AUTOMATIC ASSINCHRONOUS VIDEO DEMULTIPLEXING

Marcelo Hor-Meyll Alvares
Brazilian Federal Police
João Pessoa, PB, Brazil
hormeyll.mfmha@dpf.gov.br

Rafael Dueire Lins
Universidade Federal de Pernambuco
Recife, PE, Brazil
rdl@ufpe.br

ABSTRACT

The work on real forensics videos involving banking images lead to the method presented here: a software approach to the problem of demultiplexing time multiplexed security videos. This approach is of particular interest in cases in which, for some reason, the regular demultiplexing hardware/software is not available. Based on comparisons of the mean frame luminance and having low frequency characteristics of the scenes as references, this approach is being used with good results by video experts in Brazil to help locating sequences of particular interest in surveillance time multiplexed videos recorded in magnetic video cassettes.

Keywords — demultiplexing, video, forensics.

1. INTRODUCTION

Security surveillance systems are everywhere as a result of the feeling of insecurity and mistrust which grows in modern society, reinforced by the cost drop of equipments and systems as a result of mass production. Normally designed to develop remote surveillance of wide areas, the video provided by security systems are commonly presented to the police forces in the case of incidents. Many surveillance systems employ low cost, low resolution sensors that make use of distorting spherical lenses to survey wide areas, combined with multiplexing equipments as a way to save storage space. The lack of standardization in the multiplexing techniques makes the problem even worse once there are many proprietary incompatible techniques that make the demultiplexing process sometimes very complicated.

As a result of all the aforementioned problems, a growing number of videotapes containing inadequate images is presented to forensic experts to work on.

Conceived to demultiplex a surveillance system videotape which recorded a robbery at a bank agency, the demultiplexing solution presented here, is being tested in several videos yielding satisfactory results.

2. VIDEO CHARACTERISTICS

The video used as the starting point for this work was a gray scale interlaced video with images of a bank agency. The frames were recorded on a magnetic VHS videotape by a time multiplexed CFTV system with an unknown number of cameras. Witnesses affirmed that a certain crime was committed at the agency in a location covered by the cameras. The video was recorded in ELP mode (8hrs) in that tape, but in an unknown position. The video had no multiplexing information in the retrace lines and the demultiplexing hardware was unavailable. An initial frame-by-frame analysis revealed that the multiplexation did not follow a regular order but, somehow, it tried to sample the 8 cameras of the system uniformly as depicted in scheme shown in Figure 1.

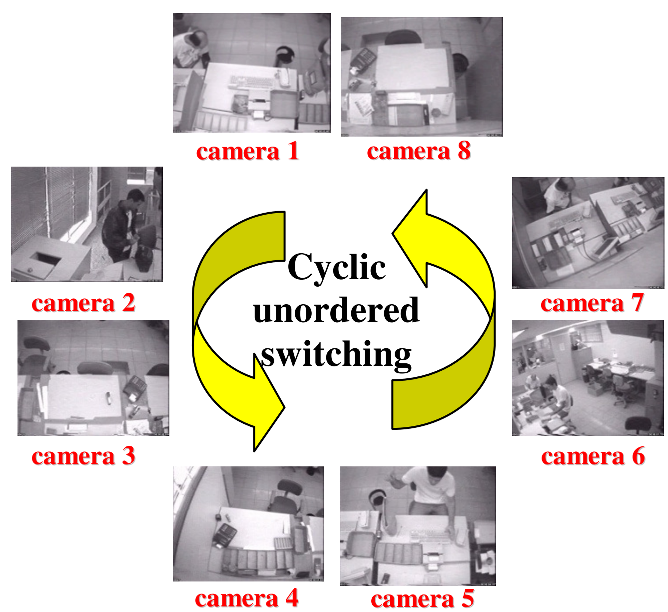

Figure 1 - Multiplexation scheme of the analyzed video

Once there was no way to demultiplex the video using the specific hardware, a software approach was the solution adopted to the problem.

