Arquitetura de resolução mista para codificação distribuída de sequências *multiview*

Diogo C. Garcia, Camilo C. Dórea, Bruno Macchiavello, Ricardo L. de Queiroz e Debargha Mukherjee

***Resumo*— Este trabalho apresenta uma nova arquitetura de codificação distribuída para sequências *multiview*, baseada no padrão H.264/AVC e em quadros de resolução mista. A arquitetura permite transferir a complexidade da codificação para a decodificação, sendo particularmente adequada para aplicações de baixo consumo de energia, como em sistemas de segurança com múltiplas câmeras para uma mesma cena. Explora-se a correlação espacial e temporal na decodificação para melhorar a qualidade final do vídeo. Os resultados demonstram potencial de ganho de qualidade objetiva em relação à codificação independente de cada vista, sem custo adicional de taxa.**

***Palavras-Chave*— Codificação distribuída de vídeo, múltiplas vistas, informação lateral.**

***Abstract*— This work presents a new distributed multiview coding architecture, based on the H.264/AVC standard and mixed resolution frames. It allows the transferring of complexity from the coder to the decoder, which is particularly suited for low-power applications, such as multiview surveillance systems. Greater quality sequences are generated by exploiting the spatial and temporal correlation at the decoder. The results show potential for objective quality improvement over simulcast coding, with no extra rate cost.**

***Keywords*— Distributed video coding, multiview, side information.**

I. INTRODUÇÃO

Nos últimos anos, diversos esforços têm sido dedicados ao desenvolvimento de técnicas de codificação distribuída de vídeo, em inglês *distributed video coding* (DVC) [1]. O principal objetivo de tais técnicas é reduzir a complexidade presente no codificador, transferindo para o decodificador a responsabilidade de explorar as correlações entre quadros. Nas técnicas convencionais de codificação, como o padrão de codificação de vídeo H.264/AVC [2], operações de estimação e compensação de movimento são usadas durante a codificação para atingir maiores ganhos de compressão. Por outro lado, nas arquiteturas DVC, os quadros são codificados de maneira independente, reduzindo assim a complexidade. No entanto, taxas de compressão similares às taxas obtidas através de codificação convencional podem ser atingidas quando os quadros são decodificados de maneira conjunta. As justificativas para tais resultados podem ser encontradas nos teoremas de Slepian-Wolf [3] e Wyner-Ziv [4].

Diogo C. Garcia e Ricardo L. de Queiroz¸ Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, Brasil. E-mails: {diogo,queiroz}@image.unb.br. Camilo C. Dórea e Bruno Macchiavello, Departamento de Ciência da Computação, Universidade de Brasília, Brasília, Brasil. E-mails: {camilo,bruno}@image.unb.br. Debargha Mukherjee, HP Laboratories Palo Alto, Media Technologies Laboratory, CA 94304 USA. Email: debargha.mukherjee@hp.com.

Em situações onde múltiplas câmeras são usadas para registrar a mesma cena, a abordagem convencional para a codificação das múltiplas vistas, ou *multiview coding* (MVC), explora tanto as correlações temporais quanto as correlações inter-vistas. Essa abordagem é capaz de gerar ganhos significativos de compressão [5], porém implica em um grande aumento de complexidade no codificador. Ademais, a predição inter-vistas impõe a necessidade de comunicação entre os codificadores, o que pode ser inviável em determinadas aplicações.

A codificação distribuída de sequências *multiview* é capaz de aliviar a carga computacional sobre os codificadores, sendo portanto de grande interesse em ambientes contendo múltiplas câmeras de baixa potência. Além da importância de reduzir complexidade durante a codificação, é também de interesse que a decodificação possa ser realizada a baixa complexidade, mesmo que ao custo de uma redução de qualidade ou de resolução da sequência de saída.

