

Métrica Pouco Referenciada para o Monitoramento de Sistemas de Transmissão de Vídeo da 4^a Geração de Comunicações Móveis

Marcio N. Bispo, Carla L. Pagliari e J. Roberto Boisson de Marca

Resumo—A transmissão de vídeo em tempo real sobre redes sem fio é desafiante devido ao canal rádio-móvel e às características desta mídia. A 4^a Geração de Comunicações Móveis (4G) promete amenizar esta tarefa ao criar a possibilidade de escolha da melhor rede dentre as disponíveis em determinada área. Esta escolha não é trivial, devido ao mapeamento entre características do canal móvel e qualidade de vídeo. Este mapeamento envolve diversos fatores, em particular a combinação entre a dependência entre quadros e o uso de ferramentas de resiliência a erros na presença de perdas de pacotes.

Este artigo propõe uma métrica que estima a distorção esperada de um vídeo devido a mudanças no canal móvel. A métrica apresenta boa precisão com ou sem o uso de ferramentas de resiliência a erros.

Palavras-Chave—Multimídia, Streaming de Vídeo, Monitoramento de Qualidade de Vídeo, 4^a Geração de Comunicações Móveis, Handoff Vertical.

Abstract—The wireless real-time video streaming is challenging due to the mobile radio channel itself and characteristics of this media. The Mobile 4th Generation technology (4G) promises to ease this task by creating the possibility of choosing the best network among all available in a given area. This selection is not trivial, due to the mapping between the channel characteristics and video quality. This mapping involves several factors, in particular the combination of inter-frame dependence and the use of error resilience tools in the presence of packet loss.

This article proposes a metric that estimates the hypothetical video distortion due to mobile channel variations. The metric presents good precision with or without the use of error resilience tools.

Keywords—Mobile Multimedia, Video Streaming, Video Quality Monitoring, 4G Wireless Communications, Vertical Handoff.

I. INTRODUÇÃO

A transmissão de vídeo em tempo real sobre redes de pacotes vem apresentando grande crescimento e espera-se que venha a ocupar parcela significativa do tráfego da Internet em futuro próximo. Outro fenômeno importante é o crescimento das tecnologias de redes móveis sem fio como integrantes das redes de acesso. As redes sem fio são notórias pelas questões relativas às grandes e imprevisíveis variabilidades do canal rádio-móvel. Tais questões podem ser minimizadas com o advento da 4^a geração das comunicações móveis (4G). A 4G será caracterizada pela capacidade de um terminal móvel

poder selecionar a melhor dentre as diversas redes sem fio disponíveis em determinada área [1]¹.

Neste cenário de transmissão de vídeo sobre redes sem fio, o conceito de "melhor" é mais complexo do que em cenários envolvendo a transmissão de outras mídias [2], devido a diversos fatores, tais como o emprego de técnicas de modulação e codificação de canal adaptativas, dependência entre quadros adjacentes e uso de ferramentas de resiliência e ocultamento de erro dos codecs (codificadores/decodificadores) modernos. A dependência entre quadros adjacentes, causada pelo uso comum de esquemas de estimação/compensação de movimento na codificação de vídeo [3], é o fator que mais acrescenta complexidade ao problema, porque pode fazer com que erros inseridos em uma parte do vídeo se propaguem para os quadros posteriores.

Para essa tarefa é necessária uma métrica de qualidade de vídeo que considere tais variáveis e possibilite não só a escolha da melhor interface sem fio disponível, mas também a melhor combinação, nesta interface, das ferramentas de resiliência e ocultamento de erros, de esquemas de modulação e de codificação de canal. O assunto "métrica de qualidade de vídeo" é uma questão importante e atual, existindo inclusive um grupo de especialistas a nível mundial, o Video Quality Experts Group — VQEG [4].

Uma das mais usadas métricas de qualidade objetivas, a MSE (Mean Square Error), provê um método de quantificação da qualidade de vídeo que opera sem a necessidade de uma avaliação humana subjetiva. Tal métrica descreve a distorção que uma seqüência de vídeo pode sofrer comparando-se quadro a quadro a seqüência original com a seqüência codificada recebida no destino. Assim a distorção sofrida pelo j -ésimo quadro pode ser calculada por:

$$D_j = \sum_{i=0}^{(N_1 \cdot N_2) - 1} (f_j^i - \tilde{f}_j^i)^2 / (N_1 \cdot N_2) \quad (1)$$

onde f_j^i e \tilde{f}_j^i indicam o i -ésimo pixel dentro do j -ésimo quadro com $N_1 \times N_2$ pixels de dimensão, antes e depois da codificação e transmissão, respectivamente. Outra métrica popular é a PSNR (Peak Signal-to-Noise Ratio) [3]:

Marcio N. Bispo, Carla L. Pagliari, Instituto Militar de Engenharia, Rio de Janeiro, Brasil, E-mails: bispo@ime.eb.br, carla@ime.eb.br; J. Roberto Boisson de Marca, Pontifícia Universidade Católica do Rio de Janeiro, Brasil, E-mail: jrbrm@cetuc.puc-rio.br.

