UNIVERSIDADE DA BEIRA INTERIOR
Engenharia

Regiões de Interesse em Vídeos 3D

Daniel Alexandre Leitão Piedade

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática
(2º ciclo de estudos)

Orientador: Prof. Doutora Maria Manuela Areias da Costa Pereira de Sousa
Co-orientador: Prof. Doutor António Manuel Gonçalves Pinheiro

Covilhã, Outubro de 2013

Dedicatória

À minha avó Maria Esperança.

Agradecimentos

Quero agradecer aos meus pais, Carlos Alberto da Ressurreição Piedade e Maria Helena Leitão Borrego Piedade pela confiança depositada em mim, pela paciência que tiveram ao longo de todo o curso e motivação constante.

Quero deixar também uma nota de agradecimento especial aos meus amigos que fizeram com que este projecto fosse realizado, que sempre me apoiaram e incentivaram todos os momentos que passei na Universidade da Beira Interior, muitos deles felizes, outros nem tanto, mas só com a sua ajuda consegui superar as dificuldades encontradas e concluir o Mestrado em Engenharia Informática.

Uma nota de agradecimento especial à Mafalda Teresa Rocha Faria e ao André Filipe Prata Ferreira pelo apoio constante, motivação e grande ajuda na realização deste projecto.

Agradeço a todos os que me ajudaram neste trabalho nomeadamente os meus Orientadores, Professora Doutora Manuela Pereira, Professor Doutor António Pinheiro e o Professor Doutor Paulo Fazendeiro pela sua disponibilidade, compreensão e paciência que tiveram ao longo deste trabalho.

Aos meus colegas no Centro de Óptica Marco, Jorge, António e Anita, também pela sua disponibilidade e paciência que tiveram ao longo do meu percurso no Centro de Óptica.

Os meus enormes agradecimentos.

Daniel Piedade

Resumo

Um estudo sobre a percepção da qualidade de vídeos 3D exibidos com mudanças de cor é apresentado. As cores de sequências de vídeos 3D foram mudadas no espaço de cor CIE 1976 (L*a*b*), com a aplicação de um erro cromático pré-definido ΔE^*_{ab}. As cores foram inicialmente divididas em clusters com o algoritmo K-Means. Cada cluster de cor é deslocado pelo erro cromático pré-definido com uma orientação aleatória nas coordenadas cromáticas a*b*. Aplicando erros ΔE^*_{ab} de 6, 12 e 18 unidades a um conjunto de seis vídeos 3D foi obtido um conjunto de vídeos 3D para a experiencia. Estes vídeos foram mostrados a vários observadores, em que lhes foi pedido para classificar a qualidade do conjunto de vídeos observados tendo em conta a sua naturalidade quanto à cor, ao mesmo tempo que era capturado a atenção do utilizador a diferentes vídeos usando um dispositivo de "eye-tracking". Foi feito um estudo para testar e quantificar a sensibilidade às alterações de cor, bem como a análise da variação de atenção com a variação da cor.

Palavras-chave

Mean Opinion Score, Cor, Avaliação, Qualidade, Vídeo 3D, EyeTracking

Abstract

A study on the perceived quality of 3D video displayed with color changes is presented . The colors of 3D video sequences have been changed in CIE 1976 (L*a*b*) color space, with the application of a predefined chromatic error ΔE^*_{ab}. The colors were initially divided into clusters with the K -Means algorithm. Each cluster is shifted by the predefined chromatic error with a random direction in a*b* chromatic coordinates. Applying the ΔE^*_{ab} errors of 6 , 12 and 18 units to the six 3D video a set of modified 3D videos been collected for the experiment. These videos were shown to individuals, where were asked to rate the quality of the observed video set based on their naturalness, at the same time it was captured the attention of the user to the different videos using an "eye-tracking"device. A study was conducted to test and quantify the sensitivity to color changes, as well as analysis of attention variance to the color variation.

Keywords

Mean Opinion Score, Color, Assessment, Quality, 3D Video, EyeTracking

Índice

Lista de Figuras

Lista de Tabelas

Lista de Acrónimos

UBI	Universidade da Beira Interior
MOS	Mean Opinion Score
QoE	Quality of experience
QoS	Qualidade de Serviço
UIT	União Internacional de Telecomunicações
VQEG	Video Quality Experts Group
CRT	Cathode ray tube
MATLAB CRS	Matlab Cambridge Research Systems
VGA	Video Graphics Array
CIE	International commision os Illumination
TCP	Transmission Control Protocol
RGB	Red Green Blue
AVI	Audio Video Interleave
sRGB	Standard Red Green Blue
Full HD	Full High Definition
SDI streams	Serial Digital Interface Streams
SAMVIQ	Avaliação Subjetiva de Qualidade de vídeo
DMOS	Diferential mean opinion scores

Capítulo 1

Introdução

1.1 Motivação

Enquanto a avaliação da qualidade objectiva e subjectiva de imagens e vídeo 3D tem sido um tema de pesquisa activa nos últimos anos, as novas tecnologias 3D emergentes requerem novas métricas e metodologias para avaliação de qualidade tendo em conta as diferenças fundamentais entre a percepção visual humana e as distorções típicas de conteúdo estereoscópico. Nesse sentido torna-se importante observar e perceber que eventuais alterações no conteúdo visual do vídeo 3D, sejam provocados por erros ou artefactos de codificação, ou por outras fontes, possam provocar na localização dos pontos de atenção.

Para isto será usado um sistema que permite avaliar a atenção do utilizador a diferentes vídeos com a aplicação de um erro cromático ΔE^*_{ab}, usando um dispositivo de "eye-tracking"e um sistema de reprodução de imagem e vídeo 3D. Será efectuada toda uma experiencia, obedecendo aos critérios exigidos pelas normas ITU. Nesta experiência, pretende-se usar o sistema atrás descrito desenvolvido para captar mapas de seguimento do olho ao longo de vários vídeos. Será usada uma base de dados de vídeos 3D criadas em conjunto pela Universidade Politécnica de Nantes e pela Universidade Politécnica de Madrid chamada Nantes-Madrid-3D-Stereoscopic-V1, NAMA3DS1-COSPAD1 [1], a base de dados dos vídeos escolhida possui várias cenas diferentes umas indoor e outdoor, rurais e urbanas. Poderá ser verificado se os pontos de interesse do observador sofrem alterações entre cada vídeo com um erro cromáticos ΔE^*_{ab} pré definido. Além disso, será feito um protocolo de qualidade subjectiva a fim de calcular o Mean Opinion Score (MOS), o que permitirá analisar se existe alguma relação entre as alterações nos pontos de interesse e o respectivo MOS.

1.2 Problema - Objectivos

O objectivo central deste trabalho é perceber a sensibilidade do utilizador a diferentes erros cromáticos ΔE^*_{ab} com magnitudes variadas, em vídeos 3D. Para isso será usado um sistema que permite avaliar a atenção do utilizador a diferentes vídeos usando um dispositivo de "eye-tracking". Visto que com a base de dados irão ser recolhidos dados para o cálculo do MOS durante a experiência, pretende-se analisar as relações pontos de interesse - MOS - Erro.

Com a apresentação do plano de trabalho presente, três objectivos principais foram pesquisados:

- Obtenção dos valores de MOS para avaliar alterações da cor em Vídeo 3D
- A análise da variação de atenção com a variação da cor.
- Relação entre a variação de atenção e a avaliação subjectiva.

Neste contexto, este projecto pretende proporcionar uma investigação valiosa na modelação de QoE em aplicações de vídeo.

1.3 Estrutura da Tese

Esta tese está estruturada em 9 capítulos. O capítulo 1 descreve, de forma geral, o propósito do trabalho. No capítulo 2 faz-se um breve estudo do estado da arte associado a este tipo de experiências e tecnologias. O capítulo 3, descreve sumariamente o sistema visual humano, em termos das percepções das dimensões visuais. No capítulo 4 é apresentado o sistema criado para a realização deste projecto. No capítulo 5 é apresentada a metodologia aplicada. No capítulo 6 são descritos os testes subjectivos de avaliação de qualidade. No capítulo 7 é feita a análise dos resultados subjectivos, no capítulo 8 é feita uma pequena análise dos mapas de atenção. Finalmente, no capítulo 9, são referenciadas as conclusões obtidas no final das experiências e análise de resultados.

Capítulo 2

Estado da Arte

Quando as aplicações visuais são consideradas, o termo comum Qualidade de Serviço (QoS) não é suficiente para fornecer uma descrição adequada do desempenho do sistema. QoS é uma medida técnica relacionada com o desempenho objetivo de um sistema. Obviamente que o QoS tem falta de evidências suficientes sobre a percepção de qualidade de um utilizador, porque o consumidor de de um sistema de uma saída visual, o observador humano, não é considerado [2]. Em oposição, Qualidade de Experiência (QoE) é uma avaliação que envolve os fatores subjetivos do utilizador. Por essa razão, a QoE é mais apropriada para avaliação do desempenho do sistema real. Em [3] e [4] uma das características gerais descritivas da qualidade de experiencia é identificada por vídeo 3D em dispositivos móveis, e sobre os sistemas de multimédia, respectivamente. Ambos os documentos concluem que toda a mistura de componentes de qualidade requerem a realização de experiências de avaliação da qualidade subjetivas com utilizadores potenciais.

A avaliação subjectiva da qualidade de áudio e visual é considerada o método mais preciso reflectindo a percepção humana [5]. A avaliação subjectiva da qualidade visual 2D de acordo com métodos padronizados, tem uma longa história. A União Internacional de Telecomunicações (UIT) emitiu várias recomendações, incluindo o amplamente utilizado ITU-R BT.500 [6]. Recentemente foi dada uma grande atenção às soluções tecnológicas que exibem informações visuais 3D, tornando-se muito importante para analisar a qualidade das imagens que são visualizadas em 3D. Percepção 3D envolve novos pontos críticos que devem ser considerados. Para a avaliação da qualidade subjectiva de dados multimédia futuros, tais como alta definição e 3DTV, métodos semelhantes [6] são recomendados em ITU-R BT.710 [7] e ITU-R BT.1438 [8]. Um dos estudos anteriores sobre atributos de percepção de vídeo 3D foi relatado em [9]. Mais tarde, outros autores realizaram testes subjetivos, a fim de identificar as principais necessidades em 3D [10, 11, 12]. O efeito na percepção de profundidade na qualidade da imagem em 3D foi relatada em [13]. Aqui, as imagens de profundidade foram quantificadas em diferentes configurações de bit rate e subjetivamente avaliada para a percepção 3D em geral. Em [14] os autores realizaram uma avaliação subjetiva na qualidade de imagem para diferentes técnicas de compressão com perdas e para diferentes monitores estereoscópicos. Em [15] os autores realizaram testes subjetivos para determinar a qualidade de imagem e percepção de profundidade de uma série de sequências de vídeo codificados de forma diferente, com diferentes taxas de perda de pacotes. [16] centra-se na percepção de profundidade na avaliação de cor mais a representação de profundidade do vídeo 3D. Além disso, a variação da avaliação da percepção é analisada com deficiências introduzidas durante a compressão das imagens em profundidade. Em [17] os autores conduziram vários testes subjectivos extensivos para estudar a influência dos parâmetros de aquisição sobre a qualidade 3D percebida. Em [19] foram realizados experiências com aparência natural e como os efeitos 3D foram afetadas pela base estereoscópica e pela distância do objeto mais próximo na cena, e também como eles dependem da base de estereoscópica e da distância focal da câmara. Em [20], são exploradas diferentes efeitos no 3D (intensidade, a profundidade, a presença, etc.) [15] conclui que os artefatos introduzidos no

vídeo a cores, dificultam a percepção de profundidade de sequências de vídeo estereoscópico. Assim, concluiu-se que existe uma alta correlação entre a qualidade da imagem percebida e profundidade percebida.

Embora os resultados dos testes subjetivos ainda sejam a avaliação mais precisa de medição da qualidade na pesquisa visual, as métricas objetivas são de extrema importância para o desenvolvimento de tecnologias relacionadas com o vídeo 3D. Assim, ITU lançou uns requerimentos para um modelo de qualidade multimédia perceptual objetiva [21]. Têm sido propostos vários métodos objectivos para medir a qualidade de vídeo percebida. Juntamente com a ITU, a VQEG produziu recomendações para a avaliação objetiva de qualidade de vídeo usando modelos de referência completa, ou seja, ITU-T J.144 [22]. Recentemente [23] tem uma pesquisa de signal-driven de áudio e vídeo perceptual com métodos de avaliação da qualidade de forma independente, e com questões relevantes investigadas no desenvolvimento conjunto de métricas de qualidade visual e de áudio. As suas experiências demonstraram que as métricas de qualidade objetivas atuais ainda não podem substituir a avaliação subjetiva da qualidade, embora o seu desempenho seja relativamente promissor. Também no 3D visual, vários autores testam diferentes métricas objetivas. Em [24, 25] foi proposta uma adaptação de métricas de 2D para 3D. Todos concluíram que não havia uma métrica de qualidade entre as testadas que pudesse ser aplicadas na imagem sozinha de profundidade, fosse capaz de correlacionar fortemente com as avaliações do observador da percepção de profundidade em vídeo 3D.

As diferentes etapas de processamento que são necessárias numa cadeia de fornecimento TV-3D podem todas introduzir artefatos que podem criar problemas em termos da percepção visual humana. Em [26] foi destacada a importância de considerar a atenção visual 3D ao abordar questões relacionadas com factores humanos ao nível do 3D. Referem que a maioria dos trabalhos existentes referem-se apenas ao vídeo 2D e que a introdução de informação de disparidade pode afetar o desenvolvimento da atenção visual. Como a percepção de profundidade desempenha um papel importante no nosso comportamento de atenção ao assistir conteúdo 3D, o entendimento e modelagem da atenção visual 3D torna-se muito relevante. Experiências de Eye-tracking podem ser conduzidas para a identificação dos locais de interesse visual no conteúdo (saliência). Pesquisas sobre atenção visual têm cada ganho cada vez mais popularidade. No entanto, em comparação com a quantidade de trabalhos em imagens fixas, há relativamente poucos estudos de investigação da atenção visual em sequências de movimento. Além disso, apenas um número muito pequeno de obras relacionadas com a atenção visual em conteúdo 3D estereoscópico podem ser encontrados actualmente na literatura [27-31]. Estas obras fornecem alguma evidência de que os resultados de estudos anteriores, utilizando estímulos 2D não podem ser automaticamente generalizados aos estímulos 3D, como a introdução de informações de disparidade pode mudar o desenvolvimento da atenção visual.

Capítulo 3

O Sistema Visual Humano

3.1 Percepção Humana da Cor

3.1.1 A luz

A luz é fundamental para a visualização de cores, sendo a energia que os nossos olhos detectam quando vemos. Quando não há luz, não somos capazes de ver. Não vemos os objectos em si, mas sim a luz que é reflectida ou transmitida pelos objectos. A luz ocorre em ondas, esta é visível com um comprimento de onda entre os 400 e os 700 nm (espectro visível), dentro deste comprimento pode ser detectada pelo olho humano e é designada por luz monocromática. Microondas e raios infravermelhos têm comprimentos de onda maiores que a luz visível, enquanto que os raios ultravioleta, raios-x e raios gama têm comprimentos de onda mais curtos do que a luz visível, podemos visualizar o espectro electromagnético na figura 3.1.

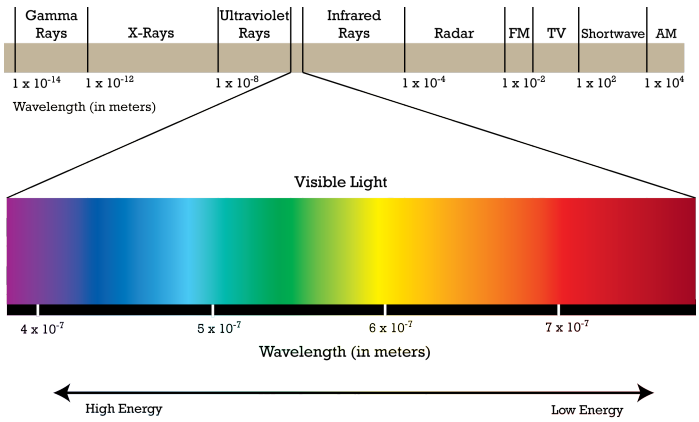


Figura 3.1: Espectro electromagnético.

3.1.2 O olho

Cabe aos nosso olhos captar a luz que nos permite ver.

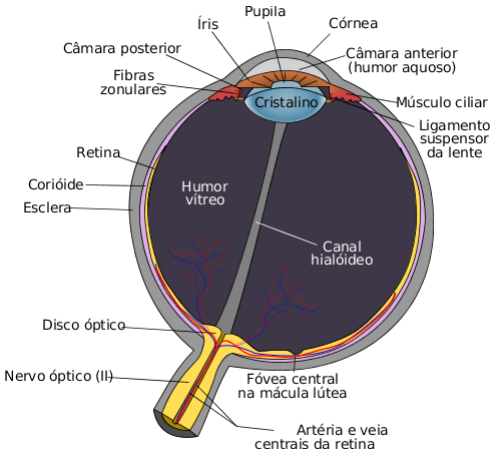


Figura 3.2: Representação esquemática da secção transversal do olho humano.

A esclera é a parte externa do olho, esta mantém a forma do olho , e protege as partes internas do olho.

A íris é a parte colorida do olho, é composta por tecidos musculares que expandem ou contraem para alterar o tamanho da pupila onde passa luz para dentro do olho, é semelhante à abertura de uma câmara. A íris ajusta o tamanho da pupila consoante a intensidade da luz.

