

Métrica objetiva para avaliação do conforto na visualização de vídeos estereoscópicos

Marcelo A. Miguel^{1,2}, Eduardo A. B. Silva¹, Carla L. Pagliari³

Resumo—Este artigo apresenta um método automático objetivo para avaliação do conforto na visualização de vídeos estereoscópicos baseado na geometria da visão estéreo, visando auxiliar o ajuste do conjunto de câmeras para captação do conteúdo. O método proposto utiliza-se da técnica SIFT para estimar, através da segmentação dos quadros dos vídeos, qual a proporção em cada cena de regiões exibidas que causam desconforto no observador.

Palavras-Chave— Estereoscopia, Qualidade, Conforto.

Abstract—This article presents a method for automatically assessing the comfort associated with viewing stereoscopic videos based on the stereo vision geometry, to assist the camera stereo rig adjustment. The proposed method estimates, using SIFT and segmentation techniques, the extent of the frame on each scene which causes eye strain to the viewer.

Keywords— Stereoscopia, Quality Assessment, Visual Comfort.

I. INTRODUÇÃO

Na última década o desenvolvimento tecnológico da indústria eletrônica permitiu o advento e comercialização de televisores capazes de reproduzir conteúdo estereoscópico a preços próximos aos dos televisores convencionais[1]. Houve também um substancial aumento na quantidade de salas de cinema que dispõem desse tipo de tecnologia.

Uma das maiores preocupações dos produtores de conteúdo é a de permitir aos usuários a melhor experiência tridimensional(3D) possível, sem causar incômodos como enjôos, dores de cabeça ou tonturas [2][3][4].

A criação de conteúdo para o cinema e televisão têm se desenvolvido rapidamente, com cada vez mais filmes lançados em 3D. Uma das dificuldades da produção de conteúdo estereoscópico é o longo tempo necessário para realizar os ajustes do arranjo de câmeras para a captação desse conteúdo [4].

Além da dificuldade do ajuste, ainda existe dificuldade na avaliação da qualidade do efeito 3D durante o processo de captação, fazendo com que a pós-produção do vídeo tenha que incluir algumas correções no conteúdo visando evitar efeitos indesejáveis da estereoscopia [4].

A percepção 3D produzida por um vídeo estereoscópico é gerada a partir da exibição de 2 diferentes sinais de vídeo: um para o olho esquerdo e um para o olho direito. As diferenças entre os vídeos exibidos devem simular as diferenças percebidas naturalmente por conta da distância entre os olhos. É

importante salientar que a percepção 3D só acontece se ao visualizar as 2 imagens, o observador seja capaz de fundí-las em algo inteligível, ou seja, que a convergência das imagens de um mesmo objeto retrate de maneira fidedigna a visualização real desse objeto [5].

Este artigo apresenta uma métrica objetiva para avaliação de conforto de visualização de vídeos estereoscópicos. Ou seja, o método proposto tem o objetivo de obter a “Quality of Experience”(QoE) de estímulos 3D. O método proposto define o afastamento do par de cameras do sistema estereoscópico que seja mais confortável para o observador humano, sem que sejam necessários testes subjetivos. Os resultados são comparados com os dos testes subjetivos reportados em [6] usando a mesma base de dados. Além disso, os resultados da métrica proposta são comparados com os resultados obtidos em [7].

II. GEOMETRIA DA PERCEPÇÃO 3D

Para que uma pessoa assistindo a um vídeo estereoscópico tenha a sensação de tridimensionalidade, é necessário que o vídeo obedeça uma série de restrições, conforme será demonstrado a seguir. A percepção de tridimensionalidade ocorre quando imagens deslocadas de um mesmo objeto são exibidas para cada olho [8]. Dependendo da profundidade do objeto na cena, o deslocamento será maior ou menor. Esse deslocamento é denominado disparidade (d_{imagem}), e é representado pela diferença entre a posição do ponto na imagem esquerda P_l e direita P_r , conforme a equação 1[9][10].

$$d_{imagem} = P_r - P_l \quad (1)$$

Nas duas imagens do mesmo objeto, os pontos que correspondem a uma mesma localização da cena são denominados pontos homólogos, ou pontos equivalentes. A fusão dessas imagens no cérebro gera a sensação de profundidade. A figura 1 exibe um exemplo da geometria da visão estéreo.