¹O conjunto de ações que vai desde a descoberta das redes disponíveis em determinada área, a seleção da melhor rede e posterior mudança para um ponto de acesso desta rede selecionada (pressupondo que o terminal já esteja conectado a algum ponto de acesso) vem sendo chamado de handoff vertical.

$$PSNR_j = 10 \log_{10} \frac{255^2}{D_j} \quad (2)$$

Duas das maiores causas de degradação na transmissão de vídeo sobre redes de pacotes são a distorção introduzida pela codificação de fonte e as perdas de pacotes. Assim a distorção total pode ser representada por:

$$D_{total} = D_e + D_v \quad (3)$$

onde D_e e D_v representam, respectivamente, a distorção causada pelo processo de codificação e aquela devido às perdas de pacotes. Em [5] é mostrado que D_e e D_v podem ser considerados praticamente descorrelacionados.

O cálculo das métricas como a MSE e a PSNR necessitam do conhecimento completo da seqüência de vídeo original, o que não é viável em sistemas de transmissão de vídeo em tempo real.

Assim a MSE e a PSNR podem ser classificadas como métricas totalmente referenciadas. Ainda, de acordo com [6], esquemas de mensuração de qualidade em sistemas de vídeo e as métricas relacionadas podem ser classificadas em mais duas categorias: pouco referenciadas — esquemas ou métricas que necessitam de poucos parâmetros do vídeo original — e não referenciadas — quando não necessitam de nenhuma informação da seqüência original.

Assim, para sistemas de transmissão de vídeo em tempo real, uma métrica adequada deverá ser pouco ou não referenciada.

Uma das principais dificuldades de se projetar uma métrica pouco ou não referenciada é a dependência temporal entre quadros adjacentes. Esta dependência é criada pela estimação/compensação de movimento dos codecs de vídeo atuais. A estimação/compensação de movimento é empregada com a finalidade de se aproveitar a correlação temporal freqüente entre quadros adjacentes das seqüências de vídeo com o objetivo de se aumentar a compressão.

Um quadro codificado sem o emprego da estimação/compensação de movimento é chamado de quadro I e este tipo de codificação é chamado de modo intra².

Importante ressaltar que a estimação/compensação de movimentos deve ser usada com precaução em ambientes de transmissão sujeitos a erros, porque erros em blocos [3] de determinado quadro podem se propagar até encontrar um quadro que foi codificado no modo intra.

Em geral, para sistemas de transmissão de vídeo em tempo real, e para os sistemas de vídeo da 4G em particular, o projeto de uma métrica não referenciada ou pouco referenciada deverá necessariamente realizar um mapeamento não só das estatísticas dos canais sem fio em qualidade de vídeo, mas

também incluir neste mapeamento a dependência entre os quadros adjacentes e ferramentas de resiliência.

Tao [2] propôs uma métrica que consegue realizar um mapeamento das estatísticas do canal em qualidade de vídeo, levando em consideração a dependência entre quadros adjacentes. Esta métrica opera de forma relativa ou diferencial, possibilitando o conhecimento do ganho ou perda da qualidade da seqüência de vídeo caso o canal seja trocado por outro de diferentes características, sem a necessidade de se conhecer a distorção absoluta.

Baseado nesta filosofia de distorção relativa ou diferencial, neste artigo é proposta uma métrica chamada Distorção Diferencial — $D_{differential}$. Esta métrica permite o mapeamento entre o ganho ou perda em distorção que uma seqüência de vídeo pode sofrer e a PER (packet error rate) de um canal de transmissão. A $D_{differential}$ também incorpora os efeitos de mudanças da configuração das ferramentas de resiliência a erros tais como a inserção periódica e randômica de macroblocos [3] codificados no modo intra.

