The Automated Lecture Capturing and Broadcasting System in MSR's 1159

Cha Zhang

Sep. 9, 2005

Joint work with Yong Rui, Jim Crawford and Li-wei He.

Motivation

- Cost, cost, cost!
- Fixed Cost
  - Computer equipments, cameras, microphones
  - One-time cost
- Re-occuring staffing cost
  - Pre-lecture (setting up the equipment)
  - During-lecture (tracking speaker/audience, switching between cameras)
  - Post-lecture (merge AV streams with slides, post the talk online)
  - This can be REALLY expensive!

Outline

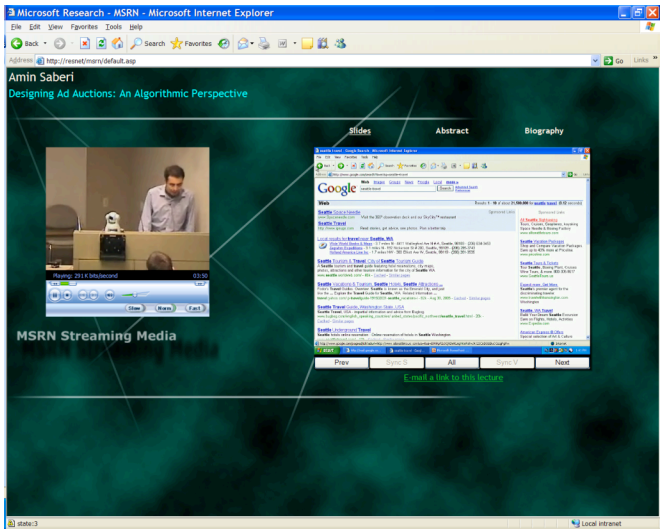
- Design goals
- System architecture
- Automated lecture capturing – iCam2
- System usage statistics
- Conclusions and future work

Outline

- **Design goals**
- System architecture
- Automated lecture capturing – iCam2
- System usage statistics
- Conclusions and future work

Design Goals

- As a user
  - High quality synchronized audio/video/slides
  - Live
    - No sacrifice on audio/video/slides quality
  - On-demand
    - Immediately available after the talk
    - Easy browsing
    - View the lecture at his/her own pace



http://resnet

Microsoft Research - MSRN - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address http://resnet/msrn/default.asp Go Links

Microsoft Research Netcasting

<< < Sep > >>

S	M	T	W	T	F	S
28	29	30	31	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	1
2	3	4	5	6	7	8

Today

Search

RSS upcoming talks

RSS now available OD

MSRN FAQ

Friday September 9, 2005

Dawn Song

Sting: an Automatic Defense System against Zero-day Worm Attacks

Available

9/9/2005 10:30:00 AM

Duration: 1:15:09

Click To Play

Abstract Biography Details

Cha Zhang

WACE 2005 - The Automated Lecture Capturing and Broadcasting System in MSR's 1159

Pending

9/9/2005 3:15:00 PM

Click To Play

Abstract Biography Details

You may already know that lectures given in 113/1159 are being recorded by an automated lecture capturing system on a daily basis. When there is a lecture in progress, you always have the option to go to http://resnet and watch the lecture live online. Or, you may watch it afterwards by visiting the same site, and have the freedom to pause/fast forward/fast backward, etc. The system has been used on a daily basis for more than 4 years, during which 522 lectures were captured with 20,000+ online viewers. In this talk, I will show you how we make such a fully automated system. I will first give an overview on the system architecture, which minimizes the pre- and post-production time, and supports capturing, broadcasting, viewing, archiving and search. Then I will focus on our second generation automated lecture capturing system called iCam2, developed in the past 2 years. ICam2 can synchronously capture all the contents of the lecture, including audio, video and visual aids. Improvement over the previous iCam system include much better portability, hybrid speaker tracking, enhanced audio quality, scripting language for virtual director configuration, etc.

Rakesh Vohra

Predicting the 'Unpredictable'

Pending

9/9/2005 3:30:00 PM

Click To Play

XML Load Completed

Local intranet

Microsoft Research - MSRN - Microsoft Internet Explorer

File Edit View Favorites Tools Help

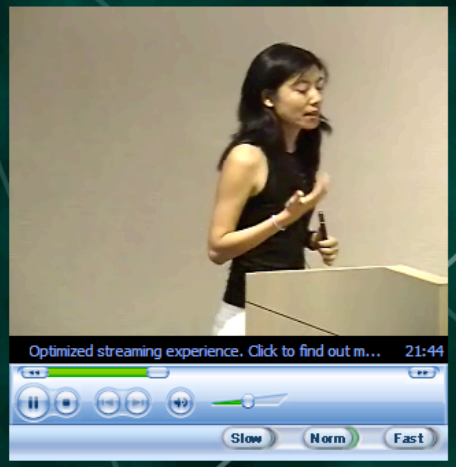
Back Search Favorites

Address http://resnet/msrn/default.asp Go Links

Dawn Song

Sting: an Automatic Defense System against Zero-day Worm Attacks

Slides Abstract Biography



MSRN Streaming Media

Comparison to Previous Mechanisms

- **Used Wilander testbed [NDSS03]**
  - **20 exploit tests**
  - **Overwrite Targets: return address, base pointer, function pointer, longjmp buffer**
  - **Overwrite Techniques: overflow to target, overflow to pointer to target**
  - **Evaluate previous run-time detection mechanisms**

20

Prev Sync S All Sync V Next

E-mail a link to this lecture

state:3 Local intranet

start Microsoft PowerP... Microsoft Researc... EN 2:13 PM

Design Goals (cont.)

- As a speaker
  - No need to send slides before or after the talk
  - No restrictions on his/her behavior during the talk
- As a system administrator
  - One click to start and stop the lecture recording/broadcasting
  - Automated capturing/broadcasting process
  - No tedious pre/post production
- As a system designer
  - Make the system portable to various room configurations

Summary of Key Solutions

Goal	Our Solution	
Synchronized audio/video/slides, no pre/post production, on-demand viewing at user's own pace	Specially designed system architecture	Automated broadcasting
High quality audio capturing	Sound source localization, beam forming, digital audio mixing	Automated content capturing
No slides from the speaker, high quality slide capturing	Live high-resolution slide capturing	
No restriction on speaker's behavior, automated lecture video capturing	Hybrid speaker tracking, sound source localization	
Portable to various room configurations	IP PTZ camera, hybrid speaker tracking, scripting language for virtual director rules, Direct Show implementation	

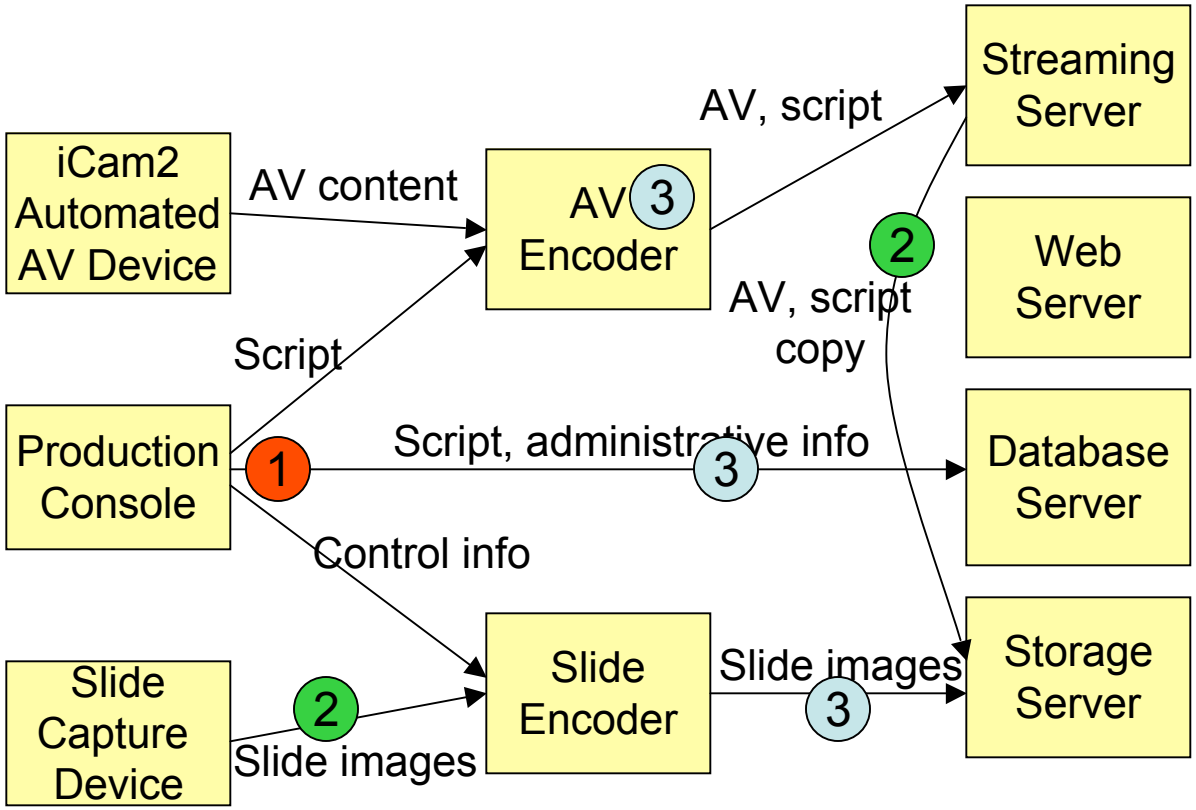
Related Work

- Automated content capturing
  - STREAMS (Cruz94)
  - AutoAuditorium (Bianchi98)
  - Classroom 2000 (Abowd99)
  - Cornell Lecture Browser (Mukhopadhyay99)
  - iCam (Rui04)
  - Many, many more…
- Automated live/on-demand broadcasting
  - BIBS (Rowe01)
  - ePresence (Baecker03)
  - Apreso Classroom (AnyStream)

