Super-Resolução no Domínio da Transformada para Imagens de Múltiplas Vistas e Resolução Mista

Edson M. Hung, Camilo Dorea, Diogo C. Garcia e Ricardo L. de Queiroz

***Resumo*— A arquitetura de resolução mista tem sido adotada tanto para redução de complexidade em codificação de vídeo quanto para compressão de dados em vídeo estereoscópico. Os quadros de alta resolução presentes na arquitetura podem também servir para realçar as imagens de menor resolução. Nesse artigo apresentamos um método de super-resolução para uso em sistemas de múltiplas vistas em resolução mista com informação de profundidade. As vistas de alta resolução são projetadas sobre a vista de baixa resolução com auxílio de mapas de profundidade. O método introduz o uso de técnicas no domínio da transformada para interpolação de imagens de baixa resolução e para agregação de conteúdo de alta frequência proveniente das vistas projetadas. A DCT é escolhida para a decomposição em frequência e os resultados apresentados demonstram ganhos de qualidade objetivos para diversas sequências de teste.**

***Palavras-Chave*— Resolução mista, super-resolução, múltiplas vistas e profundidade.**

***Abstract*— Complexity reduction in video coding and data size compression in stereoscopic video has been achieved with the use of mixed resolution formats. The high resolution frames within mixed resolution formats may also be used to enhance the low resolution frames. This paper presents a super-resolution technique for usage within a multiview video plus depth setup with mixed resolution. Available depth information is used to project high resolution views onto the view points of low resolution images. A transform domain technique is introduced to up-sample the low resolution image and to aggregate high frequency content from the projected view. The DCT is the adopted transform and results demonstrate objective quality gains for various sequences.**

***Keywords*— Mixed resolution, super-resolution, multiview plus depth.**

I. INTRODUÇÃO

Os métodos de super-resolução (SR) tem por objetivo obter uma ampliação em alta resolução de uma imagem. Através do uso de múltiplas imagens correlacionadas, tais métodos podem ultrapassar as limitações inerentes à interpolação quando esta está restrita ao uso de uma única imagem. Em geral, os métodos de SR exploram deslocamentos sub-pixel entre imagens de baixa resolução para formar uma imagem de alta resolução. No entanto, alguns métodos de SR baseiam-se em imagens disponíveis de alta resolução para assim estimar os detalhes que estão ausentes na imagem de baixa resolução [1], [2]. Por exemplo, [1] usa um conjunto de treinamento composto de imagens de alta resolução para restaurar as altas frequências ausentes em imagens sujeitas a *zoom*. De maneira semelhante, a SR é empregada em [2] com vídeo de resolução mista para recuperar quadros de baixa resolução através do uso de informação de alta frequência presente em quadros vizinhos de alta resolução. O método de SR apresentado nesse artigo assemelha-se a essas duas últimas propostas, pois é também baseado no uso de imagens disponívies de alta resolução para realçar imagens de baixa resolução.

Além do seu uso para redução de complexidade em codificação de vídeo [2], [3], as arquiteturas de resolução mista também são empregadas na redução de tamanho de arquivo em vídeo estreoscópico [4], [5]. Nesse caso, ao invés de intercalar temporalmente quadros de alta e baixa resolução, uma vista de baixa resolução é apresentada ao olho esquerdo, por exemplo, enquanto a vista de alta resolução é reservada ao olho direito. Métodos de SR tem sido empregados no processamento de vídeo em resolução mista para aliviar o problema de cintilamento (*flickering*) durante a visualização. Porém, o vídeo estereoscópico em resolução mista (também chamado de assimétrico) é geralmente visualizado em resoluções diferenciadas e sem processamento. Isso justifica-se por estudos psicovisuais [6], [7] que indicam que a agudeza e a percepção de profundidade da imagem estereoscópica são determinadas pelo canal de alta resolução. No entanto, em casos mais genéricos como os sistemas de múltiplas vistas, a resolução mista pode não ser diretamente aplicável. As diferenças de qualidade entre as diferentes vistas podem ser prejudiciais a algumas das aplicações almejadas pelos sistemas de múliplas vistas. Por exemplo, devido à sua natureza monoscópica, a navegação entre vistas dentro de um vídeo com ponto de vista livre (*free view-point video*) apresentará significativas diferenças em qualidade entre as vistas de alta e baixa resolução.