A córnea é a membrana transparente em frente do olho, tem a função de focar a luz através da pupila para a retina, como se fosse uma lente fixa.

O cristalino possui uma estrutura biconvexa, gelatinosa, possuindo grande elasticidade que diminui progressivamente com a idade. O cristalino cresce continuamente durante a vida do indivíduo. É relativamente plano quando estamos a observar objectos distantes, porque a luz é paralela, e curva-se quando estamos a observar objectos próximos, porque a luz é dispersada.

A retina é considerada como o principal mecanismo da visão, porque é onde os receptores sensoriais visuais estão localizados. Ela pode ser comparado como o filme de uma câmara, onde as imagens são gravadas. A retina contém cerca de cerca de 126 milhões de receptores sensoriais visuais de dois tipos - cones e bastonetes.

Os bastonetes são finos e longos, funcionam sob baixa intensidade de luz e só podem registrar imagens a preto e branco. Há cerca de 120 milhões de bastonetes na retina de um só olho.

Os cones são curtos e largos, funciona sob alta intensidade de luz e pode registrar imagens coloridas. Há cerca de 6 milhões de cones na retina de um só olho. A principal razão pela qual se diz que os cães são daltónicos é porque que eles não têm cones na sua retina.

A imagem captada pela nossa retina é então transduzida (ou electroquimicamente descodificada) pelas células bipolares, também localizadas na retina. Células ganglionares agem então como nervos aferentes que enviam informação sensorial visual para o cérebro. Os axônios das

células ganglionares reúnem-se e formam o nervo óptico. Como este o nervo óptico não contém cones ou bastonetes, que também é conhecido como o ponto cego.

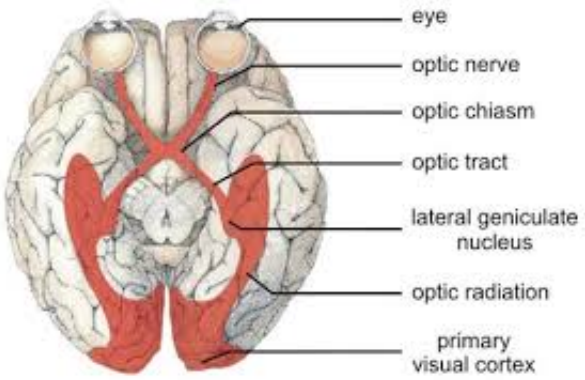


Figura 3.3: Representação esquemática da secção transversal do olho humano.

Os nervos ópticos cruzam-se logo atrás da passagem nasal, identificado como o quiasma. Estre cruzamento mostra que a informação sensorial visual registrada pelo olho esquerdo é transmitida para o córtex visual direito. Este cruzamento de nervos aferentes também é observado em sentidos auditivos, cutâneos, cinestésicos e vestibulares.

3.1.3 A visão de cores

A abordagem da psicologia evolucionária sustenta que desenvolvemos visão de cores devido à necessidade que temos de distinguir quais os alimentos que estão maduros e comestíveis, e quais não são. Explicações teóricas sobre a visão de cores foram principalmente derivadas de estudos psicológicos precoces, mas os métodos anatómicos actuais suportam essas teorias. Duas teorias principais da visão de cores são a Trichromatic Theory e a Opponent-Process Theory.

- **Trichromatic Theory -** Proposto por Thomas Young (1802) e, posteriormente, prorrogado por Hermann von Helmholtz (1852), a trichromatic theory baseia-se no pressuposto de que a visão de cor é devido ao funcionamento colectivo de sistemas receptores diferentes para as cores azul, vermelho e verde. Virtualmente todos os comprimentos de onda de luz visível individuais podem ser copiados através da combinação de comprimentos de ondas azuis, vermelhas e verdes em graus variados, segue-se que a visão de cor pode ser avaliada pela correspondência correcta de comprimentos de onda simples e complexos da mesma cor. Experiências com capacidades de correspondência de cores demonstraram que dicromatas (possuem apenas dois sistemas de cone funcionais), especialmente aqueles com mau funcionamento no sistema de cone verdes, são bastante comuns. Eles são desnecessariamente, mas normalmente, conhecidos como daltónicos.
- **Opponent-Process Theory -** A Opponent-Process Theory, proposta pelo fisiologista alemão Ewald Hering (1878), basicamente, afirma que o fenómeno pós-imagem reflecte que vemos cor em pares complementares, vermelho-verde e azul-amarelo. O fenómeno pós-imagem acontece quando se observa muito tempo a cor verde e, em seguida, observar vermelho quando se desvia o olhar. Embora um pouco diferente da trichromatic theory, é agora amplamente reconhecido que as duas teorias trabalham em diferentes níveis, isto é, os pares complementares resultam da transdução de cones tricromáticos nas células ganglionares.

3.1.4 Percepção das dimensões visuais

As dimensões visuais incluem a forma, a profundidade, o movimento e a constância, e nós percebemo-las de diferentes maneiras:

- Percepção da forma: Nós percebemos formas através da ajuda de contornos e padrões. Um contorno é onde ocorre uma mudança súbita do brilho. Por exemplo, percebemos a forma geométrica cilíndrica porque observamos que o lado esquerdo e direito diminuem gradualmente de brilho. Padrões, por outro lado, são utilizados para a percepção de organização. A psicologia "Gestalt"ressalta que o conjunto não é igual à soma das suas partes. (Nota: "Gestalt"é a palavra alemã para "forma"ou "configuração"). Princípios comuns da psicologia "gestalt"aplicados na percepção da forma são a relação figura-fundo, closure, proximidade e similaridade. Relação figura-fundo significa que nós temos a percepção de uma forma, identificando o que é um objecto e o que é um terreno, e comparando-os um com o outro. "Closure"é o preenchimento dos espaços de figuras desconexas e incompletas. Isto significa que um círculo incompleto ainda será percebido como um círculo. Naturalmente nós fechamos as aberturas ausentes para formar uma forma. Proximidade e similaridade são princípios de agrupamento. Proximidade é o agrupamento por proximidade, enquanto que similaridade é o agrupamento por semelhança. Por exemplo, uma espiral é formada pelo agrupamento de linhas circulares em proximidade, ou uma flor pode ser formada por uma semelhança na costura transversal.

- Percepção de profundidade: Percebemos a profundidade com a ajuda de pistas binoculares e monoculares. Pistas binoculares vêm da disparidade entre o olho esquerdo e o olho direito. Porque o olho esquerdo e o olho direito gravam informações sensoriais visuais diferentes devido à sua localização, o cérebro processa duas imagens diferentes para dar um sentido da profundidade de objectos visuais. Os estereogramas são adaptações de pistas binoculares. Por causa das variações individuais de posicionamento ocular, ou seja, alguns olhos são amplamente separados e alguns são muito próximos, é preciso algumas adaptações para ver imagens de estereogramas. Por outro lado, as pistas monoculares são fornecidas por um único olho. Elas também são conhecidas como pistas pictóricas porque os artistas aplicam-nas para imitar uma imagem tridimensional para uma plataforma bidimensional. É por isso que temos a percepção de profundidade, mesmo se a tela for plana. Pistas monoculares, como tamanhos familiares, altura no campo de visão, perspectiva linear, sobreposição, sombra e gradiente de textura, permite-nos perceber a profundidade com um único olho. Os tamanhos familiares vêm da experiência de cada pessoa. Sabemos com certeza pela experiência que os edifícios são mais altos do que os carros. Altura no campo de visão significa que os objectos colocados em posição mais elevada sejam considerados como estando mais longe. Esta é a razão pela qual percebemos o tamanho da lua de forma diferente em diferentes localizações. Porque não temos uma experiência familiar sobre seu tamanho real, percebemos a lua como mais distante e menor quando se encontra acima de nós do que quando está perto do horizonte. Perspectiva Linear significa que os objectos distantes ocupam menos espaço na retina, então percebemos linhas convergentes mais distantes do que linhas paralelas. Sobreposição significa que o objecto está mais próximo em relação ao objecto escondido. Sombra faz uso da iluminação e localização do objecto. Gradiente de textura significa que objectos mais densos e com uma textura mais fina estão mais longe que os objectos com uma textura mais leve e mais grossa. Os cartunistas muitas vezes usam o gradiente de textura para dar profundidade

nos seus desenhos.

- Percepção do movimento: As observações de Baylor (2001), "Quanto mais burro o animal, mais inteligente será a retina", porque as rãs e outros animais simples conseguem detectar o movimento utilizando apenas as suas retinas. Como os seres humanos são mais complexos e mais especializados, usamos extensos estímulos ambientais para perceber o movimento. Ao contrário das rãs, nós usamos um número de receptores sensoriais, visual, auditivo e outros, particularmente sentidos cinestésicos e vestibulares para perceber o movimento externo. No entanto, uma porção significativa da informação sensorial necessária para os seres humanos perceberem o movimento vem do sistema visual. A Disneylândia utiliza o conceito de como o olho humano detecta movimento, a fim de produzir a ilusão de movimento aparente. O movimento aparente vem em duas formas, estroboscópicos e efeito colateral. Movimento estroboscópico é conseguido através de estímulos rápidos nas diferentes partes da retina. Movimento colateral ocorre como um resultado da observação de movimento contínuo, onde outra superfície move-se na direcção oposta.
- Percepção de constância: Mesmo que um edifício pareça menor quando está longe, ainda sabemos que é alto. Mesmo que uma porta esteja aberta e só podemos ver sua porção fina vertical, ainda sabemos que é um rectângulo plano. Mesmo que seja escuro, ainda sabemos que as folhas são na maioria verdes. Temos a percepção de constância apesar da variação de várias sensações com a ajuda da experiência e da memória.

Capítulo 4

Sistema de Aquisição de Mapas de Atenção

4.1 Equipamento utilizado

4.1.1 Dois Computadores

Nesta experiência, devido a limitações de hardware tiveram que ser usados dois computadores.

Um Computador principal onde está ligada a Visage, com uma aplicação criada em Matlab, que também serve de servidor para outro computador se ligar e funcionarem sincronizadamente por rede, esta aplicação irá ter controlo da Visage que irá controlar o Eyetracking bem como o Switch.

Um segundo Computador que possui também uma placa Gráfica NVIDIA e que permite a reprodução de vídeos 3D, possui uma aplicação criada em Matlab, que é o cliente desta experiência, esta aplicação irá controlar a inicialização e o tempo dos vídeos 3D da experiência.

4.1.2 Visage

A Visage é um dispositivo que proporciona uma maneira simples de exibir estímulos visuais calibrados em um monitor CRT de computador com precisão de tempo, e fornece um mecanismo robusto e fiável para sincronizar a apresentação de estímulos com equipamentos externos de recolha de dados, como por exemplo Eyetrackin.

A Visage é necessária neste Projecto, já que o nosso objectivo é saber quais as zonas do vídeo em cada "frame"que são o alvo de atenção enquanto é mostrado um vídeo 3D a um observador, a precisão entre a exibição do estímulo e a recolha de dados é um ponto muito importante, sendo que a Visage oferece as seguintes vantagens:

- Um driver em tempo real para o Windows garante apresentações de frame-síncronas com frame rates superiores a 100Hz;
- Cor 14 bits e controle de luminância com suporte integrado para correcção de gama e calibração de cores;
- Possui um framestore PCI Express dedicado para guardar sequências de imagens pré-calculadas;
- Tem uma interface digital I/O integrada para controlar equipamentos de terceiros;
- Possui uma Toolbox para MATLAB CRS e está preparada para ser executada em ambiente MATLAB para flexibilidade total, que nos dá liberdade de construir a nossa própria aplicação.

Uma imagem da visage com a respectiva descrição é apresentada na figura 4.1.

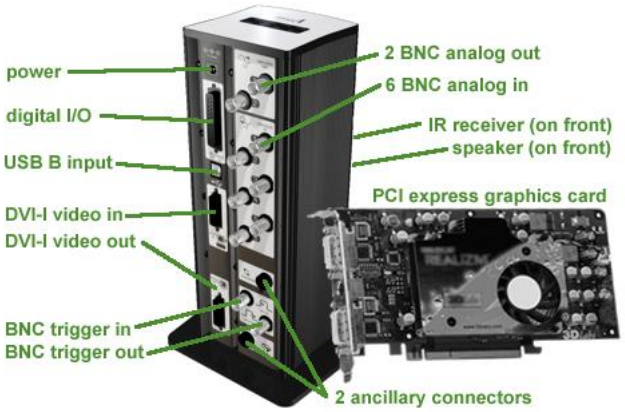


Figura 4.1: Visão global da Visage.

4.1.3 High-Speed Video Eye Tracker

A High-Speed Video Eye Tracker é uma ferramenta de pesquisa oculomotor que é ao mesmo tempo simples de usar, e faz medições precisas e com baixo ruído, sem ignorar frames. O HS-VET incorpora recursos de hardware exclusivos para este servir de interface para outros equipamentos. Possui entradas analógicas e digitais que podem ser utilizados para sincronizar os eventos externos. O HS-VET pode ser usado com a Visage, um gerador de estímulo visual. Possui uma Toolbox para MATLAB para podermos fazer uma gravação dos movimentos oculares com a nossa própria aplicação.

A imagem do sistema de eye tracking usado neste projeto está representado na figura 4.2.

Figura 4.2: High-Speed Video Eye Tracker.

4.1.4 Sistema de reprodução de conteúdo 3D

Foi adquirido hardware para ser possível a visualização de conteúdo 3D estereoscópico neste projecto. Assim é usada uma placa Gráfica NVIDIA GeForce GTX 550 Ti, esta placa gráfica juntamente com o software fornecido pela NVIDIA permite a reprodução de conteúdo 3D. Para podermos visualizar esse conteúdo 3D foi necessário adquirir uns óculos com shutter activo e

usar um monitor com certas especificações, o software da NVIDIA irá controlar o shutter dos óculos para criar o efeito 3D dos vídeos estereoscópicos. O efeito shutter dos óculos activos funcionam da seguinte forma: ao apresentar a imagem destinada para o olho esquerdo no ecrã o shutter direito dos óculos bloqueia a visão no olho direito, e depois apresentando a imagem destinada ao olho direito no ecrã o shutter esquerdo bloqueia a visão no olho esquerdo. Sendo estas repetições suficientemente rápidas para que as interrupções não interfiram com a fusão percebida das duas imagens para uma única imagem 3D. Para este efeito é necessário um monitor com uma *refresh rate* acima dos 100Hz.

4.1.5 Monitor Sony

Na experiência é utilizado o monitor Sony GDM-F520 (figura 4.3). Este é um monitor CRT de 21", que possui uma resolução máxima de 2048 x 1536 e possui também uma *refresh rate* vertical até 170Hz, sendo o mínimo 100Hz para podermos visualizar conteúdo 3D juntamente com o hardware e o software da NVIDIA este monitor é óptimo para a experiência

Figura 4.3: Monitor Sony GDM-F520.

4.1.6 Switch

Para esta experiência ser possível foi adquirido um switch (figura 4.4) para efectuar a comutação no monitor de dois computadores. Este switch possui duas portas VGA e um comando para efectuar a comutação no ecrã. Visto que um computador possui o sistema de Eyetracking e outro computador o sistema de reprodução de conteúdo 3D estes têm que estar ligados e serem visualizados no mesmo monitor, para tornar a comutação entre os dois computadores no monitor o mais sincronizada com o Eyetracking e a reprodução dos vídeos 3D um método de comutação manual não poderia ser utilizado. Então foi criado um método eléctrico que interage com a aplicação criada em MATLAB. O switch foi adaptado para funcionar como um periférico da Visage, através de inputs enviados para a Visage através da aplicação esta irá controlar o switch para ser possível a comutação dos dois computadores no monitor de forma automática.

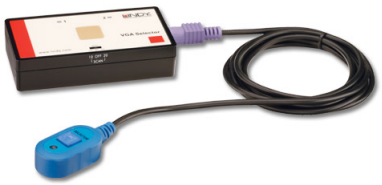


Figura 4.4: Switch com duas portas VGA.

4.1.7 PR-650 SpectraScan Colorimeter

O SpectraColorimeter PR-650 é baseado em espectro luminância telephotometer/colorímetro. Executa medições completas, fotométricos e colorimétricos são realizadas spectroradiometricamente. Supera os erros espectrais inerentes a filtros de instrumentos. O PR-650 mede a radiação óptica espectral em vez de confiar nas tecnologias de filtro. Serve administrar com precisão a medição de resultados importantes, tais como a luminância e cromaticidade - PR-650 determina esses parâmetros através da medição da intensidade absoluta em cada comprimento de onda, em seguida, calcula o valor apropriado da CIE.

Figura 4.5: PR-650 SpectraScan Colorimeter.

4.2 Linguagem e ferramentas usadas

4.2.1 Matlab

MATLAB (Matrix Laboratory) na figura 4.6, é um ambiente de computação numérica e uma linguagem de programação alto nível. Desenvolvido pela MathWorks, MATLAB permite manipulações de matrizes, fazer o plotting de funções e dados, implementação de algoritmos, a criação de user interfaces e interagir com programas escritos em outras linguagens, incluindo C, C + +, Java, e Fortran.

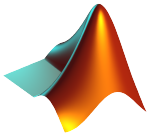


Figura 4.6: Logotipo MATLAB.

Embora MATLAB seja destinado principalmente para computação numérica, existem várias toolbox opcionais que usam o mecanismo algébrico MuPAD, permitindo o acesso a recursos de computação algébrica. Um pacote adicional, Simulink, acrescenta a simulação de multi-domínio gráfico e Design Model-Based para sistemas dinâmicos e incorporados.