3. PRE-PROCESSING THE VIDEO

As already mentioned, the video under analysis was in VHS thus, it had to be digitized before being enhanced and processed. This section focuses in the digitalization process, frame deinterlacing and finally noise removal and frame identification process.

3.1. Capturing video

Beginning by the digitalization of the video recorded in the videotape, it was necessary to pre-process the frames prior to trying the classification algorithm. As long as the analysis of video by the forensics experts aims, in general, the identification of suspects, image quality preservation is a priority. Thus, to find the adequate resolution to capture de video, the following data were analyzed: Once the video was an NTSC interlaced analog one, disregarding the 30% resolution reduction caused by the Kell effect, one has to sum up the 242.5 vertical active lines of the odd fields with the 240.5 [1] one from the even field. This way, the vertical digital resolution (system based) should be at least 483 lines. To determine the minimum horizontal resolution (source based), one should first consider the active line interval of the NTSC system 52,6μs [2] (figure 2)

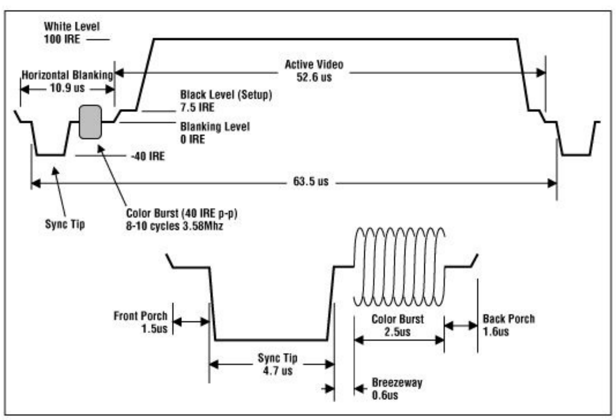

Figure 2 – Video line signal timings (ITU-R BT.470-7)

Based on this interval, it is possible to calculate the number of oscillations of a 4.2MHz video signal line:

$$horizontal \quad resolution = \frac{4.2 \cdot 10^{6} \, \frac{oscillatio\ ns}{s}}{\frac{1}{52.6 \cdot 10^{-6} \, \frac{s}{line}}} = 220.92 \, \frac{oscillatio\ ns}{line}$$

Using Shannon's sampling theory, one obtains 442 samples/line for a lossless horizontal representation. This way, a 442x485 pixels resolution should be enough not to lose any analog detail if it was not for the aspect ratio. A 442x483 pixels resolution would do a 0,91 aspect ratio, once only 640x480 (4/3 or 1.33), 720x480 (16/9 or 1.50) or more rarely 720x486 (1,48) are commercially available. A 16x9 aspect ratio capture card would add two 40x480 rows making a no information frame to the acquired video. So, as concluded by [3], this work also concluded that a 640x480 pixels frame grabber would be the best choice and do the job suitably for forensics applications.

3.2. Deinterlation

Once the time multiplexation is field based, it is common to see in security videos, hybrid frames that would introduce noise in the classification process. Thus, prior to starting any classification, deinterlacing the video is a must. Normally the deinterlation process is associated with discarding of one of the fields or merging them together. In forensic applications, once the frames will be analyzed statically, the apparent division by two of the vertical resolution caused by the frames unfolding process will bring no harm to the problem addressed herein. This way, the best option is to perform a simple field unfolding, not losing the image sharpness or resolution and separating both fields.

3.3. Noise removal and frame identification

Due to tracking misalignment of the recording heads of the VCRs used, one often finds, not only for the settling time, but sometimes for all the video, static noise and visible retrace lines in the active video interval. These problems can be introduced by various sources, but more frequently by some mechanical malfunction or magnetic tape defect. These problems are common because security system's VCRs are normally switched on 24/24hs, using continuously reutilized tapes. The results are dirty VCR heads and low quality tapes causing loss of track and synchronization problems.

These contaminated frames can introduce noise in the classification algorithm, sometimes taking it out of a recovering region. That is the reason why they should be removed in a preceding phase called here as pre-processing. Using Avisynth [4] combined with Virtualdub [5] both open source freeware software, the frames are cropped and static contaminated frames are manually analyzed and eventually discarded. Original frames sequence numbers and time-stamps are added to the frames in some quiet region of the video to allow eventual videos appending. One last step in the pre-processing phase is performed: the video codification using Indeo 5.1 codec to fulfill the input video requirements to be processed by MATLAB.

