

Sistema flexível de realidade aumentada de baixo custo com base em projecção estereoscópica

Fernando Silva

Instituto superior de Engenharia de Coimbra
Rua Pedro Nunes, 3030-199 Coimbra
fsilva@isec.pt

Nuno Rodrigues

Instituto superior de Engenharia de Coimbra
Rua Pedro Nunes, 3030-199 Coimbra
nunorod@isec.pt

Adérito Fernandes Marcos

Universidade Aberta, Dpt. Ciências e Tecnologia
Palacio Ceia, Rua Escola Politécnica, 147
1269-001 – Lisboa
marcos@univ-ab.pt

Manuel João Ferreira

Universidade do Minho
Campus de Azurém, 4800-058 – Guimarães
mjf@dei.uminho.pt.

Resumo

A Realidade Aumentada usa a renderização de modelos de objectos para complementar informação numa cena física. Para conseguir este objectivo (atingir esta meta), esses modelos necessitam de ser referenciados aos objectos físicos que é suposto serem aumentados. Esta referenciação requer uma calibração dos dispositivos usados para a projecção das imagens aumentadas, assim como dos dispositivos usados para adquirir a informação 3D da cena física. No caso de o aumento de realidade incluir nova geometria, as imagens aumentadas deverão ser projectadas e visualizadas em estéreo, de modo que a cena composta resultante faça sentido. Usualmente, bons exemplos de realidade aumentada são muito caros e as soluções implementadas são rígidas. Neste artigo descrevemos uma arquitectura e a sua implementação como um sistema flexível e de baixo custo, para criação e projecção de realidade aumentada. Na secção final apontamos alguns desenvolvimentos futuros para este sistema, requerendo mais experimentação.

Palavras chave

Realidade Aumentada, Realidade Virtual, Projecção Estéreo.

1. INTRODUÇÃO

1.1 Colocação do problema

A Realidade Virtual (RV) é uma tecnologia emergente, que se tem vindo a desenvolver em especial nos últimos vinte anos, mas que está a adquirir maturidade. Continua a ser uma área em desenvolvimento, não só no campo da investigação, mas também no campo das indústrias de hardware e de software.

Uma visão geral da utilização de RV e de ambientes virtuais é apresentada num extenso artigo de Robert Stone [Stone2002], que aponta várias aplicações na engenharia, na micro-tecnologia e nano-tecnologia, na engenharia aero-espacial, na ergonomia, na defesa, na medicina e cirurgia, na herança cultural, nas vendas a retalho, na educação, na visualização de dados e visualização científica.

Uma experiência de RV é apresentada por certos autores [Burdea2003] como sendo delimitada por um triângulo cujos vértices são a *informação*, a *imersão* e a *interacção*. Sherman e Craig [Sherman2003] são mais analíticos e tornam-se mais explícitos, ao considerarem quatro elementos chave para realizar uma experiência de

realidade virtual: *um mundo virtual*, *imersão*, *feedback sensorial* e *interactividade*.

Pode entender-se o conceito de *mundo virtual* como um conjunto de dados organizados que pretendem representar entidades físicas ou entidades imaginadas, que podem ser instanciadas recorrendo ao uso de computador equipado com os elementos adequados para esse tipo de instanciação.

Imersão é a sensação de estar num ambiente; pode ser um estado puramente mental ou pode ser complementado através de meios físicos.

Feedback sensorial é um ingrediente essencial para um sistema de RV. O sistema de RV fornece *feedback sensorial* directo aos participantes, baseado nas suas posições físicas. Esse *feedback* é frequentemente de tipo visual, mas pode ser de outro tipo, como háptico.

Para a RV parecer autêntica, deverá responder às acções do utilizador, isto é, apresentar alguma forma de *interactividade*, quer se trate de um jogo ou de um qualquer simulador.

Desde a década de 1990 que os escritores de ficção científica têm contribuído para a divulgação de visões

acerca do papel dos computadores num mundo futuro, descrevendo universos 3D sintéticos que parecerão tão credíveis como o universo físico. Os sistemas de realidade virtual permitem criar ciberespaços onde é possível interagir com objectos e com pessoas a um nível virtual. Nesses espaços, as leis convencionais de espaço e tempo não necessitam ser seguidas com rigor, tudo pode ser simulado desde que possa ser programado [Vince95].

Existe no entanto um grande número de situações em que a realidade virtual, por mais imaginativa e tecnológica que seja, precisa de ser referenciada ao mundo físico, para acrescentar informação e/ou funcionalidades essenciais a determinados objectivos. São exemplos disso, o treino de montagem de peças mecânicas, em que o utilizador faz os movimentos de instalação de uma peça virtual num dispositivo físico, ou a visualização de artefactos arqueológicos incompletos, em que é acrescentada virtualmente ao artefacto a parte em falta. Estamos nestes casos perante aquilo que se pode designar por *realidade aumentada*.

A *realidade aumentada* consiste na conjugação de elementos informativos de realidade virtual com cenas de realidade física. Com a utilização de dispositivos e técnicas adequadas (espelhos semi-transparentes, ecrãs semi-transparentes, projecção estereoscópica) pode acrescentar-se informação visual sintética para complementar informação visual existente fisicamente numa cena. A mistura de mundos virtuais com o mundo físico para produzir um novo ambiente onde objectos físicos e objectos virtuais possam coexistir e interagir em tempo real, pode ser designada por *realidade mista*. Neste caso podemos considerar que na cena estão a ser visualizados elementos de realidade física, de realidade virtual, e eventualmente de realidade aumentada. A designação de realidade mista é usada especialmente quando, para além da existência de elementos de realidade física, os elementos informativos virtuais são totalmente sintéticos. No entanto, por vezes os termos *realidade aumentada* e *realidade mista* são usados como sinónimos.

