MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Surveillance System with Object-Aware Video Transcoder

Toshihiko Hata, Naoki Kuwahara, Toshiharu Nozawa, Derek Schwenke, Anthony Vetro

TR2005-115 April 2006

Abstract

This paper presents an object-aware video surveillance system that is not only smart and friendly for users, but allows for transmission of the scene over limited bandwidth networks. The proposed system stores high-quality video data and associated object metadata, which includes ROI (Region-of-Interest) information. To satisfy bandwidth constraints and display important parts of the images fast and precisely, the stored images are efficiently transcoded in the compressed-domain based on the ROI information. At the receiver, effective display methods such as a mosaic style allow for an intuitive understanding of the important objects in the scene, as well as their movement over time.

IEEE Multimedia Signal Processing Workshop

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Copyright © Mitsubishi Electric Research Laboratories, Inc., 2006
201 Broadway, Cambridge, Massachusetts 02139

Surveillance System with Object-Aware Video Transcoder

Toshihiko Hata, Naoki Kuwahara, Toshiharu Nozawa
Mitsubishi Electric Corporation
Amagasaki, Hyogo 661-8661, Japan
{Hata.Toshihiko, Kuwahara.Naoki,
Toshiharu.Nozawa@wrc.melco.co.jp}

Derek L. Schwenke, Anthony Vetro
Mitsubishi Electric Research Laboratories
201 Broadway, Cambridge, MA 02139, USA
{schwenke, avetro@merl.com}

***Abstract*—This paper presents an object-aware video surveillance system that is not only smart and friendly for users, but allows for transmission of the scene over limited bandwidth networks. The proposed system stores high-quality video data and associated object metadata, which includes ROI (Region-of-Interest) information. To satisfy bandwidth constraints and display important parts of the images fast and precisely, the stored images are efficiently transcoded in the compressed-domain based on the ROI information. At the receiver, effective display methods such as a mosaic style allow for an intuitive understanding of the important objects in the scene, as well as their movement over time.**

Keywords— Surveillance, Object-Based, Transcoding, Region-of-Interest, JPEG2000, Streaming

I. INTRODUCTION

Recently, digital video recorder systems have come into wide use for efficient storage and transmission of surveillance contents. Many systems are beginning to employ advanced image processing technologies for extracting human faces and behavior based on the knowledge of objects in a scene [1], [2].

Considering the above trends, we have developed an object-aware video surveillance system based on JPEG 2000 that is not only smart and friendly for users, but allows for transmission of the scene over limited bandwidth networks [3]. The proposed system stores high-quality video data and associated object metadata, which includes ROI (Region-of-Interest) information.

To satisfy bandwidth constraints and display important parts of the images fast and precisely, the stored images are efficiently transcoded in the compressed-domain based on the ROI information. At the receiver, ROI and background images are synthesized in a mosaic style besides a usual video replay style. This allows for an intuitive understanding of the important objects in the scene, as well as their movement over time.

The rest of this paper is organized as follows. In the next section, concepts and construction of the object-aware video surveillance system are given. Object-aware video streaming and JPEG2000 ROI transcoding are described in section 3 and 4 respectively. In section 5, experimental results are provided and concluding remarks are given in section 6.

II. OBJECT-AWARE VIDEO SURVEILLANCE SYSTEM

The system has the following concepts.

Smart: Support high level surveillance tasks intelligently with object metadata such as human faces and behavior created by object extraction and tracking from surveillance video.

Visual: Display important objects and their behavior in more detail and intuitively based on the object metadata.

Effective: Utilize limited system resources effectively by transmitting and displaying important parts of the video in detail and quickly based on the object metadata.

Figure 1 shows a basic system construction. The main functions are outlined below.

- *Object extraction and metadata creation:* Extract objects over successive frames and create metadata such as their existing regions and movements. Our system is independent of any particular method for this function. A method based on object tracking algorithm in [6] is used in our prototype described in section V.
- *Video encoding:* Encode surveillance video into a video encoding format such as JPEG2000.
- *Data storage:* Link the video data and metadata together and record them onto hard disk drives. They are always recoded in an endless style and read out according to requests. The stored video has high quality and high frame rate independent of the object metadata since important scenes are required with high quality for detailed analysis and evidence.
- *Scene retrieval:* Retrieve video scenes using the metadata for queries such as human behavior patterns and face similarity.
- *Object-aware transcoding:* Transcode the stored images efficiently in the compressed-domain based on the metadata according to object importance, user preference and available transmission bandwidth. Important regions are given more bit rate than the background. The object-aware transcoder consists of two functions: bit rate control, which assigns the quality for each region, and ROI-based image transcoding, which transcodes an image in the compressed-domain based on the given ROI information.
- *Transmission and display*: Transmit a transcoded video stream with dynamic message exchanges between a client and a server. The received data is decoded and displayed on a screen. The decoded images are synthesized in some display styles such as a mosaic.

