Advanced Visual Analytics Interfaces

Daniel A. Keim
University of Konstanz
keim@dbvis.inf.uni-
konstanz.de

Peter Bak
University of Konstanz
bak@dbvis.inf.uni-
konstanz.de

Enrico Bertini
University of Konstanz
bertini@dbvis.inf.uni-
konstanz.de

Daniela Oelke
University of Konstanz
oelke@dbvis.inf.uni-
konstanz.de

David Spretke
University of Konstanz
spretke@dbvis.inf.uni-
konstanz.de

Hartmut Ziegler
University of Konstanz
ziegler@dbvis.inf.uni-
konstanz.de

ABSTRACT

Advanced visual interfaces, like the ones found in information visualization, intend to offer a view on abstract data spaces to enable users to make sense of them. By mapping data to visual representations and providing interactive tools to explore and navigate, it is possible to get an understanding of the data and possibly discover new knowledge. With the advent of modern data collection and analysis technologies, the direct visualization of data starts to show its limitations due to limited scalability in terms of volumes and to the complexity of required analytical reasoning. Many analytical problems we encounter today require approaches that go beyond pure analytics or pure visualization. Visual analytics provides an answer to this problems by advocating a tight integration between automatic computation and interactive visualization, proposing a more holistic approach. In this paper, we argue for Advanced Visual Analytics Interfaces (AVAIs), visual interfaces in which neither the analytics nor the visualization needs to be advanced in itself but where the synergy between automation and visualization is in fact advanced. We offer a detailed argumentation around the needs and challenges of AVAIs and provide several examples of this type of interfaces.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces; I.3.6 [**Computer Graphics**]: Methodology and Techniques; H.2.8 [**Database Applications**]: Data Mining; H.5 [**Information Interfaces and Presentation**]: Miscellaneous

Keywords

Visual Analytics, Information Visualization, Data Mining, Knowledge Discovery

1. INTRODUCTION

The rapid development of data collection technologies in the last decades led us to an ever growing capability to accumulate large quantities of data. In almost any sector of modern society we assist to a daily recording and collection of data. Medicine, marketing, public policy, public security, engineering are only few examples of areas where data is collected steadily and often at a high rate. Every institution needs to extract useful knowledge from data in order to maximize efficiency, make sound decisions, or increase understanding of complex processes.

While some of these advances are best approached through computational methods or interactive data visualization, problems exist in which neither advanced analytics nor advanced visualizations can provide adequate solutions. Advanced analytics is imperative when dealing with very large volumes of data and complex domains, as human perceptual and cognitive capabilities and visualization technologies just cannot scale to such a complexity. This is the preferred solution when a problem can be expressed in a closed format and formulated through descriptions that computers can handle. At the same time, visualization is the preferred solution for problems that require complex human judgment and interpretation, in which it is hard, if not impossible, to provide a formal description. In this case, even the most advanced computational technique just cannot compare to the ability of humans of giving a meaning and providing a context to the information extracted. But what happens when the problem we need to address requires both high computational capabilities and human judgment?

One example is the monitoring of computer network traffic for security purposes. In many of these environments data is collected at high rates, in the order of millions of flows per day, and visualization just cannot to scale to such a complexity. At the same time, however, malicious events hide like needles in the haystack, evolve with time, and are hardly discernible from non-malicious events. The challenge for an analyst lays, therefore, in the discernment between real attacks and normal traffic and in the implementation of proper countermeasures. While automated data analysis is certainly needed to reduce the search space, visual inspection, human judgment and domain knowledge are needed for proper interpretation of the observed events.

Visual Analytics has been defined as "the science of analytical reasoning supported by interactive visual interfaces" [9] with the specific intent of supporting people with prob-

lems of this type. In order to realize this vision it is necessary to develop not only advanced analytics and advanced visualization but also a proper integration of the two. But if integration of advanced analytics and advanced visualization is the answer to these challenges we are left with the core question of how to best combine them. This paper draws the lines of this integration and shows real examples of advanced visual analytics interfaces where such integration takes place.

Section 2 introduces the science of visual analytics, describes the idea of advanced visual analytics interfaces, and provides insights on how to combine analytics and visualization. Section 3 shows how these ideas can be realized in practice by presenting several examples of visual analytics problems solved through visual analytics tools. Finally, Section 4 provides some concluding thoughts.

2. ADVANCED VISUAL ANALYTICS INTERFACES

2.1 Visual Analytics

2.1.1 Visual Analytics Definition

In general, visual analytics can be described as "the science of analytical reasoning facilitated by interactive visual interfaces" [9]. To be more precise, visual analytics is an iterative process that involves information gathering, data preprocessing, knowledge representation, interaction and decision making. The ultimate goal is to gain insight in the problem at hand, which is described by vast amounts of scientific, forensic, or business data from heterogeneous sources. To reach this goal, visual analytics combines the strengths of machines with the strengths of humans. While on the one hand, methods from knowledge discovery in databases (KDD), statistics and mathematics are the driving force on the automatic analysis side, on the other hand human capabilities provide the necessary support to generate new hypotheses.

