# Enhancing Classroom and Distance Learning Through Augmented Reality

Christopher Coffin, Svetlin Bostandjiev, James Ford AND Tobias Hollerer
*University of California Santa Barbara*
*USA*
ccoffin@cs.ucsb.edu
svetlin@umail.ucsb.edu
jamesford@umail.ucsb.edu
holl@cs.ucsb.edu

**Abstract.** We present a multimedia solution for easily adding virtual annotations to class lectures through the use of augmented videoconferencing and tracked physical props. These props may be any object normally used in a lecture, such as toy cars or tops (physics), ball and stick molecules (chemistry), or frogs to be dissected (biology). In the classroom, the actions of the instructor are captured by one or more cameras. We then use a normal desktop computer to add virtual data to the camera image. Our software solution tracks the physical objects and allows for overlays of relevant information, optionally deriving information from the movement of the objects. For example, a toy car may be tracked in order to determine its velocity, which may then be displayed as a 3D arrow (vector) directly on top of the video showing the moving car. The resulting video may be sent either to a projector or monitor (to be viewed in class) or over the Internet (to be viewed by remote students). Additionally, our solution allows students to interact with the virtual data through the augmented video, even when distributed over the Internet.

## Introduction

This work addresses two main goals:

- Provide instructors with a way to strengthen students' understanding in the classroom by augmenting physical props with virtual annotations and illustrations.
- Allow students to interact with these virtual augmentations, and through the augmentations interact with the instructor and other students, even over great distances and through a variety of machines.

We are able to accomplish these goals using our software solution and a simple physical setup consisting of one or more cameras, a projector, and a desktop computer with monitor and Internet connection, for an illustration please (see Fig. 1). The Internet connection is optional and only necessary for delivering live video to remote students. Additionally, classes may be recorded and at a later time broadcast to students.

Our system supports instructors in employing physical props for illustrative purposes during lessons. A prop may be any physical object relating to the lecture material. For example, biology lectures could be enhanced by adding labels to the organs of a dissected animal and by allowing students to see animated instructions of the next steps in the dissection. In art classes, the props may take the form of a small statue over which can be drawn a high quality virtual image of a famous sculpture. The physical prop would serve as a method for orienting and inspecting the virtual sculpture displayed to the students. This would allow instructors to more easily highlight fine details or areas of interest in the artwork. Structural engineering could be illustrated through the use of a physical model of a bridge and a set of small weights. A camera would detect the position of weights placed on the bridge and a virtual model of the resulting load distribution could be overlaid on the physical bridge.
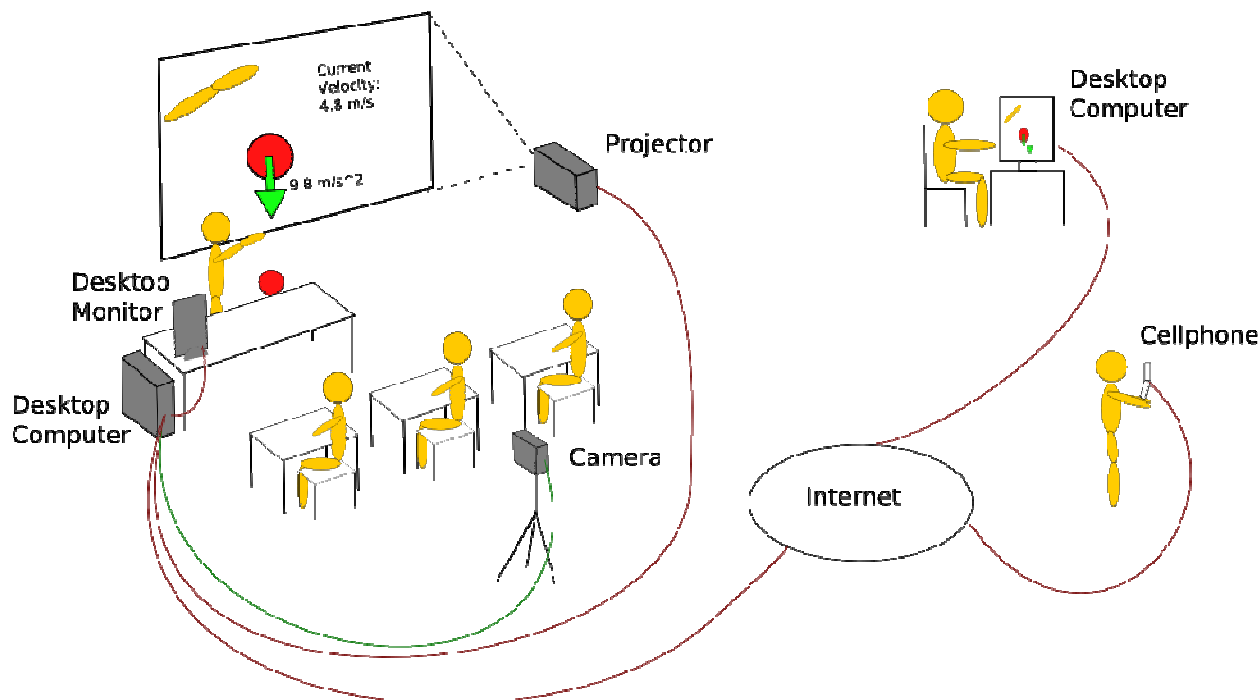
**Figure 1:** A diagram illustrating our classroom setup. The instructor is filmed by a camera while students view the augmented video on the projection screen. The same image is displayed on the desktop monitor and the projector, thereby providing the instructor with a reference without turning her back to the students. The video can also be distributed over the Internet to be viewed by students on various devices such as computers or cellphones.

