Wavelet Based Event Detection in Video
-*Falling Person Detection*-

Behçet Uğur Töreyin
Yiğithan Dedeoğlu
Supervisor: A. Enis Çetin

Funded by EU FP-6 MSUCLE

Bilkent University

Address http://edition.cnn.com/2005/WORLD/europe/11/18/internet.rescue.ap/index.html

Google Search PageRank 1360 blocke

CNN INTERNATIONAL .com

Member Center: Sign In | Register

Essential Global Intelligence. Business International, weekday mornings
Click here cnn.com/egi

SEARCH THE WEB CNN.com SEARCH

Home Page
Asia
Europe
U.S.
World
World Business
Technology

WORLD

Sons save mom overseas with webcam

Friday, November 18, 2005 Posted: 1131 GMT (1931 HKT)

Friday, November 18, 2005

OSLO, Norway (AP) -- A Web camera in a Norwegian artist's living room in California allowed her sons in Norway and the Philippines to see that she had collapsed and call for help, one of the sons said Friday.

collapsed

"But now I see the Internet as a way to save lives. It's also a wonderful tool for people who live alone in some remote area, and might need help," he said.

The mother had been unconscious for about two hours before her sons checked in, so there was also an element of luck, since they only use the camera a few times a week.

What's an "Event" in Video?

- **e·vent** *n.*

  a.A ***significant*** ***occurrence or happening.***

-www.dictionary.com-

What's an "Event" in Video?

- Event in video which is to be detected can be any pre-defined happening like a goal in a football match, or a car passing by the traffic lights that are still RED !

www.blueeyevideo.com

What's common in

1. They have significant shot boundaries (transitions) that discriminates them from *usual* actions – *non-stationarity*
2. An occurance of an event, *in a way*, resembles other occurances of the same event – *self similarity*

These properties may be apparent in both spatial and temporal domains as well as in both video and audio (maybe in other modalities, too).

Our Strategy for

1. We use wavelet domain signals to reveal the *non-stationarity* in falling down event

2. We make use of hidden Markov models for detection of the *self similarities* between the occurances of falling down event

Moving Object Detection

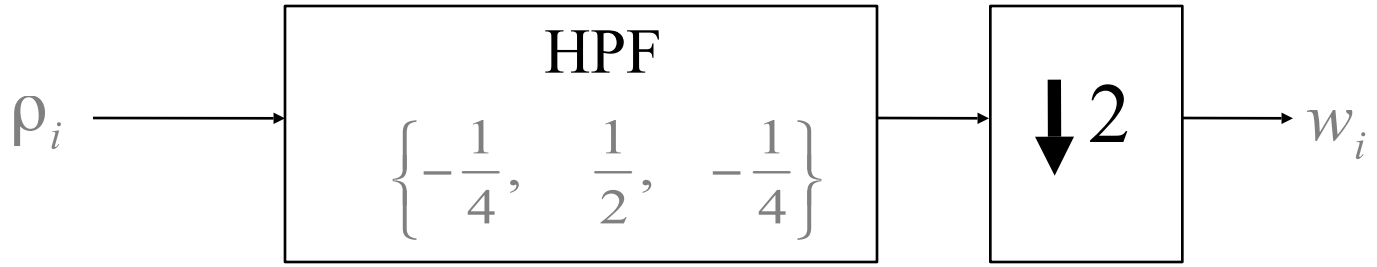
- Moving pixels are determined by adaptive background subtraction – need not be very accurate
- After a post-processing stage comprising of connecting the pixels and labeling, moving regions are encapsulated with their minimum bounding rectangles

Wavelet based

- Calculate aspect ratios as features for the i^{th} moving object at frames *n, n-1* and *n-2*:

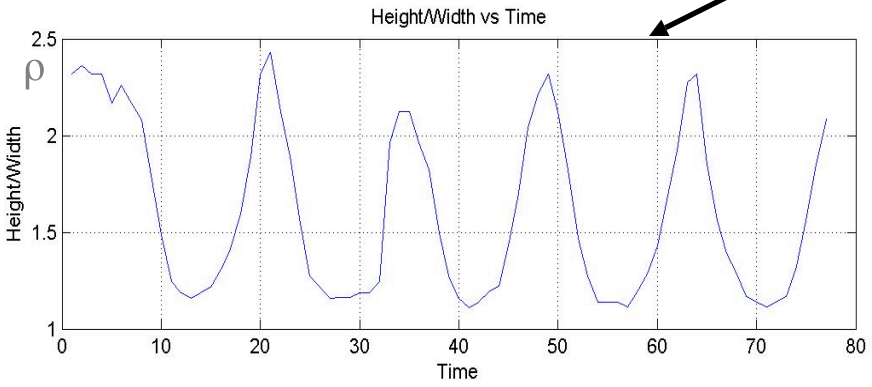
$$\rho_i(n) = \frac{h_i(n)}{w_i(n)}, \quad \rho_i(n-1) \quad and \quad \rho_i(n-2)$$

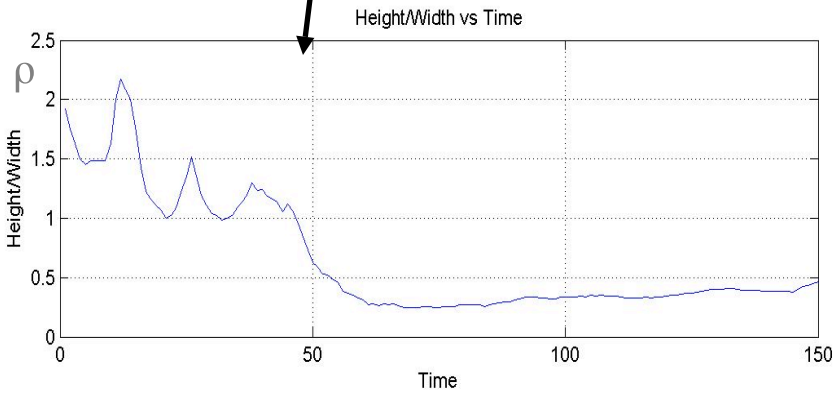
- Calculate wavelet coefficients for these aspect ratios:

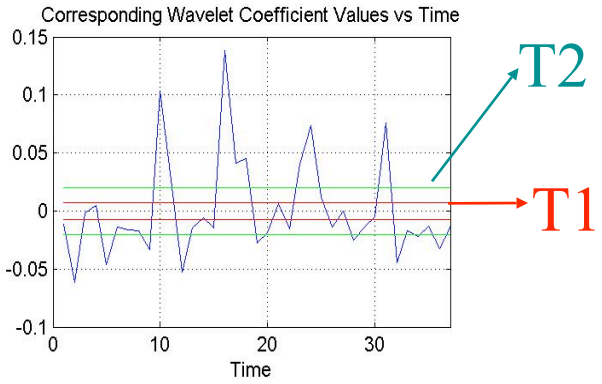


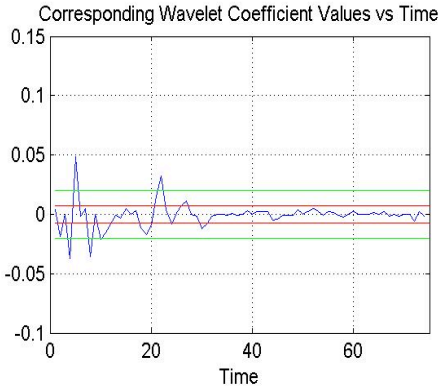
Wavelet Based Feature Extraction for Video (cont'd.)

