UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Métodos de Reversão de Vídeos Anaglíficos em Pares Estéreis de Vídeo

Aluno: Matheus Ricardo Uihara Zingarelli
(zinga@icmc.usp.br)

Orientador: Prof. Dr. Rudinei Goularte
(rudinei@icmc.usp.br)

Projeto de mestrado submetido à FAPESP

USP – São Carlos
Agosto de 2010

Resumo

A atenção voltada à produção de conteúdos 3D tem sido grande atualmente, em grande parte devido à aceitação e manifestação de interesse do público para esta tecnologia. Isso reflete num maior investimento das indústrias cinematográfica, de televisores e de games em trazer o 3D para suas produções e aparelhos, oferecendo formas diferentes de interação ao usuário. Com isso, novas técnicas de captação e codificação e modos de reprodução de vídeos 3D, denominados vídeos estereoscópicos, vêm surgindo ou sendo melhorados, visando otimizar e integrar esta nova tecnologia com a infraestrutura disponível. Entretanto, nos avanços feitos no campo da codificação, nota-se a ausência de um padrão compatível com qualquer método de visualização de vídeos estereoscópicos, sendo que para cada um há uma técnica de codificação diferente que pode causar perdas significativas se aplicada ao outro. Uma proposta é criar uma técnica de codificação que seja genérica, ou seja, que através de parâmetros adequados, se obtenha um vídeo sem nenhuma perda tanto na qualidade quanto na percepção de profundidade. Visando uma maior compressão, nesta técnica o vídeo é transformado em anaglífico. No entanto, para que ela permaneça genérica, é necessário possuir também o processo reverso, tornando o vídeo anaglífico novamente em um par estéreo. Esse processo reverso não é trivial e requer um estudo de como recuperar as informações perdidas durante a conversão do vídeo para seu formato anaglífico.

1. Introdução

Grande parte da população hoje em dia faz uso de variados tipos de mídia como forma obter informações e até mesmo interagir socialmente. Com a evolução da Web e o avanço da banda larga pode-se notar o surgimento de novos serviços como Youtube e Twitter, os quais alcançaram grande repercussão, mostrando a demanda dos usuários por novos modos de interação e visualização de informações. Dentre esses serviços encontra-se o retorno do vídeo 3D aos cinemas – com novidades tecnológicas – e os novos televisores 3D (Hutchison, 2008). Esse fato tem incentivado indústria e academia a pesquisar e desenvolver métodos e técnicas que promovam a produção e distribuição desse tipo de vídeo.

Em termos técnicos, os vídeos 3D são definidos como vídeos estereoscópicos e utilizam métodos chamados estereoscópicos, os quais consistem em apresentar duas imagens bidimensionais especiais – um par estéreo – para serem interpretadas pelo cérebro humano na criação de uma imagem única e tridimensional, provocando a sensação de profundidade e distanciamento. Tais métodos visam simular o efeito obtido na visão humana pelo fato de nossos olhos estarem distantes horizontalmente um do outro, o que faz com que cada olho tenha um ponto de vista diferente, diferença essa chamada de disparidade binocular (Azevedo & Conci, 2003).

A tecnologia 3D não é novidade, sendo que a produção de vídeos estereoscópicos já sofreu vários avanços tanto na forma de captação quanto na forma de reprodução. Câmeras especiais foram desenvolvidas visando capturar dois

pontos de vista diferentes de uma mesma imagem (gerando o par estéreo), ou então gerando um mapa de profundidade das cenas juntamente com o vídeo (Fehn et al, 2002). Também foram desenvolvidas técnicas para conversão e apresentação de vídeos 3D a partir de vídeos originalmente em 2D (Tam & Zhang, 2006). No que diz respeito à reprodução de vídeos 3D, existem tecnologias que fazem uso de óculos especiais para separar o par estéreo, direcionando a imagem correta para cada olho (Stereographics, 1997), bem como monitores denominados autoestereoscópicos, os quais permitem assistir a conteúdo 3D sem o auxílio de óculos (Fehn et al, 2006).

Apesar do recente impulso que a tecnologia 3D vem recebendo da indústria do cinema (Mendiburu, 2009; Suppia, 2007) e da televisão (Sony Corporation, 2010), ainda existe necessidade de mais pesquisa na área. Um reflexo disso é a atual falta de padronização no modo de organizar os dados estéreo para fins de armazenamento ou transmissão, sendo que tais métodos podem ser divididos em dois grandes grupos: o método de Lipton (Lipton, 1997) e os métodos aqui chamados de vinculados (Smolic et al, 2009). No método de Lipton o par estéreo é armazenado em *containers* (AVI, por exemplo), com compressão ou não. Apesar de ser mais flexível que os métodos vinculados, resulta em um volume de dados duas vezes maior, devido à necessidade de se armazenar dois *streams* de vídeo (o par estéreo).

Os métodos vinculados, por sua vez, utilizam técnicas consagradas de compressão de vídeo (como MPEG-2 e H.264) para diminuir o volume de dados e atender às demandas de armazenamento/transmissão. Contudo, tais técnicas são adaptadas para tratar vídeo 3D e funcionam apenas para casos particulares (Smolic et al, 2009), além disso, usam compressão com perdas, o que pode impossibilitar a correta percepção de profundidade em alguns casos, notadamente em vídeo anaglíficos (Andrade & Goularte, 2009). Como resultado, não existe uma técnica única de codificação para vídeo 3D que atenda a todos os atuais métodos de visualização 3D: os que necessitam de óculos especiais (anaglífico, com lentes polarizadoras e obturadores) e o autoestereoscópico.

Em um projeto de doutorado relacionado (Andrade & Goularte, 2009; Andrade & Goularte, 2010) o grupo de pesquisa está desenvolvendo uma técnica de codificação para vídeos que realiza compressão sem comprometer a qualidade de percepção de profundidade (percepção do 3D) e que, ao mesmo tempo atende a todos os atuais métodos de visualização 3D. A técnica é baseada na subamostragem de crominância em níveis adequados, aliada à aplicação de transformadas *Wavelet* com adequada parametrização do nível de quantização e de preenchimento de imagens (para imagens cujas dimensões não são múltiplo de 8). Como resultado obtém-se um par de vídeos estéreo com boa compressão, ainda abaixo dos níveis dos codificadores H.264 (referência em taxa de compressão), mas com qualidade superior em termos de percepção de profundidade. Ainda, como se tem o par de vídeos estéreo, qualquer método de visualização pode ser aplicado.