In physics classes, which are the center of focus in our prototype examples, utilizing a physical prop along with virtual objects also allows for a unique illustration of the difference between the occasionally simplified estimations given by formulas in textbooks and the results seen in reality. Take for example a teacher dropping a physical ball to demonstrate acceleration due to gravity (Fig. 1). With the use of augmentations the students are able to view both the physical ball dropped as well as a virtual ball dropped at the same time and from the same height. While the physical ball is subject to friction the virtual ball can be made to fall as if it were in a vacuum. Because we are distributing the recorded video, students following the lecture from a remote computer will also be able to rewind the video and view the balls dropping in slow motion to allow for more accurate comparisons. Students in the classroom will view the augmentations on a projector, while the instructor will be able to view and control the content from a monitor embedded in or placed on her desk. Note that the monitor placed on the instructor's desk is optional; however, it provides a useful tool by allowing the instructor to easily reference the virtual augmentations without using the projected image on the wall, thereby avoiding the need to turn her back to the class.

We also use our technique for demonstrating physical experiments and examples which are not as easily or as powerfully communicated through two dimensional drawings or illustrations. Take for example the physics lesson of precession which is usually taught using a spinning top and blackboard illustrations. The lesson demonstrates how gravity pulls on a spinning top, creating torque that points along the cross product of the angular momentum and the direction of gravity. Torque pulls the top sideways, causing it to rotate ("precess") rather than fall down. This example has two important characteristics which make it difficult to understand and teach in traditional ways. The vectors along which forces interact must be understood in 3D, and visualizing the movement of the top and resulting force vectors is critical for understanding. Our solution can not only overlay the vectors correctly in 3D, but can also slow down and replay the augmented video for closer inspection.

We have two goals for the video sent to the students. First we wish to provide every student with the highest quality images of the classroom activities possible. This provides students better access to virtual objects, text, and animation, and allows the students to focus more on the lecture, and less on attempting to interpret images. Second, we wish to provide the students with a means of interacting with the imagery. In a live broadcast this may also allow

interaction with the instructor and other students via adjustments to the virtual data.

Interaction with virtual objects which are seen by more than one person can be either shared or private. Shared interaction is much more common. In this mode the student would be able to interact with the virtual objects on their screen and any adjustments would be seen by the rest of the class. An example of this interaction is given in the following scenario. Imagine that a professor has virtual representation of several molecules and asks if there is anyone in the class who would be able to demonstrate the way in which the chemical structures would bond. A student would be able to ask the professor for permission to attempt a solution, the professor would then allow the student to interact with the virtual objects and the student would have the ability to move and rotate the structures in order to move them into the correct location, with every other student and the instructor clearly witnessing the same interaction.