2.1.2 Visual Analytics Process Model

Visual Analytics can be best described by the visual analytics process, a series of steps that integrate human and computational steps in an integrated fashion to meet an analytical goal. Figure 1 shows the main elements involved (the ovals) and the data transformations between them (arrows). Preprocessing and transformation are often the preliminary necessary steps to extract the data of interest and to format them in a way that fits the shape of the problem at hand. The analyst can then select between visual or automatic analysis methods. Mapping the data to a visual representation may directly lead to the desired knowledge, but more likely an initial visualization is not sufficient and further user interaction is needed. Several iterations of data visualization and interaction may lead to the construction of a model able to describe the process or phenomenon of interest. The construction of a model using data mining methods can also be the first step after data processing. Once a model is created the analyst has the ability to interact with the automatic methods by modifying parameters or selecting other types of analysis algorithms. Model visualization can then be used to verify the findings or to refine the model itself. Alternating between visual and automatic methods is characteristic
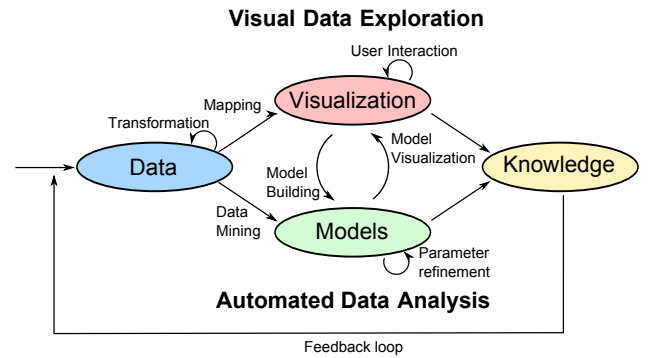

Figure 1: Visual Analytics Process Model.

for the visual analytics process and leads to a continuous refinement and verification of preliminary results.

In the visual analytics process, knowledge can be gained from visualization, automatic analysis, as well as the preceding interactions between visualizations, models, and the human analysts. The feedback loop stores this knowledge of insightful analyses in the system and enables the analyst to draw faster and better conclusions in the future.

2.1.3 Visual Analytics Vs. Information Visualization

Many people are confused by the new term visual analytics and do not see a difference between these two areas. While there is certainly some overlay, and some of the information visualization work is certainly highly related to visual analytics, traditional visualization work does not necessarily deal with an analysis task nor does it always also use advanced data analysis algorithms. Visual analytics is more than just visualization. It can rather be seen as an integral approach to decision-making, combining visualization, human factors and data analysis. The challenge is to identify the best automated algorithm for the analysis task at hand, identify its limits which cannot be further automated, and then develop a tightly integrated solution which adequately integrates the best automated analysis algorithms with appropriate visualization and interaction techniques. While some of such research has been done within the visualization community in the past, the degree to which advanced knowledge discovery algorithms have been employed is quite limited. The idea of visual analytics is to fundamentally change that. It will help to focus on the right part of the problem, i.e., the parts that cannot be solved automatically, and will provide solutions to problems that we were not able to solve in the past.

2.2 The basic idea of Advanced Visual Analytics Interfaces

If the role of visualization in visual analytics is to provide the support that automatic algorithms cannot provide, how do AVAIs look like? The main difference is that the development of visualization is no longer focused on fancy or novel visualizations. Even though some specific problems might need the development of novel visualization or interaction techniques, most of the times standard techniques work perfectly fine for the problem at hand, because the focus is on finding the most effective solution for the given application problem. The role a visualization plays in this context is ba-
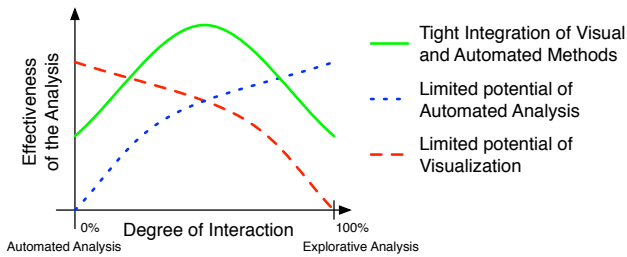
Figure 2: Type of analytical problems described through effectiveness of analysis vs. degree of interaction.

sically twofold: (1) it provides a user interface to look inside the automated process and let the user steer it to make the best out of the computational side; (2) provide an effective interface to display the results obtained as output from the automatic part. In turn, this means that neither the analytics nor the visualization in itself needs to be "advanced" rather that the combination is "advanced" (i.e., effective) in solving the application problem.

