Capa [Ver pós se há um formato específico]

Subcapa [Ver pós se há um formato específico]

Resumo

Sumário

Índice de Figuras

Índice de Tabelas

1. Introdução

As produções 3D para o cinema surgiram por volta da década de 1950 (Lipton, 1982) como forma de resgatar o público perdido com a disseminação dos televisores domésticos. Seu sucesso, porém, foi por um período muito curto de tempo, principalmente devido a problemas técnicos tanto na captura quanto na reprodução, bem como a inexperiência de diretores e produtores em criar cenas que explorassem os aspectos de profundidade, fatos que causavam produção de filmes com baixa qualidade, muitos feitos apenas com o intuito se aproveitar da fama da novidade. Com isso, surgiu um longo período de poucas produções convertidas para 3D. O sucesso voltou novamente com a estreia de Avatar em 2009, cujo custo de produção foi alto, sendo um dos fatores a tecnologia desenvolvida para captura, tratamento, produção e reprodução de imagens digitais. Do lado da indústria, é também observável o avanço nas pesquisas para criação de televisores e *displays* que reproduzam conteúdos 3D de alta qualidade e definição, com ou sem a necessidade de óculos (LG,2011; Mendiburu, 2009; Nintendo, 2011; Sony, 2011).

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 especiais – um par estéreo – para serem interpretadas pelo cérebro humano na criação de uma imagem única e tridimensional, provocando a sensação de profundidade e distanciamento. Tais métodos visam, 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 (Fehn et al., 2006).

Apesar dos avanços vistos na tanto na captura quanto na reprodução e representação de vídeos estereoscópicos, ainda existe necessidade de mais pesquisa na área da codificação de tais vídeos. 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 tais métodos podem ser divididos em dois grandes grupos: o método de Lipton (Lipton, 1997) e os métodos aqui chamados de vinculados (Smolic et al., 2009). [alterar o a divisão para métodos convencionais e métodos de vídeo+profundidade]

No método de Lipton o par estéreo é armazenado em containers (AVI, por exemplo), com compressão ou não. Apesar de ser mais flexível que os métodos vinculados, este resulta em um volume de dados duas vezes maior, devido à necessidade de se armazenar dois *streams* de vídeo (o par estéreo). Os métodos vinculados, por sua vez, utilizam técnicas consagradas de compressão de vídeo (como MPEG-2 e H.264), bem como de novos conceitos envolvendo mapas de profundidade para diminuir o volume de dados e atender às demandas de armazenamento/transmissão. Contudo, tais técnicas são apenas adaptadas para tratar vídeos estereoscópicos e funcionam somente para casos particulares (Smolic et al., 2009). Além disso, é utilizada a compressão com perdas, o que pode ocasionar a geração de artefatos que impossibilitam a correta percepção de profundidade em alguns casos, notadamente em vídeo anaglíficos (Andrade & Goularte, 2009). Por fim, pelo fato de ser necessário pelo menos um par de vídeos para a visualização estereoscópica, o volume de dados a ser armazenado é muito maior do que o necessário para vídeos monoculares. Tal característica não muda, mesmo quando utilizado técnicas de compressão mais atuais disponíveis. Como resultado, não existe uma técnica exclusiva para codificação de vídeo estereoscópico que atenda a todos os atuais métodos de visualização estereoscópica, os que necessitam de óculos especiais (anaglífico, com lentes polarizadoras e obturadores) e o autoestereoscópico, e que gere vídeos de qualidade e boa taxa de compressão.

Um dos objetivos deste trabalho é alcançar compressão adicional transformando-se o par estéreo em um único *stream* com metade do volume de dados, utilizando para isso o método anaglífico. Desse modo, o vídeo em formato anaglífico poderia ser utilizado para fins de armazenamento/transmissão (pois possuiria boa taxa de compressão) e a técnica atenderia ao método de visualização anaglífico (com diferencial em qualidade). Outro objetivo é que a técnica possa ser utilizada por qualquer método de visualização estereoscópica. Sendo assim, é necessário reverter o vídeo anaglífico 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 realizá-la. Com isso, o objetivo deste trabalho é desenvolver uma técnica de reversão de vídeo anaglífico em seu respectivo par estéreo.

