4
Webcam Estereoscópica - uma nova abordagem

Baseado nas premissas apresentadas nos capítulos anteriores, vem a proposta de usar estereoscopia através de câmeras web como uma solução para adicionar características $3D$ às imagens da câmera, e assim enriquecer a visualização, sem recorrer a sistemas com processamento complexo e caro (ainda distantes da realidade dos usuários). Então, a principal restrição será não acrescentar hardwares especiais a um sistema de computador PC popular.

4.1
Uma nova abordagem

O uso de webcam estéreo para a visualização das pessoas, em tempo real, no contexto de interações remotas, compõe uma nova abordagem. Um par de imagens estéreo será obtido a partir de duas câmeras web, e projetado na tela do computador, em tempo real, para simular a percepção de profundidade semelhante à da visão binocular humana (obtida a partir de dois olhos).

Vantagens desta proposição:

- tecnologia simples;
- tempo real;
- preço acessível;
- percepção de profundidade;
- a indicação de que isto pode aumentar o senso de presença.

Como **desvantagens** pode-se enumerar:

- imagens de câmera web têm tamanho pequeno e baixa qualidade (baixa resolução);
- baixa taxa de transmissão (fps);
- dificuldades para obter estéreo com boa qualidade;
- o desconforto causado pela visualização estereoscópica em longas exposições;
- a necessidade de usar óculos para visualizar as imagens quebra a naturalidade na comunicação.

Os problemas de baixa qualidade da imagem e baixa taxa de transmissão estão se tornando menos críticos à medida que as tecnologias de câmeras web estão sendo melhoradas. As câmeras web mais modernas disponíveis no mercado têm uma qualidade muito melhor quando comparadas com aquelas de alguns anos atrás.

Os problemas relacionados ao desconforto e qualidade do estéreo são mais críticos, e o desconforto é ainda uma das maiores deficiências da visualização estereoscópica.

A seção seguinte expõe os pontos principais da abordagem computacional para um par de imagens estéreo.

4.2
Técnicas de visualização estereoscópica

Muitas técnicas de visualização estereoscópica têm sido desenvolvidas para explorar as características do sistema visual humano e daí obter a percepção tridimensional a partir de um par de imagens. A idéia principal está em simular a separação dos olhos humanos através do uso de duas câmeras, para permitir que o olho esquerdo veja apenas a imagem esquerda e o olho direito veja a imagem direita, ambas referindo-se à mesma cena, com uma pequena diferença de posicionamento entre as imagens, assim como ocorre nas imagens retinianas.

As técnicas podem ser classificadas de uma forma genérica como estereoscópicas e autoestereoscópicas e utilizam métodos diferentes para apresentar as imagens esquerda e direita aos olhos correspondentes.

As autoestereoscópicas permitem a visualização das imagens a olho nú, mas requerem telas especiais para isso, isto é, as técnicas para separar as imagens estão incorporadas no dispositivo de exibição, permitindo assim sua visualização sem o auxílio de óculos (59).

As técnicas estereoscópicas são as que necessitam de algum auxílio visual, como óculos, para separar as imagens esquerda e direita no olho; estas técnicas, por sua vez, dividem-se em dois grandes grupos: as que apresentam as imagens esquerda e direita simultaneamente, chamados métodos de *tempo-paralelo*, e os métodos que apresentam as imagens de cada câmera em seqüência usando técnicas óticas para esconder a visão do olho direito enquanto o olho esquerdo vê a cena, e vice-versa, chamados métodos de *campo-seqüencial* ou *tempo-multiplexado* (14). As técnicas de *tempo-multiplexado* utilizam um sistema de alternância das imagens sincronizado com a tela de exibição. Quando a tela opera a 100 Hz, ela efetivamente mostra ambas as imagens esquerda e direita a 50 Hz, considerando que as imagens são mostradas alternadamente.

Visto que não é o objetivo da tese apresentar um *survey* de técnicas de estereoscopia, os aspectos pertinentes a estas técnicas serão mencionados somente quando necessário. As técnicas estão bem descritas em (14), (2) e (43), por exemplo.

