Integrating Interactive and Computational Analysis in Visualization

Helwig Hauser
Univ. of Bergen, Norway

Background = computer science (CS)
- Studied CS at Vienna Univ. of Techn., Austria; PhD on flow visualization (also there)
- Ass.-prof. at TU Wien; key res. at VRVis.at, later sci.-dir. there; prof. in visualization (CS Dept.) at UiB.no since 2007

What is visualization?
- many different interpretations available…
- Here: computer-assisted means (usually interactive) to enable insight into data (imaged phenomena)
- many different application areas, including
  - medicine (3D imaging data, patient records, etc.)
  - engineering (data from physical/chemical models, etc.)
  - business (databases, etc.)
This Talk

- **Addressing a hot research topic** (visual analytics)
  - initiative started 2004 in the US (nat. security)
  - topic hot in Europe since 2005

- **Reporting from >1 decade of own research** (interactive visual analysis, IVA)
  - started in 2000 with the new VRVis.at
  - several PhD proj.: H. Doleisch, J. Kehrer, Z. Konyha, …
  - >20 “IVA” papers, many talks, keynotes, etc.

- **Meeting HCI/CHI?** (human time constants, etc.?)
  - drawing some cautious links
  - awaiting interesting questions! :-)

The Data Analysis Challenge

- **Today = emerging new information age**
  - enormous development of (computer) technology
  - fascinating opportunities for data acquisition, incl.
    - through measurements, e.g., imaging
    - through computational simulation

- **Increasing amount and complexity of data**
  - more and more data (GB→TB→PB→EB→…)
  - heterogeneous data (“big data”, etc.)

- **Big Data = a chance & a challenge!**
  - new opportunities (advancing knowledge, better …)
  - difficult to master (getting more difficult quickly)
The “Technical” Approach(es)

- **Machine Learning, Statistics, Data Mining, ...**
  - **main idea**: exploit *computational means* to *extract information* (knowledge) from data
  - **lots of approaches** available, incl.
    - advanced **data summaries** (e.g., statistics)
    - advanced **feature extraction methods** (often application-dependent)
    - advanced **embeddings** (dimension reduction)
    - clustering
    - classification
    - etc.
  - Not really my field...

The “Human” Approach

- **Interactive visualization, visual analytics, IVA, ...**
  - **main idea**: utilize *perception & cognition* to *extract information* (knowledge) from data
  - **visualization** = show the **data** to the user (seeing = understanding)
  - **interaction** allows for **step-by-step analysis**, incl.
    - classical information drill-down (from overview to detail) – cf. Shneiderman ‘91
    - iterative analysis (show features one-by-one)
    - comparative analysis (work out relations)
    - etc.
  - our **visual sense** = data highway to the brain!
  - a picture says more than 100 words
■ The perceptual and cognitive power of users should not be left unutilized!

Matt Ward, 2010:

1. In the Beginning there were Mappings
Data values control the visual variables of points, lines, areas, surfaces, and volumes.

- Position
- Size
- Shape
- Value
- Color
- Orientation
- Texture
- Motion


Matt Ward on Visualization

■ The perceptual and cognitive power of users should not be left unutilized!

■ Matt Ward, 2010:

Dealing with Dimensions

- Many categorizations of dimension organization
(see below paper for an early one)

- My categories:
  - Subsetting
    (e.g., SPLOMs, dense pixels)
  - Reorganization
    (e.g., parallel coords, glyphs)
  - Embedding
    (dimensional stacking, stacked bar charts, trellis displays)
  - Reduction
    (PCA, MDS, RadViz)

Also from Matt Ward’s talk:

Other Challenges in Mappings

- Too many records
- Too many variables
- Non-numeric fields
- Missing values
- Streaming data

*Many partial solutions; all have limitations.*

And Then There are Relations

And What About Data Properties?

*... like data uncertainty*
After Mapping Comes Interaction

Visualization without interaction is like a sports car with no engine! Nice to look at, but not good for much! 😊

Categories of Interactions

- **Select**: mark something as interesting
- **Explore**: show me something else
- **Reconfigure**: show me a different arrangement
- **Encode**: show me a different representation
- **Abstract/Elaborate**: show me more or less detail
- **Filter**: show me something conditionally
- **Connect**: show me related items

(I)VA is about the integration of interactive visual analysis means and computational analysis

Humans and Computers

"Computers are incredibly fast, accurate, and stupid; humans are incredibly slow, inaccurate, and brilliant; together they are powerful beyond imagination."

attributed to Albert Einstein

Levels of integration:

L0: no integration – still the vast majority!
L1a: the visualization of results from some computational analysis (“for the report”, …)
L1b: making computational analysis (partially) interactive
L2: tight integration – extremely rare, still!