Para superar tais limitações um método de SR para uso em arquiteturas de múltiplas vistas em resolução mista foi proposto [8]. A arquitetura está ilustrada na Figura 1 e consiste em múltiplas sequências de vídeo de diferentes pontos de vista e resoluções e seus mapas de profundidade correspondentes. Os mapas de profundidade [9] foram mantidos a resolução máxima pois são eficientemente codificáveis e representam uma pequena porcentagem do tamanho total de dados. Os mapas são ultilizados para estabelecer as correspondências entre vistas.

O método de SR de [8] baseia-se no uso de filtros lineares para interpolar as imagens de baixa resolução e para isolar o conteúdo de alta frequência em imagens vizinhas de alta resolução. Ambas operações foram implementadas no domínio espacial, porém ambas apresentam alternativas atraentes no domínio da transformada. Ganhos objetivos de qualidade

Edson M. Hung[1], Camilo Dorea[2], Diogo C. Garcia[3] and Ricardo L. de Queiroz[2] [1]Faculdade do Gama - Engenharia Eletrônica, [2]Departamento de Ciência da Computação, [3]Departamento de Engenharia Elétrica - Universidade de Brasilia e Instituto Federal de Educação, Ciência e Tecnologia de Brasilia, Brasil E-mail: {mintsu, camilo, diogo}@image.unb.br, queiroz@ieee.org; web: divp.org

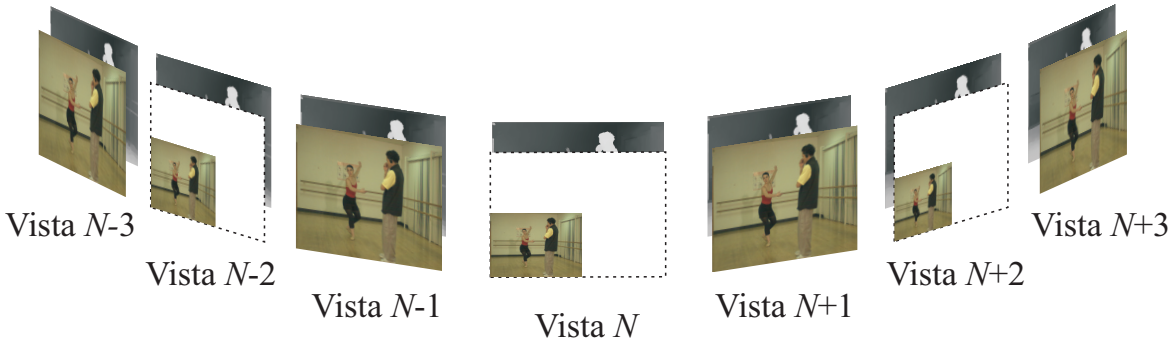


Fig. 1. A arquitetura de múltiplas vistas em resolução mista com mapas de profundidade.

obtidos através de operações de interpolação no domínio da transformada relativos à interpolação linear de parâmetros fixos, tais como bilinear, são apresentados em [10]. A qualidade visual de resultados pode ser melhorada ao combinar interpolação baseada em DCT, responsável por preservar os coeficientes de baixa frequência, com uma estimação baseada em filtros Wiener dos coeficientes de alta frequências [11]. Além de bons resultados de interpolação, o uso de transformadas constitui um domínio natural para a decomposição e isolamento de frequências em imagens.