Apesar da produção de fotografia estereoscópica ter iniciado em torno de 1850 (14), a visualização estereoscópica de imagens naturais, objeto desta tese, no entanto, ainda não foi extensivamente estudada, principalmente no contexto de visualização em tempo-real.

Pela restrição colocada, a tese não irá recorrer a técnicas de estereoscopia que usam hardwares "especiais", isto é, que ainda não são de uso popular no momento. O processamento das imagens estéreo será portanto apenas baseado em software e toda a teoria estará focada para a produção de imagens estéreo naturais, mais especificamente vídeos, e usando a técnica de anaglifo que será descrita a seguir.

4.2.1 A técnica de anaglifo

Esta será a técnica usada para visualizar as imagens da aplicação proposta por ser a de menor custo financeiro, não exigindo qualquer dispositivo especial para sua visualização além de óculos de papel.

A técnica de *anaglifo* é a mais tradicional dos métodos de tempo-paralelo e requer o uso de um óculos de papel com filtros verde e vermelho ou ciano e vermelho para visualizar as imagens (ciano = verde + azul). Existem alguns anaglifos gerados usando os filtros vermelho e azul, ou outras cores complementares.

Os anaglifos são considerados não adequados para ver imagens coloridas, e são mais conhecidos para visualizar imagens em tons de cinza. Isso é devido ao uso inicial desta técnica com filtros verde e vermelho, o que gera perda de cores na fusão das imagens pela falta do canal de cor azul. Com o uso do filtro ciano no lugar do filtro verde, esse problema fica bastante reduzido, permitindo uma visualização mais próxima das cores reais da imagem; portanto, vermelho e ciano serão os filtros adotados neste trabalho.

A imagem da câmera esquerda é projetada usando um filtro vermelho e a da câmera direita usando um filtro ciano. O observador olha para a imagem mista usando um filtro ciano no olho direito e um filtro vermelho no olho esquerdo, e assim cada imagem colorida complementar é eliminada pela filtragem. Isto resulta em uma visualização estereoscópica, onde cada parte do par estéreo é vista apenas pelo olho correspondente. Uma das vantagens deste método é que anaglifos podem ser produzidos em um PC usando software de

processamento de imagem e podem ser reproduzidos em todo tipo de mídia, como impressos e via Internet.

Como desvantagens tem-se uma imagem estéreo com tons de cor apenas aproximados do real, no anaglifo colorido, e a presença de *Crosstalk*, (ou *ghost*), que é o vazamento de uma imagem para um olho, quando ela deveria ser vista apenas pelo outro olho, gerando um efeito fantasma na imagem estéreo resultante. O crosstalk reduz a qualidade da imagem e dificulta a fusão das mesmas, quando o efeito é muito acentuado.

As possíveis fontes de crosstalk nas imagens de anaglifo são (63):

- **resposta espectral do dispositivo de exibição** (CRT, LDC, etc) - normalmente estes dispositivos funcionam emitindo luz em três faixas de cores primárias específicas (vermelho, verde e azul - RGB). O espectro real de cada faixa de luz pode variar consideravelmente entre diferentes tipos de display.
- **resposta espectral dos óculos de anaglifo** - se o filtro dos óculos deixar passar luz no domínio indesejado do espectro, a imagem também mostrará ghost.
- **compressão da imagem** - alguns formatos de compressão podem misturar informação entre os canais de cores e então introduzir ghost na imagem. A quantidade de crosstalk introduzida dependerá da quantidade de compressão usada, do tipo de compressão e, algumas vezes, do método específico de codificação usado para um determinado tipo de compressão.
- **codificação da imagem e transmissão** - diferentes formatos de vídeo codificam informações de cor de forma diferente.

O trabalho apresentado em (63) examina especificamente o efeito crosstalk em anaglifos gerados com filtros vermelho e ciano, analisando os dois primeiros pontos citados - resposta espectral do dispositivo de exibição e resposta espectral do óculos de anaglifo - e conclui que a quantidade de ghosting exibida em um anaglifo depende principalmente do tipo de óculos usado, no caso, da combinação específica dos componentes do filtro, e do dispositivo de exibição usado.

