

**VAST: A HUMAN-CENTERED, DOMAIN-INDEPENDENT VIDEO ANALYSIS
SUPPORT TOOL**

A Dissertation

by

MARLO FAYE NORDT

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2008

Major Subject: Computer Science

**VAST: A HUMAN-CENTERED, DOMAIN-INDEPENDENT VIDEO ANALYSIS
SUPPORT TOOL**

A Dissertation

by

MARLO FAYE NORDT

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Richard K. Furuta
Committee Members,	Susan A. Bloomfield
	John J. Leggett
	Frank M. Shipman
Head of Department,	Valerie E. Taylor

December 2008

Major Subject: Computer Science

ABSTRACT

VAST: A Human-Centered, Domain-Independent Video Analysis Support Tool.

(December 2008)

Marlo Faye Nordt, B.S., Stephen F. Austin State University

Chair of Advisory Committee: Dr. Richard K. Furuta

Providing computer-aided support for human analysis of videos has been a battle of extremes. Powerful solutions exist, but they tend to be domain-specific and complex. The user-friendly, simple systems provide little analysis support beyond basic media player functionality. We propose a human-centered, domain-independent solution between these two points.

Our proposed model and system, VAST, is based on our experience in two diverse video analysis domains: science and athletics. Multiple-perspective location metadata is used to group related video clips together. Users interact with these clip groups through a novel interaction paradigm – *views*. Each view provides a different context by which users can judge and evaluate the events that are captured by the video. Easy conversion between views allows the user to quickly switch between contexts. The model is designed to support a variety of user goals and expertise with minimal producer overhead.

To evaluate our model, we developed a system prototype and conducted several rounds of user testing requiring the analysis of volleyball practice videos. The user tasks

included: foreground analysis, ambiguous identification, background analysis, and planning. Both domain novices and experts participated in the study. User feedback, participant performance, and system logs were used to evaluate the system.

VAST successfully supported a variety of problem solving strategies employed by participants during the course of the study. Participants had no difficulty handling multiple views (and resulting multiple video clips) simultaneously opened in the workspace. The capability to view multiple related clips at one time was highly regarded.

In all tasks, except the open-ended portion of the background analysis, participants performed well. However, performance was not significantly influenced by domain expertise. Participants had a favorable opinion of the system's intuitiveness, ease of use, enjoyability, and aesthetics. The majority of participants stated a desire to use VAST outside of the study, given the opportunity.

DEDICATION

To my best friend and sister,

Marcie

ACKNOWLEDGEMENTS

First and foremost, I am thankful to God for the opportunity to pursue a doctoral degree and for His unfailing grace, logistic support, and guidance throughout my life. I am nothing without Him.

For my parents and their tireless love, encouragement, support, and integrity, I am forever grateful. For my sister's friendship, words cannot describe how much I treasure it. Marcie, great things await you as you set out on your own for the first time in this adventure called life. (But remember, with God you are never alone.)

Dr. Furuta, thanks for your patience in putting up with me for the past five plus years. Your academic insight and knowledge has been key to my successes as a graduate student. I am also truly grateful for the professional opportunities you have made available to me. To my committee members, Dr. Frank Shipman, Dr. John Leggett, and Dr. Sue Bloomfield, thanks for the excellent critique and feedback. Dr. Tracy Hammond, thank you for standing in the gap during my proposal presentation and prelim. This dissertation could not have been completed without you all.

Dr. Bart Childs was instrumental in my decision to pursue a Ph.D. and to follow my interest in human-computer interaction – to him I owe special thanks.

Unmil and Neal, for your brutally honest but necessary feedback and the always entertaining but seldom relevant non-computer conversations, I thank you. Neither life in the lab nor this dissertation would be the same without you guys. To Takeisha, Jie,

Michael, Haowei, Carlos, Eunyee, Paul, Juwairia, Luis, Sarah, and all the other CSDLers thanks for the memories.

I also am indebted to Drs. Chris Quick and Sarah Gatson for opening up the Cardiovascular Systems Dynamics Lab (aka Bat Lab) to a computer scientist and providing a great work and research opportunity. To all the Bat Lab participants especially Jay, Missy, Ranjeet, Waqar, Arun, Ketaki, Phuc, Josh, and Rachel, thanks for your jargon-free explanations, access to your data, and system feedback.

I am grateful for all the informal and formal user study participants including those from the Texas A&M Men and Women's club volleyball teams. VAST is and will become better because of your comments.

To all my current and past professors thank you for providing me with a solid knowledge base from which to grow my scientific knowledge. I am also exceedingly grateful to the professors at Stephen F. Austin State University for providing an undergraduate experience and education above and beyond all expectations. A special thanks is due to Dr. George Dailey for his always-open ear, regardless of the question's significance, and to Dr. Robert Strader for expecting nothing less than my very best.

Becky, Jonathan, and Garrison, thanks for your refreshing friendship. Dale and Cindi, thank you for entrusting me with one of your beloved Labrador puppies, and to all the other dog people I have met along the way – thank you. Finally, to Lady 'Mic' Machaira, thanks for your unending enthusiasm for life and not eating my dissertation.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS.....	viii
LIST OF FIGURES.....	x
LIST OF TABLES.....	xii
1. INTRODUCTION	1
2. MOTIVATION.....	3
2.1 Bat Lab Context-Demanding Tasks	4
2.2 Bat Lab Work Practices and Constraints	5
2.3 Generalizability (Volleyball)	8
3. RELATED WORK	9
3.1 Basic Video Analysis.....	9
3.2 Video Annotation	11
3.3 Hypervideo and Hypermedia	12
3.4 Video Reality	14
3.5 Digital Video Libraries.....	16
4. LOCATION AND CONTEXT	18
4.1 The Inspiration	19
4.2 Location in the Bat Lab	20
4.3 Location in Volleyball.....	21
4.4 Summary	23
5. SYSTEM DESIGN	24
5.1 Database.....	25

	Page
5.2 Database Connector	27
5.3 View Library	29
5.4 Supporting GUI	34
5.5 GUI Engine	35
5.6 Implementation	36
6. USER INTERACTION DESIGN.....	38
6.1 Search	38
6.2 Workspace.....	40
6.3 Diagram	47
7. USER STUDIES.....	49
7.1 Study Design	49
7.2 Results.....	50
7.3 Participant Feedback	66
8. DISCUSSION.....	70
8.1 Analysis (Volleyball).....	70
8.2 Implications (Bat Lab).....	72
9. FUTURE WORK.....	75
10. CONCLUSION.....	78
REFERENCES.....	81
APPENDIX A USER STUDY SUPPORTING MATERIAL	88
APPENDIX B USER STUDY DATA AND ANALYSIS	105
APPENDIX C SOFTWARE DEMO.....	124
VITA.....	126

LIST OF FIGURES

FIGURE	Page
1 Vessel networks	20
2 Volleyball location outcome.....	22
3 VAST design overview	24
4 Domain-dependent location tables.....	26
5 The video view in its independent, dependent, and associated condition..	31
6 Conceptual arrangements and associated views in the Bat Lab domain....	32
7 Conceptual arrangements and associated views in the volleyball domain.	33
8 Location-based diagrams for the Bat Lab and the volleyball domains	35
9 VAST screen shot – Bat Lab domain.....	37
10 VAST graphical user interface – volleyball domain.....	38
11 Search area.....	39
12 Basic video view without mouse focus	40
13 Basic video view with mouse focus	41
14 Annotation in the basic video view	41
15 Basic image view without and with mouse focus.....	42
16 Basic video view popup menu	43
17 Internal and external focus in a compound view	45
18 A compound view zoomed-in.....	46
19 Moving within a compound view	47

FIGURE	Page
20 Location diagrams for two different volleyball clips	48
21 User log categories for task one – analyzing foreground	54
22 User log categories for task two – identification	59
23 Task three: Identifying people in the background	61
24 Task three: Locating and identifying flags in the gym.....	61
25 User log categories for task three – analyzing background.....	63
26 User logs for task four – planning.....	65
27 Participant <i>977pdg</i> task four’s sketch	65

LIST OF TABLES

TABLE		Page
I	Task Two Average Scores.....	57
II	Task Two Individual Scores	58
III	Task Two Individual Scores by Strategy.....	60
IV	VAST User Rating Tallies by Opinion Category	66
V	VAST User Rating Averages.....	67

1. INTRODUCTION

“The picture-examining eye is the best finder we have of the wholly unanticipated”

– John Tukey in defense of exploratory data analysis [1980].

Capturing an event or phenomena with video for later study is common practice in many domains including the sciences [Grémillet et al. 2006], medicine [Guerlain et al. 2004], athletics, and psychology [Clifford et al. 2007]. But easy capture of event information, via video, does not equate to easy understanding of that event. Various solutions have been proposed to enhance this comprehension. Unfortunately, most are either complicated domain-specific tools or simple, but less powerful, basic media players.

Many domain-specific tools assume their users will be interested in coding and counting every minor event incident captured in the video. This can be a time consuming process. It also ignores the needs of semi-formal analysis users (e.g., an athlete watching film to improve his game) who watch videos with a critical eye but are not concerned with determining statistical significance.

Basic media players (e.g., Windows Media Player or Apple’s QuickTime Player) work with any domain, most video formats, and require no user setup or overhead. The price of this simplicity is limited context. Users are forced to think in terms of the

This dissertation follows the style of *ACM Transactions on Computer-Human Interaction*.

capture medium (video files) and not the event itself. No built-in support exists for comparing or organizing multiple events. And the only clue regarding the event's importance or relevance is the video's filename.

In this dissertation we outline a solution between these two extremes. Using our experience in two different domains, the Bat Lab (science) and volleyball (athletics), we have developed a flexible, domain-independent system model for supporting video analysis. It uses a location-oriented approach to organize and provide context between multiple video streams and supports a novel interaction paradigm (*views*) for these videos. Recognizing real world constraints, it keeps producer overhead (i.e., metadata collection requirements) to a minimum. To verify the feasibility and appropriateness of the model, we have implemented VAST (video analysis support tool) and conducted several user studies using the software.

The remainder of this paper elaborates on these topics and is organized as follows. Section two frames the video analysis support problem in regards to the Bat Lab and volleyball domains. Section three examines related research, including basic analysis programs, video annotations, hypervideo/media, video reality, and video digital libraries. Section four presents multi-perspective location as a tool for organizing and providing context to the video collection. Section five explains the system design model and implementation. Section six expounds on the design from the end-user's perspective. Section seven focuses on the user studies and results. Section eight discusses the analysis and implication of the results. Sections nine and ten cover future work and conclusions, respectively.

2. MOTIVATION

The Cardiovascular Systems Dynamics Laboratory in the department of Veterinary Physiology and Pharmacology at Texas A&M University and part of the Michael E. DeBakey Institute for Comparative Cardiovascular Science and Biomedical Devices is called the Bat Lab because they use the pallid bat (*Antrozous pallidus*) as an animal model for studying the cardiovascular system. The bat's transparent wings allow researchers, with the aid of a microscope, to examine arteries, veins, lymphatics, and other cardiovascular phenomena *in vivo*. No anesthesia is necessary because the bats are trained to sleep with a wing extended, permitting researchers to examine the bat while it is napping. Every study involving the bats is video recorded for record keeping and analysis.

The videos collected in the Bat Lab are a world-unique scientific resource, but the current use of the videos falls short of their impact potential in several regards. The current video analysis process employed is an ad hoc, tedious, and individualistic process. Researchers use whatever video processing or editing software is available, from the complex LabVIEW program (where researchers must program their own data analysis extensions) to the simple but limiting Widows Media Player. Sharing experimental video information between research teams is practically nonexistent. After members of the research team perform the analysis, the DVD containing the video file is stored on the team leader's desk, never to be viewed again.

To improve and support video analysis and reuse in the lab, we first examine the analysis tasks the Bat Lab participants would like to address. Subsequently, we delineate the constraints we must work within while supporting the video analysis.

2.1 Bat Lab Context-Demanding Tasks

Most questions currently investigated by Bat Lab participants are quantifying in nature: *What is the diameter variation of a lymphatic vessel in a five minute period? How many white blood cells pass through a given area in sixty seconds?* These counting and measuring tasks require little human thought once the question is formulated and are good candidates for automation.

Another category of questions, while useful to Bat Lab, go unanswered because they are more difficult to address. They require a higher level of thinking and typically require referencing two or more sections of video. These video sections may be from one single experimental recording or from several different experiments.

We will refer to this category of questions as context-demanding because they require rearranging and/or adding to the information that a single video recording provides. Below are several such questions:

- *What does the surrounding vessel structure look like?* During the course of an experiment, several different locations on the bat wing may be recorded. When reviewing the video at a later time, it is challenging to place the different locations in context with one another as the video images are arranged in time, not space.

- *What are the differences between the control and experiment events?* This question is typically answered by examining measurements that are collected during the course of an experiment. But sometimes a picture really is worth a thousand words and simultaneously viewing both video feeds priceless.
- *Did these capillaries exist six months ago?* Research is an evolving process. Sometimes good questions are not formulated until after the data is collected, and the researcher notices an unusual phenomenon. Unfortunately, the researcher may not have collected video data from the area six months ago. The current lack of support for sharing or repurposing videos in the lab makes it impossible for a researcher to “go back in time” and discover if anyone else captured that area of interest.
- *Where is a good vessel occlusion site?* Microscope time is a limited commodity in the Bat Lab. Many exploring, training, and hypothesis forming questions could be answered by synergistically combining old videos.

Our primary focus will be on supporting these more difficult, context-demanding video analysis questions.

2.2 Bat Lab Work Practices and Constraints

Support for video analysis must be grounded in the real world. We highlight these constraints in the following paragraphs.

2.2.1 Diversity of Goals and Experience

The research component of the Bat Lab is lead by a cadre of graduate students, under the management of the director, each with their own research interest and direction. In keeping with the lab's goal to provide undergraduates with research experience, the graduate students often lead teams of several undergraduates. The undergraduates have very limited, if any, research experience or bat knowledge. The beginning of each semester is devoted to teaching these undergraduates about research procedures and bat physiology. While some undergraduates spend only one semester in the Bat Lab, others spend several semesters and eventually become part of the leadership cadre.

The lab also has an extensive local and distance outreach program directed toward K-12 teachers and students. The teachers are invited to spend part of their summer in the lab and are integrated into the research groups. The rest of their time is spent collaborating with other teachers regarding the incorporation of the Bat Lab experience into the classroom. During the regular school year, the teachers and their students can access the Bat Lab remotely. (For an elaboration on the remote access tools see [Nordt et al. 2006].)

With the constant stream of new participants in the lab, support must be provided for both domain experts and novices. A variety of projects, both research and teaching oriented, must be supported as well.

2.2.2 *Producer-Consumer Paradox*

Reusing video data in the Bat Lab has the great allure of distributing the original cost of collecting the video across many consumers. It provides access to those who do not have the opportunity to collect the original information (e.g., K-12 students) and has the potential to lessen the demand on microscope time. Sharing data also increases the likelihood of making a useful discovery and correcting inaccuracies in the original analysis. However, sharing rarely occurs because the cost of understanding and repurposing old video data is very high.

In practice, the original producer or researcher operates under the *rule of expediency*, collecting only the data that fits his specific need at that time. The future needs of other people, or even sometimes himself, are unknown and unimportant. The producer is solely focused on completing the task at hand. Requiring the producer to put more effort into the collection process (i.e., collect more metadata) than necessary to achieve his results is futile.

Consumers abide by a *rule of effort*, only using previously produced data when the effort to understand the old data and associated software is less than the usefulness of the old data and the effort to create the new data. Usually, consumers find it easier and less frustrating to produce new data than to examine or reuse the old.

2.2.3 *Time Organization Problem: The One Camera Edge Case*

Some video analysis programs like [Burr 2006] are concerned with handling multiple videos capturing the same event and consequently group videos together based on time. In the Bat Lab, the antithesis of this problem exists. Usually only one video

camera, attached to the eyepiece of the microscope, captures the experiment. Buying an extra video camera is possible, but it is not the best solution because one video camera already captures most, if not all, of what the researcher needs. Adding extra cameras would increase work overhead without significantly aiding the Bat Lab worker.

With the potential of only one video feed capturing an event, organizing videos based on time may not be the best choice. The time-based context would be limiting.

2.3 Generalizability (Volleyball)

Multiple user groups and experience, the reluctance of producers to collect metadata, and the limits of time organization are not unique to the Bat Lab. These constraints exist in many other domains including athletics.

In volleyball, coaches, players, and managers all review game and/or practice video, and each has a different expertise and focus. The job of tagging the video typically falls to the managers because no one else wants to do it. (In reality, neither do the managers, but it is part of their job description.) Further, the one video stream limitation is a real constraint. At the non-collegiate level, the volleyball team may only be able to afford one camera. Also, some sports ruling committees regulate camera usage. For instance, teams are restricted to a single camera during the NCAA Volleyball Division II [NCAA 2007a] and III [NCAA 2007b] Championships.

Throughout this dissertation, we will use the two distinct and separate contexts of the Bat Lab and volleyball to extract domain-generic principles for supporting video analysis.

3. RELATED WORK

3.1 Basic Video Analysis

Two divergent approaches exist to supporting video analysis: human-aided computer analysis and computer-aided human analysis. The first approach focuses on imbuing software with intelligence to process and eventually understand video. Human input may be used to help the computer perform better, but the eventual goal is to limit or remove the human input completely. [Snoek and Worrying 2005] review many of the automatic video analysis techniques. The second approach, and the one we take, focuses on providing the proper tools and environment to assist a person in his examination of video material. Several early systems of this type are described in [Harrison and Baecker 1992].

Despite the human focus, many systems still require the tagging, coding, and annotation of events in the video for visualization purposes and/or to make searching easier (the different types of annotations are described in more detail in the following section). [Hibino and Rudensteiner 1996] propose a visualization tool based on coded video events. By progressively filtering the data via query, the user can discern important relationships between these events in the video. The system does not provide access to the original video feed; thus, the user is entirely dependent on the original, correct event coding for an accurate analysis.

DIVA [Mackay and Beaudouin-Lafon 1998] provides access to the video images and related data streams. These extra streams can originate from external data files or

from the user marking important events themselves. A maximum of thirty streams can be active at one time within the visualization, but only two can be video. I-Observe [Badre et al. 1995] does not demand any event coding by the user, but rather synchronizes computer generated software event logs with the captured video. This is feasible because they are analyzing software user-testing sessions. Not all domains lend themselves as easily to capturing external, non-video data. In both DIVA and I-Observe, non-video data streams can be used to select only interesting video segments for playback.

[Graham et al. 2003] suggest a slightly more novel paradigm for remixing video playback for analysis. In Video Paper, video key frames and associated text (i.e., transcript) are printed on a piece of paper. Underneath each key frame a barcode is printed. By scanning the bar code, a person can jump to that time location within the video. Simplicity and ease of use is the primary focus, not in-depth analysis and coding.

Other recent proposals include [Burr 2006] that supports multiple video streams (six were used during their user study). All videos are synchronized, and playback is controlled by a single widget. The video playback windows can be resized, but it is unclear whether they can be rearranged. Focus X3 [Focus 2008], a commercial system for sports analysis, supports up to four videos. These multiple feeds can be different camera angles from the same event or from different events illustrating the same skill. Dartfish [2008], another commercial system, provides picture-in-picture support for watching and analysing two video streams concurrently.

Instead of using multiple cameras to capture an interesting event, [Pea et al. 2004] propose using a panoramic camera. Their system, DIVER, supports both video analysis and repurposing video. The user operates a virtual camera moving and focusing on the parts of the panoramic video he wants to analyze in more detail. These “dives” through the video data can be annotated and uploaded to the web where others can view them. eSports [Zhai et al. 2005] is an entirely web-based system for supporting the real-time interaction and annotation of videos. Its primary motivation is supporting the remote analysis and discussion of sports footage between a coach and players.

3.2 Video Annotation

Annotations are central to many of the video analysis proposals and span the spectrum from system-centric to user-centric. System-centric annotations systems such as [Volkmer et al. 2005] exist primarily to catalog and add human semantic understanding to videos for future retrieval needs. Others [Barger et al. 1999; Aubert and Prie 2005] are more human-oriented, as they focus on supporting active learning and analysis.

The relationship between the annotations and video clip is critical. “An annotation only becomes useful because of its location and its relationship with the surrounding context” [Ramos and Balakrishnan 2003]. Consequently, the granularity at which the annotations are assigned to the video must be considered. Some annotation tools allow arbitrary assignment of annotations to the video [Davis 1993], while others only support scene-based annotations [Abowd et al. 2003]. [Costa et al. 2002] provides

for annotations at the event level. These events cannot overlap one another, but multiple annotation perspectives or tracks exist, each of which can hold a single event annotation.

A more simplistic approach (and the one we chose to take) is that used by the popular online video sharing websites such as YouTube [YouTube 2008] and Google Videos [Google 2008]. The default annotation level for these systems is the individual video clip. Producer annotations (e.g., clip title, description, category, genre) and consumer annotations (e.g., comments, ratings) are both supported but kept distinctly separate from one another.

3.3 Hypervideo and Hypermedia

The strength and novelty of hypermedia and hypervideo systems is their ability to link or connect information together. It supports the consumer in the complex task of contextualizing and developing conceptual bridges between information chunks. Unlike regular video, it “promotes deeper understandings” and “provides a better support for reflection and learning” [Chambel et al. 2004].

The link itself is critical in providing this contextual understanding. Links originating from the video medium can be divided into four groups based on functionality: unconditional, spatial, temporal, and spatio-temporal [Chambel and Guimaraes 2002]. Unconditional links may be accessed at any time during the course of the video. Spatial links, also active during the entire video, are only accessible through specific regions of the video image. While not spatially restrained, temporal links can only be followed during certain time intervals within a video. Spatio-temporal links are restricted by both space and time. (Synchronization specifications become critical when

dealing with temporal links [Buchanan and Zellweger 1992; Mujacic and Debevc 2007], but are outside the scope of our research.) Unfollowed links play just as critical a role in the user's contextual impressions as followed links [Sawhney et al. 1996].

Researchers have investigated both the human perspective of hypervideo link creation and meaning [Shipman et al. 2005] and the automated perspective. [Pattanasri et al. 2006] propose automatic link creation between video segments based on event relationships depicted in the videos (i.e., context, precondition, casual) to clarify user understanding. [Bocconi et al. 2005] automatically generate a documentary from a series of individual videos clips (implied linking), based on a pro, con, or neutral stance request. These hypermedia systems empower the consumer to repurpose individual video clips to fit their needs.

When hypermedia systems are simplified into a hierarchical "detail-on-demand" system in which only one outgoing link is available per video segment, the consumer may also save time when looking for a specific piece of information [Girgensohn et al. 2004]. Instead of watching the entire video, only the pertinent video segments need to be watched. This is similar to the concept in digital library interfaces where higher-level surrogates are connected to more detailed surrogates, providing the user with a overview-detail perspective.

While not a hypervideo system, spatial hypertexts, like VIKI [Marshall et al. 1994], suggest spatial closeness between objects is useful in implying context without the stereotypical link. The Visual Knowledge Builder [Shipman et al. 2001], VIKI's successor, supports placement of objects anywhere within a two-dimensional workspace.

Objects may be grouped in visually distinct collection objects, and collections may be included in other collections. The objects and collections also may be color-coded, giving the user another dimension, besides space, with which to create connectivity. [Atzenbeck and Nurnberg 2005; Atzenberck and Nurnberg 2006] argue for more structured collection mechanism modeling real-world bindings (e.g., drawer, binder). If a collection is limited in its capacity, the user is forced to restructure the collection when it grows beyond its bound. This makes objects easier to locate and the overall collection structure more meaningful.

3.4 Video Reality

Video reality systems focus on recreating part of the physical world visually. Unlike other video-centric interfaces, they are influenced heavily by the concept of real-world location. Video data is often collected specifically with the system in mind using multiple cameras. The systems are often developed for very specific domains and tasks such as navigational-exploration, surveillance, or GIS understanding.

Movie-Maps [Lippman 1980] was one of the earlier navigational systems. It interactively allowed users to play back image sequences previously captured to simulate a drive through town. The drive could be conducted at different speeds and during different seasons. More recent way-finding support includes [Wu et al. 2006], which superimposes navigational arrows on live video feeds for “on-road navigation”. [Nobre and Camara 2001] implement a semi-realistic campus tour by mapping multiple video tracks onto a single background image of the campus. By clicking on any section of the video track, users can advance to that section of the tour. [Uyttendaele et al. 2003]

creates interactive visual tours to model “being there” using carefully planned and captured real-world video footage.

Surveillance video interfaces include the simple [de Silva et al. 2005] and the more complex systems like [Kimber et al. 2007; Girgensohn et al. 2007] in which the user can click directly on an object in the video footage. Once the object is selected, the user can scrub the object along its path to automatically advance the video forwards or backwards. This functionality enables an event-object view of the video.

Realityflythrough [McCurdy and Griswold 2005] operates in a less constrained environment, as it is designed to work in real-time on networked, mobile cameras. It allows those in charge to assess and understand the overall tactical situation of their deployed subordinates in a crisis situation (e.g., SWAT operations). The ingenuity of the system is its reliance on closure or the “brain’s ability to fill in gaps when given incomplete information” [McCurdy and Griswold 2005]. Consequently, computing resources are saved by not requiring perfect image alignment or full live-video coverage. [Ichimura and Matsushita 2005] focus less on accurate position mapping of video to the real world, but instead rely on the notion that real world events must take place at a physical location and the assumption that people will capture interesting events with their personal video cameras. Geared toward enabling more artistic amateur movies (i.e., access to multiple camera angles from others’ cameras), Ichimura and Matsushita’s system could be used in the surveillance domain to piece together information about a particular event.

In the GIS domain, [Ghadirian and Bishop 2002] use video and augmented reality to visualize and model environmental issues (e.g., weed control intervention alternatives). [Kim et al. 2003] focus on using the spatial information present in real world video footage to create spatially accurate three-dimensional worlds.

