The Intelligent Classroom

David Franklin, Joshua Flachsbart and **Kristian Hammond**
Intelligent Information Laboratory
Northwestern University
{franklin, josh, hammond}@infolab.nwu.edu

Abstract

This paper is an adaptation of an article that appeared in the September/October 1999 issue of the IEEE Intelligent Systems journal. It provides an informal description of the Intelligent Classroom and looks at examples of what happens as the speaker writes on the board, lectures from slides, and does an anatomy lecture. Also, the paper features pretty color pictures.

1 Introduction

Computer software is being designed under the principle that the more features it has, the better it is. Consequently, most people find learning to use a new product overwhelming. What good is having several hundred commands in your word processor, if you can't find the ones you want, and aren't even certain what most of the others do?

The difficulty lies in the way people are expected to interact with their computers. All the effort lies with the users, who must decide what they want to achieve and deduce how they can do it. Intelligent systems should not restrict themselves to following this user-interaction paradigm–they should infer what their users are trying to do. In my research lab, we are developing the Intelligent Classroom, an automated presentation facility that a lecturer can interact with and control.

In the Intelligent Classroom, we are enabling new modes of user interaction through multiple sensing modes and plan recognition. The Classroom uses cameras and microphones to determine what the speaker is trying to do and then takes the actions it deems appropriate. One of our goals is to let the speaker interact with the Classroom as she would with an audiovisual assistant: through commands (speech, gesture, or both) or by just making her presentation and trusting the Classroom to do what she wants.

One way the Classroom assists speakers is by controlling AV components such as VCRs and slide projectors. Additionally, the Classroom lets speakers easily produce fair-quality lecture videos. Based on the speaker's actions, the video cameras pan, tilt, and zoom to best capture what is important. This will allow the presentation of interesting lectures on cable TV, the distribution of videos of entire classes, and the broadcasting of lectures to support distance learning–extending learning beyond the confines of a traditional classroom.

2 System overview

To effectively cooperate with the speaker, the Intelligent Classroom must act appropriately at the right moments. So, even when the Classroom understands the speaker's actions, it still must carefully synchronize its actions with the speaker's. For example, when a speaker goes to the chalkboard to write, the Classroom should use two very different camera techniques: one for when he walks and the other for when he writes. If the Classroom uses the walking technique while the speaker is writing, people viewing the video feed won't be able to read his writing.

To address this challenge, the Classroom uses plan representations that explicitly represent the speaker's actions, the Classroom's actions, and how they should fit together. These plans are intended to represent a common understanding of how a speaker and an AV assistant would interact. When the speaker is doing something, the Classroom monitors his progress through his part of the plan, waiting for the moments when the Classroom needs to act. For example, the "walk over to the chalkboard and write" plan has a process (sequence of actions) for the speaker's actions of moving to the board, stopping at it, beginning to write, and finishing. It also includes processes specifying how the Classroom should film the speaker and ad-

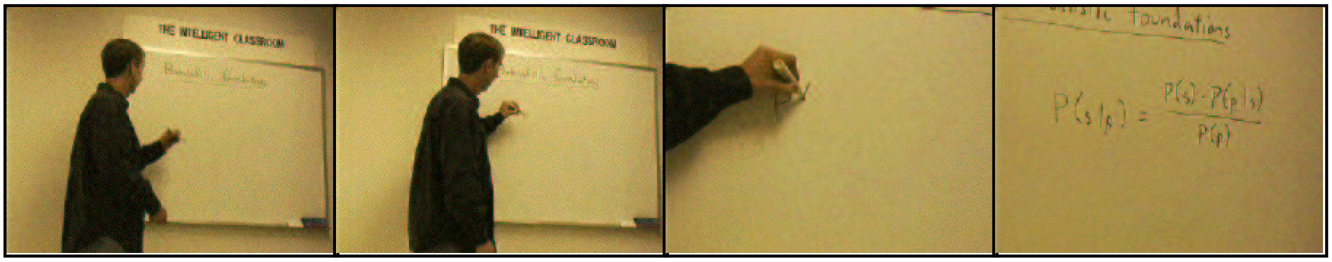


Figure 1: As the speaker writes, the Intelligent Classroom zooms in on the writing.

just the lights. Finally, the plan states that the camera technique should start changing as the speaker enters the chalkboard's vicinity.