- *Observation:* Different subband characteristics for walking and falling



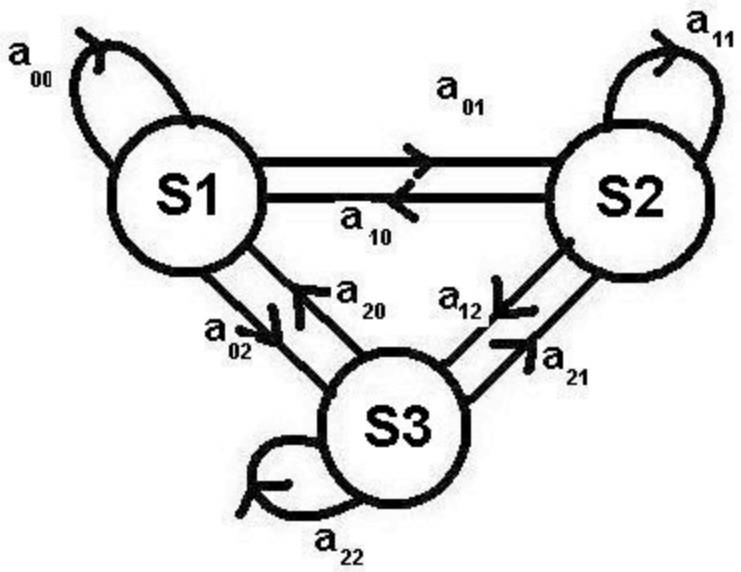




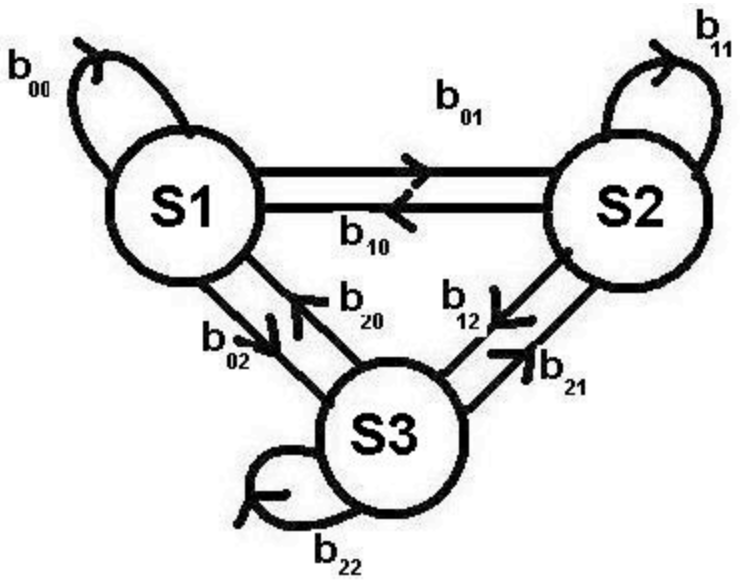


Walking & Falling Person Models

- Three state Markov models are used for walking and falling objects



Walking Model



Falling Model

- **S1:** $|w_i| < T1$
- **S2:** $T2 \le |w_i| < T1$
- **S3:** $|w_i| \ge T2$

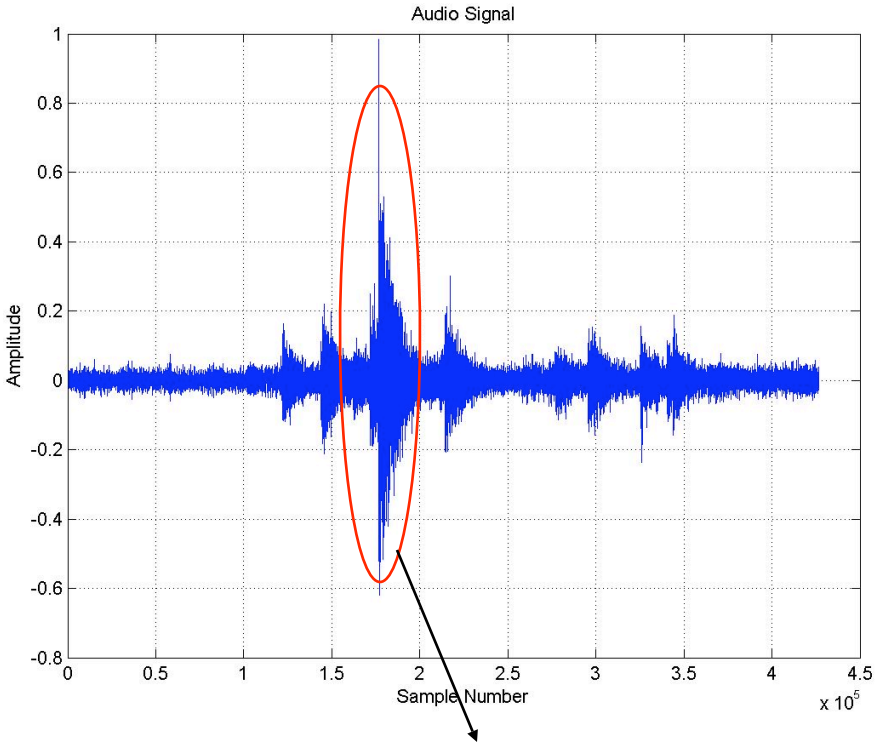