Este texto está organizado da seguinte forma: a Seção 2 traz alguns 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 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 Tabela de Índice de Cores. É também apresentado os resultados já obtidos com a técnica implementada. Na Seção 5 são discutidas as limitações da técnica apresentada, e são delineadas as atividades a serem realizadas de forma a refiná-la, juntamente com o cronograma proposto a ser seguido até o fim do Mestrado. Por fim, apresentam-se todas as referências utilizadas como apoio à produção deste texto.

1. Fundamentos da visualização estereoscópica
   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, característica classificada como disparidade binocular (Azevedo & Conci, 2003). Por essas razões era de se esperar que, ao olharmos para um objeto, ele fosse visto por nós sob duas perspectivas diferentes, e não somente uma, como ocorre em nossa visão. Além disso, dentre os vários objetos presentes no nosso campo de visão, temos a capacidade de interpretar diferentes profundidades e texturas entre eles, e tal capacidade permanece mesmo se nos movermos para um lado ou para outro. A utilização de ambos os olhos para formar uma única imagem, com diferentes níveis de profundidade entre os objetos nela presentes, é definida como estereopsia.

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

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

* + 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 pode ser dividida em: perspectiva linear, interposição, luz e sombra, perspectiva aérea, variação da densidade de textura, conhecimento prévio do objeto e paralaxe de movimento. [Considerar colocar figuras para ilustrar algumas]

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).

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 profundidade destes em comparação a outros e ao ambiente em que se encontram. Com isso, ao vermos tais objetos em uma mesma imagem, de acordo com nossas experiências e conhecimento prévio, 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).

* + 1. Informações oculo-motoras

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

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

Cada olho produz uma imagem diferente do que está sendo visto, porém, conseguimos fazer com que um objeto seja visto na mesma posição em ambos os olhos se focarmos nele. Para isso, ele deve se encontrar em um mesmo ponto para os dois olhos, chamado de ponto de convergência. De acordo com a distância em que se encontra o objeto, devemos alterar nosso ponto de convergência. O ângulo formado na movimentação dos olhos em torno do seu eixo vertical para esse ponto de convergência nos dá a informação da distância do objeto. Tanto a acomodação quanto a convergência são reproduzidas artificialmente por máquinas de captura como câmeras e filmadoras digitais, quando se altera o foco.

* + 1. Informações estereoscópicas

Como anteriormente exposto cada olho produz uma imagem diferente, devido ao fato de estarem a uma distância e ângulos diferentes (disparidade binocular). Cabe ao cérebro se encarregar de retirar as informações das distâncias relativas dos objetos e de interpretar essas duas imagens resultando na produção de uma única. 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 citada anteriormente, a estereopsia é a responsável pela sensação que temos de profundidade entre os objetos, e é obtida em virtude da disparidade binocular. Dessa forma, o requisito obrigatório para obtermos estereopsia é 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 ter fundo.

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

Diretamente ligado ao conceito de disparidade (obtida na imagem formada na retina) temos a paralaxe, que é a distância entre os pontos correspondentes nas imagens 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 os óculos ele parece estar tremido, com regiões duplicadas e sobrepostas.

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. [considerar colocar imagem de exemplo] A paralaxe zero é quando os pontos correspondentes em cada imagem estão na mesma posição, ou seja, a diferença entre eles é zero; neste caso, os pontos convergem na retina. A paralaxe positiva ocorre quando a distância entre pontos correspondentes está entre zero e uma constante t, e dão a sensação de que os objetos estão distantes; isto ocorre porque o ponto de convergência das imagens no eixo de projeção de cada olho é obtido após o plano de projeção. Já a paralaxe negativa nos passa a sensação de que os objetos estão próximos de nós, como que saindo do monitor; tal efeito é consequência do cruzamento dos eixos de projeção de cada olho ocorrer antes de chegar ao plano de projeção. Por fim, a paralaxe divergente é um caso especial da paralaxe positiva 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 similar na visão humana.

