

Um Estudo sobre Marcas Fiduciais em Realidade Aumentada: Combinando Detecção de Linhas com Calibração de Câmera

Maurício Hofmam, Felipe G. Carvalho, 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: {hofmam, kamel, szenberg, abraposo, mgattass}@tecgraf.puc-rio.br

***Abstract.** This paper presents a strategy to recognize fiducial marks in Augmented Reality, with the advantage of being able to handle some level of occlusion. This proposal for fiducial mark recognition combines a technique based on line detection, a topology identification technique, and a camera calibration method.*

***Resumo.** Este trabalho apresenta uma estratégia para reconhecimento de marcas fiduciais em Realidade Aumentada, com a vantagem de ser tolerante a níveis razoáveis de oclusão. A proposta apresentada para o reconhecimento de marcas fiduciais combina uma técnica baseada em detecção de linhas, informações topológicas e um método de calibração de câmeras.*

1. Introdução

Técnicas de visão computacional são comumente empregadas em Robótica assim como em Realidade Aumentada (RA), principalmente as relacionadas com o problema de rastreamento utilizando seqüências de vídeo. Este tipo de rastreamento consiste em, dada uma seqüência de vídeo, identificar a localização de um objeto ou da câmera quando a câmera ou o objeto está em movimento. Tal processo está basicamente dividido em duas etapas: uma inicial, para a extração de características da imagem capturada, seguida de uma segunda fase responsável pela calibração da câmera, ou seja, a estimação de seu posicionamento e orientação.

Dentre as abordagens para rastreamento baseadas em visão computacional, as que utilizam marcas fiduciais têm demonstrado resultados práticos bons e com tempos de execução aceitáveis para aplicações em tempo real [1]. Marcas fiduciais são imagens posicionadas no ambiente real que contêm características visuais fáceis de serem extraídas da imagem capturada. É importante que as características visuais dessas marcas estejam plenamente visíveis para o processo de rastreamento ser realizado com sucesso. Se, por exemplo, algum objeto estiver obstruindo parcial ou totalmente essas marcas, o processo de rastreamento pode ficar comprometido.

Entre outras abordagens para rastreamento estão aquelas que utilizam, por exemplo, rastreadores mecânicos, magnéticos, acústicos ou inerciais. Os mecânicos, apesar da boa precisão, são limitados a uma área de atuação pequena; os magnéticos também ficam limitados a uma área pequena e podem apresentar erros se um metal estiver no ambiente; os acústicos são suscetíveis a erros causados por ruídos sonoros; e os inerciais acumulam erros com o tempo [2].

Este trabalho apresenta uma abordagem de rastreamento óptico que utiliza, na fase de extração de características da imagem capturada, um processo de reconhecimento de marcas fiduciais baseado em informações topológicas e detecção de linhas [3] e, na fase de calibração de câmera, o método clássico de Tsai [4]. Tal abordagem apresenta o benefício da identificação de marcas que estejam parcialmente obstruídas, permitindo, por exemplo, o posicionamento de objetos reais em cima das marcas e favorecendo a utilização de recursos como a simulação de oclusão de objetos virtuais por objetos reais estáticos.

2. Marcas Fiduciais

Há dois grupos de marcas fiduciais: as pontuais e as planares. O primeiro tipo garante apenas um ponto de correspondência entre o ambiente real e a imagem capturada, enquanto nas marcas planares há mais pontos de correspondência — por exemplo, os cantos de uma marca com o formato de um quadrado. São necessárias várias marcas pontuais para determinar a calibração de uma câmera.

As características visuais mais relevantes nessas marcas estão relacionadas com as cores e os formatos utilizados. Os formatos mais comuns são formas geométricas como retângulos e círculos [5] [6], sendo que no caso das marcas planares é importante garantir que, através do formato escolhido, um mínimo de pontos não-colineares seja determinado para o processo de calibração da câmera. Com relação às cores existem as abordagens monocromáticas e as coloridas. As monocromáticas são normalmente escolhidas devido a algumas desvantagens das coloridas relacionadas à variação cromática de uma câmera para outra, e por também haver mais informação de luminância do que de cromaticidade nas imagens capturadas. No entanto, há trabalhos relevantes que empregam marcas coloridas ([6], por exemplo).