Walking & Falling Person Models (cont'd.)

- State transition probabilities are estimated off-line for walking and falling models
- Each objects' 20-frame-state history is fed to the models
- The model yielding higher probability is determined

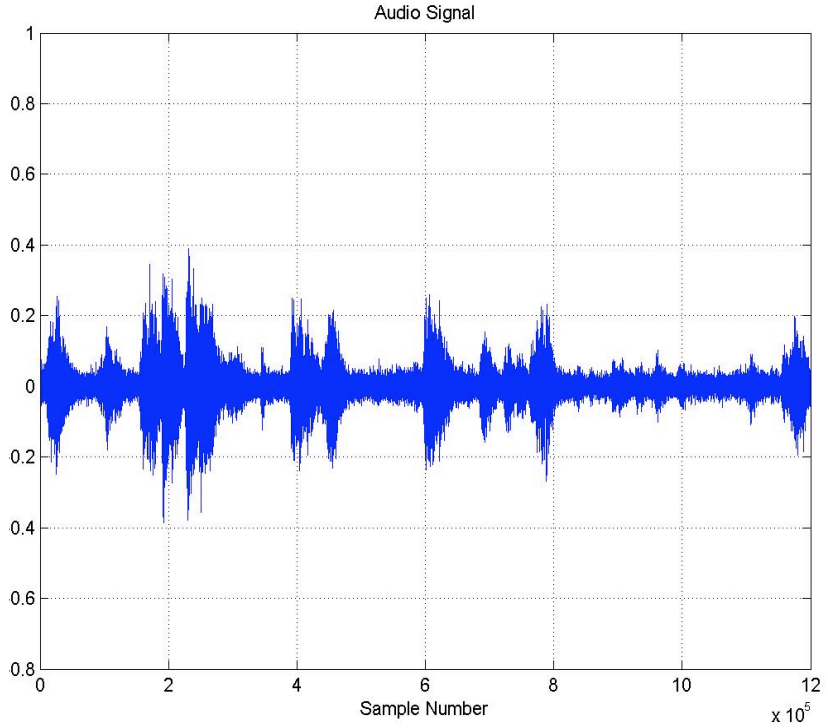
- However, video data is not enough for distinguishing between falling and bending or sitting down
- Audio should be incorporated!

