

Otimização de distorção de profundidade admissível para codificação de mapas de profundidade

Gizele Júlio, Camilo Dórea e Bruno Macchiavello

Resumo—Sistemas de múltiplas vistas são amplamente empregados na criação de vídeos 3D e de aplicações de *ponto de vista livre*. As múltiplas vistas, contendo vídeos de textura (cor) e profundidade, devem ser eficientemente comprimidas para serem transmitidas ao cliente e poderão servir para síntese de vistas no receptor. Nesse contexto, foi proposto um pré-processamento dos mapas de profundidade antes da codificação, baseado no modelo de Distorção de Profundidade Admissível (ADD - Allowable Depth Distortion). Esse trabalho explora a ADD e, adicionalmente, propõe a escolha de valores de profundidade para transmissão de acordo com a distribuição dos blocos (*macroblocks* e/ou *coding units*) empregados por codificadores padrões. Nossos resultados experimentais mostram que é possível alcançar ganhos de compressão de até 8% usando o método da mínima variância de valores dentro de um bloco de codificação, sem a introdução de perdas por distorção e preservando-se a qualidade das imagens sintetizadas.

Palavras-Chave—mapas de profundidade, Distorção de Profundidade Admissível (ADD), síntese de vistas, compressão de dados.

Abstract—Multiview systems are widely used to create 3D video as well as in Free-Viewpoint Video applications. The multiple views, consisting of texture images and depth maps, must be efficiently compressed and transmitted to clients where they may be used towards the synthesis of virtual views. In this context, the Allowable Depth Distortion (ADD) has been used in a pre-processing step prior to depth coding. This work explores ADD and, additionally, proposes the choice of depth value for transmission in accordance to the distribution of blocks (e.g., macroblocks and/or coding units) commonly employed by standardized coders. Experimental results show that our proposal can achieve compression gains of up to 8% applying the minimum variance method within a block, without introducing losses in terms of distortion and preserving synthesized image quality.

Keywords—depth maps, Allowable Depth Distortion (ADD), view synthesis, data compression.

I. INTRODUÇÃO

Os sistemas 3D tem como objetivo proporcionar uma imersão na cena, trazendo uma sensação de visualização natural do vídeo. Por meio da estereoscopia, em que é levado em consideração a separação das câmeras de captura para simular a paralaxe binocular do ser humano, pode-se proporcionar diversas intensidades de imersão a depender da preferência do usuário na experimentação 3D.

Nos últimos anos, houve uma popularização dos vídeos 3D e posterior comercialização das 3DTV's, devido, principalmente, ao grande avanço dos *displays* 3D de auto-estereoscopia e a transmissão digital de TV. Com isso, o aumento do consumo de dados, inclusive em formato de

streaming 3D, gera grande necessidade de transmissão e armazenamento eficiente [1].

Para o uso de sistemas de múltiplas vistas é incluída a informação de profundidade para cada vista da cena, os denominados mapas de profundidade, onde descreve-se a distância dos objetos em relação à câmera [2]. Para atender a este tipo de sistema é necessário o envio de não somente um único ponto de vista da cena, mas duas ou mais vistas (de diferentes posições de câmeras) de uma mesma cena, possibilitando, assim, algumas aplicações como o caso de vídeo de ponto de vista livre (*free-viewpoint*), que o usuário tenha a liberdade de escolha do ângulo de visualização da cena. Diante desta necessidade de transmitir mapas de profundidade e múltiplas vistas, o volume de dados torna-se ainda mais excessivo em relação à transmissão de vídeos 2D.