A escolha dessas características é uma etapa importante para facilitar os passos de localização e identificação das marcas na imagem capturada. A localização consiste na definição da posição espacial da marca na imagem, enquanto a identificação é responsável por analisar essa localização e identificar qual é a marca [7].

As marcas fiduciais pontuais são normalmente representadas por marcas circulares, por não haver muitas variações em relação à orientação e às distorções perspectivas. Normalmente as marcas são posicionadas de acordo com uma formação geométrica, e as suas posições no mundo são previamente conhecidas. Dado um conjunto de marcas deste tipo é então possível o procedimento de calibração da câmera [8].

As marcas planares permitem que o processamento de calibração seja feito com apenas uma marca. É necessária a visualização completa das características da marca, pois a geometria do formato da marca precisa estar consistente e sem muitas deformações. Caso seja colocado um dedo ou algum objeto em cima da marca, as etapas de localização e identificação podem ficar comprometidas. Se a marca estiver deformada, por exemplo, em uma folha de papel contorcida, poderá haver algum problema. Para solucionar a desvantagem da deformação, foram propostas marcas

fiduciais baseadas em topologia [9], uma vez que a abordagem topológica é independente de geometria.

Para o tratamento do problema de oclusão da marca, neste trabalho está sendo apresentada uma abordagem para a localização e identificação de marcas fiduciais utilizando detecção de linhas e informações topológicas.

3. Algoritmo

A técnica com utilização de detecção de linhas e topologia foi parte resultante do trabalho de Szenberg [3]. 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 da 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 (Figura 1).



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

A técnica consiste basicamente em uma discretização da imagem capturada da câmera após um processo de segmentação na forma de uma grade bidimensional, sendo cada célula analisada independentemente, com o objetivo de extrair linhas. Em seguida, um processo de união de linhas é realizado (Figura 2).

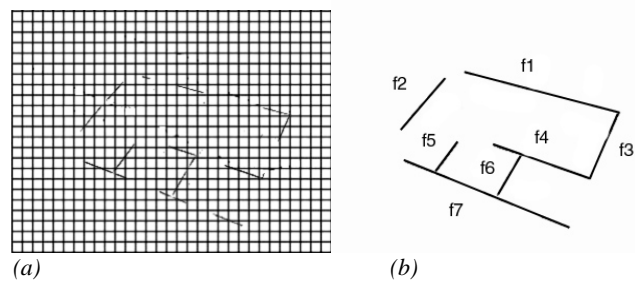


Fig. 2. Etapas iniciais do processo de reconhecimento da marca: (a) discretização da imagem na forma de uma grade para análise de cada célula independente; (b) união de linhas.

Após este estágio inicial é construída uma árvore semântica para interpretar a disposição das linhas [10]. O objetivo da árvore é obter todas as interpretações topologicamente válidas para o conjunto de linhas detectadas (Figura 3).

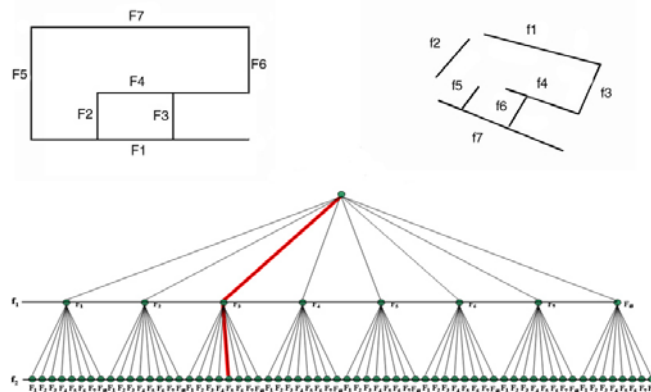


Fig. 3. Ilustração da árvore semântica utilizada para identificar o conjunto de interpretações válidas encontradas.

