A method for choosing reference images based on edge detection for video compression

Mohammed Benabdellah, Fakhita Regragui, Mourad Gharbi, El Houssine Bouyakhf
Groupe de Recherche en Informatique, Intelligence Artificielle et Reconnaissance de Formes.
Faculté des sciences de Rabat-Agdal.
med_benabdellah@yahoo.fr, regragui@fsr.ac.ma, gharbi@fsr.ac.ma, bouyakhf@fsr.ac.ma

Abstract
In this work, we propose carrying out tests on the encoding sequence of a video sequence to improve data flow and average PSNR. We experiment with the choice of reference images, in the process of video compression, by using only the intra and predicted images extracted from sequences.

For each intra and predicted image, we perform edge detection. Each image of the sequence is compared with the other images by subtracting corresponding edges. The choice of the reference image is based on the result of subtraction. We adopt the criterion of minimum pixels if the resulting images present only points and the criterion of minimum distance between the lines if they present parallel lines and possibly points.

Testing this approach on News and Kiss cool video sequences revealed an improvement in data flow and average PSNR as compared to the original encoding and choosing reference images based on the mean square error.

Keywords:
Image processing, Intra image, Predicted Image, Bidirectional Image, MSE, PSNR, RGB, YUV, DCT, GOP, MOP, ZigZag, MPEG, H.26L, JPEG.

1.Introduction

The compression of the video sequences is necessary to store them economically as well as to transmit digital video data through a limited band-width network or from a media with a limited transfer rate [11]. During the last few years, interest in multimedia and in particular diffusion of the audio-visual content involved a great amount of research in the field of video signal coding, which led to several standards such as H-263, H.26L and MPEG-4. These standards consist essentially of toolboxes for video signal processing which can be adapted to the context and desired result (flow/distortion report)[12]. Recent research aims to improve these tools and set froth new ones. In this paper, we first introduce the techniques used in the video compression process. Next, we will explain the stages followed in our experiments. The results are compared with the original sequence and the method that uses MSE to find the best reference images.

2.Methods

2.1. A video sequence representation

A rough video sequence is a continuation of fixed images, which can be characterized by three principal parameters: its resolution in brightness, its spatial resolution and its temporal resolution [14].

The resolution in brightness determines the number of nuances or possible colors in a pixel. This one is generally 8 bits for gray levels and 24 bits for coloured sequences. Spatial resolution defines the number of lines and columns in the matrix of pixels. Finally, the temporal resolution is the number of frames per second. The value of these three parameters determines the necessary memory capacity to store each image in the sequence. This memory capacity is characterized by the data flow, which is the storage cost per one second (memory size needed to store one video second). For instance, one second with a resolution of 720 by 576 pixels, a 24 bit color coding, and a frequency of 25 frames per second, will require a flow of 137 Mb/s. The flow of a rough video sequence is very high in comparison to the data flow and space provided by the current resources [10].

2.2. Coding algorithm

Data compression reduces the quantity of information by modifying its mode of representation. The techniques used to compress a video signal use space redundancy. The objective is to reduce the flow of the video sequence to be compressed, while minimizing the visible errors (MSE and PSNR)[3]. To do this, there are two principal techniques, lossless compression and lossy compression. The former makes it possible to find the initial information after decompression, while the latter will restore only an approximation of it. In the case of natural images, the lossless compression is insufficient, and the introduction of losses in the compression process makes it possible to obtain better results without preventing the interpretation of the visual content [9]. Current video standards use a hybrid coding system with compensation for movement based on blocks and a reduction of entropy by a transformer (see figure 1) [11].

