Video-as-data and digital video manipulation techniques for transforming learning sciences research, education and other cultural practices[*]

Roy D. Pea
Stanford University

"Only in the stream of thought and life do words have meaning."

Ludwig Wittgenstein (1981) *Zettel, (2nd. Ed.)*
G.E.M. Anscombe and G.H.V. Wright (Eds.). Oxford: Blackwell. (No. 173)

Abbreviated title for running headline: *Video-as-data*

(To appear in J. Weiss, J. Nolan & P. Trifonas (Eds.), *International handbook of virtual learning environments*. Dordrecht: Kluwer Academic Publishing, 2005.
Please do not quote without permission.)

Stanford Center for Innovations in Learning (SCIL)
Wallenberg Hall, Building 160
Stanford, CA 94305
Phone: 650.724.3720
Fax: 650.724.0432
Email: roypea@stanford.edu

---

[*] DIVER™, WebDIVER™, Dive™ and "Guided Noticing"™ are trademarks of Stanford University for DIVER software and affiliated services with patents pending. The DIVER project work has been supported by grants from the National Science Foundation (#0216334, #0234456, #0326497, #0354453) and the Hewlett Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of these funders. Roy Pea served as lead author with contributing authors Ken Dauber, Michael Mills, Joseph Rosen, Wolfgang Effelsberg and Eric Hoffert for portions of earlier versions of some of the text material provided here regarding DIVER. Ray Pacheone, Randall Souviney and Peter Youngs have been instrumental in our thinking of DIVER in the teacher preparation and certification context. Brian MacWhinney was an additional key collaborator in conceptualizing the Digital Video Collaboratory characterized in Section 8.0.

1. Introduction

This chapter concerns the theoretical and empirical foundations and current progress of the DIVER Project at Stanford University. The DIVER Project—Digital Interactive Video Exploration and Reflection—aspires to accelerate cultural appropriation of video as a fluid expressive medium for generating, sharing and critiquing different perspectives on the same richly recorded events and to work with others to establish a Digital Video Collaboratory that enables cumulative knowledge building from video-as-data for discovery and commentary. These uses of digital video manipulation are very distinctive from those used in virtual learning environments today across K-12, higher education and corporate training (e.g., BlackBoard, WebCT, PlaceWare), which are primarily video clips that are used to illustrate a point or concept during a lecture or a video of a faculty member teaching and using PowerPoint slides.

The DIVER system distinctively enables "point of view" authoring of video tours of archival materials (from video to animations and static imagery) in a manner that supports sharing, collaboration, and knowledge building around a common ground of reference. A fundamental goal is user-driven content re-use, prompted by the desire of content users to reinterpret content in new ways, and to communicate and share their interpretations with others, for purposes ranging from e-learning to entertainment.

DIVER makes it possible to easily create an infinite variety of new digital video clips from a video record. A user of DIVER software "dives" into a video record by input controlling—with a mouse, joystick, or other input device—a virtual camera that can zoom and pan through space and time within an overview window of the source video. The virtual camera dynamically crops still image clips, or records multi-frame video "pathways" through normal consumer 4:3 aspect ratio video or a range of parameters (e.g., 20:3) for video records captured with a panoramic camera[1], to create a dive™ (a DIVER worksheet). A dive is made up of a collection of re-orderable "panels", each of which contains a small key video frame (often called a "thumbnail") representing a clip as well as a text field that may contain an accompanying annotation, code, or other interpretation.

After creating a dive using the desktop DIVER application, the user can upload it onto WebDIVER™, a Website for interactive browsing, searching, and display of video clips and collaborative commentary on dives. In an alternative implementation, one can dive on streaming video files that are made accessible through a web server over the Internet, without either requiring the downloading of a DIVER desktop application or the media files upon which the user dives.

In the desktop DIVER implementation, the dive itself is packaged automatically as an XML document with associated media files. XML (Extensible Markup Language) is the universal language approved in 1998 by the World Wide Web Consortium (W3C) for structured documents and data on the web. The Web representation of the dive includes

[1] A panoramic video camera is created when one or more video cameras are combined with one or more mirrors to capture 360-degree horizontal data around the panoramic camera's fixed-point location. Image processing software is then used to "dewarp" the imagery that is produced from the reflective surface and to create interpretable displays on a computer monitor of the captured video data. A common method for depicting the panoramic video record is a wide-aspect ratio window that looks as if the cylinder has been sliced and "peeled-back" (see Figure 2 for illustration).

the key frames, video clips and annotations. A dive can be shared with colleagues over the Internet and become the focus of knowledge building, argumentative, tutorial, assessment or general communicative exchanges. A schematic representation of the recording, "diving" and Web-sharing phases is shown in *Figure 1*:

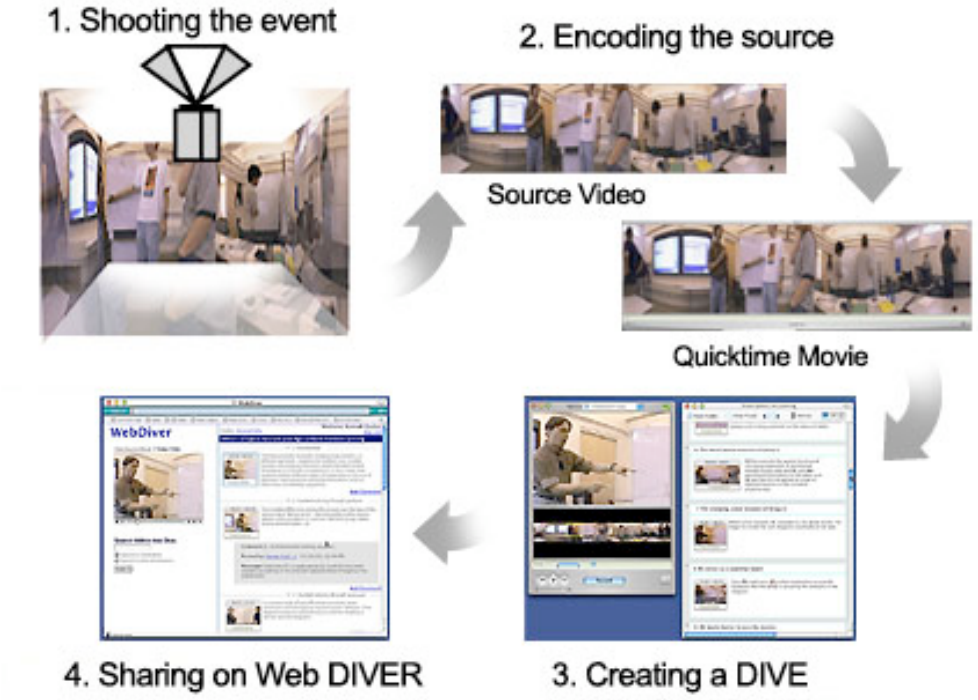

Figure 1. Overview of the DIVER video exploration and reflection system

Much of our primary DIVER work involves scenarios in which a video camera is used to record complex human interactions such as the behavior of learners and teachers in a classroom, or research group meetings. One may capture video directly into DIVER using DIVER's MovieMaker feature with a digital videocamera connected by FireWire to the computer, or use DIVER's MovieImporter feature to bring in as source material previously-captured video. A 360-degree camera may also be used to capture panoramic video of the scene—but before discussing the rationale for panoramic video recording, it is important to establish why it is that video-as-data and digital video manipulation techniques such as those that DIVER represents are critical for transforming learning sciences research and educational practices.

2. Power of video as data for human interaction analysis and reflective uses for purposes of learning, training and education

Situated learning. In recent years, there has been increasing recognition that human activities such as learning need to be understood in the context of their naturalistic situations and socio-cultural environments (e.g., Brown et al., 1989; Brown, 1992; Lave, 1993; Pea, 1993; Resnick, 1987). Over this period, the learning sciences have shifted from a view of learning as principally an internal cognitive process, towards a view of learning as "mind in context"—a complex social phenomenon involving multiple agents interacting in social and organizational systems, with one another, and with symbolic representations and environmental features (Bransford, Brown, & Cocking, 2000; Greeno & MMAP, 1998; Hutchins, 1995; Pea, 1993).

This orientation to "learning environments"—the concrete social and physical settings in which learning takes place—has led researchers to utilize tools that allow capture of the complexity of such situations, where multiple simultaneous 'channels' of interaction are potentially relevant to achieving a full understanding of learning or other human interactional phenomena. For at least eighty years, researchers have experimented with using multimedia records—first film, then analog videotape and now digital audio and video—to gather richer and more reliable data about complex social interaction than is possible with traditional alternatives like field notes, participant recollections, and transcripts of audiorecordings. These technologies are attractive for their potential in making a relatively complete record for re-use and for sharing among multiple researchers, without the inconvenience and intersubjectivity problems of recording methods like field notes. But capturing and working with video records, until the rapid technological developments of the last two decades, required access to equipment and skills that relatively few researchers sought to acquire. The consumer revolution in video technology has lowered this barrier in making sophisticated, easy to operate, cost-effective recording tools broadly available.

Uses of such audio and videorecords have been instrumental in theoretical developments by researchers contributing to this "situated shift" in studies of learning, thinking and human practices. This development was deeply influenced by conversation analysis (Goodwin & Heritage, 1990), sociolinguistic studies of classroom discourse (Mehan, 1978), by ethnographic and anthropological inquiries of learning in formal and informal settings (Erickson, 1992; Lave, 1988; Rogoff, 1990; Rogoff & Lave, 1984; Saxe, 1988, 1991) and ethnomethodological investigations (Garfinkel, 1967) as well as studies of nonverbal behavior such as gaze (Argyle & Cook, 1976), body orientation/kinesics (Birdwhistell, 1970) and gesture (Kendon, 1982). These foundations are evident in Jordan and Henderson's (1995) influential paper on interactive video analysis—in research labs throughout departments of psychology, sociology, linguistics, communication, (cultural) anthropology, and human-computer interaction, researchers work individually or in small collaborative teams—often across disciplines—for the distinctive insights that can be brought to interpretation during the analysis of video recordings of human activities.

Thanks to these socio-technical developments, interdisciplinary studies utilizing video have deepened our understanding in many learning science sub-fields such as mathematics thinking, learning and teaching (Greeno & MMAP, 1998; Lampert & Loewenberg-Ball, 1998; Schoenfeld, 1992), functions of teacher activities (Frederiksen et al., 1998), international comparative studies of videos of mathematics classrooms (Stigler et al., 1999, 2000; Stigler & Hiebert, 1999); learning of demanding topics in high school physics (Pea, 1992; Roth & Roychoudhury, 1993), informal learning in science museums (Crowley et al., 2001; Stevens & Hall, 1997), interacting with machines such as copiers, computers, and medical devices, suggesting new design needs (Nardi, 1996; Suchman, 1987, Suchman & Trigg, 1991; Tang, 1991), collaborative learning (Barron, 2000, 2003), and of specific roles for gestural communication in teaching and learning (Roth, 2001). The pervasive impact of video studies was in evidence at the 2002 American Educational Research Association meetings, which included 44 scientific panels and symposia using video for learning research, teaching, and teacher education.

The availability of such inexpensive videography equipment and promise of more

complete records of complex phenomena than earlier methods has led many researchers to adopt video recording as a primary data collection method. Yet there is a serious and persistent gap between such promise and the usefulness of video records. Video data circulates rarely and slowly within scientific research communities, even as more researchers use the medium. Video research analyses are typically restricted to text-only displays for presenting findings; original data sources are not made available for reanalysis by other researchers; and it is typically impossible for collaborators working at different sites to conduct joint analysis of shared video records. Several workshops have documented the needs for much better tools for powerful video capture and analysis technologies to support studies of learning and instruction (Lampert & Hawkins, 1998; MacWhinney & Snow, 1999; Pea & Hay, 2003). This gap between the promise and reality of digital video yields continued problems by obscuring the connection between evidence and argument, discourages sharing of video data, and impeding the development of shared examples of exemplary analyses using video that could serve training and socialization functions for novice researchers.

E-Learning. Furthermore, over the last several years, there have been a number of projects in research and industry whose goal is to advance the use of video and computer technologies to study the complex interactions underlying classroom learning. For example, the *Classroom 2000* project (Abowd, 1999) developed a system for recording many aspects of a live classroom experience—including the capture of strokes on a whiteboard—and making these recordings subsequently available to students. *Teachscape* (www. teachscape.com) and *LessonLab* (www. lessonlab.com) are commercially available products whose web-based platforms allow teachers and other professionals to study and discuss videos and other artifacts of classroom practice, as face-to-face resources or over the Internet. There are many teacher education projects that utilize some form of digital video records to provide case-based learning approaches for learning at different points in the teacher professional development continuum, including Indiana University's Internet Technology Forum (Barab et al., 2001, 2002, 2003), University of Michigan's KNOW system (Fishman, in press), San Diego State University's Case Creator (Doerr et al., 2003), and Northwestern University's VAST tool (Sherin, in press).

In a related vein, Microsoft's MRAS (Microsoft Research Annotation System) has put forward a general architecture for multimedia annotations focusing on use in higher education (Bargeron, et al., 2002). There is also a considerable body of work on video annotation that has focused on indexing and browsing video databases (e.g., Carrer, et al., 1997; Kim et al., 1996; Lee & Kao, 1993; Mills, et al., 1992; Weber & Poon, 1994). Nonetheless, as described below, many of same problems beset this work as they do research on learning and the social sciences using videorecords.

Business meeting capture and re-use. A number of projects over the past five years (e.g., Chiu et al., 2001; Fiala et al., 2004; Lee et al., 2002; Myers et al., 2001; Stiefelhagen et al., 2002; Yang et al., 1999; Young et al., 2001) have explored how to provide business and technology groups with the capability of returning to context captured from rich media records of their meetings, from automatic indexing and other methods. The general idea is that one can use video to enable persistent context for ongoing teamwork, building in a cumulative way on prior meetings and design decisions, and continuing to excavate information that may come to have subsequent value.

For application to video conferencing meetings and distance education, Sun et al. (2001; also see Foote & Kimber, 2001) have used their FlyCam system (which produces high resolution and wide-angle video sequences by stitching together video frames from multiple stationary cameras directed at the front of a seminar room) in combination with a Kalman filter to detect the speaker's motion and then use that information to steer a virtual camera for recording a region of interest within the panoramic videostream.

Although not directly concerned with education, at Microsoft Research, Rui et al. (2000) have investigated the use of panoramic recording system to allow people to "re-experience" face-to-face meetings. Interestingly, they found that users prefer having a panoramic overview – peeled-back video imagery showing 360° panorama—as a mechanism for navigating the scene and selecting detailed views.

Reflections on the need. In summary, there is substantial need, for research in the learning sciences, for e-learning purposes, and for facilitating collaborative work both in face-to-face teams and at a distance, for new methods that foster capturing, navigating, analyzing, re-purposing, and commenting on video as a medium for representing situated practices. Researchers today lack tools for sharing these video data readily with other scholars and practitioners, for building cumulative analyses of research data across investigators, and for opening up these data for public discussion and commentary.

Prospects of Digital Video Collaboratories. As in other scholarly domains of inquiry, there are fertile opportunities in education and in the learning sciences for developing new methods of work, knowledge creation, and learning that leverage the collective intelligence of the field in ways analogous to the biological, health, earth and space sciences (e.g., Cerf et al., 1993; Finholt & Olson, 1997; Finholt, 2002). Advances in the physical sciences are inexorably linked to a corpus of scientific data that can be shared, annotated, analyzed and debated by the community of physical scientists, as well as by developments in the instruments and technologies that are integral to formulating, conducting and analyzing results from scientific investigations. Apart from the notable example of *TalkBank*, to be discussed, with its emphases on digital audio files and transcripts of human talk, there is no such corpus of shareable audio and video data of human interaction in the social and behavioral science. This lack of a shareable corpus of social science data has hindered theory development. The DIVER project is devoted to fill this void by providing social and behavioral scientists with a tool and a platform for generating different perspectives on human interaction phenomena in the form of annotated audio and video recordings.

The psycholinguist Herb Clark (1996) has formulated the important social science concept of "common ground" as what it is people seek to achieve in the work that we do to coordinate what it is that we are attending to and/or referring to, so that when comments are made, what these comments refer to can be appropriately inferred. In the learning sciences literature, the "common ground" concept is usually used to examine collaborative or teaching learning discourse and pointing, bodily orientation and joint visual regard to the focus of a conversation that is being analyzed for studies of learning or teaching (e.g., Barron, 2003; Pea, 1994).

But it is not sufficient only to focus on the non-technology mediated aspects of common ground—for once we look at inscriptional systems (e.g., Latour, 1986) that lay down and layer symbolic records such as text or diagrams or other paper-based

representations, we see that they, too, become a focus of pointing and joint visual regard, and introduce new problems as transient referents. One may want to refer back to an earlier moment when only a part of a mathematical diagram was present to allude to that state or some part of that diagram, for that is what one wishes to establish common ground around. Similarly for a business meeting one may want to refer back to a state of information display on a whiteboard and what it is that the people in the room were speaking about when it was constructed.

I argue that this common ground concept must extend to the dynamics of representations, particularly for the problematic cases when these representations are computer-enabled (e.g., Pea, 1994). One often needs to refer to specific states of information display when using computer tools, so establishing a common ground for discourse and sense-making, what it is one wishes to point to, means capturing what Latour (1986) calls "immutable mobiles"[2] for replay and guided noticing.

Consider the potential but unactualized learning across learning science researchers themselves, with respect to the primary video data they collect of learning and teaching. Here too we need to facilitate the learning and cumulativity of knowledge construction that can come about through co-building on a common ground. A significant opportunity thus exists to work across researchers on a common ground of videorecords and annotation tools that is rare today. Without the dynamics of common ground, these results are improbable, and the sciences of learning and education will suffer accordingly.

Dynamic media e-publishing. Finally—a point for subsequent elaboration— the uses of digital video collaboratories in the human and learning sciences call for a new, dynamic publishing and commentary medium, one where multimedia records are integral to the presentation of the phenomena and analyses, and in which precise references to specific facets of behavior or analyses are possible to sustain the knowledge building processes of a research community (e.g., Pea, 1999). Such a call for e-journaling is commonplace (e.g., Harnad, 1991; Varmus et al., 2000) but the importance of dynamic media in such publications is less commonly raised as an issue of importance (for a notable example, see *Journal of Interactive Media Research*, but even here dynamic media are presented but not able to be indexed or commented upon by referring to their features in terms of space-time cropped selections of the interface displays).

3. Toward Video as a Scholarly Medium

Consider the core plea here to be examining what it will mean to make video as integral a scholarly medium as text is today. This aim requires treating video-as-data, and has a host of implications that will take us in different directions than are being pursued in today's broadcast-oriented approach to uses of video-as-file in e-learning, movies and sports videos-on-demand, and other cultural pursuits.

There have been diverse efforts to move video data into broader circulation so that it may be re-analyzed by others. While anthropological film's beginnings in 1898 by

[2] I refer to Latour's (1986) influential development of the concept of inscriptions as external representations of ideas that serve as "immutable mobiles" with these key properties: (1) Inscriptions are mobile; (2) They are immutable when they move; (3) They are made flat; (4). The scale of the inscriptions may be flexibly modified; (5) They can be cheaply reproduced and spread; (6) They can be reshuffled and recombined; (7) One may superimpose different images of totally different origins and scales; (8) They can be made part of a written text; and (9) Their two-dimensional character allows them to merge with geometry.

Haddon in the Torres Straits were only three years removed from the birth of the cinema (see Grimshaw, 1997), the tradition of filmmaking in visual anthropology was made salient in part by the classic works of Margaret Mead and Gregory Bateson (especially his book *Naven*, 1936, devoted to multimedia investigations of cultural ritual in Papua New Guinea). Researchers have recently taken advantage of other media to distribute video recordings together with multiple researchers' analyses of these data (CD-ROM enhanced issues of *Discourse Processes*, 1999, and *The Journal of the Learning Science*s, 2002). These important efforts nevertheless highlight the limits today to bringing video records into a broader collaborative process among researchers and practitioners. Challenging other researchers' interpretations of video analyses typically requires becoming a filmmaker oneself, and either bringing in other material or cutting and splicing materials from the original researcher's work, adding new commentary.

Compare these obstacles to the simpler task facing students of textual literary works, who are able to excerpt and juxtapose passages from their subjects with little effort, and with few limitations as to length or location within the primary work. Even in using the data-sharing innovation of a CD-ROM with video clips that were to be referenced in the articles, the special issues nonetheless decouple video data from analysis, as researchers reference video source time codes and lines in transcripts. To comment on researchers' interpretations of these video data, other scholars must engage in the same referential gymnastics, instead of referring directly, in context, to the video data in their own commentaries on others' video data commentaries.

4. Introduction: Point-of-View Video Authoring in DIVER, Guided Noticing, and Digital Video Collaboratories

Our approach to addressing these fundamental issues of making video function as a scholarly medium like text turns on several fundamental concepts, which I briefly introduce here. These concepts, illustrated in *Figure 2* are: (1) point-of-view video authoring; (2) virtual camera; (3) virtual videography; (4) "diving" into videorecords, and (5) guided noticing.

The first concept is *point-of-view authoring***.** This idea arose for me in thinking about the prospects of panoramic video for enabling fundamentally new kinds of interaction and manipulation of video records. Panoramic video, as noted above, involves using of one or more digital video cameras and mirrors to capture 360-degree horizontal imagery. Panoramic cameras are being explored for uses in sports, entertainment, surveillance, and business meetings, among other applications. The interesting issue for education and learning research is how to think about navigating and re-using such panoramic video records.

