

# Inserção Dinâmica de Objectos Virtuais no Contexto de Fotografias Tiradas por Utilizadores

Rui Nóbrega    Nuno Correia

CITI, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa  
2829-516 Caparica, Portugal

`rui.nobrega@campus.fct.unl.pt, nmc@fct.unl.pt`

---

## Resumo

A introdução de objectos virtuais em ambientes de Realidade Aumentada em tempo real requer normalmente a configuração de marcas pré-definidas e a sua posterior detecção no cenário. Neste artigo é proposto um sistema semi-automático de análise de estrutura tri-dimensional em fotografias, com o objectivo de introduzir interactivamente objectos virtuais no cenário. Estes objectos tri-dimensionais são representados com a perspectiva correcta e reagem a elementos da fotografia como se fossem entidades no mundo real. Utilizando estes elementos virtuais é possível criar aplicações de realidade aumentada em que o utilizador tira uma ou mais fotos de um determinado local e introduz interactivamente objectos ou elementos virtuais que se misturam com a foto em tempo real. Neste artigo são discutidas novas formas de adquirir imagens e informação sobre o cenário envolvendo o utilizador e também como esses dados podem ser processados para uma correcta visualização final do ambiente de realidade mista.

## Palavras-Chave

Realidade Aumentada, Visão por Computador, Computação Gráfica, Multimédia

---

## 1. INTRODUÇÃO

A maior parte das aplicações de Realidade Aumentada (RA) utilizam imagens do mundo real como base mas com limitações e constrangimentos por forma a que os elementos e grafismos virtuais sejam integrados correctamente no cenário. Muitas usam cenários fixos com modelos pré-construídos à partida, enquanto outras utilizam referências no cenário como por exemplo marcas. Este artigo propõe um sistema para construir aplicações de RA em qualquer espaço sem um conhecimento *a priori* do mesmo. Mais concretamente, pretende-se criar ambientes de RA em espaços escolhidos pelo utilizador oferecendo a possibilidade de interacção com esse espaço de forma virtual mas tendo em conta propriedades do mundo físico.

Por detrás deste projecto está a motivação de construir aplicações interactivas que permitam ao utilizador refazer, redefinir e alterar virtualmente um espaço pessoal. Possíveis implementações deste sistema seriam importantes para aplicações de decoração de quartos, reorganização de interior e mobiliário, jogos e sistemas de modelação e animação 3D.

Tendo em conta as motivações descritas o principal problema considerado neste artigo será como adquirir virtualmente um espaço designado pelo utilizador e usar o modelo detectado para adicionar informação virtual tri-dimensional de forma geometricamente correcta. Adicio-

nalmente deverá ser possível interagir com o modelo final em tempo real, respeitando as propriedades da cena representada na imagem. Mais do que construir um sistema de reconstrução 3D foto-realisticamente perfeito, pretende-se a construção de um modelo suficientemente robusto construído interactivamente por um utilizador não especializado.

A solução proposta sugere um modelo de interacção com o utilizador para adquirir uma imagem do cenário. Seguidamente são utilizados vários algoritmos para detectar elementos visuais nessa imagem. Por fim, uma aplicação de Realidade Aumentada é proposta.

## 2. TRABALHO RELACIONADO

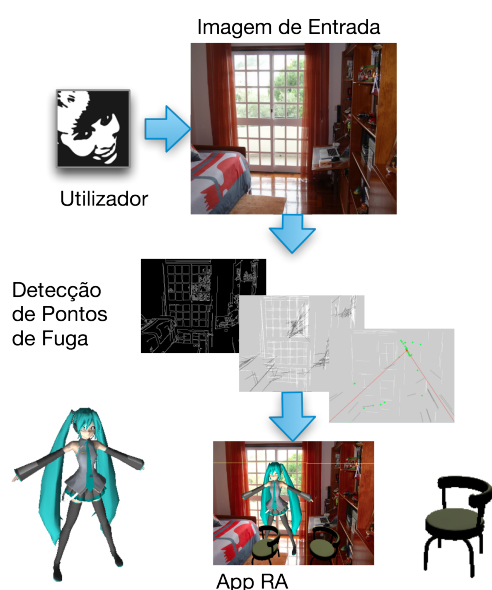
Existem vários tipos de aplicações de RA, sendo que os sistemas mais simples utilizam marcas físicas que servem de referência para a colocação de elementos virtuais em videos e imagens. Um destes sistemas é o ARtoolkitPlus<sup>1</sup> que utiliza marcas quadradas. Outros sistemas utilizam um cenário pré-definido e fixo como é o caso do jogo Tennis Real Play [Lai11]. A utilização de câmaras com sensores de profundidade pode facilitar a recuperação da estrutura da imagem como é exemplo o projecto KinectFusion [Izadi11]. Alguns projectos como o PTAM [Klein07] dependem de uma inicialização com um pequeno video em translacção do cenário. Depois de processado o video,

