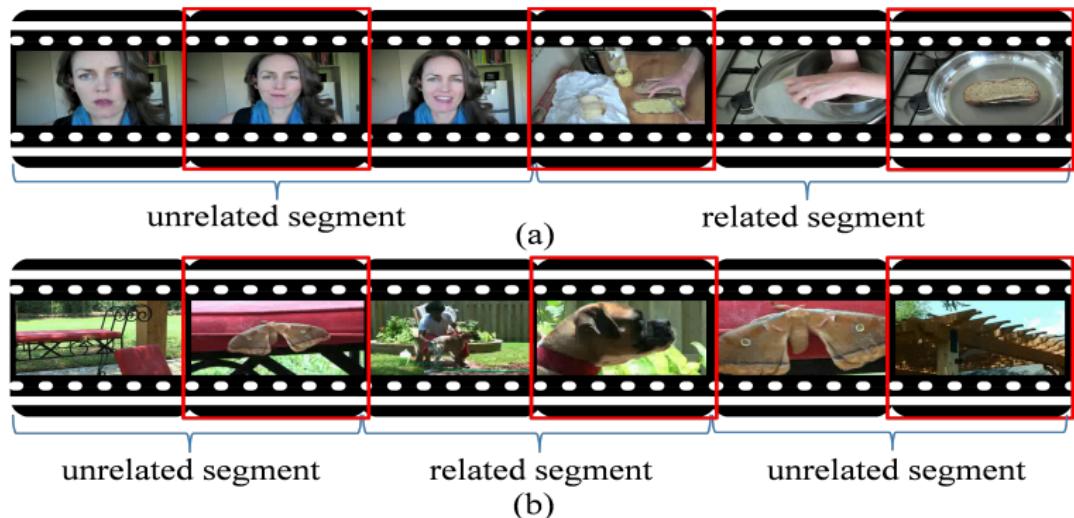


Approaches for MED

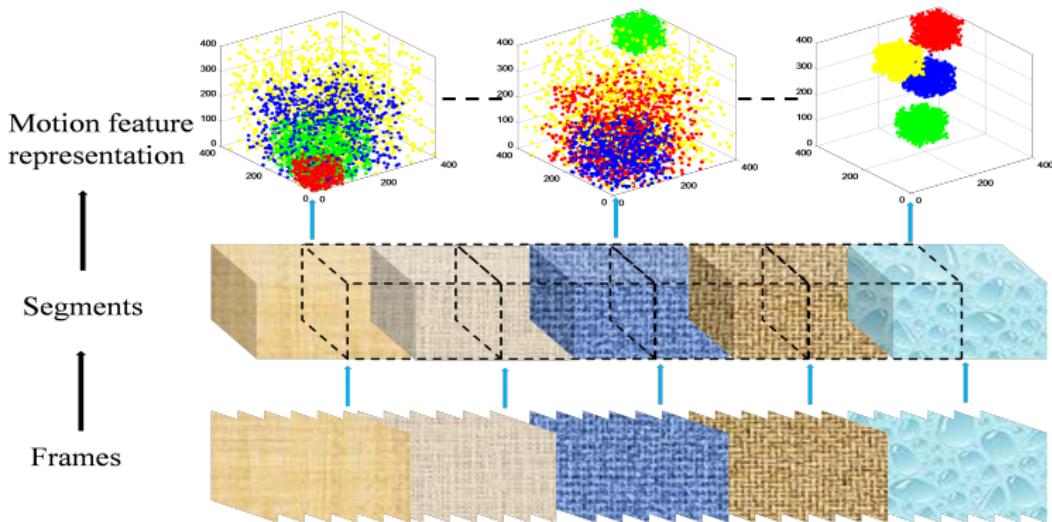
- Best MED'10 system: Columbia University, USA
 - ▶ Image features: SIFT [Lowe 2004]
 - ▶ Motion features: STIP [Laptev 2004]
 - ▶ Audio features: MFCC
- Best MED'11 system: BBN VISER, USA
 - ▶ Image features: SIFT, SURF, D-SIFT, CHOG, RGB-SIFT
 - ▶ Motion features: STIP, D-STIP
 - ▶ Audio features: MFCC, FDLP
- Common Approach: Combining multiple modalities (image, video, audio, etc.)
 - ▶ For image features: keyframe-based → image classification problems
 - ▶ For motion features: video-based approach

Problem with The Video-based Approach



- MED data is noisy → the clues to determine an event may appear within a small segment of the entire video.

Our Segment-based Approach



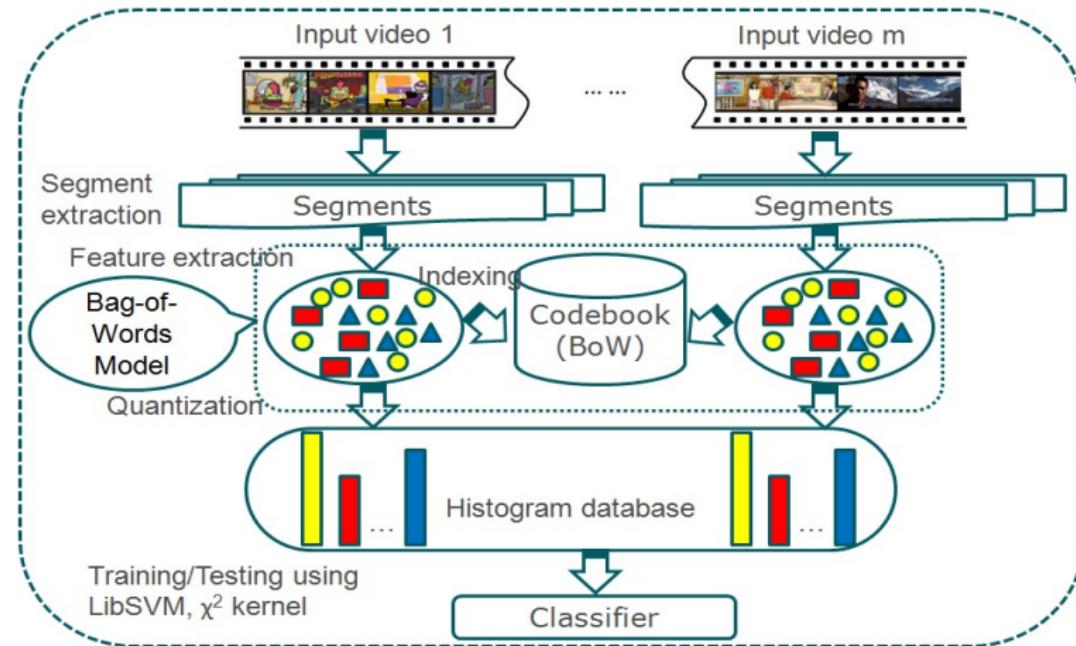
- The basic idea is to examine shorter segments instead of using the entire video.