* 1. 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, monitores autoestereoscópicos e óculos obturadores.

* + 1. Estereoscopia anaglífica

É o método mais simples para ser implementado 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 à cor vermelha (por exemplo, da imagem do lado direito), e do outro, as informações relativas às cores azul e verde (por exemplo, do lado esquerdo). Logo após, criamos um novo vídeo resultante da junção dos dois primeiros, como exemplificado na Figura 2. Na reprodução, o espectador usa um par de óculos especiais atuando como um filtro, possuindo nas lentes as cores que foram eliminadas, ou seja, uma lente vermelha (para o olho esquerdo, nesse caso) e outra lente ciano (para o olho direito). Com isso, é possível separar cada imagem diferente para cada olho, obtendo a disparidade 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, isto é, do anáglifo de volta para o par estéreo, pois a conversão envolveu eliminação de informação. A recuperação dessas informações requer investigação e é o objetivo do Mestrado.

* + 1. Luz polarizada

Neste método e nos demais métodos é utilizado 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 gravação quanto para a edição e a reprodução, fazendo-se necessário aquisição de novos equipamentos, mais robustos e por consequência, mais caros.

* + 1. Óculos obturadores

Diferente dos óculos utilizados em vídeos anaglíficos e por luz polarizada, que filtram as imagens corretas para cada olho, os óculos obturadores separam a imagens mecanicamente. Esta é uma tecnologia muito utilizada pelos televisores 3D e funciona da seguinte forma: o monitor exibe alternadamente em 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]](#footnote-1).

* + 1. 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, como o próprio nome diz, são capazes de gerar sozinhos a sensação de profundidade nas imagens reproduzidas. Tal feito é realizado criando-se diferentes visões estéreo de uma mesma cena, vista por ângulos diferentes e limitadas a certo segmento do campo de visão do espectador, fazendo com que este veja a cena de perspectivas diferentes ao movimentar-se para outro campo de visão. 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 [Colocar figura para exemplificar]. 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 está em desenvolvimento e apresenta deficiências as 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, a imagem aparece borrada. Por ainda estar em desenvolvimento, seu custo de produção é bem alto.

* 1. 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.

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]](#footnote-2). A ressalva é que isso só será possível com a produção e transmissão de conteúdos preparados para a tecnologia, que ainda é muito pequena, além da disseminação e interesse do público em obter transmissão com esse conteúdo.

O mercado de games parece ser um dos que mais serão beneficiados com a utilização de conteúdo 3D, fornecendo uma nova alternativa de interatividade e imersão dos usuários com os jogos. Os grandes fabricantes de consoles vêm se mostrando interessados em investir nessa tecnologia, como é o caso da Nintendo e seu portátil Nintendo 3DS, que utiliza duas telas, sendo uma delas autoestereoscópica e a outra sensível ao toque (Nintendo, 2011); e também o caso da Sony, que atualizou o *firmware* de seu console Playstation 3, tornando-o capaz de reprodução 3D com a utilização de televisores compatíveis com a tecnologia, e vem constantemente lançando jogos em 3D [procurar referência].

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).