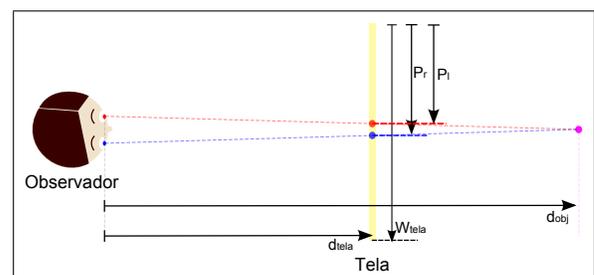


Fig. 1. Dimensões para o cálculo da geometria da visão estéreo.

¹ PEE/COPPE/DEL/EE, Univ. Fed. do Rio de Janeiro, Brazil; ² TV Globo/DGEN/DEPC/DOSRF, Rio de Janeiro, Brazil; ³ IME, Instituto Militar de Engenharia, Brazil. E-mails: marcelo.miguel@globo.com, eduardo@ips.ufrj.br, carla@ime.eb.br

Considerando-se que os pontos equivalentes de um mesmo objeto estejam alinhados na vertical, podemos dividir a localização da percepção dos objetos em cena em 3 espaços distintos:

1. Sobre a tela de projeção ou exibição.
2. Atrás da tela de projeção ou exibição.
3. A frente da tela de projeção ou exibição.

Para que os objetos sejam percebidos atrás da tela de exibição, sua imagem direcionada ao olho esquerdo deve estar localizada à esquerda da imagem direcionada ao olho direito, nesse caso, vê-se pela equação 1 que a disparidade é positiva. Já para o caso onde os objetos são percebidos à frente da tela de exibição, a imagem referente ao olho esquerdo está à direita da imagem exibida para o olho direito, logo a disparidade é negativa. Quando o objeto é percebido sobre a tela, as imagens são coincidentes e a disparidade é nula[11]. A figura 2 exhibe um exemplo dos 3 casos citados.

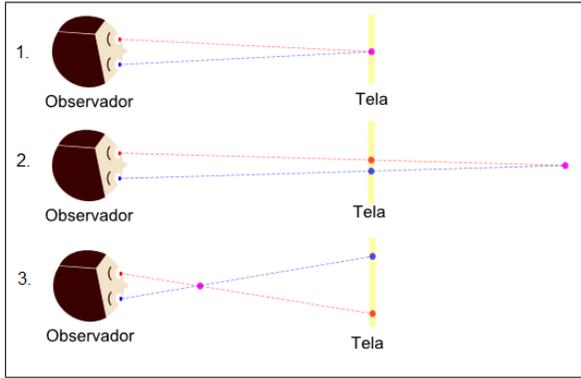


Fig. 2. 1. Objeto percebido sobre a tela, 2. Objeto percebido atrás da tela, 3. Objeto percebido a frente da tela.

Por semelhança de triângulos pode-se mostrar que a distância entre o observador e o objeto é dada pela equação 2 [12][10].

$$d_{obj} = \frac{d_{olhos} \cdot d_{tela}}{d_{olhos} - d_{imagem}} \quad (2)$$

Onde d_{objeto} é a distância do observador à imagem percebida do objeto, d_{olhos} é a distância entre os olhos do observador (cerca de 64 mm [13]), e d_{tela} é a distância entre o observador e a tela, conforme o esquema da figura 1.