2.3 Combing Analytics and Visualization

The respective amount of analytics and visualization needed to solve a problem depends very much on the nature of the problem itself. We can only assess how much visualization and how much analytics are required by assessing the users' cognitive capabilities, the analysis task and the available data. Larger data sets and well-defined problems should be solved through analytics. User involvement is extremely costly and should be avoided if not necessary. But visualization is necessary when confidence in the end results and background knowledge are needed. Figure 2 describes analytical problems according to the characteristic relationship they show between *degree of interaction* and *effectiveness of analysis*[5]. Solutions to problems such as automatic electric switching, customer scoring or credit card approval (red line), become less effective as the degree of interaction grows. In contrast, problems such as the search for the airplane of Steve Fossett in huge amounts of high resolution satellite images [1], are still better solvable by humans. Combining the best of both worlds through visual analytics applications is a very promising solution for problems that can neither be effectively solvable through automated analysis nor explorative analysis, as shown by the solid curve.

As explained by Bertini et al., the integration between visualization and analytics can take place in different ways [3]. Some solutions start from a basic visualization approach and then need the contribution of automated techniques in order to scale up to the complexity of the problem. Some others start with an automated solution and then need the support of visualization to facilitate the interpretation of the output and to steer the algorithmic process. But also, some solutions are already tightly integrated in a way that it is not possible to see a predominant role of one over the other.

---

[1]From Wikipedia (http://en.wikipedia.org/wiki/Steve_Fossett): On September 8, the first of a series of new high-resolution imagery from DigitalGlobe was made available via the Amazon Mechanical Turk beta website so that users could flag potential areas of interest for searching, in what is known as crowdsourcing.

This last class is still in its infancy in terms of research advancements and it is the one that comes closer to the described concept of advanced visual analytics interfaces. A full realization of this vision would permit a tighter collaboration between the human and the machine in a way that both can take intermediary steps in the realization of a final outcome and exploit the best features of each component: the human for complex decisions the machine for hard computation.

3. ADVANCED VISUAL ANALYTICS EXAMPLES

3.1 Visual Document Analysis

With the rapid growth of Internet technologies, there are large numbers of customer reviews on the websites. Customers are invited to comment on what they liked or disliked about a product, often in a free-text format. For some products, a thousand or more reviews are available. Those resources are interesting for both customers that are interested in buying a product and companies that would like to learn from the feedback. Often customers are asked to give a total score (see e.g. the webpage of amazon.com), yet this score does not necessarily reveal the product's true quality and may provide misleading recommendations. An attribute of a product that was important for customer A and thus had an important impact on the total score that this customer gave might be irrelevant for customer B. Thus, the latter does not mind if this feature is not available in the product or is deficient. Similarly, it is not enough for a company to know which of their products customers liked best or least. In order to improve the products, they need to know in detail which features the customers were dissatisfied with.

3.1.1 Attribute-based opinion analysis

The analysis process consists of two parts: The automatic detection of opinions and the visual analysis of a set of reviews. Figure 4 exemplifies the automatic algorithm.

Step 1: As a first step towards detecting opinions, the opinion-signal words and the attributes that are frequently commented on in the reviews are identified. This is done with the help of dictionaries that may or may not be automatically generated. In our tool, we use an automatic algorithm for detecting the attributes but let the user refine the result. This comes with the advantage that the most frequent product attributes are surely considered, but that the user can still adapt the analysis to his or her specific goals. Figure 4 shows a sample review in which attributes and opinion-signal words are highlighted.

Step 2: Next, a mapping between the attributes and the opinion-signal has to be performed. This means that we have to determine, which opinion-signal words refer to which attribute. A central assumption in this process is that the closer attributes and opinion words are, the more likely it is that they refer to each other. Equation 1 is applied to determine the opinion score for an attribute which is a sum of the polarities of all opinion-signal words that are in the same sentence S as the attribute, weighted with the distance between the attribute A and the opinion word o.

$$opinion\text{-}score(A,S) = \sum_{o \in S} dist\text{-}weight(A,o) \cdot polarity(o) \quad (1)$$
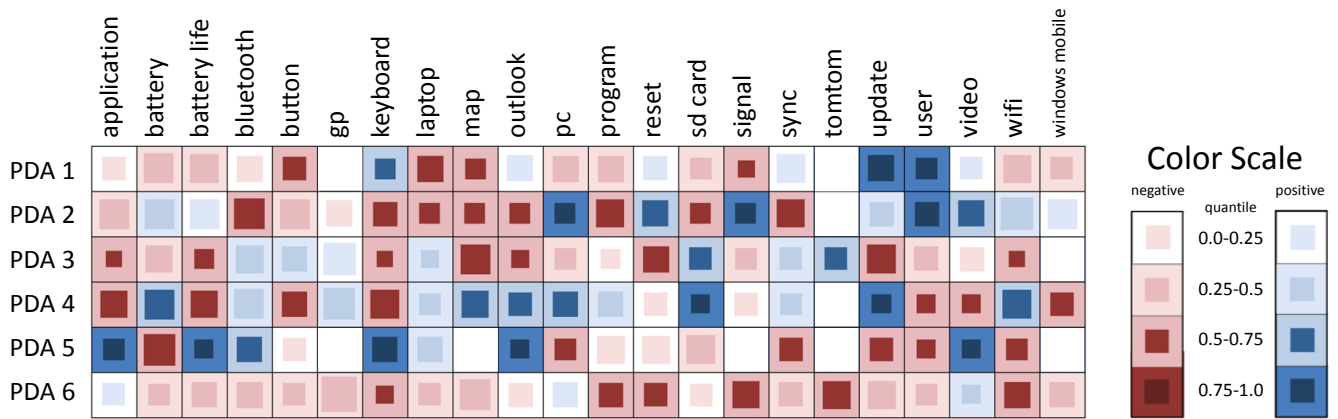
Figure 3: Summary Report of PDA reviews: Each row shows the attribute performances of a specific PDA. Blue color represents comparatively positive user opinions and red color comparatively negative ones (see color scale). The size of an inner rectangle indicates the amount of customers that commented on an attribute. The larger the rectangle the more comments have been provided by the customers.