---

<sup>1</sup>ARtoolkitPlus, <http://handheldar.icg.tugraz.at/artoolkitplus.php>

é possível gerar uma superfície virtual onde a aplicação de RA decorre. A análise de uma imagem por forma a compreender a sua estrutura foi efectuada em vários projectos. No Make3D [Saxena09] a imagem é analisada com algoritmos de segmentação para gerar um cenário tri-dimensional da figura. Em [Lee09] é apresentado um sistema que analisa linhas da imagem para reconhecer o espaço virtual da cena. Por fim, [Karsch11] apresenta uma aplicação semi-automática de reconhecimento 3D da cena com introdução virtual de objectos. Os objectos possuem texturas e iluminação baseadas na fotografia de entrada.

### 3. SISTEMA DE REALIDADE AUMENTADA



**Figura 1: Sistema de RA, o utilizador tira uma fotografia que é analisada por forma a ser construído um modelo interno do mundo real. Os objectos virtuais são posteriormente introduzidos com a perspectiva correcta.**

O sistema de Realidade Aumentada proposto tem como objectivo a criação de aplicações que tenham por base o espaço físico do utilizador. O objectivo principal é utilizar o espaço físico do utilizador sem usar marcas no mundo real e utilizando apenas sensores disponíveis em vários equipamentos. Para tal deverá ser criado um modelo interno do mundo real por forma a que a aplicação de RA possa ser executada sob as imagens capturadas. Os principais sensores utilizados serão câmaras e acelerómetros. As câmaras utilizadas terão no mínimo as características presentes em qualquer câmara disponível actualmente em qualquer telefone móvel ou computador portátil. Não incluídas nestes sensores estão as câmaras com sensores de profundidade.

Tendo em conta as ideias por detrás deste projecto realizou-se uma análise dos principais requisitos do sistema proposto. A aquisição do cenário da aplicação de

Realidade Aumentada deverá poder ser feita por um utilizador que não tenha grande experiência em fotografia ou manipulação 3D. O utilizador deverá conseguir iniciar o sistema com um mínimo de instruções básicas e pouco ou nenhum treino. O sistema deve criar o modelo espacial subjacente à aplicação que guarde todos os elementos necessários à introdução de elementos virtuais. A criação do modelo deverá ocorrer quase em tempo real. A interacção com a aplicação que use o modelo deverá ser natural, os elementos virtuais deverão revelar propriedades próprias do mundo físico (Ex: gravidade, colisões com chão).

O conceito de interacção proposto pode ser dividido em três momentos: (1) Aquisição, (2) Processamento de Informação e (3) Aplicação Interactiva. Numa primeira parte (1) temos a aquisição de imagens e informação sobre o cenário onde queremos implantar a aplicação de RA. Aqui a interface do sistema desempenha um papel fundamental pois deverá instruir o utilizador a tirar fotos com algum significado e úteis para o sistema. Adicionalmente, o utilizador poderá ser chamado a localizar certos elementos visuais como o chão ou o tecto por forma a auxiliar os algoritmos de detecção. Estes algoritmos de processamento de imagem e de visão por computador entram em cena numa segunda fase (2). Pegando na informação adquirida o objectivo é detectar visualmente elementos como chão, horizonte, pontos de fuga, orientação da cena e elementos tridimensionais presentes. Utilizando todas as propriedades que for possível detectar poderão ser criadas várias aplicações interactivas (3) que criam a possibilidade de interagir com elementos virtuais num espaço real escolhido pelo utilizador.

