Reversão de imagens e vídeos estereoscópicos anaglíficos ao par estéreo original

Matheus Ricardo Uihara Zingarelli

Orientador: ***Prof. Dr. Rudinei Goularte***

Monografia apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, para o Exame de Qualificação, como parte dos requisitos para obtenção do título de Mestre em Ciências de Computação e Matemática Computacional.

USP – São Carlos
Agosto de 2011

Reversão de imagens e vídeos estereoscópicos anaglíficos ao par estéreo original

Matheus Ricardo Uihara Zingarelli

Resumo

ZINGARELLI, M. R. U. **Reversão de imagens e vídeos estereoscópicos anaglíficos ao par estéreo original**. 2011. 59f. Monografia de qualificação (Mestrado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2011.

A atenção voltada à produção de conteúdos 3D atualmente tem sido alta, em grande parte devido à aceitação e à manifestação de interesse do público para esta tecnologia. Isso reflete num maior investimento das indústrias cinematográfica, de televisores e de jogos visando trazer o 3D para suas produções e aparelhos, oferecendo modos diferentes de interação ao usuário. Com isso, novas técnicas de captura, codificação e modos de reprodução de vídeos 3D, aqui denominados vídeos estereoscópicos, vêm surgindo ou sendo melhorados, visando aperfeiçoar 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 estereoscópica, sendo que para cada método há uma técnica de codificação diferente que pode causar perdas significativas se aplicada ao outro método. Uma proposta é criar uma técnica que seja genérica, ou seja, que através de parâmetros adequados obtenha vídeo sem nenhuma perda tanto na qualidade quanto na percepção de profundidade, característica marcante nesse tipo de conteúdo. Visando compressão, propõe-se que o par estéreo seja transformado em vídeo anaglífico, resultando em apenas um sinal, com redução de pelo menos 50% do tamanho original. No entanto, para que haja compatibilidade entre todos os tipos de visualização, é necessário possuir também o processo reverso, transformando o vídeo anaglífico novamente em um par estéreo. Tal processo não é trivial e requer um estudo de como recuperar as informações perdidas durante a conversão do vídeo para seu formato anaglífico. Este trabalho tem como objetivo propor um processo de reversão anaglífica, de modo a obter um par estéreo com qualidade e sem perda da percepção de profundidade quando reproduzido por diferentes sistemas de visualização estereoscópica.

Palavras-chave: Codificação estereoscópica. Estereoscopia Anaglífica. Vídeos Estereoscópicos.

Sumário

Índice de Figuras

Índice de Tabelas

1. Introdução

Com a estreia de Avatar em 2009, os chamados filmes 3D voltaram ao interesse do público. Acompanhando tal interesse e com o amadurecimento da tecnologia, a indústria do cinema tem investido no 3D em suas principais produções cinematográficas. Junto com este avanço, é também observável em outros ramos da indústria cada vez mais pesquisas para criação de televisores e telas que reproduzam conteúdos 3D com alta qualidade e definição, com ou sem a necessidade de óculos (LG, 2011; MENDIBURU, 2009; NINTENDO, 2011; SONY, 2011), permitindo até mesmo assistir a um vídeo em diferentes pontos de vista, de acordo com a posição que a pessoa se encontra em relação ao televisor. Na área científica, pode-se observar o uso de técnicas envolvendo 3D para visão artificial de robôs, utilizando aspectos de percepção da profundidade para o cálculo da distância entre objetos espalhados pelo ambiente (KIM et al, 2007).

Em termos técnicos, os vídeos 3D são definidos como vídeos estereoscópicos e utilizam métodos (também chamados estereoscópicos), os quais consistem em apresentar duas imagens bidimensionais de uma mesma cena, deslocadas horizontalmente – o que é chamado de par estéreo –, para serem interpretadas pelo cérebro humano na formação de uma imagem única e tridimensional, provocando a sensação de profundidade e distanciamento. Tais métodos visam, através de imagens bidimensionais, simular o efeito obtido naturalmente na visão humana: como nossos olhos estão distantes horizontalmente um do outro, cada olho tem um ponto de vista diferente, deslocado. Isso é chamado de disparidade binocular (AZEVEDO; CONCI, 2003).

Com o passar dos anos, câmeras especiais têm sido 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; SMOLIC et al., 2009). Pode ser visto também o desenvolvimento e aperfeiçoamento de técnicas para conversão e apresentação de vídeos estereoscópicos a partir de vídeos originalmente em 2D (TAM; ZHANG, 2006). No que diz respeito à reprodução, 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 estereoscópico sem o auxílio de óculos ou qualquer outro dispositivo (DODGSON, 2005).

Apesar dos avanços vistos na tanto na captura quanto na reprodução e representação de vídeos estereoscópicos, ainda existe a necessidade de mais pesquisa na área da codificação. Um reflexo disso é a atual falta de padronização no modo de organizar dados de vídeos estereoscópicos para fins de armazenamento ou transmissão, sendo que as estratégias existentes para tal organização podem ser divididas em dois grupos: o método de Lipton (LIPTON, 1997) e os métodos envolvendo vídeo e profundidade (SMOLIC et al., 2009). No método de Lipton o par estéreo é armazenado em contêineres (AVI, por exemplo), com compressão ou não, o que possibilita a reprodução de vídeos estereoscópicos com pouca ou nenhuma modificação dos sistemas de visualização. Os métodos envolvendo vídeo e profundidade, por sua vez, utilizam técnicas consagradas de compressão de vídeo (como MPEG-2 e H.264), bem como de novos conceitos envolvendo mapas de profundidade para atender às demandas de tecnologias mais atuais, como a criação de novas visões e os monitores autoestereoscópicos.

Embora simples, o método de Lipton armazena o par estéreo, o que resulta no dobro de dados comparado a vídeos monoculares (apenas um sinal de vídeo). Já os métodos baseados em vídeo e profundidade utilizam de estratégias para aumento da compressão explorando conceitos de profundidade e relacionamento entre o par estéreo. Mesmo assim, podem resultar no armazenamento de um grande volume de dados dependendo do número de sinais de vídeos envolvidos para a criação de várias visões. Além disso, as técnicas utilizadas para compressão são apenas adaptadas para tratar vídeos estereoscópicos e, devido aos diferentes tipos criados e em estudo, podem resultar em problemas de compatibilidade entre sistemas diferentes (SMOLIC et al., 2009). Por se tratar muitas vezes de compressão com perdas, ocorre também a geração de artefatos que impossibilitam a correta percepção de profundidade em alguns casos, notadamente em vídeo anaglíficos (ANDRADE; GOULARTE, 2009, 2010). Como resultado, não existe uma técnica exclusiva para codificação de vídeos estereoscópicos que produza vídeos de qualidade, com boa taxa de compressão e atendendo a todos os atuais métodos de visualização, tanto os que necessitam de óculos especiais (anaglífico, lentes polarizadas e obturadores) quanto o autoestereoscópico.

Tendo-se observado esta lacuna, em um projeto de doutorado relacionado é proposto a realizar a compressão de imagens e vídeos estereoscópicos com parâmetros adequados para que não haja perda da percepção de profundidade seja qual for o método de visualização utilizado. Seguindo esta linha de raciocínio e visando maior compressão, propõe-se reduzir o

volume de dados do par estéreo através de sua transformação em anaglífico. Desse modo, o formato anaglífico poderia ser utilizado para fins de armazenamento/transmissão (pois possuiria boa taxa de compressão) e a técnica atenderia ao método de visualização anaglífico (com diferencial em qualidade). Buscando compatibilidade com os outros tipos de visualização estereoscópica, é necessário reverter o anáglifo gerado, de forma a restaurar o par estéreo para que este possa ser utilizado pelos outros métodos. Tal reversão é uma novidade na área e necessita de mais estudos para saber como deve ser executada. Com isso, o objetivo deste trabalho é desenvolver uma técnica de reversão de anáglifos ao seu respectivo par estéreo.

O texto está organizado da seguinte forma: a Seção 2 traz fundamentos da visão humana e definições necessárias como base para o entendimento da visualização estereoscópica. A Seção 3 trata da revisão bibliográfica, se aprofundando nas pesquisas sobre codificação e compressão estereoscópica. A Seção 4 explica algumas técnicas para a avaliação da qualidade de vídeos digitais que passaram por um processo de codificação. A Seção 5 apresenta com detalhes a proposta deste trabalho e descreve as atividades já realizadas durante o primeiro ano de Mestrado, que culminaram na a criação de uma técnica de reversão anaglífica baseada na chamada “Tabela de Índice de Cores”. São também apresentados os resultados já obtidos com a técnica implementada. Na Seção 6 são discutidas as limitações da técnica criada, e delineadas as atividades a serem realizadas de forma a refiná-la, juntamente com o cronograma proposto a ser seguido até o final do Mestrado. Por fim, apresentam-se todas as referências utilizadas como apoio à produção do texto e um apêndice com um artigo submetido e aprovado para ser publicado na edição de 2011 do WebMedia.

2. Fundamentos da visualização estereoscópica

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 mesma imagem ligeiramente deslocada horizontalmente, característica classificada como disparidade binocular (AZEVEDO; CONCI, 2003). Por essas razões era de se esperar que, ao olharmos para um objeto, ele fosse visto sob duas perspectivas diferentes, e não somente uma como ocorre em nossa visão. Além disso, dentre os vários objetos presentes no campo de visão, temos a capacidade de interpretar diferentes profundidades e texturas entre eles, mesmo ao nos movermos para diferentes direções. A utilização de ambos os olhos para formar uma única imagem, com percepção de profundidade, é definida como estereopsia (LIPTON, 1982).

O principal personagem envolvido nesse fenômeno 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. Uma série de informações de profundidade está envolvida no processo de transformação tridimensional de uma imagem pelo cérebro. Tais informações podem ser divididas em três grupos: informações monoculares, informações óculo-motoras e informações estereoscópicas (AZEVEDO; CONCI, 2003).

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 é amplamente explorada 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.

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 de que a distância entre linhas paralelas que demarcam 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) e também até hoje por arquitetos no desenho de plantas e projetos.

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. Junto com a interposição, a variação de luz incidente sobre um objeto, bem como a utilização de sombras, passam 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 montanhas no horizonte, nota-se que as que se encontram mais distantes aparecem menos nítidas, como se estivessem desaparecendo. Do mesmo jeito, na ocorrência de chuvas fortes objetos distantes ficam ofuscados na paisagem. 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 que um objeto se encontra, dada pelo nível de detalhamento que obtemos. Quanto mais distante 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.

O conhecimento prévio está ligado à nossa experiência de vida. Nosso cérebro vai armazenando informações dos objetos ao passo que vamos tendo contato com eles no mundo real, criando relacionamentos de tamanho e espaço ocupado por estes 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, inferimos qual está mais próximo ou mais afastado, qual é maior ou menor.

A paralaxe de movimento é uma informação resultante de movimento, também passando a ideia de 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 óculo-motoras

As informações vistas na Seção 2.1.1 podem ser reproduzidas em imagens no papel, sendo então capturadas e formada na retina dos olhos. Já as óculo-motoras estão ligadas a aspectos fisiológicos, não sendo reproduzíveis em papel. Elas são geradas de acordo com o relaxamento e contração dos músculos envolvidos no movimento do globo ocular, sendo interpretadas pelo cérebro para relacionar a distância e profundidade entre objetos. Temos dois tipos: a acomodação e a convergência.

A acomodação está 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. Tanto a acomodação quanto a convergência são reproduzidas artificialmente por máquinas de captura como câmeras e filmadoras digitais, quando se altera o foco.

2.1.3. Informações estereoscópicas

Como anteriormente exposto, cada olho possui uma perspectiva diferente do que se está sendo observado devido à disparidade binocular. Cabe ao cérebro se encarregar de retirar as informações das distâncias relativas dos objetos e de interpretar essas duas perspectivas resultando na fusão em uma única. As técnicas que fornecem imagens diferentes, deslocadas, para cada olho tentando reproduzir esse fenômeno no cérebro são descritas como estereoscópicas e as informações utilizadas são também denominadas estereoscópicas. Destas informações, as principais são a estereopsia, disparidade e paralaxe.

Já mencionada 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 a utilização dos dois olhos. É com esta informação, em cooperação com as outras informações aqui descritas, que obtemos a fusão das imagens e percebemos 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 ou de que a tela parece ser funda.

A diferença na distância entre as posições da imagem formada em cada retina em relação ao centro desta é chamada de disparidade. Isso pode ser melhor entendido através do seguinte exemplo ilustrado na Figura 1: 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 os pontos dessas 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 projetadas por algum dispositivo 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, produzindo 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 óculos, ele parece estar tremido, com regiões duplicadas e sobrepostas.

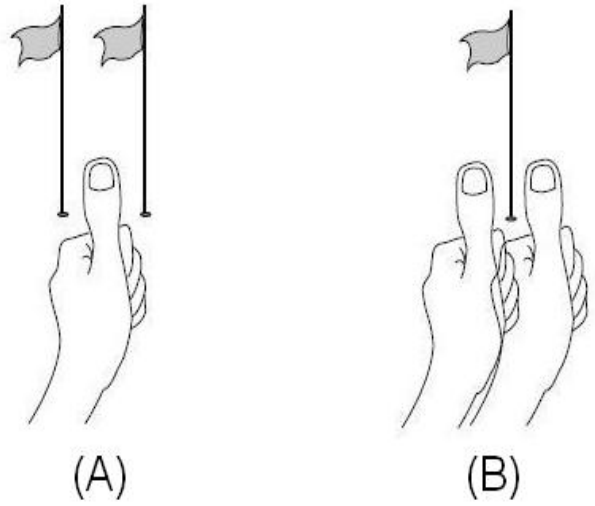

Figura 1 - Exemplo de observância da informação de disparidade (STEREOGRAPHICS, 1997). Em (A), quando focamos nossa visão no dedo polegar, a bandeira aparece duplicada ao fundo. Em (B), quando focamos nosso olhar na bandeira, o dedo polegar aparece duplicado.

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 (Figura 2(A)) ocorre 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 (Figura 2(B)) 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 (Figura 2(C)) 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 (Figura 2(D)) é um caso especial da paralaxe positiva a ser evitado, quando a distância entre os pontos correspondentes ultrapassa a constante t, causando desconforto ao usuário, já que esse tipo de fenômeno não encontra semelhante na visão humana.