Uma melhoria a essa técnica é alcançar compressão adicional transformando-se o par estéreo em um único - *stream* com metade do volume de dados, utilizando para isso o método anaglífico. Desse modo, o vídeo em formato

anaglífico poderia ser utilizado para fins de armazenamento/transmissão (pois possuiria boa taxa de compressão) e a técnica atenderia ao método anaglífico de visualização (com diferencial em qualidade). Contudo, para que a técnica continue sendo genérica, é necessário restaurar o par estéreo de vídeos a partir do vídeo anaglífico gerado, possibilitando que os outros métodos de visualização sejam empregados (por luz polarizada, com óculos obturadores ou o método autoestereoscópico). Assim, o objetivo deste projeto é desenvolver técnicas de reversão de vídeo anaglífico em par estéreo.

Este texto está organizado da seguinte forma: a seção 2 trata da revisão da bibliografia acerca de vídeos estereoscópicos bem como de aspectos biológicos da visão humana relacionados. Na seção 3 são apresentados o objetivo deste trabalho e a forma como este será realizado, junto com o cronograma proposto para a realização das atividades. Por fim, a seção 4 possui todas as referências utilizadas na produção deste projeto.

2. Síntese Bibliográfica

Nesta seção são apresentados os conceitos básicos e principais tecnologias e técnicas envolvidas na área de vídeos estereoscópicos. Aborda-se também brevemente o estado da arte das pesquisas relacionadas a esse assunto.

2.1. Aspectos da visão humana

Nossos olhos estão distantes aproximadamente 6,5cm um do outro, movimentam-se em conjunto para uma mesma direção e cada um possui um ângulo de visão limitado. Por se apresentarem em posições diferentes, cada olho observa uma imagem ligeiramente diferente um do outro, característica classificada como disparidade binocular (Azevedo & Conci, 2003). Por essas razões era de se esperar que, ao olharmos para um objeto, víssemos duas imagens e não apenas uma. Além disso, dentre os vários objetos presentes no nosso campo de visão, temos a capacidade de interpretar diferentes profundidades e texturas entre eles, e tal capacidade permanece mesmo se nos movermos para um lado ou para outro. Essa utilização de ambos os olhos para formar uma única imagem, com diferentes níveis de profundidade entre os objetos nela presentes, é definido como estereopsia.

O principal personagem envolvido nesses fenômenos é o nosso cérebro. Entretanto, ainda não é totalmente conhecido o processo que este realiza. Mesmo assim, alguns conceitos físicos e biológicos da visão humana nos ajudam a compreender melhor as tarefas envolvidas.

De acordo com (Stuart, 1996), existem três fatores visuais envolvidos no processo de transformação tridimensional de uma imagem pelo cérebro: informações monoculares, informações oculo-motoras e informações estereoscópicas.

2.1.1. Informações monoculares

As informações monoculares, do inglês *static depth cues*, são as obtidas através das imagens formadas na retina do olho. A maioria delas são amplamente exploradas pelos artistas em técnicas de pintura e podem ser divididas em: perspectiva linear, interposição, luz e sombra, perspectiva aérea, variação da densidade de textura, conhecimento prévio do objeto e paralaxe de movimento (Stereographics, 1997).

A informação da perspectiva linear está ligada à sensação que temos de que o tamanho dos objetos diminui à medida que estes se afastam de nós, valendo o mesmo para o processo inverso. Um exemplo clássico é a sensação que temos que a distância entre as linhas paralelas de uma estrada diminui até convergir no horizonte. A perspectiva é uma das principais técnicas utilizadas para expressar a noção de profundidade no papel, e foi uma das grandes descobertas no campo das Artes, sendo amplamente utilizada pelos pintores renascentistas (Azevedo & Conci, 2003).

A interposição é um conceito simples que nos dá a informação da posição relativa entre objetos. Dado que um objeto A oculta parte ou o todo de B, entendemos que A está à frente de B e mais próximo de nós. Junto com a interposição, a variação de luz incidente sobre um objeto, bem como a utilização de sombras, nos dão informações importantes sobre as características deste, tais como o volume de espaço que ele preenche, sua curvatura, sua posição em relação a outros objetos, sua solidez, transparência e textura.

A perspectiva aérea é a percepção que temos de que objetos cuja visibilidade é atrapalhada por algum fenômeno atmosférico (neblina, chuva, incidência solar) se encontram mais distantes. Por exemplo, ao olhar para uma cadeia de montanhas, nota-se que as que se encontram mais distantes aparecem menos nítidas, como se estivessem desaparecendo. Tais fenômenos atmosféricos podem enganar o cérebro e fazer com que uma imagem pareça estar mais distante do que realmente está.

A variação na densidade de uma textura também nos fornece informações sobre a distância de que um objeto se encontra, dada pelo nível de detalhamento que obtemos. Quanto mais distante se encontra um objeto, menos detalhes são vistos de sua textura. Por exemplo, ao olharmos para uma árvore, à medida que nos distanciamos dela, perdemos os pequenos detalhes de suas folhas e seu tronco.

Através do conhecimento prévio nosso cérebro vai armazenando informações dos objetos ao passo que vamos tendo contato com eles no mundo real, criando padrões de tamanho e profundidade destes em comparação a outros e ao ambiente em que se encontram. Com isso, ao vermos tais objetos em uma mesma imagem, de acordo com nossas experiências e conhecimento prévio, conseguimos inferir qual está mais próximo ou mais afastado, qual é maior ou menor.

A paralaxe de movimento, como o próprio nome sugere, é uma informação resultante de movimento que nos fornece a distância entre objetos. Observamos este fenômeno quando, por exemplo, dentro de um carro em movimento vemos objetos que se encontram mais próximos (uma cerca, por exemplo) parecendo se mover mais rápido do que objetos que se encontram mais distantes (árvores no horizonte).