Um exemplo do organigrama de interacção proposto pode ser observado na figura 1. Estas três fases serão revisitadas em maior detalhe nas próximas secções onde será apresentado o protótipo que implementa este sistema.

#### 3.1. Aquisição de Informação

A aquisição de imagens e da informação sobre o espaço onde a aplicação de Realidade Aumentada vai decorrer é um dos passos fundamentais na construção do sistema. Como já foi referido, a correcta detecção do modelo tridimensional do cenário depende de algoritmos de análise de imagem. Por este motivo é necessário que a fotografia tirada pelo utilizador tenha conteúdo suficiente para ser analisada. Isto exclui fotografias de paredes em branco, muito perto ou desfocadas. Idealmente as fotografias deverão conter uma grande quantidade de linhas rectas com bastantes superfícies lisas e texturas difusas tipo Lambert. Tipicamente este tipo de fotografias pode ser encontrado em edificações urbanas e em interiores de casas onde tipicamente as paredes, o chão, o tecto e a mobília contêm bastantes ângulos rectos. Para além de uma boa fotografia é necessário detectar a que altura do chão a imagem foi tirada. Outro problema com as câmaras é que possuem diferentes parâmetros como ângulo de visão, profundidade de campo e distorção radial de imagem. Adicionalmente é necessário ter em conta a escala do cenário e que tamanho terão que ter os objectos virtuais.

O utilizador pode inserir uma foto tirada com uma câmara fotográfica ou capturar directamente a imagem usando uma webcam. Se usada a webcam o sistema pode usar um sistema de auxílio ao utilizador que avisa quando é que a imagem tem detalhe suficiente para ser capturada. Este sistema auxiliar é baseado no número de linhas rectas detectadas (através da transformação de Hough) e através do número de pontos SURF detectados (ver implementação OpenCV<sup>2</sup>). Para obter mais dados sobre o sistema poderão ser sugeridas várias perguntas por forma a obter informações simples mas úteis sobre o espaço capturado. A perguntas ou tarefas sugeridas são:

– “Desenhe uma linha à volta do chão.” – O utilizador desenha um esboço de uma linha à volta da zona que considera o chão da cena. A localização do chão é depois efectuada com exactidão utilizando um algoritmo de segmentação como o *Graph Cuts*.

– “Desenhe uma linha à volta de objectos que estão em cima ou à frente do chão.” – Similar à anterior tarefa o objectivo aqui é inferir que tipo de objectos terão de ser representados em frente aos objectos virtuais ocludindo os mesmos.

– “A que altura foi retirada a fotografia? (1.70 m por omissão)” – Esta pergunta ajudará a inferir onde deverá ser colocado o plano virtual que representa o chão.

– “O horizonte foi correctamente detectado?” (dar duas ou três hipóteses) – O sistema de análise (explicado à frente) apresenta normalmente vários resultados com diferentes graus de probabilidade. Por vezes, nem sempre o horizonte com maior grau de probabilidade é o mais correcto, podendo assim o utilizador escolher outra possibilidade.

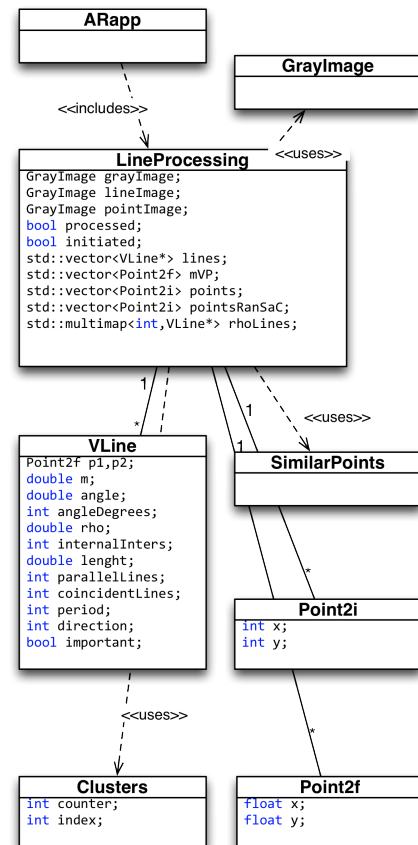
– “Desenhe um linha recta junto a uma aresta vertical de uma porta ou janela.” – Sabendo a altura média de uma porta ou de uma janela, poderá ser inferida alguma informação sobre a escala que deverão ter os objectos inseridos. Permite também inferir com maior correcção qual a direcção vertical da cena.

