

Avaliação de qualidade de vídeo sobre uma rede IP usando métricas objetivas

Dante COAQUIRA BEGAZO

Departamento de Engenharia de Sistemas Eletrônicos, Universidade de São Paulo (USP)
São Paulo, São Paulo 05508-010, Brasil

Demóstenes ZEGARRA RODRIGUEZ

Departamento de Engenharia de Computação e Sistemas Digitais, Universidade de São Paulo (USP)
São Paulo, São Paulo 05508-900, Brasil

Miguel ARJONA RAMÍREZ

Departamento de Engenharia de Sistemas Eletrônicos, Universidade de São Paulo (USP)
São Paulo, São Paulo 05508-010, Brasil

RESUMO

Atualmente, há uma ampla gama de serviços de telecomunicações que transmitem voz, vídeo e dados através de complexas redes de transmissão sendo que em alguns casos o serviço não tem uma qualidade aceitável pelo usuário final. Este trabalho apresenta de uma maneira prática de como um serviço de vídeo em uma rede IP pode ser afetado por condições adversas na rede, principalmente nos casos de perda de pacotes, atrasos e variação de atrasos. Para a realização dos testes foi montado um cenário de emulação de rede isolado, no qual são parametrizadas diferentes condições de rede. Considera-se então, um serviço de vídeo streaming para o qual são consideradas diferentes qualidades de vídeo a serem observadas pelo usuário.

Palavras Chaves: Qualidade de Vídeo, Métricas Objetivas de Vídeo, FR, PSNR, SSIM, VQM.

1. INTRODUÇÃO

A avaliação da qualidade de vídeo é realizada por métricas objetivas e subjetivas. As métricas subjetivas são baseadas na percepção humana. As métricas objetivas são baseadas em modelos matemáticos para estimar a média das opiniões dos usuários. As métricas objetivas podem ser classificadas de acordo com a disponibilidade do vídeo original, que é considerado livre de distorção ou de alta qualidade: FR (Full Reference) Referência Completa [3] quando o sinal de vídeo de referência não degradado está totalmente disponível, RR (Reduced Reference) Referência Reduzida [3] quando o sinal de vídeo original não é totalmente disponível, e os NR (No-Reference) Sem Referência [2] quando as sequências de vídeo de referência não são acessíveis, chamada de avaliação cega. O modelo de avaliação de vídeo no presente artigo é do tipo FR [3], porque utilizamos duas entradas (vídeo original e vídeo degradado) para obter a qualidade de vídeo observada pelo usuário (observador).

Para comparar o vídeo original e o vídeo degradado utilizamos as métricas objetivas: PSNR (Peak Signal to Noise Ratio) Relação Sinal Ruído de Pico [4] [5] [6], SSIM (Structural Similarity Index) Índice de Similaridade Estrutural [4] [5] [6] e VQM (Video Quality Measure) Medição de Qualidade de Vídeo [4] [6].

Métrica Objetiva PSNR

A PSNR é uma métrica amplamente utilizada pela comunidade científica, define a relação entre a máxima energia possível de um sinal e o ruído que afeta a representação do sinal entre os quadros de vídeo original e de vídeo degradado. A métrica PSNR é representada pela seguinte equação:

$$PSNR = 10 \log_{10} \frac{L^2}{\frac{1}{MNT} \sum_{m=1}^M \sum_{n=1}^N \sum_{t=1}^T [I(m,n,t) - \hat{I}(m,n,t)]^2} \quad (1)$$

em que L é a faixa dinâmica dos valores de pixel, M e N são a largura e a altura das sequências de vídeo original e vídeo degradado respectivamente, T é o número de quadros que contém as sequências, $I(m,n,t)$ e $\hat{I}(m,n,t)$ representam o pixel na posição (m,n) do quadro t para a sequência original e degradada, respectivamente. Para um sinal de 8 bits/pixel, L é igual a 255.

