The Wits Intelligent Teaching System
Detecting Student Engagement During Lectures Using CNNs

Richard Klein[1] and Turgay Celik[1,2]

[1] University of the Witwatersrand, Johannesburg, South Africa
[2] Southwest Jiaotong University, Sichuan Sheng, China

Wits Intelligent Teaching System (WITS)

Contingent Teaching [1] requires being responsive to students' needs during class.

- Lecturers need to be able to **assess the state of the audience** at a glance.
- Take steps to **positively impact learning.**

WITS is an automated system to report on behavioral engagement.

Behavioral Engagement

- Direct labelling resulted in poor inter-rater reliability (Cohen's Kappa).
- Observational checklist looks for common behaviors.

Engaged:
- Writing Notes
- Asking Questions
- Looking at the Lecturer/Notes
- Straight/Upright Posture

Not Engaged:
- Cellphone
- Sleeping
- Looking at the roof/friends/under desk/etc.
- Yawning
- Slouched Posture, Head on Desk, etc.
- Talking with Friends

Data (WITSDB)

- 3x45 Minute First Year Lectures Recorded
- UHD (4K) Resolution
- 8 FPS
- 65 Unique Subjects – varying number per lecture
- Labelled by multiple raters, labels with low inter-rater reliability were discarded.

Interest Map

WITS

- Bounding Boxes Only
- 50 subjects per frame
- HOG SVM: 1.6 fps (8 Core CPU Parallel) [2]
- AlexNet [3] Based CNN: 7 fps (GPU Acceleration, NVidia Titan X)
- Generates Live Interest Map
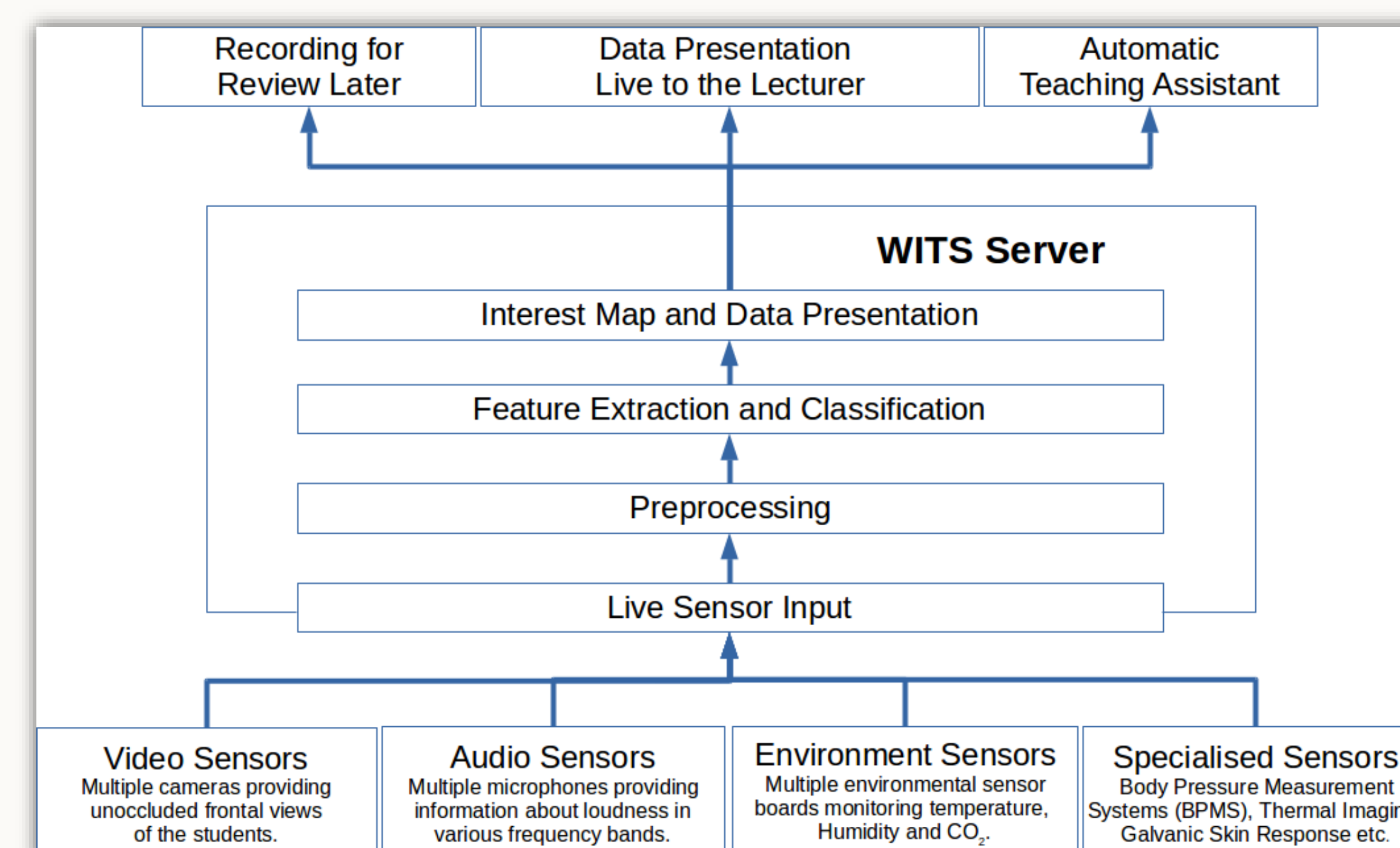
Results

Frames (All Subjects)
- Highly Correlated in Time
- HOG SVM: 72.6% testing accuracy, 73% validation accuracy
- CNN: 89.6% testing accuracy; 89.6% validation accuracy (k=0.792)

Lecture 1 Subjects Only
- Balanced Validation Set: 441 154 labelled images
- 50 Subjects, 50 Fold Cross Validation
- Random Chance: 50% accuracy
- HOG SVM: 53.5% accuracy on validation set
- CNN: 59% accuracy on full validation set
  - Max Accuracy: 91.5%
  - Min Accuracy: 50%
  - Mean Accuracy: 65.6%

Future Work
- Increase data set variety
- Increase labels
- Apply Sequence Methods (eg. LSTM) to learn image sequences

[1] Janneke van de Pol, Monique Volman, and Jos Beishuizen, "Patterns of contingent teaching in teacher–student interaction," Learning and Instruction, vol. 21, no. 1, pp. 46–57, 2011
[2] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no.3, pp. 27, 2011.
[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.