Toda a informação recolhida será incorporada no modelo do cenário da aplicação de RA. Na próxima secção é explicada sucintamente como é processada a imagem de entrada.

### 3.2. Processamento de Imagem

O protótipo actual analisa a imagem de entrada em busca de várias propriedades visuais como linhas principais, pontos de fuga principais e superfícies. A Figura 1 apresenta alguns dos passos intermédios necessários à detecção dos vários elementos. Todo o processo realizado é descrito em detalhe em [blind review]. De uma forma resumida o algoritmo analisa a imagem em três fases. Cada fase é repetida várias vezes com parâmetros diferentes para apanhar um máximo de situações possíveis (diferentes resoluções, diferentes iluminações, níveis diferentes de lixo visual).

Na Figura 2 podemos observar um esquema simplificado das várias estruturas de dados que suportam o sistema.

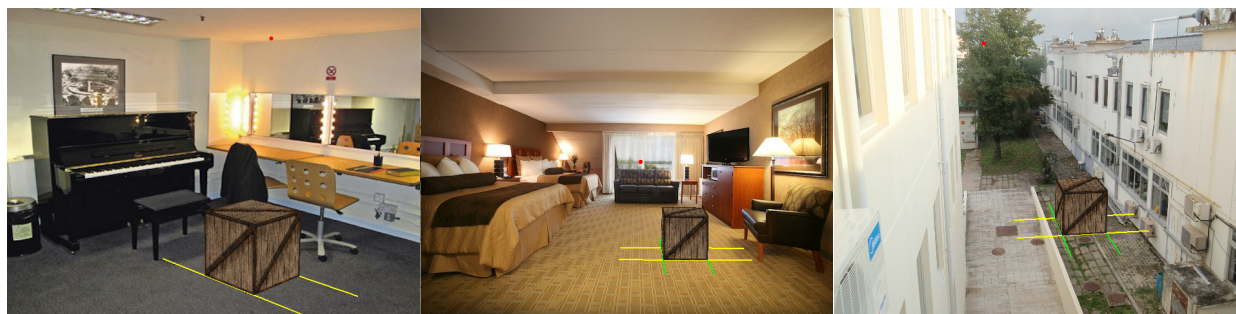


**Figura 2: Diagrama simplificado das classes que implementam o sistema de detecção de pontos de fuga. LineProcessing é a classe que centraliza e serve de repositório para toda a informação. A classe VLine representa cada linha que é detectada.**

Numa fase inicial os principais contornos da imagem são extraídos utilizando um filtro *Canny*. De seguida os principais segmentos de recta são extraídos usando a transformação de *Hough*. As linhas detectadas são guardadas em estruturas como a *VLine* presente na Figura 2. Um sistema de limpeza de lixo visual elimina as rectas que têm menos probabilidade de pertencer a estruturas e elementos sólidos da cena. Tendo em conta que são detectadas linhas na mesma imagem com parâmetros diferentes, existe uma grande quantidade de linhas que são coincidentes. Quanto mais linhas coincidentes, mais importante é uma linha. Quanto mais intersecções uma linha tiver, maior a probabilidade de essa linha ser gerada por ruído.

Usando as linhas consideradas importantes, é calculada a intersecção entre todas as linhas (ou uma amostra aleatória significativa no caso de serem muitas). Estes pontos são considerados pontos de fuga candidatos. Os vários pontos detectados são agrupados em grupos maiores. Usando algoritmos de *clustering* esse grupos de pontos são classificados em termos de probabilidade de representarem um

<sup>2</sup>OpenCV, <http://opencv.willowgarage.com/>



**Figura 3: Capturas de ecrã da aplicação de posicionamento interactivo de objectos virtuais. A caixa virtual encontra-se pousada no chão e automaticamente alinhada com o cenário onde está inserida.**

ponto de fuga válido.

Utilizando os pontos de fuga é possível encontrar o horizonte da imagem e o sitio para onde a maior parte dos elementos na fotografia estão orientados. Utilizando esta informação é definida uma rotação da cena no modelo tridimensional de forma a que quando representados os objectos virtuais tenham a mesma perspectiva que a cena original.