A métrica PSNR tem uma faixa de valores que varia de 0 a 100, sendo 100 o melhor valor possível.

Métrica Objetiva SSIM

O SSIM é uma métrica que compara cada quadro das sequências de vídeo original e vídeo degradado para quantificar a degradação sofrida. Se baseia na ideia de que imagens naturais são altamente estruturadas, em outras palavras, seus pixels apresentam forte dependência, especialmente quando estão espacialmente próximos, as dependências carregam informações importantes sobre a estrutura dos objetos na cena. O algoritmo SSIM estima a similaridade entre o vídeo original e o vídeo degradado, comparando a luminosidade $l(x,y)$, contraste $c(x,y)$ e estrutura $s(x,y)$ do vídeo original x e do vídeo degradado y de tamanho $N \times N$, utilizando as seguintes equações:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (3)$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (4)$$

em que μ_x e μ_y é a média de x e y , σ_x^2 e σ_y^2 é a variância de x e y respectivamente, σ_{xy} é a covariância de x e y , C_1 , C_2 e C_3 são

Dante Coaquira Begazo, Demóstenes Zegarra Rodríguez e Miguel Arjona Ramírez, E-mails: dante@lps.usp.br, demostenes@lps.usp.br, miguel@lps.usp.br. Este trabalho foi financiado pela CNPq (307633/2011-0) e pela FAPESP (2009/18242-5) e (2007/08288-2).

constantes que estabilizam a divisão com o denominador pico, evitando o surgimento de instabilidades quando o denominador tende a zero. Essas constantes são dadas por:

$$C_1 = (K_1 \cdot L)^2 \quad (5)$$

$$C_2 = (K_2 \cdot L)^2 \quad (6)$$

$$C_3 = \frac{C_2}{2} \quad (7)$$

em que L é a faixa dinâmica dos valores de *pixel*, para um sinal de 8 bits/*pixel*, L é igual a 255, porque $L = 2^{\#bits/pixel} - 1$, $K_1 \ll 1$, e $K_2 \ll 1$, comumente $K_1 = 0.01$ e $K_2 = 0.03$. A métrica SSIM é representada pela seguinte equação:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (8)$$

em que α , β e γ são parâmetros que definem a importância relativa das componentes de luminância, contraste e estrutura, respectivamente. Se $\alpha = \beta = \gamma = 1$, a equação acima se simplifica a:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (9)$$

A métrica SSIM tem uma faixa de valores que varia de 0 a 1, sendo 1 o melhor valor possível.

Métrica Objetiva VQM

A VQM é uma métrica que está presente na recomendação J.144 [3] da ITU, e trata da avaliação objetiva de qualidade de vídeo digital. Esta métrica foi extensivamente testada pelo Video Quality Expert Group (VQEG). A métrica VQM usa o vídeo original e o vídeo degradado como entradas para avaliar a qualidade de vídeo. O algoritmo VQM pode ser dividido nas seguintes fases:

Calibração: nesta fase inicial tem-se por objetivo calibrar o vídeo em preparação para a fase de extração de características. Com esta proposta, são estimados e corrigidos, juntamente com os deslocamentos espaciais e temporais, os deslocamentos de contraste e brilho da sequência de vídeo degradada em relação a sequência de vídeo original;

Extração de características de qualidade: nesta fase, um conjunto de características de qualidade que descreve as mudanças na percepção espacial, temporal e de crominância é extraído de sub-regiões espaço-temporais da sequência de vídeo (original e degradada). Esta fase segue os seguintes passos:

- 1) Aplicar um filtro perceptivo, com a finalidade de realçar alguns aspectos de qualidade de vídeo. A componente de luminância Y contém informações das bordas, utilizando o filtro receptivo pode-se identificar com maior precisão: o embaçamento, o efeito de bloco e outros efeitos de borda. As componentes de cor C_b e C_r são úteis para identificar distorções de matiz e erros de transmissão. A diferenciação no tempo de componentes Y de quadros consecutivos destaca movimentos não naturais e bruscos.
- 2) Dividir a sequência de vídeo em sub-regiões espaço-temporais. Cada sub-região espaço-temporal descreve um bloco de pixels $b(s, t)$ com dimensão descrita pelo número de *pixels* na horizontal, o número de linhas de quadros na

vertical, e o tempo de duração da região, dado em unidades de quadros de vídeo.

