Book Vídeo 3D

IECOM

Mestrando: Carlos Danilo M. Regis

– 2011 –

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução ao Vídeo 3D

A infra-estrutura de comunicação digital tem se desenvolvido em um ritmo muito rápido nos últimos tempos. Por um lado, este desenvolvimento tem sido positivo para as emissoras, uma vez que criou a oportunidade para oferecer serviços de televisão digital para múltiplas plataformas de mídia, que permite transformar esses serviços de forma a alcançar um público maior e mais direcionados. Por outro lado, a oferta de novos e melhorados serviços de televisão tem gerado uma competição cada vez maior pela atenção e o interesse dos telespectadores. Como resultado, as emissoras estão sendo constantemente desafiadas a inovar, a fim de atender as expectativas dos novos clientes.

Duas das mais promissoras tecnologias digitais são a televisão tridimensional (3D-TV) e o cinema digital. Em particular, o sucesso financeiro dos filmes estereoscópicos tridimensional (S3D) tem sido claramente demonstrado. Reconhecendo a oportunidade oferecida por este sucesso, a indústria de radiodifusão começou a investigar meios para oferecer programas de televisão estereoscópica e serviços [41]. Por exemplo, o *Advanced Television Systems Committee* (ATSC) na América do Norte criou recentemente uma equipe de planejamento para analisar os potenciais benefícios e desvantagens, requisitos e passos práticos que são necessários para entregar TV em 3D nas casas. Investigações semelhantes também estão sendo realizados por padrões internacionais e organizações privadas, tais como a *European Broadcasting Union* (EBU), a *International Telecommunication Union* (ITU), e o Consórcio 3D@Home.

A televisão tridimensional (3DTV) oferece melhor experiência aos seus espectadores, proporcionando a sensação de profundidade. Assim, a 3DTV é uma tecnologia baseada na exploração das propriedades da percepção da profundidade. Portanto, é importante para fornecer serviços de 3DTV de uma forma que permite alta qualidade de percepção de profundidade natural [10].

Os efeitos em terceira dimensão estão se tornando cada vez mais comuns em nosso cotidiano e estar se tornando a nova febre do mundo do entretenimento. Mas o que poucos sabem é que, embora esta tecnologia só agora tenha começado a se desenvolver, seus princípios e as primeiras experiências já têm mais de meio século.

O primeiro filme público em 3D foi exibido em 1922, recorrendo a tecnologia anaglífica (uso de óculos com uma lente vermelha e outra azul inventados por L.D. DuHaron). Nada como é apresentado nas modernas salas de hoje em dia, mas a experiência de ter a impressão de ver as imagens saindo da tela ainda que precária causou furor no público [14].

No incio dos anos 50, foram realizadas várias tentativas de popularizar os filmes tridimensionais. Essas tentativas não tiveram êxito porque a tecnologia estereoscópica limitada da época e a inclinação para ter objetos estereoscópicos longe da tela, muitas vezes produzia imagens desconfortáveis.

Assim, outras experiências foram feitas, mas na época as prioridades eram outras. Era preciso aprimorar o som, o formato de exibição de imagem, reformar as salas de cinema e aprimorar os óculos de papel com uma lente azul e outra vermelha que além de ser desconfortáveis causavam dor de cabeça e enjoo em algumas pessoas.

Muitas décadas depois, o desenvolvimento da tecnologia possibilitou enormes evoluções na qualidade do 3D produzido e assim chegaram aos cinemas filmes como Avatar que revolucionaram o pensamento do público em geral em relação ao 3D. Desta vez é da opinião pública, tanto dos fabricantes como dos consumidores que o 3D veio para ficar, quer no cinema quer na televisão.

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. Desde essa época, passando por livros com fotos [32], pelo cinema [53], as aplicações estereoscópicas têm aumentado sua aplicação em diversas áreas.

O grande interesse para a TV 3D, deriva do reconhecimento de que, quando comparado ao padro de televisão de duas dimensões (2D), esta tecnologia aumenta significativamente o valor do entretenimento de programas de televisão. Claramente, o principal benefício da TV 3D é a maior percepção da profundidade. Os benefícios da TV 3D, no entanto, inclui mais do que apenas um sentido maior de profundidade. Alguma evidência empírica sugere que a televiso estereoscópica também poderia melhorar a percepção de nitidez, sensação de presença e naturalidade. As pesquisas indicam que as pessoas preferem ver as imagens S3D do que os seus homólogos bidimensionais, desde

que as imagens estereoscópicas sejam livres de incômodos e confortáveis para ver [41].

Para a visualização estereoscópica é utilizado métodos que usam óculos especiais, a exemplo da tecnologia anaglífica (óculos com lentes de cores azul-ciano e vermelho), tecnologia da polarização (que utiliza óculos com filtros polarizadores de luz), a tecnologia imagem sequencial alternada (óculos que possuem cristal líquido e bateria) ou ainda a autoestereoscopia, que dispensa o uso dos óculos para a visualização, sendo então o método que mais desperta interesse por ser mais cômodo ao utilizador.

As tecnologias de visualização (óculos, monitores, projetores, etc.) vêm se desenvolvendo, evidenciando que a área da estereoscopia está em evolução [12], [34].

1.1 Características do vídeo 3D

Uma figura que tem ou parece ter altura, largura e profundidade é tridimensional (ou 3D). Uma figura que possui altura e largura mas não possui profundidade é bidimensional (ou 2D). Algumas delas são bidimensionais propositalmente, como por exemplo os símbolos internacionais, que indicam a porta que leva a um toalete, por exemplo. Os símbolos são projetados para que você possa reconhecê-los assim que os visualiza. É por isso que se utiliza somente os formatos mais básicos. As figuras bidimensionais são úteis para comunicar algo simples e rápido. Já os tridimensionais contam uma história mais complexa, mas precisam carregar muito mais informações para isso.

Por centenas de anos, foram estudados alguns truques que podem tornar uma figura bidimensional parecer uma janela no mundo real. Podemos ver alguns exemplos: os objetos parecem menores quando estão mais distantes; quando objetos próximos à câmera são focalizados, os que estão mais distantes tornam-se embaçados e as cores tendem a ser menos vibrantes conforme se distanciam.

A terceira dimensão não existe, pois é apenas uma ilusão da mente. Isso é possível graças a um fenômeno natural chamado estereoscopia, que apesar do nome complicado trata-se apenas da projeção de duas imagens, da mesma cena, em pontos de observação ligeiramente diferentes. O olho recebe a imagem de forma que em cada olho é projetada uma imagem do mesmo objeto e o cérebro automaticamente analisa e gera duas imagens sob perspectivas diferentes, o que produz a profundidade estereóscopica. Nesse processo é obtido informações quanto à profundidade, distância, posição e tamanho dos objetos, gerando uma ilusão de visão em 3D.

Com o advento da tecnologia 3D, o usuário tem a liberdade de escolher o ângulo de visão para assistir determinada cena e até mesmo produzir efeitos de rotação e congelamento da imagem, o que não era possível em vídeos 2D, pois só eram percebidas as alterações das imagens com relação ao tempo e não ao espaço.Além do mais, o efeito da estereóscopia melhora a percepção de nitidez, sentido de presença e a naturalidade das imagens.

1.2 Fatores Humanos na Percepção da Profundidade

A visão humana é considerada uma referência na elaboração de técnicas para processamento de imagem e a análise do seu comportamento define o Sistema de Visão Humana (SVH). A imagem tridimensional se forma desde que os olhos estejam posicionados horizontalmente na cabeça, o SVH recebe duas imagens de uma mesma cena, uma em cada olho, que se sobrepõem com alguma diferença, originando perspectivas diferentes de uma mesma imagem. Analisando do ponto de vista funsional, fixando-se os olhos em algum ponto do espaço, as imagens se enquadram em ambos os olhos sobre a retina (fóvea). Assim, o objeto é visto sob as mesmas coordenadas relativas tanto para o olho esquedo quanto para o olho direito. A imagem do objeto fixado cai sob o horópter, linha curva ou superfície que contém os todos os pontos que estão na mesma posição geométrica ou na mesma distância, e dá origem a uma percepção única fundida. Os pontos na frente a atrás do horopter são gravadas em diferentes posições relativas do olho esquerdo e direito. Essas diferenças de posições gera as disparidades horizontais da retina, que é utilizada pelo SVH para extrair a profundidade relativa dos objetos na cena, ou seja, a profundidadade em relação a posição do objeto.

Uma variedade de sugestões são exploradas pelo ser humano para perceber a profundidade do mundo em três dimensões. Estes são tipicamente classificadas em profundidade binocular e monocular. O olho humano utiliza informações monoculares como acomodação, oclusão linear e perspectiva aérea, tamanho relativo, densidade relativa e paralaxe de movimento, para contruir a percepção de profundidade com apenas um olho. Essa características já podem ser observadas em displays 2D ,por exemplo, as televisões tradicionais. Já a profundidade binocular requer que as informações de profundidade da imagem sejam exploradas com os dois olhos, de forma que as diferenças entre as imagens sejam percebidas e o efeito de profundidade seja formado. De fato, movimento de paralaxe monocular e a disparidade binocular estão

intimamente relacionadas, pois as visões uma vez separados temporalmente fornecem as mesmas informações quando separadas espacialmente.

Uma lista incompleta das sugestões de profundidade são apresentadas na Figura 1.1. A extração de informações da profundidade da imagem visa converter sugestões de profundidade monocular contidos em sequências de vídeo em valores quantitativos da profundidade de uma imagem captada.

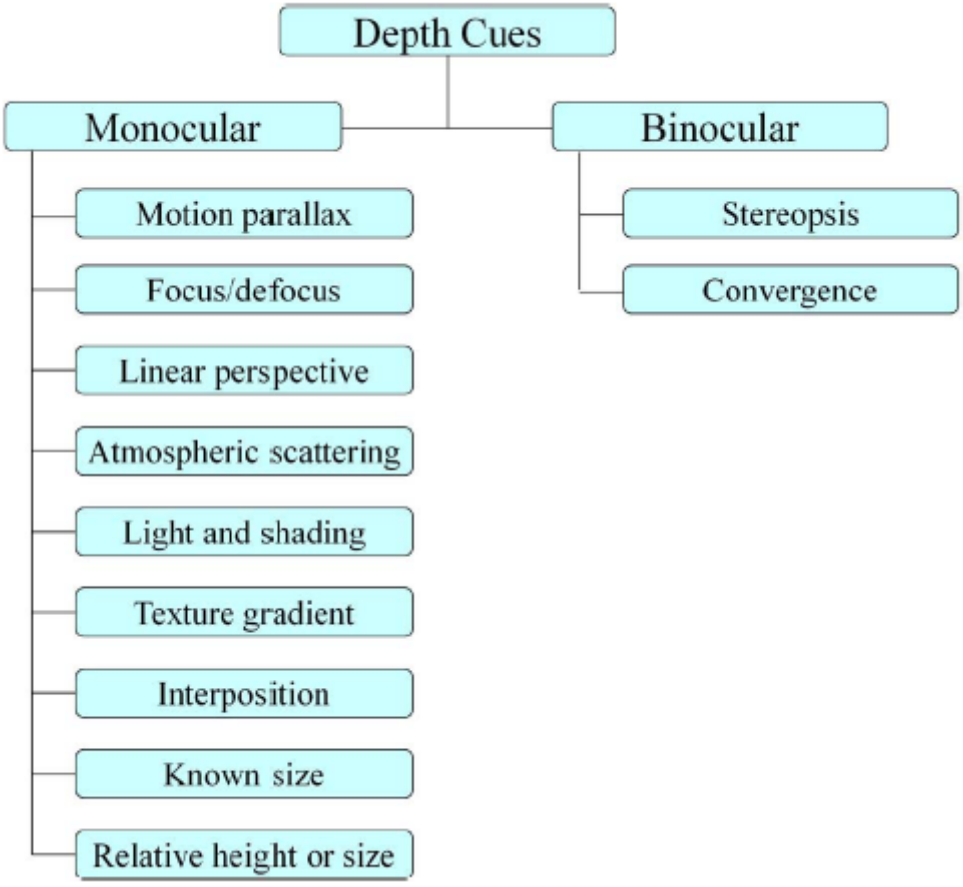


Figura 1.1: Classificação da *depth cues* [52].

1.2.1 Profundidade Binocular

Nas dicas de profundidade binocular as imagens são formadas a partir da disparidade na projeção das imagens na retina. Ela é subdividida em estereopsia, acomodação e convergência que serão melhor explicadas nas subseções a seguir.

Estereopsia

A estereopsia ocorre devido ao fato dos olhos estarem horizontalmente separados, aproximadamente 6.3 cm, proporcionando a cada olhos, um ponto

de vista único sobre o mundo. Utiliza-se duas pespectivas diferentes de uma mesma imagem vista pelos dois olhos. A imagem é formada na região central da retina (fóvea) de cada olho sob ângulos diferentes. A difrença entre os ângulos é chamado de disparidade binocular e fornece informações sobre as distancias relativas dos objetos até o observador, a estrutura de profundidade e o ambiente em geral. A Figura 1.2 ilustra geometricamente como ocorre o processo de formação da imagem nos olhos.

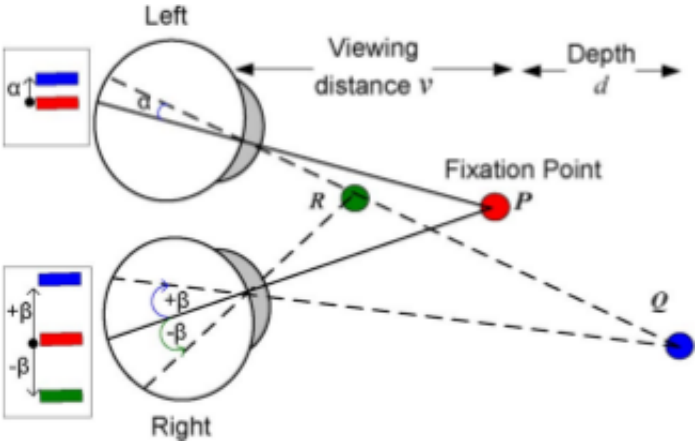


Figura 1.2: Geometria da estereopsia binocular [11].

Fixando-se um ponto P no centro de cada olho, fixamos outros dois pontos Q e R, de forma que Q esteja localizado depois do ponto central e R antes de ponto central. Cada imagem projetada por esses pontos formará angulos em relação ao ponto central, caracterizando a disparidade angular em que Q tem disparidade positiva e R negativa. Essa diferença de sinal estimula ao cérebro a perceber a profundidade dos objetos em relação a posição deles.