Nesse artigo apresentamos um método de SR para uso em arquiteturas de múltiplas vistas em resolução mista com informação de profundidade. O conteúdo de alta frequência presente nas imagens adjacentes de alta resolução é usado para realçar as imagens de baixa resolução. Introduzimos operações no domínio da transformada para realizar interpolação e isolamento de altas frequências. O método proposto preserva coeficientes DCT de baixa frequência presentes na imagem de baixa resolução e complementa os mesmos com coeficientes DCT de alta frequência provenientes de vistas de alta resolução adequadamente projetadas. A projeção é realizada com uso de informação de profundidade.

II. Método Proposto

O método proposto de SR no domínio da transformada é ilustrado na Figura 2. As imagens de alta resolução disponíveis na arquitetura de resolução mista, que utiliza vídeos com o seguinte formato: múltiplas vistas e mapa de profundidade, são inicialmente projetadas em um ponto de vista correspondente à uma imagem de baixa resolução como indica o bloco de Projeção de Vista da Figura 2. O bloco de Interpolação é responsável por aumentar o tamanho da imagem de baixa resolução em tamanhos compatíveis aos de alta resolução. Finalmente, a SR baseada em blocos DCT adiciona os coeficientes de alta-frequencia da vista projetada na imagem de baixa resolução. As próximas sub-seções descreverão com detalhes os blocos supracitados.

A. Projeção de Vista

A projeção de vista com boa qualidade é um componente essencial para a SR proposta. A técnica de projeção utilizada na arquitetura proposta faz a renderização baseada no mapa de profundidade e possui como entradas a imagem de alta

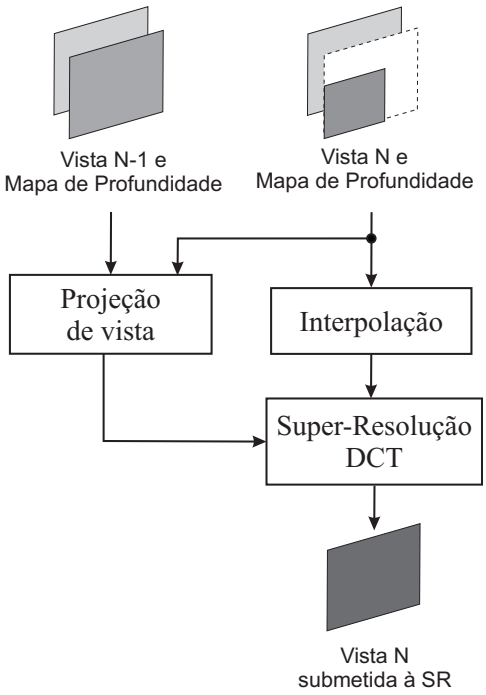


Fig. 2. Diagrama de blocos do método proposto de SR baseado em transformadas.

resolução V_{N-1}, mapas de profundidade D_{N-1} and D_N e possui como saída síntese de uma imagem na N-ésima vista $\widehat{V_N}$. O conhecimento dos parâmetros intrínsecos das câmeras $\mathbf{A}$, matriz de rotação $\mathbf{R}$, vetor de translação $\mathbf{t}$ são utilizados para projetar a localização $(\widehat{u}, \widehat{v})$ do pixel na câmera N nas coordenadas absolutas tridimensionais(x, y, z) [12]:

$$(x, y, z)^T = \mathbf{R}_N {\mathbf{A}_N}^{-1} (\widehat{u}, \widehat{v}, 1)^T D_N(\widehat{u}, \widehat{v}) + \mathbf{t}_N. \quad (1)$$

Em seguida, as coordenadas absolutas são re-projetadas na câmera $N-1$ na posição (u, v):

$$(u*w, v*w, w)^T = \mathbf{A}_{N-1} {\mathbf{R}_{N-1}}^{-1} [(x, y, z)^T - \mathbf{t}_{N-1}]. \quad (2)$$

