Interactive Visual Analysis of Rich Scientific Data

Helwig Hauser
University of Bergen

HH, ii.UiB.no/vis

HH: prof. in visualization (vis)
@ Dept. of Informatics (ii)
@ Univ. of Bergen (UiB)

UiB VisGroup

– 2007: group of 3: 
– 2009: larger projects start
– 2011: EuroVis in Bergen

– 2013: new prof.: 

[ranking from NFR's 10-year evaluation in 2011/2012]
Application-oriented basic research in visualization:

1. Researched visualization methodology (how to visualize)
   - Interactive Visual Analysis, \( n \)D data (H. Hauser et al.)
   - Visual Knowledge Discovery, 3D data (St. Bruckner et al.)
   - Illustrative Visualization (I. Viola et al.)

2. Applications at which this research is oriented (for whom)
   - Medical Visualization (partner in MedViz Bergen, etc.)
   - GeoSciences / Oil & Gas (e.g., financed by Statoil’s Akademiaavtale)
   - Biology / Bioinformatics (with CBU@ii et al.)
   - Fluid Dynamics (in collab. with FFI.no, for ex.)
   - Engineering (visual analysis of simulation data)

ii.UiB.no/vis PhDs (11 so far)

- Daniel Patel (Oct. 2009): Expressive Vis. & Rapid Interpr. of Seismic Volumes
- Jean-Paul Balabanian (Jan. 2010): Multi-Aspect Vis.: from Linked to Integrated Views
- Johannes Kehrer (May 2011): IVA of Multi-faceted Scientific Data
- Ove Daae Lampe (Nov. 2011): IVA of Process Data
- Armin Pobitzer (June 2012): IVA of Time-dependent Flows
- Paolo Angelelli (June 2012): Visual Expl. of Human Physiology
- Åsmund Birkeland (May 2013): Ultrasonic Vessel Vis.: From Extraction to Perception
- Endre Lidal (May 2013): Sketch-based Storytelling for Cognitive Problem Solving
- Çağatay Turkay (Nov. 2013): Interactive Visual Analysis of High-dimensional Data
- Mattia Natali (Sept. 2014): Sketch-based Mod. & Conceptual Vis. of Geomorphological Processes for ...
Interactive Visual Analysis

Helwig Hauser
University of Bergen

Interactive Visual Analysis (IVA)

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

- IVA is an interactive visualization approach to facilitate
  - the exploration and/or the 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
Interactive Visual Analysis ↔ Visual Analytics

- IVA ("interactive visual analysis") since 2000
- Tightly related to visual analytics, of course, e.g., integrating computational & interactive data analysis
- A particular methodology with specific components (CMV, linking & brushing, F+C vis., etc.)
- General enough to work in many application fields, but not primarily the VA fields (national security, etc.), in particular "scientific data" fields…

Integrating Interaction & Computation

- **Goal**: to combine the best of two worlds [Keim et al.]:
  - data exploration/analysis by the user, based on interactive visualization
  - and data analysis by the computer, based on statistics, machine learning, etc.
- **State of the art / levels of integration**:
  - mostly no integration, still
  - some vis. of results of computations
  - also: making comp. semi-interactive (here called "inner integration")
  - rare: tight integration
- **Outer integration** (here!):
  bundling interaction & computation in a loop
Target Data Model: “Scientific Data”

- **Characterized** by a combination of
  - **independent variables**, like *space* and/or *time* (cf. *domain*)
  - and **dependent variables**, like *pressure*, *temp.*, etc. (cf. *range*)

- So we can think of this type of data as **given as** \( d(x) \) with \( x \leftrightarrow \text{domain} \) and \( d \leftrightarrow \text{range} \) – examples:
  - CT data \( d(x) \) with \( x \in \mathbb{R}^3 \) and \( d \in \mathbb{R} \)
  - unsteady 2D flow \( v(x,t) \) with \( x \in \mathbb{R}^2 \), \( t \in \mathbb{R} \), and \( v \in \mathbb{R}^2 \)
  - num. sim. result \( d(x,t) \) with \( x \in \mathbb{R}^3 \), \( t \in \mathbb{R} \), and \( d \in \mathbb{R}^n \)
  - system sim. \( q(p) \) with \( p \in \mathbb{R}^n \) and \( q \in \mathbb{R}^m \)

- **Common property:**
  - \( d \) is (at least to a certain degree) **continuous** wrt. \( x \)

Interactive Visual Analysis of Scientific Data

- **Interactive visual analysis** (as exemplified in this tutorial) **works really well with scientific data**, e.g.,
  - results from **numerical simulation** (spatiotemporal)
  - imaging / **measurements** (in particular multivariate)
  - sampled **models**

- When used to study scientific data, **IVA employs**
  - methods from **scientific visualization** (vol. rend., …)
  - methods from **statistical graphics** (scatterplots, …), **information visualization** (parallel coords., *etc.*)
  - **computational tools** (statistics, machine learning, …)

- Applications include
  - **engineering**, **medicine**, **meteorology/climatology**, **biology**, *etc.*
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!