Em imagens capturadas em ambientes realistas, onde muitas linhas são detectadas, é possível que a interpretação realizada pela árvore encontre várias combinações topologicamente válidas, com correspondência completa entre as linhas da marca e as linhas detectadas. Assim, é necessário configurar a marca fiducial de forma conveniente para a calibração da câmera e construir um conjunto de restrições que permita selecionar a interpretação correta das linhas detectadas.

A. Design da Marca

A marca deve ser configurada 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 de interpretação. O desempenho da aplicação pode depender do tamanho dessa árvore, pois quanto mais linhas forem detectadas além das pertencentes à marca, mais a árvore cresce para baixo, aumentando o espaço de busca. Por outro lado, se o número de linhas utilizadas na

marca for muito alto, a árvore cresce para os lados, aumentando também o espaço de busca. Além disso, para restringir o número de interpretações topologicamente válidas, a configuração da marca deve evitar simetrias. Neste trabalho, para fins experimentais foi utilizada uma marca com sete linhas (Figura 4), o que permitiu obter um bom meio termo entre tolerância à oclusão e desempenho.

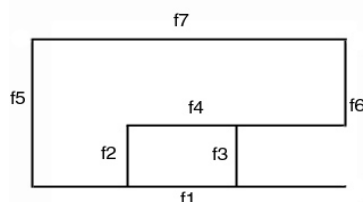


Fig. 4. Marca utilizada nos experimentos contendo linhas conforme a restrição mínima para obter os pontos necessários para o processo de calibração da câmera.

B. Detecção das Linhas

Para fazer a detecção de linhas é necessário primeiro realçar essas linhas. Isso é feito pela filtragem da imagem usando uma composição dos filtros Laplaciano e Gaussiano, também conhecida como filtro *LoG* [3]. O resultado dessa filtragem depende da espessura das linhas combinada com os parâmetros do filtro. Esses parâmetros têm de ser convenientemente ajustados de forma que as linhas sejam corretamente detectadas. A resolução da imagem e a distância da câmera à marca fiducial naturalmente influenciam nesses ajustes, pois implicam na captura de linhas mais espessas ou finas.

C. Conjunto de Restrições

A primeira restrição consistiu em rejeitar interpretações que não permitissem calibrar a câmera. Para que fosse possível realizar a calibração pelo método de Tsai [4], neste trabalho foi adotado o mínimo de 6 pontos não colineares de correspondência entre a imagem e a marca. 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, poderemos 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 à 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 aquelas das detectadas na imagem. Foi tomada como solução a interpretação com menor EQM.

D. Múltiplas Marcas

Há casos em que pode ser necessária a utilização de mais de uma marca fiducial. Marcas adicionais podem ser usadas para garantir que sempre haja pelo menos uma marca no campo visual da câmera para fazer sua calibração. Elas também podem ser empregadas para localizar e identificar objetos particulares em uma cena.

Quando são usadas várias marcas fiduciais, geralmente se deseja que elas possam ser distinguidas umas das outras. Caso sejam usadas para calibração, as coordenadas obtidas a partir de cada uma delas devem ser coerentes, ou seja, a câmera deve ser corretamente localizada a partir de qualquer uma delas e, portanto, devem ser inconfundíveis.

O processo de distinção de duas marcas fiduciais pode ser realizado em duas instâncias: topológica e de proporção geométrica.

A instância topológica é desempenhada pela árvore de interpretação, que, a partir de um conjunto de linhas detectadas, distingue subconjuntos que atendem à topologia definida pelo modelo fiducial. Vale frisar que uma topologia é identificada pela posição dos vértices de um segmento em relação aos demais segmentos da marca.

