IBM Research

The IBM Shot Boundary Detection System at TRECVID 2003

Arnon Amir

: NIST TRECVID 2003 Nov. 17th, 2003

Shot Boundary Detection

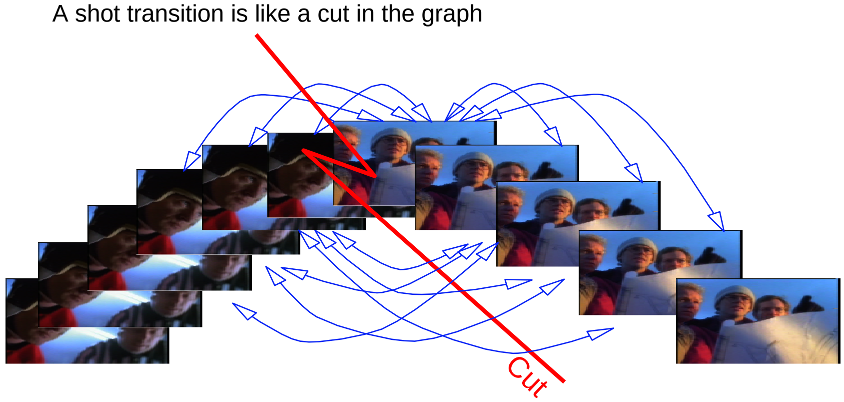
- Temporal segmentation of the video into shots
- Detect location and duration of all shot transitions of all kinds:
  - cuts, dissolves, fades, wipes, checkerboard, etc.
- Desired: accurate detection, frame-accurate location
- Challenges: It would be so much easier if there were no
  - Fast object or camera motion
  - Fast illumination changes: pop concerts, driving, backlight, …
  - Fire, flags in the wind, sea waves, …
  - Specularities, shadows, reflections from glass, water, …
  - Instantaneous illumination changes due to flash photography
  - Very short shots (up to single-frame “shots” in the Search test set)
  - Very long gradual transitions
  - Text overlay, graphics, animation
  - Screen split, video in video
  - Video artifacts: MPEG errors, compression noise, camera noise, …

IBM SBD: System Overview

- Work on fully-decoded RGB video frames
- A single pass over the video
- Graph-based, multiple pair-wise frames comparisons
- Adaptive thresholds to determine significant changes
- Finite state machine processes the video frame by frame
  - Rule-based state to state transitions
- Aids:
  - Temporal rank filtering whenever applicable
  - Detectors for various simple “events” and video artifacts
- Dedicated monitoring and debugging tools

Graph-Based, Multiple Pair-Wise Frames Comparisons

Each frame is a node in a graph
Pairs of frames, up to 13 frames apart, are connected with arcs
A shot transition is like a cut in the graph



Motivation: more pairs - a better decision [Amir & Lindenbaum, 1995]

Frames Representation and Comparison

- 3D RGB Color histogram $H(r, g, b)$
  - 3bits per color channel
  - Total 512 bins
  - Histogram comparison: $L_1(H_i, H_j)$
- 3D Localized Edges direction histogram $H(x, y, |\nabla I|)$
  - Frame is divided to 8x8 blocks (6bits)
  - Gradient magnitude (3bits)
  - Total 512 bins
  - Histogram comparison: $L_1(H_i, H_j)$
- Gray-level Thumbnails comparison
- Average frame luminance
- Black detector, monochrome detector
- Non linear, state-based fusion

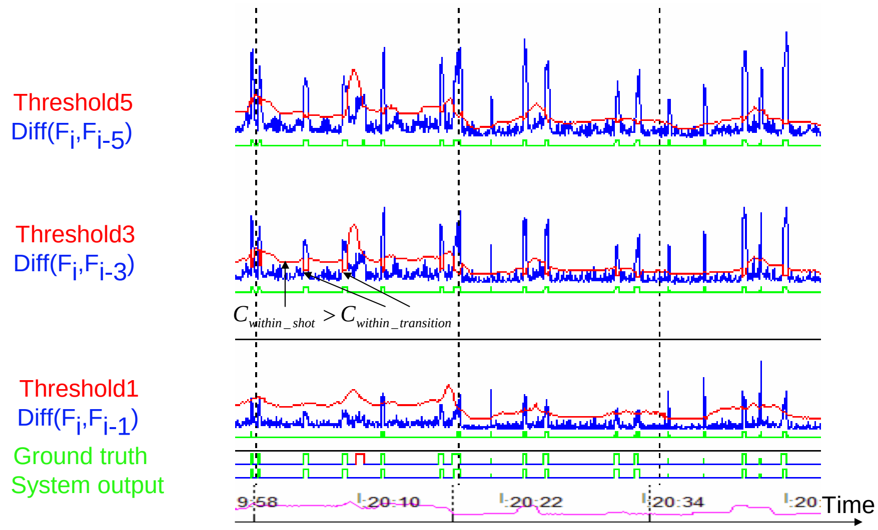
Adaptive Thresholds

- Observe:
  - Frames similarity changes significantly across different shots and videos
  - Frames dissimilarity increases as the frames are wider apart of each other