As tecnologias chave subjacentes a estes sistemas de *realidade aumentada* e de *realidade mista* são a visão computacional e os sensores para captação de interacção do utilizador, a computação gráfica em tempo real, e os dispositivos de visualização gráfica. A computação gráfica fornece a base para a criação de imagens sintéticas, que em conjugação com dispositivos de entrada e de saída específicos, cria *ambientes virtuais* (AV) que podem conter projecções de modelos 3D de objectos físicos ou de objectos abstractos. Nesses ambientes virtuais, podem também ser incorporados efeitos acústicos para simular sons, assim como outros dispositivos que forneçam sensações tácteis (toque) ou hápticas (força), para completar a sensação de ilusão [Vince95]. A utilização de várias entidades como meios de representação e de comunicação de informação, levam-nos a considerar a utilização de técnicas *multimédia* na criação de mundos virtuais. Esses ambientes virtuais com

geometria complexa e com animação, requerem poder de cálculo e capacidades gráficas especiais, que num passado recente eram características de *workstations* de custo apreciável. Para esses mundos virtuais se tornarem efectivos, ou seja, para se manifestarem, e para terem o impacto pretendido sobre o utilizador, necessitam habitualmente de dispositivos de visualização especiais, que por vários motivos apresentam um custo considerável. Há ainda a considerar o custo não desprezável de sistemas que permitem captar informação do meio físico, quer sejam para extrair informação geométrica da cena, quer seja para referenciar o utilizador no meio dessa cena, e assim permitir a *interactividade*.

1.2 Alguns sistemas de realidade virtual e realidade aumentada

Podemos referir como exemplos representativos, o sistema *CAVE* [Cruz-Neira93], o sistema *The Virtual Showcase* [Bimber01], e o sistema *Augmented Room* [Pereira04]. Todos usam relações geométricas rígidas entre a cena e o sistema de projecção, requerendo a afinação mecânica precisa dos seus elementos e a calibração correspondente. O primeiro constitui uma referência na área de realidade virtual, podendo aceitar aplicações de realidade aumentada, no entanto torna-se uma solução suficientemente cara para uma maioria dos laboratórios de investigação não terem capacidade para a sua aquisição. O segundo e terceiro sistemas constituem exemplos muito interessantes do ponto de vista de realidade aumentada aplicada a visualização de artefactos museológicos incompletos, mas são sistemas com uma aplicação em áreas restritas.

1.3 A nossa proposta

Já no início da década 2000 começou a ser possível correr, em hardware de uso corrente, aplicações com funcionalidades de tempo real que poucos anos antes necessitavam de equipamentos com custos elevados, constituindo o *Oceanário Virtual* da Expo98 um exemplo paradigmático. Também por essa ocasião Leonardis & Alter [Leonardis00] referiam que na área de software a comunidade de computação gráfica tem vindo a desenvolver cada vez melhores, mais rápidos e mais complexos métodos de animação para a criação de ambientes virtuais sintéticos tridimensionais.

Em anos mais recentes tem continuado a haver importantes desenvolvimentos nas áreas de hardware e de software, nomeadamente na integração da capacidade de processamento das GPU's (Graphic Processing Unit) nos sistemas gráficos, assim como no desenvolvimento de dispositivos de interface com o utilizador com uma melhor usabilidade e de mais baixo custo, e no desenvolvimento de projectores de vídeo em estéreo a uma fracção do custo de projectores utilizados tradicionalmente.

Tirando partido desses desenvolvimentos, fazemos uma contribuição para a generalização da aplicação dessas tecnologias, propondo um sistema que, para além de ser flexível, está ao alcance dos orçamentos da maioria dos

laboratórios de investigação, para o desenvolvimento de tecnologia genérica na área da realidade virtual e realidade aumentada.

Do ponto de vista de hardware, o sistema deverá basear-se na utilização dos seguintes elementos:

- Um computador de secretária equipado com carta gráfica com capacidade de gerar imagens em estéreo activo; se esta permitir sincronização de frame por hardware, permitirá a utilização de dois sistemas iguais para projecção de imagens estéreo compostas;
- Sistema de projecção e visualização estéreo, havendo basicamente duas soluções, uma com dois projectores a emitirem luz polarizada para ser vista com óculos polarizados passivos, e outra com um só projector estéreo a emitir frames alternadamente esquerdo e direito, para serem vistos com óculos comutados;
- Sistema de aquisição de imagem estéreo, constituído por um par de câmaras digitais; a qualidade das câmaras e das lentes utilizadas é importante para conseguir uma boa calibração total do sistema; a interface 1394 ou gigabit Ethernet devem ser preferidas, pelo melhor desempenho que permitem; também um sistema de orientação individual das câmaras em 2 eixos é aconselhável, pela flexibilidade que permitem na sua forma de utilização;
- Um sistema de seguimento (tracking) do utilizador, que permita obter a sua posição e orientação em relação a uma referência solidária com a cena; além de sistemas de detecção de infravermelhos compostos por emissor/reflector/receptor disponíveis a baixo custo, têm surgido outros elementos de interface na área dos jogos de consola ou de computador com preços muito competitivos;
- Superfícies de projecção, podendo ser usadas as próprias paredes brancas de uma sala ou painéis amovíveis que se podem posicionar conforme a aplicação; painéis semi-transparentes abrem a possibilidade de efectuar retroprojecção, assim como de experimentar outras técnicas de interacção por detecção posterior de sinais vindos do utilizador.

Existem trabalhos anteriores relativos a sistemas de visualização baseados em projecção, como os trabalhos referidos em [Hereld00], [Brown02] e [Raskar02]. Os dois primeiros referem-se em especial à composição de imagens de projecção geradas por mais de um projector. O terceiro endereça também uma forma de aumento de realidade, mas sem a inclusão de aumento de geometria, incidindo em especial na texturação de superfícies físicas já existentes na cena. No caso presente, além de alterar cores e texturas de superfícies físicas já existentes como em [Raskar02], pretende-se também incluir o aumento de realidade pela via da projecção em estéreo de nova geometria inexistente na cena física, mas calibrada e referida a essa mesma cena física.

