

# Segmentation and Appearance Features Index for Digital Video Data

Hongwon Yun, *Member, KIMICS*

**Abstract**— The numbers of digital video cameras are fast increased. Accordingly, digital video data management is becoming more important. Efficient storing method and fast browsing method still remains to be one of significant issue. In this paper, an optimized data storing process without losing information and an organized appearance features indexing method are proposed. Also, the data removing policy could be used to reduce large amount of space and it facilitates fast sequential search. The appearance features index constructs key information of moving objects to answer queries about what people are doing, particularly when, where and who they move. The evaluation results showed better performance in the transfer time and the saving in storage space.

**Index Terms**— Storing method, Digital video data, Segmentation and feature extraction, Appearance features index.

## I. INTRODUCTION

In many real IP digital video systems, huge data volume comes in continuously and presents several challenges to traditional data storing and access methods [1,2]. Data management of IP digital video is becoming more important according to the number of digital video cameras are installed widely [2-4]. The large volume of video data makes it hard work to store and browse them by just traditional way. Many researchers work and provide efficient video browsing facilities these works are two areas the one is content based video browsing systems and another one is efficient data organization [6,7]. In spite of many research efforts, how to save huge data volume and thus visual items to facilitate fast browsing without losing information still remains to be one of open issue [1,6].

In this paper, an optimized data storing process without losing information and an organized appearance features indexing are studied. The proposed system will be distributed IP cameras platform that is integrates advanced video search tool and efficient storage module. The goal of this advanced video search tool is to provide assistance to operators by retrieval of using timestamp

and features indexing. These indexing allow one to fast retrieval and play back only interesting sequences through huge amount of video data. The storage module has to deal with the large volume of data to store efficiently. It could be happened in usual that there are non-movement objects so that we can remove the number of shots to save storage space.

We wish to partition the video shots into one or more groups of movements and non-movements. Groups of non movement could be reduced large amount of space and fast sequential search is provided. Appearance feature indexing at groups of movement supports to retrieve temporal and people feature. The paper is organized as follows. We first describe the IP digital video system architecture and the structure of segmentation and feature extraction of video shots.

## II. IP DIGITAL VIDEO SYSTEM OVERVIEW

### A. IP Digital Video Architecture

The components of the IP digital video architecture are presented in Fig. 1. This system is composed of computer s connected together through a typical fast Ethernet network. The IP video cameras are plugged on the network directly. Usually the secondary storage used to browse comparatively recent a particular video contents. Large volume of video content is moved to tertiary from secondary storage when the access to secondary is rare. We adopt a deferred removing policy to save storage and to search quickly without losing any information. In Fig. 1, Level 2 storage process saves a tremendous storage space and it allows a fast search to retrieve a stored sequence.

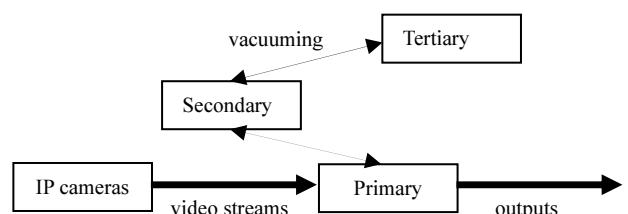


Fig. 1. The IP digital video architecture

Manuscript received October 5, 2010; revised October 22, 2010; accepted November 2, 2010.

Hongwon Yun is with the Department of Information Technology, Silla University, Busan, 617-736, Korea (Email: hwyun@silla.ac.kr)

### B. Deferred Data Removing Policy

The rare used video data to be browsed on secondary storage should be moved to tertiary. That is why volume of video data is huge to manage and waste space unnecessary. Sometimes since the video contents on tertiary storage are browsed to search a particular object to meet the purpose of installed video cameras, the storage system must be able to support loading its video contents to achieve low latency. Our goals of first to obtain an efficient storage system that can be meet the saving of storage space and second to have a system that allow users to search a particular object without losing information even though the video shots including non-movement object are removed. The basic components and simple processes of the deferred data removing structure are showed in Fig. 2.