Private interaction means that a student would be able to closely examine a virtual object without the professor or the other students being privy to that student's examination. For example, if a student was viewing the video using a small screen and wished to take a closer look at the virtual object, he would be able to enlarge it without the professor or the other students being aware of the change. This is not trivial and most research in this area has not fully considered what is necessary to achieve this on a larger scale (e.g. with hundreds or thousands of viewers) or with students viewing video on small devices such as palmtop computers or cellphones.

There are two main benefits of distributing the lecture to a variety of machines beyond desktop computers. First, we allow for greater availability of the lectures. Students are now able to view and participate in a lecture using only their cellphones, perhaps while sitting on a bus or otherwise away from their computer or the classroom. The second benefit is in allowing students in the classroom to reap the benefits of interaction with the virtual augmentation. While not every class can be equipped with computers for each student, it is financially more feasible to provide smaller "thin client" devices which can still allow students to view, rewind, or interact with the augmentations and virtual data. Additionally, students would be able to use their own cellphones or PDA's in the classroom for the same purpose.

## Previous Work

Previous works such as (Liarokapis et. al. 2002), (Fjeld 2003), (Shelton 2002), and (Kaufmann 2003) explore the use of AR in education. These works focus on presenting 3D graphical models to students, in order to assist them with complex spatial problems. Most of these projects support simple user interaction with the virtual world, allowing the exploration of an idealized universe. Our system, on the other hand, utilizes the interaction with the real world as we directly tie virtual data to physical objects. Additionally, previous work is based on interaction which involves students using head mounted displays or expensive tablet computers. That work is typically designed for only one or two students. Unlike our work, they are not designed as tools for instructors.

There has been a growing interest in distributing lectures online through websites such as YouTube (Associated Press 2007). Systems such as SciVee, DNAtube, and JOVE are all examples. Similar work, with a focus on interaction, has been either video and PowerPoint driven such as in (Yuanchun 2003), or has included digital whiteboard information along with the video. Our solution differs in that we include virtual data in the same 3D space the instructor is in, and in that we allow for interaction with both live and recorded video. The BIFS standard included in MPEG-4 (see Noimark 2003) could be used to implement a somewhat similar distribution technique. While we use the MPEG-4 standard for encoding the video, BIFS distributes the virtual data using a file format similar to VRML or XML, and therefore is not scalable to certain low-end platforms.

## Implementation

We present two technological advances for distance learning. First, we provide a software solution which allows instructors to easily add computer annotations and images to physical props used during lectures. Second, our software allows this augmented video to be displayed in the classroom or distributed to students who are then able to

interact with the virtual data.

## Augmenting Physical Props

The first phase of every demonstration is introducing any real props the instructor is interested in using into the system. This phase involves obtaining useful information about the prop, such as color histogram and shape. Then, during the experiment, we use computer vision to track the props in real time and locate their 3D positions. In order to augment physical objects, we use ARToolkit markers to establish a global coordinate system, establishing the teacher's desk (or any other planar surface) as the ground plane for our virtual world. Assuming that the chosen flat surface is parallel to the ground, we can also assume that gravity works orthogonally to that plane. Such a plane also identifies a resting place for virtual objects, meaning an instructor would be able to place virtual objects on the surface on which the marker was placed (e.g. a table or the floor). Establishing a coordinate system in which we can reliably track the 3D position of physical objects over time provides us with a powerful educational framework in which we can let the computer calculate physical forces taking place in the real world, such as speed, velocity, acceleration, centripetal and centrifugal force, pressure, friction, elasticity, and energy changes. We are then able to overlay graphics on top of the physical props to visualize these forces, invisible to the human eye.