Para o desenvolvimento de tal métrica, diversas simulações de vídeo foram realizadas usando-se o padrão de codificação de vídeo H.264/AVC [7]. O H.264/AVC foi escolhido porque é considerado o estado da arte em termos de padrões de codificação de vídeo, além de possuir um desempenho superior a outros padrões tanto em redes com fio como sem fio [7], [8].

O restante deste artigo está organizado da seguinte forma: a Seção II apresenta trabalhos anteriores relacionados ao assunto, a Seção III apresenta a métrica proposta, a Seção IV apresenta detalhes das simulações realizadas e verifica a precisão da métrica proposta e a Seção V conclui o trabalho e sugere os próximos passos.

II. TRABALHOS RELACIONADOS

O assunto “métrica de qualidade de vídeo“ é uma questão importante e atual que está sendo discutida em âmbito mundial, existindo inclusive um grupo de especialistas que trata do assunto, o VQEG [4]. A questão de como medir a qualidade objetiva de um vídeo em um esquema não referenciado ou pouco referenciado em um sistema de transmissão de vídeo sobre um canal sujeito a erros, o foco deste artigo, foi tratada de diversas formas em vários trabalhos.

Em [6] Yang propôs um esquema não referenciado que utiliza marcas d'água inseridas nos quadros das seqüências de vídeo. Tais marcas d'água, sinais conhecidos, serão possivelmente corrompidas durante a transmissão. Na recepção a marca d'água é recuperada e sua qualidade medida, possibilitando a inferência da qualidade da seqüência de vídeo de interesse.

Babu, em [9], propõe um método de baixo custo computacional que permite a avaliação da distorção em quadros reconstruídos depois da transmissão da seqüência através da análise das distorções na imagem (denominadas artefatos), causados pela perda de pacotes. Partindo de um ocultamento de erros através da reposição de fatias corrompidas por fatias de mesma posição do quadro anterior, Babu observou que o comprimento dos artefatos introduzidos pelas perdas de pacotes é proporcional à distorção resultante. Assim a métrica

²A codificação usando-se a estimação/compensação de movimento é chamada de codificação no modo inter e os quadros gerados são conhecidos como quadros P — preditivos — ou B — bidirecionais [3]. Em muitos codecs de vídeo os quadros I, P e B são combinados em uma seqüência de quadros que forma o GOP (group of pictures). O GOP se inicia com um quadro I e termina imediatamente antes da ocorrência do quadro I seguinte, que marca o início do próximo GOP.

proposta por Babu quantifica o comprimento dos artefatos através da detecção de gradientes abruptos no blocos do quadro e, baseado neste comprimento, infere a distorção causada pelas perdas de pacotes.

Em [2] Tao propõe uma metodologia simples que permite a avaliação da qualidade de vídeo em um formato não referenciado através do uso das estatísticas de perdas de pacotes. Tal métrica leva em consideração as ferramentas de resiliência, ocultamento de erros e a estimação/compensação de movimento dos codecs de vídeo atuais.

Em seu trabalho, Tao inicialmente desenvolve uma forma de avaliar a distorção causada pela perda de pacotes em um canal, sendo esta perda caracterizada por \bar{n} e P_e , respectivamente o tamanho médio da salva de perdas e a probabilidade de ocorrência de um evento de perdas de pacotes de qualquer tamanho. A estimação desta distorção é realizada tomando-se como base a distorção média causada pela perda de um único pacote chamada de D_1 . D_1 depende de diversos fatores, tais como a implementação específica de cada codificador e decodificador, seqüência de vídeo e parâmetros de codificação. Para um típico conjunto codificador-decodificador H.264/AVC, Tao mostra que a distorção devido a um regime de perdas de pacotes que tenha \bar{n} e P_e como características pode ser representada como:

$$E[D_v] = s\bar{n}P_eLD_1 \quad (4)$$

onde D_v foi definida na Equação 3 e:

s - quantidade de fatias transportadas por um pacote;

L - quantidade de pacotes necessários para transportar um quadro.

A limitação da Equação 4 está na dificuldade de obtenção do fator D_1 . Desde que D_1 é a média das distorções causadas pela possível perda de cada quadro ou fatia de uma seqüência de vídeo, para o seu cálculo é necessário a decodificação da seqüência diversas vezes (uma vez para cada quadro ou fatia perdido), inviável em sistemas de vídeo em tempo real.