As arquiteturas de DVC mais comuns utilizam os chamados quadros-chaves, codificados de maneira *Intra* [2], intercalados com quadros Wyner-Ziv (WZ) [1]. Sem a estimação de movimento ou a predição inter-vistas, a codificação é substancialmente menos complexa. O decodificador faz uso dos quadros-chave e possivelmente estimação de movimento e análise inter-vistas para formar uma estimativa do quadro WZ (denominada informação lateral). Técnicas de codificação distribuídas para *multiview* tais como [6], [7], [8] e [9] transferem ao decodificador as tarefas de estimação de disparidade inter-vistas, retificação e interpolação de vistas, análise de oclusões, entre outras. Para transferir complexidade à decodificação, essas técnicas são unicamente capazes de variar a proporção de quadros-chave para quadros WZ. Caso seja de interesse uma decodificação de baixa complexidade, deve basear-se unicamente no uso de quadros-chave.

A arquitetura DVC proposta em [10] para a codificação de uma vista simples é aplicada nesse trabalho à codificação *multiview*. Ela permite uma transferência gradual de complexidade ao decodificador, utilizando para tal quadros de resolução reduzida (decimados) além de quadros-chave e quadros WZ. Aplicação típica dessa nova arquitetura de escalabilidade espacial é a transmissão de vídeo em tempo real por telefones celulares e câmeras de segurança, onde o aumento na complexidade acarreta maior gasto de energia das baterias de alimentação. Caso o usuário final for outro celular, como em vídeo-chamadas, a decodificação pode ser feita também em baixa complexidade. Caso haja mais espaço para a complexidade na decodificação, como em computadores pessoais, a arquitetura proposta permite uma consequente melhora na qualidade do vídeo decodificado.

Com a exceção de [8], as técnicas *multiview* citadas anteriormente adotam uma vista como a principal e codificam as demais em função dessa. Essa assimetria é responsável pelo uso desigual de recursos durante a codificação e diferenças de qualidade entre vistas. A arquitetura proposta nesse trabalho caracteriza-se também por sua simetria.

Este artigo está organizado da seguinte forma. A arquitetura proposta para codificação distribuída de sequências *multiview* é descrita na seção II. Testes são propostos e resultados apresentados na seção III. As conclusões e futuras linhas de pesquisa são discutidas na seção IV.

II. Arquitetura proposta

A. Arquitetura MR-DVC

A arquitetura utilizada é baseada em [10], que apresenta um codificador de vídeo com complexidade escalonável, feito a partir do *codec* H.264/AVC. Essa arquitetura baseia-se no uso de quadros com resolução mista, ou *mixed resolution* (MR), permitindo que a complexidade seja transferida tanto para a codificação ou para a decodificação, ou inclusive diminuída em ambos.

A complexidade é diminuída no codificador com o uso alternado de quadros de resolução normal e reduzida, como mostra a Figura 1. Entre dois quadros de resolução normal, os quadros-chave, existem quadros espacialmente reduzidos (decimados), os quadros NRWZ (*non-reference Wyner-Ziv*), de forma que a estimação e compensação de movimento nesse segundo tipo de quadro é menos complexa. Além disso, uma camada de realce é codificada com técnicas Wyner-Ziv [1],[10] para quadros NRWZ.

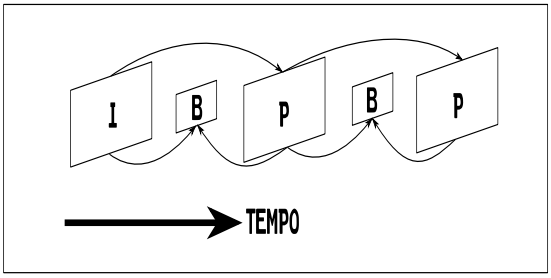

Fig. 1. Arquitetura MR-DVC com um quadro-chave para cada quadro NRWZ.

A compressão é feita com base no padrão H.264/AVC [2], que possui quadros do tipo I (predição espacial), P (predição temporal e espacial, a partir de um quadro anterior ao atual) e B (predição temporal e espacial, a partir de quadros anteriores e posteriores). A arquitetura MR-DVC permite que os quadros-chave sejam dos tipos I e P, e que os quadros NRWZ sejam dos tipos P e B.