Color Scale

	negative	quantile	positive
		0.0-0.25	
		0.25-0.5	
		0.5-0.75	
		0.75-1.0	

Attribute columns: application, battery, battery life, bluetooth, button, gp, keyboard, laptop, map, outlook, pc, program, reset, sd card, signal, sync, tomtom, update, user, video, wifi, windows mobile

Rows: PDA 1, PDA 2, PDA 3, PDA 4, PDA 5, PDA 6

Step 1: Identification of attributes and sentiment

I feel obligated to counter the bad reviews. This **printer** is just fine. I don't know what people are complaining about regarding the **software**, but it installed seamlessly and is intuitive in its operation. Even though the **paper tray** jams sometimes altogether I am happy that I bought this wonderful **printer**.

Step 2: Mapping between attributes and sentiment

[...] Even though the **paper tray** jams sometimes altogether

I am happy that I bought this wonderful **printer**.

Step 3: Determining the overall sentiment of an attribute

[...] This **printer** is just fine. [...] Even though the paper tray jams sometimes altogether I am happy that I bought this wonderful **printer**.

➔ Overall sentiment for *printer*: positive

Resulting Feature Vector

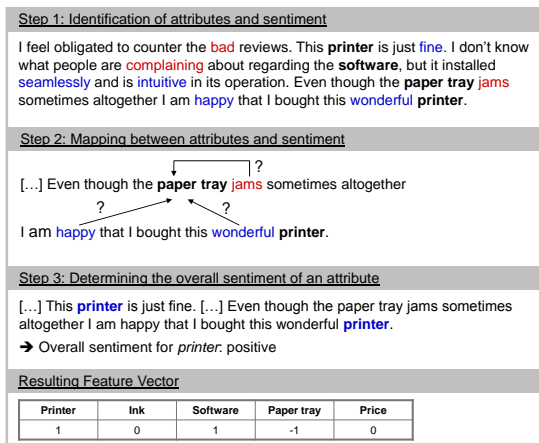| Printer	Ink	Software	Paper tray	Price
1	0	1	-1	0

Figure 4: Different steps in the automatic opinion detection process. Attributes are highlighted in bold face. opinion-signal words are colored in blue if they are positive and in red if they are negative.

Step 3: In our application scenario, we do not only want to know whether an attribute was mentioned positively or negatively in a specific sentence. Instead, we are interested in the overall opinion that was expressed about the attribute in the review. If an attribute is mentioned several times in the review, the majority vote of the sentence polarities for this attribute is determined to get its opinion value on the review-level.

Result: As a result, we get a feature vector for each review that summarizes the expressed opinions on the individual attributes. For each attribute, there is one feature dimension in the vector. The corresponding value of the vector for a particular attribute's dimension indicates whether the attribute was mentioned positively (+1), negatively (-1) or neutrally / not at all (0).

For more details on the automatic analysis process, please refer to [7].

3.1.2 Visual Summary Reports

The technique that was introduced in 3.1.1 transforms the semi-structured data into a structured format. Remember that the approach generates a feature vector for each review that holds detailed information about which attributes were liked or disliked by a customer. However, if we stopped here (as many related approaches do), this would leave the user with thousands of feature vectors. There is clearly a need for supporting the last step of the analysis process - the interpretation of the results. In [7], we introduced visual summary reports - a compact representation of thousands of reviews that yet provides enough detail to derive comprehensive insights (see figure 3 for an example).

In a summary report, each line represents one group of reviews (e.g. all the reviews for one product or brand). The table structure contains one column per attribute. For each attribute extracted by our automatic algorithm, it is shown whether it belongs to the category of attributes with a positive tendency (blue) or the category with a negative tendency (red). The size of the inner rectangles is determined by the percentage of reviews that comment on the attribute signaling the importance that the analyst should give to this attribute in his or her evaluation. Color is mapped to the percentage of positive or negative opinions, respectively. Using our automatic analysis method, we calculate the average percentage of positive comments per attribute and use this as a threshold. Attributes whose percentage of positive comments is above that threshold exhibit a positive tendency compared to the other attributes (color = blue). The ones that are below the threshold show a negative tendency (color = red). The stronger the positive or negative tendency is the darker the color value becomes. The intervals for the four shades of blue / red tones are determined by the quantiles of the set of positive or negative attributes.