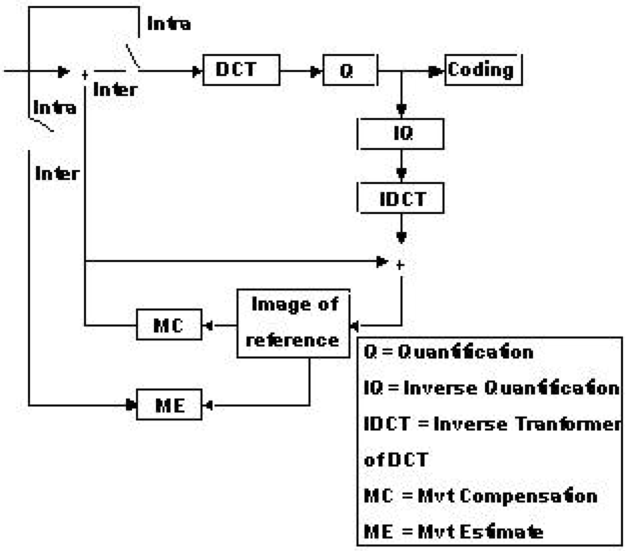

Figure 1. A general outline of a video coder.

2.2.1. Structure of coding and structure of GOP (Group Of Pictures)

The MPEG standard defines a set of coding stages that transform a video signal (digitized in standardized format) into a binary stream (a bit stream) intended to be stored or transmitted through a network. The binary stream is described according to a syntax coded in a standardized way that can be restored easily by any decoder that recognizes the MPEG standard [12]. The coding algorithm defines a hierarchical structure containing the levels described in the following figure [2] :

Sequence	Heading	GOP	....	GOP
GOP	Heading	Image	…………..	Image
Image	Heading	Slice	…………..	Slice
Slice	Heading	Macro-block	…………..	Macro-block
Macro-block	Heading	Block	…………..	Block
Block	Heading	Coefficient	…………..	Coefficient

Figure 2. Hierarchical structure of MPEG coding.

The group of pictures or GOP consists of a periodic continuation of compressed images. There are three types of the compressed images [1]: an image of type I (Intra) compressed using JPEG for the fixed images. An image of type P (Predicted) coded using a prediction of a previous image of type I or P. An image of the type B (Bidirectional) coded by double prediction (or Interpolation) by using a previous image of type I or P and a future image of type I or P as references. A GOP starts with an image I, contains a periodic continuation of the images P separated by a constant number of images B [6].

The structure of GOP is thus defined by two parameters : the number of images of GOP and the distance between images I and P [5].

2.3. The reference images

The image used to predict another image and estimate the movement to be compensated, is called a reference image. This can simply be a previously coded image in the sequence. In the latest standards, new images are used: the mosaic objects MPEG-4 standard [2] has added the coding mode sprite to Intra and Inter modes in order to construct such objects. However, nothing is mentioned with regard to the construction process in this standard [4].

2.4. The problems

Traditionally speaking, the reference image used during the encoding of a predicted image is not chosen. However, it is the previous image of the type P or I that is used instead. In certain cases, the choice of an image located further in the sequence would seem more suitable.

As a first example, let's take the case of a change of a repetitive background. If a change of background occurs between two predicted images, the prediction of the first image with the second background will be made from an image with the previous background, whereas it would seem more interesting to use a former image of the same background (if such an image exists). This problem is perfectly illustrated on a *News* sequence of figure 3 where the predicted image 69 refers to an image with background B, whereas images of background A were already coded later on (see the predicted image 63).