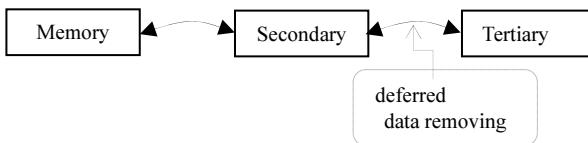


Fig. 2. Deferred data removing

## III. SEGMENTATION AND FEATURE EXTRACTION

### A. Segmentation and Feature Extraction Overview

Detection of moving objects in a video shot is very important for the clustering using data removing. Although movement is the most prominent property of video content, the recognition of moving object from video streams is still restricted. The approach to recognize motion objects is assumed in this paper and we can define the mechanism that divides input video shots two types of segmented frames. The segmented and packaged frames without motion objects can be moved to tertiary having featured extraction. The segmentation component to compose indexing requires appropriate simple statistics in a stream of shots. Fig. 3 shows the process to divide input video streams and to extract the features from segmented frames. Extracted features from segment become component of appearance features indexing.

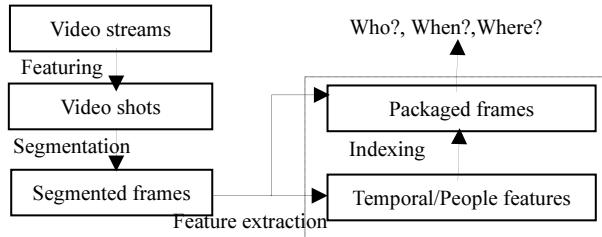


Fig. 3. Segmentation and feature extraction of video shots

### B. Process of Segmentation and Feature Extraction

Video surveillance of human activity usually requires people to be tracked. In the process leading from input video shots, two steps are significant: detection of moving objects and then segmentation and featured extraction of each segmentation. Foreground detection algorithms are normally based on background subtraction algorithm. These methods are based on extracting motion information threshold the differences between the current image and a background or the previous image respectively. We suppose that these methods are adopted to detect moving objects because they are widely used. Detection is performed by the algorithm from consecutive video shots. Detection is achieved and then it gives a check point on video shots. A pair of checkpoint is basis to produce segmentation.

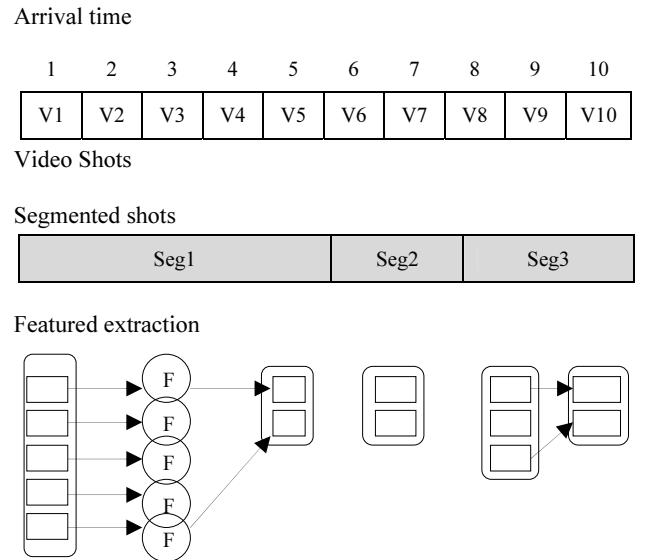


Fig. 4. Process of segmentation and feature extraction

In the segmentation and feature extraction, video shots are divided and packaged into segmented shots as the segmentation, and the segmentation may be represented by W3 information (When? Where? Who?) extracted from it. In Fig. 4, from second frame on a V2 shot to fourth frame on a V4 shot are removed when an A1 shot includes moving objects and V2, V3, V4 shots have not moving objects and moving object is appeared on a V5 shot. W3 information is extracted from a V1 shot and a V5 shot and they are used to compose appropriate index node of appearance features. Each node contains information about shot features such as W3 information as shown in Fig. 5.

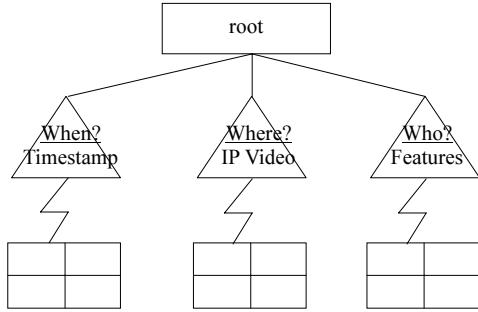


Fig. 5. Appearance features index including W3 information.

#### IV. APPEARANCE FEATURES INDEX AND CLUSTERING

##### A. Appearance Features Index

The appearance features index constructs important information of video shots including moving objects to answer queries about what they are doing, particularly when, where and who they act.