Com base nestes elementos fundamentais montámos um sistema que permite provar a exequibilidade deste tipo de solução.

2. REQUISITOS TÉCNICOS DA SOLUÇÃO

Para termos um sistema de realidade virtual temos de dispor de vários elementos, conforme já referimos anteriormente: um mundo virtual, imersão, feedback sensorial, interactividade. Para termos *realidade aumentada* vai ser necessário fazer o *registo* do mundo virtual relativamente ao ambiente físico. Na realidade, a parte virtual correspondente ao aumento de realidade, tem de se ajustar (ou justapor) perfeitamente à parte física da cena que se pretende aumentar. E esse ajuste tem de ser em tamanho, posição e orientação, de modo a haver continuidade entre uma parte e a outra, do ponto de vista da percepção visual do observador.

O registo da informação muda substancialmente o nível do problema. Enquanto em realidade virtual podemos trabalhar num espaço afim, para fazer aumento de realidade é necessário passarmos a trabalhar num espaço euclidiano. Isso obriga a fazer a calibração de todo o sistema, quer seja fazendo os alinhamentos mecânicos de pormenor no caso de termos uma geometria fixa e usar matrizes de projecção calculadas uma só vez, como acontece nos casos *CAVE* e *The Virtual Showcase*, ou é necessário calibrar todo o sistema e calcular as matrizes de projecção para cada nova configuração, no caso do sistema que estamos a propor e nos trabalhos de Bimber&Raskar [Bimber05].

O problema fundamental de calibração neste caso, consiste em calibrar o projector, isto é, saber qual a sua *pose* em relação à cena física, e calcular a transformação que o projector aplica a uma imagem que lhe é enviada, de modo que a imagem resultante ilumine determinados pontos 3D. Colocando o problema de outro modo, precisamos saber a matriz que relaciona um conjunto conhecido de pontos numa imagem passada ao projector para este projectar, com os correspondentes pontos 3D projectados numa superfície. Para isso precisamos de determinar as coordenadas desses pontos 3D, e estabelecer a correspondência, o que requer a utilização de sensores que detectem as coordenadas (X,Y,Z) desses pontos. Faz todo o sentido identificar esses pontos e determinar as suas coordenadas utilizando imagens. Mas essas imagens terão de ser adquiridas em pares estéreo, para ser possível determinar profundidades por meio de uma reconstrução 3D. Assim, o sistema necessita de um par de câmaras para adquirir imagens estéreo. Como necessitamos de calcular coordenadas euclidianas, este par de câmaras terá por seu lado de ser previamente calibrado. Estes procedimentos são implementados neste trabalho de uma forma apresentada mais à frente neste texto.

O aumento de realidade pode assumir dois aspectos: um deles consiste somente em alterar as cores que o utilizador vê em superfícies já existentes na cena física, e o outro, consiste em acrescentar geometria virtual à geometria física existente. O primeiro pode ser

conseguido utilizando um projector simples, pois o alvo da cor projectada tem exactamente a mesma profundidade em cada ponto da superfície que queremos alterar, ou seja, a profundidade virtual que queremos perceber coincide com a profundidade física. O segundo caso, com aumento de geometria, requer que a projecção seja feita em estéreo e consequentemente também a visualização, pois queremos dar ao sistema visual do utilizador a ilusão de uma profundidade da parte aumentada que não coincide com a profundidade física.

A visualização em estéreo com comutação entre imagens esquerda/direita, tem um requisito adicional quando queremos combinar imagens estéreo de duas ou mais fontes, como acontece no caso do sistema CAVE, em que é necessário combinar as partes de imagem que se sobrepõem através de arestas: a sincronização de frame. Essa sincronização é necessária para garantir que, quando um projector está a projectar a sua imagem esquerda, o outro também esteja a projectar a sua imagem esquerda, e de forma idêntica para as imagens direitas. Esse sincronismo tem de ser preciso, para garantir a qualidade da visualização. Por esse motivo as placas gráficas deverão ter a capacidade de sincronização de frame por hardware, como acontece, por exemplo com a carta gráfica NVidia Quadro 4500-G. Esta carta dispõe de um módulo de sincronização de frame por hardware que permite ligar várias placas gráficas em cascata por meio de cabos com fichas RJ-45.

A visualização em estéreo activo requer ainda a sincronização das imagens esquerda/direita com a capacidade de as visualizar com os olhos esquerdo/direito, o que é conseguido com a intercalação entre a saída da carta gráfica e o projector/monitor, de um dispositivo que active a lente esquerda/direita ao mesmo tempo que bloqueia a lente contrária.

Para que as imagens virtuais ou aumentadas sejam projectadas de modo a serem correctamente visualizadas pelo utilizador, requer a utilização de algum dispositivo que capte a posição e orientação do utilizador. Existem no mercado dispositivos de baixo custo como *Track-IR* da *Natural Point* ou o comando *Wii-Remote* da *Nintendo*.

3. DESCRIÇÃO DO HARDWARE DO SISTEMA

Apresenta-se aqui um a descrição dos equipamentos que constituem actualmente o hardware do sistema, e que foram usados na experimentação que temos vindo a realizar.

Foi usado um computador de secretária com duas portas IEEE 1394 para ligar duas câmaras de vídeo digital, uma porta RS-232-C para controlo de uma unidade de orientação das câmaras, uma carta gráfica NVidia Quadro 4500-G e um monitor com banda de passagem suficiente para permitir uma frequência vertical superior a 120Hz.

As câmaras utilizadas são do modelo Dragonfly2, da Point Grey Research, Inc, que permitem a aquisição de imagens em diversos formatos, com várias resoluções, e a uma taxa que pode variar entre 1,875 e 60 imagens por segundo, dependendo do formato e resolução

pretendidos. Têm interface IEEE 1394, cuja vantagem consiste na auto-regulação de sincronismo para aquisição em simultâneo de imagens estéreo. Podem ser adquiridas em kit, que já contém uma lente com adaptação “CS mount” (Figura 1), assim como um SDK (*Software Development Kit*), que inclui uma API (*Application Programming Interface*) que pode ser usada para simplificar o desenvolvimento das aplicações em C/C++.

