hackman	0.64	0.47	0.62	0	0	0	0.44	0.46	0.64	1	0	0	0.26	0	0	0	0	0	0	0	0.31	0.36	0.32	0	0.27	0.11	0	0.22	0.06	0	0.15	0	0	0	0	0
Mount Vernon	0.54	0.52	0.5	0.54	0.51	0.54	0.49	0	0	0	1	0.78	0.51	0.6	0.59	0	0.43	0.63	0.7	0	0.58	0.58	0.45	0.46	0.34	0	0.45	0.3	0.45	0.52	0.05	0.47	0.53	0	0.42	0.58
Martha Washington	0	0	0	0	0	0	0	0	0	0	0.78	1	0.54	0.69	0.61	0.53	0.46	0.64	0.57	0	0.54	0.44	0.43	0	0	0	0	0.27	0.08	0	0.16	0	0.31	0	0	0
Benjamin Franklin	0	0	0	0	0	0.38	0	0	0.17	0.26	0.51	0.54	1	0.57	0.65	0.48	0.62	0.56	0.37	0.41	0.47	0.29	0.31	0	0.28	0.25	0.18	0.38	0.14	0	0.15	0.22	0.09	0	0.29	0
George Washington	0	0	0	0	0	0	0	0	0	0	0.6	0.69	0.57	1	0.54	0.74	0.49	0.48	0	0	0.55	0	0.29	0	0	0.15	0	0.33	0.1	0	0.19	0.19	0.33	0	0	0
James Madison	0	0	0	0	0	0	0	0	0	0	0.59	0.61	0.65	0.54	1	0.46	0.58	0.54	0	0	0.6	0.49	0.48	0.51	0.51	0.36	0.54	0.5	0.11	0	0.57	0.7	0	0	0.34	0
Charles Willson Peale	0	0	0	0	0	0	0	0	0	0	0	0.53	0.48	0.74	0.46	1	0.45	0.6	0	0	0.35	0	0.28	0	0	0.18	0	0.14	0	0	0	0	0	0	0	0
American Revolution	0	0	0	0	0	0	0.23	0	0	0	0.43	0.46	0.62	0.49	0.58	0.45	1	0.54	0.39	0	0.41	0.32	0.35	0.19	0.26	0.24	0.24	0.44	0.05	0.11	0.18	0.19	0.07	0	0.15	0
Valley Forge	0	0	0	0	0	0	0	0	0	0	0.63	0.64	0.56	0.48	0.54	0.6	0.54	1	0.54	0.57	0.49	0.59	0.34	0	0	0	0.28	0.35	0	0.48	0.06	0.21	0	0.55	0	0.24
Fort Le Boeuf	0	0	0	0	0	0	0	0	0	0	0.7	0.57	0.37	0	0	0	0.39	0.54	1	0	0	0.52	0	0	0	0	0	0	0	0	0.22	0	0	0	0	0
Washington Circle	0	0	0	0	0	0	0	0	0	0	0	0	0.41	0	0	0	0	0.57	0	1	0.58	0.6	0.44	0	0.44	0.3	0	0.32	0	0	0	0	0	0.6	0	0
United States Capitol	0	0	0.36	0	0	0	0	0	0.22	0.31	0.58	0.54	0.47	0.55	0.6	0.35	0.41	0.49	0	0.58	1	0.69	0.61	0.5	0.44	0.17	0.41	0.52	0.27	0.41	0.31	0.36	0.15	0.31	0.25	0.3
Washington Monument	0	0	0	0	0	0	0	0	0.35	0.36	0.58	0.44	0.29	0	0.49	0	0.32	0.59	0.52	0.6	0.69	1	0.51	0.58	0.33	0.15	0.34	0.41	0.28	0.52	0.26	0.31	0	0.51	0.31	0.34
White House	0	0	0	0	0	0.29	0.21	0	0.06	0.32	0.45	0.43	0.31	0.29	0.48	0.28	0.35	0.34	0	0.44	0.61	0.51	1	0.33	0.45	0.14	0.41	0.56	0.27	0.17	0.24	0.25	0	0.17	0.11	0.25
