

# UMA ABORDAGEM PARA O TRATAMENTO DE PROBLEMAS DE OCLUSÃO EM REALIDADE AUMENTADA

Felipe G. Carvalho, Maurício Hofmam, Flávio Szenberg, Alberto Raposo, Marcelo Gattass

Grupo de Tecnologia em Computação Gráfica (Tecgraf), Departamento de Informática: PUC-Rio

CEP 22453-900, R. Marquês de São Vicente, 225, Gávea, Rio de Janeiro - RJ

Brasil

e-mails: {kamel, hofmam, szenberg, abraposo, mgattass}@tecgraf.puc-rio.br

**Resumo** – Este trabalho estuda alguns problemas de oclusão em Realidade Aumentada. São abordados, especificamente, dois problemas: oclusão do padrão de calibração por objetos reais e oclusão de objetos virtuais. A proposta apresentada combina uma técnica de identificação de padrões com outra utilizando objetos “escondidos” no buffer de profundidade para o tratamento destes problemas.

**Palavras-Chave** – Calibração de Câmera, Oclusão, Realidade Aumentada.

**Abstract** – This work studies some occlusion problems in Augmented Reality. More specifically, two problems are approached: fiducial occlusion by real objects and virtual objects occlusion. The presented proposal combines a technique to identify fiducials with a technique that uses “ghost” objects in the depth buffer to treat such problems.

**Keywords** – Augmented Reality, Camera Calibration, Occlusion.

## I. INTRODUÇÃO

A Realidade Aumentada (RA) oferece uma oportunidade para enriquecer informações do mundo real através da inclusão de elementos virtuais. Esta “inclusão”, muitas vezes, pode ser feita através de características que sejam consistentes com o mundo real, por exemplo, a exibição de sombras, perspectiva e oclusão. A oclusão juntamente com o registro são problemas clássicos de qualquer aplicação de RA. No registro estuda-se como obter as posições dos objetos reais em relação aos virtuais e na oclusão como serão ocultados os objetos virtuais em relação aos objetos reais.

A oclusão é uma característica que influencia muito na consistência da informação virtual adicionada no contexto da disposição espacial dos elementos reais no mundo, ou seja, é importante que haja uma sensação de coexistência entre o real e virtual, desta forma, a oclusão de objetos virtuais por objetos reais proporciona muito mais crédito ao complemento virtual pois adiciona aos elementos virtuais características reais daquele ambiente. Por outro lado, se for feita de maneira incorreta compromete o entendimento da cena.

O problema de registro neste trabalho está diretamente relacionado com aplicações que utilizam padrões bidimensionais (*fiducials* [1]) para resolver o problema de calibração de câmera. Este problema é resolvido

previamente, já que são conhecidas as posições e características dos objetos reais envolvidos em relação aos padrões. Esses padrões são reconhecidos analisando imagens capturadas [2] e normalmente esse reconhecimento é dependente da visualização total do padrão, ou seja, é necessário que todas as características visuais do padrão estejam na imagem capturada da câmera.

Neste contexto a oclusão por objetos reais deve levar em conta, como fatores importantes, as seguintes questões: a câmera visualizando a cena será estática ou dinâmica? Os objetos reais serão estáticos ou dinâmicos? Na questão da dinâmica do movimento da câmera, devem ser escolhidas técnicas de reconhecimento de padrão que levem em consideração mudanças no posicionamento da câmera e, dependendo do problema, até mudanças bruscas. Sobre os objetos reais estáticos, estes podem ser tratados normalmente através de um posicionamento de um objeto virtual “escondido” (também são conhecidos como “fantasmas”) idêntico no sistema de coordenadas de um padrão identificado. O termo “escondido” para qualificar o objeto virtual é utilizado assim pois o mesmo não é visto na imagem resultante mas sua informação de profundidade é utilizada para realizar um mascaramento para excluir as partes que estariam ocultas pelos objetos reais. Com objetos reais dinâmicos a questão é mais complexa, pois a posição e orientação dos mesmos podem estar mudando constantemente. Uma possibilidade seria associar padrões às faces dos objetos para que os mesmos fossem identificados independente de orientação com relação à câmera, mas surgem questões como: quantos padrões seriam necessários? O objeto é rígido, articulado ou deformável? A geometria do objeto é muito complexa?