Para responder a tal demanda, seria necessário desenvolver maiores capacidades de transmissão nas redes de comunicação, o que nem sempre é viável, dadas as limitações das tecnologias de transmissão como fibras ópticas e tecnologias de rádio-enlaces. Com isso, uma abordagem muito explorada é a redução da quantidade de dados e a possibilidade de construção da informação no lado do cliente. Para contornar esta sobrecarga dos sistemas de telecomunicações, utiliza-se muitas vezes a Renderização de Imagem Baseada em Profundidade (*Depth Image Based Rendering* - DIBR) [3]. Esta técnica utiliza a informação de textura mais profundidade por pixel advinda de uma ou mais câmeras de referência do sistema de múltiplas vistas, para a projeção e criação de uma vista sintetizada. Com isso, torna realizável a criação de vídeos 3D a partir do envio de apenas uma das vistas da cena e a concepção da segunda, a partir da técnica DIBR. Assim, torna-se desnecessário o envio de todas as vistas. Os vídeos de textura e profundidade utilizados são, ademais, codificados para compressão [4].

O trabalho de Zhao et al. [5] explora a capacidade de variação dos valores de profundidade dos mapas dentro de um determinado intervalo, ou seja, altera os valores originais, sem que isto afete a síntese da vista virtual. O trabalho mostra também que esta alteração não afeta a ordem de projeção dos pixels, pois ao utilizar o método *z-buffer*¹ para o preenchimento dos pixels, a prioridade de escolha se mantém. Além disso, é proposto uma suavização no mapa de profundidade, utilizando uma máscara gaussiana, para aumentar a eficiência de codificação INTRA, o qual não causa nenhum erro de sintetização, sendo o erro quadrático médio (MSE) entre a imagem sintetizada originalmente e a que é sintetizada

Gizele Júlio, Camilo Dórea e Bruno Macchiavello: Departamento de Ciência da Computação, Universidade de Brasília, Brasília-DF, Brasil. E-mails: gizele.abdon@aluno.unb.br, camilodorea@unb.br e bruno@cc.unb.br.

¹Técnica que leva em consideração a proximidade de determinado pixel em relação à câmera, escolhendo o mais próximo para o preenchimento da cena.

a partir dos mapas pré-processados, igual a zero. Porém, diferentemente de nossa proposta, em Zhao et al. [5] não foram exploradas durante o pré-processamento possíveis relações espaciais ou de vizinhança compatíveis com os processos de codificação a serem empregados.

Já o trabalho de Zhang et al. [6] modela a quantização como um ruído branco e o insere nos mapas de profundidade, investigando a relação entre a mudança dos valores de profundidade e seus efeitos sobre a síntese de vistas, procurando otimizar a função taxa-distorção (RDO), ao tentar fazer a melhor seleção do modo decisão de predição e de estimação de movimento em nível de Macroblocos (MB's) sendo o cálculo desta função de custo bastante complexo computacionalmente. Esta variação possível dos valores de profundidade que provocam uma mesma disparidade, empregada também por Zhao et al. [5], é denominada de Distorção de Profundidade Admissível (ADD).

No artigo de Zhang et al.[7] propõe-se uma escolha rápida do modo de seleção para a codificação dos macroblocos no codificador H.264/AVC [8]. Para isso, eles citam que existem onze variáveis a serem otimizadas através de uma função taxa-distorção, procurando os valores para as variáveis que causem menor resíduo de predição e maior poder de compressão. Portanto, o proposto é que em vez de testar todas as possíveis soluções, é possível determinar uma decisão mais cedo e com isso economizar no tempo de processamento, apresentando então uma condição de terminação rápida baseada na função RD. Os autores utilizam também a ideia do intervalo ADD para pré-processar os mapas de profundidade e variam os níveis de quantização e, portanto, a eficiência de compressão. Os métodos de Zhang et al. consideram os efeitos de distorção sobre taxa de transmissão. Em nosso trabalho, excluímos perdas por distorção e preservamos a qualidade original das imagens sintetizadas.