- 3) Extrair características de cada região. As características $f_o(s, t)$ e $f_d(s, t)$ do sinal original e degradado, são extraídas de cada um dos blocos $b(s, t)$ utilizando funções matemáticas simples, como a média e o desvio padrão. Após a extração de características, o eixo temporal t já não corresponde aos quadros individuais, e contém um número de amostras igual ao número de quadros na sequência de vídeo dividido pela extensão temporal dos blocos.
- 4) Aplicar um limite de perceptibilidade. Uma função de corte pode ser aplicada para reduzir a sensibilidade do parâmetro às distorções imperceptíveis. A função de corte é representada pela seguinte forma:

$$f_{dip} = \begin{cases} \max(f, T) & \text{se } f \text{ é positivo} \\ \min(f, T) & \text{se } f \text{ é negativo} \end{cases} \quad (10)$$

em que T é o limiar de corte.

Estimativa de parâmetros de qualidade: Nesta fase, um conjunto de parâmetros de qualidade $p(s, t)$ que descrevem as mudanças de percepção é calculado comparando características extraídas da sequência de vídeo degradado $f_d(s, t)$ com aquelas extraídas da sequência de vídeo original $f_o(s, t)$. Os parâmetros $p(s, t)$ dos blocos $b(s, t)$ formam matrizes tridimensionais abrangendo o eixo temporal e duas dimensões espaciais, ou seja, horizontal e vertical;

Estimativa da medida de qualidade: Nesta fase final tem-se como objetivo calcular uma métrica de qualidade global através de uma combinação linear dos parâmetros calculados nas fases anteriores.

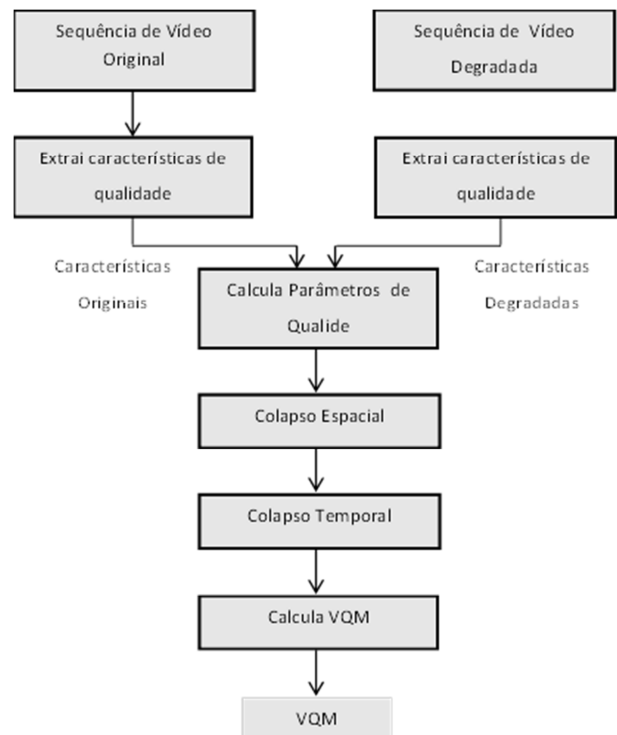


Figura 1. Diagrama de blocos do modelo VQM.

A métrica VQM tem uma faixa de valores que varia de 0 a 1, sendo 0 o melhor valor possível.

