Interactive Visual Analysis of Complex Data:
Introduction, Families of Curves

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

Overview 1
Introduction – summary once more
Coordinated Multiple Views – summary once more
Complex data: families of curves
- Traffic Surveillance
- Engineering
  - Interactive visual steering
  - A hybrid approach to visual steering
- Bio-Signals and ICU Data
- Heart reinervation
- IVA of ICU Data

Multidimensional Multivariate Data
Conventional approach deals with n-dim. Euclidian spaces
Each item is a point in n-dim. Space (n-tuple)

Coordinated multiple views
Show various dimensions simultaneously
Use mostly simple 2D views
Interaction:
- Linking & brushing
- Select items in one view and they will be highlighted in all views

Complex Data
Engineering data does not always fit to this data model
Typical simulation:
- N control parameters
- n simulation runs (n <= 2^N)
- M output values which can be mappings (f(x), x is time, crankangle, ...)

Point P_i is the same, but x_i can be a mapping

Mapping as a dimension 1
Such data model is common
Simulations, Meteorology, Traffic, Telecommunication, Financial data, ...
Such mappings themselves are common
Often described as 2D Euclidian space
Our approach differs from conventional
- Mapping is a dimension of the space
Mapping as a dimension 2

To there are N points $P_i$
And dimension $k$ is a mapping
Each $x_k$ is a function graph

There are N function graphs for dimension $k$

$A Family of Function Graphs$

Curve View displays such a family

Interaction

Curve view in a multiple linked view setup

Need for efficient brush
Line brush
- Selects all curves which cross the line
- Very intuitive
- Easy to define
- Would be much more complex using SQL

Overview 1

Introduction – summary once more
Coordinated Multiple Views – summary once more

Complex data: families of curves
- Traffic Surveillance
- Engineering
- Interactive visual steering
- A hybrid approach to visual steering
- Bio-Signals and ICU Data
- Heart reinervation
- IVA of ICU Data

Traffic Sensor Network
(Minneapolis, St. Pauls)
- 12 weeks (84 days)
- 564 sensors
- daily data: aggregated from measurements all 30 secs.
- 2·47376 = 94752 graphs
(144 f(t)-values each, one per 10 mins.)
Overview 1

Introduction – summary once more
Coordinated Multiple Views – summary once more
Complex data: families of curves
  - Traffic Surveillance
  - Engineering
    - Injection
    - Interactive visual steering
    - A hybrid approach to visual steering
  - Bio-Signals and ICU Data
    - Heart reinervation
    - IVA of ICU Data

First Success - Automotive Simulation

Krešimir Matković  InfoVis, TU Wien, 17. 04. 2013
Automotive Simulation

Simulation is increasingly employed
- fluid/gas flow (engine, interior, exterior, etc.)
- crash simulation, etc.

Methods
- computational fluid dynamics (CFD)
- finite element methods (FEM)
- discrete simulation

Issues
- simulation often costly (€, time)
- datasets often large, multi-dim./-variate, etc.

Questions, Goals

More information from data
- cues for future improvements
- improving expertise

Optimization
- seeing the unexpected, reading between lines
- utilizing expertise

Analyzing failures
- answering „why“/„how“-questions
- simulation debugging
- failure analysis

Special Challenges

Multi-variate data (many data columns)
- CFD simulation delivers dozens of attribute dims.
- Process simulation can deliver even more

Large datasets (many data rows)
- Simulation datasets: Gigabytes, ...

Spatiotemporal relations
- 3D space/object referenced
- time-referenced

Complex/nontrivial physical phenomena

Injection Design

Context:
- common rail Diesel technology
- optimization of Diesel injection

Simulation:
- simplified 1D simulation
- variational analysis wrt. boundary conditions (~10^4 sim. runs)

Simulation and Visualization 1

Common practice:
- define a model
- run simulation
- visualize results
- change the model
- run simulation ...

