Dissertation Defense Presentation

Feature Based Video Sequence Identification

by
Kok Meng Pua

Committee : Prof. Susan Gauch (Chair)
Prof. John Gauch
Prof. Joseph Evans
Prof. Jerry James
Prof. Tom Schreiber

Agenda

- Driving Problem
- The Goals
- Our Solution
- Related Work and Pilot Work
- The Video Identification Technique Design Approach
- Experimental Results and Discussion
- Conclusion and Future Work

Driving Problem

- Television and video streaming over the Internet enable us to see latest events and stories
- Large portion of these news and video sequences are repeated
- Issues:
  - inefficient use of storage media in archives
  - ineffective use of a viewer's time
- Need:
  - a fully automated topic-based video tracking system

The Goals

- Design and develop a feature-based automatic video sequence identification and tracking technique
  - real time video processing
  - video stream domain independence
- Not a topic-based tracking system
- Goals:
  - real time video sequence identification and tracking technique with high accuracy
  - lossless compression - reduce storage requirement by keeping only unique video sequences

Our Solutions

- Technique:
  - Combination of video hashing and frame by frame video comparison
- Two standalone systems were implemented

  a) Video Processing System
    - Video Feature Abstraction and video sequence segmentation

  b)Video Sequence Identification System
    - Video Sequence Hashing
    - Video Sequence Comparison
    - Video Archiving and Tracking
- Located repeated video sequences with high accuracy (>90% in recall)
- Video input domain independence

Related Work – Image Abstraction and Similarity Measure

Methods	Researchers	Description
Color histogram	Flickner ‘95 (IBM) Pentland ’95 (MIT) Huang & Pass ’97 (Cornell)	-Global color distribution of an image -Easy to compute and insensitive to small changes in viewing position
Image Partitioning with Color Moment	Stricker ’96 (Swiss Federal Institution)	- Divide an image into five overlapping regions
Color Coherent Vector	Pass & Zabih ’96 (Cornell)	-Partition pixel value (histogram bucket) based upon their spatial coherence -Fast to compute
Dominant Color	Swain ’91	-a few main colors -less storage size
Color Moments	Chang & Smith ’96 (Columbia)	-easy to compute -less storage size -more robust than histogram

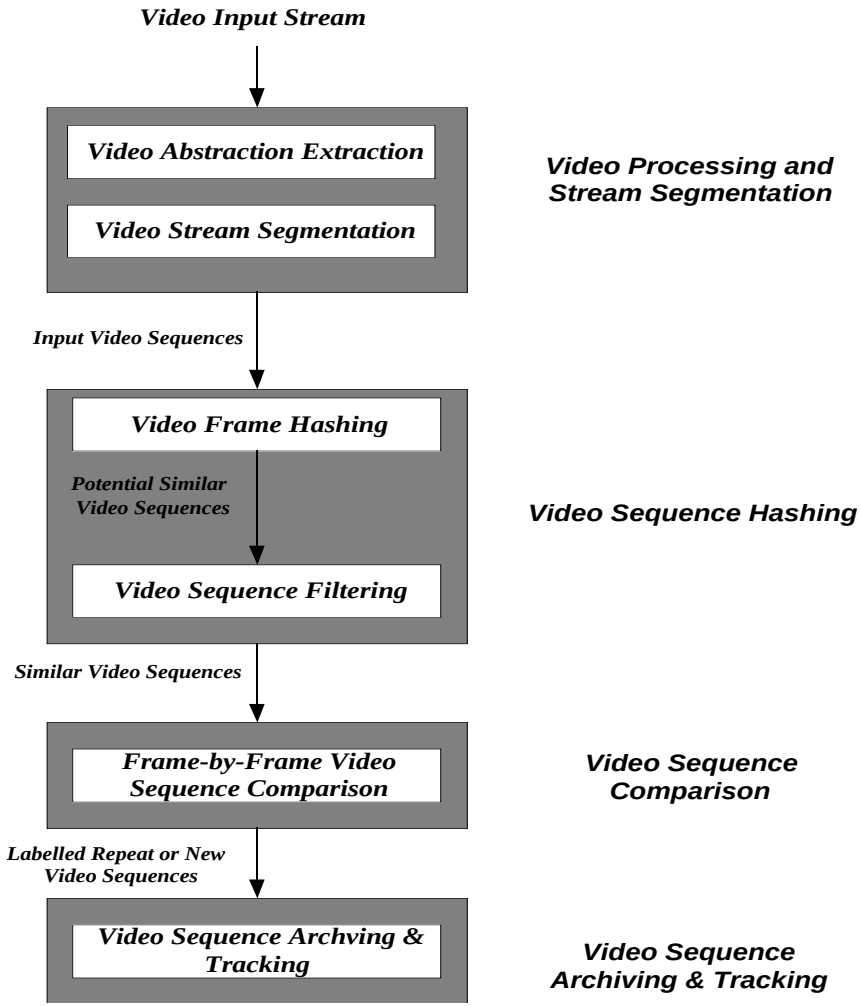
Related Work - Video Sequence Abstraction & Similarity Measure