3.5 Digital Video Libraries

Digital libraries have a slightly different focus than the previously mentioned areas regarding the support of video understanding. They are primarily concerned with getting the appropriate information to the consumer, not what the consumer does with the materials after locating it. The majority of digital video interface research has been focused on supporting the user in browsing and searching of videos (i.e., very shallow analysis).

Video surrogates (also called summaries, previews, or abstracts) are one of the primary interface mechanisms used to support this shallow analysis. A surrogate is any “representation that people scan and examine to extract meaning [about the original video source], and make rapid decisions about further processing” [Tse et al. 1998]. It acts as a substitute for the full-length video and can be lexical and/or pictorial in nature [Ding et al. 1999], using either the audio and/or visual channels [Song and Marchionini 2007] to convey information to the consumer in a static or dynamic manner [Komlodi and Marchionini 1998].

Surrogates may also represent or summarize multiple videos. [Greene et al. 2000] refer to these as overview surrogates. [Christel et al. 2002] present two such surrogates for a collection of news videos – the map collage and timeline collage. In the map

collage, key frames are arranged by location referenced in each news story. The timeline collage orders the key frames based on their broadcast date. Not only do these views provide additional information for the consumer, not easily captured in a single key frame (e.g., subject's location or broadcast date), but also help the user understand the bigger picture of the video collection and how individual newscasts relate to one another.

Other researchers, while not directly published in the digital library field, are similarly focused on supporting browsing and searching tasks through overview surrogates. [Tang and Kender 2006] arrange individual classroom lecture surrogates in a timeline display, as well as provide interface support for grouping key frames based on image similarity. [Houten et al. 2004] organize surrogates based on semantic similarity. [Chiu et al. 2005] take a novel visualization approach, creating a three-dimensional MediaMetro city in which each building represents a multimedia document. Compared to a single-video surrogate, an overview surrogate is less biased and is more insightful because it “supports comparison among related objects” and “can be used to understand an object in context” [Greene et al. 2000].

4. LOCATION AND CONTEXT

As we are focused on context-demanding questions, we must consider how to provide access to this context. *How can we synergistically combine videos together to make the user's job easier in analyzing relationships between clips?* We propose grouping and organizing related videos together based on location similarities, thus providing location-based context.

The term *context* has been subject of much research [Lee 2007]. Our use and view of context would fit best under Lee's second context category: "objective or socially constructed characteristics and conditions of the situation in which a [target entity] is, appears or occurs." This is similar to [Pauty et al. 2005] idea that context is "based on a notion of proximity."

The concept of organizing videos by space or location is not new (described in the related work Section 3.4 about video reality). But these location references are usually absolute and rely on only one perspective of location (e.g., GPS coordinates). The GIS community recognizes that location and space are much more complex. Space can be referred to at multiple resolutions and from multiple perspectives. Vangenot et al. [2002] write:

While the real world is assumed unique, the way it is represented depends on the intended use. Thus, different applications that share interest in the same real-world phenomena may have different perceptions and therefore require different representation; i.e. different sets of objects, links, properties, and/or different values. Their needs differ because they don't share the same viewpoint, they don't need data acquired at the same instant or with the same cartographic resolution.

4.1 The Inspiration

Since each producer of video data has a different perspective, it follows that our concept of location-based organization should not be one-dimensional. For inspiration handling the multiple viewpoints of location, we turn to the multiple-representative concept added to the MADS model [Parent et al. 1999] by [Vangenot et al. 2002; Vangenot 2004].

To support multi-representation of spatial data, [Vangenot et al. 2002; Vangenot 2004] introduce the perception stamp. The stamp consists of a <viewpoint, resolution> tuple where the resolution can be defined spatially or semantically. The stamp can be applied to any database component and can be used to filter data or act as an access control mechanism. Each user group (or perspective) is represented by a different stamp (e.g., stampA = <petroleum engineer, 1 meter>).

Objects represented within the system (e.g., oil well) have various attributes (e.g., well name, depth, and manager). Attributes pertinent to several groups or perspectives (e.g., well name) are marked with the appropriate stamp from each group. Other attributes, like well depth pertinent to petroleum engineers only, are stamped once (e.g., with stampA) and accessible solely to that group (e.g., engineers). In other cases, the attribute (e.g., manager) is important to several groups but holds separate values for each perspective. For example, the engineering group sees the drilling supervisor's name under the well manager attribute, while another group sees the managing company's name.

When viewpoints differ greatly regarding the same real world object, it may be necessary to create two separate object representations, each of which is stamped separately. Relationship types (i.e., identity, aggregation, set-to-set) can then be used to represent and hold information about the link between the two representations.

4.2 Location in the Bat Lab

Location perspectives differ in the Bat Lab depending on the person or group collecting the data. One or several location references may be recorded depending on the researcher's need. Four commonly used perspectives include: the vessel's network, order, relationship to other vessels, and diameter. The vessel's network (Figure 1) is the most generic of these references and refers to the parent structure of the vessel. Just as Houston is in Texas, one may say a vessel is in the Mississippi network structure.

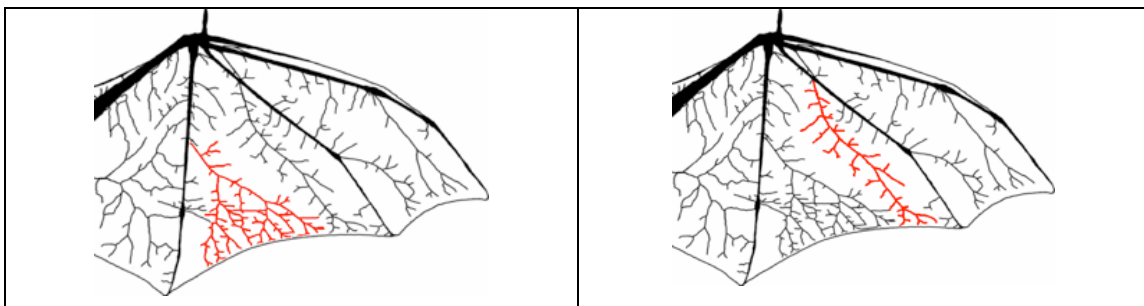


Fig. 1. Vessel networks. Mississippi network highlighted on the left, Rio Grande network highlighted on the right.

The order of a vessel is related to the number of bifurcations above the vessel. A bifurcation is the point where one vessel divides and becomes two vessels. A first order

vessel would be before the first bifurcation point. A second order vessel would be downstream of the first bifurcation but upstream of the second bifurcation.

The relationship location is also tied to bifurcations. Any vessel above a bifurcation point is called the mother vessel, and the two vessels below the bifurcation are daughter vessels.

The diameter of a vessel (minimum, maximum, average) is also a location reference as larger diameter vessels exist upstream of smaller diameter vessels. One can use the diameter of a vessel (in its natural, control state) as a method for checking or refining the other location references.

A fifth location reference, grid location, refers to the (x,y) coordinate position of the microscope when the video was collected. Within a single video stream these coordinates can be used to absolutely position the imagery, as long as the bat does not move. While useful for computer programs, grid location is not a human-friendly reference when comparing and trying to generalize information across multiple bats. Network, order, relationship, and diameter are used more frequently in every day communication between researchers.

4.3 Location in Volleyball

Location references in the Bat Lab are fairly straightforward. Less intuitive is how location relates to the other domain of interest (volleyball).

We will use three location perspectives when referencing volleyball videos: location outcome, camera, and focus. The location outcome refers to the location of the volleyball after a player has contacted it. After a set the volleyball may be in the outside,

middle, or back location (Figure 2). A serve may be in the front-end or back-end of the opposing team's court. The volleyball can certainly end up other places as well, but we do not need to list all the possibilities.

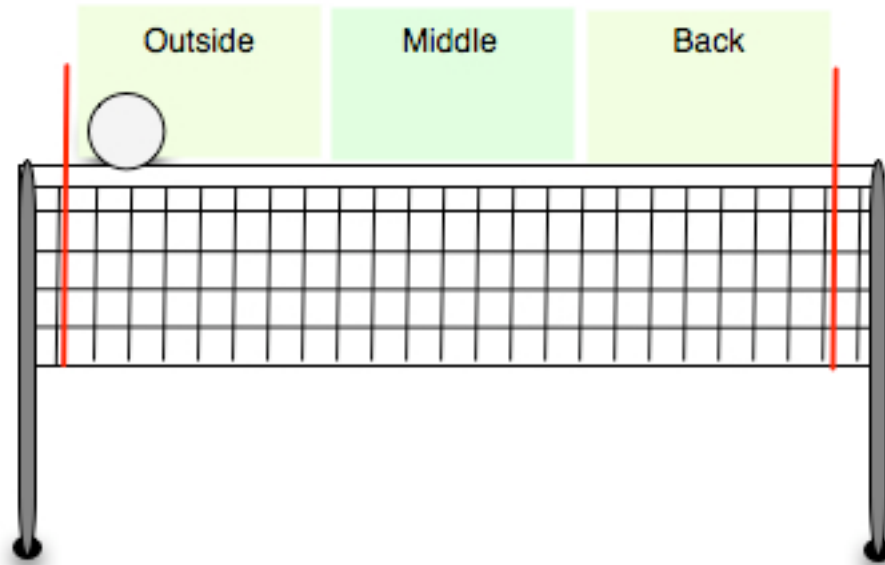


Fig. 2. Volleyball location outcome (setting).

The location of the camera capturing the volleyball game or practices refers to the camera's location with respect to the team's side (i.e., baseline, left, right, opposing baseline). Finally, the location focus refers to the part of the player the camera is focused on (i.e., whole, upper, or lower body).

4.4 Summary

The location perspectives detailed in the preceding sections for the Bat Lab and volleyball domains provide multiple ways to organize video clips. Grouping videos based on one location perspective versus another would create different contexts.

The selection of location perspectives is exceedingly important. They need to be meaningful perspectives that the producer already collects (or could be collected automatically). The producer, as we will see in the system design section, is never forced to capture any particular perspective. He is given the flexibility to capture only the location reference(s) useful to him.

One caveat exists in the general selection of the perspectives. To adopt a term from [Parent et al. 1999], the perspectives should be orthogonal. A value in one perspective should not dictate the value of another perspective. If the perspectives are not independent of one another, the location values become synonyms of one another and consistency becomes harder to maintain.

5. SYSTEM DESIGN

The VAST system design uses location-based context to provide support for video analysis tasks, but does so in a domain-generic way. It is divided conceptually into the following parts: database, database connector, view library, supporting GUI, and GUI engine (Figure 3):. The media libraries are referenced by VAST, but are not a core part of it.

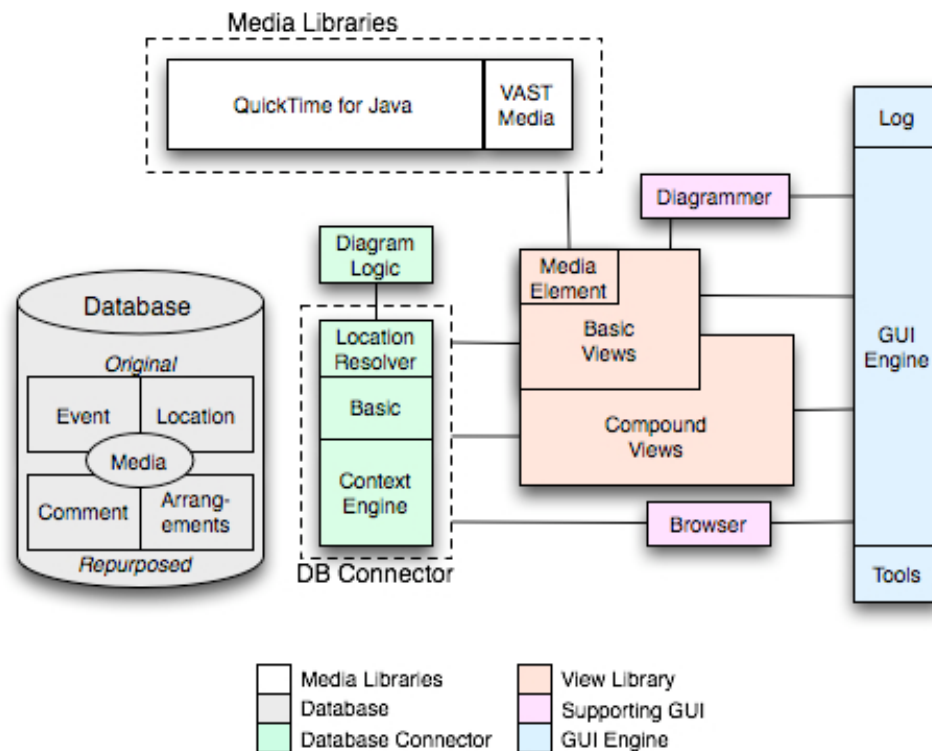


Fig. 3. VAST design overview.

5.1 Database

The database is divided into three parts: media-reference, original metadata, and repurposed metadata. It is assumed that the media items within the database are correctly segmented based on location.

5.1.1 Media-Reference Section

The media-reference section contains a record for each individual media file in the digital collection. A one-to-one mapping exists between the media-reference entry and the digital file. The file may be any media type and in any format. (Our focus is on videos, but the system can handle non-video files as well). Instead of using the term media file, we will use media item or media clip as a reminder that the media files in the library are the result of location-based segmentation and not the original files.

5.1.2 Original Metadata Section

This section stores information concerning the original context of the media items. The event subsection represents the producer's point of view when capturing the event by specifying *who* produced the event record, *what* was being recorded, *when* the event started, and *why* it was recorded. The *where* (or location) aspect of the event is not captured in detail, but a location domain reference is logged. The location domain reference acts as a pointer into the location subsection of the original metadata section.

The location subsection stores location metadata about each media clip. It is the only part of the database that is domain-dependent. Each domain has its own group of location tables. Each table represents a perspective or interpretation of location-oriented metadata. A clip may be tagged or referenced in none, one, or several of these location

perspectives depending on the availability of information or the perspectives appropriateness to the clip. Figure 4 illustrates the location subsection tables currently implemented.

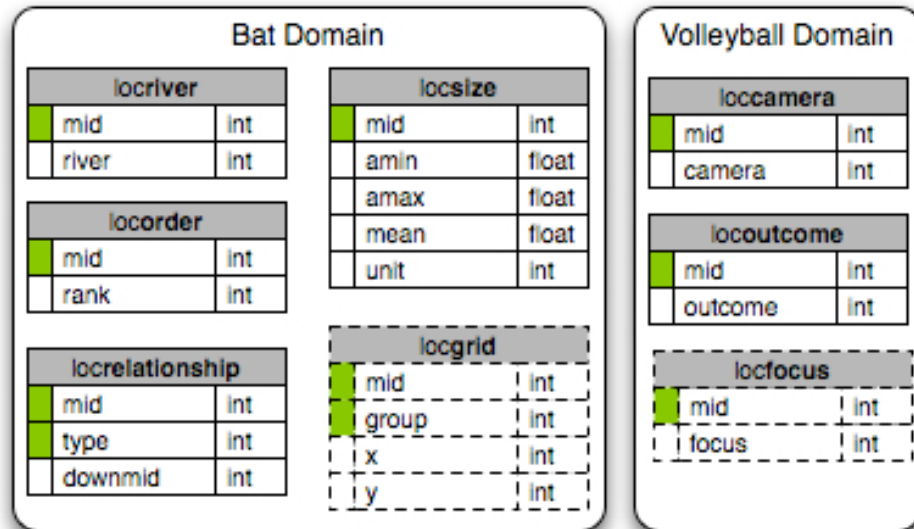


Fig. 4. Domain-dependent location tables.

In the Bat Lab domain, the vessel's network, order, relationship to other vessels, and diameter are recorded. Grid location is not implemented because it is currently determined manually and stored using arrangements. (Arrangements, the data structure for views, will be covered in more detail in the following sections.) The volleyball domain stores information about the camera's location and the location outcome of the ball. Focus location is not currently implemented, but is used conceptually during the

creation of static arrangements. The concept of location was discussed in more detail in Section 4.

5.1.3 Repurposed Metadata Section

The repurposed metadata section stores information about the media items from the consumer's perspective. Unlike the original metadata that does not change once recorded, the repurposed metadata is dynamic and changeable by the user.

The comment subsection stores textual information per media item, per user. The arrangement subsection describes how two or more media items should be grouped together for consumer consumption. A single media item may be referenced multiple times within an arrangement. Each arrangement has a native view. Arrangements and their relationships to views will be discussed in more detail in the next sections.

5.2 Database Connector

The database connector abstracts the details of database communication and logic from the other components. The connector can be divided into three sections: basic, location domain resolver, and context engine.

5.2.1 Basic

The basic section handles database connections and simple access.

5.2.2 Location Domain Resolver

The location domain resolver abstracts the domain dependent location details of the database from the other VAST components. The resolver takes a generic location request for a media clip, determines which location domain tables need to be queried, and formulates the appropriate query. Upon receiving the information from the database,

the resolver reformulates the location information into a generic wrapper. Other elements of VAST can then access the location information without specific domain knowledge.

The location resolver is also the access point for the domain specific diagram logic necessary to render the appropriate location-based diagram. After retrieving the appropriate information, the resolver surrounds the domain details in a generic wrapper for the diagrammer's use.

5.2.3 Context Engine

The context engine has two purposes: managing arrangements and converting arrangements. Arrangements are the data structure for views. Views define how a group of media items should be rendered and interacted within the graphical user interface. Each media item within an arrangement is assigned a unique (x, y, z, color) coordinate location that is used by the view to position the item. Arrangements can be defined statically, as is done in the arrangement section of the database, or can be defined dynamically as a set of queries for finding, prioritizing, and assigning coordinates to related media items.

The context engine is responsible for handling the communication between static arrangements in the database and their native compound views, as well as creating and managing dynamic arrangements. Currently, the concept of dynamic arrangements is simulated through the static arrangements. This simulation of dynamic arrangements provides greater flexibility when testing new arrangement concepts and provides quicker database response time during user testing.

The other responsibility of the context engine is determining the new arrangement needed when a conversion between views is requested. Within each arrangement one media item is designated as the internal focus or primary item. The primary item stays the same when converting between views. The context engine during a conversion is responsible for finding any arrangement whose native view corresponds to the new view requested.

If more than one candidate arrangement exists, the context engine is responsible for picking the *fit*. The *fit* currently coincides with the *first-fit* found. Other algorithms could be developed where *fits* are ranked based on similarity or dissimilarity to elements in the old or new arrangement.

5.3 View Library

A view defines how a media item (or group of media items) and associated metadata will be managed, rendered, displayed, and manipulated. The view library contains many different views.

5.3.1 Basic Views

The basic views are the building blocks for all other (compound) views. Each basic view is responsible for handling a distinct medium. The implemented views include a generic basic view (from which all basic views are derived), a video view, and an image view. Each view can be divided into five parts: media element, database hook, diagram coordinator, shared look and feel, and custom look and feel.

The media element encapsulates the low-level functionality of the media item that the view renders. It communicates with the media libraries available to VAST including QuickTime for Java (QTJ) and the VAST media utilities.

The database hook, via the database connector, retrieves information about the media clip displayed in the view. This includes media path information and event, comment, and/or location metadata. When appropriate, the diagram coordinator feeds location metadata from the database hook to the diagrammer. Location metadata is coordinated through the view ensuring synchronization between the two.

The look and feel components define display and interaction capabilities. The shared section provides consistency between basic views, regardless of media type displayed. The custom section modifies or adds to the basic look and feel component to support the unique affordances of each media type.

Within the look and feel components, three distinct render/interaction conditions are defined: independent, dependent, and associated. In the independent condition, the media item is rendered as a single, stand-alone unit. The user is given the most options for manipulating the media item and provided the most metadata information. The dependent condition indicates the basic view is contained within a compound view. Interaction capabilities are more constrained and display information is optimized to fit into a smaller space. The associated condition specifies that the basic view is connected with another view, but has a distinct display space. Display optimization is less critical in the associated condition. The three render/interaction conditions are illustrated in Figure 5.

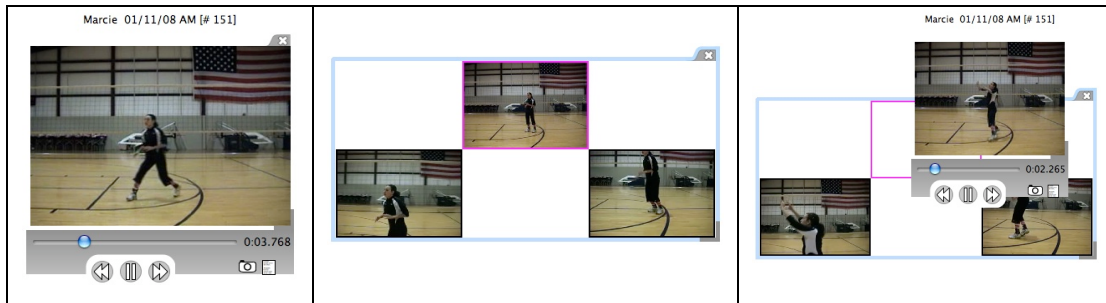


Fig. 5. The video view in its (from left-right) independent, dependent, and associated condition.

5.3.2 Compound Views

While basic views display single media items, compound views coordinate the rendering and interaction of multiple media items and associated metadata. They are the primary mechanism for displaying context (i.e., the automatic or manual arrangement of media items and metadata in an synergistic fashion). Each compound view is built using basic views, but has a separate context hook and look and feel component.

The context hook tracks the media arrangement subpart displayed in the view. The hook initiates requests to the context engine for different subsections of the arrangement (i.e., different media items). When a view conversion is requested, the context hook sends information about the original view to the context engine so a new view can be chosen.

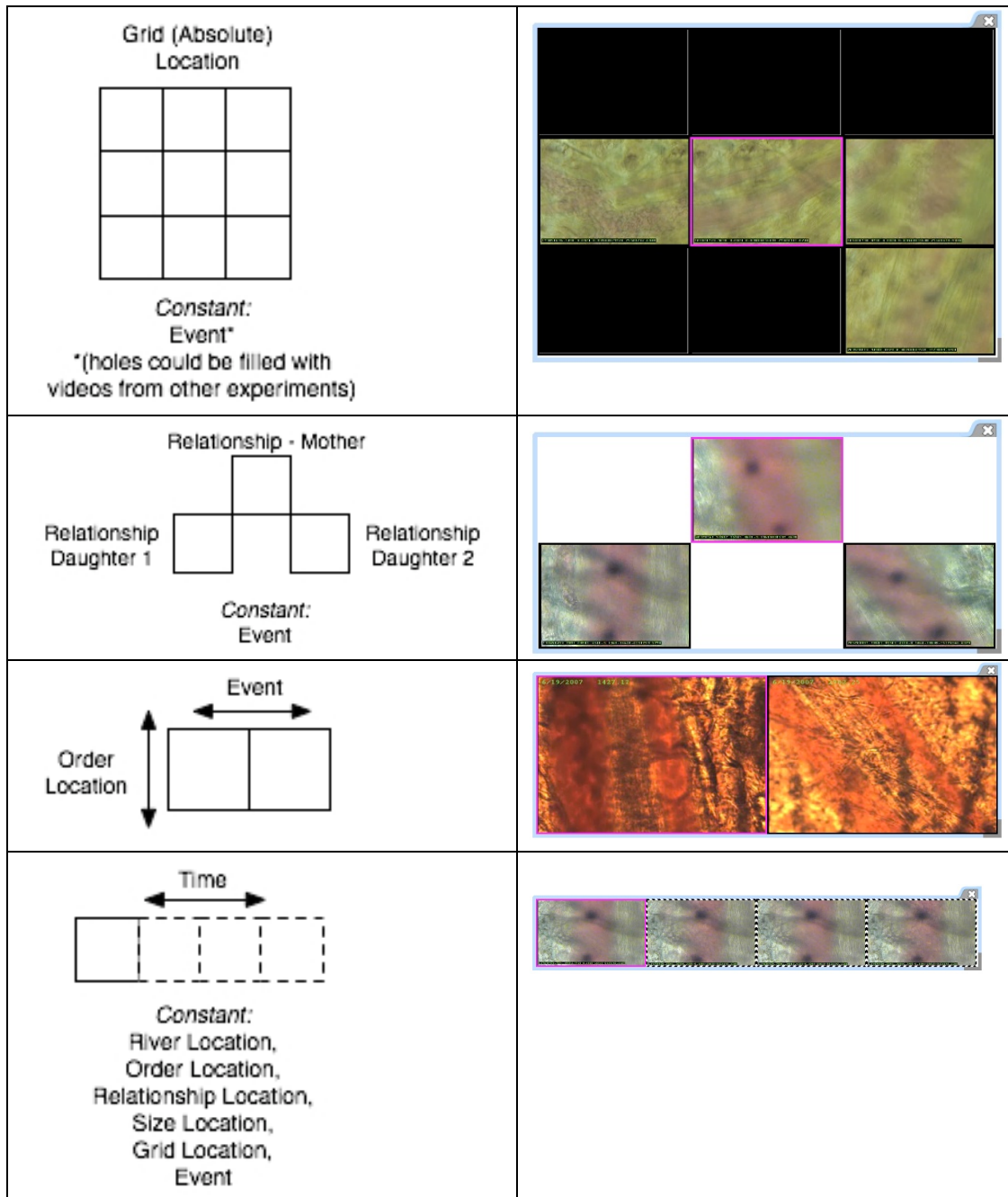


Fig. 6. Conceptual arrangements and associated views in the Bat Lab domain. From top to bottom: the area, branch, compare, and time views.