The Classroom also uses these plan representations to reason about the speaker's actions at a higher level. While the speaker gives a presentation, the Classroom monitors the processes that serve as its understanding of the activity in the environment. These include processes for both the speaker's actions and the Classroom's actions (such as playing a video or showing a slide). When the Classroom observes the speaker taking an action (such as walking, gesturing, or speaking), it tries to explain this action in the context of its understanding. That is, it looks through these processes for one that predicts that the speaker will perform that action.

However, if the Classroom doesn't find such a process, it must revise its understanding of what the speaker is doing: the speaker apparently isn't doing what the Classroom thought he was. The Classroom then hypothesizes new processes that explain the speaker's action. Initially, there may be several candidate explanations, but when the speaker's future actions contradict some of the proposed processes, they can be rejected, eventually leaving just the speaker's actual activity. (We have described this process in detail elsewhere.[1])

A two-level architecture facilitates the Classroom's physical interaction with the world (sensing through its cameras and microphones and acting through its various actuators). The higher level deals with the world at the level of the various activities (as described above), while the lower level deals with actual sensors and actuators. This lower level links reactive skills and vision modules to form tight control loops that let the Classroom reason about the world at an appropriate abstraction level.[2] The higher level can dynamically configure the lower level–telling it what to look for, what to listen for, or even what techniques to use.

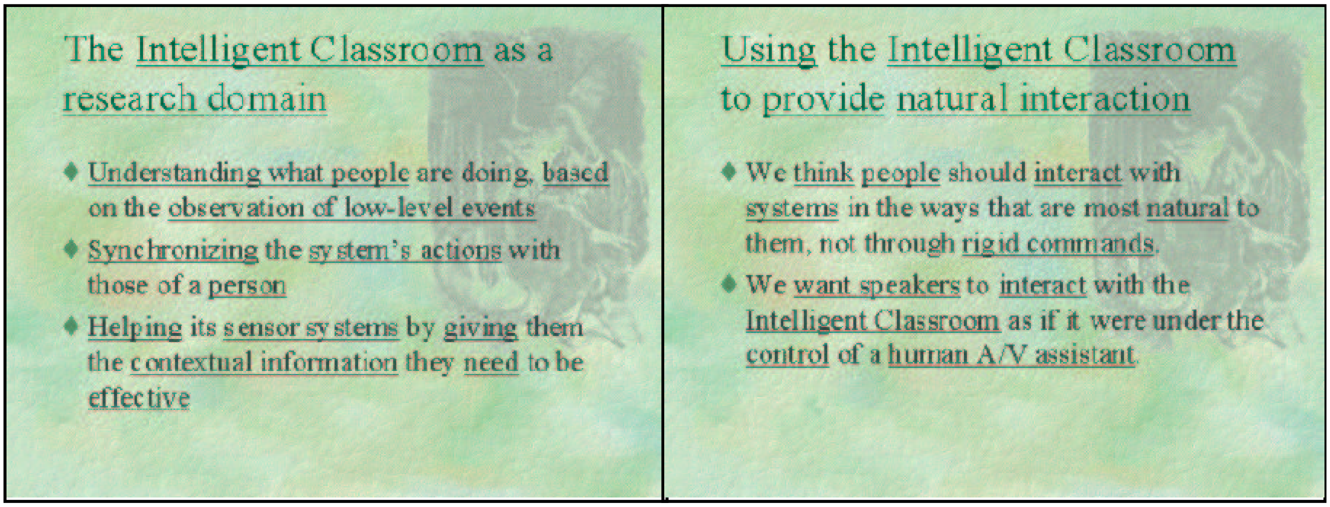
3 Working with the chalkboard

To get a taste of the Intelligent Classroom, let's look at several example interactions. First, we'll examine how the Classroom deals with the chalkboard. Even though the chalkboard isn't technologically interesting, it's still the most popular presentation medium out there. And, because people viewing a video feed of the presentation need to be able to read what is written, it provides interesting filming challenges. To be effective, the Classroom must monitor the board's contents and reason about how the speaker will use the board.