O presente trabalho foca em aplicações de RA com câmera dinâmica, envolvendo oclusão com objetos reais estáticos. Como dito anteriormente, para este problema são utilizados objetos virtuais “escondidos” posicionados no mesmo lugar de objetos reais. No entanto, essa abordagem apresenta alguns problemas que surgem da forma como são identificados os padrões. A técnica de identificação normalmente utilizada (por exemplo, no ARToolkit [3]) necessita que o padrão esteja totalmente visível para ser identificado, ou seja, nenhum objeto real pode ficar em cima ou obstruindo a visualização do mesmo. Isto implica que os objetos reais que ficarão relacionados com um dado padrão  $P$  deverão ficar posicionados um pouco distante da origem do sistema de coordenada de  $P$ . Isso é ruim porque normalmente os algoritmos utilizados para determinar este sistema de coordenada (como por exemplo, o de Tsai [4]) apresentam erros de precisão à medida que a distância do centro de

calibração é aumentada. Outras discussões sobre a utilização de padrões bidimensionais são encontradas em [1].

Um trabalho utilizando objetos estáticos foi apresentado através de um jogo chamado *Invisible Train* [5]. Neste trabalho foi utilizada a biblioteca *AR-Toolkit* [3]. Porém a identificação de padrões ficou dependente da visualização completa dos mesmos, sendo também necessário a utilização de vários padrões para evitar que houvesse objetos muito distantes.

Para contornar alguns dos problemas citados anteriormente está sendo proposta a utilização de uma técnica desenvolvida em [6] que permite que objetos reais estejam obstruindo parcialmente as características do padrão de calibração. Logo, fica garantida mais precisão no casamento de um objeto virtual com um objeto real, uma vez que poderá ser posicionado um objeto em cima do padrão e em cima da origem do sistema de coordenadas do padrão. Esta abordagem foi inicialmente desenvolvida para um problema bem específico que consistia na identificação de campos de futebol sendo depois adaptada para padrões contendo várias linhas. No caso do problema com a identificação de um campo de futebol, esta técnica foi muito eficiente, pois identificava o campo mesmo havendo jogadores em cima de partes das linhas do campo.



Fig. 1. Identificação das linhas do campo mesmo com a obstrução de alguns jogadores.

Foram escolhidas para compor um cenário de experimentação com esta abordagem as seguintes características:

- a utilização de uma câmera inicialmente estática, que depois pode ser movimentada através de movimentos suaves;
- utilização de objetos reais estáticos;

## II. ALGORITMO

Para realizar a composição na cena dos objetos virtuais foi utilizada uma técnica apresentada em [6]. Ela consiste basicamente em discretizar a imagem capturada da câmera após um processo de segmentação na forma de uma grade bidimensional, onde cada célula desta grade é analisada independentemente, com o objetivo de extrair linhas. Em seguida, um processo de união de linhas é realizado (Figura 2).

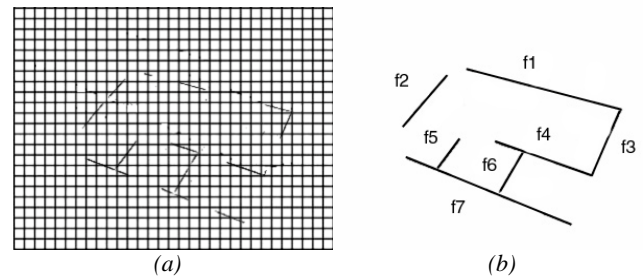


Fig. 2. Etapas iniciais do processo de identificação do padrão : (a) discretização da imagem na forma de grades para análise de cada célula independente (b) união de linhas.

Após este estágio inicial é construída uma árvore semântica para interpretar a disposição das linhas espacialmente[7]. O objetivo desta árvore é obter todas as interpretações topologicamente válidas para o conjunto de linhas detectadas (Figura 3). Dependendo da imagem capturada, é possível obter muitas interpretações válidas.

