An automatic technique for visual quality classification for MPEG-1 video

SORIN SAV, SEAN MARLOW, NOEL MURPHY, NOEL O'CONNOR
Centre for Digital Video Processing
Dublin City University
Glasnevin, Dublin 9
IRELAND
sorinsav@eeng.dcu.ie http://www.eeng.dcu.ie/~vmpg

Abstract: - The Centre for Digital Video Processing at Dublin City University developed Fischlar [1], a web-based system for recording, analysis, browsing and playback of digitally captured television programs. One major issue for Fischlar is the automatic evaluation of video quality in order to avoid processing and storage of corrupted data. In this paper we propose an automatic classification technique that detects the video content quality in order to provide a decision criterion for the processing and storage stages.

Key-Words: - video quality, Mpeg-1 video compression, decoded domain, automatic technique.

1 Introduction

The main purpose of digital media is to facilitate indexing and browsing for quick multimedia information retrieval. The quality and the amount of the rendered information are two major features of any information retrieval system. In most of the cases the low quality of video data complicates the processing steps in the system. Furthermore, in the worst cases the video content becomes imperceptible for users and therefore useless for the video retrieval concept. These cases are a waste of system's processing, storage and transmission resources. Recognition and avoidance of such cases would save more processing and storage resources for subsequent incoming data. The amount of work invested by a human operator to remove the low quality data is directly proportional to the dimensions of the retrieval system. The manual seeking and removal becomes tedious and unfeasible for large data archives. Otherwise, detection of such unusable data sequences would occur only at normal database access.

A preliminary classification of the input data quality is therefore more desirable then the processing of unusable data. This would require further allocation of system resources. Moreover, most of the time the quality of the input data is believed to be satisfactory. Therefore, the preliminary data quality classification is desirable to consume a minimum set of system resources and we chose this requirement as one major criterion for the classification technique evaluation.

In the Fischlar system television broadcasts are captured and encoded according to the MPEG-1 digital video standard. The capturing and encoding process is entirely automatic [2], therefore the degradation of reception of television programs reflected in the capturing or encoding quality could not be noticed during the system processing steps, but is reflected in the quality of audio or/and video content constituting an annoying factor for users.

2 Video quality analysis

From the Fischlar users perspective, low quality content is a video or/and audio sequence that has an incompressible content due to the presence of video and audio artifacts. During television reception or in capturing and encoding process such artifacts can affect one of the audio or video layers or indeed both layers simultaneous.

Respecting the requirements that impose a reduce influence over the normal system processing flow, we have considerated the detection of video artifacts in decompressed video domain as the most suitable solution, preferring the detection of corrupted digital image, because of the implementation simplicity. This solution not require additional hardware structure for system and not imply major changes in the processing chain. The normal processing chain for the system involves video decompression using the XIL library, a open source imaging library provided by Sun Microsystems [3]. The quality of video sequences can be evaluated without involving supplementary processing steps. Another important requirement is the implication for the processing time. A fast technique, which does not involve a substantial computation time, is desirable. Adding

an extra delay to the system is undesirable as it will increase the overall delay between when a programme is captured, analyzed, and made available to the users. In this context an aproximate but fast approach is preferable to a more exact, time consuming solution.

Our goal is detection of unintelligible content in a video sequence, such sequences being the effect of a weak reception of television signal. Relating to image quality characterization two aspects are involved:

- an objective characterization using signal processing and image quality metrics.
- a subjective appreciation by human observers.

Since Fischlar is a user-oriented system the subjective appreciation of users – human observers – is the determinant factor for image quality decision. Therefore, the evaluation of our video quality classification technique is based on visual appreciation of human observers for a large collection of television broadcast recorded video material of different quality.

2.1 Overview of possible approaches for video quality classification

As we affirmed above our goal is classification of digital video sequences, encoded according to the MPEG-1 video standard [4], in two categories:

- intelligible content – video sequences for which the image content can be easy comprehended by a human observer, even that the images quality is not perfect.
- unintelligible content – video sequences for which the images are corrupted by noise or artifacts and the content is comprehended only with difficulty.

