Exploring the Utility of Fast-Forward Surrogates for BBC Rushes

Jun Yang
juny@cs.cmu.edu
School of Computer Science
Carnegie Mellon University

TRECVID BBC Rushes Summarization Workshop
October 31, 2008

Talk Outline

- Look at a few Baseline Summarizations (MS221050, MRS158381) – simple 50x speedup
- Rationale for “Fast-Forward Surrogates”
- Two CMU variations on 50x for submitted runs:
  - Better noise shot removal
  - Incorporation of audio
- Look at the CMU submitted runs
- Results and Discussion

TRECVID 2008 BBC Rushes Summarization

- Video summary is “a condensed version of some information, such that various judgments about the full information can be made using only the summary and taking less time and effort than would be required using the full information source”
- Maximum 2% duration

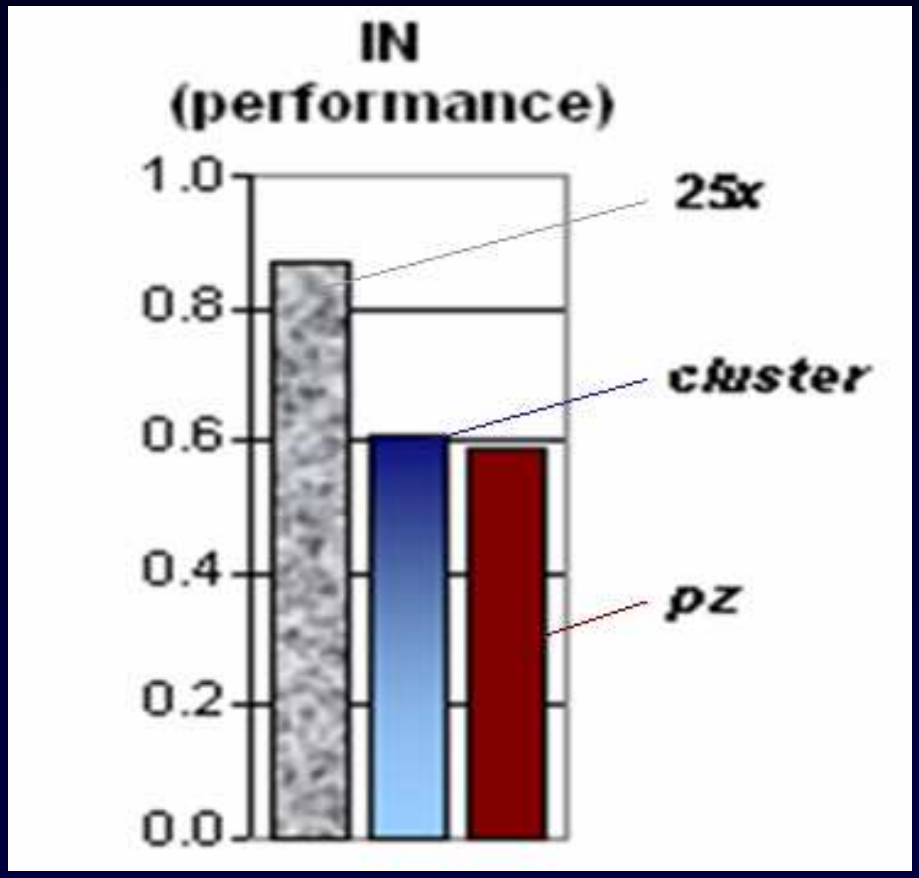
CMU TRECVID 2008 BBC Rush Summaries

- Video summary is “a condensed version of some information, such that various judgments about the full information can be made using only the summary and taking less time and effort than would be required using the full information source”
- Maximum 2% duration
- Our emphasis: Inclusion (recall) measure **MOST** important and defines purpose of the summary, so maximize IN first, then address other subjective metrics
- Do not vary length, i.e., use full 2% for baselines and our submitted runs
- Study effects of including audio into the summary

Baseline Summary
Demonstration (plain 50x)

Why 50x Baseline? (1) Speed-Up Works

CMU at *TVS'07* showed simple speedup (**25x**) significantly better than "clever" approaches **cluster** and **pz**



