

Prediction of Distortion of Transmitted Video in IP Networks

Lais Silva Gomes*, Claudio de Castro Monteiro*,

*Federal Institute of Education

Science and Technology of Tocantins, IFTO

Palmas-Brazil

laisgomes2107@gmail.com, ccm@ifto.edu.br

Abstract—This paper presents an implementation to try to predict the distortion of a video transmitted in WLAN networks. So was developed a model of video distortion prediction, which not requires access to the original video. The model is intended to be used under real conditions of video transmission in WLAN networks. The results obtained with this implementation can demonstrate the influence of packet loss in the final quality of the received video, presenting a new strategy that can be used as a selection networks criteria, in order to ensure to the user a good experience in the reception of videos during their movement between WLAN networks.

Keywords— Video distortion, Video transmission, Wireless network.

I. INTRODUCTION

Both multimedia and mobile communications technology have experimented great growth and commercial success in recent years. As the two technologies converge, video on wireless devices, such as in mobile phones and televisions with 3G/4G systems, is expected to achieve unprecedented growth and success worldwide. However, different from the traditional video coding, video transmission over wireless networks with good quality and low distortion point to point is particularly hard once the received video is not only subject to quantization error, but also an error transmission [1]. During this sudden progress of wireless technology and IP, there has been growing the emphasis on quality reviews of video in real time, in various visual communication services. For this objective metrics were developed, which can be classified according to the original video availability so there are three types of metrics: full-reference that is based on a comparison between the original video and the transmitted video, however in the absence of the original or reference videos, they are not practical for evaluation of image quality in real time applications of video transmission; reduced reference in which the comparison is made between the degraded video and with only some information from the original video; no-reference ideal in the video quality evaluating because adopted as a basis only the degraded video, and is widely recommended for monitoring video quality in real time, where the original video is not always available. Research about NR metrics has attracted much attention recently as the video quality experts group (VQEG) are working to standardize NR metrics. The NR metric was chosen to be used in this work because it goes in a real time transmission scenario, beyond converge

with the goals of application in mobile devices. In this paper, is propose a prediction of distortion model based on a NR objective metric proposed by [2]. The spatial distortion of each frame of a video is predicted using the differences between corresponding regions of two adjacent frames in the video stream. The video quality is measured through the combination of spatial distortions of all images in the sequence. Organized as follows, in the section is exposed the related researches already developed in the investigation of distortion prediction and video transmission by wireless networks. The methodology used, addressing all steps and peculiarities of the tests, scenario description and analysis of the obtained results are contemplated in section 4. In section 6 are made the concluding remarks. Both multimedia and mobile communications technology have experimented great growth and commercial success in recent years. As the two technologies converge, video on wireless devices, such as in mobile phones and televisions with 3G/4G systems, is expected to achieve unprecedented growth and success worldwide. However, different from the traditional video coding, video transmission over wireless networks with good quality and low distortion point to point is particularly hard once the received video is not only subject to quantization error, but also an error transmission [1]. During this sudden progress of wireless technology and IP, there has been growing the emphasis on quality reviews of video in real time, in various visual communication services. For this objective metrics were developed, which can be classified according to the original video availability so there are three types of metrics: full-reference that is based on a comparison between the original video and the transmitted video, however in the absence of the original or reference videos, they are not practical for evaluation of image quality in real time applications of video transmission; reduced reference in which the comparison is made between the degraded video and with only some information from the original video; no-reference ideal in the video quality evaluating because adopted as a basis only the degraded video, and is widely recommended for monitoring video quality in real time, where the original video is not always available. Research about NR metrics has attracted much attention recently as the video quality experts group (VQEG) are working to standardize NR metrics. The NR metric was chosen to be used in this work because it goes

in a real time transmission scenario, beyond converge with the goals of application in mobile devices.

In this paper, is propose a prediction of distortion model based on a NR objective metric proposed by [2]. The spatial distortion of each frame of a video is predicted using the differences between corresponding regions of two adjacent frames in the video stream. The video quality is measured through the combination of spatial distortions of all images in the sequence. Organized as follows, in the section is exposed the related researches already developed in the investigation of distortion prediction and video transmission by wireless networks. The methodology used, addressing all steps and peculiarities of the tests, scenario description and analysis of the obtained results are contemplated in section 4. In section 6 are made the concluding remarks.