Outro aspecto que também interfere na percepção de ghost na imagem é a incidência de luz ambiente sobre a tela de visualização. O processo de eliminação de ghost envolve grandes variáveis para o problema, ficando este estudo fora do escopo da tese, constituindo uma área de estudos dentro deste tipo de visualização.

4.3 Geometria do modelo convencional

Seis variáveis basicamente caracterizam a geometria de um sistema estereoscópico (62).

A configuração do sistema de câmera é determinada por:

1. distância entre as câmeras (distância interaxial), também chamada de *separação interaxial* e se refere ao eixo das lentes;

2. distância de convergência (distância das câmeras até o ponto em que os eixos óticos se interceptam);

3. o campo de visão das câmeras (FOV);

A configuração do dispositivo de exibição é determinada por:

4. distância de visualização do observador para a tela;

5. tamanho da tela de exibição (medido pela sua dimensão horizontal);

6. distância entre os olhos do observador (distância interocular)

Segundo (14), a profundidade percebida de um ponto em um par de imagens estéreo depende primariamente do *paralaxe horizontal*. Paralaxe horizontal é a distância entre os pontos similares de cada imagem, esquerda e direita, de uma cena projetada em um plano perpendicular à linha de visão do observador, chamado de *janela de estéreo* ou *plano de estéreo*.

> "Paralaxe e disparidade são entidades similares. Paralaxe é medido na tela de exibição, e disparidade é medida na retina. Ao usar óculos estereoscópicos, paralaxe se torna disparidade retiniana. É o paralaxe que produz a disparidade retiniana, e disparidade por sua vez produz visão estéreo. O paralaxe pode também ser dado em termos de medida angular, desta forma relacionando-o com a disparidade ao levar em conta a distância do observador à tela de exibição" (12).

Se as projeções de um ponto no olho esquerdo e direito estão separadas por P centímetros, e o observador está distante d centímetros da tela, então o ângulo de paralaxe α, é dado por (12)

$$\alpha = 2 \arctan \frac{P}{2d},$$

se α exceder $1,5$ graus, tende a causar desconforto.

O paralaxe positivo ocorre se o ponto parece estar atrás da janela de estéreo; o paralaxe é dito negativo quando o ponto é percebido na frente da janela de estéreo e paralaxe zero ocorre se o ponto está na mesma profundidade da janela de estéreo, neste caso, sendo esta janela o plano de disparidade zero (ZDP), onde as imagens coincidem (Figura 4.1). Normalmente a janela de estéreo é posicionada no plano da tela de visualização (14).

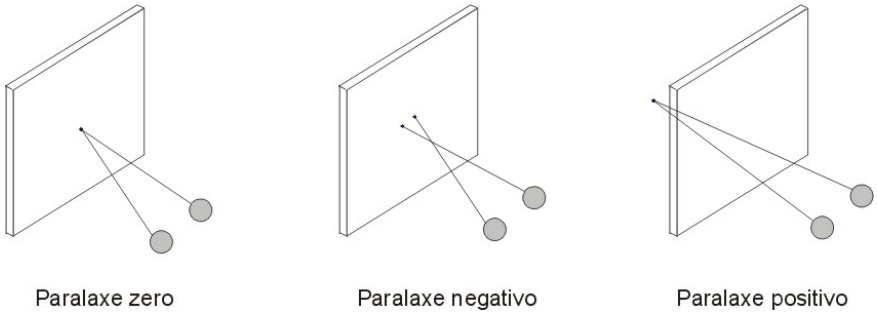

Figura 4.1: Paralaxe na tela de visualização

4.3.1 Transformação de coordenadas

A geometria de um sistema de vídeo estereoscópico, descrita em detalhes nos trabalhos de Woods, Dicherty e Koch (62) e C. H. Yang (65), dentre outros, pode ser determinada por três transformações de coordenadas, ilustradas na Figura 4.2:

1. das coordenadas X_m, Y_m, Z_m do objeto no mundo real para a posição X_s Y_s nos dois sensores de imagens das câmeras;
2. das coordenadas dos sensores para coordenadas X_t e Y_t das imagens esquerda e direita na tela;
3. para coordenadas X_o, Y_o, Z_o percebidas pelo observador.