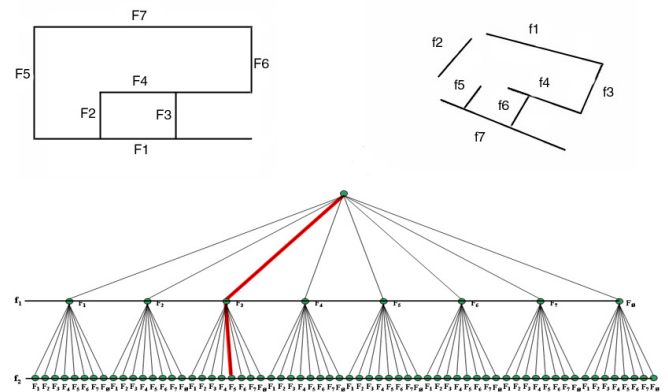


Fig. 3. Ilustração da Árvore de Interpretação utilizada para identificar o conjunto de interpretações válidas encontradas.

Em imagens capturadas em ambientes realísticos, onde muitas linhas são detectadas, é possível que a interpretação realizada pela árvore encontre várias combinações topologicamente válidas com correspondência completa entre linhas do modelo e linhas detectadas.

Assim, é necessário configurar o modelo de forma conveniente para a calibração da câmera e construir um conjunto de restrições que permita selecionar a interpretação correta das linhas detectadas.

### A. Configuração do modelo

O modelo deve ser configurado de forma a ter linhas suficientes para permitir que a câmera seja calibrada mesmo que parte delas não seja detectada, por estarem sendo parcialmente ou totalmente encobertas. Contudo, linhas em excesso tornam a interpretação lenta. Há que se balancear robustez com desempenho em tempo de execução, ou seja, utilizar um conjunto adequado de linhas que garanta uma precisão razoável sem implicar uma interpretação lenta usando a Árvore Semântica. O desempenho da aplicação pode ficar dependente do tamanho dessa árvore de interpretações, pois quanto mais linhas detectadas, além das pertencentes ao padrão, mais a árvore cresce para baixo aumentando o espaço de busca. Por outro lado, se o número de linhas utilizadas no padrão for muito grande, a árvore cresce para os lados, aumentando também o espaço de busca. Além disso, para restringir o número de interpretações topologicamente válidas, a configuração do modelo deve evitar simetrias. Neste trabalho foi utilizado um modelo com seis linhas (Figura 4), que permitiu obter um bom meio termo entre tolerância à oclusão e desempenho.

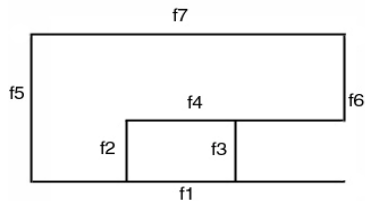


Fig. 4. Padrão utilizado nos experimentos contendo linhas conforme a restrição mínima para obter os pontos necessários para o processo de calibração da câmera.

### B. Conjunto de restrições

A primeira restrição foi rejeitar interpretações que não permitissem calibrar a câmera. Para que fosse possível fazer a calibração pelo método de Tsai [4], neste trabalho, foi adotado o mínimo de 6 pontos não colineares de correspondência entre a imagem e o modelo. Contudo, nem todos esses pontos precisam ser detectados na imagem. Se tivermos pontos suficientes para definir duas retas em uma direção e outras três em uma direção ortogonal, podemos calcular pontos de correspondência. Assim, a primeira restrição foi que a interpretação tivesse esse conjunto mínimo de 5 linhas.

A segunda restrição usada neste trabalho foi baseada no cálculo do Erro Quadrático Médio (EQM). Para todas as interpretações que atenderam a primeira restrição foram calculadas as homografias correspondentes (a homografia é calculada para verificar a exatidão da interpretação das linhas detectadas na imagem capturada). Com cada homografia, as coordenadas das interseções do modelo representadas no espaço do modelo foram transformadas para o espaço da imagem capturada. A partir daí foi calculado o EQM entre as coordenadas das interseções transformadas e das detectadas na imagem. Foi tomada como solução a interpretação com menor EQM.

