# Adapted method of slideshow processing

Tarek Ouni[1], Ismail Ktata[2] and Mohamed Abid[3]
National Engineering School of Sfax
University of Sfax, Tunisia
[1] email: tarek.ouni@gmail.com
[2] email: ktata.ismail@gmail.com
[3] email: mohamed.abid@enis.rnu.tn

***Abstract*— This paper proposes a new image-based technique adapted for the processing of the slide stream in a slideshow application. The approach takes advantage of the observable events related to slides that are visible during the conference and it takes care of specific characteristics of slide stream. In particular, the suggested approach rests on three basic points: image change detection, differential frame prediction and local adaptive coding. The proposed technique presents several advantages; it deletes immediately the similar slides, detects slide changes and provides an selective coding in order to ensure a high image quality and pertinent text legibility. These factors are the key issues for the realization of efficient adapted slideshow coding method.**

***Keywords*— Image processing, adapted coding, text detection, slide change detection.**

## I. Introduction

Recent research projects aim at broadcasting meeting, conferences... and recording them in suitable forms for later retrieval. Along such events, verbose information sources coexist: video, sound and projected slides. Retrieving, at the backend, the latter source with good quality (readability and fidelity of the received slide to the original ones) and with reasonable cost (complexity and bandwidth) is still a challenge. The professionals progressively integrated various techniques improving quality of their presentations. Remains that there does not exist yet adapted solutions perfectly adapted to broadcast certain visual contents such as slides. This solution could be a specific technique dedicated to ensure a good quality of slides. The processing and the broadcasting of the slides are possible today by the integration of various video and image processing techniques. Such techniques present several disadvantages such as image quality, legibility of the texts, bit rate, and operational complexity. The development of an economic and adapted solution is thus required. It makes it possible to simplify and thus to offer an adapted answer to problems still imperfectly studied.

In this paper, we take interest in the slide stream processing in slideshow applications.

The rest of this paper is organized as follows. In Section 2, a brief review is given for the related work on the projected documents capturing and processing. Section 3 presents a general overview about slides features on which the later work will be based. In Section 4, our slide change detection, text and edge localization and extraction and adaptive encoding are proposed. Section 5 reports some preliminary experimental results assessing the effectiveness of our approach, while Section 6 draws some conclusions.

## II. Multimedia presentation

In the slideshow applications, slide stream can be considered either as a succession of fixed images or as a video sequence [1] [2]. This generally depends on the information source and the acquisition method [3]. In fact, slides can be retrieved from a camera, from the lecturer's laptop or from the electronic file. Furthermore, the corresponding captured document images have some particular features related to the application: they hold different formats of information (text, graphics, photo, lines, textures...) and show some animations. Hence, they should have a suitable processing technique.

### A. Capture environment

For broadcasting video documents, videoconference devices are generally equipped with "composite" or S-Video inputs allowing connection of reading equipment such as VCRs. These video inputs can also be used for connecting a title bench consisting of a video camera attached vertically on a pedestal. This equipment makes it possible to present small objects or documents "papers" [4]. Currently, analog inputs, such as VGA link, are available to connect computers and videoconference equipments. They allow the broadcast of screen images with the benefit of lesser noise, in comparison with capturing video from the camera output. Some other systems diffuse the slide stream from video card memory via Ethernet support. This always ensures relevant image quality as it avoids digital-analogue conversions. Thus, the last scheme will be adopted in order to guarantee best experiment conditions and to prove the further proposed technique reliability.

### B. Image or video

For encoding video over IP, the trend in industrial field is the MPEG2 (soon MPEG4). The MPEG2 over IP provides a video quality comparable to TV quality for a 5 Mbit/s flow [5]. Actually, the implementation of such a technique is too complex and difficult to achieve (at least for nowadays). On the one hand MPEG2 is too cumbersome: one hour takes about one Giga Byte of storage capacity and for only fifty slides. Hence, the video documents, treated here as a continuous video stream, occupy an enormous bandwidth and may suffer from a remarkable deterioration in image quality in

low bit rate due to the compression. The textual documents are always unreadable and the quality of the restitution depends on the displaying properties at the backend. Additionally with a TV or video monitor, the restitution can be affected by the transcoding (XGA to H261/H263). For an ordinary video flow, we can reduce the bandwidth to the detriment of the image quality (reducing the resolution, high compression rate), but video documents as slideshow should be restored with a quality similar to the original ones. On the other hand, the used video encoder's algorithms are too complex compared with the video slide's proprieties and events complexity as the slides stream is often static. Otherwise, other software solutions are investigated in order to reduce the quantity of information in slide's flow, in the manner of ARTS (Aristote Real Time Slides) tool whose principle is explained below [6]:

- Each n seconds, a screen capture is made. The n is a generic parameter and is fixed usually from 5 to 7 seconds. If the slide N is different from the slide N-1, the former is saved in hard drive in JPEG format,

- The webserver, included in the videoconference system, refresh the page containing the old slide by substituting it by the new one,

- The way to refresh the page depends on time. Actually, a script on the server polls if a new slide is detected each s seconds,

- The name and the time of the new detected slide are then saved in a special file with OTESA format.

Such solution treats slides as a discrete succession of fixed images. This method decreases the complexity of treatment and the bandwidth considerably. The drawbacks of this method are the loss of slide animation.
Hence, the efficiency of the slides' treatment in existent slideshow systems seems to be limited. In fact, the used techniques are not based on a pertinent analysis of the nature and features of the informations to be processed, but it was a problem of finding a fast solution based on existing image and video processing techniques. A good way to solve such problems is to provide an adapted solution for slide stream coding. This needs a review of this context based on an objective analysis of the slideshow characteristics.

## III. Slideshow features

In order to perform a maximum compression level, without degrading the image quality, it is necessary to resort to a non-standard image processing. This is logical since the slide flow is neither continual video stream nor a succession of static images, but rather a heterogeneous data composed of static images with animations (Figure 1). The aim of the technique is to provide a solution for two different features: a fixed image and an animation, respecting real time constraints. This leads to a good compromise between fast treatment, operational complexity and quality.

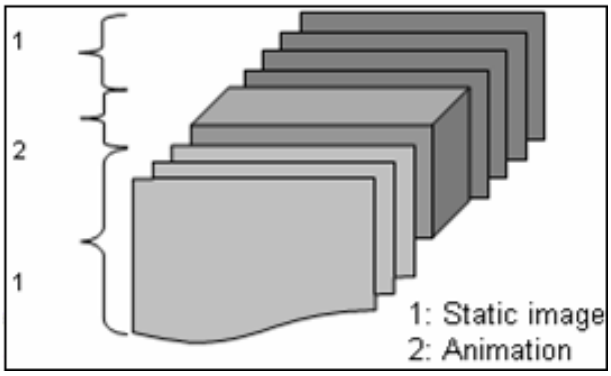


Fig. 1. Slides stream.

## IV. Proposed method

While studying video slide specificities and requirements, and existing algorithmic resolution's modes, it seems that certain existing image processing standards give solution to slides stream processing issues. Thus, we suggest an adapted technique of slides coding that inherits advantages of existing image and video processing techniques. First, it exploits intra frame coding technique advantages (JPEG, PNG). Second, it exploits the temporal redundancy as all inter frame coding techniques (MPEG) providing more significant bit rate reduction without need for complex motion estimation based techniques. Third, the method exploits characteristic of different image processing techniques in order to ensure best slide quality. Hence, the proposed technique presents a complete solution dedicated for slideshow processing; it detects slide transition, discriminates different slide features (static image or animation portion), and makes adaptive local region coding.

### *A. General description*

The proposed technique is based on the continuous comparison between slide video frames in order to detect the flow stationarity, eliminate similar frames and save only the occurred changes. Thus, the new image, as well as the time of its detection, could be saved leading to a pertinent indexing for archiving with the possibility to deduce whether the precedent portion is a fixed image or an animation. This point is important, since it solves a problem that has never been resolved yet: discriminate, whatever the context is, a fixed image from an animation.
At first, this specification is built to a pixel to pixel comparison with a threshold. Such an algorithm has the advantage of being simple to be implemented, but omit presentation features: a constant background (layout model) with a changing portion (generally text changing). Thus, images modifications are usually local. In addition, in case of the methods using a threshold of the screen modifications, the text change in a screen could not be detected with high one. In the other case of a relatively low threshold, the noise can affect the method reliability and a noisy frame could be detected as a different one.