Verizon Center	0	0	0	0	0	0	0	0	0	0	0.46	0	0	0	0.51	0	0.19	0	0	0	0.5	0.58	0.33	1	0.54	0	0.46	0.21	0.6	0	0.24	0.48	0	0	0	0.42
Georgetown University	0	0.31	0	0	0	0.35	0	0	0.13	0.27	0.34	0	0.28	0	0.51	0	0.26	0	0	0.44	0.44	0.33	0.45	0.54	1	0.34	0.52	0.36	0.42	0.46	0.45	0.47	0	0	0	0.14
George Washington U	0	0.22	0	0	0	0	0	0	0.02	0.11	0	0	0.25	0.15	0.17	0.18	0.24	0	0	0.3	0.17	0.15	0.14	0	0.34	1	0.24	0.21	0.14	0.14	0.07	0.1	0.1	0	0	0.19
Howard University	0.32	0.45	0	0	0	0	0	0	0	0	0.45	0	0.18	0	0.54	0	0.24	0.28	0	0	0.41	0.34	0.41	0.46	0.52	0.24	1	0.37	0.12	0.4	0.43	0.51	0	0	0	0.1
Washington DC	0.1	0.19	0.15	0	0	0	0	0	0.21	0.22	0.3	0.27	0.38	0.33	0.5	0.14	0.44	0.35	0	0.32	0.52	0.41	0.56	0.21	0.36	0.21	0.37	1	0.27	0.24	0.43	0.41	0.31	0.23	0.35	0.22
Seattle	0.09	0	0	0	0	0.22	0.13	0	0.09	0.06	0.45	0.08	0.14	0.1	0.11	0	0.05	0	0	0	0.27	0.28	0.27	0.6	0.42	0.32	0.12	0.27	1	0.47	0.45	0.41	0.46	0.34	0.42	0.61
Mount Rainier	0	0	0	0	0	0	0	0	0	0	0.52	0	0	0	0	0	0.11	0.48	0	0	0.41	0.52	0.17	0	0.46	0.14	0.4	0.24	0.47	1	0.43	0.39	0.7	0.63	0.56	0.61
Oregon	0	0	0	0	0	0	0	0	0.11	0.15	0.05	0.16	0.15	0.19	0.57	0	0.18	0.06	0.22	0	0.31	0.26	0.24	0.24	0.45	0.11	0.43	0.43	0.45	0.43	1	0.6	0.62	0.38	0.62	0.42
Montana	0	0	0	0	0	0	0	0	0.07	0	0.47	0	0.22	0.19	0.7	0	0.19	0.21	0	0	0.36	0.31	0.25	0.48	0.47	0.1	0.51	0.41	0.41	0.39	0.6	1	0.48	0.43	0.52	0.41
Cascade Range	0	0	0	0	0	0	0	0	0	0	0.53	0.31	0.09	0.33	0	0	0.07	0	0	0	0.15	0	0	0	0	0.1	0	0.31	0.46	0.7	0.62	0.48	1	0.54	0.74	0.56
Olympic National Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.55	0	0.6	0.31	0.51	0.17	0	0	0	0	0.23	0.34	0.63	0.38	0.43	0.54	1	0.52	0.38
Columbia River	0	0	0	0	0	0	0	0	0	0	0.42	0	0.29	0	0.34	0	0.15	0	0	0	0.25	0.31	0.11	0	0	0	0	0.35	0.42	0.56	0.62	0.52	0.74	0.52	1	0.55
Washington State	0	0	0	0	0	0	0	0	0	0	0.58	0	0	0	0	0	0	0.24	0	0	0.3	0.34	0.25	0.42	0.14	0.19	0.1	0.22	0.61	0.61	0.42	0.41	0.56	0.38	0.55	1