3.1.3 Application example

Figure 3 shows a summary report in which the reviews on PDAs are visualized. This allows a detailed analysis of strengths and weaknesses of different PDAs. For some attributes, clear differences in the average rating can be perceived. This is, for example, the case for the attributes *key-*

board and *map*. Some features are only present in a subset of the products. If an attribute was rated zero times for a product, the corresponding cell is colored in white. For example, the GPS navigation system *tomtom* only seems to be included in two of the analyzed PDAs. For those two products, the attribute was commented on many times with a very contrasting overall result.

Besides analyzing the differences between the products with respect to certain attributes, comparing different products (rows) across all lines can be interesting as well. For example, it can be perceived that PDA 6 is the one that got the lowest overall rating. The line above it (PDA 5) is eye-catching, because its rating is often reverse compared to the other lines. Attributes that were rated positively for this product are often rated negatively for the other products and vice versa.

3.2 Geo-Spatial Temporal Visual Analysis

Movement data hold and link together valuable and multifaceted information about moving objects, properties of space and time as well as events and processes occurring in space and time. In general terms within this domain, we aim at finding efficient methods to describe the spatio-temporal distribution of events. In particular, we combine automatic and visual techniques that support analysts in defining when and where events occur.

The source of the data for the current analysis is the geographically referenced photos from the Flickr photo-sharing web site [2]. The geographical positions are specified by the photo owners when they post their photos. The times when the photos were taken can be retrieved from the image files. Many Flickr users repeatedly post their geo-referenced photos taken in different places and at different times. The geographic locations and times of the photos reflect the movements of the photographers (i.e., Flickr users who published the photos). The position records correspond to the events of taking photos by Flickr users.

3.2.1 Community Contributed Space - Time Referenced Data

In this current application scenario, we explore the frequencies and temporal patterns of the photo taking events that occur at different places. The data have been automatically downloaded from the Flickr web site using a publicly available API and an approach similar to web crawling. The data are stored in a relational database. The records include the coordinates and time stamps of the photos as well as the photographers' identifiers. As a preprocessing step, we extracted photographs that have geographic and temporal information attached. To create an appropriate set of places, we apply spatial aggregation to the original positions of the photos using the method described in [1]. We perform spatial clustering of the events and build areas around the extracted clusters by means of Voronoi tessellation of the territory, as suggested by Okabe et al. [8]. The centroids of the Voronoi cells are used for generating points of interests, referring to *places* in the analysis.

3.2.2 Visualization of Spatio-Temporal Events

We apply the visualization technique called *Growth Ring Map* in order to obtain an overview of the spatio-temporal distribution of the events in a single display [2, 4]. Each

place (in the current example, the centroid of the Voronoi cells) is taken as the center point for the computation of growth rings. The pixels – representing the photos taken in this place – are placed around this place in an orbital manner resulting in so called *Growth Ring* representations. The pixels are sorted by the date the photo was taken on: the earlier the photo was taken, the closer is the pixel to the central point. When two or more neighboring growth rings are about to overlap, the layout algorithm displaces the pixels in such a way that none of them is covered by another pixel. Hence, when big clusters of photos are close in space, the corresponding growth rings will not have perfectly circular shapes but will be distorted. In this technique, each photo is represented and placed on a cartographic background under investigation. A detailed description on the computation and application of Growth Ring Maps can be acquired from Bak et al. [2].

3.2.3 Application Example

In the following example we use a subset of the Flickr-data referring to the territory of Switzerland and the period from January 1, 2005 to September 30, 2009. On this dataset, after conducting the described preprocessing steps and the methods to create places of interest, we applied the Growth Ring Map technique to assess photographers' spatio-temporal behavior. The resulting Growth Ring Map shows simultaneously the number of photos taken in different places (size of the growth ring), the seasonal differences (color hue) and the yearly periodicity of the events (color intensity). A fragment of the map is shown in Figure 5 together with a legend explaining the color mapping and a set of patterns to show the effectiveness of the method. In this particular analysis we apply a temporal aggregation: the time stamps of the photos are transformed into years and seasons. The resulting temporal information is encoded with color. To investigate seasonal differences between places in terms of the photo taking activities, we map the seasons to four distinct color hues: winter-blue, spring-green, summer-red, and autumn-orange. The year is mapped to five different intensity levels of the colors assigned to the seasons. Pale colors represent earlier years, more saturated colors more recent years. The map exposes a number of temporal patterns of event occurrences in different places.