Methods	Researchers	Description
Key frames	Zhang ’95 (NUS)	-Map the entire video sequence to some small representative images -Use still images as key frames
Centroid Feature Vector	Wu ’98 (Zhejiang University)	-created from key frames that composed of 35 numbers (32 color features and 3 texture features) -Less storage and faster similarity computation
Rframe	Arman ’94 (Princeton)	-A body (10^{th} frame of the video sequence, values of four motion tracking regions, and size of sequence -Shape (moment invariant) and color (histogram) properties of Rframes
Video Signature	Cheung ’00 (California University)	-9 signature frames per video -very expensive
Frame-by-Frame (9 Color Moments)	J. Gauch ’99 (KU)	-frame by frame comparison -less storage size but preserve uniqueness of video

Pilot Work

- The VISION Digital Library System
  - test bed for evaluating automatic and comprehensive mechanisms for library creation (video segmentation) and content-based search, retrieval, filtering and browsing of video across networks.
  - support real time content-based video scene detection and segmentation using combination of video, audio and closed caption.
- Video Seeking System (VIDSEEK)
  - web-enabled digital video library browsing system.
  - dynamic-on-demand clustering allows users to organize the video clips based on multiple user-specified video features (color, shape, category)
  - category-based browsing allows users to interactively and dynamically filter the VISION digital video library clips based on a given set of constraints (video source, keywords and date of capture)
- VidWatch Project (Video Authentication)
  - color moment based frame-by-frame video information processing technique
  - Detect and record video content difference between two television channels

Four Main Processes of The Video Sequence Identification and Tracking System



Video Processing and Stream Segmentation

- Video Processing System (VPS)
  - A standalone real time video segmentation system on NT platform
  - Use Osprey digitizer board for video stream digitization.
  - Two main functions : video sequence creation and video abstraction extraction

Video Processing and Stream Segmentation

- Video Sequence Abstraction
  - Compare abstraction for every frame in the video sequence but not key frames only
    - Video abstraction method used should require small storage and also easy comparison computation.
  - Use first 3 color moments of each primary color component (Red, Blue, and Green)
  - 3 moments are the mean, standard deviation, and skew of each color component of a video frame.
  - 9 floating points per video frame

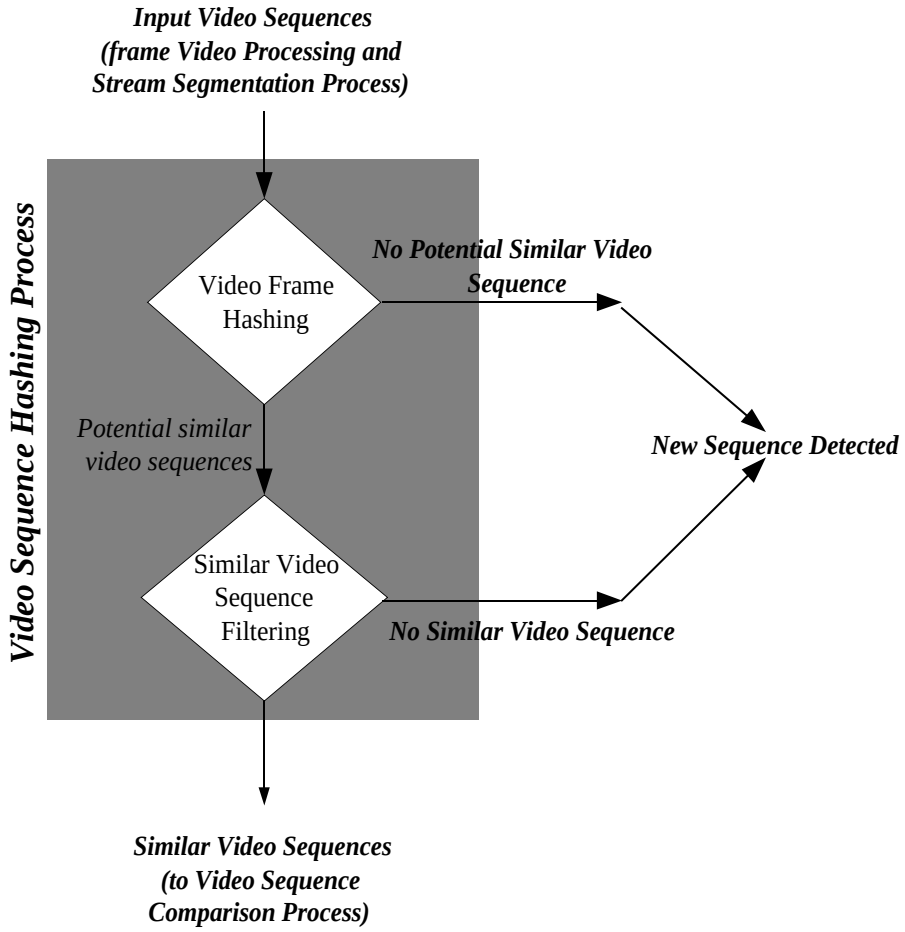
Video Processing and Stream Segmentation

- Video Sequence Creation
  - A video sequence is a video shot (an image sequence that represents continuous action and corresponds to a single action of the camera)
  - Use video segmentation technique developed in VISION project
  - Shot boundary detection : combination of average brightness, intensity difference and histogram difference of adjacent frames.
  - Look at the differences between two frames *dt* distance from each other to detect shot boundary with smoother (slow) transition
  - Meta-information captured:
    - Size of video sequence
    - Video Sequence identifier (date and time the sequence is captured)