Simulation and Visualization 2

Multiple runs:
- define a model
- define control parameters
- run simulation
  - for each combination of control parameters
  - for some of the combinations (steering)
- use interactive visual analysis
  - to understand (optimize) the model
  - to understand the phenomena
Common Rail Injection

Our example
Fast simulation
Main principle
Solenoid valve controls opening and closing (injection)

Valve body goes up
P drops
High P moves needle up

Pressure P

Common Rail Diesel Injection 1st case

Control Parameters:
- R1 = inlet / outlet
- R2 = piston / needle
- R3 = inlet / piston
- Prail = common rail pressure
- SV_{open}, SV_{close} = valve open/clos. vels.

Injection Simulation

Simulation with AVL HydSim

Variations:
- 9* R1
- 9* R2
- 5* R3
- 3* Prail
- 4* SV_{open}
- 4* SV_{close}

19440 sim. runs, multi-variate res.

Simulation Results

Computed values:
- m_{inj} = inj. fuel mass
- X_{lift} = max. needle lift
- V_{open}, V_{close} = shape params.

Goals:
- enough fuel (not too much)
- early inj. (not too slow)
- etc.

Analysis: Black Box Reconstruction

Families of curves – complex data 2

One record = one run

Various attributes
- scalar - control parameters and scalar aggregates of state parameters
- time series – state variables

Considering function graphs as resulting values

| Run | Flow resistance | Closing start | Pressure(t) | Velocity(t) | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Krešimir Matković InfoVis, TU Wien, 17. 04. 2013
**Next Step: Injection Graph Analysis**

- Injection Rate Shaping
- Engine Speed

**Analysis: Parameter Influence (box)**

**Analysis: Shape „boot“**

**Electronic Unit Injector - EUI 1**

- One injector per cylinder
- Used in Diesel engines:
  - patent from 1911
  - used since 1930s in trucks, locomotives, and ships
  - EUI in 1990s
  - mostly used in heavy vehicles
- Injector design is very important
  - emission reduction
  - engine efficiency

**Electronic Unit Injector - EUI 2**

**Model (of an EUI Injector)**

- Precise control of injection – quantity and timing
- Basic Idea:
  - SCV opened and NCV closed
  - spring pushes the needle down
  - pressure P increases
    - close SCV
    - if P &gt; spring, needle goes up
    - open NCV
    - P helps the spring
- NCV controls nozzle opening/closing

- Model definition:
  - standard building blocks are used
  - each block has control parameters
  - decide which will be varied
  - state variables are computed for each block

- A spring pushes the needle down
- Pressurized fluid helps the spring
- Needle goes up
- Needle closed

Krešimir Matković  InfoVis, TU Wien, 17. 04. 2013
Multiple simulation runs

<table>
<thead>
<tr>
<th>Block</th>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>low resistance</td>
<td>1.0; 2.0; 3.0; 4.0</td>
</tr>
<tr>
<td>I</td>
<td>low resistance</td>
<td>1.0; 2.0; 3.0; 4.0</td>
</tr>
<tr>
<td>B</td>
<td>low resistance</td>
<td>1.0; 2.0; 3.0; 4.0</td>
</tr>
<tr>
<td>C</td>
<td>closing start</td>
<td>20.25; 30.35; 40.45</td>
</tr>
<tr>
<td>G</td>
<td>opening start</td>
<td>-15.20; -25</td>
</tr>
<tr>
<td>H</td>
<td>opening start</td>
<td>-15.20</td>
</tr>
</tbody>
</table>

AVL HydSim tool used
1D CFD – fast simulation
10 simulation runs per minutes
7 parameters varied – 2880 runs
9 state variables considered in analysis (+ aggregates)

Interactive Visual Analysis

Results are available:
- use interactive visual analysis
  - to understand (optimize) the model
  - to understand the phenomena
- CMV system used
- many parameters – many views
- engineer has to know what is depicted in each view

The Model View

Engineers are familiar with model view
We want to show results and control parameters in blocks
Control and state variables
Multiple runs!
  - various values for control parameters
  - various values for state variables - time series
Limited space – three levels