There are many existing techniques related to slide change detection[7][8][9]. Histogram slide change detection is based on the comparison of the color histogram of successive frames. When an important change in color histogram is detected, slide change is signaled. This method performs well for successive slides having different background colors. However, in real slide presentations, most slides have the same background, and often only text layout changes. Thus, the histogram technique is not adapted to detect such changes.
Other techniques consist in representing the binary frames and comparing their difference with a threshold. This method works perfectly with slideshow having high contrast. However it is difficult to set a unique threshold value for various slideshow. Consequently, the global image treatment becomes irrelevant and it is more recommended to proceed with a local treatment based on local region comparison. This will permit to restrict the slides difference or the animated portions to some blocks. These blocks will be coded with the appropriate coding technique according to their content properties. Though the first main functionality of the process is detecting image changes, we can not negotiate on the quality of the latter. Then, it is convenient to look for a technique which treats mainly image's comparison issues.

### *B. Detailed description*

Hereafter is a detailed logical process of the method:

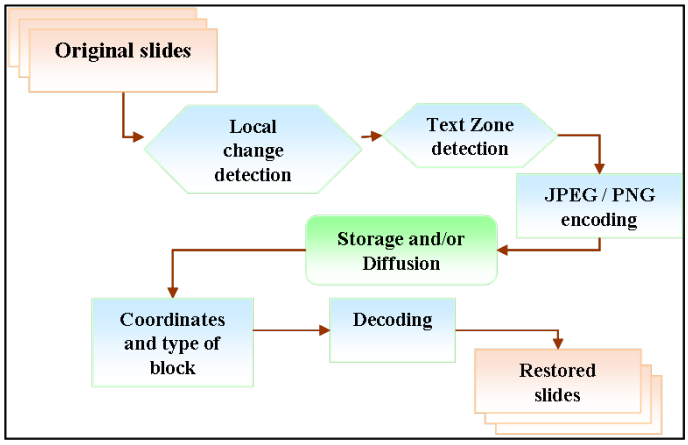


Fig. 2. Flowchart of the proposed method.

The first step consists in acquiring the first frame (reference frame), which should be saved in totality, compressed and sent to the destination. Then, the process polls the slides change. This is achieved by successively comparing the slides until a change is detected. The comparison is not applied on the successive frames, but rather between the current frame and the reference frame for the following reasons:

- The difference between successive frames is usually not detectable, especially in case of small change or low animation.

- Minimize access memory process as there is no need to update reference frame continually when no slide change is detected.

- Minimize the error accumulation (due to subsequent quantifications) after the decoding process.

The aim of such comparison is the elimination of the similar frames and detection of slide change. The changes are represented through a matrix containing differences between the old and the current frame pixel's values. This process will be later called differential frame prediction referencing to similar process in the prior works.

1- Differential frame prediction
The first step consists in calculating difference matrix whose values will take the difference between current frame pixel values and their corresponding ones in the reference frame. The resultant matrix will contain many null values which could increase the compression rate. These elements are afterward quantified according to a threshold previously chosen. The threshold's value is empiric; it was fixed following certain experiments whose aim is to guarantee visual image quality.

2- Block splitting
It consists in splitting the difference matrix into blocks in order to localize changed regions and allows to simplify further processing. The block splitting could be either dynamic or static. Dynamic means that blocks size is variable according to the modifications. However, static means that blocks size is fixed preliminary. Further, each block is processed separately.

3- Filtering (noise eliminating)
In our context, we process on the so called differential frames which have simple features without complex texture or motif as they are generally eliminated by substraction operation when making differential matrix. However, most of existing filtering methods are designed to treat original frames, having more complex features, and so are not quite efficient. The we have to find adequate filter to use. We have tested different filters on many slideshow samples captured with webcam. Those experiments proved the filter described below as the more adequate than existent ones which affect the text quality. It consists in neighborhood test: each non null pixel in the differential frame is compared with its 8 neighbor pixels. If they are null then the pixel is considered as noisy. Thus, it is cleared (set to zero).

4- Local thresholding
Pertinent changes could be generally presented by concentrated pixel value changes. However, the dispersed pixel value changes are considered as noise. Hence, this process allows determining the changed blocks to be refreshed. The total number of changed pixels in one modified block must exceed an adequate threshold, else it will be considered unchanged. Then, changed blocks will be encoded and transmitted in order to refresh last reference frame blocks having the same locations.