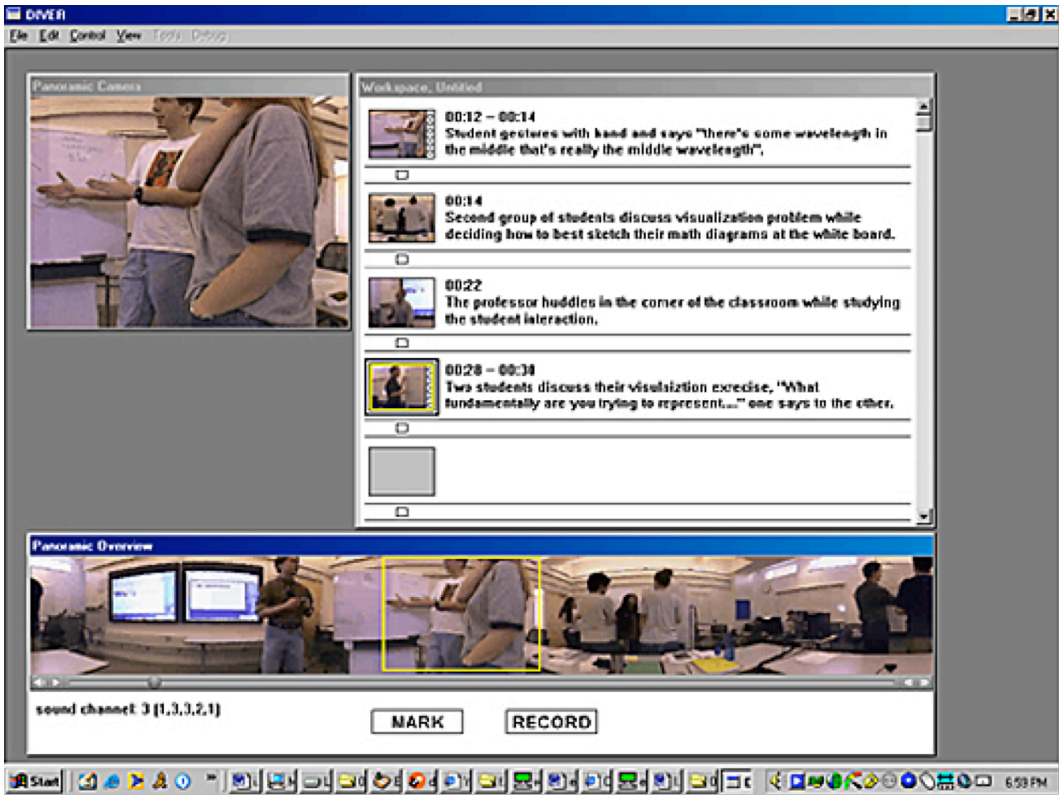

Figure 2. User interface for the DIVER desktop application, showing panoramic video. Note that we show a 20:3 wide aspect ratio panoramic video above, but that the DIVER interface can also automatically support Diving on traditional 4:3 aspect ratio video, or multiple video streams that have been synchronized from different cameras.

The innovative idea that we developed for the DIVER Project in thinking about panoramic video imagery is that of point-of-view authoring, in which a *virtual camera*, represented by a resizable rectangle (in one instantiation) can be panned and zoomed through that video, and used to 'record' a point-of-view movie through the panoramic video array—to which comments and annotations can be appended. Such a virtual videocamera makes it possible to enact *virtual videography*, creating any one of an infinite number of possible point-of-view authored films. We describe such path-movie authoring and commenting as **"***diving***"** into video. The use of a virtual camera to author point-of-view movies within a panoramic video record and to annotate these path movies performs an important action for establishing common ground that I characterize as *guided noticing*. The use of the virtual camera for the framing of a focus within a complex and dynamic visual array directs the viewer's attention to notice what it is that is thus circumscribed, and the point-of-view authoring thus guides the viewer to that noticing act. Finally, DIVER makes it possible for the author to publish their dive to a webpage, using a WebDIVER server, so that others may experience their point-of-view video document and annotations—their dive on a source video. In an alternative implementation, videos encoded for WebDIVER server accessibility may be dived upon as streaming media by authors using any simple web browser (thus not requiring the downloading of the video itself onto the diver's computer).

As it turns out, these core concepts are extensible from the case of navigating and point-of-view authoring within panoramic videorecords to the more general case of

taking as input to the authoring process I have described any video source, including normal 4:3 aspect ratio video from consumer videocameras, multiple yet synchronized 4:3 aspect ratio video streams (e.g., from different locations in a networked collaboration session), and more generally yet to any dynamic or static visual medium including art, photographs, scanned documents, animations and other imagery or visualizations, as in science, medicine or cartography (see Section 7).

I will later develop a broad account of current and prospective scenarios of DIVER use for enabling a broad range of cultural practices across and beyond learning, education, and scientific research. While not intended to be exhaustive, these scenarios will illustrate the scope of applicability of the general framework for guided noticing and point-of-view authoring. Finally, the capability that DIVER enables for any author to point to and comment on any part of a visual display and share that view so that others can experience it, by publishing it as a webpage over the Internet with commentary capabilities at its component level, opens up the prospect of *Digital Video Collaboratories*, whereby a scholarly infrastructure is established for integrating video analysis, sharing and collaboration in research communities (with potential for other cultural purposes). Making video function as a scholarly medium such as text turns on being able to refer to specific time-space coordinates in the source video, and in our work with DIVER and in collaborations underway with other researchers we are seeking to develop this vision into a reality.

5. DIVER and Guided Noticing

What is the activity of guided noticing? And how does it relate to other behaviors and concepts such as pointing, disciplined perception, and professional vision?

Guided noticing is a two-part act for a complex situation/visual scene. First, a person points to, marks out, or otherwise highlights specific aspects of that scene. Second, a person names, categorizes, comments upon or otherwise provides a cultural interpretation of the topical aspects of the scene upon which attention is focused. In a two-person (or more) interaction, there are also distinctive roles. One is the 'author' of the guidance (I guide you to notice what I notice), and the other is a 'recipient' of the notice, which is mediated by an environment in which each participant is immersed. In the case of DIVER, and related media such as video and film, such guided noticing is also time-shifted and shareable by means of recording and display technologies. Diving creates a persistent act of reference with dynamic media—which can then be experienced by others remote in time and space, and which can additionally serve as a focus of commentary and re-interpretation.

Why is guided noticing important? Because achieving "common ground" (e.g., Clark, 1996) in referential practices can be difficult to achieve, and yet is instrumental to the acquisition of cultural categories generally, and for making sense of novel experiences in the context of learning and instruction especially. Clark observes that for one person to understand another person there must be a "common ground" of knowledge between them, and his research and that of others illustrates the complex multitude of ways in which such "common ground" is inferred from immediate surroundings, mutual cultural context, and participants' own prior conversations.

Guided noticing builds on the fundamental human capacity of referring and the establishment of shared attention. As Wittgenstein (1953) observes in his *Philosophical*

Investigations in critiquing St. Augustine's view of language learning as achieved by observing pointing and naming of things by adults—unless shared attention between infant and adult is achieved, it is impossible to learn from such ostension. The roots of shared attention and what the philosopher Quine (1960) called the "ontogenesis of reference" are established early. In the 1970's, an influential set of publications empirically demonstrated the advent of shared visual regard following adult pointing or line of visual gaze at approximately nine months of age (Bruner, 1975; Scaife & Bruner, 1975; Bates et al., 1979) and its role in establishing what Trevarthen (1979) described as intersubjectivity between infant and adult (involving mutually adaptive purposeful activities).

Another reason why guided noticing is important as a human activity is that it facilitates the continual interplay between two fundamentally distinctive but complementary ways in which we make meaning in social activity—what Lemke (1999) calls the typological and topological modes. Typological representations assign a culturally meaningful category to some material form as when a pronoun is viewed as singular or plural, a verb form as expressing one type of tense or another, etc. In contrast, topological representations are visuo-spatial and often continuously variable in their qualities (examples include acoustic-vocal properties of speech such as timbre or sound quality, and visual media, ranging from gesture to graphs).

The obstacles to sharing analyses and commentaries of video data exemplified in special issues of *Discourse Processes* (1999) and *The Journal of the Learning Sciences* (2002) can be contrasted to the basic structure of human visual interpretation that builds on these human capacities for shared attention, and the interweaving of typological and topological modes of meaning-making. This structure appears in what I describe as a cycle of Looking, Noticing, and Commenting (the LNC cycle): I *look* at a visual scene, *notice* a pattern or something of interest, and *comment* upon it. If others are involved, the noticing is followed quickly by a gesture or visual regard that calls out the particular item of interest from the scene as a whole, so that others can connect the comment to the particular topical element that is being referred to. This LNC cycle—familiar to anyone in any field of work and life—is generative and recursive. Others can return to the scene as a whole to offer amendments or counter-comments, all the while noticing and gesturing in order to continue to tie the scene to the discussion and to build a common ground of understanding, or to negotiate differences in meaning. In daily life, this cycle is so well integrated into practice that it most comes to attention when it fails to function. However, in the analysis of video records, there has been no existing toolset to enable distributed communities of scholars or practitioners to readily engage in this kind of interactive commentary on a shared and accessible body of video material with anything like referential precision.

The basic Look-Notice-Comment (LNC) structure is used in early parent-child interaction in helping guide noticing of culturally significant phenomena and their meanings (e.g., Rogoff, 1990, 2003). It is also used in promoting expertise development in the exceptionally broad range of domains in which visually accessible information is critical, towards achieving what the anthropologist Chuck Goodwin (1994) calls "professional vision." Examples include learning to appreciate artistic works from specific frameworks of interpretation (Berger, 1972), recognizing diagnostic patterns in dirt at archaeological digs (Goodwin, 1994), coming to see important patterns of

classroom interactions as a teacher (Sherin & van Es, 2002), detecting tumors in radiological imagery, analyzing a motion picture, identifying mountains on a map, seeing galaxies in telescopically-enabled perception of the heavens, and so on. In each of these cases, one is faced with a complex visual array of information to *look* at, comes to *notice* aspects of that visual information (for any one of a number of reasons, such as its being distinctive, relevant to the action at hand or being pointed out in some way by a more expert other), and to *comment* on the meaning of the noticed aspect one has picked out of the looking.

Goodwin (1994) uses the phrase "professional vision" to characterize how, "central to the social and cognitive organization of a profession is its ability to shape events in the domain of its scrutiny into the phenomenal objects around which the discourse of the profession is organized." He presents compelling examples from archaeological field excavation and legal argumentation (and oceanography: Goodwin, 1995) to indicate how professions achieve this ontological work by *coding* (in which some structures of the world are captured and other possible ones ignored), *highlighting* (as in saliently establishing a figure against a ground, such as by demarcating graphically or pointing), and *producing graphical representations* to make a case (such as transcripts of talk, diagrams, and frame grabs of scenes recorded on videotape). I wish to argue that the Look-Notice-Comment cycle is the developmental base upon which the more academically focused concept of professional vision builds.

For millennia, humans conducted this central human activity as an oral discourse only, situated in the here-and-now of whatever visual information was in front of them (e.g., deLaguna, 1927) or able to be adduced from memory of shared experiences with others and talked about, and without the benefit of any inscriptional system, that is, a notation scheme that made sensible marks on a persistent and observable surface whose meaning could be shared with others (such as written words or other symbols that express a consistent system of meaning). Extensive scholarship has been devoted to the historical importance of this development in human culture, for the later emergence of logic, mathematics, science, art, and many other fields of human achievement (e.g., Cassirer, 1953-1957; Goody, 1987; Latour, 1986; Olson, 1994; Ong, 1982). For once, humans were able to inspect a persistent record of ideas and their interrelationships, and to leave behind a commentary on their topic of attentional focus from which others could learn.

In this broader framework of socio-cultural practices and technologies, the DIVER software goal is thus to provide a tool and a communications framework for augmenting the fundamental and social activity of human visual interpretation and communication about what is interesting or relevant to a purpose at hand—looking at a complex scene, noticing events of interest, and commenting on the focus of attention. As shown in *Figure 3*, the Look-Notice-Comment cycle has informed our design of the DIVER software. Specifically, the DIVER interface consists of three distinct regions, each corresponding to one of the elements in the fundamental Look-Notice-Comment cycle of visual interpretation and communication.

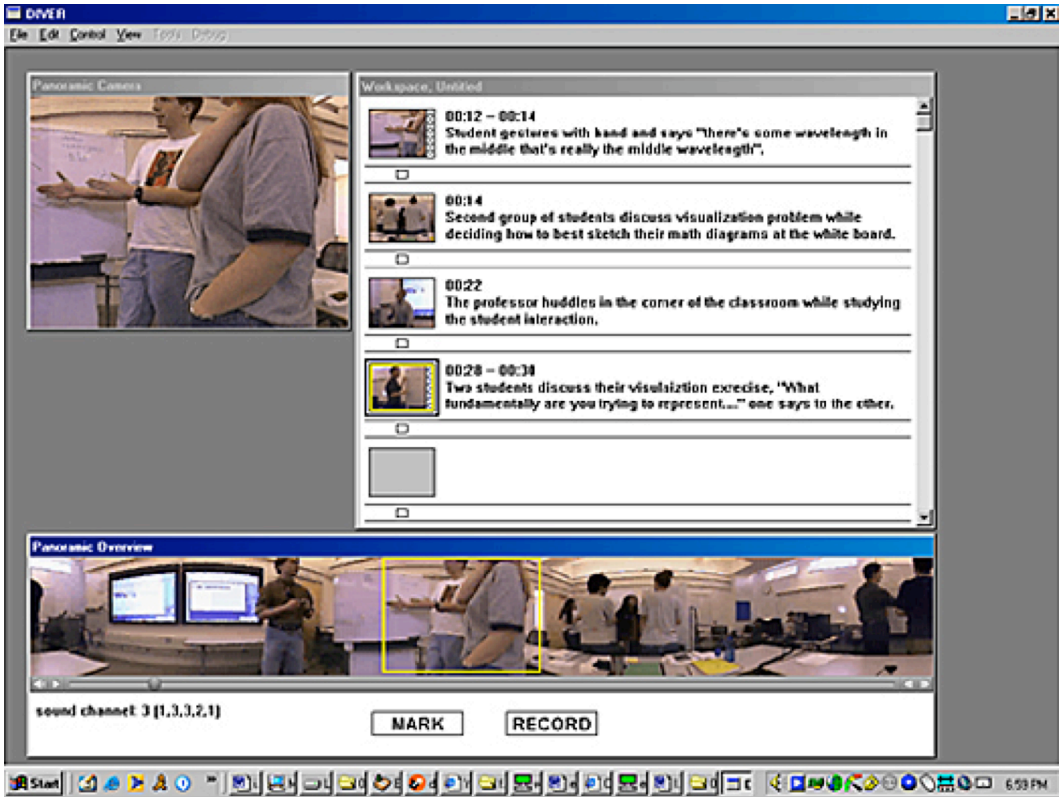

Figure 3. DIVER's windows: (1) overview ("Look") at bottom, (2) magnified viewing ("Notice") at upper left, and (3) the Dive worksheet ("Comment") comprised of panels.

The overview (Look) window provides access to the original source material that is the overall subject of the discussion or analysis, and has standard video controllers, for forward, reverse, and pause, which also allow for "time scrubbing" by moving the video player control forward or backward quickly, with a corresponding viewing of the video at accelerated speed (as a quick way to survey the content of a videorecord). In this overview window, users can return to any part of the original video source and explore it.

The magnified viewing (Notice) window depicts a selection from the original source that the user makes with the mouse by dragging a resizable selection rectangle overlay (the 'virtual camera') across the overview window, panning over and zooming into regions of interest in the scene. We sought to satisfy with this dual overview/magnified-view interface the well-known dual awareness principle of human-computer user interface design, also called "focus+context" (e.g., Card, MacKinlay & Shneiderman, 1999; Furnas, 1986; Lamping et al., 1995; Mills et al., 1992; Rao & Card, 1994). The user can thus zoom-in on an information source's details yet maintain a sense of orientation and context. Related navigational interfaces have been explored in several experimental and commercial applications of panoramic video technology (BeHere.com; Fullview.com; Kimber et al., 2001; Prihavec & Solina, 1998; Rui, Gupta & Cardiz, 2000; Teodosio & Mills, 1993).

As the virtual camera is moved across the video source overview, the viewing window provides a dynamically updated and magnified view of that region of the original video source. Two buttons–MARK and RECORD–turn this viewing area selection into

an easy-to-comprehend and flexible authoring activity. MARK takes a temporally tagged snapshot of the contents of the magnified viewing window and creates a new panel for commenting in the third region of the screen—the Dive (Comment) worksheet. A worksheet panel contains a video thumbnail representing a marked clip, as well as a text field for describing what is noteworthy about that selection[3]. If the DIVER user presses the RECORD button, another panel is created. When the record button is pressed again to end recording, the virtual camera's pathway across the video overview window completes its recording of the spatiotemporal selection shown in the virtual camera window.

Thus the user of DIVER can record a dynamic "path" through the original source video, and annotate it by entering text in the Dive panel. In creating a unique "point-of-view" virtual tour of the original video source, we say the user is creating a dive (or "diving") into the video record. He or she can then use the dive as a device for storytelling or other rhetorical purposes (e.g., video data argumentation for a research conjecture) by playing back the marked locations and dynamically cropped pathway 'recordings' through the original video source as well as the annotations of these records. A dive is subdivided into panels, which serve as containers for the data elements that constitute the dive (see prior Figure). Panels contain the following data elements:

First, static and dynamic time and space video markers: a static marker is a pointer to a frame in the video (temporal coordinate) and the location of the virtual viewfinder within that frame (spatial coordinates). Dynamic markers are recordings of the space-time path traveled by the virtual camera as the user guided it through the source video.

Secondly, a thumbnail-sized image copied from the source video at the marked time and space is used as a live link for specific time-space video coordinates. To make it easy to explore the place of the marked clip in the overview context, dragging and dropping such a marker into the viewing window resets the virtual camera to the space-time coordinates previously marked.

Thirdly, annotation text that a user typed into the panel is used to provide a commentary associated with the static or dynamic marker.

Fourth, the temporal point or range of video that is contained within a panel selection is denoted by a single time code for a static marker or a range of a time code range for a dynamic marker.

Finally, a time/space cropped video clip is depicted as a thumbnail-sized image in a panel to act as a symbolic representation of a selected portion of the source video segment. Double-clicking a thumbnail (or drag-and-dropping it back into the viewing window) repositions the source video at the corresponding time/space coordinates.

The simple and efficient creation of dives is supported by the DIVER file format. As a user drags the virtual camera viewfinder across the source overview window (see *Figure 4* below), a reference to time (the video time) and space (the position of a viewing region within

[3] The related idea of "panes" for annotating synchronized video and audio streams was first implemented by Roschelle, Pea and Trigg, 1990 in the VideoNoter system; also Roschelle & Goldman, 1991.

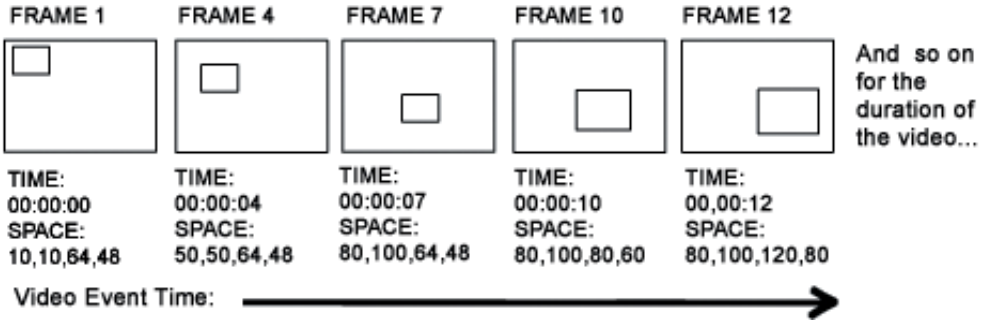

Figure 4. A graphical representation of the .dvc file format

the video's border) is recorded to a list. These time-space recordings (which we call '.dvc files', where .dvc represents "DIVER Video Collaboratory") can be played back by the user; the DIVER system reconstructs the virtual viewfinder's path through the original source video in real-time. What is recorded when the user presses the RECORD button in the DIVER desktop application is not literally a new video, but the .dvc file information needed to replay the dynamically cropped pathway which the virtual camera traced through the original sequence of video frames.

When a .dvc file is played, current video time and space is reset to the previously saved time and space coordinates. Prior to displaying each frame of video in the virtual camera window, the video's borders are scaled (in real time, on-the-fly) to match only the viewing area coordinates associated with the same frame (video time) in the recorded virtual camera path. Two key benefits inhere in using the '.dvc' list recording scheme vs. re-rendering new video clips: (1) disk storage requirements are greatly reduced since virtual video clips do not generate redundant video files; and (2) software performance is vastly improved because no rendering time is required—users can instantly create and play back dynamic path videos. Dives are thus extremely lightweight glosses on original source content. Yet they can comprise very rich and compelling viewing experiences—they truly provide the interactive experience of *making movies within movies*.

The structure and extensibility of the XML-based '.dvc' file format serves as a metadata container that can allow for the reformatting of DIVEs into many display and interaction environments. For example, on a lightweight PDA device supporting only text and image display, the '.dvc' format could be used to extract and display only text annotations and video thumbnail images.

In the desktop implementation of DIVER, a user's video dive can be uploaded to a WebDIVER™ server for sharing, discussions and collaborations. The content transformation process for the Web format takes place in a series of client- and server-side tasks invisible to the user. The user initiates an automatic series of text, image, and video clip transformation processes as he or she chooses the "Export to HTML" menu option in the desktop DIVER application (see Pea et al., 2004 for details).

When a dive is later opened on the WebDIVER site the content is reconstituted as

HTML and JavaScript, so that web pages are dynamically generated from the media files and XML data. WebDIVER browser pages look similar to desktop dives, for they contain the same data and media elements. The *Figure 5* WebDIVER window shows an example.