Acomodação e Convergência(Desfoque da Retina e Profundidade de Campo)

Na visualização dos objetos no espaço, os olhos convergem sobre ele de forma que a informação da profundidade e do ambiente caiam na área de fusão das imagens, *Panum's*. Os olhos automaticamente focam (acomodação) no objeto atualmente fixado, fazendo com que este objeto sobressaia dentre os outros objetos ao seu redor. Imagens duplas à frente ou atrás do plano de fixação tedem a ficar fora de foco e desaparecerá com o aumento da mancha optica. Acomodação é o processo responsável pela mudança do poder refrativo do olho, garantindo que a imagem seja focada no plano da retina. O poder de mudança é induzido pelos músculos ciliares que alteram a curvatura e a espessura central do cristalino, aumentando o poder dióptrico. Quando um objeto de interesse é fixado pelo olho, a acomodação é ajustada de tal forma que uma imagem nítida é percebida na retina. Uma boa acomodação

exige um tempo de fixação mínimo de um segundo ou mais. No entanto, o olho humano pode tolerar uma certa quantidade de desfocagem da retina sem reajustar a acomodação,embora os critérios para um bom foco dependam do objeto observado e variem de indivíduo para indivíduo.

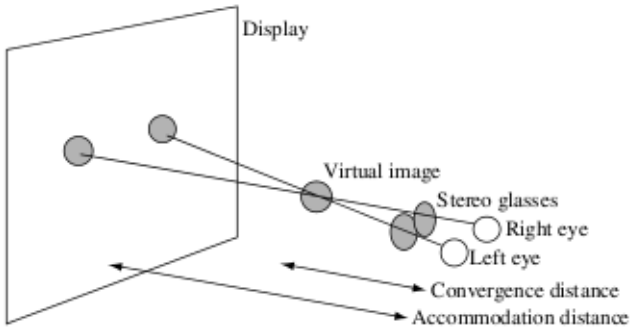


Figura 1.3: Diferença entre distância de acomodação e convergência. [39].

O processo de acomodação e convergência (CA/A) estão intimamente relacionados visto que a acomodação produz movimentos de convergência (acomodação de convergência) e convergência produz acomodação(convergência de acomodação), ou seja, objeto próximo –> acomodação <–> convergência. A acomodação é dirigido para imagens de objetos a uma distância de tela enquanto convergência está direcionado para a distâncias percebidas de ob- jetos. Em condições normais, essas distâncias se coincidem, mas a situação é diferente quando estamos assistindo vídeos 3D, pois devido ao efeito da profundidade, os objetos apresentam distâncias diferentes.

A Figura 1.4 ilustra geometricamente como acontece o processo de acomodação, o que resulta no desfoque das imagens projetadas na retina.

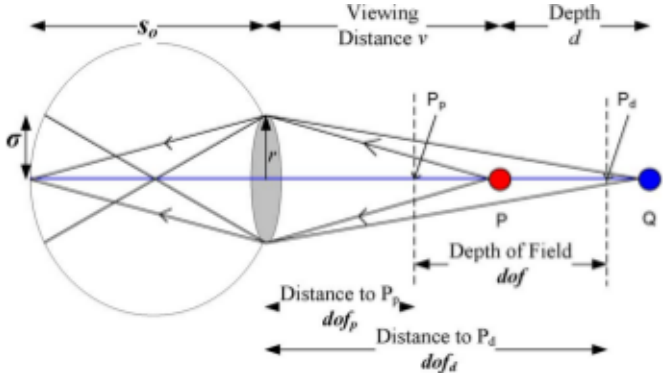


Figura 1.4: Geometria do desfoque da retina [11].

Fixando o ponto P, a potência óptica (capacidade de convergir ou divergir a luz) dos olhos é ajustada para que a imagem do ponto P esteja focada na região central da retina (fóvea). No entanto, a potência óptica dos olhos varia em uma certa frequência e, assim, o olho pode tolerar uma certa quantidade de desfoque da retina sem reajuste de acomodação. Esta diferença de potência óptica é conhecida como a profundidade de foco ocular e é apresentado em dioptrias[1] (D). Em outras palavras, o ponto P pode ser deslocado ao longo do eixo óptico dentro de um período determinado, sem perceber o desfoque de imagem. Este período é conhecido como a profundidade de campo (dof).

Os pontos mais próximos e mais distantes do limite é conhecido como o ponto "Proximal" (P_p) e do ponto "Distal"(P_d), como mostrado na Figura 1.4. Portanto, qualquer objeto dentro do limite é percebido nitidamente e os objetos fora do limite são percebidos "borrados"o que estimula a sensação de profundidade. A equação 1.1 relaciona a distância do olho ao ponto $Q(v+d)$ e a quantidade de desfoque da retina σ:

$$v + d = \frac{F \cdot r \cdot S_o}{(r \cdot S_o - F(r + \sigma))} \tag{1.1}$$

em que, v representa distância visual, d a profundidade, F tamanho focal dos olhos, r o raio de abertura dos olhos, S_o a distância do centro da retina (fóvea) aos olhos.

A magnitude do campo de profundidade (dof) difere de pessoa para pessoa, dependendo da profundidade do foco ocular do olho. A equação 1.2 representa o calculo da profundidade do foco ocular T relacionando o ponto "Proximal"e "Distal".

$$T = \frac{1}{dof_p} - \frac{1}{dof_d} \tag{1.2}$$

De acordo com o padrões de equações opticas, os valores de dof_p e dof_d são dados da seguinte forma:

$$dof_p = \frac{2v}{2 + v.T} \qquad dof_d = \frac{2v}{2 - v.T} \tag{1.3}$$

Não há nenhuma indicação na literatura sobre como calcular os valores de T. No entanto, ele varia entre $0.6D$ e $0.8D$ [11].

Em resumo, quando um objeto está em foco e é visto nitidamente, outro objeto além do campo de profundidade dof é visto "borrado". Esse fenômeno indica a sensação de profundidade da imagem. Assim, quando um objeto

[1] Dioptria é uma unidade de medida que afere o poder de vergência de um sistema óptico $(m-1)$. Na Óptica, é a unidade de medida da potência de uma lente corretiva (popularmente conhecido como grau).

é nítido e o outro é borrado tem-se a sensação além da distância do objeto nítido.Desta forma, são definidos limites para os objetos projetados em uma tela estereoscópica para minimizar o desconforto visual.

1.2.2 Profundidade Monocular

Manter a visão fixa em um ponto no espaço aciona mecanismos físicos dos olhos, ou seja, como os olhos são formados por músculos e nervos, eles são responsáveis por enviar ao cérebro informações sobre a imagem e gerar o efeito de profundidade. Profundidade vista por dicas geométricas é uma abordagem interessante para a obtenção da profundidades a partir de uma imagem 2D. As geometrias relacionadas as profundidades pictóricas sugeridas são a perspectiva linear, tamanho conhecido, tamanho relativo, altura da imagem, interposição e gradiente da textura. Alguns desses sinais são mais fortes do que outros. A interposição, por exemplo, pode nos ensinar a ordem de profundidade dos objetos, mas não a distância em profundidade entre eles. Alguns dicas podem ser difíceis de serem usadas em uma aplicação para a estimativa de profundidade. Por exemplo, as informações relacionadas com o tamanho dos objetos é difícil de serem usados, pois requer a identificação de objetos e do conhecimento de tamanhos normais para esses objetos. Os mais comuns são as dicas geométricas da perspectiva linear e a altura da imagem.

Perspectiva linear se refere à propriedade de linhas paralelas de convergir a uma distância infinita, ou equivalentemente, um objeto de tamanho fixo vai produzir um menor ângulo de visão quanto mais distante do olho. Esta característica é utilizada na estimativa da profundidade por meio da detecção de linhas paralelas nas imagens e identificação do ponto em que estas linhas convergem (ponto de fuga). Em seguida, uma atribuição adequada de profundidade pode ser derivada com base na posição das linhas e do ponto de fuga [52].

A altura na imagem indica que objetos que estão mais próximos da parte inferior das imagens são geralmente mais próximo do que objetos que estão na parte superior da imagem. Cenas ao ar livre e paisagens, principalmente, contem esta indicação de profundidade pictórica. Para extrair essa sugestão de profundidade, linhas horizontais, normalmente têm de ser identificadas para que a imagem possa ser dividida em faixas que vão desde a margem esquerda até a borda direita. Para este propósito, um algoritmo de **linha-tracing** é aplicada para recuperar as linhas divididas ideais sujeitas a algumas restrições geométricas [52].

Além da perspectiva linear e altura da imagem, também é possível recuperar a profundidade a partir da textura (chamado de forma-de-textura), que tem como objetivo estimar a forma de uma superfície com base em sugestões de marcas na superfície ou sua textura [44]. Esses métodos, no entanto, são normalmente restritos a tipos específicos de imagens e não pode ser aplicada na conversão do conteúdo de vídeo de 2D para 3D em geral.

Tamanho Relativo

Representa a profundidade da imagem levando em consideração a geometria da imagem formada na retina. A imagem projetada atua como um estímulo importante para percepção da profundidade em que o ângulo visual de um objeto projetado diminui a medida que a distância até o objeto aumenta e vice-versa. Utilizando a propriedade de semelhança de triângulo, a equação 1.4 demonstra o tamanho da imagem na retina R em função do tamanho real da imagem H, distância focal G e a distância do objeto D.

$$\frac{R}{G} = \frac{H}{D} \Rightarrow R = \frac{HG}{D} \tag{1.4}$$

O aumento na distância do objeto interfere no tamanho da imagem formada na retina, diminuindo-a de tamanho. O cérebro interpreta essas diferenças como uma mudança de profundidade ΔR. A equação 1.5 representa matemáticamente esse comportamento.

$$R - \Delta R = \frac{HF}{(D + \Delta d)} \tag{1.5}$$

É possível estimular uma mudança no tamanho da imagem projetada na retina por meio da mudança do tamanho do objeto ΔH. Com isso, a mudança no tamanho dos objetos que são fisicamente similares proporcionam a sensação de profundidade devido a uma mudança na retina do tamanho da imagem.

$$R - \Delta R = (H - \Delta H).\frac{F}{D} \tag{1.6}$$

Dicas de Profundidade de cor e intensidade

Variações na quantidade de luz que chega ao olho também pode fornecer informações da profundidade dos objetos. Este tipo de variação se reflete nas imagens captadas como variações de intensidade ou alterações na cor. Sugestões de profundidade que são baseados neste mecanismo de dispersão

atmosférica incluem, distribuição de luz e de sombra, percepção da imagem de fundo e contraste local.

Dispersão atmosférica refere-se a dispersão dos raios de luz pela atmosfera produzindo um tom azulado e menos contraste com os objetos que estão ao longe e melhor contraste com os objetos que estão em uma estreita faixa [9]. Com base em regras de cor, que são aprendidas heuristicamente usando um grande número de imagens da paisagem, a detecção da região semântica é realizada para dividir imagens da paisagem em seis regiões, como céu, montanha mais distante, montanha distante, montanha próxima, terra e outros [52].

CAPÍTULO 2

Principais Conceitos

2.1 Visão estereoscópica humana

A capacidade de apreciar uma terceira dimensão usando um monitor de TV 3D baseado nas características do sistema visual humano. Uma vez que os olhos são posicionados horizontalmente na cabeça, o sistema visual recebe duas visões da cena visual, ou seja, a visão do olho esquerdo e a do olho direito, que em grande parte se sobrepõem, mas um pouco diferente porque eles se originam a partir de duas perspectivas diferentes. O sistema visual processa a informação das duas imagens provenientes de duas perspectivas para a produção de profundidade estereoscópica [41].

Os olhos movem-se constantemente, mesmo durante a fixação. No entanto, o sistema visual binocular é extremamente bom em coordenar o movimento dos dois olhos [26]. Como resultado, a partir de um ponto de vista funcional, quando nós fixamos binocularmente um ponto no espaço, as imagens se enquadram em um ponto, de ambos os olhos esquerdo e direito, sobre a fóvea, que é a parte do fundo do olho (retina) que tem a maior acuidade. Assim, um objeto fixado é fotografado nas mesmas coordenadas em relação a vista do olho esquerdo e do olho direito e é percebido como uma simples percepção, ou seja, visto como um simples objeto.

O ponto de fixação cai no *horoptero*. O *horopter* é uma linha curva ou superfície que contém todos os pontos que estão na mesma geometria (geometria *horopter*) ou distância (*horopter* empírica) percebida do ponto de fixação [41].

Pontos localizados em frente ou atrás do *horopter* são gravadas em diferentes posições relativas as vistas do olho esquerdo e do olho direito. Estas diferenças nas posições relativas são chamadas de disparidades da retina horizontal. A magnitude da disparidade da retina de um ponto aumenta com a distância do objeto a partir do *horopter*. Pontos na frente do *horopter* dizem

ter uma disparidade negativa ou cruzada, enquanto pontos de objetos por trás dele é dito com disparidade positiva ou descruzadas (Apendice B). Diz-se que uma melhor performance na visualização estereoscópica é obtida na disparidade positiva do que na disparidade negativa. O sistema visual humano usa essas disparidades para extrair a profundidade relativa dos objetos na cena visual, ou seja, a posição em profundidade de um objeto com relação a outro objeto.

Objetos que dão origem a produzir as disparidades em imagens diferentes nas retinas direita e esquerda. No entanto, os objetos que estão localizados dentro de uma pequena região na frente e a atrás do plano de fixação da origem é a uma percepção simples fundida. A região, dentro do qual os objetos são fundidos, apesar de ter imagens diferentes nos dois olhos, é chamado de área de *Panum's Fusional*. Objetos localizados fora da área de *Panum's Fusional* resultam na visão dupla, ou seja, *diplopia*, mas ainda pode ser percebida em profundidade. O tamanho da área de *Panum's Funsional* não é fixa, mas sim depende das propriedades espaciais e temporais da meta de fixação, tais como a duração da exposição, de resolução espacial, variação da disparidade da frequência temporal [41].

Quando o ponto de fixação é alterado para olhar para um novo objeto localizado a uma distância diferente, os dois olhos se movem simultaneamente e em direções opostas, para que o novo objeto seja fotografado no centro de cada fóvea. Se o novo objeto é mais perto, os olhos se movem para dentro em direção ao outro (convergência), enquanto que se o novo objeto está mais longe dos olhos se movem para fora, longe uns dos outros (divergência). Este processo chamado de convergência e está estreitamente relacionado com a acomodação. Este último se refere ao processo pelo qual o poder óptico do olho é alterado para manter a visão clara, ou seja, uma imagem nítida, de um objeto distante.