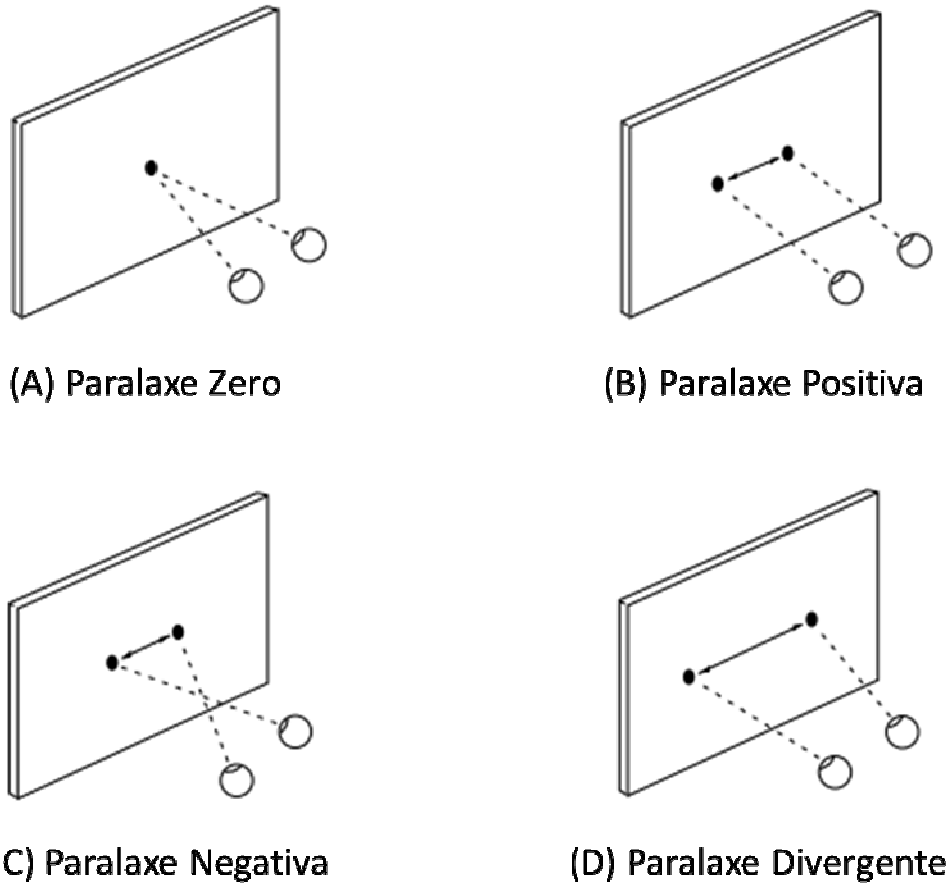

Figura 2 - Tipos de paralaxe, adaptado de Stereographics (1997). Paralaxe Zero (A) ocorre em pontos de convergência na retina. Paralaxe positiva (B) passa a percepção de que objetos estão distantes. Paralaxe negativa (C) passa a percepção de que objetos estão próximos. Paralaxe divergente (D) é um caso especial quando a paralaxe positiva ultrapassa um limiar, causando desconforto ao usuário.

2.2. Tipos de visualização estereoscópica

Foi visto na Seção 2.1 uma série de informações que auxiliam na percepção de profundidade de imagens reproduzidas por algum dispositivo. As informações estereoscópicas em especial, juntamente com a utilização de ambos os olhos, fazem com que o cérebro interprete a cena com profundidade e distanciamento. Dessa forma, um requisito para obtermos o efeito estereoscópico é a utilização de ambos os olhos.

Nas Seções de 2.2.1 à 2.2.4, os principais métodos de visualização de vídeos estereoscópicos são detalhados, sendo eles: estereoscopia anaglífica, estereoscopia por luz polarizada, óculos obturadores e monitores autoestereoscópicos.

2.2.1. Estereoscopia anaglífica

É o método mais simples e que não requer nenhum aparelho especial para reprodução. Foi utilizado na primeira tentativa dos cinemas em reproduzir filmes em 3D durante a década de 1920 (LIPTON, 1982). O método consiste em retirar de uma das imagens de um par estéreo as informações relativas a uma das cores primárias (por exemplo, a cor verde da imagem do lado direito), e do outro, as informações relativas das duas cores restantes (por exemplo, as cores vermelho e azul da imagem do lado esquerdo). Logo após, criamos uma nova imagem resultante da junção das informações retiradas das duas primeiras (para este exemplo, denominada imagem anáglifa verde-magenta), como exemplificado na Figura 3. Na reprodução, o espectador usa um par de óculos especiais atuando como um filtro, possuindo nas lentes as cores que foram eliminadas, ou seja, uma lente verde (para o olho esquerdo, nesse caso) e outra lente magenta, junção da cor vermelha com a cor azul (para o olho direito). Com isso, cada olho irá enxergar apenas uma das imagens, obtendo a disparidade

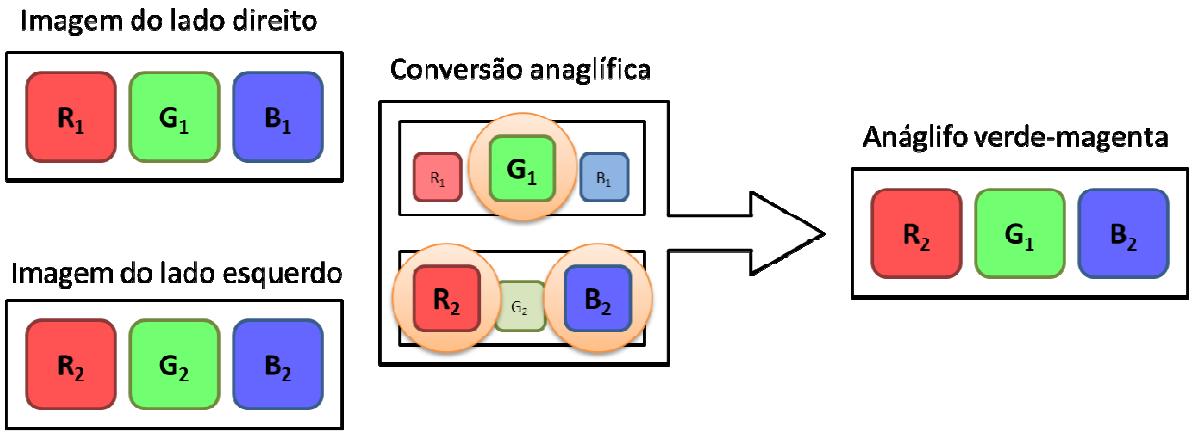

Figura 3 - Processo de conversão anaglífica verde-magenta. É retirada a cor verde da imagem correspondente à visão do lado direito e as cores vermelha e azul (magenta) da imagem correspondente à visão do lado esquerdo.

binocular.

As duas principais vantagens deste método são: o custo para a produção e reprodução desse tipo de vídeo ou imagem, que é baixo e não requer equipamentos com alta tecnologia; e o volume de dados, muito menor do que em relação aos outros métodos a serem vistos, já que neste caso é armazenado apenas metade das informações originais, garantindo boa compressão. Já a principal desvantagem é consequência direta da eliminação de metade das informações do par estéreo original. Pelo fato de retirarmos informações do canal de cores e utilizarmos aquelas presentes nas lentes dos óculos para separar cada imagem, as cores resultantes da combinação dos dois não é a real, fiel à do original. Além disso, a criação do

anáglifo seguindo este processo impossibilita a reversão deste para o par estéreo original, pois a conversão envolveu eliminação de informação. A recuperação dessas informações requer investigação e é o objetivo do Mestrado.

2.2.2. Luz polarizada

Para este método e os métodos seguintes, tem-se como requisito o par estéreo. Neste caso, cada vídeo ou imagem do par é projetado 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.

Como o par estéreo é reproduzido separadamente e de forma íntegra, não há aqui a desvantagem de se perder a cor real da cena. Por essa razão, os dispositivos que utilizam a estereoscopia por luz polarizada são os que vêm sendo comumente 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 captura quanto para a edição e a reprodução, fazendo-se necessária aquisição de novos equipamentos, mais robustos e por consequência, mais caros.

2.2.3. Óculos obturadores

Diferente dos óculos utilizados na visualização anaglífica 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 alta frequência as imagens para cada olho. 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 milhares de vezes a cada segundo, nossos olhos não notam a opacidade.

Os principais problemas desta técnica são: alto custo para a produção de cada óculos, inviabilizando seu uso em cinemas, por exemplo; a falta de um padrão para estes óculos, não sendo possível utilizar os mesmos para televisores 3D de marcas diferentes; e a perda da resolução ou brilho das imagens, dependendo do padrão de reprodução utilizado para reduzir o *flickering*[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 quando usado por muito tempo, além de quebrar o paradigma de como os espectadores estão acostumados a assistir televisão. Visando o descarte desses óculos ou qualquer outro dispositivo na visualização de vídeos 3D, temos a tecnologia envolvida na criação de monitores autoestereoscópicos, os quais são capazes de fornecer diferentes pontos de vista, chamados de visões, para cada olho, Tais visões são limitadas a certo segmento do campo de visão do espectador, fazendo com que este veja a cena de perspectivas diferentes ao movimentar-se pelo ambiente. Para isso, o monitor possui uma película especial (lenticular) formada por pequenas lentes (lentículas) capazes de direcionar a luz de cada imagem para um ângulo diferente, como ilustrado na Figura 4. 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.

Por ser o método mais atual, ainda passa por pesquisa e desenvolvimento em diversos laboratórios e fabricantes de TV, e apresenta deficiências a serem superadas. Uma delas é que o espectador deve se situar em pontos chaves para ter a percepção de profundidade, devido ao alcance limitado do campo de visão fornecido. Esses pontos são poucos e fora deles ocorre invasão de ambas as imagens do par estéreo, efeito chamado de *crosstalk* (STEREOGRAPHICS, 1997). Ainda são necessários mais alguns anos até que televisores autoestereoscópicos sejam produzidos em massa, o que torna o custo de produção elevado. Entretanto, algumas soluções para dispositivos móveis, com telas pequenas, já se encontram disponíveis no mercado estrangeiro (LG, 2011; NINTENDO, 2011).

[1] *Flickering*: fenômeno que ocorre em monitores quando sua taxa de atualização é baixa, fazendo com que apareçam piscadas rápidas durante a reprodução, o que pode se tornar incômodo na visualização.

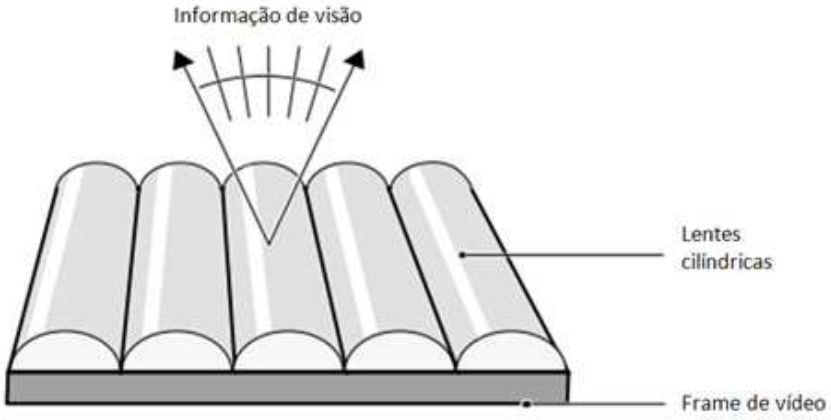

Figura 4 - Tecnologia lenticular de monitores autoestereoscópicos (HALLE, 1997). A luz, ao passar pelas lentes cilíndricas, pode ser direcionada para posições diferentes, podendo assim enviar diferentes visões para cada olho.

2.3. Aplicações

Como dito anteriormente, a presença de filmes 3D no cinema não é um fato inédito. Atualmente, eles voltaram ao centro de atenção da indústria cinematográfica por apresentar uma tecnologia mais madura, telas de alta resolução e boas estratégias de marketing, mostrando serem muito rentáveis às grandes produtoras como Disney e Warner. Mesmo assim, alguns erros do passado continuam nos filmes atuais, resultando em produção ou conversão para 3D cujos resultados são de baixa qualidade, gerando certa insatisfação do público.

Ao público doméstico a indústria vem oferecendo televisores de alta definição preparados para exibição de conteúdo 3D. Estes possuem preço elevado, que tende a diminuir conforme aumentem em escala e demanda. Pesquisas indicam que até 2014, 80% dos televisores vendidos nos Estados Unidos possuirão tecnologia 3D[2]. A ressalva é que isso só será possível com a produção e transmissão de conteúdos preparados para a tecnologia, que ainda é muito baixa, além da disseminação e interesse do público em obter transmissão deste tipo.

O mercado de games parece ser um dos que mais serão beneficiados com a utilização de conteúdo 3D para entretenimento, fornecendo uma nova alternativa de interatividade e imersão dos usuários com os jogos. Os grandes fabricantes de consoles vêm se mostrando

[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.,Acesso em: 28 jul. 2011.

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 (NINTENDO, 2011); e também o caso da Sony, cujo console Playstation 3 é capaz de reproduzir conteúdo 3D com a utilização de televisores compatíveis com a tecnologia, e vem constantemente lançando conteúdo neste formato[3].

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).

[3] Uma lista com diversos filmes e jogos em 3D para Playstation 3 pode ser visto em http://blog.us.playstation.com/2011/07/01/stereoscopic-3d-on-ps3-updated-list-of-all-3d-games-and-movies/. Acesso em: 28 jul. 2011.

3. Aspectos de codificação e compressão estereoscópica

3.1. Espaço de cores e subamostragem de crominância

A representação de imagens se dá através de tons monocromáticos ou coloridos. Computacionalmente, os tons monocromáticos podem ser representados por um byte, produzindo assim 256 níveis em escala de cinza, representando apenas informações de luminância, isto é, intensidade da luz. Quando além de luminância, se deseja também informações sobre as cores, é necessário utilizar o chamado espaço de cores, no qual cada cor é representada por uma tripla de valores (x, y, z) (SALOMON, 2008), de acordo com a teoria tricromática (AZEVEDO; CONCI, 2003). Existem vários modelos de espaço de cores, cada qual apropriado para um tipo de aplicação ou sistema de visualização. Os discutidos abaixo são os modelos RGB e YC_bC_r, que vêm sido utilizados nas atividades relacionadas à pesquisa. Mais detalhes sobre outros espaços de cores podem ser vistos nos textos de Azevedo e Conci (2003) e Feitosa-Santana et al. (2006).

O modelo RGB é baseado na tripla de cores primárias: vermelho, verde e azul. Elas são classificadas como cores aditivas, isto é, através da mistura das três são produzidas as outras cores, sendo que o branco é obtido quando misturadas em sua intensidade máxima. Este modelo é o mais popular e o comumente utilizado por dispositivos de captura, como câmeras fotográficas, e de apresentação, como as telas de LCD (RICHARDSON, 2003).

Um problema do modelo RGB é que a característica de luminância está diretamente contida no valor de cada componente de cor do modelo. Isso impossibilita que se possa explorar uma propriedade do sistema visual humano: temos mais sensibilidade à luminância do que às cores (SALOMON, 2008). Desse fato, a informação relativa às cores (crominância) pode ser representada em uma resolução menor do que a informação relativa à luminância, sem a perda de qualidade (RICHARDSON, 2003). Essa característica é explorada durante a

codificação de imagens e vídeos em uma etapa chamada de subamostragem de crominância (KERR, 2009).

O modelo YC_bC_r possui em suas componentes a separação das informações relativas à luminância (Y) das relativas à crominância (C_b e C_r), como uma tentativa de simular a visão humana. C_b e C_r representam valores de crominância das componentes azul e vermelha. A componente verde (C_g) pode ser obtida através de C_b e C_r, já que neste modelo a soma das três é sempre igual a 1, sendo por isso eliminada da representação. Os valores de Y, C_b e C_r podem ser obtidos do modelo RGB por um processo de conversão do espaço de cores através da fórmula mostrada na Equação 1, a qual é uma recomendação da ITU-T (RICHARDSON, 2003). O inverso é também possível, e se apresenta na Equação 2.

$$Y = 0{,}299 * R + 0{,}587 * G + 0{,}114 * B$$
$$Cb = 0{,}564 * (B - Y)$$
$$Cr = 0{,}713 * (R - Y)$$

Equação 1 - Conversão do espaço de cores RGB para YC_bC_r

$$R = Y + 1{,}402 * Cr$$
$$G = Y - 0{,}344 * Cb - 0{,}714 * Cr$$
$$B = Y + 1{,}772 * Cb$$

Equação 2 - Conversão do espaço de cores YC_bC_r para RGB

Como o espaço de cores YC_bC_r separa dados de luminância e cor, pode-se realizar a subamostragem de crominância. Como mencionado, o olho humano é mais sensível às variações de luminância do que de cores, com isso, os dados referentes à cor podem ser amostrados a uma taxa menor do que os dados referentes à luminância, o que pode resultar em grande redução do volume de dados final. Dependendo da taxa em que são amostrados, pode-se classificar a subamostragem em três modelos: 4:4:4, 4:2:2 e 4:2:0 (RICHARDSON, 2003). Outros modelos são sugeridos por Kerr (2009) e podem ser vistos em seu trabalho.

