

Detection of Hand Raising Gestures Utilizing an Omni-directional Video Sensor For Use in a Remote Learning Application



Bill Kapralos^{1,3}, Michael Jenkins^{1,3}, Evangelos Milios^{2,3}, John K. Tsotsos^{1,3}

{billk, jenkins, tsotsos}@cs.yorku.ca eem@cs.dal.ca

¹Department of Computer Science, York University, Toronto Ontario, Canada M3J 1P3

²Department of Computer Science, Dalhousie University, Halifax Nova Scotia, Canada B3H 3J5

³Centre for Vision Research, York University, Toronto Ontario, Canada M3J 1P3

Introduction

● Remote Classroom Interaction:

- Allows for real time transmission of voice and full motion video.
- Permits for live discussion and immediate feedback.
- Creates a social presence.

● Limitations of Existing Remote Classroom Interaction Systems:

- Use traditional cameras, thereby providing a limited number of static or manually tracked views.
- Cannot localize multiple speakers in the audio domain.
- Detecting students wishing to interact with the instructor.

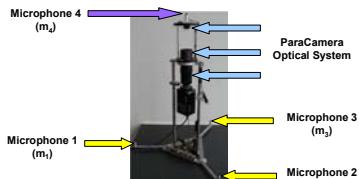
Project Goals

● Develop a teleconferencing system integrating audio and video cues.

- In multiple student setting, being capable of attending to a student signaling their intent to interact with the instructor:
 - Students may speak or raise their hand to attract instructor's attention.
 - In this work, we address signaling by hand raising gestures.

Hardware

● Novel Sensor Combining Audio and Video Systems:

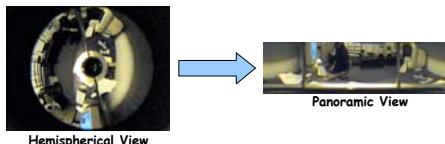


- Economical, lightweight and portable.
- Mounted in a remote classroom.

Video System

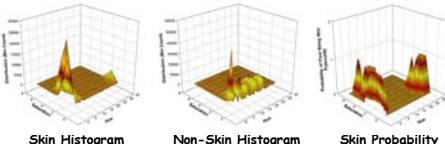
● Omni-directional Video Sensor (Cyclovision's ParaCamera) Captures a View of the Entire Visual Hemisphere from a Single Viewpoint:

- Hemispherical View is easily "un-warped", allowing for multiple dynamic views of the scene.



● Color Based Human Skin Classification:

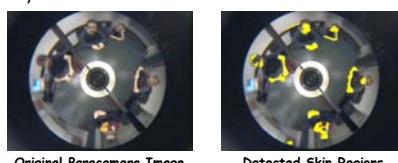
- HSV Color models for both **skin** and **non-skin** colors, constructed by manually classifying portions of ParaCamera images as either skin or non-skin.



- Bayes' rule is used to classify the pixels of each image as either skin or non-skin.

- Regions of "skin" pixels are grouped together.

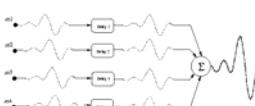
- Estimate of the direction of each skin region in the real world can be made and provided to the audio system.



Audio System

● Beamforming Focuses Four Microphones ($m_1 \dots m_4$) to the Direction of the Skin Region:

- Appropriately delaying the signal of each microphone ensures the signals are in phase.



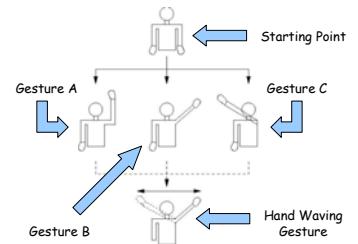
- Signal is reinforced, while noise and sound emanating from other directions are attenuated.

Raising Hand Gestures Defined

● Students wishing to interact with instructor will raise their hand:

- Difficult to detect raising hand gestures due to the large variation.
- Trajectory followed by arm doesn't follow any set pattern or path.
- May move straight up, diagonally to the left/right etc.

- Rather than a single gesture, several raising hand gestures are defined.



- **Starting Point:** Prior to Raising the hand, it is at some position below the head.

- **Gesture A:** Hand raised straight upward from starting point to above the head.

- **Gesture B and C:** Hand raised upward diagonally to the right/left respectively.

- **Hand Waving Gesture:** After hand is raised, it may possibly be waved.

Detecting Hand Raising Gestures

● Manual classification of Paracamera images in simulated classroom setting:

- Construction of statistical models capturing info. when hand is raised.
- Hand motion patterns.
- "Form" of skin region corresponding to the hand e.g. size, length vs. width.

● The Statistical Models and a Bayesian Framework are Used to Classify Detected Skin Regions of Incoming Paracamera Images:

- Students raising their hand can be detected.
- Beamformer focused in direction of the student wishing to interact.