



Avaliação da Percepção da Profundidade e Conforto Visual de Vídeos Estereoscópicos Anaglíficos Gerados a partir de Vídeos Bidimensionais.

Mikaelle Oliveira Santos² Nathália Alves Rocha Batista³, Carlos Danilo Regis⁴

¹Parte do trabalho de conclusão de curso do primeiro autor.

²Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Brasil. e-mail: mikaosantos@gmail.com

³Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Brasil. e-mail: nathaliaarb@gmail.com

⁴Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Brasil. e-mail: regis.danilo@gmail.com

Resumo: Este trabalho apresenta uma abordagem acerca da avaliação da percepção dos modelos de profundidade e conforto visual percebido pelos expectadores em vídeos tridimensionais anaglíficos. Os vídeos utilizados para tal avaliação foram vídeos em duas dimensões convertidos em vídeos tridimensionais por meio de um *software* de conversão chamado *MOVAVI Vídeo Converter*, *software* esse de conversão automática *offline*. A avaliação se deu através da métrica subjetiva utilizando-se o *método Double Stimulus Continuous Quality Scale (DSCQS)*. Os vídeos utilizados na avaliação possuíam diferentes tipos de profundidades e ambientes de cena, a exemplo de cenários de ação e natureza. O objetivo central do trabalho é avaliar como o usuário percebe aspecto dos vídeos em estudos para tipos de profundidades e cenários diferentes, para que dessa forma seja possível traçar o perfil do tipo de vídeo tridimensional de maior interesse, além de observar o conforto na visualização dos mesmos. A avaliação aconteceu em um ambiente controlado de acordo com as especificações da norma ITU-R BT. 500 trazendo resultados bastante satisfatórios além de contribuições para trabalhos futuros.

Palavras-chave: anaglíficos, conforto visual, profundidade, tridimensionais

1. INTRODUÇÃO

A capacidade do homem em interpretar pares de ilustrações ou fotos de uma mesma cena, visualizados por ângulos ligeiramente diferentes, é chamada de estereoscopia, e teve seus fundamentos lançados no século XIX (ANDRADE2009). O cérebro, automaticamente, funde as imagens capturadas para formação do vídeo em apenas uma e, nesse processo, obtém informações quanto à profundidade, distância, posição e tamanho dos objetos, gerando a ilusão de uma terceira dimensão.

A visualização estereoscópica utiliza, em alguns métodos óculos especiais, como no caso da visualização anaglífica (que utiliza óculos com filtros coloridos) e na visualização por luz polarizada (que utiliza óculos com filtros polarizadores de luz). Em outros, como no caso dos monitores auto-estereoscópicos (Monitores que dispensam o uso de óculos ou capacetes para se obter uma visão estereoscópica.), a percepção da profundidade pode ser observada sem o uso de óculos.

Os vídeos tridimensionais podem ser obtidos pela sobreposição de duas vistas de um mesmo vídeo, uma para cada olho, ou ainda, de maneira automática por meio de *softwares* de conversão que criam a segunda vista a partir de um vídeo bidimensional.

A avaliação da qualidade das imagens de vídeo 3D é alcançável por meio de avaliação subjetiva ou de métricas objetivas. Entretanto a melhor maneira de avaliar a qualidade do vídeo estereoscópico é certamente com métricas subjetivas, realizadas de acordo com protocolos de padronização definidos na norma da ITU-R BT.500, uma vez que esse tipo de avaliação esta centrada no usuário.

O Objetivo deste trabalho é avaliar o tipo de profundidade 3D preferida pelos espectadores por meio da métrica subjetiva, além da avaliação da qualidade e do conforto visual em vídeos estereoscópicos anaglíficos gerados a partir de vídeos bidimensionais(2D) .