5- Coding
This step consists in compressing the blocks to be transmitted and actualized in the reference frame. The chosen technique should be in favor of good details quality and text readability.

Many existing techniques are available: JPEG, JPEG2000, TIFF, GIF, PNG, etc. These techniques were applied to many slideshow samples. The example below (Figure 3) shows that actually JPEG and PNG are the most efficient in slide's compression. JPEG is always more efficient in slides with complex texture and motif and rich color palette (see the two slides on the top of Figure 3). Otherwise the PNG ensures a high compression rate with an invincible quality as it is a lossless compression technique. The main purpose consists in coding blocks separately with the adequate coding technique. So the block's content proprieties will be extracted in order to choose which technique will be used.

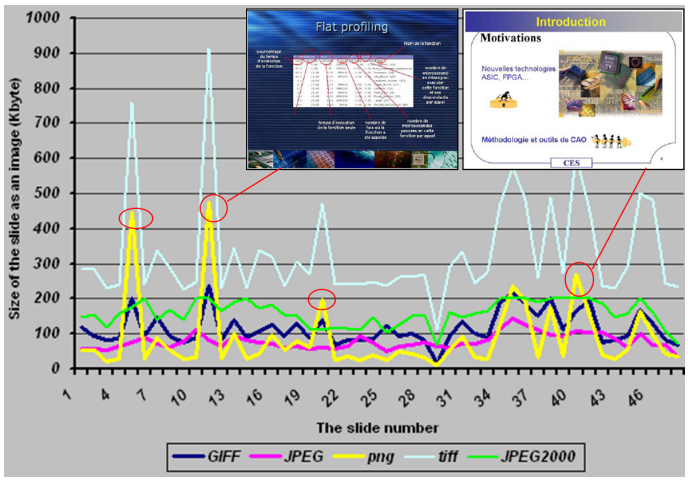


Fig. 3. Comparison of different image coding standards applied on a slideshow example.

Hence, for palletized region, as text zone and edge, the most adequate format is the PNG, which gives highest compression and best image quality. For other zones, generally continuous tone regions including photo and graphics, which tolerate loss of some image quality, JPEG is more efficient.

6- Compressed blocks transfer

All compressed blocks will be sent to the receiver. After decoding them, they will be actualized in the reference frame according to their index-number.

## V. Preliminary Evaluation

We have developed an application to automatically evaluate the various slide change detection algorithms. Many PowerPoint presentations have been collected, mainly from conferences, student projects and courses available on the web. The slides that have been accumulated this way represent many kinds of presentation styles. We have then built a corpus trying to equally balance various characteristics such as number of slides, background color, font color/size and background variability, graphics content and animation styles.

### *A. Slide change detection module*

As we mentioned earlier, a slide change is detected after a pixel by pixel comparison applied to the blocks of two slides. This method ensures gain of a considerable amount of information to be processed and transferred. This rate depends directly on the choice of the size of blocks and thresholds values.

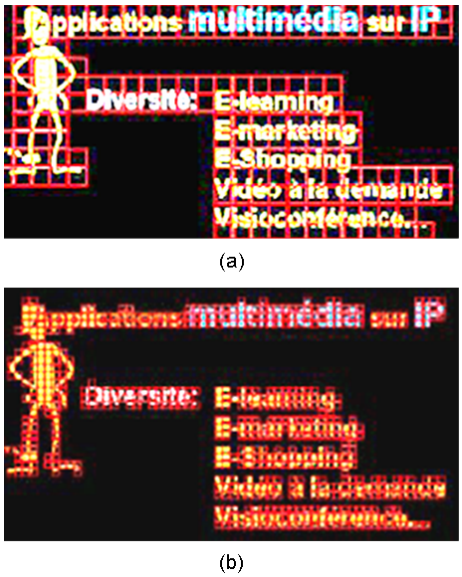


Fig. 4. Block detection.