2.1.2. Informações oculo-motoras

Diferente das informações monoculares, que podem ser reproduzidas em imagens no papel, as oculo-motoras são baseadas em aspectos fisiológicos. Elas são produzidas de acordo com o relaxamento e contração dos músculos envolvidos no movimento do globo ocular e são interpretadas pelo cérebro para relacionar a distância e profundidade entre objetos. Temos dois tipos: a acomodação e a convergência (Azevedo & Conci, 2003) (Stereographics, 1997).

A acomodação é relacionada às contrações musculares envolvidas para mudar o formato do cristalino, com o objetivo de alterar o foco nas imagens. Consegue-se obter informação sobre a distância entre objetos, de acordo com o esforço muscular envolvido para alterar o foco.

Cada olho produz uma imagem diferente do que está sendo visto, porém, conseguimos fazer com que um objeto seja visto na mesma posição em ambos os olhos se focarmos nele. Para isso, ele deve se encontrar em um mesmo ponto para os dois olhos, chamado de ponto de convergência. De acordo com a distância em que se encontra o objeto, devemos alterar nosso ponto de convergência. O ângulo formado na movimentação dos olhos em torno do seu eixo vertical para esse ponto de convergência nos dá a informação da distância do objeto.

2.1.3. Informações estereoscópicas

Como anteriormente exposto cada olho produz uma imagem diferente, devido ao fato de estarem a uma distância e ângulos diferentes (disparidade binocular). Cabe ao cérebro se encarregar de retirar as informações das distâncias relativas dos objetos e de interpretar essas duas imagens resultando na produção de uma única, fenômeno descrito como estereoscopia. Das informações estereoscópicas, as principais são a estereopsia, disparidade e paralaxe (Stereographics, 1997).

Já citada anteriormente, a estereopsia é a responsável pela sensação que temos de profundidade entre os objetos, e é obtida em virtude da disparidade binocular. Dessa forma, o requisito obrigatório para obtermos estereopsia é utilizarmos os dois olhos. É com esta informação, em cooperação com as outras informações aqui descritas, que sentimos objetos mais próximos ou mais distantes. É ela a explorada em filmes 3D para nos passar a impressão de que objetos estão saltando para fora da tela.

A diferença na distância entre as posições da imagem formada em cada retina em relação ao centro desta é chamada de disparidade. Isso é melhor entendido através do seguinte exemplo: observe um objeto a sua frente e posicione o seu polegar entre seus olhos e o objeto. Quando focalizamos no polegar, ou seja, ele se encontra no ponto de convergência das duas retinas, o objeto fica após o ponto de convergência (mais distante), aparecendo como que duplicado (Figura 1 A). Isso se dá pelo fato de as imagens fora do ponto de convergência serem formadas em posições diferentes em cada retina. A disparidade é a distância entre essas duas imagens duplicadas. O mesmo acontece se colocamos o nosso foco no objeto (Figura 1 B).

Diretamente ligado ao conceito de disparidade (obtida na imagem formada na retina) temos a paralaxe, que é a distância entre os pontos correspondentes nas imagens formadas em um monitor para cada olho. Com os valores de paralaxe, é possível dar um ponto de vista diferente de uma mesma imagem para cada olho, tendo como consequência a formação da disparidade, e esta, por conseguinte, produz o efeito de estereopsia. Uma maneira fácil de calcular a paralaxe entre dois pontos é sobrepondo uma imagem à outra e medindo a distância entre os mesmos pontos em cada imagem. É por causa da paralaxe que, por exemplo, ao assistirmos um vídeo anaglífico sem os óculos vemos partes da imagem como que duplicadas e sobrepostas.

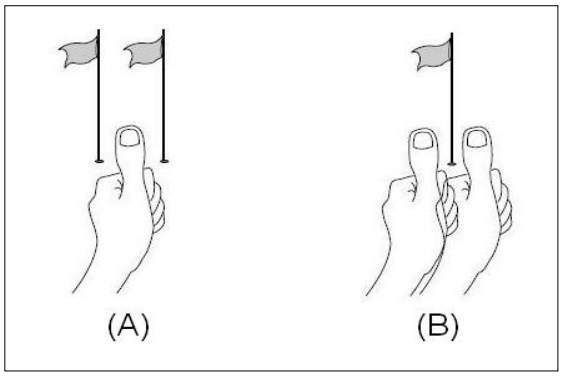

Figura 1 – Exemplo de observância da informação de disparidade. Em (A), quando focamos nossa visão no dedo polegar, a bandeira aparece duplicada ao fundo. Em (B), quando focamos nosso olhar na bandeira, o dedo polegar aparece duplicado. A distância entre as imagens duplicadas é o valor da disparidade (Stereographics, 1997).

Podemos classificar a paralaxe em quatro tipos (Stereographics, 1997), os quais afetam a nossa noção de profundidade acerca dos objetos que compõem a imagem: a paralaxe zero (ZPS - *Zero Parallax Setting*), a positiva, a negativa e a divergente. A paralaxe zero é quando os pontos correspondentes em cada imagem estão na mesma posição, ou seja, a diferença entre eles é zero; neste caso, os pontos convergem na retina. A paralaxe positiva ocorre quando a distância entre pontos correspondentes está entre zero e uma constante t, e dão a sensação de que os objetos estão distantes; isto ocorre porque o ponto de convergência das imagens no eixo de projeção de cada olho é obtido após o plano de projeção. Já a paralaxe negativa nos passa a sensação de que os objetos estão próximos de nós, como que saindo do

monitor; tal efeito é consequência do cruzamento dos eixos de projeção de cada olho ocorrer antes de chegar ao plano de projeção. Por fim, a paralaxe divergente é um caso especial da paralaxe positiva, quando a distância entre os pontos correspondentes ultrapassa a constante t, causando um certo desconforto ao usuário, já que esse tipo de fenômeno não encontra similar na visão humana.

2.2. Estereoscopia

A estereoscopia baseia-se métodos que utilizam um par de imagens planas para visualização de uma imagem tridimensional, oferecendo a cada olho do observador uma perspectiva diferente. Dessa forma, um requisito para obtermos o efeito estereoscópico é a utilização de ambos os olhos. Com imagens estereoscópicas, é resgatada uma informação muito importante que se perde em imagens bidimensionais: a sensação de profundidade entre os diferentes objetos que a compõem.