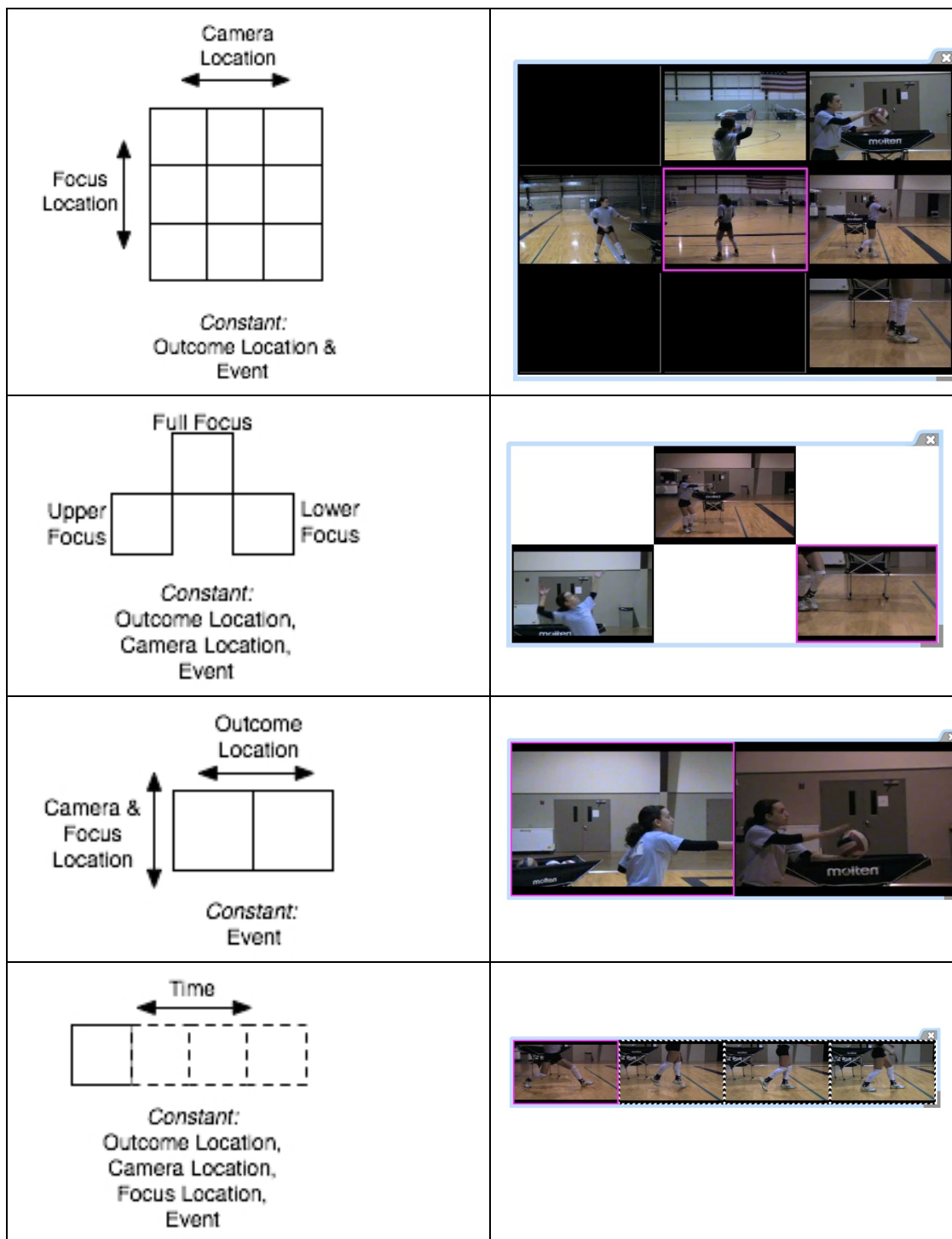


Fig. 7. Conceptual arrangements and associated views in the volleyball domain. From top to bottom the area, branch, compare, and time views.

The look and feel component is divided into a shared and a custom part. It is not responsible for handling individual media clip functionality, but defines how the basic views are put together and interacted with as a group. This functionality includes a group resize, move, close, and play/pause all media (where applicable). It also defines navigational and zooming capabilities within the view

Thus far, the library includes a generic compound view (from which all compound views are derived), an area view, branch view, compare view, and time view. Figures 6 and 7 illustrate the conceptual arrangement and associated compound interface views for the Bat Lab and volleyball domains, respectively. Changing the underlying arrangement (i.e., using a different location perspective) would change which clips were displayed in the views, but the view layout and functionality would remain unchanged.

5.4 Supporting GUI

5.4.1 Diagrammer

While views are responsible for rendering media items and associated metadata, the diagrammer uses only location-based metadata. Location information is dispatched to the diagrammer through the view, ensuring coordination between the two.

The diagrammer expresses location in three ways: base map, highlighted area, and text. Depending on the domain (determined by the location resolver), different location perspectives can be mapped to different diagram expressions. In the Bat Lab, the base map stays the same for all media items. The highlighted area provides a visual estimate of where on the bat wing the media item originated. The highlighted area is computed dynamically using information from the river, order, relationship, and size

tables. The computational specifics are abstracted from the diagrammer by the location resolver. The textual aspect of the diagram lists information from the river, order, and size table. For the volleyball domain, the base map is determined by the camera location. The highlighted and text areas both convey outcome location. Figure 8 illustrates the Bat Lab and volleyball domain diagrams.

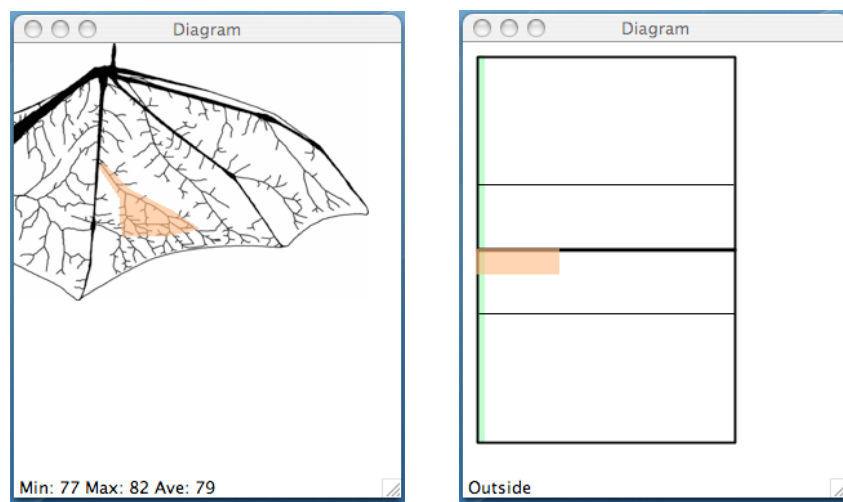


Fig. 8. Location-based diagrams for the Bat Lab (left) and the volleyball (right) domains.

5.4.2 Browser

The browser provides access to individual media clips. A user may browse the entire collection or limit the scope by searching on a particular term. Search terms are checked against the event, location, and comment metadata.

5.5 GUI Engine

The GUI Engine manages and coordinates the graphical user interface components in the system including the diagrammer, browser, and views. It also

manages the tool suite responsible for providing extra functionality to the user for quantitative data analysis. The tools in the suite, like the ruler tool, are not media specific and do not require database access. The log records user interaction with the system for research purposes. It can be accessed by any component in VAST but is managed by the GUI engine.

5.6 Implementation

To promote cross-platform compatibility the system is implemented in Java and uses QuickTime for Java (QTJ) to handle low-level video functionality. Apache Derby is used for the database because it is open source and also implemented in Java. Derby can be embedded within the application so the end-user is not burdened with the database setup. All media files reside on the local system so downloading delays and network issues do not affect the system's usability during testing. The conceptual system design does support a remote database server.

For initial testing of the system design and implementation, we used videos from the Bat Lab produced *a priori*. It was imperative we use video not produced with the system design in mind, as one of our goals was to keep producer overhead minimal. Metadata should be collected because the researcher considers it important, not because the system needs it.

The video data originated from three different research groups. These groups were picked because their work required them to collect information from multiple spots on the bat wing. Some groups spent their entire experiment time focused on a single spot. Even so, this video data can be used, as over time the stationary groups will capture

different parts of the wing. However, for testing purposes the groups who captured multiple locations within a single experiment were more interesting. Metadata for the video clips were collected from the handwritten lab notebooks and electronic logs automatically produced.

The video and metadata were manually put into the system. Figure 9 shows the implemented VAST system in the Bat Lab domain.

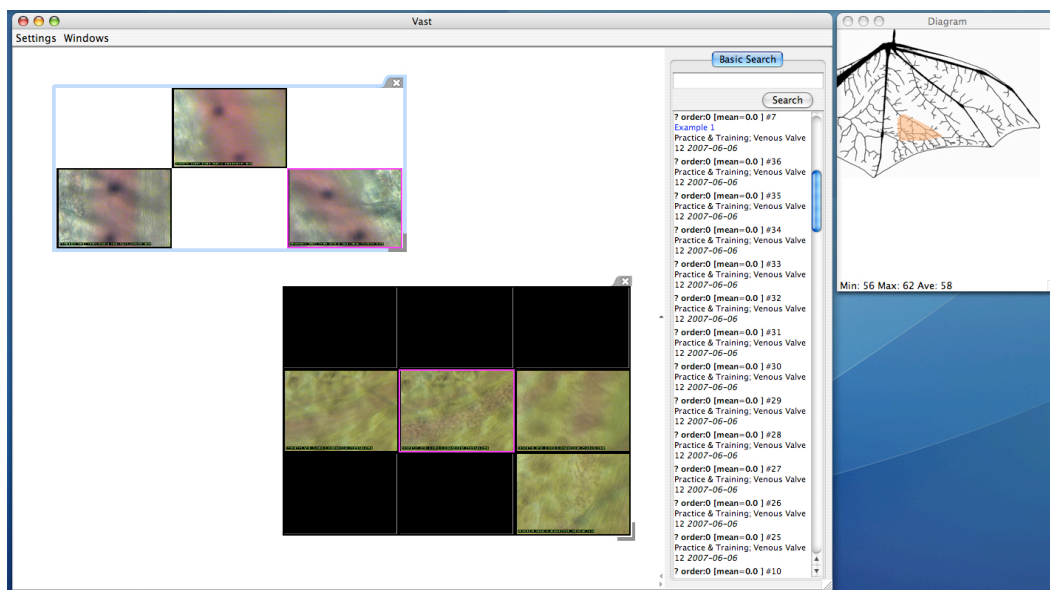


Fig. 9. VAST screen shot - Bat Lab domain.

In the following sections, VAST is discussed from the user's perspective by highlighting the human-interaction details and then by examining the user testing results. We also switch from the Bat Lab to the volleyball domain thus demonstrating the system's generalizability.

6. USER INTERACTION DESIGN

The preceding system design section was focused on VAST from a computer system's perspective; VAST is now examined from the end-user's perspective. VAST's user interface is divided into the workspace, search, and diagram areas (Figure 10). The user watches and manipulates video clips via views in the workspace. The search area provides access to the clips, and the diagram area displays location information for the clips.

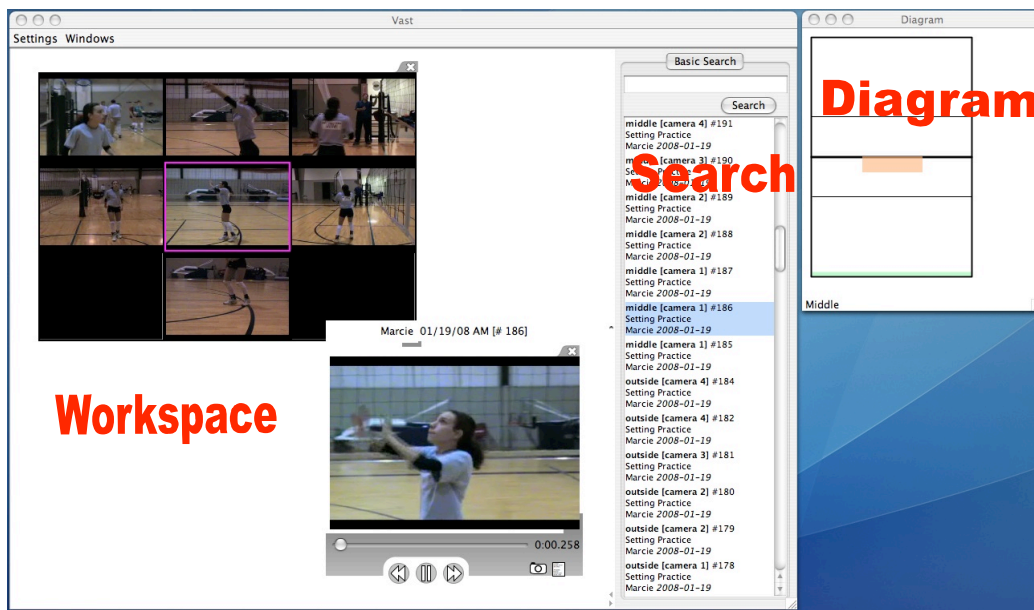


Fig. 10. VAST graphical user interface – volleyball domain.

6.1 Search

The search area is the initial mechanism for finding video clips of interest. Once a clip is in the workspace, the user can find related clips by converting the original clip

into different views. When no search term is specified, the search area displays all available clips; otherwise, only clips matching the specified term are displayed.

Each clip in the search area has the following format (Figure 11). Domain-specific location information is listed in the first line followed by a unique clip number. If applicable, the next line displays user comments. User comments are rendered in blue indicating consumer-specific metadata and distinguishing them from producer metadata rendered in black. Typically, event data is the same for multiple clips as one video captures multiple locations.

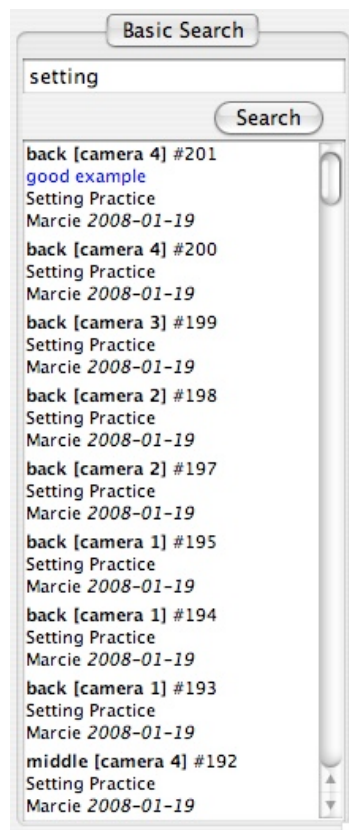


Fig. 11. Search area.

The search area is intentionally simple because our primary focus is on supporting video analysis and not browsing or searching large video collections. Improvements to the search area are discussed in future work.

6.2. Workspace

6.2.1 Basic Views

From the search area, the user can drag and drop an individual video clip into the workspace. The clip, now in a basic video view, automatically begins playing. Initially, no metadata is displayed (Figure 12).

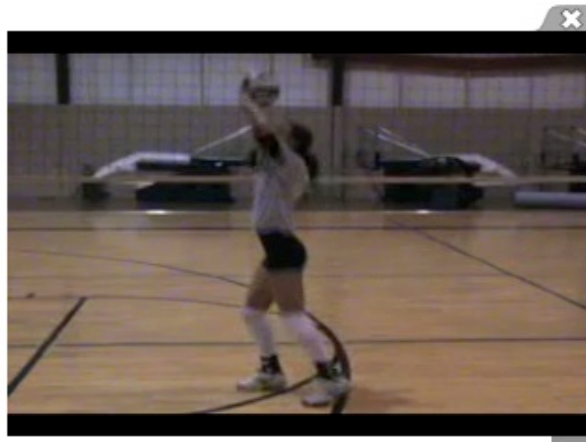


Fig. 12. Basic video view without mouse focus.

By mousing-over the video, the user can view event, clip, and time information (Figure 13). Corresponding location metadata is displayed in the diagram area (see next section). The user can adjust the movie time (i.e., directly, play/pause, next frame, or previous frame), add personal annotations to the clip (Figure 14), or take a picture of the video.

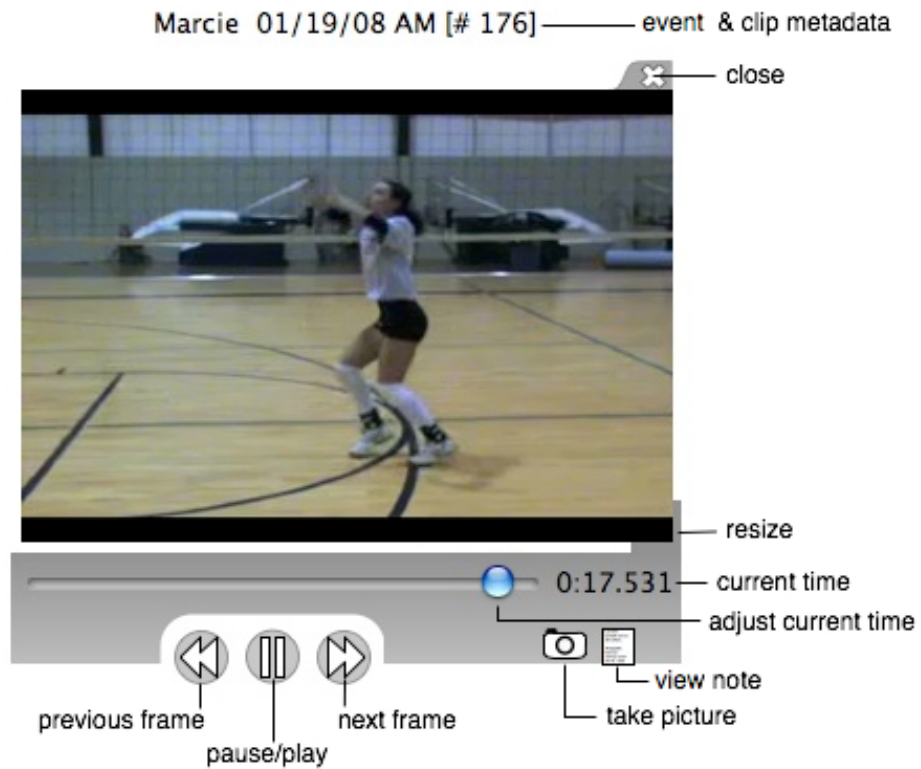


Fig. 13. Basic video view with mouse focus.

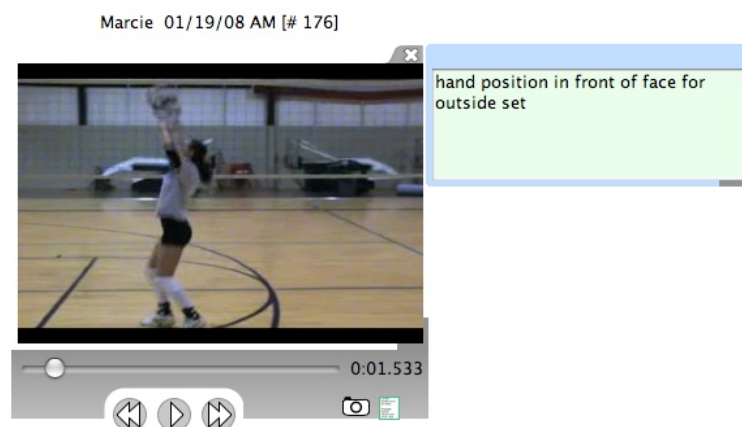


Fig. 14. Annotation in the basic video view. Annotations are assigned at the video clip granularity.

If a picture is taken from the video, the image is automatically added to the workspace. The image basic view is distinguished from the basic video view by its border. However, the image view has a similar interaction paradigm as the movie view; metadata is only displayed when the user mouse-overs the image (Figure 15).

All views place emphasis on the visual component (i.e., movie or image) of the view. Consequently, the default view state is simple, not cluttered by excessive information. Additional information and commands are accessible to the user on demand. In the image view, these commands include making the image transparent and exporting the image outside the application.

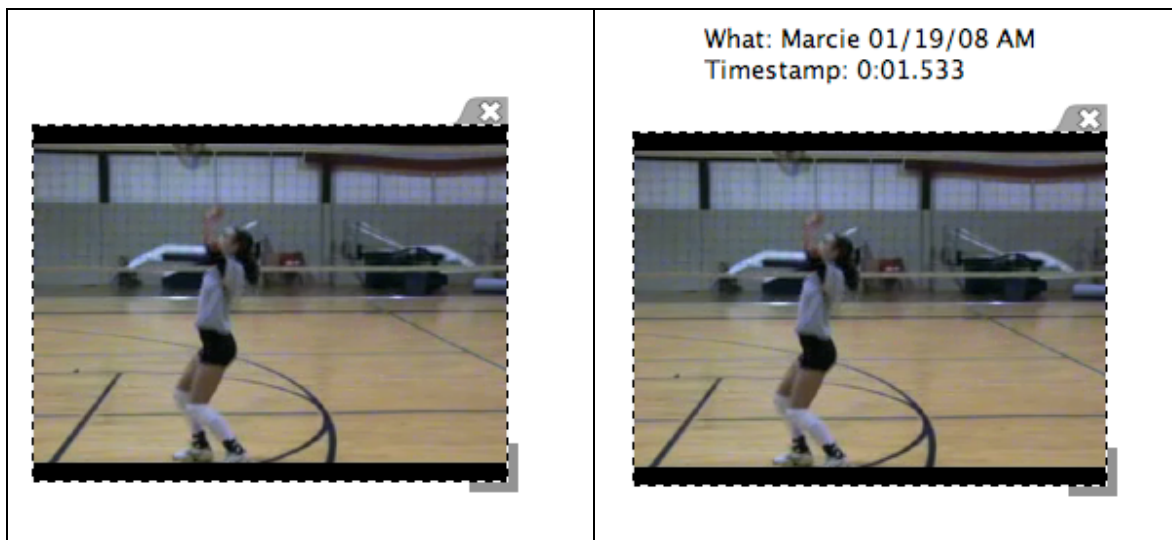


Fig. 15. Basic image view without (left) and with (right) mouse focus.

Additional video functionality is accessible via the video view's popup menu (Figure 16). The user can make the video transparent, add filters to the clips (e.g.,

sharpen, edge detection, etc.), and add or remove other video clips to the view. Changes to the video are view dependent. Once the view is closed, the clip reverts to its original state.

The basic video view can also be converted into other compound views (area, branch, compare, or time). These views present the user with different contexts to evaluate and analyze the video.

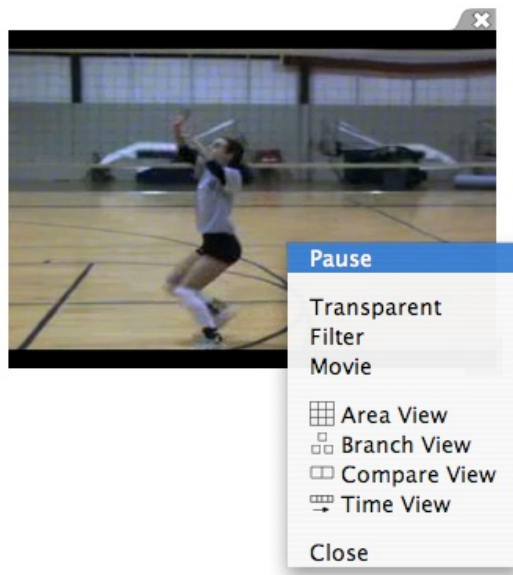


Fig. 16. Basic video view popup menu.

6.2.2 Compound Views

Compound views combine multiple videos and/or images into a single object the user can manipulate. The compound views currently available to the user include the area, branch, compare, and time views. Each compound view has a distinct look and

internal media clip arrangement (Section 5.3 View Library). By default, all video clips within a view start playing when a compound view is open. The user can pause individual video clips within a view or pause all video clips simultaneously. The clips are not synchronized with one another.

Within each view a single media item has internal focus (Figure 17). Internal focus allows a media item to be manipulated without affecting the surrounding items in a view (i.e., pausing one media item while all others continue to play). The internal focus also plays a critical role in determining how one view is converted into another (see section 5.2.3 Context Engine).

The user can change the internal focus of the view by clicking on a different media item. When the view itself is no longer active (i.e., does not have external focus), the internal focus remains providing a visual cue to the user's last focus point within the view. However, the internal focus remains in a passive state until the view gains external focus again.

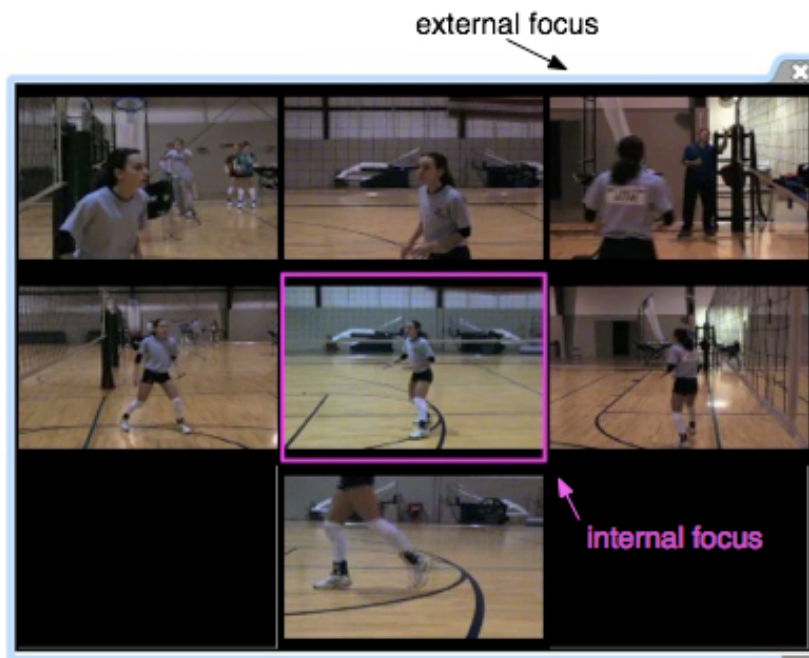


Fig. 17. Internal and external focus in a compound view. Internal focus is designated by a pink border around a single media item. External focus is designated by a light blue border around the entire view.

By double clicking on a media item, the user zooms-in on the media clip (Figure 18). All videos are paused within the compound view, except the zoomed-in clip. By zooming-in, the user gains access to additional clip information and functionality that was hidden while the clip was within its compound view. The zoomed-in clip can be resized and moved independently of its compound view. When a user moves the compound view, the zoomed-in clip moves automatically, maintaining its location relative to its parent view.