Geralmente, nem todos os pixels possuem correspondências entre as vistas. Um teste de consistência também é aplicado de forma a identificar possíveis erros de correspondência. As coordenadas (u, v) são arredondadas para os valores inteiros

(a) (b)

Fig. 3. (a) Sequencia *Ballet* (vista 0, imagem 0) e (b) sua projeção para vista 1 (buracos mostrados em branco).

mais próximos e projetados de volta à câmera N. Mas, se a distância Euclidiana entre as posições resultantes desta última projeção e as coordenadas originais $(\widehat{u}, \widehat{v})$ for menor que um limiar especificado (tipicamente 1.0) a correspondência é aceita, caso contrário ela é rejeitada. A projeção da vista é completada ao substituir $(\widehat{u}, \widehat{v})$ pela interpolação bilinear da amostra validada na posição correspondente (u, v) na vista V_{N-1}. As posições onde o teste de correspondência não logrou êxito permanecem na imagens como buracos, exemplificados pela Figura 3. Observe que no método proposto os buracos oriundos da projeção são preenchidos por pixels da imagem interpolada (originalmente com baixa resolução), portanto, não há componentes de alta frequência para essas regiões que contribuam ao processo de super-resolução descrito na próxima seção. A extensão do método de SR para uso com duas ou mais vistas adjacentes poderá diminuir a área correspondente aos buracos de projeção [8].

B. Super-resolução baseada na DCT

O método de SR por meio de transformada é baseado na DCT dos blocos da imagem, onde por conseguinte se determina os coeficientes de alta freqüência que serão adicionados às imagens de baixa resolução. Em seguida, serão revistos os métodos de interpolação e decimação baseados na DCT, que embasam a SR proposta. Seja $\mathbf{b}$ um bloco da imagem contendo $(m \times m)$ pixels. A Equação (3) expressa os coeficientes DCT de $\mathbf{b}$ como uma matriz particionada. $\mathbf{B}_{00}$ é uma sub-matriz $(n \times n)$ que contém os coeficientes de baixa freqüência. $\mathbf{B}_{01}$, $\mathbf{B}_{10}$ e $\mathbf{B}_{11}$ são matrizes de tamanhos: $(m-n \times n)$, $(n \times m-n)$ e $(m-n \times m-n)$, respectivamente, que contém os coeficientes de alta freqüência.

$$DCT\{\mathbf{b}\} = \left[\begin{array}{c:c} \mathbf{B}_{00} & \mathbf{B}_{01} \\ \hdashline \mathbf{B}_{10} & \mathbf{B}_{11} \end{array}\right]. \tag{3}$$

A decimação do bloco $\mathbf{b}$ da imagem é obtida após calcularmos a DCT inversa da sub-matriz $\mathbf{B}_{00}$, descartando assim os componentes de alta freqüência [10]. Entretanto, por conta das diferenças de tamanho entre as DCT inversas e diretas, a sub-matriz $\mathbf{B}_{00}$ deve ser multiplicada por um fator de escala $s_{dsz} = n/m$ para obtermos o bloco de imagem de tamanho $(n \times n)$:

$$\mathbf{b}_{dsp} = IDCT\{s_{dsz}[\mathbf{B}_{00}]\}. \tag{4}$$

A interpolação de um bloco da imagem pode ser obtida ao inserir zeros nos coeficientes de alta frequência e em seguida calcular a DCT inversa [10]. Por exemplo, a interpolação de $\mathbf{b}_{dsp}$ é obtida ao assumiroms que os valores das sub-matrizes $\mathbf{B}_{01}$, $\mathbf{B}_{10}$ e $\mathbf{B}_{11}$ são nulos formando assim blocos de $(m \times m)$ pixels dados por

$$\mathbf{b}_{usp} = IDCT\left\{\left[\begin{array}{c:c} \mathbf{B}_{00} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} \end{array}\right]\right\}. \tag{5}$$

Portanto a interpolação utilizando DCT é obtida ao dividirmos a imagem em blocos de tamanho $(n \times n)$, determinando assim os coeficientes de cada bloco. Em seguida adicionamos os coeficientes nulos de alta frequência e aplicamos a DCT inversa, conforme descrito na Equação (5).