Nas próximas subseções detalharemos os principais métodos de visualização de vídeos estereoscópicos: a estereoscopia anaglífica, a estereoscopia por luz polarizada, os monitores autoestereoscópicos e os óculos obturadores.

2.2.1. Estereoscopia anaglífica

É o método mais simples de se implementar e que foi utilizado na primeira tentativa dos cinemas em reproduzir filmes em 3D durante a década de 1920 (Lipton, 1982). O método consiste em termos dois vídeos, um para cada olho, e retirarmos de um deles as informações relativas à cor vermelha (por exemplo, do vídeo destinado ao olho direito), e do outro, as informações relativas às cores azul e verde (por exemplo, do vídeo destinado ao olho esquerdo). Logo após, criamos um novo vídeo resultante da junção dos dois primeiros, como exemplificado na Figura 2. Na reprodução, o espectador usa um par de óculos especiais atuando como um filtro, possuindo uma lente vermelha (para o olho direito nesse caso) e outra ciano (para o olho esquerdo). Com isso, conseguimos novamente separar cada imagem e deixamos que nosso cérebro encarregue-se de obter o efeito estereoscópico (Mendiburu, 2009).

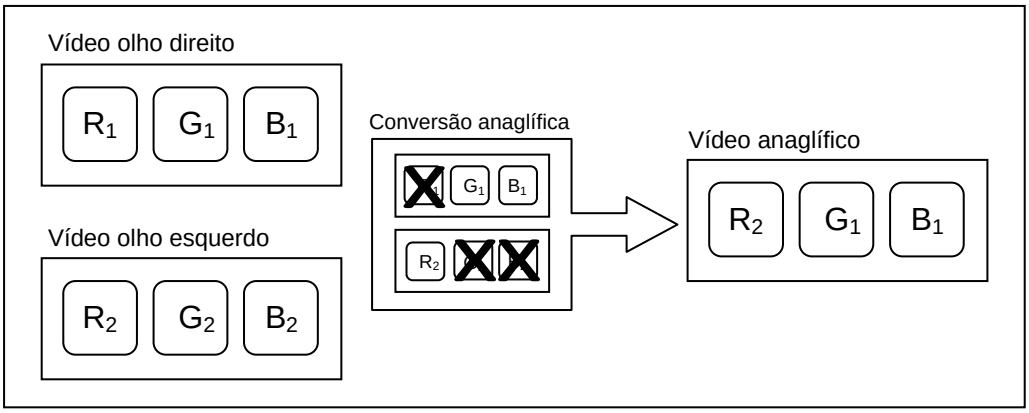

Figura 2 – Processo de conversão de um par estéreo para vídeo anaglífico. Note que os dados de R_1, G_2 e B_2 são perdidos.

As duas principais vantagens deste método são o custo para a produção e reprodução do vídeo anaglífico, que é baixo e não requer equipamentos com alta tecnologia, e o arquivo final é menor em relação aos outros métodos, já que temos somente um sinal de vídeo resultante da junção dos dois originais, garantindo uma boa compressão. Além disso, em termos de tecnologia, trata-se de um método muito simples de ser implementado.

Já a principal desvantagem é que, pelo fato de retirarmos informações do canal de cores de cada vídeo e utilizarmos as cores presentes nas lentes dos óculos para separar cada imagem, as cores resultantes da combinação dos dois não é a cor real do vídeo original. Além disso, se por um acaso for necessária a reversão do processo, isto é, de anaglífico para o par estéreo, não há uma solução trivial, pois a conversão envolveu perdas de informação de cores em ambos os vídeos. A recuperação dessas informações requer investigação.

2.2.2. Estereoscopia por luz polarizada

Neste método utilizamos dois vídeos, cada qual destinado a um dos nossos olhos, que são projetados separadamente em uma tela metalizada. Cada projetor possui um filtro polarizador, responsável por projetar a imagem em um ângulo diferente na tela. Com o auxílio de óculos possuindo esses mesmos filtros, conseguimos que cada olho veja apenas a projeção destinada a ele (Mendiburu, 2009).

Como os dois vídeos são reproduzidos separadamente e de forma íntegra, não vemos aqui a desvantagem de perdermos a cor real da cena. Por essa razão, dispositivos com estereoscopia por luz polarizada são os que vêm sendo utilizados pela indústria cinematográfica e é a tecnologia por trás dos cinemas 3D atuais. Entretanto, uma complexidade a mais é introduzida neste método: ambos os vídeos devem estar em perfeita sincronia, para que sejam reproduzidos na mesma linha de tempo. Isso é válido tanto para a gravação quanto para a edição e a reprodução, fazendo-se necessário de equipamentos mais robustos e por conseqüência, mais caros.

2.2.3. Óculos obturadores

Diferente dos óculos utilizados em vídeos anaglíficos e por luz polarizada, que filtram as imagens corretas para cada olho, os óculos obturadores separam a imagens mecanicamente. Esta é uma tecnologia muito utilizada pelos televisores 3D e funciona da seguinte forma: o monitor exibe alternadamente em uma alta frequência as imagens para cada olho e os óculos, compostos por lentes de LCD, também alternam entre si na mesma frequência o nível de opacidade de cada lente. Com isso, por uma fração mínima de tempo, uma lente se encontrará opaca e a outra não, e consequentemente, um olho vai enxergar a imagem e o outro não. Como a essa troca ocorre muitas vezes a cada segundo, nossos olhos não notam a opacidade, e o efeito adquirido é a estereopsia sem perca de qualidade de imagem.

Os principais problemas desta técnica são o alto custo para a produção de cada óculos, inviabilizando seu uso em cinemas, por exemplo; a falta de um padrão para estes, sendo assim, não é possível utilizar o mesmo óculos para televisores 3D de marcas diferentes; e a perda da resolução ou brilho das imagens, dependendo do padrão de reprodução utilizado para reduzir o *flicker*[1].

2.2.4. Monitores autoestereoscópicos

A obrigatoriedade de se utilizar óculos especiais, vista nas técnicas apresentadas anteriormente, se mostra uma abordagem invasiva que pode gerar certo desconforto ou até mesmo fadiga se usado por muito tempo.