Video Sequence Hashing Process

Purpose:

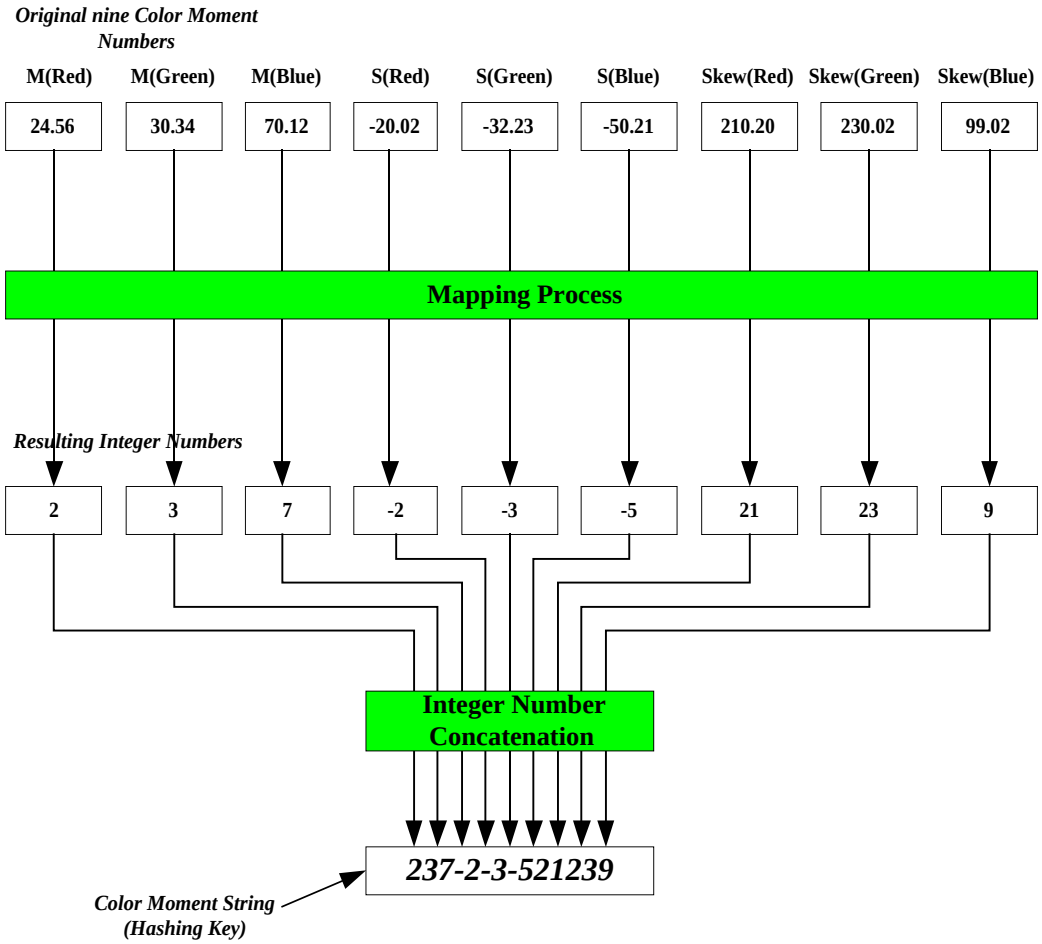
- Reduce the size of number of sequences requiring frame-by-frame comparison by selecting only similar video sequences which have many similar frames



Video Sequence Hashing Process

- Video Frame Hashing:
  - Map video frames into similar frame buckets and hash similar frames instead of “identical” video frames
  - Each video frame is represented by a color moment string
  - Color Moment String – map 9 original (10 to 1 mapping) floating numbers into 9 integers and concatenate them to become a character string
  - Use concatenation of all digits of the 9 original float numbers(moment values) as the hashing key proved to be ineffective due to noise introduced by both transmission and digitization into the video source.
    - Identical values unlikely for repeated video broadcast

Example of Color Moment String Mapping



Video Sequence Hashing Process

- Video Frame Hashing:
  - Mapping Ratio Issues:
    - Mapping ratio too small : actual identical frames fall into difference similar buckets
    - Mapping ratio too large : none-identical frames fall into same similar buckets
  - Experimental result showed 1% of mapping error (identical frames into different similar buckets)

Video Sequence Hashing Process

- Video Frame Hashing Cost Estimation:
  - Total video frame hashing cost for an input video sequence

  $Cost(m) = \sum [H(n) + L(n)]$ for $n = 1 \ldots m$

  m = size (total video frame count) of the video sequence

  H(n) = hashing cost

  L(n) = linked list traversal cost

  - Independent of hash table size (video archive size)
  - Experimental results showed an average hashing time of *500ms* for each input video sequence

Video Sequence Hashing Process

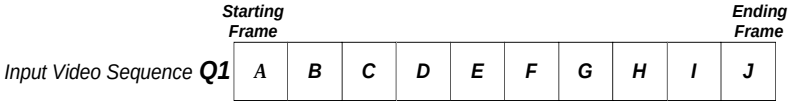
Video Sequence Filtering :