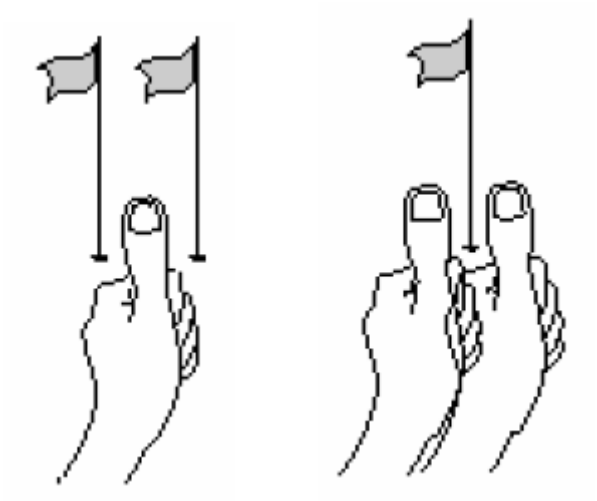
Quando os olhos se fixam em um objeto, a forma do cristalino em cada olho é alterada pelos músculos ciliares para que a imagem do objeto em foco seja fixado na parte traseira do olho, a retina. Pontos localizados mais perto ou mais longe do que os pontos acomodados no são imagens corretas na retina e, portanto, sujeitas a um grau crescente de borragem. No entanto, o sistema visual é tolerante a uma pequena quantidade de borramento [41], e pontos localizados dentro de uma pequena região ao redor do ponto de acomodação são percebidos a estar em foco.

O tamanho desta região, conhecida como a profundidade de campo (DOF – *Depth of Field*), varia inversamente com o diâmetro da pupila. A pro-

fundidade de campo tem um correspondente, conjugada, é a região em torno do plano da retina, a região é chamada de profundidade do foco.

Em condições normais, mudanças na acomodação dos dois olhos e o processo de convergência ocorre de forma integrada: mudanças na acomodação induz alterações na convergência e vice-versa [23]. No entanto, os dois processos podem entrar em conflito quando assistir alvos estereoscópicos.

Os nossos olhos estão separados de cerca de 7,5 cm entre si e isso implica que cada olho vê de uma perspectiva ligeiramente diferente a mesma cena. Isto pode ser observado com um simples experimento: alinhe o polegar da mão esquerda com uma bandeirinha e seu nariz, e foque sua visão no dedo. Você verá a bandeirinha como sendo duas, uma para cada olho (feche um olho e abra o outro e em seguida inverta), conforme a Figura 2.1a. Agora direcione sua visão para a bandeirinha, a visão que você terá com os dois olhos abertos é mostrada na Figura 2.1b, o polegar agora é visto como sendo dois.



(a) (b)

Figura 2.1: Na Figura 2.1a os olhos estão convergindo para o polegar e a bandeirinha é vista como dupla imagem. Na Figura 2.1b os olhos agora estão convergindo para a bandeirinha e o polegar é visto como uma imagem dupla [8].

Estas diferenças entre imagens geradas pelo olho direito e pelo olho esquerdo são processadas pelo cérebro nos dando uma noção de profundidade e, com isto, tem-se a idéia de imersão em um ambiente com objetos posicionados a distâncias diferentes.

2.2 Geração de Vídeos 3D

A TV-3D explora as características do sistema visual humano, recriando, embora não de forma verídica, as condições que levam percepção da profundidade relativa dos objetos na cena visual. Assim, o primeiro requisito de imagens estereoscópicas é a captura de pelo menos dois pontos de vista da mesma cena a partir de duas câmeras alinhadas horizontalmente. As imagens dos objetos na cena terá posições relativas diferentes nos dois olhos. Essa diferença é chamada de Paralaxe.

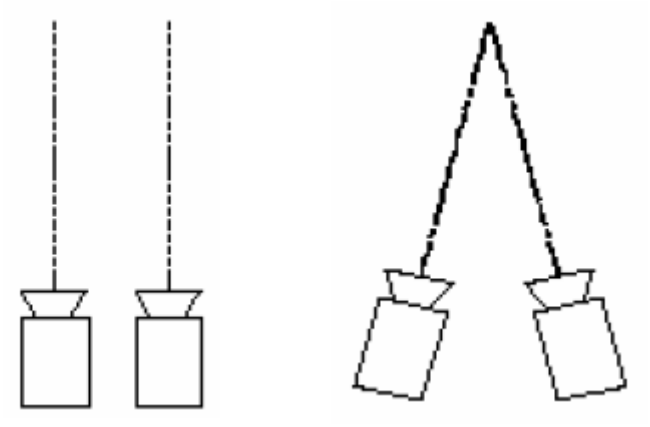
O processo de geração de vídeos em 3D pode ser simulado por meio de duas câmeras organizadas com a mesma distância interocular dos olhos humanos. Logo, colocando-se as câmeras separadas uma da outra com base nessa distância, é possível simular o sistema visual humano. Quando cada imagem das câmeras for apresentada ao seu olho correspondente, as duas imagens serão fundidas em uma única imagem pelo cérebro, produzindo a ilusão da visão estereoscópica [20].

Em [20], é apresentado dois tipos de configurações de câmeras passíveis para a captura de vídeo estereoscópico:

1. Câmeras em eixo paralelo;
2. Câmeras em eixo convergente (*toed-in*).

Na configuração de eixo paralelo, as câmeras são alinhadas de forma que os eixos centrais de suas lentes estejam em paralelo, conforme a Figura 2.2a. A convergência das imagens é alcançada por meio de um pequeno deslocamento dos sensores de captura das câmeras ou por meio de uma tradução horizontal (deslocamento horizontal das imagens para se alterar a distância ou paralaxe entre os pontos correspondentes das imagens do olho direito e do esquerdo) e do corte das imagens resultantes.

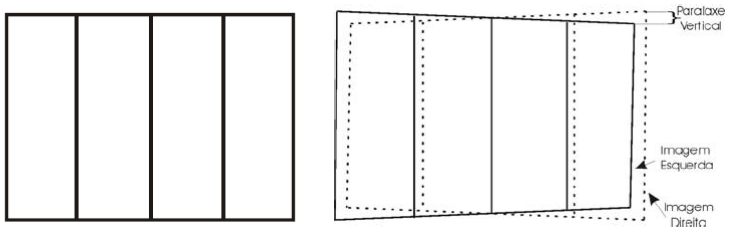
Na segunda forma de configuração, eixo convergente, as duas câmeras são rotacionadas para que seus eixos centrais sejam convergidos sobre um mesmo ponto no plano de projeção, conforme Figura 2.2b. O ponto em que a câmera converge vai ser trabalhada em relação as mesmas coordenadas na câmera da esquerda e da direita. Portanto, tem paralaxe zero. Quando exibido estereoscopicamente, este ponto é um objeto representado no plano da tela. Todos os pontos localizados a outras distâncias do objeto tem paralaxe negativo ou positivo que dependem das distâncias da profundidade dos objetos e da separação horizontal dos pontos de vista do olho esquerdo e olho direito. Objetos desses pontos aparecerão na frente ou atrás do plano da



(a) (b)

Figura 2.2: Configuração de Câmeras: 2.2a Câmeras em Eixo Paralelo e 2.2b Câmeras em Eixo Convergente (*toed in*) [20].

tela. As configurações do *Toed-in* são fáceis de configurar e permitir que um objeto de interesse possa ser posicionado no plano da tela, mas que geram distorções *keystone* na câmera esquerda e na câmera direita [43]. Distorções *Keystone* transformam as imagens em formas semi-trapezoidal de tal forma que as alturas verticais correspondam no mais entre os pontos de objetos correspondentes nas duas imagens. Essas distorções podem afetar o conforto visual [41].O problema é ilustrado na Figura 2.3.



(a) (b)

Figura 2.3: Paralaxe Vertical causada por Distorção *Keystone*: A Figura 2.3a representa a Imagem Original e a Figura 2.3b a Visão do olho esquerdo e direito sobrepostas [43].

Para prevenir a ocorrência de distorções *keystone*, as câmeras poderiam ser configuradas em uma configuração paralela. Neste caso, os eixos ópticos das câmeras não convergem, mas são paralelas. Na configuração paralela, todos os pontos do objeto tem alguma paralaxe, no entanto, as imagens obtidas com uma configuração em paralelo poderiam ser modificadas para se alinhar horizontalmente as imagens de um objeto-alvo de interesse para ter paralaxe zero antes de exibir, gerando assim uma distribuição de paralaxe comparável as distorções do *toed-in*, mas sem distorção de *keystone*. No caso da confi-

guração de câmeras com eixos paralelos não ocorre paralaxe vertical, mas há necessidade de uma tradução horizontal das imagens resultantes. Por causa dessa tradução, as imagens não são perfeitamente sobrepostas. Esse fato requer cortes nas imagens, de forma que somente um campo de visão comum seja apresentado. Dependendo de como as imagens são traduzidas, os planos de convergências podem ser posicionados em profundidades de percepção diferentes [20].

Um requerimento básico para geração de vídeos 3D, é que o posicionamento das câmeras seja feito de tal forma, que o olho esquerdo veja somente a imagem da câmera esquerda e o olho direito a imagem da câmera direita. Com as câmeras alinhadas horizontal e verticalmente, a separação do intereixo deve ser de aproximadamente 65 milímetros, tornando a imagem produzida mais realista. Essa configuração é comum à todas as tecnologias estereoscopicas, incluindo a 3DTV [18].

Capítulo 3

Conversão de vídeos 2D para 3D

A adoção bem sucedida de TV-3D pelo público em geral dependerá não só dos avanços tecnológicos em *displays* 3D e em sistemas de radiodifusão de TV-3D [52], mas também da disponibilidade de uma ampla variedade de conteúdo no formato estereoscópico para os serviços 3D (S3D) [16]. A oferta de conteúdos S3D será especialmente importante nos estágios iniciais de implantação da TV-3D, de forma a garantir que o público tenha interesse em adiquirir os *displays* 3D e os serviços de TV-3D. No entanto, um certo período de tempo será necessário para os provedores de conteúdo gravar e criar com câmeras estereoscópicas o material S3D suficiente.

Consideramos que a conversão de imagens/vídeos de 2D para 3D é uma maneira de aliviar este problema difícil. 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.

As técnicas de conversão de 2D para 3D podem ser rentável para os provedores de conteúdo que estão sempre à procura de novas fontes de receita para a sua vasta biblioteca de materiais de vídeo. Este mercado potencial está atraindo muitas empresas a investir seus recursos humanos e dinheiro para o desenvolvimento de técnicas de conversão de 2D para 3D.

O princípio fundamental das técnicas de conversão de 2D para 3D se baseia no fato de que o sistema visual humano transforma as pequenas diferenças na distancia da informação da imagem (pixel desloca horizontal) do olho esquerdo e do olho direito tal que os objetos são percebidos em diferentes profundidades e fora do plano 2D. Assim, a conversão de imagens 2D para imagens estereoscópicas em 3D envolve o princípio subjacente de deslocamento horizontal dos *pixels* para criar uma nova imagem, de modo que existem disparidades horizontal entre a imagem original e uma nova versão

dele. A extensão do deslocamento horizontal do *pixel* depende não apenas da distância de um objeto para a câmera estereoscópica, mas também sobre a separação inter-lente que determina o ponto de vista da nova imagem.

Várias abordagens para a conversão de 2D para 3D têm sido propostas. Estas abordagens podem ser classificados em três esquemas: conversão manual, humana assistida e automática [52]. 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 [16], mas é muito demorado e caro.

O esquema humana assistida converte imagens 2D para 3D estereoscópico com algumas correções feitas "manualmente"por um operador [52]. Mesmo que este esquema reduza o tempo consumido em comparação com o regime de conversão manual, uma quantidade significativa de esforço humano ainda é necessário 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 originado 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). Pode ser feito em tempo real ou em um processo mais demorado (*off-line*). A qualidade do produto resultante está relacionada com o nível de processamento envolvido, por isso os sistemas de tempo real normalmente produzem a conversão de menor qualidade.

Há duas questões importantes a serem considerados para as técnicas automáticas de conversão de 2D para 3D : recuperar a profundidade de uma imagem ou vídeo 2D [52], e a forma de gerar imagens de alta qualidade estereoscópica em novos pontos de vista virtual [51].

3.1 *Framework* para a Conversão de Vídeo de 2D para 3D

A conversão de vídeo de 2D para 3D pode ser visto, pelo menos conceitualmente, como um caso especial de modelagem de imagem baseadas em técnicas de renderização, desenvolvido para fornecer novos pontos de vistas virtuais de um determinado conjunto de imagens. Com base na modelagem de imagens e das técnicas de renderização podem ser classificadas em três categorias principais, de acordo com a quantidade de informação sobre a geometria explicitamente utilizados no processo [7]:

1. Métodos que usam um modelo da imagem 3D completo: Esta categoria exige a reconstrução completa e precisa de um modelo geométrico para a imagem capturada. Esse modelo irá conter todas as informações necessárias para a prestação de uma nova visão virtual a partir de um ponto de vista dado. Estrutura da *silhouette*, por exemplo, é uma técnica comumente usada para construir modelos de objetos 3D. Dado o modelo 3D e as condições de iluminação da imagem, uma nova visão virtual pode ser facilmente construida a partir de um ponto de vista desejado, usando técnicas convencionais de computação gráfica. No contexto de conversão de vídeo 2D para 3D, geralmente é extremamente difícil e propenso a erros a recuperação da estrutura da imagem completa em 3D a partir de uma sequência de imagens ou vídeo único, exceto se o vídeo for capturado em condições rigorosas. É, portanto, impraticável usar uma abordagem de modelo 3D completa para a conversão automática de vídeo 2D para 3D.

2. Métodos que usam apenas imagens e nenhuma informação explícita da geometria: Esta categoria diretamente torna novas vistas virtuais a partir de um conjunto de imagens capturadas, geralmente centenas de milhares de imagens são necessários, com nenhuma ou muito pouca informação geométrica, por exemplo, *Lightfields* e *Lumigraph*. Na conversão de vídeo de 2D para 3D, o número de imagens disponíveis para renderização é normalmente pequena, fazendo esta abordagem impraticável para a conversão automática de vídeo 2D para 3D.