There are growth rings, in which one color hue strongly prevails, meaning that the events in these places mostly occurred in one season and year, as shown in Pattern A. Such a pattern indicates a festival that took place only in one specific year and place.

The growth ring marked B demonstrates one more pattern: dominance of particular color intensity, while the hues differ (middle intensity green, red and orange). This means that the photos were taken in the spring, summer and autumn of one year. The photographers were attracted by a particular happening for a limited duration of time that do not occur regularly in these places. A closer look revealed a "car cemetery", which represented a main attraction in the area for around one year until it was removed.

The presence of different shades of the dominant color means that the events occurred in the same season during different years. For example, Pattern C consists mainly of blue-colored rings with different color intensities, and in fact it represents a ski resort where most probably people take pictures only during winter.
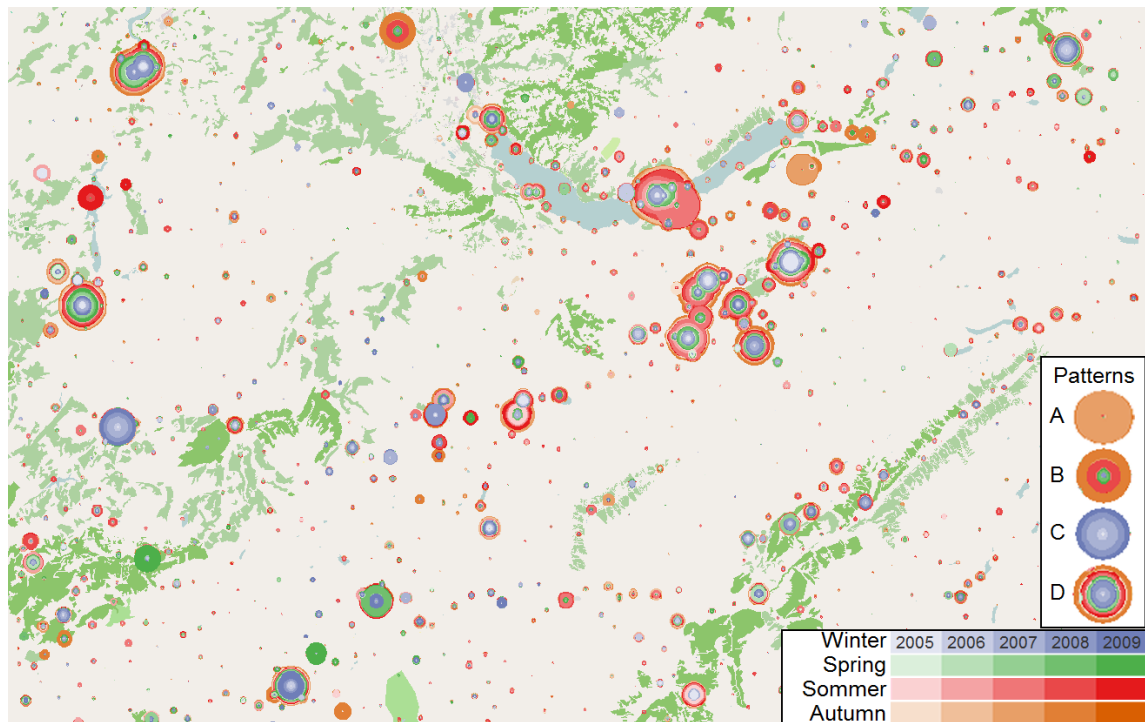
Figure 5: A fragment of the Growth Ring Map showing the spatio-temporal distribution of the Flickr photos taken in Switzerland. Each pixel represents a photo taking event, where its temporal properties are mapped to color. A set of user selected patterns is shown in the legend.

The growth rings, in which many of the color hues and shades are present, indicate a stable interest of photographers to the respective places: the photos were taken during (almost) all years and seasons. Pattern D shows such an example, in fact it represents a main city where people take pictures throughout years and seasons.

3.3 Visual Financial Analysis

3.3.1 Application example

Large parts of our economy and wealth depend on the functionality of the financial sector, as we recently experienced during the crash of the real estate market and the banking and economy crisis. The financial market, with all its interconnected components, is a highly complex system in which it is very difficult to uncover relevant relationships and extract reliable patterns. One reason is the many different products that are offered, such as: stocks, bonds, treasure bills, options, mutual funds, futures, or commodities, and often several of them are combined to even more complex financial products. In addition, factors such as different currencies, market sectors, and geographic regions make these data even more complex and entangled. This complex network with more than a hundred thousand different assets and hundreds of currencies and countries is used by millions of participants worldwide on a daily basis, generating data streams of millions of transactions every day. With modern computer technology that evolved during the last decade, financial data suppliers today transmit data of the stock market with an impressive rate of 50.000 data updates per second at peak times by connecting the stock exchanges and other heterogeneous sources with each other. This massive amount of data is too large to visualize with common visualization techniques, and too complex and too unstructured to only use automatic methods. Therefore, in order to generate meaningful knowledge and to obtain insight and understanding into the financial market, a combined approach of automatic and visual measures that overcomes the limitations of each single field seems to be an adequate and promising solution.