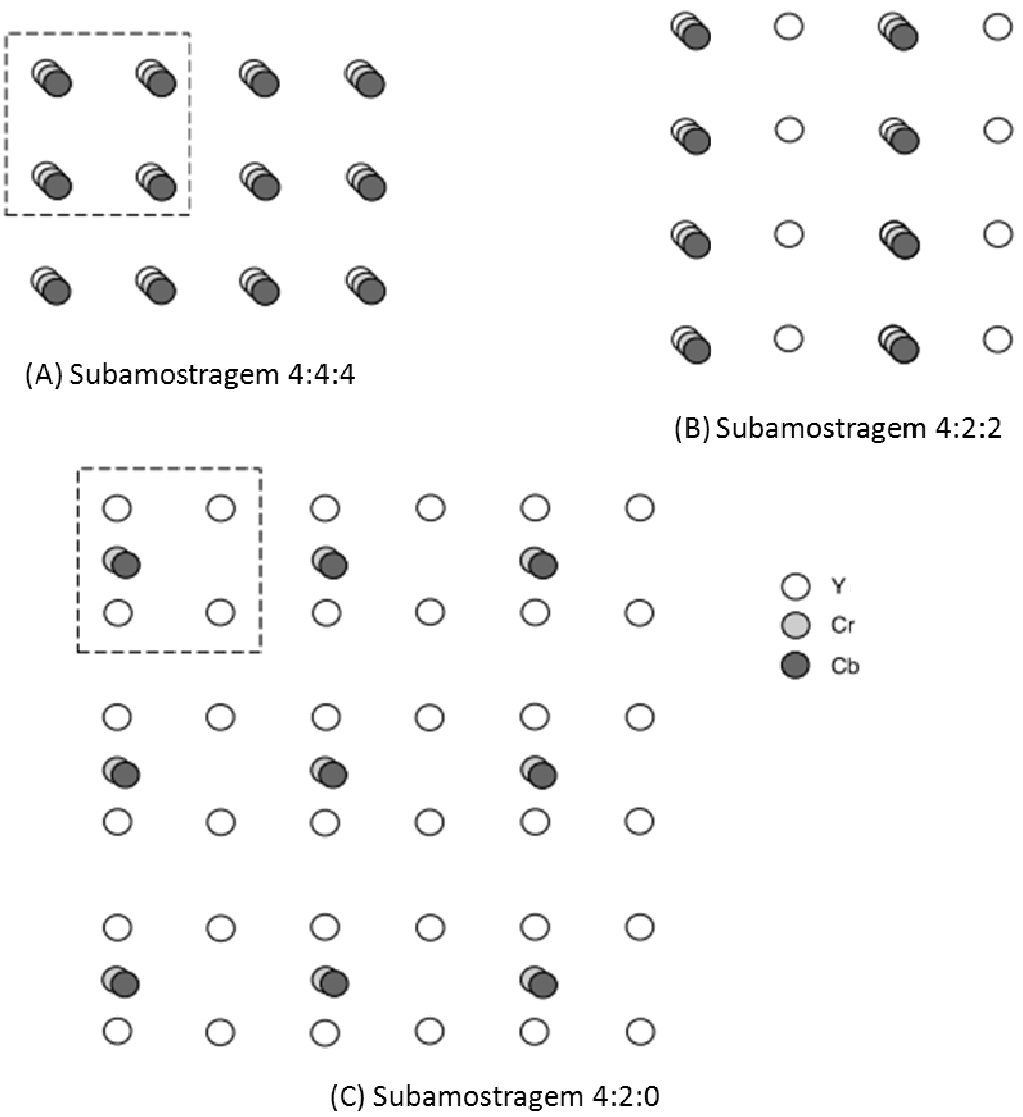

Figura 5 - Tipos de subamostragem de crominância, dependendo da quantidade de redução da informação de cor, adaptado de Richardson (2003). No modelo 4:4:4 não há redução de informação. No modelo 4:2:2 há redução de informação na horizontal. No modelo 4:2:0 há redução de informação tanto na horizontal quanto na vertical.

No modelo 4:4:4 (Figura 5 (A)) não há redução da resolução das cores, isto é, para cada amostra de Y, há uma amostra de C_b e uma de C_r. Este modelo mantém a fidelidade das cores da imagem, porém, não contribui na compressão. No modelo 4:2:2 (Figura 5 (B)), para cada quatro amostras horizontais de Y, há duas amostras de C_b e duas de C_r, reduzindo-se com isso 1/3 de informação. Já no modelo 4:2:0 (Figura 5 (C)), a redução é feito tanto horizontal quanto verticalmente, havendo variações na escolha de qual pixel a amostragem deve ou não ocorrer.

Cabe lembrar que o processo de eliminação de cores da etapa de subamostragem de crominância, utilizando-se os modelos 4:2:2 e 4:2:0, é irreversível. Com isso, na conversão de uma imagem YC_bC_r (4:2:2 ou 4:2:0) para seu similar em RGB, o retorno aproximado ao modelo 4:4:4 pode ser obtido copiando-se o valor dos pixels vizinhos (ou uma média deles) a

cada pixel não amostrado. Deve-se retornar ao modelo 4:4:4 já que o espaço de cores RGB não possui dados de luminância e crominância separados.

3.2. Codificação estereoscópica

Com o passar do tempo, novas codificações para vídeo digital vão surgindo, tendo em vista diminuição do volume de dados sem perda da qualidade do vídeo. Com a utilização de vídeos digitais estereoscópicos, o desafio aumenta, pois o volume de dados a ser armazenado tende a ser o dobro de um vídeo digital monocular, já que são necessários dois sinais de vídeo, um para cada olho. Com isso, novas estratégias de codificação vêm sido estudadas, algumas visando adaptar as técnicas já conhecidas, outras explorando características específicas encontradas em vídeos estereoscópicos. Essas estratégias podem ser divididas em dois tipos: a codificação convencional, baseada no método de Lipton, e a codificação baseada em vídeo e profundidade, que explora características de profundidade do par estéreo para aumentar a taxa de compressão.

3.2.1. Codificação convencional

Em seu trabalho, Lipton (1997) criou dois formatos para vídeos estereoscópicos em uso até hoje, os quais foram feitos de modo que pudessem ser utilizados com pouca ou nenhuma alteração na infraestrutura de hardware disponível para visualização. Nestes formatos o par estéreo é armazenado (em um contêiner AVI, por exemplo), sendo que cada quadro contém tanto o quadro do vídeo esquerdo quanto o do direito, sendo posicionados sobrepostos (formato acima-abaixo) ou lado a lado (formato lado-a-lado), dependendo do sistema em que serão reproduzidos. Dessa forma, cada vídeo pode então ser codificado utilizando as técnicas já conhecidas.

Quando apenas dois sinais de vídeo são armazenados, isto é, o par estéreo, o formato do vídeo pode também ser classificado como CSV – *Conventional Stereo Video* (SMOLIC et al., 2009). Entretanto, novas tecnologias de telas e monitores são capazes de gerar mais de uma visão ao espectador, dependendo da posição em que ele se encontra em

relação à tela. Para cada visão, é necessário um par estéreo diferente, os quais podem ser armazenados tanto como lado-a-lado quanto acima-abaixo. Neste caso, o formato é chamado de MVC (*Multiview Video Coding*) e já possui seu padrão pelo grupo MPEG – MPEG-2 *Multiview Profile* e também para o H.264/AVC (SMOLIC et al., 2009).

O problema claro desse tipo de codificação é o tamanho final do arquivo, já que é necessário armazenar ao menos dois sinais de vídeo, o que torna a taxa de compressão obtida limitada.

3.2.2. Codificação baseada em vídeo e profundidade

Esse tipo de codificação busca explorar características dos vídeos estereoscópicos em relação à profundidade. Ao invés de se armazenar o par estéreo, armazena-se apenas um dos sinais de vídeo, junto com seu respectivo mapa de profundidade de pixels, o qual pode ser entendido como um sinal de vídeo auxiliar, com dados apenas de luminância, em que o valor de cada pixel significaria sua distância em relação à câmera de captura (Figura 6). Através deste mapa de profundidades, seria possível recriar o segundo vídeo do par estéreo, ou até mesmo novas visões. Por conter apenas dados de luminância, possui tamanho menor em relação a um vídeo colorido (o segundo vídeo do par estéreo, neste caso), o que possibilita maior compressão.

Figura 6 - Processo de codificação utilizando vídeo e mapa de profundidades para a formação de um vídeo estereoscópico (SMOLIC et al., 2009).

Na pesquisa realizada por Smolic et al. (2009), codificações baseadas em vídeo e profundidade foram classificadas em três tipos: V+D (*Video plus Depth*), cujo funcionamento é o mencionado no parágrafo anterior e que possui uma especificação no MPEG-C Parte 3; MVD (*MultiView plus Depth*), utilizado quando mais de um sinal de vídeo é enviado, possibilitando a geração de múltiplas visões, semelhante ao MVC visto na Seção 3.2.1; e LDV (*Layered Depth Video*), mais complexa, envolvendo além do sinal de vídeo e mapa de profundidades, camadas adicionais de vídeo contendo informações auxiliares retiradas de capturas feitas por outras câmeras, que seriam utilizadas para a geração de novas visões, sem a necessidade do armazenamento do vídeo completo. Os autores também propõem outro formato, chamado de DES (*Depth Enhanced Stereo*), o qual seria um apanhado de todos os outros: o par estéreo seria armazenado junto com respectivas camadas de profundidade e auxiliares, promovendo um formato genérico que poderia ser utilizado por diferentes sistemas estereoscópicos.

Por serem mais flexíveis e possibilitarem a criação de visões virtuais, a codificação baseada em vídeo e profundidade mostra possuir um papel importante para o futuro da tecnologia 3D. Pesquisas vêm sendo desenvolvidas em cima desta abordagem, como a criação de câmeras que capturam a cena e já geram o mapa de profundidades (FEHN et al., 2002), ou a conversão de vídeos 2D para 3D através de mapas de profundidades (TAM; ZHANG, 2006). Entretanto, os algoritmos tanto para criação de mapa de profundidades quanto para criação de visões virtuais ainda são complexos e propensos a erros.

Em outra pesquisa, esta realizada por Vetro (2010), os formatos aqui mencionados são discutidos em relação às diferentes técnicas de compressão e tecnologias de representação de cada um. Ao final do artigo, o autor observa a falta de adoção de formatos que garantam interoperabilidade entre diferentes sistemas estereoscópicos. Smolic et al. (2009) também aponta este problema e tenta solucioná-lo propondo o formato DES. Entretanto, tal formato agrega uma grande quantidade de informações que podem não ser utilizadas dependendo do dispositivo para o qual são transmitidas, o que leva ao armazenamento de dados desnecessários.

3.2.3. Compressão

Um *stream* de vídeo é na verdade uma sequência de imagens (chamadas de quadros) que, mostradas em conjunto a certa frequência, nos dá 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 redução do espaço de cor, tendo em vista diminuir a quantidade de informação cores para promover compressão, como visto na Seção 3.1. Logo após, há aplicação de uma transformada, uma função matemática que vai mudar a forma de representação dos dados em função da sua frequência, e posterior quantização, que visa eliminar as frequências mais altas do que certo limiar, sendo, portanto, com perdas. 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 Cosine Transform*) e DWT (*Discrete Wavelet Transform*) (GONZALEZ; WOODS, 2008). Com isso, são eliminadas as redundâncias espaciais e psicovisuais. Por fim, é feita a redundância estatística, sem perda, a qual atribui um número de bits para cada dado conforme a frequência em que aparecem, garantindo compressão. Destas, as mais conhecidas são Huffman, LZW e por carreira (*run-length*).

Além de aplicar a compressão em cada imagem, temos nos vídeos 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, armazenando somente as diferenças entre eles.

Para a remoção da redundância temporal, baseado no padrão MPEG-1, os quadros são classificados em I, P ou B (CHAPMAN; CHAPMAN, 2004; SAYOOD, 2005). 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 anterior 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 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, 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.

3.2.4. Limitações na codificação de imagens e vídeos estereoscópicos

Um problema na compressão de vídeos estereoscópicos utilizando a compressão de vídeo monocular é que o nível de compressão obtido utilizando técnicas atuais já não é suficiente, levando em conta que dependendo do tipo de visualização estereoscópica utilizado, podemos ter o dobro ou mais de informações do que um vídeo monocular. Além disso, como discutido na Seção 3.2.2, há a falta de um padrão de codificação específico para imagens e vídeos estereoscópicos. Isso traz como consequência uma série de pesquisas em andamento para obter melhores resultados na codificação de mídias estereoscópicas, a maioria das vezes adaptando-se os padrões de codificação existentes.

Pode-se observar adaptações visando melhoria de desempenho para transmissão, como visto no trabalho de Li et al. (2009), em que os autores buscam melhorar a eficiência da codificação MVC presente como extensão do H.264, propondo uma nova estrutura de criação de visões adaptável. Porém, o que é mais encontrado na literatura são pesquisas explorando um novo tipo de redundância encontrada em imagens estereoscópicas, chamada de correlação de imagens (*worldline correlation*)[4] (ADIKARI et al., 2005; BALASUBRAMANIYAM, EDIRISINGHE e BEZ, 2005). Como há uma grande semelhança entre as imagens do par estéreo, é proposto que uma das imagens sirva de base para a predição da outra, parecido com a estimativa de movimento realizada na etapa de remoção de redundância temporal. Com isso, o par estéreo poderia ser codificado como apenas um sinal de vídeo, sendo o segundo sinal reconstruído pelas estimativas obtidas da correlação de imagens, obtendo-se assim boa taxa de compressão.

Embora obtenha uma boa taxa de compressão, o que se nota em todas essas pesquisas é que as técnicas estudadas são voltadas cada uma para um tipo específico de visualização

[4] Alguns autores usam o termo *inter-view correlation* (LIN et al., 2009; MERKLE et al., 2007)

estereoscópica, não havendo uma técnica genérica que seja compatível para todos os tipos. Além disso, em nenhuma delas é considerada a visualização anaglífica. Testes feitos por Andrade e Goularte (2009, 2010) mostram que a compressão de um par de vídeos estéreo, através das estratégias utilizadas pelos codificadores atuais para subamostragem de crominância e aplicação de transformadas com posterior quantização, pode incluir ruídos no vídeo resultante que impossibilitam a percepção de profundidade quando utilizado o método anaglífico. Os autores também encontram nestes trabalhos parâmetros adequados para ambas as etapas. Entretanto, por ainda armazenar o par estéreo, a taxa de compressão é baixa.

4. Avaliação de qualidade de vídeos digitais

Durante a codificação de vídeos digitais podem surgir diversos artefatos tais como *blockiness*[5] e *blurring*[6], causados pela aplicação de transformadas e posterior quantização. Do mesmo modo, a utilização de subamostragem de crominância elimina dados de cor, com a intensidade dependendo do tipo aplicado. Tudo isso influencia na qualidade final do vídeo codificado, podendo dificultar sua visualização. A qualidade de um vídeo digital pode ser avaliada de duas maneiras: objetivamente, através de cálculos e métricas, e subjetivamente, observando-se a qualidade visual.

Para este trabalho, a avaliação objetiva é focada na utilização do *Peak Signal-to-Noise Ratio* (PSNR). O PSNR é uma métrica de qualidade muito utilizada na análise de compressão de imagens (WINKLER, 2005). Sua fórmula não é complicada e está baseada na comparação pixel a pixel de duas imagens, sendo uma a fonte (imagem original) e a outra a imagem que passou por algum processo (neste caso, a imagem codificada). O PSNR retorna um valor em decibéis, num intervalo de 0 a 100, resultante do nível de similaridade das imagens comparadas, sendo que quanto maior o valor, maior a similaridade encontrada. Segundo Ebrahimi, Chamik e Winkler (2004), o PSNR não possui a palavra final em termos de qualidade de imagem, já que não leva em conta a percepção visual humana, apenas uma análise matemática entre pixels correspondentes. Com isso, uma imagem com baixo PSNR não significa necessariamente ser de baixa qualidade quando visualizada por uma pessoa.