In addition to writing on the board, a speaker will often erase writing, make changes, or refer to particular points. To deal with these, the Classroom requires a notion of how the board's contents are arranged. When the speaker changes a sentence, the video feed must show not only the changes but also the entire sentence. Otherwise, someone viewing the video feed might be unable to understand the change's effects. When the speaker points at an equation on the board, the video feed should focus on the equation. In both these scenarios, the presentation camera has to zoom in on something particular (otherwise people won't be able to read what's important). But to determine what's important, the Classroom needs to know what's on the board. However, the Classroom isn't able to read what the speaker has written.

So, the Classroom maintains a representation of the regions of the board on which the speaker has written and uses those to understand the speaker's actions. If the speaker is writing in an area in which she has already written, the Classroom understands her to be revising what she wrote, and the camera should show the entire area. Similarly, if she points at something on the board, the Classroom can determine at which area she is pointing and show it.

Figure 1 shows what happens as a result of the Classroom's representation of the board as a speaker writes. When the speaker begins writing, the Classroom recognizes that he is writing in a clear area of the board,



The Intelligent Classroom as a research domain

- Understanding what people are doing, based on the observation of low-level events
- Synchronizing the system's actions with those of a person
- Helping its sensor systems by giving them the contextual information they need to be effective

Using the Intelligent Classroom to provide natural interaction

- We think people should interact with systems in the ways that are most natural to them, not through rigid commands.
- We want speakers to interact with the Intelligent Classroom as if it were under the control of a human A/V assistant.

Figure 2: Two slides with key words and phrases underlined. The Classroom finds these words and phrases through syntactic analysis.

so it zooms in on the area in which he is writing and keeps track of the boundaries of his writing. Once he is finished, the Classroom has framed all that he wrote. All this happens without the speaker giving any explicit commands–because the Classroom knows what the speaker is doing, it knows how to respond.

4 Lecturing from slides

The Classroom can also assist a speaker by controlling Microsoft PowerPoint slide presentations. The speaker can tell the Classroom what to do by just saying what he wants (“Next slide, please”), making a gesture to go on or go back, or touching the Next button that is drawn on the slide. Also, because the Classroom can access the contents of the slides, the speaker can lecture from the slides as the Classroom follows along, changing the slides at the appropriate moments.

Behind this last approach is the idea that the Classroom uses a shallow (syntactic) understanding to match the speaker's words to important phrases in the slides. The slides in Figure 2 show underlined key phrases that the Classroom found through syntactic analysis. As the speaker talks, the Classroom listens for these phrases (and a number of simple grammatical variants) to keep track of which slide the speaker is discussing and where she is in the slide. When the speaker starts speaking phrases from the next slide, the Classroom knows to display the next slide. (In Figure 2, the Classroom switched to the slide on the right after the speaker said “We want people to be able to naturally interact with the Classroom.”) If the speaker skips around in the slides (to answer a question or to otherwise deviate from the planned slide order), the Classroom will wait until it is certain which slide she wants to skip to before switching slides.

Fortunately, this purely syntactic understanding appears sufficient for knowing when to switch slides. This isn't entirely unexpected: you can expect a human AV assistant to match what the speaker is saying to the contents of the slides without any technical knowledge of the presentation's subject matter. An additional benefit of using a shallow understanding is that the Classroom can better deal with the inevitable speech-recognition errors. Even when the Classroom mishears half the speaker's words (not unreasonable with a conversational speaker), the Classroom will still hear enough of the important phrases to advance the slides appropriately.

After achieving some encouraging results with the Classroom's slide controller, we have made a standalone version, called Jabberwocky, which can run on a laptop computer. Jabberwocky gets its name from the Lewis Carroll nonsense poem from “Through the Looking Glass” in which, even though most of the words are made up, the reader can still get the gist of what's happening. As Alice said, after reading the poem, “Somehow it seems to fill my head with ideas–only I don't exactly know what they are! However, somebody killed something: that's clear at any rate.”