3.3.2 Computation of financial performance measures

One common question investors pose for their analyses is about the performance of a share S over several years compared to the market. Even for this very simple question one needs to compare the time series of S with all other shares in the database, and thus have to include the whole database. If we have the time series on a daily basis for 15 years, the time series consists of 3750 days, which results in $n*(n-1)/2$ = 7.029.375 combinations of "point of purchase" and "point of sale". For each of these seven million time interval combinations, we have to compare the performance of S with about 50.000 other shares, and to determine how well S has performed compared to them. For this, we compute a ranking function, and determine the position of S in the sorted list of 50.000 growth rates. Instead of only computing the performance (growth) over time, it is also possible to use other measures such as the volatility or sharp ratio, which increase the amount of computation by one more orders of magnitude.

3.3.3 Visual analysis of advanced computations

The massive computation of statistical and mathematical measures such as those mentioned in 3.3.2 can only be
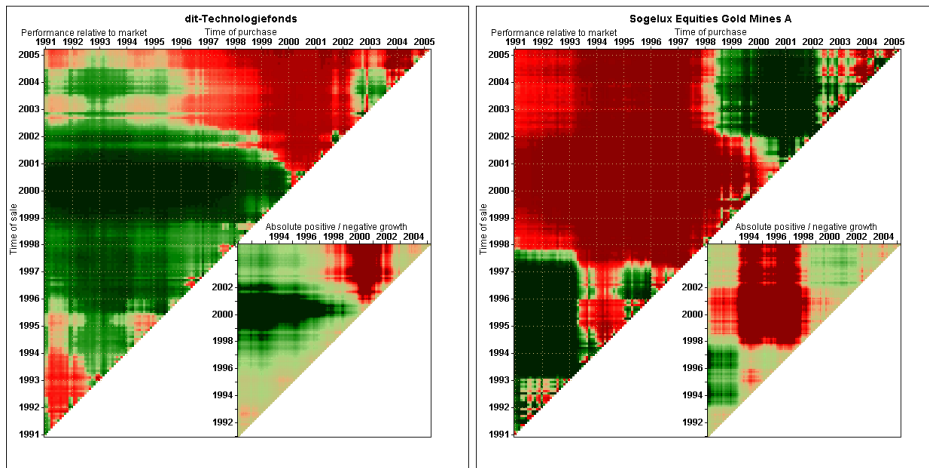
Figure 6: Visual analysis of financial data with Growth Matrices[6]. On the left side, we can see the impact of the dot.com-crisis in 2002 on technology funds. On the right side, we can see that at the same time of the impact of the dot.com-crisis, the shares for gold mines have been the top performers on the market. Both fingerprints reveal an overall reverse characteristic.

performed by a computer. However, the second question is how to generate a meaningful output for the human that answers the question that we posed in the beginning of 3.3.2 in an understandable way. In Fig. 6, we solved this problem using a pixel-based triangulation matrix technique that covers all possible time intervals between "point of purchase" on the x-axis and "point of sale" on the y-axis[6]. Such a pixel-based triangle can easily represent millions of time intervals, if enough resolution is available. In the large triangle, we visualize how well a given share S performed to all other shares in the market, by using a ranking function. Dark green means that share S performed better than most of the other 50.000 shares in the database, dark red means that share S performed worse than most of the shares in the database.

The small triangle represents the absolute performance of a stock (green=profit, red=loss). With this technique, each share generates a unique fingerprint that reveals its relative performance compared to all other shares on the market.

Fig. 6 shows an example of a financial analysis application. In this example, we compare a technology fund and a fund for gold mines. By inspecting the large triangle we can easily see that the dot.com-crisis in 2000 hit the technology fund, in fact it has been one of the worst performing funds in the database. In contrast, the funds for gold mines turned out to be one of the best funds during that period. However, if we take a closer look at the small triangles (the absolute revenue), we can see by the light green color that even the gold mines funds did not return much for the investment. We can also see that large differences in time tend to create horizontal or diagonal shapes from which we can derive how long it took to compensate a loss or to destroy a profit. We can also see many small red and green areas directly along the diagonal axis reflecting the short time volatility. The tool allows to select shares and funds from a database by name, market sector, or country, and to compare them one to each other. The application is fully scalable, so it is possible to compare several hundred triangles at a time, to refine the search, to add or remove triangles, and to in-
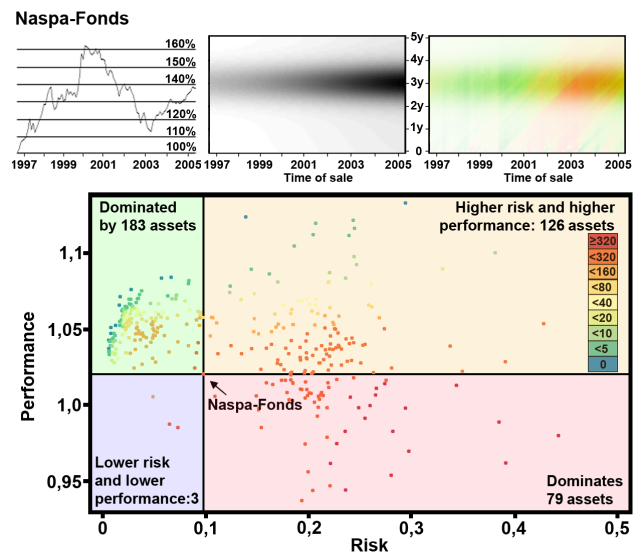