3. Métodos híbridos que, explicitamente, utilizam as informações geométricas: Esta categoria utiliza uma abordagem geométrica híbrida e baseado em imagem. Novas exibições virtuais são renderizadas a partir de um número limitado de imagens com a ajuda de informações geométricas incompletas da cena. Nesta categoria os métodos incluem profundidade de imagem baseada em processamento (DIBR) [16], profundidade de imagens em camadas (LDI), e reconstrução intermediária da visualização (IVR). A maior parte dos algoritmos de conversão de vídeo propostos de 2D para 3D usam uma estrutura que se enquadra nessa categoria, uma abordagem de geometria híbrido e baseado em imagem.

O quadro comumente utilizado para a conversão automática de vídeo 2D para 3D, basicamente, consiste de dois elementos (Figura 3.1): a extração de informações de profundidade e de geração de imagens estereoscópicas, de acordo com informações da profundidade estimada e das condições de visualização esperada.

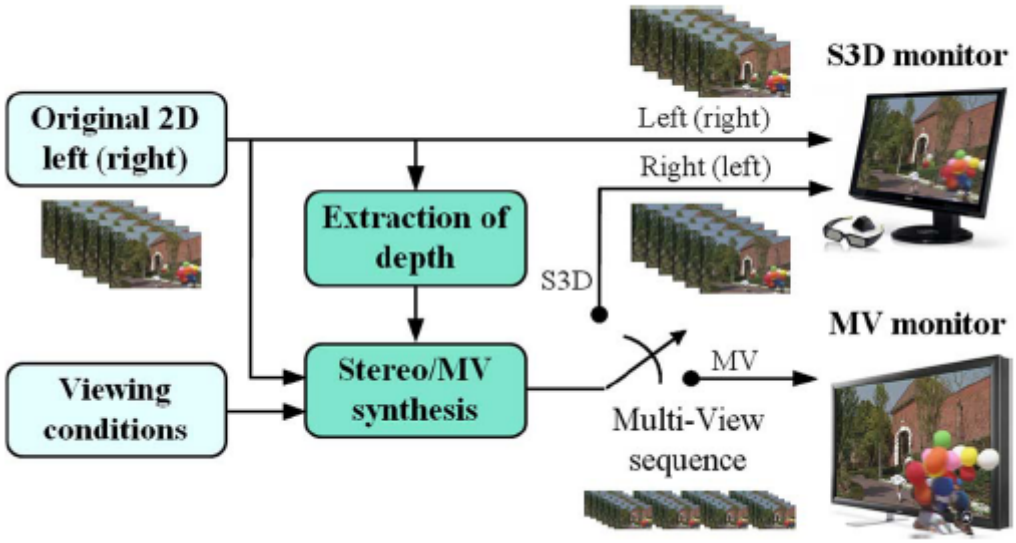


Figura 3.1: Diagrama de conversão automática de 2D para 3D [52].

A extração de informações de profundidade visa explorar pistas pictóricas[1] e a paralaxe de movimento, contido em uma única imagem ou vídeo em 2D, para recuperar a estrutura de profundidade da imagem. A informação de profundidade recuperada é então convertida em uma representação adequada para uso no processo de conversão de vídeo 2D para 3D.

3.2 Modelos de extração de profundidade

3.2.1 Profundidade De Foco/Defoco

Como visto no capítulo de introdução, acomodação é o mecanismo do olho humano usado para se concentrar em um dado plano de profundidade. Câmeras de abertura real fazem o mesmo,centrando-se em um dado plano. Este mecanismo pode ser explorado para a geração de informações detalhadas a partir de imagens captadas, que contêm um plano focado e objetos fora do plano focado. Este tópico é conhecido na literatura como *depth-from-focus/defocus*, que é um dos primeiros mecanismos a serem empregados para recuperar a profundidade das imagens individuais [13], [15].

Na prática, existem duas abordagens principais que são usados para implementar este mecanismo. A primeira emprega várias imagens com enfo-

[1]Uma imagem é pictórica quando produzida por ordenação de pigmentos sobre algum suporte, geralmente utilizando técnicas de fotografia, desenho, pintura, gravura e outras das Artes Visuais. A imagem pictórica pode ser figurativa, se representar algo existente materialmente na natureza (ou supostamente existente, como no caso de figuras mitológicas, ou abstrata, se não se prender a nenhuma representação material).

ques em diferentes características a fim de extrair a variação do borrado para uma característica determinada da imagem por meio das imagens disponíveis. Esta variação pode ser traduzida em profundidade, encontrando o ponto em que o recurso especial deve ser o foco [52]. Embora essa abordagem seja confiável e forneça estimativa de boa profundidade, a exigência de ter várias imagens da mesma cena, capturadas com diferentes sistemas ópticos simultaneamente é uma restrição para qualquer aplicação prática no problema de conversão 2D para 3D.

A segunda abordagem tenta extrair as informações da borragem a partir de uma única imagem, medindo a quantidade do desfoco associado a cada *pixel* e então mapear as medidas de borragem da profundidade desse *pixel*. Um processo de deconvolução no domínio da frequência usando a filtragem inversa foi introduzido em [33] para recuperar a quantidade de desfoco em uma imagem. Para resolver a instabilidade relacionada com a filtragem inversa no domínio da frequência uma abordagem de regularização foi proposto um método de controle local para detectar bordas em diferentes níveis de desfoco e para calcular o desfoco associado a essas arestas.

A indefinição gaussiana de kernel foi usada para modelar o desfoco das bordas e sua segunda derivada foi usado para medir a **propagação da borda**, a fim de extrair o nível de desfoco. Mais recentemente foi proposta uma abordagem baseada em *wavelet*, em que uma decomposição de *wavelet* em macroblocos dentro de uma imagem foi realizada para recuperar o conteúdo de alta frequência desse macro-bloco e o número de coeficientes *wavelet* de alto valor foi contado para ser usado como uma medida de borragem. Uma abordagem semelhante foi usada para a análise *wavelet* 2D para a detecção e análise de bordas e usar a cor baseado em segmentação para adicionar consistência ao mapa de profundidade. Estatísticas de ordem superior também tem sido utilizado para estimar a quantidade de desfoco em imagens convertidas de 2D para 3D [52].

Embora a abordagem da recuperação da profundidade pelo foco/desfoco seja relativamente simples, ela sofre uma grande desvantagem, em distinguir o primeiro plano do fundo, quando a quantidade de desfoco é semelhante. Em muitos casos o primeiro plano corresponde ao plano de focagem, mas quando isso não é o caso, então é impossível distinguir uma região fora de foco no primeiro plano de uma região fora de foco no segundo plano.

3.2.2 Profundidade a partir de dicas Pictorial

As pistas da profundidade pictórica são os elementos em uma imagem que nos permitem perceber a profundidade em uma representação 2D da imagem. Esta tem sido conhecida há séculos e tem sido amplamente aplicada em artes visuais para melhorar a percepção de profundidade. Percepção de profundidade pode está relacionada as características físicas do Sistema Visual Humano (HVS), tais como a percepção de profundidade por acomodação ou pode ser aprendido com a experiência adquirida com a percepção da altura relativa dos objetos na imagem, perspectiva, sombras e outros sinais pictóricos [52].

A geração de informações de profundidade a partir de pistas pictóricas embutido em uma imagem pode ser subdividida em duas abordagens. O primeiro se refere à extração de informações de profundidade "real"a partir de pistas disponíveis em uma imagem pictórica. Por "real", queremos dizer profundidade relativa entre os objetos na cena. É impossível obter mais profundidades absolutas, sem o conhecimento da posição e das características ópticas do dispositivo de captura. A segunda abordagem cria informações detalhadas artificial ou não-verídicas, explorando pistas pictóricas que são comumente encontrados em todas as cenas de uma determinada categoria, como paisagens ou interiores. Discutiremos três categorias de sinais pictóricos comumente usado para extração de informações detalhadas nas subseções a seguir.

Capítulo 4

Codificação

4.1 A codificação com H.264

O padrão para codificação de vídeo H.264/AVC, também conhecido por MPEG-4 Part 10, foi desenvolvido pelo JVT (Joint Video Team), composto por especialistas dos grupos VCEG (Video Coding Experts Group) e MPEG (Moving Picture Experts Group), respectivamente das organizações ITU-T e ISO/IEC [**?**]. A estrutura básica do padrão H.264/AVC é similar a de padrões anteriores (H.261, MPEG-1, MPEG-2 / H.262, H.263 ou MPEG-4 part 2) [2], porém com uma maior eficiência na codificação, reduzindo o número de bits resultantes no vídeo codificado além de maior robustez a falhas na transmissão. Estas melhorias implicam em uma maior complexidade e consumo de recursos computacionais para o processo de codificação [**?**].

4.1.1 O padrão H.264

O padrão H.264 define a sintaxe e a semântica do vídeo codificado, tendo como foco o decodificador e as atividadesenvolvidas na decodificação [**?**]. Cada quadro na entrada do codificador é dividido em macroblocos. Cada macrobloco é composto pelas componentes de luminância (Y) e de crominância (Cr e Cb). Para a componente de luminância, o macrobloco é composto de 16x16 amostras e para as componentes de crominância, o macrobloco é composto por 2 blocos contendo 8x8 amostras, e que são sub-amostrados por um fator de 2, na direção horizontal e vertical. O número de amostras das componentes pode variar dependendo do profile ou de extensões do padrão. Os macroblocos são processados em slices que geralmente são conjuntos de macroblocos [**?**]. Um diagrama em blocos simplificado de um codificador H.264 é apresentado

na figura abaixo. Na sequência apresenta-se uma breve descrição de cada etapa representada pelos blocos.

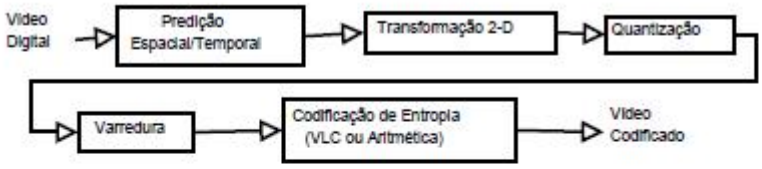


Figura 4.1: Diagrama em bloco da codificação H.264/AVC.

Após a captura dos macroblocos na entrada do sistema, as amostras passam pela etapa de predição, que pode ser temporal ou espacial. Essas etapas são chamadas de predição Inter e Intra respectivamente. Na predição Inter, são consideradas as semelhanças entre macroblocos de quadros consecutivos de um fluxo de vídeo e a predição Intra realiza a predição de amostras do macrobloco com base em informações de macroblocos já transmitidos de um mesmo quadro, de forma análoga ao processo utilizado na codificação de imagens estáticas. A predição Inter utiliza-se de um mecanismo conhecido como compensação de movimento e gera um vetor de deslocamento que é transmitido com o vídeo codificado e posiciona o macrobloco atual em relação a um quadro de referência previamente codificado. A etapa de transformação permite diminuir a redundância espacial das amostras [**?**] convertendo as amostras do sinal do domínio espacial para o domínio da frequência, utilizando a Transformação de Inteiros, em substituição a DCT (Discrete Cosine Tranform) utilizada em padrões anteriores.

Na etapa de quantização, cada amostra é dividida pelo parâmetro de quantização (Qp) que pode ser diferente para cada macrobloco e que para amostras de 8-bits pode assumir 52 possíveis valores [**?**]. Após a transformação e quantização, a varredura é realizada sobre a matriz resultante das etapas anteriores com o objetivo de colocar os coeficientes de maior variância no inicio da sequência de coeficientes.

A varredura também tem por objetivo, maximizar o número de coeficientes zerados consecutivos e assim representa-los de uma forma bastante compacta na etapa de codificação de entropia. O tipo de varredura também varia com o tipo de predição utilizada [**?**]. A codificação de entropia é uma etapa que não gera perda de informação do vídeo e é responsável pela efetiva compactação dos dados. O objetivo principal é representar os dados de maior ocorrência na sequência pela menor representação binária possível. Diferente dos padrões anteriores, no H.264 utilizase uma codificação de entropia adaptativa ao contexto, sendo assim muito mais eficiente [**?**]. Durante a predição do macrobloco, o H.264 testa todos os possíveis modos de codificação na seguinte ordem : SKIP, 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4 [**?**], o que demanda um alto custo computacional.

Capítulo 5

Avaliação da Qualidade de Vídeos 3D

Avaliação da qualidade é um componente chave de qualquer sistema de serviço relacionado a vídeo. A avaliação da qualidade de vídeo 3D e imagens estéreo é alcançável através de avaliação subjetiva ou através de métricas objetivas. A melhor maneira de avaliar a qualidade do vídeo seria executar testes subjetivos de acordo com protocolos padronizados, que são definidos a fim de obter correta avaliação da qualidade e confiabilidade de vídeo através de participantes. Enquanto os resultados dos testes subjetivos continuam a ser os melhores padrões de medição de qualidade em pesquisa de vídeo, as métricas objetivas são de extrema importância para o desenvolvimento de tecnologias de vídeo 3D.

Os métodos de avaliação da qualidade de imagem 2D não são adequados para medir a qualidade da imagem em 3D, uma vez que a profundidade (o fator mais importante em um sistema 3D) e as distorções típicas da estereoscopia (por exemplo, *crosstalk*) não são incorporadas [28]. Até agora, poucos métodos objetivos de avaliação de imagens estéreo foram apresentadas [3], que usaram o mapa de profundidade para avaliar o sentido estéreo. O Mapa de profundidade, trará três questões: primeiro, o processo de encontrar os mapas de profundidade das imagens é um processo altamente intensivo e demorado. Em segundo lugar, não é fácil calcular com precisão o mapa de profundidade, por causa da oclusão, do ruído da câmera, e assim por diante. Terceiro, é difícil decidir qual o grau de profundidade é bom, e quanto mais, melhor ou menos é melhor.

A avaliação subjetiva da qualidade do vídeo pode produzir medidas de qualidade perfeita e desejada. No entanto, o uso do testes subjetivos é demorado devido aos preparativos para o teste, requisitos, custo e consumo de

tempo. Além disso, a análise dos dados coletados não é tão simples e há uma demanda crescente para o uso de avaliações objetivas de qualidade(AOQ).

A tecnologia 2D atualmente existente mede a qualidade objetiva de cores individuais e as sequências de mapa de profundidade não representam a verdadeira qualidade da imagem percebida pelos usuários. Embora alguns pesquisadores usem relação sinal-ruído de pico (PSNR) para avaliar vídeos 3D, as limitações de PSNR também foram analisadas para vídeos 2D.