Se houver mais de um subconjunto topologicamente aceitável, a instância de proporção geométrica escolhe o que melhor corresponde à marca fiducial transformado pela homografia obtida a partir de cada subconjunto.

Para aumentar a probabilidade de que as marcas sejam corretamente distinguíveis, é conveniente que sejam topologicamente diferentes, como mostrado nas Figuras 5a e 5b. Note que as Figuras 6a e 6b também são topologicamente diferentes, mas podem levar a interpretações ambíguas pois a Figura 6b é um subconjunto da Figura 6a e, como abordado neste trabalho, as marcas podem estar parcialmente cobertas, podendo levar a erros na sua classificação.

As Figuras 7a e 7b são topologicamente iguais e podem ser classificadas corretamente na instância de proporção geométrica do processo, mas com maior suscetibilidade a erros.

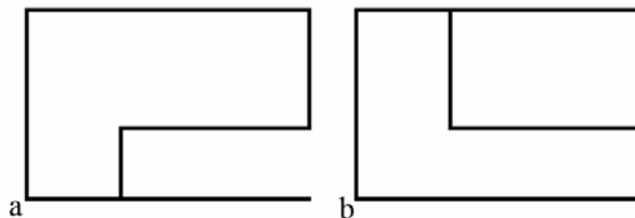


Fig. 5. Marcas fiduciais distinguíveis na instância topológica do processo de classificação.

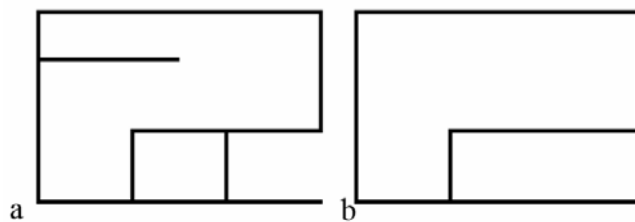


Fig. 6. Marcas fiduciais sujeitas a serem confundidas na instância topológica do processo de classificação.

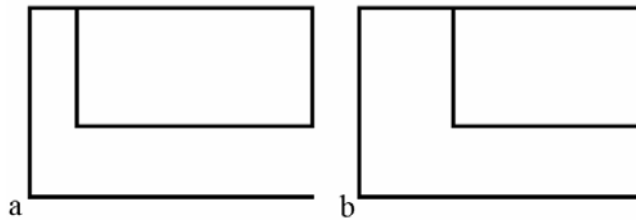


Fig. 7. Marcas fiduciais distinguíveis apenas na instância de proporção geométrica do processo de classificação.

E. Oclusão

A RA oferece a oportunidade de enriquecer informações do mundo real através da inclusão de elementos virtuais. Esta “inclusão” muitas vezes pode ser feita lançando mão de características que sejam consistentes com o mundo real — por exemplo, com a exibição de sombras, perspectiva e oclusão. A oclusão e o registro são problemas clássicos de qualquer aplicação de RA. Com respeito ao 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 a 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 no experimento realizado neste trabalho está diretamente relacionado com as marcas fiduciais utilizadas para resolver o problema de calibração de câmera. Este é resolvido previamente, já que são conhecidas as posições e características dos objetos reais envolvidos em relação às marcas.

Neste contexto, a oclusão por objetos reais deve levar em conta, como fatores importantes, as seguintes questões: a câmera que visualiza a cena será estática ou dinâmica? Os objetos reais serão estáticos ou dinâmicos? Quanto à questão da dinâmica do movimento da câmera, devem ser escolhidas técnicas de reconhecimento da marca que levem em consideração mudanças no posicionamento da câmera e, dependendo do problema, inclusive mudanças bruscas. No caso de objetos reais estáticos, estes podem ser tratados normalmente através do posicionamento de um objeto virtual “escondido” (também conhecido como “fantasma”) idêntico no sistema de coordenadas de uma marca identificada. O objeto virtual é denominado “escondido” pois não é visto na imagem resultante, mas sua informação de profundidade é utilizada para realizar um mascaramento de modo a 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 pode mudar constantemente. Uma possibilidade seria associar marcas às faces dos objetos para que eles fossem identificados independentemente da orientação com relação à câmera, mas surgem questões como: quantas marcas seriam necessárias? O objeto é rígido, articulado ou deformável? A geometria do objeto é muito complexa?