Considerando a exibição de um vídeo estereoscópico com resolução de W_{pixels} pixels de largura, e a disparidade em pixels dada por d_{pixels} podemos verificar a influência do tamanho da tela de exibição (representado aqui pela largura da tela, W_{tela}) na percepção do vídeo.

$$d_{imagem} = \frac{W_{tela} \cdot d_{pixels}}{W_{pixels}} \quad (3)$$

Substituindo a equação 3 em 2, vem:

$$d_{obj} = \frac{d_{olhos} \cdot d_{tela}}{d_{olhos} - \frac{W_{tela} \cdot d_{pixels}}{W_{pixels}}} \quad (4)$$

Além disso, nos casos em que um vídeo estereoscópico pré-gravado vai ser exibido, d_{pixels} e W_{pixels} são fixos. Neste

caso, a equação 4, quando d_{pixels} é positivo, nos mostra que há um limite no tamanho da tela para a percepção de distância do objeto, uma vez que o denominador da expressão tem que ser um número positivo. Se as dimensões da tela extrapolam esse limite, o observador tem uma cena que o cérebro não será capaz de reconstruir, pois os olhos teriam que ficar em posição divergente, conforme a figura 3 [14]. A equação 5 expõe esse limite.

$$W_{tela} < \frac{d_{olhos} \cdot W_{pixels}}{d_{pixels}} \quad (5)$$

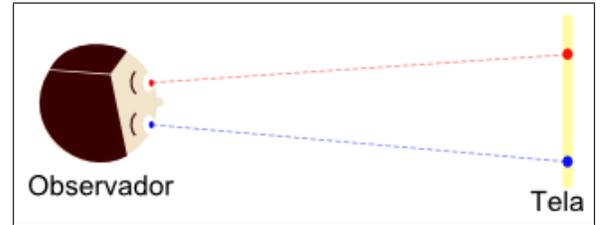


Fig. 3. Visão Divergente

Inversamente, podemos pensar na captação do filme. Levando em consideração a maior tela que o filme será exibido, temos um valor máximo que podemos ter de disparidade no vídeo (ou seja, da distância máxima que podemos ter em pixels entre a imagem do olho esquerdo e do olho direito).

$$d_{pixels} < \frac{d_{olhos} \cdot W_{pixels}}{W_{tela}} \quad (6)$$

Em certos casos há deslocamento vertical entre os pontos equivalentes das duas vistas. Esse tipo de distorção deve ser evitada, mantendo o alinhamento vertical das câmeras durante a captação. Quanto maior o deslocamento, pior é a percepção do observador. Outro fator que influencia negativamente a qualidade é a variância do deslocamento vertical entre diferentes pontos de uma mesma imagem, normalmente causado pela rotação das câmeras, ou ainda por distorções causadas pelas lentes[12].

Existem outros fatores citados por [15] que também influenciam no conforto do observador, dentre eles, a relação entre vergência, que é o quanto os olhos têm de convergir um em direção ao outro para observar os pontos equivalentes, e acomodação, que é o ajuste dos cristalinos para a focalização das imagens. Ao observar uma cena 3D real, as distâncias de vergência e a acomodação são iguais. Essa relação é denominada linha de Donder[16]. Porém nos vídeos estéreo, a acomodação é constante e fixa na tela de exibição e a vergência varia de acordo com o conteúdo da cena.

A definição de um limite de conforto para vergência e acomodação, denominado CVR (*comfortable viewing range*) foi alvo de diversos estudos [10][17][18][19][20], de forma que verificou-se que há desconforto visual quando a distância percebida do objeto (d_{objeto}) está fora da região de foco dos olhos. Essa região é de 0,2 dioptra [19][20], de forma que as distâncias máxima e mínima de percepção do objeto são dadas pelas equações 7 e 8, considerando que todas as distâncias são dadas em metros[16][21].

$$d_{obj} > \frac{1}{\frac{1}{d_{tela}} + 0,2m^{-1}} \quad (7)$$

$$d_{obj} < \begin{cases} \frac{1}{\frac{1}{d_{tela}} - 0,2m^{-1}} & , \text{ se } d_{tela} < 5m \\ \infty & , \text{ caso contrário} \end{cases} \quad (8)$$