Visando o descarte desses óculos ou qualquer outro equipamento na visualização de vídeos 3D, temos a tecnologia envolvida na criação de monitores autoestereoscópicos, que como o próprio nome diz, são capazes de gerar sozinhos a sensação de profundidade nas imagens reproduzidas. Tal feito é realizado criando-se diferentes visões estéreo de uma mesma cena, vista por ângulos diferentes e limitados a certo segmento do campo de visão do espectador, fazendo com que você veja a cena por outra perspectiva ao movimentar-se para outro campo de visão. Para isso, coloca-se no monitor uma película especial, chamada película lenticular, que é formada por pequenas lentes, as lentículas, capazes de direcionar a luz de cada imagem para um ângulo diferente. Além disso, o par de imagens estéreo é submetido a uma técnica chamada *interlacing*, na qual as imagens são fatiadas em pequenas partes do tamanho das lentículas e são intercaladas. Com isso, cada fatia é direcionada pelas lentículas para o respectivo olho (Sony Corporation, 2010) (Hutchison, 2008).

Um problema ainda em estudo e enfrentado em monitores autoestereoscópicos é que o espectador deve se situar em pontos chaves para ver a imagem 3D, devido ao alcance limitado do campo de visão fornecido. Esses pontos são poucos e fora deles, a imagem aparece borrada. Além disso, ainda é uma tecnologia a ser aprimorada e de alto custo de produção.

2.3. Aplicações de conteúdo estereoscópico

A presença de vídeos estereoscópicos no cinema não é um fato inédito. Houve um crescimento cinematográfico na década de 1950 utilizando-se do método de luz polarizada, como uma forma de trazer o público novamente ao cinema, o qual naquela época começava a experimentar um declínio de audiência, devido à popularidade das TVs (Lipton, 1982). Entretanto, devido à baixa qualidade e tecnologia apresentada, rapidamente caíram em desuso. Atualmente, o vídeo 3D voltou ao centro de atenção da indústria cinematográfica partir da estreia do filme Avatar, utilizando-se tecnologia mais

[1] *Flicker*: fenômeno obtido em monitores quando sua taxa de atualização é baixa, fazendo com que apareçam na reprodução piscadas rápidas, que podem se tornar incômodas durante a visualização.

madura, telas de alta resolução e uma boa estratégia de marketing, mostrando serem muito rentáveis às grandes produtoras como Disney e Warner, em mais uma tentativa de atrair o público.

Dirigindo-se para o lado doméstico, a indústria vem oferecendo aos poucos televisores de alta definição e preparados para exibição de conteúdo 3D. Estes se apresentam com um preço elevado, que tende a diminuir conforme a escala e a demanda aumentem. Pesquisas indicam que até 2014, 80% dos televisores vendidos nos Estados Unidos possuirão tecnologia 3D[2]. Porém, tem-se a ressalva de que isso só será possível com a produção e transmissão de conteúdos preparados para a tecnologia, além da disseminação e interesse do público em obter uma transmissão com esse conteúdo.

O mercado de games parece ser um dos que mais serão beneficiados com a utilização de conteúdo 3D, fornecendo uma nova alternativa de interatividade e imersão dos usuários com os jogos. Os grandes fabricantes de consoles vêm se mostrando interessados em investir nessa tecnologia, como é o caso da Nintendo e seu portátil Nintendo 3DS, que utiliza duas telas, sendo uma delas autoestereoscópica e a outra sensível ao toque[3]; e também o caso da Sony, que atualizou o *firmware* de seu console Playstation 3, tornando capaz de reproduzir jogos em 3D com a utilização de televisores compatíveis com a tecnologia.

Na parte científica, os vídeos estereoscópicos têm grande relevância em aplicações médicas, tais como a visualização de estruturas complexas em 3D, permitindo ao médico fazer uma melhor análise na hora de uma cirurgia, por exemplo. A área de robótica também pode se beneficiar de técnicas estereoscópicas para reconhecimento de imagens e rastreamento de objetos por robôs, como estudado por (Kim et al, 2007).

2.4. Codificação de vídeo digital estereoscópico

A atenção e o apelo comercial voltado ao vídeo 3D têm novamente crescido, como é visto principalmente na produção de filmes em 3D para o cinema e no crescimento de ofertas de televisores de alta definição habilitados para sua reprodução. Com isso, foi preciso a criação de outras codificações de vídeo para acomodar as novas tecnologias que vêm surgindo. Atualmente, novas formas de codificação vêm sendo definidas e outras ampliadas, porém, cada uma visa atender um tipo específico de aplicação ou técnica estereoscópica. Apresenta-se nas seções 2.4.1 e 2.4.2 algumas dessas codificações como expostas por (Smolic et al, 2009), divididas em convencional, na qual os vídeos não sofrem alteração no formato de representação, e a baseada em vídeo e profundidade, em que novas camadas de dados estão presentes juntos com o sinal de vídeo.

[2] Pesquisa publicada em http://idgnow.uol.com.br/computacao_pessoal/2010/08/02/pesquisa-80-das-tvs-vendidas-nos-eua-em-2014-terao-3d, último acesso em 07 de agosto de 2010.
[3] Mais informações em: http://e3.nintendo.com/3ds/, último acesso em 07 de agosto de 2010.

2.4.1. Codificação convencional

É a mais comum e utilizada na produção de vídeos para canais de TV e filmes. Pode ser subdividida em CSV (*Conventional Stereo Video*) e MVC (*Multiview Video Coding*) (Smolic et al, 2009).

O CSV é o que se utiliza de dois vídeos de uma mesma cena cujas imagens apresentam diferentes ponto de vista, sendo codificados e processados tendo como objetivo apresentar um vídeo distinto para cada olho. É aqui que se encaixam os processos de criação de vídeos anaglíficos e por luz polarizada, por exemplo. A desvantagem dessa codificação é que a mesma cena é vista pelo usuário independente da posição onde ele se encontra frente à tela.

O MVC surge como uma extensão do CSV, com o objetivo de utilizar mais de 2 vídeos para uma mesma cena, conseguindo assim fornecer diferentes pontos de vista baseado na localização do usuário em frente à tela. O problema se encontra no número de pontos de vista limitado e que acarreta um aumento do arquivo final dependendo do número de vídeos comportados.

