Linking Video Segments to Relevant Wikipedia Content
Victor Lavrenko, Johanna Moore, Sean Moran

Goal: facilitate access to the BBC archive
- BBC video archive
  - ~10^6 feet of film, ~10^6 hours of video
  - unique content: British cultural heritage
- Mostly inaccessible to the public
  - will take 80+ years to digitise
  - long, monolithic programs
  - no metadata / transcripts
    - some available in analog form
- Turn archive into set of LEGO bricks
  - break programs into topically-coherent chunks
  - annotate each chunk using external resources
  - cross-link chunks by topic, flow
  - serve as enabling technology for down-stream applications
    - intelligent search and navigation over the archive
    - personalized TV programs assembled on-the-fly

Streaming architecture
- Massive volumes of data, need low latency (on-demand)
- Computationally-intensive, but mostly space-localized
  - similarity of nearby frames
  - motion/transition detection
  - FFT on audio
  - agglomeration (adjacent)
- Good fit for GPU-based implementation

Approach
- Identify scenes using visual/audio similarity, motion
- Detect & transcribe speech, align to script (if available)
- Agglomerate into hierarchy using text, visuals, prosody
- Align with best-matching portions of Wikipedia articles
  - phonemic matching to overcome OOV errors in recognizer
  - use hierarchy for context (helps with vocabulary mismatch)

Outcomes (jointly with the BBC)
- Roll out interactive platform for searching / browsing / annotating the BBC archive
- TSB application for next-generation intelligent television viewing platform