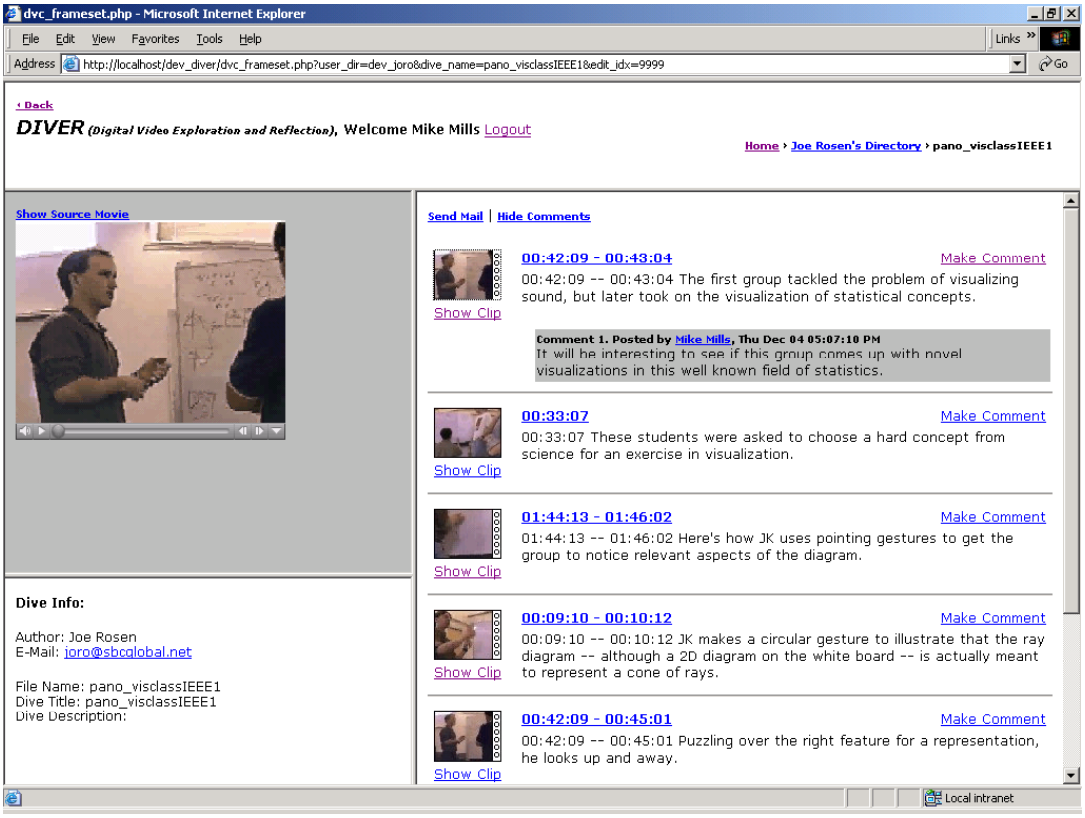

Figure 5. A dive from the WebDIVER site, as seen in a browser window

In addition to the original author of the dive, other WebDIVER users can now both view a dive and explore its full source video. Moreover, users can collaborate on a dive by entering their own comments on the dive annotations made by the dive's creator for each panel. WebDIVER users can also perform keyword and metadata searches within and across all annotations.

The paradigm of desktop DIVER use and WebDIVER uploading of both the media file and the user's dive for access and commentary by others—while very useful and powerful—has several key limitations. One of these is that the time for uploading large video files will often be prohibitive. A second concern is that many use scenarios where one may wish to allow users to dive on videos (e.g., research collaboration, entertainment) recommend a scheme for protecting intellectual property in which a user does not have the video file itself residing on his or her own computer, but where the video simply streams from a web server during the video playing and diving processes. A third concern is that the requirement to download and install the desktop DIVER application (and regularly update it with new versions) may make DIVER far less widespread in its uses than a web-browser based approach would allow.

To achieve this more demanding objective, we developed a new WebDIVER design

that provides many of the key desktop DIVER application functions, but in a fully browser-based approach. The design as of December 2004 is as follows (*Figure 6):*

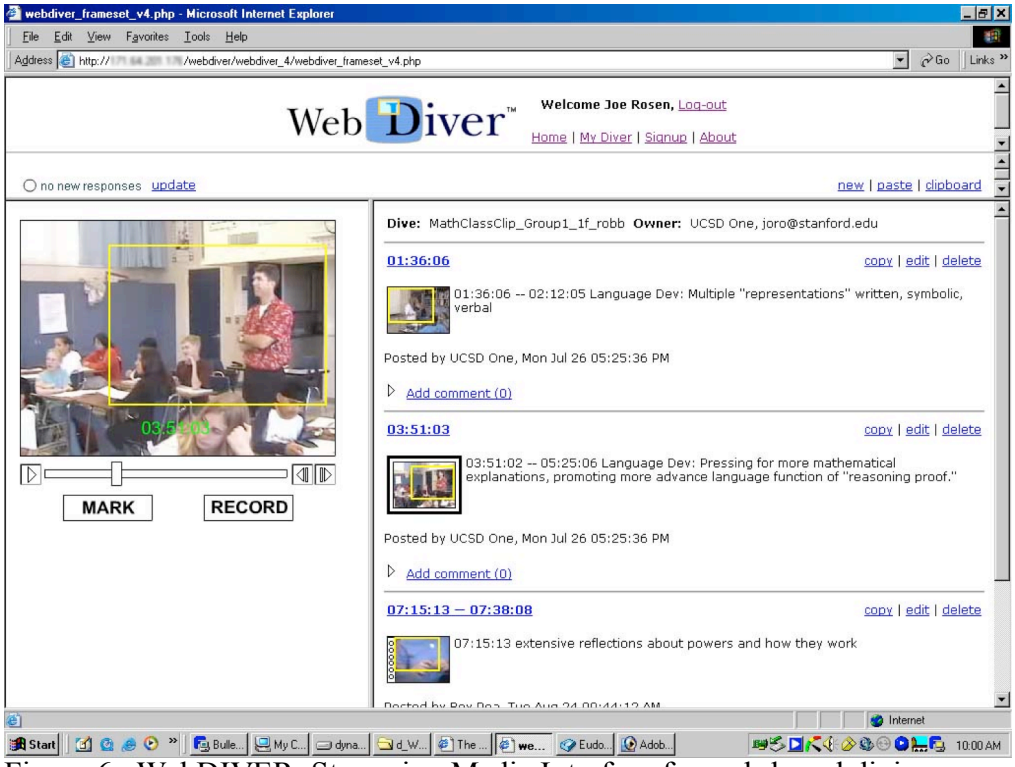

Figure 6. WebDIVER: Streaming Media Interface for web-based diving

One or more remote users use a web browser on their computers to go to a specific URL where a videofile has been encoded and made accessible so that it can be "dived" upon as described above. The new WebDIVER design still represents a dive as made up of a set of panels, each of which has an author and editing capabilities affiliated with it. The panel is a data container for the DIVER keyframe that is marked or the DIVER movie that is made with the virtual camera. Multiple users can be connected to the URL at once if they have permissions to do so, and can make their own dive panels for the videofile, with the MARK and RECORD buttons and the use of the virtual camera viewing rectangle under mouse control in the web-browser version, as on the desktop DIVER application. When they do so, they see their own work appearing in the dive in real-time, and if others are also making dive panels for that video, if one presses the 'update' button on the webpage (see *Figure 6*), the different panels of the DIVE that are being created collaboratively—whether locally or by distributed users—are made visible. Thus, users may be either face to face in a meeting room, or connected to the same webpage remotely via networking, as they build a collaborative video analysis.

In principle and in practice, there is no need for the users to be watching the same portions of the video at the same time when they work collaboratively; as the video is streamed to them through their web-browser, they may mark and record and comment at their own pace and as a function of their own interests. The collaborative video analysis activity that uses DIVER can be as planful or emergent as the participants choose to make it; constraints on focus, intent, duration of sessions, and so forth are not built in to the technology but a matter of negotiated social practice. Note one important difference:

whereas in the desktop DIVER application two video windows are part of the user experience (one an overview window of the whole videofile and the second the zoomed video contents circumscribed by the virtual camera rectangle), the WebDIVER system allows for just one window of the whole scene. The diver achieves guiding noticing by directing the dive user's attention to where the diver points in their dives with the virtual camera rectangle. Thus, when one plays back a dive on the WebDIVER browser version, the viewing rectangle moves across the video and shrinks or expands to represent what the diver pointed to during their dive creation.

Having now provided a basic account of how the DIVER and WebDIVER software systems work and the connections of their functions to socio-cultural purposes of guided noticing and point-of-view authoring, it will now be worthwhile to examine the profound implications of panoramic video when brought into this context.

6. DIVER and Panoramic Video as a New Medium for Learning Sciences Research and Educational Practices

The need. Because 4:3 aspect ratio digital videos captured with digital video cameras are so pervasive, they will often be the source material for DIVER. However, the use of such video records as data in social science research is hampered by their partial quality. One or two video recorders and microphones produces a record that is partial (since limited by what the operator chooses to capture and what to ignore), and often of low quality, particularly in the audio dimension (Lampert & Hawkins, 1998). The original researcher or other researchers who may be interested in different aspects of the events that transpired cannot recover events relevant to the analysis occurring off camera. These issues drew us to developing for research and educational uses the application of panoramic video and audio capture, navigation, and the DIVER application for creating point-of-view authored tours of videorecords.

Today, panoramic video is often employed for omni-directional vision/sensing applications such as teleconferencing, navigation and security or military surveillance (e.g., Fullview, 2002; Nayar, 1997). Bapna et al. (1998) provided a dramatic demonstration of panoramic video-guided tele-operated navigation of the Nomad robot for 220 km across the Atacama Desert of southern Chile, anticipating planetary rovers. Other panoramic video applications include making available a remote window onto a panoramically recorded event such as a lecture (Kimber et al., 2001), or as one component in an effort to produce an immersive simulation of activities such as a sports game or a concert (Neumann, et al., 2000; Rizzo et al., 2001).

We were attracted to exploring the feasibility and practicality of panoramic videorecording for research and education because one can place such a camera in a room, shoot the video, save it digitally, and in post-production, create with DIVER a virtually infinite number of different experiential paths through the video—as if one had multiple cameras in the room at the time of the video shoot. Panoramic video is of special interest in any context where the purpose and application of the video resources is not well defined before recording.

It is hard to wrap one's head around this when seeing and thinking about it initially. The reason may well be that it is easy to have a misleading question, that the videocamera is like an eye and therefore must have a "point of view." But in acquiring the images in the 360-degree surround, the panoramic camera at every moment in time,

every frame saved, provides the raw materials for an infinite number of points of view at once. And for every next frame, one can pan, tilt and zoom to any one of a very large number of other points of view, and so on, ad infinitum. There are infinitely many pathways that one can explore and then express, using DIVER's point of view authoring methods, what one sees and notices in the pathways one crafts through the panoramic recordings.

For uses of video in research on learning, our goal is to capture much more complete event recordings than is possible with usual means, by using techniques such as panoramic high definition video, multiple phased microphone arrays, and a software and hardware infrastructure to support variable numbers of synchronized data streams from sources such as handheld devices and video displays that may be part of the context of the activity filmed.

Synoptic video capture. We use DIVER as a software analysis tool for selecting and annotating clips from what we have called a "synoptic," or all-encompassing, panoramic view of an event, in order to allow users to gain control over the high volume of data thus produced. Our ultimate purpose is to record events in a way enabling researchers to return as often as they wish to the original event, refining, modifying, or rejecting their conjectures in the light of new interpretations of the recorded activities. This goal stems from the high degree of complexity of learning environments (such as classrooms), defined by the interaction of multiple simultaneous channels of behavior and communication in context of a complex physical and representation-rich environment. It is difficult to determine, either in real time or in a single viewing, what is important about a learning event. Making sense of an event typically requires repeated viewings, isolating in turn individual components of an event (see Jordan & Henderson's 1995 paper on video interaction analysis). Conversely, any real-life learning situation can sustain multiple different interpretations, focusing on different factors of theoretical interest and on different aspects of the recorded behavior.

We would thus like to produce as complete a record of an event as is practicable from panoramic recording, so that the DIVER software can support repeated interrogations of the original event from differing perspectives (and with differing hypotheses) about what aspects of the event are significant. Our aim with this synoptic approach is in part to overcome faulty presumptions about inherent limitations of "video as data" in how ethnographic records have traditionally been collected with video cameras, for example:

> "Although video technology can clearly capture more than one observer could possibly notice and record, in another sense video technology may capture less. Videographers have options as they decide how to tape a classroom: they might pan across the room or fix on the teacher, depending on the structure and goals of the data collection and the videographer's training. Tapes cannot capture everything that happens in a classroom, and what they miss is typically determined by the position of the camera rather than by a trained observer's instinctive reactions to events as they unfold. Moreover, while video data are less "processed" than other kinds of data, they are not equivalent to direct observational data. The view of someone watching videotape is constricted; he or she is experiencing what is taped through the frame of a host of decisions made before and as the data were collected" (Hall, 2000).

These assumptions are challenged with a "synoptic" approach to video capture, that is, the use of multiple perspectives/instruments on the same event in order to compensate for the partial quality of any one instrument. In post-capture interrogations of the

panoramic digital video data from a classroom situation, the analyst can choose different fields of view and pathways through them depending on their purposes—in effect, going "beyond being there" (Hollan & Stornetta, 1992) in ways that a single videographer could not have done at the time with a traditional camera.

The completeness of the record is further addressed in the spaces instrumented for panoramic recording, which may consist of a room set up with multiple audio and video-recorders, with the goal of producing a synoptic record of events. We obviate the need to make real time decisions about what is "important" to shoot by using a panoramic video recording system. We have also demonstrated that it is possible to equip a recording space with multiple microphones to capture with equal clarity interactions occurring at any point in the room (see below). This synoptic architecture is deliberately designed to be open to future enhancements, both by enhancing the recording ability of any particular recording modality (e.g., additional video cameras to deal with problems of occlusion and differential resolution based on spatial position, or moving from standard definition to high definition video), and by adding new modalities (e.g., recording, in addition to the audio and video, the content of displays being used by participants in the space, or other behavioral data streams).

An example of "being beyond there" that ethnographers have found particularly compelling in using DIVER is the capability of creating grouped clips, that are made up of selected regions of the video scene which can then be played in unison to explore relationships between what was happening in different parts of a learning situation at the same time, and annotated accordingly.

A major technical challenge in recording panoramic video is *image resolution* (see Benosman & Kan, 2001 for an excellent edited volume on problems and recent solutions in panoramic vision). In order to be able to zoom into an area of the panoramic video as we do with DIVER's virtual camera and still see sufficient detail, the video must be captured at the highest possible resolution. In principle, panoramic video can be recorded with a single-lens system: a spherical mirror, somewhat similar to a fisheye lens, collects the light from the 360 degrees surround and redirects it onto the CCD chip of a single camera. The digital video stream of the camera can then be processed (stored, compressed, edited) like any other video stream. Such systems are relatively inexpensive and readily available; examples include the BeHere® camera (behere.com) and the camera used in Ricoh's Portable Meeting Recorder (Lee et al., 2002). But single-lens systems have very low image resolution because the number of pixels on the CCD chip, designed for a small angle of vision, must now be distributed across a much wider angle for surround vision. Thus zooming into an area within the video leads to a very low image quality due to the pixelization of the video data.

To solve this problem, several multi-lens systems have recently emerged. They capture a scene through multiple lenses and CCDs with slightly overlapping fields of vision and then construct the panoramic image by "stitching" the frames together (Tan et al., 2004). Examples are the Fullview® camera (fullview.com), the FlyCam developed at the Fuji Xerox Palo Alto Laboratory (Sun et al., 2001) and the RingCam used in Microsoft's meeting capture and broadcasting system (Cutler et al., 2002). Multi-lens systems address the image resolution problem, but they introduce the challenge of stitching and de-warping of multiple frames in the computer, perhaps in real-time. Also, multi-lens systems running at full resolution generate a very high bit rate, often too high

to be transferred over a single PCI bus.

In the DIVER project we use a Fullview® camera system with five mirrors and five cameras, whose streams are fed into a PC where they are stitched and de-warped at a resolution of 1800x240 pixels, at about 15 frames/s. The panoramic video is written to a file for subsequent processing by the DIVER toolset. We use three parallel PCI buses on the PC doing the processing to accommodate the very high data rate generated by these five digital video streams. We have developed a tape-based alternative recording system (see below) capable of even higher resolution, which is desirable for the re-purposing of panoramic video. We have also devised a system for capturing multi-channel sound with a microphone array so that panning and zooming with the virtual camera into the overview window on the resultant panoramic video record provides a corresponding audio focusing mechanism[4] (see below).

It is important to ensure compatibility with 4:3 aspect ratio video and panoramic video content at a variety of aspect ratios. As a result, DIVER has been architected to support a general approach for video analysis, collaboration, and interaction allowing for use of a diversity of aspect ratios and resolutions for recorded video material. This provides a flexible approach for users to record, analyze, and share a broad range of video material in the DIVER environment, independently of the specifics of the video material that has been captured.

In order to accommodate even higher resolution video capture than is possible with native FullView® live preview and direct-to-disk video, we developed an alternative full-resolution video tape-based recorder. The output from the FullView® camera array is recorded at full-resolution (720 x 5 x 480) and full frame rate (~30 frames/s) to five consumer-quality Digital Video Tape Recorders. Desired video tape segments are later transferred to a PC, where in a software assisted post-process the multiple recordings are reassembled into seamless full-resolution 3500x460 @ ~30 frames/s QuickTime encoded panoramic movies ready for DIVER.

While there are a number of systems that provide panoramic still image capture (e.g., Apple QuickTime VR: Chen, 1995), and fewer systems that capture panoramic video, they are all oriented to providing an immersive remote or recorded experience (e.g., Boult, 1999; Neumann et al., 2000; Rizzo et al., 2001), not enabling authoring by the user of point of view perspectives on their interests in the visual array of the recorded data. One interesting exception is the FlyAbout system where panoramic video captured by moving a 360-degree camera along continuous paths is spatially indexed for subsequent interactive navigation so that users can interactively replay video by traveling along paths using a map-like interface (Kimber et al., 2001).

The DIVER software system provides users with capabilities for adding perspective on the digital files that have been collected with a panoramic camera. Diving is designed to be a liberating authoring medium for video, perhaps most evident with panoramic video records, as it is constructivist in nature since any user can create their own annotated point of view on the video that has been captured. Creating one's personal point of view in establishing a dive through a panoramic videorecording amounts to what

[4] In addition to Joe Rosen's fundamental contributions on design and implementation of DIVER audio focusing, I'd like to make special thanks to Dan Nelson, Bob Smith, and Brian Luehrs of Stanford's Center for Innovations in Learning.

I call Personal Reality Authoring—as a new complement to the recent decade's developments in Virtual Reality (Rheingold, 1992) and Augmented Reality (Feiner, 2002), neither of which are created by the intended user of the medium—because the diver can save his or her path through the panoramic videorecord. Once saved such a record provides a mechanism for replaying the user's authored path, or guided noticing tour, through the panoramic videorecording.

Panoramic audio and DIVER zooming into the Panoramic Audio-Video Array. We have highlighted the common problem in use of one or two videocameras for learning research of poor audio quality for select parts of the videorecord. DIVER's ability to focus attention on a particular visual area within a panoramic (or non-panoramic) video image—the unique capability for interacting with video which we call "guided noticing"—raises the question of how to handle navigation of the sound field, once captured from multiple microphones in an instrumented recording space. Our current solution is associating coordinates within the video visual field with sound tracks, which locations then function like individual physical microphones (although they could also be produced by mixing microphone inputs in post-processing, and vary from the number of physical microphones). In this manner, we can switch to the corresponding sound track as the user navigates the virtual camera through the visual field (as in *Figure 2*).

DIVER panoramic movies may currently contain up to ten independent sound tracks. Each is mapped to an associated field of view captured by the FullView® panoramic camera. These multi-channel sound recordings are embedded as navigable QuickTime audio tracks within the post-processed panoramic movies. The result is that panoramic sound can be steered from inside DIVER in such a way that only the tracks associated with the current focus of the virtual camera become audible. This mechanism allows for the easy repurposing of audio recordings and the dynamic creation of infinite virtual audio mixes. The audio remapping process happens continually and on-the-fly, so the user is not required to wait through a time consuming sound re-rendering phase.

7. DIVER Visions and Developing Scenarios of Use: Research, Diagnosis, Training and Communication

Introduction. We distinguish several key forms of point of view authoring made possible using DIVER which we believe will prove useful: (1) tool for *learning research use* for expert analysis and discussion of video data (and other media); (2) *assessment of novices and trainees* in diverse fields, regarding details that they notice in complex visual media and the qualities of their commentary about it; (3) *the "scaffolded" training of novices*, who could be guided in their analysis of a video record by comments designed to stimulate their own perceptions; and (4) *expressive uses* for communication of perspective on complex media records by users in virtually any life sphere. The details of diving will depend on the nature of the cultural purposes of the media on which one dives. I then exemplify these categories for a broad variety of cultural purposes in research in the learning and other social sciences, and then consider prospective DIVER uses across this taxonomic span of purpose types for the physical sciences, art, film and popular culture, sports and other entertainments, e-commerce, and training of all kinds including sales and law. It is worth noting that one can distinguish DIVER use by type of media (dynamic or static), and the dynamic media may be either videos, animations, or dynamic records captured from running a simulation or model, whereas the static media may be a radiology image or a work of art or a photograph of a landscape scene or a

microstructure of a plant, to name but a few examples.

Tools for Research on Learning and other Media. DIVER has been developed first for use as an expert tool involving selecting and annotating video clips of human learning in diverse contexts from classroom learning settings, to the workplace, to home, community and more informal learning settings—for the purposes of empirical investigations of patterns and relationships in behaviors, talk, and other levels of analysis. With its use for these purposes, we may study knowledge-in-the-making and examine the ways in which the physical, inscriptional and social environments contribute to learning and knowledge construction. As applied to teaching, one can do empirical investigations of patterns and relationships in the work practices of teachers, and how they relate to student behaviors.

In one notable DIVER extension to be discussed, when applied to the physical, biological and medical sciences, it is often observed that to investigate and understand many of the interesting phenomena of biological growth, or cellular activities, or dynamics of fluid flow or mechanics in physics, one needs video or simulations that can 'play' at different spatial and temporal scales. In this use of DIVER, once one has video in hand, one can select and annotate video clips of the physical or biological phenomena at hand—for the purposes of empirical investigations of patterns and relationships.

Diagnostic Tool. DIVER also provides an experimental tool for teasing out differences in the ways that novices and experts in the learning sciences "see" or "notice" social and behavioral phenomena captured in the videorecordings. As applied to teachers, we can use DIVER as a means of identifying differences in the ways that novice and expert teachers "see" or "notice" patterns in learning, classroom interaction, or teaching. We know that there are extraordinary differences in the perceptions and cognitions of expert and novice teachers. Expert teachers literally notice very different things in the unfolding drama of classroom interactions than do novices, and their perception of these patterns contributes to their ability to make nuanced decisions about what to do next to enhance the learning that will take place (Berliner, 1988, 1994). Teachers need to develop expert models for perceiving instructional opportunities and how the challenges of meeting them are achieved.

