

## Thesis

---

# Multi-camera vision for smart environments

Chen Wu

Stanford University, USA

E-mail: chenwu@stanford.edu

**Abstract.** On September 21, 2010, the author successfully defended her Ph.D. thesis entitled *Multi-Camera Vision for Smart Environments* at Stanford University.

### 1. Thesis summary

Technology is blending into every part of our lives. Instead of having them as intruders in the environment, people prefer the computers to go into the background, and automatically help us at the right time, on the right thing, and in the right way. A smart environment system senses the environment and people in it, makes deductions, and then takes actions if necessary.

Sensors are the “eyes” of smart environments. In this dissertation we consider vision sensors. Image and video data contain rich information, yet they are also challenging to interpret. From obtaining the raw image data from the camera sensors to achieving the application’s goal in a smart environment, we need to consider three main components of the system, i.e., the hardware platform, vision analysis of the image data, and high-level reasoning pertaining to the end application. A generic vision-related problem, e.g., human pose estimation, can be approached based on different assumptions. In our approach the system constraints and application’s high-level objective are considered and often define boundary conditions in the design of vision-based algorithms.

A multi-camera setup requires distributed processing at each camera node and information sharing through the camera network. Therefore, the computation capacity of camera nodes and the communication bandwidth can define two hard constraints for algo-

rithm design. We first introduce a smart camera architecture and its specific local computation power and wireless communication constraints. We then examine how the problem of human pose detection can be formulated for implementation on this smart camera, and describe the steps taken to achieve real-time operation for an avatar-based gaming application.

We then present a human activity analysis technique based on multi-camera fusion. The method defines a hierarchy of coarse and fine level activity classes and employs different visual features to detect the pose or activity in each class. The camera fusion methods studied in this dissertation include decision fusion and feature fusion. We show the results of experiments in three testbed environments and analyze the performance of the different fusion methods.

Although computer vision already involves complicated techniques in interpreting the visual data, it still constitutes just the sensing part of a smart environment system, and further application-related high-level reasoning is needed. Modeling such high-level reasoning will depend on the specific nature of the problem. We present a case study in this dissertation to demonstrate how the vision processing and high-level reasoning modules can be interfaced for making semantic inference based on observations. The case study aims to recognize objects in a smart home based on the observed user interactions. We make use of the relationship between objects and user activities to infer the

objects when related activities are observed. We apply Markov logic network (MLN) to model such relationships. MLN enables intuitive modeling of relationships, and it also offers the power of graphical models. We show different ways of constructing the knowledge base in MLN, and provide experimental results.

## 2. Ph.D. Oral Defense Exam

The Ph.D. Oral Defense took place on September 21, 2010, at Stanford University. The defense committee members included Prof. Hamid Aghajan (principle advisor), Prof. Benjamin Van Roy, Prof. Fei-Fei Li, and Prof. James Harris (chair). I am deeply grateful to Prof. Aghajan for his advice throughout my Ph.D. studies, and I am thankful for the insightful comments provided by the committee members.



Fig. 1. Prof. Hamid Aghajan introduces Chen Wu at the Defense Exam.