Substituindo d_{objeto} nas equações 7 e 8 pela expressão da equação 5, e colocando em função de d_{pixels} , vêm:

$$d_{pixels} > -\frac{0,2m^{-1} \cdot d_{olhos} \cdot d_{tela} \cdot W_{pixels}}{W_{tela}} \quad (9)$$

$$d_{pixels} < \frac{0,2m^{-1} \cdot d_{olhos} \cdot d_{tela} \cdot W_{pixels}}{W_{tela}}, \text{ se } d_{tela} < 5m \quad (10)$$

III. PROPOSTA DE MÉTRICA OBJETIVA

A equação 4 mostra que a percepção de profundidade da estereoscopia durante a exibição do conteúdo está relacionada a 2 grandezas: o tamanho da tela de exibição, e a distância do observador para a tela.

Durante o processo de captação das imagens, nem sempre será possível para a equipe técnica dispor de uma tela de tamanho suficiente para verificar durante a gravação se o produto gravado gera os efeitos de profundidade aos quais ele se propõe. Observar o mesmo vídeo em um monitor de 9 ou 14 polegadas (tamanhos comuns em estúdios de gravação), podem esconder determinados efeitos indesejáveis que aparecerão na exibição nas telas de 40 a 50 polegadas normalmente usadas nas residências.

O técnico responsável, conhecendo a geometria da exibição, deve ajustar a estrutura das câmeras e lentes de forma a garantir a inexistência de disparidades proibidas assim como de distorções verticais causadas pelo desalinhamento vertical das câmeras ou ainda da rotação das câmeras em relação ao plano da tela [4].

Uma forma de tentar minimizar a presença de conteúdo inadequado para a percepção estereoscópica é gerar um modelo de análise do vídeo, de forma a identificar automaticamente as regiões que não se adequam às restrições de exibição.

Se para cada pixel de cada objeto da cena captada para o olho direito for possível determinar o seu correspondente na cena captada para o olho esquerdo, usando a geometria da visão estereoscópica será possível verificar se o vídeo possui ou não áreas ou objetos que gerariam problemas ao observador, dado que tem-se o tamanho da tela e a distância do observador.

Vale ressaltar que não se pode usar na análise quaisquer informações do arranjo das câmeras, ou ainda das lentes utilizadas, para a solução do problema, visto que o objetivo do algoritmo é exatamente encontrar automaticamente falhas na elaboração desse arranjo.

Uma solução para o problema seria utilizar um método automático para encontrar as correspondências entre pixels dos objetos dos vídeos da esquerda e da direita. Porém, a maior parte dos métodos desenvolvidos utiliza informações relacionadas à geometria das câmeras e das lentes utilizadas, de maneira a reduzir o espaço de busca do problema, usando a

geometria epipolar. Sem essa redução o custo computacional do método o torna proibitivo.

Como os métodos tradicionais de correspondência estéreo violam a principal premissa do método que se deseja propor, não se pode utilizá-las diretamente, sendo necessário recorrer a outras formas de abordar o problema.

O método proposto para a elaboração da métrica de qualidade de experiência de vídeos estereoscópicos é baseado na busca de correspondências através da técnica SIFT (Scale Invariant Feature Transform), o SIFT irá verificar as semelhanças entre as imagens da vista esquerda e direita, marcando os pontos correspondentes (representados na figura 1 pelos pontos vermelho e azul na tela), Depois a técnica de segmentação e busca do vizinho mais próximo irá aproximar a correspondência SIFT, que é esparsa, para uma estereocorrespondência densa.

A técnica SIFT é de um “Método para extrair descritores distintos e invariantes de imagens que podem ser usados para executar uma correspondência confiável entre diferentes vistas de um mesmo objeto ou cena”, conforme escreveu David G. Lowe, no artigo em que se descreve o método [22].