For either application, one can do free-field selection of parts/clips of the videorecording or annotation on their part but can also provide highly structured assessment templates, in which the DIVER user's attention is directed to particular spatiotemporal coordinates in the video array and about which specific questions can be asked. Teaching mentors may also mediate the novice in role-playing the decision-making role of the teacher depicted in the video—"what would you do next?"

Finally, as discussed later, DIVER may find utility in teacher certification processes, used for teachers to build video-based evidence from their teaching practices and other documents that they use to argue that they meet specific rubrics of evaluation.

Training Tool. We know from the model-scaffold-fade instructional approach that has been called "cognitive apprenticeship" (Collins, Brown & Newman, 1989) that a powerful mode of learning a complex skill and knowledge base is to experience an expert in the discipline modeling their problem formation and problem solving processes, 'thinking aloud', and to then be scaffolded, i.e., supported as a learner, in the performance of that complex activity. Finally, the instructor then 'fades' the scaffolding

so that the learner can function autonomously without the aides initially required (Pea, 2004). In applying DIVER for such a scenario, learners who seek domain expertise would be provided with inquiry modeling in structured video dives provided by experts. These dives will highlight with guided noticing and commentary how the expert perceives and interprets the complex environment in which their developed skills are deployed. Then, following experiences with such a modeling phase in the cognitive apprenticeship approach, the DIVER user would be provided with scaffolding support in pursuing their own perceptual activities in the task domain. Such scaffolding can take various forms, including reducing the degrees of freedom affiliated with an interpretive response, or other forms of hinting of prospective solution approach to the problems at hand. In any respect, the DIVER user's dives, once posted to WebDIVER, can be provided with asynchronous commentary/critique from experts concerning the user's learning efforts to detect or notice the patterns in the videorecords that the experts recognize.

Expressive Communicative Tool. In this scenario, anyone who can use DIVER and desires to communicate a point of view about a media record (static or dynamic) can author a dive. These purposes can be exceptionally diverse, from day to day dialogs with peers or family, to formal certification activities, to entertainment—as in making dives of a favorite movie to share fun parts of it with a friend.

A particularly powerful use in this vein would be to ask children in a classroom to dive on a panoramic video of a lesson and to provide the point of view path movie and commentary on what they noticed and thought about as it was happening (e.g., Goldman-Segall, 1994, 1998, explores what she calls "point of viewing" by learners about their school experiences). This stimulated-recall scenario is likely to yield results that will surprise the teacher and others who view the students' dives, because we so rarely understand their points of view on the instruction they receive. Yet what matters is their experience, not simply the experience of watching the teaching an adult provides.

There is a potentially radical activist agenda that could be pursued with point-of-view authoring of DIVER videos as well, that builds on work from the 1960's that began to provide cameras and videocameras to children in poverty conditions for their own documentation of their reality (see *Figure 7*), and in general to provide personal alternatives to mass produced studio television programming (e.g., Halleck, 2002; Shamberg's "Guerilla Television," 1971).

Figure 7. Photo courtesy of Paper Tiger Television (http://www.museum.tv/archives/etv/A/html/activisttele/activisttele.htm)

7.1 Tools for Multimodal Studies of Classroom Discourse

Researchers who study learning and teaching in classroom interactions often capture multimedia records and seek to use them for understanding their phenomena of interest. Problems of the selectivity of videorecording are compounded with an additional problem. As Rogers Hall (2002, personal communication) has argued, the process of moving from a concrete video to a general proposition with scientific status should be transparent and reversible, in order to preserve the status of video as evidence. He suggested that providing 'viewers' with access to the original video and the means to author their own perspectives might be a useful way to maintain focus on the process of fact and argument construction out of video evidence.

If digital video is to become widely used and accepted as evidence in the learning sciences, it is necessary to develop practices for video-based argumentation, as well as statistical uses of video. Otherwise, oft-heard arguments will remain that the selective capture of video is anecdotal, and used primarily for buttressing the beliefs of an investigator rather than for systematically testing scientific conjectures. For example, with panoramic videorecords that do not bias a point of view in selecting specific students to focus on, and an algorithm for data-sampling, one could do randomized time-sampled or other designs for systematic experimental comparisons of student behaviors in an environment employing designed interventions versus a control classroom.

Ken Hay's ITMD Project (Integrated Temporal Multimedia Data) at the University of Georgia uses a variety of cameras, audio recorders, and computing devices to produce multiple streams of data from event recording (Hay & Kim, in press). This approach ensures that the event is fully documented, but poses difficult issues in data management and integration. Hay and Kim describe some aspects of a response to these issues, such as real-time coding of 'nodes' of event activities. The amount of data generated by the ITMD research approach has led Hay's group to a concept with very broad applicability: the amount of "friction" involved in creating and, most importantly, storing, retrieving, and using video records. Friction refers to the amount of time and, secondarily, expense that is required to make a video record useful. If friction is high, researchers will tend to make relatively less video and to make less use of it in their work. If friction is reduced, then we can expect digital video to be more heavily used. Hay urges that friction be quantified as much as possible, and that attention be directed within the community to reducing the amount of friction associated with digital video inquiry.

There are a plethora of tools for video analysis, editing, and reflection, and it is important to characterize some of them briefly and how they relate to our DIVER work. We found in a recent workshop bringing together leading video researchers in the learning sciences and teacher education that there were ten different functions of video research that are supported (or not) across these different tools (Pea & Hay, 2003): (1) Acquisition; (2) Chunking; (3) Transcription; (4) Way-finding; (5) Organization/Asset Management; (6) Commentary; (7) Coding/Annotation; (8) Reflection; (9) Sharing and Publication; and (10) Presentation.

Many of the tools used by research communities have focused on developing only a few of these capabilities, and as several examples illustrate, they vary considerably in their strengths of support for the functions. VideoPaper Builder is designed primarily to facilitate the creation of web-based "video-papers," educational research publications that

incorporate videoclips (Nemirovsky et al., 2001). The CLAN tools that MacWhinney and colleagues (2000) have developed for TalkBank provide an exceptional suite of transcription, coding, and annotation tools but are not oriented to supporting reflection, sharing, or commentary. In contrast, the Teachscape platform for providing school districts and teachers with video case studies of exemplary teaching integrated with on-line community of interpretation focuses on chunking (there are highlighted sections of each of the videos, for particular instructional purposes), and reflection (as supported by use models of study groups of teachers who exploit the online community features at home, e.g., Lu & Rose, 2003). But while providing transcripts of videos used, it does not offer transcription tools, as it is designed more as a teacher professional development environment than as a research tool.

Other tools are tuned to analyzing videorecords by coding, creating transcripts for and/or annotating videoclips (e.g., *Constellations and Orion*: Goldman-Segall, 1994, 1998; *Transana*: Derry et al., 2003; *Video Analysis Support Tool*: Sherin 2001), and several commercial tools used for coding and transcripts in behavioral sciences research are available in Nudist/nVivo, and atlas.ti. (There is a long history to video annotation and analysis systems I will not survey here, e.g., Harrison & Baecker, 1992; Mackay & Beaudouin-Lafon, 1998; Roschelle, Pea & Trigg, 1990).

Video editing and chunking are often accomplished with commercial tools such as Apple Computer's *iMovie* or Adobe *Premiere* but these are not oriented to coding or reflection, among needed functions noted above.

Reed Stevens' desktop *VideoTraces* system (Stevens et al., 2002; Stevens & Toro-Martell, in press; Cherry et al., 2003) is oriented to reflection and presentation, in enabling users of his software to lay down a reflective "trace" on top of a videorecord (the "base" layer that can be played at variable speeds), consisting of voice annotation and a pointing gesture implemented as a pointing hand cursor. A VideoTraces file may then be replayed so that one hears the audio trace overlay and can see the "pointing at" specific aspects of the videorecord upon which comments are being made. Stevens and colleagues have used this system in science education museums and in diverse higher education courses including rowing and dance composition, illustrating the value of this methodology for providing a time and space based interpretive layer on video records. In Cherry et al. (2003), uses of VideoTraces in an undergraduate class in dance composition were revealing. They illustrated how providing a concrete representation upon which pointing gestures and voice annotations could be made: (1) supported students in new ways of seeing dance works in terms of concepts and techniques of choreography; and (2) enabled the instructor to better understand students' intentions in their dance work from their VideoTraces annotation of that work, and to provide individual feedback. Students considered viewing the instructor's traces on their dance work video as the course component that contributed most to their learning.

VideoTraces is guided by Stevens and Hall's (1998) insights on the importance in "learning to see" in specific disciplines, what they call "disciplined perception," in the spirit of Goethe (e.g., Seamon & Zajonc, 1998; Shotter, 2000) and Humboldt from an earlier era who used the phrase to describe the possible objective histories of nature that science enables.

Stevens' use of virtual pointing and voice-recorded commenting within a video

record provides a complementary but different mechanism to our use in DIVER of guided noticing for achieving common ground in a referring act in the complexity of a video datastream. As yet the VideoTraces system is a stand-alone desktop application written in Macromedia Director, but at a VideoTraces installation site, a use community can also make a response on a VideoTraces file in the manner of a threaded discussion. Unlike DIVER and WebDIVER, VideoTraces does not involve any video transcoding or path authoring activities for the user, or Web-based access to such traces.

Diverse computer-based tools are also used for compiling video materials for teacher professional development (Shrader et al., 2002), including commercial products Teachscape and LessonLab. LessonLab (owned by Pearson Education) provides a client-server solution for K-12 schools to construct their own video case training materials from video that they capture themselves. Related research efforts include the Carnegie Knowledge Media Lab (Young, 2001; see http://www.carnegiefoundation.org/KML/), the Case Creator Project (Doerr, Masingila & Bowers, 2003), Indiana University's Internet Teaching Forum (Barab et al., 2003), STEP (Derry et al., 2003) and Schrader et al.'s work on pre-service teachers' web-based video cases in elementary literacy (2003).

At a more advanced level of teacher competencies, elaborated documentations of professional vision in teaching have been developed at the Carnegie Foundation for the Advancement of Teaching. The Carnegie Foundation, through the Carnegie Academy for the Scholarship of Teaching and Learning (CASTL) and the foundation's Knowledge Media Laboratory, provides exemplary teachers with resources and technical support to fully document an extensive aspect of their teaching, such as a course. CASTL provides fellowship support to allow teachers to devote themselves for a period of one or two years to the documentation and improvement of their teaching, in part expressed in richly documented web-accessible video (http://kml2.carnegiefoundation.org/html/gallery.php).

In the Case Creator Project (http://www.sci.sdsu.edu/mathvideo/cc/), teacher educators can create interactive video case studies of teaching by importing QuickTime videos and transcripts, creating an "issues matrix" of many different pedagogical issues that are relevant to the case, add web hyperlinks and/or supplementary text.

As interesting as this distinct efforts are, the most important lesson from our video research workshop (Pea & Hay, 2003) was that, without the use of meta-data coding and affiliated XML schema exposing such video analysis metadata to browser search, analyses developed with any of these tools will be stranded in data islands that can only be used and understood within the tool in which they are created.

While the issues affiliated with establishing broadly applicable meta-data for video analyses and video cases in learning and teaching are significant, since generic and discipline-specific categories for meta-data will be needed, the broader goals of establishing distributed communities that can communicate about their video data productively will be unachievable without such efforts. Productive work in this direction for coding meta-data for instructional materials is evident in GEM's instructional topics hierarchy, and pedagogy elements/values (http://www.geminfo.org/), building on the Dublin Core (http://dublincore.org/), and in the metadata development activities of the OLAC (Open Language Archives Community, www.language-archives.org) initiative, which in turn conforms to the larger OAI (Open Archives Initiative, www.oai.org). The stated goal of OLAC/OAI is that "any user on the Internet should be able to go to a single

gateway to find all the relevant language resources available at all participating institutions, whether the resources be data, tools, or advice. The community will ensure on-going interoperation and quality by following standards for the metadata that describe resources and for processes that review them."

These efforts are founded on the emergence of the Semantic Web and its uses of the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs (Universal Resource Identifiers, which include URLs) for naming. "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." (Berners-Lee et al., 2001). As noted on the World Wide Web Consortium site (http://www.w3.org/RDF/): "The Resource Description Framework (RDF) integrates a variety of applications from library catalogs and world-wide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events using XML as an interchange syntax. The RDF specifications provide a lightweight ontology system to support the exchange of knowledge on the Web."

7.2 Distributed Collaborative Video Interaction Analysis

We believe that video dives will come to play special functions in interpretive activities when people are working to do sense-making together of the records captured in video, whether it is human interactions, teaching behaviors, or other phenomena of interest. Like designers collaboratively working to design objects (Geisler & Rogers, 2000; Hindmarsh et al., 1998), the makers and users of dives take on special relationships with these point-of-view authored video tours as "boundary objects" (Star & Griesemer, 1989) between their special terms and conceptual frameworks of engagement—but opening up the prospects of building common ground (Clark, 1996). Henderson and Jordan (1995)'s paper on the Interaction Analysis Laboratory (IAL) methodology for developing multi-disciplinary analyses of video through group viewing of tapes in face-to-face settings illustrates the power of doing better collective thinking when researchers can build video analyses by sharing their perspectives on the observed phenomena and theoretical frameworks that they use to make sense of interactions captured in videos.

Building on the IAL methodology, WebDIVER enables distributed communities to contribute distinctive perspectives that, for example, individual researchers have on a videorecorded learning interaction, and to then use the contrasts and convergences to advance their collective understanding of the empirical materials. These needs and an approach for tackling them that we have underway will be taken up in Section 8 on video collaboratories.

7.3 Teacher Learning and Teacher Education

U.S. processes of teacher education and certification are in major transformation given concerns with increasing levels of student performance in standards-based instruction spurred in part by the federal *No Child Left Behind Act*. Beyond coursework and student teaching, over 40 states mandate that teaching candidates pass tests of basic skills, subject matter knowledge, and/or pedagogical knowledge for certification (Youngs, Odden, & Porter, 2003). Since these certification processes are mostly remote from the actual teaching practices, based on written tests and coursework completion, many teacher education programs and states are turning to performance assessments that include videorecordings of candidates' classroom teaching practices to make their

certification decisions. Video is a potentially more effective assessment method than a written test in providing an unfiltered (even if selective) view of the teacher candidate directly interacting with students.

However, as already noted for learning research, video suffers from a number of obstacles that impede its usefulness, e.g., such as requiring complex knowledge of video-editing software and techniques. And for assessment purposes, video simply does not support the kind of fine-grained access and reference possible with text. It is challenging for a teaching candidate to reference a particular teaching interaction in a video to illustrate how her teaching meets a specific evaluation rubric, e.g., "I promoted conceptual understanding by supporting student inquiry following the question about nonlinear functions from the student in the blue shirt in the front left of the room that occurs about five minutes into the video." A scorer of the teacher's video portfolio materials must then search for this incident, and may find it difficult to uniquely identify the candidate's intended event. Scorers of video have the same problem in efficiently and unambiguously pointing to the elements of teaching practice involved in reaching their judgments. The same features that make DIVER uniquely useful for research purposes—using a 'virtual camera' to precisely refer to events bounded in space and time within a video, and then attach text annotations to them—make it a potentially transformative technology in the context of teacher assessment and evaluation.

The effort to use DIVER for performance assessment of video records of teaching events builds on other responses to strengthen prospective and practicing teachers' knowledge of both content and content pedagogy, across a broad variety of subject disciplines and age levels, and on portfolio approaches more generally for documenting the work of teaching (e.g., Anderson & DeMeulle, 1998; Athanases, 1994; Porter et al., 2001; Snyder et al., 1998; Stone, 1998). These responses include videocases portraying teaching dilemmas related to content (e.g., Barnett, 1998; Crismond, 2003; Galvis & Nemirovsky, 2003; Horvath & Lehrer, 2000; Koehler, 2002; Mumme, 2003; Schrader et al., 2003), "video clubs" for teachers to review and discuss one another's work (Frederiksen et al., 1998), computer files containing artifacts of an entire year of teaching (Lampert & Loewenberg-Ball, 1998), and videotapes enabling international comparisons of teaching practices (Ma, 1999; Stigler & Hiebert, 1999; Stigler, Gallimore & Hiebert, J., 2000; Ulewicz & Beatty, 2001; Hiebert et al., 2003). Multimedia records portray teaching that is less filtered by interpretation than vignettes or observations, and avoid the complexity and inefficiency of making sense of multiple observations of different teaching actions that occur when students observe different classrooms.

The use of DIVER provides an opportunity for teacher education programs to reform their practices in ways that increase candidates' ability to construct and apply pedagogical content knowledge, analyze student learning, and reflect on video and other records of practice.

Can video analysis software that permits precise reference into a video and other records (such as scanned student work records) improve the quality of assessment over the use of video and other artifacts without such a tool? Can the use of such software reduce the cost (and thus increase the feasibility) of implementing a performance assessment system for credentialing that includes video? Can teacher candidates' use of DIVER software improve their ability to reflect upon their teaching practices, and to recognize the connection between their practice and state and national standards? Uses

of DIVER software for teacher education and certification provide a powerful research opportunity to produce knowledge that can guide teacher education institutions and certification agencies.

We have begun to explore these questions in a partnership with the California PACT (Performance Assessment for California Teachers) initiative, designed to respond to 1998 state legislation (California SB 2042) aimed at ensuring that schools employ only "highly qualified" teachers. SB 2042 required that teacher education institutions use a teaching performance assessment certified by the state Commission on Teacher Credentialing to align with California's Teaching Performance Expectations. Twelve California colleges and universities created the PACT consortium to pilot and implement Teaching Event portfolios in this context, and in 2002-03, more than 600 candidates (and over 125 trained scorers) piloted PACT Teaching Events in the areas of elementary education, English/language arts, history/social science, mathematics, and science. For each PACT Teaching Event, or unit of instruction, candidates complete several entries including: teaching context narrative, lesson plans, 1-2 videotapes of subject-specific teaching and learning tasks during the unit, analyses of their students' work, and written reflections. These assessments are presented in terms of a Planning, Instruction, Assessment, and Reflection (PIAR) model whereby the teacher certification candidate uses knowledge of students' skills and abilities, coupled with content knowledge and pedagogical knowledge—how best to teach that content in planning, implementing, and assessing instruction. In 2003-04, all 12 PACT institutions continued using the Teaching Events, 1000 candidates piloted PACT Teaching Events, and pilot uses of DIVER were deployed in this context. Unfortunately, California's budget crisis has put the SB 2042 requirements on hold, and with it, our prospects for undertaking these collaborative studies with the necessary funding.

Nonetheless, there is great promise inherent in California's development of the PACT Teaching Events, as it is consistent with a national trend in which several states and national organizations have created performance-based assessments for teachers including teacher-developed portfolios, classroom observations, and structured interviews (i.e., the Educational Testing Service [ETS], the Interstate New Teacher Assessment and Support Consortium [INTASC], the National Board for Professional Teaching Standards [NBPTS], and institutions seeking accreditation from the National Council for the Accreditation of Teacher Education [NCATE]). PACT Teaching Events could eventually be used in combination with the NBPTS assessments to establish a professional continuum whereby teachers have comparable assessment and professional development experiences throughout their careers. These assessments have many common aspects in being standards-based, content-specific, and involving videotaped instruction, student work products, and reflective analyses of planning, teaching, and student learning.

We hope to have the opportunity to experiment with DIVER use in embedded assessments during teachers' coursework, which would provide an opportunity to capture additional information about candidates' teaching fitness as they proceed through their teacher education program. Such work would provide the occasion for investigating how digital video provides value added over text in the case studies used today (Shulman, 1992). There is relatively little research about how video works in learning: What is the process by which teachers learn from digital video case materials about what types of

events it is important to notice, and how to reason about what should be done in such situations, in ways that lead to improvements in their teaching actions and in student learning outcomes? Our plan in the PACT work was to have teacher candidates use a modified version of DIVER to reflect upon and annotate the videos and other media that they submit as part of their PACT Teaching Event performance documentation. DIVER now presents a blank worksheet to be populated with video clips and annotation, but it can be modified to establish a scaffolded worksheet requiring teacher candidates to identify, through video examples and affiliated media, including a description of their teaching context, lesson plans, student work samples, and written reflections, those facets of their teaching that satisfy criteria or tasks that are part of the Teaching Event.

This approach builds on a model of the teacher education process as developing prospective teachers' ability to think analytically and reflectively about their teaching performance (Schön, 1987; Grossman, 1990; King, 2002; Zeichner & Wray, 2001). Locating specific classroom events in terms of a body of professional knowledge about teaching is a crucial capability in integrating the general content of teacher education into consistently effective practices of teaching. A performance assessment system thus provides an evaluation of the teacher candidate's ability to both construct an effective teaching performance and to reflect upon that performance in professional terms.

This design approach closes the gap between the assessment criteria and the teaching fitness evidence—the video record of the teaching performance and supporting materials of plans and student work. Evaluators would see a worksheet populated by the candidate with clips, document highlights, and text annotations that express precisely how the candidate believes that he or she has satisfied PACT assessment criteria. The evaluator would use the DIVER worksheet to quickly but accurately assess the candidate's understanding of the standards and his or her performance, and can easily refer back to the full video record of the Teaching Event to settle questions that arise.

7.4 User Studies

A large field of use for video is in the human-computer interaction field, a design-oriented discipline in which records of users interacting with software, information systems, or computing devices are analyzed to understand problems users are having that could be improved through user-centered design accounting for forms of support and resources given discoveries about the user's information needs during tasks (e.g., Allen, 1996; Dourish & Button, 1998; Ehn, 1988; Laurel, 1990; Schuler & Namioka, 1993; Preece et al., 2002; Winograd, 1996). The concept for using DIVER in this line of research is tightening the loop between design prototyping, use study and iterative re-design driven by what is learned from use studies. Participatory design methods could also incorporate stimulated recall sessions with users who reflect using DIVER on video captured of their use of prototype or new technologies in context (inspired by Davis & Masten, 2003; Iacucci et al., 2002). The DIVER researcher can capture video, and after uploading it to the WebDIVER communication server, can either create dives in the DIVER desktop application or directly on streaming video from the WebDIVER site through a web browser. These dives could highlight specific facets of the user interactions that reveal problems, and with their web accessibility could immediately enable the designer or design team to review the dive to learn about problems with their design and, if offered by the diver, suggested resolutions.