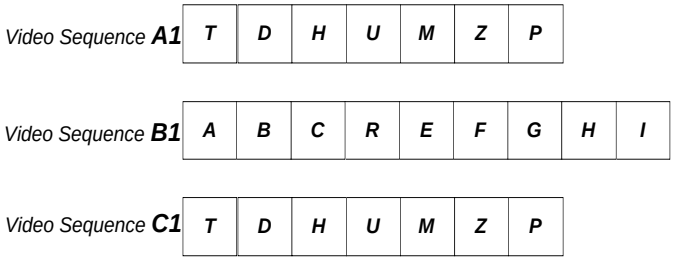
- Second component of similar video hashing process
- Purpose:
  - Identify truly similar video sequences
- Potential similar video sequences are filtered to remove sequences whose degree of similarity is below some threshold
- Two video sequences are similar if:
  1. The size difference of the two video sequences is less than 10%
  2. The percentage of frames in the two video sequences that have identical color moment strings is > 30% (overlap threshold)

Video Sequence Comparison Process

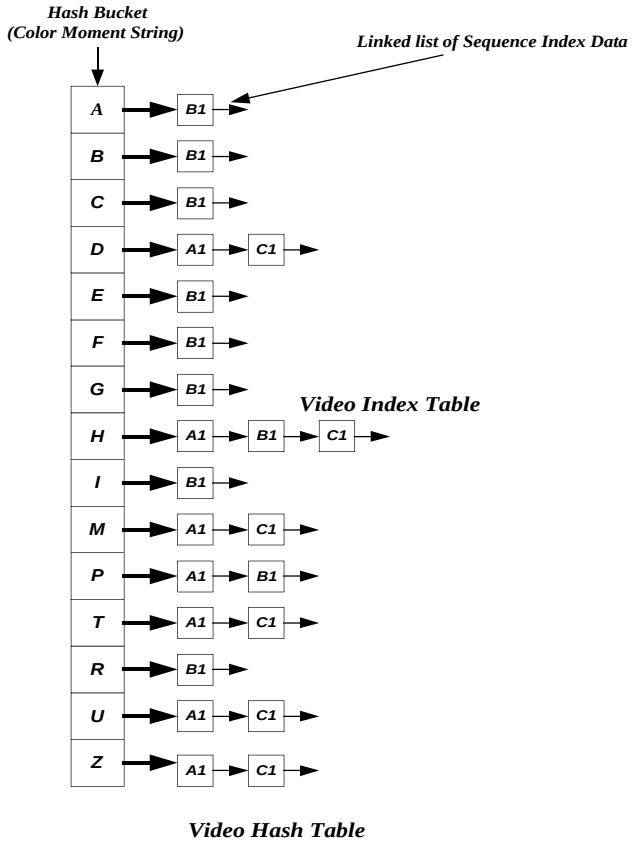
- Video Sequence hashing results in false positive matches:
    - Color moment strings used for hashing are built from approximate color moment values
    - consider only the percentage of similar video frames and ignore their temporal ordering
- A more accurate frame-by-frame comparison is required
- Sum of the absolute moment difference of two sequences is calculated
- One video sequences is detected as a repeat of the other if sum of their absolute moment difference < moment difference threshold (10.0)

An Example of Video Identification Process

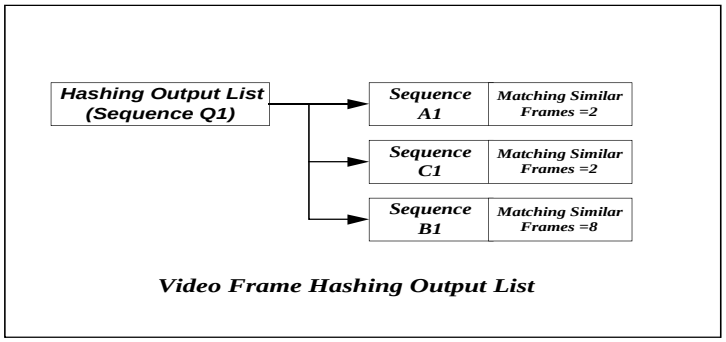




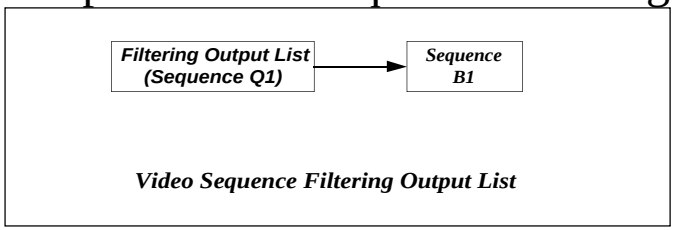
Note :
Each block is a video frame with its color moment string represented by an alphabet



Step 1 : Video Frame Hashing



Step 2 : Video Sequence Filtering



Video Sequence Archiving and Tracking

- Purposes
  - Record sequence identification results
    - Number of video sequences processed
    - Number of detected new and repeat sequences
  - Control and enable video sliding window function
    - Allow the total video archive to grow until it contains 24 hours worth of video sequences
    - Oldest/expired sequences are dropped from video archive as the newer sequences are added
- Use a video index table to capture identification results and keep the total video archive within the 24 hours window size