Os testes subjetivos envolvem um grupo de pessoas avaliando a qualidade de um conjunto de vídeos. Para manter coesão dos resultados é necessário seguir um padrão tanto para a apresentação dos vídeos quanto para a classificação da qualidade de cada um. Neste trabalho, pretende-se seguir as recomendações da ITU-T (2004), utilizando o método *Double Stimulus Continuous Quality Scale* (DSCQS) para apresentação e classificação de qualidade

[5] *Blockiness* é um efeito de distorção no qual se observa a formação de pequenos blocos quadriculados na imagem, geralmente causados pela aplicação de uma transformada DCT.
[6] *Blurring* é um efeito que resulta em perda de detalhes finos nas bordas de objetos de uma imagem, geralmente causados pela aplicação de uma transformada DWT.

dos vídeos, e o *Mean Opinion Score* (MOS) para análise dos dados obtidos. Seguindo a metodologia DSCQS, são montadas estruturas de vídeos alternados ABAB, em que A é o vídeo original e B é o vídeo após um processo de codificação, sendo separados por intervalos com trechos de tela cinza (Figura 7). Com isso, ambos o vídeo original e o codificado de cada sequência são passados duas vezes. Logo após os avaliadores classificam a qualidade de cada um utilizando a tabela de classificação ilustrada na Figura 8. Os resultados são anotados e ao final é feita uma média da classificação dada por cada avaliador, o MOS. Como pode ser observado na Figura 8, a tabela de classificação tem uma escala numérica de 0 a 100, subdivida por intervalos que classificam o vídeo desde "Péssimo" até "Excelente" (WINKLER, 2005). Os avaliadores classificam cada vídeo fazendo uma marcação horizontal na tabela correspondente. Desse modo, é possível converter a classificação para um valor numérico e consequentemente obter um valor final para a avaliação subjetiva.

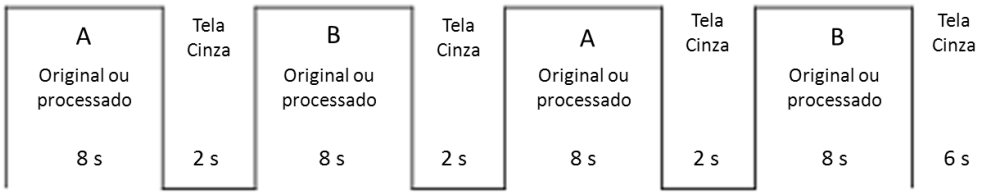

Figura 7 - Processo de teste de qualidade subjetiva de imagens ou vídeos. Adaptado de ITU-T (2004).

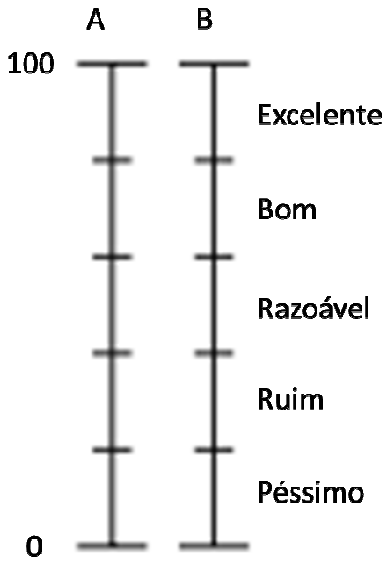

Figura 8 - Tabela de classificação de vídeo utilizando o método de avaliação DSCQS. Adaptado de Winkler (2005).

5. Proposta de trabalho

5.1. Apresentação da proposta

Pelo exposto na Seção 3.2, observa-se dois problemas na codificação de vídeos estereoscópicos. O primeiro é o grande volume de dados a ser armazenado, já que se trabalha com dois ou mais sinais de vídeo, dependendo da tecnologia de visualização a ser empregada. O segundo é a falta de uma técnica de codificação específica para vídeos estereoscópicos e independente do tipo de visualização. Percebe-se que as técnicas tradicionais de compressão de vídeo monocular com perdas produzem artefatos que prejudicam a percepção de profundidade quando aplicadas a vídeos estereoscópicos; da mesma forma, novas técnicas criadas especificamente para codificação estereoscópica produzem boa taxa de compressão, entretanto, são exclusivas para um método ou sistema de visualização, não sendo aplicável a todos. Tendo isso em vista, o objetivo do trabalho é desenvolver uma técnica de reversão de vídeos anaglíficos ao seu correspondente par estéreo. Deste modo, é possível que seja posteriormente desenvolvida uma técnica de codificação na qual o par estéreo é transformado em anaglífico, já que neste formato apenas um sinal de vídeo é armazenado, reduzindo pela metade o volume de dados. Com isso, utiliza-se a técnica de reversão para obter novamente o par estéreo, o qual pode ser então utilizado por outros métodos de visualização estereoscópica, tornando a codificação genérica.

O processo de reversão do vídeo anaglífico para o par estéreo requer uma estratégia bem elaborada, uma vez que a geração do anáglifo implica em perda de informação tanto espacial quanto de cor. Como visto na Figura 3, dos seis canais de cores existentes no par estéreo, apenas três são utilizados sendo os outros três descartados. Uma simples duplicação das informações dos canais presentes no anáglifo não bastaria para recuperar o par estéreo com qualidade, já que as imagens no par original são deslocadas em cada lado.

Durante o primeiro ano de Mestrado, algumas atividades foram realizadas tendo em vista recuperar os dados perdidos durante a transformação anaglífica. Estas atividades estão detalhadas na Seção 5.2.

5.2. Atividades realizadas

A abordagem estudada foi não eliminar nenhum dado de cor do par estéreo durante a transformação anaglífica, e sim armazenar aqueles não utilizados em uma estrutura de dados que chamamos de "Tabela de Índice de Cores". Da Figura 3, podemos ver que esta tabela seria então formada pelos dados dos canais de cores R_1, G_2 e B_1. Juntos, estes três canais formam um novo anáglifo, que chamamos de "anáglifo complementar", deixando a denominação de "anáglifo principal" para aquele a ser de fato utilizado em combinação com os óculos. Observa-se que desta forma um decodificador possuiria todos os dados necessários para reconstruir o par estéreo com qualidade e fidelidade de cores. Entretanto, nenhuma compressão é obtida, já que foi feito apenas uma reorganização dos canais de cores do par estéreo.

Como um requisito necessário para a reconstrução com qualidade do par estéreo são as informações de cor de ambos seus componentes, uma estratégia visando compressão é converter o espaço de cores do anáglifo complementar de RGB para YC_bC_r e armazenar somente as informações referentes à crominância (C_b e C_r), descartando informação de luminância (Y), já que esta pode ser obtida do anáglifo principal. Além disso, o anáglifo complementar, já no espaço YC_bC_r, pode passar pela etapa de subamostragem de crominância, reduzindo ainda mais o volume de dados a ser armazenado na Tabela de Índice de Cores.

De posse dessas informações, foi realizado o processo de conversão anaglífica que está ilustrado na Figura 7. Primeiro, o par estéreo é transformado em dois anáglifos, o principal (verde-magenta) e o complementar. O anáglifo verde-magenta foi escolhido por ter se mostrado com os melhores resultados pelo trabalho de Andrade e Goularte (2010). Começa então o processo de construção da Tabela de Índice de Cores, através da conversão do anáglifo complementar do espaço de cores RGB para YC_bC_r, passando pela subamostragem de crominância 4:2:2, também testada por Andrade e Goularte (2010) como a melhor alternativa em conjunto com o anáglifo verde-magenta. Logo após, descartamos as informações de Y e armazenamos somente C_b e C_r juntamente com o anáglifo principal. Observe que as informações de Y podem ser descartadas, pois trazem apenas dados relacionados à luminância, o que não impacta tanto quanto a perda de dados de cor. Além disso, dados de Y podem ser recuperados através do anáglifo principal durante o processo de reversão, a ser explicado a seguir. Vale também ressaltar que tanto a Tabela de Índice de

Cores quanto o anáglifo principal podem ainda passar por um processo de compressão de dados sem perdas, reduzindo ainda mais o tamanho final.

O processo de reversão está ilustrado na Figura 8. Nesta etapa, o anáglifo principal também passa pelo processo de conversão do espaço de cores de RGB para YC_bC_r. Com isso, obtemos um Y', os dados de luminância do anáglifo principal. Em conjunto com os dados da

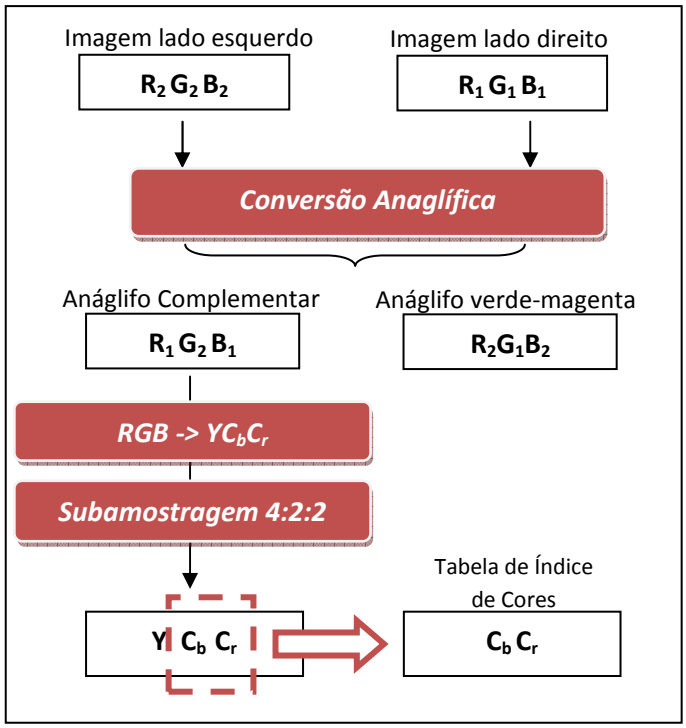

Figura 9 - Conversão anaglífica utilizando a Tabela de Índice de Cores

Tabela de Índice de Cores, utilizamos o Y' para reconstruir o anáglifo complementar, neste caso na forma de $Y'C_bC_r$, através do processo para retornar à amostragem 4:4:4 e então ser revertido para o espaço de cores RGB. De posse novamente dos dois anáglifos, basta apenas reordenar seus canais de cores para obter o par estéreo.

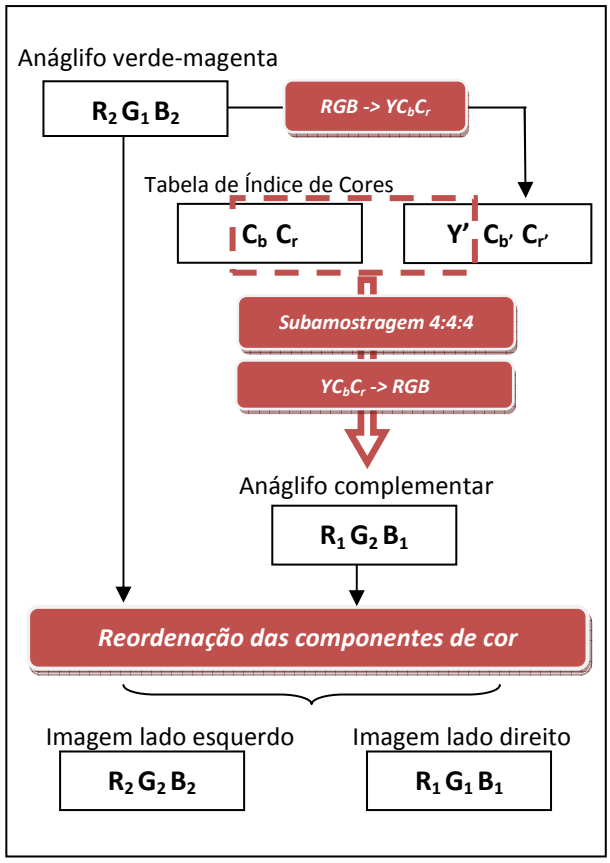

Figura 10 - Reversão anaglífica utilizando a Tabela de Índice de Cores

5.3. Resultados obtidos

A Tabela de Índice de Cores foi formada eliminando os dados de Y e utilizando os dados de C_b e C_r do anáglifo complementar, após a subamostragem de crominância 4:2:2. Isso significa que de cada 12 pixels (do formato 4:4:4), estamos descartando 4 pixels de luminância e 4 de crominância (2 de C_b e 2 de C_r). Matematicamente, espera-se que isto resulte em uma adição de 33% de dados ao arquivo final (que contem o anáglifo principal), o que pode ser reduzido ainda mais após passar pelo processo de compressão de dados sem perdas.

O processo descrito na Seção 5.2 foi implementado em C++ com suporte da biblioteca OpenCV[7], e aplicado a uma base de testes contendo 32 imagens de par estéreo. Estas imagens foram retiradas da base construída por Andrade, Cordebello e Goularte (2010), disponível em http://200.136.217.194/videoestereo/. Dos resultados obtidos, foi analisado o tamanho final do arquivo em relação à imagem original e o PSNR das imagens obtidas após o processo de reversão.

Os resultados obtidos podem ser vistos na Tabela 1, que possui cinco colunas. A primeira é a identificação de cada imagem, seguida da taxa de redução da imagem original em relação à imagem anaglífica, seguida da taxa de redução ao se adicionar os dados armazenados na Tabela de Índice de Cores, seguido do quanto de informações adicionais (*overhead*) foi inserido no arquivo final pelo processo, e por fim, o PSNR médio medido. Na última linha da Tabela 1, temos a média aritmética de cada um desses valores.

Tabela 1 - Resultados dos testes da compressão de imagens estereoscópicas usando conversão anaglífica com a Tabela de Índice de Cores (continua)

ID	Redução sem a Tabela	Redução com a Tabela	Overhead da Tabela	PSNR médio (dB)
arv01.bmp	62,80%	52,12%	10,68%	30,284
corr01.bmp	75,66%	67,65%	8,01%	35,037
cruz01.bmp	69,69%	60,49%	9,19%	34,803
do01.bmp	75,10%	67,60%	7,50%	36,484
do02.bmp	72,09%	63,78%	8,31%	34,239
do03.bmp	76,39%	68,99%	7,40%	33,386
do04.bmp	81,27%	75,57%	5,70%	36,888
do05.bmp	70,87%	62,30%	8,57%	33,777
dz01.bmp	86,10%	81,23%	4,87%	34,610
dz02.bmp	67,33%	58,46%	8,87%	36,766
dz03.bmp	68,75%	59,66%	9,09%	36,026
dz04.bmp	70,85%	61,90%	8,95%	37,126
fw01.bmp	79,17%	73,71%	5,46%	36,822

[7] O código do OpenCV pode ser obtido em http://sourceforge.net/projects/opencvlibrary/ e a Wiki contendo documentação e suporte ao uso pode ser vista em http://opencv.willowgarage.com/wiki/ (Acesso em: 28 jul. 2011).