Nesse aspecto, este trabalho tem como intuito mitigar ainda mais o volume de dados a ser transmitido ao investigar a capacidade de compressão dos dados de profundidade que são necessários para a síntese de vista. Propõe-se um pré-processamento nestes mapas de profundidade com a finalidade de explorar a propriedade de Distorção de Profundidade Admissível (ADD) apresentada, inicialmente, por Zhao et al. [5] e modelada, posteriormente, por Zhang et al. [6], [7]. Adicionalmente, é proposto neste trabalho um método que, utilizando o conceito ADD, escolhe o valor ótimo de substituição da profundidade dentro do intervalo ao qual pertença, considerando uma codificação INTRA quadro sem perdas. Avaliamos o impacto dessa escolha, denominada critério de Mínima Variância, entre o valor de profundidade a ser pré-processado e a média de profundidade do bloco ao qual está espacialmente inserido este pixel, com o intuito de alcançar maiores níveis de compressão. O método proposto é então comparado em relação à escolha de um único valor de substituição para todas as profundidades que pertencem a um mesmo intervalo ADD, sem levar em consideração sua localização espacial e sua vizinhança.

Revisamos a seguir alguns conceitos sobre síntese de vistas na Seção II. Apresentamos o modelo de Distorção de Profundidade Admissível e nossa proposta de escolha através de Mínima Variância na Seção III. Resultados e Conclusões são

discutidos nas Seções V e VI, respectivamente.

II. SÍNTESE DE VISTAS

A síntese de vistas busca criar uma imagem para uma dada posição virtual de câmera. Para isso, utiliza a técnica DIBR [3] para realizar as projeções das câmeras de referência para o ponto sintetizado, utilizando as informações do vídeo de textura mais o de profundidade. A informação de profundidade têm finalidade apenas para a projeção geométrica do pixel de textura. O mapa de profundidade pode ser obtido por meio de aferições diretas por pulsos de luz infra-vermelha ou por técnicas de estimação de disparidade. O padrão MPEG-3DV adota um esquema de conversão dos valores de profundidade reais z em um conjunto de valores v que variam de 0 a 255. O valor real de profundidade é calculado a partir da equação inversa da quantização do valor de profundidade v :

$$z = \frac{1}{\frac{v}{255} \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}}} \quad (1)$$

onde Z_{near} e Z_{far} são os valores mais próximo e mais distante da câmera, respectivamente [9].

No momento de captura da cena, as câmeras podem estar retificadas ou não. No caso retificado, a distância entre as câmeras é geralmente uniforme e a distribuição linear resultando em variações em apenas uma das direções, normalmente, a horizontal.

Quando em sistema de múltiplas vistas é realizada a projeção de uma vista a partir de uma referência, é necessário relacionar todos os sistemas coordenadas: câmera, global e da imagem. A partir das imagens capturadas (textura + profundidade) e os parâmetros de configuração da câmera é possível realizar a projeção da vista em dois passos: criar a projeção 2D para 3D e, a seguir, realizar a projeção para a posição sintetizada escolhida. Para realizar a projeção de qualquer pixel p_1 de coordenada m na vista 1 para sua correspondente posição p_2 de coordenada m' na vista 2, considerando duas câmeras no sistema, e assumindo que as coordenadas globais são iguais às coordenadas do sistema da câmera, a equação de disparidade entre estes dois pontos é dada por [3]

$$z'm' = zA'RA^{-1}m + A't \quad (2)$$

sendo z' , m' e A' , o valor de profundidade do pixel, a coordenada 2D da imagem em coordenadas homogêneas e a matriz intrínseca da câmera virtual, respectivamente. Enquanto que z , m e A são as respectivas informações da câmera de referência. Já R e t são os valores das matrizes de rotação e translação que fazem a deformação dos pixels da imagem de referência para a virtual.

Em sistemas retificados os eixos óticos são paralelos e as duas câmeras possuem os mesmo parâmetros intrínsecos, logo $A_1 = A_2$, $R_1 = R_2$ e $t_1 - t_2 = [L, 0, 0]^T$, sendo L a *baseline* e a disparidade vertical igual a zero. A disparidade horizontal pode ser dada, simplificada, por [5]:

$$d = Disp(z) = \frac{f_x \cdot L}{z} \quad (3)$$

onde f_x é a distância focal da câmera na direção x , horizontal, e z a profundidade real, a qual é dada pela Equação (1). Portanto, a coordenada de p_2 relaciona-se com p_1 através da disparidade, ilustrada na Figura 1:

$$m' = m \pm d. \quad (4)$$

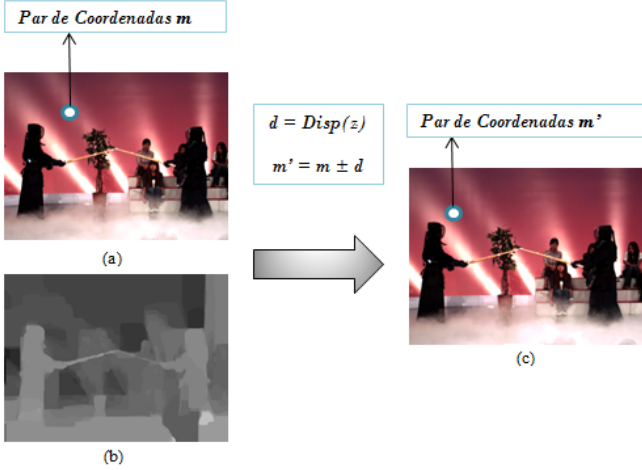


Fig. 1. Ilustração do processo de síntese de vista para a sequência Kendo. (a) Pixel na vista 1 (esquerda), com correspondente (b) valor de profundidade, é projetado para a (c) vista 2 (direita).

III. DISTORÇÃO DA PROFUNDIDADE ADMISSÍVEL

A partir das Equações (1) e (3) pode-se definir a Equação (5) da disparidade para um sistema retificado com as câmeras dispostas no eixo x [6]

$$d_x = \Phi(f_x(C_1v + C_2)) \quad (5)$$

sendo $C_1 = \frac{1}{2^n}(\frac{1}{Z_{near}} - \frac{1}{Z_{far}})$ e $C_2 = \frac{1}{Z_{far}}$. Φ é o processo de arredondamento descrito em [6], em que considera-se a precisão de pixel, podendo ser inteira, $\frac{1}{2}$ pixel ou $\frac{1}{4}$ de pixel. Durante o processo de quantização que ocorre na compressão do vídeo de profundidade, distorções na codificação podem ocorrer, introduzindo um valor Δv ao valor original v , o que conseqüentemente, poderá levar a uma distorção na disparidade dada por

$$\Delta d_x = g(v, \Delta v) = \Phi(Lf_x(C_1v + C_2)) - \Phi(Lf_x(C_1(v + \Delta v) + C_2)). \quad (6)$$

De acordo com Zhang et al. [6], diversos valores de distorções Δv inseridos em v_i (posição i) poderão provocar um mesmo valor de variação Δd_x por conta do processo de arredondamento, caracterizando um mapeamento “muitos-para-um”. Desta forma, muitos valores de profundidade v pertencem ao intervalo ADD definido como $[v^{min}, v^{max}]$ onde serão todos mapeados para uma mesma disparidade. Os limites do intervalo são definidos pelo máximo valor Δv^{max} tal que $v^{min} = v - \Delta v^{max}$ e $v^{max} = v + \Delta v^{max}$. Esta propriedade é chamada de Distorção da Profundidade Admissível (ADD) [6], [7].

A. Agrupamento de profundidade usando ADD

O pré-processamento proposto em [5], [6] e [7] baseia-se na propriedade ADD que realiza o mapeamento “muitos-para-um”. Como mostra a Figura 2, haverá vários valores de profundidade, variando entre 0-255, que originarão idênticos valores de disparidade, entre 0-30 normalmente [6]. Pode-se observar que a relação caracteriza uma função monotonicamente crescente. Com isso, vários valores de profundidade podem ser agrupados, de forma que para um determinado grupo seus valores podem ser substituídos por um único que os representará.

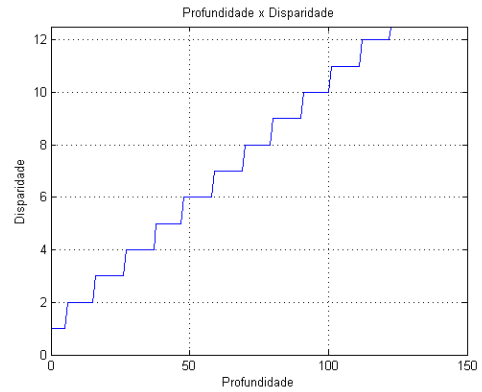


Fig. 2. Comportamento do cálculo de disparidade, Equação (5), em função de profundidade para a sequência Kendo, câmera 1.