Wavelet Based Modelling of Audio

- Falling down has a non-stationary characteristic sound, whereas bending or sitting down have no distinguishable sound from the background



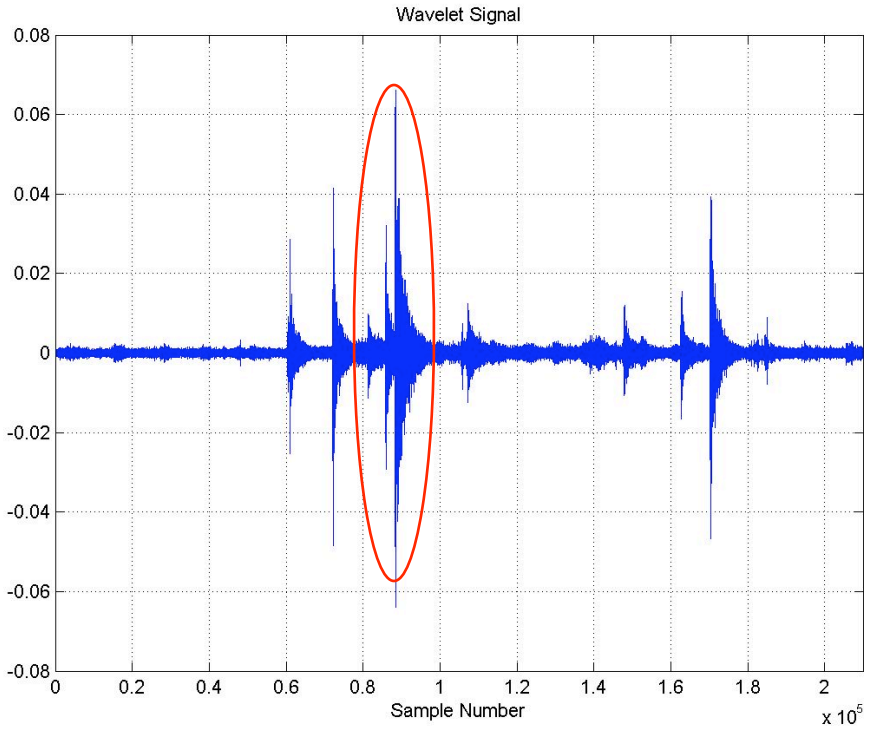
Falling
Audio Signal



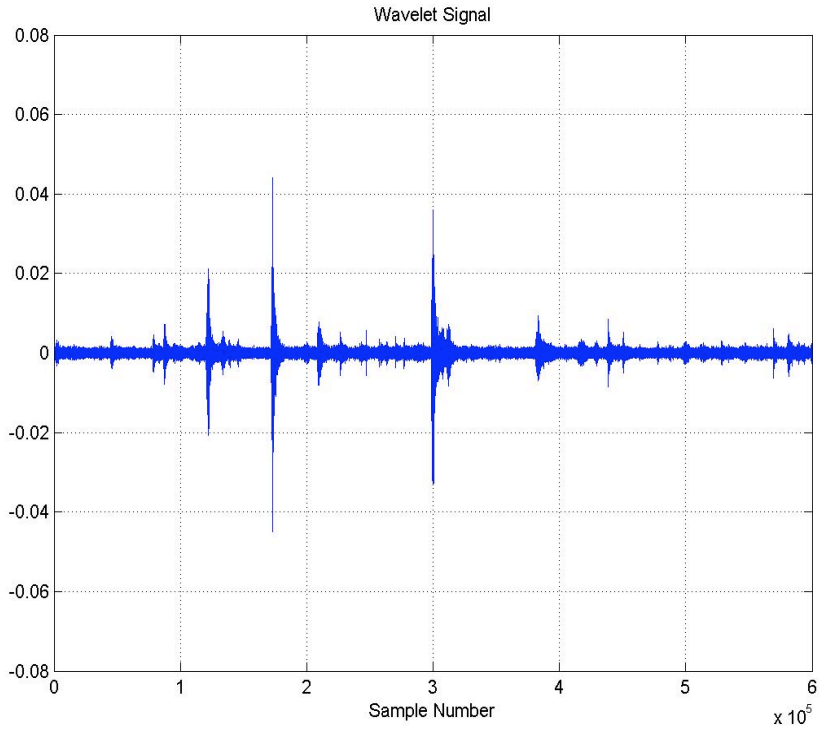
Walking + Talking +Bending
Audio Signal

Wavelet Based Modelling of Audio (cont.'d)