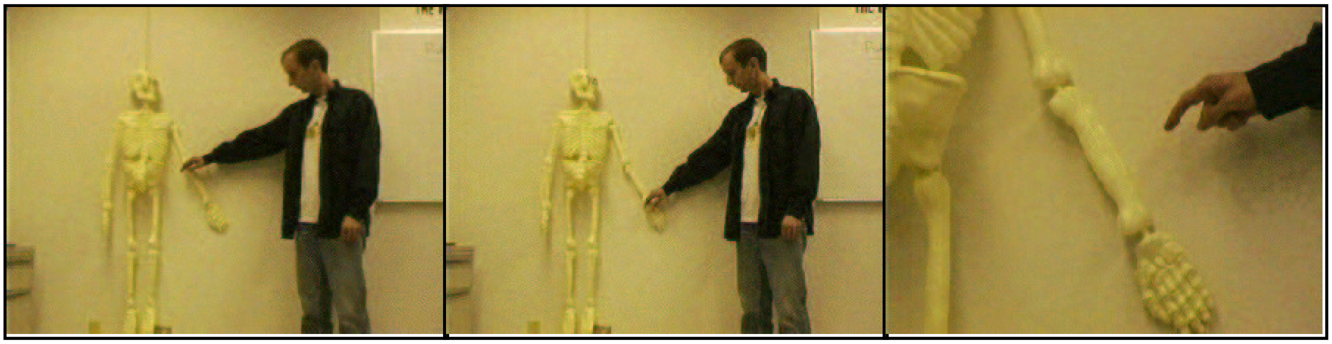


Figure 3: The speaker specifies the camera's focus area by pointing to each end of a bone.

5 Adding lecture-specific knowledge

While we hope that the Intelligent Classroom's default behaviors will suffice for typical presentations, we also are working on ways of letting speakers tell the Classroom how to deal with certain events in their presentations. For example, in an anatomy lecture, the speaker might want the Classroom to zoom in on the appropriate bones in a skeleton. Without a knowledge of skeletal structure and without the ability to see the ends of bones, the Classroom cannot do a good job of showing the skeleton's parts.

In these situations, the speaker and the Classroom must agree on how the speaker can aid the Classroom. Figure 3 shows one way that the speaker could help the Classroom frame the bones. When the speaker wants the Classroom to zoom in on a particular bone, he touches each end of the bone, giving the Classroom the information it needs to accurately frame the bone.

6 Conclusion

The Intelligent Classroom embodies a promising philosophy of human-computer interaction. Although it will respond when the speaker commands it, the Classroom encourages the speaker to just go about his presentation, while the Classroom determines how it can assist. In building intelligent systems that try to understand what their users are doing, we are building cooperative systems with which people can naturally and intuitively interact.

7 About the authors

David Franklin is a doctoral candidate at the Intelligent Information Laboratory in Northwestern University's Computer Science Department. His research interests include AI, multimodal human-computer interfaces, plan recognition, and robotics. He earned his BS in computer science from the University of Washington and his MS in computer science from the University of Chicago. He is a member of the AAAI. Contact him at the Intelligent Information Laboratory, Northwestern Univ., 1890 Maple Ave., Evanston, IL 60201; franklin@ils.nwu.edu.

Joshua D. Flachsbart is a graduate research assistant in the Intelligent Information Laboratory at Northwestern University. His research interests include using goals and context to improve real-time active computer vision, and the more general problem of combining sensing and action. He received his BA in physics and MS in computer science from the University of Chicago. Contact him at the Intelligent Information Laboratory, Northwestern Univ., 1890 Maple Ave., Evanston, IL 60201; josh@ils.nwu.edu.

Kristian J. Hammond is a professor in the Department of Computer Science at Northwestern University and the director of the university's Intelligent Information Laboratory. His technical interests are in the application of AI to problems of information capture, management, and retrieval. He received his BA in philosophy and his MS and PhD in computer science, all from Yale University. Contact him at the Intelligent Information Laboratory, Northwestern Univ., 1890 Maple Ave., Evanston, IL 60201; hammond@ils.nwu.edu

8 References

1. D. Franklin, "Cooperating with People: The Intelligent Classroom," Proc. AAAI '98: 15th Nat'l Conf. Artificial Intelligence, AAAI Press, Menlo Park, Calif., 1998, pp. 555-560.

2. R.J. Firby et al., “An Architecture for Vision and Action,” Proc. IJCAI ’95: Int’l Joint Conf. Artificial Intelligence, Vol. 1, Morgan Kaufmann, San Francisco, 1995, pp. 72-81.