ON-LINE ADAPTIVE VIDEO SEQUENCE TRANSMISSION BASED ON GENERATION AND TRANSMISSION OF DESCRIPTIONS*

Juan Carlos San Miguel, José M. Martínez

Grupo de Tratamiento de Imágenes
Escuela Politécnica Superior, Universidad Autónoma de Madrid, SPAIN
E-mail: {JuanCarlos.SanMiguel, JoseM.Martinez}@uam.es

ABSTRACT

This paper presents a system to transmit the information from a static surveillance camera in an adaptive way, from low to higher bit-rate, based on the on-line generation of descriptions. The proposed system is based on a server/client model: the server is placed in the surveillance area and the client is placed in a user side. The server analyzes the video sequence to detect the regions of activity (motion analysis) and the corresponding descriptions (mainly MPEG-7 moving regions) are generated together with the textures of moving regions and the associated background image. Depending on the available bandwidth, different levels of transmission are specified, ranging from just sending the descriptions generated to a transmission with all the associated images corresponding to the moving objects and background.

Index Terms – Surveillance, MPEG-7, video analysis, object extraction, adaptive video transmission.

1. INTRODUCTION

Currently, surveillance systems[1] have more demand, specially for outdoor/indoor security in buildings. Traditionally surveillance systems are setup as closed circuit television (CCTV) systems with systems with a large number of cameras and switches. With recent advances, these systems are now becoming networked and issues such as adaptation, scalability and usability (how information needs to be given to the right people at the right time) become very important [2]. In order to obtain more robust surveillance systems, bandwidth reduction consumption will be improved. .

The motivation of this paper is to contribute to the above described scenario with the application of technologies of generation and transmission of descriptions, both for reducing transmission and storage requirements as well as for providing descriptions capable of helping to the automatic interpretation. This paper focuses in the generation of descriptions (following the MPEG-7 standard[3]) in order to provide on-line adaptive transmission of video sequence allowing to send the most relevant information (instead of the original content) and conditions of use (in exchange for greater processing resources both in the sender and receiver terminals, for analysis and synthesis, respectively).

The paper is structured as follows: section 2 gives a brief overview of the state of the art, before sections 3 and 4 describe the proposed system and section 5 explains the generation of descriptions. Then in section 6, experimental results are shown, while section 7 closes the paper with some conclusions.

2. STATE OF THE ART

The work presented in this paper covers mainly three aspects: visual analysis, adaptive transmission and scene reconstruction.

In video-surveillance context, there are two main techniques for motion-based object segmentation: based in the noise introduced by the camera[4]and based in shape detection[5]. In background generation/update, there are a lot of well known techniques based on pixel analysis, like running average, mixture of gaussians, kernel density estimators. A brief review of these techniques can be found in[7]. In object tracking, there are two main approaches: pixels content analysis and model-based. A short review to more detailed information about these techniques can be found in [8].

In adaptive video transmission, we didn't find similar approaches to our proposed system. In [4] and [5] systems for semantic analysis (similar to our proposal) are presented but no adaptive transmission (based in the results of the analysis) is performed.

Many algorithms have been studied during the last years for automatic scene reconstruction in different fields. In [9] a system based on object-extraction and object-coding is described to storage surveillance video data. All extracted data is coded using MPEG-4 tools that reduce the surveillance data available. In [10] a visualization system for remote surveillance is developed using traditional image analysis techniques and fusing dynamic object-extraction to do a dynamic 3D (or 2D) reconstruction. Another system for remote surveillance tasks based on static cameras is presented in[11], where an object-extraction using background-subtraction technique and a reconstruction by fusing these objects into a preset background are done.

* This work is partially supported by Cátedra Infoglobal-UAM para Nuevas Tecnologías de video aplicadas a la seguridad. This work is also supported by the Ministerio de Ciencia y Tecnología of the Spanish Government under project TIN2004-07860 (MEDUSA) and by the Comunidad de Madrid under project P-TIC-0223-0505 (PROMULTIDIS).

3. SYSTEM FUNCTIONALITIES

The main objective of the work presented in this paper is the design and implementation of a system to transmit relevant information of sequences of surveillance cameras at adaptive binary rate based on the generation of descriptions. The work is inspired by the work presented in[4]. The system is based on a server/client model.

The primary objective of the server is the moving object detection and tracking. Then the descriptions of moving objects can be generated at different levels of detail, ranging from the trajectory of the centroid of the moving region to the description of trajectory of the exact object shape, and including bounding boxes trajectory and associated textures, among others. Depending on the level of quality/detail (closely related with the available bandwidth), textures and background can be transmitted either when there are significant changes, periodically, or on demand. The moving object's texture does not imply to overcome privacy issues, as if the application requires it, the texture can be not shown in regular use, being only available on demand in special authorized situations.