Para resolver este problema, Tao, em [2], propôs uma métrica chamada rPSNR. Esta métrica opera de forma relativa, possibilitando o conhecimento do ganho ou perda da qualidade da seqüência caso o canal seja trocado por outro de diferentes características. Tao definiu então:

$$rPSNR = 10\log_{10} \frac{255^2}{\{E[D_v]\}'} - 10\log_{10} \frac{255^2}{\{E[D_v]\}^o} \quad (5)$$

onde $\{E[D_v]\}^o$ é a distorção média causada por um canal de referência e $\{E[D_v]\}'$ é a distorção causada pelo novo canal que está sendo considerado. Aplicando a Equação 4 na Equação 5 resulta:

$$rPSNR = 10\log_{10} \frac{n^0 P_e^0}{n' P_e'} \quad (6)$$

É interessante destacar que as duas distorções comparadas na Equação 5 estão sendo combinadas em um formato multiplicativo. A rPSNR foi testada neste trabalho usando-se as simulações descritas na Seção IV.

A Figura 1 mostra alguns resultados selecionados destas simulações. Nesta figura pode ser visto a PSNR média da componente de luminância (YPSNR) [3] versus PER da seqüência Foreman. As diversas estruturas de GOP nomeadas como GOP = N são descrições de GOP's com o primeiro quadro do tipo I e os N-1 quadros restantes do tipo P. O GOP aberto se refere a uma seqüência de vídeo codificada com o primeiro quadro do tipo I e todos os outros quadros do tipo P.

Desde que a rPSNR opera de forma relativa, os testes executados buscaram estimar, partindo de um ponto (isto é, de uma determinada PER) das curvas descritas na Figura 1, outros pontos de diferentes PER usando a Equação 6.

Na Figura 2 pode ser visto a comparação da YPSNR obtida nas simulações (aqui, neste caso particular, selecionou-se a curva correspondente ao GOP aberto) e a YPSNR calculada através da rPSNR, tomando-se como referência a PER=2%. A estimação parece razoável. Repetindo o mesmo procedimento para a curva referente ao GOP = 5, pode ser visto, na Figura 3, que a estimação não é tão boa quanto à da Figura 2. Um exame na Figura 1 mostra que a rPSNR somente funciona bem no caso do GOP aberto.³

Um dos desafios da transmissão robusta de vídeo é a proteção da seqüência codificada em ambientes de transmissão sujeitos a erros. Em esquemas de vídeo que usam estimação/compensação de movimento, tal como o H.264, quadros I são normalmente seguidos de quadros P e B, dentro do GOP.

Porém, em ambientes de transmissão propensos a erros, quadros corrompidos em um GOP podem propagar o erro para todos os quadros seguintes até a próxima ocorrência de um quadro I. Assim, dependendo das características do ambiente de transmissão com erros, pode-se melhorar a qualidade do vídeo recuperado no receptor diminuindo-se a quantidade de quadros P dos GOP's, eliminando-se os quadros B e inserindo-se macroblocos codificados no modo intra nos quadros P remanescentes. Desta forma a alteração da estrutura do GOP pode ser considerada uma ferramenta de resiliência. A contrapartida da inserção de componentes da imagem codificados no modo intra é o aumento da quantidade de bits necessários à codificação de determinada seqüência ou a aceitação de uma maior distorção devido à codificação, caso esteja-se em um regime CBR (constant bit rate).

Portando, o uso de GOP's abertos (codificações que empregam somente o primeiro quadro no tipo I e o restante dos quadros do tipo P ou B) não é uma boa opção para a transmissão de vídeo sobre redes sem fio. Isto implica que a rPSNR apresenta utilização limitada.

III. MÉTRICA PROPOSTA

Uma abordagem mais adequada poderia ser o uso de um modelo aditivo ao invés do modelo multiplicativo usado na Equação 6. Assim este trabalho propõe uma forma relativa de calcular a distorção com o formato abaixo, definindo-se assim a métrica Distorção Diferencial ($D_{differential}$):

³Esta afirmação é baseada na observação de comportamentos similares das curvas YPSNR x PER nas diversas simulações realizadas com diversas seqüências de vídeo e diversas configurações de GOP.