II. RELATED WORK

In recent years many efforts are made around the research on video quality. Researches that influenced and formed the basis for several moments of doubt. In [1] was presented three algorithms, all using HVS characteristics. The first two are full reference algorithms. The third is a null reference algorithm. The main scenario for the application of the algorithms is in video calls where are few changes in the context of the images. The differential of the research is in present algorithms for the assessment of video quality with low computational cost. For computational cost meant memory and processor usage. Two of the algorithms use PSNR and the third use NR metric and obtained better results in the tests, served for us as the basis for the distortion model used in this research. The algorithm 3 has a much closer result of subjective than the other algorithms including the PSNR, which had the worst performance. Featuring worse results than the PSNR and the other two algorithms only when there are frequent scenes changes. This occurs because the algorithm loses the reference of the last frame. Furthermore, the third presented processing consumption slightly higher than the other two. Other methods of video quality assessment based on distortion can be found in [2].

In the article [3] was developed an analysis about the application scenarios to estimate distortion in real time, using it to obtain performance gain using priorities datagrams of video. In order to design an Unequal Error Protection (UEP) adaptive of video data, a method for assigning a set of priorities for video datagrams was needed. This set must be adaptive in the sense that is designed to deal with handle packet losses in a channel subject to errors. This prioritization attribution algorithm can be used to obtain a performance gain in many different scenarios. But in the scenarios discussed in the paper, which are: First fulfilling requirements of QoS by adaptation mechanisms and bandwidth limit. The second improve the video quality; received in multi-hop environment. In these scenarios the gain in performance was not as satisfactory as those obtained in the other two application scenarios in other papers in the literature, however stimulated further research in video distortion and adaptivity can be used.

In the article [4] was presented a mathematical analysis to video quality prediction through formulas derivation, which seeks to contribute in video and transmission codification.

In [5] was made an analysis of the relationship between packet and video quality loss. Developed across a real-time analysis package system designed to monitor the packet loss, in order to evaluate the network congestion.

A. Distortion Prediction Method

The prediction method used in the development of this work is based on the third algorithm developed by [1], who uses NR metric type.

This algorithm was chosen to be developed, for having shown good results, and by the interest in the investigation of metrics that can be used in real time applications. According to [1] the algorithm use the pixels luminance varying between different frames as evaluation factor. If the pixels luminance variation is similar to the variation of the area around you (your neighborhood ([6])), so this variation is considered movement; if the variation is different in relation to the variation in the region around, it is considered distortion. In the algorithm ref alg: class demonstrates the step-by-step, to make clear the role of each.

Step 1: Take frame n and apply Sobel filter over the image and select the regions it has delimited. Apply the maximum likelihood algorithm to select the regions for evaluation. Inside a region, determine (x, y) for each pixel.

Step 2: For each (x, y) pixel inside a region, determine the pixels $(x + 1, y)$, $(x - 1, y)$, $(x, y + 1)$, $(x, y - 1)$;

Step 3: Take frame $(n+1)$, the next frame.

Step 4: Take the average luminance of pixels $(x + 1, y)$, $(x - 1, y)$, $(x, y + 1)$, $(x, y - 1)$ and (x, y) of frame n (we call it $Lum(n)$).

Step 5: Calculate $Lum(n)$ and $Lum(n + 1)$;

Step 6: If

$$|(X_n, Y_n) - (X_{n+1}, Y_{n+1})| \leq 2 * |Lum(n) - Lum(n+1)| \quad (1)$$

, where (X_n, Y_n) are the pixel (x, y) in frame n , we consider that there is distortion and, to weight movement, distortion is considered as

$$\frac{|(X_n, Y_n) - (X_{n+1}, Y_{n+1})|}{\max(Lum(n), Lum(n + 1))} \quad (2)$$

;

Step 7: Repeat it for every pixel of a frame and sum all distortions. Do it for every frame.

Algorithm 1: Taken from [1]

This model is used three video sequences with different characteristics, that for obtain values of distortion within an order of magnitude, and through the statistical method of normal distribution to be able to predict the threshold of video distortion streamed on WLAN with the same characteristics.