- Non-stationary characteristic is apparent also from the wavelet coefficients



Falling
Wavelet Signal

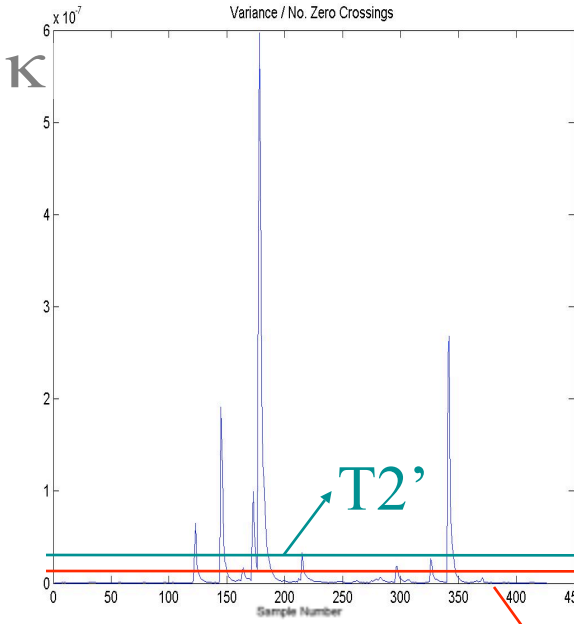


Walking + Talking +Bending
Wavelet Signal

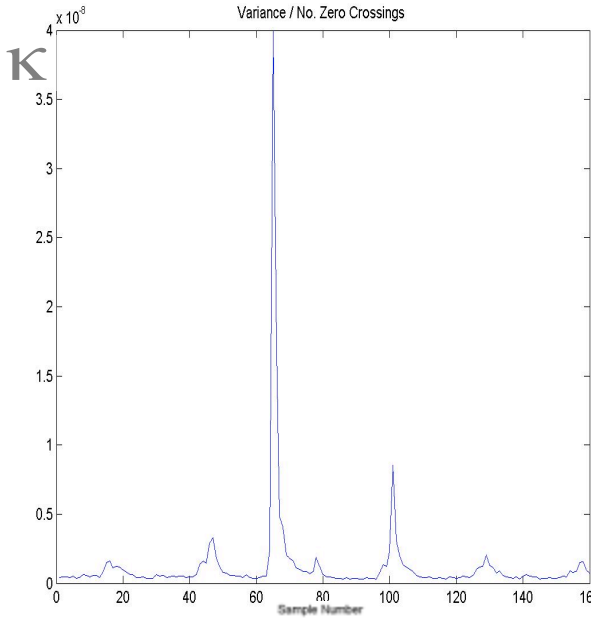
HMM Modelling of Audio

- Analyze the wavelet coefficients of the audio signal in fixed length (500 sample) windows
- Determine σ^2_i and number of zero crossings (Z_i) in i^{th} window
- *Observations:*
  1. Walking has a quasi-periodic sound characteristic in terms of σ^2 and number of zero crossings
  2. When falling down:
  - Number of zero crossings *decreases*
  - Variance *increases*
- Model with a similar HMM
- Use $\kappa = \frac{\sigma^2_i}{Z_i}$ as the feature for the i^{th} window