Why 50x Baseline? (2) Prior Research

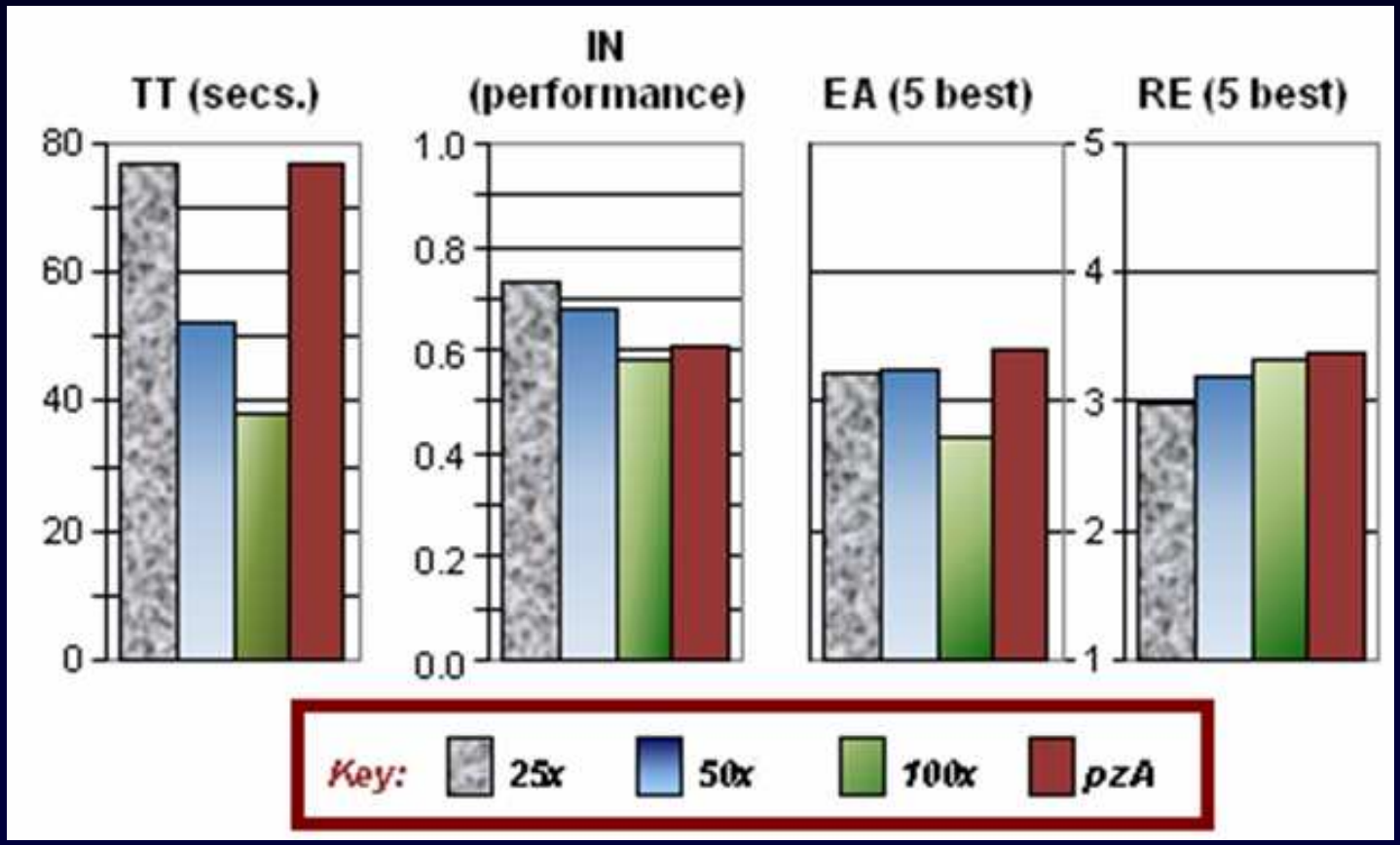
UNC researchers looked specifically at frame rate speedup for documentary videos:

How Fast Is Too Fast? Evaluating Fast Forward Surrogates for Digital Video. In *Proc. ACM/IEEE JCDL 2003*.