Este conjunto de restrições garantiu desempenho satisfatório para a aplicação com até cerca de 40% de oclusão do modelo de calibração utilizado, dentro das condições do experimento (descrito a seguir).

## III. EXPERIMENTO

Para validar o uso dessa abordagem foi feito um experimento utilizando 8 objetos obstruindo partes do padrão de calibração. Desta forma foram criados 8 objetos virtuais “escondidos” para obter a informação de profundidade necessária para criar o efeito de oclusão quando fossem inseridos outros objetos virtuais que estariam visíveis.

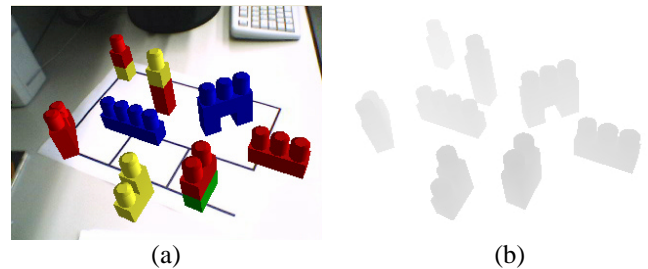


Fig. 5. Objetos virtuais “escondidos” desenhados no *color buffer* (a) e *depth buffer* (b).

O posicionamento dos objetos reais e seus respectivos correspondentes virtuais ficou visualmente satisfatório conforme a Figura 6. Para ilustrar melhor esse posicionamento foram desenhados no *color buffer* os objetos virtuais na forma *wireframe*.

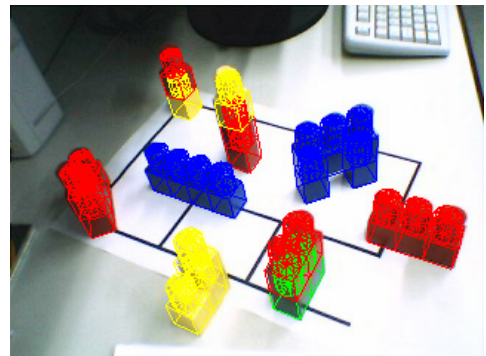


Fig. 6. Visualização do posicionamento dos objetos virtuais com os seus respectivos pares reais.

A Figura 7 mostra as partes do padrão que estavam obstruídas e, mesmo assim, o processo de identificação do mesmo permaneceu satisfatório, assim como o casamento dos objetos reais e virtuais. O importante foi garantir que indícios de linhas existissem indicando a topologia do padrão, pois depois através destes indícios, no processo de união de linhas, seriam criadas novas linhas que seriam usadas na árvore de interpretação.

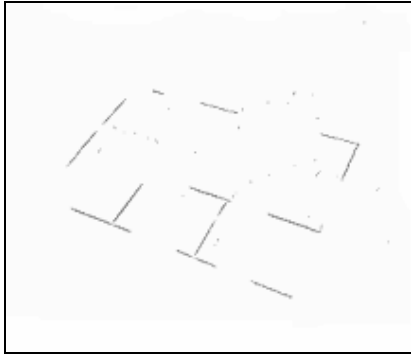


Fig. 7. Áreas do padrão obstruídas pelos objetos reais.

Na composição final (Figura 8), já exibindo o efeito de oclusão, foram incluídos 7 objetos virtuais em posições estratégicas, para um posicionamento e orientação de câmera, onde ficasse evidente a obstrução causada pelos objetos reais. Esta imagem foi obtida com o OpenGL desenhando-se os objetos reais apenas no mapa de profundidade de forma a garantir o funcionamento da oclusão dos fragmentos dos objetos virtuais que estão atrás dos reais.

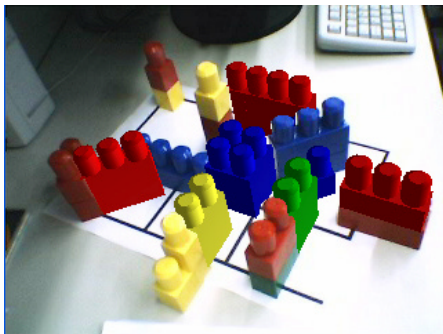


Fig. 8. Resultado do experimento compondo os objetos virtuais com os reais.