Figure 7: Interactive dominance plot[10]. After computing performance and risk of funds and dividing the result set into four sectors, the application allows to interactively investigate each fund. By moving the mouse cursor onto the dots the application shows additional information for each fund.

vestigate the remaining result. The combination of high computational capabilities with human judgment in an interactive tool results in a typical Visual Analytics interface where the user can investigate the data, and generate new domain knowledge.

In addition to that, the application also allows to perform performance/risks analysis [10]. The Dominance Plot in Fig. 7 visualizes the distribution of performance and risk values for a sample set of 392 funds from our database, for a user-specified Weight Matrix displayed in the upper middle of

the image. The selected fund in the middle, in this case the
"Naspa-Fonds", divides the image into four sectors regarding
higher/lower performance, and higher/lower risk. The 183
funds in the upper left sector dominate the "Naspa-Fonds"
and are better investments because they have higher perfor-
mance with less risk. By moving the mouse cursor onto the
points, the user can investigate each fund in detail.

4. CONCLUSIONS

In this paper we presented Advanced Visual Analytics In-
terfaces (AVAIs) as a new paradigm to consider in the de-
velopment of visual interfaces that help users make sense
of large data in complex environment. We argue that it
is necessary to go beyond the concept of advanced visual
interfaces and embrace the idea of a tighter collaboration
between the machine and the human to ensure a proper in-
tegration of computational and reasoning capabilities. The
main message consists in the idea that in AVAIs neither the
analytics nor the visualization are necessarily advanced but
their integration is advanced because it solves problems that
these components alone just cannot solve. The examples we
show in the paper are initial realizations of this concept but
much more research is needed to realize the full vision of
visual analytics and AVAIs. Especially, a tighter integration
between computational and interactive methods is foreseen
as a way to create more powerful solutions able to solve the
highly complex and critical problems we are confronted with
in modern society.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] N. Adrienko and G. Adrienko. Spatial generalisation
    and aggregation of massive movement data. *IEEE
    Transactions on Visualization and Computer
    Graphics*, 2010 (to appear).

[2] P. Bak, F. Mansmann, H. Janetzko, and D. Keim.
    Spatiotemporal Analysis of Sensor Logs using Growth
    Ring Maps. *IEEE Transactions on Visualization and
    Computer Graphics*, 15(6):913–920, 2009.

[3] E. Bertini and D. Lalanne. Investigating and reflecting
    on the integration of automatic data analysis and
    visualization in knowledge discovery. *SigKDD
    Explorations Journal*, December 2009 (to appear).

[4] D. Keim, M. Hao, U. Dayal, H. Janetzko, and P. Bak.
    Generalized scatter plots. *Information Visualization*,
    2009 (to appear).

[5] D. A. Keim, F. Mansmann, and J. Thomas. Visual
    analytics: How much visualization and how much
    analytics? *SigKDD Explorations Journal*, December
    2009 (to appear).

[6] D. A. Keim, T. Nietzschmann, N. Schelwies,
    J. Schneidewind, T. Schreck, and H. Ziegler. A
    spectral visualization system for analyzing financial
    time series data. In *EuroVis 2006:
    Eurographics/IEEE-VGTC Symposium on
    Visualization*, 2006.

[7] D. Oelke, M. Hao, C. Rohrdantz, D. A. Keim,
    U. Dayal, L.-E. Haug, and H. Janetzko. Visual
    Opinion Analysis of Customer Feedback Data. In
    *IEEE Symposium on Visual Analytics Science and
    Technology (VAST 2009)*, pages 187–194, 2009.

[8] A. Okabe, B. Boots, and K. Sugihara. *Spatial
    tessellations: concepts and applications of Voronoi
    diagrams*. John Wiley & Sons, Inc., New York, NY,
    USA, 1992.

[9] J. Thomas and K. Cook. *Illuminating the Path:
    Research and Development Agenda for Visual
    Analytics*. IEEE Computer Society Press, 2005.

[10] H. Ziegler, T. Nietzschmann, and D. A. Keim.
    Relevance driven visualization of financial
    performance measures. In *EuroVis 2007:
    Eurographics/IEEE-VGTC Symposium on
    Visualization*, 2007.