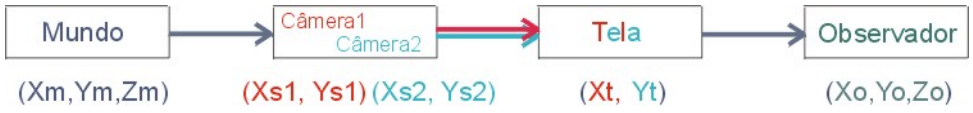

Figura 4.2: Transformações de coordenadas

Estas coordenadas serão descritas a seguir, para os modelos de câmeras paralelas e convergentes, respectivamente.

O modelo de câmeras paralelas

Este modelo computacional assume que as câmeras - como "olhos" - estão olhando para o infinito, i. e., que os eixos através das lentes das câmeras estão paralelos, tendo uma distância interaxial entre elas, como mostrado na Figura 4.3.

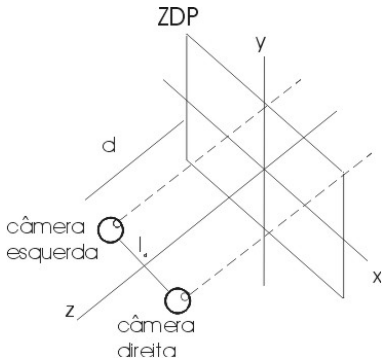

Figura 4.3: Modelo de câmeras paralelas para o estéreo computacional

A convergência das imagens é obtida, no entanto, pelo deslocamento dos sensores das câmeras, variando-se a distância interaxial, ou pela translação horizontal das imagens. Este deslocamento das câmeras ou a translação das duas visões perspectivas ao longo do eixo-x, que é também chamada de translação horizontal de imagem (HIT) (12), muda a posição do plano de disparidade zero (ZDP), onde as imagens coincidem; esta translação horizontal das imagens esquerda e direita controla o paralaxe (Figura 4.4).

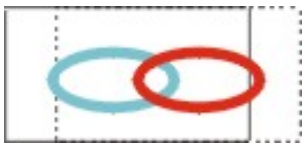

Figura 4.4: Translação horizontal das imagens esquerda e direita

Muitas vezes, esta translação, em busca de um paralaxe adequado para a visualização estereoscópica, faz com que a borda de uma imagem ultrapasse a outra, mostrando uma faixa de imagem com apenas um dos filtros, i.e., a imagem de apenas uma das câmeras. Para solucionar o problema aplica-se a operação de *cropping* sobre a imagem.

É importante observar que todas as operações efetuadas, sejam: translação, projeção, cropping, são efetuadas diretamente sobre o objeto gráfico imagem.

Neste modelo de câmeras, a disparidade horizontal como função da profundidade Z, é obtida pela subtração da coordenada X de projeção da câmera esquerda, x_l, da coordenada X de projeção da câmera direita , x_r, (59).

$$d_{h,p}(Z) = x_r(X, Z) - x_l(X, Z),$$

e como função da separação interaxial das câmeras, I, e de Z, é dada por:

$$d_{h,p}(Z) = \lambda \frac{-I_c}{\lambda - Z},$$

sendo λ é o comprimento focal da câmera.

No *StereoGraphics Developers' Handbook* (12) encontra-se uma discussão detalhada sobre as considerações geométricas para o modelo de câmeras paralelas, discussões dos limites para paralaxe mínimo e máximo, e variação da distância interaxial das câmeras.

O modelo de câmeras convergentes

No modelo de câmeras convergentes, os eixos óticos de ambas as câmeras se interceptam em um ponto de convergência, escalado pela distância interaxial das câmeras, I, e pelo ângulo de convergência, α, tal que a distância Z ao ponto de convergência é dada por (59):

$$Z = \frac{I}{2 \tan \alpha}$$

A Figura 4.5 mostra o modelo de câmeras convergentes.