Our prototype has been used to enhance physics lessons with a simple mono-colored ball. In this case, we initially retrieve information about the color, shape, and real size of the ball. Then, at every video frame during real-time playback, we employ color segmentation to approximate where the center pixel of the ball is. After a first rough approximation of the location of the center pixel, we grow a region of similarly colored pixels around that pixel. That region eventually covers the whole ball and then its center of mass defines the 2D center pixel of the ball. Assuming that the ball moves on the established working plane, we use simple trigonometry to find the 3D position of the center of the ball relative to the coordinate system defined by the ARToolkit marker. This technique provides us with a simple way of tracking the 3D position of regular single colored objects. Once we know where the center of the ball is at every frame we calculate the ball's instantaneous velocity (blue), acceleration (green), and centripetal force (yellow), and overlay vectors on top of the ball, see (Fig. 2).
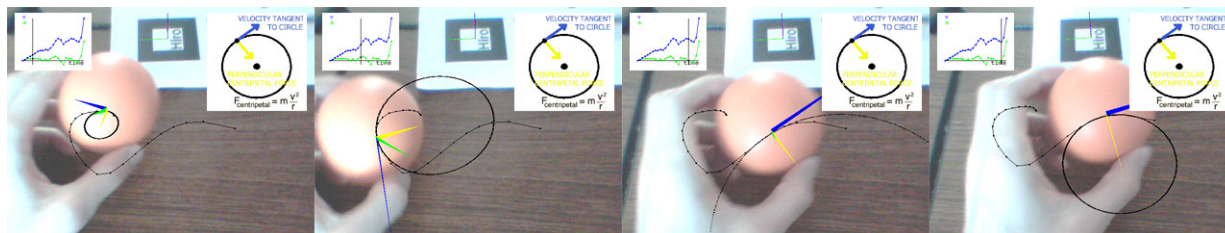


**Figure 2**: The following sequence was taken from a demonstration of centripetal force. The instantaneous velocity (blue), acceleration (green), and centripetal force (yellow) can be seen as arrows originating from the center of the ball. The black arc illustrates the movement of the ball over the video sequence and the black circle is a projected path based on the most recent movement. Graphs related to the lecture can be seen in the upper corners of the image.

Another type of augmentation is to display particular graphs and charts relevant to the demonstration. This can provide visual explanations of fundamental concepts and formulas which are often confusing for students without a strong mathematical background. Also we can dynamically draw diagrams on the screen to illustrate changes of involved entities over time.

In order to enhance teaching of the physical concepts, the instructor may need to reproduce the demonstration in slow motion, or go through it frame by frame, or just display a particular snapshot of interest. Therefore, we have implemented video recording and playback, as well as frame by frame access. While showing particular frames or sequences, the instructor may display different types of augmentations (e.g. vectors, graphs, formulas) as appropriate.

Another benefit of dealing with a previously recorded video stream is that it provides us with global information about the entire experiment, no matter the current viewing frame. We can use this information to correct for inaccuracies in the real time computer vision algorithms and to better illustrate global concepts such as overall

trajectories. In the example we have developed, we use the information from all frames to accurately display the trajectory of the ball over the whole demonstration period. We achieve this by fitting a Bezier curve to the set of points representing the ball's position at each frame. Being able to display the trajectory of the ball over time helps teaching that the net force on the ball moving along a curve can be decomposed into a tangential velocity component that changes the speed of the ball, and a perpendicular centripetal force component that changes the direction of motion. Moreover, the instructor can give a visual explanation of the concept of instantaneous acceleration by showing how the acceleration vector is the difference of two consecutive velocity vectors.

### Distributing the Video

Our next advancement is in distributing the video and computer graphics to a large range of students. We would like to allow as many students to connect to the video feed as possible. Additionally, all of these students should be able to interact with the things they see. As previously discussed, not all students will have access to a high-speed Internet connection or a powerful computer with advanced rendering capabilities. We have included (Tab 1) to illustrate our proposed solutions for a combination of network and computation resources:

|  | Low Rendering Capabilities | High Rendering Capabilities |
|---|---|---|
| Low Bandwidth | E: Cell Phone<br>S: Thin Client or Meta approach | E: WiFi-enabled PDA<br>S: Meta approach |
| High Bandwidth | E: Laptop with cellular modem<br>S: Structure approach | E: Desktop computer w. LAN<br>S: Structure approach |

**Table 1:** The above table gives four classifications based on the graphical capabilities of the student's machine and the available bandwidth. Example machines for each class are given (E) as well as our proposed solutions (S).