Experimental Results

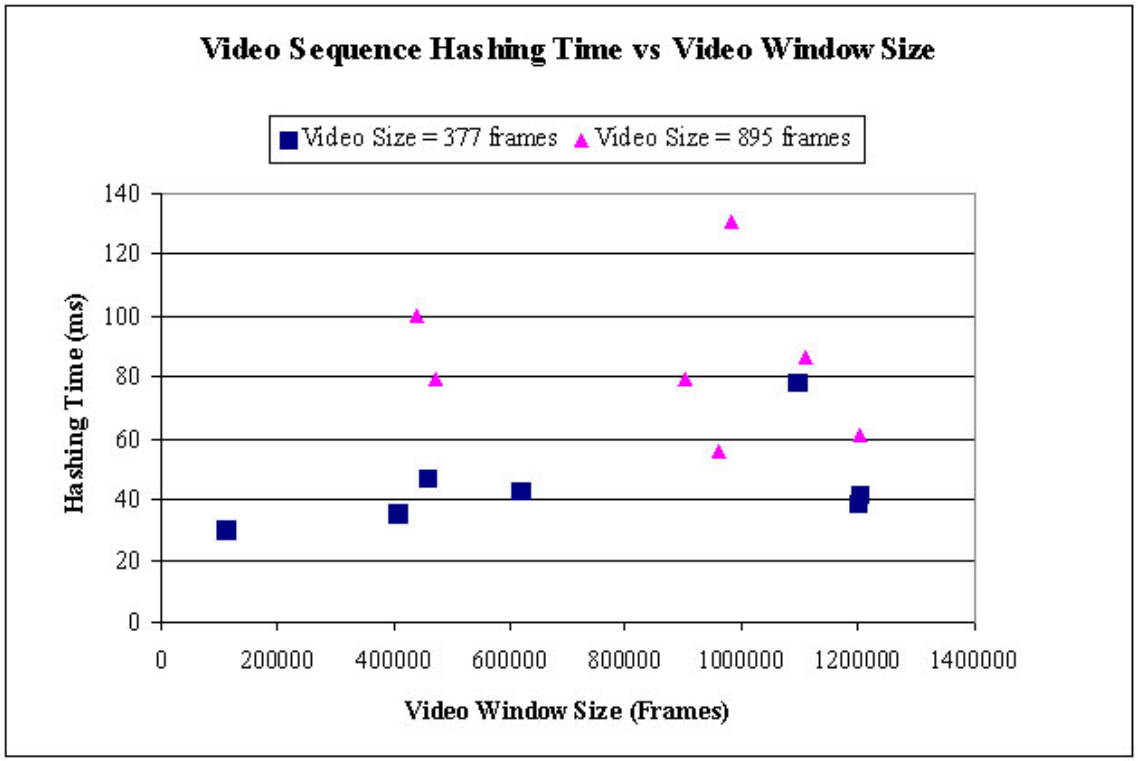
Input video source:

- 32 hours of video stream as primary input source
- A second 24 hours of difference video source for video identification technique input domain independent validation

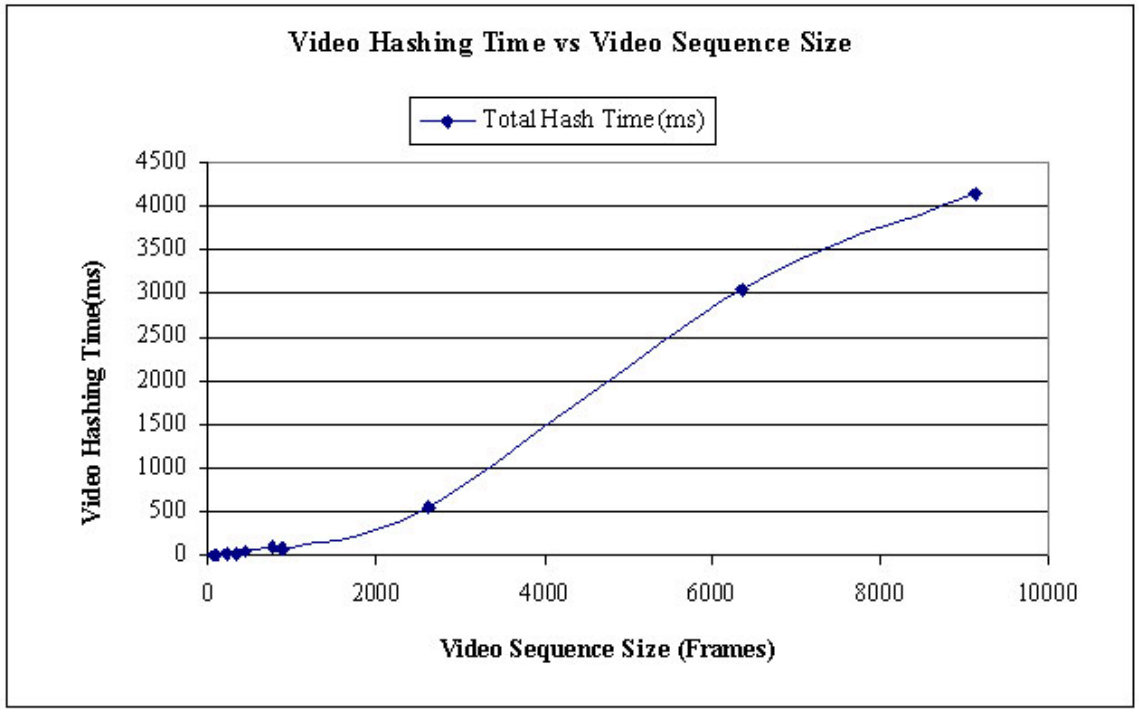
Measurements :