Our Segment-based Approach

- Non-overlapping
 - ▶ Uniform sampling
 - ▶ Segment length: 30, 60, 90, 120, 200, 400 seconds
 - ▶ Compare with the video-based approach (using the whole video)
- Overlapping sampling
 - ▶ Uniform sampling, 50% overlapping
 - ▶ Segment length: 30, 60, 90, 120, 200, 400 seconds
 - ▶ Compare with the video-based approach (using the whole video)
- Segment sampling based on shot boundary detection
 - ▶ Take into account the boundary information of each segment
 - ▶ Employ the technique proposed by Guimaraes et al. (2003)

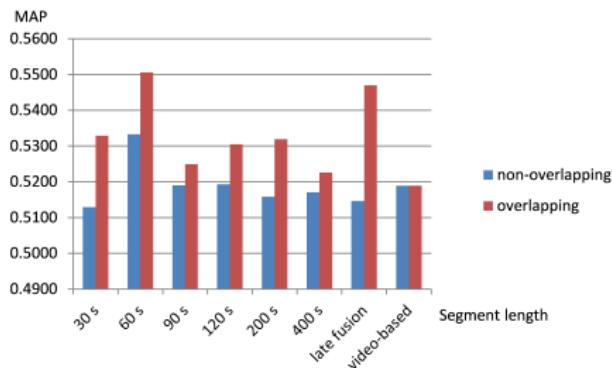
Guimaraes, S.J.F., Couprie, M., Araujo, A.d.A., Leite, N.J: Video segmentation based on 2d image analysis. Pattern Recognition Letters, 2003, 24(7), 947-957.

Evaluation Framework

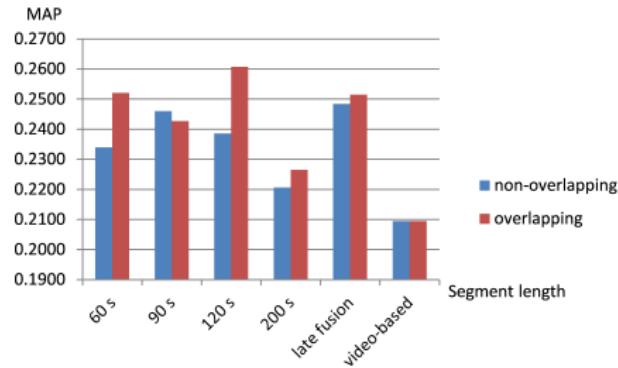


- Evaluation framework for our baseline MED system

Non-Overlapping vs. Overlapping Sampling



(b) On the MED 2010 dataset



(b) On the MED 2011 dataset

In most cases, the overlapping sampling performs the best.

Comparison

Table: Comparison of different segment-based approaches with the video-based approach on the MED 2010 dataset.

Event/MAP	Best non-overlapping	Best overlapping	SBD segments	Video-based
Assembling shelter	0.4511	0.4781	0.4284	0.4911
Batting in a run	0.7852	0.7918	0.7866	0.7902
Making a cake	0.3636	0.3819	0.1918	0.2755
All	0.5333	0.5506	0.4689	0.5189

Comparison

Table: Comparison of different segment-based approaches with the video-based approach on the MED 2011 dataset.

Event/MAP	Non-overlapping sampling			Overlapping sampling			Video-based
	Best (at 90 s)	Late fusion (all lengths)	Late fusion (60, 90, 120 s)	Best (at 120 s)	Late fusion (all lengths)	Late fusion (60, 90, 120 s)	
E006	0.1277	0.1217	0.1244	0.1151	0.1086	0.1083	0.0959
E007	0.1521	0.1419	0.1369	0.1552	0.1610	0.1616	0.1303
E008	0.4923	0.4975	0.4973	0.4969	0.4903	0.4871	0.4766
E009	0.2072	0.2145	0.2064	0.2160	0.1954	0.1958	0.0943
E010	0.0916	0.0771	0.0753	0.1008	0.1108	0.1109	0.1020
E011	0.0698	0.0805	0.0813	0.1591	0.0819	0.0845	0.0609
E012	0.3560	0.3309	0.3277	0.3150	0.3293	0.3341	0.2858
E013	0.6030	0.6033	0.6096	0.6188	0.5872	0.5910	0.5385
E014	0.2008	0.2585	0.2579	0.2744	0.2706	0.2694	0.2138
E015	0.1599	0.1583	0.1622	0.1562	0.1795	0.1795	0.0964
All	0.2460	0.2484	0.2479	0.2607	0.2515	0.2522	0.2095