Fig. 18. A compound view zoomed-in.

Internal focus is important for manipulating a single video clip in a compound view, but external focus is important for manipulating all clips in a view. When a view has external focus, a user can manipulate all media items within a view together (i.e., moving the view in the workspace or moving around within a view) or can activate the view's internal focus. Only one view can have external focus in the workspace at a time.

Once a compound view has external focus, the user can move around in a compound view using the arrow keys (Figure 19). Moving around within a view allows the user to explore additional media contexts.

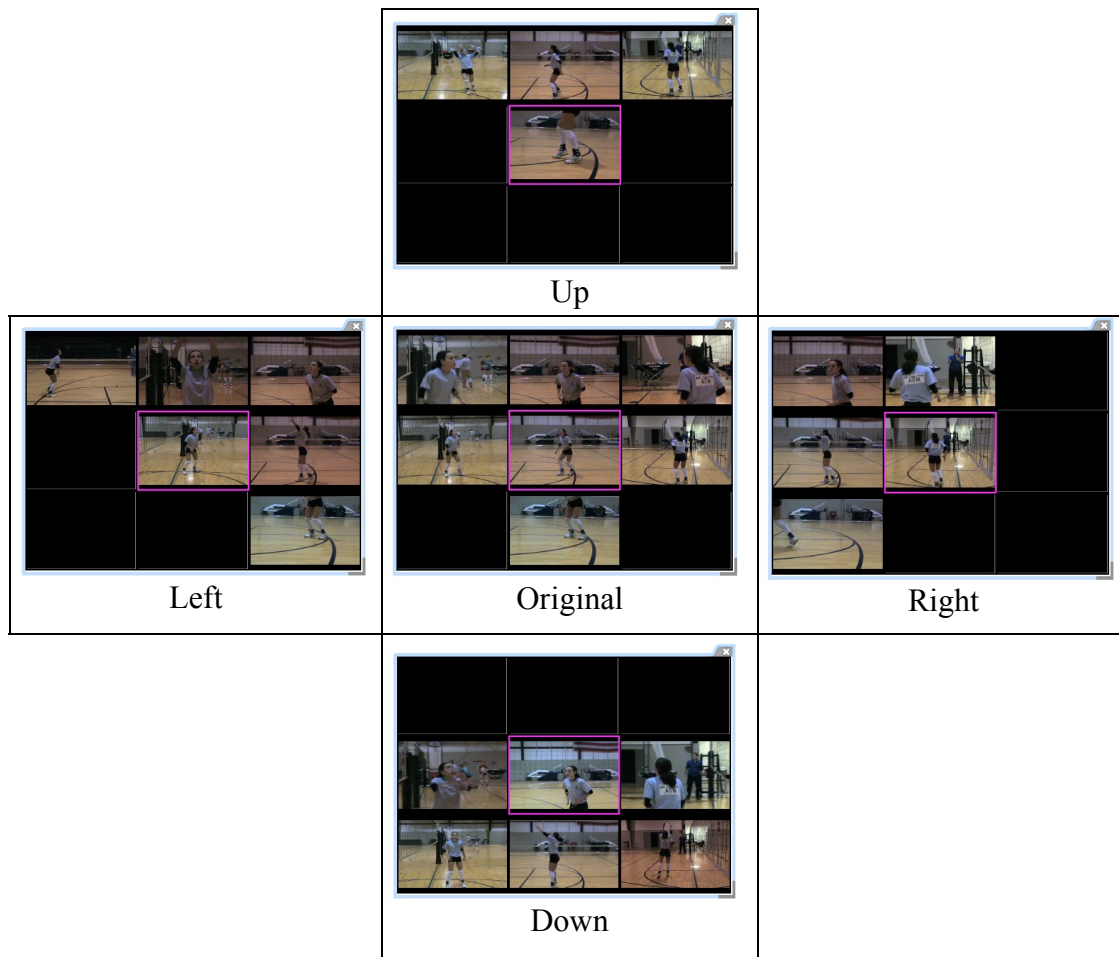


Fig. 19. Moving within a compound view.

6.3 Diagram

Regardless of view type, video location information is displayed in the diagram. Location metadata is displayed in both text and graphic form. Depending on the domain, specific metadata may be represented in only one or both formats. Figure 20 illustrates the diagrams for two different video clips in the volleyball domain.

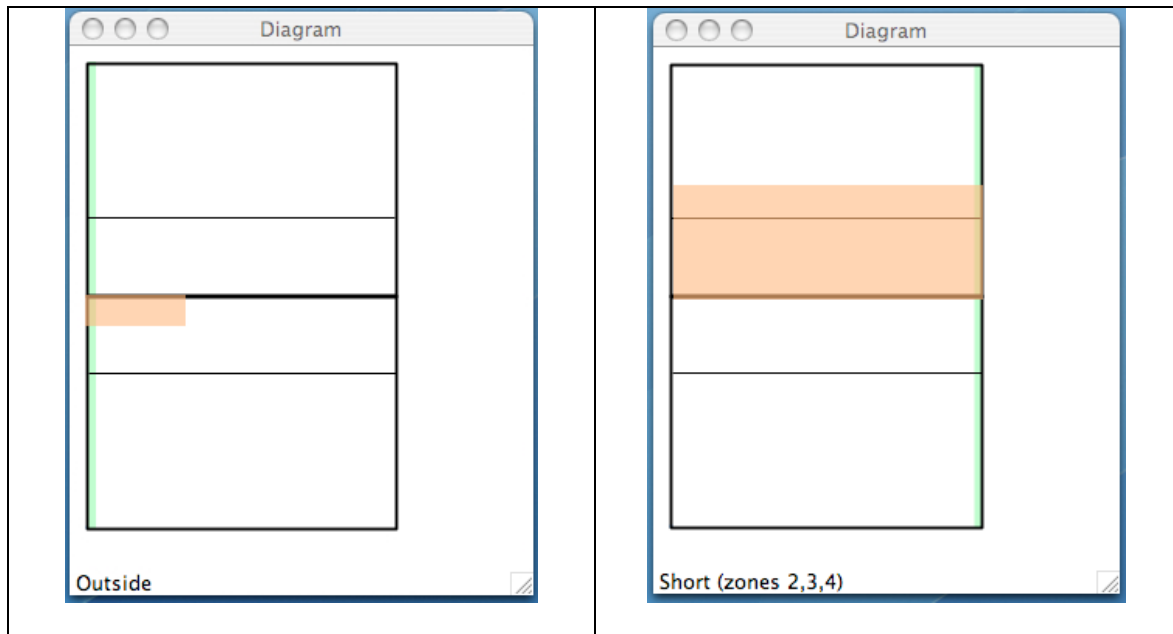


Fig. 20. Location diagrams for two different volleyball clips. The left diagram denotes an outside set (location outcome) captured by a camera on the left side of the court (location camera). The right diagram specifies a short serve (location outcome) captured by a camera on the right side of the court (location camera). Location focus is not represented in the diagrams, as this is discernable from the video clip itself.

7. USER STUDIES

Following the implementation of VAST, we conducted several user studies. We choose the volleyball domain, as opposed to the Bat Lab domain, for testing purposes to increase the potential user pool and test the generalizability of the VAST concepts. All studies were conducted on a 2.33 GHz Intel Core 2 Duo 15' Widescreen (1440x900) MacBook Pro with 3GB of RAM running Mac OS X (10.4.11).

7.1 Study Design

Three rounds of user testing were completed in the volleyball domain. Informal discussions were held with two computer scientists and two volleyball experts in the first round. Each volunteer had the opportunity to work with the system and offer feedback. The main focus of the informal sessions was to improve the usability of VAST and develop the formal study questions and format.

In the two subsequent formal rounds, participants used VAST to complete several video analysis tasks. Questionnaires were used to gauge their volleyball expertise and solicit their opinions about VAST. The pre-task questionnaire requested basic demographic, general computer and TV usage habits, and volleyball experience. (See Appendix A for the questionnaire and all other study material.) The post-task questionnaire asked participants for feedback regarding VAST.

All participants were given the option to read a basic volleyball tutorial and watch the associated volleyball clips before starting the video analysis tasks. This was

followed by a required VAST tutorial that stepped the participant through the general concept and functionality of the software.

After the tutorials, participants were given four tasks (each task consisted of one or several related questions) to complete using the provided VAST software. Users were given one task at a time and had to turn in their answers before receiving the next task questions.

Participants could opt-in (all did) to be video recorded during the study. The video recordings captured the user activity on the computer screen. A user log recorded key interactions with the software and the time of their occurrence.

7.2 Results

7.2.1 Demographics

Four people, all males, with varying degrees of volleyball playing experience from recreational to collegiate club participated in the first formal round of the testing. All participants completed the required tasks, but for two users the software had to be restarted three or more times. Because of the excessive number of restarts, we stopped testing and addressed the causes of the crashes (see next section).

The second formal study round included ten participants, five with recreational or no volleyball playing experience (novices) and five with considerable volleyball playing experience (experts) including four who played at the collegiate club level. Four of the five volleyball experts also had coaching experience, ranging from recreational to the collegiate varsity level.

Six out of the ten participants were male and four were female. They came from various educational backgrounds including engineering, business, and the pure sciences. Five had some graduate level education or higher, while the other five had at least some undergraduate training. All participants were fairly young (35 years or younger) with most falling into the 18-25 year range. In this second formal round, all participants completed the tasks.

7.2.2 Round One vs. Round Two

Minor changes were made to the system between rounds one and two. Most of the changes were internal and did not affect the user experience with the exception of the zoom functionality. Instead of all videos continuing to play when zoomed-in on a particular clip within a compound view, all videos within the view are paused except the zoomed-in clip. This modification sped up computer response time when zooming in and out of several views in quick succession and also helped stabilize the software.

The changes reduced the restart rate per user from 2 in round one to 0.6 in round two. Between the fourth and fifth participant in round two, another bug was fixed and QuickTime was updated. Since both of these changes were internal and did not change the user interaction with the software, besides increasing its reliability, all results from the second round are grouped together.

It might have been possible to reduce the restart rate further by limiting the number of concurrent videos a user could play at once. Not limiting the user was deemed more important than providing a foolproof implementation.

We also used the pause between rounds to improve the study design. Steps 47-49 were added to the VAST tutorial (Appendix A) instructing the user to convert a compound time view to a single view. The change was motivated by participants using the *zoom-in* functionality more than the *convert to single* command. This is noteworthy because both commands allow the participant to focus on a single movie, but one does so with context (zoom-in) and the other without context (convert to single). Since the convert to single command was not explicitly covered in the tutorial, it was impossible to determine if the zoom-in dominance was the result of participant preference for context or the result of command familiarity. The tutorial modification eliminated this ambiguity for the second round.

Task four was also modified between rounds to reduce ambiguity. The question “*identify the difference between Marcie’s short and deep serve in the following areas: arm, footwork, overall,*” was eliminated because it was not a planning task.

7.2.3 Task One: Analyzing Foreground (Comparison)

The setter in volleyball, like a good quarterback in football, wants to minimize body cues indicating to whom he will give the ball. The opposing team tries to identify the subtle changes in form. The first task asked participants to identify these differences between the outside and middle set and the middle and back set in reference to the setter’s arms, footwork and overall form. The nature of the task required participants to compare multiple video clips together focusing on the foreground of the video.

All participants were able to identify at least one difference between the outside and middle set and at least one difference between the middle and back set. Participants

had the hardest time distinguishing between the footwork of the middle and back set. Only seven of the fourteen participants found a difference. Three people, all volleyball novices, were able to identify differences in all six categories. On average the novices identified differences in 4.8 of the six categories, slightly out performing the experts (4.5 out of six). The novices had greater variability ($s=1.33$) compared to the experts ($s=0.53$). The smaller expert variability suggests that the trained volleyball players had a repertoire of common differences they were looking for. The novices, with no preconceived notions to limit their comparison, were sometimes able to outperform the experts.

To determine problem-solving strategies employed, we analyzed user log data from round two. (Round one log data is not included because of the changes between rounds.) Figure 21 is a visualization of the logs. A six-character code identifies the test subject in the visualization. In the figure, time progresses along the x-axis and every 100-second time interval is denoted by a black tick mark. The red lines on the x-axis indicate a software restart.

A horizontal bar represents each object. A change in bar color indicates an object was converted from one view to another view. For example, when a single movie view is converted to an area view, the underlying object remains the same, but the object's view or context changes.

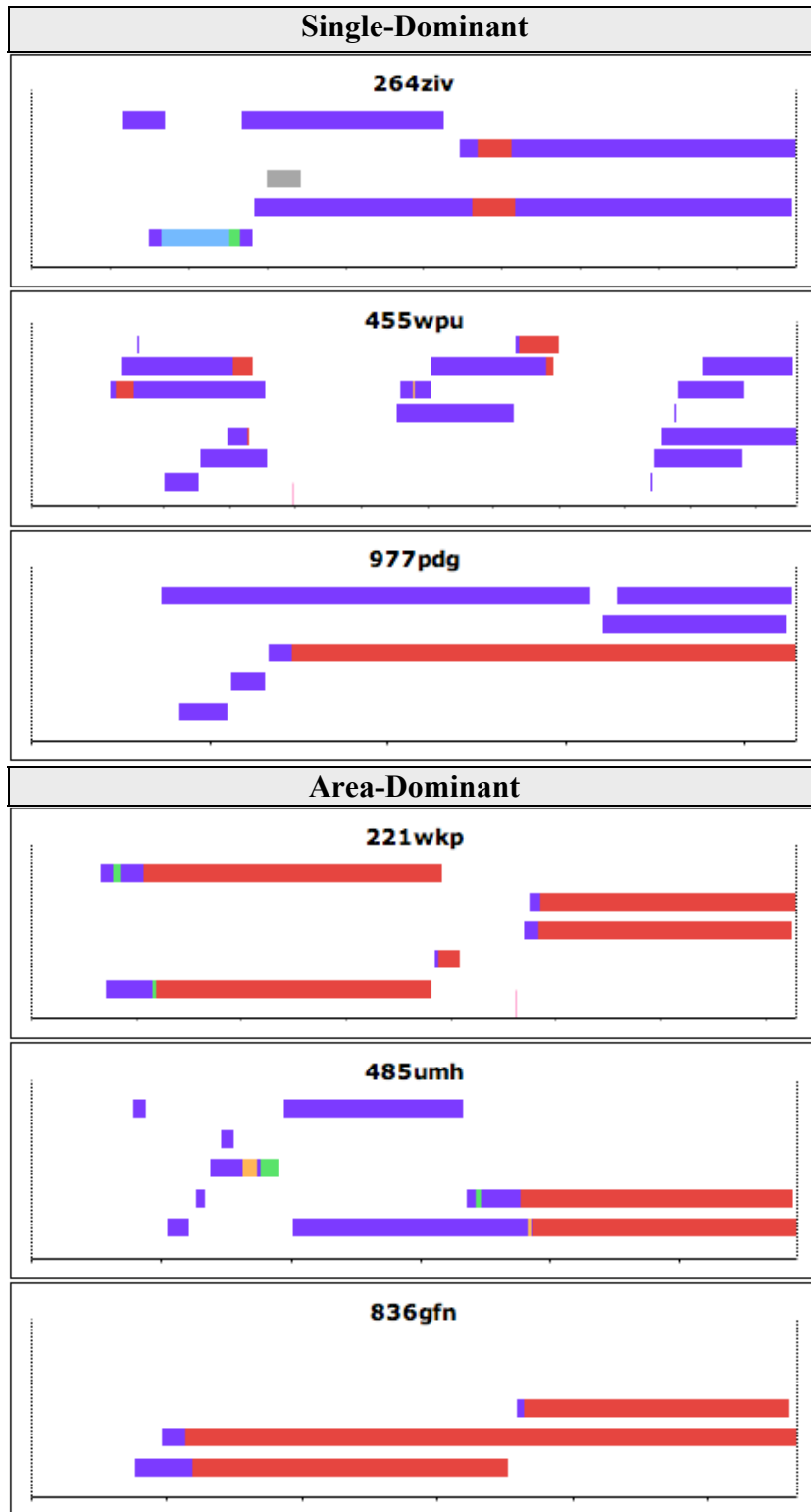


Fig. 21. User log categories for task one – analyzing foreground. (Key on next page.)

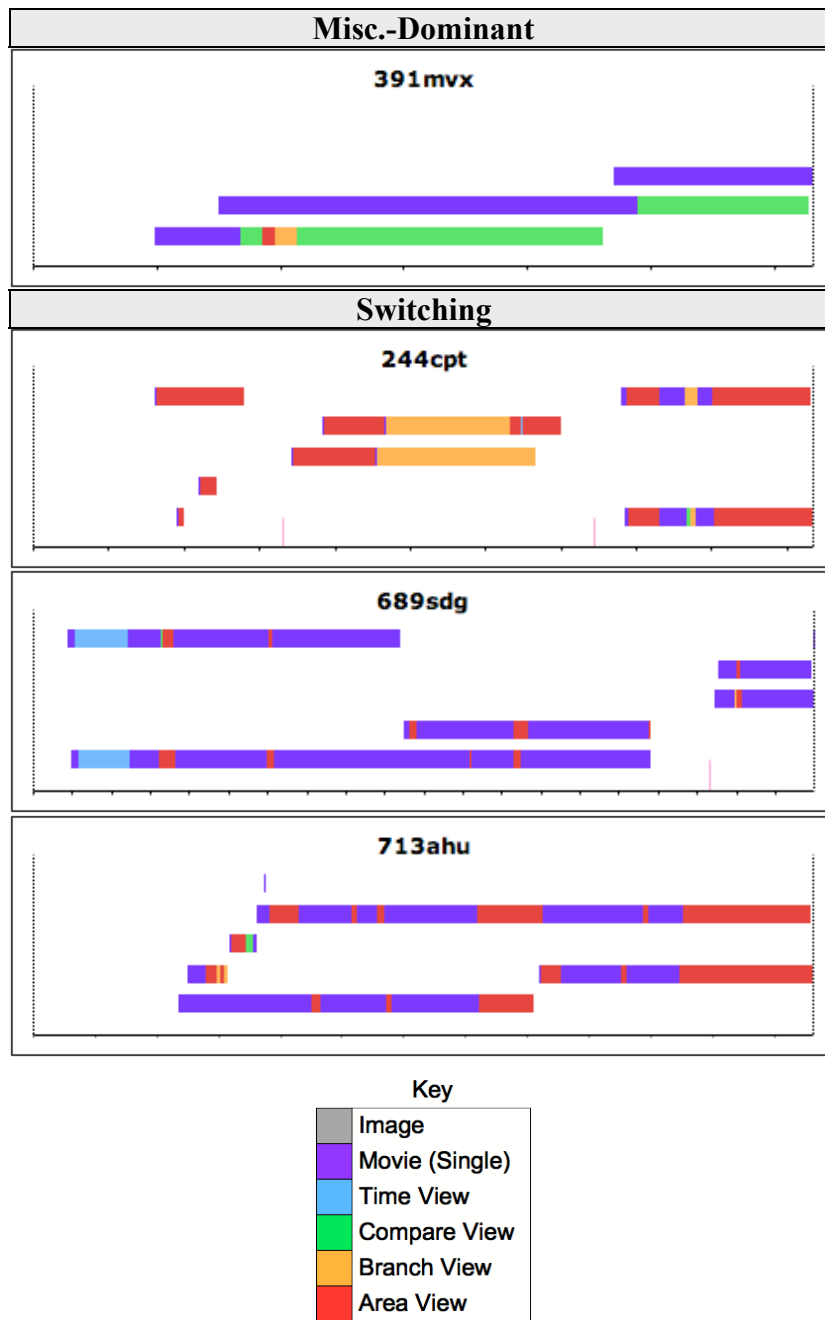


Fig. 21. Continued.

From the logs we see four general strategies employed by participants to complete the task: single-dominant, area-dominant, miscellaneous-dominant, and switching. Those employing a dominant strategy stuck with one particular view. Other views may have been used for a short time, but the participants generally used their preferred view. Most participants using the dominant strategy relied primarily on the single movie view (single-dominant) or the area view (area-dominant). However, one participant preferred the compare view, which we categorize as miscellaneous-dominant. Unlike the participants employing a dominant strategy, the switchers did not rely on any particular view and often made multiple switches between views during the lifetime of a single object.

All users, regardless of strategy, kept multiple objects open at one time. The multiple objects were not limited to two single movies. Often two area views were opened at once potentially allowing eighteen videos to play simultaneously. (Each area view has a placeholder for nine clips, but the data used for the study never used all nine placeholders at once.)

7.2.4 Task Two: (Ambiguous) Identification

Expanding on the knowledge gained in task one, task two required participants to identify fifteen video clips as examples of an outside, middle, or back set. Each clip stopped at the moment the volleyball player contacted the ball making it harder to distinguish between the sets. It also simulated the “reading” skills of a volleyball player trying to anticipate where the setter on the opposing team will put the ball.

Combining both rounds, 71.9% (10.79 out of 15) video clips were identified correctly (Table I). Second round participants had a slightly lower average (71.3%) than the first round (73.3%). Experts in the second round outperformed novices identifying 11.4 compared to 10 clips correctly. (A first round comparison is inappropriate because only one novice participated.) These results are not statistically significant as determined by t-tests assuming unequal variance in the samples.

Table I. Task Two Average Scores (Max Score = 15)

	Novices	Experts	Average
Round 1	14 (1)	10 (3)	11.00 (4)
Round 2	10 (5)	11.4 (5)	10.70 (10)
Average	10.67 (6)	10.88 (8)	10.79 (14)

Note: Number in parenthesis indicates number of users in category.

The margin between experts and novices is diminished when data is combined from both rounds. On average, the novices identified 10.7 clips correctly and the experts 10.9 clips. The smaller margin between novices and experts is due, in part, to the inclusion of volleyball players in round one at the expert level who had some playing experience above recreational, but not at the consistent high level of round two experts. Eliminating the two players with marginal expert experience (one with non-varsity high school and the other with 15 and under club) the cream-of-the-crop experts on average

identified 11.7 clips correctly while the novices identified only 10.7 clips. The difference, however, is not significant.

The high score of 14 (out of 15) was only attained twice (Table II), both times by novices. These two novices (*341ski* and *713ahu*) were also two of the three people who found differences in all six categories of task one.

Table II. Task Two Individual Scores (Max Score = 15)

Novices	Experts
14	13
14	13
11	13
10	12
8	12
7	11
	7
	6

Note: Gray highlight indicates marginal expert experience.

To solve task two, participants followed either a single-open or multiple-open strategy (Figure 22). Those employing a single-open strategy kept only one object, usually a single video, in the workspace. This group of users did not reference any clips outside the test collection. Two users (*391mvx* and *713ahu*) tried to convert one of the test clips to a different view. The test clips were not linked to other clips so the compound views, except the time view, were empty.

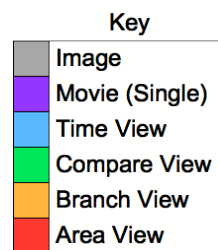
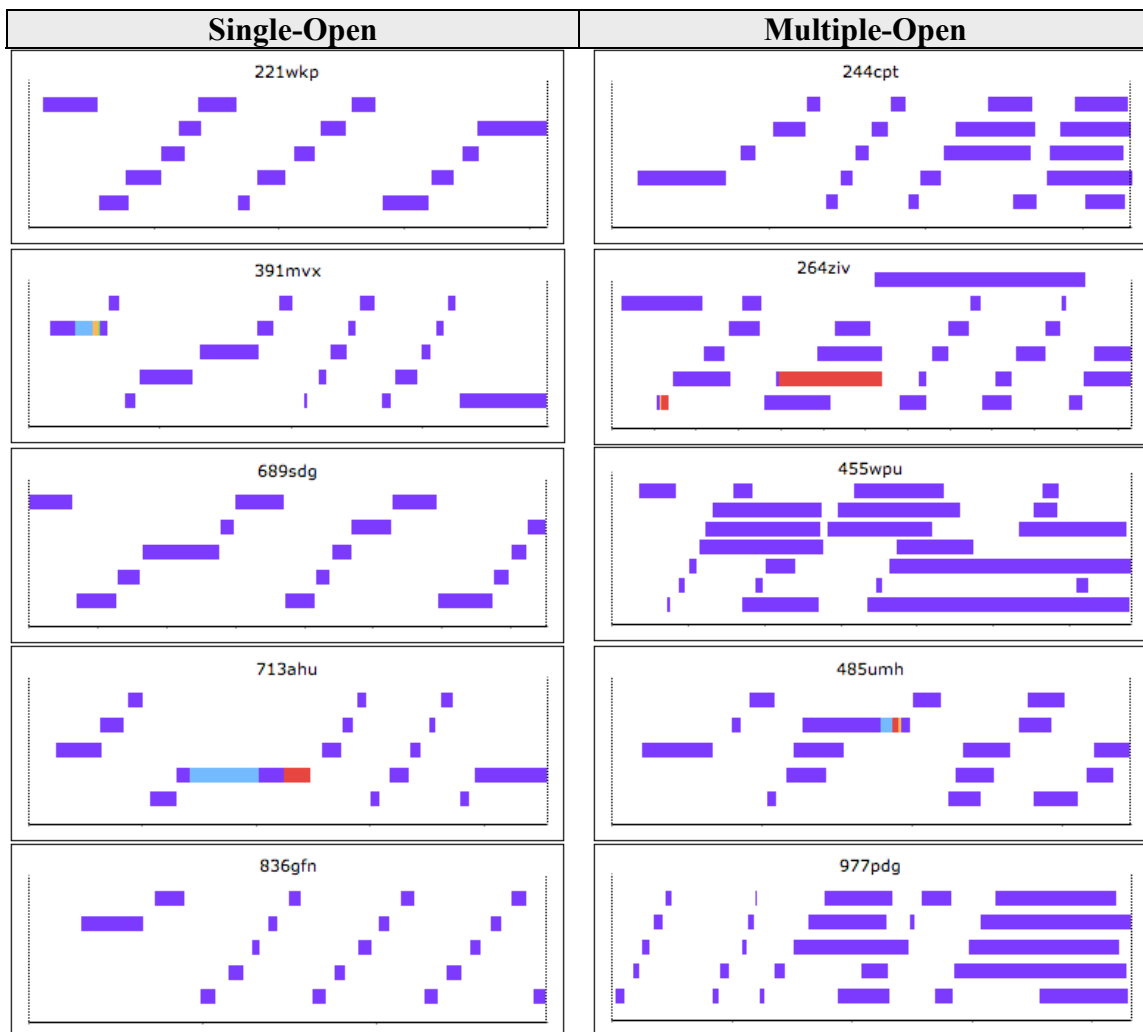


Fig. 22. User log categories for task two – (ambiguous) identification.