2. MATERIAL E MÉTODOS

2.1 VÍDEO 3D

O vídeo 3D oferece uma sensação de profundidade da cena observada. Essa dita "nova funcionalidade" não é tão jovem assim (SMOLIC2005), apesar de esta tecnologia só agora tenha começado a se difundir, seus princípios e as primeiras experiências já tem mais de meio século.

O primeiro filme foi rodado em um cinema dos Estados Unidos em 1952. O grande impasse na época era o fato do grande desconforto ocular gerado pelo uso dos óculos de lentes anaglíficas, o que causou grande confusão na época. Além disso, era preciso aprimorar as salas, os dispositivos de áudio e a qualidade da tela.

Existem vários tipos de telas 3D disponíveis, com e sem óculos e, portanto, também diferentes tipos de algoritmos de renderização 3D específicos. Em visores autoestereoscópicos um sinal de vídeo e um mapa de profundidade por *pixel* são transmitidos para o usuário. A própria tela se encarrega de unir as duas imagens em apenas uma e o usuário percebe uma visão de profundidade (3D) da cena.

2.2 ESTEREOSCOPIA ANAGLÍFICA

O termo Estereoscopia é oriundo da justaposição dos termos gregos *stereo*, relativo a dois (duplo), e *scopos*, relativo a visão (observador). Em linhas gerais, a estereoscopia ou visão binocular, trata-se da simulação de duas imagens de uma cena as quais são projetadas nos olhos em pontos de observação ligeiramente diferentes. O cérebro funde as duas imagens, e nesse processo, obtém informações quanto à profundidade, distância, posição e tamanho dos objetos, gerando uma sensação de visão tridimensional.

De acordo com a literatura, existem vários tipos de estereoscopia. O método mais simples é a estereoscopia anaglífica. Essa técnica surge de uma mesma imagem plana, cujo relevo se obtém por dois canais separados por cores complementares, criando o se que chama de anaglifo.

Esta técnica caracteriza-se por colorizar com uma cor primária diferente cada uma das imagens referentes a cada olho, de modo que o espectador possa separar cada uma das imagens que se encontram misturadas na tela utilizando óculos com uma lente vermelha e outra ciano (mancini1994).

Para a codificação deste tipo de vídeo estereoscópico é necessário separar os canais RGB dos vídeos do par estereoscópico, sendo que do vídeo que corresponde à visão do olho esquerdo é extraída a informação do canal vermelho, e do vídeo que corresponde à visão do olho direito extrai-se os canais azul e verde, como mostrado na Figura 1. Compõe-se então com a componente vermelha da visão do olho direito e das componentes azul e verde da visão do olho esquerdo um novo vídeo RGB. A imagem anaglífica resultante pode ser observada na Figura 2.

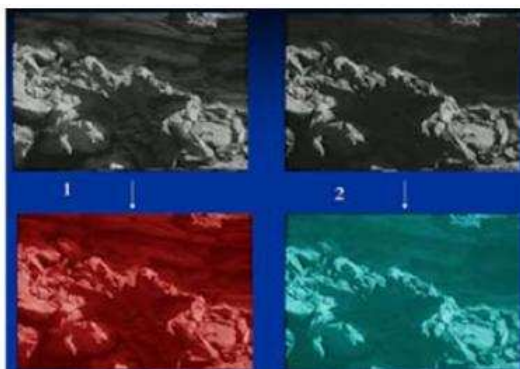


Figura 1 - Processo de extração do Canal Vermelho e dos Canais Verde e Azul do Vídeo 3D.

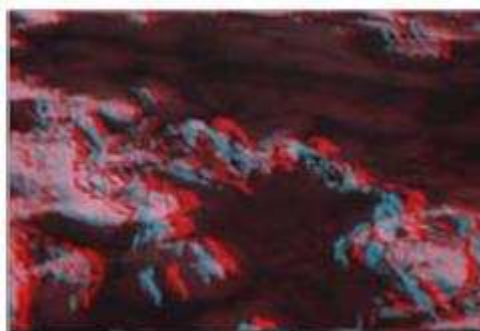


Figura 2 - Exemplo de um quadro anaglífico.