Several in visual analytics / IVA / … aim currently at conquering L2!
**Goal**: to combine the best of two worlds [Keim et al.]:

#### Integrated Methods

- **Clustering**
  - k-means
  - hierarchical clustering methods
  - etc.
- **Projections (embeddings), e.g., for dimension reduction**
  - PCA
  - MDS
  - etc.
- **Classification, regression**
  - decision trees
  - SVM
  - etc.
- Etc.

**Outer integration (here!):**

bundling interaction & computation in a loop
Interactive Visual Analysis

- Given data – too much and/or too complex to be shown all at once:

- IVA is an interactive visualization methodology to facilitate
  - the exploration and/or analysis of data (not necessarily the presentation of data), including
    - hypothesis generation & evaluation, sense making,
    - knowledge crystallization, etc.
  - according to the user’s interest/task, for ex., by interactive feature extraction,
  - navigating between overview and details, e.g., to enable interactive information drill-down [Shneiderman]

- through an iterative & interactive visual dialog

The Iterative Process of IVA

- Loop / bundling of two complementary parts:
  - visualization – show to the user!
    Something new, or something due to interaction.
  - interaction – tell the computer!
    What is interesting? What to show next?

- Basic example (show – brush – show – …), cooling jacket context:
  1. show a histogram of temperatures
  2. brush high temperatures (>90°[±2°])
  3. show focus+context vis. in 3D
  4. locate relevant feature(s)

- KISS-principle IVA:
  - linking & brushing, focus+context visualization, …
SHOW & BRUSH

**Tightest IVA loop**
- **show data** (explicitly represented information)
- **one brush** (on one view, can work on >1 dims.)

**Requires:**
- multiple views (≥2)
- interactive brushing capabilities on views (brushes should be editable)
- **focus+context visualization**
- linking between views

**Allows for different IVA patterns** (wrt. domain & range)

A TYPICAL (start into an) IVA SESSION of this kind:
- bring up multiple views
  - at least one for \( x, t \)
  - at least one for \( d_i \)
- I see (something)!
- brush this “something”
- linked F+C visualization
- first insight!

... leads to...
... requires...
... is realized via ...
IVA: Multiple Views

- One dataset, but multiple views
- Scatterplots, histogram, 3D(4D) view, etc.

Interactive Brushing

- Move/alter/extend brush interactively
- Interactively explore/analyze multiple variates
Interactive Brushing

- Move/alter/extend brush interactively
- Interactively explore/analyze multiple variates

[Doleisch et al., '03] (SimVis)
IVA: Focus+Context Visualization

- Traditionally space distortion
  - more space for data of interest
  - rest as context for orientation
- Generalized F+C visualization
  - emphasize data in focus (color, opacity, ...)
  - differentiated use of visualization resources

F+C Visualization in IVA Views

- Colored vs. gray-scale visualization
- Opaque vs. semi-transparent visualization

In a scatterplot (left) or histogram (right): brushed data in red...
F+C Visualization in IVA Views

In parallel coordinates (above): brushed data in red & over …

In 3D (above): less transp. & colored, in illustrative context …
IVA: Linked Views

- Brushing: mark data subset as especially interesting
- Linking: enhance brushed data in linked views consistently (F+C)

[Doleisch & Hauser, '02]
IVA: Degree of Interest (DOI)

- $\text{doi}(\cdot)$: data items $tr_i$ (table rows) $\rightarrow$ degree of interest $\text{doi}(tr_i) \in [0,1]$
  - $\text{doi}(tr_i) = 0 \Rightarrow tr_i$ not interesting ($tr_i \in \text{context}$)
  - $\text{doi}(tr_i) = 1 \Rightarrow tr_i$ 100% interesting ($tr_i \in \text{focus}$)

- Specification
  - explicit, e.g., through direct selection
  - implicit, e.g., through a range slider