For example, with a block of 20x20 size (Figure 4-a), there are 237 modified blocks detected, which means 237 calls to the coding functions, but also less fine detection. On the contrary, with 4x4 block size (Figure 4-b), more treatment will be required and there will be lesser black areas (unmodified). The following table shows the results of block detection module with different parameters: it gives the number of changed blocks with different block size and local thresholds and it provides the reduction size rate RSR (Table 1). RSR= size of transmitted data / size of total data. Size of transmitted data= number of changed blocks x block size x 8 bits + number of changed blocks x size of position vector.

Table 1. Experimental results.

| **Block Size** | **Nb of blocks** | **Nb of changed blocks** | **Local Th** | **Size of position vector** | **Transmission rate** |
|---|---|---|---|---|---|
| 8x8 | 7500 | 321 | 10 | 13 | 4,388 % |
| 20x20 | 1200 | 85 | 10 | 11 | 7,107 % |
| 20x20 | 1200 | 93 | 0 | 11 | 7,110 % |

After many experimental tests, a local threshold equal to ten pixels has been chosen . The used blocks are of 8x8 size, because it gives a finer detection. We show, just below, the results corresponding to experiments done on a presentation example. This presentation contains 45 slides and lasts 20 minutes. It is 800x600 resolution and the majority of its slide contains different effects animation and transitions. Then the following results have been recovered:

- Total number of images (for sequence of 25 img/s): 30000,

- Total number of different images: 895, thus 2.98 %,

- Total number of coded blocks: 1 594 051 blocks of 6 712 500 blocks of different images, thus 23.74 %.

The bit rate was 774.826Ko/s for MJPEG, 186,026 Ko/s for MPEG1, 500,053 Ko/s for MPEG2 and 32,8 Ko/s with our method.

### *B. Text detection module*

The text represents the most important information to be treated in a slideshow. For that, we have considered a module that detects blocks containing text and/or edges. Currently, algorithms can be classified in two categories. There are those working on the compressed domain and those working on the spatial domain [10][11]. Every method takes into account different features depending on the type of sequence (e.g. sport event or news) and its quality. We have tested two methods. The first processes in spatial domain, it is based on color histograms. The second processes in frequential domain and it is based on texture features extracted from DCT coefficients. Many experiments had been conducted to study the two method performance. They proved that the DCT based technique is more efficient considering text legibility criteria as it is shown in The figure 5. In fact, (Figure 5-a) shows the result relative to the DCT based method, (Figure 5-b) shows the result relative to color histograms based one.

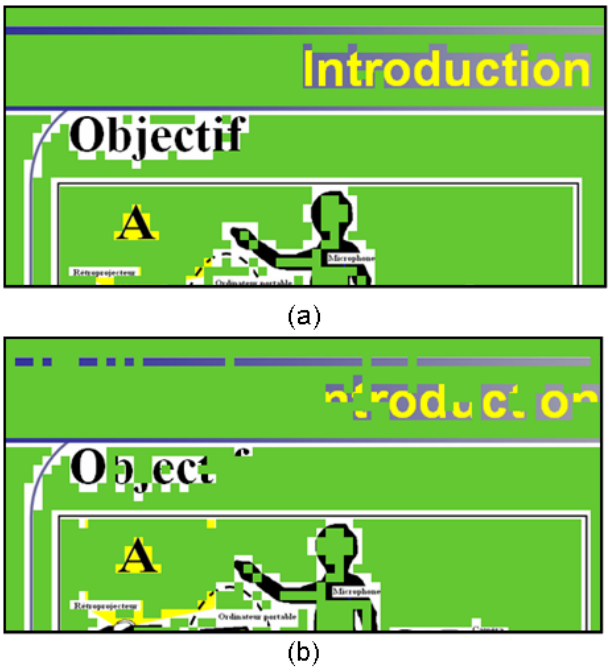


Fig. 5. Text detection techniques.

### *C. Coding module*

After the identification of the nature of each changed block, we proceed to the coding as follows: if the block contains text, then it will pass over DCT, quantification, RLE and Huffman modules. Thus, it will be JPEG encoded. Otherwise, it will be encoded in PNG format ensuring pertinent quality for text and edge region details.

### *D. Metrics for evaluating quality*

Several experiments were undertaken in order to show the proposed method performances in term of quality and compression. Others were carried out in order to show just the quality provided by hybrid compression. For the same example of presentation, we have proceeded to a PSNR quality measure. Figure 6-(a) presents the measure of the flow coded by only JPEG; and Figure 6-(b) shows the measure of the flow coded by JPEG and PNG. In the first case, the PSNR is about 33 dB and it was increased by 9 dB in the second case. As well as the objective measure of PSNR, the subjective measures (by the naked eye) show the advantage of the second method in keeping text quality and readability.