Ao extrair os descritores, as principais características apresentadas são invariância à escala e rotação, e correspondência robusta mesmo em situações que apresentem variação de iluminação, mudança de ponto da câmera, ruído, e variações de perspectiva.

Essa técnica se encaixa bem como método para o início da análise de qualidade. Isto acontece porque ela consegue extrair os descritores das imagens para determinar as correspondências estéreo de uma forma em que as mesmas estejam de acordo com as premissas definidas na formulação do problema descrita anteriormente.

A correspondência através do SIFT é realizada para cada par de quadros correspondentes do vídeo estéreo. Os pares de pontos-chaves, determinados pelo SIFT, são utilizados para a verificação da conformidade na percepção do observador. Porém, em alguns casos, a quantidade de pontos em um quadro é pequena e pouco representativa da cena como um todo. Para melhorar a quantidade de pontos da cena, os quadros vizinhos (anterior e posterior) são comparados com o atual, e em caso de não haver movimento de um determinado ponto-chave do frame vizinho, o quadro atual passa a possuir também esse ponto-chave.

Após a listagem dos pares, é verificado se cada par de pontos está localizado em uma região permitida ou proibida em relação à percepção pelo observador.

Conforme foi visto no ítem anterior, a percepção do vídeo estereoscópico está associada às dimensões da tela de exibição, e à distância do observador à tela. Com isso, podemos dividir a profundidade em três regiões diferentes, conforme ilustrado na figura 4:

1. Região divergente (DIV) - é composta pelos pontos onde não é possível recriar a imagem percebida através da convergência dos olhos, ou seja, são classificados como DIV os pontos que não atendam a restrição da equação 6.
2. Região de conflito de vergência e acomodação (NCVR) - é composta pelos pontos onde há convergência, ou

seja, onde a restrição da equação 6 é atendida, porém as equações 9 e 10 não são atendidas. Essa região gera desconforto quando há movimento em profundidade e espaço, fazendo com que a vergência varie rapidamente, forçando os músculos dos olhos [15].

3. Região de conforto (CVR) - é composta pelos pontos em que a convergência ocorre normalmente, ou seja, são classificados como OK os pontos que atendam a todas as restrições anteriores.

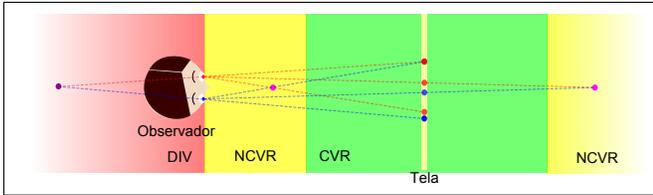


Fig. 4. Diagrama com as regiões de profundidade

A imagem da vista esquerda é submetida a uma segmentação, utilizando o software Edison, desenvolvido no Robust Image Understanding Laboratory da Rutgers University, NJ - EUA, baseado nos artigos [23], [24] e [25], configurado de forma que não haja regiões com menos de 2.000 pixels.

Para cada ponto-chave, verifica-se em qual segmento ele se encontra. Se o ponto é o único no segmento, todos os pontos do segmento passam a pertencer à mesma região do ponto original. Caso haja mais de um ponto no segmento, todos os pontos daquele segmento são distribuídos na proporção dos pontos que ali apareçam, por exemplo: se há um segmento com X pixels, e nesse segmento há L pontos CVR, J pontos NCVR e K pontos DIV, teremos então $X*L/(L+J+K)$ pontos CVR, $X*J/(L+J+K)$ pontos NCVR, e $X*K/(L+J+K)$ pontos DIV.

Algumas regiões não terão nenhum ponto-chave associado. Nesse caso, o centróide da região é calculado e verifica-se a classificação do ponto-chave mais próximo a ele. A região recebe então a classificação desse vizinho mais próximo.