Estas limitações são similares as avaliações da qualidade de imagem em vídeos 3D, e o PSNR não fornece informações sobre a percepção de profundidade. Com a abordagem AOQ, algoritmos computacionais são usados para estimar a qualidade dos vídeos sem a necessidade de usar seres humanos. Assim, utilizamos a avaliação da qualidade subjetiva (ASQ) para confirmar a avaliação da qualidade objetiva (AOQ) das sequências de vídeo 3D e imagens.

Muitas acadêmias e as comunidades industriais tem concentrado esforços para desenvolver modelos de testes subjetivos que avaliem ambas as sequências de vídeo 2D e 3D, de forma que esse modelos tenham uma melhor qualidade no processamento e no recebimento dos dados.

Com o advento da tecnologia 3D aplicada a diferentes áreas como esportes, eventos corriqueiros, filmes, séries de TV e documentários, aplicações médicas, jogos, as imagens 3D e vídeos precisam ser processadas, transmitidas e distribuídas para vários usuários. Portanto, é importante definir procedimentos subjetivos e objetivos para avaliar a qualidade do processamento de vídeos estéreos.

Os atributos multidimensionais, tais como bloqueio e ruído associado com vídeo 2D não pode ser usado para medir as qualidades perceptualmente importantes do vídeo 3D, como noção de percepção de presença, nitidez naturalidade e profundidade que está relacionado com a sensação de imersão no cenário 3D.

As diferentes tecnologias de *displays* 3D que são baseados nessas diferentes características levam em consideração o impacto de cada tecnologia na visualização pelo usuário e que fatores devem ser analisados para medir a qualidade de vídeo 3D e como eles impactam na percepção visual.

5.1 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 através de canais sem fio, como redes celulares.

Investigação no campo da avaliação de vídeo 3D de qualidade depende sobre a disponibilidade de pontos subjetivos, *Mean Opinion Score* (MOS), coletados por meio de experimentos onde grupos de pessoas são convidadas a avaliar a qualidade do 3D nas sequências de vídeo. A fim de recolher dados significativos que são confiáveis e estatísticos, testes subjetivos devem ser cuidadosamente projetados e realizadas, e requerem um número relevante de participantes.

Estes testes são bastante demorado, no entanto, são fundamentais para testar e comparar o desempenho do algoritmos objetivos,ou seja, que tentam prever a qualidade de vídeo 3D através 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 [36]. 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 [1].

Além de auxiliar na melhora do desempenho de avaliações objetivas, as métricas de qualidade subjetiva são extremamente importantes no que diz respeito ao tempo de resposta, ou seja, a opinião do usuário é automaticamente produzida após o teste.

Selecionado os grupos de usuários para o experimento, eles serão submetidos a um questionário de perguntas referente aos vídeos visualizados, em que o questionário é baseado na escala definida logo a seguir.

- Excelente, Bom, Razoável, Não Sabe, Ruim;
- Impacto Alto, Impacto moderado, Pouco Impacto, Não Sabe, Nenhum impacto;
- Muito importante, Importante, Pouco importante, Não Sabe, Nenhuma importância;
- Concordo totalmente, Concordo, Discordo, Discordo totalmente;

Após o preenchimento do questionário e a sumarização dos pontos, é calculado o MOS para cada condição de teste de acordo com a fórmula 5.1.

$$MOS_j = \frac{\sum_{i=1}^{N} m_{ij}}{N} \tag{5.1}$$

em que, N é o número de indivíduos, m_{ij} a pontuação por assunto i para cada teste j.

A relação entre os valores médios estimados com base em uma amostra da população (isto é, os indivíduos que participam e completam o questionário da pesquisa) e os verdadeiros valores médios de toda a população é dada pelo intervalo de confiança da média estimada. A média é estimada utilizando alguma distribuição de probabilidade. Em [1], foi utilizado a distribuição de probabilidade T-Student.

Contudo, a medição de qualidade de vídeo subjetiva também pode ser um desafio, porque ele pode exigir pessoas capacitadas e treinadas para julgar a sua qualidade. Muitas medidas de qualidade subjetiva de vídeo são descritos na seção de recomendação do ITU-T BT.500. Sua ideia principal é o mesmo que no *Mean Opinion Score* para sequências de vídeo que são apresentados ao grupo de telespectadores e, em seguida, sua opinião é gravada e gerada a média para se avaliar a qualidade de cada sequência de vídeo.

Otimização de sistemas de vídeo 3D em tempo hábil é muito importante, portanto,é necessário que as medidas subjetivas de confiança sejam calculadas com base na análise estatística. Testes subjetivos são realizados para verificar a qualidade de vídeo 3D e a percepção de profundidade de uma série de sequências de vídeo codificados de maneira diferente, com taxas de perda de pacotes que variam de 0% a 20%. Os resultados de qualidade subjetiva são usados para determinar com mais precisão a qualidade das métricas de avaliação objetiva para sequências de vídeo 3D, como o média PSNR, as semelhanças estruturais (SSIM), *Mean Square Error (MSE)*.

Na avaliação subjetiva das imagens em 3D e sequências de vídeo, a qualidade do vídeo é mais próximo do conceito de qualidade da experiência e deve ser considerada multi-dimensional: a qualidade visual, qualidade da profundidade/percepção e conforto. A primeira dimensão pode ser considerada a qualidade visual, no sentido de 2D. O valor acrescentado de profundidade foi, muitas vezes proposta como um segundo critério, e o termo "naturalidade"foi proposto para expressar a combinação da profundidade percebida e a qualidade geral [21].

Como a tecnologia de exibição 3D ainda está em avanço, existem diferentes tecnologias e nenhuma pode ser recomendada como uma referência. O ângulo de visão, o campo de visão, a quantidade de *crosstalk* e o brilho são muitas vezes fatores limitantes.

5.2 Métricas Objectivas: De 2D para 3D

Em comparação com o vídeo 2D, a avaliação objetiva de vídeo em 3D é mais complexa:

- Opinião do observador pode ser considerado como multidimensional, incluindo fatores como a fadiga visual e a percepção de profundidade;
- Mais aspectos do HVS precisam ser abordadas, por exemplo, rivalidade binocular e a supressão binocular.

Na cadeia de transmissão em 3D, artefatos visíveis ocorrem em vários locais. Na captura da câmera ou nas etapas de conversão e renderização tem-se a introdução das degradações geométricas [46, 19].

Deve ser mencionado que o conteúdo do vídeo em si tem um impacto maior sobre a qualidade percebida visual em 3D do que em 2D.

Estudos têm indicado que os telespectadores tendem a focar sua atenção em áreas específicas de interesse na imagem e modelos de atenção visual têm sido propostas [24]. Há um interesse crescente na utilização de modelos de atenção visual (mapas saliência) no interior de modelos de avaliação de qualidade de vídeo, a fim de melhorar sua precisão. Atenção visual é sem dúvida também um fator crucial na percepção de vídeo 3D [19].

5.3 Métricas Objetivas de qualidade em 3D

Em [19] é apresentado que existe uma falta de metodologias confiáveis para a avaliação subjetiva de vídeos em 3D e que essa falta de metodologias pode afetar o valor de validação das métricas objetivas. No entanto, novas métricas estão sendo desenvolvidas afim de que se produzam técnicas subjetivas mais confiáveis e qualidade dos vídeos 3D seja melhorada.

Um algoritmo de avaliação objetiva, que utiliza o mapa de profundidade, bem como a visão estereoscópica é proposto em [40]. Inclui as partes do índice de similaridade estrutural (SSIM) e a detecção de degradações borda e cor.

A supressão binocular indica que um ponto de vista do par estéreo pode ser transmitido a uma qualidade visual pior do que outro. Esta investigação é realizada em [31] usando um modelo de taxa de distorção com base na estimativa da qualidade visual com um sinal de pico adaptado para ruído (PSNR) e uma métricas de *jerkiness*.

A aplicabilidade da PSNR e de modelos de vídeo 2D (SSIM e VQM) para o 3D foi investigada em um conjunto de dados pequeno, tanto para o caso de vídeo estereoscópico, quanto para vídeo com informações de profundidade monoscópico [50]. Os resultados mostram que a qualidade de vídeo 3D pode ser estimada a partir de avaliação separada de cada vista, enquanto modelos para 2D também poderia ser usado para estimar a qualidade de percepção de profundidade. Uma análise similar é feita em [17], na qual foram realizados os

experimentos subjetivos com um dispositivo Philips com tela de 42 polegadas auto-estereoscópica para apresentar o conteúdo 2D mais profundidade. No entanto, um estudo mostrou que os indivíduos preferem desligar o efeito 3D no *display* [19].

5.4 Avaliação da profundidade

Avaliação da qualidade de vídeo 3D representa novos desafios para a comunidade científica 3DTV, dentre elas a avaliação da percepção da profundidade permanece sendo um problema sem resposta. Enquanto os seres humanos usam diferentes cálculos fisiológicos e psicológicos para perceber a profundidade [10], a estereoscopia binocular é a sugestão de profundidade fornecida pelos modernos *displays* 3D estereoscópico, em comparação aos tradicionais *displays* 2D. Estereoscopia binocular é a sensação pela qual o cérebro interpreta as informações de profundidade, fazendo uso dos dois pontos de vista ligeiramente diferentes visto pelos dois olhos. Assim, o cérebro humano é capaz de perceber a profundidade por meio da análise das disparidades de diferentes objetos nas vistas estereoscópicas.

Um dos primeiros estudos explorou a variação da experiência humana em três atributos específicos do vídeo 3D. Esses atributos são a percepção da presença de profundidade e a naturalidade da profundidade. Foi mostrado que com o aumento da profundidade se obtém um maior sentido de presença, desde que a profundidade é percebida como natural [10].

O efeito da qualidade da percepção da profundidade da imagem 3D foi apresentado em [25]. A profundidade foi quantizada em diferentes taxas de *bits* e avaliados subjetivamente para a percepção 3D em geral. Concluiu-se em [25], que a imagem de profundidade pode ser significativamente quantizada sem afetar a qualidade da percepção 3D.

Em [17], a correlação entre as métricas de qualidade existentes de vídeo 2D e os resultados subjetivos da percepção do atributo da cor *plus* do vídeo com profundidade estereoscópica, foi analisada. E foi concluído que, das métricas 2D avaliadas, o *Video Quality Metric* (VQM) [35] da componente de cor é fortemente correlacionada com a percepção geral da profundidade.

Em [42], foi relatado que as distorções na profundidade do vídeo 3D são menos significativas do que as distorções de cor e que a percepção da profundidade não muda com diferentes níveis de quantização da imagem de profundidade, não explicando a razão.

CAPÍTULO 6

Exibição/Renderização

6.1 Visualização de Vídeos 3D

Na visualização do vídeo 3D para a televisão é necessário que o cérebro trabalhe da mesma maneira, com as 2 imagens. Para exibição do vídeo 3D existe duas vertentes: a estereoscopia que se refere a vídeo 3D (com óculos) e auto-estereoscopia que representa vídeo 3D sem óculos.

Na estereóscopia são utilizados óculos, passivos ou ativos. Os ativos devem ser alimentados normalmente por baterias, ao passo que os passivos não requerem alimentação.

6.1.1 Óculos 3D Ativos

Os óculos obturadores (*shutter glasses*) são os óculos ativos 3D mais comuns. As lentes são basicamente pequenos tela de LCD e por isso quando uma tensão é aplicada nas mesmas, as lentes escurecem, ou seja, o obturador fecha.

Os óculos de cristal líquido obturadores ativos são usados para bloquear e desbloquear alternativamente a visibilidade das imagens do olho esquerdo e do olho direito na sincronização com uma taxa de exibição do monitor. Assim, quando a vista esquerda está sendo apresentada na tela, o obturador está aberto, deixando o obturador direito fechado, e vice-versa. Como resultado, cada olho somente visualiza seu respectivo.

Taxas mais lentas (quadros/s) de atualização pode introduzir *flicker* (Cintilação indesejável na imagem que aparece na tela do monitor). Este, porém, geralmente não é visível, desde que a frequência da mudança seja maior do que a frequência de *flicker* crítica (CFF) do sistema visual humano [41].

As imagens sucedem-se com uma velocidade tal que o cérebro interpreta as duas imagens recebidas por cada olho em instantes de tempo consecutivos, em conjunto, gerando uma imagem 3D.

Os óculos obturadores necessitam apresentar uma combinação transmissor-receptor que utilize tecnologia de infra-vermelhos, Bluetooth ou rádio. A televisão tem de enviar um sinal para os óculos se sincronizarem com a apresentação de imagens na tela. Usualmente é usado um feixe de infra-vermelhos que é difundido tal como nos comandos das TV's. Esse sinal é depois captado pelo receptor eletrônico presente nos óculos. A obturação de cada lente é comandada por sinais eléctricos alternados, o que está de acordo com o funcionamento dos LCD's [14].

Figura 6.1: Óculos obturadores da XpanD (únicos que podem ser usados tanto em cinema digital como em televisores) [14].

A maior desvantagem do 3D ativo é que o conteúdo vai ser alternado entre os dois olhos ao ritmo em que o conteúdo é transmitido, ou seja, por exemplo se forem emitidas para o televisor 50 imagens por segundo, a cada olho apenas chegam 25 imagens por segundo.

A frequência de imagens em quadros por segundo (*frame rate*) do ponto de vista do utilizador passou para metade. Em vez de 50 imagens 2D o utilizador passa a visualizar 25 imagens 3D. Esta quebra do ritmo tem consequências. Isto porque quanto maior esse ritmo, mais suaves são as transições entre imagens e portanto mais fluida é a visualização de movimento nos conteúdos emitidos.

Assim todo o equipamento na cadeia de transmissão e recepção tem de ser capaz de processar imagens ao dobro do ritmo anterior para se manter a qualidade do vídeo. Isto significa que os requisitos de *hardware* do equipamento a utilizar duplicam. O ritmo de atualização dos elementos de imagem (*refresh rate do televisor*) é ao contrário do que acontece com o *frame rate*, o suportado pelo televisor.

Nos LCD's que suportam esta tecnologia o *refresh rate* é usualmente elevado (de 100 Hz na Europa e de 120 Hz na América). Isto para tornar as imagens mais nítidas e reduzir o efeito de *motion blur* (o borramento da

imagem quando as transições entre imagens são muito bruscas, devido ao movimento muito rápido da cena).

Ainda que o *refresh rate* seja elevado, a quebra do *frame rate* nesta tecnologia pode ainda assim provocar ligeira trepidação (*flicker*) em conteúdo apresentado em câmara lenta ou com rápido movimento [14].

Esta é a tecnologia de vídeo 3D com óculos que produz a melhor qualidade, sendo a tecnologia mais vulgar e mais cara no mercado dos televisores 3D.