O decodificador tem a escolha de codificar os quadro-chave e NRWZ e ignorar a camada de realce, ou então gerar versões de melhor qualidade dos quadros NRWZ usando os quadros-chave e a camada de realce. No primeiro caso, a decodificação é de baixa complexidade, e no segundo, de alta complexidade. Assim, a compressão do vídeo final possui complexidade escalonável, visto que o codificador pode escolher usar quadros NRWZ, e o decodificador pode subsequentemente escolher realçar estes quadros.

A camada de realce corresponde à codificação Wyner-Ziv do resíduo do quadro NRWZ. O codificador tira a diferença entre o quadro original e o quadro NRWZ interpolado às dimensões originais, realiza a transformada DCT deste resíduo, quantiza os coeficientes de saída e calcula os índices de *cosets* sem memória [10] correspondentes.

No decodificador, a camada de realce só pode ser aproveitada se houver alguma aproximação do quadro original, visto que os *cosets* indicam a diferença entre este e o quadro NRWZ. Para isso, é utilizada a semi super resolução, que consiste em acrescentar informação de alta frequência dos quadros-chave aos quadros NRWZ, aproximando o quadro original. Este processo recebe o nome de *geração de informação lateral*.

Os quadros de referência para a semi super resolução são o quadro NRWZ interpolado às dimensões originais (quadro *LR-NRWZ*) e os quadros-chave decimados e interpolados de volta às dimensões originais (quadros *LR-CH*), que representam versões passa-baixa dos quadros-chave. A alta frequência destes é o resíduo entre eles e os quadros *LR-CH*. Para acrescentar alta frequência aos quadros-NRWZ, é feita uma estimação de movimento dos quadros *LR-CH* para cada quadro *LR-NRWZ*. Os vetores de movimento correspondentes apontam para onde nos quadros-chave se deverá obter a informação de alta frequência que será somada aos quadros NRWZ.

B. Arquitetura MR-DMVC

Semelhante à arquitetura MR-DVC, a proposta MR-DMVC pretende comprimir sequências *multiview* de vídeo transferindo a complexidade da codificação para a decodificação. Dessa forma, a arquitetura proposta realiza a codificação de forma *simulcast*, isto é, cada sequência é codificada independentemente das outras. Na decodificação, a semi super resolução gera aproximações dos quadros originais NRWZ utilizando quadros-chave da mesma sequência e também de outras câmeras. Assim, a correlação entre as vistas é explorada para melhorar a semi super resolução, em relação à decodificação *simulcast*.

Aplicações de codificação *multiview* buscam melhorar a predição entre vistas aproveitando-se das relações geométricas entre as câmeras. Entre as contribuições já propostas, destacam-se a compensação de iluminância e a correção de vistas e oclusões [11],[12]. Na arquitetura proposta, as correlações temporais e espaciais entre blocos são exploradas.

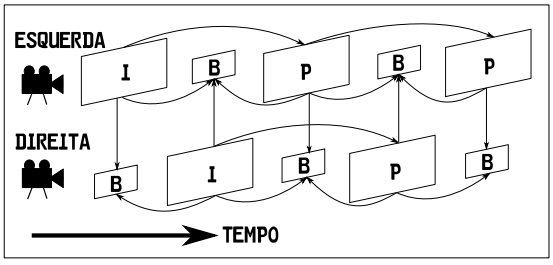

Fig. 2. Sequência estéreo decodificada com RWZ de 1:1

A Figura 2 apresenta uma aplicação desta arquitetura para uma sequência estéreo (duas vistas simultâneas), com a razão 1:1 entre quadros-chave e NRWZ. Já a Figura 3 apresenta uma razão 1:2 para o mesmo contexto. Chamaremos esta razão entre quadros de RWZ (*razão Wyner-Ziv*).