2. METODOLOGIA

A base de dados utilizada no artigo foi obtida do VQEG (Video Quality Experts Group) [7]. Os vídeos são denominados de SRC alguns dos quais são comprimidos usando MPEG-2 para taxas de compressão diferentes, as características do vídeo são descritas (número de quadros por segundo, resolução, taxa de bit, espaço de cor e bits por pixel) para realizar testes. O VQEG tem ferramentas de processamento de imagens como conversores de formato, que podem ser descarregados livremente.

Para realizar os testes se montou um cenário de rede totalmente controlado, ou seja, sem nenhum tráfego de dado adicional que possa influenciar nas avaliações realizadas. Este cenário de teste esta composto de 3 PC's, como pode ser observado na Figura 2, onde a PC-A atua como um servidor Web. O PC-B emula um canal de transmissão, sendo aqui configurados diferentes cenários de rede baseados principalmente nos fatores de degradação da rede: perda de pacotes e atraso ponto a ponto. O PC-C é um usuário que utiliza o servidor Web para descarregar arquivos de vídeo e onde também roda os softwares de avaliação objetiva de qualidade de vídeo e softwares de gravação e edição de vídeo.

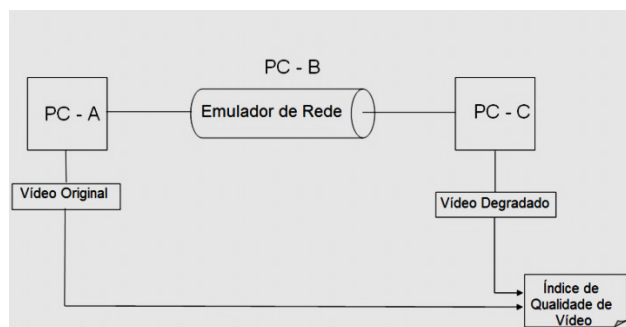


Figura 2. Cenário de teste para um modelo FR.

O modelo de avaliação de vídeo empregado neste trabalho é do tipo full reference, pois para a determinação da qualidade do vídeo observado pelo usuário são utilizadas duas entradas: o arquivo de vídeo original e o arquivo de vídeo degradado.

Os softwares que rodam em cada um dos PC's são descritos a seguir: Apache Web Server [8], software livre para administração de servidores web onde, para o cenário descrito, são alocados os arquivos de vídeos que serão visualizados pelo usuário. NISTNET [9], emulador de rede, onde pode-se configurar os fatores de degradação da rede como: probabilidade de perda de pacote, atraso ponto a ponto, variação de atraso, largura de banda. Adicionalmente funciona como um roteador. NETEM [10], software emulador de rede, onde pode-se configurar os fatores de degradação da rede como: probabilidade de perda de pacotes e atraso. CamStudio [11], software de gravação de vídeo, onde a gravação é realizada em tela cheia. VideoPad Vídeo Editor [12], já que a realização dos testes é realizado de forma manual e por não contar com uma sincronização exata, os vídeos são tratados com a finalidade de ter a mesma duração. MSU Video Quality Measurement Tool [13], é uma ferramenta para determinar qualidade de vídeo. Suporta diferentes formatos de vídeo com entrada como: avi, avs, yuv, bmp, mov, wmv, mp4, mpeg, flv e outras.

Na Figura 3 é representada a configuração dos endereços IP na rede emulada utilizada neste trabalho.

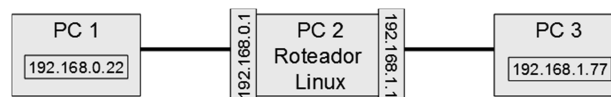


Figura 3. Endereços IP configurados na rede emulada.

É importante ressaltar que o cenário foi montado em base de softwares livres, sendo um cenário de fácil reprodução.

Os vídeos utilizados no artigo tem uma duração de 11.6 segundos, com uma taxa de quadros de 25 quadros/segundo, o que nos dá um valor de 290 quadros, com resolução de 1024x576 e com uma taxa de transferência de 3072kbps.