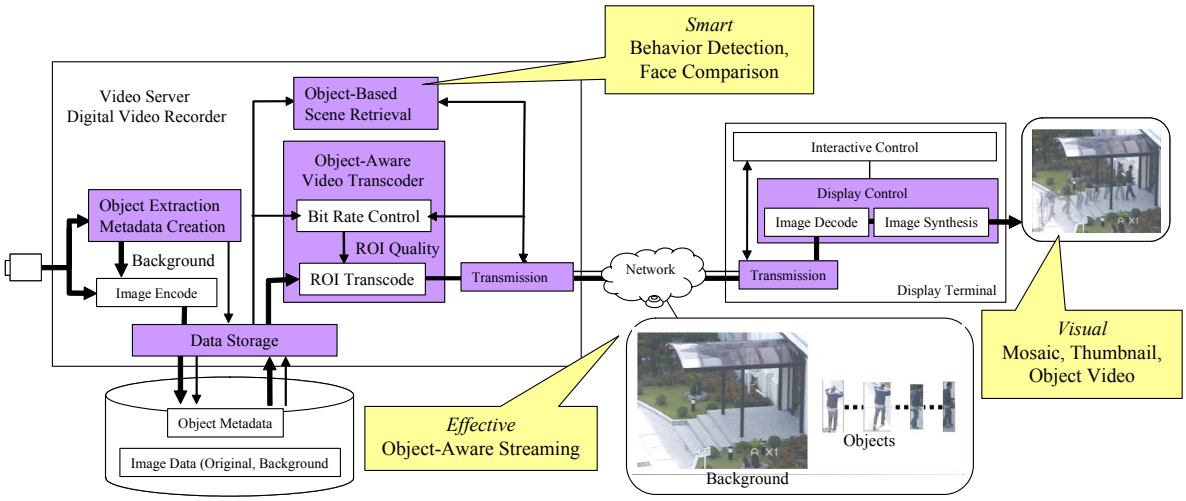


Figure 1. Object-aware video surveillance system

III. Object-Aware Video Streaming

As described above, more bit rate is assigned to ROI's during transcoding for efficient transmission and effective display. The object-aware transcoder controls spatial and temporal quality for each image region. The spatial quality refers to the quality of a still image, which is a function of the quality layers, resolution levels and color components in JPEG2000 The temporal quality refers to the smoothness of an object's movement. The transcoder assigns higher quality and/or higher frame rate to ROI's compared to the background. In the following, the different methods for object-aware video streaming are described; examples are shown in Figure 2.

Frame by frame streaming (FFS)

FFS controls the spatial quality of ROI's and a background in a frame-by-frame manner (see Figure 2b). For each region, a part of the encoded data with higher quality than a specified value is removed. FFS transmits a background every frame, so the spatial quality is lower than that of the other methods described below, but changes in the background can be seen dimly. This method does not require synthesis of the decoded images for display and its implementation is simple.

Successive ROI with occasional background streaming (RBS)

RBS controls the temporal quality of ROI's and a background (see Figure 2c and 2d). First, ROI images and a background image with high spatial quality are transmitted. After that, only ROI images with high spatial quality are transmitted. At the receiver, the background image is decoded and held in a work memory. The successive ROI images are decoded and superimposed on the background. The background is renewed with lower frame rate according to its importance and available network bandwidth. RBS does not transmit a background every frame, therefore the subjective quality is fine in low bit rate. The overall bit rate can be significantly lowered with a low quality background.