We distribute the augmented video using three basic techniques: *Structure Approach*, *Thin Client*, and *Meta Approach*. The first is simply to distribute the structure of the virtual objects, and then allow the students' computer to render the virtual object to the screen, overlaid on top of the video in the appropriate location. We term this technique the *structure approach*, as the student receives only structural information in addition to the video. The structure approach allows for the largest amount of private interaction, and it is very easy for the student to examine or change the virtual objects without affecting the class. The main disadvantage is that the student's machine must be sufficiently powerful to render the virtual objects on top of video textures.

Our second approach is to have the instructor's computer draw the virtual objects on top of the video. The resulting image is then sent to the student's computers. This technique is named the *thin client approach* after the computing model it closely resembles. There are some advantages for using this approach. Performing the rendering on the professor's machine means that the student's machine does not need to have any graphical capabilities. That is, students may use cell phones or other small devices that can stream video. Additionally, the professor's computer may be able to produce a higher quality image than any student's machine. In such cases, this technique produces better imagery for viewing than our previous solution. The disadvantage is that interaction becomes more difficult. With this approach, there is no technique for allowing a scalable number of students to have private interaction with the augmentations. Additionally, students may find that there is a large time difference between a student's interaction event (such as a button press to move an object) and when the student is able to see the desired result (the object appears moved). This is due to the need for such messages to travel from the student's computer to the professor's computer and back again.
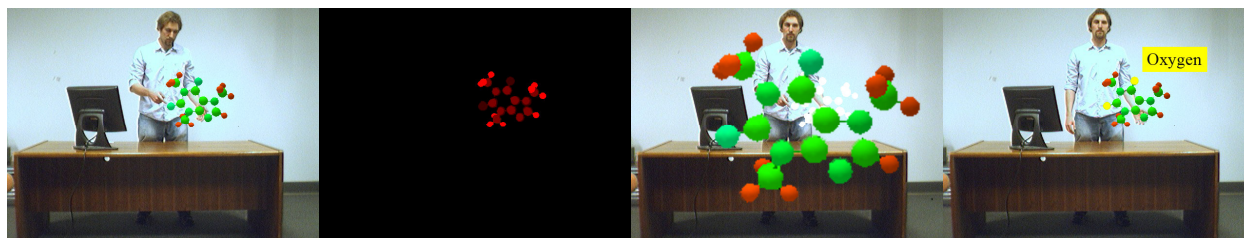


**Figure 3**: Chemistry lecture with a virtual caffeine molecule and interaction using meta image. From left to right, the composite image sent to the student's machine, the flat color meta image of the augmentation (contrast has been

enhanced to show color difference to naked eye), the enlargement filter applied (original molecule colored white during enlargement, using same filter), and the highlighting filter identifying oxygen atoms in yellow with additional textual key.

Our third technique is the *meta approach*. This approach allows students with machines without high-end graphics capabilities (such as cellphones) to perform both shared and private interaction with the virtual objects. In addition, this technique is just as scalable as the second approach. In order to allow the student to interact with the video, we provide an additional internal image frame which can be combined with the image produced by the instructor's computer from the previous technique. This second image frame, termed the meta image, contains information regarding the virtual objects in the scene and allows the student to have at least a limited amount of instantaneous interaction. We accomplish this by using the meta image along with some simple image filters in order to produce adjustments to the image such as:

- enlarging virtual objects in order to give students a closer look (Fig. 3)
- highlighting of virtual objects (Fig. 3)
- moving a virtual object
- adding additional text to accompany highlights (Fig. 3)