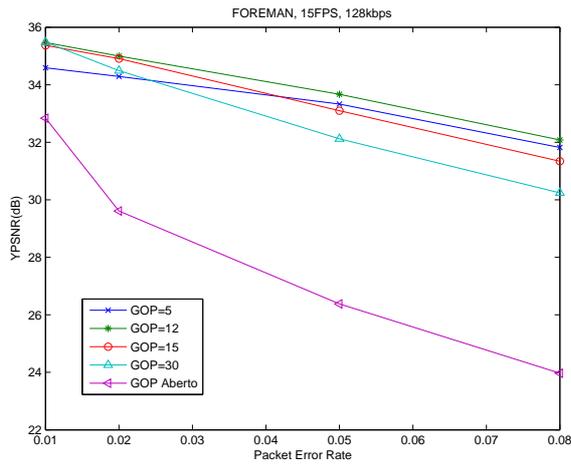


Fig. 1. Resultados da simulação: YPSNR x PER da seqüência Foreman, codificada a 15 fps, 128kbps, comparação de diversos configurações de GOP's

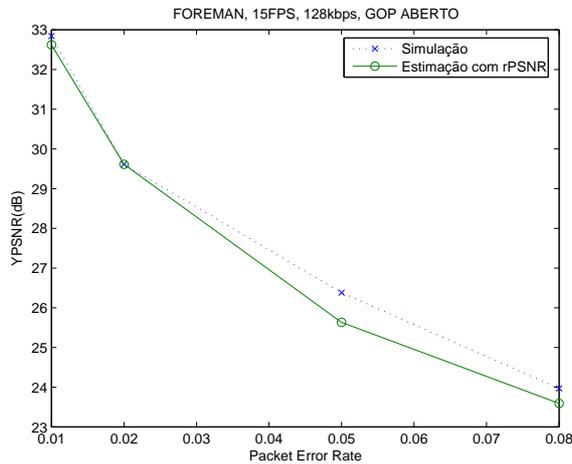


Fig. 2. Teste da rPSNR: YPSNR x PER da seqüência Foreman (15 fps, 128kbps, GOP aberto), simulação (linha pontilhada) e estimada pela rPSNR (linha cheia)

$$D_{differential} = \{E[D_v]\}^0 - \{E[D_v]\}' \quad (7)$$

É importante notar que o subscrito “v” mostra que esta distorção é resultante exclusivamente das perdas de pacotes. E $\{E[D_v]\}^0$ e $\{E[D_v]\}'$ se referem, respectivamente, à distorção média em um canal de referência e em um novo canal sendo testado. A principal questão é como fazer $\{E[D_v]\}^0$ e $\{E[D_v]\}'$ refletirem os parâmetros relevantes do canal de transmissão, codec e seqüência de vídeo em uma forma conveniente.

Em [5] He desenvolveu um esquema que permite a estimação da distorção média de uma seqüência de vídeo com base em poucos parâmetros do vídeo e do codec, além da PER do canal. Este esquema é baseado na premissa de que a distorção devido à perda de um pacote é função da diferença entre dois quadros consecutivos, pressupondo-se que o esquema de ocultamento de erro usado se dá através de

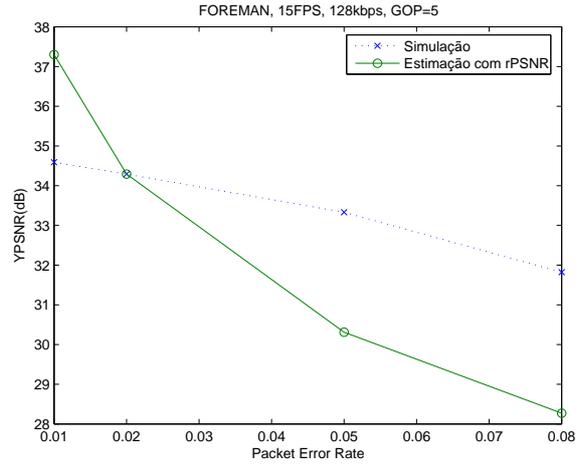


Fig. 3. Teste da rPSNR: YPSNR x PER da seqüência Foreman (15 fps, 128kbps, GOP = 5), simulação (linha pontilhada) e estimada pela rPSNR (linha cheia)

reposição de fatias corrompidas por fatias de mesma posição do quadro anterior⁴. He mostra ainda que o seu esquema permite a estimação da distorção com razoável acurácia.