A aplicação MATLAB é construída com base na linguagem MATLAB, é possível executar código na janela de comandos da aplicação ou executar ficheiros de texto com código MATLAB incluindo scripts e/ou funções. Há a possibilidade de chamar directamente bibliotecas escritas em Java a partir do MATLAB o que era essencial para a criação da aplicação deste projecto.

Esta linguagem foi escolhida porque foi idealizado e implementado com sucesso um sistema automático para avaliação e captura de mapas de atenção de vídeos 3D. Como o software fornecido pelo fabricante do "EyeTracker"e da Visage é na linguagem Matlab, não houve outra alternativa senão criar o sistema nesta linguagem.

4.2.2 Socket

Normalmente, um servidor é executado em um computador específico e tem um socket associado a uma porta específica. O servidor fica à espera que um cliente faça um pedido de conexão .

No lado do cliente: O cliente sabe o hostname da máquina na qual o servidor está a ser executado e o número da porta na qual está à escuta. Para fazer um pedido de conexão, o cliente tenta conectar-se com o servidor nesse hostname e porta específica. O cliente também se identifica ao servidor. Para isso ele liga-se a porta local que irá utilizar durante esta ligação. Isto é geralmente designado pelo sistema, esquema na figura 4.7.

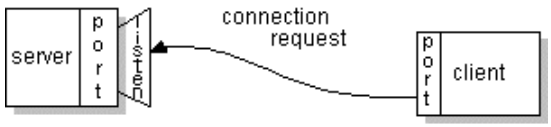


Figura 4.7: Pedido de conexão do cliente ao servidor.

O servidor aceita a conexão, figura 4.8. Após a aceitação, o servidor recebe um novo socket associado à mesma porta local e também tem o seu endpoint remoto definido para o endereço e porta do cliente. É preciso um novo socket para que possa continuar à escuta no socket original para outros pedidos de conexão enquanto processa as necessidades do cliente conectado.

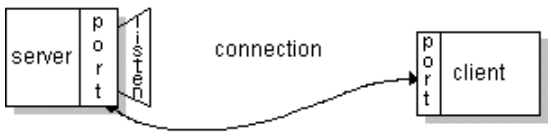


Figura 4.8: Conexão aceite entre o servidor e o cliente.

No lado do cliente, se a conexão for aceita, é criado um socket com sucesso e o cliente irá usar esse socket para comunicar com o servidor.

O cliente e o servidor podem agora comunicar através de escrita e leitura dos sockets associados.

Um socket é um endpoint de uma comunicação bidirecional entre dois programas em execução na rede. Um socket é ligado a uma porta de para que a camada de TCP possa identificar a aplicação são destinados os dados.

Um endpoint é uma combinação de um endereço IP e um número de porta. Cada conexão TCP pode ser identificada exclusivamente por dois endpoints. Dessa forma, poderemos ter várias conexões entre a nossa máquina e o servidor.

O pacote java.net na plataforma Java fornece uma classe, denominada Socket , que implementa um lado de uma conexão bidirecional entre um programa Java e outro programa na rede. A classe Socket fica no topo da implementação de uma plataforma dependente, escondendo os detalhes de qualquer sistema específico de um programa Java. Ao usar a classe java.net.Socket em programas Java podemos comunicar através da rede.

Além disso, java.net inclui a classe ServerSocket , que implementa um socket para que os servidores possam utilizar para ficar à escuta e aceitar conexões de clientes.

4.2.3 Class Robot

Essa classe é usada para gerar eventos de input do sistema para fins de teste de automação, por exemplo, aplicações onde é necessário controlar o rato e o teclado. O objectivo principal da classe Robot é facilitar testes automatizados de implementações na plataforma Java. Permite simular as acções de um utilizador, como mover o rato, pressionar teclas continuamente ou não, pressionar os botões do rato, etc. As acções mencionadas anteriormente são as necessárias para o correcto funcionamento da aplicação cliente deste projecto.

4.2.4 Detecção de eventos (Teclas Pressionadas)

Para a execução deste teste foi necessário a utilização, por parte do observador, de um dispositivo que interagisse com o software. O observador dispunha de um teclado wireless situado entre o monitor e o mesmo.

As funcionalidades gerais do teclado foram desactivadas, todas as teclas quando pressionadas ficavam sem efeito através do controlo das teclas pressionadas. Apenas três teclas ficavam activas quando pressionados, é o caso da tecla arrow left, a tecla arrow right e a tecla espaço. O observador usando as teclas arrow left e arrow right iria orientar para a esquerda e para a direita a sua avaliação na barra, quando o utilizador seleccionava o seu valor pretendido para a avaliação, validava a resposta pressionando a tecla espaço, caso o observador não avaliasse num espaço de tempo de 10 segundos, o sistema continuava e iria guardar o vídeo visualizado anteriormente e repetir a sua visualização no fim da experiência para uma nova avaliação sem conhecimento do observador.

4.3 Funcionamento Geral

Com os recursos de hardware , linguagens e ferramentas usadas sumarizadas procedemos à descrição do funcionamento geral do sistema. Como já foi referido os dois computadores que controlam o hardware necessário à realização da experiência estão ligados em rede, desta forma foi criado um sistema automático e quase autónomo. Para a recolha de dados feita com o "EyeTracker"ser precisa, o sistema precisa de uma intervenção humana para controlar a calibração do "EyeTracker", fora isso o sistema é completamente automático.

Para iniciar o sistema temos que conectar o cliente (PC 3D) ao servidor (PC da Visage), após a ligação ser feita é pedido no servidor o vector aleatório da sequência de vídeo a ser visualizada pelo observador, ao mesmo tempo a sequência aleatória a ser visualizada tem que ser inicializada no player estereoscópico no cliente em modo janela para não activar o efeito "shutter"dos óculos. De seguida é feita a calibração pupilar inicial do observador.

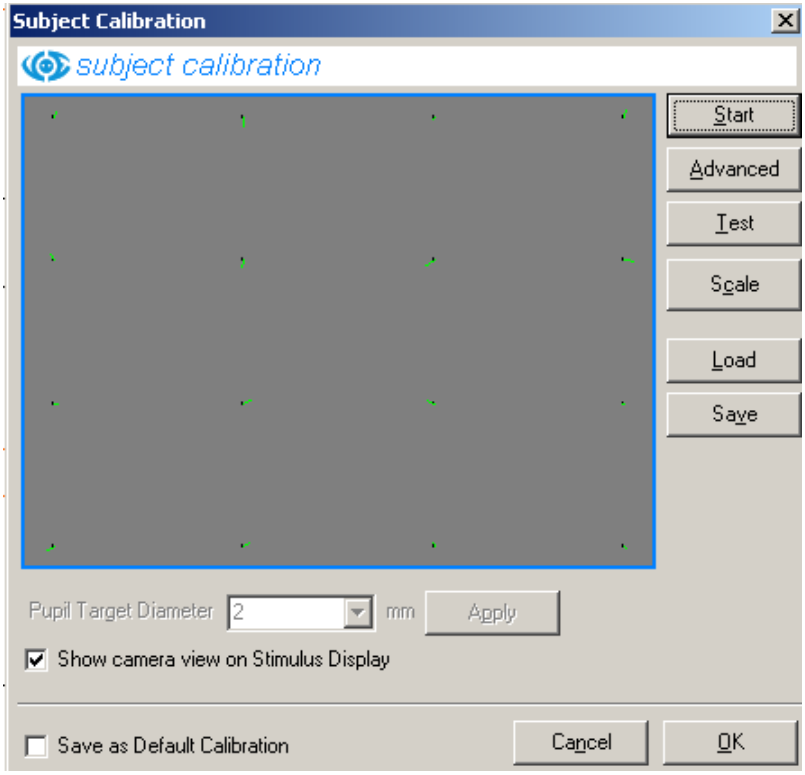


Figura 4.9: Exemplo de uma boa calibração.

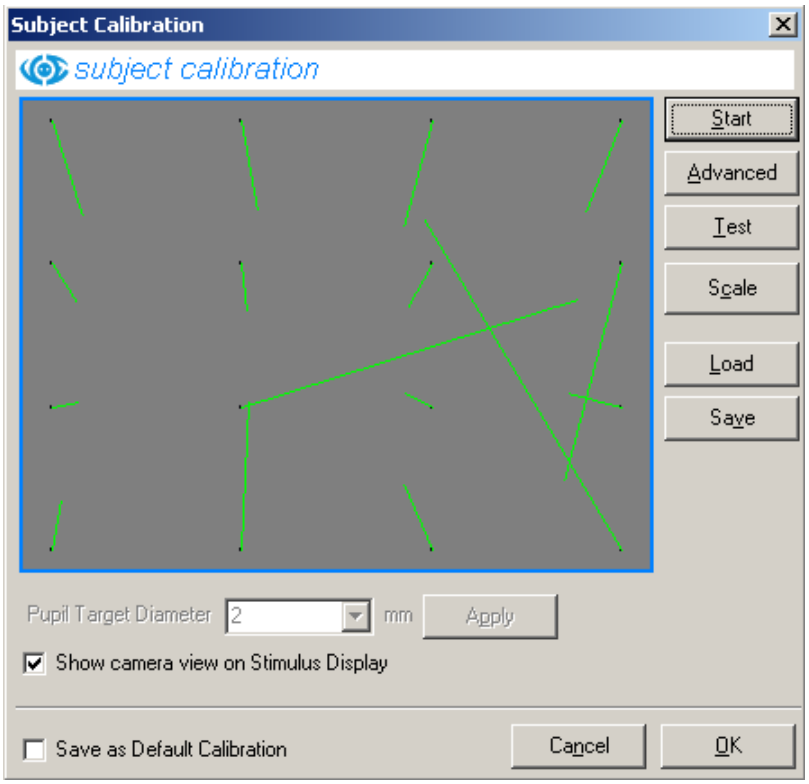


Figura 4.10: Exemplo de uma má calibração.

Se a calibração for boa, então o sistema deverá prosseguir, no entanto se a calibração for má deverá repetir-se até que se obtenha uma boa calibração. Quando o sistema prossegue o sistema faz a comutação dos monitores e maximiza o vídeo automaticamente para ser visualizado em full screen. Antes de dar inicio ao vídeo, o rato é deslocado para a extermidade inferior do ecrã para não influenciar a atenção do estímulo. Quando o vídeo começa, ao mesmo tempo começa a recolha de dados do "EyeTracking". Quando o tempo de visualização do estímulo termina, o sistema faz a comutação do monitor novamente e irá aparecer a janela com a escala de avaliação do estímulo visualizado, ao mesmo tempo o player estereoscópico passa para o modo de janela para o shutter não ficar activo durante a observação na parte da visage.

Se o observador não avaliar em 10 segundos, o sistema continua e irá repetir o estímulo por avaliar no final da sequência aleatória. Se o obsevador avaliar, o tempo restante dos 10 segundos é respeitado com um ecrã cinzento para limpar a memória do estímulo visualizado anteriormente. De seguida é mostrada a imagem 4.11 para verificar a calibração do "EyeTracker", é pedido ao observador que olhe para as letras alternadamente, é dada uma opção para fazer uma nova calibração do "EyeTracker", se os pontos de fixação coincidirem com as bolas das letras o sistema não precisa de ser calibrado, senão coincidirem teremos que efectuar uma nova calibração.

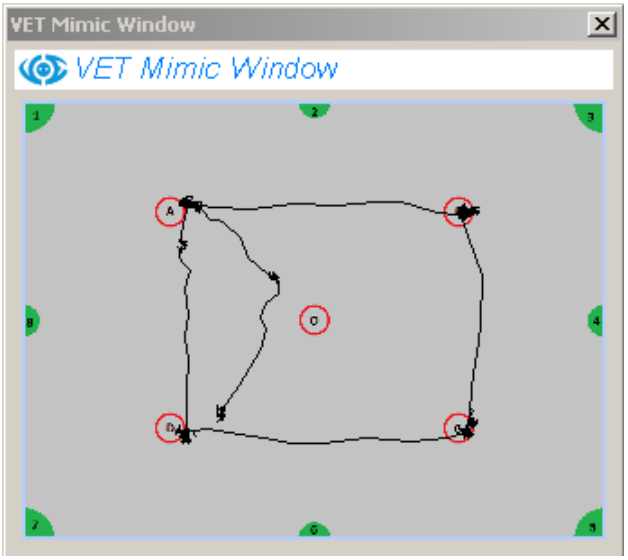


Figura 4.11: Imagem usada para verificar a calibração do "EyeTracker".

O processo descrito é repetido até a sequência aleatória do observador terminar, se houve vídeos que não receberam avaliação estes serão repetidos sem que o observador dê conta disso para não influenciar as avaliações subjectivas.

Inicialmente os vídeos com uma recolha de tracking inferior a 60% eram repetidos no fim da sequência aleatória com o intuito de obter uma recolha de tracking superior a 60%. Mas testes comprovaram que este método acabaria por ser mau, visto que aumentava em muito o tempo de uma sessão de teste, o cansaço das pessoas reflectivo por uma sessão viria a influenciar os dados dos mapas de atenção. Por isso, este método foi removido para as experiências.

Também foi criada uma forma diferente de o observador avaliar o estímulo, o observador em vez de usar o teclado para classificar a sua avaliação na barra, este usaria o olhar. Este método parecia ser o mais eficaz, porém os problemas do "EyeTracker"descritos na secção **??** iriam influenciar a avaliação do observador se durante a avaliação o sistema perde-se o tracking do olho.

Capítulo 5

Metodologia

Para criarmos a nossa base de dados de vídeos com diferentes erros cromáticos foi necessário converter os vídeos do espaço de cor RGB para o espaço de cor CIELAB. Para tal convertemos primeiro do espaço de cor RGB para o espaço de cor XYZ e posteriormente para o espaço de cor CIELAB.

Na conversão do espaço de cor XYZ para o espaço de cor CIELAB usámos um iluminante diferente do comum D65. Foi usado o ponto branco de referência do monitor da experiência (Sony FD250) como iluminante de referência.

Num estudo anterior, Aldaba et al. [34] segmentou o espaço CIELAB em cubos. Em cada cubo era aplicado um erro ΔE^*_{ab} e foi concluído que a percepção da existência de um erro ΔE^*_{ab} era maior que 5 unidades. Contudo, este método produzia artefactos espaciais, que tinham grande influência na avaliação dos observadores maior que as alterações da cor. Assim, é possível quantificar se a avaliação fornecida depende exclusivamente da mudança de cor, que foi o objectivo do estudo.

Para evitar os efeitos de artefactos espaciais, as cores das imagens foram subdivididas em clusters aplicando o algoritmo do K-Means em vez de cubos, feito num estudo anterior, Marco et al. [35] Para encontrarmos o valor de K que nos levaria a melhores resultados, foram feitos vários testes, o número de K iria variar entre 3 e 9. Os resultados finais mostraram que K = 4 formava grupos de cores similares, reduzindo a influência de artefactos espaciais. Cada pixel $(L^*_i a^*_i b^*_i)$ corresponde um cluster de cor e foi adicionado um erro ΔE^*_{ab} com uma magnitude pré-definida e uma direcção randómica para obter o novo pixel $(L^*_i a^*_{ei} b^*_{ei})$.

$$\Delta E^*_{ab} = \sqrt{(L^*_i - L^*_i)^2 + (a^*_i - a^*_{ei})^2 + (b^*_i - b^*_{ei})^2} \quad \textbf{(5.1)}$$

Este procedimento foi aplicado a todos os clusters de cores, mantendo o erro cromático, mas aplicando direcções aleatórias. Assim, foi garantido que grupos de cores similares foram alterados na mesma direcção. As magnitudes de erro ΔE^*_{ab} variaram entre 6 e 18, com saltos de 6 níveis. De um conjunto de 6 vídeos foram gerados 42 vídeos para cada grandeza de erro pré-definida para cobrir um grande número de direcções no espaço de cor CIELAB.

Considerando-se que o sistema visual humano é mais sensível à luminância e considerando as limitações da gama de cores do monitor, a luminância L* das frames de cada vídeo permaneceram sempre com o mesmo valor. Contudo, nem todos os pixéis poderiam ser reproduzidos com precisão. Então, eles foram reproduzidos aplicando um clipping para a cor mais próxima da superfície da gama de cores no espaço CIELAB do monitor. O número de pixéis clipados era, em média, inferior a 2%. No fim foram obtidos 48 vídeos, incluindo 6 amostras dos vídeos originais (ΔE^*_{ab}=0).

5.1 Pré-Processamento da Base de Dados

A base de dados de vídeos da experiência passou inicialmente por uma fase de pré-processamento. Isto porque os vídeos originais da base de dados de NAMA3DS1-COSPAD1 são vídeos Full HD progressivo com uma resolução de 1920x1080 e 25 frames por segundo. Cada vídeo encontra-se no formato Uncompressed AVI com um tamanho de 1.5GB por vídeo. Para podermos representar uma sequência de um vídeo seria preciso reproduzir no Stereoscopic Player dois vídeos (vídeo esquerdo e vídeo direito) totalizando um total de 3GB. Com o nosso software tal seria impossível, a reprodução do vídeo apresentava um efeito de slowmotion.

Alteramos então o tipo de compressão de cada vídeo para AVI com perdas com uma qualidade de 100% para reduzir o máximo de perdas de informação de cada vídeo, as perdas de informação ao mudar o tipo de compressão de cada vídeo foram verificadas e foi provado não haver alterações a olho nu, visto que também ao efectuar os cálculos das perdas a nível computacional provou-se que essas perdas eram muito reduzidas.

Posteriormente redimensionámos os vídeos para uma resolução de 1152x648 e acrescentámos uma parte em cima e em baixo de cada frame com a cor preta em RGB (0, 0, 0), no fim cada vídeo apresentava a resolução de 1152x864, sendo esta a resolução do display da experiência. Ao termos os vídeos com uma resolução espacial de 1152x864 podemos relacionar os pontos espaciais obtidos por tracking de forma directa, após a conversão de mm para pixéis, visto que a resolução espacial dos vídeos é a mesma que a resolução do display.