---

**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)
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

[Doleisch et al., ’03]
Interactive Brushing

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

[Doleisch et al., '03]
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

[Hauser... 2001, 2003]

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…

[Matković et al., '09]
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...

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

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

[Novotný & Hauser, '06]

[Muigg et al., '07]
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)

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

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

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

(continues 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 \cite{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 \(c = \min(a, b)\)
\(c = \max(a, b)\)
\(c = 1\)

Matches the smooth nature of the data
Goes well with F+C visualization, e.g.,
opacity varies gradually with DOI

[Doleisch & Hauser 2002]
Three Patterns of SciData IVA

- Preliminary: domain $x$ & range $d$ visualized (≥2 views)

1. **Brushing on domain visualization**, e.g., brushing special locations in the map view
   - "$x$" local investigation
   - "... from $x$ to $d$ ...

2. **Brushing on range visualization**, e.g., brushing outlier curves in a function graph view
   - "$d$" feature localization
   - "... from $d$ to $x$ ...

3. Relating multiple range variates
   - "$d$" multi-variate analysis
   - "... within $d$ ...

IVA – Levels of Complexity

- 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, ...

```
show & brush

temp.  vel.  ...  data

between the lines...

??

buried deeper...

features in application terms

... vort.
```
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 (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

- **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 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]
- percentile brush [new]

**3rd level IVA, adv. brushing example**

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

Example prepared by Konyha, Zoltan
3rd level IVA, adv. brushing example

- A simple line brush is not enough

- Combining line brushes does not work, either

feature of interest: not explicitly available
The *angular line brush* (a specialized brush) selects the intended function graphs that it intersects, and the angle is in a given threshold.

*3rd level IVA, adv. brushing example*

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] $d_\varphi$)
  - brush $d_\varphi$ to get $\Phi$

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

- **IVA w/ $T$ ↔ visual computing**

**Attribute Derivation ↔ 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! 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%  
      ⇒ **use**, e.g., **normalization**
  - **change** (instead of current values),
    ex.: show me regions with increasing temperature  
      ⇒ **derivative estimation**
  - some **non-local property**, 
    ex.: show me regions with high average temperature  
      ⇒ **numerical integration**
  - **statistical properties**, 
    ex.: show me outliers  
      ⇒ **descriptive statistics**
  - **ratios/differences**, 
    ex.: show me population per area, difference from trend  
      ⇒ **calculus**
  - **etc.**  
      ⇒ **data mining**
      (fast enough?)

- **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 $t_{\text{change}} = (t_{\text{future}} - t_{\text{past}}) / \text{len(future−past)}$
    - 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 $t = t_{\text{smooth}}$ data attribute
    - selecting from a **histogram** of $t_{\text{change}}$ (computed like above) is then less sensitive to temperature fluctuations
A lot can be done with basic IVA, already! [Pareto rule]

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

1. **IVA, level 2:** logical combinations of brushes, utilizing the feature definition language [Doleisch et al., 2003]

2. **IVA, l. 3:** attribute derivation; advanced brushing, with interactive formula editor; e.g., similarity brushing

3. **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.
Interactive Visual Analysis – delivery

- Understanding data **wrt. range** \( d \)
  - which distribution has data attribute \( d_i \)?
  - how do \( d_i \) and \( d_j \) relate to each other? (**multivariate analysis**)
  - which \( d_k \) discriminate data features?

- Understanding data **wrt. domain** \( x \)
  - **where** are relevant features? (**feature localization**)
  - **which** values at specific \( x \)? (**local analysis**)
  - how are they related to parameters?

---

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!

Alexander Lundervold and HIB!

Collaborators:
H. Doleisch, R. Fuchs/Bürger,
J. Kehrer, Ç. Turkay, Z. Konyha,
Kr. Matković, P. Filzmoser,
et al.

Funding agencies!