We can either employ a two-stream approach (meta image and augmented video), or a three-stream approach, in which we send the augmentations and the meta image separately from the original plain video image, enabling the video recipient to use filters involving transparency and remove virtual objects altogether. The meta image is simply a rendering of the virtual objects in the image, but with each virtual object draw using only a single unique color. The advantage of using an image to send this data is in the use of the MPEG compression. Typically, meta images are on the order of 100 bytes for a 320 by 240 image with a bitrate which is near-lossless quality. This can be much smaller if the augmentations are stable as in the case of statically positioned graphs, or tables. Meta images are created by rendering each object (or individual section) as a unique color (see Fig. 3). The color used is based on a numeric index. While the default method for determining the index is based on the structural (rendering) order of the elements, the concept can be extended to help classify the data being viewed. For example, all of the charts and graphs may be given an index of one while augmentations are given an index of two. This would allow users to easily remove only the augmentations or only the graphs being displayed. The same technique can be used to automatically distinguish between annotations (e.g. vector arrows) to physical models and annotations of virtual models, as in our earlier example of demonstrating gravity. It can also be used to distinguish various types of virtual objects, such as the oxygen atoms in the molecular example in (Fig. 3).

Determining the flat color seen in the meta image from the indices is fairly straight forward. Using our system there are 256 intensity values for each of our three color values (red, green, and blue). This provides a 256 base numbering system with a possible 3 digits (4 if using the byte for transparency) with red being the least significant digit and blue the most significant. For example, the first index would produce a very dark shade of red (black is used to distinguish background), while the greatest index would produce a bright white. Using the reverse method this value is then converted to a base 10 value by the student's computer to be used for indexing.

Using color values, we are able to further reduce the size of the meta image by allowing for lossy encoding. We then compensate for any error in the encoding by reducing the number of intensity values we assume (i.e. from 256 to 128, 64, 32, etc...). As the intensity range is used to distinguish individual elements seen in the image, the quality of the encoding can be allowed to degrade to match the number of visible elements. This means for a small number of objects the encoding can be extremely poor. As a reference point a lossless encoding (256 intensity values) would allow for more objects than there are pixels in eight frames of high definition video, meaning we have a great deal of flexibility.

Please note that the choice for one of the above techniques is flexible and the data sent to the student can be changed at any time. This may allow the student to save resources on their own computer, when not interacting, or not interacting heavily. Take for example a student viewing a chemical demonstration involving a million atoms. The student's machine may not be able to produce as crisp of an image of the real-time simulation as the professor's computer is able to produce. The student would then prefer to use the professor's image and therefore would use the thin client approach. If the student is interested in making small changes, such as enlarging the simulation or moving some atoms to another location, they could use the meta image technique and still see the higher quality image. If they are interested in more detailed interaction, such as rotating the simulation to get a view from another angle, or

removing several atoms from the simulation (not just their view) then they would need to use the structure technique.


## Tests and Discussion

We are currently developing this work to the point where a larger usability test would be desirable, and this is the next step in our agenda. For now, expert evaluation of our current work by physics educators is extremely promising and encouraging. We implemented a test scenario involving an instructor at UCSB illustrating a simple physics lesson using both virtual and real objects. The physical setup for this scenario consists of a single camera (in this case a Unibrain Fire-i) filming the instructor. For the instructor's computer, controlling overlays and distribution schemes, we used a normal desktop computer (Dell Dimension E510). Before the lecture began, the instructor printed a single ARToolkit marker which was briefly placed on the wall behind him and served to determine the position of the camera relative to the working plane (vertical in this case). The instructor then gave a simple physics lecture demonstrating two simple physical properties, linear and rotational velocity. The instructor demonstrated linear velocity using a simple ball which was identified and tracked using the camera. Students were able to view vectors overlaid on top of the ball demonstrating how quickly the ball was moving. Rotational velocity was then demonstrated using a simple virtual model of a top, on which additional virtual data was overlaid.

The software support has been well tested. We have utilized the distribution method for our augmented video teleconferencing work for collaboration between the University of California Santa Barbara and the Korean Institute of Science and Technology, with interactive framerates.

Our system exhibits several benefits, which make it advantageous for normal classroom use. Our system is very low cost. The computing power needed to run our software is very reasonable, and virtually any web-camera can be used with this system. The only other cost is that of any physical props which are again fairly negligible. Additionally, the setup time for our solution is very small as the only necessary configuration involves setting up the working plane using ARToolkit.