Table III. Task Two Individual Scores by Strategy (Max Score = 15)

	Single- Open	Multiple- Open
	14	13
	13	12
	12	11
	10	7
	8	7
Average	11.4	10

The other round two participants used a multiple-open strategy in which multiple objects were kept open in the workspace. The objects were usually single test clips. This strategy allowed participants to compare several unknown videos concurrently. Several users (*264ziv*, *455wpu*, *977pdg*) opened up non-test setting videos. Participant *264ziv* used these known videos to compare to the unknown test clips. It appears from the screen video capture that *455wpu* and *977pdg* may have opened the non-test videos by mistake.

Neither strategy produced superior results. The five top scores were split between strategies (Table III).

7.2.5 Task Three: Analyzing Background

Task three required participants to switch their focus from foreground activities to background objects. In the first question, participants were asked to identify the shirt color and gender of the two people assisting volleyball player Marcie in the videos from January 19, 2008. The second question was more open-ended and required participants to find and identify any flags (if any) in the gym.

	Shirt Color	Gender	Shirt Color	Gender	Correct
221WKP	Blue	M	Blue	F	4
244CPT	Dark blue	M	Blue	F	4
264ZIV	Blue	M	Blue	F	4
391MVX	Blue	M	Blue	F	4
455WPU	Dark blue	M	Blue	F	4
485UMH	Blue	M	Blue	F	4
689SDG	Blue	M	Blue	F	4
713AHU	Blue	M	Blue	F	4
836GFN	Blue	M	White	M	2
977PDJ	Black	M	Blue	F	3
341SKI*	Blue	M	Light Blue	F	4
391JQT*	Blue	M	Blue	F	4
423TIX*	Blue	M	Blue	F	4
725EDI*	Blue	M	Green	M	2
Total Correct	13	14	12	12	

Fig. 23. Task three: Identifying people in background. (Incorrect answers are highlighted in pink. Round one participants are indicated by an *).

	Flags?		Correct
221WKP	USA	Texas	2
244CPT	USA	X	1
264ZIV	USA	X	1
391MVX	USA	X	1
455WPU	USA	X	1
485UMH	USA	X	1
689SDG	USA	%	1
713AHU	USA	X	1
836GFN	X	X	0
977PDJ	USA	X	1
341SKI*	USA	Texas	2
391JQT*	X	X	0
423TIX*	one	X	0.5
725EDI*	USA	X	1
Total Correct	11.5	2	

Fig. 24. Task three: Locating and identifying flags in the gym. (Incorrect answers are highlighted in pink. Yellow highlight indicates partial answer and orange highlight indicates flag was found after task was complete. Round one participants are indicated by *).

Participants performed well on the first task (Figure 23 lists individual results) with eleven of the fourteen users correctly identifying both people. Of those who did not receive a perfect score, one user (*836gfn*) did not reference any videos during the task.

The second question proved more challenging for participants with only two people finding both the American and the Texas flags (Figure 24). A third participant found the Texas flag while working on task four.

To answer these questions most users relied heavily on the area views (Figure 25) in order to scan multiple videos at one time. Often after finding a particular video of interest, the users would zoom-in on a particular video to examine the image in more detail. A few of the participants preferred using the single movie views and used the area-views sparingly, or not at all.

Unlike the analyzing foreground task in which multiple objects were kept open in the workspace, users during this task primarily worked in a sequential nature keeping one object open in the workspace at a time. The notable exception was user *244cpt* who, at one time, had three area views open in the workspace together.

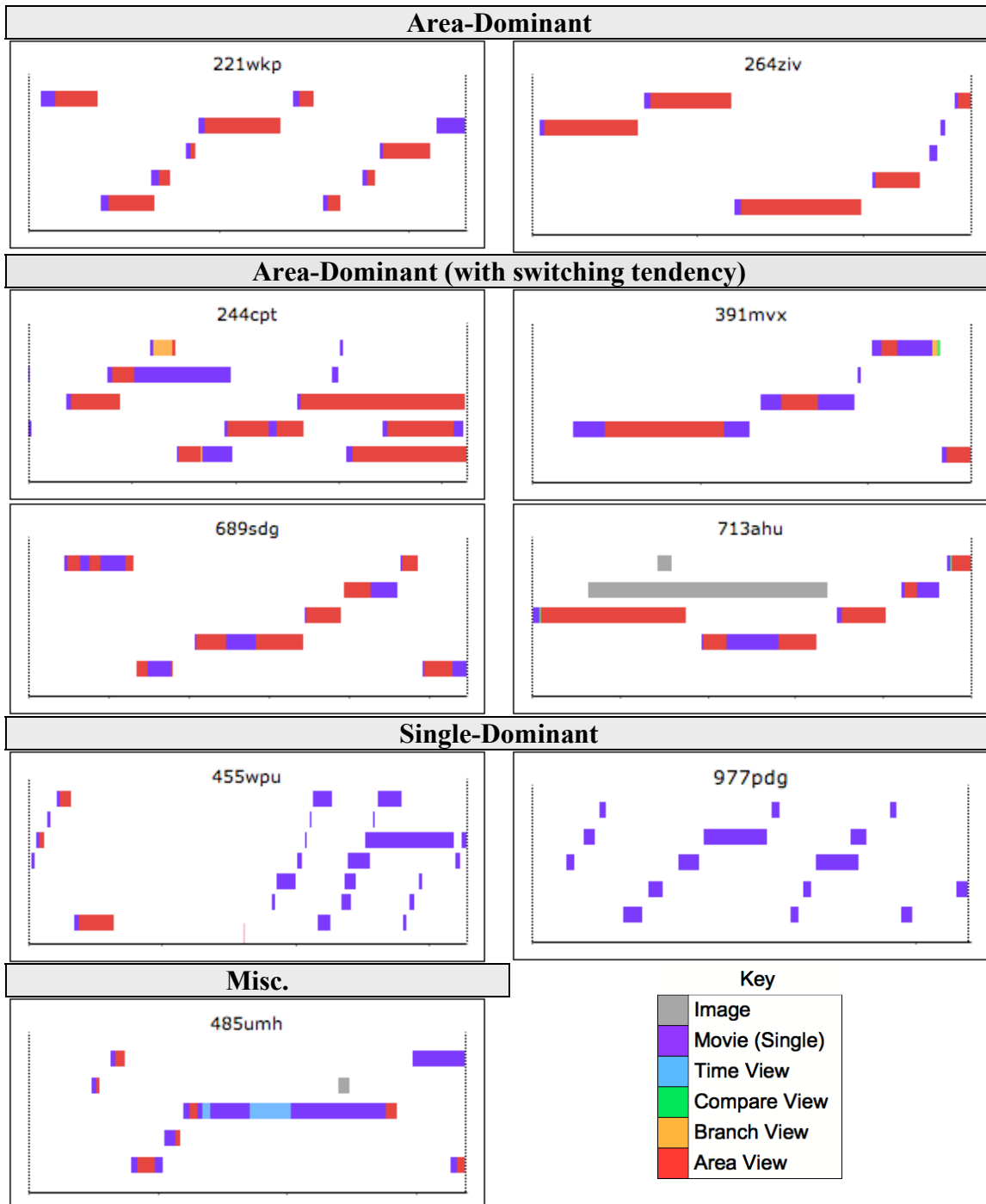


Fig. 25. User log categories for task three – analyzing background. (User *836gfn* did not reference any videos.)

7.2.6 Task Four: Planning

In the final task participants were asked to pick a camera angle, not necessarily one they had already seen, that would best capture a particular serving phenomena. Only three participants, all from round two, picked a new camera angle. Two of those participants suggested the camera be placed overhead.

Less than half of the round two participants accessed any videos during the task (Figure 26). Three of those four users were volleyball novices, suggesting the more experienced volleyball participants relied on prior knowledge and did not see a purpose in checking that knowledge against the videos. The only volleyball expert in round two to access videos (*977pdg*) drew an elaborate sketch (Figure 27) to answer the questions. The other participants drew stick figure variants. This expert also took several pictures from the videos, indicating the video material was used primarily for artistic reference and not directly for answering the question.

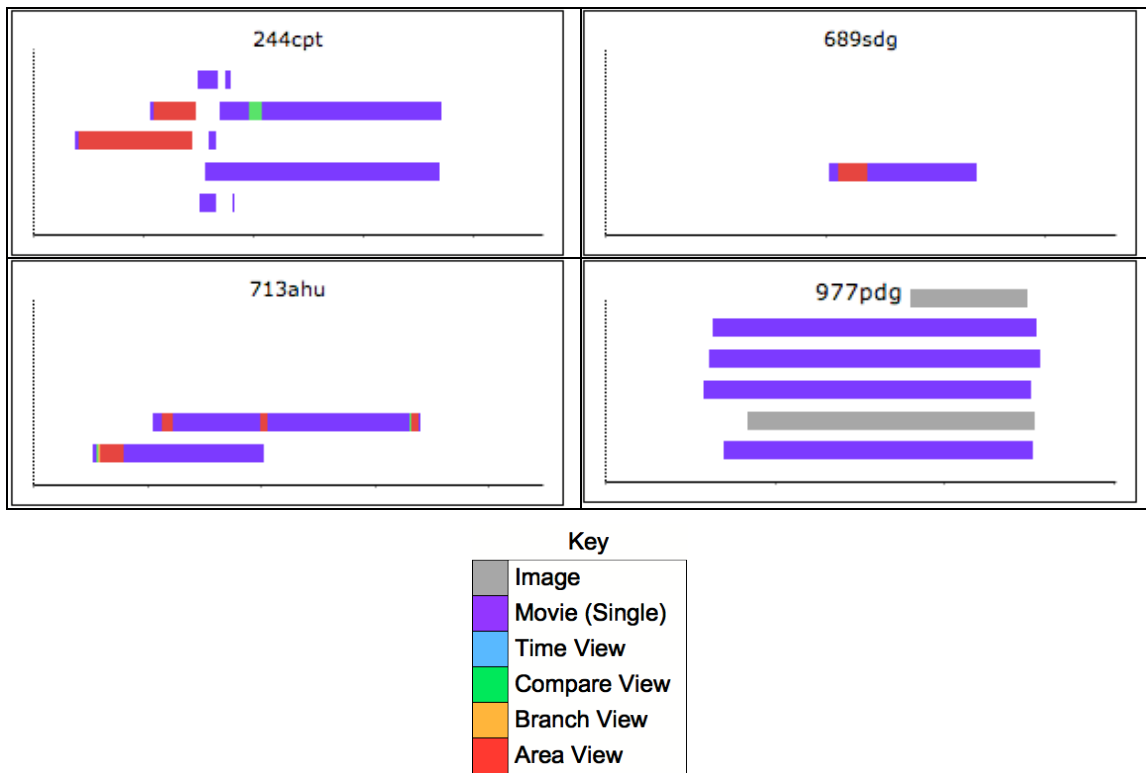


Fig. 26. User logs for task four - planning. Most participants did not reference any videos during this task.

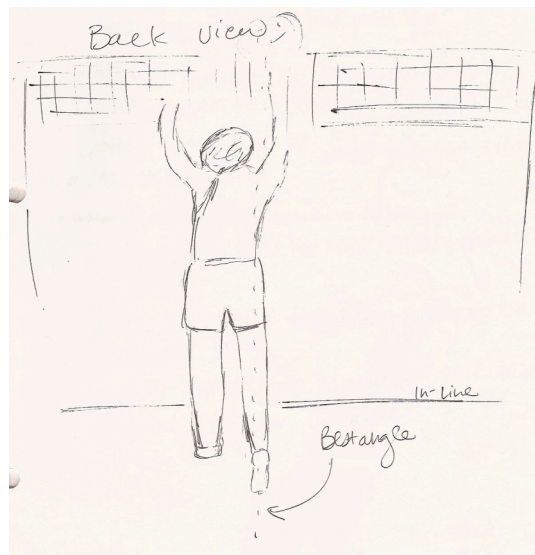


Fig. 27. Participant 977pdg task four's sketch.

7.3 Participant Feedback

At the end of the study, participants were asked to evaluate VAST. Both open-ended and ranking questions were asked. In the ranking questions users were asked to indicate their level of agreement or disagreement with these statements regarding VAST:

- It is easy to use.
- It is fun to use.
- It is aesthetically pleasing.
- It is intuitive to use.

Table IV. VAST User Rating Tallies by Opinion Category

	Easy	Fun	Aesthetic	Intuitive	Total
Strongly Disagree	0	0	0	0	0
Disagree	0	0	0	0	0
Neutral	1	1	3	1	6
Agree	8	7	7	8	30
Strongly Agree	5	6	4	5	20

All participants ranked the system at neutral or better in all categories (Table IV). And only 6 out of the 56 ratings (14 users x 4 questions) were at the neutral level.

To determine a numerical mean score for each question, opinion categories were assigned a value on the following scale: strongly disagree (-2), disagree (-1), neutral (0), agree (+1), and strongly agree (+2). VAST received its highest mean score (1.36) for being *fun to use*, followed closely by *easy* (1.29) and *intuitive* (1.29) to use (Table V). *Aesthetic* received the lowest score (1.07), but it still fell slightly above the *agree* rating.

Table V. VAST User Rating Averages

	Easy	Fun	Aesthetic	Intuitive
Round 1	1.2	1.5	1.0	1.2
Round 2	1.5	1.0	1.3	1.5
All	1.29	1.36	1.07	1.29

(-2)Strongly Disagree; (-1)Disagree; (0)Neutral; (+1)Agree; (+2)Strongly Agree

The participant comments were similarly positive. One participant wrote, “I enjoyed using the program and would love to use it more if I had the chance. I would really like to have some of the coaching staff at A&M participate because I think they would really like it too!” Another thought VAST “would be great to teach the fundamentals of volleyball.” And another, “I am considering coaching so it would be great for that!” It was called a “great program!!” by one user, and another stated that “it has an intuitive interface which is actually easy to use.”

When asked what they liked most about VAST, people most frequently mentioned the multiple camera angles (“lots of angles”, “multiple angles”, “all of the different angles at once”, “how all the angles are able to be viewed...allow[s] the viewer to see things that just one camera may have missed”). The area view was also mentioned several times by participants, including one person who felt it was “cool & beneficial to be able to see an action from several different point[s] of view.” Another liked the area view because it “allowed me to focus on separate parts and the whole view.” Participants

also appreciated the ability to switch between views, stating it was “very easy to switch between different views,” and one liked the “multiple comparison options.”

While one user praised the “intuitive position of the different video angles in the Area View,” another was less complimentary of the video organization in the branch and compare views, stating “I have no idea what branch & compare did.” Organization was also an issue for another user who stated the labeling was “a little disorganized” and another suggest VAST “may need good documentation.” Most of the other system negative comments, however, were easily corrected usability issues.

The suggestions to improve search capabilities were expected, as we intentionally focused on analysis task, and not the search task. Improving the search engine and visualization will be addressed in the future.

Other functionality requests include the ability to speed up video playback. This feature was removed from the original system to simplify the interface. The ability to “freeze frame and view the exact frame from all the angles at once” was also requested. (This request was the only video synchronization related-comment by any user). Many of the other suggestions were usability-related or requests for functionality that existed, but the user had not located.

Despite shortcomings, the majority of participants (5 experts and 3 novices, eight total) said they would definitely use VAST if it was available to them. Several of these participants cited using it for sports-related analysis, but others suggested using it for surveillance, crime scene investigation, following multiple people in a movie or computer game, “family reunion videos”, studying the “motion of people/animals for

animation purposes”, “data visualization,” class presentations, and “how to videos.” Only three participants answered negatively (“no”, “probably not”, “not right now”), citing a lack of relevance or need in their current activities. The remaining three users fell between the definitive yes and no.

8. DISCUSSION

8.1 Analysis (Volleyball)

Overall, participants had a favorable opinion of VAST and considered working with the system an enjoyable experience during the user studies in the volleyball domain. The learning curve for basic system functionality seems fairly low, as participants completed all tasks and rated the system's intuitiveness and ease of use high. Participants displayed a wide variety of strategies and ingenious uses of VAST when solving the tasks demonstrating the system's flexibility.

Based on user performance, VAST is better suited for comparison (foreground analysis) and ambiguous identification tasks, than open-ended background analysis tasks. However, part of the difference in user performance probably reflects a variation in task difficulty and not a reflection of the system itself. The planning task had no right or wrong answer, so VAST's success in helping participants in this area is difficult to judge objectively. Domain experts tended to view VAST unnecessary for answering the planning question, while the majority of domain novices used it for the task.

Users appeared to grasp the concept of views and conversion between views quickly. They found viewing multiple related clips simultaneously within a view useful. The desire for good documentation by one user and puzzlement concerning the branch and compare views' purpose by another participant suggest that clip organization within certain views was not readily apparent to all users. The different views were not equally popular. Participants requested the single and the area view conversions most frequently.

(Conversion requests do not include the single video view each clip initially begins as when entering the workspace.) These two views are on opposite sides of the context-providing spectrum. In regards to related video clips, the single view provides the least context and the area view the most. However, the single view provides easily accessible video functionality and event metadata, while all other views require the user to zoom-in on a particular video clip before accessing similar functionality. Both contexts serve different needs of the user during video analysis.

Also, the lack of sophistication of the search interfaces could contribute to the area view's popularity. Unlike the rest of VAST, the search interface is not visually oriented. It relies on textual information to help the user find pertinent video clips. The area view affords browsing the most clips simultaneously thus potentially acting as a default browser.

The branch and compare views were used less frequently than the area view. Also, despite participants manipulating video time frequently (i.e., adjusting the movie time bar, using the next frame or previous frame button), the time view was used surprisingly infrequently. Allowing customization of the time view's granularity between snapshots and addressing usability issues might increase the view's use.

Navigation functionality within a view (i.e., zooming-in and out, moving up, down, left, and right) was used fairly often. Each navigation command essentially modifies the view's context, allowing the user to tweak the view to suite his purposes. Other system features (e.g., take picture, transparency) were used on a sporadic basis. Participants left some features (e.g., adding and removing video clips from views)

completely unexplored. The variation in system feature usage is partially a result of tutorial focus and partially based on the nature of the assigned tasks. Only by observing participants using VAST over an extended period of time and a wider range of tasks could the reason users chose not to use certain features become clear.

8.2 Implications (Bat Lab)

Based on the user study results in the volleyball domain, VAST should successfully support the wide range of user expertise and goals in the Bat Lab. Although a user study was not conducted in the Bat Lab domain, two domain experts participated in separate, formal, demo-and-discuss sessions. Both interviewees spent multiple years in the Bat Lab as doctoral students. (Informal demo-and-discuss sessions were conducted with a variety of Bat Lab participants during the earlier stages of development.)

According to one of the formal interviewees, VAST would be suitable for teaching, training, planning, and presentation purposes. The other interviewee stated VAST would be useful for teaching and presentations, as well as during experiments. If VAST was available during experiments he said, a user could compare what happened in the past (during the control part of the experiment) with the live video data being collected. He also thought the time view would be useful for analyzing changes in a vessel's diameter and the branch view helpful in observing changes in a vessel network.

Both interviewees expressed a desire for measuring tools to quantify observed phenomena in the video clips. Their enthusiasm for VAST's analysis support ability was tempered by the lack of these tools. In addition, one expert tended to compare

VAST's features (or lack thereof) with the current Bat Lab software focused on quantifying, not context-demanding, tasks. The areas where VAST was considered affective (e.g., training) require video analysis but not in a manner traditionally considered "analyzing" by the interviewees.

It is also possible to apply lessons learned about VAST, during the volleyball domain user studies, to judge VAST's ability to support context-demanding analysis tasks in the Bat Lab (such as those found in Section 2.1). In the volleyball domain user studies, participants were successful in the foreground analysis (comparison) and ambiguous identification task. VAST should support similar questions in the Bat Lab domain. (E.g., *what are the differences between the control and experiment events?*)

Questions similar to *what does the surrounding vessel structure look like*, should also be fairly well supported by VAST since the vessels of interest are comparatively large and usually in the foreground focus. When the area of interest is fairly small, the entire structure could be viewed within one (or several) compound view(s). If the area of interest becomes too large, users may struggle with the task.

However, VAST may not be useful in answering questions such as *did these capillaries exist six months ago?* The question should be a simple comparison tasks, but capillaries are small and tend to disappear in the background. Unless a person can pinpoint the exact location in both clips, the task could devolve into an open-ended background analysis task, which participants struggled with during the user study.

As noted before, planning-oriented tasks are hard to judge for success but strategies employed during the tasks can be observed. In the volleyball-planning task,

many user study participants did not use VAST but relied on previous knowledge when answering the question. The greater complexity of the bat's cardiovascular system (compared to the game of volleyball) suggests a greater need to rely on VAST while addressing planning questions in the Bat Lab domain. (E.g., *where is a good vessel occlusion site?*) Deploying VAST in the Bat Lab for a formal, extended user study would be the next step in judging VAST's analysis support ability in this domain.

9. FUTURE WORK

In the future, we will address the usability improvement suggestions that surfaced during the user studies. We will incorporate a better search interface in VAST, including an interactive timeline on which all events will be marked. The location diagram will also be interactive. Users will be able to request video clips from a particular area of interest by selecting a portion of the diagram.

VAST's view library will be expanded so that we can test other view configurations. We are particularly interested in increasing the support for visually complex documents [Audenaert et al. 2008] and investigating their role in providing context for the videos. Handwritten lab notes and sketches could accompany the videos of the related science experiment or coaching notes could be included with the practice video. These types of documents are already produced during the course of the event and could provide further insight with limited overhead.

We would also like to support users creating their own compound views from the existing basic views. In addition to creating different view layouts, users with a set of pre-determined location perspectives would be able to determine how videos within a view are related to one another. This query-by-template capability would give users more flexibility to repurpose the videos to answer a specific question.

It is unclear whether these user-created views would be understandable or usable to other people. Probably the views would be partially usable by others (e.g., participants in the user studies did not create their own views but were still able to

complete the required tasks), but they may not explicitly be able to state the relationship between the videos without additional information.

The intelligence of the context engine, which manages the arrangements of the views, will also be expanded in the future. Arrangements will become truly dynamic, determined on the fly by the context engine at the user's request. Additional algorithms will be developed and tested for converting one view (and underlying arrangement) to another view. Currently, only the primary video within a view/arrangement is used to determine the next arrangement.

Additional user studies will address several open questions including: does the grouping of videos together via views help people manage more videos than if they were not grouped together? Requesting compound views and switching between views occurred frequently during the user studies, suggesting participants found the context provided by the views helpful. Does a better search interface, providing more context via the interactive timeline and diagram, increase or decrease the participants' reliance on the view contexts?

A long-term study would lend insight into how people use VAST after becoming more familiar with the software. What tasks do people use VAST for when not given a particular problem to solve? Do people start using the same problem-solving strategy for all tasks (e.g., employing the one or two views they are most comfortable with) or do the strategies for solving different tasks remain distinct?

We will also expand and test VAST in several other domains. In meteorology, it could help scientists analyze weather phenomena using satellite and radar feeds.

Amateur satellite image readers could use it to compare past imagery of a particular location with new images or to learn the basics of image/video analysis with a vetted information set. City managers or the average driver could use it to analyze traffic patterns and conditions from highway video cameras and thus aid emergency planning, route familiarization, or alternative path decisions.

10. CONCLUSION

We have presented a domain-generic philosophy and design for supporting video analysis. Based on our experience in the Bat Lab, a typical science community, we have identified several critical constraints to be addressed within the design. The producers' overhead (i.e., metadata input) should be kept to a minimum. The design should be flexible to support consumers with a variety of end goals and expertise. Video clips should be contextualized to support the harder tasks (i.e., context-demanding questions) that are not easily answered with a single video.

To provide context, we proposed a location-oriented solution. Location is one of the basic methods by which people organize and contextualize information. It lends itself well to human-interaction navigational metaphors [Arbeeny and Silver 2001]. Also, it can be described from several different perspectives, providing a natural flexibility to any design incorporating location.

We also proposed a novel paradigm for viewing and manipulating the contextualized video. *Views* allow the user to interact with a group of related videos. Each view has a native *arrangement* structure defining how the clips are grouped together. Each view and associated arrangement provides a different location-based context for the video clip of interest. Switching between views, the user can examine a video clip in multiple contexts. A context engine handles the switching between views.

VAST is an implementation of the proposed design. To test the flexibility of the design, video data collected a priori by the Bat Lab was successfully put in the system.

To prove applicability to other domains, we switched from the scientific to the athletic (volleyball) domain to conduct the user studies.

Both domain novices and experts participated in the formal evaluation. Three tasks (1,2, and 4) in the study were selected to reflect questions a volleyball player or coach might need to answer. The other task, background analysis, was a realistic (e.g., security application) non-volleyball task. All tasks required participants to examine volleyball practice videos they did not produce personally.

Domain expertise did not contribute significantly to the success or failure of participants. Working style trends were evident from the VAST user logs, but no strategy produced superior results. Nor did the problem solving-strategies employed by the users split along expert and novice lines. The use of one strategy in a particular task did not indicate that the same strategy would be employed for a different task.