4. PROCESSING THE VIDEO

4.1. Platform and frames behavior

As already mentioned, MATLAB was chosen to process the images. The main reason for this choice was that all the demultiplexing solution was developed during a real case in which there were arrested people and time was rushing out. MATLAB offers some really interesting functions that speeded up the process.

The initial frame-by-frame analysis showed some characteristics of the frames behavior as can seen in the following graph (Figure 3):

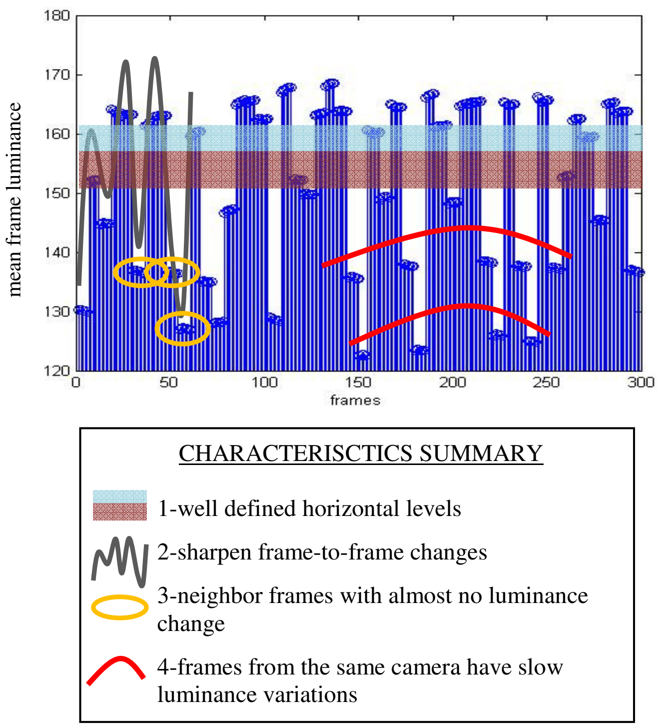

Figure 3 – Graph of the mean luminance of a 300 frames time multiplexed video test file

Based on these characteristics, a MATLAB script was developed to perform the time demultiplexation, transforming a single multi-camera video file into N single-camera video files, where N represents the total number of cameras in the system.

4.2. Classification and frame comparison

The comparison variable chosen was the normalized mean luminance of the frame. The main idea is to group the frames as long as they show little difference in luminance, actually smaller than a pre-defined threshold empirically found. When the mean luminance differences increases beyond a limit, the algorithm assumes that that frame belongs to a different camera. The process works automatically creating a "new camera", that in fact is a structure with various information, when the comparison between the current frame and all the "existing cameras" is greater than the threshold. For comparison means, each camera is represented by a reference frame, obtained by a low-pass filtering of a limited window containing the last frames classified as belonging to that camera. One example of this filtering is shown in Figure 4.

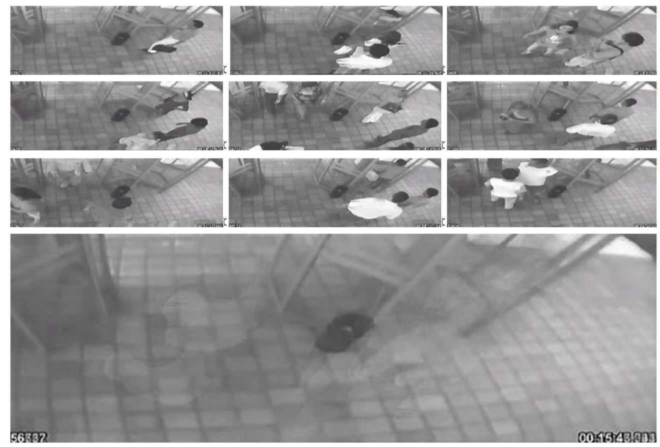

Figure 4 – Example of some camera frames (up) and the respective camera reference frame

As can be seen above, the frames show a bank entrance revolving door. The noise caused by the constant flux of people wearing clothes of different colors is a typical example of how this filtering technique can be valuable. The filter time constant has to be slow enough to reject the unwanted noise but not too slow not to be able, for example, to adapt to the dimming daylight.