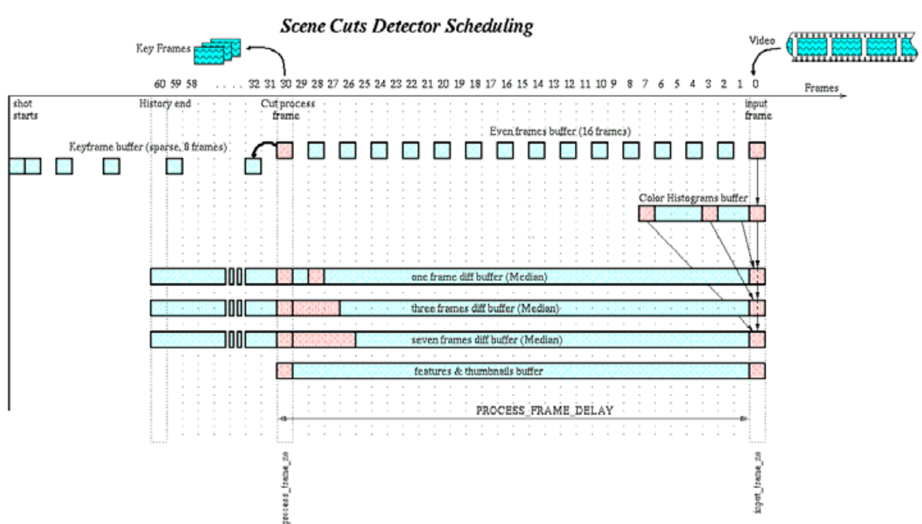
☞ Adaptive thresholds are required

- Collect statistics of frames similarity in a 61-frames buffer around the processed frame
  - Assumption: at least 25% of the buffer content is within-shot
  - Use rank filtering to determine the threshold: $C \cdot (Buff @ 25\%)$
  - Apply a hysteresis when change between shot and transition states. $C_{within\_shot} > C_{within\_transition}$

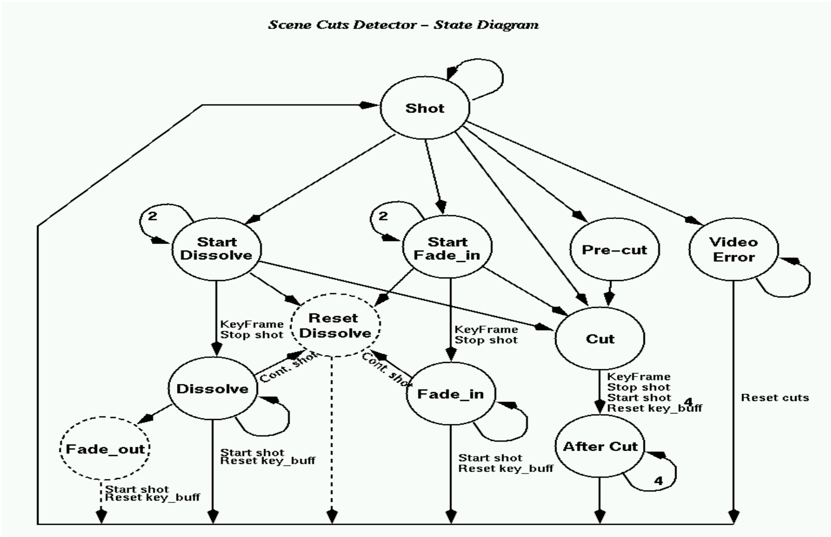
Example: Adaptive Thresholds



SBD - Processing Schedule



SBD – Finite State Machine



Development Methodology

- Apply to development data with ground truth
- Analyze causes for errors, classify them
- Address one class of errors at a time:
  - Use a visual monitoring tool to analyze system behavior at errors
  - Modify the finite state machine - modify/add rules and states
  - Add new detectors
- Test again