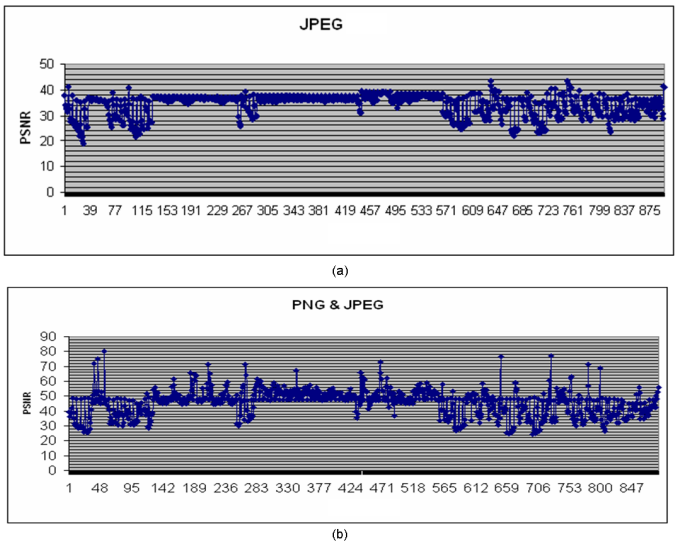


Fig. 6. PSNR measures.

## VI. Conclusion

In this paper, we proposed a method based on image processing, perfectly dedicated to slideshow processing. This method is a complete solution for coding, broadcasting and/or archiving a slideshow. It ensures slide acquisition, efficient local slide change detection and adaptive hybrid coding based on the slide proprieties. It respects slide stream characteristics as it permits to discriminate static frames and animated portions. In addition, the method ensures pertinent compromise quality/bit-rate and relevant matchless text readability without needing complex processing. This is confirmed by objective evaluations provided by PSNR values and subjective ones showed in image visual quality.

This technique developed as software application in C/C++ language, is designed to be included in a complete audio-visual system whose final aim is to automate storage, processing

and broadcasting of different multimedia data in a slideshow system.

In order to prepare for the following stage of the design, we plan to make a detailed analysis of the performance of the various functions. It is about the HW/SW partitioning which must take account of the operational complexity of each function (execution time).

## REFERENCES


[1] A. Behera and D. Lalanne. Looking at projected documents: Event detection & document identification. In *Intl. Conf. on Multimedia Expo (ICME' 04)* , 2004.

[2] S. Mukhopadhyay and B. Smith. Passive capture and structuring of lectures. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, Orlando, Florida, 1999.

[3] J. Revertera. Système de détection de changement de slides. Master, Architects and Enginneering School of Fribourg, 2008.

[4] P. Gasser. Visioconférence : les technologies d'aujourd'hui. In *MSH Paris nord - Plate forme Arts, Sciences, Technologies*, Paris, 2005.

[5] J. Prévost. La vidéo numérique sur IP et la communauté Renater. In *JRES*, 2001.

[6] G. Bisiaux et al. Projet DIM : Télé-enseignement pour les DESS Nouvelles Technologies. In *JRES*, 2001.

[7] R. Brunelli, O. Mich and C.M. Modena. A survey on the automatic indexing of video data. *Journal of Visual Communication and Image Representation*, 10(2):78–112, 1999.

[8] F. Idris. and S. Panchanathan. Review of image and video indexing techniques. *Journal of Visual, Communication and Image Representation*, 08(2):146–166, 1997.

[9] G. Ahanger and T.D.C. Little. A survey of technologies for parsing and indexing digital video. *Journal of Visual Communication and Image Representation*, 07(1):28–43, 1996.

[10] X. Qian et al. Text detection, localization, and tracking in compressed video. *Signal Process. Image Commun. (ELSEVIER 2007)*, doi:10.1016/j.image, 2007.

[11] M. Leon and A. Gasull. *Text detection in images and video sequences* Image processing group, Department of Signal Theory and Communications, Technical University of Catalonia, 2005.