2.4.2. Codificação baseada em vídeo e profundidade

Para este tipo, além do vídeo, é enviado junto com o sinal um mapa de profundidade, obtido através de cálculos complexos que mapeiam a cena fazendo a estimativa de disparidade e profundidade dos objetos nela contidos. Esses cálculos oneram o dispositivo por adicionarem processos de síntese e *rendering* tanto na codificação quanto no processo reverso. Além disso, os algoritmos são complexos e ainda propensos a erros. Pode ser dividido em três: V+D (*Video plus Depth*), MVD (*MultiView plus Depth*) e LDV(*Layered Depth Video*) (Smolic et al, 2009).

O V+D foi o primeiro disponível e cuja funcionalidade foi supracitada: junto ao sinal do vídeo segue um mapa de profundidade que habilita o dispositivo à criação do segundo vídeo tendo em vista a produção da estereopsia.

Uma extensão do anterior, o MVD combina enviar no sinal de vídeo múltiplas visões de uma mesma cena, cada qual com seu próprio mapa de profundidade. Novas visões podem sem criadas combinando-se duas outras existentes. Com isso, temos a possibilidade de disponibilizar várias visões ao usuário, sendo um bom candidato a ser utilizado por monitores autoestereoscópicos.

O LDV inclui no sinal, além da camada do vídeo e seu mapa de profundidade, nomeadas como visão principal, novas camadas responsáveis por outras visões, como dados contendo informações referentes à cena vista de outras direções. Tudo isso é processado para a criação de diferentes visões. A complexidade dos algoritmos aumenta, porém, o arquivo final é menor do que o do MVD (as camadas conseguem eliminar visões que elas mesmo conseguem processar).

Como se pode observar, as desvantagens gerais dessas codificações são os algoritmos complexos e ainda propensos a erros, passíveis de um melhor estudo. Além disso, temos um processamento pesado tanto no lado transmissor quanto no receptor, exigindo equipamentos mais poderosos e caros.

2.5. Compressão de vídeo digital estereoscópico

O resultado final da gravação de um vídeo resulta em arquivos muito grandes. Com isso, durante a codificação deve-se levar em conta a aplicação de técnicas de compressão visando obter um arquivo menor e com um mínimo de qualidade. Para o caso de vídeos estereoscópicos, uma atenção especial deve ser tomada, visto que nestes a quantidade de dados pode ser bem maior.

Apresenta-se na seção 2.5.1 o processo de compressão utilizado em vídeos monoculares, ou seja, compressão baseada em apenas um sinal de vídeo, e na seção 2.5.2 são discutidas as limitações apresentadas por esta compressão quando aplicada a vídeos estereoscópicos.

2.5.1. Compressão de vídeo monocular

Um *stream* de vídeo é na verdade uma sequência de imagens (chamadas de quadros) que mostradas em conjunto a certa frequência nos passam a sensação de movimento. Tendo isso em vista, o primeiro passo na compressão de vídeo digital é utilizar em cada quadro a compressão aplicada em imagens para eliminar as informações de redundância que estas apresentam. Isso pode envolver tanto métodos de compressão sem perdas quanto com perdas, o que influencia na qualidade da imagem resultante.

O processo de compressão de imagens envolve aplicar uma redução do espaço de cor, tendo em vista diminuir a quantidade de cores para promover compressão, sendo portando com perdas. Logo após, há aplicação de uma transformada, uma função matemática que vai mudar a forma de representação dos dados em função da sua freqüência, e posterior quantização, que visa eliminar as freqüências mais altas do que um certo limiar. Dependendo do limiar estabelecido, o olho humano pode não perceber diferenças significativas, ou seja, obtém-se maior ou menor qualidade. Exemplos de transformadas comumente utilizadas são a DCT (*Discrete Cossine Transform*) e DWT (*Discrete Wavelet Transform*). Vale lembrar que a compressão é feita na etapa de quantização, a qual elimina dados – método com perdas. Por fim, é feita a redundância estatística, sem perda, a qual atribui o número de bits para cada dado conforme a freqüência em que aparecem, garantindo compressão. Destas, as mais conhecidas são Huffman, LZW e *Run-length* (Gonzales & Woods, 2002).

Além de aplicar a compressão em cada imagem, temos nos vídeos um outro tipo de redundância a ser explorada: a redundância temporal. Esta é representada pela similaridade entre quadros vizinhos de uma sequência, resultando em

dados que podem ser eliminados. Como os quadros são similares, o proposto é codificar apenas alguns e predizer como serão os próximos, excluindo assim a codificação destes.

Para a remoção da redundância temporal, no formato clássico de compressão, os quadros são classificados em I, P ou B (Chapman & Chapman, 2004). Os quadros I (*Intracoded frames*) são aqueles que sofrem apenas a compressão espacial, através dos algoritmos de compressão de imagens. Os quadros P (*Predictive frames*) são codificados em relação a um quadro I ou P anteriores a ele, obtendo-se uma estimativa do que mudou entre ele e seu antecessor (estimativa de movimento), ou seja, excluímos este quadro e ficamos apenas com os dados da estimativa de movimento para posterior reconstrução deste. Como essa predição envolve erros, é também codificada uma tabela de compensação de movimento, contendo a diferença entre a posição estimada e a posição real dos objetos. Como outros quadros P podem ser codificados a partir de um quadro P anterior, há uma propagação de erros, e por essa razão, deve-se estabelecer um limite de criação de quadros P consecutivos, chamado de *Prediction Span*. Por fim, os quadros B (*Bidirectional frames*) são codificados tanto em relação ao um quadro P ou I anterior a eles quanto em relação a um quadro P ou I posterior a eles, obtendo-se uma taxa maior de compressão, porém impactando o tempo de processamento, já que precisamos esperar os quadros P ou I posteriores serem processados para o cálculo.

2.5.2. Limitações na compressão para vídeos digitais estereoscópicos

Um problema na compressão de vídeos estereoscópicos utilizando a compressão de vídeo monocular é que o nível de compressão obtido já não é suficiente, levando em conta que dependendo do tipo de método de visualização estereoscópica utilizado podemos ter o dobro ou mais de informações do que um vídeo monocular.