The most challenging task for participant was not volleyball related, but the open-ended background analysis task (*Is there a flag in the gym? If yes, how many and what kind?*). Despite a fairly small corpus of video clips, most users missed the second (*Texas*) flag. As these results emphasize, no guarantee exists that all instances of an object will be identified, especially if the person looking for the object is not sure it exists. Most likely, participants would have been more diligent looking for the Texas flag if they had known two flags existed. To increase the odds of answering the open-ended question correctly, VAST would need to support a more systematic approach to video exploration. The flexibility of VAST in supporting different users' working strategies is beneficial in most tasks, but not this task.

In all other tasks, participants performed well. The overall success in the tasks and positive participant feedback indicates VAST provided enough context for users to understand and use video they did not produce. Participants also grasped the concept of views easily and employed them frequently to observe multiple videos at once. Most users did not express confusion about the internal arrangement of the videos in the views, but it did cause concern for some.

Many participants kept multiple views open in the workspace and had little trouble juggling the visual attention demands. The area view (with the greatest potential for active videos) was used more frequently than the branch, the compare, or the time view. When users wanted to examine a video clip in more detail within a compound view, they chose to zoom-in more often than selecting the convert to single command, suggesting that the users wanted the context provided by the surrounding video clips within easy access. The lack of video synchronization did not bother users from either a usability standpoint nor prevented them from successfully completing the task, suggesting synchronization for some tasks is not necessary.

Overall, VAST provided effective support to both domain novices and experts as they analyzed video. The underlying model of VAST was able to support two very distinct domains without burdening the producer or frustrating the consumer.

REFERENCES

- ABOWD, G. D., GAUGER, M., AND LACHENMANN, A. 2003. The family video archive: An annotation and browsing environment for home movies. In *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR '03)*. ACM, New York, 1-8.
- ARBEENY, S. AND SILVER, D. 2001. Spatial navigation of media streams. In *Proceedings of the Ninth ACM International Conference on Multimedia (MULTIMEDIA '01)*. ACM, New York, 467-470.
- ATZENBECK, C. AND NÜRNBERG, P. J. 2005. Constraints in spatial structures. In *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia (HYPERTEXT '05)*. ACM, New York, 63-65.
- ATZENBECK, C. AND NÜRNBERG, P. J. 2006. Structural analysis for the knowledge management domain. *The New Review of Hypermedia and Multimedia*, 12, 2, 191-216.
- AUBERT, O. AND PRIÉ, Y. 2005. Advene: Active reading through hypervideo. In *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia (HYPERTEXT '05)*. ACM, New York, 235-244.
- AUDENAERT, N., LUCHESSÉ, G., AND FURTUA, R. 2008. CritSpace: Using spatial hypertext to model visually complex documents. In *Digital Humanities 2008* (Oulu, Finland).
- BADRE, A. N., GUZDIAL, M., HUDSON, S.E., AND SANTOS, P.J. 1995. A user interface evaluation environment using synchronized video, visualizations and event trace data. *Software Quality Journal* 4, 2, 101-113.
- BARGERON, D., GUPTA, A., GRUDIN, J., AND SANOCKI, E. 1999. Annotations for streaming video on the web. In *Extended Abstracts on Human Factors in Computing Systems (CHI '99)*. ACM, New York, 278-279.
- BOCCONI, S., NACK, F., AND HARDMAN, L. 2005. Vox populi: A tool for automatically generating video documentaries. In *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia (HYPERTEXT '05)*. ACM, New York, 292-294.
- BUCHANAN, M. C. AND ZELLWEGER, P. T. 1992. Specifying temporal behavior in hypermedia documents. In *Proceedings of the ACM Conference on Hypertext*. D. Lucarella, J. Nanard, M. Nanard, and P. Paolini, Eds., ACM, New York, 262-271.

- BURR, B. 2006. VACA: A tool for qualitative video analysis. In *Extended Abstracts on Human Factors in Computing Systems (CHI '06)*. ACM, New York, 622-627.
- CHAMBEL, T. AND GUIMARÃES, N. 2002. Context perception in video-based hypermedia spaces. In *Proceedings of the Thirteenth ACM Conference on Hypertext and Hypermedia (HYPERTEXT '02)*. ACM, New York, 85-94.
- CHAMBEL, T., ZAHN, C., AND FRINKE, M. 2004. Hypervideo design and support for contextualized learning. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies*. IEEE Computer Society, Los Alamitos, CA, 345-349.
- CHIU, P., GIRGENSOHN, A., LERTSITHICHAJ, S., POLAK, W., AND SHIPMAN, F. 2005. MediaMetro: Browsing multimedia document collections with a 3D city metaphor. In *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*. ACM, New York, 213-214.
- CHRISTEL, M. G., HAUPTMANN, A. G., WACTLAR, H. D., AND NG, T. D. 2002. Collages as dynamic summaries for news video. In *Proceedings of the Tenth ACM International Conference on Multimedia (MULTIMEDIA '02)*. ACM, New York, 561-569.
- CLIFFORD S., YOUNG, R., AND WILLIAMSON, P. 2007. Assessing the early characteristics of autistic disorder using video analysis. *Journal of Autism and Developmental Disorders*, 37, 2, 301-313.
- COSTA, M., CORREIA, N., AND GUIMARÃES, N. 2002. Annotations as multiple perspectives of video content. In *Proceedings of the Tenth ACM International Conference on Multimedia (MULTIMEDIA '02)*. ACM, New York, 283-286.
- DARTFISH. 2008. Dartfish software. *Dartfish Video Software Solutions*. <http://www.dartfish.com/en/index.htm> Accessed May 13, 2008.
- DAVIS, M. 1993. Media streams: An iconic visual language for video annotation. In *Proceedings of the 1993 IEEE Symposium on Visual Languages (Bergen, Norway)*. 196-202.
- DE SILVA, G.C., YAMASAKI, T., AND AIZAWA, K. 2005. Evaluation of video summarization for a large number of cameras in ubiquitous home. In *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*. ACM, New York, 820-828.
- DING, W., MARCHIONINI, G., AND SOERTEL, D. 1999. Multimodal surrogates for video browsing. In *Proceedings of the Fourth ACM Conference on Digital Libraries (DL '99)*. ACM, New York, 85-93.

- FOCUS. 2008. Focus software. *Elite Sports Analysis – Sports Statistics and Performance Analysis*. <http://www.elitesportsanalysis.com> Accessed May 13, 2008.
- GHADIRIAN, P., AND BISHOP, I.D. 2002. Composition of augmented reality and GIS to visualize environment changes. In *Proceedings of the Joint ARISA and Institution of Surveyors Conference* (Adelaide, South Australia). 25-30.
- GIRGENSOHN, A., WILCOX, L., SHIPMAN, F., AND BLY, S. 2004. Designing affordances for the navigation of detail-on-demand hypervideo. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)*. ACM, New York, 290-297.
- GIRGENSOHN, A., KIMBER, D., VAUGHAN, J., YANG, T., SHIPMAN, F., TURNER, T., RIEFFEL, E., WILCOX, L., CHEN, F., AND DUNNIGAN, T. 2007. DOTS: Support for effective video surveillance. In *Proceedings of the 15th International Conference on Multimedia (MULTIMEDIA '07)*. ACM, New York, 423-432.
- GOOGLE. 2008. Google video. <http://www.video.google.com> Accessed February 12, 2008.
- GRAHAM, J., EROL, B., HULL, J. J., AND LEE, D. 2003. The video paper multimedia playback system. In *Proceedings of the Eleventh ACM International Conference on Multimedia (MULTIMEDIA '03)*. ACM, New York, 94-95.
- GREENE, S., MARCHIONINI, G., PLAISANT, C., AND SHNEIDERMAN, B. 2000. Previews and overviews in digital libraries: Designing surrogates to support visual information seeking. *Journal of the American Society for Information Science* 51, 4, 380-393.
- GRÉMILLET, D., ENSTIPP, M.R., BOUDIFFA, M., LIU, H. 2006. Do cormorants injure fish without eating them? An underwater video study. *Marine Biology* 148, 5, 1081-1087.
- GUERLAIN, S., GREEN, K.B., LAFOLLETTE, M., MERSCH, T.C., MITCHELL, B.A., POOLE, G.R., CALLAND, J.F., LV, J., AND CHEKAN, E. G. 2004. Improving surgical pattern recognition through repetitive viewing of video clips. *IEEE Transactions on Systems, Man and Cybernetics, Part A* 34, 6, 699-707.
- HARRISON, B. L. AND BAECKER, R. M. 1992. Designing video annotation and analysis systems. In *Proceedings of the Conference on Graphics Interface '92*. K. S. Booth and A. Fournier, Eds., Morgan Kaufmann Publishers, San Francisco, CA, 157-166.
- HIBINO, S. AND RUNDENSTEINER, E. A. 1996. MMVIS: A multimedia visual information seeking environment for video analysis. In *Conference Companion on Human Factors in Computing Systems: Common Ground (CHI '96)*. M. J. Tauber, ed. ACM, New York, 15-16.

- HOUTEN, Y., SCHURMAN, J. G., AND VERHAGEN, P. 2004. Video content foraging. *Image and Video Retrieval*. Lecture Notes in Computer Science, Vol. 3115, Springer, Berlin, 15-23.
- ICHIMURA, S., AND MATSUSHITA, Y. 2005. Web-based video editing systems for sharing clips collected from multi-users. In *Proceedings of the Seventh IEEE International Symposium on Multimedia (ISM '05)*. IEEE, Los Alamitos, CA.
- KIM, S., LEE, S., KIM, K., LEE, J. 2003. A unified visualization framework for spatial and temporal analysis in 4D GIS. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS '03)* 6, 3715-3717.
- KIMBER, D., DUNNIGAN, T., GIRGENSOHN, A., SHIPMAN, F., TURNER, T., AND YANG, T. 2007. Trailblazing: Video playback control by direct object manipulation. *2007 IEEE International Conference on Multimedia and Expo* (Beijing, China). IEEE, 1015-1018.
- KOMLODI, A. AND MARCHIONINI, G. 1998. Key frame preview techniques for video browsing. In *Proceedings of the Third ACM Conference on Digital Libraries (DL '98)*. I. Witten, R. Akscyn, and F. M. Shipman, Eds., ACM, New York, 118-125.
- LEE, C. A. 2007. Taking context seriously: A framework for contextual information in digital collections. *UNC SILS Technical Report 2007-04*. October 18, 2007. http://sils.unc.edu/research/publications/reports/TR_2007_04.pdf.
- LIPPMAN, A. 1980. Movie-maps: An application of the optical videodisc to computer graphics. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '80)*. ACM Press, New York, 32-42.
- MACKAY, W. E. AND BEAUDOUIN-LAFON, M. 1998. DIVA: Exploratory data analysis with multimedia streams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. C. Karat, A. Lund, J. Coutaz, and J. Karat, Eds., ACM Press/Addison-Wesley Publishing Co., New York, 416-423.
- MARSHALL, C. C., SHIPMAN, F. M., AND COOMBS, J. H. 1994. VIKI: Spatial hypertext supporting emergent structure. In *Proceedings of the 1994 ACM European Conference on Hypermedia Technology (ECHAT '94)*. ACM, New York, 13-23.
- MCCURDY, N. J. AND GRISWOLD, W. G. 2005. A systems architecture for ubiquitous video. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys '05)*. ACM Press, New York, 1-14.

- MUJACIC, S., AND DEBEVC, M. 2007. A formal approach to hypervideo design. In *Proceedings of 2007 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference Focused on Speech and Image Processing, Multimedia Communications and Services* (Maribor, Slovenia). IEEE, 189-192.
- NCAA. 2007a. *2007 Division II Women's Volleyball Championship Handbook*. http://www.ncaa.org/library/handbooks/volleyball/2007/2007_d2_w_volleyball_handbook.pdf Accessed June 23, 2008. 27-28.
- NCAA. 2007b. *2007 Division III Women's Volleyball Championships Handbook*. http://www.ncaa.org/library/handbooks/volleyball/2007/2007_d3_w_volleyball_handbook.pdf Accessed June 23, 2008. 27-28.
- NOBRE, E. M. N., AND CAMARA, A. S. 2001. Spatial video: Exploring space using multiple digital videos. In *Preliminary Proceedings of the 6th Eurographics Workshop on Multimedia*. <http://virtual.inesc.pt/egmm2001/proceedings/nobre.pdf> Accessed August 9, 2008.
- NORDT, M., MEISNER, J., DONGAONKAR, R., QUICK, C. GATSON, G. KARADKAR, U., AND FURUTA R. 2006. eBat: Technology-enriched research community model for the life sciences. In *Proceedings of the American Society for Information Science and Technology Annual Meeting (ASIS&T 2006)* (Austin, TX).
- PARENT, C., SPACCAPIETRA, S., AND ZIMÁNYI, E. 1999. Spatio-temporal conceptual models: Data structures + space + time. In *Proceedings of the 7th ACM International Symposium on Advances in Geographic Information Systems (GIS '99)*. C. B. Medeiros, ed. ACM, New York, 26-33.
- PATTANASRI, N., JATOWT, A., AND TANAKA, T. 2006. Enhancing comprehension of events in video through explanation-on-demand hypervideo. *Advances in Multimedia Modeling*. Lecture Notes in Computer Science, Vol. 4351, Springer, Berlin, 535-544.
- PAUTY, J., COUDERC, P., AND BANÂTRE, M. 2005. Using context to navigate through a photo collection. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (MobileHCI '05) 111*, ACM, New York, 145-152.
- PEA, R., MILLS, M. ROSEN, J., DAUBER, K., EFFELSBERG, W., AND HOFFERT, E. 2004. The diver project: Interactive digital video repurposing. *IEEE Multimedia* 11, 1, 54-61.
- RAMOS, G. AND BALAKRISHNAN, R. 2003. Fluid interaction techniques for the control and annotation of digital video. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)*. ACM, New York, 105-114.

- SAWHNEY, N., BALCOM, D., AND SMITH, I. 1996. HyperCafe: Narrative and aesthetic properties of hypervideo. In *Proceedings of the Seventh ACM Conference on Hypertext (HYPERTEXT '96)*. ACM, New York, 1-10.
- SHIPMAN, F., HSIEH, H., MALOOR, P., AND MOORE, J. M. 2001. The visual knowledge builder: A second generation spatial hypertext. In *Proceedings of the Twelfth ACM Conference on Hypertext and Hypermedia (HYPERTEXT '01)*. ACM, New York, 113-122.
- SHIPMAN, F., GIRGENSOHN, A., AND WILCOX, L. 2005. Hypervideo expression: Experiences with hyper-hitchcock. In *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia (HYPERTEXT '05)*. ACM, New York, 217-226.
- SNOEK, C.G. M. AND WORRING, M. 2005. Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications* 25, 1, 5-35.
- SONG, Y. AND MARCHIONINI, G. 2007. Effects of audio and visual surrogates for making sense of digital video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, 867-876.
- TANG, L., AND KENDER, J. R. 2006. Designing an intelligent user interface for instructional video indexing and browsing. In *Proceedings of the 11th International Conference on Intelligent User Interfaces (IUI '06)*. ACM, New York, 318-320.
- TSE, T., MARCHIONINI, G., DING, W., SLAUGHTER, L., AND KOMLODI, A. 1998. Dynamic key frame presentation techniques for augmenting video browsing. In *Proceedings of the Working Conference on Advanced Visual interfaces (AVI '98)*. T. Catarci, M. F. Costabile, G. Santucci, and L. Taranfino, Eds., ACM, New York, 185-194.
- TUKEY, J. W. 1980. We need both exploratory and confirmatory. *The American Statistician* 34, 1, 23-25.
- UYTTENDAELE, M., CRIMINISI, A., KANG, S.B., WINDER, S., HARTLEY, R., AND SZELISKI, R. 2003. High-quality image-based interactive exploration of real-world environments. *Microsoft Research Technical Report MSR-TR-2003-61*, October 2003.
- VANGENOT, C., PARENT, C., AND SPACCAPIETRA, S. 2002. Modelling and manipulating multiple representations of spatial data. In *Symposium on Geospatial Theory, Processing and Applications* (Ottawa, Canada).
- VANGENOT, C. 2004. Multi-representation in spatial databases using the MADS conceptual model. In *ICA Workshop on Generalisation and Multiple Representations*. (Leicester, United Kingdom).

VOLKMER, T., SMITH, J. R., AND NATSEV, A. 2005. A web-based system for collaborative annotation of large image and video collections: An evaluation and user study. In *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*. ACM, New York, 892-901.

WU, W., YANG, J. AND ZHANG, J. 2006. A Multimedia system for route sharing and video-based navigation. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2006)* (Ontario, Canada). IEEE, 73-76.

YOUTUBE. 2008. YouTube. <http://www.youtube.com> Accessed February 12, 2008.

ZHAI, G., FOX, G.C., PIERCE, M., WU, W., BULUT, H. 2005. eSports: Collaborative and synchronous video annotation system in grid computing environment. In *Proceedings of the Seventh IEEE International Symposium on Multimedia (ISM '05)*. IEEE, Los Alamitos, CA.

APPENDIX A
USER STUDY SUPPORTING MATERIAL

Brief Volleyball Tutorial

Overview of the Game

Volleyball is a team sport played with a soft leather ball. A net separates the two teams from one another. The objective of the game is to get the ball over the net and into the opposing team's side (marked by lines on the gym floor). A team scores a point when ball lands on the opposing team's side or the opposing team cannot return the ball in three touches or contacts.

For a complete overview of the game of volleyball you may visit the volleyball wiki page at <http://en.wikipedia.org/wiki/Volleyball> However, for the study you will only need to know the following:

The SERVE (serving)

The serve is used to initiate the ball into play to the opposing team after a point has been awarded (i.e. the ball has hit the ground). It is very similar to a serve in tennis except the player contacts the ball with the palm of his hand instead of a racquet. The player serving the ball must stand behind his side of the court when serving.

Strategy

The player serving often wants to place the serve in a particular section or zone on the opposing team's side (in order to make the other team's job of sending the ball back over the net more difficult). A short serve would not travel very far after going over the net, falling in the front section of the opposing team's court. A deep serve on the other hand would land in the back section of the opposing team's court.

Video Example

Now watch the video entitled `serve_example.mp4` to see a demonstration of the serve

The SET (setting)

The set refers to the skill in which a player contacts the ball with both hands above his head and pops the ball up in the air. The purpose of the set is to redirect the ball to the team's hitters (spikers / attackers / offense). The player doing the setting is called a setter, and like the quarterback of a football team, is responsible for getting the ball to his offense.

Strategy

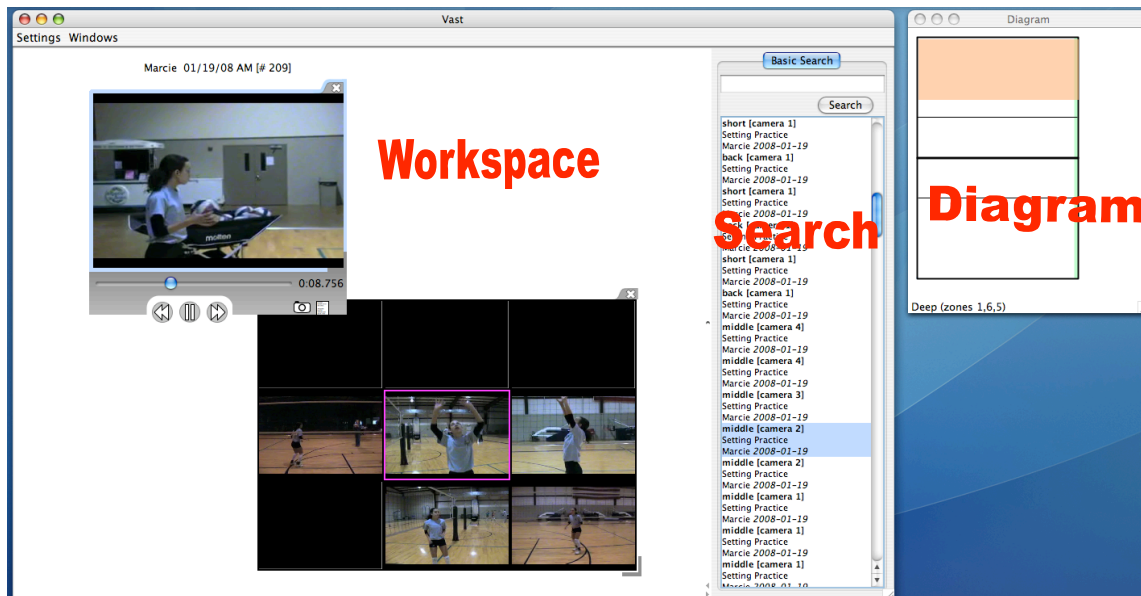
Three common sets are *outside*, *middle*, and *back*. Regardless of which set is performed, the setter (like a good quarterback) does not want to give away where he will be directing the ball. The more deceptive a setter is in making all his movements prior to contacting the ball look similar, the harder it is for the opposing team to set up their defense.

Video Example

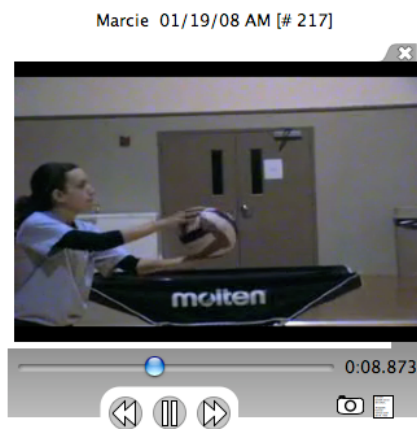
Now watch the video entitled `set_example.mp4` to see a demonstration of the set.




VAST Tutorial


1. The VAST interface is divided into three parts: workspace, search, and diagram areas.

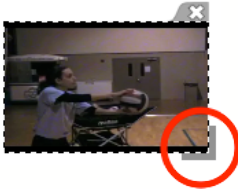


2. Click the search button in the search area. (This will display all available video clips)
3. In the search area select the second item in the list.
4. Drag this clip to the workspace.
5. Move your mouse so that it is hovering directly over the video. You should see something like this:

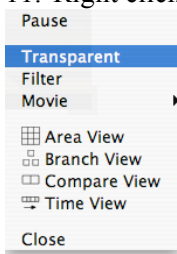


6. Click on the pause button  to stop the video.
7. Now click the frame forward button  to advance the video one frame.
8. Click on the play button  to start the video again.

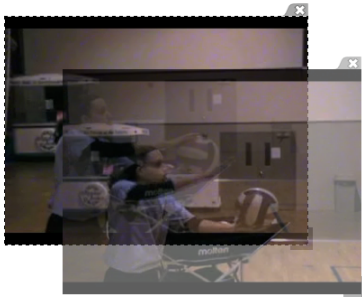
9. Now click on the picture button  to capture a still image from the video. Notice that the still image now appears in the workspace area.
10. Click on the resize button of the still image while dragging your mouse down to enlarge the picture to approximately the same size of the video.



11. Right click on video and select the *Transparent* menu option.



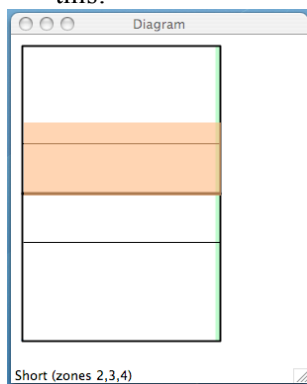
12. Now drag the video so that it now partially overlaps the still image.



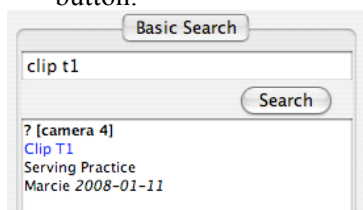
13. Right click on the video again and unselect the *transparent* menu option to make the video opaque again.
14. Close the still image by clicking on its close button




15. Hover over the video again and look at the diagram area. You should see something like this:



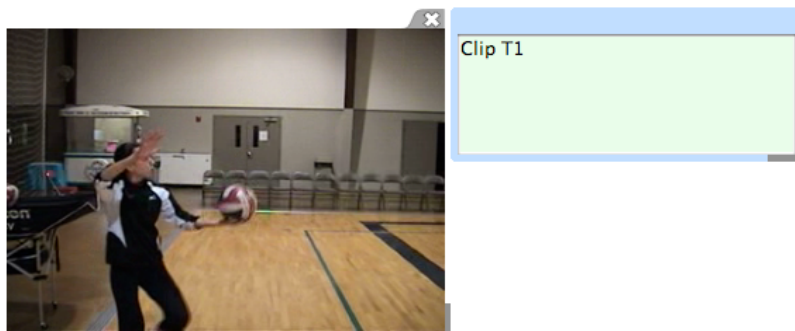
16. The black lines in the diagram represent the volleyball court. (The horizontal-middle black line that divides the rectangle into two squares represents the net. The top square is the opposing team's side, and the bottom square represents the home team's side.)
17. The orange highlighted area represents the outcome of the volleyball. (In this case, the serve landed in the front half of the opposing team's side.)
18. The thin green line represents the camera's location (In this video, the camera is located on the right side of the court.)
19. Now close the video clip.
20. Now find *Clip T1* by typing 'clip t1' in the search box and then clicking on the search button.



21. Now drag *Clip T1* into the workspace.

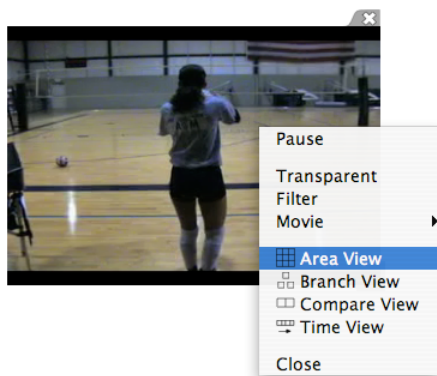
22. Hover the mouse over the video and then click on the note button .

