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, feitos apenas para se aproveitar da fama. 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 e 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 simular o efeito obtido na visão humana pelo fato de nossos olhos estarem distantes horizontalmente um do outro, o que faz com que cada olho tenha um ponto de vista diferente, algo que 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).

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 maior do que o necessário para vídeos monoculares. Tal característica não muda, mesmo ao se utilizar as técnicas de compressão mais atuais. 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). Além disso, 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 métodos de luz polarizada, óculos obturadores e monitores autoestereoscópicos. Tal reversão é uma novidade na área e necessita de mais estudos para saber como realiza-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 refinar a técnica, juntamente com o cronograma proposto para se seguir 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
      1. Informações monoculares
      2. Informações oculo-motoras
      3. Informações estereoscópicas
   2. Tipos de visualização estereoscópica
      1. Estereoscopia anaglífica
      2. Luz polarizada
      3. Óculos obturadores
      4. Monitores Autoestereoscópicos
   3. Aplicações

Falar da importância da visualização estereoscópica encontrados em diversas áreas do conhecimento e também comercialmente.

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

Falar apenas de (RGB e YCbCr)

* 1. Subamostragem de crominância

Falar das principais: 4:4:4, 4:2:2, 4:2:0, 4:1:1

* 1. Codificação estereoscópica
     1. Codificação convencional

Método de Lipton, CSV, MVC

* + 1. Codificação baseada em vídeo e profundidade

V+D, MVD, LDV, DES

* 1. Compressão
     1. Compressão de vídeo monocular

Falar genericamente da codificação MPEG

* + 1. Limitações para compressão de vídeo estereoscópico

Problema do tamanho do arquivo, problema do anáglifo, problema do head motion parallax

* + 1. Abordagens de compressão de vídeo estereoscópico

Artigos utilizando de extensões do MPEG-2 e H.264. Resultados do Léo

1. Proposta de trabalho
   1. Apresentação da proposta

Pelo que foi exposto na Seção 3.4, 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 a ser utilizada. 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 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 compressão de vídeos estereoscópicos com qualidade, que não apresentem perda de percepção de profundidade e sejam compatíveis com qualquer método de visualização disponível. 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, ao invés do par estéreo, 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 X, dos seis canais de cor existentes no par estéreo, três foram descartados. Uma simples duplicação das informações dos canais presentes no anáglifo não bastaria para recuperar o par estéreo, já que as imagens no par original não são exatamente iguais, o que afetaria a qualidade.

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 X, 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 o anáglifo a ser de fato utilizado para a visualização. 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 Léo. 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. 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[[1]](#footnote-1), e aplicado a uma base de testes contendo 32 imagens de par estéreo. Estas imagens foram retiradas da base construída por Andrade e Goularte, 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 [Referência Winkler] 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.

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 da 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)[[2]](#footnote-2). 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) [Livro do Winkler], 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 (ITU, 2004; ITU, 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 e trabalhos 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.
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 um elo da cadeia. Mais compressão pode ser obtida em outras etapas do processo de compressão de imagens e vídeos, tais como os aspectos da 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

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

(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>

(LG, 2011) LG Electronics – 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.

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

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

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

(Smolic et al.) 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>

Glossário [é necessário?]

1. 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-1)
2. 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-2)