Mosaic streaming (MS)

MS controls the temporal quality in the same way as RBS, but it superimposes successive ROI images on a background in a mosaic style (see Figure 2e). Human behavior can be

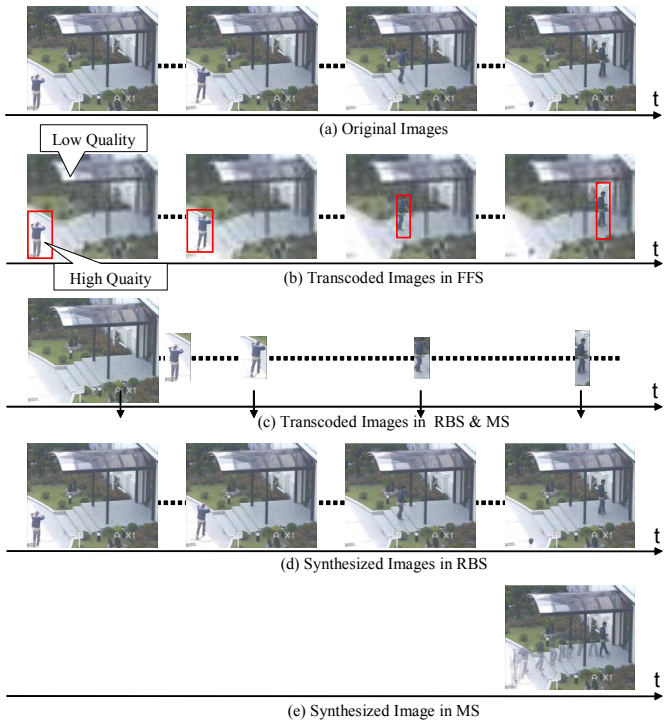


Figure 2. Examples of object-aware video streaming methods

understood immediately and intuitively in a mosaic image, and it is very effective for behavior analysis and scene browsing as well as efficient to transmit the surveillance video over a narrow band network.

IV. JPEG2000 ROI Transcoding

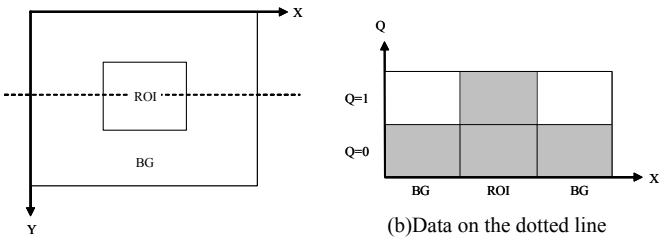
JPEG2000 is known to have high compression efficiency, as well as various scalability and error resilience features [5]. Recently, JPEG 2000 has been adopted for use in surveillance equipment such as network cameras and digital video recorders. In our system, we exploit the built-in scalability of JPEG 2000 for ROI-based quality scalable image transcoding [4]. The key features and algorithm are described below.

***Precinct-based**:* Spatial scalability with fine granularity is necessary in surveillance application, for example a 64x64 pixel unit specifies the ROI in a 640x480 image. It also requires real time transcoding even on very low-cost processors. JPEG2000 tile-based transcoding is very simple and has low computational complexity, but there is quality degradation such as block distortion when a tile size is small. On the other hand, a precinct is a group of neighboring code blocks that are the minimum spatially accessible units in EBCOT (Embedded Block Coding with Optimal Truncation) and it generates one packet in a JPEG2000 code stream. Since images with various progressive types are created by changing order of the packets and deleting some packets without decoding and re-encoding of arithmetic entropy coding and DWT (Discrete Wavelet Transform), precinct-based transcoding has very low complexity.

***Inserting null packets**:* The packets with higher quality than a specified quality layer for each region are deleted and replaced with null packets. Tag-tree decoding and re-encoding not necessary in this case, so the complexity is very low.

An example of the transcoding with Qroi=1 and Qbg=0 is shown in Figure 3. Figure 3b and 3c show data structures on a dotted line of Figure 3a in x-axis and JPEG2000 code stream, respectively. An input image is encoded in LRCP (Layer-Resolution-Component-Position) progression and the boxes in Figure 3 include the packets for each level and each component. Gray boxes are parts of the image that are unchanged and white boxes are parts of the image that are deleted and replaced with null packets during the transcoding.

The transcoding scheme consists of two processes, data analysis and ROI transcoding. The data analysis module extracts indexing information about the structure of the code stream. The ROI is specified by a rectangular bounding box surrounding the extracted object. Since the ROI coordinates may not match the precinct corners, the ROI coordinates are outwardly adjusted to the nearest precinct corner location. The ROI transcoding module performs a dynamic manipulation of the code stream according to the ROI coordinates and quality layer value that are provided as input.



