Interactive Visual Analysis of Complex Data 2: Families of Curves, Families of Surfaces

Krešimir Matković
Vienna University of Technology, VRVis Research Center, Vienna
Overview

Coordinated Multiple Views

Complex data: families of curves

- Traffic Surveillance
- Engineering
- Bio-Sygnals and ICU Data
  - Heart reinervation
  - IVA of ICU Data
- Animal Paths
- Trajectories
- Image Collection
The Domain - Ethology

Scientific study of animal behavior

- neural mechanisms of learning processes
- physiological processes
- influence of drugs
- ...

Types of ethological studies

- open field (OF)
- open field with object placement task (OPT)
- mazes
- ...

Krešimir Matković  InfoVis, TU Wien, 15. 04. 2016
The Study - Object Placement Task

Observation I
- open field (120cm x 120cm) with two identical objects
- observation time: 5 minutes

Observation II
- one object of observation I setup is displaced
- observation time: 5 minutes

Evaluation
- spatial reference memory
- locomotor activity, anxiety related behavior
Data Acquisition

Video Tracking System
- records path of animal
- parameters:
  - total distance
  - average speed
  - time spent in area X
  - field crossings
  - turn arounds
  - ...
Entry: Interactive Visual Analysis

Conventional approach
- statistical evaluation
- a single path can be examined at a time

New approach
- compare tracks of all animals (whole ensemble) interactively
  => integrate the path itself in the analysis
- coordinated multiple views (CMV)
Coordinated Multiple Views

- scatter-plot, histogram, parallel coordinates, ...
- curve view
- line brush
- linking and brushing
- composite brushing
- focus + context
Coordinated Multiple Views

- scatter-plot, histogram, parallel coordinates, ...
- curve view
- line brush
- linking and brushing
- composite brushing
- focus + context
Coordinated Multiple Views

- scatter-plot, histogram, parallel coordinates, ...
- curve view
- line brush
- linking and brush
- composite brush
- focus + context
Coordinated Multiple Views

- scatter-plot, histogram, parallel coordinates, ...
- curve view
- line brush
- linking and brushing
- composite brushing
- focus + context
Data Model

Multivariate, multidimensional data

- a record contains
  - animal parameters: ID, age, ...
  - aggregated path parameters: total path length, average speed, time spent in a specific area, ...
  - time series parameters (Konyha et. al): path, distance traveled

- number of records: 814
Interactive Analysis
Validating evaluation parameters
Interactive Analysis
Identifying Patterns in Paths
Interactive Analysis
Identifying Patterns in Paths

Krešimir Matković
InfoVis, TU Wien, 15. 04. 2016
Interactive Analysis
Identifying Patterns in Paths
T-Maze 1
T-Maze 2
Overview

Coordinated Multiple Views

Complex data: families of curves

- Traffic Surveillance
- Engineering
- Bio-Sygnals and ICU Data
  - Heart reinervetion
  - IVA of ICU Data
- Animal Paths
- Trajectories
- Image Collection
VAST challenge 2015

**INPUT**
Row data: movements and check-ins, more than 25 millions of entries:
- 2014-6-06 08:00:59, 903769, check-in, 99.77
- 2014-6-06 08:00:59, 1696241, check-in, 99.77
- 2014-6-06 08:01:03, 531318, movement, 64.97
- 2014-6-06 08:01:04, 179396, movement, 64.97

Park map: Time threshold for movement data: 20 sec

**PROCESS**
Semi-automatic Data processing:
1. Match coordinates to the facilities using the map
2. Create a record for each unique ID containing:
   a. events and trajectories
   b. basic scalars: rides count, moves count,...
   c. time series: distance travelled,...