III. PROPOSAL

In the experiment with video sequences, we analyzed three characteristics of movement: the Highway little movement, the Container slow movements, and the Football high movement. Starting the initial proposal three results were found. Aiming model accuracy the commands that were a priori manually executed were structured as testbed for performing more automated way. The developed testbed has well-defined functions:

- Acquisition of server multimedia content and stores them in the client directory;
- Separation of video frames received in .mp4 format;
- Calculation of the distortion of the video;

In content acquisition is the step where the videos are available on the server and purchased via network transmission to the client. For this the VLC tool was used and were performed the following commands on the server:

Listing 1. File cfgVLC.sh code

```
#!/bin/bash
vlc -I telnet --telnet--password videolan
--rtsp--host 0.0.0.0:5554

new container vod enabled
setup container input "file:///home/
gredes/mp4/container.mp4"

new highway vod enabled
setup highway input "file:///home/gredes/
mp4/highway.mp4"

new football vod enabled
setup football input "file:///home/gredes
/mp4/football.mp4"

telnet localhost 4212
```

The commands performed on the client are purchased through the IP of the server and the number of required tests, in this case 542. The transmitted sequences are stored in their respective directories on the client computer, for later reading and processing, which can be seen in list 2.

Listing 2. File receiver.sh code

```
#!/bin/bash
num_testes=$1

mkdir football
mkdir container
mkdir highway

for ((j=1; j<$num_testes;j++))
do

cvlc rtsp://10.10.20.14:5554/football --
sout file/mp4:football/football$j.mp4
vlc://quit
```

```
sleep 1
cvlc rtsp://10.10.20.14:5554/container --
sout file/mp4:container/container$j.
mp4 vlc://quit
sleep 1
cvlc rtsp://10.10.20.14:5554/highway --
sout file/mp4:highway/highway$j.mp4
vlc://quit
sleep 1
done
```

After the pre-processing step, where the videos are stored on the client, the sequences must be separated into frames, which was carried through the ffmpeg tool. The list 3 is the executed script in the separation of the received video frames in mp4 format:

Listing 3. File ffmpeg.sh code

```
#!/bin/bash
for ((j=1; j<542;j++))
do

ffmpeg -i football/football$j.mp4
-r 25 -f image2
FramesFootball/fr$j/frame%01d.
jpg
sleep 1
ffmpeg -i container/container$j.
mp4 -r 25 -f image2
FramesContainer/fr$j/frame%01d.
jpg
sleep 1
ffmpeg -i highway/highway$j.mp4 -
r 25 -f image2 FramesHighway/
fr$j/frame%01d.jpg

done

exit
```

After executing these three files, the 542 directories with their respective frames for each video sequence, which will be the input parameter for the distortion. Because have generated several files on developing, only the most important will be exposed. File 4 gather the necessary information to start the distortion calculation which is the directory where are the frames and the number of frames that are in this directory because depending on the state of the network is possible to obtain different number of frames in each transmission.

Listing 4. File Main.m code

```
dir='FramesFootball';
for u=1:542
aux=[ './nfiles.sh ',num2str(u)];
[status,frames]=unix(aux);
nframes=str2num(frames);

distorcao = getframes(u,nframes,dir);
```

```

file=[ dir , '/fr ', num2str(u) , '/dist ',
      num2str(u) , '.mat '];
save(file , 'distorcao ');
end

```

The file nfiles.sh which running the ls command in the directory with the option wc-l to return the number of files inside the directory. With the number of frames the next step is the distortion calculation which is held by getframes function in file 5.

Listing 5. File getframes.m code

```

function result= getframes(t,nframes,dir)
dist = zeros(1,nframes);
for f=1:nframes-1
n = [ dir , '/fr ', num2str(t) , '/frame ',
      num2str(f) , '.jpg ']; % Converts the
      name string in frame n
disp('lendo ... ');
disp(n);
i = f+1;
m = [ dir , '/fr ', num2str(t) , '/frame ',
      num2str(i) , '.jpg ']; % Converts the
      name string in frame n+1
disp('lendo ... ');
disp(m);
dist(f) = procframes(n,m);%
      Processes Frames
end
result = dist;
end

```

As can be observed in file 5 has the procframes function where the distortion calculating is realized, in the function are defining the regions of interest through the filter. Within this function has the Apixel function where the distortion calculations are performed according to the algorithm, as may be seen in file 6.