[Procurar mais aplicações científicas]

1. Aspectos de codificação e compressão estereoscópica
   1. Espaço de cores e subamostragem de crominância

A representação de imagens se dá através de tons monocromáticos ou coloridos. Computacionalmente, os tons monocromáticos são 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 YCbCr, que vêm sido utilizados nas atividades relacionadas à pesquisa. Mais detalhes sobre outros espaços de cores podem ser vistos nos textos de Azevedo & 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 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 YCbCr possui em suas componentes a separação das informações relativas à luminância (Y) das relativas à crominância (Cb e Cr), como uma tentativa de simular a visão humana. Cb e Cr representam valores de crominância das componentes azul e vermelha. A componente verde (Cg) pode ser obtida através de Cb e Cr, já que a soma das três é sempre igual a 1, sendo por isso eliminada da representação. Os valores de Y, Cb e Cr 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.

**Equação 1 – Conversão do espaço de cores RGB para YCbCr**

**Equação 2 – Conversão do espaço de cores YCbCr para RGB**

Como o espaço de cores YCbCr 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.

No modelo 4:4:4 (Figura W-a) não há redução da resolução das cores, isto é, para cada amostra de Y, há uma amostra de Cb e uma de Cr. Este modelo mantém a fidelidade das cores da imagem, porém, não contribui na compressão. No modelo 4:2:2 (Figura W-b), para cada quatro amostras horizontais de Y, há duas amostras de Cb e duas de Cr, reduzindo-se com isso 1/3 de informação. Já no modelo 4:2:0 (Figura W-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 YCbCr (4:2:2 ou 4:2:0) para seu similar em RGB, o retorno 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.

* 1. Codificação estereoscópica [melhorar e atualizar referências]

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 (Lipton, 1997), e a codificação baseada em vídeo e profundidade, que explora características de profundidade do par estéreo para aumentar a taxa de compressão.

* + 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 container 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.

* + 1. 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. Através deste mapa de profundidades, seria possível recriar o segundo vídeo do par estéreo, ou até mesmo novas visões. Por conter apenas dados luminância, possui tamanho menor em relação a um vídeo colorido (o segundo vídeo do par estéreo, neste caso), o que possibilita maior compressão.

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: seria armazenado o par estéreo, e respectivas camadas de profundidade 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 se mostra um ator 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 e quanto para criação de visões virtuais ainda são complexos, propensos a erros.

Em outra pesquisa, esta realizada por Vetro (2010), os formatos aqui mencionados são discutidos em relação a 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.

* + 1. Compressão [melhorar e atualizar referências] (Smolic et al., 2009)

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 uma 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 Cossine 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 o número de bits para cada dado conforme a freqüência em que aparecem, garantindo compressão. Destas, as mais conhecidas são Huffman, LZW e por carreira.

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 os dados das 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 anteriores a ele, obtendo-se uma estimativa do que mudou entre ele e seu antecessor (estimativa de movimento), ou seja, excluímos este quadro e ficamos apenas com os dados da estimativa de movimento para posterior reconstrução deste. Como essa predição envolve erros, é também codificada uma tabela de compensação de movimento, contendo a diferença entre a posição estimada e a posição real dos objetos. Como outros quadros P podem ser codificados a partir de um quadro P anterior, há uma propagação de erros, e por essa razão, deve-se estabelecer um limite de criação de quadros P consecutivos, chamado de *Prediction Span*. Por fim, os quadros B (*Bidirectional frames*) são codificados tanto em relação ao um quadro P ou I anterior a eles quanto em relação a um quadro P ou I posterior a eles, obtendo-se uma taxa maior de compressão, porém impactando o tempo de processamento, já que precisamos esperar os quadros P ou I posteriores serem processados para o cálculo.

* + 1. 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 *worldline correlation[[3]](#footnote-3)* (Balasubramaniyam et al., 2005; Adikari et al., 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 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 *worldline correlation,* obtendo-se assim boa taxa de compressão.

Embora obtenham 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 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 & 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.