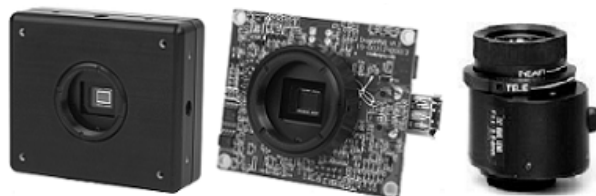


Figura 1. Câmara Dragonfly2 com caixa e lente

As câmaras são montadas em unidades “*Computer-Controlled Pan-Tilt Unit*” (Figura 2), fornecidas pela *Directed Perception, Inc*, que permitem realizar movimentos de rotação sobre o eixo vertical (movimento de *Pan*) ou sobre o eixo horizontal (movimento de *Tilt*). Este equipamento é passível de ser controlado a partir de um computador por recurso à comunicação via norma RS-232. Disponibiliza um modo de controlo baseado em *strings* ASCII, vantajoso em situações de teste, pois permite o envio de comandos do utilizador por inserção directa usando por exemplo o *Hyperterminal*.



Figura 2. Unidade Pan & Tilt com controlador

O projector de vídeo utilizado é um modelo *DepthQ*, representado na Figura 3, que suporta projecção estereoscópica activa, por comutação de imagens esquerda e direita, até um máximo de 2x60 imagens por segundo, com uma resolução nativa de 800x600 (1280x720 no modelo mais recente). A utilização deste equipamento não obriga a nenhuma configuração especial, bastando ligar a respectiva entrada à saída da carta gráfica e fazer a continuação da ligação para um monitor.

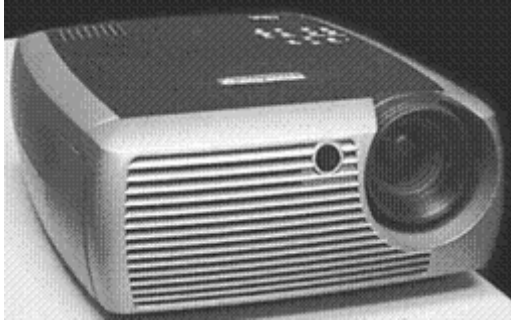


Figura 3. Projector Stereo InFocus DepthQ

Para uma visualização de imagens estéreo com óculos comutados, é necessário intercalar entre a saída da carta gráfica e o projector, um dispositivo que capta os sinais de controlo da carta gráfica e os emite para óculos com lentes comutadas. Na **Figura 4**, apresenta-se o modelo *E-D Wireless 3D Glasses* da *e-dimensional*, conjunto constituído pelos óculos com comutação, o dispositivo de sincronização de vídeo e um emissor de infra-vermelhos.



Figura 4. Óculos estéreo, sincronizador e emissor

Como sistema de seguimento (tracking) do utilizador, que permita obter a sua posição e orientação em relação a uma referência solidária com a cena, foi usado o *Track-IR* da *Natural Point* (**Figura 5**), um dispositivo com detecção por infravermelhos composto por emissor-reflector-receptor, com 6 graus de liberdade.



Figura 5. Emissor-reflector-receptor de infravermelhos

Como superfícies de projecção foram usadas uma parede de sala, um painel de projecção, uma superfície planar horizontal e uma caixa 3D para “pintar” com luz projectada.

4. IMPLEMENTAÇÃO DE SOFTWARE

4.1 Comunicação com o hardware

De modo a facilitar a tarefa de implementação do software, quer ao nível do sistema final, quer ao nível da construção de protótipos diversos para teste de conceitos, é importante criar uma camada intermédia de software que permita que a comunicação com os elementos de hardware que compõem o sistema seja feita a um nível de abstracção elevado, escondendo assim os detalhes de mais baixo nível característicos da comunicação com aqueles elementos.

Assim, avançou-se para a criação de um conjunto de módulos de software, seguindo uma abordagem orientada a objectos, cujas funcionalidades incluem:

- Configuração inicial da comunicação com os equipamentos.
- Controlo da orientação das unidades *Pan & Tilt* no eixo vertical (*pan*) e horizontal (*tilt*).
- Captura de imagens pelas câmaras.
- Conversão de formato das imagens capturadas.

4.2 Calibração das câmaras

O modelo matemático de partida para representar o comportamento das câmaras é o modelo de projecção em perspectiva [Hartley00], que traduz a transformação de um ponto X do mundo tridimensional, representado em coordenadas homogéneas, num ponto x no plano da imagem (a duas dimensões, portanto) também com representação em coordenadas homogéneas. Esta transformação pode ser escrita como:

$$\tilde{x} = K \cdot [R \ t] \cdot \tilde{X} = P \cdot \tilde{X},$$

onde:

- $K = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$ é a matriz que contém os

parâmetros intrínsecos da câmara,

- $[R \ t]$ é uma matriz 3x4 representativa da *pose* da câmara, ou a relação entre o seu sistema de coordenadas e o sistema de coordenadas global ou do mundo, sendo R a matriz 3x3 que contém a rotação que transforma um sistema de coordenadas no outro e t é um vector coluna de 3 elementos que contém a componente de translação da dita transformação.

De modo a introduzir no modelo uma forma de representação dos efeitos de distorção da imagem provocados, mais notoriamente, por imperfeições em lentes de baixo custo, consideram-se como parâmetros adicionais os coeficientes de distorção radial k_1 e k_2 e de distorção tangencial p_1 e p_2 [Heikkila97].