A nota final é calculada a partir da quantidade de pontos classificada em cada uma das regiões, re-escalada para a escala de 0 a 100, de acordo com a equação 11. Os pontos classificados como DIV, por serem a falha mais severa[15] recebem pontuação nula. Os pontos classificados como NCVR recebem pontuação k (onde $0 < k < 1$), e os pontos classificados como CVR recebem pontuação 1. O valor de k será estimado a partir da base de testes.

$$N_{OBJ} = 100 * \frac{n_{CVR} + k * n_{NCVR}}{n_{DIV} + n_{CVR} + n_{NCVR}} \quad (11)$$

IV. TESTE E RESULTADOS

O método foi avaliado usando a base de vídeos estereoscópicos da École Polytechnique Federale de Lausanne (EPFL) [6][26].

A base é formada por 30 vídeos de 10 segundos cada, sendo divididos em 5 diferentes vídeos representando a mesma dentre 6 cenas. Cada cena foi gravada 5 vezes, a primeira vez com

as câmeras com uma distância horizontal entre elas de 10 cm, e a cada novo take a distância entre as câmeras era aumentada em 10 cm.

A resolução dos vídeos é de 1920 x 1080 pixels, e taxa de 25 quadros por segundo. As cenas possuem conteúdo gravado em áreas internas e externas, com variedades de cor, movimentos, texturas e profundidades.

Os vídeos foram submetidos pelos autores de [6], para a realização de testes subjetivos de acordo com as normas ITU-R BT.500[27], ITU-R BT.710[28] e ITU-R BT.1438[29] com um monitor estereoscópico de 46" e resolução de 1920 x 1080 pixels, com o observador posicionado a 2 metros da tela. A partir dos testes subjetivos, foram obtidas as notas MOS (*mean opinion score*) relativa à qualidade dos vídeos da base.

Os 30 vídeos foram divididos em 25 vídeos de treinamento e 5 vídeos de teste para que fosse calculado o valor ótimo de k . Todas as C_{25}^{30} combinações foram calculadas. A otimização buscou o valor de k que maximizasse a correlação de Pearson entre as notas N_{OBJ} e MOS dos vídeos pertencentes ao conjunto de treinamento. O valor de k determinado em cada uma das combinações foi aplicado ao conjunto de teste, onde foram calculados a correlação de Pearson, a correlação de Spearman, o erro médio absoluto normalizado (NMAE) e o erro médio quadrático normalizado (NMRSE). O valor médio de k é de 0,7225, com desvio de 0,0428. A figura 5 mostra o histograma dos valores de k calculados, com intervalo de 0,01 e a figura 6 mostra o histograma da correlação de Pearson calculada para cada combinação, com intervalo de 0,01. A média e o desvio das correlações e dos erros médios estão exibidos na tabela I.

TABELA I
CORRELAÇÃO E ERRO MÉDIO DAS NOTAS OBTIDAS PELO MÉTODO PROPOSTO E DAS NOTAS OBTIDAS NOS TESTES SUBJETIVOS DE [6].

	Média	Desvio
Pearson	0,8590	0,1644
Spearman	0,8058	0,2049
NMRSE	0,1753	0,0427
NMAE	0,1530	0,0428

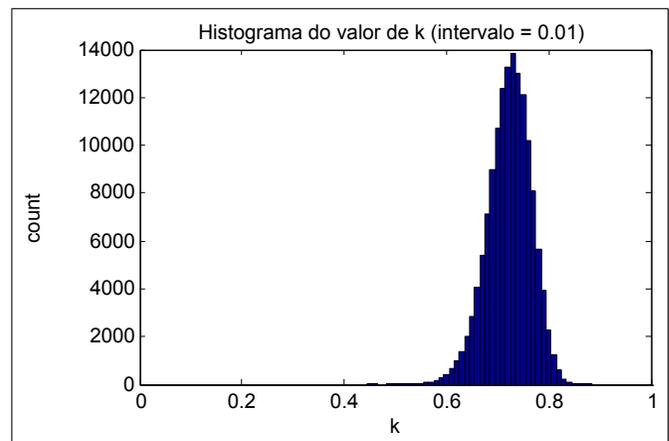


Fig. 5. Histograma dos valores de k estimados para as C_{25}^{30} combinações, com intervalo de 0,01.

