A Comparison of Several Key Information Visualization Systems for Secondary Use of Electronic Health Record Content

***Francisco S. Roque[1], Laura Slaughter[2,3], Alexandr Tkatšenko[4,5]**

[1]Center for Biological Sequence Analysis, The Technical University of Denmark, Lyngby, Denmark
[2]The Interventional Center, Oslo University Hospital, Oslo, Norway
[3]Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[4]Institute of Computer Science, University of Tartu, Tartu, Estonia
[5]Software Technology and Applications Competence Center, Tartu, Estonia
http://dsv.su.se/hexanord
*All three authors contributed equally to this work.

Abstract

An overview is provided of six information visualization systems designed specifically for gaining an overview of electronic health records (EHR). The systems discussed all make use of timelines: Lifelines, Lifelines2, KNAVE II, CLEF Visual Navigator, Timeline, and AsbruView. With the exception of Lifelines2, the main user groups targeted are physicians involved in direct patient care. Little attention has been paid towards supporting true secondary use of EHR contents, for activities such as assessing quality of care, patient health and safety monitoring, and clinical trial recruitment. Future work on such systems needs to address the complexity of EHR data, missing and incomplete information, and difficulties in displaying data with differing levels of granularity.

1 Introduction

This paper provides an overview of several information visualization (infovis) systems that have been built for exploring abstracted information from Electronic Health Records (EHR). EHRs are systems that are used to document care of patients. The records can include a wide range of data and information, including medications prescribed and administered, immunization history, laboratory test results, allergies, radiology images, treatment plans, and care notes. Currently, most EHR systems implemented are proprietary and highly customized when used by larger care institutions.

It is usually the case that only clinicians and other healthcare professionals with direct responsibility for providing care have access to patient data. The suggestion of secondary use of health data is not new and has been handled separately from the issue of creating user interfaces and visualizations. Safran et al. discuss the purpose of clinical data repositories in their white paper and point towards the goal of a national framework for the secondary use of health data in the U.S. (Safran et al., 2007). According to their definition, secondary use includes activities such as analysis, research, quality and safety measurement, public health, payment, provider certification and accreditation, marketing, and general business applications, while at the same time taking into account the ethical, political, technical and social implications of such re-use. De Lusignan and van Weel highlight the challenges of making use of clinical data for research, stating, "The available research methods for working with large data sets are limited; it is difficult to infer meaning from data; there is a rapid pace of change in both medicine and technology; and integrating data without reliable unique identifiers is difficult." (de Lusignan and van Weel, 2006). Prokosch and Ganslandt have recently summarized the latest advances in enabling clinical data re-use for research purposes (Prokosch and Ganslandt, 2009). They identify as key challenges the establishment of comprehensive clinical data repositories, the establishment of professional IT infrastructure to sup-

Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents, pages 76–83,
Los Angeles, California, June 2010.

port clinical data capture, and the integration of medical record systems and clinical trial databases. As discussed in these articles, aggregated, abstracted and manipulable information is underutilized and hard to come by.

The emerging field of *Visual Analytics* (Keim, 2008) is relevant to this review. This field is focusing on combining related research areas such as visualization, data mining and statistics to handle large and heterogeneous volumes of data, such as EHR. The systems we encountered are integrating human judgment with automated analysis, suggesting that future work will be related to handling massive amounts of data that contains missing elements - including the results of textual analysis of records content.

1.1 Purpose

Our motivation for creating this overview is to compare and discuss some of the available information visualization/visual analytics tools and how are these used for secondary, i.e. for purposes other than direct patient care. This is a first step towards infrastructure and coordinating efforts to produce systems that are based on standard input formats, and meet the needs of specifically defined users. The reader of this overview is most likely working on information extraction, temporal abstraction, and summarizing EHRs.

Source	*Search Keywords*
Pubmed	visualization health records Medical Records Systems, Computerized Computer Graphics User-Computer Interface
ACM DL	electronic health records or medical record information visualization or visualization healthcare or health care user interface
IEEE DL	visualization medical records
Google Scholar	electronic medical records or EHR information visualization visual analytics

Table 1. Keywords searched.