A utilização de objetos virtuais “escondidos” posicionados no mesmo lugar de objetos reais pode apresentar alguns problemas em função da forma como são identificadas as marcas. A técnica de identificação normalmente utilizada (por exemplo, pelo *ARToolkit* [1] [5]) necessita que a marca esteja totalmente visível para ser identificada, ou seja, nenhum objeto real pode ficar em cima ou obstruindo sua visualização. Isto implica que os objetos reais que ficarão relacionados com uma dada marca fiducial P deverão estar posicionados um pouco afastados. Isso é ruim porque normalmente os algoritmos utilizados para calibração da câmera — o de Tsai [4], por exemplo — apresentam erros de precisão à medida que a distância da marca fiducial, utilizada para a calibração, é aumentada [11]. Um trabalho utilizando objetos estáticos foi apresentado através de um jogo chamado *Invisible Train* [12]. Nele foi utilizada a biblioteca *ARToolkit*, porém a identificação das marcas depende de sua visualização completa, sendo também necessária a utilização de várias marcas para evitar que houvesse objetos reais muito distantes.

A abordagem utilizando detecção de linhas, empregada neste trabalho, permite que a marca fiducial esteja obstruída parcialmente e, portanto, que os objetos reais fiquem em cima da marca, proporcionando a simulação de obstrução de objetos virtuais por reais.

F. Desempenho

Como descrito na seção 3A, o desempenho do processo de classificação de marcas fiduciais depende da quantidade de linhas detectadas na cena e da quantidade de linhas da marca. Quanto mais linhas em uma ou outra, pior o desempenho, pois maior se torna a árvore de interpretação.

A Figura 8 relaciona os tempos médios consumidos na identificação correta de marcas fiduciais com 6, 7 e 8 linhas e as quantidades de linhas detectadas na cena. Chamamos a atenção para o fato de que esses valores podem flutuar em função da ordem em que as linhas da marca fiducial forem detectadas, pois isso irá influenciar o caminhamento da árvore de interpretação.

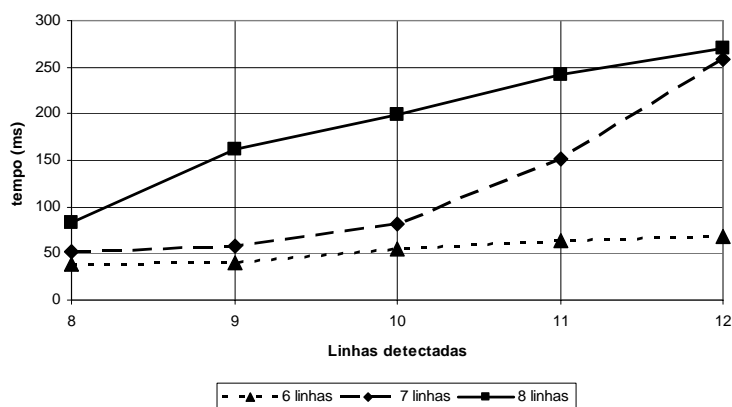


Fig. 8. Gráfico de desempenho do processo de identificação de uma marca fiducial.