- Video hashing accuracy and efficiency measures
  - Video hashing time
  - Select an optimum overlap threshold value
- Video sequence comparison cost
- Overall video identification accuracy and efficiency
  - Recall and precision measurement
- Achievable storage compression
- Validating video identification technique with different video source

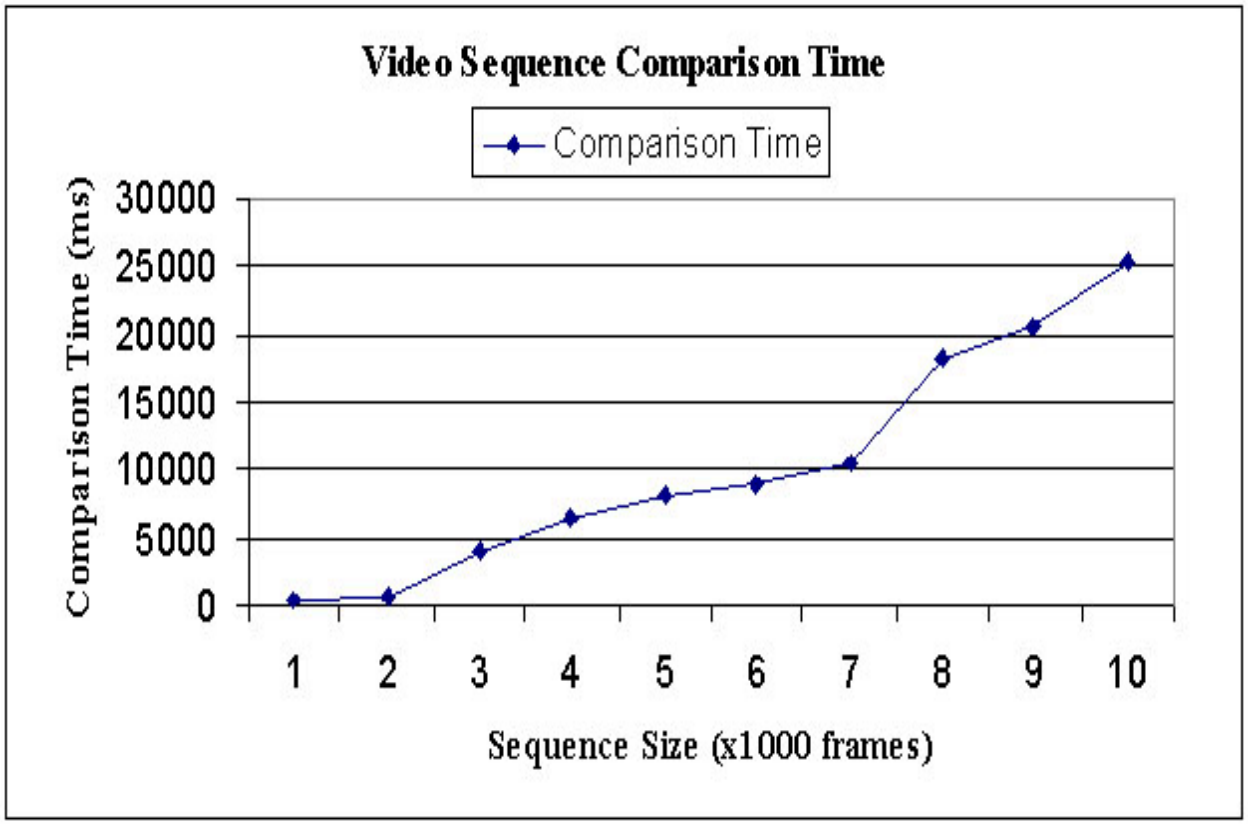
Video Hashing Time vs. Video Archive Size



Video Hashing Time vs. Video Sequence Size



Frame-by-Frame Video Sequence Comparison Cost



Experiment Results

- Recall and Precision Calculation
- Recall
  - The ratio of the number of correctly detected repeated sequences to the total number of true repeated sequences in the video archive
  - Recall = true positives / (true positives + false negatives)
- Precision
  - The ratio of the number of correctly detected repeated sequences to the number of detected repeated sequences in the video archive
  - Precision = true positives / (true positives + false positives)
- Term definition:
  - True positive : detect repeat, and it is repeat
  - True negative : detect new, and it is new
  - False positive : detect repeat, and it was new
  - False negative : detect new, and it was repeat

Example of Recall and Precision Calculation

- Inputs : S1.1, S1.2, S1.3,S1.4, S2.1, S2.2, S2.3
- S1.2, S1.3 & S1.4 are repeat of S1.1 Also, S2.2 & S2.3 are repeat of S2.1

Outputs:

S1.1 : Detected New
S1.2 : Detected Repeat of S1.1
S1.3 : Detected New
S1.4 : Detected Repeat of S1.1, S1.2 & S1.3
S2.1 : Detected New
S2.2 : Detected Repeat of S2.1
S2.3 : Detected Repeat of S2.1 & S1.2