3. RESULTADOS

Pode-se determinar segundo a Tabela 1 que existem perdas em nosso sistema, mesmo que não apresente alterações nos fatores de degradação da rede IP implementada, com valores de 0 para perda de pacotes e para atraso.

Perda de Pacotes (%)	Atraso (ms)	Y PSNR Média	Y SSIM Média	Y VQM Média
0	0	22.12219	0.81124	6.05202
0	1	22.12219	0.81124	6.05208
0	2	22.12219	0.81124	6.05207
0	3	22.12244	0.81108	6.05227
0	4	22.12219	0.81124	6.05209
0	5	22.12222	0.81125	6.05198
0	6	22.12218	0.81125	6.05204
0	7	22.12219	0.81125	6.05205
0	8	22.12217	0.81123	6.0521
1	0	22.11998	0.81117	6.05331
2	0	22.12219	0.81124	6.05207
3	0	22.1222	0.81125	6.05202
4	0	22.12208	0.81121	6.05225
5	0	22.12219	0.81124	6.05205
6	0	22.12221	0.81125	6.05203
7	0	22.12222	0.81125	6.05201

Tabela 1. Médias das métricas.

A Tabela 1 representa a medição do quadro (frame) 0, na comparação do vídeo original com o vídeo degradado para as diferentes métricas objetivas, especificamente para o componente de luminância Y. Os valores das médias para quadros iguais são 100 para a métrica PSNR, 1 para a métrica SSIM e 0 para métrica VQM. No caso das métricas PSNR e SSIM valores altos são melhores, e para a métrica VQM valores baixos são melhores.

Pode-se observar, conforme a Tabela 1, que na medida em que o atraso é incrementado a média das métricas PSNR e SSIM diminui e a média da métrica VQM aumenta.

Na Figura 5 é apresentado o resultado que utiliza a métrica PSNR para a componente de luminância Y. As cores para PSNR são apresentadas em ordem crescente: vermelho, amarelo, verde, azul e preto, a maior valor da média da PSNR, menor é a diferença. Na Figura 6 é apresentado o resultado que utiliza a métrica SSIM para a componente de luminância Y, onde as áreas com maior brilho correspondem à maior diferença. Na Figura 7 é apresentado o resultado que utiliza a métrica VQM para a componente de luminância Y, onde os blocos com maior brilho correspondem à maior diferença.



Figura 7. Imagem resultado da comparação utilizando VQM.

As diferentes métricas objetivas apresentam médias com muitas diferenças. A que obteve melhor resultado ocorre quanto menores forem a perda de pacotes e o atraso.

A qualidade de vídeo é menor em função do aumento do atraso e da perda de pacotes, conforme os resultados mostrados nas Figuras 8a e 8b para a métrica PSNR e também nas Figuras 9a e 9b para a métrica SSIM. O fator de degradação de atraso apresenta uma maior influência no resultado de qualidade de vídeo.

O gráfico da métrica VQM nas Figuras 10a e 10b é interpretado diferentemente nas outras métricas, ou seja, enquanto esta aumenta com o aumento de número de quadros as outras diminuem.



Figura 4. Imagem do vídeo original.

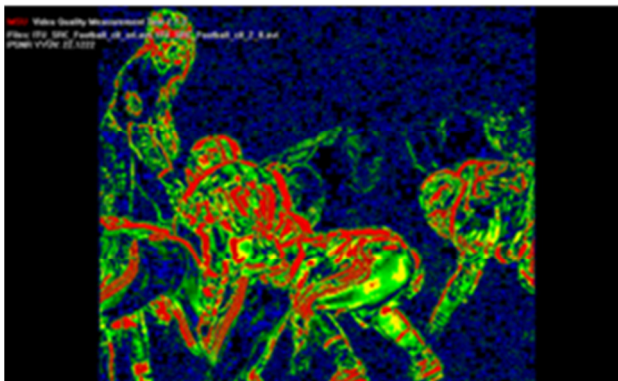


Figura 5. Imagem resultado da comparação utilizando PSNR.