- Fractional DOI values: $0 \leq \text{doi}(tr_i) \leq 1$
  - several levels (0, low, med., ...)
  - a continuous measure of interest
  - a probabilistic definition of interest

(Cont’d on next slide)

IVA: Smooth Brushing $\rightarrow$ Fractional DOI

- Fractional DOI values esp. useful wrt. scientific data: (quasi-)continuous nature of data $\leftrightarrow$ smooth borders

- Goes well with gradual focus+context vis. techniques (coloring, semitransparency)

- Specification: smooth brushing [Doleisch & Hauser, 2002]
  - “inner” range: all 100% interesting (DOI values of 1)
  - between “inner” & “outer” range: fractional DOI values
  - outside “outer” range: not interesting (DOI values of 0)
Fuzzy Classification

DOI 

\(0, 1\) – 0… not interesting 

1… 100% interesting

Requires fuzzy logic for combination, we use 

\[ c = \frac{a}{b} \]

\[ c = \min(a, b) \]

\[ c = \max(a, b) \]

\[ c = \frac{1}{a} \]

Matches the smooth nature of the data

Goes well with F+C visualization, e.g.,

opacity varies gradually with DOI

[Doleisch & Hauser 2002]
IVA – Levels of Complexity (1/4)

- A lot can be done with basic IVA, already! [pareto rule]

- We can consider a layered information space:
  from explicitly represented information (the data) to implicitly contained information, features, ...

IVA – Levels of Complexity (2/4)

- A lot can be done with basic IVA, already! [pareto rule]

- For more advanced exploration/analysis tasks, we extend it (in several steps):
  - IVA, level 2: logical combinations of brushes, e.g., utilizing the feature definition language [Doleisch et al., 2003]
  - IVA, l. 3: attribute derivation; advanced brushing, with interactive formula editor; e.g., similarity brushing
  - IVA, l4: application-specific feature extraction, e.g., based on vortex extraction methods for flow analysis

- Level 2: like advanced verbal feature description
  - ex.: “hot flow, also slow, near boundary” (cooling j.)
  - brushes comb. with logical operators (AND, OR, SUB)
  - in a tree, or iteratively (((b_0 \ op_1 b_1) \ op_2 b_2) \ op_3 b_3) ...
IVA – Levels of Complexity (2/4)

A lot can be done with basic IVA, already! [pareto rule]

For more advanced exploration/analysis tasks, we extend it (in several steps):

- IVA, level 2: **logical combinations of brushes**, e.g., utilizing the **feature definition language** [Doleisch et al., 2003]
- IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; e.g., similarity brushing
- IVA, l4: **application-specific feature extraction**, e.g., based on vortex extraction methods for flow analysis

Level 2: like advanced verbal feature description
- ex.: “hot flow, also slow, near boundary” (cooling j.)
- brushes comb. with **logical operators** (AND, OR, SUB)
- in a tree, or iteratively (((b₀ op₁ b₁) op₂ b₂) op₃ b₃) ...

IVA – Levels of Complexity (3/4)

- A lot can be done with basic IVA, already! [pareto rule]
- For more advanced exploration/analysis tasks, we extend it (in several steps):
  - IVA, level 2: **logical combinations of brushes**, e.g., utilizing the **feature definition language** [Doleisch et al., 2003]
  - IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; e.g., similarity brushing
  - IVA, l4: **application-specific feature extraction**, e.g., based on vortex extraction methods for flow analysis

Level 3: using **general info extraction** mechanisms, two (partially complementary) approaches:
  1. derive additional attribute(s), then show & brush
  2. use an **advanced brush** to select “hidden” relations
IVA – Levels of Complexity

- A **lot** can be done with basic IVA!
- For more **advanced** exploration/analysis tasks, we extend it (in several steps):
  - IVA, level 2: **logical combinations** of brushes utilizing the **feature definition language**
  - IVA, l. 3: **attribute derivation** with interactive formula editor
  - IVA, l4: **application-specific feature extraction** based on vortex extraction methods

- Level 3: using **general info extraction** mechanisms, two (partially complementary) approaches:
  1. derive **additional attribute(s)**, then show & brush
  2. use an **advanced brush** to select “hidden” relations

IVA (level 3): Advanced Brushing

- **Std. brush:** brush 1:1 what you see
- **Adv. brush:** executes additional function (“intelligent”?)

