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

Falar do desenvolvimento da tecnologia 3D: desenvolvimento da captura tornou possível a criação de vídeos estereoscópicos digitais de alta qualidade que têm sido utilizados intensamente pelas produtoras de filmes. Seguindo essa onda, temos as produtoras de televisores, investindo em pesquisa para criar televisores preparados para o 3D, tanto com óculos quanto sem, o estado da arte, possibilitando até que o ponto de vista da pessoa mude conforme a posição dela em relação à TV. Mostrar que apesar desses avanços na captura e visualização, o campo da codificação fica um pouco atrás. Podem-se utilizar os melhores codificadores para vídeo monocular atuais, porém, o tamanho do arquivo ainda é grande, já que mais de um sinal de vídeo está sendo utilizado, e as perdas na compressão geram artefatos que prejudicam a percepção de profundidade na visualização estereoscópica, principalmente na anaglífica. Falar então da proposta de diminuir o tamanho do arquivo ao se armazenar o par estéreo em anáglifo, e realizar o processo de reversão, obtendo novamente o par estéreo para ser utilizado por outros tipos de visualização. Explicar que é um trabalho pioneiro e contribuição para a área, e que não é um processo trivial, pois a codificação anaglífica implica em perda de metade das informações de cores do par estéreo, porém, já obtido alguns avanços. Explicar também que o método anaglífico é o que mais sofre perdas na compressão, por isso, se produzimos uma imagem com qualidade para este, consequentemente estaremos contribuindo também para os outros métodos. Finalizar falando da organização do trabalho.

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 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 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.

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 se 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 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.

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 será 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 estes valores para serem armazenados em variáveis de dados que utilizam menos espaço de armazenamento. Isso resulta em perda tanto da precisão quanto dos valores que sejam maiores do que o limite permitido na 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* e é 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 ponto conjugado

Falar como o ghosting influencia tanto no valor do psnr quanto na qualidade visual da imagem.

Mostrar uma figura da base para ilustrar os efeitos de ghosting.

* 1. Avaliações Objetiva e Subjetiva

Dar mais detalhes do MOS.

* 1. Cronograma
  2. 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?]

1. Referências

Glossário [é necessário?]