The classification of images in one of these categories is based on human observer evaluation. For development of the classification technique were selected a set of reference video sequences of different quality from a wide spectrum of television broadcast sequences covering many categories of program content: news, movies, sports, cartoons, talk shows as well as gray-levels images from black and white movies.

Simplicity is an important goal and we chose this requirement as one of the major criteria in development and evaluation of the visual classification technique. Other proposed metrics presented in [5], [6], [7] and [8] require complex computational resources that are evidently time consuming and thus inadequate for our application-imposed criteria.

As a starting base were selected as plausible the following approaches:

- image reconstruction based on digital filtering in real or Fourier space.
- image characterization based on statistical proprieties of image.
- video sequence characterization based on pixel differences between successive frames.

In the following sections we treat the above approaches.

2.2 Video quality classification using digital filtering

Video quality classification using digital filtering is based on idea that passing a deteriorate image (image with low visual quality) through a proper digital filter most of the disturbing effects will be removed obtaining an image with an improved quality. For an initial clean image the effect of digital filtering operation must be minim, if possible the image should pass unchanged. Comparing the initial and the resulted images through a common metric like PSNR – peak to noise signal ratio [9], substantial differences should be found between the deteriorate image and its filtered version, and minimum, or ideal no differences, between a clean image and its filtered version.

This technique is similar with image reconstruction techniques the common goal being the recover of noise and artifacts affected images. The target for this technique is the development of the digital filter with the required transfer characteristic.

The MPEG-1 video standard uses the YUV color space to represent digital images, every image has three-color components: luminance component Y and two chrominance components U and V. The image representation in three-color components gives to two problems connected with the digital filtering operations:

- a digital filtering operation must be effectuated for every component, involving three times more computing time.
- the convolution kernel for every component should compute the new value for that component based on values from all three components of the image, involving a three dimensional kernel.

A simple solution is the reduction of filtering operation only to the luminance component Y. Using this approach for the common 3x3 low-pass filter we expected to have satisfactory preliminary results. Despite our expectations the result haven't been very promising, for low quality images the

PSNR being close to the PSNR for clean images. Changing the convolution kernel to a 3x3 nearest neighbor kernel we obtained similar results as is shown in Figure 1.

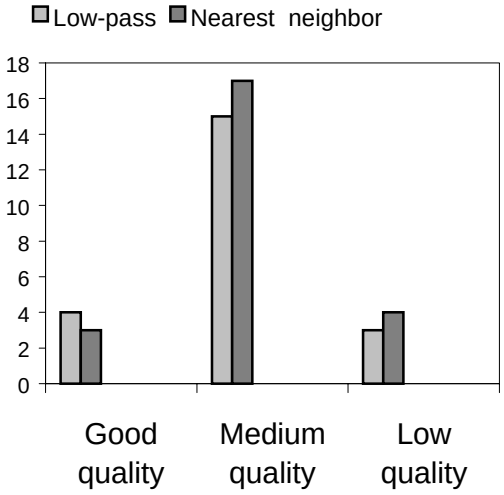

Fig 1. PSNR average values for digital filtering

The reason for the inability of the filter to improve substantially the visual quality for corrupted images resides in the image distortion characteristic. Due to improper reception of television channel the encoded image presents blockiness effects, an MPEG specific artifact, and the corrupted noisy pixels are grouped in small formations of 3 – 5 pixels with similar values. This specific configuration of deteriorated images makes reconstruction a hard task for any convolution kernel.

From this we conclude that the computational effort required by an optimal convolution kernel excess the constraints imposed in our techniques. In this context, the implementation of an automatic quality classification technique using digital filtering is computationally inefficient.

2.3 Video quality classification based on successive frames differences

The video quality classification based on successive frames differences exploits the differences between the pixels from the same location in two consecutive frames of a video sequence.

Watching a video sequence, for a human observer the differences between two consecutive frames is generally imperceptible, during the same camera shot. This corresponds to the presumption that between two corresponding pixels from consecutive frames is a minimum difference. Presuming the disturbing effects as having time random distribution, therefore every frame from the video sequence is affected independently from any other frame and the affected pixels values are uncorrelated in successive frames. Thus, for a clean, unaffected video sequences the pixels values from successive frames are strongly correlated as well the difference between corresponding pixels from successive frames is negligible. Obviously for noise corrupted video sequences the difference between corresponding pixels should be significant. Using this approach and PSNR as metric we have expected relevant indications for visual quality classification.