Besides the filtering one, other techniques showed to be useful whenever processing multiplexed videos. In the current version, for example, a camera creation inertia, is employed to help in the camera creation process. The idea was to introduce a progressive difficulty in the creation of new cameras: easier in the beginning of the video and making it more difficult as the video continues.

5. RESULTS

Some real cases have already been solved using the presented technique. Despite MATLAB 7 limitation in processing video files not bigger than 2GB, TRUECOLOR videos and Y8 grayscale ones have been tested with good results. Table 1 summarizes the tests performed.

Some classification problems were identified in the videos processed, normally they are associated to the loss of synchronization (Figure 5), the static contamination comprising considerable portion of the frame (Figure 6) or extremely low illumination in the scene (Figures 7 and 8).

Table 1 – Tests performed data	
CPU used	
Processor	Intel Core 2 Duo 6300, *clock* 1.86GHz
MB	Intel DG965SS – *chipset* Intel G965
RAM	Kingston KVR667D2N5/1G – 1GB
HD	Samsung HD160JJ – 160GB/7200RPM/8MB *cache*
Processing data	
Mean pre-processing frame rate (deinterlacing, cropping, insertion of frame number and time-stamp, and Indeo 5.1 re-encoding) – Avisynth and Virtualdub	10-11fps
Cameras in the video	7-15
Mean processing frame rate - MATLAB	3-6fps
Classification efficiency	98-100%
Number of frames processed	12,500-57,000

Figure 5 – Loss of synchronization

Figure 6 – Static contamination

Figures 7 and 8 – Similar low illumination frames classified as belonging to the same camera

Contrariwise to what was concluded by the Cognitech team in references [6] and [7], the experiments performed in this work found that the mean luminance can represent quite well the frame content, as long as other techniques are associated with the luminance information. The solution developed was used in several other forensic videos under analysis at the Federal Police of Brazil (Polícia Federal do Brasil). If this paper were accepted, the executable code for the solution described herein may be freely obtained by contacting the first author of this paper, provided it is for non-commercial use. Besides that, the solution described may also be demonstrated at ICIP 2009.

6. REFERENCES

[1] GROB, BERNARD, Basic Television, EUA: New York, 1964;

[2] ITU-R BT 470-7, Conventional analog television systems;

[3] RUDIN, LEONID I. *et al*, Sampling theory for digital video acquisition: the guide for the perplexed user;

[4] AVREY, LEE, AviSynth available at http://www.avisynth.org ;

[5] AVREY, LEE, VirtualDub, available at http://www.virtualdub.org;

[6] RUDIN, LEONID I. *et al*, Meaningful Automatic Video Demultiplexing with Unknown Number of Cameras, Contrast Changes, and Motion, EUA: Cognitech Inc., Pasadena, CA, Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference, Publication Date: 15-16 Sept. 2005;

[7] RUDIN, LEONID I. *et al*, Software-based universal demultiplexing: threshold-free energy minimization approach, SPIE (The International Society for Optical Engineering). Proc. SPIE Vol. 4232, p. 513-520, Enabling Technologies for Law Enforcement and Security, Publication Date: 2/2001;

[8] GOODRIDGE, STEVEN G., Content-based software demultiplexing of surveillance video, Proc. SPIE Vol. 4232, p. 489-496, Enabling Technologies for Law Enforcement and Security, Simon K. Bramble; Edward M. Carapezza; Leonid I. Rudin; Eds.

[9] COLMENAREZ, ANTONIO J.*et al*, *Video* monitoring and surveillance systems capable of handling asynchronously multiplexed *video*, United States Patent 6,999,613, February 14, 2006, Maracaibo, Venezuela;