### 3.3. Aplicação Interactiva e Resultados

Por forma a validar os conceitos introduzidos nas secções anteriores, um protótipo foi desenvolvido que valida uma parte significativa das ideias propostas. A aplicação foi implementada recorrendo às bibliotecas OpenFrameworks<sup>3</sup> e OpenGL para visualização gráfica e interacção. Para suporte dos algoritmos de análise de imagem foi utilizada a já referida biblioteca OpenCV. A aplicação pode ser observada nas Figuras 1 (em baixo) e 3 e permite a introdução de objectos que podem ser movidos interactivamente. Os objectos colidem com o chão e estão inseridos na perspectiva da cena. Ao posicionar o objecto o utilizador sente que o objecto se alinha automaticamente com a fotografia aparecendo duas linhas guia verdes. Da mesma forma quando o utilizador pousa o objecto no chão duas linhas guia amarelas aparecem indicando que atingiu o plano virtual. A interacção é feita usando os dedos ou o rato.

Testes realizados e apresentados em [BFRP] mostram que o sistema de criação do modelo a partir de imagens funciona em 70 a 80% das situações consoante o conjunto de fotografias que se testa. Isto quer dizer que quando são colocados, os objectos encontram-se visual e geometricamente correctos. Também foi possível concluir que ao utilizarem este sistema os utilizadores colocaram 32% melhor os objectos a nível tri-dimensional.

## 4. DISCUSSÃO E CONCLUSÕES

O protótipo actual implementa um sistema que permite introduzir objectos 3D pré-definidos dentro de imagens fornecidas pelo utilizador [Nóbrega12]. É possível mover es-

tes objectos mantendo estes a perspectiva correcta e interagindo com o chão e com a orientação natural do mundo fotografado. As actuais limitações prendem-se com a pouca quantidade de elementos virtuais que podem ser actualmente inseridos e com o facto de o factor de escala não ser neste momento automático. No futuro pretende-se testar este sistema em sistemas móveis que permitam tirar fotografias em qualquer lugar e imediatamente interagir com os objectos virtuais.

## 5. AGRADECIMENTOS

This work was funded by FCT grants SFRH/BD/47511/2008 and PEst-OE/EEI/UI0527/2011.

## 6. REFERÊNCIAS

- [Izadi11] Shahram Izadi, Richard A. Newcombe, David Kim, Otmar Hilliges, David Molyneaux, Steve Hodges, Pushmeet Kohli, Jamie Shotton, Andrew J. Davison, and Andrew Fitzgibbon. KinectFusion: real-time dynamic 3D surface reconstruction and interaction. In *ACM SIGGRAPH 2011 Talks*, UIST '11, page 23. ACM, ACM, 2011.
- [Karsch11] Kevin Karsch, V. Hedau, and David Forsyth. Rendering synthetic objects into legacy photographs. *ACM Trans. Graph.*, 30(6):1–12, 2011.
- [Klein07] Georg Klein and David Murray. Parallel Tracking and Mapping for Small AR Workspaces. *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, 07:1–10, 2007.
- [Lai11] JH Lai, CL Chen, PC Wu, and CC Kao. Tennis real play: an interactive tennis game with models from real videos. In *Proceedings of the 19th ACM international conference on Multimedia - MM '11*, volume 3, pages 483–492, 2011.
- [Lee09] David C. Lee, Martial Hebert, and Takeo Kanade. Geometric reasoning for single image structure recovery. In *Computer Vision and Pattern Recognition (CVPR'09), 2009 IEEE Conference on*, volume 0, pages 2136–2143. IEEE Computer Society, 2009.
- [Nóbrega12] R Nóbrega and N Correia. Interactive Insertion of Virtual Objects in Photos and Videos. In *Eurographics 2012*, volume 0, pages 0–1. Eurographics Association, 2012.
- [Saxena09] Ashutosh Saxena, Min Sun, and Andrew Y. Ng. Make3D: Learning 3D Scene Structure from a Single Still Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):824–40, 2009.

<sup>3</sup>OpenFrameworks, <http://openframeworks.cc/>.