Prós: Excelente qualidade de imagem.

Contras: televisor e óculos caros.

6.1.2 Óculos 3D Passivos

A tecnologia passiva é alcançada por polarização da luz e sinais de disparo eletrônico ou óptico, não obrigatórios. Neste método os filtros diferenciais das imagens da tela do olho esquerdo e do olho direito usam a polarização da luz, essas imagens são então vistas com filtros de polarização correspondentes colocado na frente do olho esquerdo e do olho direito. Os filtros circularmente polarizados são geralmente preferidos porque permitem movimentos a mais na cabeça sem afetar a separaçao da vista.

Óculos com filtros de cor

Os primeiros óculos 3D inventados usam um método denominado de Anaglífico de Cores Complementares. A tinta atua como um filtro de cor que consegue de alguma maneira filtrar uma imagem para o seu respectivo olho.

O método de visualização estereoscópica anaglífica, é o mais simples dos métodos. Essa técnica caracteriza-se em 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 [27]. Para a visualização o espectador necessita utilizar óculos com um lado com lente vermelha (esquerda) e o outro com lente ciano (luz ciano = luz verde + luz azul), como apresentado na Figura 6.2.

Para a codificação do vídeo estereoscópico é necessário separar os canais RGB dos vídeos de cada umas das vistas (olho direito e olho esquerdo) do par estereoscópico, Figura 6.3. Compondo-se a componente do olho direito e a componente do olho esquerdo em um novo vídeo RGB. A imagem anaglífica resultante pode ser observada na Figura 6.4.

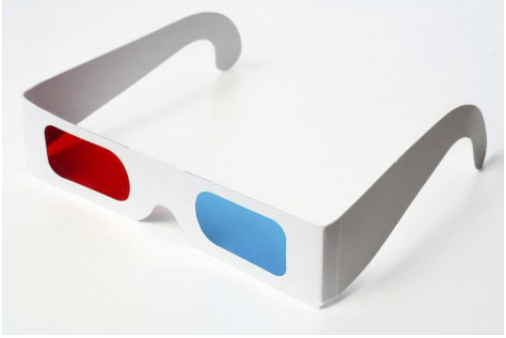


Figura 6.2: Óculos com filtros de cor vermelho e cyan [14].

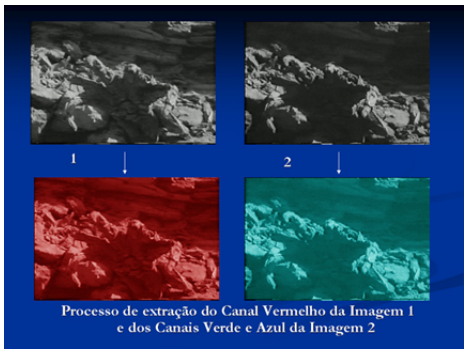


Figura 6.3: Processo de extração do Canal Vermelho do vídeo 1 e do Canal Ciano do vídeo 2.

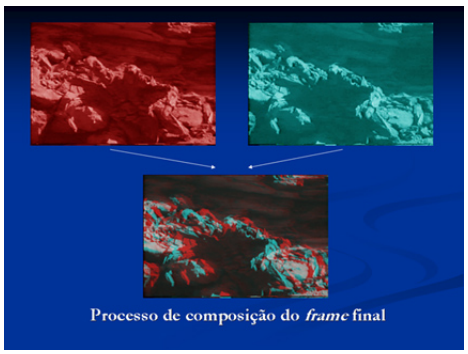


Figura 6.4: Exemplo de um quadro anaglífico (a percepção de profundidade desta imagem pode ser observada com óculos anaglífico ciano-vermelho)

A lente vermelha deixa passar as "partes vermelhas"da imagem como partes claras (brancas) e a lente ciana deixa passar as "partes azuis e verdes"da imagem que aparecerão como claros e bloqueia as partes vermelhas da imagem que aparecerão como escuros.

Deste modo cada olho vê uma imagem (claros e escuros) diferente, que interpretadas em conjunto pelo nosso cérebro dão origem a uma imagem 3D. O cérebro em todo este processo fornece uma adaptação de modo a dar cor à imagem.

Este formato de emissão 3D é limita a quantidade de cores que podem ser usadas para criar conteúdo, o que conduz a que esta técnica não seja muito realista, sendo por isso pouco imersiva, razão pela qual já não se recorre a ela quer nos cinemas quer nas TV's [14].

Prós: Barato e fácil de criar a ilusão 3D.

Contras: Pode causar dores de cabeça ou náuseas e a qualidade do 3D é má devido à perda e falta de variedade de cor.

Óculos de polarização 3D

Lentes linearmente polarizadas usam polarização vertical em uma lente e polarização horizontal na outra. Duas imagens são captadas de dois ângulos ligeiramente diferentes e são projetadas com um projetor cada. Cada imagem tem de ter sido previamente polarizada da mesma forma que as lentes dos óculos, de forma que cada imagem chegue ao seu respectivo olho. A superfície na qual as imagens são projetadas é coberta com elementos químicos especiais para não afetar o efeito da polarização [14].

Assim produz-se um efeito 3D desde que os utilizadores mantenham a cabeça na posição correta. Inclinar a cabeça irá "estragar"o efeito 3D, o que se entende observando a Figura 6.5. Observe que se o utilizador inclinar a cabeça, as lentes polarizadas já não coincidem com as respectivas imagens polarizadas [14].

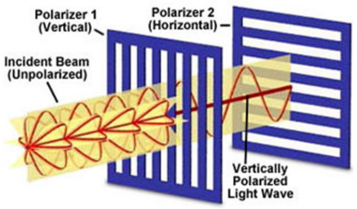


Figura 6.5: Luz a passar por polarizadores [14].

Nas lentes com polarização circular, uma lente é polarizada no sentido dos ponteiros do relógio, a outra no sentido contrário aos ponteiros do relógio. Com esta tecnologia, o efeito 3D é mantido se o utilizador inclinar a cabeça e além disso é apenas necessário um projetor. Um polarizador circular tem de ser colocado à frente do projetor. Este polarizador alterna rapidamente entre os dois sentidos de polarização para criar o efeito 3D, polarizando cada duas imagens consecutivas em sentidos inversos uma da outra.

Os cinemas estão agora a fornecer óculos com lentes polarizadas para a visualização de filmes 3D. Foi esta a tecnologia que foi utilizada na transmissão e nos óculos que permitiram ver o filme Avatar nos cinemas em 3D [14].

Essa emissão foi de 24 imagens (*frames*) por segundo para cada olho, ou seja, 48 imagens 2D correspondendo a 24 imagens 3D por segundo. Para reduzir o efeito de *flicker* projeta-se cada imagem 3 vezes na tela [14].

Todo o equipamento envolvido em transmissões deste tipo no cinema é muito caro, excepto os óculos que são baratos. Esta tecnologia tem maior probabilidade de ser afetada por *crosstalk* (Apêndice A) que a tecnologia ativa (imagem destinada a um olho ser parcialmente ou totalmente captada pelo outro, o que causa efeito fantasma).

Figura 6.6: Óculos de polarização 3D no cinema [14].

Ao ser usado apenas um projetor, os sistemas 3D de visualização de filmes no cinema que utilizam luz polarizada causando uma perda de brilho na tela ainda maior que no caso de uso de óculos ativos. Isto deve-se à distribuição do conteúdo entre os dois olhos e aos próprios filtros polarizadores nos óculos que bloqueiam ainda mais radiação. Isto pode ser corrigido recorrendo a projetores mais brilhantes.

Figura 6.7: Óculos de polarização 3D [14].

Prós: Óculos baratos e de peso leve, imagens com grande nível de detalhe e cor, boa qualidade de imagem [14].

Contras: Maior probabilidade de ocorrer *crosstalk*.

Apenas a LG conseguiu até agora tornar a tecnologia de polarização circular viável para televisores. As restantes usam polarização linear e reduzem a resolução do conteúdo 3D a metade da resolução do conteúdo 2D equivalente. Isto porque é feito o entrelaçamento das duas vistas 2D na tela. Além disso estes televisores contêm o problema da inclinação da cabeça estragar o efeito 3D, embora este tipo de visualização seja menos afetado pela perda de brilho [14].

6.1.3 Tecnologias Sem Óculos

Como se sabe a necessidade de usar óculos 3D é uma das maiores barreiras para a aceitação em massa da televisão 3D como um meio de entretenimento. Os óculos podem ser caros, desconfortáveis para alguns e a necessidade de usá-los significa que existe a obrigação de possuir múltiplos pares para se visualizar conteúdo 3D com a família ou amigos.

Nos *displays* autoestereoscópicos, as visões esquerda e direita são multiplexadas espacialmente, permitindo ao observador visualizar uma imagem tridimensional sem a necessidade óculos especial. Cada imagem do par estéreo é "fatiada" e reside sobre as colunas pares e ímpares do monitor. As fatias são direcionados para o olho do observador por meio de uma película lenticular colocada na superfície do monitor (Figura 6.8) ou pelo cálculo de distância e posicionamento dos olhos do observador. Maiores detalhes podem ser encontrados em [43].

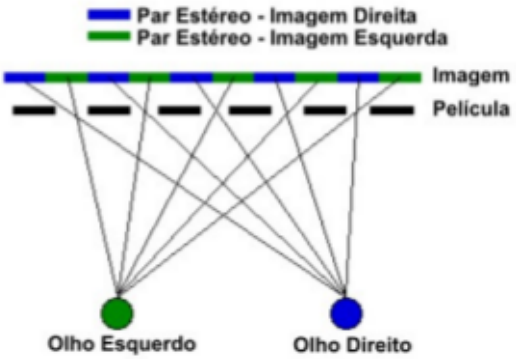


Figura 6.8: Display Autoestereoscópico. [43].

Tecnologia de Barreira de Paralaxe

Esta tecnologia funciona da seguinte forma: pequenas lentes são integradas na tela da TV. Essas lentes consistem em camadas de LCD's. Cada uma dessas camadas contém pequenas tiras que escondem pixels específicos (criando fendas de precisão) de forma que um conjunto de pixels só possa ser visualizado pelo olho esquerdo e outro conjunto de pixels só possa ser visualizado pelo olho direito [14].

Devido ao fato das lentes serem embutidas na tela, o uso de óculos para a visualização do 3D deixa de ser necessário. A principal desvantagem da tecnologia de barreira de paralaxe é que esta apenas funciona se o utilizador permanecer num local bem definido ("*sweet spot*") sendo que o efeito 3D desvanece caso este se desloque para um outro local. Esse local é sensivelmente, em frente da tela, a uma certa (curta) distância. Por essa mesma razão os pri-

meiros equipamentos do tipo fabricados pela Toshiba por exemplo, têm sido relativamente pequenos. O brilho da tela é também afetado nesta tecnologia como acontece em todas as outras [14].

Para se obter compatibilidade direta, é possível visualizar conteúdo 2D com esta tecnologia bastando tornar a barreira de paralaxe transparente, de forma que a luz a atravesse e os dois olhos vejam a mesma imagem 2D. Isto pode ser feito com sinais elétricos uma vez que a barreira consiste em camadas LCD [14].

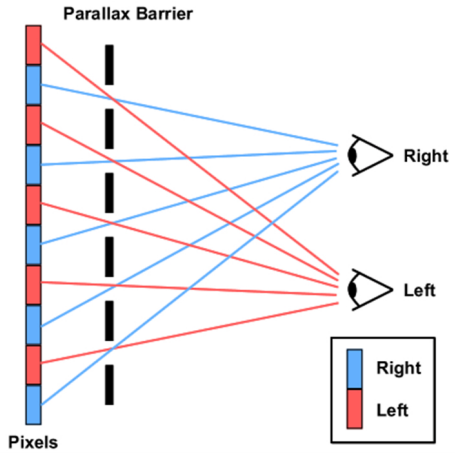


Figura 6.9: Tecnologia de barreira de paralaxe [14].

Tecnologia de lentes lenticulares

Outro método de proporcionar 3D TV sem óculos é utilizar lentes lenticulares (lentes convexas), que são projetadas e realizadas para que uma imagem diferente seja recebida por cada olho dependendo do ângulo de visualização do utilizador.

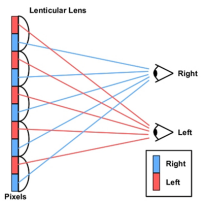


Figura 6.10: Tecnologia de lentes lenticulares [14].

As duas imagens 2D ligeiramente diferentes, a ser transmitidas para gerar uma imagem 3D, são alternadas em linhas (tiras) na tela. As lentes referidas são alinhadas com as imagens a ser transmitidas de tal forma que a luz refletida por cada tira é refratada numa direção ligeiramente diferente, mas a luz refletida pelos *pixels* que emitem a mesma imagem é enviada na

mesma direção (ângulo). O resultado final é que cada olho vê uma imagem inteira diferente, o que conduz à visualização de uma imagem 3D.

Com esta tecnologia consegue-se uma menor redução de brilho do que na tecnologia de barreira de paralaxe e consegue-se também uma maior abrangência de ângulos de visualização (permitindo que mais do que uma pessoa visualize a emissão 3D ao mesmo tempo), que no caso do uso de tecnologias de barreira de paralaxe é muito restrito.

A abordagem de vistas múltiplas com lentes lenticulares, reduz a resolução da imagem, porque muitos *pixels* são utilizados para mostrar a mesma imagem 3D a ângulos diferentes. Ou seja, existe o dilema: a televisão necessita de suportar o efeito 3D para muitos ângulos de visualização para tornar a 3D TV sem óculos prática para vários utilizadores. No entanto, a resolução irá degradar-se à medida que mais ângulos são adicionados [14].

Tecnologia de monitorização de utilizador

Esta tecnologia (denominada de *head tracking* na linguagem anglo-saxônica) funciona por meio do uso de uma webcam que segue os olhos do utilizador e ajusta as imagens enviadas da tela para cada olho à medida que o utilizador se move para que este veja em 3D.

Esta tecnologia só funciona claramente para um utilizador de cada vez, razão pelo qual parece mais adequada a utilizar em aparelhos mais pessoais, com telas pequenas como celulares (*smartphone*) ou computadores portáteis. De fato, já foram mostrados computadores portáteis 3D que recorrem a esta tecnologia. Repare-se que assim já não existe mais o problema do ângulo de visualização do utilizador, como no caso do uso de barreira de paralaxe [14].