Input Sequence	Total Detected Repeated Sequences	Correctly Detected Repeated Sequences	True Repeated Sequence		True Positive	True Negative	False Positve	False Negative
S1.1	0	0	0		0	1	0	0
S1.2	1	1	1		1	0	0	0
S1.3	0	0	2		0	0	0	2
S1.4	3	3	3		3	0	0	0
S2.1	0	0	0		0	1	0	0
S2.2	1	1	1		1	0	0	0
S2.3	2	1	2		1	0	1	0
			TOTAL		***6***	***2***	***1***	***2***

Recall = true positives / (true positives + false negatives) = 6 / (6+2) = 0.75
Precision = true positives / (true positives + false positives) = 6 / (6+1) = 0.85

Choosing An Optimum Overlap Threshold Value

Overlap Threshold (%)	0	10	20	30	40	50	60	70	80
True Positive	6108	6082	6044	5953	5779	5439	4781	3600	1886
False Positive	49734	4788	2264	1648	1154	833	505	263	126
True Negative	202	872	1067	1148	1194	1245	1346	1552	1963
False Negative	222	248	286	377	551	891	1549	2730	4444
Recall	0.96	0.96	0.95	0.94	0.91	0.86	0.76	0.57	0.30
Precision	0.11	0.56	0.73	0.78	0.83	0.87	0.90	0.93	0.94
Average Hashing Time Per Sequence(ms)	541	540	544	536	549	522	521	494	518

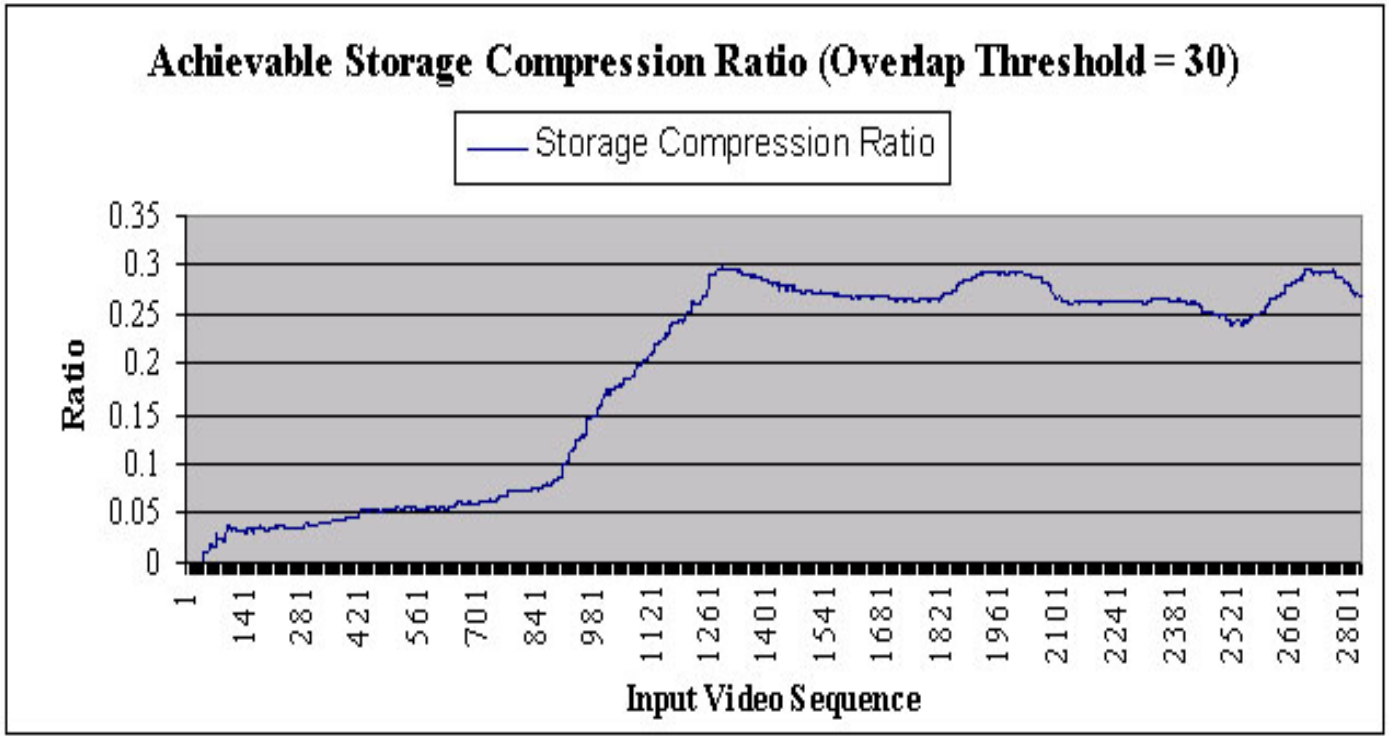
- Overlap threshold :
  - percentage of total similar video frames (identical color moment string) over the total video frames of a video sequence
- Recall = true positives / (true positives + false negatives)
  = total correctly detected repeated sequence / total true repeated sequence
- Precision = true positives / (true positives + false positives)
  = total correctly detected repeated sequence / total detected similar video sequences

Overall Video Identification Performance Measurement

-Total of 32 hours of video source (2831 sequences)
-1228 new and 1603 repeat in a 24 hour sliding window