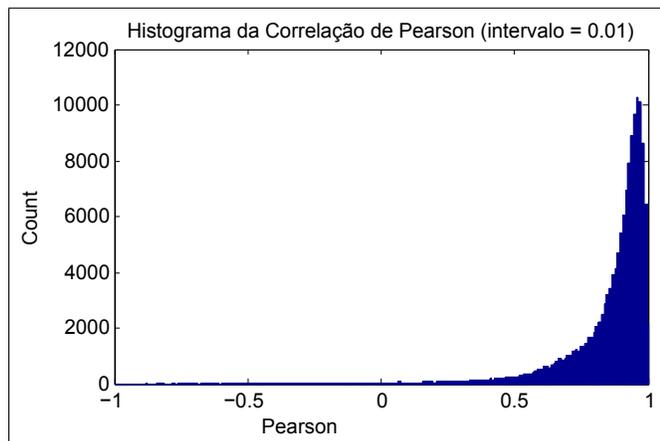


Fig. 6. Histograma da correlação de Pearson do conjunto de testes calculado para as C_{25}^{30} combinações, com intervalo de 0,01.

O método apresentado em [7] propõe utilizar as características estatísticas do mapa de disparidade e do gradiente da disparidade, além de indicadores de atividade espacial e de movimento para determinar de forma objetiva o conforto na visualização de vídeos estereoscópicos, sendo que os resultados dos testes apresentados foram obtidos para a mesma base de vídeos utilizada nesse artigo. O coeficiente de correlação de Spearman obtido em [7] tem média de 0,76 e desvio de 0,25 para o método proposto utilizando PCA, e média de 0,68 com desvio de 0,28 para o método proposto utilizando FFS. Já no método proposto nesse artigo, o coeficiente de correlação de Spearman obtido é de 0,81, com desvio de 0,20, obtendo um desempenho melhor que os métodos propostos em [7], mostrando que vale a pena continuar investigando técnicas baseadas no método proposto.

V. CONCLUSÕES

Os resultados dos testes com a base de vídeos do EPFL mostram que o algoritmo é eficiente para a detecção automática das falhas do ajuste horizontal das câmeras. Como a base não possui outros tipos de falhas, o desenvolvimento do algoritmo ficou limitado a esse problema. Porém, é importante ressaltar, que o ajuste da distância entre as câmeras é um dos pontos críticos do ajuste de cena durante a captação do conteúdo, uma vez que a distância entre as câmeras é variável de acordo com as lentes utilizadas e do conteúdo a ser captado, e é também dependente do tamanho da tela e da resolução para o qual o conteúdo é destinado (os ajustes verticais e de rotação das câmeras são fixos e não variam de cena para cena).

AGRADECIMENTOS

Agradecemos à Robson Ferreira Raiol e Lilian R. Geraldino Miguel pelo apoio na execução da pesquisa.