1. Proposta de trabalho
   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. Foi visto 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 que vêm sido 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 mestrado é atacar estes dois problemas, realizando a codificação de vídeos estereoscópicos com qualidade, sem apresentar perda de percepção de profundidade, compatíveis com qualquer método de visualização disponível e com uma boa taxa de compressão. Visando compressão, a proposta é que o par estéreo seja convertido e armazenado em formato anaglífico, já que neste formato apenas um sinal de vídeo é armazenado, reduzindo pela metade o volume de dados. Visando a compatibilidade, deve-se também criar a técnica reversa, ou seja, fazendo com que do vídeo anaglífico seja obtido o par estéreo, o qual pode ser então utilizado por outros métodos de visualização.

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 1, dos seis canais de cores existentes no par estéreo, três foram 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 4.2.

* 1. 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 1, podemos ver que esta tabela seria então formada pelos dados dos canais de cores R1, G2 e B1. 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 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 YCbCr e armazenar somente as informações referentes à crominância (Cb e Cr), descartando informação de luminância (Y), já que esta pode ser obtida do anáglifo principal. Além disso, o anáglifo complementar, já no espaço YCbCr, 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 Y. 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 & 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 YCbCr, passandopela subamostragem de crominância 4:2:2, considerada por Andrade & Goularte (2010) como a melhor alternativa em conjunto com o anáglifo verde-magenta. Logo após, descartamos as informações de Y e armazenamos somente Cb e Cr 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, 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 H. Nesta etapa, o anáglifo principal também passa pelo processo de conversão do espaço de cores de RGB para YCbCr. Com isso, obtemos um Y’, os dados de luminância do anáglifo principal. Em conjunto com os dados da Tabela de Índice de Cores, utilizamos o Y’ para reconstruir o anáglifo complementar, neste caso na forma de Y’CbCr, 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.

* 1. Resultados obtidos

A Tabela de Índice de Cores foi formada eliminando os dados de Y e utilizando os dados de Cb e Cr 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 Cb e 2 de Cr). Matematicamente, espera-se que isto resulte em uma adição de 33% de dados ao arquivo final, o que pode ser reduzido ainda mais após passar pelo processo de compressão de dados sem perdas.

O processo descrito em 4.2 foi implementado em C++ com suporte da biblioteca OpenCV[[4]](#footnote-4), e aplicado a uma base de testes contendo 32 imagens de par estéreo. Estas imagens foram retiradas da base construída por Andrade et al. (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. O PSNR (*Peak Signal-to-Noise Ratio*)é uma métrica muito utilizada na análise de compressão de imagens (Winkler, 2005) que calcula a similaridade entre duas imagens (no caso, o par estéreo original e o obtido na reversão anaglífica), avaliando quantidade de ruído introduzida após um processo de compressão com perdas. O PSNR retorna um valor em decibéis, num intervalo de 0 a 100, sendo que quanto maior o valor, maior a similaridade encontrada.

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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |
| 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. Analisando a média do PSNR encontrado, 34,524 dB, vemos que este se encontra baixo. Isso provavelmente é devido a arredondamentos feitos durante o processo, já que a conversão de RGB para YCbCr resultaem dados de ponto flutuante, que são truncados na criação do arquivo final. Segundo 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 fórmulas matemáticas envolvendo sinais. Por isso, foi também analisada a qualidade subjetiva das imagens originadas no processo de reversão anaglífico, ou seja, elas foram verificadas visualmente e se mostraram de boa qualidade, sendo inclusive possível utilizá-las para gerar um novo anáglifo, sem a perda de percepção de profundidade, como pode ser visto na Figura A.

Nas imagens revertidas, é perceptível a presença de *ghosting*, isto é, como se alguns elementos da imagem aparecessem duplicados. 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.

Esse processo de conversão e reversão anaglífica, bem como os resultados obtidos, foram condensados em um artigo submetido ao XVII Simpósio Brasileiro de Sistemas Multimídia e Web – WebMedia, tendo sido aprovado [otimismo].

1. Metodologia de Trabalho
   1. Limitações da técnica criada

Como citado na Seção 4.3, 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. 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 *ghosting*, bem como realizar testes em uma base de dados bem 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.