1.2 Scope

The review is non-systematic. We didn't expect to find large numbers of articles, since this is a relatively narrow area of interest. The search was confined to user interfaces and visualizations for EHR data, we searched pubmed, ACM digital library, IEEE library, and Google Scholar, using basic keywords and checked references in found articles. We also looked for papers on work we had read or known about previously from conferences or other sources. The literature search covered articles in English only. Keywords used are listed in Table 1.

2 Systems

In this section, we give an overview of the state-of-the-art systems related to visualization of temporal information in EHRs. Our intention is to cover broad areas of application including representation of medical histories, visual data query and aggregation, generation of temporal abstractions and visualization of treatment plans. Due to the limitations in space, we focus only on the most representative systems, which feature interesting and potentially reusable visualization techniques.

Lifelines

LifeLines uses a timeline visualization technique to represent personal histories, medical records and other types on biographical data (Plaisant et al., 1996). In LifeLines, horizontal bars are used to depict temporal duration and location of events on a horizontal time axis. Similar events are organized into facets, which can be expanded and collapsed to provide increasing or decreasing level of detail. Color notations and line thickness are used to indicate the importance and relationship of events. To handle regions with high data density, LifeLines provides zooming functionality allowing users to compress and stretch the time scale at any location. Additional content (e.g., multimedia) can be added in a linked fashion. Authors apply LifeLines in the analysis of complex patient medical records to visualize temporal relationships between treatments, consultations, disorders, prescriptions, hospitalizations and other events.

Lifelines2

LifeLines2 (Wang et al., 2008) is an extension of LifeLines, allowing the user to analyze records from multiple patients at a time. The system facilitates comparative visualization of records by means of aligning, filtering and sorting operations. By aligning patient records on some common reference event (e.g., the first heart attack), users can easily spot co-occurring and neighboring events. Ranking and filtering operations complement alignment by interactively reordering or narrowing the set of records to suit a user's changing focus. The system proved to be particularly suitable for observational research, where researchers analyze data from different studies in order to better understand health problems or study the effect of treatments, and in finding patients for clinical trials. Evaluation studies showed that the system significantly simplifies typical analytical tasks and that medical specialists can quickly learn the interface. LifeLines2 is currently used to display EHR data provided by the Informatics for Integrating Biology & the Bedside (i2b2) Project (Murphy et al., 2006).

While in LifeLines2 the main focus is on visualizing temporal ordering of events, Wang et al. emphasizes practical need in viewing multiple records as an aggregate in order to study frequency of event data over time (Wang et al., 2009). For instance, a user might be interested to analyze blood pressure of all patients who have had an open-heart surgery within 3 months of their first heart attack. As a solution, authors complement LifeLines2 framework with a new visualization technique, called temporal summaries, which represents distributional trends of events over a set of records in a histogram-like chart. Furthermore, the system allows splitting the whole dataset of records into multiple subsets and use temporal summaries to compare event patterns between these groups.

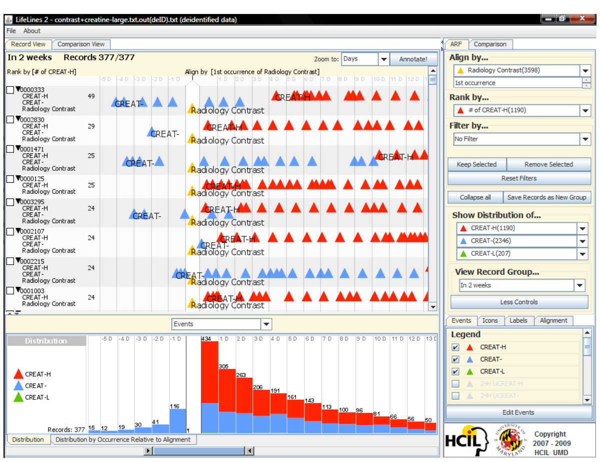

Figure 1. The Lifelines2 main window, with focus on timelines.