## Future Work

The first area of improvement will be in increasing the amount of interaction possible when using the physical props. So far, we can reliably track single colored regular shaped props. This allows us to easily integrate a variety of physical phenomena that can be simulated using simple moving props. For example, we can track a pendulum to assist teaching the concepts of angular momentum and torque, or we can augment Newton's cradle to teach conservation of momentum. In order to increase the capabilities of our system to account for more complex physical events, our next step would be to allow instructors to introduce props of irregular shape and different color patterns into the system. This involves incorporation of machine learning and more advanced computer vision techniques, so that we can introduce new objects on the fly, without having to re-program the system.

The second area where we plan future development is in improving the distribution system. Ideally, as many students as possible should be able to view a lecture. In practice, there are limitations due to the computing power and network bandwidth available to the instructor. This is true for any distance learning application. Our solution is scalable with respect to computer hardware and it brings the network cost of local interaction down to the coast of two or three broadcast video streams, but it does not solve the bandwidth issue of broadcast video itself. Fortunately, there are several existing algorithms specially designed for increasing the rate of compression for situations similar to ours. Object based encoding can be used to reduce the size of the video sent to the student's machines. Additional work such as (Cohen-Or 2001) has used optical flow to improve the encoding process. While we have not yet attempted implementation of any of these techniques, future development in this area promises increased performance.


## References

Associated Press. (2007). Scientists make videos for the Web. Retrieved December 10, 2007, from CNN Website: http://www.cnn.com/2007/TECH/12/03/science.youtube.ap/index.html

Cohen-Or, D., Noimark, Y., and Zvi, T. 2002. A serverbased interactive remote walkthrough. *In Proceedings of the Sixth Eurographics Workshop on Multimedia 2001* (Manchester, UK, September 08 - 09, 2001). J. A. Jorge, N. Correia, H. Jones, and M. B. Kamegai, Eds. Springer-Verlag New York, New York, NY, 75-86.

Cooperstock, J.R. 2001. The classroom of the future: enhancing education through augmented reality. *Proc. HCI Inter. 2001 Conf. on Human-Computer Interaction*, New Orleans, USA, 688-692 (2001).

Fjeld, M. P. Juchli and B. M. Voegtli (2003): Chemistry Education: A Tangible Interaction Approach. *Proceedings of INTERACT 2003*, pp. 287-294.

Kaufmann, H., Collaborative Augmented Reality in Education, *Imagina 2003 conference*, Feb. 3rd, 2003.

Liarokapis, F., Petridis, P., Lister, P.F., White, M., Multimedia Augmented Reality Interface for E-Learning (MARIE), *World Transactions on Engineering and Technology Education*, UICEE, 1(2): 173-176, (2002). ISSN: 1446-2257.

Liarokapis, F., Mourkoussis, N., White, M., Darcy, J., Sifniotis, M., Petridis, P., Basu, A., Lister, P.F., Web3D and Augmented Reality to support Engineering Education, *World Transactions on Engineering and Technology Education*, UICEE, 3(1): 11-14, (2004). ISSN: 1446-2257.

Noimark, Yuval. Daniel Cohen-Or. Streaming Scenes to MPEG-4 Video-Enabled Devices. *IEEE Computer Graphics and Applications*, vol. 23, no. 1, pp. 58-64, January/February, 2003.

Shelton, B. E. 2002. Augmented reality and education: Current projects and the potential for classroom learning. *New Horizons for Learning*, 9(1).

White, M., Jay, E., Liarokapis, F., Kostakis, C., Lister, P.F., A Virtual Interactive Teaching Environment (VITE) using XML and Augmented Reality, *The International Journal of Electrical Engineering Education, Manchester University Press*, 38(4): 316-329, October (2001). ISSN: 0020-7209.

Yuanchun Shi, Weikai Xie, Guangyou Xu, Runting Shi, Enyi Chen, Yanhua Mao, Fang Liu. (2003). The Smart Classroom: Merging Technologies for Seamless Tele-Education. *IEEE Pervasive Computing*, vol. 02, no. 2, pp. 47-55, Apr-Jun, 2003