Desde que o interesse é o cálculo da distorção causada por perdas de pacotes, em [5] é mostrado que:

$$E[D_v] = \frac{a}{(1-b+b\beta)} \cdot \frac{p}{1-p} E\{F_d(n, n-1)\} \quad (8)$$

onde:

$E\{F_d(n, n-1)\}$ - diferença média entre dois quadros consecutivos da seqüência de vídeo original (ainda não codificada).

p - PER (packet error rate)

a - razão entre diferença média entre dois quadros consecutivos da seqüência de vídeo codificada e $E\{F_d(n, n-1)\}$

b - variável relacionada à mobilidade das cenas do vídeo em questão⁵

β - taxa de inserção de macroblocos no modo intra⁶

É importante ressaltar que β reflete não só a estrutura do GOP como também outras ferramentas de resiliência, como por exemplo a inserção randômica de macroblocos no modo intra[7].

Este trabalho propõe o uso da Equação 8 para o cálculo da distorção presente na Equação 7, assim criando-se uma métrica batizada aqui de Distorção Diferencial — $D_{differential}$:

$$D_{differential} = \left[\frac{a^0}{(1-b^0+b^0\beta^0)} \cdot \frac{p^0}{1-p^0} - \frac{a'}{(1-b'+b'\beta')} \cdot \frac{p'}{1-p'} \right] E\{F_d(n, n-1)\} \quad (9)$$

⁴É interessante destacar que em [9] emprega-se o mesmo esquema de ocultamento de erros e que o uso deste esquema de ocultamento de erro é muito comum. Assim esta hipótese pode ser considerada bastante geral para boa parte das aplicações de transmissão de vídeo em tempo real.

⁵Esta variável está atualmente em investigação, dado que [5] oferece limitada informação sobre a mesma. Baseado nas simulações descritas na sessão IV, alguns valores de “b” foram estimados para possibilitar a obtenção de resultados.

⁶

$$\beta = \frac{nr \text{ total de macroblocos no modo intra}}{nr \text{ total de macroblocos}}$$

Uma análise detalhada das variáveis presentes na Equação 9 mostra que o único termo que não irá variar durante uma mudança de canal é $E\{F_d(n, n-1)\}$, desde que ele somente depende da seqüência de vídeo original. Este termo pode ser calculado no codificador com pouco esforço.

As outras variáveis podem variar ou não, dependendo da variação do canal ou das ferramentas de resiliência:

- β - irá mudar caso a quantidade de macroblocos do modo intra e inter variar;
- p - a PER do canal

Desde que "a", "b", β e $E\{F_d(n, n-1)\}$ possam ser calculados pelo codificador e transmitidos para o decodificador, a métrica descrita na Equação 9 pode ser classificada como pouco referenciada.

IV. RESULTADOS

Com a finalidade de se analisar o desempenho da $D_{\text{differential}}$, algumas simulações de transmissão de vídeo foram realizadas usando-se o esquema chamado Evalvid [10]. O Evalvid permite, em conjunto com o simulador de redes NS-2 [11], a simulação da transmissão de vídeo sobre topologias de rede arbitrárias ou canais caracterizados por modelos teóricos tais como modelos de Gilbert ou Markov. O codec H.264/AVC de referência empregado foi o JM 11.0[12].

As simulações foram realizadas usando-se um canal simples caracterizado por um modelo de erro de Bernoulli descrito através de uma PER (packet error rate), onde um erro em um pacote leva à perda do mesmo.

Um conjunto de seqüência de referência foi testado neste trabalho, todas no formato QCIF (176 x 144), a 15 fps (frames per second ou quadros por segundo). Estas seqüências apresentam diferentes características de textura e graus de movimento (baixo, médio e alto). Foram escolhidos, para servirem como base deste artigo, alguns resultados selecionados da seqüência Foreman, devido ao elevado grau de movimento presente nas cenas.

O codificador foi configurado para operar em modo CBR, nas seguintes taxas: 64kbps, 96kbps e 128kbps. Os resultados obtidos são bem similares, assim optou-se por mostrar aqui somente os resultados referentes a 128kbps.

Todos os GOPs usados se caracterizam pelo primeiro quadro ser do tipo I seguidos por quadros do tipo P até o próximo quadro I que caracteriza o início do GOP seguinte.

Os tamanhos de GOP escolhidos foram: 5, 12, 15, 30 e o GOP aberto. As PER escolhidas foram 1%, 2%, 5% e 8%. Com a finalidade de se construir uma base de dados com boa riqueza estatística, cada combinação {taxa do CBR, tamanho do GOP} foi decodificada diversas vezes para cada PER. O número que cada combinação {taxa do CBR, GOP, PER} foi decodificada foi função da PER: por exemplo, para a PER de 1%, a seqüência foi decodificada 200 vezes, para a PER de 2%, 100 vezes, e assim por diante.