CLEF

Hallet (Hallett, 2008) proposes a visualization architecture for browsing medical histories, which integrates visual navigation tools and automatically generated textual summaries. While the graphical interface facilitates interactive navigation, textual descriptions can, in addition, convey complex temporal information and display details that would otherwise be too complex for visualization components. Within the system, the patient's medical history is represented as a network of semantically and temporally organized events, which serves as an input for visualization and natural language generation components. The visual navigator depicts a high level overview of a patient's medical history by plotting events along three parallel timelines, corresponding to diagnoses, treatments and investigations. In addition to zooming time scale and detail-on-demand functionality, the navigator provides interactive visualization of semantical relationships between events (e.g., caused-by, has-locus, indicated-by, etc.). Having different features from the LifeLines interface, the navigator also allows the user to visualize numerical data (e.g., results of blood tests) by plotting results of measurements on separate line charts. Natural language generation is used for two purposes: 1) to create customized textual reports for printing or exchange purposes and 2) as a support tool for the visual navigator, to enable better description of complex events and relationships between them.

KNAVE-II

KNAVE-II (Goren-Bar et al., 2004) is an interface enabling knowledge-based visualization and interactive exploration of time-oriented data at different levels of temporal abstractions (e.g., abstraction of periods of bone marrow toxicity from raw individual hematological data). Users can navigate through the links of a semantic network while simultaneously navigating visually through multiple degrees of temporal abstraction of the dataset under observation. The evaluation results have shown that users of KNAVE-II were able to perform queries both faster and more accurately than with other standard tools.

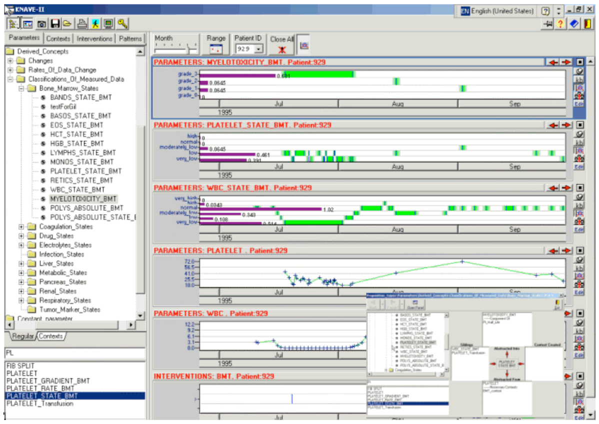

Figure 2. The Knave-II system.

TimeLine

The TimeLine system (Bui et al., 2007) is a problem-centric temporal visualization of patient records. The contents of the EHR are integrated, reorganized, and displayed within the user interface (UI) along a timeline. It is similar to Lifelines in the way that the different elements of the EHR are grouped along the y-axis: imaging, reports, lab tests, etc are collapsible categories. However, unlike Lifelines, the TimeLine system uses an XML data representation to handle data from distributed, heterogeneous medical databases. Data elements that are displayed in the UI are classified based on a knowledge base that guides both data inclusion rules and the visualization metaphors used to render the data.

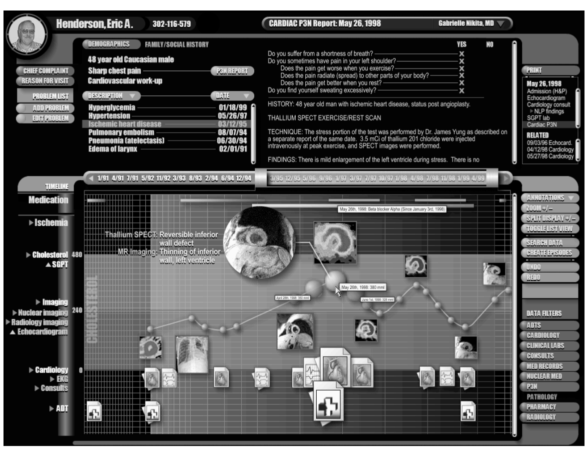

Figure 3. TimeLine system.

ASBRUVIEW

AsbruView (Kosara and Miksch, 2001) is a visualization and user interface on top of Asbru language (Shahar et al., 1996) designed to represent treatment procedures as structured time-oriented plans. AsbruView represents hierarchical and temporal relationships between treatment plans using a 3D visualization perspective. Plans are aligned along the time axis and can be stacked on top of each other and laid out in different ways. To simplify the interface, all graphic elements are represented by well-known real world objects (e.g., track, traffic light, etc.). Also a 2D view is available which focuses on temporal aspects of plans in greater detail. To depict uncertainty of future events, AsbruView extends the timeline by using time annotation glyphs (Chuah, 1997).