Outline

- Design goals
- **System architecture**
- Automated lecture capturing – iCam2
- System usage statistics
- Conclusions and future work

System Architecture

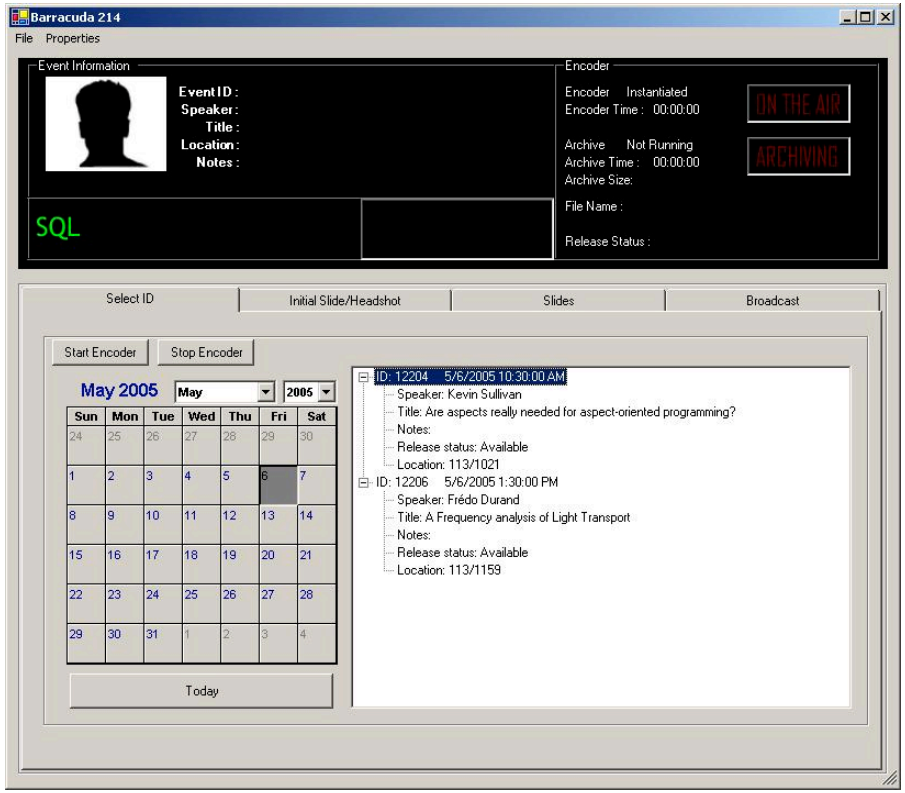


1 Synchronized audio/video/slides

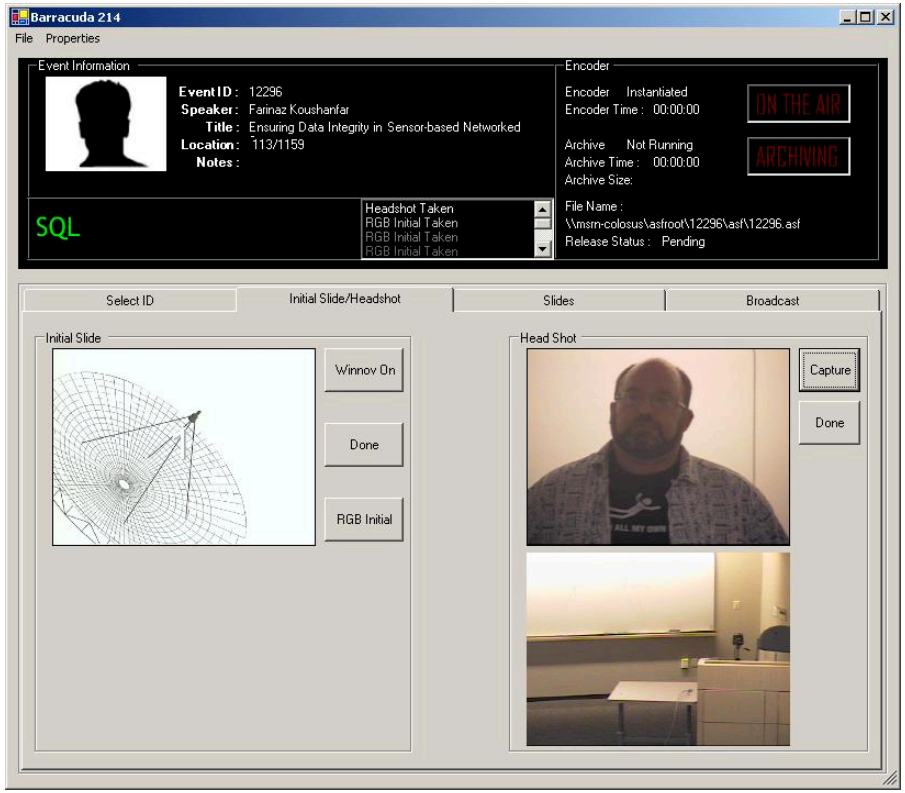
2 No pre/post production

3 On-demand viewing at user's own pace

Management Console



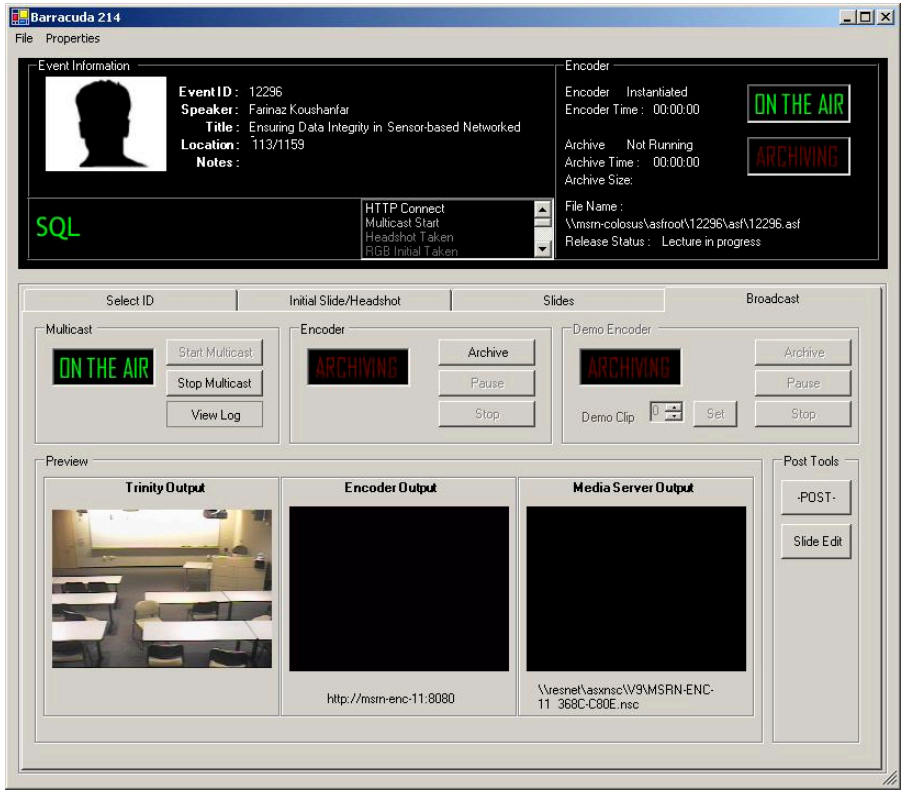
Select a lecture



Initial head shot

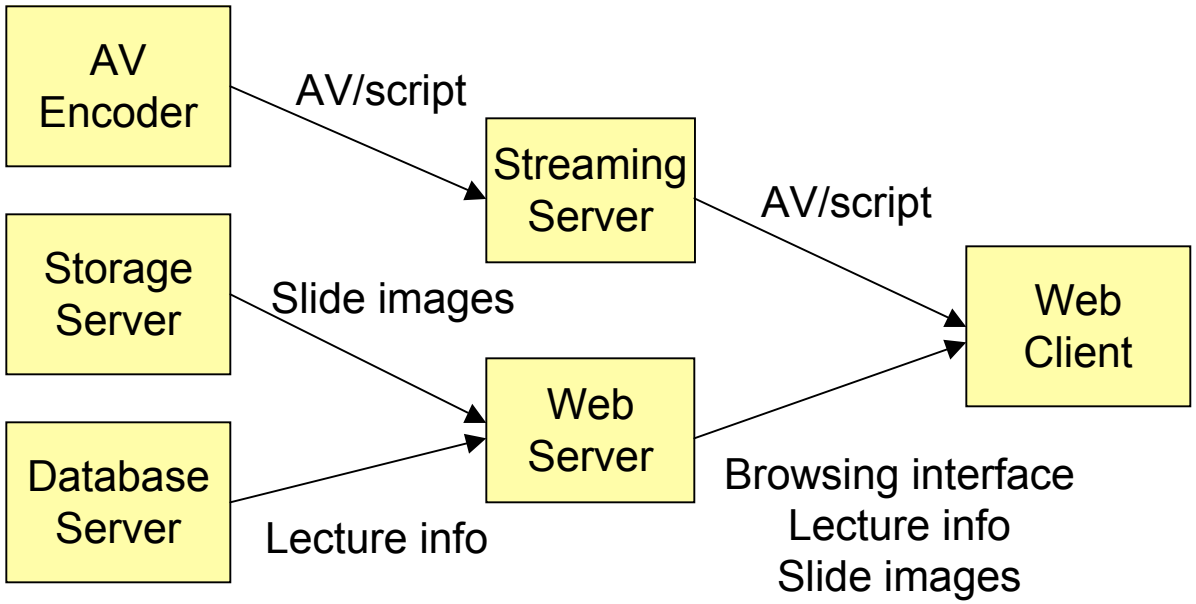