Tabela 1 - Resultados dos testes da compressão de imagens estereoscópicas usando conversão anaglífica com a Tabela de Índice de Cores (conclusão)

ID	Redução sem a Tabela	Redução com a Tabela	Overhead da Tabela	PSNR médio (dB)
fw02.bmp	84,88%	75,17%	9,71%	35,040
hei01.bmp	67,63%	58,31%	9,32%	32,010
hei02.bmp	66,60%	56,89%	9,71%	32,124
hei03.bmp	68,70%	59,01%	9,69%	31,846
hei04.bmp	66,20%	55,89%	10,31%	31,960
mp01.bmp	74,73%	67,12%	7,62%	37,389
old01.bmp	69,22%	59,85%	9,37%	34,637
old02.bmp	66,20%	55,95%	10,26%	32,684
old03.bmp	66,12%	55,59%	10,53%	31,314
old04.bmp	64,06%	52,62%	11,44%	29,382
rv01.bmp	76,48%	69,71%	6,76%	36,395
rv02.bmp	73,83%	65,88%	7,95%	32,802
rv03.bmp	71,44%	62,55%	8,89%	35,439
rv04.bmp	71,45%	63,11%	8,34%	36,717
rv05.bmp	63,52%	52,90%	10,61%	34,724
rv06.bmp	70,94%	62,96%	7,97%	39,625
sky01.bmp	74,14%	66,18%	7,95%	35,404
sky02.bmp	73,40%	65,48%	7,92%	34,807
trave01.bmp	69,67%	60,36%	9,31%	34,212
MÉDIAS	**71,73%**	**63,09%**	**8,63%**	**34,524**

Com as informações da Tabela 1, pode-se observar que a quantidade de dados adicionais inseridos pela utilização da Tabela de Índice de cores é bem abaixo do esperado, numa média de 8,63%, já considerando que esta passou pela etapa de compressão sem perdas. Isso mostra a possibilidade de se adquirir uma boa taxa de compressão (média de redução de 63,09%), com a vantagem de que agora é possível reverter o anáglifo para o par estéreo original. O PSNR obtido cujo valor foi de 34, 524 dB foi medido comparando-se o par estéreo original e o obtido pela reversão anaglífica. A qualidade visual em todas as imagens se

mostrou boa, sendo inclusive possível utilizá-las para gerar um novo anáglifo, sem a perda de percepção de profundidade, como pode ser visto na Figura 11.

O processo de conversão e reversão anaglífica bem como os resultados obtidos foram condensados em um artigo com o título "*Reversing Anaglyph Videos Into Stereo Pairs*", submetido ao XVII Simpósio Brasileiro de Sistemas Multimídia e Web – WebMedia 2011, tendo sido aprovado para apresentação e posterior publicação. Mais detalhes sobre os resultados obtidos podem ser vistos no artigo, que se encontra ao final deste trabalho no APÊNDICE A.

(A)

(B)

Figura 11 - Comparação qualitativa do anáglifo verde-magenta obtido a partir do par estéreo original (A) com o obtido a partir do par estéreo revertido (B). Figura utilizada da base de teste com ID old01.bmp

6. Metodologia de Trabalho

6.1. Limitações da técnica criada

Como citado na Seção 5.3, o processo de conversão e reversão anaglífica utilizando a Tabela de Índice de Cores mostrou possuir resultados bastante positivos e com baixo acréscimo de informações ao arquivo comprimido. Entretanto, este processo precisa ser refinado em busca de resultados ainda melhores em relação à qualidade subjetiva e objetiva do arquivo revertido. Por exemplo, no par estéreo revertido, é perceptível a presença de *crosstalk*. Tal efeito é mais visível nas bordas dos elementos. Isso se deve a estarmos utilizando dados de luminância do anáglifo principal para reconstruir o complementar, uma vez que estes não são exatamente iguais para os dois, devido ao deslocamento presente entre as duas imagens que formam o par estéreo, ou seja, os dados de paralaxe positiva e negativa.

Os próximos passos do Mestrado serão guiados visando tal refinamento. Para isso, serão estudadas formas de como melhorar o PSNR obtido, estratégias para eliminar ou suavizar a presença de *crosstalk*, bem como realizar testes em uma base de dados maior e com o envolvimento de mais pessoas, tendo em vista obter uma avaliação subjetiva mais completa.

Nas próximas subseções são dados mais detalhes dos procedimentos a serem seguidos, bem como é apresentado o cronograma das atividades a serem desenvolvidas até o término do Mestrado.

6.2. Melhoria de PSNR

Nos resultados obtidos e apresentados na Seção 5.3, o PSNR se mostrou baixo, apresentando o valor de 34,524 dB numa escala de 0 a 100 dB. Entretanto, em uma análise subjetiva, as imagens se mostraram de boa qualidade visual. Mesmo assim, o PSNR é um bom indicador quando utilizado para fazer comparação e análise da técnica proposta em

relação a outras técnicas de compressão disponíveis. Por isso, melhorar seu resultado é importante e pode ser conseguido.

No processo de conversão e reversão anaglífico mencionado na Seção 5.2 e ilustrado nas Figuras 9 e 10, há uma etapa de mudança de espaço de cores do RGB para YC_bC_r e vice-versa. Tal mudança envolve uma fórmula matemática aplicada a cada pixel que resulta em valores de ponto flutuante. O armazenamento destes valores em ponto flutuante acarreta em um aumento expressivo do arquivo final e, portanto, compromete a compressão desejada. Dessa forma, é necessário truncar tais valores para serem armazenados em variáveis de dados que utilizem menos espaço de armazenamento. Isso resulta em perda tanto da precisão quanto dos valores que sejam maiores do que o limite permitido pela variável. Uma hipótese a ser estudada é se tal truncamento é uma das causas do baixo valor de PSNR (outra causa é presença de *crosstalk*, citada na Seção 6.1). Para isso, devem ser estudadas novas estratégias e estruturas de dados que consigam armazenar mais valores e com mais precisão, buscando encontrar uma que resulte em um bom balanço entre o PSNR e a taxa de compressão.

6.3. Análise de correlação de imagens

A Figura 12 mostra a comparação de qualidade visual de uma imagem estéreo sem compressão (Figura 12 (A)) e sua correspondente após passar pela reversão anaglífica utilizando a Tabela de Índice de Cores (Figura 12 (B)). Pode-se observar boa qualidade visual em (B), com algumas imperfeições, notadamente encontradas nas bordas de alguns elementos, correspondendo a regiões de paralaxe positiva mais acentuada. Tais imperfeições aparecem como regiões duplicadas, o *crosstalk*. Observando os contornos do trem e a copa da árvore ao fundo na Figura 12 (B) é possível notar a presença das cores magenta clara e verde no lado esquerdo da figura. O lado direito possui menos imperfeições, sendo mais notável a presença das cores verde e branca.

O par estéreo é formado por imagens semelhantes, deslocadas uma da outra pela distância do dispositivo de captura, de modo a simular o sistema visual humano. Este deslocamento se encontra presente nas componentes Y, C_b e C_r de cada anáglifo. Como estamos utilizando a componente Y do anáglifo principal para reconstruir o complementar (ver Figura 10), tais deslocamentos são também incorporados a este. Essa é a razão do

crosstalk no par estéreo revertido. Tal efeito afeta não somente a qualidade visual, como também o resultado do PSNR.

Para eliminar o *crosstalk*, uma estratégia é estudar a aplicação da correlação de imagens (Seção 3.2.4). Como o deslocamento aparece apenas em certas regiões do par estéreo, seria utilizada uma janela de busca a fim de achar os pontos que se encontram em posições diferentes em cada imagem, com relação à componente de luminância Y. Seria então calculado o quanto cada ponto se encontra deslocado e armazenado os valores encontrados. Na etapa de reversão anaglífica, esses valores de deslocamento seriam utilizados para replicarmos os dados de um ponto na posição correta.

Para esta parte do trabalho, faz-se então necessário um estudo de pesquisas relacionadas para encontrar o estado da arte, para depois incorporá-la ao processo de conversão e reversão anaglífica e analisar os novos resultados.

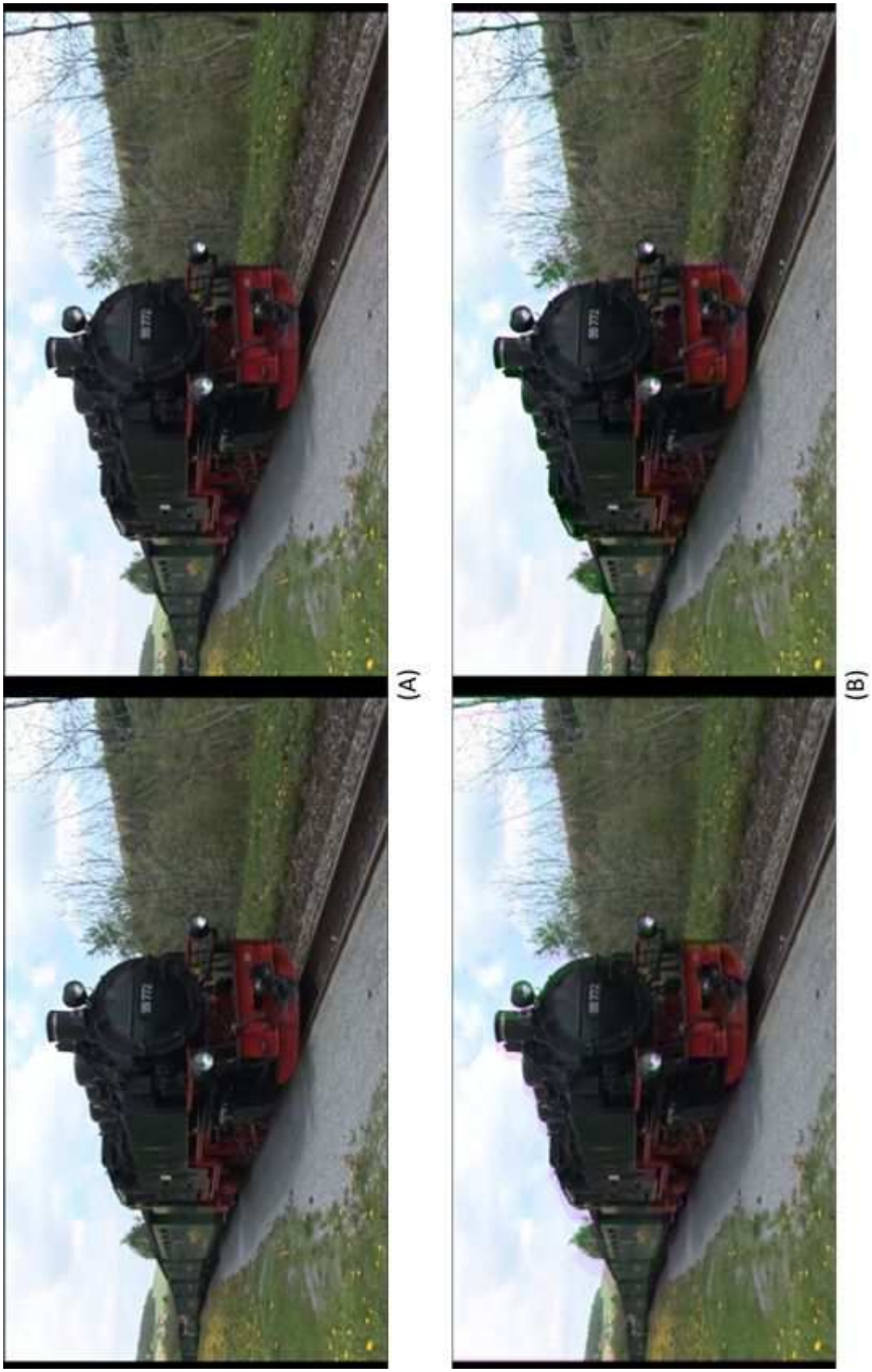

Figura 12 - Comparação qualitativa do par estéreo original (A) e o obtido pelo processo de reversão anaglífica com o uso da Tabela de Índice de Cores (B). Figura utilizada da base de teste com ID old01.bmp

6.4. Avaliações objetiva e subjetiva

A avaliação objetiva continuará sendo feita através do cálculo do PSNR. O cálculo é feito utilizando-se a versão gratuita do programa chamado MSU *Video Quality Measurement Tool* (VMQT)[8]. Este programa contém uma fórmula otimizada para o cálculo do PSNR, fornecendo valores individuais para cada componente, tanto no espaço de cores RGB quanto no YC_bC_r, além de fornecer uma imagem com as diferenças encontradas nas imagens comparadas, servindo como uma boa referência visual.

Para a avaliação subjetiva, os vídeos serão apresentados utilizando-se o método DSCQS e os dados serão analisados através do MOS (WINKLER, 2005) para o cálculo da média das notas dadas por avaliadores em uma sessão de testes. Com isso, é necessário o envolvimento de usuários reais. Durante o Mestrado, serão abertos chamados no Instituto para conseguir usuários voluntários a realizar a sessão de testes. Além disso, os professores do grupo de pesquisa ministram aulas de Multimídia e de Interação Usuário-Computador para os cursos de Graduação e Pós-graduação do ICMC-USP, nos quais temas como métodos de avaliação de qualidade e percepção humana são comuns. Assim, pretende-se realizar as avaliações também com os alunos desses cursos.

A base de dados citada na Seção 5.3, continuará sendo utilizada. Para a avaliação objetiva, será estendido o número de amostras a serem avaliadas, visando cobrir um número maior de resultados e imperfeições a serem analisadas.

Espera-se que com a avaliação subjetiva seja possível identificar pontos de falha no vídeo e imagem e ao mesmo tempo medir a severidade das possíveis falhas com os testes objetivos. Esse processo dará subsídios para análises das possíveis causas das falhas, o que poderá incentivar novas pesquisas. Ainda, os testes objetivos e subjetivos a serem aplicados possibilitarão avaliar se o processo de reversão afetou a qualidade do vídeo em relação ao vídeo original – e o quanto afetou –, assim como possibilitará medir o quanto a percepção de profundidade foi afetada e se isso constitui um problema real para visualização por parte dos usuários.

[8] O software MSU VQMT pode ser baixado em http://compression.ru/video/quality_measure/vqmt_download_en.html#free. Acesso em 28 jul. 2011.

6.5. Cronograma

Segue abaixo a proposta de atividades a serem realizadas. A Tabela 2 contém as atividades divididas nos períodos em que serão desenvolvidas.

1. Qualificação do Mestrado.
2. Análise contínua da literatura: revisão de livros, artigos, teses e dissertações relacionados ao projeto via fontes de pesquisa confiáveis, envolvendo as áreas de codificação e compressão de imagens e vídeos estereoscópicos, processamento e correlação de imagens.
3. Estudo de novas estruturas de dados que ajudem na melhoria do PSNR, sem afetar a taxa de compressão obtida.
4. Estudo da correlação de imagens e criação do algoritmo visando remover ou atenuar as imperfeições encontradas nos resultados já obtidos com atividades realizadas.
5. Implementação das melhorias encontradas ao código já desenvolvido em atividades anteriores.
6. Elaboração, aplicação e análise de testes dos resultados obtidos.
7. Revisão do projeto e possíveis alterações. Com base nos testes obtidos, fazer correções necessárias e revisar as técnicas criadas e/ou utilizadas.
8. Submissão de artigos para conferências e periódicos da área. Durante o Mestrado, serão submetidos artigos com os resultados parciais ou finais do projeto para conferências e periódicos relacionados com a área de aplicação, tais como WebMedia e ACM Multimedia e ACM SAC. As datas de submissão na Tabela 2 são apenas estimadas.
9. Integração com a técnica de compressão em desenvolvimento em projeto de doutorado relacionado.
10. Defesa do Mestrado

Tabela 2 - Cronograma de atividades para a conclusão do Mestrado

	2011					2012							
A t i v i d a d e s	**Ago.**	**Set.**	**Out.**	**Nov.**	**Dez.**	**Jan.**	**Fev.**	**Mar.**	**Abr.**	**Mai.**	**Jun.**	**Jul.**	**Ago.**
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													

6.6. Considerações finais

Os resultados deste trabalho pretendem contribuir na área de compressão digital, em especial a compressão de imagens e vídeos estereoscópicos. A técnica apresentada é inovadora, pois se utiliza do modelo anaglífico para gerar grande compressão no volume de dados, e é pioneira na criação de uma técnica de reversão, até então não estudada. O arquivo comprimido gerado pode ser decodificado e utilizado pelos diferentes tipos de visualização estereoscópica atuais, possibilitando tanto independência quanto interoperabilidade na utilização da técnica por qualquer sistema de visualização.