Figura 6. Imagem resultado da comparação utilizando SSIM.

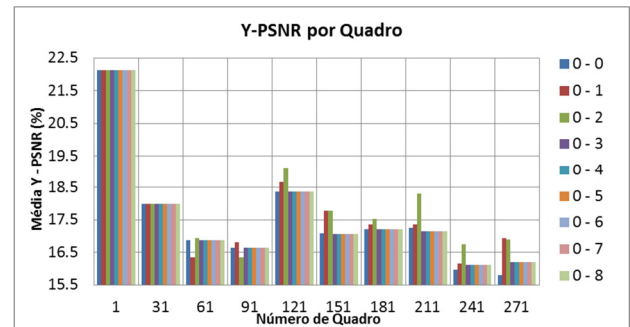


Figura 8a. Gráficos da métrica PSNR por número de quadro para o atraso.

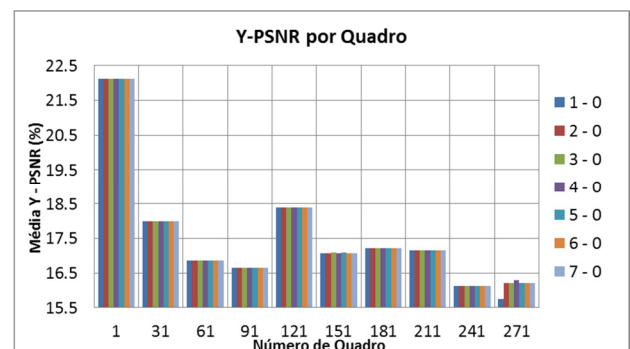


Figura 8b. Gráficos da métrica PSNR por número de quadro para a perda de pacotes.

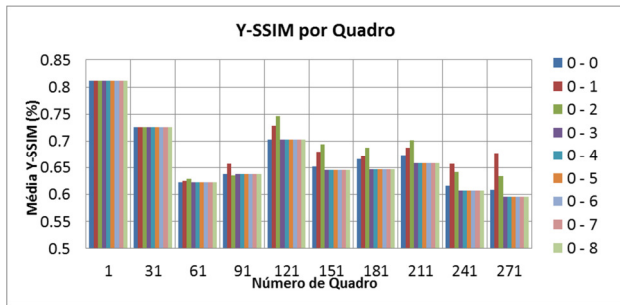


Figura 9a. Gráficos da métrica SSIM por número de quadro para o atraso.

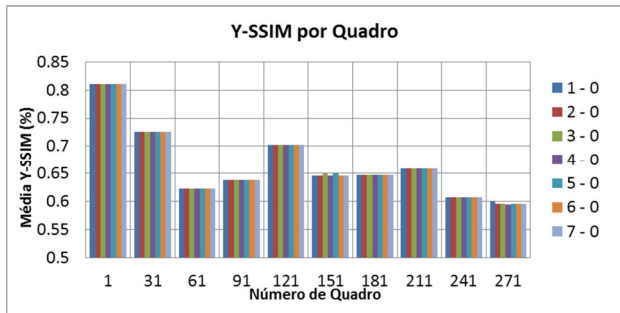


Figura 9b. Gráficos da métrica SSIM por número de quadro para a perda de pacotes.