7.5 Scientific imagery, visualization and models

We have begun to explore DIVER as a tool for diving ever more deeply into the study of visual phenomena apart from human behavior and interaction. The different kinds of applications that we see as fertile fall into a number of different higher-order classes, with extensive branching below in the hierarchy. For each class, we see utility in two broad types of DIVER use—for static and dynamic (time-varying) materials. The higher order classes I will differentiate here are: (1) image analysis at any scale, (2) visualization analysis at any scale and (3) visual model or diagram analysis.

Geo-spatial and semio-spatial representations. An additional distinction of note is that between representations that are "geo-spatial" vs. "semio-spatial " (see Roschelle & Pea, 2002). Both class (1)—*images* (such as photos)—and class (2)—*visualizations* (such as geographic maps)—are geo-spatial. But class (3)—*visual models and diagrams*—also important scientific representations, are commonly semio-spatial. We define geo-spatial representations (geo = "of the world") by formally specifiable mapping functions from measurable spatial parameters of the physical world (distance and direction, as in terms of height, depth, width) and their representational system counterparts (i.e., inscriptions: e.g., 2D and 3D maps, drawings, photos). These mapping functions are not always isomorphic, as in the renowned Mercator projection for the earth's surface, producing distortions at the poles. Geo-spatial representations are geo-gridded, a common property of many scientific visualizations (e.g., Goodchild, 1992; Gordin & Pea, 1995; Pea, 2002), in that they incorporate latitude and longitude grid "cells," which may vary in their resolution from coarse to minute.

The important differentiation for semio-spatial representations is that their spatial attributes *are not mappable* to spatial attributes of the physical world. Consider several familiar semio-spatial representations by which we may illustrate this important difference: *graphs* of all sorts, and *concept maps* (which represent taxonomic and other typological relations in the form of directed graphs), *flowcharts,* and *organization charts*. Yet, importantly for the human activity of guided noticing, both geo-spatial and semio-spatial representations are usefully exploited for supporting reasoning, argumentation, and deictic functions that are important for establishing co-reference and attentional alignment of different individuals in their communicative purposes.

With this distinction as backdrop, let us now consider the three categories of diving for scientific inquiries that I believe are useful to distinguish: (1) image analysis; (2) visualization analysis; and (3) visual model or diagram analysis.

Image analysis at any scale. Whether one is looking at a nanotechnology scale at atom assemblies or a cosmological scale involving astronomical images from the origins of the universe, the guided noticing framework of DIVER provides a powerful facility for establishing referential common ground—whether among scientific experts or in educational communications. We have been looking at DIVER use for a broad range of different scientific applications involving static images: here we will briefly describe prospects concerning medical, astronomical, and satellite imagery.

Consider **medical imagery**. Whether images are captured with microscopes, x-ray machines, CAT scans, f-MRI technology, or simply cameras, the images are complex and multifaceted in nature, revealing differentiation and structure. In radiology, a common issue is diagnosing whether a tumor is revealed, say in a mammogram (see *Figure 8*):

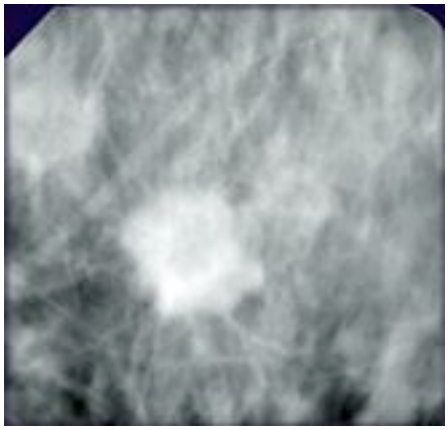

Figure 8. A malignant tumor revealed in a digital mammogram (Electronic Radiation Laboratory.

Use of guided noticing in DIVER provides a focusing and naming mechanism for depicting the malignant organ part in a radiological image, and a DIVER collection of clear cases of tumors, clear cases of non-tumors, and differentiating properties of boundary cases (tumor, non-tumor) could be used for training up diagnostic vision among mammogram analysts.

Consider **astronomical imagery**. In the Sloan Digital Sky Survey (http://www.sdss.org/, http://skyserver.sdss.org), the field of astronomy is collaboratively developing an exceptional resource for aggregating all observations and imagery of the skies visible from the Northern Hemisphere for advancing scientific discovery in the field, with over 100 million astronomical objects already logged in over 200 Terabytes of data. The virtual observatory in development as SkyQuery (http://www.skyquery.net/) enables interactive queries of astronomical data from the SDSS, so that one can zoom in or out, investigate adjacent phenomena, link through to publications on the astronomical object referred to and so on. *Figure 9* illustrates the exquisite spiral of an astronomical object called "NGC 3521."

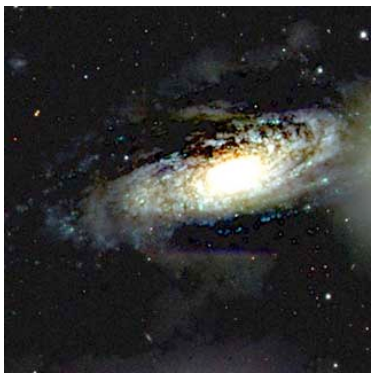

Figure 9. The famous "spiral" astronomical object NGC-3521 from the Sloan Digital Sky Survey (http://skyserver.sdss.org/dr1/en/tools/places/page2.asp)

We can foresee DIVER use for assembling a database of dives to illustrate for educative or research purposes multiple examples of a particular astronomical categorical type (e.g., spiral galaxies, edge-on galaxies, interacting galaxies), where particular differentiating features of these entities are established using DIVER's path movies, guided noticing mechanism, and annotations.

Finally, consider static **aerial, satellite and topographical images of the earth's surface**. As in the case of the SDSS, a vast online atlas has been developed—the TerraServer website (http://terraserver-usa.com/ provides United States imagery and http://terraserver.com/ offers European and Asian imagery)—which integrates over 15 terabytes (TBs) of aerial imagery and 1.5 TBs of topographic maps from the U.S. Geological Survey. These photographic images, one per location and averaging five years old, are captured with medium resolution cameras so as to identify geographical features (not people, autos, or houses), as illustrated in Figure 10.

Figure 10. An aerial photo of Stanford University Quad from 1991, downloaded from the TerraServer, with 2-meter resolution.

Using such images with DIVER, one can create "dives" for a variety of purposes involving geographic information. Urban planners and landscape designers can depict flow routes and landmarks, and create dive commentary on their interrelationships—for historical teaching, for designing based on prior art (or overcoming its limitations), for discussions with citizenry to explore issues and environmental impact concerns of anticipated changes in a city or town's landscape.

My intent here is be illustrative in sketching medical, astronomical, and geographical imagery that is static in nature but nonetheless provides a rich resource for diving with DIVER, across the expert research, diagnostic use, training, and expressive communicative use categories that were distinguished earlier.

Although I have been focusing here on the values of diving on *static* imagery, dynamic, animated sequences of images, at different time scales, can yield crucially new and different kinds of information to support theory-driven inferences about developmental, causal, and other relationships that lie latent in the static image on its own. In my work during the 1990's on scientific visualizations for learning, I was struck by the power of a time-lapsed video sequence of San Francisco skies captured by the Exploratorium Science Museum; it revealed different strata of clouds that were moving in different directions at wildly different speeds, an effect I had never observed in the normal time-course of experience. MIT's Harold Edgerton, through his invention of high-speed photography and film using stroboscopes, brought a powerful new machinery to "seeing the unseen" in the visual world, from hummingbird flight to bullets bursting balloons to the atomic blast (Bruce, 1994; Edgerton et al., 2000). When coupled with DIVER's use of the virtual camera guided noticing mechanism, science investigations of dynamic imagery provide an exciting arena for further development and research.

Visualizations at any scale. While still geo-spatial in their fundamental nature,

scientific visualizations (as distinct from imagery), utilize a geo-gridded framework and superimpose upon it representations of data values for variables that have been either directly captured through observational instruments or inferred in terms of a model for interpolating between the distinct locations where data is collected. A reference point is provided by the commonly experienced graphics from the newspaper *USA Today* and TV news weather reports: the use of a false-color map representing the range of temperature in degrees Fahrenheit or Celsius on a geographic map. While this weather data is useful at a particular point in time, for reading out the actual or predicted temperature at specific locations on the map, it is only from animating a sequence of such visualizations that the patterns of change that indicate 'fronts' (highs and lows of pressure that influence temperature and precipitation changes) become evident, among other emerging interrelationships between the data quantities visualized.

Scientific visualization dramatically changed scientists' practices in recent years by "using vision to think" (Card et al., 1999), exploiting human visual pattern perception for complex investigations within large datasets (e.g., Brodie et al., 1992; Cerf et al., 1993; Nielsen et al., 1997; Tufte, 1997; Wolff & Yaeger, 1993). Such visualization provides an image rendered through high-speed computer graphics that is based on a numerical dataset that describes some quantity in the world (e.g., global temperatures). The field of scientific visualization came to broad visibility after a major report provided an integrated account of the productive collaborations that were emerging from the disciplines of science, computer science, and the visual arts (McCormick, DeFanti, & Brown, 1987). Such scientific visualizations commonly leveraged uses of color, shape, and motion to provide new visions into the structure and patterns that could be teased out of large and complex data sets. Gordin and Pea (1994) observed that scientific visualizations are generally characterized by how they: (1) incorporate massive amounts of quantitative data, (2) aim for verisimilitude with the phenomena they represent, (3) attempt to represent entire phenomena holistically by interpolating from data, (4) employ color and shape to encode the magnitude of variables, (5) use animated sequences to show progression over time and (6) rely on high speed computation to generate images.

Gordin and Pea noted that scientific visualizations, while often similar to digital photographs in containing a set of values that can be rendered by mapping each number to a particular color, vary from digital imagery in that the numerical values composing a scientific visualization image commonly represent abstract quantities (e.g., precipitation), not formed from the variable intensities of visible light used in photography. A color range is defined according to which various values of precipitation are assigned specific colors, so that a geographic map can depict a digital image of variation in precipitation across the region displayed. The variations and patterns of color allow one using scientific visualizations to discover underlying patterns more straightforwardly than by scanning over data tables for different geo-locations. They may reveal to the human senses what could readily go unnoticed in a typological representation such as a data matrix or function. Underlying processes may also become visible if successive images over time are animated. By design, scientific visualizations strategically engage color and motion in order to tap the human visual system's capabilities.

In our CoVis Project (Learning through Collaborative Visualization, see Pea, 2002; http://www. covis.nwu.edu) at Northwestern University, and the related WorldWatcher software development and curriculum projects (Edelson & Gordin, 1998; Edelson,

Gordin & Pea, 1999; http://www.worldwatcher.northwestern.edu/), we developed a powerful open framework in which static and animated visualizations could be created for project-based scientific inquiries from scores of public domain geo-gridded datasets as well as personal data. WorldWatcher provided a learning and inquiry environment for students to explore datasets created by the scientific community (and to also create their own data using built-in arithmetic operations and climate models) while providing key features for data display and analysis commonly used in the powerful, general-purpose visualization environments of scientists. For example, the mapping of variable values to colors is used to display 2-D global temperature data (as in *Figure 11*) on a geographical information grid with latitude and longitude border marks and an overlay of a

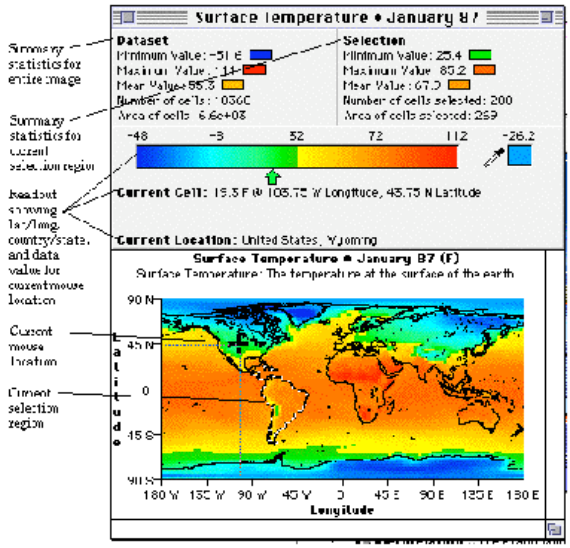

Figure 11. A WorldWatcher visualization window with key to what is represented.

continent outline (optional). When a user interacts with the WorldWatcher visualization, a continually updating readout tracks the user's mouse location and displays current latitude, longitude, country or state/province, and data values.

In their inquiries, users may tailor visualizations by modifying color-scheme, mapping of colors to numerical values, spatial resolution, and magnification. Statistical summaries are provided for regions a user selects, for whole maps by specifying geographic areas (by name), or data values (by ranges). WorldWatcher also offers select operations for mathematical data analysis within an image so that users can add, subtract, multiply, or divide all the values in a region or an entire image by a constant (as well as normalizing values for an image). WorldWatcher users can apply a function at each location in two different images using any of a set of binary operators— addition, subtraction, multiplication, division, maximum, minimum, and correlation—which then yields a new visualization as a result.

WorldWatcher datasets for global climate change for examining the transfer of energy through the earth-atmosphere system (e.g., absorbed/reflected solar energy, surface temperature) are complemented by human and physical geography data so students can examine sources and impacts of climate change. Global climate data on precipitation and relative humidity are provided, as are physical geography data on elevation/bathymetry, soil type, dominant vegetation and ground cover, and plant energy

absorption (FPAR). Students may also access global information on population magnitude and density, carbon emissions, and national boundaries.

What is the relevance of WorldWatcher for DIVER's guided noticing framework? It is the pre-eminent example of scientific visualization software tools and datasets used for pre-college education. Yet many practical challenges affiliated with its use in an educational setting arise from the difficulty of establishing common ground to what is being referred to in a learning conversation between individuals that has WorldWatcher scientific visualizations as its source. WorldWatcher users have powerful display and analysis capabilities, akin to those of scientists. They also have very useful notebook facilities that enable users to author text, and paste in visualization window results, animations and hyperlinks to specific visualizations. But even with these facilities, the complex visual arrays of information represented by its visualizations cry out for a complementary tool of guided noticing such as DIVER could provide.

How might this work? WorldWatcher already supports the creation of animations in the form of Apple QuickTime movies from a meaningful sequence of visualizations. Such animation capability is especially useful for time-series data, as in the 12 different visualizations of global temperature data (or insolation), one per month, so as to highlight the patterns of temperature (or incoming solar energy) and their distribution over the earth's surface. But a learner can easily need to refer to specific time-space pathways through such an animation—as DIVER would support for such animated movies—rather than the animation as a whole representing the whole earth's surface. These movies can only be played within the WorldWatcher application today, but we can see how diving into both the dynamic and static scientific visualizations that WorldWatcher enables users to create could provide a crucial enhancement to their meaning-making and expressive capabilities concerning their beliefs and evidence for them in the data analyses that they are constructing. These same arguments apply to other scientific visualization environments that may be used to greater effect for learning and establishing common ground in conversations by use of DIVER's guided noticing capabilities.

Visual model or diagram analysis. In Section 7.5, I observed that whereas both scientific imagery and scientific visualization are geo-spatial and turn in part for their meaning on the mapping between the spatial coordinates of the world that they represent and features of the representation, visual models and diagrams do not necessarily share this property. Instead, they are semio-spatial, in that the use of spatial distinctions in the external representation (or inscription, as Latour would call it) has a utility defined by social convention but that does not map onto a geographical grid according to any mathematical function. The spatial relations depicted in the visual models or diagrams are often not to scale (e.g., diagrams of the planets in our solar system never are). Furthermore, visual models and diagrams often introduce arbitrary conventions for the uses of space; as in Cartesian x-y coordinate graphs, or flow charts in operations research or computer science.

Concept maps are used as a technique for representing knowledge in graphs, or networks of concepts, which are made up from nodes and links that connect the nodes. Typically, nodes in concept maps represent concepts, and links represent their interrelationships (e.g., causes, is-a, is-part-of, son-of, comes-after). These representations are often used as an instructional technique but also for learning assessment and for brainstorming in a group (e.g., Lawson, 1994; Novak, 1998). It

should be clear from this characterization that the spatial relationships in concept map representations are not geo-gridded in nature.

This category of representation, as with the categories of scientific imagery and scientific visualizations, encompasses a great plethora of examples. And many of the tokens of this type also have their static and dynamic, time-varying types (e.g., one can play out different successive states in such a diagram model of a set of conceptual relationships, as in moving an object progressively toward a concave mirror and depicting when the image "flips" in orientation). In many scientific fields, visual explanatory models are established that are semio-spatial in structure and not intended to convey spatial mapping onto the world's coordinates. These models may be two-dimensional, three-dimensional, and four-dimensional (with time as the fourth dimension), and can encompass much harder to visualize 4+n dimensions.

There are hybrid representations that incorporate both geo-spatial and semio-spatial aspects, as in physics force-diagrams, particularly when used in educational software programs such as *Interactive Physics (*as in *Figure 12*) or *Working Model* (http://www.krev.com/), or in simulations for teaching physics such as the Java applets of the Constructing Physics Understanding Project (http://cpucips.sdsu.edu/). In such

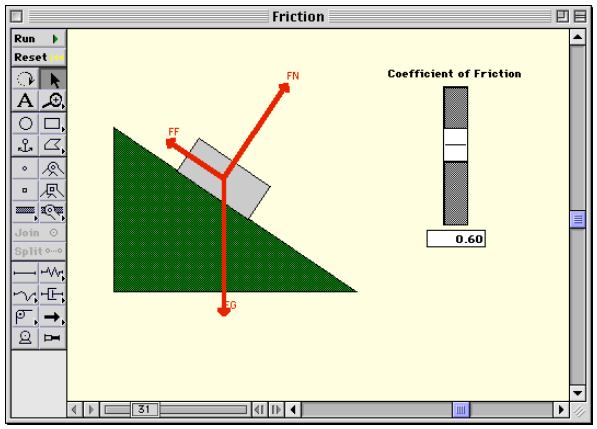

Figure 12. Physics diagram illustrating geo-spatial and semio-spatial properties.

examples, representations of conceptual entities such as forces are superimposed on *schematic* diagrams of a *general* type of physical situation, which has spatial properties although how they map onto any specific geographical space is not of primary concern; they are intended to represent the general case of such a situation, wherever it is embodied in the world. Other similar examples are diagrams used to depict ray-tracing in geometrical optics to explain how images are formed (Pea, 1992, 1994), energy diagrams depicting how heat, light and sound are propagated through different materials (Linn & Hsi, 2000), and schematics of electrical circuits and how current flows in them with batteries and bulbs.

As in the case of scientific imagery and scientific visualizations, we would argue that DIVER's guided noticing provides an important capability for use with such representations, whether for purposes of instruction, assessment, or the needs of the expert in the field communicating about their beliefs or results to colleagues. In many of these complex representations, specific part-to-part relationships are the subject of a commentary or interpretation, and it is difficult to refer at the 'whole model' or diagram level. While pointing arrows or circled regions are commonly used to refer to a specific

component of a visual model or diagram for labeling or interpretation in the acts of communication or teaching, it is challenging to make reference to multi-part connections within such representations—and harder still to respond to them in a precise way.

These issues could potentially be resolved by the use of a DIVER worksheet that takes as inputs either static visual models or diagrams, or their depicted changed states over time, and in use scenarios where the user dives on these representations to make a point or raise a question that involves connecting diverse components within the DIVER input representation.

7.6 Art, Photography and DIVER: The Virtual Docent

What does the art expert see in the works that he or she views? How do experts disagree about what there is to see in a given painting? Most of us have toured an art museum and been led on walks through exhibits of paintings, sculpture, photography and other cultural artifacts. The docent who leads such tours, whether a seasoned expert art critic or historian, or simply a knowledgeable hobbyist who works for the institution, has a stock of professional vision that they bring to bear in this social practice. They know where to focus their attention on the figures or features of works that bear close scrutiny, and how to use the categories of interpretation and interrelationships among artists, periods, historical events, and other factors to provide any one of a number of hermeneutic accounts of what is in view. There are many art books in print, as well—exemplified in works such as John Berger's (1972) *Ways of Seeing*—that seek to provide support for others to learn to notice properties of artistic works and how they illustrate specific techniques, influences, or emblematic characteristics of a school such as Surrealism or Impressionism. Picasso's painting Guernica is a favored example of a famous and influential work requiring such interpretive gymnastics to bring its senses to the viewer (e.g., Martin, 2002).

DIVER makes it easy to take such transient practices as the art museum tour, today repeated again and again and primarily face-to-face in the art museum itself with original works, and to create "virtual docents" (a concept suggested by Rayward & Twidale, 1999, among others). Indeed, the WebDIVER database established from art analysts' dives on media recordings of art originals (or scanned versions of their reproductions) may also encourage a form of comparative activity in which multiple interpretations of an artistic work could be established, and form the grist for either online or seminar-based reflective discussions. Art museums could provide docents with DIVER to use to craft reflective tours using guided noticing and annotations for remote visitors of different ages and interests, not only a one-size-fits-all form like today. And more imaginatively, it would be possible for WebDIVER visitors for art museums to experience different dives about a significant work and make their own comments and ratings of these dives, much as user communities today can do for books in Amazon.com or sellers in Ebay.com. Then one could come to view, depending on one's preferences, only the dives of greatest perceived value for any given art work.

The same categories of use we have seen for videorecords of teaching practice and scientific imagery and visualization apply in this case. The art expert can use DIVER for professional analyses; the art instructor may use it for guiding the development of professional vision for students and affiliated assessment activities, while the public may engage in informal communicative exchanges surrounding their interests in art.