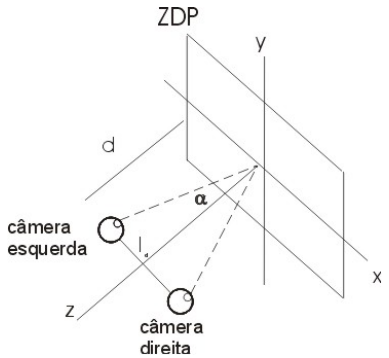

Figura 4.5: Modelo de câmeras convergentes

Como os eixos das câmeras não são mais paralelos, a transformação de um ponto em coordenadas do mundo para as coordenadas dos sensores das

câmeras requer uma translação horizontal ao longo do eixo-x e uma rotação em torno do eixo-y, seguidas de uma projeção perspectiva.

Em aplicações de realidade virtual, por exemplo, o modelo não inclui qualquer ângulo de convergência. A simulação desta convergência é feita pelo HMD (*Head-Mounted Display*), empregando o princípio ótico de divergência ou convergência (60).

Padrão proposto para controle do paralaxe

Na literatura de estereoscopia, a manipulação de teclas é um padrão proposto para auxiliar os usuários de software de estereoscopia a controlar as duas variáveis consideradas fundamentais para a visualização das imagens estéreo, distância interaxial (DI) e translação das imagens (HIT). As teclas mostradas na Figura 4.6, (N - norte, S - sul, L - leste e O - oeste), são as sugeridas para controlar estes parâmetros.

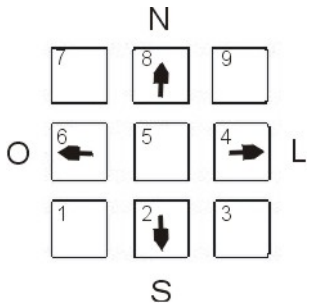

Figura 4.6: Teclas para controle dos parâmetros de estéreo (12)

As setas verticais manipulam a distância interaxial das câmeras; a tecla N aumenta a distância entre os eixos das câmeras, e S diminui esta distância. Para controlar HIT existem duas formas possíveis: afastar as imagens ou aproxima-las através das teclas L e O. Esta variação na translação vai fazer com que o ZDP seja posicionado em diferentes planos. Este parâmetro é melhor controlado interativamente sem o uso de óculos porque a posição relativa entre as imagens é mais facilmente observada olhando-se para a superposição das imagens esquerda e direita (14), (12).

É importante observar que a variação da distância interaxial das câmeras é viável para câmeras virtuais. No caso de câmeras reais, webcams especificamente, as teclas N e S poderiam ser usadas para manipular o ajuste de paralaxe vertical das imagens, por exemplo. Por outro lado, o fato de analisar a translação das imagens ser melhor sem óculos, dificulta para a visualização de imagens interativas e em tempo real, sobretudo enfatizando que o usuário leigo não terá conhecimento dos conceitos de paralaxe e ZDP.

A maioria dos algoritmos de estéreo ou experimentos é baseada na variação da distância entre as câmeras e o paralaxe das imagens, para determinar o ponto ideal para a visualização, como exemplos tem-se (59), (32).

Os autores em (32) mencionam os principais problemas do estéreo computacional e afirmam: "boas imagens são difíceis de obter e a **variável chave** que deve ser determinada para a criação de qualquer imagem estereoscópica é a separação das câmeras, já que isto afeta diretamente a quantidade de profundidade percebida na imagem resultante". A solução apresentada por eles varia a separação das câmeras, para cada vez que o observador se move, para ajustar a cena em relação à distribuição de profundidade e distorção. Isto é muito aplicável para câmeras virtuais, que podem ser totalmente controladas pelo programa dinamicamente, ou na abordagem de fotografia.

A pesquisa apresentada em (59) fez uma análise precisa com experimentos para encontrar um intervalo de percepção de profundidade ideal, e a distância interaxial das câmeras foi a principal variável.