Management Console (cont.)

Initialize slide detection

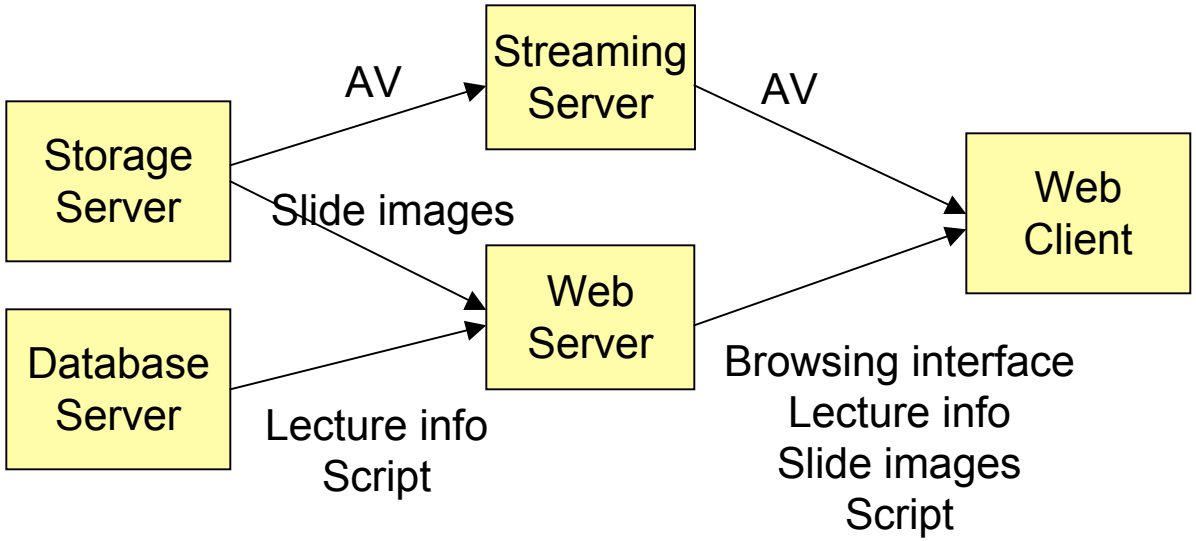


Broadcasting in progress

Live Broadcasting



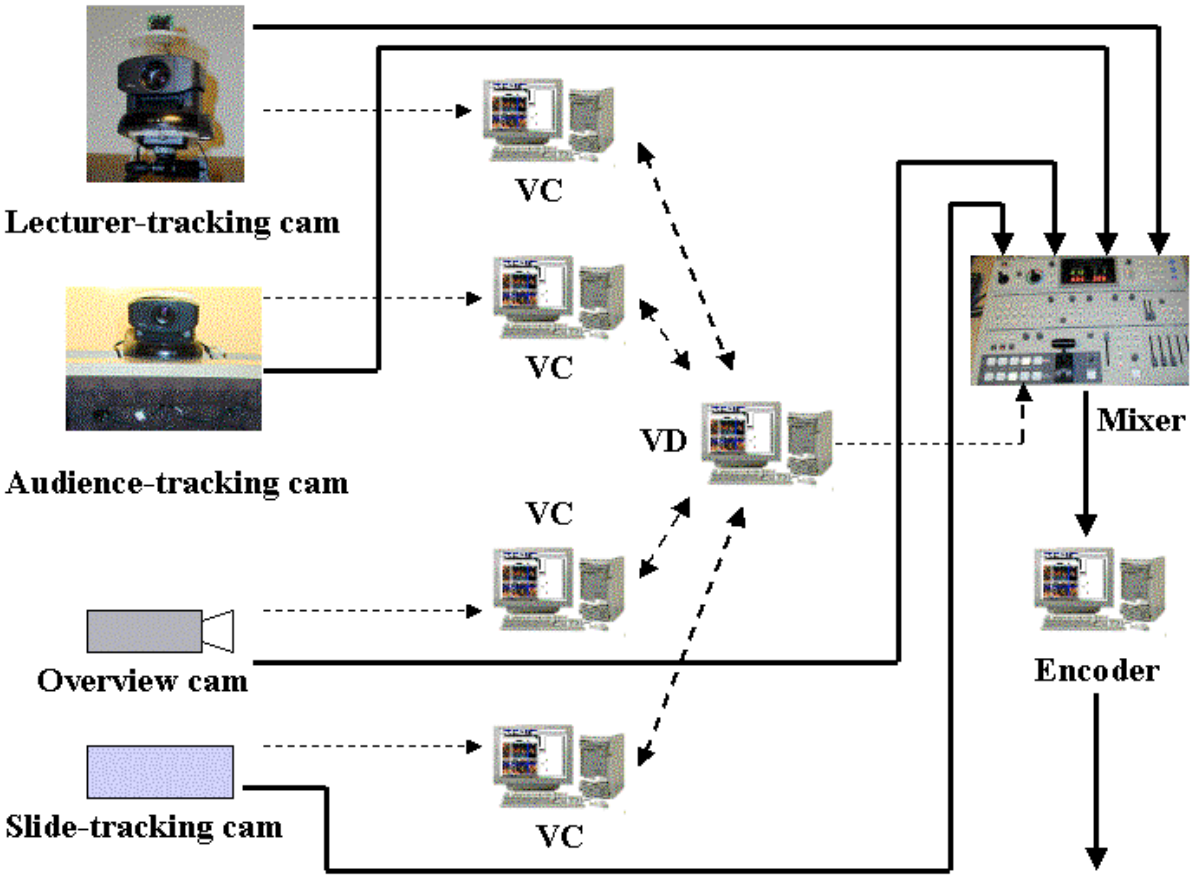
On-Demand Broadcasting



Outline

- Design goals
- System architecture
- **Automated lecture capturing – iCam2**
  - **The 1st generation iCam system**
  - **Audio**
  - **Slides (visual aids)**
  - **Video**
- System usage statistics
- Conclusions and future work