As we enter a new media era in which more and more digital representations of artistic works are being established so that a world of computer users with Web access may experience them in some manner from afar (e.g., http://www.archimuse.com/), the needs of being able to refer to specific features of these works as an anchor for annotations and interpretive discussions will also grow. The Getty Museum among others has begun to recognize this need in its use recently of a web software-authoring tool called *Zoomify* (http://www.zoomify.com/), which makes it possible to interactively pan and zoom atop still images (http://www.getty.edu/art/exhibitions/flemish/zoom.html), akin in spirit to the DIVER guided noticing mechanism.

With the development of methods that extend the user experience of an art work temporally by providing interactive methods for exploring its representation in a digital medium, such as interactive 3-D depictions of sculptures that one can virtually walk around, or interactive panoramic static imagery of large spaces such as churches or city streets captured in immersive media like Apple QuickTimeVR using 360-degree rotational filming techniques (e.g., http://www.multimedialibrary.com/), we are increasing the challenges of making precise references to the facet of the art work for interpretive focus. What DIVER can bring to these interactive art experiences is a storytelling, guided noticing technology for one to be able to re-purpose those fascinating representations to author one's own point-of-view perspective on the artistic works thus represented.

7.7 Film and Cultural Studies: Motion pictures, documentaries, television and advertising

If the need for guided noticing to establish common ground for interpretive discussions and teaching of professional vision is evident for static art works, it is certainly a paragon need once we enter the semiotic realms of film and television. Whether the medium is fictional, documentary or art film in focus, a soap opera show or simply a television ad during the SuperBowl, the path movie techniques of DIVER provide an indispensable tool for film and cultural studies.

Film school students spend considerable time studying major filmmakers, film genres such as *film noir* or *new wave cinema*, the grammar of cinematography (Metz, 1974) including shot segmentation, camera rotational movements (*pan* on x-axis, *tilt* on y-axis, and *dolly* on z-axis) and transition effects (cut, fade and cross-fade) as well as narrative techniques such as montage and flashbacks, and the new computer morphing and animated special effects that have defined recent developments of this cultural medium, as in *The Matrix* films. And then they create film works to bring to light their own creative talents, inevitably grounded in how they come to be interpreted by others in the works of an earlier era that created the categories we now view as significant (when we "see," for example, the use of Alfred Hitchcock's techniques for establishing terror in the films of Brian DePalma).

DIVER provides a new tool for the film school faculty member and the film student to develop the hermeneutical talents and web of perceptive knowledge that ties together the history of films, filmmakers, film methods and techniques and film criticism. Being able to create annotated dives about a film, its makers' works, and the perceptible influences of other film works would be an exceptional resource for the film industry and filmmaker education.

The practical challenge is to establish the digitized films for use in DIVER software and to ensure that the digital rights are preserved for the film owners. It may also be controversial among filmmakers to have people diving on their films, for derivative works resulting from such techniques as the colorizing of black and white films by Ted Turner on TNT Cable have not been popular. But because of the focus plus context interface for DIVER film analysis (i.e., the full film view is always available so that the virtual camera focus window contents can be seen in context), and the fact that one is not creating new movies during Diving but simply making a replayable view onto the original movies, it can plausibly be argued that DIVER is providing a kind of conceptual spectacles for seeing movies, and not making derivative works.

Although I have emphasized film, the same points apply for media-based studies of popular culture through mass communications including TV programs, news and advertising, whose forms and techniques also have cultural histories and interpretive needs for guided noticing (e.g., Goffman, 1979; Williamson, 1994). There are also many thousands of historical film archives among the million digital artifacts represented in the American Memory Project of the Library of Congress (http://memory.loc.gov/), and the Internet Archives (http://www.archive.org/), which could be used in DIVER to illustrate specific genres, attitudes, behaviors, or phenomena of a particular historical period.

7.8 Sports, concerts, travel and other forms of entertainment

Films and shows of sporting events, musical concerts, plays, dances and other performances provide compelling media for DIVER use as well. For example, even with the rapidly growing prevalence of people on chat channels and online forums discussing television shows even as they are being watched, it is impossible to anchor these discussions in specific time-space segments of particular shows. Viewers use awkward circumlocutions to tie their questions, approvals, and complaints to specific parts of these media events. DIVER would allow not only video-anchored discussions of popular media from broadcast television of sports and music and other events, but path-movies and annotations that reference very specific segments of what happened for purposes of discussion.

Sports events are becoming an increasing realm of technological innovation to enhance the user experience for that event—new interactive television methods provide a pay-on-demand user with the capabilities to switch among different cameras trained on a football game (from the coach to players wearing helmet cameras, as in *OpenTV's* service for over 20 Million World Cup Fans in 2002 enabling viewers to select camera angles, and view statistics on players and teams). And *DartFish*, a Swiss company with the slogan "Making the invisible visible," uses video image processing to provide dramatic image overlays of different skiing trials during the Winter Olympic Games and other sporting events. In another example of interactive applications, Dartfish's 3D Visualization technology made its TV debut during the 2003 Alpine Ski World Championships. It provides a 3D virtual reality landscape of a real racing environment by precisely reconstructing the course traveled during a race by each competitor and then situating them in the 3D landscape so that spectators can experience an unrestricted view of the actual race, watching it from any angle at any playback speed. But much more interactivity is possible if the sports lover could dive on different video records of the sporting event and send links to their uniquely authored dives to friends and others—their perspective, path, and commentary on the dynamic events of sporting, from the great play

to the botched catch. Such an application of DIVER, if it could be established among the sporting program producers, could become a new cultural medium for sharing sporting experiences.

Another example is the television travelogue—the tour to Rome, the Yucatan's Aztec ruins, nature experiences in the jungles or African Sahara, or Darwin's Galapagos Islands. Today's travelogues tend to be monologues, but with point of view authoring many different travelers could utilize a database of panoramic video (or other videos) and make their own DIVER movies to communicate about what they found compelling. These dives could be share as DIVER video blogs with family and friends at home but more broadly they could be posted online with the likelihood of stimulating greater tourism, and be classified according to the demographics and interests of those who are making the dives (show me Cairo from the point of view of the ancient Egyptian scholar). The advent of videocamera cellphones could accelerate this vision substantially.

7.9 Training of all kinds

DIVER's guided noticing approach for any user to author and annotate path movies within movies for web-publishing and reflective discourse has potentially broad applicability for training in virtually any work sector. Typically learning is affiliated with the development of professional vision and comparative discrimination of higher and lower forms of performance that learning from dives created by instructors may enable. Of course watching Tiger Wood golf videos is no substitute for playing golf oneself, and the importance of learning from comparative analyses of one's own performances and those of an expert is well-recognized already in commercial video software tools for sports training available from such companies as SportsTec, Dartfish and Pinnacle. Enhanced perception in a professional sphere alone will not usually suffice for expertise (with judging of sports such as gymnastics, skating and diving providing interesting exceptions)—it is the application of that enhanced professional vision when coupled with the actions appropriate to expertise in the domain that is the intention of video-enhanced instructional processes.

It is perhaps less well-known that video is extensively used for high-stakes training in applications such as mock jury trial preparation and witness deposition, police handling of crisis events, and making effective sales closings for costly products (e.g., ultra-fast Internet backbone switches). Whether in applications such as these, or in routine forms of workplace or corporate training ranging from running quality meetings to using factory machines correctly, providing DIVER as a tool for simply creating instructional dives on in-situ performances, and enabling diving on the learner's own performances could serve on-the-job training.

8. Towards Digital Video Collaboratories

In the last section of this chapter before considering future directions, I will outline our approach to developing an infrastructure for integrating video analysis, sharing and collaboration in research communities, although it can have broad applicability in the other cultural spheres sketched throughout this chapter.

8.1 A Digital Video Collaboratory user scenario

The following user scenario of behavioral scientists using video as data illustrates the capabilities we are developing with our colleague Brian MacWhinney of Carnegie Mellon University to enable a new form of scientific dialog based on detailed

examinations of video. Using current practices and tools, the scientific work described below would not be feasible. Consider a scenario where a researcher learns about a new finding in a peer-reviewed journal in her field, and has a competing interpretation for the phenomenon (e.g., four-year olds use of their fingers in solving counting problems). She searches the Digital Video Collaboratory for examples of such counting phenomena and finds six relevant datasets, physically located in video repositories at a number of different universities, all contributed to a distributed data registry when other researchers provided open access to their data when they published their related work. She analyzes these cases of hand gestures children used during counting by exploiting an existing coding scheme for video analysis of hand-gestures, loading it into the DIVER interface as a plug-in. Her hunch appears warranted; working from these open source data, she finds significantly more gestures supporting her theory over the claims of the published work she has been reading. She and her students collaborate to develop a dive—a worksheet containing annotated video selections from the datasets she has found through the Collaboratory—that sets forth her theory of hand-gestures, and she contributes the dive to the Video Collaboratory. She sends an email with a hyperlink to her group's new video analysis to the journal article's author and a network of other researchers who are interested in children's counting behavior.

The Video Collaboratory e-print posting of her claims as a commentary to the published work sparks a lively discussion, engaging the article's author, who dissents from her conclusions on the gestural data in children's counting. He responds by reinterpreting the same video cases that she has analyzed, and after several cycles of data analyses, interpretations and researcher communications, she submits her team's revised e-print as a formal submission to the journal, hyperlinking it to the digital video records and dives comprising their analyses. During the article's peer review process, one reviewer calls for better reliability data for her gesture categorizations from the videos; and she responds with a new round as part of a revise-and-resubmit cycle with the journal's editor. Soon her paper is accepted and published, with its multimedia data analysis history in the Video Collaboratory made accessible to the interested readers.

This scenario illustrates how a number of central research activities will be fundamentally enabled by a digital video collaboratory infrastructure. After outlining our conceptual analysis of this future, I will briefly sketch the technology we are building for enabling this new mode of research.

8.2 Three obstacles to a Digital Video Collaboratory

Research communities constitute the social infrastructure for scientific knowledge building, as the human networks through which data are shared among colleagues located at different institutions, research results are circulated and critiqued, journal editing positions are staffed, and manuscripts and publication decisions are vetted. Enabling these communities to use video, not only text, requires removing three obstacles:

1. Video data and analyses must be universally accessible independently of its physical location. Today video data reside in many heterogeneous repositories, with separate access control, user interface, and querying syntax. We aim to remove this obstacle by developing a virtual repository, building on grid-enabled middleware in combination with an API (application program interface) and metadata scheme designed to support research use of video as data.

2. Research communities must have access to video analysis tools that support discipline-specific analytic practices of annotation, coding, and reference, while enabling communities to make use of existing applications and infrastructure where possible. Today video analysis tools make few provisions for interoperability, forcing researchers to commit to one tool or toolset rather than allowing use of multiple tools. We seek to remove this obstacle by developing generic and discipline-specific XML metadata schema for video analysis, and flexible desktop and Web-based video analysis tools.

3. Analyses and commentary must be available for public participation, collaboration, and publishing, including circulation and feedback among smaller groups of researchers. It is extremely difficult now to publish video analyses through the Web, and we lack a platform for sustaining video-based collaboration and critique. We are removing this obstacle by developing DIVER as a Web-based video analysis and dive-authoring tool.

By removing these three obstacles, we will enable the creation of a Digital Video Collaboratory (DVC): a network of shared video resources, analyses, and collaboration opportunities located within scientific research communities. The following sections briefly address each of these DVC elements in turn.

8.3 Virtual repository for video data and analyses

We use the phrase "virtual repository" to characterize a distributed set of heterogeneous video and metadata repositories that appear to users and client tools as a single searchable repository. A virtual repository is a key element in a Collaboratory as it provides a research community with a touchstone corpus of empirical materials and analyses to which members may have access (Berman et al., 2003). There currently exists no such virtual repository for video data in the human sciences, although the *Open Video Project* has developed a large testbed resource for digital video research work on such problems as automatic segmentation, summarization, creation of content surrogates, and developing face recognition algorithms (Geisler et al., 2002). The closest analogue is *TalkBank*, which provides a few heavily used and oft-cited data corpora (particularly audio data) in a number of language-related subdisciplines. However, TalkBank is currently a single repository with a site-specific user and software interface, requiring local knowledge, such as what kinds of materials are likely to be found.

Even if a research community could be built around a single repository, video storage requirements mitigate for distributed storage. The storage needs are vast for even 100 researchers contributing a few hundred hours of video (a common corpus size for a single study) at a variety of resolutions and different compression ratios (a 'small' corpus relative to the many thousands of human scientists using video integrally in their research). A moderate size research community will need to store and manage tens and even hundreds of terabytes of video (with petabytes and exabytes close within view). As a benchmark, consider that TV worldwide is generating about 31 million hours of original programming (~70,000 terabytes) each year, while worldwide production of filmed original content for cinema, and made-for-TV films or series is about ~25,000 terabytes per year (Lyman & Varian, 2003). The estimated global total of motion pictures made with surviving stock from 1890 to 2003 is about 370,000 and if digitally stored, would require 740.7 petabytes of storage. Several research centers already serve data from petabyte-size storage archives (http://en.wikipedia.org/wiki/Petabyte).

Insulating users and applications from idiosyncratic features of multiple repositories

requires an intermediate software layer for translating generic requests into requests that can query each specific repository and translate whatever data is returned into a standard form. We envision a middleware layer of software architecture providing repository services to client applications via a public interface, and in turn interacting with repository-specific translation components that map generic method calls for access, search, and retrieval into repository-specific interfaces. Repositories can expose their contents to the collaboratory without altering their practices for storing and retrieving video and metadata, so long as they implement a version of the translation layer for their repository.

The NSF Middleware Initiative (NMI: www.nsf-middleware.org) and the incorporated Globus Toolkit provides a solid base of common protocols and low-level services for implementing this vision using Grid computing technology (Foster & Kesselman, 2003), although it does not deal directly with video resources. Some NMI components can be used directly (e.g., Grid Security Infrastructure for access control and authentication, GridFTP for file transfers), but others will require extension for use with video virtual repositories. Treating *video as data for discovery*, in contrast to content to be located and delivered (as in Internet-2 videoconferencing or file-serving today), imposes a heavier burden on the repository and its search functionality. Delegating basic functionality to the NMI components such as user authentication and resource registration enables us to focus on search functions, perhaps the greatest challenge in establishing a virtual repository. Here the gap is widest between the ideal of a single searchable repository and the reality of repositories with heterogeneous metadata schemes, some standardized and others ad hoc. Our approach will not require repositories to re-index data to a common metadata scheme, or engage in an unrealistic translation project, but focuses on assuring that all repositories support a core metadata set, while exposing the range of metadata schemes used beyond that core for the user of the virtual video data repository.

We aim to support three different types of search, each with unique implications for metadata development and the functions of the software layer that translates between the generic virtual repository software interface and the specific interfaces exposed by each local repository:

1. *Full text search* of all metadata. The translation layer for each repository will map a generic full text search into the appropriate local search action.

2. *Core metadata search.* The repository will support a core set of metadata (for which we will adopt the Dublin Core as a first pass, to be amended as experience proves necessary) that are guaranteed to apply to all local repositories. This ensures that all resources from participating repositories have a base level of visibility. The translation layer will map the core metadata into the local repository-specific metadata and syntax.

3. *Extended metadata search.* The repository will expose to the user information about the metadata schema available across the local repositories to which the repository has access (e.g., the Gateway to Educational Materials' (GEM) instructional topics hierarchy, and pedagogy elements/values, http://www.thegateway.org/). Users can select metadata schema for use in searches, at which point only those repositories supporting the selected metadata schema will be queried. The translation layer will pass the queries through to the local repository, with whatever modifications are required to conform to

local syntax. This use of metadata has the advantage of exposing all of the metadata available, while not requiring a mapping to a common *lingua franca* outside the core metadata. It requires only that the local repository publish to the virtual repository a standard description of the metadata schema in use (e.g., using XML Schema).

Many video data repositories will provide scheme-specific metadata to allow researchers to locate classroom interactions matching specific topical and grade-level criteria. Whereas the virtual repository must support a range of metadata schemes at the level of the local repository, the development of a scientific field is enhanced by the common ground established by a shared meta-data scheme. Therefore, based on its importance to the multi-disciplinary learning research community in our work toward a Digital Video Collaboratory, collaborator Brian MacWhinney's group at CMU is extending the TalkBank XML Schema (xml.talkbank.org) in building a system for classifying interactional structures in classrooms by developing metadata characterizers based on the vocabulary of interaction analysis methods such as conversational analysis, speech act theory, discourse analysis and classical rhetoric. This metadata development will be compliant with OLAC (Open Language Archives Community, http://www.language-archives.org/) and the larger Open Archives Initiative (http://www.openarchives.org/).

8.4 Interoperable Application Infrastructure for Video Analysis

Research communities relying solely on text have access to a pre-existing, richly developed research and communications medium that is flexibly adaptable to the requirements of the communities' analytic work. Consider a Shakespeare scholarly community: they have access to the full apparatus of print publishing, word processing, and so on, while also having the capability to extend this apparatus by adding discipline-specific conventions such as shorthand notations for play titles, acts, scenes, and line numbers. This text infrastructure ubiquity is nearly invisible, yet it provides critical support for the community's work in conducting and publishing research.

Today we do not have a comparable infrastructure for supporting video analysis, publication, collaboration and critique using video within distributed communities. Existing tools maroon researchers on isolated islands of data, with no easy way to bring to bear other tools or to circulate results to other researchers.

For our Digital Video Collaboratory work, we bring together the TalkBank and DIVER projects as a concrete base for reconciling this tension. Based on over twenty years of work in the Child Language Data Exchange System (CHILDES: MacWhinney, 2000), the NSF-funded TalkBank project has constructed a full set of quantitative programs for concordance, reliability, part of speech tagging, parsing, coding, and code tracking. These interdisciplinary tools, collectively called CLAN, are designed to support numerous research communities in the area of linguistics and speech analysis, and are being rewritten in Java to work with new XML schema developed at TalkBank for the CHILDES data. As this chapter explains, our work with DIVER provides a user interface paradigm and various applications for making persistent pointing references to video source materials that can be annotated and commented upon. DIVER supports researchers in extracting elements of video in the service of analysis and argumentation for many different disciplines. The ability a DIVER user has to impose a theoretical and empirical point of view on a video source, previously available only through general-

purpose video editing software, makes video potentially as pliable and portable a communicative resource as text.

Our collaborative Digital Video Collaboratory project with MacWhinney is extending these works by:

(1) Extending and disseminating the XML metadata schema at TalkBank to provide for interoperability among video analysis applications (e.g., DIVER and CLAN initially; and later today's linguistic research tools such as SignStream, ATLAS [Architecture and Tools for Linguistic Analysis Systems], Praat, MediaTagger), including a standard method for extending a core set of functionality.

(2) Extending DIVER using an XML-based schema for representing data dimensions and categories and transcription tools that can accept multiple markup languages via XML configuration to incorporate functionality with broad relevance for the human sciences, using the CLAN software as a reference.

(3) Extend WebDIVER services through a lightweight server-based API (Application Program Interface) architecture that will enable users (and developers) to flexibly invoke a number of software components with specific functionality running in a web browser client.

(4) Extend DIVER to enable analysis of video located remotely so as to remove the need to distribute video copies for local analysis, a functionality we now have in place.

The Digital Video Collaboratory will thus contain not a single application, but a loosely coupled set of related programs that are fully interoperable because of their common use of standardized XML schema for representing metadata.

8.5 Community-Based Collaborative Video Publishing, Commentary, Critique

Our project focuses on a Digital Video Collaboratory for research communities rather than for individual researchers. This emphasis stems from the recognition that interaction, collaboration and critique are at the heart of the scientific enterprise, even though much of the visible research work happens in isolation. In the human sciences today video is generally confined to the research phase, and discussion and reporting phases are conducted entirely in text (Pea, 1999). This short-circuits the process of critique or secondary analyses, since other researchers do not have access to the underlying evidence in its original form. This state may be contrasted with fields such as *biology* (where GenBank®, the NIH genetic sequence database [Benson et al., 2002], provides an annotated collection of all publicly available DNA sequences, over 40 million as of late 2004), in *neuroscience* (see the Human Brain Project, Koslow, 2002, Gardner et al., 2003); or in the 2003 National Institutes of Health policies requiring data-sharing for all grants with over $500K direct costs: NIH-OD-03-032.

As noted earlier, there were recent efforts (both involving TalkBank) to address these problems by experimenting with linking publication and collaborative commentary on video: a special issue of the *Discourse Processes* published with a CD-ROM (Koschmann, 1999) had multiple researchers analyzing a six-minute medical school problem-based learning episode. A TalkBank conference in Fall 1999 reflected on that experience and led to TalkBank's organization and production of the CD-ROM for a special issue of the *Journal of the Learning Sciences* (Sfard & McClain, 2002), where six different researchers analyzed a video documented the role of designed artifacts in secondary mathematics learning. For that JLS special issue, authors had access to

transcripts linked to the video in the TalkBank CLAN program, final PDF versions of the articles were edited with Acrobat to create links from specific portions of text to segments of the QuickTime video, and in the final stage of CD-ROM production, authors' analyses were linked back to the original CLAN transcripts and video segments. This JLS issue represented a first case of a digital video collaborative commentary, but even here analysis was disconnected from video data, as authors referenced time codes and transcripts to re-engage with the video source. Such work cannot be done at all on the web today. When considering the challenge of presenting researcher works using multimedia data and enabling web-based commentary, the problem grows. We need the consistent and open form of data organization provided by the global XML standard.

Current video analysis tools are strong individually in various aspects, such as coding, editing, or transcription, as outlined earlier. But none of these tools directly addresses the core challenge of supporting the broader use, sharing, publishing, commentary, criticism, hyper-linking and referencing of the multimedia data produced and output by the tools. Video-using researchers need to be able to make standardized, accessible and direct contact with competing analyses of audio-video data to enable critical collaborative commentary and cross-referencing for advancing disciplinary knowledge. The integration of video analysis for the work of communities of researchers and practitioners poses technical and design issues that transcend those inherent in developing video analysis tools toward providing a general infrastructure for collaboration.

WebDIVER is designed to go beyond the posting of video data and analyses to the Web (now uncommon) so that video may be used as an interactive communication medium in the process of scientific interchange, including collaboration and critique of scientific argument and research evidence. Our first version of WebDIVER provided basic functionality for publishing, viewing, and commenting on analyses authored within the client desktop DIVER application. DIVER users uploaded their source video, analytic clips, and annotations to the WebDIVER site, which then organized the material for display to users through a browsable, searchable community-based website. Other researchers viewing a dive could respond to the author's annotations by posting their own textual comments linked to the video in question, which could then be seen by the author and by other researchers. The key challenge to this approach is that it provides only incomplete Digital Video Collaboratory features, for a researcher cannot respond to a video comment using source video, but only with textual comments.