Após estas transformações todas pudemos reproduzir de maneira correcta cada vídeo no Stereoscopic Player, em termos de tamanho houve uma redução em média de 80% para cada vídeo, ou seja, em vez de reproduzirmos uma sequência de um vídeo com 3GB a sequência no total teria 300MB, o que é uma grande diferença em termos computacionais.

5.2 Conversão para o espaço de cor CIE LAB

Os vídeos inicialmente encontravam-se no espaço RGB, mas para aplicarmos o erro cromático pré-definido teríamos que convertê-los para o espaço CIE LAB. Para tal, os vídeos passaram por várias transformações.

Primeiramente os valores do espaço RGB de cada vídeo foi convertido para o espaço sRGB que possui uma gama de valores entre [0.0 1.0], foi feita esta transformação para posteriormente a gama de valores do espaço CIE LAB de cada vídeo se encontrarem entre [0.0 100.0]. (L*), e [-100.0 100.0]. (a* e b*). Ao para o espaço sRGB, a conversão para o espaço de cor XYZ terá os seus valores compreendidos entre [0.0 1.0].

$$R'_{sRGB} = R_{8bit}/255.0 \quad (5.2)$$

$$G'_{sRGB} = G_{8bit}/255.0 \quad (5.3)$$

$$B'_{sRGB} = B_{8bit}/255.0 \quad (5.4)$$

Se R'_{sRGB}, G'_{sRGB}, $B'_{sRGB} \leq 0.04045$

$$R_{sRGB} = R'_{sRGB}/12.92 \quad (5.5)$$
$$G_{sRGB} = G'_{sRGB}/12.92 \quad (5.6)$$
$$B_{sRGB} = B'_{sRGB}/12.92 \quad (5.7)$$

se R'_{sRGB}, G'_{sRGB}, $B'_{sRGB} > 0.04045$

$$R_{sRGB} = \left[(R'_{sRGB} + 0.055)_{/1.055}\right]^{2.4} \quad (5.8)$$
$$G_{sRGB} = \left[(G'_{sRGB} + 0.055)_{/1.055}\right]^{2.4} \quad (5.9)$$
$$B_{sRGB} = \left[(B'_{sRGB} + 0.055)_{/1.055}\right]^{2.4} \quad (5.10)$$

Antes de ser feita a conversão para XYZ teremos que calcular a matriz de transformação [M] que é calculada a partir das referências primárias RGB do sistema.

Dadas as coordenadas cromáticas de um sistema RGB (x_r, y_r), (x_g, y_g) e (x_b, y_b) e sua referência do branco (X_W, Y_W, Z_W), este é o método para calcular a matriz para a conversão de RGB para XYZ:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \Big[M\Big] \begin{bmatrix} R_{sRGB} \\ G_{sRGB} \\ B_{sRGB} \end{bmatrix} \quad (5.11)$$

onde,

$$\Big[M\Big] = \begin{bmatrix} S_r X_r & S_g X_g & S_b X_b \\ S_r Y_r & S_g Y_g & S_b Y_b \\ S_r Z_r & S_g Z_g & S_b Z_b \end{bmatrix} \quad (5.12)$$

com,

$$X_r = x_r/y_r; \quad (5.13)$$
$$Y_r = 1; \quad (5.14)$$
$$Z_r = (1 - x_r - y_r)/y_r; \quad (5.15)$$
$$X_g = x_g/y_g; \quad (5.16)$$
$$Y_g = 1; \quad (5.17)$$
$$Z_g = (1 - x_g - y_g)/y_g; \quad (5.18)$$
$$X_b = x_b/y_b; \quad (5.19)$$
$$Y_b = 1; \quad (5.20)$$
$$Z_b = (1 - x_b - y_b)/y_b; \quad (5.21)$$
$$(5.22)$$

$$\begin{bmatrix} S_r \\ S_g \\ S_b \end{bmatrix} = \begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix}^{-1} \begin{bmatrix} Xw \\ Yw \\ Zw \end{bmatrix} \quad (5.23)$$

Para calcularmos a referência do branco (X_W, Y_W, Z_W) precisamos das coordenadas cromáticas do branco do sistema (x_w, y_w) e a luminância L em cd/m^2. Foi usado telespectroradiometro (PR-650 $SpectraColorimeter^{TM}$ - Photo Research, Inc., CA) para determinarmos as coordenadas cromáticas RGB e do branco e as luminâncias L em cd/m^2 respectivas do monitor usado na experiência (Sony F520), os dados recolhidos foram os seguintes:

Tabela 5.1: Dados do monitor recolhidos com o telespectroradiometro.

	x	y	L em cd/m^2
White	0.303	0.328	89.6
Red	0.624	0.342	20.0
Green	0.287	0.606	61.1
Blue	0.149	0.073	8.4

Durante a recolha de dados apenas a iluminação do monitor era presente no laboratório. A resolução do monitor era de 1152x854 com uma frame rate de 85 Hz e um scan rate de 77.10 Hz.

Tendo todos os dados necessários podemos prosseguir com os cálculos do ponto branco de referência do monitor (X_W, Y_W, Z_W).

$$X_W = \frac{x_w * Lw}{y_w} \quad (5.24)$$

$$Y_W = Lw \quad (5.25)$$

$$Z_W = \frac{(1 - x_w - y_w) * L_w}{y_w} \quad (5.26)$$

Depois de todos os cálculos terem sido efectuados foi obtida a seguinte matriz de transformação [M].

$$\Big[M\Big] = \begin{bmatrix} 0.4194 & 0.3216 & 0.1861 \\ 0.2298 & 0.6790 & 0.0912 \\ 0.0228 & 0.1199 & 0.9716 \end{bmatrix} \quad (5.27)$$

Por fim, com a normalização dos valores RGB e obtida a matriz de transformação [M] vamos então converter para o espaço de cor XYZ.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4194 & 0.3216 & 0.1861 \\ 0.2298 & 0.6790 & 0.0912 \\ 0.0228 & 0.1199 & 0.9716 \end{bmatrix} \begin{bmatrix} R_{sRGB} \\ G_{sRGB} \\ B_{sRGB} \end{bmatrix} \quad (5.28)$$

Ao termos os nossos vídeos no espaço de cor XYZ podemos então por conversão directa, converte-los para o espaço de cor CIE LAB, a seguinte transformação é mostrada nas funções abaixo. (X_r, Y_r, Z_r) são as coordenadas ponto branco de referência do monitor (X_W, Y_W, Z_W)

$$L = 116 f_y - 16 \quad \textbf{(5.29)}$$

$$a = 500\,(f_x - f_y) \quad \textbf{(5.30)}$$

$$b = 200\,(f_y - f_z) \quad \textbf{(5.31)}$$

$$f_x = \begin{cases} \sqrt[3]{x_r} & x_r > \varepsilon \\ \frac{\kappa x_r + 16}{116} & x_r \leq \varepsilon \end{cases} \quad \textbf{(5.32)}$$

$$f_y = \begin{cases} \sqrt[3]{y_r} & y_r > \varepsilon \\ \frac{\kappa y_r + 16}{116} & y_r \leq \varepsilon \end{cases} \quad \textbf{(5.33)}$$

$$f_z = \begin{cases} \sqrt[3]{z_r} & z_r > \varepsilon \\ \frac{\kappa z_r + 16}{116} & z_r \leq \varepsilon \end{cases} \quad \textbf{(5.34)}$$

$$x_r = \frac{X}{X_r} \quad \textbf{(5.35)}$$

$$y_r = \frac{Y}{Y_r} \quad \textbf{(5.36)}$$

$$z_r = \frac{Z}{Z_r} \quad \textbf{(5.37)}$$

Com:

$$\varepsilon = \begin{cases} 0.008856 & Actual\,CIE\,Standard \\ 216/24389 & Intent\,of\,the\,CIE\,Standard \end{cases} \quad \textbf{(5.38)}$$

$$\kappa = \begin{cases} 903.3 & Actual\,CIE\,Standard \\ 24389/27 & Intent\,of\,the\,CIE\,Standard \end{cases} \quad \textbf{(5.39)}$$

5.3 Cálculo das Key frames

Encontrando os nossos vídeos no espaço de cor CIE LAB, foi necessário efectuar outro passo importante nesta experiência, o cálculo das Key frames. Para podermos prosseguir para a divisão das cores de todas as frames de cada vídeo em clusters, precisamos seleccionar as Key frames de cada vídeo para serem usadas no algoritmo do K-means.

Embora fosse possível aplicar o K-means em cada frame de um vídeo vários testes demonstraram que este procedimento não seria a melhor escolha. A razão é simples, a decisão do K-means no espaço CIE LAB é feita usando as coordenadas L*, a* e b*, mas neste espaço de cor a variável que terá mais peso na decisão do algoritmo será a variável L*.

A variável L* corresponde à luminosidade de cada vídeo, na nossa base de dados a luminosidade varia muito em cada vídeo. Por exemplo, em um vídeo com um cenário de um carro

e uma cancela, as sombras criadas pela cancela e pela árvore fazem variar a luminosidade do carro em movimento, ou até mesmo o carro faz variar a luminosidade da estrada enquanto se move, então ao aplicarmos o algoritmo do K-means a cada frame para decidir os clusters, a olho nu notar-se-ia grandes variações nas áreas definidas pelos clusters em frames seguidas após a aplicação do erro cromático ΔE^*_{ab}.

Então para superar isto, foram seleccionadas as Key frames mais relevantes de cada vídeo para calcular os centroids de cada cluster do vídeo correspondente usando o algoritmo K-means, posteriormente foi calculada pixel a pixel de cada frame, a distância euclidiana entre cada centroid e feita a atribuição ao cluster correspondente com a menor distância euclidiana, assim cada frame teria os clusters definidos pelos centroids das Key frames mais relevantes do vídeo correspondente, tornando a decisão dos clusters de cada frame mais uniforme entre frames seguidas resolvendo assim o problema referido anteriormente.

As Key frames foram calculadas da seguinte forma, primeiramente calcula-se a diferença absoluta de cada frame com a frame anterior, com todas as diferenças absolutas calculadas iremos calcular a sua média e o seu desvio padrão. Depois com a média e o desvio padrão das diferenças iremos ficar o seguinte threshold:

$$threshold = \alpha STD * \beta MEAN \quad (5.40)$$

Foram feitos alguns testes para calcular o melhor threshold, para tal os valores de α e β variaram entre 1 e 4. No fim foi determinado que $\alpha = 1$ e $\beta = 2$ selecionavam as melhores Key frames.

As Key frames selecionadas para as sequências de vídeos na secção **??** foram as seguintes:

- Para a sequência Basket [297 298 299 300 301 302 303 304];
- Para a sequência Car [385];
- Para a sequência Hall [36];
- Para a sequência Umbrella [176 176 233 239 240];

É de notar que a sequência Basket possui mais Key frames que as outras por ter uma mudança abrupta da cena (rotação da câmera).

5.4 Kmeans

Nesta experiência como já foi referido foram feitos vários testes para verificar o número de K a usar para obtermos os melhores resultados possíveis. O número de K variou entre 4 e 9, com os testes feitos verificou-se que o número ideal seria K = 4, podemos observar na imagem que à medida que o número de clusters aumenta é mais provável ocorrerem artefactos espaciais, isto porque ao aplicarmos o erro cromático ΔE^*_{ab} em cada cluster, em clusters adjacentes verifica-se uma variação muito grande da cor, pelo que estes possuem cores muito semelhantes sem a aplicação do erro, levando assim a formação de artefactos espaciais.

5.5 Aplicação do erro cromático ΔE^*_{ab}

Com os clusters definidos em todas as frames de cada vídeo é a vez do processo da aplicação do erro cromático correspondente a cada cluster. Para tal, a cada pixel é adicionado um erro ΔE^*_{ab} com uma magnitude predefinida e uma direcção aleatória (α) atribuída ao cluster a que pertence. Como foram escolhidos 4 clusters para cada frame, foram gerados 4 $\alpha' s$ aleatórios sendo um para cada cluster n. Assim o erro adicionado a cada pixel será o seguinte.

$$L_i^{*\prime} = L_i^* + (\Delta E^*_{ab} * \cos(\Theta)) \quad (5.41)$$
$$a_i^{*\prime} = a_i^* + (\Delta E^*_{ab} * \cos(\alpha_n) * \sin(\Theta)) \quad (5.42)$$
$$b_i^{*\prime} = b_i^* + (\Delta E^*_{ab} * \sin(\alpha_n) * \sin(\Theta)) \quad (5.43)$$

Com:

$$\Theta = 90 \quad (5.44)$$
$$\alpha[0\ 360] \quad (5.45)$$

Assim, podemos verificar que o erro ΔE^*_{ab} apenas afectará as coordenadas a^* e b^*, mantendo assim os valores de L^* de cada píxel. Para cada vídeo foram recriados 3 vídeos com as magnitudes de erro ΔE^*_{ab} igual 6, 12 e 18. Foi feito um estudo para decidir as magnitudes dos erros ΔE^*_{ab} a aplicar, testes comprovaram que erros de diferença de 3 de magnitude não existia percepção da diferença da cor entre eles, ou seja, um vídeo com magnitude de erro 6 e um vídeo com magnitude de erro 9 não tinha diferenças da cor perceptíveis em relação ao observador. Então a magnitude dos erros escolhidos para a experiência foram 6, 12 e 18, o erro 18 em si já é um erro grande e muito perceptível então foi definido como o limite das magnitudes seleccionadas.

5.6 Cliping dos pixéis

Como já foi referido atrás, o monitor apresentar um gamut de cores limitado, ou seja, o monitor só consegue representar determinadas cores dentro da gamut dele. Por isso, para podermos ter a certeza que todos os pixéis dos vídeos poderiam ser reproduzidos com precisão foi aplicada uma escala ao L* de cada frame para reduzir o número de pixéis que não poderiam ser representados pelo monitor. Cada frame CIELAB foi convertida para o espaço de cor XYZ e posteriormente para o espaço de cor xyY.

Conversão de CIELAB para XYZ com (X_r, Y_r, Z_r) iguais ao ponto branco de referência do monitor (X_W, Y_W, Z_W):

$$X = x_r X_r \quad (5.46)$$
$$Y = y_r Y_r \quad (5.47)$$
$$Z = z_r Z_r \quad (5.48)$$

$$x_r = \begin{cases} f_x^3 & f_x^3 > \varepsilon \\ (116 f_x - 16)/\kappa & f_x^3 \leq \varepsilon \end{cases} \quad (5.49)$$

$$y_r = \begin{cases} ((L+16)/116)^3 & L > \kappa\varepsilon \\ L/\kappa & L \leq \kappa\varepsilon \end{cases} \quad (5.50)$$

$$z_r = \begin{cases} f_z^3 & f_z^3 > \varepsilon \\ (116 f_z - 16) & f_z^3 \leq \varepsilon \end{cases} \quad (5.51)$$

$$f_x = \frac{a}{500} + f_y \quad (5.52)$$

$$f_z = f_y - \frac{b}{200} \quad (5.53)$$

$$f_y = (L+16)/116) \quad (5.54)$$

Com:

$$\varepsilon = \begin{cases} 0.008856 & Actual\, CIE\, Standard \\ 216/24389 & Intent\, of\, the\, CIE\, Standard \end{cases} \quad (5.55)$$

$$\kappa = \begin{cases} 903.3 & Actual\, CIE\, Standard \\ 24389/27 & Intent\, of\, the\, CIE\, Standard \end{cases} \quad (5.56)$$

Conversão de XYZ para xyY.

$$x = \frac{X}{X+Y+Z} \quad (5.57)$$

$$y = \frac{Y}{X+Y+Z} \quad (5.58)$$

$$Y = Y \quad (5.59)$$

Através dos dados recolhidos com o telespectroradiometro podemos calcular e representar o gamut do monitor, que podemos ver na figura 5.1.

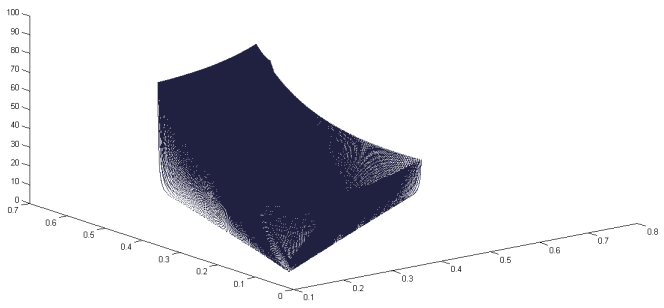
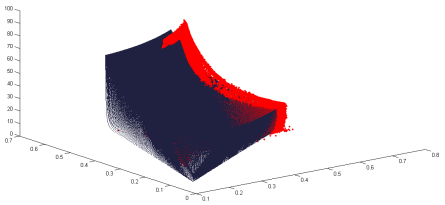


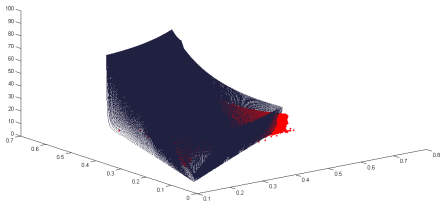
Figura 5.1: Gamut do monitor.