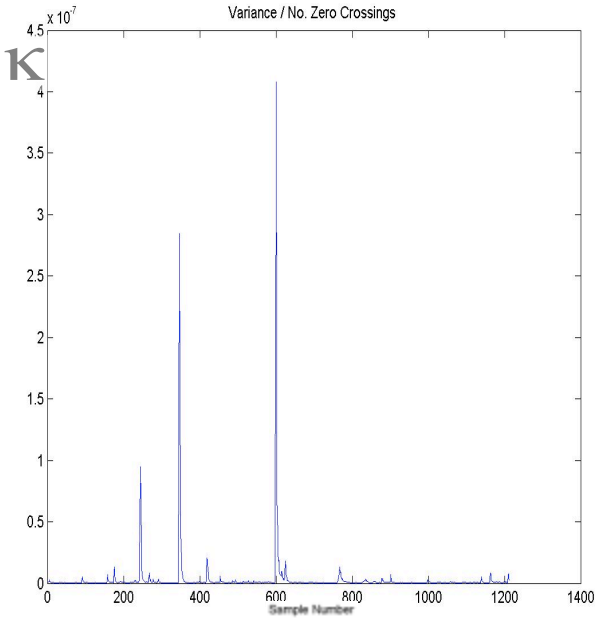
HMM Modelling of Audio (cont'd.)



Falling



Walking

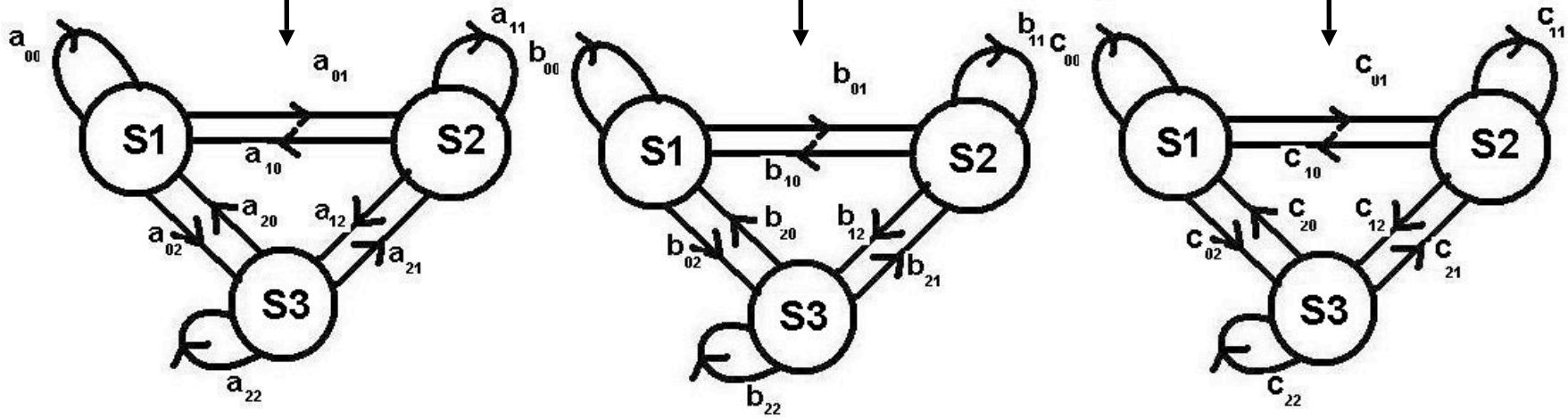


Walking + Talking + Bending down

- Thresholds (states) are defined in this 'κ' domain

HMM Modelling of Audio (cont'd.)

- Three state Markov models are trained with walking, talking and falling persons



- **S1:** $\kappa < T1'$
- **S2:** $T2' \leq \kappa < T1'$
- **S3:** $\kappa \geq T2'$

Experimental Results

- Detection results for the test set
  - 64 video clips with 15,823 frames

Video Content	Include Audio	No. of Clips	No. of Clips in which Falling is Detected	
			Video	Audio+Video
Walking + Talking	Yes	16	0	0
Sitting down + Talking	Yes	5	5	0
Sitting down	Yes	4	4	0
Walking + Falling	Yes	25	25	25
Walking + Falling	No	14	14	14

Summary

- Audio-visual falling person detection, which is an “important event in video” using ***wavelet based hidden Markov modelling.***
- Trained with ***wavelet data***
- We combine ***both audio and video***