(a) ROI on a XY plane

(b)Data on the dotted line

(c) JPEG2000 code stream on the dotted line

Figure 3. Examples of ROI transcoding with empty packets

V. Experimental Results

We developed a prototype as shown in Figure 1 to evaluate the object-aware video streaming methods. Subjective quality, data size and computational complexity are evaluated with the prototype. A set of outside surveillance image sequences are used for the evaluation; a sample image is shown in Figure 4a. The sequences consist of 1487 frames with 7.5 fps. Each image has resolution 640x480 and color format 4:4:4 and is coded with 4 quality layers and 3 resolution levels with LRCP progression. The rate for each layer is 1.0, 0.5, 0.25, 0.125, respectively. The precinct sizes are set as 64x64, 32x32 and 16x16. The encoded image size is 38KB.

Figure 4c - 4e show examples of transcoded images with Qroi=3 and Qbg=0, 1, 2 respectively. The image with Qbg=0 has very noticeable degradation in the background such as degradation with precinct size. The data size is 9.4KB and 25% of the input image. The quality with Qbg=1is not so good, but better than that of Q=0. The data size is 13.0KB and 34%. In the image with Qbg=2, the background looks a little less sharp than the original and it is hardly noticeable. The data size is 20.9KB and 55%. Observing the transcoded image sequences as moving pictures, visual changes over time in the background with Qbg=0 and 1 are very noticeable.

Figure 5b shows an example of ROI image with Qroi=3 that is used in RBS and MS. The data size is only 6.0KB and 16%. It is very useful for a mobile phone because of its narrowband transmission and low-resolution display. The frame rate of background depends on its importance and available bandwidth. For example, when the frame rate of ROI is 7.5fps and that of background is 1fps, the bit rate is 644Kbps and 29% of the original image sequences.

Figure 4f shows an example of mosaic image in which ten ROI's with 1.5fps are superimposed on the background. The walking trajectory is understood intuitively and immediately.

Table I shows average time per frame of the data analysis, ROI transcoding, decoding and display respectively. They were measured for the above image sequences in a notebook PC with Mobile Pentium 1.6GHz, 1GB memory and Windows XP. It takes only 9.7ms to transcode an image. We found that it takes a half of the time for memory allocation and free in the data analysis, though it is not shown in the table. It should be clear that our transcoding techniques are suitable for implementation on very low-cost processors.

TABLE I. Processing Time

Methods	**Transcoding** ms		**Decoding** ms	**Display** ms
	Data Analysis	*ROI Trans.*		
FF	9.5	0.2	39.5	7.9
BR	9.5	0.1	33.2	7.9

Mobile Pentium 1.6GHz, 1GB memory, WindowsXP

VI. Conclusions

This paper presented an object-aware video surveillance system that transcodes JPEG2000 images stored with high quality efficiently in the compressed-domain based on the ROI information, transmits them over limited bandwidth networks and displays the objects and their movements effectively. The experimental results showed that the proposed streaming methods realize efficient transmission and effective display of surveillance video. It was also shown that the proposed JPEG2000 ROI transcoding techniques are suitable for implementation on very low-cost processors.

References

[1] K. Sato, T. Haga, and T. Nozawa, "A study of human action search system," *IEICE General Conference,* D-11-132, April 2005.