**FINAL DATA**
Final data (2 out of 11374 rows):

<table>
<thead>
<tr>
<th>USER ID</th>
<th>TRAJECTORY</th>
<th>RIDES COUNT</th>
<th>MOVE COUNT</th>
<th>DISTANCE (T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>scalar</td>
<td>complex</td>
<td>scalar</td>
<td>scalar</td>
<td>complex</td>
</tr>
<tr>
<td>903769</td>
<td>30</td>
<td>1345</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1696241</td>
<td>42</td>
<td>1987</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
VAST challenge 2015
VAST challenge 2015
1. Visitors who visited the park on Friday and Saturday, but left early on Saturday. One of them moved very little, let's drill down further.
VAST challenge 2015

2. Details on demand reveal that individual (ID 657863) entered the park on Friday at 16:11:36 and left it on Saturday at 08:10:03. The individual stayed over night! His activity graph shows enter, movement, one very long (overnight) ride.
3. Let’s check the map! The individual walked directly to the Alvarez Beer Gardens (33), entered it at 17:04:28 and stayed there until 21:50:50. It is very likely that after drinking beer for over 4 hours, the individual was more or less drunk.
4. After leaving the Beer Garden, the visitor checked in to Sholtz Express ride (20) at 21:53:11, and did not move until 08:00:27 on Saturday, then. It is plausible to reason that the individual was drunk and found a place in the ride area to sleep.
VAST challenge 2015

Epilogue. The individual moved again on Saturday morning at 08:00:27, probably awaken by the noise of the park opening at 08:00:00. The individual then quickly exited the park at 08:10:03 on Saturday.
Overview

Coordinated Multiple Views

Complex data: families of curves

- Traffic Surveillance
- Engineering
- Bio-Sygnals and ICU Data
  - Heart reinervation
  - IVA of ICU Data
- Animal Paths
- Image Collection
Large Image Collections - Comprehension and Familiarization by Interactive Visual Analysis

Krešimir Matković, D. Gračanin, W. Freiler, J. Banova, and H. Hauser

SmartGraphics 2009
Motivation

Large internet image collections – new phenomena

Already used in research (3D reconstruction, e.g.)

We want to understand collection itself by means of visual analysis

Many potential uses: marketing, sociology, etc.
Data Set

Flicker, more than 3000 images uploaded per minute

>5 billion images as of November 2011!

API available

26 000 + images downloaded

Parameters downloaded & computed
Parameters

External

- 6 image parameters: date taken, number of comments, number of views,…
- 10 user parameters: name, gender, singleness, home town,…

Derived

- More than 60: hue, saturation, lightness, colorfulness, number of frontal faces,…
- Scalar + series (curves)

We use coordinated multiple views for analysis
Derived Parameter - Lightness
Views for families of curves

Curve view

Segmented curve view
Derived Parameter - Lightness

a.

b.

0 1 2 3
4 5 6 7
8 9 10 11
12 13 14 15

c.

d.
Dark Corners Example
Picture View

Parameters are not enough

We cannot imagine an image from parameters
Image as a Brush

Fuzzy specification

Visual image query as basis

Any query can be used

Novel way of brush specification

Tangible brushing
A screenshot from analysis
Example 1 – highly saturated M/F
Examples – taken/single
Conclusion – Image Collection

Large image collections offer unlimited opportunities for analysis

New phenomena which will certainly be exploited in the future

Interactive visual analysis can help the user to understand image collection

Image view and image as a brush are necessary when dealing with images
Overview Families of Surfaces

Complex data: families of surfaces

- Meteorology
- Engineering
- Car racing
Overview

Complex data: families of surfaces

- Meteorology
- Engineering
- Car racing

InfoVis wrap-up
Meteorology model 1

Climate research at the PIK institute

Lake Agassiz (North America)
- 4000 years lifespan
- 8000 years ago lake drained due to climate warming