Tecnologia de 3D automático

Para solucionar o problema de falta de conteúdos 3D disponibilizados pelo mercado, numa fase inicial alguns televisores 3D, da Samsung por exemplo, realizam conversão de conteúdos emitidos em 2D para 3D em tempo real. Este método funciona com DVD's, Blu-Ray, e TV (*broadcast*). No entanto, a precisão obtida neste tipo de vídeo 3D nomeadamente no que toca à profundidade, está longe do verdadeiro 3D estereoscópico filmado com duas câmaras.

O processo limita-se a simular um efeito 3D não contendo qualquer informação da profundidade da cena a ser exibida. Logo não se oferece a verdadeira informação de profundidade da cena ao contrário do que acontece com o verdadeiro 3D.

6.2 Conclusão

Garantindo que as diferentes visualizações são corretamente apresentados para os diferentes olhos, que provou ser um grande desafio. Na verdade, com ambas as tecnologias passivas e ativas da separação dos olhos estão longe de ser perfeito, pois algumas das informações destinados a um olho é visto pelos outro olho. Este vazamento de informações de um olho para o outro olho é conhecido como *crosstalk* (Apendice A) [41]. Um dos requisitos mais importantes de um sistema estereoscópico é a capacidade de limitar a quantidade de *crosstalk*. Sabe-se que mesmo uma pequena quantidade de *crosstalk* pode ter um efeito negativo sobre a qualidade da imagem. No entanto, menos claro qual a quantidade de *crosstalk* tem um efeito negativo sobre o conforto [41].

CAPÍTULO 7

Representação de vídeos 3D

A tecnologia atual de exibição de vídeo 3D consiste de telas planas, oferecendo apenas a ilusão de profundidade, representando as imagens que são vistas pelos dois olhos com um ângulo de paralaxe [22]. A paralaxe é o ângulo entre as linhas de visão que leva à disparidade entre as duas imagens da retina.

Hoje a maioria dos *displays* 3D são *displays* estéreos que requer exatamente dois pontos de vista a cada instante. Um display estéreo para vários visualizadores especiais requer óculos 3-D que filtram a visão correspondente para os olhos esquerdo e direito de cada espectador [4]. De um ponto de vista da compressão, uma abordagem simples para representar efetivamente um sinal de vídeo estéreo é dada tratando como dois sinais de vídeo com dependência estatística. A dependência estatística entre as duas visões podem ser exploradas por meio de técnicas de compressão conhecidas.

Para alguns *displays* stereos, a disparidade na configuração da câmera estéreo pode não coincidir com a melhor paralaxe de visualização natural no *display* 3-D. Assim, um dos dois pontos de vista precisa ser reposicionado. Este processo é chamado estéreo *repurposing*.

Normalmente, os dois pontos de vistas adquiridos constituem uma boa base para o calculo de um outro ponto de vista entre eles, utilizando informações de geometria da cena adicionais, como dados de profundidade ou disparidade.

Visto que não são in-between os pontos de vista adquiridos são mais críticas como conteúdo de fundo é revelado, onde nenhuma informação a partir de qualquer ponto de vista é disponível.

Além disso, o problema é mais grave da geração da visão é quando precisa ser feito utilizando vistas comprimidas com o ruído de quantização, como tipicamente afeta o processo de estimativa de profundidade ou a disparidade

dos dados que são necessários para a síntese de vista. Outro problema com a estimativa da profundidade no receptor é que ele geralmente requer consideráveis recursos computacionais e algoritmos diferentes que produzem resultados diferentes. Uma última consideração deve ser dada ao fato de que o proprietário do conteúdo teria controle limitado sobre a qualidade resultante exibida em caso de uma estimação diferente da profundidade e da visualização de algoritmos de síntese serem usados nas extremidades da recepção [29].

A adição do *display* estéreo e também do *display multiview* estão cada vez mais disponíveis [4, 22]. Como o *display multiview* normalmente não exige óculos 3-D, um dos maiores obstáculos na aceitação do usuário da 3DV é superada. No entanto, um *display multiview* exige a disponibilidade de muitos pontos de visualização. Por exemplo, protótipos atuais emitem oito ou nove pontos de visualização. Espera-se que a melhoria da qualidade de *displays multiview* será regido por um aumento de pontos de visualização e podemos esperar *displays* com 50 ou mais pontos de visualização no futuro. Por isso, é necessário um formato que permite a geração de números arbitrários de pontos de visualização, enquanto a taxa de *bits* de transmissão é constante.

A multi visualização com base em um sinal de vídeo estereo no receptor sofre do mesmo problema que o *repurposing stereo*. Aqui o problema é realmente pior do que o estéreo *repurposing*, como muitos pontos de vista precisam ser gerados para exibe a multi visualização [22]. Assim, um usuário percebe muitos pares de visualização, que consistem em duas visões sintetizadas por uma série de posições de visualizações, enquanto o par de visualização do *stereo repurposed* consiste de uma visão original e uma sintetizada.

Uma abordagem para superar os problemas com a geração de visualização no receptor é estimar os pontos de vista dos usuários e transmitir um sinal que permita a síntese de visão direta ao receptor. Tal sinal tem que estar relacionada com a geometria da cena. Neste trabalho, consideramos mapas de profundidade, em combinação com os sinais de vídeo estéreo como uma representação eficiente para a síntese de visão no receptor.

7.1 Soluções em 3DV com base em sinais estereo

Um sinal de vídeo estéreo capturado por duas câmaras de entrada é primeiro pré-processada, como apresentado na Figura 7.1. Isto inclui uma possível retificação do alinhamento da vista esquerda e direita [29], assim como cor e correção de contraste, por possíveis diferenças entre as câmeras de entrada. O formato 3-D é chamado de vídeo estéreo convencional (CSV – *conventional stereo video*) para a vista esquerda e direita. Este formato é

codificado por métodos de codificação de *multiview*, como especificado no perfil high do H.264/AVC, em que as dependências temporais de cada ponto de vista, bem como dependências entre ambas as vistas são explorados para uma compressão eficiente.

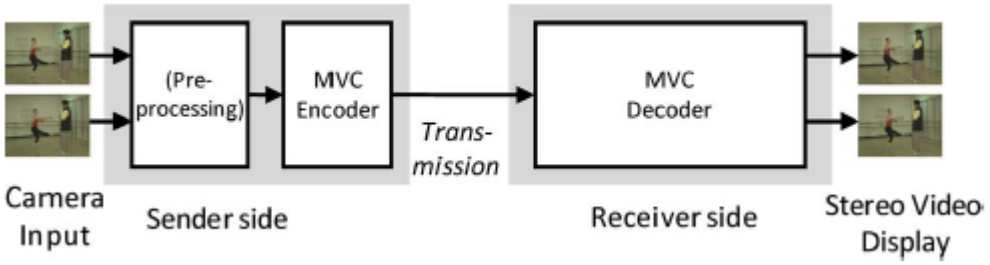


Figura 7.1: Primeira geração do sistema 3DV baseado em vídeo colorido estereoscópico [29].

Soluções padronizadas para o CSV tem seu mercado no cinema 3-D, *Blu-Ray Disc*, e *broadcast*. Enquanto o cinema 3-D é baseado em JPEG-2000, a especificação de Disco Blu-Ray 3-D é baseado no perfil high do H.264/AVC. Estes sistemas *stereo* oferecem uma solução 3-D robusta e fácil, visto que os formatos e a codificação 3-D incluem apenas dados de vídeo estéreo. Assim, o processamento complexo, como o fornecimento e estimativa da geometria 3-D da cena ou as adicionais síntese de exibição não são necessários e métodos de codificação podem ser otimizados para as estatísticas de dados de cor do vídeo. Por outro lado, esses sistemas são restritos a mostra estéreo que necessitam de óculos.

7.1.1 Comparação do *display* 3DV

Uma grande força para a tecnologia 3DV será a disponibilidade de alta qualidade autostereoscópica (sem óculos) *display multiview*. Aqui, um claro benefício do *displays multiview* é a exibição estereoscópica, que pode ser obtida para múltiplos usuário. Atualmente, tipos de visualização estéreo e multiview mostram vantagens e desvantagens específicas, que são resumidas na Tabela 7.1.

Aqui, as principais propriedades de aceitação do usuário são dadas e as entradas em negrito indicam que tipo de exibição melhor atende-los. Olhando para estas preferências do usuário, a exibição para múltiplos usuários têm o potencial para se tornar a primeira escolha para 3DV, como eles não exigem ajuda na visualização, como óculos estéreo para cenários multiusuário, dando uma impressão mais natural 3-D. Se os usuários se movem em frente da tela,

Tabela 7.1: Comparação das propriedades dos displays Multi e stereo usuários

Propriedade	*Display Stereo*	*Display* Multi-usuário
Ajuda Visualização	Requerido principalmente	Não requerido
Efeito **olhar ao redor**	Não	**Sim**
Resolução por visualizador	A**lto**	baixo
Profundidade de cena percebida	**Alta**	baixa

eles esperam o efeito do "olhar ao redor", ou seja, eles querem ser capazes de ver o fundo recém-revelada por trás de objetos em primeiro plano. Isso só pode ser oferecido pela exibição para múltiplos usuários, por eles fornecem múltiplos pares estéreo com conteúdo um pouco diferente para cada posição de visualização. Note que estas duas primeiras propriedades são de natureza sistemática, isto é, relacionadas com as limitações de vídeo estéreo convencional e, portanto, apenas suportado pela exibição para múltiplos usuários.

No entanto, para que os displays multiview se tornem amplamente aceitável, as desvantagens precisam ser eliminadas. Cada display geralmente sofre de uma resolução de tela limitada em geral, sendo o número de amostras disponíveis ser dividida em todas as vistas N. Isto leva a um problema de otimização para o número escolhido de pontos de vista, uma vez que por um lado, apenas alguns pontos de vista dão uma resolução mais alta *per-view* e, por outro lado, pontos de visualização são mais necessários para melhorar a visualização 3-D. A solução para isso é a fabricação de *displays* de ultra alta definição 3-D para múltiplos usuários, em que, por exemplo, 50 pontos de vista pode ser oferecido com cada visualização em alta resolução. Isso também melhora o problema do ângulo de visualização de telas *multiview* atuais, pois a faixa de visualização torna-se mais amplo.

APÊNDICE A

Efeito Crosstalk

Cross-talk ou imagem com fantasma é causada principalmente por: (i) persistência de fósforo em monitores CRT e (ii) técnicas imperfeitas de separação da imagem no qual a visão do olho esquerdo vaza até o ponto de vista do olho direito e vice-versa.

Cross-talk é percebida como um fantasma,sombra ou contornos duplos e é provavelmente um dos principais fatores que causam a má qualidade de imagem e desconforto visual.No entanto, para as técnicas de polarização linear o posicionamento incorreto da cabeça do observador (cabeça, por exemplo, inclinado) também causa o efeito fantasma de imagens.

Infelizmente, devido às limitações da tecnologia atual em dispositivos de apresentação,como monitores de vídeo, a separação das imagens esquerda e direita pode originar dois problemas. No primeiro, como os monitores são usados em alta frequência (de 100 a 120 Hz), os fósforos da tela não têm tempo suficiente para retornar ao seu estado de baixa energia entre as apresentações da imagem esquerda e da imagem direita. No segundo problema, os obturadores de cristal líquido dos óculos não podem bloquear 100% a passagem da luz. Parte da luz (aproximadamente 10%) pode passar por meio dos obturadores, permitindo que o olho veja, parcialmente, a outra imagem apresentada.

O efeito crosstalk não está restrito apenas a imagens estereoscopicas que necessitem de óculos para serem visualizadas, podendo acorrer em imagens autoestereóscopicas dependendo das técnicas de endereçamento utilizadas para cada olho.

Ambos os problemas acabam possibilitando que cada olho veja sua própria imagem, mais uma sobreposição, ou um "fantasma", da imagem do outro olho. Esse defeito, conhecido como efeito *Crosstalk*, não impede a visualização estereoscópica, mas causa desconforto visual no observador [43].

Algumas soluções já foram desenvolvidas para que esse problema seja minimizado. Algoritmos foram implementados para serem adaptados a cada tipo de exibição e/ou de óculos, aumentando o conforto visual do observador.

APÊNDICE B

Fundamentos Matemáticos

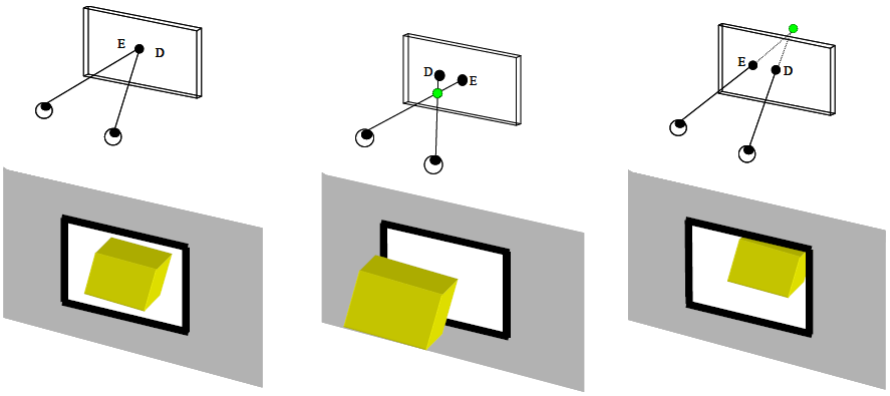
Existem diferenças entre imagens formadas nas retinas de cada olho quando sobrepostas. Estas diferenças são na direção horizontal. A disparidade é zero para objetos em que os olhos convergem.

Já a paralaxe é a distância entre os pontos correspondentes das imagens do olho direito e do esquerdo na imagem projetada na tela. Em outras palavras, disparidade e paralaxe são duas entidades similares, com a diferença que paralaxe é medida na tela do computador e disparidade, na retina. É a paralaxe que produz a disparidade, que por sua vez, produz o estéreo. Os três tipos básicos de paralaxe são:

- Paralaxe zero: conhecida como ZPS (*Zero Paralax Setting*). Um ponto com paralaxe zero se encontra no plano de projeção, tendo a mesma projeção para os dois olhos (Figura B.1a).
- Paralaxe negativa: significa que o cruzamento dos raios de projeção para cada olho encontra-se entre os olhos e a tela de projeção, dando a sensação de o objeto estar saindo da tela (Figura B.1b).
- Paralaxe positiva: o cruzamento dos raios é atrás do plano de projeção, dando a sensação de que o objeto está atrás da tela de projeção (Figura B.1c).