Os tempos obtidos (ocasionalmente superiores a 250 ms) não comprometem o funcionamento do sistema em tempo real, pois lembramos que o processo de classificação é realizado apenas esporadicamente e não em todos os quadros capturados pela câmera. Feito o reconhecimento da marca em um quadro, ela é rastreada nos quadros seguintes e a calibração da câmera é ajustada de acordo com as novas posições

das linhas. A Figura 9 mostra o desempenho do processo na calibração de uma câmera ao longo de 2000 quadros capturados. Os picos do gráfico indicam instantes em que o acompanhamento das linhas foi perdido e as linhas detectadas tiveram de ser reinterpretadas.

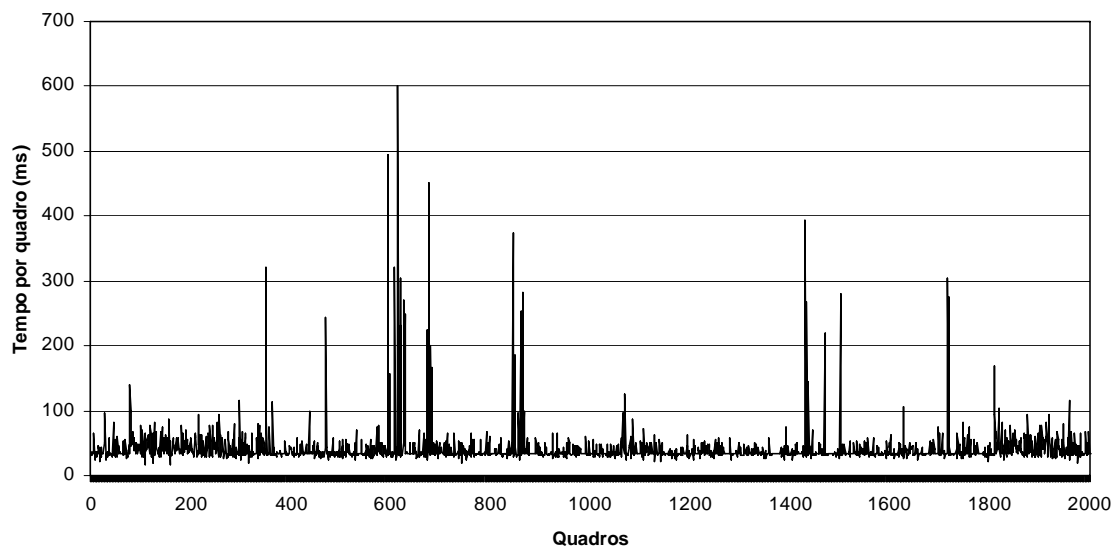


Fig. 9. Gráfico de desempenho do processo de calibração de uma câmera ao longo de 2000 quadros capturados.

4. Validação

Para validar o uso desta abordagem foi feito um experimento com 8 objetos obstruindo partes da marca fiducial utilizada para calibrar a câmera. 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.

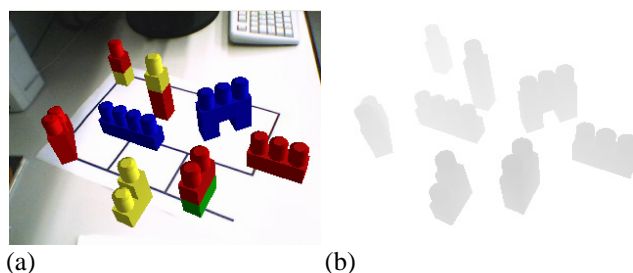


Fig. 10. Objetos virtuais “escondidos” desenhados no *color buffer* (a) e *depth buffer* (b).

O posicionamento dos objetos reais e de seus respectivos correspondentes virtuais ficou visualmente satisfatório, conforme mostra a Figura 11. Para ilustrar melhor esse posicionamento, os objetos virtuais foram desenhados no *color buffer* na forma *wireframe*.