As of Autumn 2004, we have made a major advance in providing video researchers with new WebDIVER capabilities that enable anyone to do their diving on streaming video files, so that video no longer needs to reside on the diver's own computer. Divers simply use a web-browser to mark and record space-time segments of videos and to make text annotations, and they may draw on a large archive of videos that have been made dive-accessible. Video files are encoded by users in a specific format (Macromedia Flash,.flv) and then uploaded via a webpage to the WebDIVER communications server, where our technology makes them accessible for diving. Our DIVER team is extending these developments to realize our vision of the Digital Video Collaboratory for robust access control, group formation, email notifications of changes in dives one has authored or subscribed to, and so on. We will also integrate WebDIVER with the Digital Video Collaboratory virtual data repository concept, so WebDIVER users can store and retrieve video data and analyses without regard to the underlying physical storage locations.

9. Ethical Issues

A variety of ethical issues are raised by the prospects of broad scale access to video data for research and repurposing, including issues of human subjects, security and publishing rights. The basic issues raised by point-of-view authoring of digital records for research are privacy and access issues affiliated with human subjects requirements in research policy (Pea & Hay, 2003). Unlike privacy protection for research subjects involving textual or numerical data representations of research data, educational video records potentially include substantial information about specific children and teachers, and in need of careful reviews to protect the privacy of subjects and maintain confidentiality of data whenever appropriate.

These issues are encountered most obviously in the context of protecting the rights of human research subjects in human subject reviews, where research involving video records often faces heightened scrutiny in the United States by Institutional Review Boards (IRBs) for research institutions receiving federal research funding. It is apparent that informed consent for uses of videorecords of learning and teaching can mean very different things to IRBs and their members, as different personal values come to play in judgments concerning potential risks and benefits from research participation. Many IRBs do not consider uses of research videorecords of learning and teaching problematic in their repurposing for teacher education, or in their use for illustrating learning phenomena. But some institutions consider video data to be 'secondary' and require destruction of videorecords after research studies transcribe and code such data, which makes impossible any data sharing or re-visitation of original source data for re-interpretation. Other IRBs require reconsideration of video data use for each and every repurposing, in that all individuals involved in the recordings must approve each new specific use of the videoclips.

The key issue in human subjects review compliance is disclosing foreseeable risks. The Human Subjects Panels of IRBs typically have oversight responsibility for the review of all University projects that involve human subjects in non-medical research to ensure that the rights and welfare of the subjects are adequately protected with informed consent review. Panel reviews commonly involve approval of a clearly worded consent form which assures that the subject (or responsible parent) is fully informed of the risks inherent in participation and of the benefits which might be reasonably expected. As the basic age for participation in research is 18 years, parental permission is requested for their child's participation in educational research.

There is currently little precedent for identifying the range of risks. One option would be to identify the worse case scenario, to disclose such a possibility, and to minimize the likelihood of such risks. For example, a child may become a standard example for the wrong answer to a question. Or if a specific school is identifiable in a video, it may be possible for a criminal to encounter or seek out a video of a child in that school on the Web for nefarious purposes. Correlatively, there is also insufficient common practice and understanding concerning the benefits that might be reasonably expected from uses of videorecords of learning and teaching for promoting advances in scientific understanding of education and enhanced educational practices.

The universality of Web access also leads to the question of security. One way to mitigate the risks to subjects of being included in a video is to limit access to known

users, and to ensure that the video cannot circulate beyond that group (for example, by using streaming technology rather than file downloads to make the video available). Another method in the biological and medical sciences is to provide access through a user contract to sensitive data via a controlled site, commonly called a data enclave.

The TalkBank site maintains a large collection of audio and video data, based at Carnegie Mellon University, and addresses this problem by requiring users of its collection to agree to an extensive code of ethics. This code obligates anyone using TalkBank data to avoid criticism of individuals depicted in the materials. Anyone who violates the code of ethics is subject to community censure. This code is an attempt to balance the privacy and other interests of research subjects with the scientific needs of researchers, and represents the judgment that personal criticism has sufficient potential for social harm that it should be impermissible.

Another approach is the use of alternatives to videorecordings of situated behaviors as they occur naturally in real classroom settings, such as dramatizations and labeling of performances-as-acted allowing illustrations of teachers and students engaged in both desirable and undesirable practices. In such circumstances, media releases rather than human subjects releases are used, since the point is not human subject research but media capture as in a play or movie that can then be shown to others. A central problem here is that the performed versions of learning and teaching may not sufficiently resemble the real thing so as to serve the educational and reflective learning purposes for which they are intended. Research is needed to examine this question. Furthermore, teacher educators who have worked extensively with video case studies of teaching practice commonly emphasize how crucial it is for the teachers learning from these cases that they are filmed in real classrooms, and not enacted behaviors.

Another technical possibility is the use of digital masks of identity, such as using blurring or pixelization filters for faces (e.g., Boyle et al., 2000) and digital transformation of voices to prevent identification of individuals (as in news shows protecting crime witnesses). This approach has the advantage of redacting participant-identifying characteristics of the data in order to protect privacy. It is unclear whether such a method could be made commonly useful, as research is necessary to identify the range of research questions that could best be addressed with such masking, and which could not. For example, discourse and interaction analysis frequently relies upon information about gaze direction, joint attention, and such information in the video would not be available in data records that utilize digital masking.

10. Emerging trends

In one sense the concepts and technologies reviewed throughout this chapter are already exotic with respect to everyday learning and teaching practices and uses of video in virtual learning environments. But there are a number of technological frontiers that promise to bring interesting developments worth watching.

With the increasing reduction in cost, size and network connectivity of video devices, digital video sensors and cameras will become ubiquitous, and video data will proliferate through wireless data transmission from locations throughout cities, streets and in nature. We already see the exponentially growing use of digital photography with camera-enabled cellphones, as people find it compelling to share visual experiences with friends and family. Video cameras and video file sharing are emerging as supported services in

cellphones in Japan, Korea and Scandinavia, if uncommon in the United States (as of late 2004). In principle, the video that is captured and shared through such mechanisms provides extensive media for DIVER guided noticing and repurposing.

Many research groups are also pursuing automated video and voice recognition, to enable parsing of the video streams and meta-data tagging components of scenes, such as recognizing specific people and objects, and identifying motion segments and speech conversation segments, for facilitating information retrieval and media re-use[5]. These developments will be of benefit for speeding the process of research use of videorecords, but also raise challenging problems if they become broadly used and come to endanger human privacy, as some anticipate, in surveillance and security applications.

11. Conclusions

Eco (1989) writes about modern music, writing, art and science that "'open' works, insofar as they are in movement, are characterized by the invitation to make the work together with the author and that (2) on a wider level (as a subgenus in the species 'work in movement') there exist works, which though organically completed, are 'open' to a continuous generation of internal relations that the addressee must uncover and select in his act of perceiving the totality of incoming stimuli" (p. 21).

To the extent that DIVER use can make video and other rich media 'open' to diving—the authoring of guided noticing using path movie-making and annotation—there is without question an active role for the reader, who becomes an author in bringing the work of the video or other medium to a more completed state in his or her interpretations of it. DIVER also provides a tool for evidence-based argumentation, in which one uses what one notices in the medium to make a case around it, and thus extends the work in significant ways with the act of authoring the dive.

In this chapter, I have characterized the theoretical foundations of the DIVER Project, which has created a new software system for capturing, annotating and sharing perspectives (dives) on human activities videorecorded in real-world spaces. DIVER couples a Looking overview window with a magnified Noticing window that can be controlled by the user. This viewing mechanism is also an authoring tool, allowing a user to capture and share his or her unique point-of-view and commentary on a source video. Users annotate their selections, share them on the Web, and receive comments by others, promoting dialogue and collaboration around video-records of common interest. Supporting this user activity sequence of looking at a full record, noticing events of interest and recording them for later use, and then commenting on what has been seen is fundamental for both research and learning from video sources. For the constructivist educator or more generally for those who want a more active voice in media uses for communication and knowledge production, DIVER provides a platform for moving away from today's broadcast-centric and asymmetric uses of video to the communicative empowerment of the video user, who can easily craft point-of-view movies within movies with commentaries to share with others. Whether such methods are used to elaborate collaborative knowledge building in the learning sciences from video sources or for everyday consumer video communications in the future, this shift from consumption to authorship is a fundamental transformation in the use of the video medium.

---

[5] The IEEE International Conference on Image Processing (ICIP), and ACM's SIGMMM (Multimedia) SIGGRAPH conferences are vital forums for learning about these developments.

References

Abowd, G.D. (1999). Classroom 2000: An experiment with the instrumentation of a living educational environment. *IBM Systems Journal*, *38*(4), 508-530.

Allen, B. L. (1996). Information tasks: Toward a user-centered approach to information systems. New York: Academic Press.

Anderson, R.S. & DeMeulle, L. (1998). Portfolio use in twenty-four teacher education programs. *Teacher Education Quarterly, 25*(1), 23-32.

Apple QuickTime 6, including MPEG-4 (2002). (http://www.apple.com/mpeg4/)

Argyle, M. & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, U. K.: Cambridge University Press.

Athanases, S. Z. (1994). Teachers' reports of the effects of preparing portfolios of literacy instruction. *The Elementary School Journal, 94*(4), 421-439.

Bapna, D., Rollins, E., Murphy, J., Maimone, E., Whittaker, W., & Wettergreen, D. (1998). The Atacama Desert Trek: Outcomes. *IEEE International Conference on Robotics and Automation (ICRA-98),* pp. 597-604

Barab, S. A., Barnett, M. G., & Squire, K. (2002). Developing an empirical account of a community of practice: Characterizing the essential tensions. *The Journal of the Learning Sciences*, 11(4), 489-542.

Barab, S., MaKinster, J. G., Moore, J., & Cunningham, D. (2001). Designing and building an online community: The struggle to support sociability in the Inquiry Learning Forum. *Educational Technology Research and Development*, 49(4), 71-96.

Barab, S., MaKinster, J., & Schecker, R. (2003). Designing System Dualities: Characterizing An Online Professional Development Community. In Barab, S. A., Kling, R., & Gray, J. (2003). (Eds.). *Designing for Virtual Communities in the Service of Learning*. Cambridge, MA: Cambridge University Press.

Bargeron, D., Grudin, J., Gupta, A., Sanocki, E., Li, F. & LeeTiernan, S. (2002). Asynchronous collaboration around multimedia applied to on-demand education. *Journal of MIS.* 18(4), 117-145.

Barnett, C. (1998). Mathematics Teaching Cases as Catalyst for Informed Strategic Inquiry. *Teaching and Teacher Education*, 14(1), 81-93.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3*)*, 307-359.

Barron, B. (2000). Achieving coordination in collaborative problem solving groups. *The Journal of the Learning Sciences, 9*(4), 403-436.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Voltera, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.

Bateson, G. (1936). *Naven: A survey of the problems suggested by a composite picture of the culture of a New Guinea Tribe drawn from three points of view*. Stanford: Stanford University Press (2nd Edition published 1981).

BeHere Corporation (http://www.behere.com/)

Benosman, R., & Kan, S. B. (Eds.). (2001). *Panoramic vision: Sensors, theory and applications*. New York: Springer Verlag.

Benson D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., & Wheeler, D.L**.** (2002, Jan. 1). GenBank. *Nucleic Acids Research* 30(1): 17-20.

Berger, J. (1972). *Ways of seeing*. Harmondsworth, England: Penguin.

Berliner, D.C. (1988). *The development of expertise in pedagogy.* New Orleans: American Association of Colleges for Teacher Education.

Berliner, D.C. (1994). Expertise: The wonder of exemplary performances. In J.M. Mangier & C. C. Block (Eds.), *Creating powerful thinking in teachers and students: Diverse perspectives (pp. 161-186).* Fort Worth, TX: Holt, Rinehart, & Winston.

Berman, F., Fox, G. C., & Hey, A.J. G. Hey (2003). (Eds.). *Grid computing: Making the global infrastructure a reality*. New York: John Wiley & Sons.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001, May). The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*. (http://www.sciam.com/2001/0501issue/0501berners-lee.html)

Birdwhistell, R. L. (1970). *Kinesics and context: essays in body motion communication.* Philadelphia: University of Pennsylvania Press.

Boult, T. (1999). Personal panoramic perception. In *Proc. Int. Conf. on Imaging Science Systems* (CISST '99).

Boyle, M., Edwards, C., & Greenberg, S. (2000). *The effects of filtered video on awareness and privacy*. Proceedings of the 2000 ACM conference on Computer Supported Cooperative work (CSCW2000), pp. 1 - 10. Philadelphia PA.

Bransford, J. D., Brown, A., & Cocking, R. (2000). (Eds.), *How people learn: mind, brain, experience and school (expanded edition)*. Washington, DC: National Academy Press.

Brodie, K.W., Carpenter, L.A., Earnshaw, R.A., Gallop J.R., Hubbold, R.J., Mumford, A.M., Osland, C.D., & Quarendon P. (1992). *Scientific visualization*. Berlin, Germany: Springer-Verlag.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of The Learning Sciences, 2*, 141-178.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18, 32-42.

Bruce, R. R. (1994). (Ed.). *Seeing the unseen: Dr. Harold E. Edgerton and the wonders of Strobe Alley*. Cambridge, MA: MIT Press.

Brugman, H. & Kita, S. (1998). CAVA: Using a relational database system for a fully multimedial gesture corpus. *Workshop: Constructing and Accessing Multi-media Corpora: Developments in and around the Netherlands*. Nijmegen, The Netherlands. February 1998.

Bruner, J. (1975). From communication to language—A psychological perspective. *Cognition*, *3*, 255-287.

Card, S. K., Mackinlay, J. D., Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco, CA: Morgan Kaufmann.

Carrer, M., Ligresti, L., Ahanger, G., and Little, T.D.C. (1997). An annotation engine for supporting video database population. *Multimedia Tools and Applications*, *5*(3), pp. 233-258.

Cassirer, E. (1953-1957). *Philosophy of symbolic forms, trans. Ralph Manheim, 3 vols*. New Haven, CT: Yale University (originally published 1923-1929).

Cerf, V. G., Cameron, A.G.W., Lederberg, J., Russell, C.T., Schatz, B.R., Shames, P.M.B., Sproull, L.S., Weller, R.A., & Wulf, W.A. (1993). *National collaboratories: Applying information technology for scientific research*. Washington, DC: National Academy Press.

Chen, S.E. (1995, August). QuickTime VR – an image-based approach to virtual environment navigation. *Proc. of ACM SIGGRAPH'95*, pp. 29-38.

Cherry, G., Fournier, J., & Stevens, R. (2003). Using a digital video annotation tool to teach dance composition. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*. (http://imej.wfu.edu/articles/2003/1/01/)

Chiu, P., Boreczky, J. S., Girgensohn, A., & Kimber, D. (2001). LiteMinutes: an Internet-based system for multimedia meeting minutes. *WWW 2001*, 140-149.

Clark, H.H. (1996). *Using language.* Cambridge, MA: Cambridge University Press.

Collins, A., Brown, J.S. & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. Resnick (Ed.), *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Crismond, D. (2003). Approaches to using video cases in teacher professional development. Proceedings of the The 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'03).

Cutler, R., Rui, Y., Gupta, A., Cadiz, JJ., Tashev, I., He, L., Colburn, A., Zhang, Z., Liu, Z., Silverberg, S. (2002, December). Distributed meetings: A meeting capture and broadcasting system. *Proc. ACM Multimedia 2002*, Juan-les-Pins, France, pp. 503-512.

Crowley, K., Callanan, M.A., Tenenbaum, H.R., & Allen, E. (2001). Parents explain more often to boys than to girls during shared scientific thinking. *Psychological Science, 12* (3), 258-261.

DeLaguna, G. (1927). *Speech: Its function and development.* New Haven: Yale University Press.

Derry, S. J. and the STEP Team (2003). *The STEP System for Collaborative Case-Based Teacher Education: Design, Evaluation & Future Directions*. University of Wisconsin, Madison. Unpublished report.

Doerr, H., Masingila, J., & Bowers, J. (2003). *The Case Creator Project: A video-based case creation tool for teacher education*. (http://www.sci.sdsu.edu/mathvideo/cc/)

Dourish, P., & Button, G. (1998). On "technomethodology": Foundational relationships between ethnomethodology and system design. *Human-Computer Interaction, 13*, 395-432.

Eco, U. (1989). *The open work*. Cambridge, Massachusetts: Harvard University Press.

Edelson, D. C., & Gordin, D. (1998). Visualization for learners: A framework for adapting scientists' tools. *Computers and Geosciences, 24* (7), 607-616.

Edelson, D. C., Gordin, D.N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *The Journal of the Learning Sciences, 8*(3&4), 391-450.

Edgerton, H., Kayafas, G. (Ed.), & Jussim, E. (2000). *Stopping time: The photographs of Harold Edgerton*. San Francisco, CA: Harold N. Abrams.

Ehn, P. (1988). *Work-oriented design of computer artifacts*. Stockholm: Arbetslivscentrum (Distributed by Lawrence Erlbaum Associates, Hillsdale, NJ).

Erickson, F. (1992). Ethnographic microanalysis of interaction. In M. LeCompte, et. al. (Eds.), *The handbook of qualitative research in education.* San Diego: Academic Press.

Feiner, S. K. (2002). Augmented reality: a new way of seeing. *Scientific American*, April 24, 2002. 8 pp. (http://www.sciam.com/2002/0402issue/0402feiner.html.)

Fiala, M., Green, D., & Roth, G. (2004, October). A panoramic video and acoustic beamforming sensor for videoconferencing. *IEEE International Workshop on Haptic Audio Visual Environments and their Applications (HAVE'2004)*. Ottawa, Ontario, Canada.

Finholt, T. A. (2002). Collaboratories. In B. Cronin (Ed.) *Annual Review of Information Science and Technology*, Vol. 36. Medford, NJ: Information Today.

Finholt, T., & Olson, G.M. (1997). From laboratories to collaboratories: a new organizational form for scientific collaboration. *Psychological Science. 8*, 28-36.

Fishman, B. (in press). Linking on-line video and curriculum to leverage community knowledge. In J. Brophy (Ed.), *Advances in research on teaching: Using video in teacher education (Vol. 10)*. New York: Elsevier Science.

Foote, J., & Kimber, D., (2000, August). FlyCam: practical panoramic video. *Proc. IEEE International Conference on Multimedia and Expo, vol. III*, pp. 1419-1422.

Foote, J., & Kimber, D. (2001). Enhancing distance learning with panoramic video. *Proceedings of the 34th Hawaii International Conference on Systems Sciences.*

Foster, I., & Kesselman, C. (2003). (Eds.). *The Grid 2: Blueprint for a new computing infrastructure*. San Francisco, CA: Morgan Kaufmann Publishers.

Frederiksen, J., Sipusic, M., Sherin, M., & Wolfe. E. (1998). Video portfolio assessment: Creating a framework for viewing the functions of teaching. *Educational Assessment*, *5*(4), 225-297.

Frege, Gottlob (1970/1892). On sense and reference. In P. Geach & M. Black (Eds.), *Translations from the Philosophical Writings of Gottlob Frege, pp.56-78.* Oxford: Basil Blackwell.

Fullview (2004). (http://www.Fullview.com.)

Furnass, G.W. (1986). Generalized fisheye views. In Proceedings of the ACM SIG-CHI 86 Conference on Human Factors in Computing Systems, 16-23.

Galvis, A. H., & Nemirovsky, R. (2003). Video case studies: grounded dialogue matters most. *@Concord, 7*(1), 8.

Gardner, D., et al. (2003). Towards effective and rewarding data sharing, *Neuroinformatics Journal,* 1(3), 289-295.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.

Geisler, C., & Rogers, E. H. (2000). Design collaboration and mediating technologies. *Proceedings of the 18th annual ACM International conference on computer documentation: technology & teamwork*, pp. 395-405. Cambridge, Massachusetts.

Geisler, G., Marchionini, G., Wildemuth, B., Hughes, A., Yang, M., Wilkens, T. & Spinks, R. (2002*).* Video browsing interfaces for the Open Video Project. *Conference Extended Abstracts of CHI 2002* (pp. 514-515). New York, NY: ACM Press.

Globus Project (2000, September 5). GridFTP: Universal Data Transfer for the Grid. *Globus Project White Paper*, published by the University of Chicago and the University of Southern California. (www.globus.org/datagrid/deliverables/C2WPdraft3.pdf)

Goffman, E. (1979). *Gender advertisements*. Cambridge, MA: Harvard University Press.

Goldman-Segall, R. (1994). Challenges facing researchers using multimedia tools. *Computer Graphics* 28(1), 48-52.

Goldman-Segall, R. (1998). Points of viewing children's thinking: A digital ethnographer's journey. Mahwah, New Jersey: LEA.

Goodchild, M.F. (1992). Geographical information science. *International Journal of Geographical Information Systems, 6*(1), 31-45.

Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96, 606-633.

Goodwin, C. (1995). Seeing in depth. *Social Studies of Science*, 25, 237-274.

Goodwin, C. and Heritage, J. (1990). Conversation analysis. *Annual Review of Anthropology, 19*, 283-307.

Goody, J. (1987). *The logic of writing and the organization of society*. New York: Cambridge University Press.

Gordin, D., & Pea, R. D. (1995). Prospects for scientific visualization as an educational technology. *The Journal of the Learning Sciences, 4*(3), 249-279.

Greeno, J. G., & MMAP (Middle School Mathematics Through Applications Project Group). (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*, 5-26.

Grimshaw, A. (1997). The eye in the door: anthropology, film and the exploration of interior space. In M. Banks & H. Morphy (Eds.), *Rethinking visual anthropology*, pp. 36-52. New Haven, CT: Yale University Press.

Grossman, P.L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.

Hall, R. (2000). Videorecording as theory. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 647-664). Mahwah, NJ: Lawrence Erlbaum Associates.

Halleck, D. (2002). *Hand held visions: the impossible possibilities of community media*. New York: Fordham University Press.