At the other side, the client application synthesizes and visualizes the whole video sequence at different levels of detail.

4. SYSTEM DESCRIPTION

Two parameters are configurable in the system (their values having major impact in system performance and results):

- Background Update Time (BUT): this parameter indicates how frequently the background is updated.
- Object Update Time (OUT): this parameter indicates how frequently the motion detection is performed.

The server application is composed by different functional modules (see Fig. 1):

Image Acquisition: performs the acquisition and adaptation of inputs (video or set of images) to make them available for use in the system.

Moving object detection: performs motion based object segmentation, based in the temporal information and the noise introduced by the camera[4].

Background Model: updates the background of the sequence periodically, using a combination of Average and Running Average methods[7]. This module does an analysis of the latest history for each pixel and determines, if the pixel belongs (with a given probability) to the background. If the pixel belongs to background, running average method is applied, if not an average of last background pixels is done.

Object & Feature Extraction: extracts different features from the relevant moving object (such as shape description, bounding box, motion trajectory, dominant colour, motion duration...).

Object Association Module: performs a simple tracking of the relevant objects identified in the previous module. It identifies different objects in the current analyzed frame and tries to associate it, with the objects detected in the previous frames. A similar approach can be found in [12].

MPEG-7Description Generation: generates the "master" description of all the information obtained from the previous modules. MPEG-7 is a good tool to generate descriptions of generic video objects for scene reconstruction (as we can see in [13]).

The client application is composed by three different functional modules (see Fig. 2):

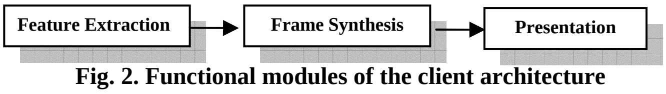

Fig. 2. Functional modules of the client architecture

Feature Adquisition: reads the MPEG-7 description received and extracts the features needed to synthesize the sequence.

Frames synthesis: synthesizes the sequence frame by frame using the information in the description, as well as the objects's textures and the background image, if available.

Presentation: displays the synthesized sequence

5. DESCRIPTIONS' LEVELS OF DETAIL

From the master description the system generates different levels of detail, but in a real service implementation, this can be customized in order to avoid unnecessary analysis and adaptation, reducing the computing cost and resources requirements, or increasing performance for the same resources.

One parameter of the level of detail is the granularity of the description, that is, if the moving regions shape and trajectory are more or less detailed. With respect to regions shape, it can range from a detailed shape description (result of pixel accuracy segmentation) to the centroid of the object. With respect to the motion trajectory, it can range from the trajectory of each pixel in the region's shape to a set of sampled points of the trajectory, including in the range associated interpolation functions. The second parameter deals with the textures associated to each moving object in the description and the background image.

Based in these two parameters, we have currently implemented four levels:

- Level 1. For each background analysis slot (BUT),

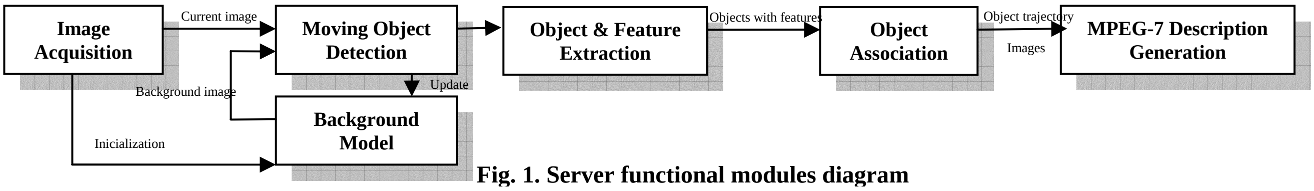

Fig. 1. Server functional modules diagram

the description includes an image representing the background, objects images associated to the object shapes (one each OUT) and the trajectory for each object (considered as rigid objects).

- *Level 2*. In this level, the difference with level 1 is that object images are reduced to one for the complete background analysis slot. The most representative of the set of available images is selected as object key-image.
- *Level 3*.In this level, images are not included in the description, which consists only of moving regions descriptions. The feature that is used to synthesize the object texture at the client is the dominant colour. The background image is updated each BUT.
- *Level 4*. In this level the updated background image is requested on demand.

5.1. MPEG-7 description tools

The proposed system generates one master MPEG-7 compliant description file for each group of pictures analyzed in a BUT. The MPEG-7 description tools currently incorporated are:

- *StillRegion DS*. It's used to describe the location of the pictures, format, coding,...
- *MotionTrajectoryDS*. It's used to describe the trajectory for each object. The trajectory is calculated using a "bounding box" that fits it.
- *DominantColor DS*. With this description, we have information about the tracked object (important in synthesizing level 3 and 4) without the requirement of transmitting the whole associated texture.