Logo, a eficiência de codificação aumenta na medida que a variação dos valores de profundidade diminui. Portanto, dado que é possível trocar o valor original por qualquer outro no intervalo ADD ao qual pertença, escolheu-se substituir todas as profundidades que recaem num mesmo intervalo por um único valor dele. Com isso, a entropia do mapa de profundidade diminui significativamente em relação ao mapa original. Adotou-se, então, o meio do intervalo ADD para a substituição. Os valores de profundidades v'_i são, conseqüentemente, escolhidos baseados no conjunto

$$V_i = \{v^{min}, v^{min} + 1, \dots, v^{max}\} \quad (7)$$

conforme

$$v'_i = \text{mediana}(V_i). \quad (8)$$

Desta forma, é montado um mapa de profundidade com a menor entropia de profundidade.

B. Minimização da Variância

Na codificação de imagens com, por exemplo, o codec *High Efficiency Video Coding* (HEVC) [10], um menor valor de resíduo por unidade de codificação (UC's) é desejável. Portanto, este trabalho propõe diminuir a variância dentro de uma mesma UC, tornando-a mais homogênea, com isso mostra que uma menor entropia nem sempre resulta em maior eficiência de compressão. Essa proposta é válida para qualquer codec baseado em codificação de blocos, não importando a escolha pelo codificadores específicos para múltiplas vistas,

uma vez que este pré-processamento atua somente nos mapas de profundidade e tem como objetivo a otimização da codificação local destes mapas. Utilizamos a propriedade ADD para escolher um valor de profundidade em função de seus vizinhos espaciais dentro do bloco a ser codificado.

Desta forma, escolhe-se o valor que cause a mínima variância em relação ao valor mediano do bloco (ou UC) a qual aquela profundidade está localizada, respeitando os limites do intervalo ADD, ou seja, causando a mesma disparidade. Esta abordagem difere da Seção III-A que realiza sua escolha com o intuito de se diminuir a entropia total do mapa, sem levar em consideração as redundâncias espaciais que ocorrem no momento da codificação por blocos. Denotemos, então, o valor mediano do bloco como m_i

$$m_i = \frac{\sum_{j \in B_i} v_j}{N \times N} \quad (9)$$

onde B_i é o bloco que contém o pixel i . Em nossa implementação os blocos são de dimensão (N) 64×64 .

Dado que temos um conjunto V_i dentro do qual a profundidade v_i poderá variar, o valor baseado na menor variância será dado conforme

$$v_i'' = \arg \min_{a \in V_i} |m_i - a|. \quad (10)$$

IV. METODOLOGIA

Num primeiro momento, foi criada a síntese de vistas a partir das equações da seção II. A Equação (4) foi então aplicada a cada um dos pixels da imagem da esquerda e da direita, sendo projetada a imagem entre as duas câmeras. Com isso, construiu-se as imagens sintetizadas que serão comparadas utilizando os mapas de profundidade originais e pré-processados.

O próximo passo é então pré-processar os mapas de profundidade de todas as sequências para que os valores dentro de um mesmo macrobloco estejam o mais uniforme, para que assim o erro de predição esteja também mais uniforme, causando uma maior taxa de compressão. Então, esta abordagem é comparada com o método de se escolher apenas um valor no intervalo, substituindo todas as profundidades por este único valor. Isso fará, conseqüentemente, com que a entropia total do mapa de profundidade diminua, em oposição ao método da mínima variância, que apresenta uma entropia maior dos dados.

V. RESULTADOS

O método proposto neste artigo, que procura escolher o valor que cause Mínima Variância no Bloco (MVB) de acordo com a Equação (10) é comparado à simulação implementada da proposta definida como ADD-M dada pela Equação (8).