Com isso, novas técnicas de compressão foram criadas, baseadas em alguns aspectos específicos de vídeos estereoscópicos. Por exemplo, observa-se um novo tipo de redundância no par estéreo envolvendo a presença de muitas informações correlatas entre seus quadros (Siegel et al, 1994). Como cada um dos quadros de um dos vídeos é muito semelhante ao quadro do vídeo correspondente, diferenciando-se nas distâncias promovidas pela disparidade, propõe-se a codificação de somente um destes e a transmissão junto ao sinal de um mapa de disparidade, tornando possível recriar o segundo vídeo. Outra abordagem é a codificação utilizando um vídeo e seu mapa de profundidades (seção 2.4.2).

O que se observa é que as técnicas de compressão para vídeo estereoscópico já existentes e as que estão sendo desenvolvidas são voltadas cada uma para um tipo específico de visualização estereoscópica (anaglífica, luz polarizada, óculos obturadores ou monitores autoestereoscópicos), não havendo uma técnica genérica que seja compatível para todos os tipos. Por exemplo, testes feitos por (Andrade & Goularte, 2009) mostram que na compressão de um par de vídeos estéreo, durante o processo da aplicação de transformadas e posterior quantização, dependendo da transformada utilizada

pode-se incluir ruídos em um deles que não estarão presentes no outro, diminuindo assim a qualidade final e a percepção da estereopsia quando utilizado o método de visão anaglífica, como pode ser visto na figura 3.

MJPEG	x264	Rududu
28,30 Db	23,24 Db	30,79 Db

Figura 3 - Médias de PSNR[4] resultantes da codificação de um vídeo anaglífico. Pelos dados observa-se que o vídeo codificado com a técnica Rududu, que utiliza transformadas *Wavelet*, foi melhor classificado frente às outras duas técnicas que utilizam DCT (Andrade & Goularte, 2009)

3. Proposta de trabalho

O objetivo deste trabalho é desenvolver uma técnica que faça a conversão de um vídeo anaglífico em seu correspondente par estéreo. Com isso, pode-se propor um processo alternativo de codificação de vídeos estereoscópicos que não seja dependente de um método específico de visualização, possibilitando ser genérico o suficiente para codificar o vídeo e obter resultados com qualidade de percepção de profundidade melhor que as técnicas atuais e com boa taxa de compressão.

3.1. Metodologia

Essa proposta consiste em uma extensão dos trabalhos iniciados por (Andrade & Goularte, 2009) e (Andrade & Goularte, 2010). Neles os autores encontraram parâmetros que, ao serem utilizados no processo de codificação de um par estéreo, possibilitam deixar os vídeos resultantes com uma boa qualidade e utilizáveis para qualquer técnica estereoscópica. Esses parâmetros são para redução do espaço de cor (Figura 4 A) e aplicação de transformada com posterior quantização (Figura 4 B).

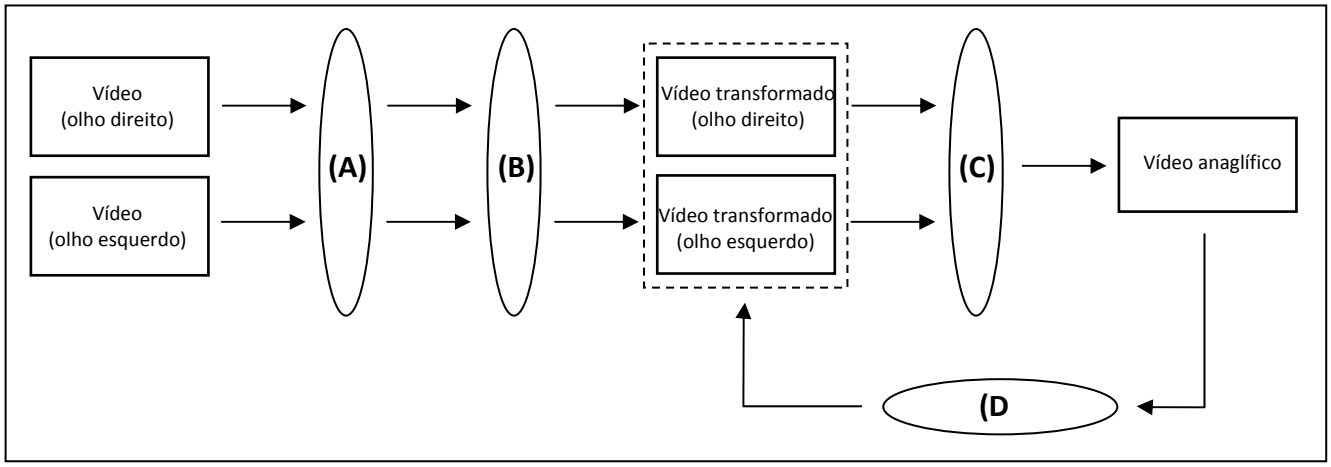

Figura 4 – Esquematização do processo proposto, envolvendo redução do espaço de cor (A), aplicação da transformada *Wavelet* e quantização (B), geração do vídeo anaglífico (C) e também do passo de reversão (D)

[4] PSNR é uma métrica medida em decibéis que relaciona a quantidade de ruído introduzida no vídeo após o processo de compressão com perdas. Quanto maior o valor do PSNR, diz-se que o vídeo apresenta uma maior qualidade.

O proposto é continuar na ponta final deste processo, tendo em vista obter uma maior compressão do arquivo sem perder a qualidade obtida. Para isso, propõe-se utilizar o método anaglífico (Figura 4 C) e produzir um único *stream* de vídeo a ser transmitido. Visando continuar a ser genérico, como o vídeo anaglífico é incompatível para outros métodos de visualização como luz polarizada, deve-se estudar e aplicar o processo reverso (Figura 4 D), isto é, a partir do vídeo anaglífico, remontar o par estéreo do passo anterior, o que requer investigação de como recuperar as informações perdidas na conversão anaglífica, como discutido na seção 2.2.1.