A paralaxe positiva P pode apresentar problema quando comparada à distância interaxial (t_c), distância entre os olhos. Quando P tem valor menor, mas próximo a t_c o resultado é ruim, a menos que se queira posicionar o objeto no infinito. Se P for maior que t_c, significa que há um erro, pois é um caso degenerado. Estes casos estão ilustrados na Figura B.2.

Deve-se tomar cuidado para que as projeções sempre caiam no retângulo que define o campo de visão no plano de projeção. Caso isto não ocorra,



(a) (b) (c)

Figura B.1: Tipos de paralaxe: B.1a Paralaxe zero (ZPS), B.1b Paralaxe negativa e B.1c Paralaxe positiva [8].

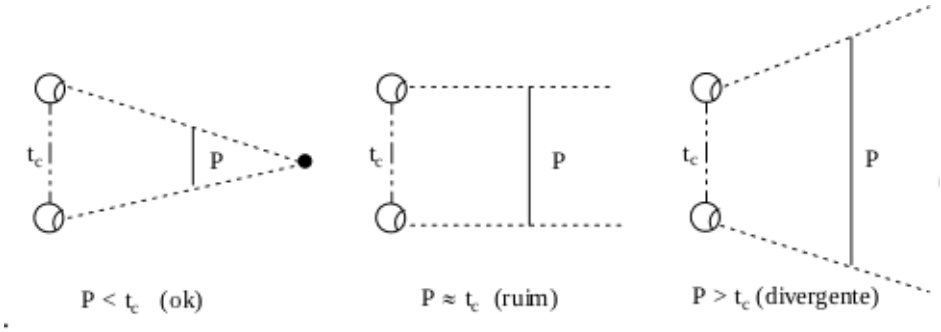


Figura B.2: Problemas com paralaxe positiva [8].

significando que apenas um dos olhos esteja vendo o ponto, a noção de estereoscopia é perdida. Isto apenas é aceitável para pontos que se movam rapidamente.

Um fator importante que deve ser levado em consideração é que a distância do observador à tela afeta o efeito de estereoscopia. Quanto maior a distância à tela, maior será o efeito estereoscópico (tanto positivo quanto negativo).

Um grande desafio da estereoscopia é gerar maior efeito de profundidade com menor valor de paralaxe devido ao espaço físico limitado da tela e da distância máxima que um ambiente comporta para os observadores. Em regra geral, o ângulo de paralaxe (β) deve estar no intervalo [-1,5°, 1,5°], definindo paralaxes mínimas e máximas. O esquema de controle da paralaxe é ilustrado na Figura 11.23, em que d é a distância do observador à tela.

Portanto $P = 2 * d * tan(\frac{\beta}{2})$.

Em uma situação com um *desktop*, normalmente $d = 60cm$. Portanto o valor máximo de paralaxe P_{max} é 1,57 cm. Em uma sala de visualização com distância média de 3 m, o valor máximo de paralaxe P_{max} é 7,85 cm.

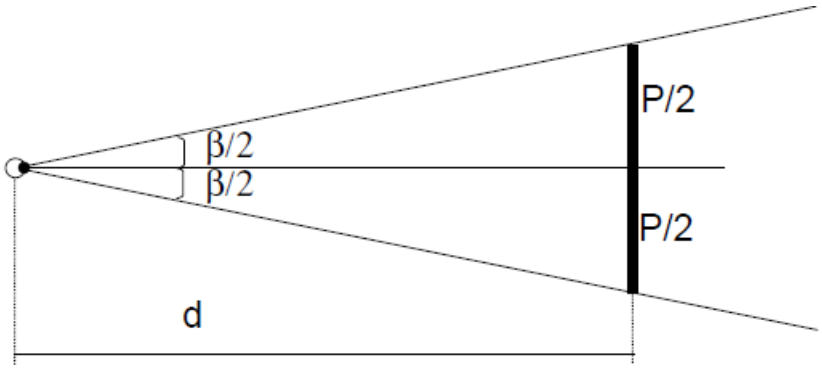


Figura B.3: Intervalo de controle do ângulo de paralaxe [8].

Esta paralaxe é relativa ao mundo físico, em que são feitas as projeções. Para transformar para o mundo virtual temos que dividir esta grandeza pela largura da tela de projeção e multiplicar pelo número de pontos horizontais.

A distância interaxial também influencia a paralaxe. Quanto maior a distância interaxial, maior é a paralaxe e, consequentemente, maior a sensação de estéreo. Contudo sempre se deve obedecer aos limites para a paralaxe positiva, mencionados anteriormente.

APÊNDICE C

Teste de Ishihara

Antes de qualquer avaliação subjetiva dos vídeos, é necessário que os usuários façam o teste de cores de Ishihara, sendo ele um teste para detecção do daltonismo, já que pessoas que possuem essa deficiência podem não enxergar o efeito 3D [2] utilizando a técnica anaglífica, visto que pessoas daltônicas não conseguem diferenciar entre as cores verde e vermelho, e uma das lentes dos óculos anaglíficos é vermelha.

O exame consiste na exibição de uma série de cartões coloridos, cada um contendo vários círculos feitos de cores ligeiramente diferentes das cores daqueles situados nas proximidades. Seguindo o mesmo padrão, alguns círculos estão agrupados no meio do cartão de forma a exibir um número que somente será visível pelas pessoas que possuírem visão normal.

O grupo selecionado para avaliação foi submetido a esse teste e nenhum dos avaliadores foi considerado daltônico, o que otimizou o trabalho da avaliação. Das 20 pessoas selecionadas, todas foram aptas a avaliação, eliminando a necessidade de busca por novos avaliadores.

Referências Bibliográficas

[1] A. H. Sadka A. S. Umar, R. M. Swash. Subjective quality assessment of 3d videos. IEEE Transactions on Multimedia, September 2011.

[2] L. A. ANDRADE and R. GOULARTE. Percepção estereoscópica anaglífica em vídeos digitais comprimidos com perda. XV Simpósio Brasileiro de Sistemas Multimídia e Web - WebMedia, 2009, Fortaleza.

[3] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau. Using disparity for quality assessment of stereoscopic images. In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pages 389 –392, oct. 2008.

[4] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. von Kopylow. A survey of 3dtv displays: Techniques and technologies. IEEE Transactions on Circuits and Systems for Video Technology, 17(11):1647 –1658, nov. 2007.

[5] S.C. Chan, Heung-Yeung Shum, and King-To Ng. Image-based rendering and synthesis. IEEE Signal Processing Magazine, 24(6):22 –33, nov. 2007.

[6] StereoGraphics Corporation. Stereographics Developers' Handbook: Background on Creating Imagens for CrystalEyes and SimulEyes. http://www.stereographics.com/support/downloads_support/handbook.pdf, 1997.

[7] F. Cozman and E. Krotkov. Depth from scattering. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997., pages 801 –806, jun 1997.

[8] D.V.S.X. De Silva, W.A.C. Fernando, G. Nur, E. Ekmekcioglu, and S.T. Worrall. 3d video assessment with just noticeable difference in depth evaluation. In 17th IEEE International Conference on Image Processing (ICIP), pages 4013 –4016, sept. 2010.

[9] V. De Silva, A. Fernando, S. Worrall, H.K. Arachchi, and A. Kondoz. Sensitivity analysis of the human visual system for depth cues in stereoscopic 3-d displays. IEEE Transactions on Multimedia, 13(3):498 –506, june 2011.

[10] N. A. Dodgson. Autostereoscopic image compressiom. http://www.cl.cam.ac.uk/nad/compr/compr.html Acesso em: 10 mai. 2009., 1998.

[11] J. Ens and P. Lawrence. An investigation of methods for determining depth from focus. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(2):97 –108, feb 1993.

[12] Miguel Fragoso, Pedro Cruz, and Vasco Marcelino. Televisão 3D.

[13] Ge Guo, Nan Zhang, Longshe Huo, and Wen Gao. 2d to 3d convertion based on edge defocus and segmentation. In IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008., pages 2181 –2184, 31 2008-april 4 2008.

[14] P. Harman. Home based 3d entertainment-an overview. In Proceedings International Conference on Image Processing, 2000., volume 1, pages 1 –4 vol.1, 2000.

[15] C.T.E.R. Hewage, S.T. Worrall, S. Dogan, S. Villette, and A.M. Kondoz. Quality evaluation of color plus depth map-based stereoscopic video. IEEE Journal of Selected Topics in Signal Processing, 3(2):304 –318, april 2009.

[16] N.S. Holliman, N.A. Dodgson, G.E. Favalora, and L. Pockett. Three-dimensional displays: A review and applications analysis. IEEE Transactions on Broadcasting,, 57(2):362 –371, june 2011.

[17] Quan Huynh-Thu, P. Le Callet, and M. Barkowsky. Video quality assessment: From 2d to 3d – challenges and future trends. In 17th IEEE International Conference on Image Processing (ICIP), 2010, pages 4025 –4028, sept. 2010.

[18] Mathias Johanson. Stereoscopic video transmission over the internet. IEEE Workshop on Internet Applications, 0:0012, 2001.

[19] Ronald G. Kaptein, André Kuijsters, Marc T. M. Lambooij, Wijnand A. IJsselsteijn, and Ingrid Heynderickx. Performance evaluation of 3d-tv systems. Proc. SPIE 6808, 680819, 2008.

[20] J. Konrad and M. Halle. 3-d displays and signal processing. IEEE Signal Processing Magazine, 24(6):97 –111, nov. 2007.

[21] M.L. Kung, T.L. Alvarez, and J.L. Semmlow. Interaction of disparity and accommodative vergence. In Bioengineering Conference, 2003 IEEE 29th Annual, Proceedings of, pages 29 – 30, march 2003.

[22] O. Le Meur and P. Le Callet. What we see is most likely to be what matters: Visual attention and applications. In 16th IEEE International Conference on Image Processing (ICIP), 2009, pages 3085 –3088, nov. 2009.

[23] G. Leon, H. Kalva, and B. Furht. 3d video quality evaluation with depth quality variations. In 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, pages 301 –304, may 2008.

[24] Simon P. Liversedge, Nicolas S. Holliman, and Hazel I. Blythe. Binocular coordination in response to stereoscopic stimuli. Proc. SPIE 7237, 2009.

[25] A. Mancini. Disparity Estimation and Intermediate View Reconstruction for Novel Applications Stereoscopic Video. Thesis of mestre, Engineering McHill University, Montreal, Canadá, 1994.

[26] L.M.J. Meesters, W.A. IJsselsteijn, and P.J.H. Seuntiens. A survey of perceptual evaluations and requirements of three-dimensional tv. IEEE Transactions on Circuits and Systems for Video Technology, 14(3):381 – 391, march 2004.

[27] K. Mu andller, P. Merkle, and T. Wiegand. 3-d video representation using depth maps. Proceedings of the IEEE, 99(4):643 –656, april 2011.

[28] N. Ozbek and A.M. Tekalp. Unequal inter-view rate allocation using scalable stereo video coding and an objective stereo video quality measure. In IEEE International Conference on Multimedia and Expo, 2008, pages 1113 –1116, 23 2008-april 26 2008.

[29] J. I. Parente. A Estereoscopia no Brasil 1850-1930. Sextante, Rio de Janeiro, Brasil, 1999.

[30] Alex Paul Pentland. A new sense for depth of field. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-9(4):523 –531, july 1987.

[31] Ken Perlin, Chris Poultney, Joel S. Kollin, Daniel T. Kristjansson, and Salvatore Paxia. Recent advances in the

nyu autostereoscopic display. Proceedings of the SPIE, <http://www.mrl.nyu.edu/publications/autostereo/spie2001.p df>. Acesso em: 10 mai. 2009 2001.

[32] M.H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. IEEE Transactions on Broadcasting, 50(3):312 – 322, sept. 2004.

[33] M. Polonen, T. Jarvenpaa, and J. Hakkinen. Comparison of near-to-eye displays: Subjective experience and comfort. Journal of Display Technology, 6(1):27 –35, jan. 2010.

[34] O. Schreer, P. Kauff, and Sikora T. 3D Video Communication. Wiley, Berlin, Alemanha, 2005.

[35] Hang Shao, Xun Cao, and Guihua Er. Objective quality assessment of depth image based rendering in 3dtv system. In 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009, pages 1 –4, may 2009.

[36] W.J. Tam, Speranza, F. Yano, S., K. Shimono, and H. Ono. Stereoscopic 3d-tv: Visual comfort. IEEE Transactions on Broadcasting, 57(2):335 –346, june 2011.

[37] A. Tikanmaki, Gotchev, A., A. Smolic, and K. Miller. Quality assessment of 3d video in rate allocation experiments. In IEEE International Symposium on Consumer Electronics, (ISCE), pages 1 –4, april 2008.

[38] Romero Tori, Claudio Kirner, and Robson Siscoutto. Fundamentos e Tecnologia de Realidade Virtual e Aumentada. http://www.ckirner.com/realidadevirtual/?%26nbsp%3B_LIVROS_E_CAP%CDTULOS:Livro_de_RV_2006, 2006.

[39] A. Torralba and A. Oliva. Depth estimation from image structure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(9):1226 – 1238, sep 2002.

[40] Andrew Woods, Tom Docherty, and Rolf Koch. Image distortions in stereoscopic video systems. In STEREOSCOPIC DISPLAYS AND APPLICATIONS, 1993.

[41] S.L.P. Yasakethu, D.V.S.X. De Silva, W.A.C. Fernando, and A. Kondoz. Predicting sensation of depth in 3d video. Electronics Letters, 46(12):837 –839, 10 2010.

[42] S.L.P. Yasakethu, W.A.C. Fernando, B. Kamolrat, and A. Kondoz. Analyzing perceptual attributes of 3d video. IEEE Transactions on Consumer Electronics, 55(2):864 –872, may 2009.

[43] S.L.P. Yasakethu, C. Hewage, W. Fernando, and A. Kondoz. Quality analysis for 3d video using 2d video quality models. Consumer Electronics, IEEE Transactions on, 54(4):1969 –1976, november 2008.

[44] Liang Zhang, W.J. Tam, and D. Wang. Stereoscopic image generation based on depth images. In International Conference on Image Processing, 2004. ICIP '04, volume 5, pages 2993 – 2996 Vol. 5, oct. 2004.

[45] Liang Zhang, C. Vazquez, and S. Knorr. 3d-tv content creation: Automatic 2d-to-3d video conversion. IEEE Transactions on Broadcasting, 57(2):372 –383, june 2011.

[46] R. Zone. Stereoscopic cinema and the origins of 3-D film, 1838-1952. The University Press of Kentucky, Kentucky, USA, 2007.