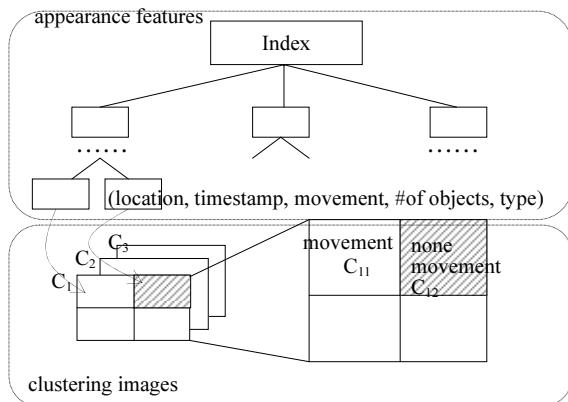


Fig. 6. Appearance features and clustering images

We denote a leaf node by the set of attributes  $S = (L, T, M, N, Q)$ , where  $L = \{l_1, l_2, \dots, l_i\}$ , where  $l_i$  represents the  $i^{\text{th}}$  IP digital camera,  $T = \{t_1, t_2, \dots, t_j\}$ , where  $t_i$  denotes timestamp each video shots,  $M = \{0, 1\}$ , where 0 means none movement and 1 means movement,  $N = \{1, 2, \dots, k\}$ , where  $i$  is number of objects, and  $Q = \{x, y\}$ , where pair  $(x, y)$  points the type of movement on the 2 dimensional matrix.

We supposed that a tracking module is adapted to our platform. This tracking module is derived from the method proposed by Piater and Crowley [8] and the derived detection module is presented at [10]. A detection module takes the list of zones as input and computes a list of blobs. The detection image  $D$  is the threshold difference between the current image  $I$  and the background image  $B$ , computed the following equation [10]:

$$D = \text{threshold}(\min(\sum_i |I_i - B_i|, I_{max}), t) \quad i = r, g, b \quad (1)$$

where  $t$  is a parameter of the algorithm. The background is updated to take small illumination changes into account:

$$B_i = \alpha I_i + (1 - \alpha) B_{i-1} \quad (2)$$

where  $\alpha$  is a parameter of the algorithm.

The supposed system adopts this tracking people algorithm to detect movement of objects on the shots. We adopt the same strategy used in scalable a video stream into shots and process them shot by shot.

**Algorithm:** simple sketch of segmentation and clustering

**Input:** A set of video shot  $V = \{v_1, v_2, \dots, v_n\}$ .

**Output:** A set of movement cluster  $CM = \cup v_i$ , a set of non-movement cluster  $CN = \cup v_j$  and a set of appearance feature  $S = \{s_1, s_2, \dots, s_k\}$  for each element of set of  $CM$ .

**Step:**

1. Detection of no movement for each shot;
2. Gathering appearance features for each movement shot and clustering for each movement shots;
3. Remove video shots for non-movement data and clustering for each non-movement shots.

We show that using a simple segmentation and clustering method works equally well as the more complicated segmentation and clustering. In the simple approach, after detection of moving objects together with extracted features, each video shot is clustered into one clustering.

**Algorithm:** segmentation and clustering

$X = \{x_{i,j}\}$  is sequential video streams of  $i^{\text{th}}$  IP digital camera.

**Input:** A video streams of unit-length data  $\{v_1, v_2, \dots, v_j\}$

**Output:** A set of video segment  $C = \{c_1, c_2, \dots, c_n\}$  and a set of attribute  $S = \{s_1, s_2, \dots, s_n\}$

**Steps:**

1. Initialize the  $S$  attribute ( $L, T, M, N, T$ ) as empty and Set  $D$  as threshold for detection;
2. Reading in next  $s$  until the data are exhausted.
  - (a) Set  $i = m = n = 0$ ,  $i = i + 1$ ,  $S = \text{null}$ ,  
 $m = m + 1$ ,  $n = n + 1$ ;
  - (b) if unit data  $s_i$  is  $D$  then gathering feather,  
Save  $s = (L, T, M, N, Q)$  for the each unit data;
  - (c) if unit data  $s_i$  is not  $D$  then
    - (i) set  $c = s_{i-m} \cup \dots \cup s_i$  for the each unit data;
    - (ii) remove all unit data  $v_{i+1}, \dots, v_{i+n-1}$  until  $s_i$  is  $D$  and a set  $c = v_i \cup s_n$ ;

There are two features with our segmentation and clustering algorithm. The first is the feature extraction from input video streams. This is straightforward given appearance features index, which assists fast browsing. The second feature is the removing data which have no

movement based on a detection module and then does clustering.