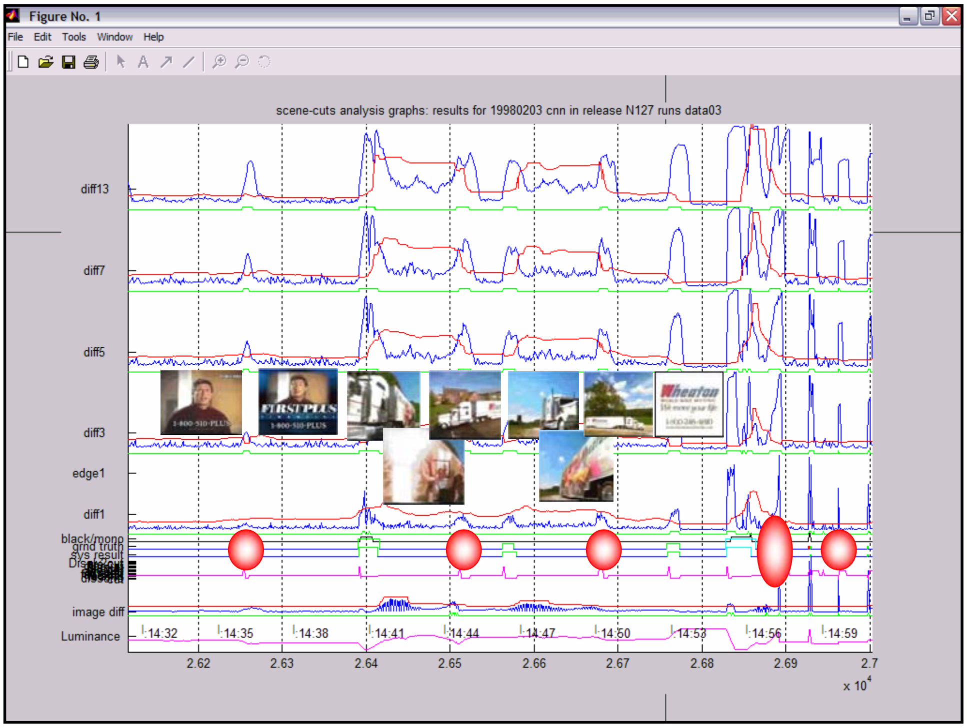
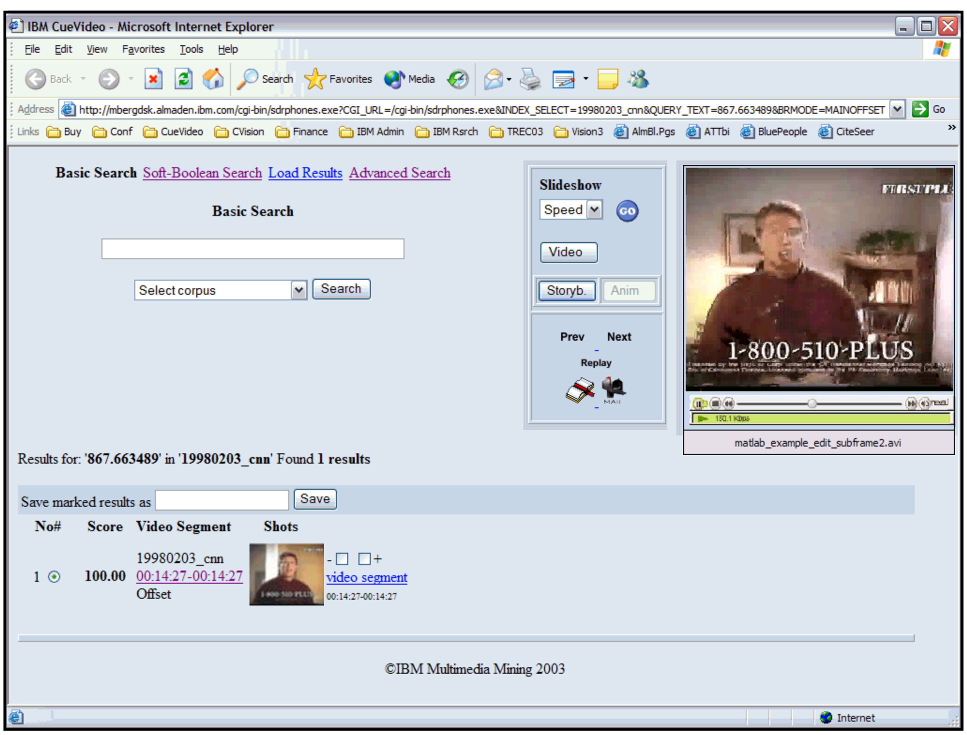
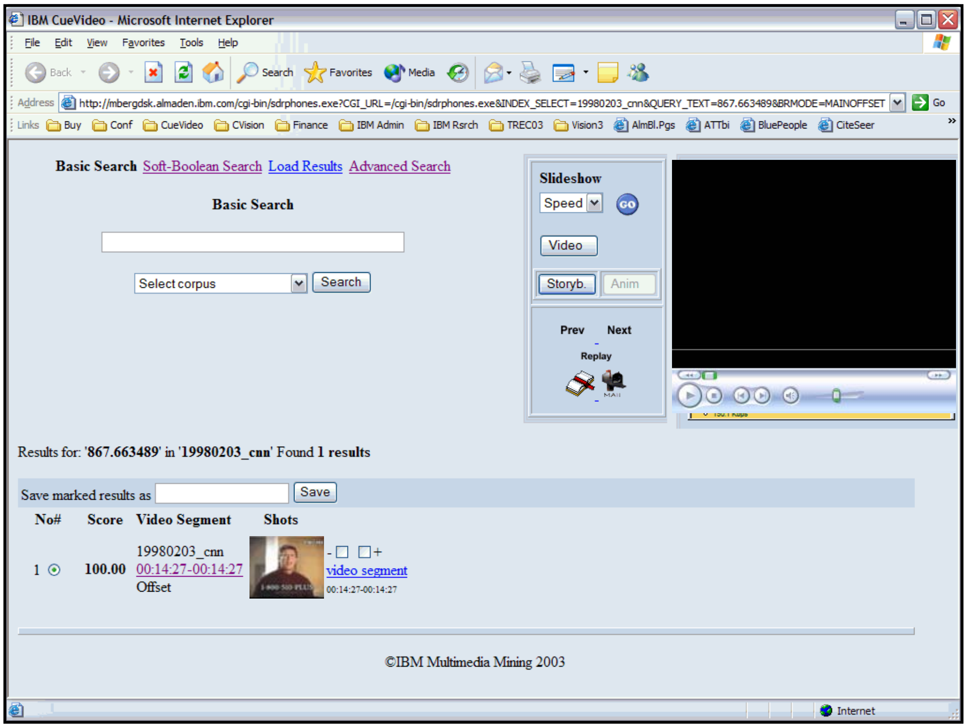