O método de SR proposto tem por objetivo melhorar a qualidade da interpolação, utilizando-se da estimação dos coeficientes de alta frequência baseados em outras vistas, ao invés de assumir que as sub-matrizes de alta freqência são nulas. Como descrito na sub-seção II-A, a vista contendo imagens de alta resolução são consistentemente projetadas para o mesmo ponto de vista das imagens de baixa resolução. As vistas projetadas servirão como fonte de coeficientes de alta frequência que são utilizados para preencher os coeficientes de detalhe para aplicar a SR nas imagens de baixa resolução. Para cada bloco $\widehat{\mathbf{b}}$ na imagem com a vista projetada, os coeficientes são descritos como

$$DCT\left\{\widehat{\mathbf{b}}\right\} = \left[\begin{array}{c:c} \widehat{\mathbf{B}}_{00} & \widehat{\mathbf{B}}_{01} \\ \hdashline \widehat{\mathbf{B}}_{10} & \widehat{\mathbf{B}}_{11} \end{array}\right]. \tag{6}$$

As sub-matrizes de alta frequência $\widehat{\mathbf{B}}_{01}$, $\widehat{\mathbf{B}}_{10}$ e $\widehat{\mathbf{B}}_{\mathbf{11}}$ são utilizados para completar a informação inexistente de alta frequência em blocos co-localizados $\mathbf{b}_{usp}$ da imagem interpolada, que originalmente era de baixa resolução. Portanto, o bloco da imagem que sofreu o processo de SR $\mathbf{b}_{SR}$ é dado por

$$\mathbf{b}_{SR} = IDCT\left\{\left[\begin{array}{c:c} \mathbf{B}_{00} & \widehat{\mathbf{B}}_{01} \\ \hdashline \widehat{\mathbf{B}}_{10} & \widehat{\mathbf{B}}_{11} \end{array}\right]\right\} \tag{7}$$

onde a sub-matriz $\mathbf{B}_{00}$ são coeficientes DCT de baixa frequência remanescentes do processo de interpolação, como descrito na Equação (5). Ao aplicar a SR em cada bloco de baixa resolução, obtemos como resultado uma imagem com alta resolução cujos coeficientes de alta frequência são oriundos das imagens de alta resolução de outras vistas.

III. Resultados Experimentais

O desempenho do método proposto foi avaliada com um conjunto publicamente disponível de imagens reais e sintéticas. Devido à falta de conteúdo de alta frequência, as imagens reais *Ballet* e *Breakdancers* foram redimensionadas para 512×384 e 256×192 *pixels*, respectivamente, antes de serem testadas. As imagens em múltiplas vistas, disponibilizadas em conjunto com os mapas de profundidade ou de disparidade, foram reduzidas de forma a comporem a arquitetura de resolução mista apresentado na Figura 1. O método de SR é aplicado a cada imagem em baixa resolução, baseado na vista mais próxima em alta resolução. As operações utilizam a DCT do tipo II em todos os casos, com blocos de tamanho

original 8×8 e tamanho reduzido 4×4 ($m = 8$ e $n = 4$), o que resulta em um fator de redução de 2.

Os primeiros testes comparam os resultados de interpolação da imagem em baixa resolução utilizando um filtro linear de alto desempenho (com um *kernel* Lanczos) e com o método baseado em DCT apresentado na Sub-seção II-B. A Tabela I mostra uma ligeira piora do método baseado em DCT em relação à interpolação com *kernel* Lanczos, de $-0,32$ dB, em média. Note que o método baseado em DCT simplesmente acrescenta zeros como uma estimativa dos componentes de alta frequência.

