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
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*
- **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…

---

**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 $x \leftrightarrow$ **domain** and $d \leftrightarrow$ **range** – examples:
  - **CT data** $d(x)$ with $x \in \mathbb{R}^3$ and $d \in \mathbb{R}$
  - **unstead 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**  
(IVA level 1)

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

**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**  
(IVA level 1)

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

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

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

... leads to...
... requires...
... is realized via...

(Next slide)
IVA: Multiple Views

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

IVA: 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

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)

- **dois(.)**: data items $tr_i$ (table rows) $\rightarrow$ degree of interest
  - $dois(tr_i) \in [0,1]$
  - $dois(tr_i) = 0 \Rightarrow tr_i$ not interesting ($tr_i \in$ context)
  - $dois(tr_i) = 1 \Rightarrow tr_i$ 100% interesting ($tr_i \in$ focus)

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

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

(contr’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)
Three Patterns of SciData IVA

1. **Preliminary**: domain $x$ & range $d$ visualized ($\geq 2$ views)
   - **brushing on domain visualization**, e.g., brushing special locations in the map view
   - **local investigation**
   - **brushing on range visualization**, e.g., brushing outlier curves in a function graph view
   - **feature localization**
   - Relating multiple range variates
   - **multi-variate analysis**

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**

<table>
<thead>
<tr>
<th>temp.</th>
<th>vel.</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>between the lines...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>buried deeper...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>features in application terms</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

vort.
IVA – Levels of Complexity

- A lot can be done with KISS-principle IVA!
- 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 (level 2) Example

---

Layered information space
data between the lines…
buried deeper…
features in application terms

temp. vel.
vort.

brush combinations
IVA – Levels of Complexity

A lot can be done with KISS-principle IVA! [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]
**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” $\mapsto$ in the middle
  - so? $\Rightarrow$ 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%*  
    ⇒ **use**, e.g., *normalization*
  - **change** (instead of current values)
    ex.: show me *regions with increasing*  
    ⇒ **derivative estimation**
  - some **non-local property**,  
    ex.: show me regions with *high average*  
    ⇒ **numerical integration**
  - **statistical properties**,  
    ex.: show me *outliers*  
    ⇒ **descriptive statistics**
  - **ratios/differences**,  
    ex.: show me population per area, difference  
    ⇒ **calculus**
  - *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 (1/2)

- **The Iterative Process of 3rd-level IVA:**
  - **Example 1:**
    - you look at some temp. distribution over some region
    - you are interested raising temperatures, **but not temperature fluctuations**
    - you use a temporal derivate estimator, for ex., central differences $t_{\text{change}} = (t_{\text{future}} - t_{\text{past}}) / \text{len}(\text{future} - \text{past})$
    - you plot $t_{\text{change}}$, e.g., in a **histogram** and **brush** what ever change you are interested in
    - maybe you see that 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
The Iterative Process of 3rd-level IVA:

- Example 2:
  - 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.

Visualizing / analyzing lots of statistics

- Useful statistical measures include:
  - moments ($\mu$, $\sigma$, …), robust versions (median, IQR, …)
  - quartiles, octiles, and quartiles $q(p)$

- Useful views allow the interactive visual analysis
  - quantile-plot $q(p)$ vs. $p$, here for numerous $x$
  - detrending (e.g., $-q_2$), normalization (e.g., $z$)
Brushing of Attribute Clouds for the Visualization of Multivariate Data

Heike Jänicke, Michael Böttinger, and Gerik Scheuermann, Member, IEEE

IVA – Levels of Complexity

A lot can be done with KISS-principle IVA! [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, e.g., similarity brushing feature extraction, e.g., methods for flow analysis traction mechanisms, oaches:

s), then show & brush select “hidden” relations
A lot can be done with KISS-principle 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 [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.

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...

...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.)
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, …)