Figure No. 1
File Edit Tools Window Help
scene-cuts analysis graphs: results for 19980203 cnn in release N127 runs data03
diff13
diff7
diff5
diff3
edge1
diff1
black/mono
grnd truth
sys result
image diff
Luminance
14:32
14:35
14:38
14:41
14:44
14:47
14:50
14:53
14:56
14:59
2.62
2.63
2.64
2.65
2.66
2.67
2.68
2.69
2.7
x 10^4


IBM CueVideo - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://mbergdsk.almaden.ibm.com/cgi-bin/sdrphones.exe?CGI_URL=/cgi-bin/sdrphones.exe&INDEX_SELECT=19980203_cnn&QUERY_TEXT=867.663489&BRMODE=MAINOFFSET
Basic Search Soft-Boolean Search Load Results Advanced Search
Basic Search
Select corpus
Search
Slideshow
Speed
GO
Video
Storyb.
Anim
Prev
Next
Replay
1-800-510-PLUS
matlab_example_edit_subframe2.avi
Results for: '867.663489' in '19980203_cnn' Found 1 results
Save marked results as
Save
No# Score Video Segment Shots
1 100.00 19980203_cnn 00:14:27-00:14:27 Offset
video segment
00:14:27-00:14:27
©IBM Multimedia Mining 2003
Internet


IBM CueVideo - Microsoft Internet Explorer
Basic Search Soft-Boolean Search Load Results Advanced Search
Basic Search
Select corpus
Search
Slideshow
Speed
Video
Storyb.
Anim
Prev
Next
Replay
Results for: '867.663489' in '19980203_cnn' Found 1 results
Save marked results as
Save
No# Score Video Segment Shots
1 100.00 19980203_cnn 00:14:27-00:14:27 Offset
video segment
©IBM Multimedia Mining 2003