Experiments have proved that the above considerations are correct for relatively static sequences but for dynamic video sequences, which involve objects in rapid movement or camera zooms, the differences between successive frames are closer to results obtained for corrupted video sequences. It is therefore difficult to establish the threshold for classification criteria. The experimental results are plotted in Figure 2.

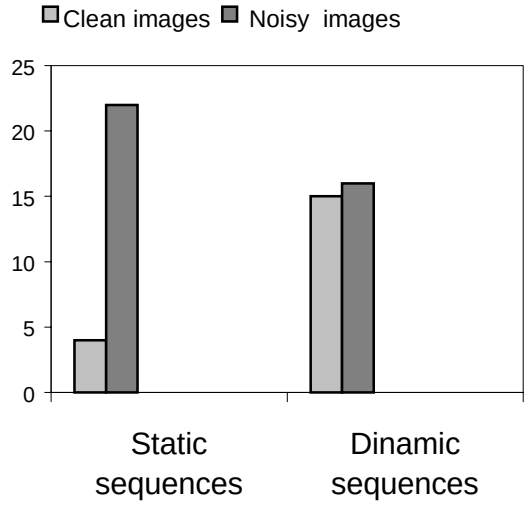

Fig 2. PSNR average values for successive frame difference

The explanation for these results is simple; an object movement or a camera zoom is a spatial shift of image that change dramatically the pixels positions from a frame to the next one introducing errors values for every moved pixel. The way to deal with dynamic video sequences is to estimate the changes for actual frame relative to previous one,

but this process being computationally expensive is unfeasible for our application.

2.4 Video quality classification based on statistical proprieties of images

Any statistical aspect common for a group of images can constitute a criterion for classification of images. Defining a category of images implies to find any common proprieties which describe the given image category and the methods to distinguish if an image poses or not those proprieties. The most used proprieties for image classifications are: histogram and histogram distribution, average intensity, skewness of intensity, maximum and minimum pixels values, etc.

For visual quality classification of video sequences by statistical proprieties the target is to find the characteristic proprieties of clean and respective noise and artifacts affected sequences and the methods to make them evident. Visual observations have revealed the presence of strong primary colors of the RGB three-color space in images corrupted by noise and video artifacts. Practically this means that a numerous percentage of pixels from every frame have an intense red, green or blue color more intense than in the usual cases of good quality images. Based on this visual observation of low quality images of MPEG-1 sequences we presumed that transposing the image from the MPEG-1 implicit YUV color space to RGB color space, those pixels which have strong primary colors, poses in the RGB space values near the maximum intensity for that color component which color they have and zero values for the rest of the components. If a corrupted pixel is intense blue it will have in RGB color space a value near 255, maximum value for the component intensity in MPEG-1 video standard, for the blue component and values 0 for the red and the green components.

Converting the images from the YUV color space to the RGB color space and experimenting on the video sequences from the selected test set, the presumption has proved correct. One single issue is necessary to complete the visual quality classification technique: to establish the threshold values for a proper classification. Using an extended test set we have investigated the ranges of possible zero values in images with diverse content and diverse visual quality. A summary of the results of investigation for the given test set is presented in Table 1. The results are not separated by categories of contents because we used the video recordings of television broadcast and every recording contains more then only one single content category, most of the television programs being in fact a mixture of contents (some contain advertisements).

Quality level	*Number of zero values*
Good visual quality	0 – 4200
Low visual quality	3800 – 49000
Pure noise	6300 – 12000

Table 1. The number of zero values

From Table 1 is easy to observe that the interval of values for pure noise images (images encoded in absence of television signal) is included in the interval of values corresponding to the images with low visual quality. For the current application requirements is not necessary to distinguish between these levels of quality, both being images with unintelligible content, unable for use. However the distinction between these categories of images is simple, for pure noise images all pixels are in gray scale characterized by presence of equal intensity values for each color component in the RGB color space. By simply checking the equality of intensity values for all three-color components, we could differentiate the pure noise images from the low quality images.