4.3.2 Distorções geradas pelo estéreo

> "Distorções estereoscópicas são formas nas quais uma imagem estereoscópica difere da visão real da cena"(62)

As distorções estereoscópicas podem ser causadas pela geometria do modelo, neste caso pode ser na primeira transformação de coordenadas (do mundo para as câmeras) ou na segunda transformação (das câmeras para a tela de exibição), por características do display (no caso o monitor ou tela de exibição), e por último pela inadequação entre a exibição das imagens (limitações dos displays, etc) e o processo visual humano, ocorrendo na terceira transformação de coordenadas, do display para o espaço de visualização do observador.

O trabalho clássico de Woods et al. (62), um dos mais citados da literatura, apresenta as principais distorções em sistemas de vídeo estereoscópico, descritas a seguir.

1. Curvatura do plano de profundidade

   Este trabalho (62) desenvolveu um programa de computador para gerar e exibir, em diagrama, a transformação de coordenada do espaço do mundo (do objeto) para o espaço da imagem. O diagrama mostra a forma como o espaço do objeto, em frente ao sistema de câmeras, é transformado para a tela (espaço da imagem). Os dois círculos na Figura 4.7 representam os olhos do observador e a linha em negrito é a tela de visualização.

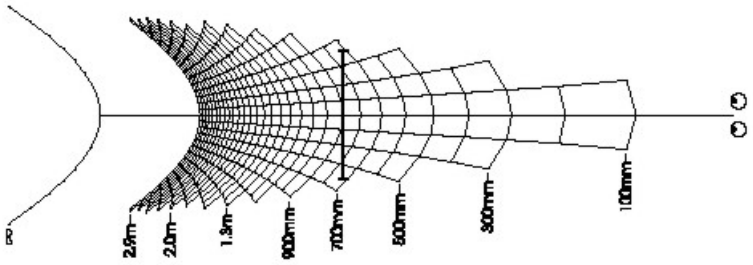

Figura 4.7: Curvatura do plano de profundidade (62)

A configuração de câmeras convergentes resulta na curvatura dos planos de profundidade, mostrado na Figura 4.7. Isto faz com que os objetos nas bordas da imagem pareçam mais afastados do observador do que objetos no centro da imagem. A configuração de câmeras paralelas resulta em planos de profundidade paralelos à superfície da tela. A curvatura do plano de profundidade está proximamente relacionada com a distorção de *keystone* que será descrita.

2. Cisalhamento

   Como mostra a Figura 4.8, um deslocamento lateral do observador resulta em um cisalhamento da imagem estereoscópica em relação à superfície da tela do computador. Imagens que saltam da tela (paralaxe negativo) parecem cisalhar na direção do movimento do observador e imagens atrás da superfície da tela cisalham na direção oposta. Esta distorção resulta na percepção errada de distância relativa entre os objetos da cena.

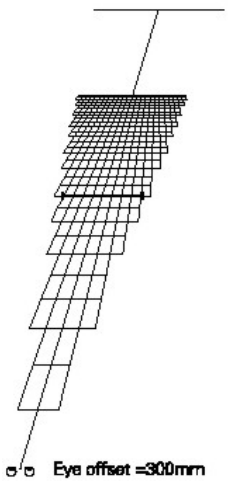

Figura 4.8: Cisalhamento - câmeras paralelas (62)

A Figura 4.8 mostra os planos de profundidade para a configuração de câmeras paralelas (62).

3. *Keystone*

   Um efeito conhecido da configuração de câmeras convergentes é a distorção de keystone. Esta distorção gera paralaxe vertical na imagem estereoscópica devido à rotação entre as câmeras e assim as imagens dos sensores das câmeras caem em planos diferentes. O efeito do keystone é mostrado na Figura 4.9.