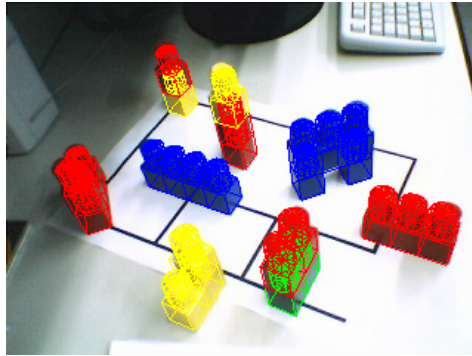


Fig. 11. Visualização do posicionamento dos objetos virtuais com os respectivos pares reais.

A Figura 12 mostra as partes da marca que estavam obstruídas e, mesmo assim, o processo de identificação da mesma permaneceu satisfatório, assim como a combinação dos objetos reais e virtuais. O importante foi garantir que existissem indícios de linhas indicando a topologia da marca, pois através destes indícios, no processo de união de linhas, seriam criadas novas linhas para serem usadas na árvore de interpretação.

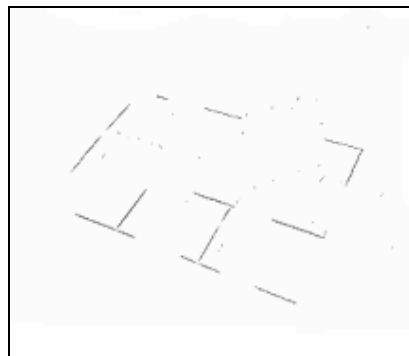


Fig. 12. Áreas da marca obstruídas pelos objetos reais.



Fig. 13. Resultado do experimento compondo os objetos virtuais com os reais.

Na composição final (Figura 13), já exibindo o efeito de oclusão, foram incluídos 7 objetos virtuais em posições estratégicas segundo o posicionamento e a orientação da 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 das partes dos objetos virtuais que estão atrás dos reais.

A aplicação de teste disponibiliza várias visões informativas do experimento (Figura 14), como a visão final da composição, a visão do *depth buffer* e as partes da marca obstruídas.

Durante os testes foi constatado que é possível calibrar corretamente a câmera mesmo com o encobrimento de 34% a 63% do somatório das extensões das linhas da marca fiducial. Esse nível de tolerância depende tanto de características da imagem capturada como de parâmetros do algoritmo de detecção das linhas. Quanto menos fragmentos de linhas espúrias forem detectados sobre a marca fiducial, maior a tolerância ao encobrimento. O ajuste do tamanho das células usadas na detecção das linhas e a razão mínima entre o maior e menor autovalor que caracterizam as direções principais de cada célula também influenciam o nível de tolerância.

Um vídeo ilustrando o experimento encontra-se em <http://www.tecgraf.puc-rio.br/~hofmam/ra/ra.avi>.