TABELA I
COMPARAÇÃO EM PSNR DE MÉTODOS DE INTERPOLAÇÃO.

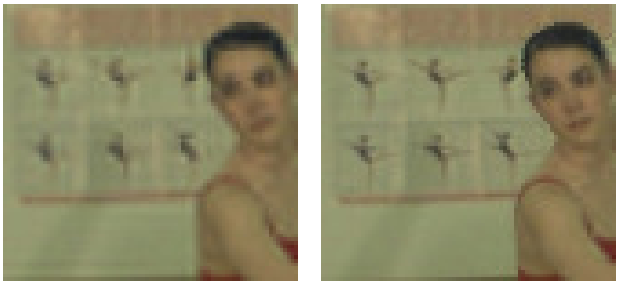
Sequência	*Kernel* Lanczos	DCT	Ganho em PSNR
Ballet	34,01 dB	33,71 dB	-0,30 dB
Breakdancers	35,47dB	34,95 dB	-0,52 dB
Barn1	27,76 dB	27,49 dB	-0,27 dB
Barn2	31,06 dB	30,74 dB	-0,32 dB
Bull	32,46 dB	32,23 dB	-0,23 dB
Map	28,00 dB	27,56 dB	-0,44 dB
Poster	26,46 dB	26,06 dB	-0,40 dB
Sawtooth	28,32 dB	27,93 dB	-0,39 dB
Venus	28,63 dB	28,40 dB	-0,23 dB

O segundo conjunto de testes compara o método proposto de SR no domínio da transformada com um método de SR no domínio espacial. Este segundo método utiliza uma interpolação linear baseada no *kernel* Lanczos e a extração espacial de alta frequência assim como descrito em [8]. Porém, para efeito de comparação, a projeção da vista adjacente é idêntica em ambos os métodos (nos domínios espacial e da transformada). A Tabela II indica que o método no domínio da transformada é superior ao método no domínio espacial em todas as sequências, exceto por uma. Verifica-se um ganho médio de $0,16$ dB, chegando a $0,5$ dB para a sequência *Bull*. Observe que o método no domínio da transformada utiliza uma interpolação de desempenho objetivo inferior, como indicado na Tabela I, mas que se mostra mais adequado no processo de SR.

TABELA II
COMPARAÇÃO EM PSNR DE MÉTODOS DE SUPER RESOLUÇÃO NOS DOMÍNIOS ESPACIAL E DA TRANSFORMADA.

Sequência	*Kernel* Lanczos	DCT	Ganho em PSNR
Ballet	36,18 dB	**36,31 dB**	0,15 dB
Breakdancers	38,69 dB	**38,84 dB**	0,15 dB
Barn1	35,83 dB	**36,22 dB**	0,39 dB
Barn2	38,40 dB	**38,50 dB**	0,10 dB
Bull	37,96 dB	**38,46 dB**	0,50 dB
Map	31,20 dB	**31,24 dB**	0,04 dB
Poster	33,93 dB	**34,09 dB**	0,16 dB
Sawtooth	**33,72** dB	33,32 dB	-0,40 dB
Venus	35,61 dB	**35,99** dB	0,38 dB

Uma avaliação subjetiva do método proposto pode ser realizada por meio das imagens da Figura 4. Para a sequência *Ballet*, a SR da vista 1 baseada em DCT usando uma imagem de alta resolução correspondente à vista 2 pode ser comparada com a interpolação da vista 1 baseada em DCT. Detalhes de alta frequência foram inseridos pelo método de SR, ressaltando o contorno na face da bailarina e na textura do fundo da imagem. Note que nesse experimento projetou-se a vista 2 e não a vista 0 como ilustrado no exemplo da Figura 3. Estas melhoras refletem o ganho atingido em termos de PSNR, de $2,60$ dB, que pode ser obtido comparando a segunda coluna das Tabelas I e II.