* 1. Melhoria de PSNR

O PSNR é uma métrica de qualidade muito utilizada na comparação de imagens e vídeos que passaram por algum tipo de codificação. Sua fórmula não é complicada e está baseada na comparação pixel a pixel de duas imagens, retornando como resultado a similaridade entre elas. Por fazer apenas uma análise matemática entre pixels correspondentes, seu resultado é limitado em relação à qualidade subjetiva da imagem, isto é, uma imagem com baixo PSNR não significa necessariamente ser de baixa qualidade quando visualizada por uma pessoa. [Verificar se removo este parágrafo, já que há um parecido na Seção 4.3]

Nos resultados obtidos e apresentados na Seção 4.3, 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 4.2 e ilustrado nas Figuras Y e H, há uma etapa de mudança de espaço de cores do RGB para YCbCr 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 *ghosting*, analisada na Seção 5.3). 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.

* 1. Análise de correlação de imagens

A Figura Z (usar old1.bmp) mostra a comparação de qualidade visual de uma imagem estéreo sem compressão (A) e sua correspondente após passar pela reversão anaglífica utilizando a Tabela de Índice de Cores (B). Pode-se observar boa qualidade visual em (B), com algumas imperfeições, notadamente encontradas nas bordas de alguns elementos, correspondendo a regiões de paralaxe negativa mais acentuada. Tais imperfeições aparecem como regiões duplicadas, chamadas de *ghosting*, sendo que na imagem do lado esquerdo de (B) é mais notável a presença das cores magenta clara e verde (observe os contornos do trem). O lado direito de (B) 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, Cb e Cr de cada anáglifo. Como estamos utilizando a componente Y do anáglifo principal para reconstruir o complementar (ver Figura H), tais deslocamentos são também incorporados a este. Essa é a razão do aparecimento de ghosting no par estéreo revertido. Tal efeito afeta não somente a qualidade visual, como também o resultado do PSNR.

Para eliminar o *ghosting*, uma estratégia é realizar a análise do que é chamado “correlação entre as imagens”. 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 mestrado, faz-se então necessário um estudo das pesquisas relacionadas à correlação de imagens para saber o estado da arte nesta área, para depois incorporá-la ao processo de conversão e reversão anaglífica.

* 1. 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 VMQT (Video Quality Measurement Tool)[[5]](#footnote-5). 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 YCbCr, além de fornecer uma imagem com as diferenças encontradas nas imagens comparadas, servindo como uma boa referência visual.

Para a avaliação subjetiva, será utilizado o MOS (Mean Option Score) (Winkler, 2005), isto é, a 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.

Pretende-se seguir as recomendações da ITU-T (2004; 2008). Inicialmente montam-se estruturas de vídeos ou imagens ABAB, onde A é o original e B é o obtido através da reversão anaglífica, sendo que nos intervalos entre cada um existem trechos de tela cinza (Figura B). Os avaliadores recebem então esta estrutura e dão uma nota de 0 a 5 para o vídeo ou imagem revertida.

A base de dados citada na Seção 4.3, 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.

* 1. Cronograma

Segue abaixo a proposta de atividades a serem. 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. Desenvolvimento de um software com interface amigável para realizar a compressão e descompressão de vídeos e imagens utilizando a técnica criada. [ainda precisamos conversar sobre esta parte]
10. Defesa do Mestrado

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2011** | | | | | **2012** | | | | | | | |
| **Atividades** | **Ago.** | **Set.** | **Out.** | **Nov.** | **Dez.** | **Jan.** | **Fev.** | **Mar.** | **Abr.** | **Mai.** | **Jun.** | **Jul.** | **Ago.** |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |  |  |  |

* 1. Considerações finais

Os resultados deste mestrado 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 deste 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 este é apenas uma parte do processo completo de codificação e compressão de imagens e vídeos. Mais compressão pode ser obtida em outras etapas do processo, tais como os aspectos envolvendo redundância espacial e temporal. [Seria bom falar do modelo como um todo que está sendo pensado, utilizando tanto os trabalhos do professor e do Léo quanto o meu?]