Semantic Distance between 36 Concepts

Measured using Wikipedia Link Measure.

George Washington U.	Howard University	Verizon Center	Oregon	Olympic National Park	Cascade Range
Washington Circle	Georgetown University	Washington DC	Montana	Mount Rainier	Washington State
American Revolution	James Madison	United States Capitol	Washington Monument	Seattle	Columbia River
Benjamin Franklin	Valley Forge	White House	Mount Vernon	Russell Crowe	Gene Hackman
George Washington	Martha Washington	Out of Time	The Great Debaters	Inside Man	Training Day
Fort Le Boeuf	Charles Willson Peale	For Queen and Country	Cry Freedom	Denzel Washington	The Bone Collector

Organization Result

More intuitive visualization of similarities between articles.

Global Cities

With size indicating the importance.

Conclusions and Future Work

- The problem of placing data into structured layout is studied.
  - Three algorithms are incrementally developed.
- The techniques visualize the data and their relations in an intuitive way.
  - Especially useful when user interaction is unavailable (public billboard) or undesirable (mobile device).

Contributions

North America Weather Map

North America Weather Conditions

Calgary	3°	Mexico City	7°
Chicago	-4°	Monterrey	15°
Ciudad Juarez	12°	Montreal	-12°
Ciudad Nezahualcoyotl	10°	New York	-5°
Dallas	14°	Phoenix	15°
Detroit	-7°	Puebla	8°
Guadalajara	10°	Tijuana	9°
Houston	15°	Toronto	-10°
Leon	7°	Vancouver	6°
Los Angeles	12°	Zapopan	11°

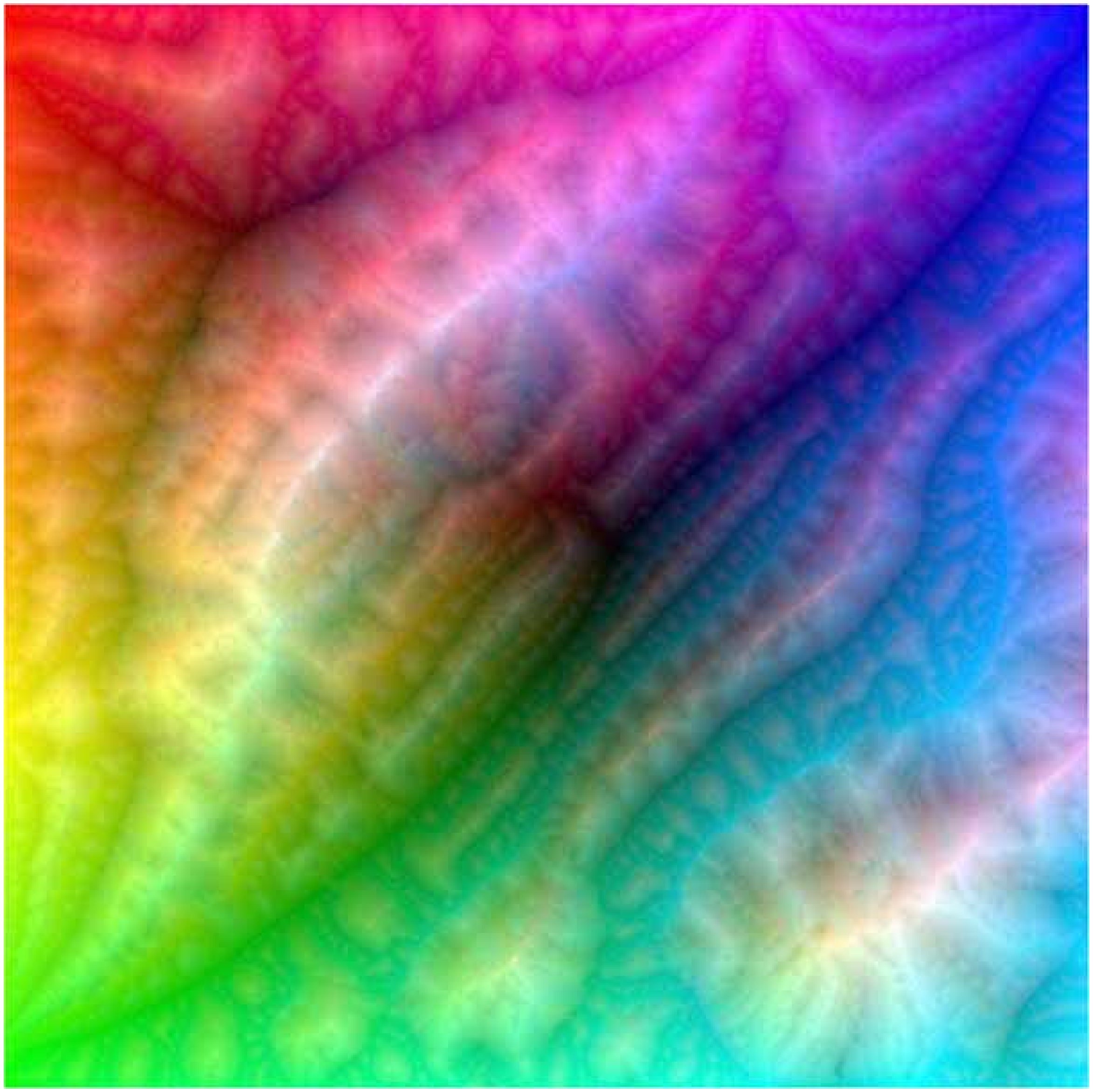
Vancouver	6°	Calgary	3°	Chicago	-4°	Toronto	-10°	Montreal	-12°
Los Angeles	12°	Phoenix	15°	Dallas	14°	Detroit	-7°	New York	-5°
Tijuana	9°	Ciudad Juarez	12°	Monterrey	15°	Mexico City	7°	Houston	15°
Guadalajara	10°	Zapopan	11°	Leon	7°	Ciudad Nezahualcoyotl	10°	Puebla	8°