6. EXPERIMENTAL RESULTS

In this section, experimental results of the proposed system are presented, illustrating the subjective effect of object (region) level analysis at each different reconstruction level as well as the size of the associated description (data to be transmitted and used for reconstruction). Results are evaluated in terms of original data size reduction due to visual analysis (that implies a transmission bandwidth reduction) and client-side quality reconstruction (where description's levels of detail provide an adaptation to bandwidth requirements). Due to space constraints and it order to provide a well-known sequence based example, sample results are shown for the MPEG-4 test sequence Hall Monitor (CIF resolution at 30 fps). Additional results can be found at http://www-gti.ii.uam.es/publications/OnLineAdaptiveVideoSequenceTransmissionBasedOnGenerationAndTransmissionOfDescriptions/. Results shown in Figures 3 to 5, correspond to analysis with 1 second for background update (BUT) and 0.2 seconds for objects analysis update (OUT).

Fig. 3 shows an example for motion-detection stage. In the next analysis stage, object filtering (e.g., small objects, artifacts with small appearance times,…) and tracking is done. For each object, the system (currently) generates a bounding box that fits it (Fig. 3.d). These bounding boxes are tracked during each BUT (each OUT the object trajectory is updated) and the system generates the information corresponding to each level.

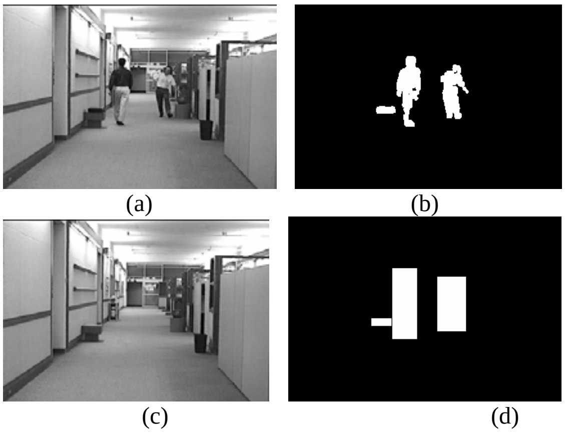
(a) (b)
(c) (d)

Fig. 3. Example of the segmentation module output for frame 160. (a) original frame, (b) motion detection, (c) background frame and associated bounding boxes (d)

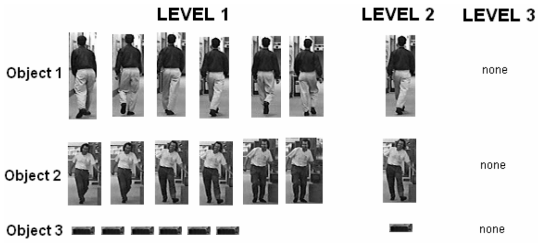

Fig. 4. Example of objects images generated for the frames 150-180 (one BUT analysis slot).

To describe the object and motion information in the subsequence (each BUT), descriptions are automatically generated. The motion information is described by means of the motion of each vertex of the bounding box, as the bounding box size can change each OUT (this a very rough approximation to a non-rigid object). Finally, object texture images are incorporated to the description depending on the level of description desired. Results for the three first levels are shown in Fig. 4.

Data size of the descriptions and images generated is shown in Table 1. A comparison for data sizes with different compression techniques is shown in Table 2. In the proposed system, all images (for background and objects) are transmitted in JPEG format. In these two tables, we can see that the original data size is highly reduced (it should be noted that no temporal redundancy is taken into consideration) and the correct selection of the parameters allows to modify the size of data generated. Data for level 4 is not listed, as it depends on user interaction. The obtained results show that the performance of compression techniques at low bitrates is under the proposed system results. Additionally, the motion and tracking analysis stage could provide data to other surveillance functionalities at the server side (e.g., alarms, detection of intrusion in predefined zones).

PARAMETERS		LEVEL 1 Size	LEVEL 2 Size	LEVEL 3 Size
BUT (s)	OUT (s)			
1	0.2	980KB	580KB	470KB
2	0.2	792KB	388KB	270KB
3	0.2	712KB	298KB	190KB
1	0.4	696KB	544KB	410KB
1	0.8	544KB	486KB	390KB

Table 1 – Example of data size (images + description) generated for each analysis level (entire sequence analyzed)

Original Data compression	Average Transmission rate (Higher-Lower in kbps)	Total Data Size (KB)
No compression	50600	75900
MJPEG	4876-1283	7314-1924
MPEG-2	3300-320	4900-480
MPEG-4	2860-312	4290-450
Proposed system	650-350	980-390

Table 2 –Transmission rates comparison for Hall Monitor sequence (without overhead transmission packets).