Three Levels

Show control parameters on left, state variables on right
Up to 3 variables (user selects these)

- Level one
  - histograms used
  - if state variables are time series – use aggregates

- Level two
  - double width and height
  - up to six histograms, or fewer larger
  - scatter plot also possible (not used in our case)
  - aggregates of state variables still used

- Level three
  - block can not be larger
  - floating view with map and time series data

Analysis

Data derivation and aggregation integrated in the system
Successfully used to analyze and tune EUI for different operation modes

Complex Interaction – Simple Data
vs.
Simple Interaction – Complex data

Analysis – the High Power Mode 1

Square shaped injection curve
High injection pressure
- we want a lot of fuel

Blocks C and D are of main interest
3rd level view configured using model view
3 families of curves
Analysis – the High Power Mode 2

Select curves with steep rise
- line brush with limited crossing

Refine selection
- only high pressure at the injection start
- stronger fuel penetration – higher power

Refine once more
- subtract too slow rising
- or change limit of the crossing

Analysis – the High Power Mode 3

Second needle opening
- undesired behavior
- can be seen in the curve view
- maybe additional curves are hidden

Use a derived family of function
- first derivative
- more curves detected
- explore why (see paper)

Analysis – the High Power Mode 4

Select curves with step rise
- line brush with limited crossing

Refine selection
- only high pressure at the injection start
- stronger fuel penetration – higher power

Refine once more
- subtract too slow rising
- or change limit of the crossing

Exclude second needle opening

Analysis – the High Power Mode 5

All parameters are always highlighted in the model view

Unlimited Possibilities

Low Consumption mode
Instead of brushing curve view

Use histograms of min and max aggregates of 1st derivative

Integration of Model view into CMV

Helps users in:
- figuring out the basic behavior based on aggregates
- identifying important elements for a given scenario
- configuring views, select what is displayed by one click
- connecting views to originating models

Very positive feedback from domain experts
- will be included in commercial software in near future
Overview 1

Introduction – summary once more
Coordinated Multiple Views – summary once more
Complex data: families of curves
- Traffic Surveillance
- Engineering
- Injection
- Interactive visual steering
- A hybrid approach to visual steering
- Bio-Signals and ICU Data
- Heart reinervation
- IVA of ICU Data

Motivation

Simulation
- widely used in science and engineering
- helps in understanding various phenomena
- recent trend:
  - multiple simulation runs
  - variation of various control parameters
  - study influence on output
  - understanding of simulation model
  - understanding of physical phenomena

Simulation and Visualization 1

Common practice:
- define a model
- run simulation
- visualize results
- change the model
- run simulation ...

Simulation and Visualization 2

Multiple runs:
- define a model
- define control parameters
- run simulation
  - for each combination of control parameters
  - for some of the combinations (steering)
- use interactive visual analysis
  - to understand (optimize) the model
  - to understand the phenomena

Interactive Visual Steering

Model, parameters, many simulation runs
How many is many?
- 11 parameters / 10 variations each
- $10^{11}$ combinations
- 10 simulations per minute (very fast) - $10^{10}$ minutes
- 19 000 years!

Basic Idea

Start with a simple model, tune, expand, tune, expand... It is always possible to go back!
Interactive Steering

3 loops
- Explore initial results set
- Change parameter values
- Change model

Tabular data
- Initial table
- Add / remove rows
- Add / remove columns

Interaction

3 loops
- A – real time interaction
- B – “live” parameter refinement
  (A loop functioning)
- C – model change - after initial runs and data management update same as B

First Step

Start with a simple model
Assume the rest to be ideal
Tune the first model
4 parameters, 750 combinations
12 minutes
Two targets

Interactive Visual Analysis

Second Step

Different responses

Second Step

Krešimir Matković  InfoVis, TU Wien, 17. 04. 2013
Parameter refinement

Additional investigation needed
– refine parameters

Output at various steps

Final model

Everything is set now
We wanted to check all output curves
Actuator variations
Interesting peeks detected!