Simulation of cooling of 3.6 K of North Atlantic due to melt water pulse
Meteorology model 2

| Time step | Diff_h | Diff_v | output 1 | output2 | ...
|-----------|--------|--------|----------|---------|--------
| 1         | Diff_h | Diff_w |          |         |        |
| 2         | Diff_h | Diff_w |          |         |        |
| 3         | Diff_h | Diff_w |          |         |        |
| 4         | …      | …      | …        | …       | …      |
| …         | …      | …      | …        | …       | …      |
Multiple runs

Multiple runs:

- two diffusivity parameters
  - 10 steps each - 100 runs
- 500 time steps (1 step = 1 year) – per run

35 different results aggregated from the more detailed raw simulation data (temperature, …)

Common way of storing the data:

- multiply previous table 100 times, 50 000 rows now
- 3 independent variables now timestep, diffh, diffv
<table>
<thead>
<tr>
<th>Time</th>
<th>Diff_h</th>
<th>Diff_v</th>
<th>output1</th>
<th>output2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>dh1</td>
<td>dv1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>dh1</td>
<td>dv1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>dh1</td>
<td>dv1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>dh1</td>
<td>dv2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>dh1</td>
<td>dv2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>dh1</td>
<td>dv2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>dh1</td>
<td>dv3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>dh1</td>
<td>dv3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>dh1</td>
<td>dv3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multiple runs - exploration

-4.19  24.50  2.77

C0  C1

-5.45 Greenland temp.  24.01 Tropical temp.  2.70 Global preci.

9.5E-5  7.5E-5

1250  Diff_V  2500  Diff_H
Mappings as dimensions - curves

Mapping as a dimension [Konyha et al. 2006]:

- group results for one run
- outputs are not scalar any more
- but functions of time
- 100 rows now

<table>
<thead>
<tr>
<th>Run</th>
<th>Diff_h</th>
<th>Diff_v</th>
<th>Output1(t)</th>
<th>Output2(t)</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>dh1</td>
<td>dv1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>dh1</td>
<td>dv2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>dh1</td>
<td>dv3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>dh10</td>
<td>dv10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Family of curves
Mappings as dimensions - surfaces

Curves – we are interested about output as \( f(\text{time}) \)

What if we are interested in outputs as \( f(x,y) \)?

Group outputs from all runs in a “surface”
- e.g. temperature = \( f(\text{diffh, diffv}) \)

500 rows now, one for each time step

Offers new possibilities for analysis

<table>
<thead>
<tr>
<th>Timestep</th>
<th>Output1(diffh, diffv)</th>
<th>Output2(diffh, diffv)</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td></td>
</tr>
</tbody>
</table>
Visualization of family of surfaces

Family of curves
- curve view shows them all
- works great

Family of surfaces
- show them all!
- it does not work!

We can show single surface but

Cannot show whole family at once

If we cannot see it we will grasp it thorough interaction
Analysis of family of surfaces

Look at the data at different levels

Three levels of detail:

- analysis through single scalar aggregates (overview)
  - reduction by 2 dimensions
- analysis through aggregated profiles (drill down)
  - reduction by 1 dimensions
- analysis through surfaces (details on demand)
  - no dimension reduction but selective visualization

level 1
level 2
level 3
Single Scalar Aggregates

Overview, statistics

- first step
- reduce surface to minimum, maximum, median, ...
- use common views to explore the data
- get the first impression
- explore the family at the first level
- family – not single surface
Single Scalar Aggregates

Overview, statistics:

- first step
- reduce surface to minimum, maximum, median, ...
- use common views to explore the data
- get the first impression
- explore the family at the first level
- family – not single surface

level 1

1. Year
2. MinGreen|Temp
3. AvgGreen|Temp
4. MaxGreen|Temp
5. SpanGreen|Temp
6. MedGreen|Temp

n,...
Aggregated Profiles

General concept in data exploration – projection
- reduce dimensionality if you cannot show it

Profiles
- reduction operators
- project the surface to the plane
- min, max, median, average,... profiles

Analysis through aggregated profiles (drill down)
- advanced analysis
- very complex, challenging, and powerful
Aggregated Profiles