Harnad, S. (1991). Post-Gutenberg galaxy: the fourth revolution in the means of production of knowledge. *Public-Access Computer Systems Review*, *2*(1), 39-53.

Harrison, B., & Baecker, R. (1992). Designing video annotation and analysis systems. *Proceedings Graphics Interface '92,* 157-166.

Hay, K. E., & Kim, B. (in press, 2005). The Integrated Multimedia Data (ITMD) research system. In R. Goldman, R. D. Pea, B. Barron & S. Derry (Eds.) *Video research in the learning sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, H. G., Hollingsworth, H., Jacobs, J., Chui, A.M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., &

Stigler, J. (2003, March). *Teaching mathematics in seven countries: results from the TIMSS 1999 video study*. Washington, DC: U.S. Department, Institute of Education, NCES Report #2003-013.

Hindmarsh, J., M. Fraser, Heath, C., Benford, S., & Greenhalgh, C. (1998). Fragmented interaction: establishing mutual orientation in virtual environments. In *Proceedings of CSCW '98*, Seattle, WA, 217-226.

Hoffert, E., Krueger, M., Mighdoll, L., Mills, M., Cohen, J., Camplejohn, D., Leak, B., Batson, J., Van Brink, D., Blackketter, D., Arent, M., Williams, R., Thorman, C., Yawitz, M., Doyle, J. & Callahan, S. (1992, January). QuickTime: an extensible standard for digital multimedia. In *Proceedings of the 37th International Conference for the IEEE COMPCON*, pp.15-20. San Francisco, California.

Hollan, J., & Stornetta, S. (1992). Beyond being there. *Proceedings of CHI-92 (*pp. 119-125). New York, NY: ACM Press.

Horvath, J., & Lehrer, R. (2000). The design of a case-based hypermedia teaching tool. *International Journal of Computers for Mathematical Learning,* 5, 115-141.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Iacucci, G., Iacucci, C., & Kuutti, K. (2002). Imagining and experiencing in design, the role of performances. *ACM International Conference Proceeding Series, Proceedings of the 2nd Nordic conference on Human-computer Interaction*, pp. 167 - 176. Aarhus, Denmark.

IMS Global Learning Consortium (2004). (http://www.imsglobal.org/metadata)

Internet2 (2004) (http://www.internet2.edu)

iPix (2004). (http://www.ipix.com/)

Jordan, B., & Henderson, A. (1995). Interaction analysis: foundations and practice. *The Journal of the Learning Sciences, 4,* 39-103.

Kendon, Adam (1982). The study of gesture: Some observations on its history. *Recherches Sémiotiques/Semiotic Inquiry* 2(1): 45-62.

Kim, K.W., Kim, K.B., Kim, H.J. (1996). VIRON: an annotation-based video information retrieval system. In *Proceedings of COMPSAC '96* (Seoul, Aug. 1996), pp. 298-303. New York: IEEE Press.

Kimber, D., Foote, J., & Lertsithichai, S. (2001, November). FlyAbout: Spatially indexed panoramic video. In *Proc. ACM Multimedia 2001*, Marina-del-Rey, pp. 339–347.

King, M.B. (2002). Professional development to promote schoolwide inquiry. *Teaching and Teacher Education*, 18, 243-257.

Kipp, M., (2001). ANVIL: a generic annotation tool for multimodal dialogue. *Proceedings of Eurospeech*, pp.1367-1370.

Koehler, M.J. (2002). Designing case-based hypermedia for developing understanding of children's mathematical reasoning. *Cognition and Instruction, 20* (2), 151-195.

Koschmann, T. (1999). (Ed.). Special issue: meaning making. *Discourse Processes, 27*(2).

Koslow, S.H. (2002). Sharing primary Data: a threat or asset to discovery? *Nature Reviews Neuroscience,* 3 (4), 311-313.

Lampert, M., & Hawkins, J. (1998, September). *New technologies for the study of teaching*. Report to the National Science Foundation from a workshop held June 9-11, 1998, Ann Arbor, MI (NSF Grant #REC-9802685).

Lampert, M., & Loewenberg-Ball, D. (1998). *Teaching, multimedia and mathematics: investigations of real practice.* New York: Teachers' College Press.

Lamping, J., Rao, R., & Pirolli, P. (1995). A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. *Proceedings of the SIGCHI conference on human factors in computing systems,* pp. 401-408. Denver, Colorado.

Latour, B. (1986). Visualization and cognition: Thinking with eyes and hands. In H. Kuklick & E. Long (Eds.), *Knowledge and Society: Studies in the Sociology of Culture Past and Present-A Research Annual, vol. 6, pp. 1-40*. Greenwich, CT: JAI Press.

Laurel, B. (1990). (Ed). *The art of human-computer interaction*. Reading, MA: Addison-Wesley.

Lave, J. (1988). *Cognition in practice.* Cambridge: Cambridge University Press.

Lave, J. (1993). Situating learning in communities of practice. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 17-36). Washington, DC: American Psychological Association.

Lawson, M. J. (1994). Concept mapping. In T. Husén & T. N. Postlethwaite (Eds.), *The international encyclopedia of education (2nd ed., Vol. 2, pp. 1026-1031)*. Oxford: Elsevier Science.

Lee, D., Erol, B., Graham, J., Hull, J., & Murata, M. (2002, December). Portable meeting recorder. *Proc. ACM Multimedia 2002*, Juan-les-Pins, pp. 493-502.

Lee, S.Y., & Kao, H.M. (1993). Video indexing: an approach based on moving object and track. *Proceedings of the SPIE*, Vol. 1908, 25-36.

Lemke, J. (1999). Typological and topological meaning in diagnostic discourse. *Discourse Processes*, *27*(2), 173-185.

Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lu, J., & Rose, R. (2003). Seeing math through multimedia case studies. *@Concord, 7*(1), 1, 4-5.

Lyman, P., & Varian, H. R. (2003). *"How Much Information", 2003*. Retrieved from http://www.sims.berkeley.edu/how-much-info-2003 on December 30, 2004.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ma, X., Lee, H., Bird, S., Maeda, K., (2002). Model and tools for collaborative annotation, *Proceedings of the Third International Conference on Language Resources and Evaluation*, Paris: European Language Resources Association.

Mackay, W., & Beaudouin-Lafon, M. (1998). DIVA: exploratory data analysis with multimedia stream. *CHI 1998 Proceedings*. New York, NY: ACM Press.

MacWhinney B, (2000). *The CHILDES Project: tools for analyzing talk.* Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B., & Snow, C. (1999). *Multimodal studies of instructional discourse.* Report to the National Science Foundation. (http://www.nsf.gov/sbe/tcw/events_991203w/)

Martin, R. (2002). *Picasso's war: The destruction of Guernica and the masterpiece that changed the world*. New York: E.P. Dutton.

Masten, D., & Plowman, T.M.P. (2003). Digital ethnography: The next wave in understanding consumer experience. *Design Management Journal, 14*(2), 75-81.

McCormick B.H., DeFanti T.A., & Brown, M.D., (1987, November). (Eds.) Visualization in scientific computing (special issue). *Computer Graphics. 21*, 6.

Mehan, H. (1978). Structuring school structure. *Harvard Educational Review*, 48(1), 32-64.

Metz, C. (1974). *Film language: a semiotics of the cinema*. New York: Oxford University Press.

Mills, M. I., Cohen, J., & Wong Y. Y. (1992). A magnifier tool for video data. *Proceedings of CHI '92*. pp. 93-98. New York, NY: ACM Press.

MPEG-4 Standard (2004). (http://www.m4if.org/mpeg4/)

Mumme, J. (2003). *Videocases for mathematics teacher professional development* (http://www.wested.org/cs/we/view/pj/361).

Myers, B., Casares, J.P., Stevens, S., Dabbish, L., Yocum, D., & Corbett, A. (2001). A multi-view intelligent editor for digital video libraries. In Proceedings of the First *ACM/IEEE Joint Conference on Digital Libraries, pp. 106-115*. Roanoke, VA.

Nardi, B. (Ed.). (1996). *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, MA: The MIT Press.

Nayar, S. K. (1997, May). Omnidirectional video camera. *Proc. DARPA Image Understanding Workshop*, New Orleans.

Neidle, C., Sclaroff, S., & Athitsos, V. (2001). A tool for linguistic and computer vision research on visual-gestural language data. *Behavior Research Methods, Instruments, and Computer*s, 33(3), 311–320.

Nemirovsky, R., Lara-Meloy, T., Earnest, D., & Ribeiro, B. T. (2001). *Videopapers: Investigating new multimedia genres to foster the interweaving of research and teaching.* Paper presented at the 25th Meeting of the International group for the Psychology of Mathematics Education, Utrecht University. The Netherlands.

Neumann, U., Pintaric, T., & Rizzo, A. (2000). Immersive panoramic video. In *Proceedings of the 8th ACM International Conference on Multimedia*, pp. 493-494. Marina del Rey, California.

Nielson, G. M., Muller, H., Mueller, H., & Hagen, H. (1997). *Scientific visualization: overviews, methodologies and techniques*. New York: IEEE Computer Society.

Novak, J. D. (1998). *Learning, creating, and using knowledge: Concept maps® as facilitative tools in schools and corporations*. Mahwah, NJ: Lawrence Erlbaum Associates.

NSF Middleware Initiative (NMI). (2004). (http://www.nsf-middleware.org)

Olson, D. (1994). *The world on paper: the conceptual and cognitive implications of writing and reading*. Cambridge: Cambridge University Press.

Ong, W. J. (1982). *Orality and literacy: technologizing the word*. New York: Methuen.

Open QuickTime (2004). *Open standards for QuickTime*. (http://www.openquicktime.org.)

Pea, R. D. (1992). Augmenting the discourse of learning with computer-based learning environments. In E. de Corte, M. Linn, & L. Verschaffel (Eds.), *Computer-based learning environments and problem-solving (NATO Series, subseries F: Computer and System Sciences)*. New York: Springer-Verlag GmbH (pp. 313-343).

Pea, R. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations (pp. 47-87)*. Cambridge, England: Cambridge University Press.

Pea, R. D. (1994). Seeing what we build together: Distributed multimedia learning environments for transformative communications. *The Journal of the Learning Sciences*, *3*(3), 283-298.

Pea, R. D. (1999). New media communication forums for improving education research and practice. In E. C. Lagemann & L. S. Shulman (Eds.), *Issues in Education Research: Problems and possibilities* (pp. 336-370). San Francisco, CA: Jossey Bass.

Pea, R.D. (2002). Learning science through collaborative visualization over the Internet. In N. Ringertz (Ed.), *Nobel Symposium: Virtual museums and public understanding of science and culture*. Stockholm, Sweden: Nobel Academy Press.

Pea, R. D. (2004). The social and technological dimensions of "scaffolding" and related theoretical concepts for learning, education and human activity. *The Journal of the Learning Sciences*, *13*(3), 423-451.

Pea, R.D., & Hay, K. (2003, March). *CILT Workshop on Digital Video Inquiry in Learning and Education, November 25-26, 2002* (Report to the National Science Foundation based on NSF #0124012). Palo Alto, CA: Stanford Center for Innovations in Learning.

Pea, R., Wulf, W., Elliot, S.W., & Darling, M. (2003, August). (Eds.). *Planning for two transformations in education and learning technology* (Committee on Improving Learning with Information Technology). Washington, DC: National Academy Press.

Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W., & Hoffert. E. (2004, Jan-Mar). The DIVER™ Project: interactive digital video repurposing. *IEEE Multimedia, 11*(1), 54-61.

Porter, A.C., Youngs, P., & Odden, A. (2001). Advances in teacher assessment and their uses. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th ed., pp.259-297). Washington: American Educational Research Association.

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: beyond human-computer interaction*. New York, John Wiley & Sons.

Prihavec, B., & Solina, F. (1998). User interface for video observation over the Internet. *Journal of network and computer applications*, 21, 219-237.

Quine, W.V.O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Rao, R., & Card, S.K. (1994, April 24-28). The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information, *Proceedings of the SIGCHI conference on human factors in computing systems*, pp.318-322. Boston, Massachusetts.

Rayward, W.B., & Twidale, M. B. (1999). From docent to cyberdocent: education and guidance in the virtual museum. *Archives and Museum Informatics*, *13*, 23-53.

Resnick, L. (1987). Learning in school and out. *Educational Researcher, 16*(9), 3-21.

Rheingold, H. (1992). *Virtual reality: the revolutionary technology of computer-generated artificial worlds-and how it promises to transform society*. New York: Simon & Schuster.

Rizzo, A. A., Neumann, U., Pintaric, T., & Norden, M. (2001). Issues for application development using immersive HMD 360 degree panoramic video environments. In M. J. Smith, G. Salvendy, D. Harris, & R.J. Koubek (Eds.), *Usability evaluation and interface design, Vol. 1*, pp. 792-796.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Rogoff, B. (2003). *The cultural nature of human development*. New York: Oxford University Press.

Rogoff, B., & Lave, J. (Eds.). (1984). *Everyday cognition: its development in social context*. Cambridge, MA: Harvard University Press.

Roschelle, J., & Goldman, S. (1991). VideoNoter: A productivity tool for video data analysis. *Behavior Research Methods, Instruments, and Computers*, *23*, 219-224.

Roschelle, J., & Pea, R. D. (2002). A walk on the WILD side: How wireless handhelds may change computer-supported collaborative learning (CSCL). *The International Journal of Cognition and Technology, 1*(1), 145-168.

Roschelle, J., Pea, R. D., & Trigg, R. (1990). *VideoNoter: A tool for exploratory video analysis*. Palo Alto, CA: Institute for Research on Learning, Technical Report, No. 17.

Roth, Wolff-Michael (2001). Gestures: Their role in teaching and learning. *Review of Educational Research, 71* (3), 365-392.

Roth, W.- M. (2001). Situating cognition. *The Journal of the Learning Sciences*, 10, 27-61.

Roth, W.- M.,& Roychoudhury, A. (1993). The concept map as a tool for the collaborative construction of knowledge: A microanalysis of high school physics students. *Journal of Research in Science Teaching*, *30*, 503-534.

Rui, Y., Gupta, A., & Cadiz, J.J. (2001). Viewing meetings captured by an omni-directional camera. *Proceedings of the 2001 ACM Conference on Human Factors in Computing Systems (CHI 2001).*

Saxe, G. B. (1988). Candy selling and math learning. *Educational Researcher*, *17*(6), 14-21.

Saxe, G. B. (1991). *Culture and cognitive development: Studies in mathematical understanding.* Hillsdale, NJ: Erlbaum.

Scaife, M. & Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature, 253*, 265–266.

Schoenfeld, A. H. (1992). On paradigms and methods: What do you do when the ones you know don't do what you want them to? Issues in the analysis of data in the form of videotapes. *The Journal of The Learning Sciences, 2*, 179-214.

Schon, D. (1983). *The reflective practitioner*. New York: Basic Books.

Schon, D.A (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco, Jossey-Bass.

Schrader, P.G., Leu, Jr., D. J., Kinzer, C.K., Ataya, R., Teale, W. H., Labbo, L.D., & Cammack, D. (2003, July). Using Internet delivered video cases, to support pre-service teachers' understanding of effective early literacy instruction: An exploratory study. *Instructional Science*, 31(4), 317-340.

Schuler, D., & Namioka, A. (1993). (Eds). *Participatory design: principles and practices*. Hillsdale, NJ: Erlbaum.

Seamon, D., & & Zajonc, A. (Eds.) (1998). *Goethe's way of science: a phenomenology of nature*. Albany, New York: State University of New York Press.

Sfard, A., & McClain, K. (2002). Special issue: Analyzing tools: Perspectives on the role of designed artifacts in mathematics learning. *The Journal of the Learning Sciences*, *11*(2 & 3).

Shahabi, C., Zimmermann, R., Fu, K., & Didi Yao, S-Y. (2002). Yima: a second-generation continuous media server. *IEEE Computer, 35*(6), 56-64.

Shamberg, M. & Raindance Corporation (1971). *Guerrilla television*. New York: Holt, Rinehart and Winston.

Sherin, M.G. (2001). Developing professional vision of classroom events. In T. Wood, B. S. Nelson, & J. Warfield (Eds.)*, Beyond classical pedagogy: Teaching elementary school mathematics, pp. 75-93*. Mahwah, NJ: Erlbaum.

Sherin, M. G. (in press). Redefining the role of video in teacher education. In J. Brophy (Ed.), *Using video in teacher education*. New York: Elsevier Science.

Sherin, M. G., & van Es, E. A. (2002). *Using video to support teachers' ability to interpret classroom interactions.* Paper presented to the Society for Information Technologies in Education (SITE). http://www.aace.org/conf/site/pt3/paper_3008_1031.pdf

Shotter, J. (2000) Seeing historically: Goethe and Vygotsky's 'enabling theory-method'. *Culture and Psychology, 6*(2), pp.233-252.

Shrader, G., Fishman, B., Barab, S. A., O'Neill, K., Oden, G., & Suthers, D. D. (2002). Video cases for teacher learning: Issues of social and organizational design for use. In G. Stahl (Ed.), *Computer Support for Collaborative Learning: Foundations for a CSCL Community (pp. 708-709)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shulman, J. H. (Ed.). (1992). *Case methods in teacher education*. New York: Teachers College Press.

Snyder, J., Lippincott, A., & Bower, D. (1998). The inherent tensions in the multiple uses of portfolios in teacher education. *Teacher Education Quarterly, 25*(1), 45-60.

Star, S.L., & Griesemer, J.R. (1989). Institutional Ecology, "Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science. 19*, 387.

Stevens, R., Cherry, G., & Fournier, J. (2002). Video Traces: rich media annotations for teaching and learning. *Proceedings of the CSCL 2002 Conference on Computer Supported Collaborative Learning*, Boulder, CO.

Stevens, R. and Hall, R. (1997). Seeing Tornado: How VideoTraces mediate visitor understandings of (natural?) spectacles in a science museum. *Science Education*, *18*(6), 735-748.

Stevens, R. & R. Hall (1998). Disciplined perception: Learning to see in technoscience. In M. Lampert & M. L. Blunk (Eds.), *Talking mathematics in school: Studies of teaching and learning*. New York: Cambridge University Press.

Stevens, R. & Toro-Martell, S. (in press). Leaving a trace: Digital Meta-exhibits for supporting visitors to represent and exchange their ideas about museum exhibits. *Journal of Museum Education*.

Stiefelhagen, R., Yang, J. & Waibel, A. (2002). Modeling focus of attention for meeting indexing based on multiple cues. *IEEE Transactions on Neural Networks, 13*(4), pp. 928-938.

Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S., and Serrano, A. (1999). *TIMSS Videotape Classroom Study*, U.S. Department of Education, Washington, DC, NCES 1999-074.

Stigler, J., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, *35*(2), 87-100.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Stone, B.A. (1998). Problems, pitfalls, and benefits of portfolios. *Teacher Education Quarterly, 25*(1): 105-114.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge, England: Cambridge University Press.

Suchman, L., & Trigg, R. H. (1991). Understanding practice: video as a medium for reflection and design. In J. Greenbaum & M. Kyng (Eds.), *Design at work*, pp. 65-90. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sun, X., Foote, J., Kimber, D., & Manunath, B.S. (2001, September). Recording the region of interest from FLYCAM panoramic video. *Proc. IEEE International Conference on Image Processing (ICIP 2001),* pp. 409-412. Thessaloniki, Greece.

Sun, X., Foote, J., Kimber, D., & Manjunath, B.S. (2001, November). Panoramic video capturing and compressed domain virtual camera control. *Proc. ACM Multimedia 2001,* pp. 329-338. Marina del Rey, CA.s

Sun, X., & Manjunath, B. S. (2002, September). Panoramic capturing and recognition of human activity. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2002),* Vol. II: pp. 813-816. Rochester, NY.

Tan, K-H, Hua, H., & Ahuja, N. (2004). Multiview panoramic cameras using mirror pyramids. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*(7), 941-946.

Tang, J. C. (1991). Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, 34 (2), 143-160.

Teodosio L.A., & Mills, M. (1993). Panoramic overviews for navigating real-world scenes. In *Proceedings of the First ACM International Conference on Multimedia (*pp. 359-364). Anaheim, CA.

Transana (2004). *Transana: a tool for the transcription and qualitative analysis of audio and video data* (http://www.transana.org)

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication, pp. 321–347*. New York: Cambridge University Press.

Tufte, E. R. (1997). *Visual explanations: images and quantities, evidence and narrative*. Cheshire, CT: Graphic Press.

Ulewicz, M. & Beatty, A. (2001). (Eds.). *The power of video technology in international comparative research in education*. Washington DC: National Research Council, Board on International Comparative Studies in Education, Board on Testing and Assessment, Center for Education.

Varmus, H., Lipman, D., Ginsparg, P., & Markovitz, B.P. (2000). The impact of open access on biomedical research. *Genome Biol.* 1(3).

Virage (2004). (http://www.virage.com)

Weber, K., & Poon, A. (1994, April). Marquee: a tool for real-time video logging. In *Proceedings of CHI-94*, pp. 58-64. New York, NY: ACM Press.

Williamson, J. (1994). *Decoding advertisements: Ideology and meaning in advertising*. London: Marion Boyars.

Windows Media Technologies (2004). (http://www.microsoft.com/windows/windowsmedia/)

Winograd, T. (1996). (Ed.). *Bringing design to software*. Reading, MA: Addison-Wesley.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

Wolff, R., & Yaeger, L. (1993). *Visualization of natural phenomena*. New York: Springer-Verlag.

Yang, J., Zhu, X., Gross, R., Kominek, J., Pan, Y., & Waibel, A. (1999). Multimodal people ID for a multimedia meeting browser. In *Proceedings of the 7th ACM International Conference on Multimedia*, pp. 159-168. Orlando, FL.

Young, J. R. (2001, July 19). Professors use the web to publish portfolios of teaching techniques. *The Chronicle of Higher Education*.

Youngs, P., Odden, A., & Porter, A.C. (2003). State policy related to teacher licensure. *Educational Policy. 17*(2), 217-236.

Zeichner, K. & Wray, S. (2001). The teaching portfolio in U.S. teacher education programs: What we know and what we need to know. *Teaching and Teacher Education, 17*(5), 613-621.