Para calcularmos a razão da escala a ser aplicada no Y que representa o L* (luminância), temos que encontrar a distância maior entre os valores da coordenada Y da imagem xyY e a coordenada Y do gamut do monitor. Esta distância tem que ser calculada verticalmente, ou seja, x e y não podem variar muito no cálculo desta distância, terá que haver algumas variações em x e y porque os valores x e y tanto da imagem como do gamut encontram-se valores decimais entre 0 e 1, por isso nunca iriamos encontrar x e y correspondestes para todos os pontos de comparação entre o gamut e a imagem, já o Y encontra-se entre 0 e 100. Depois de encontrarmos a distância máxima iremos aplicar o resultado obtido e multiplica-lo na coordenada Y de todos os pontos da imagem xyY, basicamente o que é feito é um resize do Y, podemos ver o seu resultado nas figuras 5.2a e 5.3b.

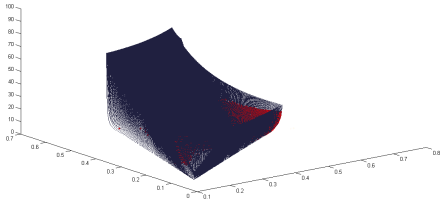
Depois de ser feita a escala do Y há que verificar se ainda existem pontos que se encontram fora da representação do monitor, após verificação é criada uma máscara lógica com os pontos a clipar. Nas figuras 5.3b, 5.3c e 5.3d podemos verificar as máscaras geradas ao longo de todo este processo de clipping. Após a localização desses pontos eles são clipados para o ponto mais próximo do gamut do monitor, ou seja, irão assumir esses valores. No fim os pixeis por clipar é inferior a 2%.



(a) Imagem xyY sem ajuste do Y e sem clipping.



(b) Imagem xyY com o ajuste do Y mas sem clipping.

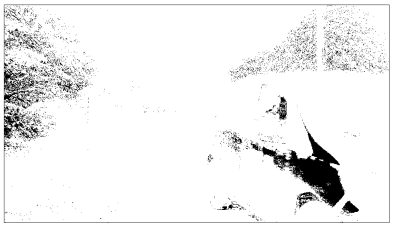


(c) Imagem xyY com o ajuste do Y e com clipping.

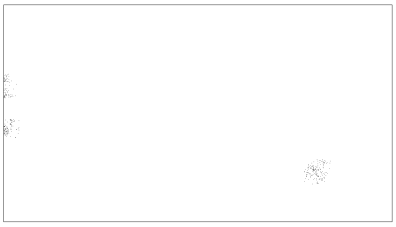
Figura 5.2: Evolução do processo de clipping.

(a) Imagem em RGB a ser tratada.

(b) Máscara da imagem xyY sem ajuste do Y e sem clipping.



(c) Máscara da imagem xyY com o ajuste do Y mas sem clipping.



(d) Máscara da imagem xyY com o ajuste do Y e com clipping.

Figura 5.3: Evolução do processo de clipping.

Depois de acabado o processo de clipping iremos converter as frames xyY de cada vídeo para XYZ:

$$X = \frac{xY}{y} \quad (5.60)$$

$$Y = Y \quad (5.61)$$

$$Z = \frac{(1-x-y)Y}{y} \quad (5.62)$$

Posteriormente iremos converter as frames XYZ de cada vídeo para sRGB, sendo [M] a matriz de transformação calculada mais acima:

$$\begin{bmatrix} R_{sRGB} \\ G_{sRGB} \\ B_{sRGB} \end{bmatrix} = \begin{bmatrix} M \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (5.63)$$

Se R_{sRGB}, G_{sRGB}, $B_{sRGB} \leq 0.0031308$

$$R'_{sRGB} = 12.92 * R_{sRGB} \quad (5.64)$$

$$G'_{sRGB} = 12.92 * G_{sRGB} \quad (5.65)$$

$$B'_{sRGB} = 12.92 * B_{sRGB} \quad (5.66)$$

se R_{sRGB}, G_{sRGB}, $B_{sRGB} > 0.0031308$

$$R'_{sRGB} = 1.055 * {R_{sRGB}}^{(1.0/2.4)} \quad (5.67)$$
$$G'_{sRGB} = 1.055 * {G_{sRGB}}^{(1.0/2.4)} \quad (5.68)$$
$$B'_{sRGB} = 1.055 * {B_{sRGB}}^{(1.0/2.4)} \quad (5.69)$$

Por fim:

$$R_{8bit} = 255.0 * R'_{sRGB} \quad (5.70)$$
$$G_{8bit} = 255.0 * G'_{sRGB} \quad (5.71)$$
$$B_{8bit} = 255.0 * B'_{sRGB} \quad (5.72)$$

5.7 Selecção dos vídeos

Após estes processos todos para a criação dos vídeos falta um passo muito importante na experiência, este passo é a selecção dos vídeos criados. Apesar de ser feito com sucesso a aplicação do erro cromático ΔE^*_{ab} nos vídeos e o seu respectivo clipping, muitos dos vídeos criados para esta experiência foram rejeitados. A razão é simples, como o processo do erro cromático ΔE^*_{ab} é aleatório não há qualquer controlo quanto ao aparecimento de artefactos espaciais que sejam criados durante este longo processo. A única forma de controlar isto foi a observação completa de cada vídeo criado e verificar se os artefactos espaciais criados eram mínimos ou imperceptíveis a olho nu. Foram criados em média 70 vídeos para cada sequência da base de dados. Apenas 3 vídeos foram escolhidos para cada sequência tendo em conta a relevância do deslocamento espacial do erro cromático ΔE^*_{ab} nas coordenadas $a*$ e $b*$ e a ausência de artefactos espacias, na figura 5.4 podemos visualizar as 3 sequências Basket com o erro cromático $\Delta E^*_{ab} = 18$ juntamente com a sequência original.

(a) Sequência Basket original.

(b) Deslocamento cromático 1.

(c) Deslocamento cromático 2.

(d) Deslocamento cromático 3.

Figura 5.4: Sequência Basket original e diferentes deslocamentos com erro cromático $\Delta E^*_{ab} = 18$.

Na figura 5.5, podemos analisar os diferentes deslocamentos em CIELAB obtidos para a sequência Basket. Como podemos ver, todos os deslocamentos são distintos, o que vemos na imagem é a representação dos pontos CIELAB agrupados por clusters, para tal, em cada gráfico são representados os 4 clusters tendo cada um a cor em RGB do seu respectivo centroid.

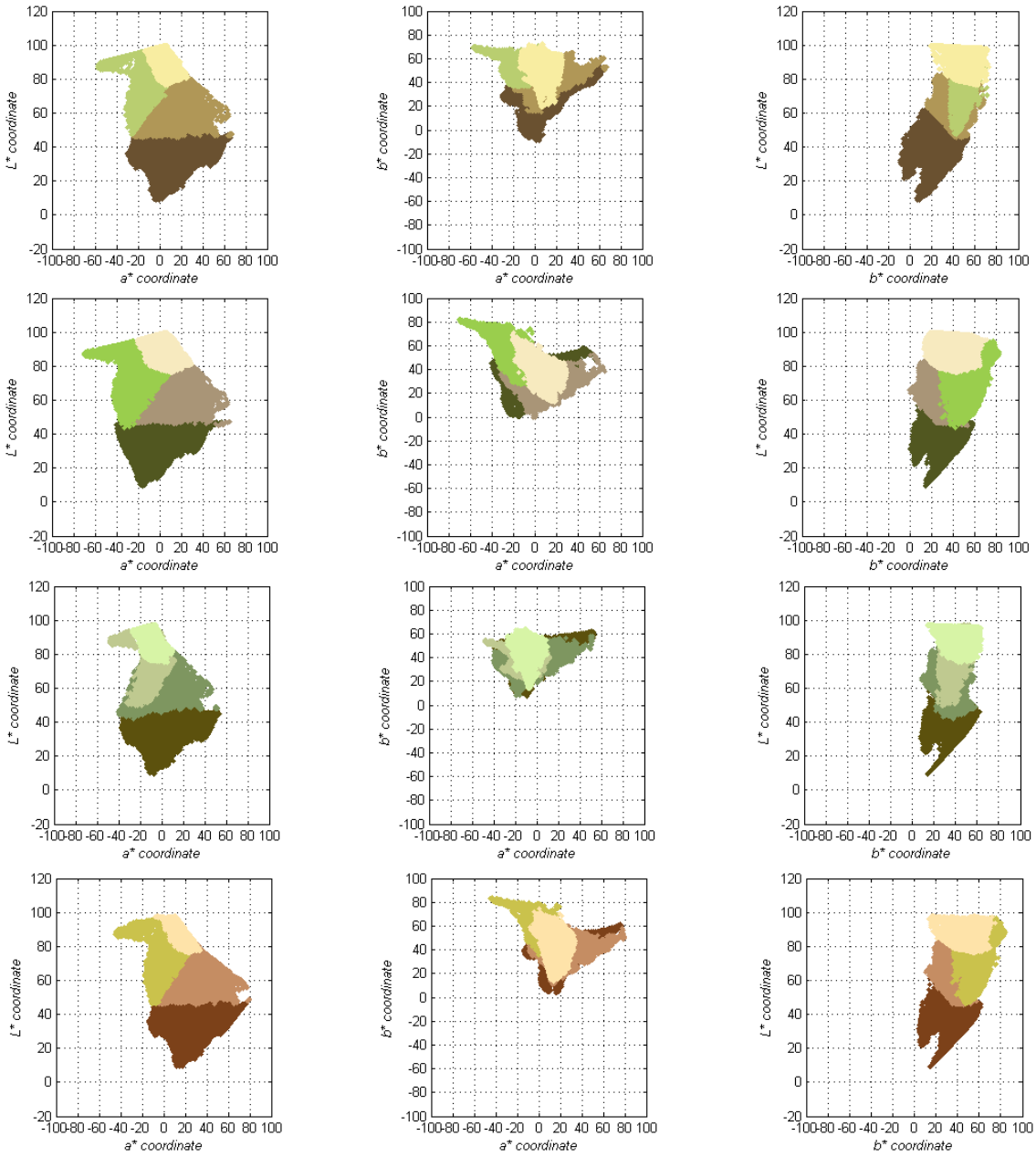


Figura 5.5: A primeira linha de gráficos é a representação em CIELAB da sequência Basket original, a segunda linha é a representação em CIELAB do deslocamento 1, a terceira linha é a representação em CIELAB do deslocamento 2 e finalmente a quarta linha é a representação em CIELAB do deslocamento 3.

Capítulo 6

Testes subjetivos de avaliação de qualidade

6.1 Laboratório

As experiências para avaliação de qualidade subjectiva foram realizadas no Centro de Óptica da Universidade da Beira Interior (UBI), que está em conformidade com as recomendações emitidas pelo ITU-R [36] para a avaliação subjectiva de dados visuais.

Os vídeos utilizados neste estudo foram exibidos num monitor CRT colorido de 21 polegadas (Sony, GDM-F520). Todos os vídeos de teste tinham uma resolução espacial de 1152 x 648 pixéis. Como já foi referido, foi acrescentada uma parte em cima e em baixo em todas as frames dos vídeos com a cor preta em RGB (0, 0, 0), no fim cada vídeo apresentava a resolução de 1152x864. Os vídeos encontravam-se à mesma resolução que o monitor para podermos relacionar os dados espaciais obtidos por tracking de forma directa. Estes vídeos foram observados pelos sujeitos sentados em linha com o centro do monitor a 1 metro de distância e com a cabeça centrada no equipamento do EyeTracker tanto verticalmente como horizontalmente com o centro do monitor, dentro de um quarto escuro apenas com a iluminação criada pelo monitor e dois projectores Kaiser RB1 usados na experiência necessários para ajudar na obtenção de dados com o EyeTracker, estes projectores usam uma lâmpada halógena com uma potência máxima de 650 watts cada.

6.2 Base de dados

6.2.1 Sequências NAMA3DS1-COSPAD1

As sequências originais da experiência foram criadas em conjunto pela Universidade Politécnica de Nantes e pela Universidade Politécnica de Madrid, pelo que o conjunto da base de dados se chama Nantes-Madrid-3D-Stereoscopic-V1, NAMA3DS1-COSPAD1 [1]. Na figura podemos ver uma frame de cada sequência.

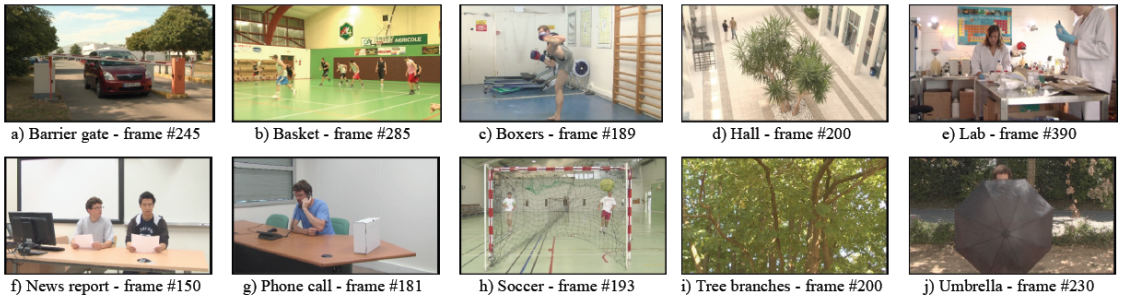


Figura 6.1: Visualização de todas as sequências da base de dados NAMA3DS1-COSPAD1.

6.2.1.1 Configuração da câmera e as condições de captura

As sequências foram capturados com uma câmera Panasonic AG- 3DA1E com lentes duplas [37], os eixos ópticos são separados por 60 mm, a uma distância perto da distância interpupilar, proporcionando um conteúdo de aparência natural. As lentes são equipadas com um zoom 5,6x motorizado, com uma distância focal que varia entre 4,2 mm e 23,5 mm, e uma a abertura de F1.8 a F2.4, respectivamente. As lentes duplas são ajustadas e sincronizadas, assim evitando desvios angulares e rotacionais verticais (< 1,2%), o brilho tem um desencontro entre a visão esquerda e a visão direita : não são necessários ajustes de som para a maioria das utilizações, ao contrário de [38]. Devido aos diferentes cenários e condições de captura, os parâmetros da câmera foram escolhidos individualmente para cada captura, incluindo o balanço do branco. A Tabela 1 na figura reporta nomeadamente o factor de zoom (Z) e a convergência (C) da câmara, para cada sequência. O Z varia entre 0 (angulo grande, 1x) e 99 (foco longo, 5,6). O C varia entre 0 (2,2 m) e 99, o que indica capturas em paralelo convergem no infinito. Embora estes parâmetros da câmara possam não ser suficientes para algoritmos de processamento de multi-view ou stereoprocessing de alta precisão, tais como a estimação da profundidade de benchmarking, eles podem ser úteis em estudos e contextos menos exigentes.

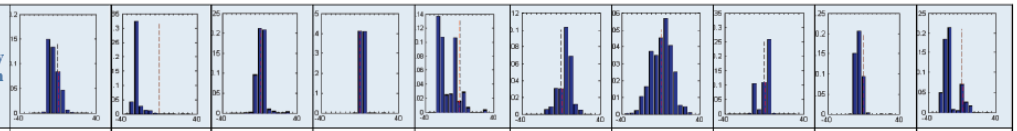
Sequence	Barrier gate	Basket	Boxers	Hall	Lab	News report	Phone call	Soccer	Tree branches	Umbrella
Duration	16s	16s	16s	16s	16s	16s	16s	16s	16s	13s
Scenes	1	1	2	1	1	1	1	2	1	1
Condition	Outdoor	Indoor	Indoor	Indoor	Indoor	Indoor	Indoor	Indoor	Outdoor	Outdoor
Description	Car and barrier gate	Basket ball training	1. Boxer warms up 2. Boxing training	Persons meeting in a hall.	Two lab assistants working	News report mimic	Phone call in an office	1. 2 players score 2. Goal keeping	Tree leaves and wind	Person playing with an umbrella
Shooting conditions	Medium	Long distance	Close distance	Long distance, high angle	Close distance	Close distance	Close distance	Long distance	Medium distance, low angle	Close distance
C	73	n. a.	n. a.	99	n. a.	50	36	n. a.	61	60
Z	0	n. a.	n. a.	0	n. a.	0	n. a.	n. a.	0	0
Source compression	No	Panasonic H.264 HP	Panasonic H.264 HP	No	Panasonic H.264 HP	No	No	No	No	No
SI	59	71	50 (50)	82	53	53	36	89 (89)	101	74
TI	21	41	56 (19)	5	12	4	13	38 (21)	14	19
DSI	20.42	11.72	24.40	17.02	17.77	21.58	21.56	24.70	23.02	17.02
DTI	15.43	9.69	18.03	6.97	10.04	8.71	11.78	18.08	13.63	15.24
Coding α	7.89	11.66	11.36	8.20	8.94	4.68	6.63	10.73	8.52	6.57
Coding β	9.88	2.97	3.01	7.94	9.34	25.99	19.98	3.76	3.09	9.38
Disparity histogram										
D+	6	-14	9	7	6	16	17	7	3	5
D-	9	26	3	-3	22	6	15	10	9	17

Figura 6.2: Lista das sequências fornecidas pela base de dados: descrição e descritores calculados.

6.2.1.2 Cadeia de gravação

As sequências capturadas possuem uma resolução de 1920x1080 em Full HD progressivo por cena e 25 frames por segundo. Quando possível, as sequências uncompressed foram transmitidas como dual SDI streams da câmera para um sistema de Clearview Extreme de Video Clarity. Esta cadeia tem sido utilizada para as sequências Barrier gate, Hall, News report, Phone call, Soccer, Tree branches, and Umbrella. No entanto, em alguns cenários não foi possível transmitir as sequências gravadas no Clearview por razões práticas. Especificamente, as sequências Basket, Boxer, e Lab foram registradas na própria câmera em 1080p25 por cena, usando dois cartões separados SD SanDisk Extreme classe 10. Para armazenar as sequências, a câmera AG-3DA1E Panasonic comprime-os utilizando um codificador H.264/AVC, de alto perfil, e uma taxa

máxima de 24 Mbps (média das sequências: 21Mbps).