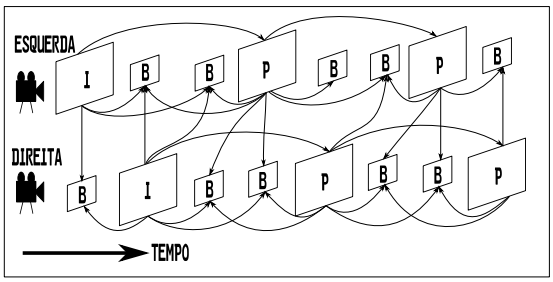

Fig. 3. Sequência estéreo decodificada com RWZ de 1:2

Nas Figuras 2 e 3, as setas indicam quais são os quadros-chave usados como referência na geração de informação lateral para os quadros NRWZ. Observe que o período dos quadros-chave foi deslocado entre as vistas esquerda e direita, de forma que no mesmo instante de tempo, não existem dois quadros-chave. Assim, a correlação entre as vistas é melhor explorada, pois os quadros NRWZ terão quadros-chave mais próximos como referência.

III. Resultados experimentais

A fim de verificar a coerência da arquitetura proposta, foram feitos testes com duas sequências *multiview* disponibilizadas publicamente em [13], *'ballroom'* e *'vassar'*. Assumiu-se um arranjo estéreo de trinta quadros para cada câmera. A RWZ utilizada foi de 1:1, com quadros-chave do tipo I-P-P-P e quadros NRWZ do tipo B. Os passos de quantização (QPs) utilizados foram 22, 27, 32 e 37. O fator de decimação empregado nos quadros NRWZ foi de 2. Para efeito de comparação, as mesmas sequências foram decodificadas nas mesmas condições de forma *simulcast*. Neste trabalho, consideramos somente os ganhos na geração da informação lateral. A principal meta é explorar o uso da informação de múltiplas vistas no processo de semi super resolução, para posteriormente incluir o realce devido à camada Wyner-Ziv.

Considerando as PSNRs quadro a quadro, para diferentes QPs e diferentes vistas, observou-se um ganho médio de 0,065dB para a sequência *'ballroom'*, e de 0,05dB para *'vassar'*. As Figuras 4, 5, 6 e 7 apresentam, a título de ilustração, as PSNRs dos quadros NRWZ (em relação aos originais) para algumas vistas e valores de QP. As Figuras 8 e 9 apresentam as PSNR globais (isto é, considerando todos os quadros e todas as vistas) para as duas sequências teste. Note que a taxa não foi considerada nos gráficos, pois ela é igual para as decodificações *simulcast* e *multiview* - não é necessária nenhuma informação extra para mudar de um tipo de decodificação para outro. Em outras palavras, a geração de informação lateral não influi na taxa final.

IV. Conclusões

Este trabalho apresentou uma nova arquitetura de codificação distribuída de múltiplas vistas, baseada no padrão

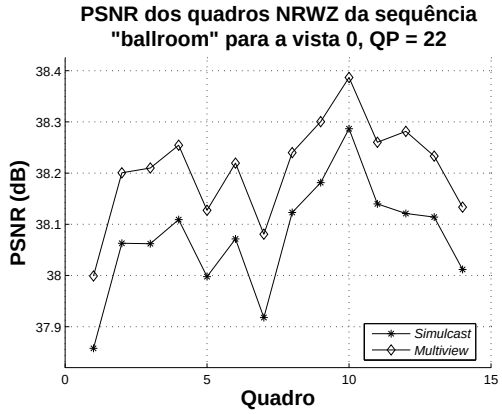

Fig. 4. PSNR dos quadros NRWZ decodificados com QP 22 de forma *simulcast* e *multiview*, para a vista 0 da sequência *'ballroom'*.