Ao término do Mestrado, pretende-se obter uma técnica bem testada e que gere imagens e vídeo comprimidos e com boa qualidade. Pretende-se também divulgar os resultados em periódicos e revistas conhecidos da área. Por fim, pretende-se criar um software a ser disponibilizado para que qualquer usuário possa utilizá-lo para comprimir imagens e vídeos estereoscópicos utilizando da técnica desenvolvida.

Vale lembrar que esta é apenas uma parte do processo completo de codificação e compressão de vídeos. Mais compressão pode ser obtida em outras etapas do processo, tais como os aspectos envolvendo redundância temporal, fora do escopo do projeto do Mestrado.

Referências[9]

ADIKARI, A.B.B. et al. A H.264 compliant stereoscopic video codec. **Canadian Conference on Electrical and Computer Engineering**, Saskatoon, p. 1614-1617, may 2005. DOI:10.1109/CCECE.2005.1557292.

ANDRADE, L. A.; CORDEBELLO, P. D.; GOULARTE, R. **Construção de uma base de vídeos digitais estereoscópicos**. São Carlos: ICMC-USP, 2010. 35p. Relatório técnico. Disponível em: <http://www.icmc.usp.br/~biblio/BIBLIOTECA/rel_tec/RT_351.pdf>. Acesso em: 28 jul. 2011.

ANDRADE, L. A.; GOULARTE, R. Percepção Estereoscópica Anaglífica em Vídeos Digitais Comprimidos com Perda. **Proceedings of the XV Brazilian Symposium on Multimedia and the Web (WebMedia '09)**, New York, p. 226-233, 2009. DOI:10.1145/1858477.1858506.

ANDRADE, L. A.; GOULARTE, R. Uma Análise da Influência da Subamostragem de Crominância em Vídeos Estereoscópicos Anaglíficos. **Proceedings of the XVI Brazilian Symposium on Multimedia and the Web (WebMedia '10)**, [S.l.], p. 1-8, 2010.

AZEVEDO, E.; CONCI, A. **Computação gráfica**: teoria e prática. Editora Campus, 2003.

BALASUBRAMANIYAM, B.; EDIRISINGHE, E.; BEZ, H. An Extended H.264 CODEC for Stereoscopic Video Coding. **Proceedings of SPIE**, San Jose, p. 116-126, 2005. DOI:10.1117/12.587583.

CHAPMAN, N. P.; CHAPMAN, J. **Digital Multimedia**. 3rd ed. Wiley, 2004.

[9] De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

DODGSON, N. A. Autostereoscopic 3D Displays. **Computer**, [S.l.], v. 38, n. 8, p. 31-36, aug. 2005. DOI:10.1109/MC.2005.252.

FEHN, C. et al. An Evolutionary and Optimised Approach on 3D-TV. **Proceedings of International Broadcast Conference**, [S.l.], p. 357-365, 2002.

FEITOSA-SANTANA, C. et al. Espaço de cores. **Psicologia USP [online]**, v.17, n.4, p. 35-62, 2006. Disponível em: <http://www.revistasusp.sibi.usp.br/scielo.php?script=sci_arttext&pid=S1678-51772006000400003&lng=pt&nrm=iso>. Acesso em: 28 jul. 2011.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 3rd ed. Upper Saddle River: Prentice-Hall, 2008.

HALLE, M. Autostereoscopic displays and computer graphics. **ACM SIGGRAPH 2005 Courses (SIGGRAPH '05)**, New York, p. 104-109, 2005. DOI:10.1145/1198555.1198736.

ITU-T. **Objective perceptual assessment of video quality**: Full reference television. Geneva: ITU-T – Telecommunication Standardization Bureau (TSB), 2004. Disponível em: <http://www.itu.int/dms_pub/itu-t/opb/tut/T-TUT-OPAVQ-2004-FRT-PDF-E.pdf>. Acesso em: 28 jul. 2011.

KERR, D. A. **Chrominance Subsampling in Digital Images**. 2009. Disponível em: <http://dougkerr.net/pumpkin/articles/Subsampling.pdf>. Acesso em 28 jul. 2011.

KIM, IH. et al. An embodiment of stereo vision system for mobile robot for real-time measuring distance and object tracking. **International Conference on Control, Automation and Systems**, Seoul, p. 1029-1033, oct. 2007. DOI:10.1109/ICCAS.2007.4407049.

LG ELECTRONICS. **LG Optimus 3D P920**. Apresentação de celular com tela 3D. Disponível em: <http://www.lg.com/uk/mobile-phones/all-lg-phones/LG-android-mobile-phone-P920.jsp>. Acesso em: 28 jul. 2011.

LI, S. et al. Stereoscopic Video Compression Based on H.264 MVC. **2nd International Congress on Image and Signal Processing, 2009 (CISP '09)**, Tianjin, p. 1-5, oct. 2009. DOI:10.1109/CISP.2009.5301218.

LIN, Z. et al. An Improved Stereo Video Coding Scheme Based on Joint Multiview Video Model. **First International Workshop on Education Technology and Computer Science**, Wuhan, p. 1091-1095, mar. 2009. DOI:10.1109/ETCS.2009.249.

LIPTON, L. **Foundations of the Stereoscopic Cinema**: a study in depth. New York: Van Nostrand Reinhold Company Inc., 1982.

LIPTON, L. Stereo-Vision Formats for Video and Computer Graphics. **Proceedings SPIE**, San Jose, p. 239-244, feb. 1997. DOI:10.1117/12.274462.

MENDIBURU, B. **3D Movie Making**: Stereoscopic Digital Cinema from Script to Screen. Oxford: Elsevier, 2009.

MERKLE, P. et al. Efficient Prediction Structures for Multiview Video Coding. **IEEE Transactions on Circuits and Systems for Video Technology**, [S.l.], v. 17, n. 11, p. 1461-1473, nov. 2007. DOI:10.1109/TCSVT.2007.903665.

NINTENDO OF AMERICA INC. **Nintendo 3DS**. Apresentação de videogame com tela estereoscópica. Disponível em: <http://www.nintendo.com/3ds/hardware>. Acesso em: 28 jul. 2011.

RICHARDSON, I. E. G. **H.264 and MPEG-4 Video Compression** – Video Coding for Next-generation Multimedia. West Sussex: Wiley, 2003.

SALOMON, D. **A Concise Introduction to Data Compression** (Undergraduate Topics in Computer Science). London: Springer, 2008.

SAYOOD, K. **Introduction to Data Compression**, 3rd ed. San Francisco: Elsevier, 2005.

SMOLIC, A. et al. An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic-solution. **Picture Coding Symposium**, Chicago, p. 1-4, may 2009. DOI:10.1109/PCS.2009.5167358.

SONY CORPORATION. **Sony 3D TV Technology**. Apresentação da tecnologia de televisores 3D da Sony. Disponível em: <http://www.sony.net/united/3D/#technology/3dtv/>. Acesso em: 28 jul. 2011.

STEREOGRAPHICS CORPORATION. **Stereographics® Developers' Handbook**: background on creating images for CrystalEyes® and SimulEyes®, [S.l.], Stereographics Corporation, 1997.

TAM, W. J.; ZHANG, L. 3D-TV Content Generation: 2D-to-3D Conversion. **IEEE International Conference on Multimedia and Expo**, Toronto, p. 1869-1872, jul. 2006. DOI:10.1109/ICME.2006.262919.

VETRO, A. Representation and Coding Formats for Stereo and Multiview Video. **Studies in Computational Intelligence**, [S. l.], v. 280, p. 51-73, 2010. DOI: 10.1007/978-3-642-11686-5_2.

EBRAHIMI, F.; CHAMIK, M.; WINKLER, S. JPEG vs. JPEG2000: An Objective Comparison of Image Encoding Quality. **Proceedings of SPIE Applications of Digital Image Processing**, Denver, v. 5558, n. 300, p. 300-308, 2004. DOI: 10.1117/12.564835.

WINKLER, S. **Digital Video Quality**: vision model and metrics. West Sussex: Wiley, 2005.

APÊNDICE A – Artigo submetido e aprovado para o WebMedia 2011

Reversing Anaglyph Videos Into Stereo Pairs

Matheus Ricardo U. Zingarelli
Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Av. Trabalhador São Carlense, 400
São Carlos, São Paulo, Brasil
zinga@icmc.usp.br

Leonardo Antonio de Andrade
Universidade Federal de São Carlos
Departamento de Artes e Comunicação
Rod. Washington Luiz, km 235
São Carlos, São Paulo, Brasil
landrade@ufscar.br

Rudinei Goularte
Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Av. Trabalhador São Carlense, 400
São Carlos, São Paulo, Brasil
rudinei@icmc.usp.br

ABSTRACT
There is a diversity of strategies for stereoscopic video coding, commonly known as 3D videos. Each strategy focuses on one type of 3D visualization format, which may lead to incompatibility or depth perception problems if one strategy is applied to another format. A generic approach may be based on the proper coding of the stereo pair – every visualization format knows how to deal with stereo pairs – however, to keep the stereo pair, even a coded one, demands a high data volume. This paper proposes a method for reversing an anaglyph video back into a stereo pair. This way, a coder may convert a stereo pair into an anaglyph video, reducing the data volume by half, at least. Moreover, a correspondent decoder may use the proposed reversing method to restore the stereo pair, ensuring the visualization independence of the coding method. Tests showed that both conversion and posterior reversion resulted in videos with good quality obtaining an average of 63.09% of compression and 34,52dB of PSNR.

Categories and Subject Descriptors
E.4 [**Coding and Information Theory**]: Data compaction and compression; I.4.2 [**Image Processing and Computer Vision**]: Compression (Coding) – *approximate methods*

General Terms
Algorithms, Performance, Standardization.

Keywords
Anaglyph video, stereo video coding, stereoscopy, digital video coding.

1. INTRODUCTION
Stereoscopic videos, commonly known as 3D videos, are formed by a pair of videos – called stereo pair (right eye, left eye) – and are reproduced in a way that gives a depth perception for a person watching them, mimicking the human stereo vision [4] (Section 2). Over the last few years, there's been an increased boost of 3D content production by the movie industry, largely due to the acceptance and expression of public interest for this technology. Besides that, 3D technology is being gradually incorporated at homes in forms of 3D television [15], cell phones [8] and video games [12] with each device supporting different kinds of visualization. Consequently, new techniques for capturing, coding and playback modes of stereoscopic videos are emerging or being improved in order to optimize and integrate this technology with the available infrastructure.

In the stereo video content production field, new cameras were developed for recording two views of the same scene, in order to produce the stereo pair, with the possibility of also generating a depth map of the scene that can be used to create new views [7]. There are also techniques developed for converting 2D content into 3D [18]. In the field of visualization, we have techniques for viewing stereo content with the support of specific glasses – anaglyph stereoscopy [9], polarized light [9] and shutter glasses [16] –, and also autostereoscopic displays, or glasses free, that allow us to view 3D content without wearing any type of glasses or specific devices [14].

In spite of the advances made in the field of visualization and representation of stereoscopic videos, it's noticeable that advances in the coding field are slower. On one hand, we have Lipton's Method [10], which describes formats for stereoscopic video representations, being the stereo pair stored in a single video container, without compression, with double of data than a regular 2D video stream. Since Lipton's Method keeps the stereo pair, it can be used by any visualization system – it is generic. On the other hand, we have what we call "adapted methods" that use well-known compression techniques, like MPEG-2 or H.264, to reduce the amount of data to be transmitted. However, such techniques are only adapted to work on stereoscopic videos, lacking of a standard compression technique specific designed for this kind of video. Moreover, each adapted method is designed for a particular visualization system, which brings two problems: they may not be compatible for all formats and types of stereo visualization [14], and since lossy compression is used, depth perception may be compromised in some cases, especially with anaglyph videos [2][3]. With that said, one can realize that there's a lack of a generic method specific for coding stereoscopic videos, compatible with different types of 3D visualization and providing good quality without loss of depth perception.

Related work has been done looking for compression of stereo video pairs without significant quality loss. Results like [3] demonstrate that it's possible to develop coding methods to reduce the data volume having no depth perception loss and, more important, being independent of the visualization method. In spite of the reduction of data volume achieved by those coding methods, the compression rate still remains low since they keep the stereo pair. A straightforward solution is the conversion of the stereo pair into a single anaglyph video stream (Section 4.1), resulting in a stereo video with higher compression rate, superior quality and, obviously, compatible only with the anaglyphic visualization method.

This way, in order to keep the generality of such coding methods, making them also compatible with other kinds of visualization, anaglyph to stereo pair reversion methods are needed. A reversion method is not trivial though, since the anaglyph conversion causes loss of color components in the stereo pair, and they must be retrieved somehow. This paper demonstrates that this reversion is possible and proposes a technique based on a compressed chroma sub-sampled color index table. With this technique we were able to achieve 63.09% of compression and to create a reversed stereo pair with no loss in depth perception.

This paper is organized in the following sections. Section 2 presents details about stereo vision and depth perception needed for a better understanding of the proposed technique. Section 3 describes related works about stereo video coding and formats. Section 4 presents the technique we propose to tackle the problem of reversing an anaglyph video. Section 5 presents the experimental tests and results of using our technique. Finally, in Section 6 we present our conclusions and future work.

2. STEREO VISION AND DEPTH PERCEPTION

Our eyes are approximately 6.5 cm distant from each other, move together in the same direction and each one has a limited viewing angle. By presenting themselves in different positions, each eye sees a slightly different image [4]. For these reasons it was expected that when we look at an object, we would saw two images and not just one. However, the brain takes charge of calculating information from relative distances from objects and to interpret these two images, resulting in production of a single image, phenomenon known as stereoscopy. The main stereoscopic information are stereopsis, (binocular) disparity and parallax [16].

The stereopsis is responsible for the depth sensation that we have between objects and is obtained due to binocular disparity. Thus, the mandatory requirement to obtain stereopsis is to use both eyes. With this information we feel objects closer or farther away. It is explored in 3D movies to give us the impression that objects are "bouncing off the screen" [16].

The binocular disparity is the difference in distance between the positions of the image formed on each retina. This is best understood through the following example: observe an object in front of you and place your thumb between your eyes and the object. When we focus on the thumb, i.e., it is at the point of convergence of the two retinas, the object is past the point of convergence (farther), appearing as doubled (Figure 1 (A)). This happens because the images off the focal point are being formed at different locations in each retina. The disparity is the distance between these two duplicate images. The same happens if we put our focus on the object (Figure 1 (B)).

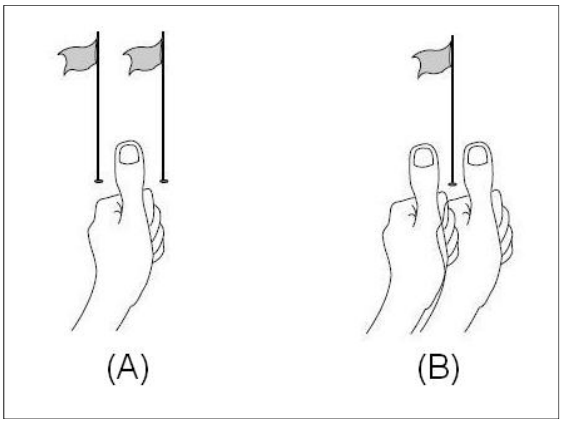

Figure 1 – Retinal Disparity [16].