6.2.1.3 Geração dos mapas de profundidade

Os mapas de profundidade foram gerados a partir das vistas da esquerda e da direita, utilizando um algoritmo de estimativa de disparidade com base em um algoritmo de optimização primal-dual convex de primeira ordem proposto por Chambolle et al. [39], o qual foi adaptado para a correspondência estéreo. Em particular, as áreas de oclusão foram tidas em conta no processo de optimização. Para cada frame, foram calculados dois mapas de profundidade: um utilizando a vista esquerda como referência da imagem, e outro usando a vista direita como referência da imagem. Os mapas de profundidade gerados foram então armazenados em ficheiros Y16 individualmente, onde as disparidades são introduzidas com valores de 16 bits (1 bit para o sinal, 9 bits para a parte inteira, 6 bits para a parte fraccionária).

6.2.2 Vídeos de teste

A base de dados original é composta por 10 sequências diferentes, mas apenas 6 foram usadas nesta experiência, 4 sequências usadas no período avaliação e 2 sequências no período de treino.

Das sequências originais usadas no período de avaliação foram escolhidas 2 sequências indoor, uma delas com vário movimento e com uma rotação da cena (Sequência Basket - Imagem 1), já a outra sequência indoor possui movimento mas mais suave em relação à primeira, com uma cena semelhante à de uma camara de vigilância (Sequência Hall - Imagem 3), foram também escolhidas para o período de avaliação 2 sequências outdoor, a primeira possui um carro em movimento em direcção a uma cancela (Sequência Car - Imagem 2), e a segunda sequência possui uma pessoa a abrir um chapéu de chuva numa paisagem mais rural (Sequência Car - Imagem 4).

Das sequências originais usadas no período de treino foram escolhidas 2 sequências, uma indoor e outra outdoor. A sequência outdoor é a captura da cena de uma árvore sendo esta uma paisagem rural (Sequência Tree - Imagem 5) e a sequência indoor é a captura de um treino de Kickboxe dentro de uma sala (Sequência Boxe - Imagem 6).

6.3 Metodologia da avaliação

Como referido no estudo [Philipe],entre os diferentes protocolos standard de avaliação de qualidade subjectiva, Absolute Category Rating com Hidden Reference (ACR -HR) [40] e Metodologia de Avaliação Subjetiva de Qualidade de vídeo (SAMVIQ) [40] têm sido amplamente utilizados para avaliar o conteúdo 2D e também para avaliar conteúdos relacionados com aplicações de vídeo 3D [41, 42] . Tais metodologias são geralmente escolhidas pela sua fiabilidade conhecidas no âmbito da avaliação de média 2D. Com efeito, Brotherton et al. [43] investigaram a adequação das metodologias de ACR e SAMVIQ para avaliar 2D . O estudo mostrou que ACR permite mais sequências de ensaio (pelo menos o dobro) a serem apresentadas para a avaliação em comparação com a metodologia SAMVIQ. ACR também provou ser de confiança nas condições de teste. Rouse et al. [44] estudaram a troca destas duas metodologias no contexto

de imagens fixas e sequências de vídeo de alta definição. Eles concluíram que a adequabilidade dos dois métodos poderia depender de aplicações específicas.

Huynh -Thu et al . [45] realizaram um estudo para comparar diferentes metodologias de acordo com suas diferentes escalas de avaliação (5 pontos discretos, 9 pontos discretos, uma escala de 5 pontos contínuos, e escalas contínua de 11 pontos). Os testes foram realizados no contexto de vídeo de alta definição. Os resultados mostraram que a metodologia do ACR produziu resultados subjectivos fiáveis , mesmo em diferentes escalas. Além disso, esta metodologia de base é conhecida pela sua facilidade de execução. Com base nestes resultados anteriores, seleccionamos a metodologia ACR-HR com uma escala de qualidade contínua de 11 pontos.

A metodologia ACR-HR consiste na apresentação de objectos de teste (ou seja, imagens ou sequências de vídeo) para os observadores, um de cada vez. Os objectos são classificados de forma independente com uma escala de classificação. A referência de cada objecto tem de ser incluída no procedimento da experiência e classificado como qualquer outro estímulo. Isso explica o termo usado de "hidden reference". A partir dos resultados obtidos, pode ser calculado o diferential mean opinion scores (DMOS) que calculado entre os mean opinion scores (MOS) de cada objecto de teste e sua referência escondida associada. ACR-HR requer muitos observadores para minimizar os efeitos contextuais (estímulos apresentados anteriormente influenciam a opinião do observador, ou seja, a ordem de apresentação influencia avaliações da opinião), para tal cada observador irá visualizar uma sequência aleatória, sendo diferente entre cada observador minimizando assim também os efeitos contextuais. A precisão aumenta também com o número de participantes.

6.4 Observadores

As avaliações subjectivas foram realizadas em um ambiente de teste conforme a ITU. Os estímulos foram apresentados em um monitor Sony FD250 (1152x864p), e de acordo com a ITU [36].

Participaram 25 voluntários no teste de avaliação subjectiva de qualidade em uma sessão 30 minutos. Todos dentro da faixa etária de 16-33 anos, com uma idade média de 23 anos e um desvio padrão de 3,26 anos. Neste conjunto de voluntários 64% eram do sexo masculino (16 voluntários) e 36% eram do sexo feminino (9 voluntários). Todos os voluntários eram naive quanto ao objectivo da experiência. Todos os observadores foram submetidos a uma triagem para avaliar a sua visão a nível da cor, da estereopsia e da dominância ocular. Para tal cada observador fez o teste Ishiara para testar a sua visão a nível da cor, o teste Randot Stereotests para testar a sua visão estéreo e o teste Miles para determinar a sua dominância ocular.

6.5 Procedimento

Antes do início de cada teste, foi mostrado aos observadores um protocolo experimental. O ensaio foi composto por duas partes: uma primeira, designada por Período de Treino (usando

um conjunto de vídeos diferentes), com fim de o observador ficar familiarizado com o procedimento, seguido da sessão de teste. Somente as respostas dadas pelo observador durante a sessão de teste foram consideradas para os resultados deste estudo.

Cada observador visualizava uma sequência de vídeos aleatória única composta por 44 vídeos, 4 deles destinavam-se para o Período de Treino e apenas estes 4 vídeos eram apresentados de maneira igual para cada observador (Boxe com erro 18 - Tree com erro 12 - Boxe com erro 6 - Tree com erro 0). O teste resultante tinha em média uma duração de 30 minutos. Sessões de teste mais longas não são aconselháveis por causa dos efeitos da fadiga que poderiam influenciar os resultados finais tanto na avaliação como na recolha de dados para os pontos de atenção.

(a) Visão lateral de um teste. (b) Visão frontal de um teste.

Figura 6.3: Ambiente do teste subjectivo.

Durante os testes, o observador visualiza um vídeo durante um período correspondente a esse vídeo, seguida pela avaliação do vídeo apresentado. Para a avaliação do vídeo, os observadores dispunham de um teclado wireless situado entre o monitor e o mesmo. Ao usar esse teclado os observadores poderiam seleccionar uma nota entre 0 (baixa qualidade / artificial) e 10 (alta qualidade / natural) em uma escala contínua simulada (100 níveis) . Ver figura 6.4 com a janela de selecção.

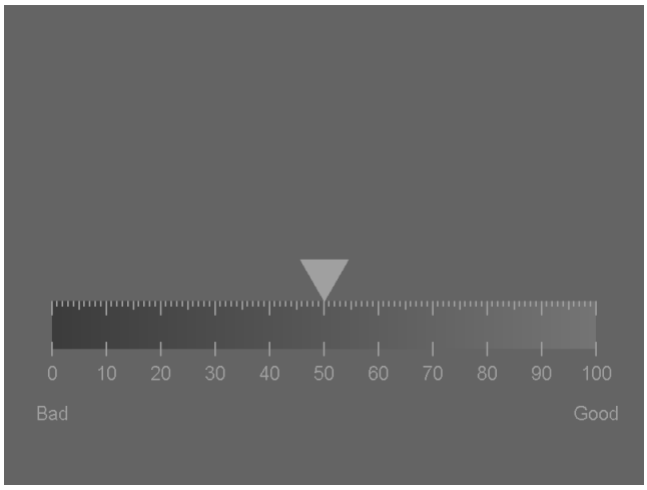


Figura 6.4: Escala de avaliação usada neste trabalho.

6.6 Testes de triagem

Como foi anteriormente referido todos os voluntários da experiência estiveram sujeitos a uma triagem para testar a sua visão a nível da cor, da estereopsia e da dominância ocular. O sucesso e a execução destes testes preliminares são essenciais para seleccionarmos os voluntários de modo a que a experiência seja o mais precisa possível. Esta experiência tem como objectivo uma avaliação da qualidade da cor em vídeos 3D, logo os voluntários que falhassem o teste Ishiara ou o teste Randot Stereotests não estariam aptos a participar na experiência.

Apenas um voluntário foi rejeitado no processo de triagem da experiência por ter falhado no teste Ishiara, verificou-se que tinha deficiência no vermelho-verde.

Quanto ao teste Randot Stereotests [46] os voluntários que tiveram uma classificação positiva no teste Ishiara tiveram uma média de resposta de 50 segundos de arco, em que se verificava que estariam aptos para participar na experiência.

Quanto ao teste de dominância ocular foi realizado a fim de posicionar o Eyetracker no olho dominante de cada pessoa para obtermos resultados mais precisos a nível de captação de informação da posição espacial para qual o observador estava a olhar, em 25 casos apenas dois se verificam ter o olho esquerdo como olho director.

Nas secções mais abaixo podemos encontrar uma informação mais detalhada de cada teste feito aos voluntários desta experiência.

6.6.1 Teste Ishiara

O teste de cores de Ishihara é um exemplo de um teste de percepção de cor para as deficiências das cores verde e vermelho. Foi concebido pelo Dr. Shinobu Ishihara, um professor da Universidade de Tóquio, que primeiro publicou seus testes em 1917.

O teste consiste na exibição de uma série de placas coloridas, chamadas placas Ishihara, cada placa contém um círculo de pontos com tamanhos e cores ligeiramente diferentes, em relação às cores situadas nas proximidades. Dentro do padrão estão pontos que formam um número ou uma forma claramente visível para observadores com uma visão normal das cores, e as invisíveis, ou difíceis de ver, para aqueles com um defeito de visão das cores verde e vermelho, ou ao contrário. O teste completo consiste em 38 placas, mas a existência de uma deficiência geralmente é clara após algumas placas. Há também o teste menor que consiste apenas 24 placas, na figura 6.5 podemos observar uma placa do teste como exemplo.

As placas formam vários modelos de teste diferentes:

- Placas de transformação: os indivíduos com deficiência na cor devem ver um número diferente de indivíduos com visão normal das cores.
- Placas de desaparecimento: apenas indivíduos com visão normal das cores poderá reconhecer a figura.

- Placas com dígitos ocultos: apenas indivíduos com defeito deficiência na cor podem reconhecer a figura.

- Placas de diagnóstico: destina-se a determinar o tipo de defeito de visão das cores (ou protanopia deuteranopia) e da gravidade da mesma.

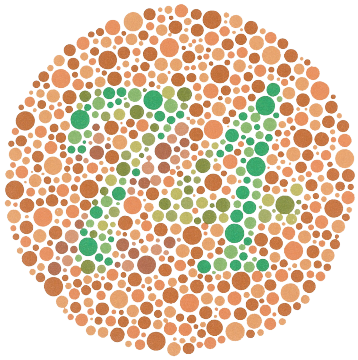


Figura 6.5: Exemplo de uma placa de cor do teste Ishihara. O número "74"deve ser claramente visível para os indivíduos com visão normal de cor. Indivíduos dicromatas ou tricomatas anómalos poderam lê-lo como "21".

Na tabela 6.1 podemos observar as diferentes respostas dos indivíduos que se sujeitam a este teste consoante a sua percepção nas cores. A marca x mostra que a placa não poderá ser lida. Espaços em branco significam que o tempo de leitura poderá ser indefinido. Os números entre parêntesis mostra que podem ser lidos mas são relativamente pouco claros.

Tabela 6.1: Tabela de respostas do teste Ishiara.

Número da placa	Pessoa normal	Pessoa com deficiência no vermelho-verde				Daltonismo total e fraqueza
1	12	12				12
2	8	3				x
3	6	5				x
4	29	70				x
5	57	35				x
6	5	2				x
7	3	5				x
8	15	17				x
9	74	21				x
10	2	x				x
11	6	x				x
12	97	x				x
13	45	x				x
14	5	x				x
15	7	x				x
16	16	x				x
17	73	x				x
18	x	5				x
19	x	2				x
20	x	45				x
21	x	73				x
		Protan		Deutan		
		Forte	Fraco	Forte	Fraco	
22	26	6	(2)6	2	2(6)	
23	42	2	(4)2	4	4(2)	
24	35	5	(3)5	3	3(5)	
25	96	6	(9)6	9	9(6)	

6.6.2 Randot Stereotests

Este teste é produzido pela empresa Stereo Optical, Randot é uma marca registrada da empresa Stereo Optical.

O Randot Stereotests na figura 6.6, é destinado para o teste de estereopsia adulta, mas também inclui uma porção de animais para os testes a nível pediátrico. Os observadores são convidados a identificar seis formas geométricas durante o teste. Este teste ajuda a testar a percepção de profundidade do observador juntamente com a visão estéreo normal. Os testes adultos encontram-se entre os 400 e 20 segundos de arco, e o teste pediátrico encontra-se entre os 400-100 segundos de arco. Este teste ajuda na detecção de ambliopia, estrabismo e supressão.

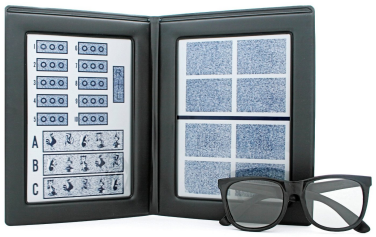


Figura 6.6: Exemplo do Teste Randot Stereo usado na triagem para cada observador.

6.6.3 Dominância ocular

A dominância ocular [47], às vezes chamado de olho director, é a tendência a preferir o input visual de um olho em relação ao outro. É algo análogo à lateralidade da destreza manual, direita ou esquerda. Os dois hemisférios controlam ambos os olhos, mas cada um se encarrega de uma metade diferente do campo de visão e, portanto, uma metade diferente de ambas as retinas. Assim, não há analogia directa entre a "imparcialidade "e "eyedness "como fenómenos laterais.

Aproximadamente dois terços da população são dominantes no olho direito e um terço no olho esquerdo, no entanto, uma pequena parte da população não possui um olho director. A dominância parece mudar dependendo da direcção do olhar porque o tamanho da imagem muda nas retinas.

Na visão binocular normal há um efeito de paralaxe, e, portanto, o olho dominante é o que está principalmente posicionado para informação precisa. Isso pode ser especialmente importante em deportos que exigem destreza ocular, como tiro com arco, dardos e tiro desportivo.

6.6.3.1 Determinação da dominância ocular

- O teste de Miles: O observador estende ambos os braços, junta as duas mãos e posiciona-as de forma a criar uma pequena abertura (exemplificado na figura 6.7), de seguida, com os dois olhos abertos observa um objecto distante através da abertura. O observador fecha os olhos alternadamente, se o objecto permanece no ponto de vista o olho aberto é o olho director, caso contrário será o olho fechado.

Figura 6.7: Teste Miles de dominância ocular.

Capítulo 7

Análise dos resultados subjectivos

7.1 Análise de distribuição

Nesta secção, são apresentados os resultados dos testes e a sua análise. Com uma probabilidade de 95%, o valor absoluto da diferença entre a pontuação média experimental e a pontuação média experimental "verdadeira"(para um número muito elevado de observadores) é menor do que o intervalo de confiança de 95%, com a condição de que a distribuição das pontuações individuais atende a determinados requisitos.

A relação entre a pontuação média experimental com base em uma amostra da população (ou seja, os indivíduos que participaram na experiência) e a pontuação média experimental "verdadeira"de toda a população é dada pelo intervalo de confiança da média estimada. Os intervalos de confiança de 95% foram calculados.

Uma vez que pode ser difícil de interpretar em detalhe os valores da DMOS mais à frente na secção 7.3 , sugerimos contar com ferramentas estatísticas para uma melhor compreensão. A partir destas avaliações, a análise estatística foi realizada utilizando a análise de variância (ANOVA [48]). Com efeito, a análise ANOVA permite saber se as diferenças entre as médias dos grupos são consideradas significativas (isto é, se devido à influência do método de síntese de vista, as características de conteúdo, ou a métodos de indução do erro na cor) ou simplesmente devido ao acaso.

Tabela 7.1: Tabela dos resultados ANOVA para a sequência Basket.

ANOVA sequência Basket					
'Source'	'SS'	'df'	'MS'	'F'	'Prob>F'
'Groups'	76.77333	3	25.5911	5.7043	8.6326e-4
'Error'	1103.62	246	4.4863		
'Total'	1180.40	249			

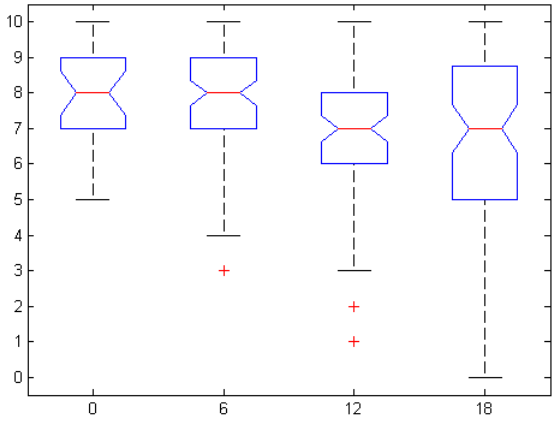


Figura 7.1: Gráfico Basket ANOVA de cada erro e respectiva referência.