2.3 CONVERSÃO 2D PARA 3D

A conversão de imagens/vídeos de 2D para 3D vem como uma alternativa para solucionar o problema da geração de conteúdos, necessária para se implantar a 3DTV. Desta forma, a vasta coleção de materiais em 2D que existe atualmente, na forma de programas de televisão e filmes para cinema, e a sua conversão em imagens estereoscópicas deve minimizar esse efeito.

Várias abordagens para a conversão de 2D para 3D têm sido propostas. Estas abordagens podem ser classificadas em três esquemas: conversão manual, humana assistida e automática (ZHANG2011). O sistema manual é para mudar os *pixels* na horizontal com um valor de profundidade escolhidos para diferentes regiões/objetos na imagem, gerando uma nova imagem (HARMAN2009), o que demanda muito tempo.

O esquema humana assistida converte imagens 2D para 3D estereoscópico com correções feitas "manualmente" por um operador (ZHANG2011). Mesmo que este esquema reduza o tempo consumido em comparação com o regime de conversão manual, uma quantidade significativa de esforço humano ainda é necessária para concluir a conversão.

Para converter a vasta coleção de materiais disponíveis de 2D para 3D de uma forma econômica, um esquema de conversão automática é desejada. O esquema de conversão automática explora informações detalhadas dos objetos da cena originadas de uma única imagem ou de um fluxo de imagens, para gerar uma nova projeção da cena com uma câmera virtual de um ponto de vista um pouco diferente (na horizontal deslocado).

2.4 AVALIAÇÃO SUBJETIVA

Em vídeos 3D, a extração de cores e mapas de profundidade tem atraído grande atenção nos últimos 10 anos, já que pode reduzir os requisitos de armazenamento e largura de banda para a transmissão de conteúdo estereoscópico em canais sem fio, como redes celulares.

A fim de recolher dados significativos que são confiáveis e estatísticos, testes subjetivos devem ser cuidadosamente projetados e realizados, e requerem um número relevante de participantes.

Estes testes são bastante demorados, no entanto, são fundamentais, pois tentam prever a qualidade de vídeo 3D por meio da percepção humana, analisando o vídeo em 3D. Durante a fase de testes, o conforto torna-se essencial, pois alguns observadores apresentam fadiga visual com sintomas como cansaço visual, dor de cabeça ou náuseas. Este efeito é muitas vezes medido por meio de questionários (POLONEN2010).

Esses questionários ajudam a identificar as necessidades, exigências, expectativa de vídeo 3D em geral qualidade, conforto visual, a percepção de preferência, e satisfação dos usuários 3D (UMAR2011).

Neste trabalho, a avaliação se dá através do método Double Stimulus Continuous Quality Scale (DSCQS) (ITU_500-12), onde será apresentada ao espectador a exibição de dois vídeos ou dois estímulos, no entanto, não será utilizada uma referência do mesmo vídeo sem codificação, e sim um

outro vídeo também codificado, de forma a obter um comparativo entre a qualidade dos mesmos.

2.5 O EXPERIMENTO

O primeiro ponto foi a escolha dos vídeos que fariam parte da base de avaliação, todos em 2D e no formato avi. A escolha se deu de forma analítica e de acordo com a norma ITU-R BT.500, em que foram observadas as características de profundidade das cenas, realismo e computação gráfica.

Foram escolhidos 6 (seis) vídeos, contemplando três aspectos diferentes: 2 (dois) vídeos sobressaltando profundidade externa, 2 (dois) vídeos representando as cenas da natureza, um com profundidade externa e outro com profundidade interna e externa e 2 (dois) vídeos evidenciando a profundidade interna, de modo que o avaliador pudesse visualizar vídeos estereoscópicos de naturezas diversas.