A Figura 1 mostra resultados selecionados, onde tem-se a PSNR da componente Y (luminância) versus PER.

Com o objetivo de se testar a capacidade de estimação da $D_{\text{differential}}$, o primeiro teste realizado será a estimação

esperada caso o vídeo seja transmitido por canais com diversas PER, sem nenhuma alteração nas características das ferramentas de resiliência a erro. O alvo será a estimação da curva correspondente ao GOP = 5 da Figura 1, partindo-se do ponto com PER = 2% desta mesma curva como referência.

A estimação da PSNR pode ser vista na Figura 4 na linha cheia. Nesta mesma figura a linha pontilhada é o resultado das simulações. O resultado parece ser razoável para todos os valores de PER selecionados, com um erro menor que 1dB. Importante dizer que a única variável da Equação 9 que muda de um canal para outro é o "p".

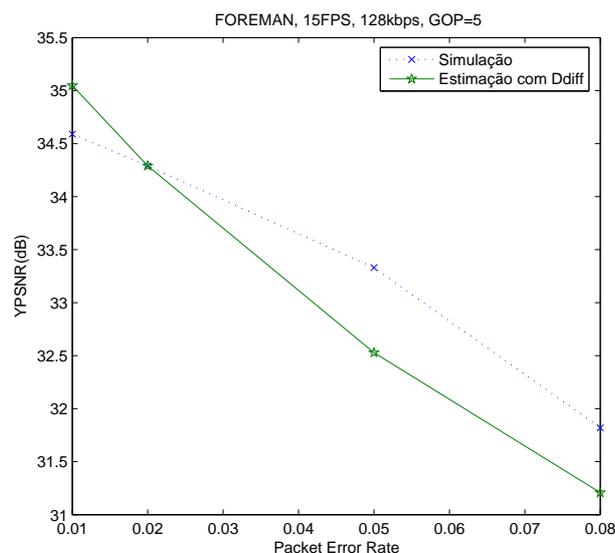


Fig. 4. Teste da capacidade de estimação da $D_{\text{differential}}$: YPSNR x PER da seqüência Foreman (15 fps, 128kbps, GOP = 5), simulada (linha pontilhada) e estimada pela $D_{\text{differential}}$ (linha cheia), usando um canal com PER = 2% como referência.

A seguir a nova medida foi utilizada para estimar a distorção para um GOP diferente do teste anterior, ainda sem alteração das ferramentas de resiliência. Assim o alvo é a estimação da curva correspondente ao GOP = 30 da Figura 1, usando-se a PER = 2% como referência. A Figura 5 mostra o resultado, onde poder-se ver que o erro de estimação é menor do que 0,5dB.

Finalmente foi testada a capacidade de estimação da métrica no caso de variação da configuração das ferramentas de resiliência a erros. Mais especificamente, a variação da quantidade de macroblocos no modo intra, implementada neste exemplo através da variação da estrutura do GOP. O alvo é a estimação da curva correspondente ao GOP = 15 da Figura 1, usando-se a curva correspondente ao GOP = 5 e seu ponto de PER = 2% como referência.

A Figura 6 mostra o resultado, onde pode ser visto que o erro de estimação é menor que 1dB. Importante dizer que a única variável da Equação 9 que muda da situação representada pelo GOP = 5 para o GOP = 15 é o β .

Um cenário que poderia se beneficiar do esquema mostrado na Figura 6 seria a necessidade de se saber, dado que existem outros canais com diferentes características de erro, qual seria a distorção resultante da combinação do handoff para

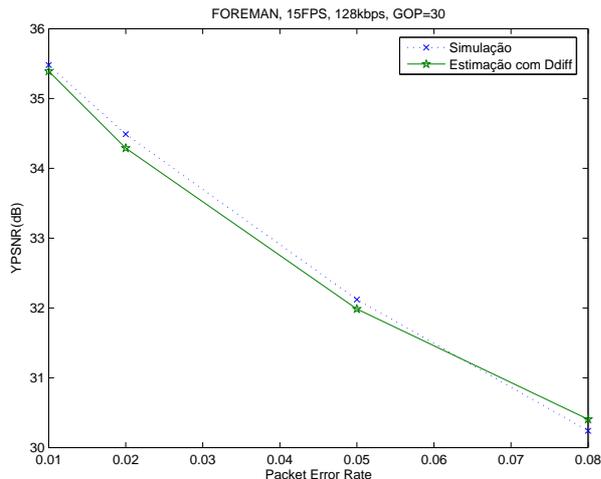


Fig. 5. Teste da capacidade de estimação da $D_{\text{differential}}$: YPSNR x PER da seqüência Foreman (15 fps, 128kbps, GOP = 30), simulada (linha pontilhada) e estimada pela $D_{\text{differential}}$ (linha cheia), usando um canal com PER = 2% como referência.