Referências

(Adikari et al., 2005) Adikari, A.B.B.; Fernando, W.A.C.; Arachchi, H.K.; Loo, K. A H.264 compliant stereoscopic video codec. *Canadian Conference on Electrical and Computer Engineering*, 1614-1617, 2005. DOI = <http://dx.doi.org/10.1109/CCECE.2005.1557292>

(Andrade et al., 2010) Andrade, L. A.; Cordebello, P. D.; Goularte, R. . Construção de uma base de vídeos digitais estereoscópicos. *Relatório Técnico ICMC-USP*, 2010. Disponível em <http://www.icmc.usp.br/~biblio/BIBLIOTECA/rel_tec/RT_351.pdf>. Último acesso feito em 24/06/2011.

(Andrade & Goularte, 2009) 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)*, 226-233, 2009. DOI = <http://doi.acm.org/10.1145/1858477.1858506>

(Andrade & Goularte, 2010) 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), 1-8, 2010.[Já possui DOI?? Onde eu procure??]*

(Balasubramaniyam et al., 2005) Balasubramaniyam, B.; Edirisinghe, E.; Bez, H. – An Extended H.264 CODEC for Stereoscopic Video Coding. *Proceedings of SPIE*, 116-126, 2005. DOI = <http://dx.doi.org/10.1117/12.587583>

[Verificar referência](Azevedo & Conci, 2003) Azevedo, E.; Conci, A. – Computação gráfica: teoria e prática. Editora Campus, Brasil, 2003.

(Chapman & Chapman, 2004) Chapman, N. P.; Chapman, J. – Digital Multimedia. Wiley, England, 2004.

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

(Fehn et al., 2006) Fehn, C.; de la Barré, R.; Pastoor, S. – Interactive 3-DTV – concepts and key technologies. *Proceedings of the IEEE*,524-538, 2006. DOI = <http://dx.doi.org/10.1109/JPROC.2006.870688>

(Feitosa-Santana et al., 2006) Feitosa-Santana, C.; Oiwa, N. N.; da Costa, M. F.; Tiedemann, K. B.; Silveira, L. C. L.; Ventura, D. F. – Espaço de cores. *Psicologia USP [online],* vol.17, n.4, 35-62, 2006 . Disponível em: <http://www.revistasusp.sibi.usp.br/scielo.php?script=sci_arttext&pid=S1678-51772006000400003&lng=pt&nrm=iso>. Último acesso feito em 08/07/2011.

(Gonzalez & Woods, 2008) Gonzalez, R. C.; Woods, R. E. – Digital Image Processing. 3ª Edição. Prentice-Hall, USA, 2008.

(ITU-T, 2004) ITU-T – ITU-T Recommendation J.144. Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference. *International Telecommunication Union*, Suiça, 2004. [Não tenho essas referências, onde conseguir??]

(ITU-T, 2008) ITU-T – ITU-T Recommendation J.247 : Objective perceptual multimedia video quality measurement in the presence of a full reference. *International Telecommunication Union*, 2008. [Não tenho essas referências, onde conseguir??]

(Kerr, 2009) Kerr, D. A. - Chrominance Subsampling in Digital Images. 2009. Disponível em <http://dougkerr.net/pumpkin/articles/Subsampling.pdf>. Último acesso feito em 11/07/2011.

(LG, 2011) LG Electronicst – LG Optimus 3D P920, 2011. Disponível em <http://www.lg.com/uk/mobile-phones/all-lg-phones/LG-android-mobile-phone-P920.jsp>. Último acesso feito em 24/06/2011.