The 1st Generation iCam System



iCam and iCam2

Dual camera for speaker tracking

Analog mixer

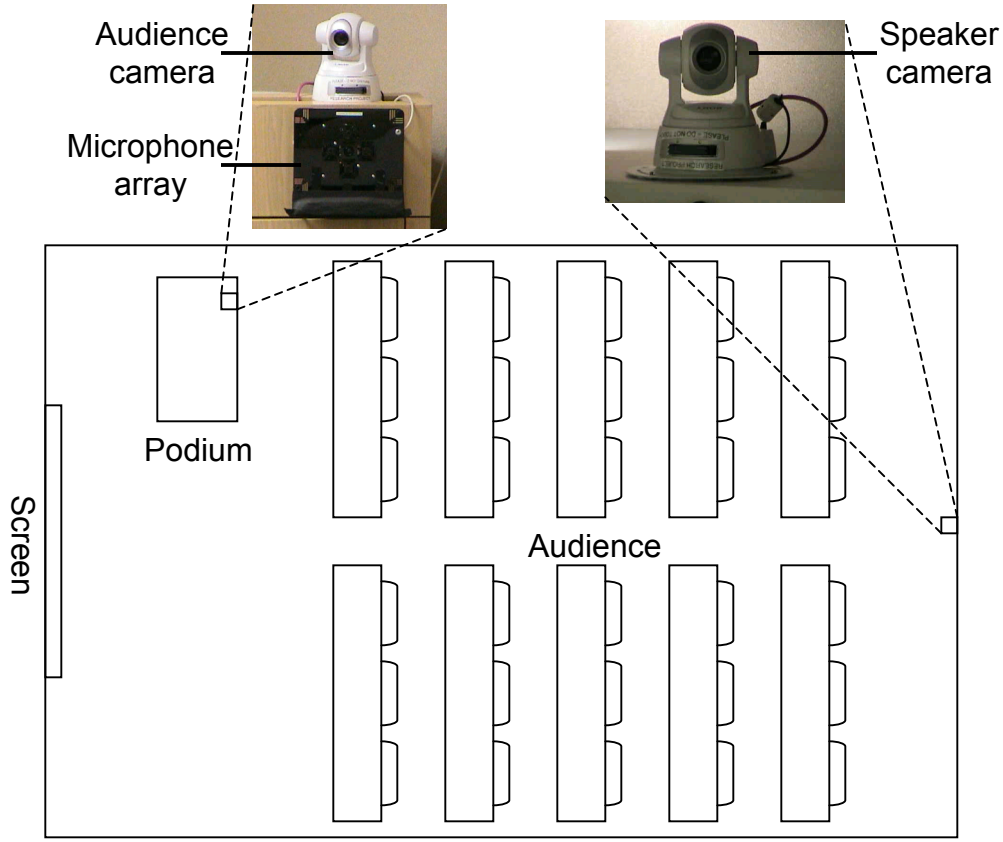
Miles of wires

Single PTZ network camera for speaker tracking

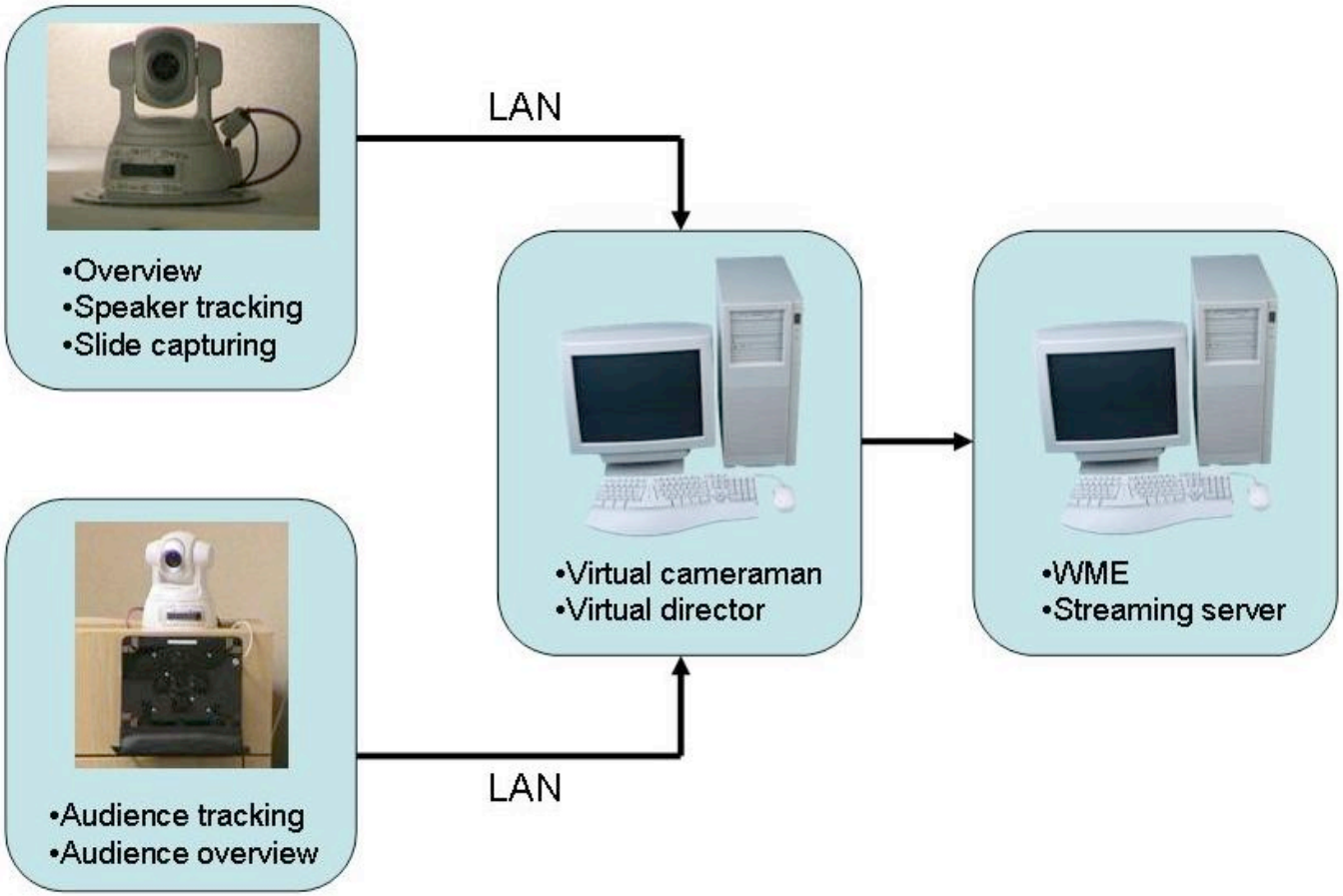
Digital mixing

Great portability!

iCam2 Configuration



The 2nd Generation iCam2 System



What's New in iCam2

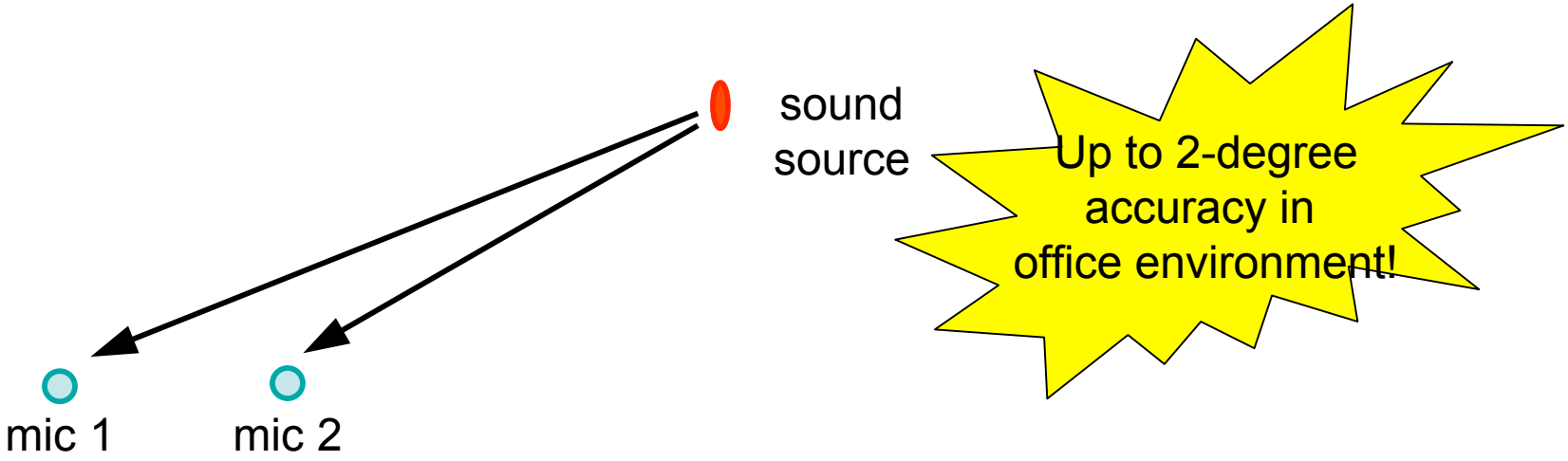
- Audio
  - High quality audience voice capturing with 8-element microphone array
    - Sound source localization
    - Beam forming
    - Intelligent digital audio mixing
- Visual Aids
  - Automated slide change detection
- Video
  - Portable design with network cameras
  - Hybrid speaker tracking
  - Scripting Language for virtual director rules
  - Direct Show implementation also enhances portability

What's New in iCam2

- **Audio**
  - **High quality audience voice capturing with 8-element microphone array**
    - **Sound source localization**
    - **Beam forming**
    - **Intelligent digital audio mixing**
- Visual Aids
  - Automated slide change detection
- Video
  - Portable design with network cameras
  - Hybrid speaker tracking
  - Scripting Language for virtual director rules
  - Direct Show implementation also enhances portability

Audio – Sound Source Localization

- Estimate where the sound source is coming from
- Why?
  - Identify who is asking a question
  - Selectively amplify the questioner's voice
- How? (TDOA based, Rui04)



Audio – Sound Source Localization

- Two microphone case: handle ambient noise and room reverb.