O conjunto de teste foi composto por bases de dados públicas onde foram testadas as sequências Kendo, Balloons [11], Poznan Street [12], Pantomime [13] [11] e Lovebird1 [14]. As sequências estão regularmente dispostas em um arranjo 1D, contendo várias resoluções e distância de *baseline* diferentes, descritas na Tabela I. São utilizados apenas o primeiro quadro das sequências pois não há relações temporais

nos métodos. A codificação de todas as imagens foi feita usando HEVC [10] em modo sem perdas (*lossless*). Utilizou-se para a síntese de vistas, a técnica DIBR implementada no Software de Referência para Síntese de Vistas (VSRS 3.5) [15], com precisão inteira em modo 1D paralelo.

Para medir a eficiência dos testes realizados foi utilizada a taxa média de bits gastas por pixel, por câmera [bppc] para os mapas de profundidade comprimidos após processamento. Consideramos o tamanho médio entre as duas câmeras de referência, usadas para síntese, após compressão. Os resultados são apresentados na Tabela II. Estes valores são bem menores do que 8 bits/pixel porque como os mapas de profundidades possuem em sua maioria poucos valores, quando codificados são necessários menos símbolos, mesmo que seja uma codificação *lossless* ou entrópica.

A porcentagem de ganho é calculada em relação ao quanto os métodos foram capazes de comprimir em relação à compressão dos mapas originais e são apresentados na Tabela III. As sínteses geradas pelas técnicas a partir da compressão são sem perdas, ou seja, mantém-se a propriedade citada na Seção II. Portanto, os PSNRs das sínteses geradas tanto pelos mapas originais quanto pelos mapas pré-processados, quando comparados em relação à imagem *ground truth* (imagem da câmera real naquela posição) são idênticos.

O ganho médio do método MVB em relação ao método ADD-M para as sequências testadas foi de 3.61%. Sendo a sequência Kendo a que obteve menor ganho em relação ao método ADD-M, com ganho de 0.51 %. Enquanto que o maior ganho foi da sequência Pantomime com 8.37 %, que possui um fundo bastante homogêneo, em sua maioria de um único valor.

TABELA I
DESCRIÇÃO DAS SEQUÊNCIAS USADAS PARA TESTE

Sequência	Resolução	Vistas Referência	Vista Sintetiz.
Kendo	1024x768	1,3	2
Balloons	1024x768	2,4	3
Poznan St.	1920x1088	3,5	4
Pantomime	1280x960	37,39	38
Lovebird1	1024x768	4,6	5

TABELA II
COMPRESSÃO, EM BPPC, DOS MAPAS DE PROFUNDIDADE ORIGINAIS E PROCESSADOS COM ADD-M E MVB.

Sequência	Original	ADD-M	MVB
Kendo	0,111	0,085	0,084
Balloons	0,167	0,114	0,113
Poznan St.	0,477	0,331	0,318
Pantomime	0,099	0,068	0,060
Lovebird1	0,292	0,141	0,124

VI. CONCLUSÃO

A proposta apresentada consegue atingir ganhos até 8% superiores em relação às taxas de compressão dos mapas de

TABELA III

GANHO DE COMPRESSÃO, EM %, DOS MAPAS DE PROFUNDIDADE PROCESSADOS COM ADD-M E MVB EM RELAÇÃO AOS ORIGINAIS.

Sequência	ADD-M	MVB
Kendo	23.64	24.15
Balloons	31.45	32.02
Poznan St.	30.46	33.23
Pantomime	30.97	39.34
Lovebird1	51.70	57.52

profundidades utilizando o método ADD-M. O ganho médio para as sequências testadas foi de 3.61 %.

Vale ressaltar que apesar de os mapas de profundidade apresentarem menor entropia quando pré-processados com o método ADD-M, os resultados do pré-processamento com o método MVB mostrou-se mais eficiente em compressão dos dados, isto se deve ao fato de se explorar as vizinhanças dos pixels, sabendo que se trata de um codificador de blocos, é possível tirar proveito das redundâncias espaciais que o mesmo utiliza no momento da compressão. Ao tornar o mapa mais homogêneo em relação aos seus blocos de codificação ganha-se mais em eficiência de compressão.