## V. EVALUATION RESULTS

To evaluate the approach we used the five classes video data and each class had five rates of no moving objects, the rate of non-movement in each video class were 0.1, 0.2, 0.3, 0.4 and 0.5. We used digital video data by varying size from 720MB to 3600MB. Unless otherwise mentioned, the algorithm parameters were set at average latency = 4.16 ms, random read seek time = 8.5 ms, and I/O data transfer rate = 600 MB/s. Storing efficiency is measured by transfer time and seek time.

One novel feature of our algorithm is that it can reduce a storage space of video shots with sparse movement. We expect proposed algorithm to be more effective than current browsing methods at video shots rarely movement.

Figure 7 shows the effective performance of the transfer time using the proposed algorithm. When the non-movement rate is 50%, transfer time is prominently reduced less than that of zero percentage rates. The reason is that there is some removed video shots to reduce overhead. We compare the performance of proposed algorithm with that of video size for different horizons at different non movement rates. Figure 8 shows the effective speedup of the algorithm for a varying rate of non movements. When the video size is increasing gradually, the proposed method is significantly better in 50% of the 3 cases. This is because with this method, the detection module to detect the moving objects can usually dedicate to reduce a space of video shots.

The use of current browsing method results in high storage space requirement, which results in waste times. Figure 9 shows the main performance advantage of the proposed approach comes because of the reduction in storage utilization costs. The reduction in storage space with the use of proposed approach shown in Fig. 9 is actually a reduction in processing time per video shot. It is a significant issue in processing time of digital video data. We have evaluated seek time to compare different search algorithm i.e. Sequential browsing and Feature index &. The Feature index & is prominently better than Sequential on all five digital video data set as shown in Fig. 10.