Testing 32x, 64x, 128x and 256x with human subjects, they concluded that 64x supports good performance and user satisfaction.

Why 50x Baseline? (3) 50x Works for Rushes

CMU at *CIVR'08* showed empirically (15 subjects (8 female, 7 male; average age 25.7) that **50x** keeps high ***IN***, fast ***TT***



Summary on Baseline, and What CMU Did

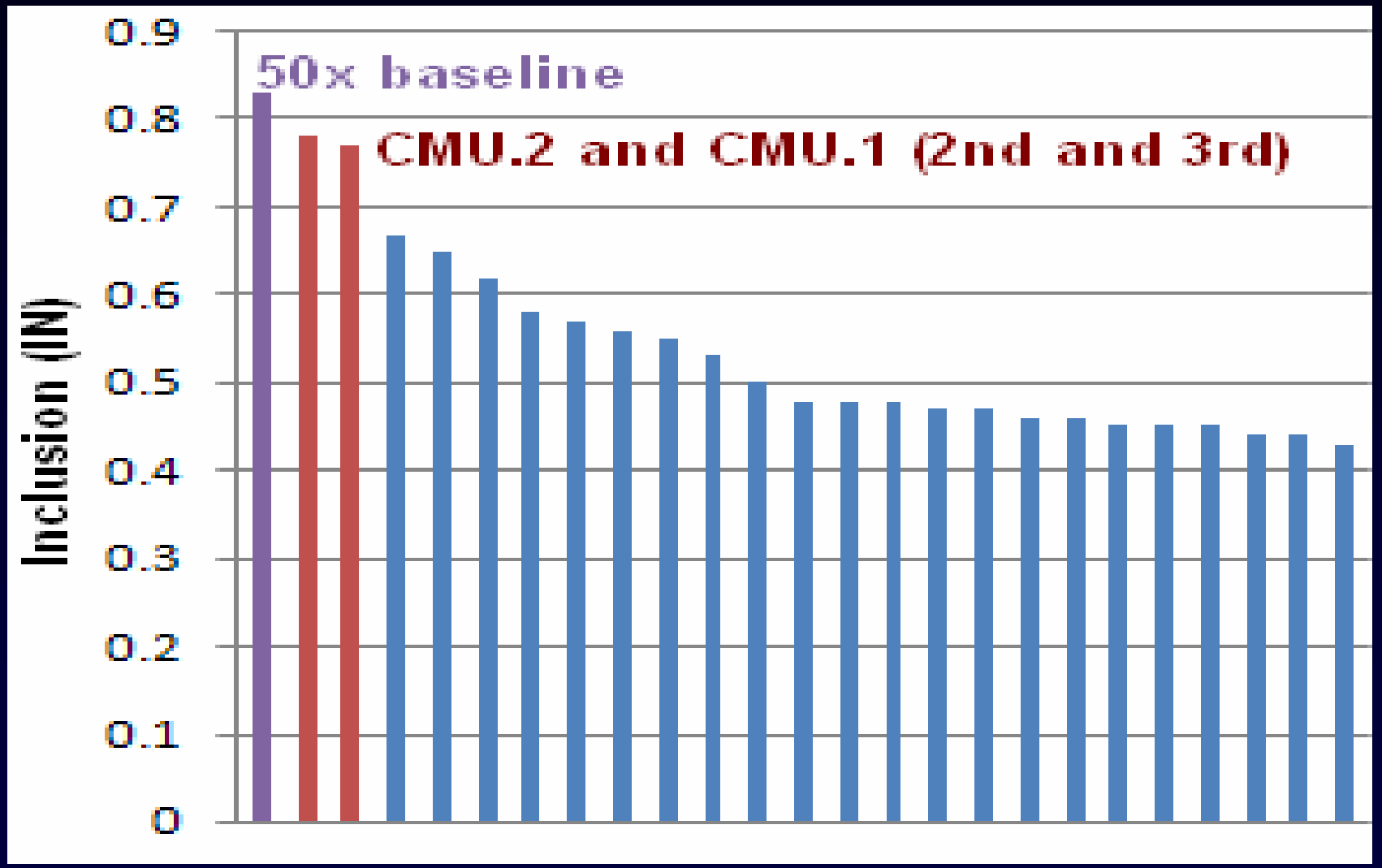
- To maximize ***IN*** for 2% summary, simple 50x approach will work well, based on:
  1. Speed-up works (25x produces superior ***IN*** for *TVS'07*)
  2. Prior work by UNC (see reference 10 in paper)
  3. 50x tested well with BBC Rushes in 2007 (see *CIVR'08*)
- "Baseline" kept simple, plain vanilla sampling every 50th frame, with no junk shot removal, no audio
- Points for study in our CMU submissions:
  - Will adding in junk shot removal improve summary?
  - Will adding in an audio component improve summary?