O grupo de avaliadores foi composto de 15 pessoas, todos submetidos ao teste de acuidade visual (Teste de Ishihara). O tempo de duração da avaliação foi de aproximadamente 20 minutos, os vídeos foram exibidos e avaliados um por vez. Os vídeos utilizados foram Resident Evil Afterlife (Vídeo 1), A Origem (Vídeo 2), Flores Desabrochando (Vídeo 3), The Matrix Revolutions (Vídeo 4), O Senhor dos Aneis. O Retorno do Rei (Vídeo 5) e Flores se abrindo (Vídeo 6), sendo estes escolhidos por possuírem os critérios conforme os padrões do (ITU, 2009).

Os vídeos 1 e 4 evidenciam profundidade para fora, como uma espécie de *zoom* para o espectador, os vídeos 2 e 5 tem profundidade interna, e os vídeos 3 e 6 exaltam cenas da natureza, sendo o vídeo 3 produzido com profundidade externa e o vídeo 6 com os tipos de profundidades.

O experimento foi realizado em um dispositivo Samsung 32". Os Avaliadores foram posicionados em frente da tela. A distância de visualização foi de 2 m a partir do plano da tela conforme a recomendação da ITU-R BT.710.

3. RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os gráficos que reúnem os resultados da avaliação. Cada um, referente a um critério avaliado de acordo com a norma da ITU-500: Qualidade do Vídeo, Nível de Conforto, Profundidade Percebida, Impacto da Dominância do Olho e Nível de Satisfação.

O critério Impacto da Dominância do Olho, objetiva observar se as lentes anaglíficas dos óculos produzem algum tipo de predileção de visão, ou seja, se o espectador consegue enxergar melhor a cena por uma apenas uma das visões.

Esses critérios têm por finalidade avaliar como o usuário percebe aspecto dos vídeos em estudos para tipos de profundidades e cenários diferentes, para que dessa forma seja possível traçar o perfil do tipo de vídeo tridimensional de maior interesse, além de observar o conforto na visualização dos mesmos.

Em relação ao critério Qualidade do Vídeo os vídeos 1 e 4 obtiveram os melhores resultados com médias 4,13 e 4,2 respectivamente como apresentado na Figura 3 e desvio padrão 0,74 e 0,68 respectivamente. Esses vídeos possuem as mesmas características 3D com profundidade externa (semelhante ao *zoom*), além de apresentarem cenas de ação.

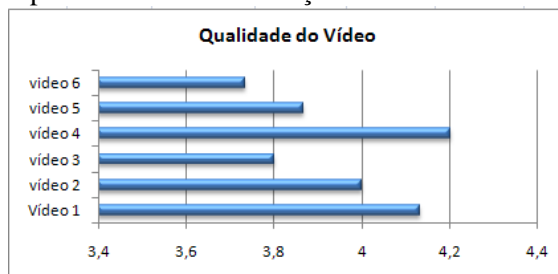


Figura 3 - Avaliação do Critério Qualidade do Vídeo.

Já os vídeos 3 e 6 foram considerados os piores com médias 3,8 e 3,73 respectivamente como apresentado na Figura 3 e desvio padrão 0,56 e 0,80. Esses vídeos também possuem características 3D semelhantes, o vídeo 3 possui profundidade externa e o vídeo 6 profundidades interna e externa, no entanto diferentemente dos vídeos considerados os melhores (1 e 4), estes remetem a cenas da natureza.

Em relação ao critério Nível de Conforto o vídeo 4 também se destaca com média 3,53 e desvio padrão 0,64, seguido do vídeo 5 com média 3,2 e desvio padrão 0,56, como apresentado na Figura 4. Estes vídeos possuem características 3D diferentes em que o vídeo 4 ressalta profundidade externa enquanto o vídeo 5 revela profundidade interna, no entanto os dois vídeos apresentam cenas de ação.