3 Comparisons

Infovis techniques are a way of augmenting human cognitive capabilities, to help humans find patterns in large volumes of data. The systems described above target specific user types that will benefit from the visualization methods. While some user interfaces were developed in close dialog with medical practitioners, like Lifelines2 and Knave-II, others, such as the first Lifelines, Clef and Asbruview have had only minimal input from their intended audience.

3.1 Users, Goals and Tasks

Most of the tools were directed at clinicians and clinical practice, although they were not always developed in close relation to them. Table 2 gives an overview of intended users for each of the named systems, and their proposed goals/tasks. From the user point of view, a number of tasks and goals can be defined for each tool. Some are very specific and tend to care for niche usages, while others provide more general visualization methods that can be applied to a number of situations.

System	*Users, Goals, Tasks*
Lifelines	Clinician Patient care Use EHR content in temporal time-based view
Lifelines2	Clinical researchers Research Compare patterns of events, detecting trends
CLEF	Clinician, Biomedical researcher Patient care Visualize timelines, use NLP to extract complex temporal data, aggregate numerical data
KNAVE-II	Clinician

	Patient care Generation and exploration of context sensitive abstractions of temporal data
TimeLine	Clinician Patient care Use EHR content in temporal time-based view, with additional filters on data based on NLP techniques
AsbruView	Clinician Patient care Medical therapy planning and execution

Table 2. Users, Goals, Tasks.

These systems were designed with input from only a few medical personnel involved in the project. In general, articles we read concerning these systems that have a more guided development process, i.e. closely related with physicians, have more specific goals and tasks, because they were designed with these in mind. Visualizing data for decision-making and analyzing treatment outcome is often a general goal in many of the tools developed in interaction with medical staff (Aigner and Miksch, 2006; Mamykina et al., 2004; Portet, 2009). There is an emphasis on pre-processed patient data, specifically numeric, such as lab tests, heart rate, and blood pressure. Systems mainly try to help physicians answer questions about correlations in the patient's data, and provide a means for supporting quick decision-making making when combining several types of highly heterogeneous data. Physicians can follow a specific treatment plan and check the patient's physiological variables over time. This also enables the practitioner to check the influence of certain variables in the treatment process and change the protocol if needed. CLEF, for example, allows the physician to discover events during specific time spans, such as searching for past specific liver problems. Lifelines2 is specifically geared towards research uses and towards answering complex queries. In Lifelines2, a case study involved verifying the results of a clinical study with real-life EHR data to see if clinical care data differ from the study results.

The systems we discuss have conducted evaluation studies as a part of the end-stages of development. The Lifelines evaluations were conducted in another domain (use of pattern searching related to monitoring graduate student progress), with limited interviews and input from experts in the medical domain. The KNAVE system conducted a cross-over study with doctors, comparing KVAVE with existing tools. TimeLine was evaluated following the development of the interface by five radiologists- focusing on questions related to data integration and the temporal display. Asbruview was evaluated using questionnaires sent to clinicians.

3.2 Visualization Methods

The focus of this paper is on temporal visualization methods since this has been the primary visualization type studied for aiding humans in organizing and exploring patterns in abstracted EHR content. All the systems that are compared in this paper display some type of timeline with time running from the left part the screen to the right, time being on the x-axis, and categories of events along the y-axis. Various techniques for graphically representing specific events are used (e.g. icons, shapes), AsbruView makes use of 3D, while all the others are flat 2D.