- More than two microphone: hypothesis-testing approach (single step) by decomposing the weighting function to each individual microphone

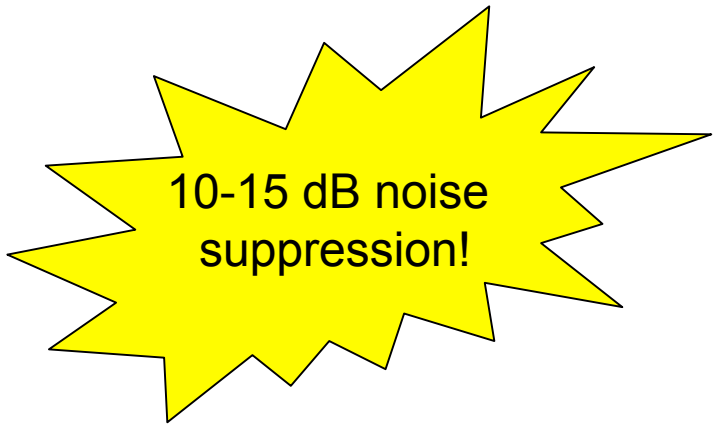
Audio – Beam Forming

- Make the microphone array to listen to a given location (from SSL), suppressing the signals coming from other locations
- Weighted sum of the input audio signals

$$Y(f) = \sum_{i=0}^{M-1} W(f,i) X_i(f)$$

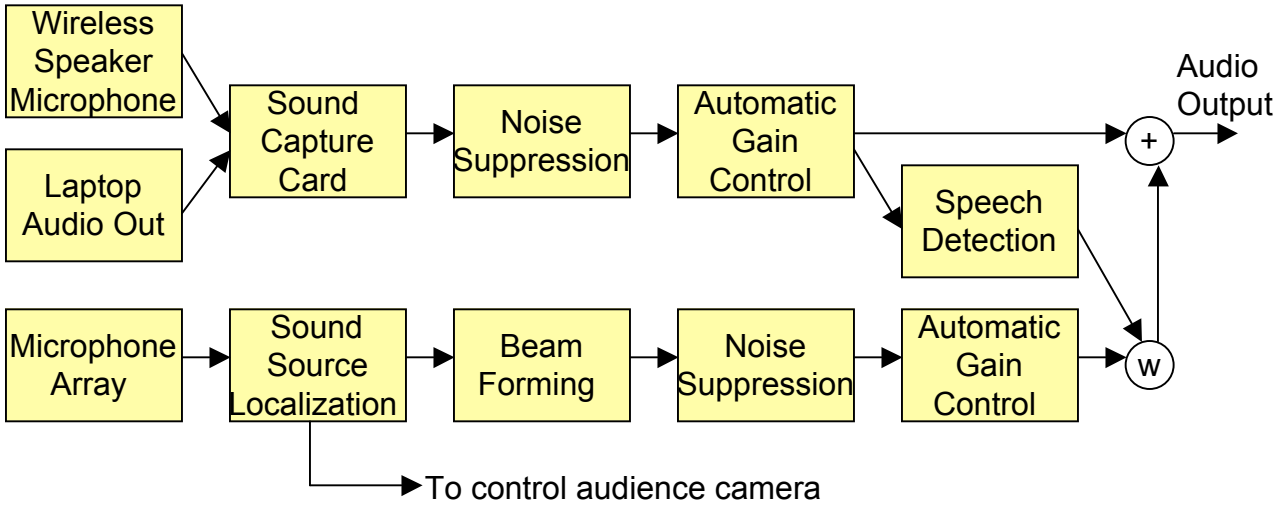
M – number of microphones
Xi(f) – spectrum of *i*-th channel
W(f,i) – weight coefficients matrix
Y(f) – output signal

(Tashev05)



Audio – Digital Mixing

- Two unsynchronized audio input
  - Wireless microphone
  - Microphone array
- Problem of regular mixing: echoes
- Our solution

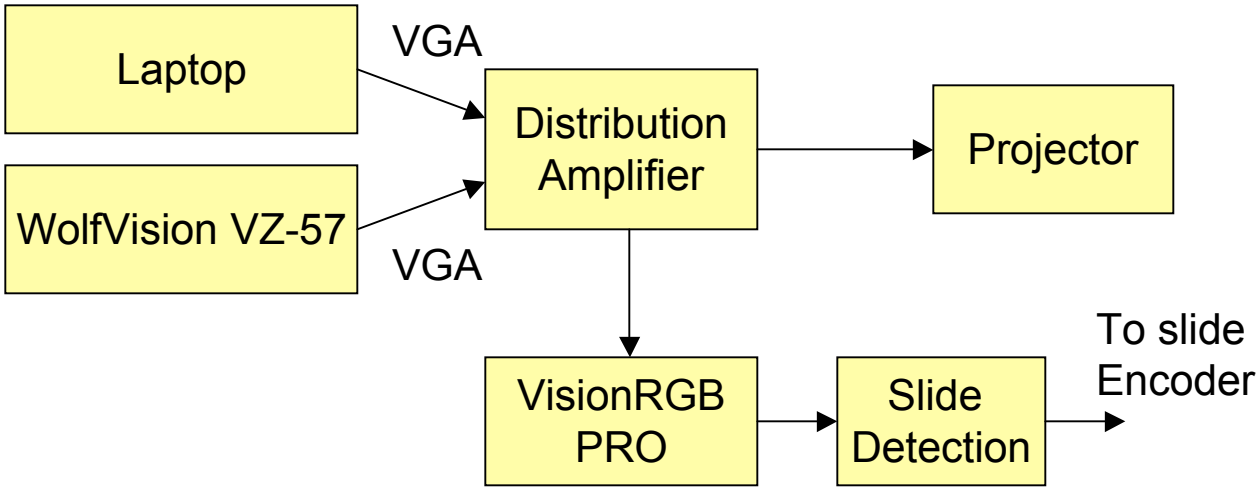


What's New in iCam2

- Audio
  - High quality audience voice capturing with 8-element microphone array
    - Sound source localization
    - Beam forming
    - Intelligent digital audio mixing
- **Visual Aids**
  - **Automated slide change detection**
- Video
  - Portable design with network cameras
  - Hybrid speaker tracking
  - Scripting Language for virtual director rules
  - Direct Show implementation also enhances portability

Slides Capturing

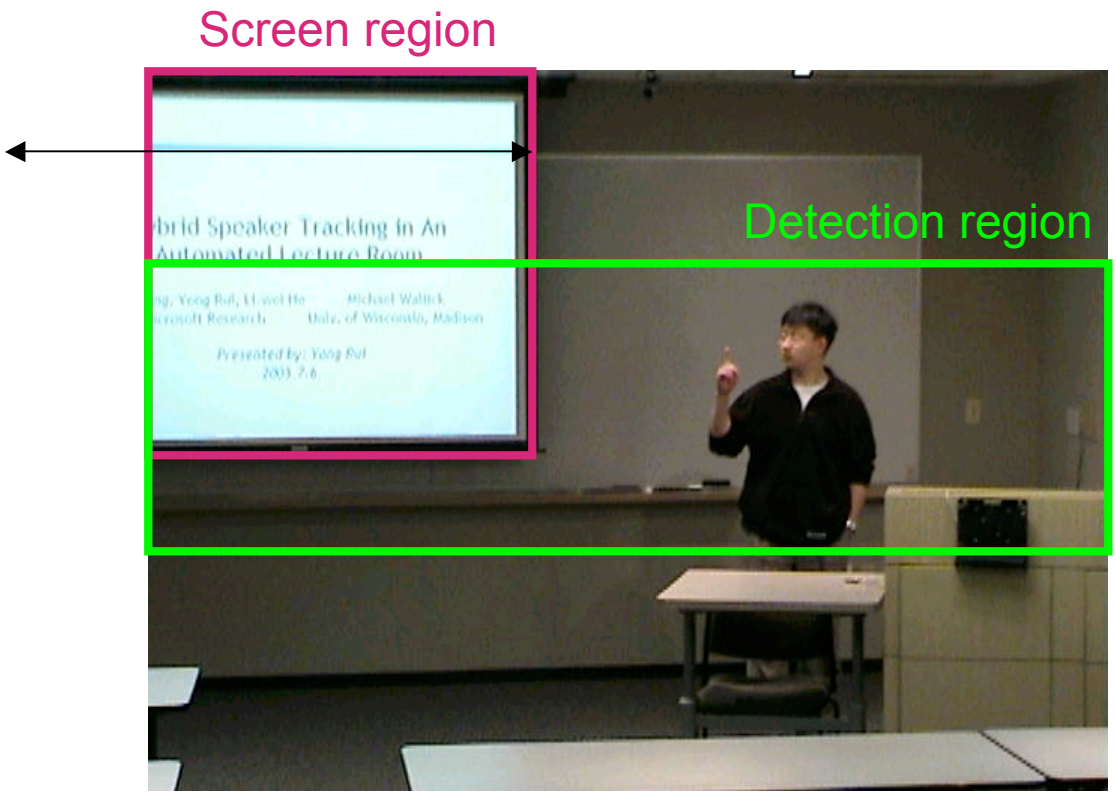
- No PowerPoint file before or after the talk
- We use a video capture card