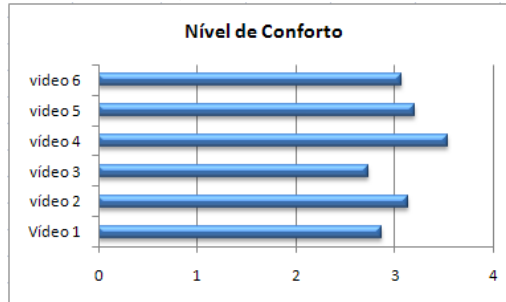


Figura 4 - Avaliação do Critério Nível de Conforto.

Para o critério Profundidade Percebida o vídeo 4 se destaca com média 4,66 como mostrado na Figura 5 e desvio padrão 0,49, seguido do vídeo 2 com média 4,4 e desvio padrão 0,51. Os vídeos tidos como menos profundos foram novamente os vídeos que remetem cenas da natureza (vídeos 3 e 6) com média 3,66 e 3,8 respectivamente como mostrado na Figura 5 e desvio padrão 0,49 e 0,68, vídeos estes que apresentam um baixo nível de profundidade.

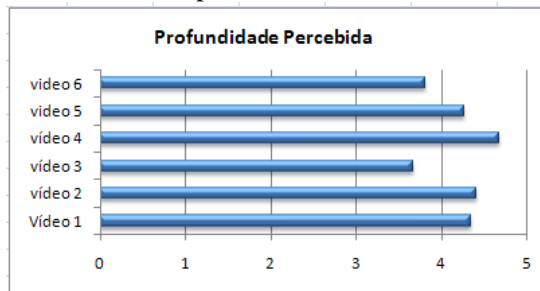


Figura 5 - Avaliação do Critério Profundidade Percebida.

Em relação ao critério Impacto da Dominância do Olho o vídeo 3 mostrou-se melhor com média 1,6 e desvio padrão 0,51, seguido do vídeo 2 com média 1,8 e desvio padrão 0,77, como apresentado na Figura 6. Este critério tem por finalidade observar se as lentes anaglíficas produzem algum tipo de predileção de visão, ou seja, se o espectador consegue enxergar melhor a cena por uma apenas uma das visões.

Os vídeos com maior impacto foram o 4 e 5 com média 2,8 e 2,26, respectivamente, e desvio padrão 0,67 e 0,80. Apesar de estas vídeos tratarem de tipos de profundidades diferentes (vídeo 4 com profundidade externa e vídeo 5 com profundidade interna), ambos possuem cenários com grande quantidade de cores escuras.

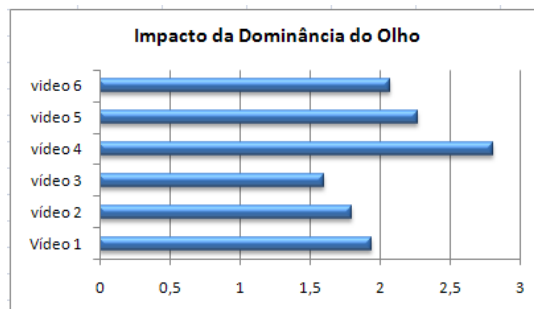


Figura 6 – Avaliação do Critério Impacto da Dominância do Olho.

Por fim, o vídeo com maior Nível de Satisfação por parte dos avaliadores foi o vídeo 1 com média 4,4 e desvio padrão 0,63, seguido do vídeo 4 com média 4,33 e desvio padrão 0,72 como apresentado na Figura 7. Os vídeos que apresentaram os menores níveis de satisfação foram os vídeos 3 e 6 com média 3,86 e 3,73, respectivamente, e desvio padrão 0,74 e 0,70, por tratarem de cenas com baixo nível de profundidade.