User Study

Which presentation can best facilitate information refinding?

Photo Collage

How to embed characters into collage?



Art?

Organization for 1 million random colors

QUESTIONS?

Thanks for Your Interests!

References

- G. Strong & M. Gong: Self-sorting map: An efficient algorithm for presenting multimedia data in structured layout, submitted to *IEEE Trans. Multimedia*.
- G. Strong: Arranging arbitrary data into structured layouts, Ph.D. Dissertation, Memorial University of Newfoundland, St. John's, NL, Canada, 2013.
- E. Hoque, O. Hoeber, G. Strong, & M. Gong: Combining conceptual query expansion and visual search results exploration for web image retrieval, *J. Amb. Intell. Hum. Comput.(in press), 2013.*
- G. Strong & M. Gong: Data organization and visualization using self-sorting map, *Proc. Graphics Interface: 199-206, 2011.*
- G. Strong & M. Gong: Similarity-based image organization and browsing using multi-resolution self organizing map, *Image Vis. Comput. 29(11): 774-786, 2011.*
- E. Hoque, G. Strong, O. Hoeber, & M. Gong: Conceptual query expansion and visual search results exploration for web image retrieval, *Adv. Intell. Web Mastering: 73-82, 2011.*
- G. Strong, E. Hoque, M. Gong, & O. Hoeber: Organizing and browsing image search results based on conceptual and visual similarities, *Adv. Vis. Comput.: 481-490, 2010.*
- G. Strong, O. Hoeber, & M. Gong: Visual image browsing and exploration (vibe): User evaluations of image search tasks, *Active Media Tech.: 424-435, 2010.*
- G. Strong & M. Gong: Organizing and browsing photos using different feature vectors and their evaluations, *Proc. Int. Conf. Image Video Retrieval: 1-8, 2009.*
- G. Strong: Similarity-based image organization and browsing, M.Sc. Thesis, Memorial University of Newfoundland, St. John's, NL, Canada, 2009.
- G. Strong & M. Gong: Browsing a large collection of community photos based on similarity on GPU, *Adv. Vis. Comput.: 390-399, 2008.*

- **The goal of this work is to organize data into structured layouts such that proximity reflects similarity. Each data item can be as simple as a binary bit or as complex as a structure containing multiple data types. Different data items may or may not be comparable for ordering, but are assumed to have dissimilarities among them well-defined. The organization problem is formulated as maximizing the correlation between the dissimilarities among the data items and their placement distances in the structured layouts. The organization results are meant to provide simple and occlusion free presentations for the data, which facilitate users to understand the underlying relationship among the data or to refind the desired data item at a later time.**
- **Three algorithms for the above problem are incrementally developed. The first extends upon the Self-Organizing Map algorithm to make the generated layouts occlusion free. It performs admirably but requires the input data to be represented as vectors. The second algorithm, named as the Self-Sorting Map approaches the problem from a sorting perspective and generates occlusion free layouts for all types of data directly. It is fast, scalable, and implementable on Graphics Processing Units. Instead of minimizing dissimilarities among data items locally like the first two approaches, the third one maximizes the global correlation between data dissimilarities and placement distances directly. Entitled the Max Correlation Map, it is simpler in concept, yet more robust than its counterparts. Creative organizations of images, articles, and world cities show the practical applications of organizing data into structured layouts.**

Abstract