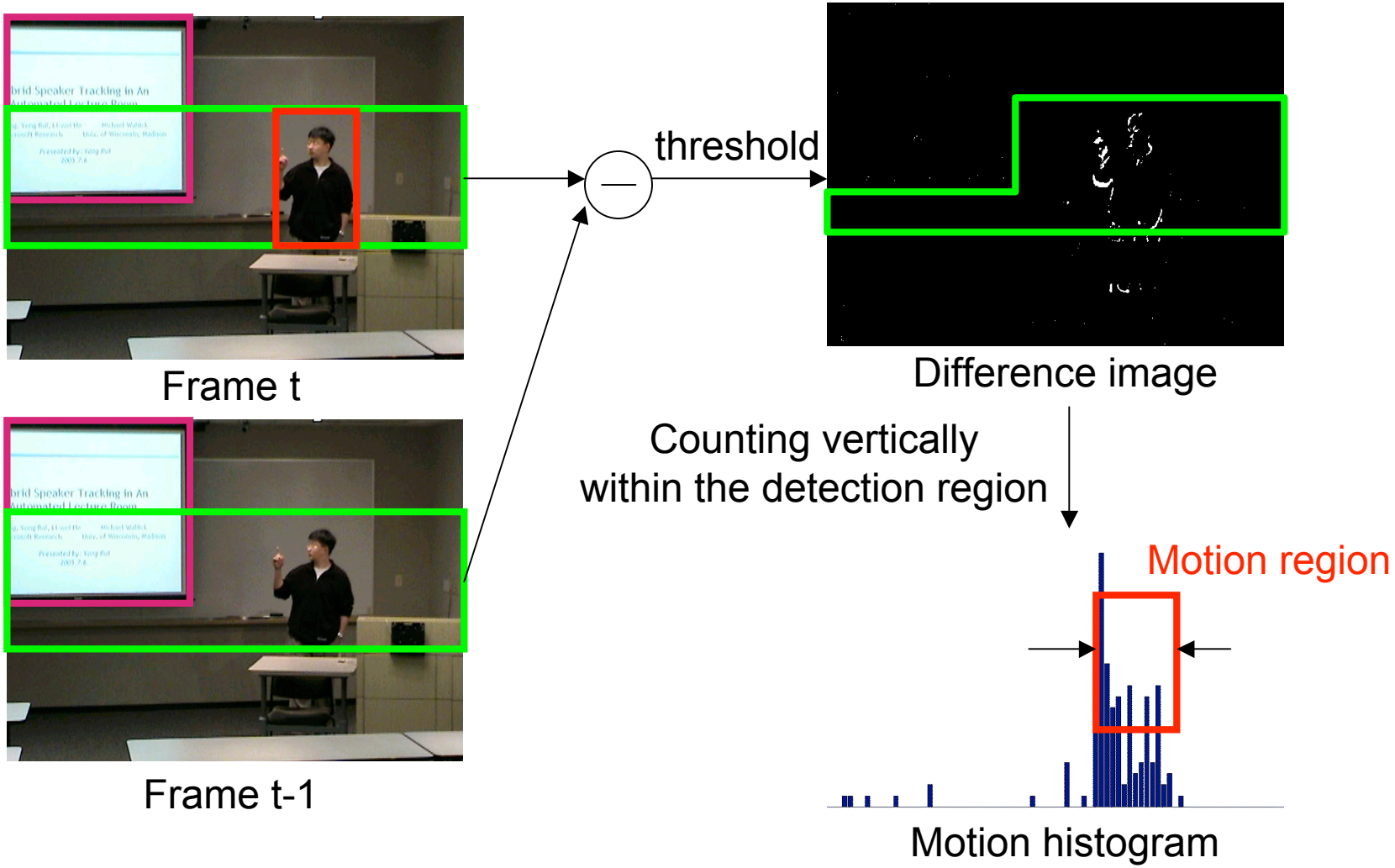
What's New in iCam2

- Audio
  - High quality audience voice capturing with 8-element microphone array
    - Sound source localization
    - Beam forming
    - Intelligent digital audio mixing
- Visual Aids
  - Automated slide change detection
- **Video**
  - **Portable design with network cameras**
  - **Hybrid speaker tracking**
  - **Scripting Language for virtual director rules**
  - **Direct Show implementation also enhances portability**

Video – Speaker Detection

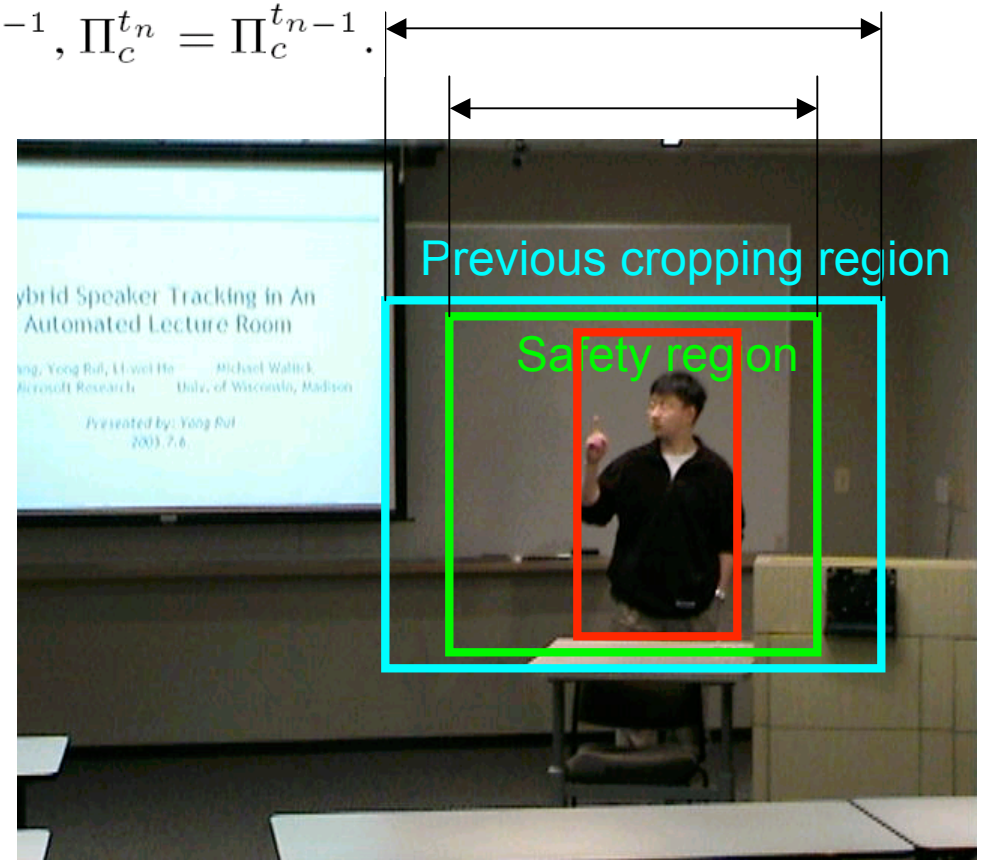


Video – Speaker Detection



Video – Digital Tracking

Rule 1 If $\Pi_m^{t_n} \subset \Pi_s^{t_{n-1}}$, $\Pi_c^{t_n} = \Pi_c^{t_{n-1}}$.