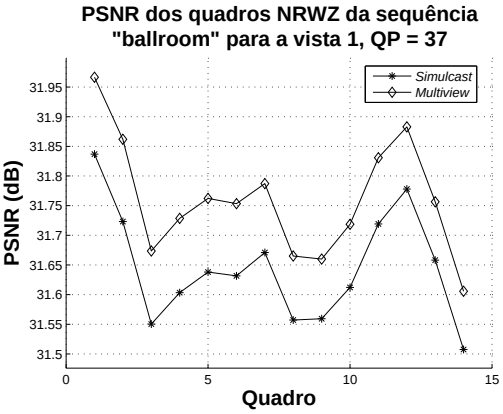

Fig. 5. PSNR dos quadros NRWZ decodificados com QP 37 de forma *simulcast* e *multiview*, para a vista 1 da sequência *'ballroom'*.

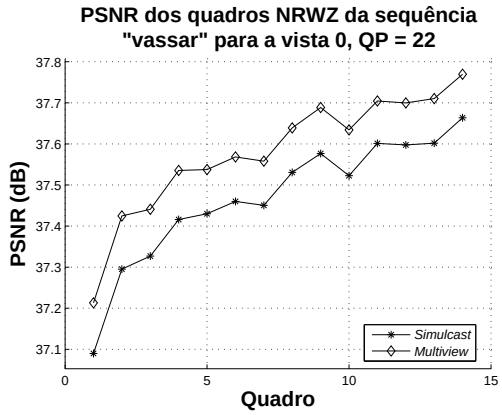

Fig. 6. PSNR dos quadros NRWZ decodificados com QP 22 de forma *simulcast* e *multiview*, para a vista 0 da sequência *'vassar'*.

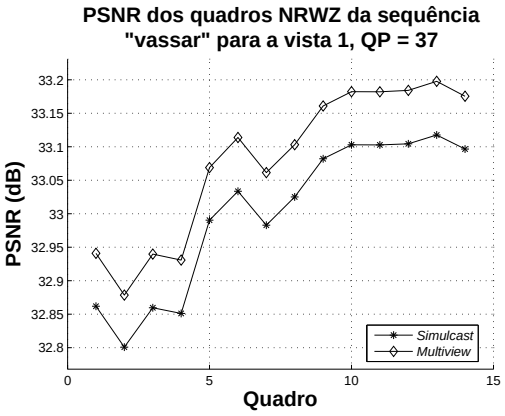

Fig. 7. PSNR dos quadros NRWZ decodificados com QP 37 de forma *simulcast* e *multiview*, para a vista 1 da sequência *'vassar'*.

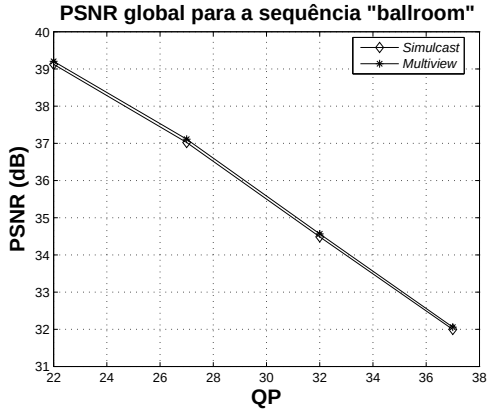

Fig. 8. PSNR global (todos os quadros e todas as vistas) da sequência *'ballroom'*.

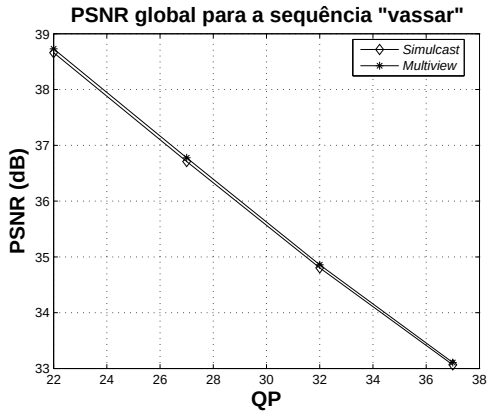

Fig. 9. PSNR global (todos os quadros e todas as vistas) da sequência *'vassar'*.