Infovis has been the keyword used to describe these systems, with the idea of presenting a method for human users (most often stated as being clinicians), to recognize patterns and thereby "amplify cognition" (Chittaro, 2006). Other methods for recognizing patterns in EHR for secondary use are purely automated and conducted through data mining techniques. Bertini and Lalanne (2009) wrote about the complementary role of automatic data analysis and visualization in knowledge discovery. They discuss "visual analytics", an outgrowth of infovis that can be seen as an integrated approach combining visualization, human factors, and data analysis. They suggest 4 categories for classifying approaches: Pure Visualization (VIS), Computationally-enhanced Visualization (V++), Visually enhanced Mining (M++), and Integrated Visualization and Mining (VM). In the systems we have compared, there is a spectrum of ideas about how to visualize EHR contents, including movements towards "enhanced" or "intelligence" in the processing of the underlying EHR data. In Lifelines2, the data visualized was obtained from anonymized EHRs though cooperation with the i2b2 Project (Murphy et al., 2006) The input form of the data is a simple 3-column table containing "ID", "Event Type", and "Time". Each ID can have multiple events happening at various times. Lifelines2 allows sorting of the data so that records with the most incidents of one type of event are shown at the top of the screen. This type of infovis relies on human pattern recognition only and would be considered as "VIS" by Bertini and Lalanne (2009). In the CLEF project, the CLEF Chronicle, which underlies the visualizations, is a semantic network modeling of what happened to the patient, why, and how. Semantic relations are: causality, reason, finding, and consequence. The types of events modeled are: problem, investigation, and treatment. The CLEF Visual Navigator might be considered as "V++", computationally enhanced visualization because some sort of automated computation supports the visualization. In CLEF, the visual display is "enhanced with visual techniques for highlighting relationships between events on the timeline." None of the systems so far that we have seen, would qualify as "visually enhanced mining" or "integrated visualization and mining." Table 3 provides a full overview for all systems reviewed.

System	*Category*	*Notes*
Lifelines	VIS	
Lifelines 2	VIS	
CLEF	V++	• automated generation of summaries • semantic network of EHR record events
KNAVE-II	V++	• semantic (ontology-based) navigation and exploration of the data • knowledge base is used to interpet raw data
TimeLine	V++	• data mapping and reorganization • content-based techniques to elucidate predominant subject of reports for classification
AsbruView	VIS	

Table 3. Visual Analytics of Systems using Bertini and Lalanne's (2009) classification.

The papers we have read that cover EHR visualization, as seen in the systems presented, express the complexity of abstracted EHR data. Missing and inconsistent data, dealing with hierarchical data, and problems with granularity are all concerns that become readily apparent through attempting to build infovis systems. Wang (2008) summed it up best "Clinical data tend to be messy with aspects that become only obvious when the data is visualized. The same heart attack might be recorded three times in three days (by the emergency room physician, a cardiologist, and a clerk from the billing office) and it can be hard to differentiate it from 3 separate events. Even if medical event information is carefully recorded at the time of the doctor visit or during a hospitalization, the time stamp is usually inaccurate by nature." Future work on visualizations needs to adequately address the complexity of the data rather than work with test data that is too simplistic.

3.3 Text Mining Tasks

All mentioned systems, except the CLEF and TimeLine, operate with readily available lists of type- and time-tagged events. However, clinical records are often stored in textual form what makes them inaccessible for machine processing. Text mining techniques need to be applied to automatically transform textual data into structured, normalized form. Key tasks involve event extraction, classification and normalization.

The CLEF system uses an advanced information extraction engine to identify pre-defined classes of entities (e.g. diseases, investigations, problems, drugs, etc.) and semantic relationships between them (e.g. investigation indicates problem) in natural language texts. The information extraction process involves lexical and terminological analysis, syntactic and semantic analysis, and discourse analysis. To address the complexity of medical language, the system makes use of language resources including the Unified Medical Language System and the Gene Ontology. Extracted information is stored in templates, which can be queued or used to generate textual summaries. The TimeLine system makes use of both textual contents of the EHR as well as numerical data and codes. An NLP-based system is used in conjunction with the TimeLine UI, for example, performing section analysis in radiology reports to determine whether specific subsections exist within the reports that are related to certain medical problems (Bui et al. 2007).

4 Conclusions

The infovis systems analyzed allow secondary use of EHR content data especially aimed at clinicians documenting patient care. All of them are focused on visualizing temporal data in a timeline, while displaying specific events from the patient data.

Although directed at medical practitioners in their daily patient care routine, they were not always developed with user feedback. Evaluation of the different tools was often based on situations outside of the clinical setting, and might not reflect reality. A more intimate dialog with clinicians would benefit the creation of targeted systems addressing specific needs of the medical community.