Directly related to the disparity concept (obtained in the retinal image) we have the parallax [16][11], which is the distance between corresponding points in the images projected on a monitor. With the values of parallax, it is possible to give a different point of view of the same image to each eye, resulting in the formation of the disparity, and it therefore has the stereopsis effect. An easy way to calculate the parallax between two points is superimposing an image to another and measuring the distance between the same points in each image. It is because of parallax that, for example, when watching an anaglyph video without the proper glasses, we see parts of the image as duplicated and overlapped.

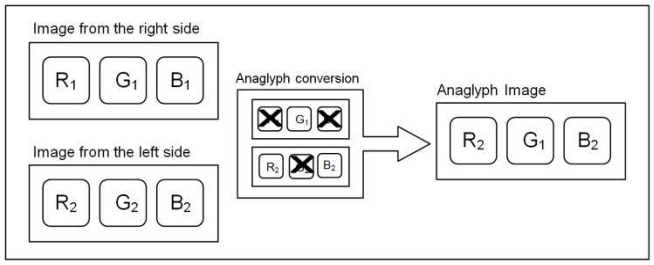

Figure 2 - Anaglyph Conversion to create a green-magenta anaglyph.

This way, stereoscopy is based on methods that present to an observer a pair of flat (2D) images from the same object, each one providing each eye with a slightly different perspective – the stereo pair. The stereo pair causes retinal disparity, then stereopsis, which, in turn, provides depth perception. As previously mentioned at Introduction, there are four main stereoscopic visualization methods: anaglyphic, polarized light, shutter glasses and autostereoscopic. The focus of this work is the anaglyphic method, which will be briefly described in the following.

The anaglyphic method is an important low cost and simple way of coding a stereoscopic pair to be visualized using proper glasses. In this method, we simply remove some color components from each image and then merge them into a single one. Afterwards, by using glasses with lenses that mimics the color components removed, thus acting as filters, images are separated again and each eye sees only one of them, leading to the binocular disparity and resulting in the stereoscopic effect.

Figure 2 illustrates the anaglyph conversion. The RGB channels of each image are combined in a way to have information from both images (the left eye and the right eye) [11]. Let the stereo pair be formed by $R_1G_1B_1$ (the right eye image) and $R_2G_2B_2$ (the left eye image). The conversion takes advantage of the green channel (G_1) of the right eye image and channels R_2 and B_2 of the left eye image, forming a third image (the anaglyph image) with information of both images of the stereo pair: $R_2G_1B_2$. As the combination of red (R_2) and blue (B_2) results in the magenta color, the anaglyphic method illustrated in Figure 2 is known as the green-magenta. Other possible methods are: red-cyan ($R_1G_2B_2$) and blue-yellow ($R_2G_2B_1$).

Dubois [6] and Andrade & Goularte [3] reported that green-magenta method presents better results for anaglyphic visualization than the other methods. Therefore, in this work, we use the green-magenta color combination.

Besides being simple, the anaglyphic method also doesn't require expensive or complex equipment to be executed or visualized, and the compression rate obtained is high, since only one video stream is transmitted/stored. The other visualization methods may have to keep the stereo pair, i.e., two video streams. Such advantages show that the anaglyphic method can be a potential candidate to be used in a generic stereoscopic coding process. Instead of sending the stereo pair, we could transform it into an anaglyph, reducing the data to be transmitted by half. This, however, comes with a cost: we are losing color components, and as mentioned in Section 1, that leads to a problem of recreating the stereo pair using a reversion method.

3. RELATED WORK

The MPEG group has well-known standard methods for video codification, even with extensions for stereoscopic videos, like MPEG-2 Multiview Profile [14]. However, there is no standard codification method specific for stereoscopic videos only. With that, different authors have created different coding strategies, each designed to attend requisites of one or another type of 3D system, which makes the implementation device-dependent and may result in incompatibility of content between different systems.

Among the different strategies, we can cite Lipton's Method [10] which describes several ways for presentation of a stereo pair of videos with concern of having little or no modifications in the hardware already available. They can be classified into two groups: field sequential scheme, in which right and left fields are alternated in sequence and the user, with proper glasses that synchronizes with the display, views only one of them on each eye. And pixel sequential scheme, with the above-and-below and side-by-side formats, in which the right and left subfields are united in a single field either horizontally or vertically.

Smolic et al. also stated in [14] the diversity of stereoscopic video formats, each directed to a specific system, thus requiring different implementations and structures. The authors classify these formats based on the number of video signals (called views), order of complexity and types of data involved, resulting in six classes. Conventional Stereo Video (CSV) is the simplest one, similar to Lipton's Method. Multiview Video Coding (MVC) is an extension to when more than two views are used. Video plus depth (V+D) is a format with more complexity, in which a depth map is sent together with the video signal to create the stereo pair, also enabling the possibility to generate a limited number of other views. These first three classes can be implemented using available video codecs. For advanced video applications like autostereoscopic televisions, the next three classes are used.

Multiview plus depth (MVD) is a combination of MVC and V+D properties, which means that multiple views and multiple depth maps are sent. The next format is layered depth video (LDV), in which besides a video signal and depth map, it is also sent a set of layers and associated depth maps used to generate virtual views. Finally, the last format is called depth enhanced stereo (DES), proposed by the authors as a generic 3D video format. It's an extension of the CSV, with the additional of depth maps and layers, providing compatibility among different formats, since each one uses only the types of data needed. Even though DES is designed to be a generic format, there are two major drawbacks that need a deep study: depending on which format data will be represented, it may be necessary the storage of a great amount of data to hold both video signal and additional depth maps and layers. We also have an increase in the system's complexity and errors that may arise from depth calculations.

Notice that in Lipton's Method both videos from the stereo pair are stored, resulting in a video file twice as big as a normal 2D video file, while Smolic et al. describe formats in which the stereo pair are not necessarily needed, what we call "adapted methods". Some authors study compression techniques to reduce the amount of data to be transmitted in these adapted methods, either using well-known compression techniques or designing new ones [17]. Vetro [19] performed a survey over the different formats and representation of stereoscopic and multiview videos, with their corresponding compression techniques and several types of displays in which they can be visualized. This survey clearly demonstrates the challenge in creating a generic method for stereoscopic video representation and coding: each surveyed method has specific types of compression, representation, storage, and plays only specific types of media and displays (visualization methods).

Andrade & Goularte [2] [3] studied how the usage of well-known lossy compression techniques on stereoscopic videos might affect quality and depth perception in different types of stereo visualization. They showed that the data lost during compression compromises depth perception in the anaglyph stereoscopy, and discovered suitable parameters for color space reduction, used together with Wavelet transform and quantization, that could compress stereoscopic videos with good quality and no depth perception loss regarding the anaglyph stereoscopy. In that work authors demonstrate the viability of a generic stereoscopic coding method, however, their method stores a stereo pair of videos, lowering the compression rate.

Thus, through related work, one can conclude that there is a lack of stereoscopic coding methods that: a) are independent of visualization methods, making the coding process generic enough to be possible to achieve stereo depth perception by the means of any visualization method; b) achieve high compression rates preserving image quality and depth perception, while keeping the visualization independence.

4. THE ANAGLYPH REVERSION TECHNIQUE

The reversion process is not trivial because during the anaglyph video generation some information is discarded. The stereo pair has six color channels (Figure 3): three (R_1, G_1 and B_1) from the right eye video/image, and three (R_2, G_2 and B_2) from the left eye

video/image. When generating the anaglyph video, three channels are discarded, one channel from one video and two channels from the other video (at any combination). The challenge here is to recover lost information without significant compromise of depth perception and achieving good compression rates.

In order to recover the stereo pair from an anaglyph video, a first attempt is to store the discarded color information into some data structure, let's call it "Color Index Table". Following Figure 2 example, this table will be formed by color information from channels R_1, G_2 and B_1 and stored together with the anaglyph video $R_2G_1B_2$. This way, a decoder will have all the needed information to rebuild the stereo pair. In spite of this approach to keep color quality (it will preserve the color data), it does not present any compression advantage.

As only the color information is required in order to build the Color Index Table, a better approach is to use a color space conversion, from RGB to YC_bC_r [13]. This way, it is possible to separate luminance information (Y) from color information (C_b and C_r), using just C_b and C_r to compose the table. Moreover, C_b and C_r channels may be sub-sampled, reducing even more the amount of data needed to be stored in the table.

There are some possible color (sub) sampling combinations, presenting different tradeoffs between compression and color fidelity [13]. The 4:4:4 method is the best in quality, but the worst in terms of information reduction. The 4:1:1 method is exactly the opposite. Since colors greatly influence the anaglyphic method [2], Andrade & Goularte [3] have developed a study concluding that the 4:2:2 chrominance sub-sampling method offers a good tradeoff without affecting depth perception. Therefore, in this work, we use the 4:2:2 method.

The next three sections (4.1, 4.2 and 4.3) present how the anaglyph video is produced using chrominance sub-sampling in order to build the Color Index Table, how this table is used in order to reverse the anaglyph video back into a stereo pair, and a discussion about the method used.

4.1. The color index table

The Color Index Table is built following 4 steps, depicted in Figure 3: (I) creation of a green-magenta anaglyph video from the uncompressed stereo pair; (II) creation of another anaglyph with the remaining color components, which we call "complementary anaglyph"; (III) conversion of the complementary anaglyph from RGB to YC_bC_r color space using 4:2:2 sub-sampling and (IV) compression and storage of C_b and C_r components to create the table.

The production of a green-magenta anaglyph video from a stereo pair follows the scheme depicted in Figure 2. Each image from the stereo pair, named right image and left image, has its R, G and B color components separated. Then a new image is created, whose green color component is from the right image and red and blue (magenta) color components are from the left image ($R_2G_1B_2$). Doing this for every frame will result in the anaglyph video. Here, for the sake of clarity, we will explain the entire process using images as examples.

Figure 3 illustrates the creation of the green-magenta anaglyph image, which is kept apart meanwhile, and a complementary anaglyph image. This last one is generated by the color components not taken at first, that is, the red and blue color components from the right image and the green one from the left image ($R_1G_2B_1$).

Then, the complementary anaglyph is converted from RGB to YC_bC_r color space, using Equation 1 [5] to calculate color conversion. Equation 1 is an ITU-T recommendation.

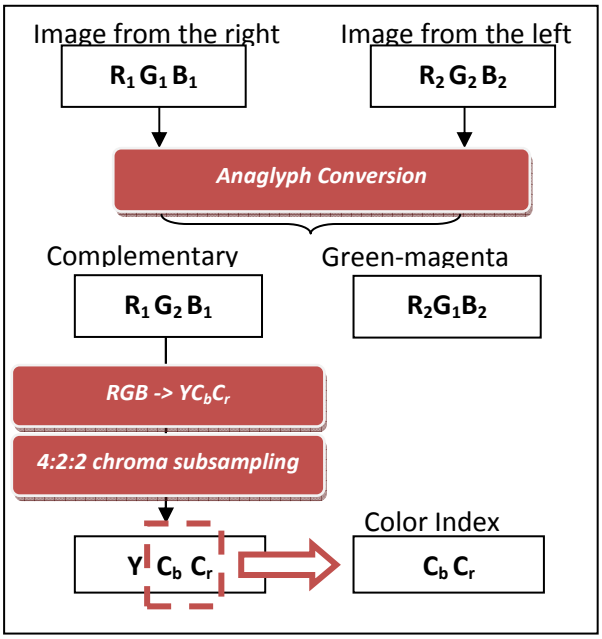

Figure 3 – Anaglyph conversion of a stereo pair of images and creation process of the Color Index Table

```
Y  ← 0.299 * R + 0.587 * G + 0.114 * B
Cb ← (B - Y) * 0.564 + delta
Cr ← (R - Y) * 0.713 + delta,

                {128 for 8-bit images,
where delta =   {32768 for 16-bit images,
                {0.5 for floating-point images
```

Equation 1 – RGB to YC_bC_r conversion

The result is a 4:4:4 YC_bC_r image, and the complementary anaglyph is discarded. The 4:4:4 YC_bC_r image has its Y component discarded and its C_b and C_r components are 4:2:2 sub-sampled. It means that, for each 12 pixel samples (4:4:4) of the original image, 8 were discarded: 4 from the Y component, 2 from the C_b component and 2 from C_r component. This two sub-sampled components form the Color Index Table (Figure 3), which will be stored together with the anaglyph video. It should be noticed that, in spite of the Color Index Table being an overhead, it has only 33% of the complementary anaglyph data volume and, more compression will be achieved after applying Huffman lossless compression technique - Section 5 presents the results.

4.2. Anaglyph reversion

The anaglyph reversion consists of recreating the stereo pair using the Color Index Table and the anaglyph image, as depicted in figure 3. In order to obtain the stereo pair, we need the color components lost during the anaglyph conversion. These can be extracted from the Color Index Table by recovering the luminance component (Y) and applying a formula to convert from YC_bC_r to RGB. The luminance component can be calculated from the green-magenta anaglyph image using Equation 1. The conversion

from YC_bC_r to RGB color space can be done by using Equation 2 [5]. Again, this formula is an ITU-T recommendation. Since we sub-sampled the image during the anaglyph conversion, we need to first duplicate each chrominance component to every 4 samples of luminance component and then apply Equation 2.

```
R <- Y + 1.403 *(Cr - delta)
G <- Y - 0.344 *(Cr - delta) - 0.714 *(Cb - delta)
B <- Y + 1.773 *(Cb - delta),

                {128 for 8-bit images
where delta =   {32768 for 16-bit images
                {0.5 for floating-point images
```

Equation 2 – YC_bC_r to RGB conversion

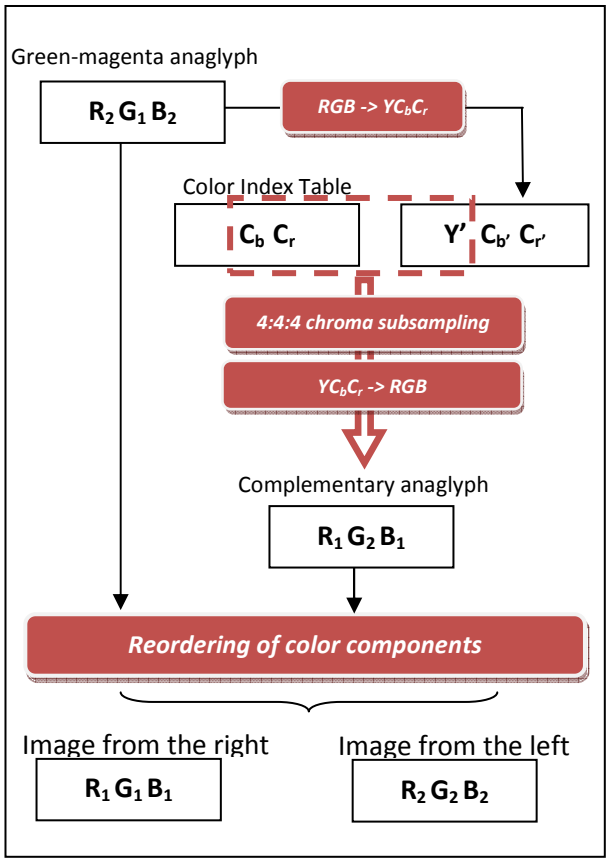

Figure 4 - Anaglyph reversion using a Color Index Table

Now that we have the missing color components, we can recreate the complementary anaglyph and then just need to reorganize the color components pertinent to each image of the stereo pair. It means that we take the red and blue color components from the green-magenta anaglyph and the green from the complementary anaglyph to recreate the left-eye image of the stereo pair. Likewise, with the green color component from the green-magenta and the red and blue from the complementary, we are able to recreate the right-eye image. Figure 4 summarizes the reversion process.

4.3. Method discussion

An important question, worth to discuss here, is the reason behind the use of the complementary anaglyph in order to build the Color Index Table.