De forma a atingir o objetivo, será feito um estudo da literatura relacionada à compressão de vídeos, com destaque aos trabalhos atuais que vêm sendo feitos com vídeos estereoscópicos, com uma possível análise da viabilidade da aplicação de compressão temporal como forma de aumentar a compressão do vídeo resultante do processo proposto. Também será levado em conta um estudo envolvendo tratamento e processamento de imagens, visando obter conhecimento e possíveis técnicas a serem aplicadas para recuperação do par estéreo a partir de um vídeo anaglífico.

Para analisar os resultados obtidos com os estudos, pretende-se desenvolver uma aplicação que implemente as técnicas a serem criadas tanto para a codificação do vídeo anaglífico quanto para o processo reverso. Para isso, será escolhida uma linguagem de programação capaz de lidar com arquivos de vídeo e imagem efetuando boa performance, bem como de um computador munido de um bom poder de processamento e armazenagem.

Como forma de avaliar a qualidade do arquivo final pretende-se realizar testes objetivos e subjetivos que possam ser empregados com o envolvimento de usuários reais nos casos em que forem possíveis, para tornar a avaliação imparcial. Prevê-se a utilização das medidas objetiva PSNR (*Peak Signal-to-Noise Ratio*) e subjetiva MOS (*Mean Opinion Score*) (Winkler, 2005).

3.2. Cronograma

Segue abaixo a proposta de atividades a serem seguidas e a divisão destas durante o mestrado. Tal divisão visa tanto cumprir as obrigações do curso de mestrado quanto o trabalho a ser desenvolvido no projeto.

1. Obtenção dos créditos do curso: através do cumprimento de disciplinas obrigatórias pelo programa do curso.
2. Exame de inglês: a ser realizado no segundo semestre de 2010 ou primeiro semestre de 2011.
3. Exame de qualificação: a ser agendado e realizado entre julho e agosto de 2011
4. Análise da literatura: revisão de livros, artigos e trabalhos relacionados ao projeto, retirados de fontes de pesquisa confiáveis, envolvendo codificação e compressão de vídeos estereoscópicos e processamento de imagens.

5. Estudo dos trabalhos em andamento no grupo de pesquisa: análise do trabalho e resultados obtidos por (Andrade & Goularte, 2009) e (Andrade & Goularte, 2010) e estudo para adaptação das novas técnicas a serem desenvolvidas.
6. Implementação da aplicação para criar vídeos anaglíficos: desenvolvimento ou adaptação de uma aplicação que transforme um par de vídeos estéreo em um vídeo anaglífico.
7. Estudo e criação do algoritmo de reversão do método: estudo de maneiras de realizar a conversão de um vídeo anaglífico para seu correspondente par estéreo.
8. Implementação do algoritmo do processo reverso: desenvolvimento da aplicação que realize a conversão do vídeo anaglífico para seu correspondente par estéreo.
9. Elaboração, aplicação e análise de testes dos resultados obtidos.
10. Revisão do projeto e possíveis alterações: com base nos testes obtidos, fazer correções necessárias e revisar as técnicas criadas e/ou utilizadas.
11. Submissão de artigos para conferências e periódicos da área: durante a duração do trabalho, serão submetidos artigos com os resultados parciais ou finais do projeto para conferências e periódicos relacionados com a área de aplicação.
12. Defesa da dissertação: a ser realizada no final do primeiro semestre de 2012.

Tabela 1. Cronograma de realização de atividades

Atividades	**2° semestre de 2010**		**1° semestre de 2011**		**2° semestre de 2011**		**1° semestre de 2012**	
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								

4. Referências

(Andrade & Goularte, 2009) Andrade, L. A.; Goularte, R. – Percepção Estereoscópica Anaglífica em Vídeos Digitais Comprimidos com Perda. Webmedia – Brazilian Symposium on Multimedia and the Web, 2009.

(Andrade & Goularte, 2010) Andrade, L. A.; Goularte, R. – Uma Análise da Influência da Subamostragem de Crominância em Vídeos Estereoscópicos Anaglíficos. Webmedia – Brazilian Symposium on Multimedia and the Web, 2010.

(Azevedo & Conci, 2003) Azevedo, E.; Conci, A. – Computação gráfica: teoria e prática. Editora Campus, Elsevier, 2003.

(Chapman & Chapman, 2004) Chapman, N. P.; Chapman, J. – Digital Multimedia, 2nd edition. Wiley, 2004.

(Fehn et al, 2002) Fehn, C.; Kauff, P.; Op de Beeck, M.; Ernst, F.; IJsselsteijn, W.; Pollefeys, M.; Van Gool, L.; Ofek, E.; Sexton, I. – An Evolutionary and Optimised Approach on 3D-TV. Proceedings of International Broadcast Conference, 2002.

(Fehn et al, 2006) Fehn, C.; de la Barré, R.; Pastoor, S. – Interactive 3-DTV – concepts and key technologies. Proceedings of the IEEE, 2006.

(Gonzales & Woods, 2002) Gonzales, R. C.; Woods, R. E. – Digital Image Processing. Prentice Hall, 2002.

(Hutchison, 2008) Hutchison, D. – Introducing DLP 3-D TV. White Paper, 2008. Disponível em: http://dlp.com/downloads/Introducing%20DLP%203D%20HDTV%20Whitepaper.pdf. Último acesso feito em 26 de agosto de 2010.

(Kim et al, 2007) Kim, IH.; Kim, DE.; Cha, YS.; Lee, K.; Kuc, TY. – An embodiment of stereo vision system for mobile robot for real-time measuring distance and object tracking. International Conference on Control, Automation and Systems, 2007

(Lipton, 1982) Lipton, L. – Foundations of the Stereoscopic Cinema: a study in depth. Van Nostrand Reinhold Company Inc., 1982.

(Lipton, 1997) Lipton, L. – Stereo-Vision Formats for Video and Computer Graphics. White Paper, 1997. Disponível em http://www.cours.polymtl.ca/inf6802/stereo/body_stereo_formats.html. Último acesso feito em 26 de agosto de 2010.

(Mendiburu, 2009) Mendiburu, B. – 3D Movie Making Stereoscopic Digital Cinema from Script to Screen. Ed. Focal Press, Elsevier, 2009.

(Siegel et al, 1994) Siegel, M. W.; Gunatilake, P.; Sethuraman, S.; Jordan, A. G. – Compression of stereo image pairs and streams. Stereoscopic Displays and Virtual Reality Systems, 1994.