(a) (b)

Fig. 4. Detalhe parcial da vista 1 da sequencia *Ballet*: (a) Imagem interpolada com DCT (33.71 dB) e (b) Imagem submetida à SR baseada na DCT (36.31 dB).

Para a sequência sintética *Barn1*, a diferença de PSNR entre a SR baseada em DCT e a interpolação baseada em DCT é de $8,73$ dB. A Figura 5 permite uma comparação subjetiva. Observe que a inserção de componentes de alta frequência pelo método proposto resulta em uma imagem mais detalhada e definida.

(a) (b)

Fig. 5. Vista 1 da sequencia *Barn1*: (a) Imagem interpolada com DCT (27.49 dB) e (b) Imagem submetida à SR baseada na DCT (36.22 dB).

IV. CONCLUSÕES

Este artigo apresenta um novo método de SR no domínio da transformada, para uso em sistemas de resolução mista para sequências com múltiplas vistas com informação de profundidade. Uma ténica baseada em DCT é introduzida para realizar a interpolação da imagem em baixa resolução. O método proposto procede projetando a vista em alta resolução para o ponto de vista da imagem de baixa resolução. Coeficientes de alta frequência da vista projetada são utilizados para preencher os coeficientes de alta frequência que estão ausentes na imagem em baixa resolução. A imagem submetida ao processo de SR no domínio da transformada alcança ganhos de qualidade significativos sobre a imagem interpolada, tanto em termos objetivos como subjetivos.

Trabalhos futuros incluem investigar a remoção no domínio da transformada de ruído das vistas projetadas, assim como a redução de artefatos e o aguçamento da imagem submetida à SR por meio da manipulação dos componentes DCT de alta frequência. A arquitetura proposta também permite o uso de outras técnicas de decomposição em frequência além da DCT como, por exemplo, as wavelets.

REFERÊNCIAS

[1] W.T. Freeman, T.R. Jones, and E.C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, Vol. 22, pp. 56-65, 2002.

[2] F. Brandi, R. de Queiroz, D. Mukherjee, "Super-resolution of video using key frames and motion estimation,", *Proc. IEEE Intl. Conf. on Image Processing*, San Diego, EUA, Outubro 2007.

[3] D. Mukherjee, "A robust reversed complexity Wyner-Ziv video codec introducing sign-modulated codes," *HP Labs Technical Report*, HPL-2006-80, Maio 2006.

[4] H. Brust, A. Smolic, K. Mueller, G. Tech and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices,", *Proc. 3DTV Conference*, Potsdam, Alemanha, Maio 2009.

[5] M. G. Perkins, "Data compression of stereopairs," *IEEE Trans. on Communications*, Vol. 40, pp. 686-696, 1992.

[6] B. Julesz, "Foundations of cyclopean perception," University of Chicago Press, 1971.

[7] L. Stelmach, W. J. Tam, D. Meegan and A. Vincent, "Stereo Image quality: effects of mixed spatio-tamporal resolution," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 10, no. 2, pp. 188-193, Março 2000.

[8] D. C. Garcia, C. C. Dorea, and R. L. de Queiroz, "Super-resolution for multiview images using depth information," *Proc. IEEE Intl. Conf. on Image Processing*, ICIP, Hong Kong, China, Setembro 2010.

[9] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM SIGGRAPH*, Los Angeles, EUA, Agosto 2004.

[10] R. Dugad and N. Ahuja, "A fast scheme for image size change in the compressed domain," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 11, no. 4, pp. 461-474, Apr. 2001.

[11] Z. Wu, H. Yu e C. W. Chen, "A New Hybrid DCT-Wiener-Based Interpolation Scheme for Video Intra Frame Up-Sampling," *IEEE Signal Processing Letters*, vol. 17, issue 10, pp. 827-830, Outubro 2010.

[12] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic e R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Image Communication*, pp. 217-234, 2007.