Another common interval of values is between the low visual quality images and the good visual quality images, the values from 3800 to 4200 cover both categories of images. Generally the values over 2800 are specifics to news journal images, in which for the news reader's scenes are used strong intensity backgrounds, usually artificial generated backgrounds.

3 Thresholds and decision criteria

From the above sections it is evident that the only solution that respects all the necessary requirements for our application is the last presented method, the video visual quality classification based on statistical proprieties of images, the other two approaches being computationally inefficient. For implementation we have chose the only viable technique at this moment, manually setting the adequate thresholds and decision criteria.

Important information for thresholds settings is revealed from the values in Table 1 and from the above considerations. It is evident that the thresholds for every image category must be chosen near the minimum experimental observed value for

the corresponding image category. We have chosen the following thresholds as shown in Table 2. These thresholds represent the number of zero values in the values of intensity components of the image.

Quality level	*Thresholds*
Good visual quality	0 – 3000
Low visual quality	over 4500
Pure noise	over 6000

Table 2. The image categories thresholds

Limiting the visual quality categories at those two categories defined in section 2.1 corresponding thresholds for them are the following:

- intelligible content images: 0 – 3000 zero values
- unintelligible content images: over 4500 zero values

It can be observed that the thresholds chosen don't cover the case of news journal images above mentioned, additional decision criteria being necessary to avoid wrong predictions in the common interval for the good visual quality images and the low visual quality images. The advantage of video sequences is that can be used the average zero value for a number of frames for quality classification. Usually every news story has two distinct parts, a preamble presented by the speaker in the news studio on a possible artificial background and a reportage composed by natural scenes. This provides a useful method to discern the image quality, we chose the average value for one hundred frames, decreasing in this way the average values for the video sequence.

Concluding, the above-presented thresholds combined with the average of zero values for a number of frames provide a powerful and simple automatic technique for visual quality classification.

4 Conclusion

In this paper we have presented a summary of our investigation in designing an automatic technique for visual quality classification for MPEG-1. In the development and the evaluation of the presented technique were used a set of representative video sequences of different visual quality covering many categories of television broadcast content: news, movies, sports, cartoons, talk shows and as well gray-levels images from black and white movies.

Our visual quality classification technique is able to discern between video sequences with intelligible content or unintelligible content in the sense defined in section 2.1. We have evaluated the performance of the technique by comparing its predictions to quality classifications made by human observers for over 2 hours of diverse video sequences.

However the technique presented in this paper is just approximative, realizing only a rough classification of visual quality, but suitable for our application. For more fine classification further improvements are necessary. A further work will cover more complex methods to refine the prediction scale.

References:

[1] Fischlar website. *Centre for Digital Video Processing. http: //www.fischlar.com*

[2] N. O'Connor, S. Marlow, N. Murphy, A. Smeaton, P. Browne, S. Deasy, H. Lee, K. Mc Donald, Fischlar: An On-line system for indexing and browsing of broadcast television content, *ICASSP 2001 – International Conference on Acoustics, Speech and Signal Processing. Salt Lake City, UT, 7-11 May 2001*.

[3] *XIL Programmer's Guide – August 1997*, SunSoft, Sun Microsystems, 1997.

[4] J. L. Mitchell, W.B. Pennebaker, C.E. Fogg, D.J. LeGall, *MPEG video compression standard*, Chapman & Hall, New York, 1996.

[5] A.B. Watson, Toward a perceptual video quality metric, *Human Vision, Visual Processing and Digital Display*, Vol. VIII, No. 3299, 1998, pp. 139 – 147.

[6] K.T. Tan, M. Ghanbari, D.E. Pearson, A video distortion meter, *Picture Coding Symposium*, 1997, pp. 119-122.

[7] T. Hamada, S. Miyaji, S. Matsumoto, Picture quality assessment system by three-layered bottom-up noise weighting considering human visual perception, *society of Motion Picture and Television Engineers*, 1997, pp. 179 – 192.

[8] C.J.B. Lambrecht, Color moving pictures quality metric, *International Conference on Image Processing*, Vol. I. 1996, pp. 885 – 888.

[9] A.N. Netravali, B.G. Haskell, *Digital Pictures: Representation, Compression and Standards (2nd Ed)*. Plenum Press, New York, 1995.