The goal of the index is to store color information lost during the anaglyphic conversion. Since color information may take space, our approach is based on using chrominance sub-sampling in order to reduce the amount of data needed to be stored.

A first method one may think, trying to keep color quality at the best possible, is to apply the YC_bC_r conversion and sub-sampling directly to the stereo pair storing the chrominance components. This method will store 4 chrominance components together with the anaglyph image/video, instead of the two in our approach. In addition, in the reversion phase (Section 4.2), the Y component will have to be restored based on the anaglyphic version of the stereo pair (the stereo pair does not exist anymore), which has different information from those originally used to generate Y. This will result in distortions in the recovered image/video.

A second attempt is not to use a complementary anaglyph and to apply the YC_bC_r conversion and sub-sampling to the anaglyph image/video, generating the Color Index Table as described in Section 4.1. At the reversion phase the Y component may be properly restored based on the anaglyph image/video (Section 4.2) since it was also generated based on the anaglyph image/video. However, the C_b and C_r color components retrieved in this way have color information coded mostly from just one of the images in the stereo pair. For example, C_b and C_r may represent colors from a $R_2G_1B_2$ anaglyph image, which means there is no color information about R_1, G_2 and B_1 components in order to properly recreate the stereo pair. This will also result in distortions in the recovered image/video.

By using the proposed approach (Figure 3), the anaglyph image has color information that came from three of the six stereo pair color channels: R_2, G_1 and B_2, and the Color Index Table (C_b and C_r sub-sampled color components) carries color information obtained from the complementary anaglyph ($R_1G_2B_1$ in Figure 3). So, as we have color information that came from all the six RGB color channels of the stereo pair, we have conditions to properly recreate the stereo pair. This is done, as explained at Section 4.2, applying RGB-to-YC_bC_r conversions in order to recreate the complementary anaglyph image.

This way, we have R_1, G_2 and B_1 color channels from the complementary anaglyph and R_2, G_1 and B_2 color channels from the anaglyph image. Reordering the color channels, we get back the six channels of the stereo pair without depth distortions in the recovered image/video.

5. EXPERIMENT AND RESULTS

In our experiment, we focused on evaluate quantitatively the quality of stereo images outputted by our anaglyph reverse technique, based on 3 criteria: brightness, saturation and contrast, summarized in Table 1. Brightness has to do with intensity levels of luminance. Since each image from the stereo pair has a different point of view of a same scene, depending on the environment that they were captured, different intensity of brightness may appear. Saturation means the purity of a color, i.e. how much of white light there is in this color, where low saturation means higher amount of white light and vice-versa. Last criterion is contrast, the difference between adjacent colors in the image. The more two colors are different, the higher is the contrast, which allows better visualization in the details of an image. We've used a set of 32 stereo images classified between these criteria. Since the criteria are not mutually exclusive, one of more of them may appear in the same image. Images were extracted from a test database of stereoscopic videos created in [1]

[3]. The database is available online and can be visited in http://200.136.217.194/videostereo/.

Our analysis on the quality of the stereo images reverted from our technique were based on calculating the PSNR – Peak Signal-to-Noise Ratio – of each pair of images: the original and the reverted one. The PSNR is a metric widely used to evaluate how similar is an image compared to another one [20]. It makes a pixel-by-pixel comparison and returns a value measured in decibels (dB), in the range of 0 to 100, with 0 meaning no similarities and 100 meaning total similarity.

Each PSNR was calculated using a free verion of a software called MSU VQMT[10] (Video Quality Measurement Tool). This software implements several evaluation metrics for image and video assessment, PSNR being one of them. For each criteria, we've calculated the PSNR of each color component of a pair of image in the RGB color space and then the average of the three results.

Table 1. Criteria used in the evaluation of the anaglyph reversion technique

Criterion	Types of images
1. Brightness	Images with high levels (brighter) or low levels (darker) of luminance
2. Saturation	Images with the presence of one predominant color and different levels of purity.
3. Contrast	Images with great or little variety of colors.

5.1. Results

For brightness, we've tested 18 images with high, medium and low levels of luminance. The average of PSNR among all images was 33,70dB, with a maximum of 37,39 dB belonging to an image of high luminance contrast – very brighter in some regions and darker in others –, and a minimum of 29,38 dB belonging to an image with high levels of luminance and little contrast. Difference in the values of PSNR between R, G and B color components on each image stayed on the range of 2,55 dB. The average of the compression rate was of 63.12%. Figure 5 shows the PSNR values obtained for each image.

For saturation, we've tested 20 images with different levels of color purity for the predominant colors. The average of PSNR among all images was 38,43 dB, with a maximum of 39,62 dB belonging to an image of medium contrast and presence of saturation levels of green and brown colors, and a minimum of 30,28 dB belonging to an image with the predominance of color green in different levels of saturation. Difference in the values of PSNR between R, G and B color components on each image stayed on the range of 2,55 dB. The average of the compression rate was of 63.61%. Figure 6 shows the PSNR values obtained for each image.

[10] Available at http://compression.ru/video/quality_measure/index_en.html

For contrast, we've tested 17 images with great, medium and little variety of colors. The average of PSNR among all images was

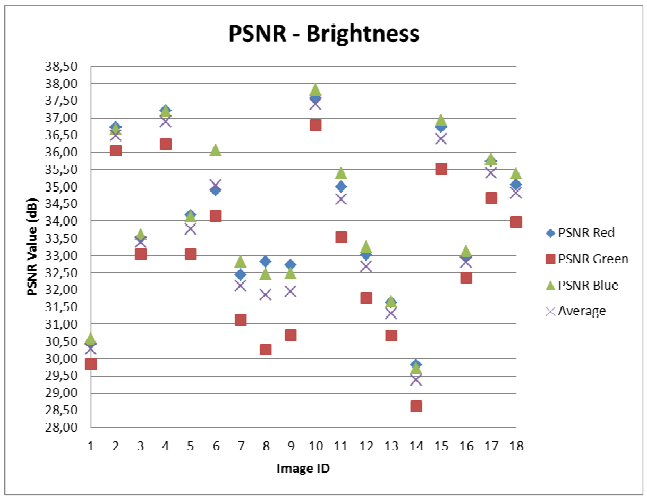

Figure 5 - PSNR values of images from the brightness criterion.

33,05 dB, with a maximum of 39,62 dB belonging to the same image tested for saturation, and a minimum of 29,38 dB belonging to the same image tested for brightness. Difference in the values of PSNR between R, G and B color components on each image stayed on the range of 2,05 dB. The average of the compression rate was of 64.17%. Figure 7 shows the PSNR values obtained for each image.

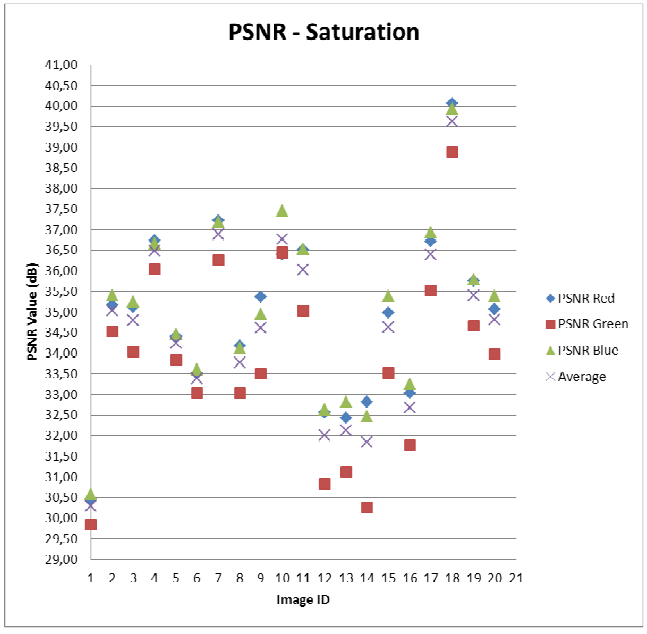

Figure 6 - PSNR values of images from the Saturation criterion.

From the results, we can observe that the PSNR value and the compression rate were similar in the three criteria defined, with a general average of 34,52 dB of PSNR and 63.09% of compression. We can also observe that the difference between the PSNR value of each RGB component of an image did not exceed 2,55 dB, which, accordingly to [2] is an acceptable value, in which the

depth perception is not affected – difference of values greater than 5 dB are prohibitive, since it affects depth perception.

5.2. Image quality and depth perception

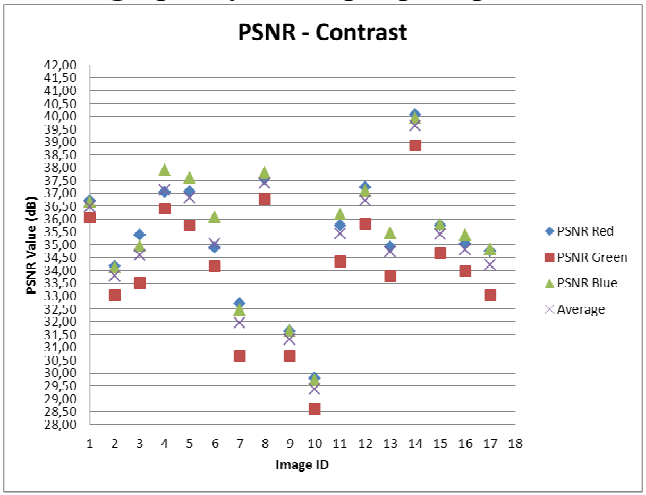

Figure 7 - PSNR values of images from the contrast criterion.

In our experiment, we took an anaglyph image and used our reversion technique to obtain a stereo image. This stereo image was compared to the original image and we calculated its PSNR. Even though the average of PSNR obtained was low, when comparing qualitatively both the reverted and original image, one can notice only a few differences between then, mainly caused by the pixel displacement that is present between the two images of the stereo pair. That means that even with a low PSNR value, the quality of the reverted image is still good. The PSNR value was affected by the conversion between RGB and YC_bC_r color spaces, which results in float values. Our technique was implemented in a way that each pixel is stored in an unsigned char variable, thus we have rounding errors involved in the process. As a future work, we will be reviewing our implementation, in order to achieve better PSNR results.

We took another final experiment regarding the reverted images obtained by our technique. Each one was converted into a green-magenta anaglyph and we compared qualitatively this anaglyph with the anaglyph formed by the original image. The anaglyph from the reverted image presented good quality and depth perception was not lost, which proves that the maximum difference of 2,55 dB of PSNR values between each RGB component obtained in our previous experiments does not affect depth perception.

6. CONCLUSION

In this work, we reported several formats available for stereoscopic representation and visualization, each one designed for a specific device or system. From that, we observed that the compression methods available for 2D videos could also be used by stereoscopic videos, but that could affect depth perception depending on the type of visualization technique used, highlighting the lack of a standard coding method exclusive for stereoscopic videos that would be generic and compatible among different visualization systems. We then showed that the anaglyphic is a stereoscopic visualization method that is simple to implement, does not require expensive or complex equipment to be visualized and can reduce data from a stereo video by half. Only if we could reverse it to its original stereo pair, the anaglyphic method would be a potential candidate for a generic stereoscopic compression process. Therefore, we proposed a technique for anaglyph reversal, with the creation of a Color Index Table that stores data from the color components discarded during the anaglyph process.

Our experiment showed that this reversion process is viable indeed. We were able to recreate a stereo pair of images from its respective green-magenta anaglyph with an average of 34,52 dB of PSNR and good quality when compared to the original one. The overhead of data with the addition of the Color Index Table in the anaglyph process was little, with an average reduction of 63.09% of the file size generated by it. Moreover, we found that the image resulted from the reversion process could still be transformed in anaglyph with no loss of depth perception.

The reversion technique involves converting images from RGB to YC_bC_r color spaces, which leads to rounding of floats values. That affected PSNR measurement. As a future work, we'll be restructuring our implementation in order to get better rounding values. We will also add more complexity in the technique to explore pixel displacement between the images of the stereo pair, in order to increase the quality of the reverted image.

7. REFERENCES

Andrade, L., Dolosic, P., Goularte, R. 2010. *Construção de uma Base de Vídeos Estereoscópicos*. Technical Report. ICMC-University of São Paulo, São Paulo, Brazil. Available at http://www.icmc.usp.br/~biblio/BIBLIOTECA/rel_tec/RT_351.pdf.

Andrade, L. A., Goularte, R. 2009. Anaglyphic stereoscopic perception on lossy compressed digital videos. In *Proceedings of the XV Brazilian Symposium on Multimedia and the Web* (WebMedia '09). Fortaleza, v.1, n.1, 226-233. DOI=http://doi.acm.org/10.1145/1858477.1858506.

Andrade, L. A., Goularte, R. 2010. Uma Análise da Influência da Subamostragem de Crominância em Vídeos Estereoscópicos Anaglíficos. In *Proceedings of the XVI Brazilian Symposium on Multimedia and the Web* (WebMedia '10).

Azevedo, E., Conci, A. 2003. *Computação gráfica: teoria e prática*. Campus, Elsevier, Brazil.

Bradski, G., Kaehler, A. 2008. *Leaning OpenCV*. O'Reilly, United States.

Dubois, E. 2001. A projection method to generate anaglyph stereo images. *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on* (May, 2001), 1661-1664. DOI= http://dx.doi.org/10.1109/ICASSP.2001.941256.

Fehn, C., Kauff, P., Op de Beeck, M., Ernst, F., IJsselsteijn, W., Pollefeys, M., Van Gool, L., Ofek, E., Sexton, I. 2002. An Evolutionary and Optimised Approach on 3D-TV. In *Proceedings of International Broadcast Conference*, 357-365.

LG autostereoscopic mobile phone, available at http://www.lg.com/uk/mobile-phones/all-lg-phones/LG-android-mobile-phone-P920.jsp.

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

Lipton, L. 1997. Stereo-Vision Formats for Video and Computer Graphics. *Proc. SPIE*, 3012, 239 (February. 1997). DOI=http://dx.doi.org/10.1117/12.274462.

Mendiburu, B. 2009. *3D Movie Making Stereoscopic Digital Cinema from Script to Screen*. Focal Press, Elsevier, Brazil.

Nintendo 3DS, available at http://www.nintendo.com/3ds/hardware.

Richardson, I. E. 2003. *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*. Wiley, England.

Smolic, A.; Mueller, K.; Merkle, P.; Kauff, P.; Wiegand, T. 2009. An Overview of Available and Emerging 3D Video Formats and Depth Enhanced Stereo as Efficient Generic Solution. *Proceedings of the 27th conference on Picture Coding Symposium, 2009* (May, 2009), 1-4, 6-8. DOI= http://dx.doi.org/10.1109/PCS.2009.5167358.

Sony 3D TV Technology, available at http://www.sony.net/united/3D/#technology/3dtv/.

StereoGraphics Corporation. 1997. *Stereographics® Developers' Handbook: Background on Creating Images for CrystalEyes and SimulEyes*.

Sumei, L., Chunping, H., Yicai, Y., Xiaowei, S., Lei, Y. 2009. Stereoscopic Video Compression Based on H.264 MVC. *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on* (October, 2009), 1-5, 17-19. DOI= http://dx.doi.org/10.1109/CISP.2009.5301218.

Tam, W. J., Zhang, L. 2006. 3D-TV Content Generation: 2D-to-3D Conversion. *Multimedia and Expo, IEEE International Conference on* (July 2006), 1869-1872. DOI= http://doi.ieeecomputersociety.org/10.1109/ICME.2006.262919.

Vetro, A. 2010. Representation and Coding Formats for Stereo and Multiview Video. In Studies in Computational Intelligence. Springer Berlin / Heidelberg, 51-73. DOI= http://dx.doi.org/10.1007/978-3-642-11686-5_2.

Winkler, S. 2005. *Digital Video Quality: vision model and metrics*. Wiley, England.