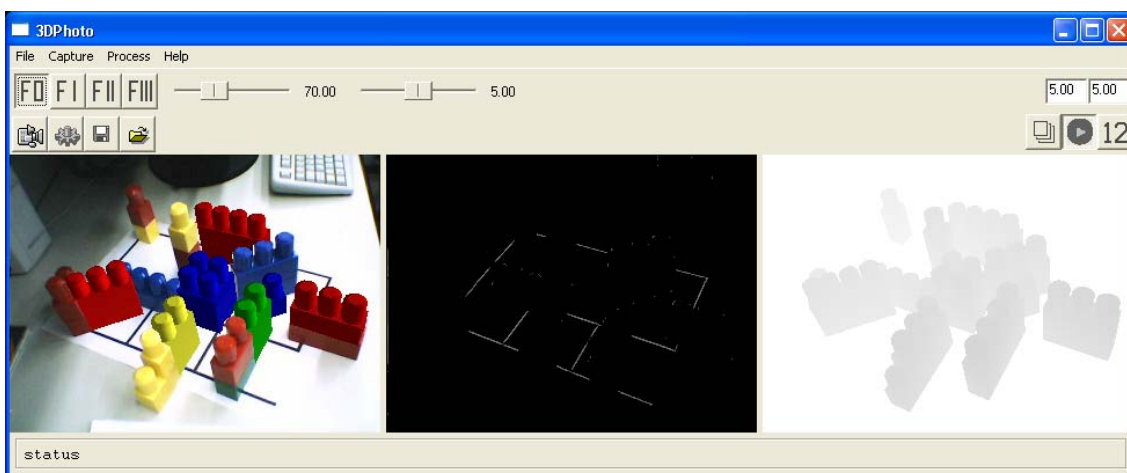


Fig. 14. Aplicação criada para realizar o experimento. Disponibiliza a imagem resultante da composição (à esquerda), as partes da marca obstruídas (no centro) e a imagem do *depth buffer* mostrando os objetos virtuais “escondidos” (à direita).

5. Conclusão

A utilização de marcas fiduciais para o problema de rastreamento em aplicações de RA é uma abordagem robusta e têm demonstrado resultados satisfatórios para aplicações em tempo real. Entretanto, normalmente é necessária a visualização completa destas marcas para a realização do procedimento de calibração de câmera. Este trabalho apresentou um experimento com rastreamento óptico baseado em marcas fiduciais utilizando recursos de detecção de linhas e informações topológicas associados ao método clássico de Tsai para calibração de câmera. Tal abordagem permite a localização e identificação de marcas que estejam parcialmente obstruídas, com resultados comparáveis aos sistemas que utilizam múltiplas marcas fidúcias, como o ARTag [13], assim como a simulação de efeitos de oclusão de objetos virtuais por reais. Estas vantagens permitem vislumbrar a utilização dessa abordagem em cenários

externos, com marcas fiduciais grandes, onde poderá haver vários elementos reais (objetos, pessoas, animais, etc.) obstruindo as marcas.

Foi realizado um primeiro experimento para validar a abordagem e demonstrar a sua viabilidade prática, tendo sido atingidos tempos de execução aceitáveis para aplicações em tempo real.

5. Referências

- [1] ARToolkit, <http://www.hitl.washington.edu/artoolkit/>
- [2] Bowman, D., Kruijff, E., Lavila, J., Poupyrev, I., *3D User Interface Theory and Practice*, Addison-Wesley, 2004.
- [3] 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.
- [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] Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., Tachibana, K., *Virtual Object Manipulation on a Table-Top AR Environment*, Proceedings of the International Symposium on Augmented Reality, pp.111-119, 2000.
- [6] Cho, Y., Lee, J., Numann, U., *A Multi-ring Color Fiducial System and An Intensity-invariant Detection Method for Scalable Fiducial-Tracking Augmented Reality*, IEEE International Workshop on Augmented Reality, 1998.
- [7] Owen, C., Xiao, F., Middlin, P., *What is the best fiducial?*, The First IEEE International Augmented Reality Toolkit Workshop, pp. 98-105, 2002.
- [8] Lepetit, V., Fua, P., *Monocular Model-Based 3D Tracking of Rigid Objects: A survey*, Foundations and Trends in Computer Graphics and Vision, Vol 1, No 1, 2005.
- [9] Costanza, E., Robinson, J., *A Region Adjacency Tree Approach to the Detection and Design of Fiducials*, Proceedings of Vision, Video and Graphics, Bath, UK, July 2003.
- [10] Grimson, W.E.L., *Object Recognition by Computer: The Role of Geometric Constraints.*, Massachusetts Institute of Technology, 1990.
- [11] Kato, H., Billinghurst, M., *Marker tracking and HMD calibration for a video-based augmented reality*, Conferencing System, IEEE and ACM International Workshop on Augmented Reality, pp.85-94, 1999.
- [12] Wagner, D., Pintaric, T., Ledermann, F., Schmalstieg, D., *Towards Massively Multi-User Augmented Reality on Handheld Devices*, Proceedings of the Third International Conference on Pervasive Computing, 2005.
- [13] ARTag, <http://www.cv.iit.nrc.ca/research/ar/artag/>.