Para os parâmetros intrínsecos α_x , α_y , s , x_0 e y_0 , os coeficientes de distorção k_1 , k_2 , p_1 e p_2 , a matriz R e o vector t de *pose* da câmara, a projecção de um ponto $Q=(X, Y, Z)^T$ do mundo tridimensional num pixel $q=(u,v)^T$ da imagem será obtida da seguinte forma:

1. $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t.$
2. $\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{1}{z} \begin{bmatrix} x \\ y \end{bmatrix}.$
3. $\begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} x' + x'(k_1 r^2 + k_2 r^4) + 2p_1 x' y' + p_2 (r^2 + 2x'^2) \\ y' + y'(k_1 r^2 + k_2 r^4) + p_1 (r^2 + 2y'^2) + 2p_2 x' y' \end{bmatrix},$
com $r^2 = x'^2 + y'^2.$

$$4. \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \alpha_x x'' + sy'' + x_0 \\ \alpha_y y'' + y_0 \end{bmatrix}.$$

O processo de calibração passa pela determinação dos parâmetros intrínsecos (e coeficientes de distorção) característicos de cada câmara e pela determinação da respectiva *pose*, relativamente ao sistema de coordenadas global definido.

A implementação adoptada para este processo é baseada em [Zhang98] e [Liu04], e inclui os seguintes passos:

1. Imprimir um padrão de calibração com pontos conhecidos (por exemplo um tabuleiro de xadrez) e colá-lo a uma superfície plana;
2. Obter algumas imagens desse padrão a partir de *poses* diferentes, movendo o objecto ou a câmara;
3. Detectar pontos relevantes do padrão nas diversas imagens;
4. Obter uma estimativa inicial para os parâmetros intrínsecos da câmara bem como dos parâmetros extrínsecos (*pose*) referentes a cada imagem obtida.
5. Refinar os valores obtidos através de um processo de minimização não linear (algoritmo de Gauss-Newton [Hartley00]).

De acordo com [Zhang98], os dados utilizados na calibração, nomeadamente as coordenadas dos pontos relevantes nas imagens devem ser previamente normalizados para que se possam evitar problemas de instabilidade numérica. Assim, os pontos detectados são transformados de modo a que: a) o seu centróide seja deslocado para a origem e b) a sua distância média à origem seja $\sqrt{2}$ [Hartley00].

4.3 Reconstrução

Uma vez calibradas as câmaras, podemos tentar reconstruir informação tridimensional a partir de pares de imagens obtidas através delas.

Recorrendo, por exemplo, a métodos simples de triangulação linear [Hartley00], é possível estimar as coordenadas tridimensionais de um ponto a partir das coordenadas das projecções desse ponto nas imagens capturadas por cada uma das câmaras. Se considerarmos $\tilde{x}_e = k(x_e, y_e, 1)^T$ e $\tilde{x}_d = k(x_d, y_d, 1)^T$ as representações em coordenadas homogéneas das projecções desse ponto em cada uma das imagens (capturadas com a câmara esquerda e direita, respectivamente e devidamente corrigidas dos efeitos da distorção das lentes), e $\tilde{X} = (X, Y, Z, 1)^T$ as suas coordenadas homogéneas no mundo tridimensional, o modelo matemático da câmara estabelece que:

$$\tilde{x}_e = P_e \tilde{X}$$

$$\tilde{x}_d = P_d \tilde{X},$$

em que P_e e P_d são as matrizes de projecção da câmara esquerda e direita, respectivamente.

Detectando as coordenadas de um mesmo ponto projectado em cada uma das imagens e sabendo os valores das matrizes P_e e P_d (obtidos no processo de calibração), podemos resolver as equações em ordem a \tilde{X} e assim reconstruir as coordenadas 3D do ponto:

$$\begin{bmatrix} x_e P_e^3 - P_e^1 \\ y_e P_e^3 - P_e^2 \\ x_d P_d^3 - P_d^1 \\ y_d P_d^3 - P_d^2 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0 \Leftrightarrow A \tilde{X} = 0,$$

onde P^i representa a *i*ésima linha da matriz P .

A solução para o sistema de equações pode ser obtida por cálculo do vector singular unitário correspondente ao menor valor singular da matriz A [Hartley00].

4.4 Calibração do projector

A utilização de projectores é uma das formas de disponibilizar imagens de realidade aumentada. Nomeadamente, é possível projectar imagens directamente sobre os objectos cujas características se desejam manipular (cor, iluminação ou textura, por exemplo).

Consideremos a tarefa de iluminar um determinado ponto X do mundo tridimensional, com uma determinada cor, recorrendo a um projector. A informação que pode ser passada ao projector consiste numa imagem a duas dimensões que será então projectada no espaço à frente da lente. O problema a resolver consiste, portanto, em determinar qual o ponto x na imagem que deve conter a informação de cor a ser projectada em X . À determinação do mapeamento das coordenadas dos pixels da imagem em coordenadas dos pontos tridimensionais por eles iluminados vamos chamar *calibração do projector*.

De acordo com o defendido em [Bimber05], o modelo de formação de imagem de um projector pode ser aproximado por um modelo semelhante ao usado para aproximar o modelo de formação de imagem das câmaras. Assim, podemos escrever este modelo como:

$$\tilde{x} = K[R \ t] \tilde{X} = P \tilde{X},$$

onde $\tilde{x} = k(x, y, 1)^T$ é a representação em coordenadas homogéneas do pixel da imagem do projector, $\tilde{X} = (X, Y, Z, 1)^T$ é a representação, também em coordenadas homogéneas, do ponto tridimensional iluminado na superfície de projecção, R e t representam os parâmetros extrínsecos e K é a matriz de parâmetros intrínsecos, semelhantes ao modelo matemático das câmaras.

O processo de calibração passa por obter as coordenadas de um conjunto de N pontos na imagem passada ao projector e as coordenadas respectivas no mundo 3D dos pontos por ele iluminados, que possibilitem a resolução simultânea de N pares de equações, representantes do modelo do projector, em ordem aos elementos da matriz P . Cada par de equações será:

$$\begin{bmatrix} -\tilde{X}^T & 0^T & x\tilde{X} \\ 0^T & -\tilde{X}^T & y\tilde{X} \end{bmatrix} \begin{bmatrix} P^{1T} \\ P^{2T} \\ P^{3T} \end{bmatrix} = 0,$$

onde P^i é a i ésima linha da matriz P .