H.264/AVC e em quadros de resolução mista. A arquitetura apresenta complexidade escalonável na codificação e na decodificação, simetria de uso de recursos no codificador, e explora a correlação espacial e temporal na decodificação. Os resultados demonstram o potencial de ganho de qualidade objetiva para aplicações de baixa taxa e complexidade computacional.

Espera-se encontrar maiores ganhos de qualidade objetiva com a inclusão de uma maior quantidade de vistas, melhor explorando a correlação entre vistas. As sequências de teste utilizadas, por exemplo, apresentam até oito vistas da mesma cena. Além disso, será investigada a influência da RWZ sobre o resultado final, pois o aumento do número de quadros NRWZ acarreta diminuição da taxa de transmissão, sendo que a correlação entre vistas poderá compensar a consequente piora de qualidade objetiva.

Em seguida, a arquitetura proposta será modificada para aproveitar as relações geométricas entre as diversas câmeras. Técnicas como correção entre vistas, interpolação de quadros e o tratamento de oclusões poderão contribuir na geração de informação lateral de melhor qualidade.

REFERÊNCIAS

[1] B. Girod, A. Aaron, S. Rane e D. Rebollo-Monedero, "Distributed video coding", *Proceedings of the IEEE*, v. 93, pp. 71–83, Janeiro 2005.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard e A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Trans. Circuits Syst. Video Technol.*, v. 13, pp. 560–576, Julho 2003.

[3] D. Slepian e J. K. Wolf, "Noiseless coding of correlated information sources", *IEEE Trans. Inform. Theory*, v. 19, pp. 471–480, Março, 1973.

[4] A. D. Wyner e J. Ziv, "The rate-distortion function for source coding with side information at the decoder", *IEEE Trans. Inform. Theory*, v. 22, pp. 1–10, Janeiro 1976.

[5] P. Merkle, A. Smolic, K. Muller e T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding", *IEEE Trans. Circuits Syst. Video Technol.*, v. 17, pp. 1461-1473, Novembro 2007.

[6] X. Artigas, E. Agneli e L. Torres, "Side information generation for multiview distributed video coding using a fusion approach", *7th Nordic Signal Processing Symposium*, Reykjavik, Islândia, Junho 2006.

[7] M. Ouaret, F. Dufaux e T. Ebrahimi, "Fusion-based multiview distributed video coding", *Fusion-based multiview distributed video coding*, Santa Barbara, EUA, Outubro 2006.

[8] M. Tagliasacchi, G. Prandi e S. Tubaro, "Symmetric distributed coding of stereo video sequences", *Proc. IEEE Intl. Conf. on Image Process.*, Atlanta, EUA, Setembro 2007.

[9] X. Guo, Y. Lu, F. Wu, D. Zhao e W. Gao, "Wyner-Ziv-Based Multiview Video Coding", *IEEE Trans. Circuits Syst. Video Technol.*, v. 18, pp. 713–724, Junho 2008.

[10] B. Macchiavello, D. Mukherjee e R. L. de Queiroz, "Iterative Side-Information Generation in a Mixed Resolution Wyner-Ziv Framework". IEEE Trans. Image Process., 2009.

[11] J.-H. Hur, S. Cho e Y.-L. Lee, "Adaptive local illumination change compensation method for H.264/AVC-based multiview coding", *IEEE Trans. Circuits Syst. Video Technol.*, v. 17, pp. 1496–1506, Novembro 2007.

[12] K. Yamamoto et al., "Multiview video coding using view interpolation and color correction", *IEEE Trans. Circuits Syst. Video Technol.*, v. 17, pp. 1436–11450, Novembro 2007.

[13] Mitsubishi Electric Research Laboratories, "MERL multiview video sequences", `ftp://ftp.merl.com/pub/avetro/mvc-test seq.`