Chapter XX
Rationale, Design
and Implementation of
a Computer Vision-Based
Interactive E-Learning System

Richard Y. D. Xu
Charles Sturt University, NSW, Australia

Jesse S. Jin
University of Newcastle, NSW, Australia

ABSTRACT
This article presents a schematic application of computer vision technologies to e-learning that is synchronous, peer-to-peer-based, and supports an instructor’s interaction with non-computer teaching equipments. The article first discusses the importance of these focused e-learning areas, where the properties include accurate bidirectional interaction and low cost hardware; system portability and versatile vision technology are emphasized. In the subsequent sections, we present some results aiming to achieve these goals. In particular, we highlight the most recent advancements in the interactive PTZ camera control from both the instructor and remote student. We also illustrated how these results have successfully addressed the challenges.

INTRODUCTION
In recent times, there has been an increase in the research activities aiming to apply computer vision (CV) technologies to various automated e-learning multimedia systems.

In this article, we will discuss our progressive work in Intelligent Video Detection Agent (IVDA), which is a scheme of hardware design and computer vision software algorithms to assist e-learning systems that are synchronous, peer-to-peer based and an instructor uses non-computer based teaching equipment.
Before we present our prototype system, we need to explain the importance of these three properties, and illustrate how it differs from the other existing CV e-learning systems’ focuses.

Synchronous E-Learning

In synchronous e-learning, the video and teaching multimedia are exchanged between the students and instructors in real-time, creating a highly interactive teaching and learning environment. Synchronous e-learning provides the students with instantaneous feedbacks and provides the instructor with a platform from which to monitor and adapt to student’s activities (Soreanu & Saucan, 2003).

Computer vision technologies are used in various synchronous e-learning systems to replace many manual, laborious and time-consuming tasks, which made real-time automated camera control and instantaneous multimedia authoring possible.

Peer-to-Peer E-Learning

E-learning application can be classified according to its participant numbers, into peer-to-peer (P2P) and institutional ones. In P2P e-learning, there is usually one instructor and one student. Both participants can be of a variety of types, including home-based computer users. The learning can be both formal and informal.

Our work is focusing on P2P e-learning, since it allows for various types of participants, more importantly, the average computer users to take part in either or both the student and instructor role. It is an inexpensive learning platform and many studies can be identified to argue its advantages, such as literature in Jokela (2003) and Fletcher (2004).

However, the widespread of P2P e-learning participants also means that there will be additional considerations that need to be taken into account when designing computer vision software to support it. These considerations include:

1. The CV software must be able to intelligently assist the instructor when he works alone.
2. The CV software must adapt to standard equipment of an average computer user.

Non-Computer Based Interaction

In the current synchronous e-learning applications, the instructor and student(s) usually communicated via standard computer interfaces, using chat window, computer whiteboards, sharable text and drawing pads, for example (Deshpande & Hwang, 2001).

However, many researchers have argued that e-learning cannot replace traditional learning altogether. In a similar argument, Liu & Kender (2004) mentioned that, based on a recent survey carried out at University of South Carolina, students consider traditional blackboard presentation as “essential” and “indispensable.”

Therefore, allowing the instructor, at least, to use traditional teaching equipments and video-capturing and streaming such information to student in real-time becomes a compromise solution between modern technology and traditional pedagogy.

However, most of the operator-less instructional video-capturing is achieved through the use of static camera(s) without computer vision processing. While this method is sufficient for some applications, the crude streaming suffers from an uninteresting presentation of teaching materials and is dependent on the instructor’s verbal instruction and remote student’s visual perception to determine what the student should pay attention to. In addition, the instructor is required to constantly check the streaming video results, such as the position of his/her body with regard to the camera view, or the size of his/her whiteboard writing. This constant distraction makes uninterruptible teaching difficult during a synchronous e-learning session.

For this reason, computer vision has played an important role in detecting classroom events