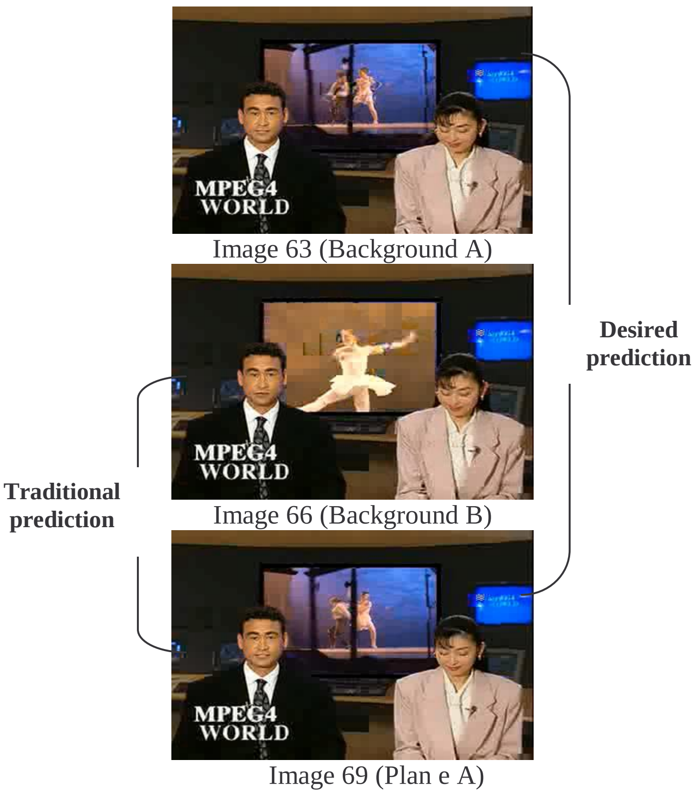

Figure 3. Prediction during a background change (the News sequence).

The second example consists of the entry of an object into a scene of a *Kiss cool* video sequence consisting of 713 frames, shown below in figure 4. In this case, it would better to choose as a reference image the one where the object is completely returned to the field of view. Therefore, the prediction would be more effective for the visible parts of the object for each image where it appears.

Image 75

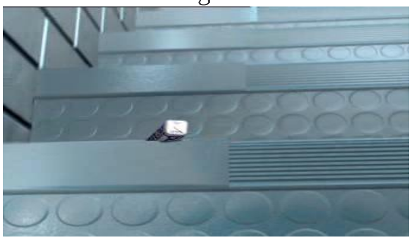

Image 132

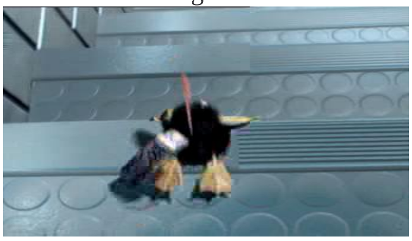

Image 144

Figure 4. The *kiss cool* sequence.

The objective of this work is thus to experiment on possible relevant references and identify a criteria to determine which images should be used as reference images.

2.5.Diagram of the proposed method

The proposed method is described in figure 5 below.

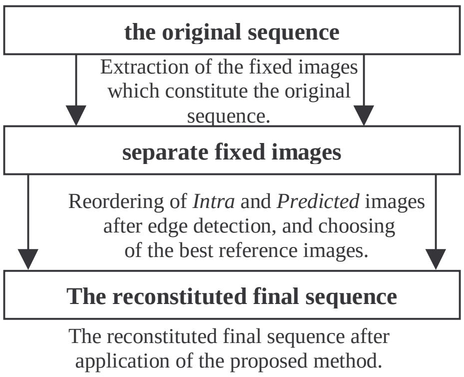

Figure 5. The proposed methodology.

3. Results

3.1. Automatic edge detection

There are numerous methods of edge detection. Therefore, we chose to use two types of filters : the exponential filter and the Gaussian filter, both of which are skeletal. On the one hand, these filters are used to eliminate the parasitic noise present on the image, and on the other hand to calculate the first and the second directional derivatives of the variations in the image gray levels. The first derivative makes it possible to obtain the amplitude of the image gradient and therefore gives a initial image of the edge gray levels. The second derivative, especially the zero passages, helps to determine the maximum gradient (figure 6). The edges of the objects present on the initial image are deduced from this information. These results are presented in the form of a binary image with white edges of 1 pixel thickness on a black background. We will carry out these calculations several times by using different parameters for each filter and we retain the best result based a statistical test. However, these edges present a small noise in the form of isolated points or short branches. We carry out an automatic verification on edges in order to optimize their detection. This verification is based on a follow-up method and classification of branches. As a result, we obtain a noiseless image with smooth edges [13].