Final solution

One operation point
Choose one and test
If everything OK – done!

Overview 1

Introduction – summary once more
Coordinated Multiple Views – summary once more
Complex data: families of curves

- Traffic Surveillance
- Engineering
  - Interactive visual steering
  - A hybrid approach to visual steering
- Bio-Signals and ICU Data
  - Heart reinervation
  - IVA of ICU Data

Complex system

A complex system
4 injectors
AVL BOOST HydSim
Many parameters
Main Goals

To understand the model
To understand the regime
To optimize the model

Possible approaches

Automatic – use automatic optimization methods to determine optimum point of operation
- a regression model is needed

Our approach:
Combine automatic and interactive approach
- human in the loop needed
- time consuming
- besides optimum user gets insight in the system operation

Interactive Visual Steering

Computational steering and interactive visualization
- simulation and visualization often decoupled
- integrated system is much more powerful
- integration of: modeling, simulation, visual analysis
  - run simulation
  - explore results
  - change the simulation model or parameters from visualization

Multiple Simulation Runs

Needed for interactive and automatic approach
Define a model
Define control parameters
Run simulation
  - for each combination of control parameters
  - for some of the combinations
Use interactive visual analysis
  - to understand (optimize) the model
  - to understand the phenomena

Data 1

Scalar parameters define each simulation model block
Blocks are connected in a model
Simulation computes state variables for each block
State variables are mostly time (crank-angle) dependent
We do not have a conventional data with a multidimensional point

Data 2

Optimization methods and regression model building methods expect a
multidimensional point
We need to parameterize curves, to represent them with a set of scalars
Impossible to capture all possible variations
Many parameters – good approximation, complex optimization
Interactive methods can use aggregates as well
Basic Idea

Define a model
Define control parameters that will be varied
Determine design points – combinations of control parameters
Run simulations for initial set of points (a sparse set in the continuous parameter space)

Krešimir Matković, InfoVis, TU Wien, 17.04.2013

Illustrative example – initial data

Six parameters varied – 2700 simulation runs

Regression model built
Interactive visual analysis used to explore the data

Krešimir Matković, InfoVis, TU Wien, 17.04.2013

First goals

We want balanced system and balanced individual injectors

Some of the goals:
- small difference between injection pressures of individual pilot injections for each cylinder
- small difference between amounts of fuel injected of individual pilot injections for each cylinder
- as fast as possible needle opening and closing velocities for pilot and main injections
- etc.

Krešimir Matković, InfoVis, TU Wien, 17.04.2013

Brushing the goals

Krešimir Matković, InfoVis, TU Wien, 17.04.2013

Setting Constraints

Based on the selection user sets constraints
“Orders” an optimum from automatic system
Optimum is based on the regression model, there are no curves

Krešimir Matković, InfoVis, TU Wien, 17.04.2013

The optimum is computed

Load the optimum
Run simulation for optimum points – get full data
Compute points in the neighborhood – optimum might be wrong, it’s based on the regression
Explore the neighborhood

Krešimir Matković, InfoVis, TU Wien, 17.04.2013
Explore the neighborhood

P3_pt_diff was minimized
There are better points in the neighborhood!

Another example – changing the regression model

Final decision – a compromise

Conclusion

Close coupling of simulation, optimization, and visualization
Complex systems – complex data
- too complex for pure automatic or pure interactive approach
An integrated system and workflow presented
Supports experts in understanding the system, automatic optimization speeds up the process
Very positive feedback from engineers

Overview 1

Introduction – summary once more
Coordinated Multiple Views – summary once more
Complex data: families of curves
- Traffic Surveillance
- Engineering
  - Interactive visual steering
- A hybrid approach to visual steering
- Bio-Signals and ICU Data
- Heart reinervation
- IVA of ICU Data