São necessárias, no mínimo, 11 equações (6 pontos) para que o sistema possa ser determinado, já que a matriz P tem 11 graus de liberdade, mas, para fazer face às imprecisões inerentes ao processo, devem ser usadas mais equações (o que implica a utilização de mais pontos).

As coordenadas dos pontos podem ser obtidas através da projecção de uma imagem com um conjunto de pontos conhecidos, como a presente na **Figura 6**, e pela reconstrução da posição tridimensional desses mesmos pontos, a partir das imagens capturadas pelas câmaras, de acordo com o método exposto na secção anterior.

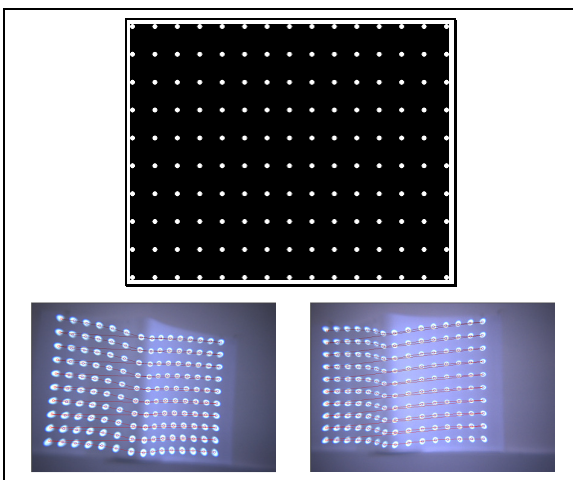


Figura 6. Imagem passada ao projector para calibração e imagens capturadas pelas câmaras.

4.5 Projecção

O sistema proposto pode ser usado como base para a criação de realidade aumentada de vários pontos de vista. Como exemplos, consideremos os seguintes propósitos:

- Alterar características de um objecto existente na cena fisicamente (a sua cor ou textura, por exemplo);
- Aumentar a cena física com objectos virtuais criados sinteticamente.

Para demonstrar a possibilidade de utilização do sistema para a satisfação destes propósitos, construiu-se um ambiente de testes constituído por duas superfícies planas de projecção, uma aproximadamente vertical e outra aproximadamente horizontal, aos quais se juntou uma pequena caixa branca que servirá como objecto físico cujas características se pretendem alterar (ver **Figura 7**).



Figura 7. Ambiente de projecção.

Assumindo que conhecemos os modelos de projecção e as poses das câmaras e do projector, obtidos através dos processos de calibração referidos atrás, podemos dirigir a nossa atenção à satisfação dos dois propósitos avançados.

Para alterar as características do objecto colocado fisicamente na cena (a caixa), recorreremos à iluminação do projector, fazendo-o projectar uma imagem de modo que as faces visíveis desse objecto, do ponto de vista do projector, sejam devidamente iluminadas de acordo com a cor e a textura que se lhe pretende sobrepor. Para isso vamos assumir que conhecemos a geometria desse objecto e a sua posição no mundo tridimensional. O problema resume-se então à projecção do modelo 3D daquele objecto numa imagem 2D, a fornecer ao projector, de acordo com o seu modelo de projecção e a sua pose.

Como facilmente se compreende, esta problemática é resolvida por diversos sistemas gráficos existentes, dos quais o *OpenGL* é um dos exemplos. Temos apenas que definir como parâmetros da “câmara virtual” do *OpenGL* os mesmos parâmetros obtidos por calibração do projector. No caso presente, isso traduz-se na utilização das matrizes homogéneas *Projection* e *Modelview* daquele sistema gráfico, extraídas a partir da matriz P deduzida na secção anterior, usando o procedimento referido em [Bimber05].

Para acrescentar à cena alguns objectos criados de forma sintética, a abordagem é um pouco mais complexa, devido à necessidade de dar resposta a uma série de novos problemas que se levantam.

Em primeiro lugar, para que os objectos sintéticos sejam visualizados de forma minimamente credível, é necessário que a sua projecção seja geometricamente correcta, ao nível do tamanho, posição e orientação. Isto implica que se leve em conta a posição do observador, já que, na prática, esta dificilmente irá coincidir com a posição do projector.

Em segundo lugar, os planos onde os objectos serão projectados não são necessariamente perpendiculares em relação ao eixo óptico do projector. Por isso, uma imagem rectangular causará uma forma não rectangular,

ao ser projectada nas superfícies alvo da projecção. Assim, para que a imagem seja visualizada de forma geometricamente correcta pelo observador, esta deverá sofrer uma transformação, de modo a compensar a distorção introduzida pelo próprio sistema de projecção.

A adicionar a este problema surge a utilização de dois planos de projecção distintos, em que a distorção da imagem projectada resultante da projecção não ortogonal, é diferente para cada um deles.

Por fim, para ser possível a visualização estereoscópica, há a necessidade de projectar imagens distintas para serem vistas pelo olho esquerdo e pelo olho direito do observador.

Para fazer face a estas condicionantes, a abordagem aqui proposta começa por focar-se na determinação da geometria dos dois planos de projecção utilizados. Como já se referiu, as áreas iluminadas em cada um dos dois planos de projecção são quadriláteros que, dadas as suas posições relativas ao projector, serão tendencialmente irregulares. Assim, a primeira tarefa será obter as coordenadas tridimensionais dos quadriláteros que são iluminados pelo projector, em cada um dos planos. Para isso, projecta-se uma imagem totalmente branca (**Figura 8**) e, através do processo de reconstrução referido atrás, partindo dos vértices da área mais clara nas imagens obtidas pelas duas câmaras, obtêm-se as respectivas coordenadas 3D.