23. Double click in the green note area to make the note editable.

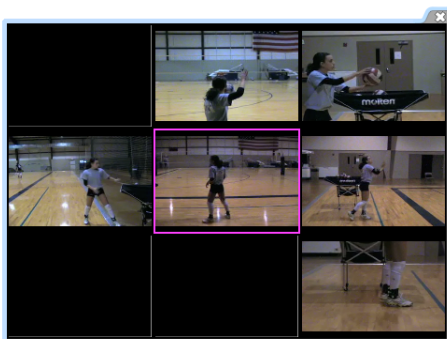


24. Change the note to the following "player focused on hand ball contact."

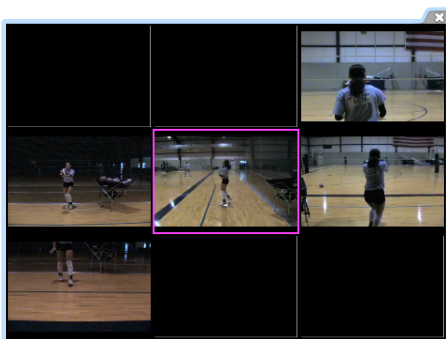
25. Hide the note by clicking on the note button again. (The comment is automatically saved, but will not show up in the search area until your next search)
26. Close the video clip.
27. Find all examples of deep serving by typing 'serving deep' in the search box and then clicking on the search button.
28. Select the last clip in the list and drag it to the workspace.
29. Right click on the video and select the *Area View*.



30. Click anywhere on the *Area View*. It should now be highlighted like below:

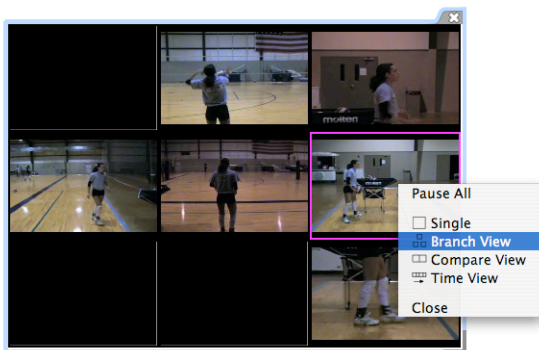


31. Now that the *Area View* is highlighted, press the left arrow key on the keyboard. You should now see the following:

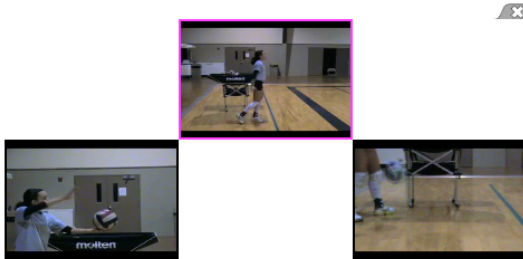


32. Use the right arrow key to move back to the original view.
33. Mouse over each of the videos and notice how the diagram information changes.

34. Now right click on the middle right video in the view and select *Branch View* from the menu.



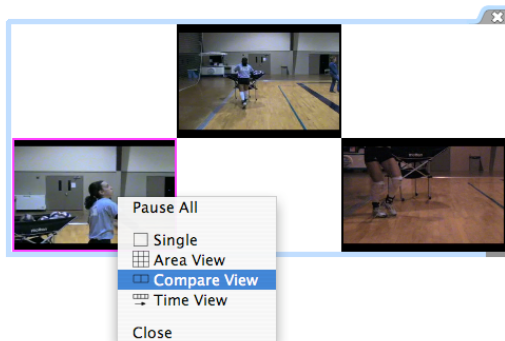
35. You should now see the following *Branch View*:



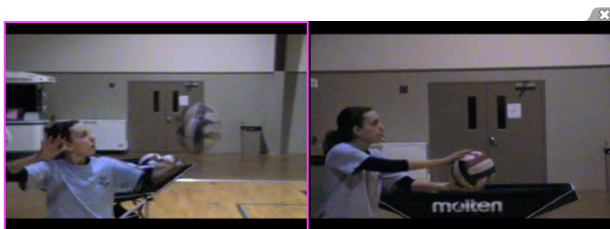
36. Double click on the top video to zoom in.

37. Click any other place in the branch view (besides the zoomed video) to zoom back out.

38. Right click on the bottom left video and select *Compare View* from the menu.

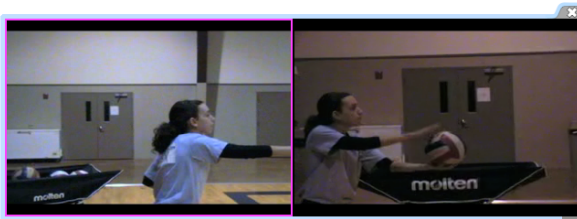


39. You should now see the following *Compare View*:

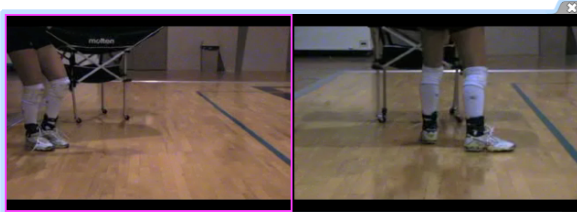


40. Mouse over each of the videos and notice how the diagram information changes.

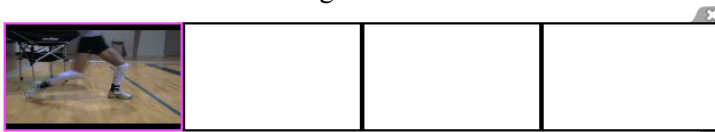
41. Click anywhere on the *Compare View*. It should now be highlighted in blue.



42. Press the down arrow key on the keyboard. You should now see the following:

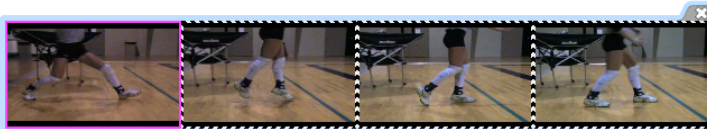


43. Now right click on the left video in the view and select *Time View* from the menu. You should see the following:



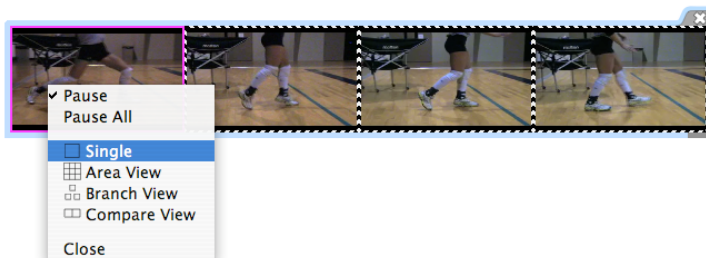
44. Click anywhere on the *Time View*. It should now be highlighted in blue.

45. Press the right arrow key on the keyboard. You should see the following:

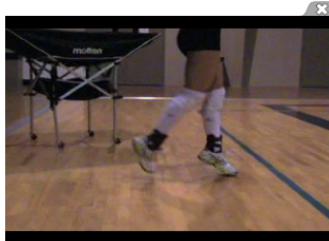


46. Press the right arrow key again to step another half second in time.

47. Right click on the far left video and select *Single* from the menu.



48. You should see the following:



49. Now close the video clip.

VAST Pre-Task Questionnaire

Alias: _____

Date: _____

Instructions:

- Please circle the letter/number that you think is the best answer for a given question.
- Please do not answer a question if it makes you uncomfortable, or you would not like to answer it for any other reason.

Personal Information

1) Gender

- a) Male
- b) Female

3) Nationality

5) Education Level

- c) High School
- d) Some Undergraduate
- e) Bachelors
- f) Some Graduate
- g) Masters
- h) Doctorate

2) Age Group

- i) 18-25
- j) 26-35
- k) 36-45
- l) 56-65

4) Educational Background

- m) Business
- n) Education
- o) Engineering
- p) Pure Sciences
- q) Social Sciences
- r) Other _____

General

6) Rank your general computer skill level:

Novice

Amateur

Intermediate

Advanced

Expert/
Master

1



3



5



7



9

7) How much time do you spend on the computer daily?

- a) 0-1 hours per day
- b) 2-4 hours per day
- c) 5-7 hours per day
- d) 8-10 hours per day
- e) 11 or more hours per day

8) How often do you watch sporting events live/in person?

- a) Daily (almost everyday)
- b) Weekly (2 to 3 times a week)
- c) Monthly (2 to 3 times a month)
- d) Less than once a month
- e) Never

- 9) How often do you watch sporting events on TV?
- Daily (almost everyday)
 - Weekly (2 to 3 times a week)
 - Monthly (2 to 3 times a month)
 - Less than once a month
 - Never

Volleyball

- 10) Rank your overall knowledge about volleyball:

Novice		Amateur		Intermediate		Advanced		Expert/ Master
↓		↓		↓		↓		↓
1	2	3	4	5	6	7	8	9

- 11) Have you ever PLAYED volleyball?
- Yes
 - No

If yes, please answer the following questions.

- 12) What is the highest level of volleyball you have played?

- Recreational
- High School (Non-Varsity)
- High School (Varsity)
- Club (under 15s)
- Club (15s-18s)
- Collegiate (Club)
- Collegiate (Varsity)
- International/Professional

- 13) How many years have you played volleyball?

- 0-2
- 3-5
- 6-8
- 9-11
- 12-14
- 15+

- 14) Circle your primary position:

OH RS MB S Libero DS

- 15) Circle your secondary position:

OH RS MB S Libero DS None

16) Have you ever COACHED volleyball?

- a) Yes
- b) No

If yes, please answer the following questions.

17) What is the highest level of volleyball you have coached?

- a) Recreational
- b) High School (Non-Varsity)
- c) High School (Varsity)
- d) Club (under 15s)
- e) Club (15s-18s)
- f) Collegiate (Club)
- g) Collegiate (Varsity)
- h) International/Professional

18) How many years have you coached volleyball?

- a) 0-2
- b) 3-5
- c) 6-8
- d) 9-11
- e) 12-14
- f) 15+

19) Do you ever WATCH volleyball film/video?

- a) Yes
- b) No

If yes, please answer the following questions.

20) How often do you watch volleyball film/video for the following purposes

	Daily (Almost every day)	Weekly (2 to 3 times a week)	Monthly (2 to 3 times a month)	Less than once a month	Never
Evaluating Performance					
Scouting					
Recruiting					
Fun/Enjoyment					

Rate the following sentences about VAST from strongly disagree to strongly agree.

Question	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
5) It is easy to use.					
6) It is fun to use.					
7) It is aesthetically pleasing.					
8) It is intuitive to use.					

9) Any other comments? Suggestions?

Alias: _____

Date: _____

VAST User Study

Instructions: Use the provide software and videos to answer the following questions.

- 2) Identify the differences between Marcie's outside and middle SET in the following areas:

Arms

Footwork

Overall

- 2) Identify the differences between Marcie's middle and back SET in the following areas:

Arms

Footwork

Overall

3) Identify the following clips as being either an outside, middle or back set:

Clip Number			
Q1	outside	middle	back
Q2	outside	middle	back
Q3	outside	middle	back
Q4	outside	middle	back
Q5	outside	middle	back
Q6	outside	middle	back
Q7	outside	middle	back
Q8	outside	middle	back
Q9	outside	middle	back
Q10	outside	middle	back
Q11	outside	middle	back
Q12	outside	middle	back
Q13	outside	middle	back
Q14	outside	middle	back
Q15	outside	middle	back

4) Two people were helping Marcie (i.e. tossing, catching the volleyballs) on day 1/19/2008. List the color shirt each person was wearing and their gender.

Person 1

Shirt Color:

Gender:

Person 2

Shirt Color:

Gender:

5) Is there a flag in the gym? If yes, how many and what kind?

(Formal User Study Round 1 – Task 4)

6) Identify the differences between Marcie's short and deep SERVE in the following areas:

Arm

Footwork

Overall

7) For stability while SERVING a volleyball player should have his/her hand, elbow, shoulder, and foot in-line at the moment he/she contacts the ball. Assume you only have one camera to record the player's serve. Where should you place the camera in order to best determine if the player's hand, elbow, shoulder, and foot are indeed in-line. (It does not have to be at one of the camera angles you have seen today).

Sketch a picture/diagram below indicating the player and where the camera should be placed in relationship to the player:

(Formal User Study Round 2 - Task 4)

6) For stability while SERVING a volleyball player should have his/her hand, elbow, shoulder, and foot in-line at the moment he/she contacts the ball. Assume you only have one camera to record the player's serve. Where should you place the camera in order to best determine if the player's hand, elbow, shoulder, and foot are indeed in-line. (It does not have to be at one of the camera angles you have seen today).

Sketch a picture/diagram below indicating the player and where the camera should be placed in relationship to the player:

APPENDIX B
USER STUDY DATA AND ANALYSIS

Task One: Identify the Difference between Outside and Middle Set –
Volleyball Novices

Gray background indicates no difference between sets was found.

An * indicates round one participants.

	Arms	Footwork	Overall
244CPT	It seems that there is more force put on the outside set.	There is more movement in the outside set . Marcie is hopping more.	Marcie is putting more effort on the outside set.
341SKI*	on outside set it seems that see to make contact with the ball out in front of her body where as in the middle set she makes contact more above her head. The outside set her arms go more out, and middle set arms go more straight up.	Take bigger steps going into the outside set, more knee bend with the outside set, even a small jump on outside set. With middle set, small step, little to no knee bend, never leaves the ground	seem more relaxed will doing middle set, eyes, do not follow the ball as long to see where it went. Seem like the outside is harder for her.
391MVX	outside - moves forward / with trajectory middle - moves forward / upward	outside - moves several steps within a radius middles - less momentum	outside - the setting motion is with speed/momentum middle - less momentum
455WPU	has more power when she hit the ball	footwork seems almost same	Her overall motion has more power and looks faster in outside than in middle
689SDG	Arms extend further away from the body for outside set. More force is applied to outside set.	more emphasis on plant/jump for outside set. Middle set seems more relaxed. Marcie positions herself to jump into the motion of the ball before outside sets.	Forward body motion for outside set is almost vertical body motion for middle set.
713AHU	Outside - arms out in front of body & extended, deep follow through Middle - arms up over head, shallow follow through	same basic steps, but outside set brings feet almost off the ground. Middle set just raises onto toes, but no jump	Middle is a shorter, more jerky set so not as much power is put in the ball. Outside set had more follow through & leg power to send it further.

Task One: Identify the Difference between Outside and Middle Set –
Volleyball Experts

Gray background indicates no difference between sets was found.

An * indicates round one participants.

	Arms	Footwork	Overall
221WKP	middle set has higher hands; outside deeper consistent with follow through	uses legs for more power on the outside set good footwork, consistent crouches more before setting the outside	It's pretty obvious that she is setting the outside, just because of the spring in the legs and arms before
264ZIV	finishes higher on middle set finishes out on outside set	feet are farther apart on outside set	takes the set in a very similar place on forehead. Body posture is fairly upright slightly forward with legs bent
391JQT*	outside set Marcie has full arm extension	There a small jump on the outside set, not present on the middle set	There is no much difference in technique overall
423TIX*	when setting outside Marcie extends her arms more & keeps the ball slightly more in front of her.	Her foot work seems to be very similar in both video clips	Marcie is using more of her body & extending through the set when setting the outside. When setting the middle she uses more power from her hands.
485UMH	OH - fully extends arms M - lightly sets	OH - much more footwork to get to the ball M - stationary both left then right footwork	OH - much more effort & push w/ legs M - less movement overall
725EDI*	Hands are more inclined to front in outside but in the middle less inclined.	I couldn't find any bit differences.	she is bending while setting ball in outside more compared to middle set.
836GFN	for outside set Marcie extends her arms outward with a hard push. For middle sets the push is much softer and the arms extend more upward.	The footwork is generally the same. Both have a left-right hop keeping the hips facing the target.	The outside set uses more bending of the knees so the set will push farther
977PDJ	outside set - arms move much quicker and further out middle - arms are slower and push up.	footwork still the same. Left, Right before making contact. Feet are a little more spread out on outside set.	outside - quicker harder set middle - calmer, slower overall same form & technique

Task One: Identify the Difference between Middle and Back Set –
Volleyball Novices

Gray background indicates no difference between sets was found.

An * indicates round one participants.

	Arms	Footwork	Overall
244CPT	Marcie's arms are slightly tilted towards the front in the middle SET. In the back SET she sends the ball completely upwards.	I couldn't find considerable differences	Apart from the difference on where the ball is set, the two sets are much alike.
341SKI*	ball contact made at the same point, right above the forehead, back serve hands go straight up, middle serve they go forward a little.	middle set one foot is in front of the other, with back set the two feet are right together	seems like the footwork and body are more similar with these two sets then the outside and middle. Here though with the back set she can not follow the ball with her eyes after the shoot.
391MVX	the orientation of the palm is different thus changing the velocity/direction of the ball	the footwork is very similar	both middle & back set are similar except the resulting direction of the ball
455WPU	Her arms are more straighten towards the sky (90 degrees) when she does back SET, and in the middle set her arms go towards front (around 60 degree)	footwork seems almost same	Her arms seems moving faster in the back SET than middle set. Overall movement looks similar.
689SDG	Whereas arms extend slightly forward for middle set. They extend directly up for back set.	Feet are slightly closer together for back set.	Legs are slightly less extended for back set. Back set has less overall body motion then middle sets, and is even more vertical.
713AHU	Both are w/ hands above head, but middle set is slightly in front. Back set is with arms straight up above shoulder to send ball behind.	Back set - feet closer together (toes flush w/ each other) & very little lift off the floor Middle - one foot in front of the other, slightly more lift in heels	Seem to be pretty similar. Back w/ slightly less power in legs & arms completely vertical.

Task One: Identify the Difference between Middle and Back Set –
Volleyball Experts

Gray background indicates no difference between sets was found.

An * indicates round one participants.

	Arms	Footwork	Overall
221WKP	pretty consistent in keeping hands high. Takes the ball back just a little to far on the backside	good footwork. Almost identical	Again, consistent. Not too many differences, only very subtle w/ the hands.
264ZIV	taking the set on a middle slightly in front of forehead on back set she is more underneath the ball	she is more prone to having her feet side by side on back sets	has a tendency to lean backwards and take the set from more on top of her than on her middle set
391JQT*	on back set arms are extended straight up. Middle set are leaning forward.	legs gets more extension on back set but no much difference	Marcie's middle set is more relaxed than back set.
423TIX*	In the middle she keeps her arms in front of her forehead after making contact. When back setting she flips her wrist back & takes the pass higher on her head.	Her footwork is almost the exact same motion for both front & back sets.	The only difference is Marcie's follow through with the set. She pushes the back set back behind her with her hands above her head whereas when setting middle she pushes the ball right off her forehead.
485UMH	M - arms pushed out in front of her B - arms pushed straight above head	M & B - good, still left then right footwork	M - easier for her B - knees more bent & back arched more
725EDI	No big difference I observed.	In back set more steady legs.	She is more steady and always looking back to check in back. But in middle, she is moving around.
836GFN	The back set has a stronger thrust of the arms and a directly verticle movement. The middle set is softer and the arms extend slightly outward.	The footwork is the same. The left-right hop right before the set.	In the back set, the hips are pushed forward
977PDJ	Arms go directly above her head, a little behind on back set.	Same, - left right - feet always facing outside	Arms go straight up to push ball backwards. Footwork / technique the same.

O = Outside Set

M = Middle Set

B = Back Set

Incorrect answers are highlighted in red.

An * indicates round one participants.

Q	None	Recreational					High School (Non-Varsity)	Club (under 15s)	Club(15s-18s)	Collegiate (Club)					
	Novices						Experts								
	455 WPU	244 CPT	391 MVX	689 SDG	713 AHU	341 SKI*	391 JQT*	725 EDI*	485 UMH	264 ZIV	836 GFN	221 WKP	977 PDJ	423 TIX*	
1	B	M	M	M	M	M	M	B	M	B	O	B	M	M	
2	B	B	O	B	B	B	B	B	B	B	B	B	B	B	
3	M	O	B	M	M	M	O	B	B	B	O	M	M	M	
4	M	M	O	M	M	O	M	M	O	M	O	O	O	O	
5	O	M	O	O	O	O	O	M	O	M	O	O	O	O	
6	O	B	M	M	M	B	B	B	M	B	M	M	M	B	
7	O	O	M	M	O	O	O	M	B	M	O	O	O	O	
8	M	M	B	B	M	M	M	M	B	B	M	B	M	M	
9	M	O	B	B	M	M	M	M	M	B	B	M	M	B	
10	M	B	M	B	B	B	B	M	B	B	B	B	B	B	
11	O	M	M	O	O	O	O	O	O	O	O	O	O	O	
12	M	B	M	M	M	M	M	M	M	M	M	M	B	M	
13	B	B	B	B	B	B	M	M	B	B	B	B	M	B	
14	O	M	O	O	O	O	O	O	O	O	O	O	O	O	
15	B	B	B	M	B	B	B	M	B	B	B	B	B	B	
	11	7	8	10	14	14	11	6	12	7	12	13	13	13	

Task Two: Identifying clips as examples of an outside, middle, or back set.

User rankings of VAST

	Easy	Fun	Aesthetic	Intuitive	Average
221WKP	1	2	1	1	1.25
244CPT	1	2	1	2	1.50
264ZIV	1	2	0	1	1.00
391MVX	1	1	1	1	1.00
455WPU	1	2	2	2	1.75
485UMH	2	2	1	2	1.75
689SDG	2	1	1	1	1.25
713AHU	0	1	2	0	0.75
836GFN	1	1	0	1	0.75
977PDJ	2	1	1	1	1.25
341SKI*	1	1	0	1	0.75
391JQT*	1	1	2	1	1.25
423TIX*	2	2	2	2	2.00
725EDI*	2	0	1	2	1.25
Average	1.29	1.36	1.07	1.29	1.25

Round 1 participants are indicated by an *

- (-2) Strongly Disagree
- (-1) Disagree
- (0) Neutral
- (+1) Agree
- (+2) Strongly Agree

User Free Response Feedback on VAST

1) What did you like most about VAST?	
221WKP	Very versatile, lots of angles and active videos
244CPT	Can have multiple videos at the same time running with multiple comparison options.
264ZIV	I like the transparency and being able to overlay videos as well as the multiple angles and the ability to go frame by frame
391MVX	the area view - it's cool & beneficial to be able to see an action from several different point of view
455WPU	1 can have different view type of video (specially area view)
485UMH	all of the different angles at once
689SDG	Intuitive position of different video angles in the Area View. Very easy to switch between different views.
713AHU	linking of camera angles in area view, easy to choose single clip & enlarge
836GFN	I liked the area view. It allowed me to focus on separate parts and the whole view. Very easy to navigate.
977PDJ	Easy to work, easy to search for specific videos.
341SKI*	Being able to have many video running at once, also the ability to take it one frame at a time. The area-view was also very good/useful. The Note was also, good, because it helped be keep track of which video was which.
391JQT*	The Area View and the compare functionality
423TIX*	I really like how all angles are able to be viewed! This can allow the viewer to see things that just one camera may have missed.
725EDI*	Comparison of videos is awesome. We can even compare picture of videos

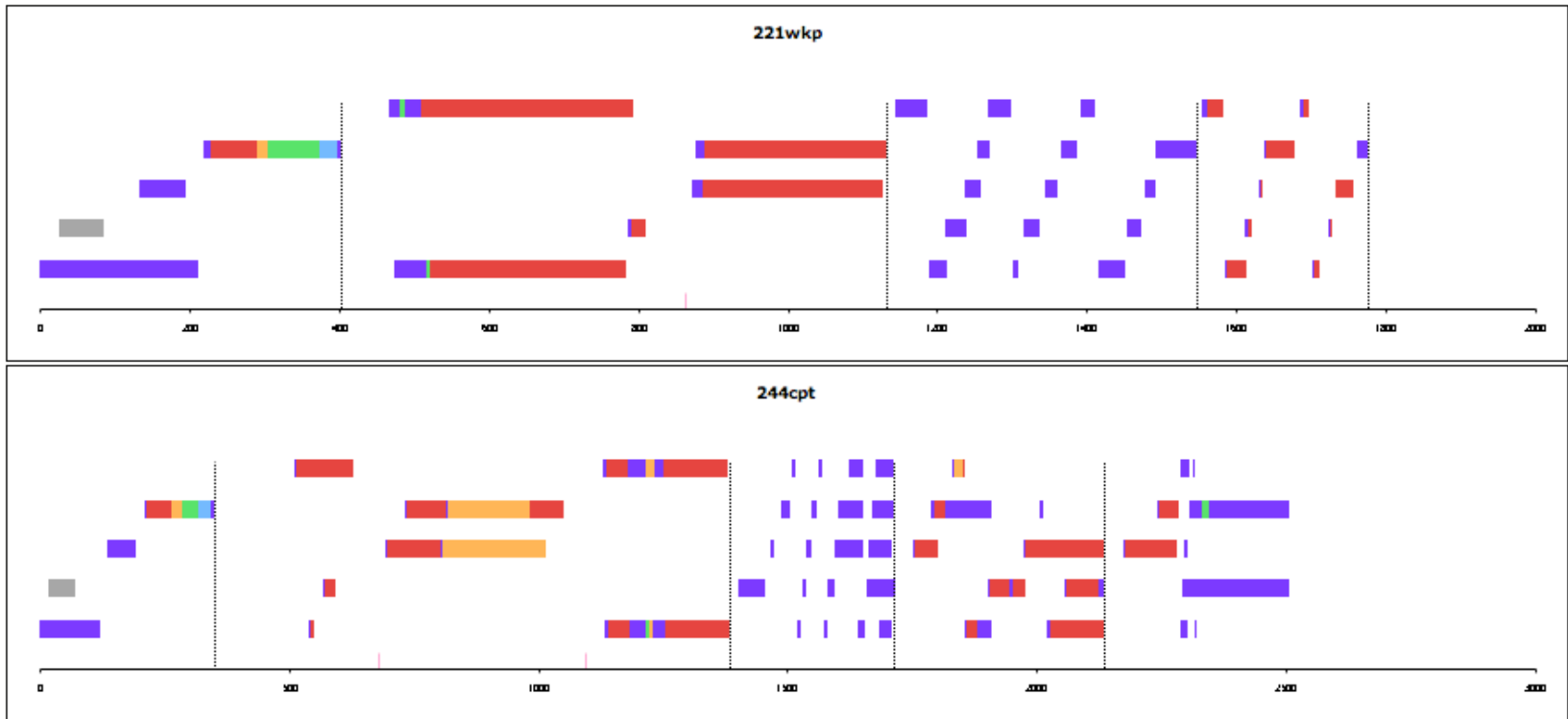
2) What do you dislike most about VAST?	
221WKP	a little disorganized with the labeling
244CPT	The black rectangles whenever there is no video are distracting. A lighter, less intrusive color would be better.
264ZIV	Having to manually restart or loop the Q section of setting videos
391MVX	none
455WPU	X
485UMH	you have to drag the picture over instead of double-clicking it
689SDG	The Time View samples are too far apart. I would prefer a frame-by-frame time view.
713AHU	I have no idea what branch & compare did. It was hard to resize clips.
836GFN	The videos were a bit small especially when in area view.
977PDJ	It is very simple, maybe I didn't see all of it but more could be added to it. Didn't tell camera angles so you had to click on all of them.
341SKI*	I wish the video would have be sight larger as a default, found myself increasing the size a lot.
391JQT*	The focus (light blue) highlight at the bigginig, when I started to use program. I could not navigate in the Area View because I did not realize that I lost focus of the window. It was not a problem later on.
423TIX*	I found some parts a little tricky to use (ex. time view was a little challenging to get to work.)
725EDI*	May need good documentation