Heart Reinervation 1
Heart Reinervation 2

ICU Data

Modern Intensive Care Unit

- a lot of high technology equipment
- routinely logging of
  - many physiological measurements
  - laboratory results
  - interventions information
- these information can help intensive care physicians to understand physiological, pathological and therapeutic processes and so, consequently, to improve medical care

Usage of Data

Huge amounts of data available

Great potential but:

- many proprietary formats, data bases rarely merged
- statistical models are complex and hard to develop without merged DBs
- communication between medical and statistics experts

State of the art

- various regression models and advanced statistics
- Protocols (researchers – daily practice)
- APACHII Score, for example

Interactive Visual Analysis

- novel concept for ICU medical researchers

Data 1

~1500 patients

<table>
<thead>
<tr>
<th>Time</th>
<th>Patient</th>
<th>Age</th>
<th>Height</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:00</td>
<td>4005</td>
<td>72</td>
<td>173</td>
<td>70</td>
</tr>
<tr>
<td>01:00</td>
<td>1404</td>
<td>42</td>
<td>165</td>
<td>70</td>
</tr>
<tr>
<td>02:00</td>
<td>2211</td>
<td>73</td>
<td>172</td>
<td>70</td>
</tr>
<tr>
<td>02:00</td>
<td>4989</td>
<td>72</td>
<td>173</td>
<td>70</td>
</tr>
<tr>
<td>02:00</td>
<td>1404</td>
<td>42</td>
<td>165</td>
<td>70</td>
</tr>
<tr>
<td>02:00</td>
<td>2211</td>
<td>73</td>
<td>172</td>
<td>70</td>
</tr>
<tr>
<td>04:00</td>
<td>4989</td>
<td>72</td>
<td>173</td>
<td>70</td>
</tr>
<tr>
<td>04:00</td>
<td>1404</td>
<td>42</td>
<td>165</td>
<td>70</td>
</tr>
<tr>
<td>04:00</td>
<td>2211</td>
<td>73</td>
<td>172</td>
<td>70</td>
</tr>
<tr>
<td>08:00</td>
<td>2211</td>
<td>72</td>
<td>172</td>
<td>70</td>
</tr>
</tbody>
</table>

- complex processing and analysis
- C-reactive protein concentrations
nD space

Conventional approach deals with n-dim. Euclidian spaces
Each item is a point in n-dim. Space (n-tuple)

\[ P_i = [x_1, \ldots, x_n] \]

Complex Data

Medical data does not always fit to this data model
We want data that belongs to one patient to be in one record
- more natural way of thinking of data
- new possibilities of analysis

Point \( P_i \) is the same, but \( x_i \) can be a mapping
Mapping can generally be \( f(s), f(x,y), \text{set, ...} \)

Family of Curves

Interaction

Curve view in a multiple linked view setup
Need for efficient brush
Line brush
- selects all curves which cross the line
- very intuitive
- easy to define
- would be much more complex using SQL

Composite brushing
- iteratively combine brushes using Boolean operations

Workflow

Prepare the data according to the new data model
Derive data if necessary
Domain Experts have to familiarize with the tool
Add specific features needed by domain experts depending on their tasks
More complex data exploration and analysis
Case Study 1 – Confirming a Hypothesis

We are exploring sodium concentration

Brushing hypernatremia patients
- Use curve view and a long horizontal brush
- Compute maximum aggregate – curve -> scalar and use conventional views

Additional Features

Case Study 2 – Exploring Variability

Case Study 3 – Interactive drill down

New data model makes advanced drill down possible
The timing of fall/rise can be easily specified

Animal paths
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

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)

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

Conclusion and Future Work
- inclusion of animal paths enable interactive drill down
- explanations for new discoveries

Conclusion
- Mappings as dimension - very powerful concept
- Can be applied almost everywhere
- Very positive feedback on line brush from Vis community
- Very positive feedback from domain experts in various domains
Thank you!

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

Krešimir Matković  InfoVis, TU Wien, 17. 04. 2013