Listing 6. File APixel.m code

```

function resLum = APixel(i,j,R,F)
r=zeros(1,5);
f=zeros(1,5);
r(1)=R(i,j);
r(2)=R(i+1,j);
r(3)=R(i-1,j);
r(4)=R(i,j+1);
r(5)=R(i,j-1);

L1= sum(r)/5;%Mean luminance in
      frame n
f(1)=F(i,j);
f(2)=F(i+1,j);
f(3)=F(i-1,j);
f(4)=F(i,j+1);
f(5)=F(i,j-1);

```

```

L2 = sum(f)/5;%Mean Luminance in
      frame n+1

aux=r(1)-f(1);
aux2=abs(L1-L2);
if aux <= 2*aux2
Res= abs(r(1)-f(1))/max(L1,
      L2);

end

resDist = Res;
end

```

An interesting feature of this implementation is that can access the distortion values between frames as this value is saved in a distortion vector with the same size of the number of frames in the video sequence of a given transmission.

IV. METHODOLOGY

To find a satisfactory prediction model was necessary to study the algorithm proposed in [1] and its adaptation to the video transmission environment. Each video is analyzed frame by frame, separating only the interest areas (objects edges) enclosed by Sobel filter. Example of regions bounded is the figure 1 which illustrates a frame and it result after filtering.



Fig. 1. Results of the Sobel filter

A. Test Environment

In this work, we used three sets of video sequences for comparison. These sets of sequences represent a range of video content typical for low and high latency. Were used all the frames of each video sequence for codification. The set is consists of three sequences with characteristics of low, medium and high motion respectively: Container, Highway and Football. These sequences are in CIF format (352 x 288), and in 25 fps frame rate. Besides the videos used in the paper has lasts of 13 ", 1'19" and 9 "respectively, with a frame rate of 25 frames/second, which gives us a value of approximately 275, 1975 and 225 frames.

The test scenario in which the implementation was underwent, follows the steps defined by the algorithm, thus tests

were performed with three video each one of a modality. The video sequences used were obtained by Xiph.Org Foundation [7]. The videos are using MPEG-2, the tool used for manipulating the videos was FFmpeg and VLC.

The first step was to set up the environment for videos transmission, thus it was assembled using the software VLC which is a player of multimedia content able to provides them to other users on the network. A computer was configured as server, and others as client and thus transmitted in a period of normal use of the network.

The second step was to prepare the videos for the distortion calculation by separating the frames. As the distortion calculation is done by analyzing the video frame by frame, for both the three sequences were separated using the commands of the software ffmpeg, generating JPEG images in grayscale.

The third step was the development of the algorithm using the software Matlab, where some adaptations were made to converge to the expected result.

In carrying out the tests it was mounted a network scenario completely free, ie, without blocking data traffic so that it can come closer to an real environment of networks use.

This test scenario is composed of two computers, as can be seen in figure 2, where PC-1 acts as a video streaming server and PC-2 acts as the video receiver. It is noteworthy that the whole scenario was mounted based on free software, being so a scenario of accessible reproduction.

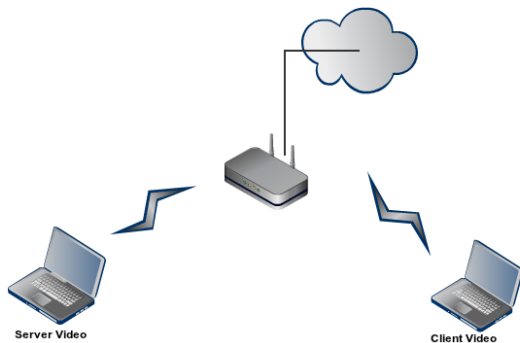


Fig. 2. Test scenario.

Various automation scripts were used for tests performed with minimally costly. In tests, the distortion average of videos MPEG transmitted over wireless network was observed. Also, for the prediction, we used statistical method with normal distribution, so that from a pre-sample of 10 transmissions for each type of video reach the value of a reliable sample. To calculate the sample size was broadcast the video sequences 10 times and reaching the value: 542 which was the value found for the Football video, that by having more movement has a high standard deviation and therefore more samples is needed, as other sequences the values resulted below, was adopted the main because is enough for the three sequences. In each test is performed 542 transmissions of each video sequences. We present the results in a confidence interval of 95 %.