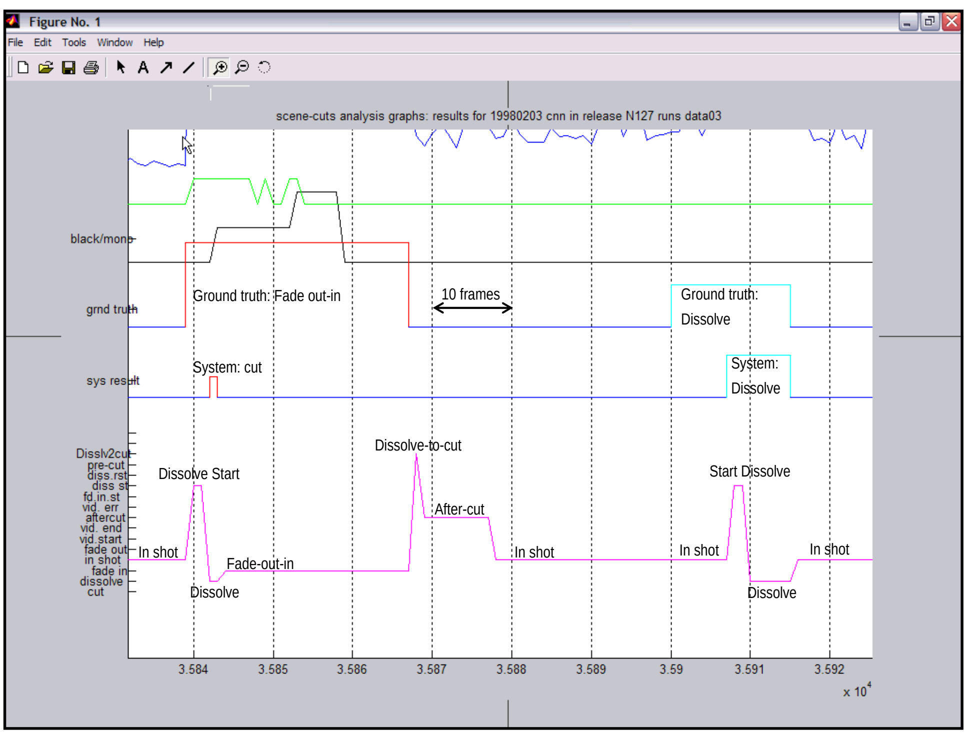


Figure No. 1
scene-cuts analysis graphs: results for 19980203 cnn in release N127 runs data03
black/mono
grnd truth
sys result
Ground truth: Fade out-in
10 frames
Ground truth: Dissolve
System: cut
System: Dissolve
Dissolve Start
Dissolve-to-cut
After-cut
Start Dissolve
In shot
Fade-out-in
Dissolve
In shot
In shot
In shot
Dissolve
3.584
3.585
3.586
3.587
3.588
3.589
3.59
3.591
3.592
x 10^4

Shot Boundary Detection - System Improvements 2003

- Baseline system is the IBM *CueVideo* SBD best run at TREC 2002
- Algorithmic improvements for TREC 2003
  - Flash photography detector. Eliminates false detections on rapid flashlights. 19980426_CNN.mpg , 19980422_CNN.MPG
  - Better handling of fades (better detection, linking fade out-in, abrupt fades)
  - Improved detection of graduals' boundaries (improves gradual accuracy)
  - Detection & handling of partial-frame MPEG errors (reduces insertion errors)

IBM Research | TRECVID03 Team

Results on SBD Development Set

- Changes were tested on a development set with manually labeled frame-accurate ground truth: 10 video segments, 5 minutes each. Total of 363 cuts, 145 graduals.

10 Runs on development set

	All Transitions		Cuts		Gradual transitions		Graduals accuracy	
	Recl	Prec.	Recl	Prec.	Recl	Prec.	Recl	Prec.
alm1	0.951	0.881	0.99	0.94	0.86	0.74	0.65	0.94
N047	0.947	0.887	0.98	0.94	0.87	0.76	0.74	0.92
N110	0.950	0.905	0.98	0.95	0.88	0.8	0.74	0.92
N119	0.941	0.938	0.98	0.98	0.85	0.84	0.81	0.92
N120	0.944	0.938	0.98	0.98	0.86	0.84	0.82	0.92
N122	0.944	0.938	0.98	0.98	0.86	0.84	0.82	0.92
N123	0.950	0.930	0.98	0.98	0.88	0.81	0.86	0.91
N126	0.941	0.940	0.98	0.97	0.85	0.87	0.77	0.93
N127	0.941	0.947	0.98	0.98	0.85	0.87	0.81	0.93
N128	0.944	0.941	0.98	0.98	0.86	0.85	0.85	0.92