Moment Difference Threshold	Initial	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	15.0	20.0
Tpositive	5953	4090	4756	5134	5456	5598	5743	5809	5831	5889	5917
Fpositive	1648	314	385	438	532	580	595	627	656	741	1039
Tnegative	1148	1498	1406	1358	1326	1299	1273	1259	1247	1222	1179
Fnegative	377	2240	1574	1196	874	732	587	521	499	441	413
Recall	0.94	0.65	0.75	0.81	0.86	0.88	0.91	0.92	0.92	0.93	0.93
Precision	0.78	0.93	0.93	0.92	0.91	0.91	0.91	0.90	0.90	0.89	0.85

Achievable Compression Ratio

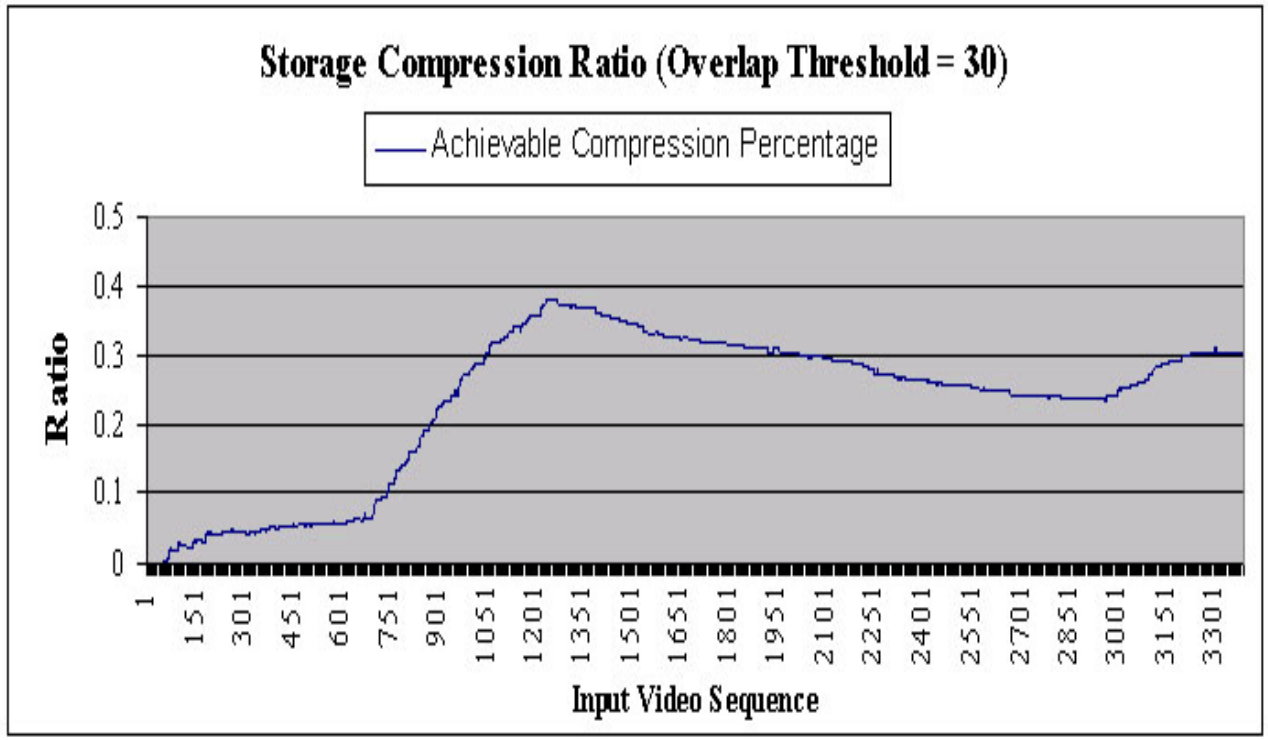


Validating Video Identification Technique with difference Video Source

-Total of 24 hours of video source (3394 sequences)
-1938 new and 1456 repeat

Moment Difference Threshold	Initial	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	15.0	20.0
Tpositive	5341	4148	4666	4953	5119	5198	5219	5236	5245	5325	5341
Fpositive	1324	188	246	308	331	358	371	401	440	613	865
Tnegative	1833	2141	2068	2042	2011	1978	1973	1967	1961	1913	1878
Fnegative	554	1747	1229	942	776	697	676	659	650	570	554
Recall	0.91	0.70	0.79	0.84	0.87	0.88	0.89	0.89	0.89	0.90	0.91
Precision	0.80	0.96	0.95	0.94	0.94	0.94	0.93	0.93	0.92	0.90	0.86

Achievable Storage Compression Ratio



Conclusion

Summary

- This thesis reports on a video sequence identification and tracking technique that can be used to process continuous video streams, identify unique sequences and remove repeated sequences
- The algorithm described is efficient (runs in real time) and effective
- Accurately locate repeated sequences (recall >90% and precision >89%) for two different video streams
- Achieve a compression gain factor of approximately 30% for both video streams

Future Work

- Technique Improvement
  - Partial Matching
    - Detection of subset of a known video sequence or superset of few sequences overlapping one another
  - Disk based Hashing
    - Handle a very large video window size to detect and track repeated video sequences that occur farther apart in time
- User Application
  - Web-enabled Video Stream Browsing System
    - Enable users to search and browse the video archive and view selected video sequences
  - Story based Video Sequence Identification
    - Group video sequences into different stories using video abstraction such as closed caption, audio and video content.