V. RESULTS

All sample data were calculated based on the normal distribution. The prediction model is based on the distortion values obtained for each type of video sequence, based on these results it is possible to reach a confidence interval ie predicts that video with a particular motion characteristic has distortion within that range. Table 1 illustrates the tests performed with the sample of 542 transmissions, in the first column have the video sequence, in the second the tests average values, in the third and fourth maximum and minimum reached values, already in the fifth the standard deviation between the values founded.

TABLE I
RESULT DISTORTIONS

542 Transmissions				
Video	Mean	Minimum	Maximum	Standard Deviation
Container	189.44	0.00	600.27	1.57
Highway	255.13	0.00	1693.83	2.90
Football	1073.51	0.00	2725.62	28.18

From the tabulated data with a confidence interval of 95% referring to the normal distribution table has $z_{\alpha/2} = 1,96$. The confidence interval is given by the formula:

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (3)$$

According to the confidence intervals found, the model becomes able to predict the videos distortion with the same motion characteristic. Applying the formula 3 in the results of table 1 we reached at the following table of confidence intervals for each type sequence.

Analyzing the generated results we can observe the level of complexity of the implementing this metric of distortion calculation, for it to be executed in the most optimized way possible using the maximum of available resources. Also another fact that is worth mentioning was the runtime calculations and transmissions, which reached an average of 1 day for every 542 transmissions and 3 days for distortion calculation of each sequence. This runtime occurred by the volume of data to be analyzed, and because the metric analyze the videos frame by frame.

The video Container of little motion, the behavior of the metric was an average of 189.44 and a low standard deviation 1:57 the lowest among the three sequences. Already on the Highway with slow movements the average had an important value and standard deviation accompanied also with an increase. In the high motion sequence the results were very higher reaching up to 3 times higher compared to the Highway. The use of these three types of sequences aims to explore the metric in various types of videos, furthermore was the basis for the prediction of distortion in video sequences with the same characteristics.

VI. FINAL CONSIDERATIONS

This article investigated the distortion of videos streamed in WLAN. In the work, we propose a application scenario using

distortion image estimation algorithm previously proposed by the authors in [1]. Using this algorithm adapted to images sequences, a set of videos sequences were submitted to it for the distortion calculate. The first consideration observed in the used metric is that it is not sensitive to the network variables. An example of this can be observed if in a transmission of a sequence, the packet loss is very high. The results shows no effect of this packet loss in the metric, while the influence of the continuity and motions are noticed. The model developed here has opened doors to other challenges and improvements, that can be done with the evolution of the study to others videos sequences in a future work, in order to decrease the interval between video that has more motion for which has less, choosing intermediate videos, to further validate the model.

REFERENCES

- [1] C. de Farias and P. de Aguiar Rodrigues, "Objective metrics for measuring perceptual video quality," *17th International Conference on Systems, Signals and Image Processing - IWSSIP*, June 2010.
- [2] C. Yang, L. Zhao, and Z. Liao, "Objective quality metric based on perception for video," in *Computer Engineering and Technology, 2009. ICCET'09. International Conference on*, vol. 2. IEEE, 2009, pp. 20–23.
- [3] F. Babich, M. D'Orlando, and F. Vatta, "Distortion estimation algorithms for real-time video streaming: An application scenario," in *Software, Telecommunications and Computer Networks (SoftCOM), 2011 19th International Conference on*. IEEE, 2011, pp. 1–5.
- [4] Z. Chen and D. Wu, "Prediction of transmission distortion for wireless video communication: Analysis," *Image Processing, IEEE Transactions on*, vol. 21, no. 3, pp. 1123–1137, March 2012.
- [5] N. Chen, X. Jiang, C. Wang, and J. Su, "Study on relationship between network video packet loss and video quality," in *Image and Signal Processing (CISP), 2011 4th International Congress on*, vol. 1, Oct. 2011, pp. 282–286.
- [6] R. Gonzales and R. Woods, "Digital image processing," *Prentice-Hall*, vol. 1, p. 40, 2002.
- [7] X. Foundation. Xiph.org home page. [Online]. Available: <http://media.xiph.org/video/derf/>