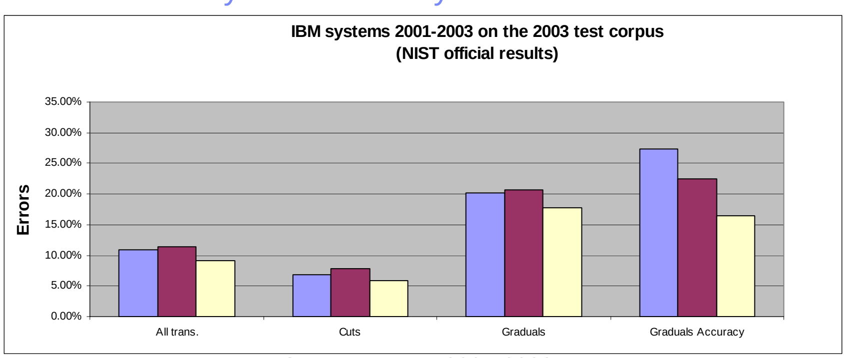
NIST Results on the Test Set

- Trends are similar to the results on the Development set
- Absolute values are typically 0 to 0.06 lower
- Cuts – one frame shift

10 Runs on NIST Test set

	All Transitions		Cuts		Gradual transitions		Graduals accuracy	
	Recl	Prec.	Recl	Prec.	Recl	Prec.	Recl	Prec.
alm1	0.915	0.870	0.947	0.916	0.84	0.761	0.597	0.928
N047	0.895	0.872	0.944	0.9	0.778	0.808	0.681	0.899
N110	0.893	0.879	0.939	0.911	0.784	0.804	0.681	0.905
N119	0.893	0.918	0.936	0.942	0.792	0.865	0.784	0.894
N120	0.897	0.911	0.938	0.942	0.802	0.84	0.793	0.894
N122	0.897	0.911	0.938	0.942	0.802	0.84	0.793	0.894
N123	0.898	0.905	0.936	0.946	0.809	0.809	0.823	0.875
N126	0.886	0.916	0.932	0.943	0.776	0.856	0.723	0.913
N127	0.892	0.922	0.937	0.947	0.784	0.865	0.771	0.913
N128	0.894	0.916	0.935	0.951	0.798	0.833	0.794	0.893

Shot Boundary Detection - System Results 2001-2003



System Errors 2001-2003

System	All Trans.	Cuts	Graduals	Gr. Accuracy
2001 (alm1)	10.86%	6.88%	20.14%	27.34%
2002 (N047)	11.46%	7.85%	20.73%	22.50%
2003 (N127)	9.18%	5.80%	17.75%	16.40%
Rel. Improvement	19.9%	26.1%	14.4%	27.1%

(Error=1-Fnumber)

SBD Results TREC-VID2003 Sorted by F# Per Each Measure (top30 runs)