- **Examples:**
  - angular brushing [Hauser et al., 2002]
  - similarity brushing [Muigg et al., 2008]
3rd level IVA, adv. brushing example

- Considering a visualization of a family of function graphs:
  - select the steeply rising graphs

![Graph example with fuel injection simulation][1]

3rd level IVA, adv. brushing example

- A simple line brush is not enough

![Graph example with fuel injection simulation][2]

---

[1]: example-prepared-by-Konyha-Zoltan
[2]: example-prepared-by-Konyha-Zoltan
A simple line brush is not enough

Combining line brushes does not work, either

**feature of interest:** not explicitly available

---

3rd level IVA, adv. brushing example

- The *angular line brush* (a specialized brush) selects the intended function graphs
  - that it intersects, and
  - the angle is in a given threshold

---

example prepared by Konyha, Zoltan
**IVA (level 3): Attribute Derivation**

- **Principle** (in the context of iterative IVA):
  - see some data feature $\Phi$ of interest in a visualization
  - identify a mechanism $T$ to describe $\Phi$
  - execute (interactively!) an attribute derivation step to represent $\Phi$ explicitly (as new, synthetic attribute[s] $d_\Phi$)
  - brush $d_\Phi$ to get $\Phi$

- **Tools** $T$ to describe $\Phi$ from:
  - numerical mathematics
  - statistics, data mining
  - etc.
  - scientific computing

**IVA w/ $T \leftrightarrow$ visual computing**

---

**Attribute Derivation $\leftrightarrow$ User Task / example**

- The tools $T$, available in an IVA system, must reflect/match the **analytical steps of the user**:

- **Example:**
  - first vis.: $\leftrightarrow$ user wishes to select the “band” in the middle
  - so? an advanced brush? a lasso maybe?
  - ah! $\rightarrow$ let’s normalize $y$ and then brush (a)

- leading to the wished selection:
What user wishes to reflect?

- Many **generic wishes** – users interest in:
  - something **relative** (instead of some absolute values),
    example: show me the top-15%
  - **change** (instead of current values),
    ex.: show me regions with increasing temperature
  - some **non-local property**, 
    ex.: show me regions with high average temperature
  - **statistical properties**, 
    ex.: show me outliers
  - **ratios/differences**, 
    ex.: show me population per area, difference from trend
  - **etc.**

- **Common characteristic** here:
  - **questions/tools generic**, not application-dependent!

How to reflect these user wishes?

- Many **generic wishes** – users interest in:
  - something **relative** (instead of some absolute values),
    example: show me the top-15%
  - **change** (instead of current values),
    ex.: show me regions with increasing temperature
  - some **non-local property**, 
    ex.: show me regions with high average temperature
  - **statistical properties**, 
    ex.: show me outliers
  - **ratios/differences**, 
    ex.: show me population per area, difference from trend
  - **etc.**

- **Common characteristic** here:
  - **questions/tools generic**, not application-dependent!
Some useful tools for 3rd-level IVA

From analysis, calculus, num. math:

- **linear filtering** (convolve the data with some linear filter on demand, e.g., to smooth, for derivative estimation, etc.)
- **calculus** (use an interactive formula editor for computing simple relations between data attributes; +, −, ∙, ⁄, etc.)
- **gradient estimation, numerical integration** (e.g., wrt. space and/or time)
- **fitting/resampling** via interpolation/approximation

From statistics, data mining:

- **descriptive statistics** (compute the statistical moments, also robust, measures of outlyingness, detrending, etc.)
- **embedding** (project into a lower-dim. space, e.g., with PCA for a subset of the attribs., etc.)

Important: executed on demand, after prev. vis.

---

3rd-level IVA – Sample Iterations

The Iterative Process of 3rd-level IVA:

Example 1:

- you look at some temp. distribution over some region
- you are interested in raising temperatures, but not temperature fluctuations
- you use a temporal derivate estimator, for ex., central differences $t_{\text{change}} = (t_{\text{future}} - t_{\text{past}}) \div \text{len(future−past)}$
- you plot $t_{\text{change}}$, e.g., in a histogram and brush whatever change you are interested in
- maybe you see some frequency amplification due to derivation, so you go back and
- use an appropriate smoothing filter to remove high frequencies from the temp. data, leading to a derived new $t_{\text{smooth}}$ data attribute
- selecting from a histogram of $t_{\text{change}}$ (computed like above) is then less sensitive to temperature fluctuations
The Iterative Process of 3\textsuperscript{rd}-level IVA:

- Example exploiting PCA:
  - you bring up a scatterplot of $d_1$ vs. $d_2$: (from an ECG dataset [Frank, Asuncion; 2010])
  - obviously, $d_1$ and $d_2$ are correlated, our interest: the data center wrt. the main trend
  - we ask for a (local) PCA of $d_1$ and $d_2$:
  - then we brush the data center
  - we get the wished selection
  - from here further steps are possible…, incl. study of other PCA-results, etc.