Tabela 7.2: Tabela dos resultados ANOVA para a sequência Car.

ANOVA sequência Car					
'Source'	'SS'	'df'	'MS'	'F'	'Prob>F'
'Groups'	129.3973	3	43.1324	11.286	5.8031e-07
'Error'	940.1067	246	3.8216		
'Total'	1069.5	249			

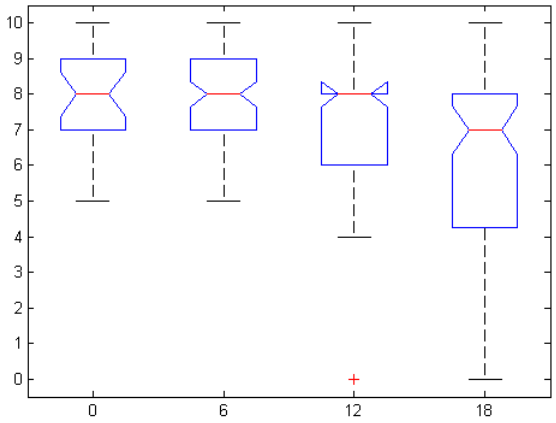


Figura 7.2: Gráfico Car ANOVA de cada erro e respectiva referência.

Tabela 7.3: Tabela dos resultados ANOVA para a sequência Hall.

ANOVA sequência Hall					
'Source'	'SS'	'df'	'MS'	'F'	'Prob>F'
'Groups'	154.6440	3	51.5480	7.4632	8.4011e-05
'Error'	1699.1	246	6.9070		
'Total'	1853.8	249			

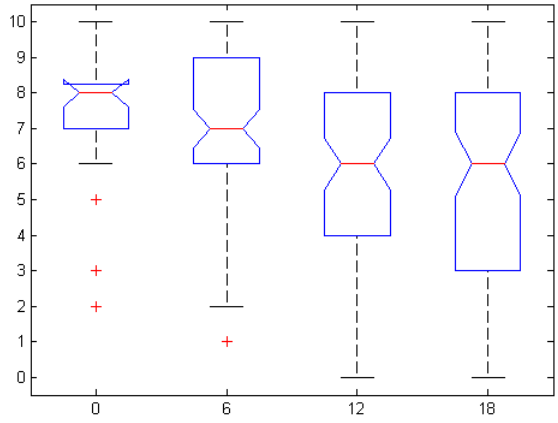


Figura 7.3: Gráfico Hall ANOVA de cada erro e respectiva referência.

Tabela 7.4: Tabela dos resultados ANOVA para a sequência Umbrella.

ANOVA sequência Umbrella					
'Source'	'SS'	'df'	'MS'	'F'	'Prob>F'
'Groups'	193.0507	3	64.3502	12.4461	1.3219e-07
'Error'	1271.9	246	5.1703		
'Total'	1464.9	249			

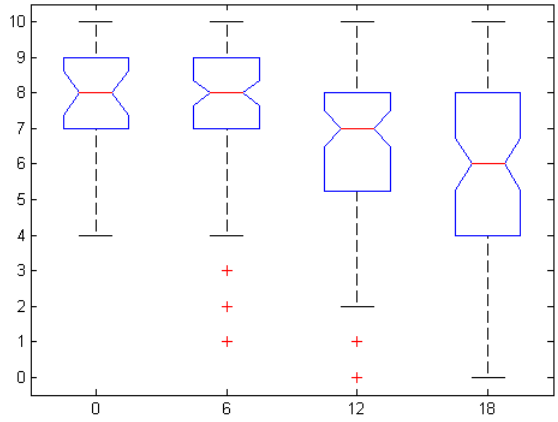


Figura 7.4: Gráfico Umbrella ANOVA de cada erro e respectiva referência.

Analisando os resultados da ANOVA podemos concluir que os resultados obtidos são estatisticamente significativos, uma vez que o valor de P é sempre inferior a 0.05 e o valor de F é sempre superior a 3.354.

Vários autores discutem a influência da remoção dos outliers detectados nos resultados. Então, é necessário detectar eventuais outliers entre observadores. Foram aplicadas ao nosso conjunto de dados os critérios de rejeição do padrão ITU [36]. Segundo este padrão nenhum observador foi detectado e rejeitado como outlier, por isso todos os dados dos observadores da experiência foram considerados.

7.2 Mean opinion scores

Foram calculadas medidas estatísticas para descrever a distribuição da pontuação através dos observadores.

O MOS é calculado para cada condição de teste com a seguinte fórmula:

$$MOS_j = \frac{\sum_{i=1}^{N} s_{ij}}{N} \quad (7.1)$$

onde N é o número de indivíduos e sij é a pontuação por objecto i para a condição do teste. Os resultados são apresentados nas figuras 7.5, 7.6 e 7.7.

Pela análise do gráfico 7.5, podemos observar que o MOS de vídeos com o erro $\Delta E^*_{ab} = 6$ são relativamente altos, em alguns casos até supera o MOS do seu vídeo de referência. No geral, vídeos com o erro $\Delta E^*_{ab} > 6$ foi notada a degradação da qualidade da cor, pelo que os seus valores de MOS tendem a descer quando o erro é aumentado.

Apenas em alguns casos únicos se pode observar que vídeos com erro ΔE^*_{ab} possuem um MOS mais elevado que o MOS do seu vídeo de referência, isto poderá ser explicado pelo efeito do erro aplicado, os vídeos com uma certa magnitude a nível visual seriam mais atractivos em relação ao seu vídeo de referência. Por exemplo, na sequência de vídeo Car, alguns erros aplicados a nível perceptual poderiam parecer mais naturais porque causava um efeito de mais saturação na cena.

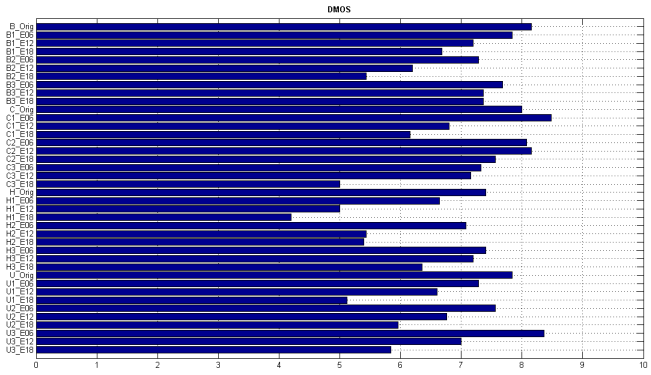


Figura 7.5: MOS de cada vídeo.

Pela análise do gráfico seguinte 7.6, como era de esperar, há uma redução do valor do MOS à medida que a magnitude do erro ΔE^*_{ab} aumenta, o que é verificado em todos os vídeos. Os valores de MOS para o erro $\Delta E^*_{ab} = 6$ como já foi referido, apresentam valores altos. É de notar que o valor do MOS com erro $\Delta E^*_{ab} = 6$ para cenas mais rurais é muito próximo do MOS do vídeo de referência, pelo que se pode concluir que os observadores são menos sensíveis em vídeos com erro $\Delta E^*_{ab} = 6$ em cenas mais rurais do que em cenas urbanas, Marco et. al [35] conclui em estudos feitos anteriormente.

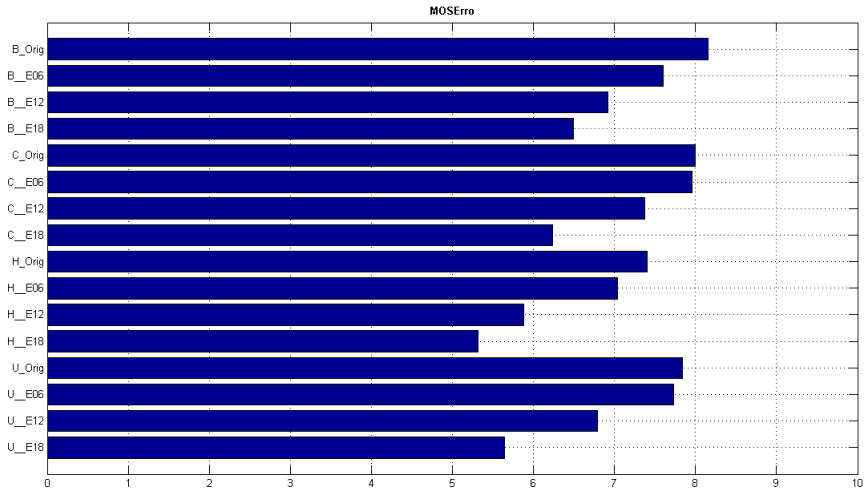


Figura 7.6: MOS para cada erro.

Pela análise do gráfico do MOS por erro com os respectivos intervalos de confiança 7.7, pode-se observar que os valores intervalos de confiança são pequenos, o que releva a similaridade de resultados entre os observadores, pelo que os valores são mais ou menos constantes para os diferentes erros.

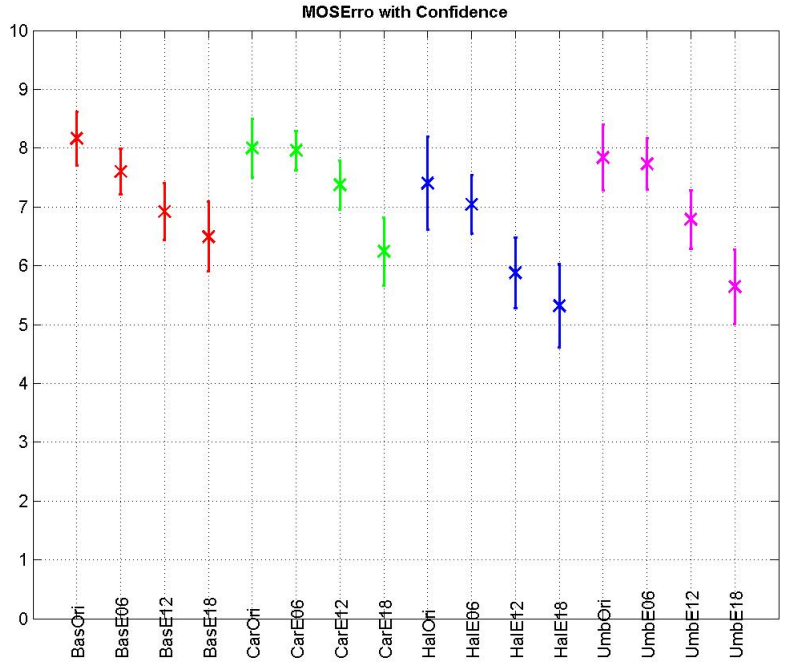


Figura 7.7: MOS de cada erro com o respectivo intervalo de confiança.

7.3 Differential mean opinion scores

O MOS e o DMOS foram calculados entre cada estímulo e sua referência correspondente (oculta). Conforme recomendado no plano de teste de multimédia VQEG [49], o DMOS foi

calculado por observador e por estímulo. O DMOS é calculado subtraindo a classificação do vídeo de referência a cada estímulo correspondente, podemos observar pela seguinte fórmula 7.2, onde i é representa um vídeo com erro ΔE^*_{ab} e ref o vídeo da referência correspondente.

$$DMOS_i = MOS_i - MOS_{ref} + 10 \quad \textbf{(7.2)}$$

Pelo seguinte gráfico 7.2, podemos mais facilmente, fazer uma análise de quais os vídeos que obtiveram melhores resultados a nível do MOS em relação à sua referência. Os vídeos que neste gráfico 7.2 possuem um DMOS superior a 10 correspondem aos vídeos com melhor MOS em relação à sua referência. É de notar também em geral os vídeos $\Delta E^*_{ab} = 18$ tiveram uma classificação muito inferior à da sua referência, pelo que podemos concluir que erros ΔE^*_{ab} muito elevados são perceptíveis.

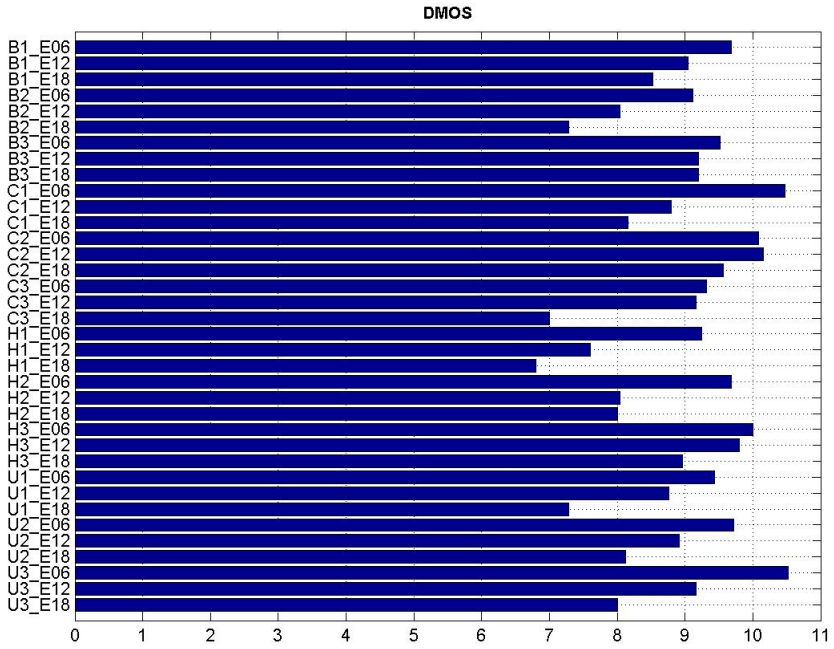


Figura 7.8: DMOS de cada vídeo.

Foi criado também o seguinte gráfico 7.9 para melhor análise do DMOS num nível geral. Neste gráfico podemos observar a variação do MOS em relação a sua referência. Como foi antes referido, o MOS das cenas mais rurais é ficaram muito próximas do MOS do vídeo de referência, como é o caso da sequência Car e da sequência Umbrella. Neste gráfico também é visivel claramente a variação do MOS em relação à magnitude do erro ΔE^*_{ab}.

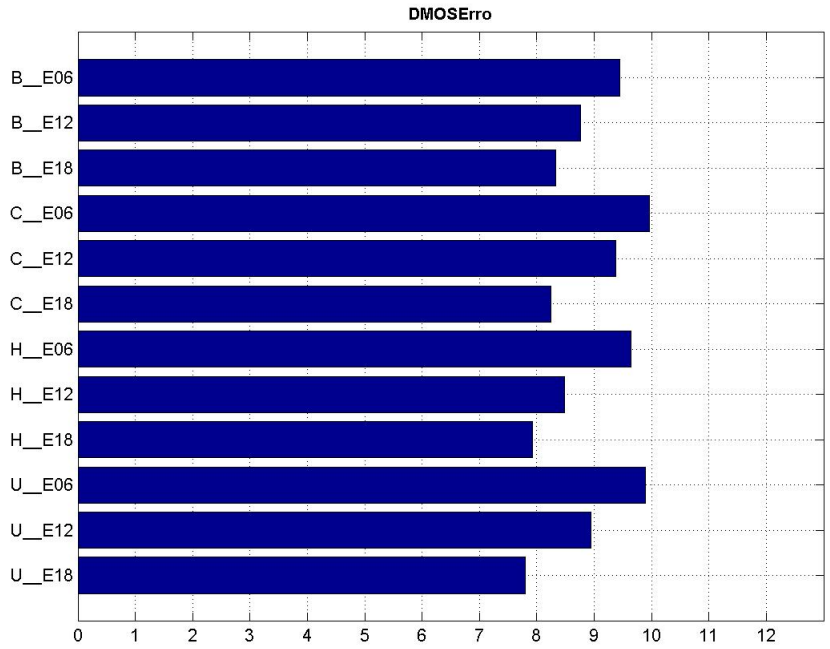


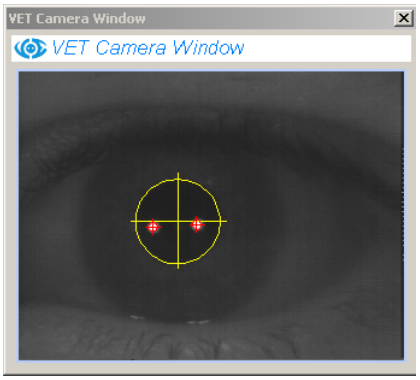
Figura 7.9: DMOS para cada erro.