Run#	All Transitions				Cuts				Graduals				Graduals accuracy			
	Sys	Rcl	Prc	F#	Sys	Rcl	Prc	F#	Sys	Rcl	Prc	F#	Sys	Rcl	Prc	F#
2003	n127	0.892	**0.925**	**0.908**	n128	0.935	0.951	**0.943**	n119	0.792	0.865	**0.827**	n123	0.823	0.875	**0.848**
	n119	0.894	0.921	**0.907**	n127	0.937	0.947	**0.942**	n127	0.784	0.865	**0.823**	n128	0.794	0.893	**0.841**
	n120	0.899	0.913	**0.906**	n123	0.936	0.946	**0.941**	n120	0.802	0.840	**0.821**	n120	0.793	0.894	**0.840**
	n122	0.899	0.913	**0.906**	n120	0.938	0.942	**0.940**	n122	0.802	0.840	**0.821**	n122	0.793	0.894	**0.840**
	n128	0.895	0.917	**0.906**	n122	0.938	0.942	**0.940**	n128	0.798	0.833	**0.815**	n127	0.771	0.913	**0.836**
	n126	0.887	0.919	**0.903**	n119	0.936	0.942	**0.939**	n126	0.776	0.856	**0.814**	n119	0.784	0.894	**0.835**
	n123	0.899	0.906	**0.902**	n126	0.932	0.943	**0.937**	n123	0.809	0.809	**0.809**	n126	0.723	0.913	**0.807**
2001	nalm1	0.916	0.868	**0.891**	z	0.893	**0.976**	**0.933**	nalm1	**0.840**	0.761	**0.799**	a	**0.835**	0.759	**0.795**
baseline	[illegible]0	0.894	0.881	**0.887**	nalm1	0.947	0.916	**0.931**	n110	0.784	0.804	**0.794**	c	0.742	0.844	**0.790**
2002	n047	0.896	0.875	**0.885**	n110	0.939	0.911	**0.925**	n047	0.778	0.808	**0.793**	c	0.748	0.835	**0.789**
	c	0.882	0.881	**0.881**	n047	0.944	0.900	**0.921**	c	0.737	0.849	**0.789**	c	0.761	0.812	**0.786**
	c	0.863	0.898	**0.880**	c	0.942	0.891	**0.916**	c	0.713	0.882	**0.789**	n110	0.681	0.905	**0.777**
	c	0.892	0.858	**0.875**	c	0.960	0.872	**0.914**	c	0.693	0.913	**0.788**	n047	0.681	0.899	**0.775**
	c	0.829	0.921	**0.873**	c	0.905	0.917	**0.911**	c	0.742	0.827	**0.782**	c	0.762	0.788	**0.775**
	c	0.904	0.832	**0.867**	s	0.958	0.868	**0.911**	c	0.672	0.924	**0.778**	a	0.812	0.737	**0.773**
	o	0.845	0.868	**0.856**	r	0.961	0.855	**0.905**	c	0.755	0.784	**0.769**	a	0.789	0.754	**0.771**
	c	0.909	0.809	**0.856**	c	0.966	0.851	**0.905**	c	0.645	0.934	**0.763**	a	0.784	0.750	**0.767**
	o	0.839	0.871	**0.855**	o	0.910	0.892	**0.901**	c	0.762	0.745	**0.753**	c	0.755	0.770	**0.762**
	o	0.855	0.854	**0.854**	o	0.911	0.889	**0.900**	o	0.698	0.805	**0.748**	a	0.771	0.735	**0.753**
	o	0.855	0.839	**0.847**	o	0.905	0.890	**0.897**	o	0.717	0.762	**0.739**	c	0.756	0.749	**0.752**
	o	0.863	0.826	**0.844**	v	0.940	0.855	**0.895**	c	0.783	0.697	**0.738**	a	0.798	0.704	**0.748**
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.911	0.880	**0.895**	o	0.734	0.733	**0.733**	v	0.612	0.952	**0.745**
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.835	0.963	**0.894**	o	0.664	0.808	**0.729**	c	0.749	0.733	**0.741**
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.972	0.827	**0.894**	o	0.643	0.834	**0.726**	v	0.598	0.950	**0.734**
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.905	0.882	**0.893**	o	0.749	0.699	**0.723**	c	0.750	0.718	**0.734**
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.918	0.869	**0.893**	o	0.608	0.885	**0.721**	v	0.596	0.947	**0.732**
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.918	0.868	**0.892**	c	0.787	0.645	**0.709**	nalm1	0.597	0.928	**0.727**
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.910	0.868	**0.889**	o	0.657	0.768	**0.708**	v	0.586	0.944	**0.723**
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.924	0.855	**0.888**	o	0.760	0.654	**0.703**	v	0.583	0.944	**0.721**
	c	0.916	0.753	**0.827**	a	0.924	0.855	**0.888**	o	0.564	0.907	**0.696**	v	0.583	0.941	**0.720**

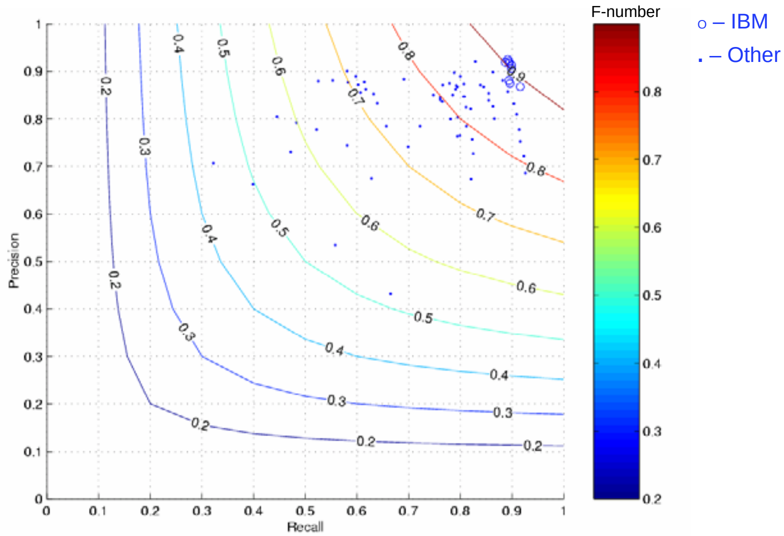
Four of the ten IBM runs are highlighted: IBM's best system of TRECVID 2001, 2002, and 2003, and the baseline 2003 system.