Brushing of Attribute Clouds for the Visualization of Multivariate Data

Helke Jänicke, Michael Böttinger, and Gerik Scheuermann, Member, IEEE
IVA – Levels of Complexity

A lot can be done with basic IVA, already! [pareto rule]

For more advanced exploration/analysis tasks, we extend it (in several steps):

- IVA, level 2: logical combinations of brushes, utilizing the feature definition language [Doleisch et al., 2003]
- IVA, l. 3: attribute derivation; advanced brushing, with interactive formula editor; e.g., similarity brushing
- IVA, l4: application-specific feature extraction, e.g., based on vortex extraction methods for flow analysis

Level 4: application-specific procedures
- tailored solutions (for a specific problem)
- “deep” information drill-down
- etc.
The Iterative Process of IVA...

...leads to an interactive & iterative workbench for visual data exploration & analysis (compare to visual computing, again)

- A really important question is: how fast is one such loop?

- Jean-Daniel Fekete, 2012:

<table>
<thead>
<tr>
<th>TIME CONSTANT</th>
<th>VALUE</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptual processing</td>
<td>0.1 s</td>
<td>[5]</td>
</tr>
<tr>
<td>Immediate response</td>
<td>1 s</td>
<td>[21]</td>
</tr>
<tr>
<td>Unit task</td>
<td>10 s</td>
<td>[5, 21]</td>
</tr>
</tbody>
</table>

THE INFORMATION VISUALIZER, AN INFORMATION WORKSPACE
Stuart K. Card, George G. Robertson, Jock D. Mackinlay
Xerox Palo Alto Research Center
Palo Alto, California 94304
(415) 994-4382, Card.PARC@Xerox.COM

CHI '91

Categories of Interaction Pace

- Separate ▶ unit task ▶ immediate ▶ continuous
- separate: offline processing
- unit task [Card et al., '91]: ≈10s – before attention breaks!
- immediate: ≈1s – maintains an interplay, a conversation
- continuous: ≈0.1s – smooth in the eye (perception)

Really important differences on the user side!

Response Times

- 0.1 sec - animation, visual continuity, sliders
- 1 sec - system response, conversation break
- 10 sec - cognitive response


- Beyond 20 sec, users wait and loose attention
  - Forget their goals and plans
  - Progress bar needed!
The Iterative Process of IVA...

...leads to an interactive & iterative workbench for visual data exploration & analysis
(compare to visual computing, again)

Different levels of complexity (show & brush, logical combinations, advanced brushing & attribute derivation, etc.)...

...lead to according iteration frequencies:
- on level 1: smooth interactions, many fps, for example during linking & brushing
- on level 2: interleaved fast steps of brush ops., for example when choosing a logical op. to cont. with
- on level 3: occasionally looking at a progress bar, for example when computing some PCA, etc.

These frequencies limit the spectrum of usable tools
- New res. work will help to extend this spectrum!

The Iterative Process of IVA...

...is a very useful methodology for data exploration & analysis

...is very general and can be (has already been) applied to many different application fields (in this talk the focus was on scientific data)

...meets scientific computing as a complementary methodology (with the important difference that in IVA the user with his/her perception/cognition is in the loop at different frequencies, also many fps)

...is not yet fully implemented (we’ve done something, e.g., in the context of SimVis, ComVis, etc.) – from here: different possible paths, incl. InteractiveVisualMatlab, IVR, etc.)
Acknowledgements

You!

Krešimir Matković & Giuseppe Santucci!

Helmut Doleisch, Raphael Fuchs, Johannes Kehrer, Çağatay Turkay, et al.!


All around SimVis and ComVis and …

Funding partners (FFG, AVL, EU, UiB, …)