A aplicação teste disponibiliza várias visões informativas do experimento (Figura 9), como a visão final da composição, a visão do *buffer* de profundidade e as partes do padrão obstruídas. Um vídeo ilustrando o experimento é encontrado em [www.tecgraf.puc-rio.br/~hofmam/ra/ra.avi](http://www.tecgraf.puc-rio.br/~hofmam/ra/ra.avi).

#### IV. CONCLUSÕES

A oclusão é um problema clássico no campo da Realidade Aumentada e é tratado neste trabalho em uma abordagem utilizando duas soluções. A primeira solução busca a identificação de padrões que estejam com partes obstruídas. A segunda solução busca o tratamento da obstrução de objetos virtuais por objetos reais através da utilização da informação de profundidade de objetos virtuais no *buffer* de profundidade. Esta segunda solução é beneficiada pela primeira, que permite aumentar a precisão do algoritmo de calibração de câmera, uma vez que os objetos reais podem ser colocados mais próximos à origem do sistema de coordenadas determinado pelo marcador.

Foi realizado um primeiro experimento para validar a abordagem, mas ainda é necessário aprofundar mais o estudo

levando em conta questões sobre a complexidade×eficiência em relação a quantidade de linhas no padrão utilizado, assim como as questões sobre a porcentagem da oclusão permitida para a identificação satisfatória do padrão.

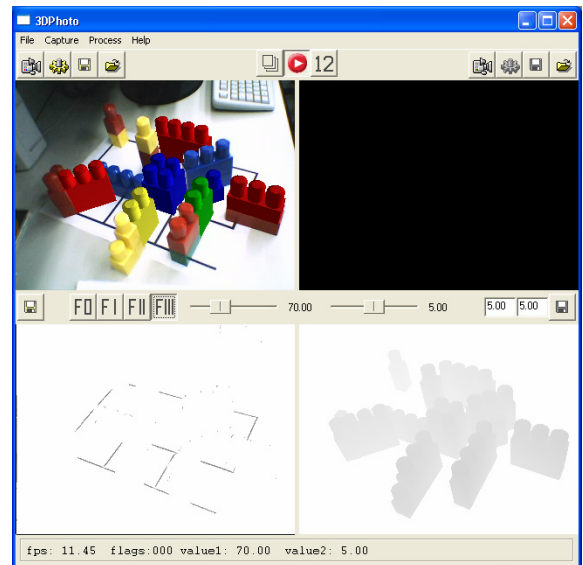


Fig. 9. Aplicação criada para realizar o experimento. Disponibiliza a imagem resultante da composição, as partes do padrão obstruídas (embaixo à esquerda) e a imagem do *depth buffer* mostrando os objetos virtuais “escondidos” (embaixo à direita).

#### REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Owen, C., Xiao, F., Middlin, P., *What is the best fiducial?. The First IEEE International Augmented Reality Toolkit Workshop*, pp. 98-105, 2002.
- [2] Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., Tachibana, K. *Virtual Object Manipulation on a Table-Top AR Environment. In proceedings of the International Symposium on Augmented Reality*, pp.111-119, 2000.
- [3] ARToolkit, <http://www.hitl.washington.edu/artoolkit/>
- [4] Tsai, Roger Y., *An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 364–374, 1986.
- [5] Wagner, D., Pintaric, T., Ledermann, F., Schmalstieg, D., *Towards Massively Multi-User Augmented Reality on Handheld Devices. Proc. of the Third International Conference on Pervasive Computing*, 2005.
- [6] Szenberg F. *Acompanhamento de Cenas com Calibração Automática de Câmeras. Tese apresentada ao Departamento de Informática da PUC-Rio como parte dos requisitos para a obtenção do título de Doutor em Ciências em Informática*, 2001.
- [7] Grimson, W.E.L., *Object Recognition by Computer: The Role of Geometric Constraints. Massachusetts Institute of Technology*, 1990.