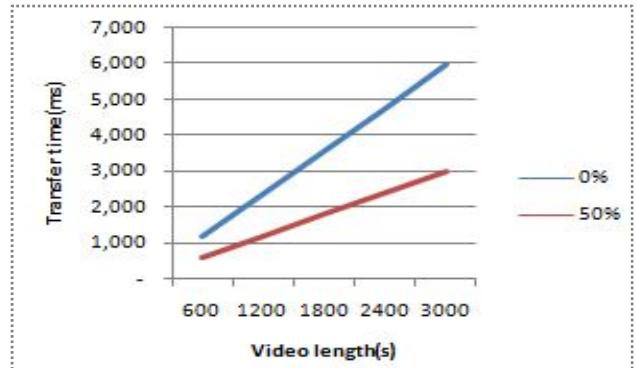


Fig. 7. Transfer time with video data length

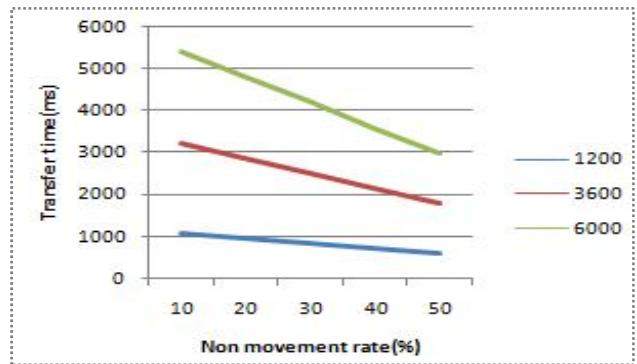


Fig. 8. Transfer time with non-movement

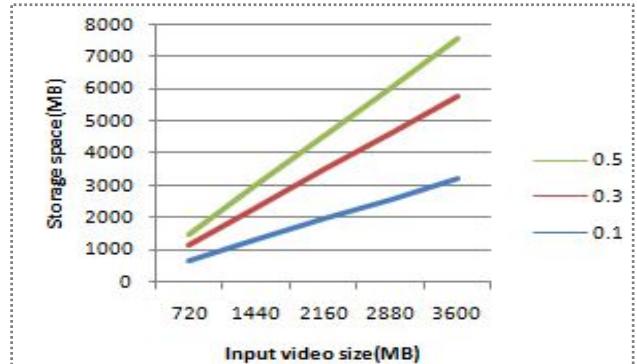


Fig. 9. Storage space utilization

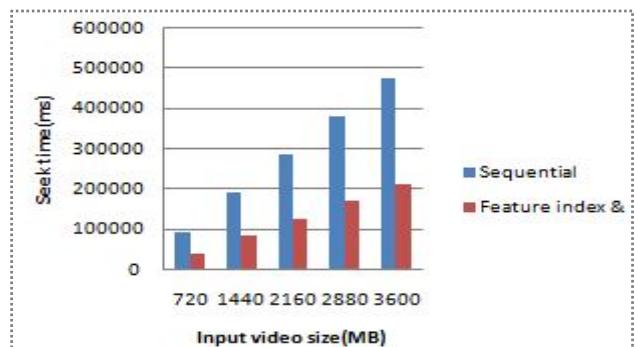


Fig. 10. Sequential and Feature index &

## VI. CONCLUSIONS

Digital video data management is becoming more important according to the number of IP digital video cameras are installed widely. The large volume of digital video data makes it a difficult work to browse and manage them by just traditional way. Efficient storing methods to large data volume and thus visual items to facilitate fast browsing without losing information still remains to be one of significant issue. In this paper, we proposed an optimized data storing process without losing information and an organized appearance features indexing. We wish to partition the video shots into one or more groups of movements and non-movements video shots through the detection of moving objects. The deferred data removing policy could be reduced large amount of space and fast sequential search is provided without losing information. The segmented and packaged video shots without moving objects can be moved to tertiary remaining information of extracted features into the appearance features index. The appearance features index constructs important information of moving objects to answer queries about what they are doing, particularly when, where and who they move. To evaluate the approach, we measured the transfer time and seek time using the five classes' video data. The evaluation results showed the effective performance of the transfer time and the reduction in storage space.

## REFERENCES

- [1] D. Zhong, H. J. Zhang and S F. Chang, "Clustering Method for Video Browsing and Annotation," *Proc. SPIE Storage Retrieval Image Video Database IV*, pp. 239 – 246, 1996.
- [2] I. Haritaoglu, D. Harwood and L. S. Davis, "A Real Time System for Detecting and Tracking People," *3rd IEEE Int. Conf. Automatic Face and Gesture Recognition*, Japan, 1998.
- [3] S. Zhong, 'Efficient Streaming Text Clustering,' *Neural Networks* 18, pp. 790-798, 2005.
- [4] M. Stonebraker, U. Cetintemel and S. Z donik, "The 8 Requirements of Real-Time Stream Processing," *SIGMOD Record*, Vol. 34, No. 4, Dec. 2005.
- [5] B. Rich and D. Thain, "DataLab: Transactional Data-Parallel Computing on an Active Storage Cloud," *IEEE/ACM High Performance Distributed Computing*, pp. 233-234, 2008.
- [6] T. Deselaers, D. Keysers and H. Ney, "Clustering Visually Similar Images to Improve Image Search Engines," *Informatiktage 2003 der Gesellschaft fr Informatik*, Bad Schussenried, Germany., 2003.
- [7] G. Qiu, "Image and Feature Co-clustering," *ICPR* (4), pp. 991-994, 2004.
- [8] J. Piater and J. Crowley, "Multi-modal Tracking of Interacting Targets using Gaussian Approximations," *2nd Int. Workshop on PETS*, pp. 141-147, 2001.
- [9] C. C. Aggarwal, J. Han, J. Wang and P. S. Yu, "A Framework for Clustering Evolving Data Streams," *Proc. 2003 Int. Conf. on Very Large Data Bases*, Germany, Sept. 2003.
- [10] B. Georis, X. Desuamont, D. Demaret, S. Redureau, JF. Delaigle and B. Macq, "IP-Distributed Computer-Aided Video-Surveillance System," *Proceeding of the Intelligent Distributed Surveillance Systems Workshop*, 26<sup>th</sup>, February 2003.
- [11] G. Garcia-Mateos, A. Garcia-Merono and C. Vicente-Chicote, "Time and Date OCR in CCTV Video," *Image Analysis and Processing*, Vol. 3617, pp. 703-710, 2005.

- [12] L. M. Fuentes and S. A. Velastin, "People Tracking in Surveillance Applications," *Proc. 2<sup>nd</sup> IEEE Int. Workshop on PETS*, U.S.A, Dec. 2001.
- [13] R. Jin and G. Agrawal, "Efficient Decision Tree Construction on Streaming Data," *Proc. 9<sup>th</sup> ACM SIGKDD Int. Conf. on KDD*, pp. 571-576, 2003.



**Hongwon Yun**

He received his B.S. and the Ph.D. degrees at the Department of Computer Science from Pusan National University, Korea, in 1986 and 1998, respectively. He is a professor at the Department of Information Technology, Silla University in Korea. His research interests include database, temporal database and data management.