3) How can VAST be improved?	
221WKP	It could have more heierarchy w/ videos (a serving folder w/ sub branches; a setting folder)
244CPT	In the search it would be nicer to have not the individual clips but rather a tree like classification. Back set -> clip 1 Middle set -> clip 2
264ZIV	having an option for video loop or selectable button to start from the beginning again
391MVX	- add the capability to add note to videos when in the compare mode (or any other mode in addition to the single mode) - add the metadat for characters / people in each video (I.e. allowing a search for videos with a coach)
455WPU	not quite. When I open two video with an area view the system seems having hard time.
485UMH	a better search engine -> more words in the title I guess to allow that
689SDG	Run it on something with a larger screen. Enable scrubbing on the video timeline.
713AHU	Context clues (cursor change) when over clip resize region so I don't end up moving the clip. It's hard to guess where the resize hot spot is.
836GFN	I think a back view should be show to show the alignment and also the ball crossing the net.
977PDJ	If it is used to teach - vocal explanation of skills would help.
341SKI*	Well it did crash once well I ran it, so that could be improved.
391JQT*	1) Hints for common error, detecting if user have navigation errors. 2) Double click on video in search Result places video in workspace.
423TIX*	If you can freeze frame and view that exact frame from all the angles at once. (It may already be able to do this. I didn't know)
725EDI*	It would be better if ther is functionality to compare videos frame by frame and also by moving with 2x, 4x speeds both at the same time

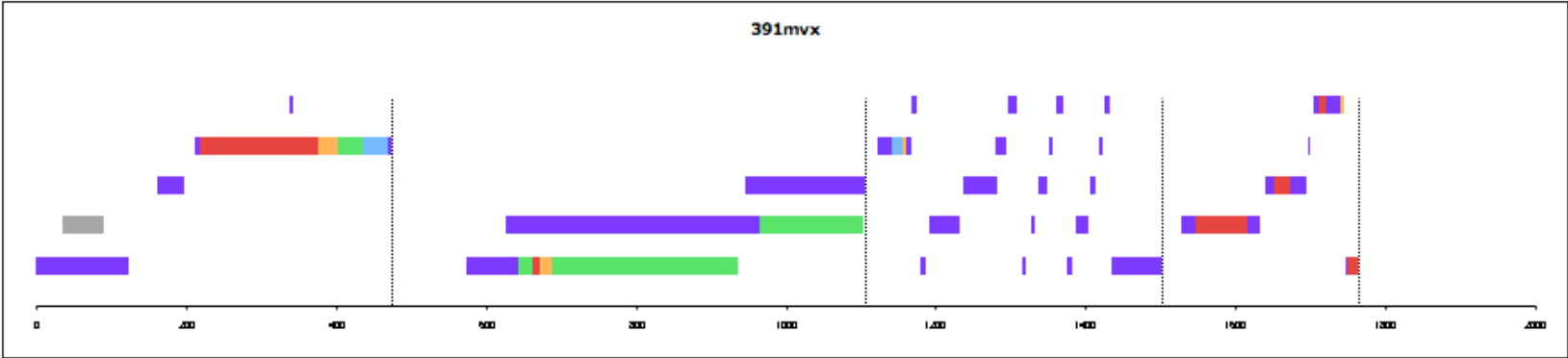
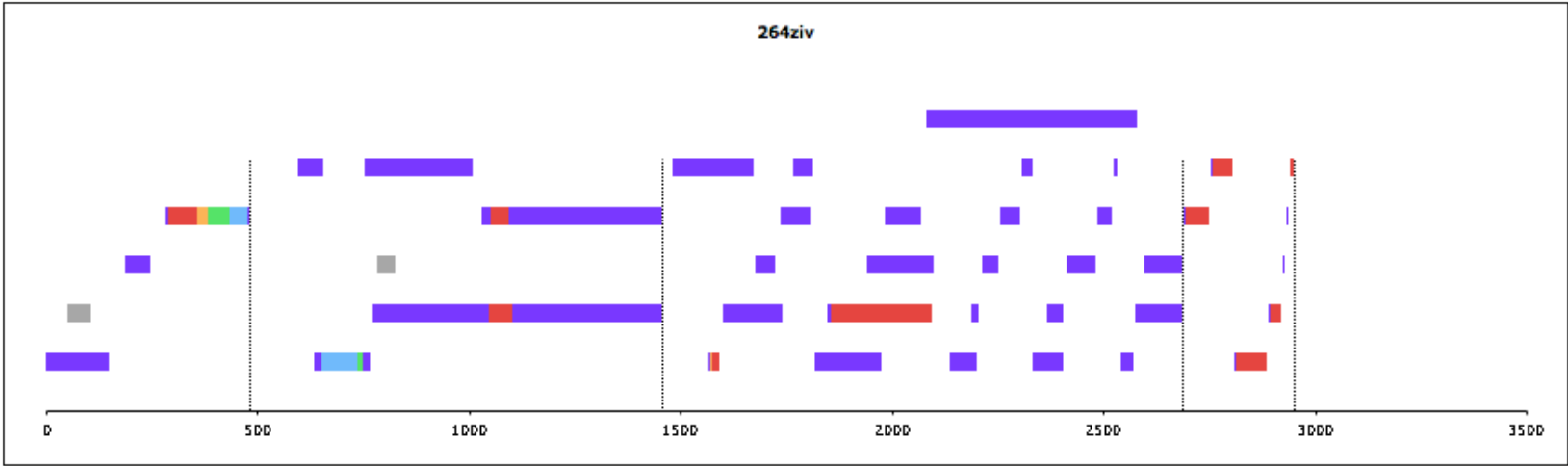
4) If VAST was a software program on your computer that worked with videos of your choosing (not necessarily volleyball)...		
	<i>Would you use it (Why/why not?)</i>	<i>What type of tasks would you use it for?</i>
221WKP	Maybe. Only if I had a bunch of angles of a shot. Otherwise it loses a bit of its effectiveness. Seems like it would be good for recording plays or something similar	Evaluating skills. Probably would not use it for matches but maybe, depending on the angles.
244CPT	I would probably use it, since it has an intuitive interface which is actually easy to use.	Coveoving surveillance videos.
264ZIV	Yes, I think being able to overlay transparent videos can be used to highlight differences for a wide variety of sports/functions	Sports analysis any type of comparison between similar video sources to highlight differences
391MVX	Yes!!! I would use it given there are video collections that I can download (or stream online). I don't think I'll use it with video from my own camera (I won't have multiple view points)(if I do, that'll be great!)	- Watching a movie following multiple characters - watching live sports cast - family reunion videos looking for specific things in video (ie evidence for something...crime scene)
455WPU	Yes, it would be useful for me to improve my skills in some sports, specially golf for me.	I will record my movements while I swing when I play golf to improve skills.
485UMH	no, because its not relevant to anything but sports	none
689SDG	Absolutely. It would be great for analyzing motion of people/animals for animation purposes	animation, studying motion
713AHU	Probably not. I don't have a need to see multiple angles of the same video.	If I did, however, I might link computer game gameplay footage to get a "big picture" of what happened. I would be neat to combine multiple player's POVs
836GFN	Yes, I would use it.	Many classes require presentations and I could use videos to show graphics instead of just pictures.
977PDJ	Yes. It is easy to search, compare, take snapshots.	I am considering coaching so it would be great for that!

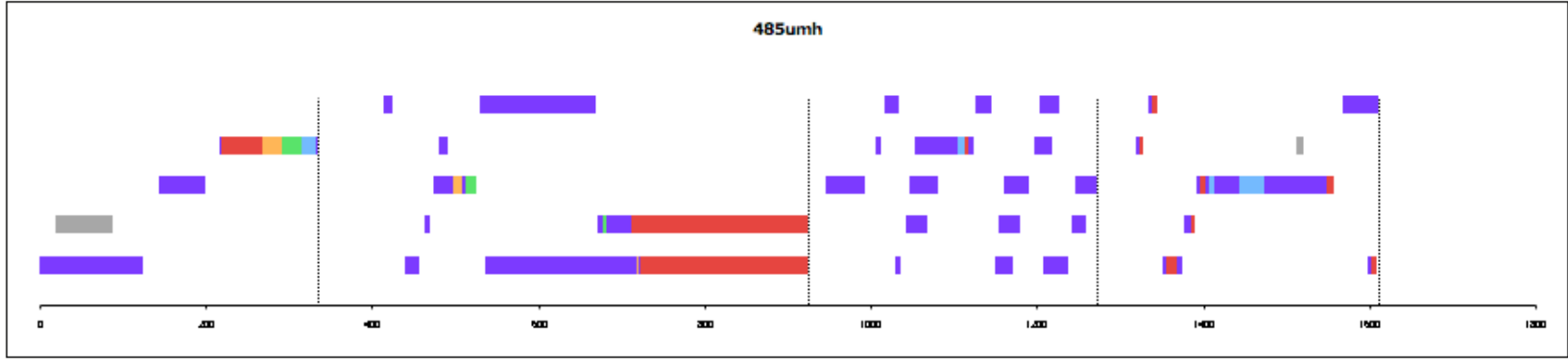
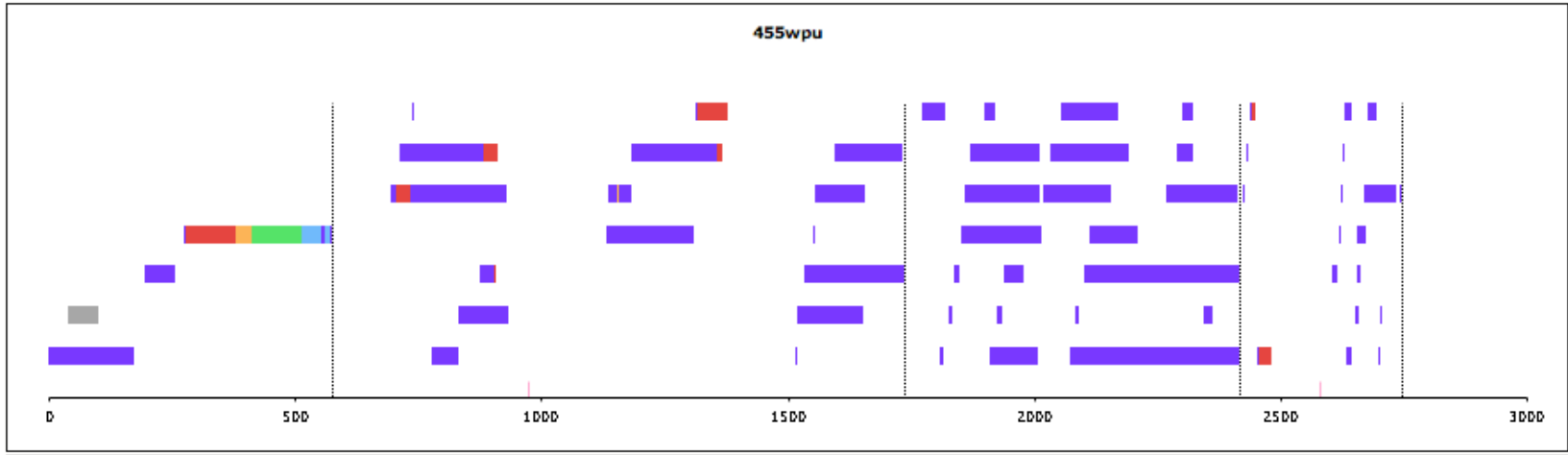
	<i>Would you use it (Why/why not?)</i>	<i>What type of tasks would you use it for?</i>
341SKI*	I do not have any videos to put on it or use it with. So probably not right now.	I think that it would/could be a good coach tool for any sport. It seems to work best for task that would repeat, and then be able to compare these task.
391JQT*	Yes, I would use it, if I am doing video editing with simultaneous recording. But I think it is more suitable for training, coaching, and sport scouting.	Another task I would use this tool is for data visualization, (multiple videos of multiple representations/dimensions)
423TIX*	Yes & for volleyball! It gives a way of viewing things that can heighten the visual experience greatly! Especially as a training tool for athletes.	Sports for sure. Maybe like educational things like "how to" videos such as how to build something
725EDI*	I find use it for picture snapshots. Next can be used to observe patterns in two diff related videos.	Two find similarity or coherence b/w two dif videos

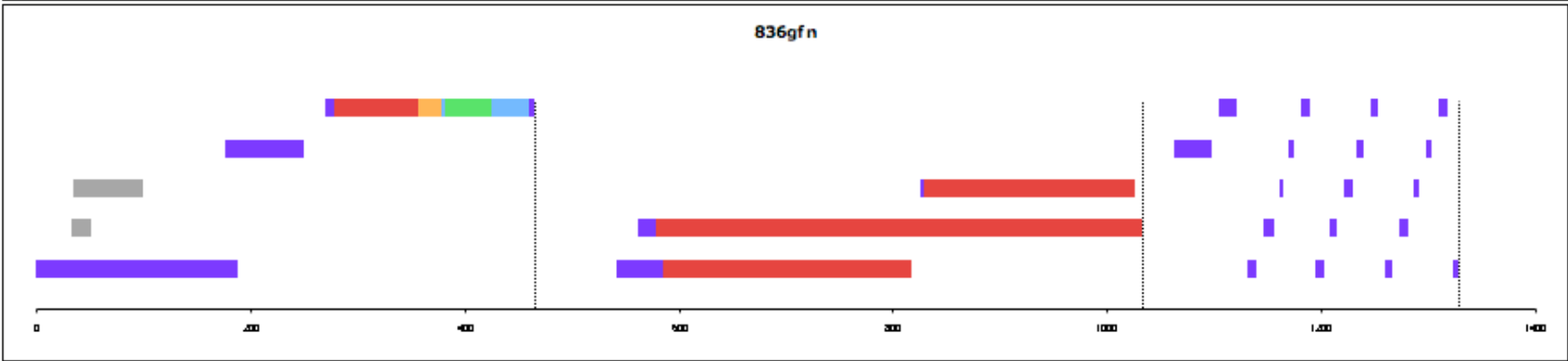
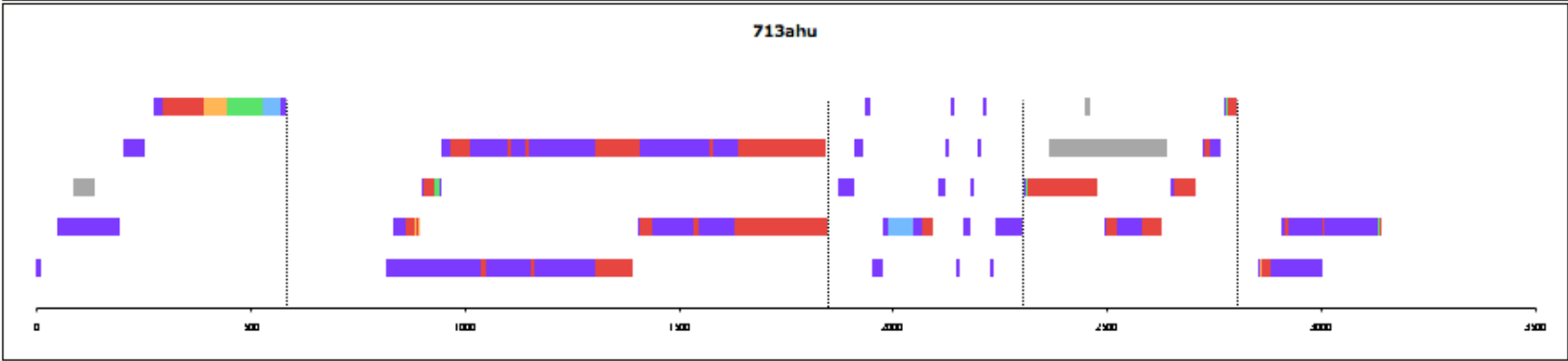
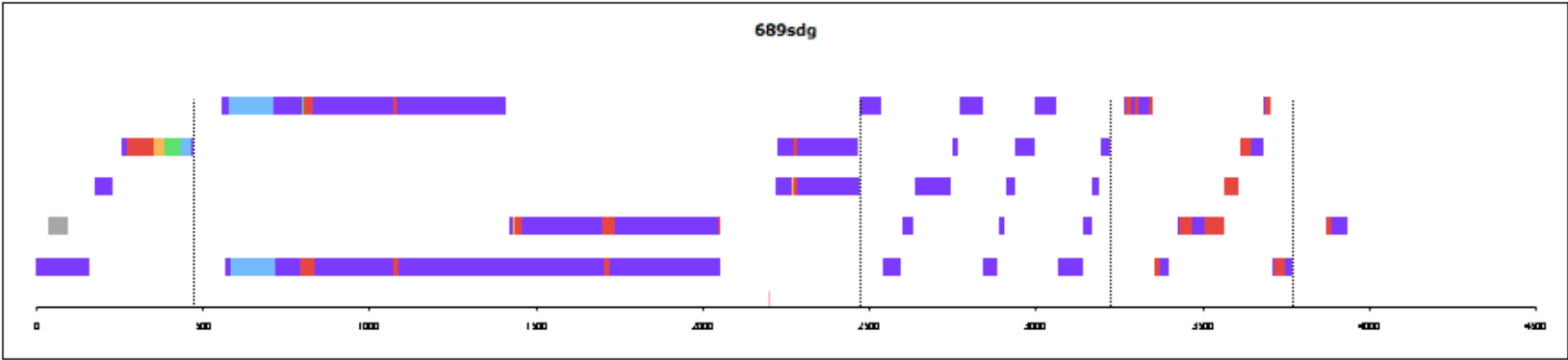
9) Any other comments? Suggestions	
221WKP	Other than organization, I think it is pretty great.
244CPT	<ul style="list-style-type: none"> - The exact position of the camera would be nice to be shown - hierarchical view in the search box of the clips - Ability from a single video to choose other camera - When inserting a video, it should have focus - If on inserting a video, it overlaps with an existing one the previous one should be made transparent - Ability to rewind The video - Preview of videos before inserting them
264ZIV	I enjoyed using the program and would love to use it more if I had the chance. I would really like to have some of the coaching staff at A&M participate because I think they would really like it too!
391MVX	X
455WPU	It was nice experience to interact with video clips. Maybe tagging an object inside the video clip would be also helpful.
485UMH	Marcie is a good player
689SDG	more than anything, larger screen or dual monitor would make it easier to compare things.
713AHU	Branch & compare were not intuitive to me. Don't know what they did. Very slick & clean interface. Easy to get what I wanted with Area/Single views. Ease of use - hard to resize videos w/o some context clues.
836GFN	I think this software would be great to teach the fundamentals of volleyball.
977PDJ	X
341SKI*	No.
391JQT*	good luck!
423TIX*	Great program!!
725EDI*	X

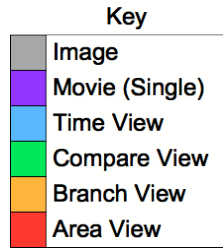
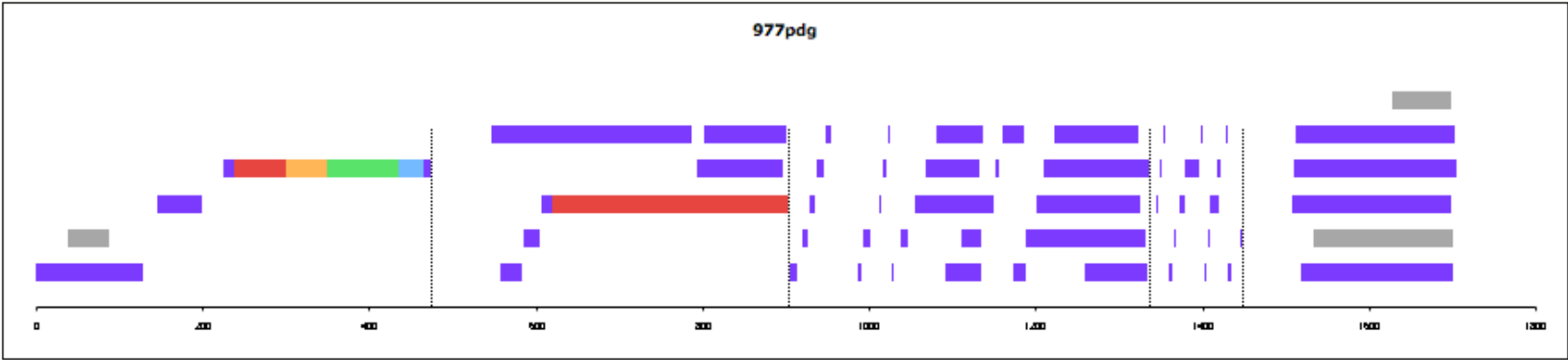


User log visualization for tutorial and all tasks – formal round two.









User log statistics for tutorial and all tasks – formal round two.

Round 2	221 wkp	244 cpt	264 ziv	391 mvx	455 wpu	485 umh	689 sdg	713 ahu	836 gfn	977 pdg	Ave.	Std. Dev.
Distinct Objects												
Movies	33	51	38	30	67	34	31	33	24	55	39.6	13.533
Images	1	1	2	1	1	2	1	4	2	3	1.8	1.033
<i>Total</i>	<i>34</i>	<i>52</i>	<i>40</i>	<i>31</i>	<i>68</i>	<i>36</i>	<i>32</i>	<i>37</i>	<i>26</i>	<i>58</i>	<i>41.4</i>	<i>13.476</i>
Search												
Search	27	20	37	31	33	30	28	34	24	32	29.6	5.016
Conversions												
To Area	15	23	10	6	9	11	25	27	4	2	13.2	8.967
To Branch	1	7	2	4	2	4	3	4	1	1	2.9	1.912
To Compare	3	3	2	6	1	3	2	6	1	1	2.8	1.874
To Time	1	2	3	2	2	4	3	2	2	1	2.2	0.919
To Single (Movie)	2	12	5	5	4	9	20	20	1	1	7.9	7.249
<i>Total</i>	<i>22</i>	<i>47</i>	<i>22</i>	<i>23</i>	<i>18</i>	<i>31</i>	<i>53</i>	<i>59</i>	<i>9</i>	<i>6</i>	<i>29.0</i>	<i>18.221</i>
Navigation												
Area	5	68	2	25	10	2	81	82	2	2	27.9	34.786
Branch	0	57	0	0	0	0	0	0	0	0	5.7	18.025
Compare	15	1	1	16	1	1	9	1	1	1	4.7	6.219
Time	2	9	18	6	3	5	4	14	7	15	8.3	5.539
<i>Total</i>	<i>22</i>	<i>135</i>	<i>21</i>	<i>47</i>	<i>14</i>	<i>8</i>	<i>94</i>	<i>97</i>	<i>10</i>	<i>18</i>	<i>46.6</i>	<i>45.427</i>
Navigation Blocked	0	5	0	5	0	0	2	6	0	0	1.8	2.530
Zoom												
Zoom In	21	2	11	6	2	2	1	17	14	3	7.9	7.310
Zoom Out	18	2	11	6	1	1	1	16	14	2	7.2	6.877

Round 2	221 wkp	244 cpt	264 ziv	391 mvx	455 wpu	485 umh	689 sdg	713 ahu	836 gfn	977 pdg	Ave.	Std. Dev.
Movie Manipulation												
<i>Time Related</i>												
Pause	16	10	12	10	10	20	55	47	2	5	18.7	17.845
Play	8	6	6	7	7	20	43	26	2	5	13.0	12.901
Pause All (Group)	0	6	0	0	0	0	0	1	0	0	0.7	1.889
Play All (Group)	0	2	0	0	0	0	0	1	0	0	0.3	0.675
Adjust Timebar	73	11	43	21	19	0	139	23	4	7	34.0	42.760
Previous Frame	12	2	36	36	1	0	443	182	14	4	73.0	141.042
Next Frame	85	0	233	58	2	18	1185	163	5	4	175.3	363.431
<i>Misc.</i>												
Show Note	1	3	1	6	9	1	1	2	1	1	2.6	2.757
Object Manipulation												
Resize	3	4	13	7	1	1	39	18	1	4	9.1	11.921
Transparent On	2	2	8	1	1	1	1	1	2	1	2.0	2.160
Transparent Off	2	2	1	1	1	1	1	1	2	1	1.3	0.483
Session												
Program Restarts	1	2	0	0	2	0	1	0	0	0	0.6	0.843

APPENDIX C
SOFTWARE DEMO

Software Demo: Separate video file containing a demo of the VAST software.

VITA

Name: Marlo Faye Nordt

Address: Dept. of Computer Science
Texas A&M University
TAMU 3112
College Station, TX 77843-3112

Email Address: mnordt@tamu.edu

Education: B.S., Computer Science, Stephen F. Austin State University, 2003
Ph.D., Computer Science, Texas A&M University, 2008