The overall goal of these tools is to present users temporal information contained in a record, improving their ability to recognize patterns for knowledge discovery and following treatment. They introduce simple visualization tools, but some include automated computational enhancements supporting it.

EHR contain missing and inconsistent data, which is in general messy. Due to the complexity of the underlying data, future work needs to address these intricacies rather than using simplistic approaches.

Acknowledgments

We would like to thank Nordforsk and the Nordic Council of Ministers for the funding of our research network HEXAnord - Health text Analysis network in the Nordic and Baltic countries. This work was partially supported by a grant from the Villum Kann Rasmussen fund.

References

Aigner, W., Miksch, S., 2006. CareVis: integrated visualization of computerized protocols and temporal patient data. Artif Intell Med 37, 18.

Bertini, E., Lalanne, D., 2009. Surveying the complementary role of automatic data analysis and visualization in knowledge discovery, Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration. ACM, Paris, France, pp. 12-20.

Bui, A., Aberle, D.R., Kangarloo, H., 2007. TimeLine: Visualizing Integrated Patient Records. IEEE Transactions on Information Technology in Biomedicine 11(4): 462-473.

Chittaro, L., 2006. Visualization of patient data at different temporal granularities on mobile devices, AVI '06: Proceedings of the working conference on Advanced visual interfaces, pp. ACM--487.

Chuah, M.a.E., S., 1997. Glyphs for software visualization. 5th International Workshop on Program Comprehension (IWPC ’97) Proceedings, 183-191.

de Lusignan, S., van Weel, C., 2006. The use of routinely collected computer data for research in primary care: opportunities and challenges. Fam Pract 23, -263.

Goren-Bar, D., Shahar, Y., Galperin-Aizenberg, M., Boaz, D., Tahan, G., 2004. KNAVE II: the definition and implementation of an intelligent tool for visualization and exploration of time-oriented clinical data, AVI '04: Proceedings of the working conference on Advanced visual interfaces, pp. ACM--174.

Hallett, C., 2008. Multi-modal presentation of medical histories, IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces, pp. ACM--89.

Keim, D.A., Mansmann, F., Schneidewind, J., Thomas, J., Ziegler, H., 2008. Visual analytics: Scope and challenges. Visual Data Mining: Theory, Techniques and Tools for Visual Analytics, Springer, 76-90.

Kosara, R., Miksch, S., 2001. Metaphors of movement: a visualization and user interface for time-oriented, skeletal plans. Artif Intell Med 22, 111-131.

Mamykina, L., Goose, S., Hedqvist, D., Beard, D.V., 2004. CareView: analyzing nursing narratives for temporal trends, CHI '04: CHI '04 extended abstracts on Human factors in computing systems, pp. ACM--1150.

Murphy, S.N., Mendis, M.E., Berkowitz, D.A., Kohane, I., Chueh, H.C., 2006. Integration of clinical and genetic data in the i2b2 architecture. AMIA Annu Symp Proc.

Plaisant, C., Milash, B., Rose, A., Widoff, S., Shneiderman, B., 1996. LifeLines: Visualizing Personal Histories, CHI, pp. -227.

Portet, F., Reiter, E., Gatt, A., Skyes, C., 2009. Automatic generation of textual summaries from neonatal intensive care data. Artificial Intelligence, 789-816.

Prokosch, H.U., Ganslandt, T., 2009. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods Inf Med 48, -44.

Safran, C., Bloomrosen, M., Hammond, W.E., Labkoff, S., Markel-Fox, S., Tang, P.C., Detmer, D.E., Panel, E., 2007. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc 14, -9.

Shahar, Y., Miksch, S., Johnson, P., 1996. An intention-based language for representing clinical guidelines. Proc AMIA Annu Fall Symp, 592-596.

Wang, T.D., Plaisant, C., Quinn, A.J., Stanchak, R., Murphy, S., Shneiderman, B., 2008. Aligning temporal data by sentinel events: discovering patterns in electronic health records, CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, pp. ACM--466.

Wang, T.D., Plaisant, C., Shneiderman, B., Spring, N., Roseman, D., Marchand, G., Mukherjee, V., Smith, M., 2009. Temporal Summaries: Supporting Temporal Categorical Searching, Aggregation and Comparison. IEEE_J_VCG 15, -1056.