Details on CMU Submitted Runs

- Common Features:
  - 50x frame rate
  - Keep some audio representation in the summary (coherent, recognizable audio segments)
  - Remove junk frames
  - *Quad chart of 4 “people” shots added to start and end frames to perhaps tune viewer to people shown in video*
- Differences between CMU.1 and CMU.2
  - CMU.2 more aggressively removes junk frames (better recall in junk frame clapper detector, but less precision, so non-junk frames sometimes removed too)
  - CMU.2 backfills freed up space with pan-zoom automatically detected sequences

CMU Submitted Summary Demonstrations:
1.50x plus junk removal plus audio
2.Same, but more aggressive junk removal

Results


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Inclusion (IN)
50x baseline
CMU.2 and CMU.1 (2nd and 3rd)

Discussion of Time and Performance

- 50x strategy provides excellent coverage of video
- Our goal of maximizing performance (***IN*** score) achieved
- Excellent performance comes at a cost: slowest, 3rd and 4th slowest times of all the graded runs, but:
  - We backfilled to keep summaries 2% to focus study on summary make-up, not summary length
  - Task time was about double summary time, but getting 75% or better inclusion required only a 4% investment of user time – tremendous savings over watching full video!

Discussion of Subjective Metrics *RE* and *TE*

- 50x will never score well on “contains lots of duplicate video” or “had a pleasant tempo/rhythm”, and the baseline, CMU.1 and CMU.2 scored at the bottom of the judged runs for these metrics, but:
  - What is “duplicate video?” If duplication occurs very fast at subsecond rate to reinforce something shown too fast the first time, is it still bad?
  - Do end users care about tempo/rhythm? Was this question motivated by real-world user concerns?
- Concern that we TRECVID researchers on this task abandoned a metric used a year earlier, ***EA***
  - Difficult to compare ***EA*** of 2007 with ***RE***, ***TE*** of 2008
  - EA was closer to real-world concerns: **ease of use**, without introducing bias regarding tempo or redundancy

Discussion of Subjective Metric *JU*

- We attempted separation between baseline (should have lowest ***JU*** score as it implemented no junk removal), CMU.1, and CMU.2 (latter having aggressive junk frame removal)
- CMU runs did distinguish themselves from baseline, but not from each other

	JU	*RE*	*TE*
50x baseline	**2.66**	**2.02**	**1.44**
CMU.1	**3.02**	**2.28**	**1.76**
CMU.2	**2.96**	**2.25**	**1.64**

Conclusions

- If ***IN*** should be maximized for BBC rushes 2% summaries, 50x techniques work quite well
- Empirically, does tempo (***TE***) matter, or redundancy (***RE***), or just ease of use (the missing ***EA***)?
  - Improve detector for “significant” pans/zooms
  - Sacrifice coverage for pan/zoom inclusion
- Audio snippet selection still needs work; perhaps that is why it did aid subjective metrics for CMU.1 and CMU.2

Thanks to NIST, BBC, and TRECVID organizers for making this investigation possible. This work supported by the National Science Foundation under Grant Nos. IIS-0205219 and IIS-0705491