Capítulo 8

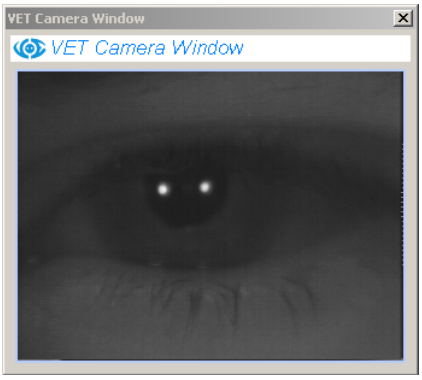
Mapas de atenção

8.1 Sistema 'EyeTracker'

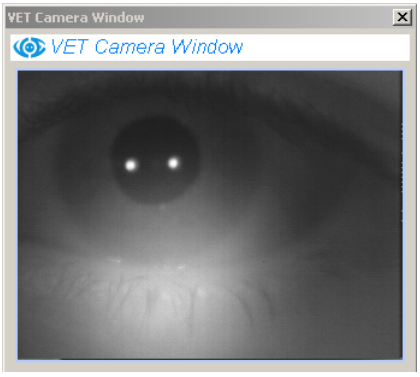
Neste capítulo é feita uma pequena análise dos mapas de atenção recolhidos na experiência. Note-se que devido à limitação de hardware a recolha de dados não foi a mais completa possível. Devido à utilização dos óculos 3D Vision para visualizar conteúdo 3D, o "EyeTracking"tinha dificuldade em detectar o olho do observador. Apesar da iluminação já existente, era aconselhável não haver um aumento de iluminação na sala de testes para não causar estímulos prejuciais à experiência. Outro problema encontrado no reconhecimento pupilar é a fisionomia do olho. Por exemplo, se a pálpebra estiver sobre alguma parte da pupila o sistema não consegue reconhecer a pupila do observador. Por vezes ocorriam reflexos ocasionais nos óculos o que dificultava o reconhecimento da pupila. Também se verificou que observadores com olhos mais claros ao usar os óculos, o sistema não conseguia detectar o limiar da pupila, resultando só na experiência de avaliação de qualidade subjectiva. Infelizmente não foi possível obter um sistema de "eyetracking"mais sofisticado devido aos preços elevados destes sistemas.



(a) Pupila detectada.



(b) Fisionomia do olho a dificultar a detecção.



(c) Reflexo dos óculos a dificultar a detecção.

Figura 8.1: Várias situações de reconhecimento pupilar.

8.2 Análise dos Mapas de atenção

Vários mapas de atenção foram recolhidos durante a experiência. Foi feita uma selecção dos mapas a serem usados para análise. Apenas os mapas dos observadores com uma percentagem de tracking superior a 50% eram considerados. Apenas se o tracking verifica-se a condição anterior em todas os vídeos de uma sequência, ou seja, se um observador tivesse um tracking superior a 50% em todos os vídeos visualizados do Car, então os mapas de atenção eram usados como objecto de estudo.

Nas figuras podemos observar os mapas de atenção de todos os observadores seleccionados para estudo da sequência Car. É de notar que cada cor utilizador corresponde à marcação com uma cor diferente em todos os vídeos. Os mapas de atenção estão organizados da seguinte forma.

	Frame original	
Deslocamento 1 Erro 6	Deslocamento 1 Erro 12	Deslocamento 1 Erro 18
Deslocamento 2 Erro 6	Deslocamento 2 Erro 12	Deslocamento 2 Erro 18
Deslocamento 3 Erro 6	Deslocamento 3 Erro 12	Deslocamento 3 Erro 18

Foram criadas 3 imagens de todos os erros incluindo a referência, cada uma correspondente a um terço da sequência de vídeo.

Feita uma pequena análise às figuras A.1, A.2 e A.3 alusivas à sequência Car, a um nível geral podemos dizer que com erros $\Delta E^*_{ab} = 18$ as pessoas tendem a focar mais a parte central do vídeo, ou seja, irão focar mais o carro em movimento juntamente com a cancela a abrir e não focam tanto o ambiente em redor. As nuvens de pontos de atenção tendem a dispersar menos na cena em geral e concentram-se mais no centro da cena.

Analisando as figuras A.4, A.5 e A.6 alusivas à sequência Hall notamos que há uma grande uniformidade dos mapas de atenção mesmo quando os erros ΔE^*_{ab} são elevados. As pessoas tendem a observar o movimento e as acções das pessoas da cena, pelo que reduz a atenção dos outros aspectos da cena.

Feita a análise das figuras A.7, A.8 e A.9 alusivas à sequência Umbrella, reparámos que a maioria da atenção deste vídeo é na acção da pessoa a abrir o chapéu de chuva. Porém com o aumento dos erros ΔE^*_{ab} é de notar que as zonas de atenção ficam reduzidas a duas praticamente, uma sendo a pessoa da cena e outra a folhagem da àrvore no lado direito juntamente com as flores de um arbusto logo abaixo. Aqui pudemos concluir que o aumento do erro ΔE^*_{ab} influenciou a atenção da sequência Umbrella.

Feita a análise da sequência Basket nas figuras A.10, A.11 e A.12, reparámos que toda a atenção deste vídeo se foca nos jogadores e na bola em movimento, mesmo com o aumento do erro ΔE^*_{ab} é de verificar que o movimento neste vídeo é o factor principal dos resultados obtidos nesta experiência.

Capítulo 9

Conclusões e Trabalho Futuro

9.1 Conclusões Principais

Foi realizado um estudo sobre a percepção de variações de cor em vídeos 3D. Um estudo passado revelou que a visão humana é perceptível a mudanças de cor para $\Delta E^*_{ab} \geq$ 2,2 unidades [50]. No entanto, podemos verificar que a percepção dos observadores nas alterações de cor ΔE^*_{ab} = 6 unidades é bem tolerada, independentemente do conteúdo. Nesse caso verificou-se que certas cenas rurais com ΔE^*_{ab} = 6 eram atribuídas melhores classificações em relação às suas referências. Isto está de acordo com o estudo do Marco, el al. (QoMex) onde se refere que cores naturais levam a uma menor sensibilidade, equanto cores criadas que existem em ambientes não naturais, levam a menos sensibilidade ao erro. Concluímos, então, que os observadores revelam uma baixa sensibilidade na visualização de vídeos 3D com um erro cromático ΔE^*_{ab}. Quanto aos mapas de atenção pudemos verificar que em vídeos $\Delta E^*_{ab} > 12$ estavam sujeitos a alterações nas zonas de atenção, porém foi também verificado que vídeos com vário movimento e com características que sejam apelativas à curiosidade das pessoas, como por exemplo pessoas em cena, irão influenciar os mapas de atenção em relação ao erro cromático ΔE^*_{ab} introduzido no vídeo.

9.2 Trabalho Futuro

Como trabalho futuro, seria apropriado efectuar mais testes de qualidade subjectiva com diferentes estímulos. Por exemplo, usando os erros ΔE^*_{ab} desta experiência em conteúdo 2D e comparar com os resultados obtidos em conteúdo 3D. Será também interessante estudar a influência da aplicação do erro ΔE^*_{ab} em um só cluster em vez de todos os clusters tanto em conteúdo 2D como em contúdo 3D e no fim fazer uma comparação de resultados e observar se há relações. Estes estudos embora interessantes não foi possível a sua realização devido à grande dificuldade em fazer testes subjectivos, considerando que todos os estudos feitos até agora foram realizados com a ajuda de voluntários.

Bibliografia

[1] M. Urvoy et al, NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences.

[2] REICHL, P., FABINI, J., HAPPENHOFER, M., EGGER, C. From QoS to QoX: A charging perspective. In Proceedings of the 18th ITC Specialist Seminar on Quality of Experience. Karlskrona (Sweden): Blekinge Insitute of Technology, 2008, p. 35 - 44.

[3] Satu Jumisko-Pyykko(1), Dominik Strohmeier(2),Timo Utriainen(1),Kristina Kunze(2), "Descriptive Quality of Experience for Mobile 3D Video", Proceedings: NordiCHI 2010, Outubro 16â20, 2010.

[4] QUALINET QoE definition, Dagstuhl seminar, 2012, QUALINET whitepaper.

[5] Recommendation ITU P.911, "Subjective audiovisual quality assessment methods for multimedia application", Tech. Rep., ITU Telecommunication Standardization Sector, Dezembro 1998.

[6] ITU-R, "Methodology for the subjective assessment of the quality of television pictures", Tech. Rep. BT.500- 11, ITU-R (2002).

[7] ITU-R, "Subjective assessment methods for image quality in high-definition television", Tech. Rep. BT.710-4, ITU-R (1998).

[8] ITU-R, "Subjective assessment of stereoscopic television pictures", Tech. Rep. BT.1438 (2000).

[9] W. IJsselsteijn, H. de Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman, "Perceived depth and the feeling of presence in 3DTV", Displays, vol. 18, Maio 1998, pp. 207-214.

[10] Lambooij M.T.M.; IJsselsteijn W.A.; Heynderickx I.; "Visual Discomfort in stereoscopic Displays: A Review", SPIE-IST, Volume 6490, Janeiro 2007.

[11] Meesters, L.M.J.; IJsselsteijn, W.A.; Seuntiens, P.J.H., "A survey of perceptual evaluations and requirements of three-dimensional TV", IEEE Trans. on Circuits and Sys. For Video Techn., Vol. 14, No.3, MarÃ§o 2004 Page(s): 381- 391.

[12] Fehn C.;"3D TV Broadcasting", chapter from 3D Video communication, pages: 23-38, Jan. 2006

[13] G. Leon, H. Kalva, and B. Furht, "3D Video Quality Evaluation with Depth Quality Variations", 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008, pp. 301-304.

[14] Benoit, A., Le Callet, P., Campisi, P., Cousseau, R. Using disparity for quality assessment of stereoscopic images Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, p. 389-392., Outubro, 2008.

[15] C. Hewage, S. Worrall, S. Dogan, S. Villette, and A. Kondoz, "Quality Evaluation of Color Plus Depth Map-Based Stereoscopic Video", IEEE Journal of Selected Topics in Signal Processing, vol. 3, 2009, pp. 304-318.

[16] D.V.S.X. De Silva, W.A.C. Fernando, G. Nur, E.Ekmekcioglu and S.T. Worrall , "3D VIDEO ASSESSMENT WITH JUST NOTICEABLE DIFFERENCE IN DEPTH EVALUATION", Proceedings of 2010 IEEE 17th International Conference on Image Processing, Setembro 26-29, 2010, Hong Kong.

[17] Goldmann, L., De Simone, F., Ebrahimi, T., "A Comprehensive Database and Subjective Evaluation Methodology for Quality of Experience in Stereoscopic Video", Proceedings of SPIE, San Jose, 7526, 2010.

[18] Milos Klima, Karel Fliegel, Petr Pata, Stanislav Vitek, Martin Blazek, Petr Dostal, Lukas Krasula, Tomas Kratochvil, Vaclav Ricny, Martin Slanina, Ladislav Polak, Ondrej Kaller, Libor Bolecek, "DEIMOS An Open Source Image Database", RADIOENGINEERING, vol. 20, num. 4, Dezembro 2011.

[19] Fliegel, K., Vitek, S., Klima, M., Pata, P. Open source database of images DEIMOS: high dynamic range and stereoscopic content. In Proc. SPIE 8135, 2011.

[20] Martin SLANINA, Tomas KRATOCHVIL, Vaclav NY, Libor BOLEEK, OndEj KALLER, Ladislav POLAK, "Testing QoE in Different 3D HDTV Technologies", RADIOENGINEERING, Vol. 21, num. 1, Abril 2012.

[21] ITU Recommendation J.148, "Requirements for an objective perceptual multimedia quality model", Tech. Rep., ITU Telecommunication Standardization Sector, Maio 2004.

[22] ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference", Tech. Rep., ITU Telecommunication Standardization Sector, Mar. 2001.

[23] J. You, U. Reiter, M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio-visual services: A survey.", Signal Processing: Image Communication, pp. 482â501, 2010.

[24] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Muller, "Quality assessment of 3D video in rate allocation experiments", IEEE International Symposium on Consumer Electronics, 2008, pp. 1-4.

[25] N. Katoh and M. Ito, "Gamut mapping for computer generated images", in Fourth Color Imaging Conference: Color Science, Systems and Applications. IST/SID, 1996, pp. 126â129.

[26] Quan Huynh-Thu, Member, IEEE, Marcus Barkowsky, Member, IEEE, and Patrick Le Callet, Member, IEEE, "The Importance of Visual Attention in Improving the 3D-TV Viewing Experience: Overview and New Perspectives", IEEE TRANSACTIONS ON BROADCASTING, VOL. 57, NO. 2, Junho 2011.

[27] L. Jansen, S. Onat, and P. Konig, "Influence of disparity on fixation and saccades in free viewing of natural scenes", J. Vis., vol. 9, no. 1, pp. 1â19, Jan. 2009.

[28] D. A.Wismeijer, C. J. Erkelens, R. van Ee, and M.Wexler, "Depth cue combination in spontaneous eye movements", J. Vis., vol. 10, no. 6, pp. 1â15, Jun. 2010.

[29] J. HÃkkinen, T. Kawai, J. Takatalo, R. Mitsuya, and G. Nyman, "What do people look at when they watch stereoscopic movies?", in Proc. SPIE Conf. Stereoscopic Displays Appl. XXI, San Jose, CA, Jan. 2010, vol. 7524.

[30] C. Ramasamy, D. House, A. Duchowski, and B. Daugherty, "Using eye tracking to analyze stereoscopic filmmaking", in Proc. SIGGRAPH 2009: Posters, 2010.

[31] Q. Huynh-Thu and L. Schiatti, "Examination of 3D visual attention in stereoscopic video content", in Proc. SPIE Conf. Human Vis. Electron. Imaging XVI, San Francisco, CA, Jan. 2010.

[32] Sockets em java, `http://docs.oracle.com/javase/tutorial/networking/sockets/definition.html`, 21 Outubro 2013.

[33] Java, `http://docs.oracle.com/javase/`, 21 Outubro 2013.

[34] M. Aldaba, J. Linhares, P. Pinto, S. Nascimento, K. Amano, and D. Foster. Visual sensitivity to color errors in images of natural scenes. Visual Neuroscience, 23(3-4):555559, 2006.

[35] M. Bernardo, M. Pereira, A. Pinheiro, P. Fiadeiro. A Study on the User Perception to Color Variations, MMâ12, 29 Outubro a 2 Novembro, 2012.

[36] ITU-T Recommendation BT 500-13: Methodology for the subjective assessment of the quality of television pictures. Technical report, ITU Telecom. Standard. Sector, 2012.

[37] Panasonic Corporation, "Panasonic AVCCAM AG-3DA1integrated twins-lens 3D camera recorder", 2011.

[38] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video", Electronic Imaging, 3D Image Processing and Applications, USA, Jan. 2010.

[39] A. Chambolle, T. Pock, "A first-order primal-dual algorithm for convex problems with application to imaging", Journal of Mathematical Imaging and Vision, vol. 40, no. 1, pp. 120-145, Maio 2011.

[40] ITU-T P.910, "Subjective video quality assessment methods for multimedia applications", International Telecommunication Union, Abr. 2008.

[41] H. Kalva, L. Christodoulou, and B. Furht, "Evaluation of 3DTV service using asymmetric view coding based on MPEG-2", in 3DTV Conference, Maio 2007, pp. 1 a 4.

[42] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment", in 15th European Signal Processing Conference (EUSIPCO â07), Poznan, Poland, Set. 2007.

[43] M. Brotherton, Q. Huynh-Thu, D. Hands, and K. Brunnstrom, "Subjective multimedia quality assessment", IEICE Transactions on Fundamentals of Electronics Communications and Computer Science, vol. 89, no. 11, pp. 2920â2932, 2006.

[44] D. M. Rouse, R. Pepion, P. Le Callet, and S. S. Hemami, "Tradeoffs in subjective testing methods for image and video quality assessment", Proc. SPIE 7527, Human Vision and Electronic Imaging XV, Fevereiro 2010.

[45] Q. Huynh-Thu, M.-N. Garcia, F. Speranza, P. Corriveau, and A. Raake, "Study of rating scales for subjective quality assessment of high-definition video", IEEE Transactions on Broadcasting, vol. 57, no. 1, pp. 1â14, Mar. 2011.

[46] Randot Stereo Test, `http://precision-vision.com/index.cfm/product/255_7/randot-stereo-test.cfm`, 21 Outubro 2013.

[47] DominÃncia ocular, http://en.wikipedia.org/wiki/Ocular_dominance, 21 Outubro 2013.

[48] Miller, R. G. Beyond ANOVA: Basics of Applied Statistics. Boca Raton, FL: Chapman Hall, 1997.

[49] VQEG, "Final report from the video quality experts group on the validation of objective models of multimedia qualiy assessment, Phase 1", 2008.

[50] M. Aldaba, J. Linhares, P. Pinto, S. Nascimento, K. Amano, and D. Foster. Visual sensitivity to color errors in images of natural scenes. Visual Neuroscience, 23(3-4):555 559, 2006.

[51] Electromagnetic-spectrum, http://www.pion.cz/_sites/pion/upload/images/a14cf10a5583d19f7cfdebd63cf64382_electromagnetic-spectrum.png, 21 Outubro 2013.

[52] Matalab logo, http://upload.wikimedia.org/wikipedia/commons/2/21/Matlab_Logo.png, 21 Outubro 2013.

[53] Ishihara9, http://upload.wikimedia.org/wikipedia/commons/e/e0/Ishihara_9.png, 21 Outubro 2013.

[54] Randot Stereo Test imagem, http://ecx.images-amazon.com/images/I/91N7ZScKXPL._SL1500_.jpg, 21 Outubro 2013.

[55] PR650, http://www.photoresearch.com/current/images/pr650.jpg, 21 Outubro 2013.

[56] Eye dominance, http://www.topendsports.com/testing/images/eye-dominance.jpg, 21 Outubro 2013.

[57] Sinal visual, https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRlbz6FQTJpwq4UBWj3gpeHy5-QaQIK7BuHtVHPSZQkCWp6sKXe, 21 Outubro 2013.

[58] Schematic diagram of the human eye, http://upload.wikimedia.org/wikipedia/commons/8/8a/Schematic_diagram_of_the_human_eye_pt.svg, 21 Outubro 2013.

[59] Connect, http://docs.oracle.com/javase/tutorial/figures/networking/6connect.gif, 21 Outubro 2013.