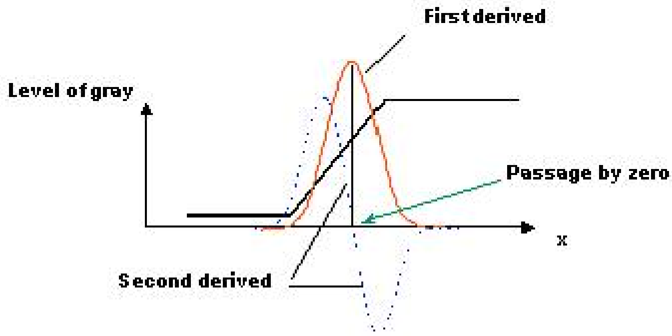

Figure 6. Presentation of the first and second derived of a transition type "Amplitude Jump".

These combined pieces of information (amplitude of the gradient and zero passages) enables us to deduce the desired edges by applying an adapted segmentation method.

3.2. The search for the best intra image

Realizing that the reconstruction of sequences allows us to choose the Intra image of a GOP, we would like to observe the influence of the choice of the Intra image on compression quality (in the flow/distorsion direction). A script is provided, for a given GOP, to test the encoding with each image of the GOP as an Intra image and plot the storage costs of the images P and the average PSNR of the decoded images of each encoding. The result of the execution of this script on the *Kiss cool* sequence for the GOP from frame 78 to 90 can be seen in the figure 7. A gain is noticed in storage using the second image (image 79) as an Intra image. Good results are not obtained from the following images. The use of the image 81 produces an equivalent result (a quite higher storage, but lower distortion). To obtain better results, we reorder the sequence completely.

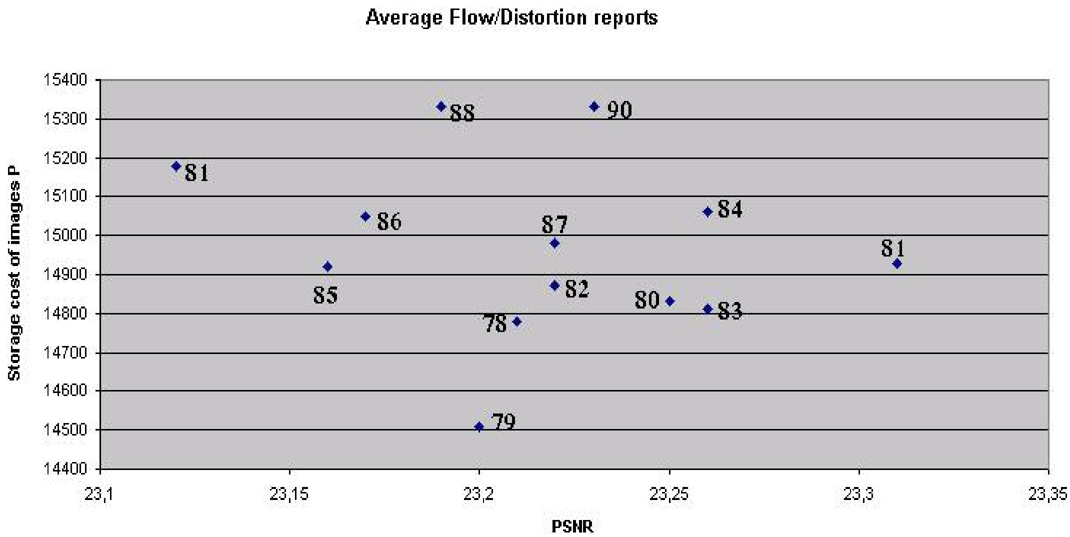

Figure 7. Flow/distortion report of the various Intra images of the GOP 78-90 in the Kiss Cool sequence.

3.3. Reordering of images and applications

The process consists of first gathering the similar images based on edge detection of *Intra* and *Predicted* images. Second, we perform a subtraction between the images in question and the other *Intra* and *Predicted* images of the GOP. The selected image of reference is the one whose edge subtraction with the image in question produces either the fewest points or maintains the same minimal distance between lines of points. A *new* sequence is then reconstructed image by image, through comparing each image with the others to select the nearest one. Later on, a search for the best *Intra* image is carried out for each GOP of the sequence. Once all of this information is gathered, the sequence to be encoded is then reconstructed. The results of this method on the sequence *News* (images 60-81) and on the sequence *Kiss cool* (images 132-153) are given as follows:

***News* sequence :**

Encoding with reordering and choice of the best Intra images :

Reordering of the sequence : 0 1 7 6 5 4 2 3

The best Intra image :

GOP 1 : image 1

GOP 2 : image 1

***Kiss Cool* sequence :**

Encoding with reordering and choice of the best Intra images:

Reordering of the sequence : 0 1 5 4 3 2 7 6

The best Intra image :

GOP 1 : image 1

GOP 2 : image 1

	Original encoding of the sequence		Encoding by the method using MSE		Encoding by the introduced method	
	Flow (Kbit/s)	Average PSNR (dB)	Flow (Kbit/s)	Average PSNR (dB)	Flow (Kbit/s)	Average PSNR (dB)
The News sequence	1367.73	38.036927	1055.32	38.051030	1038.39	38.056025
The Kiss Cool sequence	1137.33	40.015128	823.32	40.032173	807.47	40.037121

Table 1: The results, given in flow and average PSNR, obtained after the application of the original encoding, the encoding by the method uses the MSE and the introduced method.

The re-ordering of the sequence was effecient since the images were gathered by background. For same visual quality, the flow obtained with the regrouping of the sequence is much better than that obtained by original sequence and that obtained by Nicolas DUMOULIN’s method which uses the MSE to deduce the similar images and to reconstruct the sequence of the images.

4. Conclusion

The main idea behind this paper is to carry out tests on the modification of the encoding sequence of the video sequence images to produce a gain in the result. Some tests highlighted a possible gain for certain sequences through the choice of the reference images. Our method improves upon the MSE method in terms of being able to determine similar images and gather them before applying the encoding. We hope to apply this method to any type of image in the GOPs including *bidirectional* images.

Bibliography

1. UIT-T. Recommandation H.263, version 2 “Codage vidéo pour communication à faible débit”, 1998.
2. Rob Koenen. “Overview of the MPEG-4 standard” iso/iec jtc1/sc29/wg 11 n4030. Mars 2001.
3. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. Joint Final Commitee Draft (JFCD) of Joint Vidéo Specification (ITU-T Rec. H.264 – ISO/IEC 14496-10 AVC).
4. Majid Rabbani, Paul W. Jones. “Digital Image Compression Techniques”, volume 7 of tutorial texts in optical engineering. SPIE Optical Engineering Press, Bellingham, WA, USA, 1991.
5. Henri Nicolas. “Contribution à la création et à la manipulation des objets vidéo”, PhD thesis, Université de Rennes 1, Institut de Formation Supérieure en Informatique et Communication, 2001.
6. M. Irani, P. Anandan, J. Bergen, R. Kumar et S. Hsu.“Efficient representations of vidéo sequences and their applications. Signal Processing : Image Communication”, 8:327-351, 1996.
7. Guy Côté et Lowell Winger. ‘‘Progrès récents dans le domaine de la compression vidéo’’. IEEE canadian Review, 2002.
8. Nicolas DUMOULIN, “Compression de séquences vidéo et choix des images de références”, Rapport de stage DEA Informatique, Université de Rennes 1-IFSIC, Juin 2003.
9. Charles WAGNER , “De l’image vers la compression”, Rapport de Recherche de l’INRIA, Septembre 1993.
10. Monique COLINET, “Multimédia : Images-Sons-vidéos”, CEFIS-Université Notre Dame de la Paix, Février 2001.
11. Jacques GUICHARD, Dominique NASSE, “Traitement des images Numériques pour la réduction du débit binaires”, CENT-Paris, CCETT, France.
12. Grégoire MERCIER, Christian ROUX, Gilbert MARTINEAN, “ Technologie du Multimédia”, ENST Bretagne, F-29280 Brest, France, 15 Janvier 2003.
13. J.P. Cocquerez, S. Philipp, “Analyse d’images : Filtrage et segmentation”, MASSON.
14. M. Shen, “Cours de traitement d’image”, DESS Ingénierie de l’Image année 1999 – 2000.

Article received: 2005-09-23