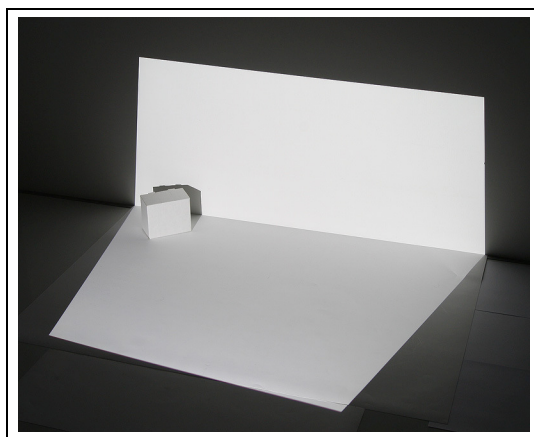


Figura 8. Iluminação dos planos de projecção.

Conhecendo a geometria das áreas iluminadas pelo projector, podemos então definir uma estratégia de produção das imagens dos objectos virtuais assente em dois passos:

1. Para cada plano de projecção, desenhar os objectos do ponto de vista do utilizador e guardar o resultado numa textura;
2. Aplicar cada uma das texturas ao respectivo quadrilátero 3D representativo do plano de projecção e passar a imagem assim formada ao projector.

Para executar o primeiro passo, isto é, desenhar os objectos do ponto de vista do utilizador, fazendo uso do *OpenGL*, começa por ser definido, para cada plano de projecção, um volume de visualização (ou *frustum*) que determine que parte da cena vista pelo utilizador deve ser projectada nesse plano. Cada *frustum* deve ser construído tendo como centro de projecção a posição do observador (ponto **O** da **Figura 9**), e tendo como arestas os vértices do rectângulo envolvente ao quadrilátero da área de projecção, alinhado com os eixos do sistema de coordenadas do mundo.

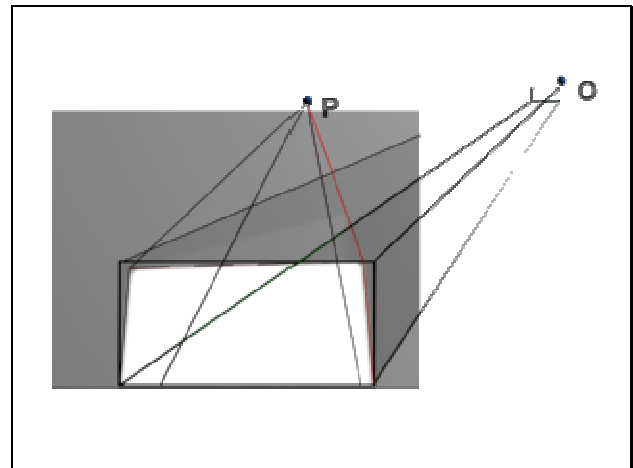


Figura 9. Projector (P) e volume de visualização do observador (O) para o plano de projecção vertical.

Definidos os volumes de visualização, a cena pode ser desenhada do ponto de vista do utilizador, usando cada um desses volumes e guardando cada resultado numa textura.

Para visualização estereoscópica é necessário duplicar o procedimento e, para cada um dos planos, definir dois volumes de visualização a partir de duas posições do observador ligeiramente diferentes, correspondentes às imagens vistas pelos olhos esquerdo e direito, respectivamente.

No segundo passo, as texturas criadas no primeiro passo são aplicadas aos quadriláteros correspondentes às áreas iluminadas em cada um dos planos. Neste passo, o volume de visualização deixa de ser considerado relativamente à posição do utilizador e passa a ser calculado do ponto de vista do projector. Isto implica que, na prática, estamos a aplicar uma textura a um objecto e a abordagem é semelhante à descrita acima, quando se referiu a problemática de alteração de características de objectos existentes fisicamente na cena.

A **Figura 10** mostra um exemplo de utilização do sistema em que pode ser visualizado o resultado das duas formas de criação de realidade aumentada descritas. Por um lado, podemos verificar que a pequena caixa foi “pintada” e “texturada” de modo a ter o aspecto de um relógio duplo. Por outro lado, foram projectados objectos virtuais (tabuleiro e peças de xadrez) para complementar a cena. Esses objectos virtuais foram projectados de modo a

parecerem ao utilizador como estando pousados na mesma superfície física que a caixa, e com uma profundidade que começa antes do plano de projecção vertical, e termina depois dele, efeito esse conseguido pelo recurso à estereoscopia activa.

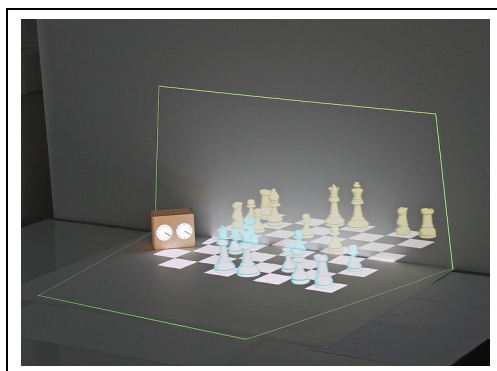


Figura 10. Visualização da cena contendo objecto físico alterado e adição de objectos virtuais.

Para obter a projecção da cena anterior, a imagem resultante da abordagem proposta e passada ao projector está representada na Figura 11.

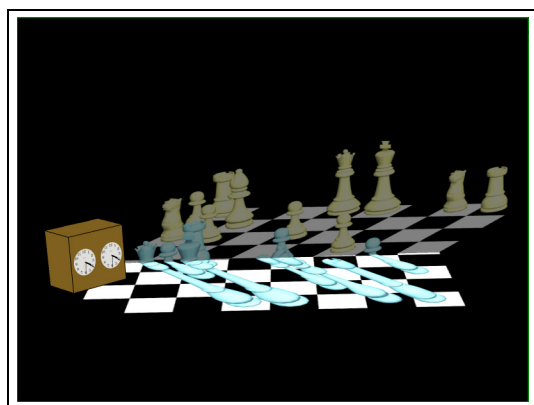


Figura 11. Imagem passada ao projector.