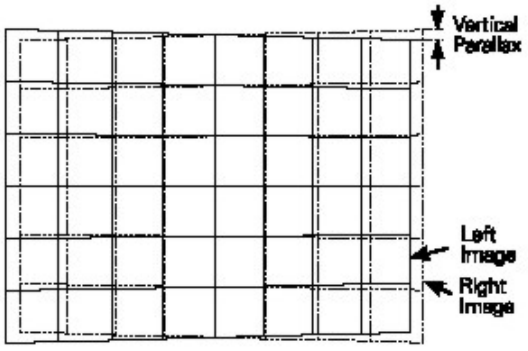

Figura 4.9: Paralaxe vertical causado por keystone (62)

   Em uma das câmeras a imagem parece mais larga de um lado do que do outro. Na outra imagem o efeito é contrário. Isto resulta em diferença vertical entre mesmos pontos, chamado de paralaxe vertical. A quantidade de paralaxe vertical é maior nas bordas da imagem, e aumenta com a separação das câmeras e com menor distância de convergência. A configuração de câmeras paralelas não gera distorção de keystone.

4. Relacionamento Acomodação/convergência

   Esta é uma limitação relacionada com os dispositivos de exibição e o olho humano. Woods et al. (62) dizem que, por experiência, o paralaxe excessivo na tela pode levar à percepção das imagens fora de foco e/ou que o observador não seja capaz de fazer a fusão, e isto pode também estar associado ao relacionamento de acomodação e convergência.

4.3.3
Restrições do modelo convencional

De acordo com (12), existem algumas restrições para aplicações que permitem o usuário mudar a configuração do observador. Estas são:

- se uma translação ocorre nos eixos x ou y, os dois centros de projeção (olhos) precisam permanecer centrados em relação a um ponto perpendicular ao centro da tela;
- se uma translação ocorre no eixo z, a coordenada z do ZDP (Zero Disparity Plane) precisa ser mudada;
- se qualquer escala ocorre, ou quaisquer operações que mudem os intervalos em x, y ou z, então o ZDP, assim como a distância das câmeras ao ZDP, d, e a separação interaxial precisam ser mudados.

Estas restrições estabelecem muitas limitações para a obtenção de imagens estéreo a partir de imagens naturais, em tempo real, quando o observador é uma pessoa que está na frente do computador com a expectativa de mover-se livremente.

Em uma aplicação em tempo real, como o uso de webcams, não é possível ficar ajustando a separação das câmeras a cada vez que o observador se move, como acontece com o uso de câmeras virtuais; da mesma forma, não é prático reajustar o paralaxe das imagens, manualmente, assim como também não há a possibilidade de restringir o posicionamento do usuário. Por exemplo, seria fisicamente impossível posicionar as webcams com uma distância interaxial de 2,78cm, como o resultado obtido em (32).

Na proposta aqui apresentada, o estéreo é gerado e visualizado em curta distância, portanto, existem considerações específicas a serem feitas, tanto na geometria quanto no processamento das imagens.

Seguindo a abordagem da área de Presença, isto conduz a buscar apoio em estudos psicofísicos, fisiológicos e perceptuais relacionados à visão humana, especificamente à visão estéreo, para daí obter critérios importantes a serem considerados na proposição e implementação de uma solução viável.

A tese defende que a inclusão de propriedades da visão binocular e de estereopsia no processamento computacional, assim como resultados de análises psicofísicas existentes na literatura do olho humano, podem conduzir a resultados melhores no cálculo do estéreo computacional, e assim eliminar algumas restrições clássicas como as citadas acima.

Para isso, enunciam-se as seguintes hipóteses da tese:

HIPÓTESES:

- $H1$ - Uma geometria adequada para as câmeras web pode permitir a visualização estereoscópica de imagens naturais, em tempo real, sem ajuste contínuo de distância interaxial das câmeras.
- $H2$ - A inclusão de parâmetros inerentes à visão binocular e estereopsia no processamento computacional da visualização estereoscópica de imagens naturais pode melhorar a qualidade da imagem estéreo.
- $H3$ - As hipóteses $H1$ e $H2$ implementadas juntas podem vir a influenciar na redução do desconforto na visualização estereoscópica de imagens naturais.

Com base nisto, esta pesquisa irá fazer considerações a partir dos aspectos psicofísicos e perceptuais do olho humano, da estereopsia, e da visão binocular com o intuito de obter melhores resultados para a visualização estereoscópica, a partir de imagens capturadas por duas câmeras web simples, sincronizadas em tempo real, almejando a redução dos problemas que ocorrem nesta técnica de visualização.