(Li et al., 2009) Li, S.; Hou, C.; Ying, Y.; Song, X.; Yang, L. – Stereoscopic Video Compression Based on H.264 MVC. *2nd International Congress on Image and Signal Processing, 2009 (CISP '09)*, 1-5, 2009. DOI = <http://dx.doi.org/10.1109/CISP.2009.5301218>

(Lin et al., 2009) Lin, Z.; Xiang, Z; Luo, L.; Zhao, L. An Improved Stereo Video Coding Scheme Based on Joint Multiview Video Model. *First International Workshop on Education Technology and Computer Science*, 1091-1095, 2009. DOI = <http://dx.doi.org/10.1109/ETCS.2009.249>

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

(Lipton, 1997) Lipton, L. – Stereo-Vision Formats for Video and Computer Graphics. *Proceedings SPIE*, 239-244, 1997. DOI = <http://dx.doi.org/10.1117/12.274462>.

(Mendiburu, 2009) Mendiburu, B. – 3D Movie Making: Stereoscopic Digital Cinema from Script to Screen. Elsevier, United Kingdom, 2009.

(Merkle et al., 2007) Merkle, P.; Smolic, A.; Muller, K.; Wiegand, T. Efficient Prediction Structures for Multiview Video Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 1461-1473, 2007. DOI = <http://dx.doi.org/10.1109/TCSVT.2007.903665>

(Nintendo, 2011) Nintendo of America Inc. – Nintendo 3DS, 2011. Disponível em <http://www.nintendo.com/3ds/hardware>. Último acesso feito em 24/06/2011.

(Richardson, 2003) Richardson, I. E. G. – H.264 and MPEG-4 Video Compression – Video Coding for Next-generation Multimedia. Wiley, England, 2003.

(Salomon, 2008) Salomon, D. – A Concise Introduction to Data Compression (Undergraduate Topics in Computer Science). Springer, England, 2008.

(Sayood, 2005) Sayood, K. – Introduction to Data Compression, 3ª Edição. Morgan Kaufmann Publishers Inc., USA, 2005.

(Smolic et al., 2009) Smolic, A.; Mueller, K.; Merkle, P.; Kauff, P.; Wiegand, T. An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic-solution. *Picture Coding Symposium*, 1-4, 2009.

DOI = <http://dx.doi.org/10.1109/PCS.2009.5167358>

(Sony, 2011) Sony Corporation – Sony 3D TV Technology, 2011. Disponível em <http://www.sony.net/united/3D/#technology/3dtv/>. Último acesso feito em 24/06/2011.

(Stereographics, 1997) StereoGraphics Corporation – Stereographics® Developers’ Handbook: background on creating images for CrystalEyes® and SimulEyes®, 1997.

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

(Vetro, 2010) Vetro, A. – Representation and Coding Formats for Stereo and Multiview Video. *Studies in Computational Intelligence*, 51-73, 2010. DOI= <http://dx.doi.org/10.1007/978-3-642-11686-5_2>.

(Winkler, 2004) Winkler, S. JPEG vs. JPEG2000: An Objective Comparison of Image Encoding Quality. *Proceedings of SPIE Applications of Digital Image Processing*, 300-308, 2004.

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

Glossário [é necessário?]

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. [↑](#footnote-ref-1)
2. Pesquisa publicada em <http://idgnow.uol.com.br/computacao_pessoal/2010/08/02/pesquisa-80-das-tvs-vendidas-nos-eua-em-2014-terao-3d>, último acesso em 26/06/2011. [↑](#footnote-ref-2)
3. Alguns autores usam o termo *inter-view correlation* (Merkle et al., 2007; Lin et al., 2009) [↑](#footnote-ref-3)
4. 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/> (último acesso em 24/06/2011). [↑](#footnote-ref-4)
5. O software MSU VQMT pode ser baixado em <http://compression.ru/video/quality_measure/vqmt_download_en.html#free>, visitado pela última vez em 24/06/2011. [↑](#footnote-ref-5)