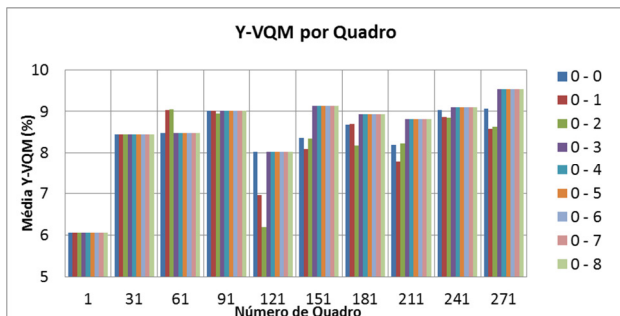


Figura 10a. Gráficos da métrica VQM por número de quadro para o atraso.

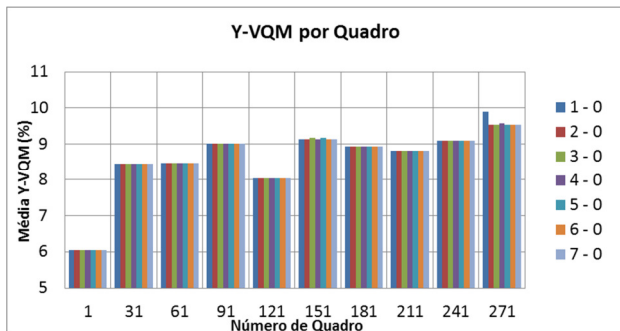


Figura 10b. Gráficos da métrica VQM por número de quadro para a perda de pacotes.

4. CONCLUSÕES

Nós chegamos na conclusão de que a qualidade de vídeo é afetada diretamente pelos fatores de degradação de rede como o atraso e a perda de pacotes, com uma maior influência do primeiro, as diferentes métricas utilizadas são ferramentas que ajudam a estabelecer uma melhora no serviço de vídeo numa rede IP. A rede IP implementada apresenta perdas na transmissão mesmo sem fazer as alterações dos fatores de degradação de perda de pacotes e de atraso. O fator de degradação de perda de pacotes, não apresenta variação significativa já que o sinal de vídeo utilizado foi de curta duração.

Em trabalhos futuros temos como objetivo fazer a avaliação usando métricas subjetivas e acrescentar o número de fatores de degradação.

5. REFERÊNCIAS

- [1] F. Xue, V. Markovski and L. Trajkovic, **Packet Loss in Video Transfers over IP Networks**, IEEE Conference on Circuits and Systems, vol. 2, pp. 345-348, May 2001.
- [2] Y. Tian and M. Zhu, **Analysis and Modelling of No-Reference Video Quality Assessment**, IEEE Conference on Computer and Automation Engineering, pp. 108-112, March 2009.
- [3] Z. Wang, H. R. Sheikh and A. C. Bovik, **Objective Video Quality Assessment**, The Handbook of Video Databases: Design and Applications, ch. 41, pp. 1041-1078, September 2003.
- [4] S. Pasqualini, F. Fioretti, A. Andreoli and P. Pierleoni, **Comparison of H.264/AVC, H.264 with AIF, and AVS based on different video quality metrics**. IEEE Conference on Telecommunications, pp. 190-195, May 2009.
- [5] A. Horé and D. Ziou, **Image quality metrics: PSNR vs. SSIM**, IEEE Conference on Pattern Recognition, pp. 2366-2369, August 2010.
- [6] M. Vranjes, S. Rimac-Drlje and K. Grgic. **Locally Averaged PSNR as a Simple Objective Video Quality Metric**, IEEE Conference ELMAR, vol. 1, pp. 17-20, September 2008.
- [7] Video Quality Experts Group, <http://www.its.bldrdoc.gov/vqeg>
- [8] The Apache Software Foundation, <http://www.apache.org>
- [9] National Institute of Standards and Technology, <http://www-x.antd.nist.gov/nistnet>
- [10] Linux Fundantion, <http://www.linuxfundantion.org/collaborate/>
- [11] CamStudio Open Source, <http://www.camstudio.org>
- [12] NCH Software, <http://www.nchsoftware.com/videopad>
- [13] Graphics & Media Lab (Video Group), http://compression.ru/video/quality_measure