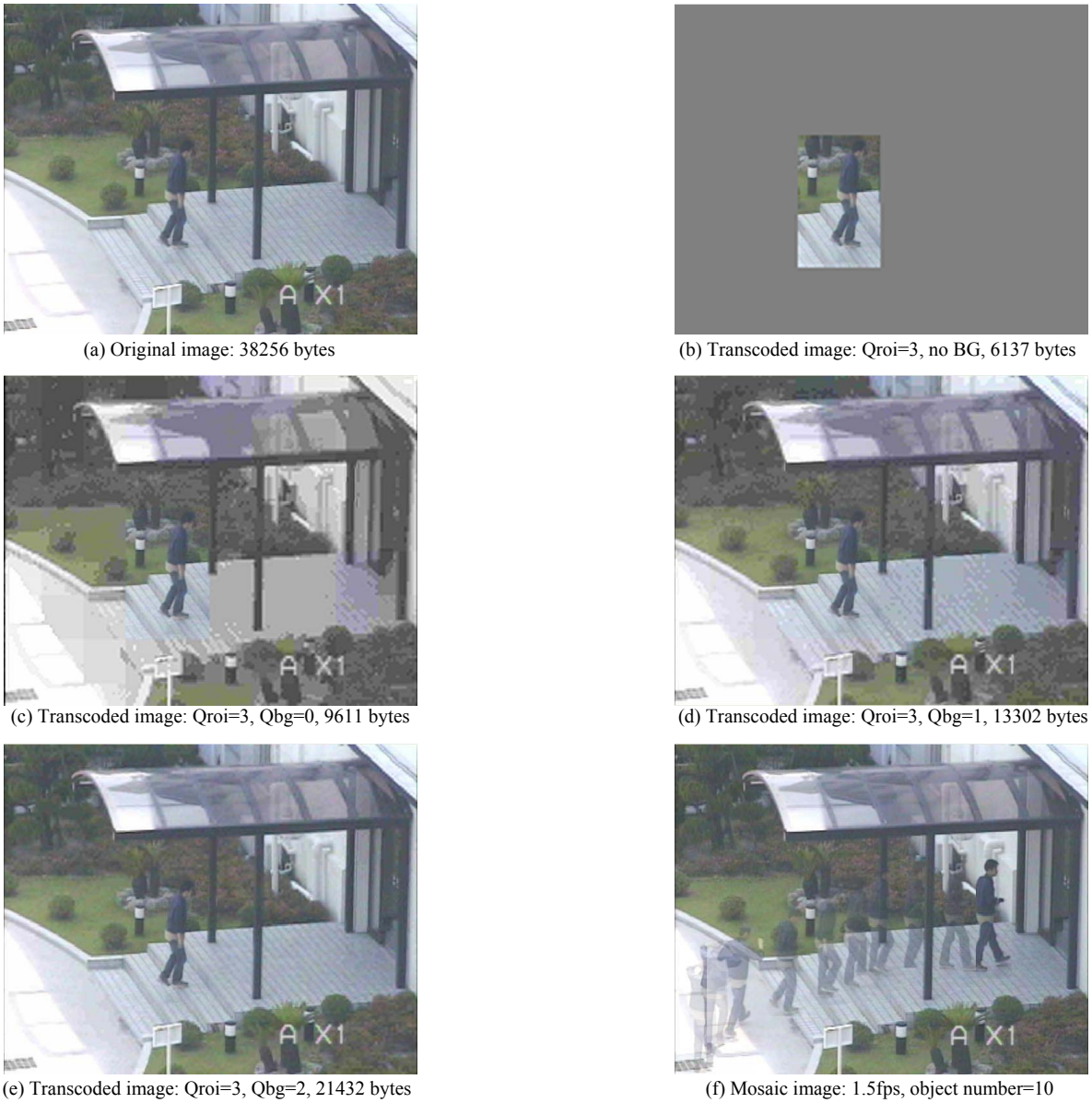
[2] K. Hashima, S. Miwa, H. Kage, M. Hashimoto, M. Jones and J. Thornton, "High speed best appearance facial image recording system," *IEICE General Conference,* D-12-56, April 2004.

[3] T. Hata, N. Kuwahara, T. Nozawa, H-S Kong and A. Vetro, "Object-based video surveillance system -System and JPEG2000 ROI transcoder-," *Technical Report of IEICE,* May 2005.

[4] H-S Kong, A. Vetro, T. Hata, N. Kuwahara, "ROI-based SNR scalable JPEG2000 image transcoding," *Proc. Visual Communications and Image Processing 2005,* Beijing, China, July 2005.

[5] ISO/IEC 15444-1,"Information technology – JPEG2000 image coding system – part1: Core coding system," March 2000.

[6] F. Porikli and O. Tuzel, "Human body tracking by adaptive background models and mean-shift analysis", *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, Graz, Austria, March 2003.



(a) Original image: 38256 bytes

(b) Transcoded image: Qroi=3, no BG, 6137 bytes

(c) Transcoded image: Qroi=3, Qbg=0, 9611 bytes

(d) Transcoded image: Qroi=3, Qbg=1, 13302 bytes

(e) Transcoded image: Qroi=3, Qbg=2, 21432 bytes

(f) Mosaic image: 1.5fps, object number=10

Figure 4. Example images of object-aware video streaming