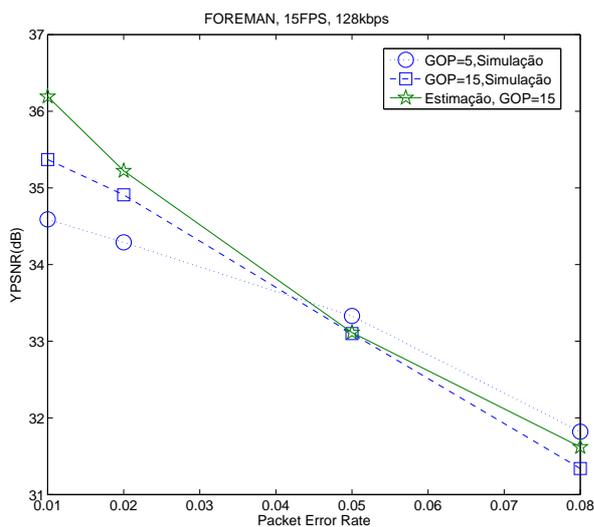


Fig. 6. Teste da capacidade de estimação da $D_{\text{differential}}$ para GOP's diferentes: YPSNR x PER da seqüência Foreman (15 fps, 128kbps), simulada com GOP = 5 (linha pontilhada), simulada com GOP = 15 (linha tracejada) e estimada, para o GOP = 15, pela $D_{\text{differential}}$ (linha cheia), usando um canal com PER = 2% e o GOP = 5 como referência.

um destes novos canais mais a alteração da percentagem de macroblocos codificados no modo intra. Assim poderia-se escolher, por exemplo, para cada PER, o tamanho de GOP que maximizasse a qualidade de vídeo.

V. CONCLUSÕES E TRABALHOS FUTUROS

Este artigo propôs uma nova métrica chamada Distorção Diferencial — $D_{\text{differential}}$ — para a estimação da qualidade de vídeo em um ambiente de propagação propenso a erros. Foi mostrado que a $D_{\text{differential}}$ é capaz de prever a diferença em qualidade que uma seqüência de vídeo experimentará no caso de mudança das características do canal de transmissão

e das ferramentas de resiliência a erros e estrutura do GOP, com precisão melhor que 1dB.

Trabalhos futuros expandirão a $D_{\text{differential}}$ procurando a combinação desta com algum método de estimação da distorção devido à codificação de vídeo, a fim de incluir a variação da taxa de bits dos diversos canais de transmissão.

VI. AGRADECIMENTOS

O primeiro autor agradece o apoio do Instituto Militar de Engenharia. O segundo autor agradece o apoio da FINEP sob o convênio FINEP/FAPEB no. 2645/06 do Centro Tecnológico do Exército e do Instituto Militar de Engenharia. O terceiro autor agradece o apoio do CNPq.

REFERÊNCIAS

- [1] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Communications*, pp. 49–55, Feb 2003.
- [2] S. Tao, J. Apostolopoulos, and R. A. Guerin, "Real-time monitoring of video quality in IP networks," in *Proceedings of NOSSDAV*, pp. 129–134, 2005.
- [3] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia*. Wiley, 2003.
- [4] <http://www.its.bldrdoc.gov/vqeg/>.
- [5] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, pp. 511–523, Jun 2002.
- [6] Y. Fu-zheng, W. Xin-dai, C. Yi-lin, and W. Shuai, "A no-reference video quality assessment method based on digital watermark," in *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, vol. 3, pp. 2707–2710, Sep 2003.
- [7] T. Wiegand, G. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560–576, Jul 2003.
- [8] T. Stockhammer, M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, p. 657–673, Jul 2003.
- [9] R. V. Babu, A. Bopardikar, A. Perkis, and O. I. Hillestad, "No-reference metrics for video streaming applications," in *Proceedings of International Packet Video Workshop*, Dec 2004.
- [10] www.tkn.tu-berlin.de/research/evalvid.
- [11] <http://www.isi.edu/nsnam/ns/>.
- [12] <http://iphome.hhi.de/suehring/tml/>.