General
- reduce dimensionality if you cannot show it

Profiles
- reduction operators
- project the surface to the plane
- min, max, median, average,… profiles

Analysis
- advanced analysis
- very complex, challenging, and powerful
Aggregated Profiles

General concept in data exploration – projection

- reduce dimensionality if you cannot show it

Profiles

- reduction operators
- project the surface to the plane
- min, max, median, average,… profiles

Analysis through aggregated profiles (drill down)

- advanced analysis
- very complex, challenging, and powerful
Surfaces

Analysis through surfaces (details on demand)

- no dimension reduction but selective visualization
- functions for few surface only
Surfaces

- no dimension reduction but selective visualization
- functions for few surfaces only
Overview

Complex data: families of surfaces
- Meteorology
- Engineering
- Car racing

InfoVis wrap-up
EHD bearing

Important part in IC engine design

Durability, performance, wear, noise

Two step analysis

225 simulation runs – 4 to 5 days on a standard PC
Analysis snapshot 1

level 2  level 2  level 2  level 2
Analysis snapshot 2

level 2  level 2  level 3  level 3
Additional Views

Lighting Example
Additional Views
Overview

Complex data: families of surfaces

- Meteorology
- Engineering
- Car racing

InfoVis wrap-up
Race Car Setup

The Open Car Race Simulator - TORCS

Parameters varied

- Wing Angle: 8–18 degrees (six steps)
- Brake Ratio: 0.41–0.50 (five steps)
- Max. Brake Pressure: 6,500–19,000 (six steps)
- Gear Sets: Six different gearbox sets (six steps)
- Front Spring: 1,000–2,000 (three steps)
- Rear Spring: 1,000–2,000 (three steps)

Telemetry data saved
## Single run

<table>
<thead>
<tr>
<th>Lap</th>
<th>Sector</th>
<th>Wing Angle</th>
<th>Brake Ratio</th>
<th>...</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
<td>0.41</td>
<td></td>
<td>120</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>8</td>
<td>0.41</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>8</td>
<td>0.41</td>
<td></td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>23</td>
<td>8</td>
<td>0.41</td>
<td></td>
<td>89</td>
</tr>
</tbody>
</table>
Multiple runs

Multiple runs:
- >9000 variations of car parameters
- 20 laps per simulation, 23 sectors

Telemetry data stored

Common way of storing the data:
- multiply previous table 9000 times
- 7 independent variables now lap, sector, car parameters
<table>
<thead>
<tr>
<th>Lap</th>
<th>Sector</th>
<th>Wing Angle</th>
<th>Break Ratio</th>
<th>...</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>23</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>23</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>23</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>23</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mappings as dimensions - surfaces

Group outputs from all runs in a “surface”
- e.g. \( \text{speed} = f(\text{lap, sector}) \)

9000 rows now, one for each simulation run

Offers new possibilities for analysis

<table>
<thead>
<tr>
<th>Run</th>
<th>Wing Angle</th>
<th>Speed(lap,sector)</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td><img src="image1" alt="Surface" /></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td><img src="image2" alt="Surface" /></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td><img src="image3" alt="Surface" /></td>
<td></td>
</tr>
</tbody>
</table>
Analysis of family of surfaces

Look at the data at different levels

Three levels of detail:
1. Analysis through single scalar aggregates (overview)
2. Reduction by 2 dimensions
3. Analysis through aggregated profiles (drill down)
4. Reduction by 1 dimensions
5. Analysis through surfaces (details on demand)
6. No dimension reduction but selective visualization
Analysis of family of surfaces
Analysis of family of surfaces
Family of surfaces - conclusion

- (Almost) Everyone
  - but they are simplified because of lacking of technology

Interactive visual analysis offers new possibilities for analysis
- Preconfigured views
- 3D views
Thank you!

Special thanks for used materials to H. Hauser, and colleagues from VRVis!