All top seven runs in all measures are IBM.

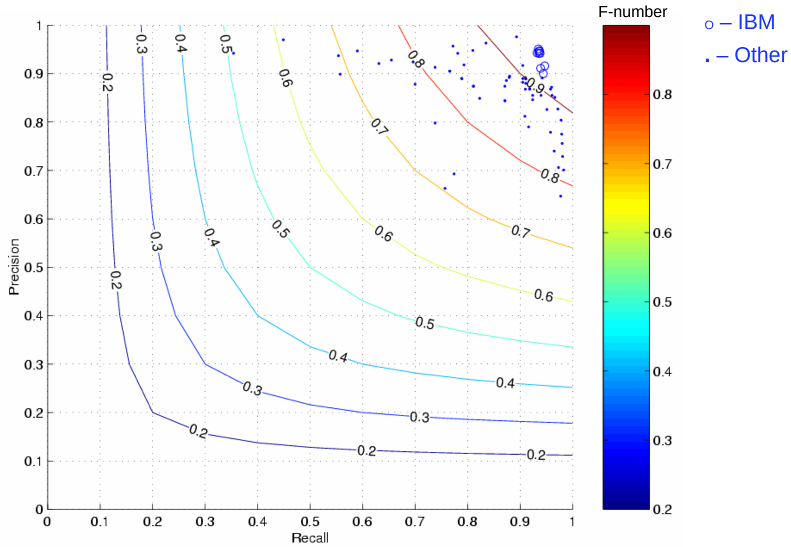
Summary

- Multiple pairs of frames
- Finite State Machine
- Frame comparison: RGB, Edges, Thumbnails
- Adaptive thresholds
- System monitoring tools
- Systematic work with an SBD development set

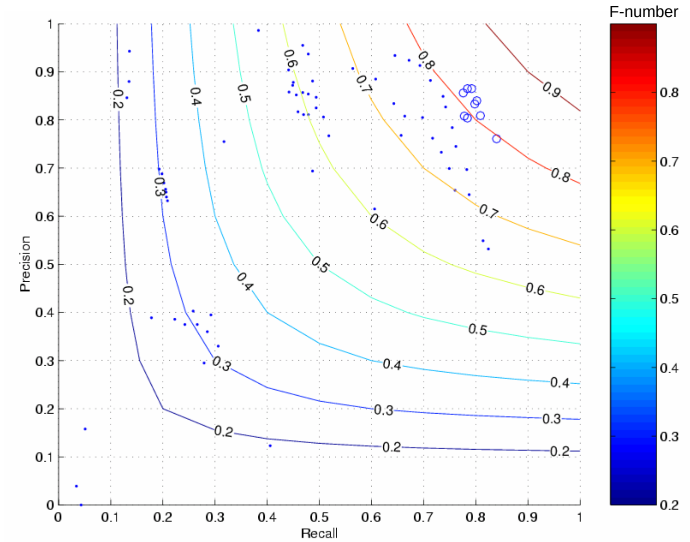
Detection of All Transitions



Detection of Cuts



Detection- Gradual Changes

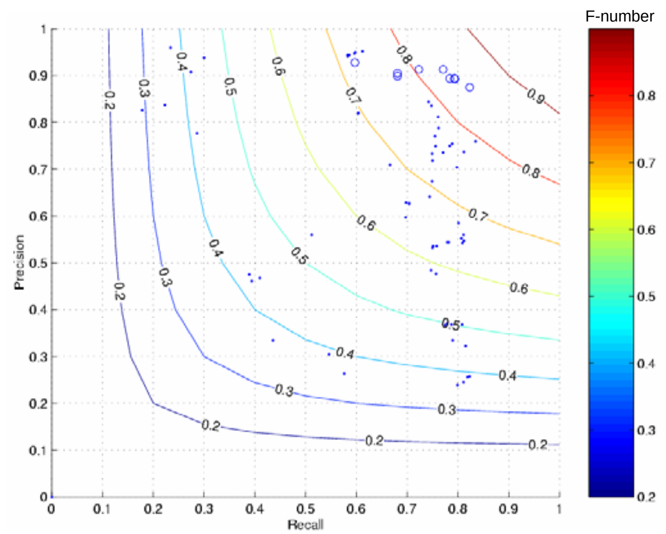


o – IBM

. – Other

IBM Research | TRECVID03 Team

Gradual Changes – Frame Accuracy



o – IBM

. – Other