Considerando-se que foi realizado os testes somente para o primeiro quadro dos vídeos, sendo que alguns mapas comprimiram mais que outros, devido à sua estrutura e distribuição de profundidades, este ganho deverá aumentar ainda mais com a compressão do vídeo inteiro. Com isso, têm-se como trabalhos futuros a implementação e teste para o pré-processamento e codificação das sequências inteiras e a variação do tamanho do bloco de codificação para buscar uma melhor otimização de homogeneidade de valores.

AGRADECIMENTOS

Agradecimento à Fundação de Apoio a Pesquisa do Distrito Federal (FAP-DF) que possibilitou a apresentação deste trabalho.

REFERÊNCIAS

- [1] Peter Kauff, Nicole Atzpadin, Christoph Fehn, Marcus Müller, Oliver Schreer, Aljoscha Smolic, and Ralf Tanger, "Depth map creation and image-based rendering for advanced 3d tv services providing interoperability and scalability," *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 217–234, 2007.
- [2] Cheon Lee and Yo-Sung Ho, "View synthesis using depth map for 3d video," in *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*. Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, International Organizing Committee, 2009, pp. 350–357.
- [3] Christoph Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 93–104.
- [4] Philipp Merkle, Aljoscha Smolic, Karsten Müller, and Thomas Wiegand, "Multi-view video plus depth representation and coding," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. IEEE, 2007, vol. 1, pp. I–201.
- [5] Yin Zhao, Ce Zhu, Zhenzhong Chen, and Lu Yu, "Depth no-synthesis-error model for view synthesis in 3-d video," *Image Processing, IEEE Transactions on*, vol. 20, no. 8, pp. 2221–2228, 2011.
- [6] Ye Zhang, Sam Kwong, Song Hu, and C-C Jay Kuo, "Efficient multiview depth coding optimization based on allowable depth distortion in view synthesis," *Image Processing, IEEE Transactions on*, vol. 23, no. 11, pp. 4879–4892, 2014.
- [7] Yun Zhang, Zhaoqing Pan, Yang Zhou, and Linwei Zhu, "Allowable depth distortion based fast mode decision and reference frame selection for 3d depth coding," *Multimedia Tools and Applications*, pp. 1–20, 2015.
- [8] *ITU-T Recommendation and International Standard of Joint Video 670 Specification*, ITU-T Rec H.264/ISO/IEC 14496-10 AVC, March 2005.
- [9] Dong Tian, Po-Lin Lai, Patrick Lopez, and Cristina Gomila, "View synthesis techniques for 3d video," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2009, pp. 74430T–74430T.
- [10] *High efficiency video coding*, ITU-T Recommendation H.265, October 2014.
- [11] M. Tanimoto, M. Fujii, and K. Fukushima, "1D parallel test sequences for MPEG-FTV," in *ISO/IEC JTC1/SC29/WG11 MPEG 2008/M15378*, Archamps, France, April 2008.
- [12] M. Domanski, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K. Wegner, "Poznan multiview video test sequences and camera parameters," in *ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050*, Xian, China, October 2009.
- [13] M. Tanimoto, M. Fujii, T. Senoh, T. Aoki, and Y. Sugihara, "Test sequences with different camera arrangements for call for proposals on multiview video coding," in *ISO/IEC JTC1/SC29/WG11 MPEG 2005/M12338*, Poznan, Poland, July 2005.
- [14] *ETRI(Electronics and Telecommunications Research Institute)*, MPEG-Korea Forum 161 Gajeong, Yuseong Gu, Daejeon, 305-700, Republic of Korea.
- [15] M. Tanimoto, M. Fujii, T. Suzuki, K. Fukushima, and N. Mori, "Reference softwares for depth estimation and view synthesis," in *ISO/IEC JTC1/SC29/WG11 MPEG 2008/M15377*, Archamps, France, April 2008.