5. CONCLUSÕES

Neste trabalho apresentámos uma arquitectura para sistemas de realidade virtual e realidade aumentada com duas características fundamentais: flexibilidade, e baixo custo. Foram conjugadas duas formas de aumento de realidade: alteração de características de cor e textura de um objecto existente fisicamente na cena, e aumento de geometria, projectando sobre mais de um plano imagens sintéticas em estéreo, registadas em relação a uma referência em coordenadas globais. O protótipo de hardware e o software já desenvolvidos permitem fazer a prova do conceito, e perspectivam uma boa contribuição para o desenvolvimento de projectos de aplicação na área de realidade virtual e realidade aumentada. Serve também como um sistema privilegiado para estudo e teste de técnicas em computação gráfica, processamento de

imagem e visão computacional. O sistema apresenta boas possibilidades de evolução e de receber novos desenvolvimentos, não tendo ainda sido detectadas limitações de funcionamento. No entanto, quando o sistema for extensivamente testado com uma gama variada de aplicações, é possível virem a surgir, entre outros, problemas de oclusão visual, de dificuldade de ocultação de objectos brilhantes, e dificuldades na detecção robusta e interpretação de cenas 3D. Esses e outros problemas, não tendo uma solução genérica, terão de ser analisados e resolvidos no contexto em que possam eventualmente surgir.

6. DESENVOLVIMENTOS FUTUROS

Pretende-se continuar o desenvolvimento deste sistema em mais de um ponto de vista.

Do ponto de vista de hardware, será desenvolvido um novo sistema de orientação das câmaras, de modo a ser possível efectuar ambas as rotações de pan&tilt em torno do centro óptico das câmaras, assim como dotar o projector de movimento de rotação em torno do eixo vertical. Os suportes das câmaras serão montados num veio rotativo de modo a permitir a utilização de diferentes valores para a chamada linha-base. Será ainda acrescentado um *observador independente*, constituído por uma terceira câmara.

Do ponto de vista de software, serão desenvolvidos módulos de visualização panorâmica em estéreo, de aquisição e interpretação robusta de cenas 3D, e de projecção estéreo sobre superfícies não planares.

Do ponto de vista de arquitectura, pretende-se utilizar painéis com retro projecção e testar novas formas de interacção do utilizador, assim como implementar uma mini-CAVE.

A par destes desenvolvimentos, serão estudadas soluções de aplicação do sistema a casos concretos.

7. REFERÊNCIAS

- [Bimber05] Bimber, O. and Raskar, R., "Spatial Augmented Reality – Merging Real and Virtual Worlds", A K Peters, 2005.
- [Bimber01] Bimber, O., Fröhlich, B., Schmalstieg, D., and Encarnação, L.M., "The Virtual Showcase", In IEEE Computer Graphics & Applications, vol. 21, no.6, pp. 48-55, 2001.
- [Brown02] Michael S. Brown, W. Brent Seales, "A Practical and Flexible Tiled Display System", IEEE Proceedings of the 10th Pacific Conference on Computer Graphics and Applications, 2002.
- [Burdea03] Grigore C. Burdea, Philippe Coiffet; "Virtual Reality Technology, Second Edition"; Wiley-Interscience - A John Wiley & Sons, Inc., Publication, 2003.
- [Cruz-Neira93] Cruz-Neira C., D. Sandin and T. DeFanti; "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE"; Com-

puter Graphics (Proc. Siggraph 1993), ACM Press, 1993, pp.135-142.

- [Hartley00] Hartley, R. and Zisserman, A., Multiple View Geometry in Computer Vision, Cambridge University Press, 2000..
- [Heikkila97] Heikkila, J. and Silven, O., A Four-step Camera Calibration Procedure with Implicit Image Correction. In Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97) (June 17 - 19, 1997). CVPR. IEEE Computer Society, Washington, DC, 1106, 1997.
- [Hereld00] Mark Hereld, Ivan R. Judson, Rick L. Stevens, "Introduction to Building Projection-based Tiled Display Systems", IEEE Computer Graphics and Applications, (July/August 2000).
- [Leonardis00] A. Leonardis, F. Solina and R. Bajcsy, "Confluence of Computer Vision and Computer Graphics", NATO Science Series, 3. High Technology – Vol.84, Kluwer Academic Publishers, 2000.
- [Liu04] Liu, Y., Wu, Y., Wu, M. and Hu, X. Planar Vanishing Points Based Camera Calibration. In Proceedings of the Third international Conference on Image and Graphics (December 18 - 20, 2004). ICIG. IEEE Computer Society, Washington, DC, 460-463, 2004.
- [Pereira04] Pedro Pereira, Nuno Matos, Adérito Marcos; "Augmented room: a device for visualization of museum artifacts"; In book "Virtual Design and Automation", Ed. Weiss Z., published by Publishing House of Poznan - University of Technology 2005, pp.389-398, ISBN: 83-7143-201-1., 2004.
- [Raskar02] Ramesh Raskar "Projector-based Three Dimensional Graphics", PhD thesis, University of North Carolina, 2002.
- [Sherman03] Wiliam R.Sherman, Alan B.Craig; "Understanding Virtual Reality - Interface, Application, and Design", Morgan Kaufmann Publishers, 2003.
- [Stone02] Robert J. Stone; "Applications of Virtual Reality Environments: An Overview" in: Handbook of Virtual Environments: Design, Implementation, and Applications, edited by Kay M. Stanney, 2002, Lawrence Erlbaum Associates, ISBN 0-8058-3270-X.
- [Vince95] John Vince; "Virtual Reality Systems"; Addison-Wesley; 1995.
- [Zhang98] Zhang, Z., A Flexible New Technique for Camera Calibration, Technical Report, MSR-TR-98-71, 1998.