REFERÊNCIAS

- [1] NDP Group, *Awareness of 3D TVs and Blu-ray Players Grows, According to The NDP Group*, Port Washington, NY, USA, 11 de abril 2011.
- [2] ITU-R, *Stereoscopic television based on R-and L-eye two channel signals*, Rec. BT.1198, 1995.
- [3] L. Meesters, W. IJsselsteijn, P. Seuntjens, "Survey of perceptual quality issues in three-dimensional television systems," *Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems X*, v.5006, pp. 313–326, Janeiro 2003.
- [4] B. Mendiburu, *3D Movie Making - Stereoscopic Digital Cinema from Script to Screen*, 1st edition, Focal Press, 2009.
- [5] J. Mansson, "Stereovision: A Model Of Human Stereopsis," *Lund University Cognitive Studies*, v. 64, 1998.
- [6] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," *IS&T/SPIE Electronic Imaging, 3D Image Processing (3DIP) and Applications*, San Jose, USA, 2010.
- [7] A. Mittal, A. K. Moorthy, J. Ghosh, A. C. Bovik, "Algorithmic assessment of 3D quality of experience for images and videos," *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*, 2011 IEEE, pp 338 - 343, Sedona, AZ, USA, 4-7 Jan. 2011.
- [8] B. Julesz, *Foundations of Cyclopean perception*, The University of Chicago Press, Chicago, IL, USA, 1971.
- [9] I. Howard, B. Rogers, *Binocular Vision and Stereopsis*, New York: Oxford University Press, 1995.
- [10] D. A. Southard, "Viewing model for virtual environment displays," *Journal of Electronic Imaging*, pp. 413?-420, v. 4(4), Outubro 1995.
- [11] E. Trucco, A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, Englewood cliffs, NJ, USA, 1998.
- [12] A. Woods, T. Docherty, R. Koch, "Image Distortions in Stereoscopic Video Systems", *Proc. SPIE, Stereoscopic Displays and Applications IV*, v. 1915, pp. 36–48, San Jose, CA, EUA, 1993.
- [13] N. Dodgson, "Variation and Extrema of Human Interpupillary Distance," *Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems XI*, v. 5291, pp. 36–46, 2004.
- [14] R. Spottiswoode, N. Spottiswoode, *The theory of stereoscopic transmission & its application to the motion picture*, University of California Press, Los Angeles, CA, USA, 1953.
- [15] M Lambooi, W. IJsselsteijn, I. Heynderickx, "Visual discomfort in stereoscopic displays: A review," *Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems XIV*, v. 6490, 2007
- [16] F. Devernay and P. Beardsley, "Stereoscopic Cinema," *Image and Geometry Processing for 3-D Cinematography, volume 5 of Geometry and Computing* Springer Berlin Heidelberg, 2010.
- [17] N. Valys, *Stereoscopy*, Focal Press, London, UK, 1966.
- [18] Y. Yeh, L. Silverstein, "Limits of Fusion and Depth Judgment in Stereoscopic Color Displays," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, v. 32, pp. 45-60, Fevereiro 1990.
- [19] S. Yano, M. Emoto, T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images", *Displays*, v. 25, Issue 4, pp. 141–150, Elsevier, Novembro 2004.
- [20] D. M. Hoffman, A.R. Girshick, K. Akeley, M.S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision* v. 8(3), pp. 1–30, 2008.
- [21] W. Chen, J. Fournier, M. Barkowsky, P. Callet, "New requirements of subjective video quality assessment methodologies for 3DTV," *Video Processing and Quality Metrics 2010 (VPQM)*, Scottsdale, USA, 2010.
- [22] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, v. 60, number 2, pp. 91–110, 2004.
- [23] D. Comanicu, P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, v. 24, pp. 603–619, 2002.
- [24] P. Meer, B. Georgescu, "Edge detection with embedded confidence," *IEEE Trans. Pattern Anal. Machine Intell.*, v. 23, pp. 1351–1365, 2001.
- [25] C. Christoudias, B. Georgescu, P. Meer, "Synergism in low-level vision," *16th International Conference on Pattern Recognition*, v. IV, pp. 150–155, 2002.
- [26] L. Goldmann, F. De Simone, and T. Ebrahimi, "Impact of acquisition distortions on the quality of stereoscopic images," *5th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Scottsdale, USA, 2010.
- [27] ITU-R, *Methodology for the subjective assessment of the quality of television pictures*, Rec. BT.500-11, 2002.
- [28] ITU-R, *Subjective assessment methods for image quality in high-definition television*, Rec. BT.710-4, 1998.
- [29] ITU-R, *Subjective assessment of stereoscopic television pictures*, Rec. BT.1438, 2000.