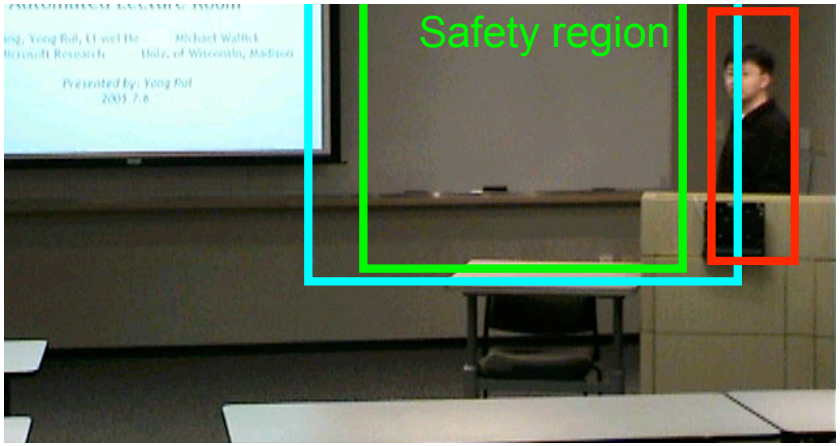


Rule 1: If motion region is within safety region, keep the previous cropping region.

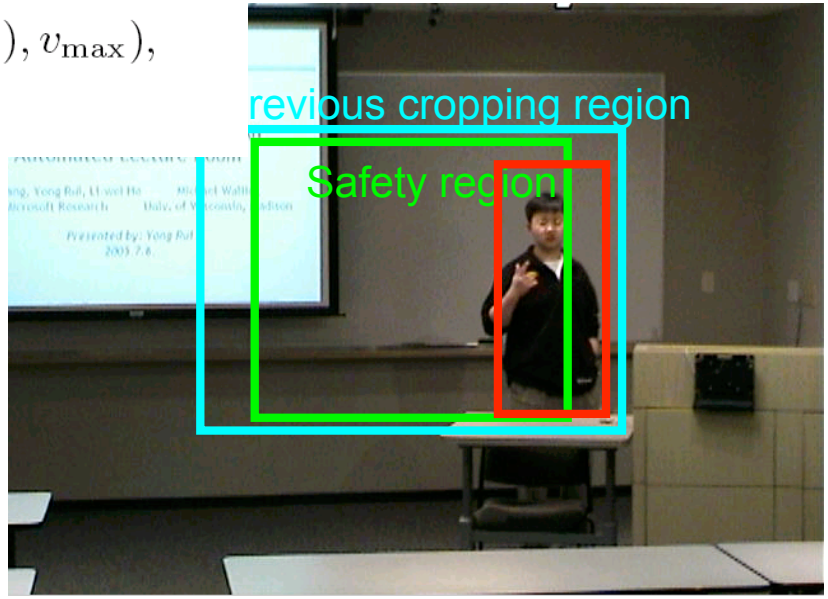
Video – Digital Tracking

Rule 2 If $\Pi_m^{t_n} \cap \Pi_s^{t_{n-1}} = \emptyset$ or for a period greater than T_0 we have $\Pi_m^{t_n} \cap \Pi_s^{t_{n-1}} \neq \Pi_m^{t_n}$, digital panning is performed which follows a CALS model:

$$
\begin{aligned}
s^{t_n} &= \text{sign}(d^{t_n}), \\
v^{t_n} &= \min(v^{t_{n-1}} + \alpha s^{t_n}(t_n - t_{n-1}), v_{\max}), \\
\Pi_c^{t_n} &= \mathcal{S}(\Pi_c^{t_{n-1}}, v^{t_n}(t_n - t_{n-1})).
\end{aligned}
$$



Speaker completely outside the safety region

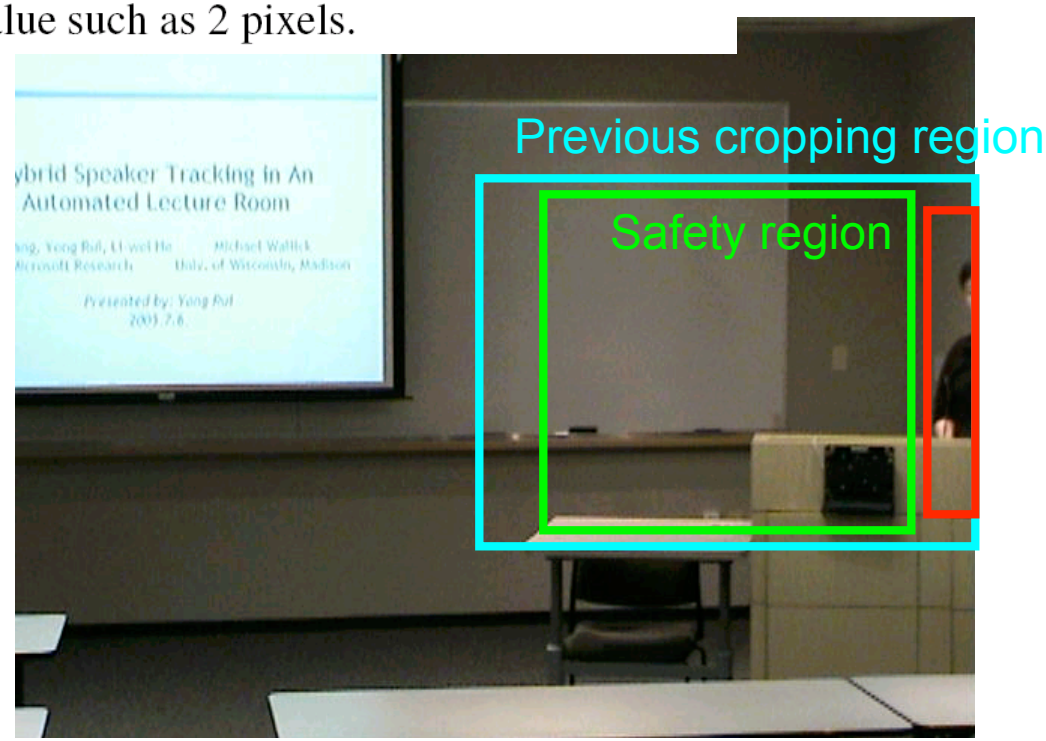


Part of the speaker outside safety region for a certain amount of time

Rule 2: If the above two cases happen, digitally pan the cropping region according to the constant acceleration, limited speed (CALS) model.

Video – Mechanical Tracking

Rule 3 Mechanical panning of the camera should be issued if $\Pi_m^{t_n} \cap \Big[(0, \epsilon) \cup (640 - \epsilon, 640)\Big] \neq \emptyset$.
Here ϵ is a small value such as 2 pixels.

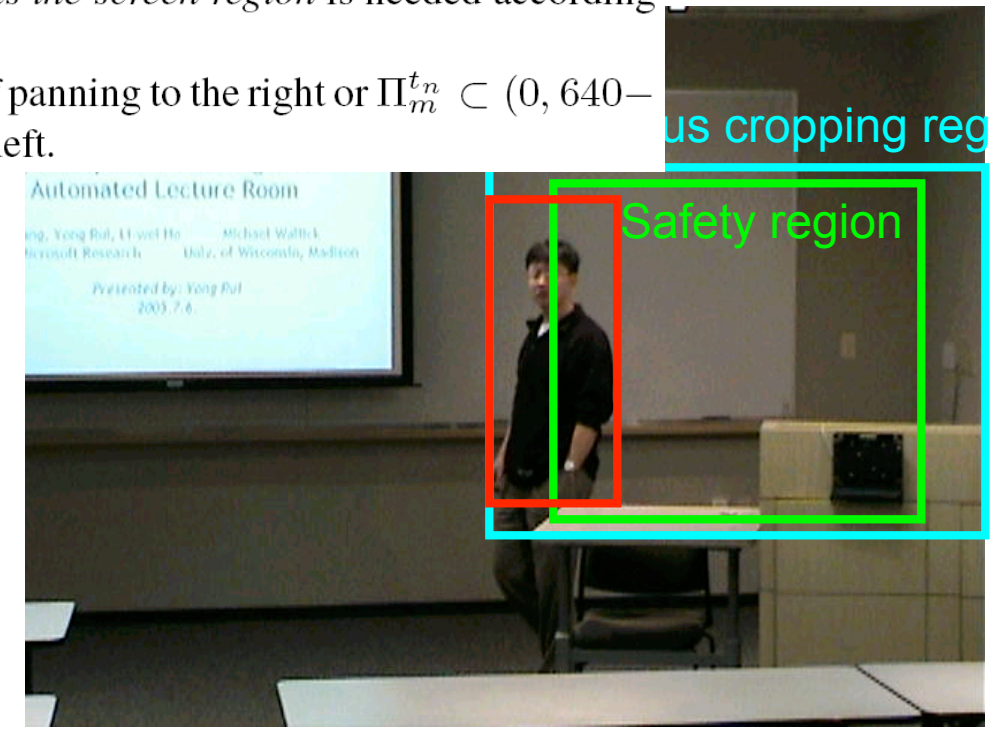


Rule 3: If motion region is near the view boundary, mechanically pan the camera, while the cropping region stays at its relative position.