At the receiving side, the client application synthesizes a reconstruction of the original sequence. Depending on the selected level, different quality levels are obtained (see Fig. 5). It can be observed that between the original and level 1 reconstruction have a good quality and levels 2 and 3 have a progressive loss of quality (due to the bounding box approach). Due to space constraints, time processing issues are not discussed here in detail. In the tests done, real-time operation was reached, obtaining good performance results showing that the critical stage in server is motion detection).

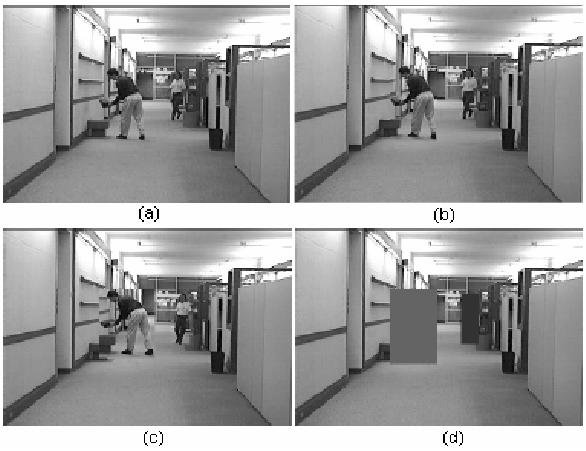

Fig. 5. Example of the three levels of reconstruction at the client. (a) original frame and reconstructed frame for level 1 (b), level 2 (c) and level 3 (d)

7. CONCLUSIONS

In this paper a proposal for enabling adaptive transmission of relevant information from video-surveillance sequences have been presented. The system is based on traditional image analysis techniques and reduces highly the amount of original data. Additionally it provides usually a good object feature extraction for the surveillance application domain.

Transmission rate can be customized to the particular application requirements (generating descriptions with different levels of detail), always focusing on what is happening in the scene at a semantic level. A reduction of original data size (without incorporating image compression) shows that the objective of reducing the binary rate has been reached, as well the provision of adaptive reduction based on different levels of detail (providing a good quality reconstruction for level 1).

8. REFERENCES

[1] K.N. Plataniotis, C.S. Regazzoni (eds.), “Special Issue in Visual-centric Surveillance Networks and Services”, IEEE Signal Processing Magazine, 22(2), March 2005

[2] D. Doermann, A. Karunanidhi, N. Parkeh, M.A. Khan, S. Chen, H.T. Ozdemir, M. Miwa and K.C. Lee, ”Issues in the transmission, analysis, storage and retrieval of surveillance video”, IEEE ICME, 2003.

[3] ”Introduction to MPEG-7: Multimedia Content Description Language”, B.S. Manjunath, P. Salembier, T. Sikora (eds.), John Wiley & Sons, Ltd., 2002.

[4] A. Cavallaro, O. Steiger, T. Ebrahimi, “Semantic Video Analysis for Adaptive Content Delivery and Automatic Description”, IEEE Transactions of Circuits and Systems for Video Technology, 2005.

[5] M. Bramberger, J. Brunner, B. Rinner, H. Schwabach: "Real-Time Video Analysis on an Embedded Smart Camera for Traffic Surveillance". IEEE Real-Time and Embedded Technology and App. Symposium 2004

[6] Tadashi Nakanishi and Kenichiro Ishiim “Automatic vehicle image extraction based on spatio-temporal image analysis”, NTT HIL 2003 Japan

[7] M. Piccardi. “Background subtraction techniques: a review”, Proc. IEEE Conference on Systems, Man and Cybernetics, 2004.

[8] Q. Zang and R. Klette. “Object classification and tracking in video surveillance”, Proc. Computer Analysis of Images and Patterns, , 2003.

[9] A. Vetro, T. Haga,, K. Sumi, “Object-based coding for long-term archive of surveillance video”, Proc. Int. Conference on Multimedia, 2003

[10] A. Calbi, C.S. Regazzoni, L. Marcenaro, “Dynamic Scene Reconstruction for Efficient Remote Surveillance”, IEEE Proc. Int. Conference on Video and Signal Based Surveillance, 2006.

[11] X. Yao, L. Xu, “Object-based Video Reconstruction for Route-oriented Intercity Highway Tasks with Static Cameras”, IPCV 2006.

[12] V. Rehrmann, M. Rothhaar, “Detection and Tracking of Moving Objects in Color Outdoor Scenes”, Proc. Int. Symposium on Automotive Tech. and Automation, 1997

[13] 0. Steiger, A. Cavallaro, T. Ebrahimi. "MPEG-7 Description of Generic Video Objects for Scene Reconstruction", Proc. SPIE Electronic Imaging, 2002