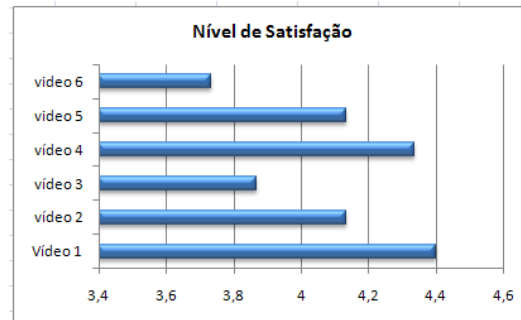


Figura 7 – Avaliação do Critério Qualidade da Experiência.

6. CONCLUSÕES

Ao fim da avaliação pode-se concluir que avaliar vídeos em 3D com métricas subjetivas, demanda esforços tanto na preparação da cena que se deseja avaliar, no cenário escolhido para exibição, na seleção dos vídeos a serem exibidos, quanto do próprio espectador humano, pois é preciso estar atento a vários critérios estabelecidos no questionário de acordo com a norma (ITU_500-12).

O indivíduo convidado à avaliação teve a sua disposição vídeos com características 3D diferentes, e avaliando os resultados obtidos podemos observar que os vídeos que apresentavam cenas mais dinâmicas à exemplo de cenas de ação e uma profundidade maior, vídeos 1, 4, e 5 e vídeos foram os que tiveram melhor aceitação, obtendo assim as melhores pontuações, Ficando constatado que os usuários preferem cenas com maior nível de profundidade, principalmente a externa.

Em relação ao desconforto ocasionado pelo uso dos óculos, e o impacto da dominância do olho, observamos que houve poucas queixas, alguns usuários sequer sentiram algum tipo de desconforto.

O Resultado final da avaliação subjetiva mostrou-se bastante satisfatório, uma vez que nenhum vídeo obteve média inferior a 3, e de um modo geral o nível de satisfação por parte dos usuários foi bastante positivo para todos os vídeos exibidos, com médias variando entre 3,73 e 4,4.

Dando continuidade ao trabalho desenvolvido, pretende-se realizar avaliações subjetivas em vídeos com interferência de vários tipos de ruído a fim de observar o impacto dos mesmos para o espectador, além de desenvolver métricas para avaliações objetivas, na tentativa de uma avaliação mais simples e rápida, no entanto, igualmente eficiente.

REFERÊNCIAS

ANDRADE, L. A.; GOULARTE, R. **Percepção Estereoscópica Anaglífica em Vídeos Digitais Comprimidos com Perda**. Webmedia – Brazilian .Symposium on Multimedia and the Web, Fortaleza, 2009.

HARMAN, P. **Home based 3D entertainment-an overview**. In Proceedings International Conference on Image Processing, 2000., volume 1, pages 1 - 4 vol.1, 2000.

ITU-R. **Recommendation BT.500-11: Methodology for the subjective assessment of the quality of television pictures**, 2002.

ITU-R. **Recommendation bt.500-12: Methodology for the subjective assessment of the quality of**



television pictures, 2009.

MANCINI, A. **Disparity estimation and intermediate view reconstruction for novel applications stereoscopic video.** Master's thesis, McHill University., 1994.

POLONEN, M.; JARVENPAA, T.; HAKKINEN, J. **Comparison of near-to-eye displays: Subjective experience and comfort.** Journal of Display Technology, 6(1):27 - 35, jan. 2010.

SMOLIC, A.; KIMATA, H. ; VETRO, A. **Development of MPEG Standards for 3D and Free Viewpoint Video.** Technical report, Mitsubishi Electric Research Laboratories, October 2005.

UMAR, A. S.; SWASH, R. M.; SADKA, A. H. **Subjective quality assessment of 3d videos.** IEEE Transactions on Multimedia, September 2011.

ZHANG, L.; VAZQUEZ, C.; KNORR, S. **3D-TV Content Creation: Automatic 2D-to-3D Video Conversion.** IEEE Transactions on Broadcasting, 57(2):372 - 383, june 2011.