Video – Panning for Mimicking a Screen Camera

Rule 4 A mechanical panning of the camera is issued if:

1. $\Pi_{scr}^{t_n} \cap (0, 640) \neq \Pi_{scr}^{t_n}$;
2. A panning *towards the screen region* is needed according to Rule 2 or 3;
3. $\Pi_m^{t_n} \subset (\eta, 640)$ if panning to the right or $\Pi_m^{t_n} \subset (0, 640 - \eta)$ if panning to the left.



Rule 4: In order to bring the screen inside the FOV, favor a mechanical panning if panning is needed.

Video – Automatic Zoom Level Control

Rule 5 The zoom level of the camera is changed at the end of each time period T_1, following:

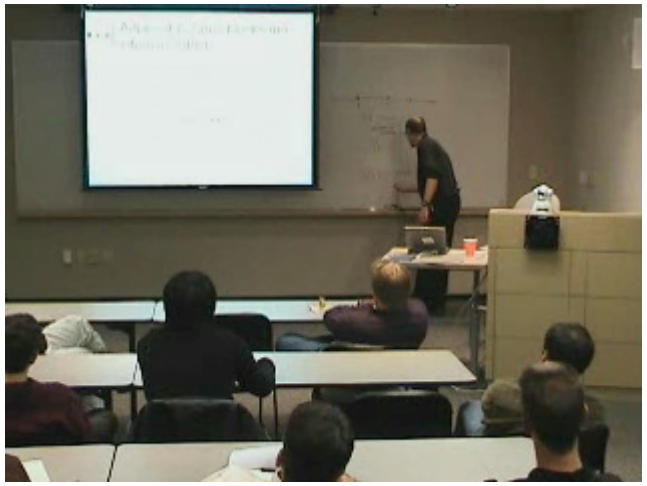
$$z_{\text{new}} = \begin{cases} \max(z_{\text{old}} - \Delta z, z_{\min}) & \text{if } u > U_1 \\ \min(z_{\text{old}} + \Delta z, z_{\max}) & \text{if } u < U_2 \\ z_{\text{old}} & \text{otherwise} \end{cases}$$

High activity, low zoom level

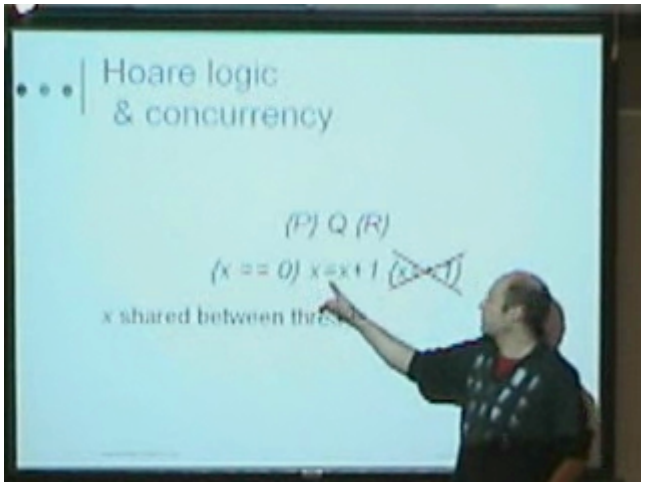
Low activity, high zoom level

Rule 5: Adjust the zoom level of the speaker camera according to the speaker's activity level.

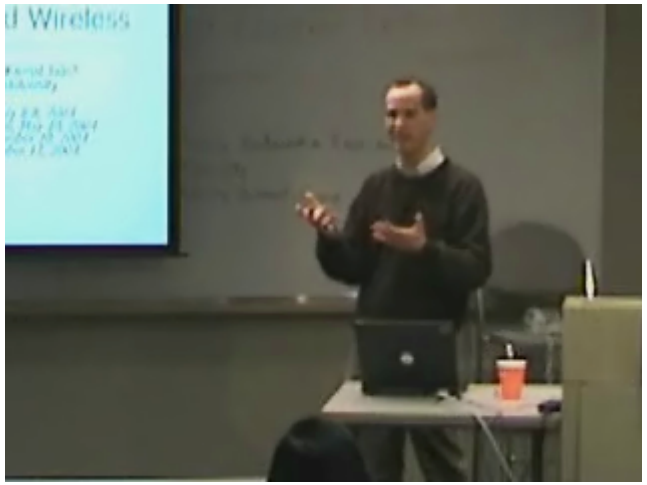
Results



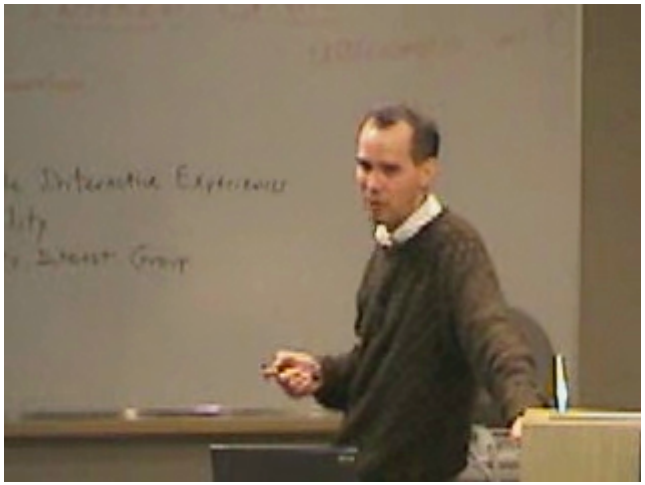
Global view



Screen view



Zoomed out speaker view



Zoomed in speaker view

Results

Scripting for Cinematography Rules

01 CAMERAS
02 CIPCamera SpeakerCam 0
03 CAudience AudienceCam 1
04
05 STATES
06 Audience AudienceCam
07 Speaker SpeakerCam
08
09 TRANSITIONS
10 # if audience camera has low confidence, switch back to speaker camera
11 Audience 3
12 CONFIDENCE SpeakerCam 2 G CONFIDENCE AudienceCam 5 L TIME 5 G
13 Speaker 1.0
14 # if audience camera has been up for more than 12 secs without a high confidence,
15 # forced to switch to the speaker camera (most likely for a global view)
16 Audience 2
17 CONFIDENCE AudienceCam 5 L TIME 12 G
18 Speaker 1.0
19 # if SSL found something, switch to audience camera
20 Speaker 2
21 CONFIDENCE AudienceCam 4 G TIME 3 G
22 Audience 1.0
23 #if speaker camera has a low confidence, switch to the audience camera
24 Speaker 2
25 CONFIDENCE SpeakerCam 2 L TIME 5 G
26 Audience 1.0
27 # if speaker camera has been up for more than 90 secs, randomly choose next view
28 Speaker 1
29 TIME 90 G
30 Speaker 1.0 Audience 2.0
31
32 INITALSTATE Speaker
33 MINSHOT 3
34 MAXSHOT 10000

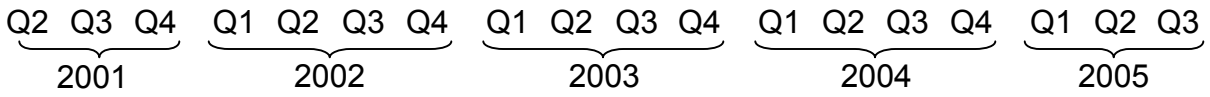
Outline

- Design goals
- System architecture
- Automated lecture capturing – iCam2
- **System usage statistics**
- Conclusions and future work

System Usage

- iCam launched June 2001, iCam2 launched Oct. 2004
- Till mid-Aug. 2005
  - 522 lectures recorded
  - 20,383 live/on-demand sessions
    - 11,115 live sessions (54.4%)
    - 9,268 on-demand sessions

Lecture and Viewing Statistics



Distribution of Lectures

When Do On-Demand Sessions Happen after the Lecture

When Do People Often Watch Lectures On-Demand

Outline

- Design goals
- System architecture
- Automated lecture capturing – iCam2
- System usage statistics
- **Conclusions and future work**

Conclusions

- iCam/iCam2 is the only existing system that automates both lecture capturing and broadcasting
- A well-designed system architecture that enables audio/video/slide synchronization, and minimum pre/post production
- iCam2 has many enhancements over iCam in AV processing and implementation

Future Work

- Distributed iCam for distributed classrooms (integration with CXP)
- 2D SSL for better audience speaker localization
- Better indexing using audio features

Thanks!