

## Metodologia de geração de dados de referência para rastreamentos ópticos

Lucas Teixeira, Manuel Loiaza, Alberto Raposo, Marcelo Gattass  
*Tecgraf, Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro*  
*{lucas,manuel,abraposo,mgattass}@tecgraf.puc-rio.br*

### Resumo

*Neste trabalho apresentamos uma metodologia para geração de dados de referência para algoritmos de rastreamento óptico. Nós definimos um tipo de marcador especial para tal tarefa, propomos duas maneiras de fazer a medição da posição 3D dos marcadores e apresentamos um algoritmo para localizar os marcadores na imagem semi-automaticamente. Nós resultados apresentamos uma comparação entre as duas maneiras de realizar a medição dos pontos 3D no cenário real e um caso de uso usando um marcador ARToolKitPlus.*

### 1. Introdução

Na área de realidade virtual e aumentada uma ferramenta fundamental tem sido os sistemas de rastreamento óptico. Este tipo de sistema tem o objetivo de processar, identificar e rastrear constantemente um conjunto de características, naturais ou artificiais, encontradas dentro de um cenário específico. É justamente baseado nessas características que a correspondência e o mapeamento 3D da posição de diversos objetos no cenário são realizados. Esse mapeamento e a sua precisão são as chaves mestras para poder adicionar os elementos virtuais no cenário real com maior qualidade visual e coerência espacial. O resultado esperado é que esses objetos virtuais sejam confundidos com os reais, enquanto o usuário navega pelo cenário e o vê através das imagens capturadas e posteriormente processadas de uma câmera de vídeo.

No entanto, garantir a corretude de um sistema de rastreamento óptico é bastante difícil, principalmente quando a área de rastreamento é muito longa. Pensando nisso alguns métodos baseados em padrões fiduciais têm sido propostos. Como exemplo deste tipo de método podemos citar a reconhecida biblioteca de realidade aumentada “ARToolkit” ou clássicos

padrões em formato de xadrez, os quais possuem formatos que permitem a sua identificação de uma forma rápida e com altas taxas de acerto em uma área de rastreamento. Porém, esses tipos de marcações e padrões não possuem pontos de referência suficientes para mensurar a qualidade de um rastreador projetado para funcionar em ambientes extensos. Por este motivo, muitas vezes, é preciso espalhar no ambiente um número grande destes padrões para poder manter um rastreamento preciso dentro do cenário.

Nossa proposta para mensurar a qualidade do rastreamento é desenvolver um método composto por duas etapas. Uma primeira etapa consiste em medir alguns pontos manualmente na área de rastreamento, usando para isto uma estação total, equipamento muito utilizado na engenharia civil para medir com precisão a posição 3D de pontos específicos num cenário real. A segunda etapa consiste em capturar o mesmo ponto no espaço para o qual fizemos a medição com a estação total e, usando os sistemas de rastreamento óptico, reprojeter a posição 3D do ponto na imagem da câmera usando os parâmetros de calibração fornecidos pelo sistema de rastreamento. O erro que desejamos medir para usar como base de comparação é a distância em pixels do ponto correto medido visualmente na imagem com o ponto reprojeterado pelo sistema de rastreamento. Esse erro pode ser usado como uma métrica para mensurar a qualidade de um sistema de rastreamento óptico.

No entanto, esta técnica apresenta o problema de ter que realizar de forma manual a marcação dos pontos de referência na imagem para depois compará-los com os pontos projetados pela calibração da câmera fornecida pelo sistema de rastreamento. Supondo que um vídeo de teste com duração de 10 segundos tenha sido capturado a uma taxa de 30 frames por segundo, se existirem 10 pontos visíveis na imagem seria necessário marcar cerca de 3000 pontos manualmente. Além do trabalho de marcação, a possibilidade de uma marcação errada é diretamente proporcional ao número

de pontos a serem marcados. Por esse motivo, na metodologia proposta neste trabalho, também é implementada uma técnica para detectar pontos na cena de forma semi-automática. Para isso propomos um novo tipo de marcador e um algoritmo capaz de acompanhá-lo durante a navegação na cena. Após o marcador ter sido detectado uma primeira vez em um quadro específico, o algoritmo proposto o rastreará durante os próximos quadros nos quais ele esteja presente, só precisando ser reiniciado quando ele sair e voltar ao campo de visão da câmera que o esteja rastreando.

Este artigo está organizado da seguinte forma. A seção 2 apresenta trabalhos relacionados. A seção 3 descreve o método proposto. A seção 4 apresenta os resultados. A seção 5 apresenta algumas conclusões.

## 2. Trabalhos Relacionados

A verificação de um sistema de rastreamento óptico é normalmente realizada de forma muito limitada. Para isso é necessário uma outra forma de medir a posição da câmera ao longo do tempo. Atualmente, as únicas formas de se fazer isso é usar um sistema mecânico como o da FARO mostrado na figura 1 ou usando um sistema óptico como o da VICON mostrado na figura 2, que garantem erro de medição de décimos de milímetro a uma distância maior de 50cm entre a câmera e os pontos medidos. Usar marcadores tipo ARToolKit é inviável pela sua limitada precisão e por ocuparem uma área muito grande da cena.

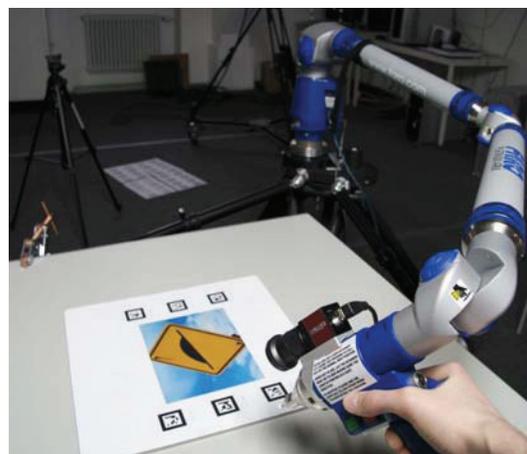
Ambos os sistemas são precisos, mas têm áreas de utilização muito pequenas. O braço extensível da FARO tem um alcance de 1.8 a 3.7 metros dependendo do modelo e o sistema VICON tem uma área útil, em alta precisão, de cerca de 0.5 a 2 metros quadrados.

Outra forma mais genérica para verificar se um algoritmo de rastreamento está correto é usar uma base de dados onde a posição da câmera tenha sido previamente calculada. Lieberknecht et al [1] apresentam uma base de dados com a posição da câmera já calculada mecanicamente. No entanto, esta base de dados só serve para algoritmos de rastreamento baseados em texturas, pois suas imagens são passeios de câmera olhando para uma textura planar.

Outra base de dados muito utilizada é a Middlebury [2]. Além de usar imagens sintéticas, eles conseguiram criar pontos de referência bastante densos usando uma textura fluorescente escondida atrás dos objetos da cena. Esse mesmo grupo fez uma segunda base de dados para reconstrução baseada em imagens estéreis usando um scanner a laser. No entanto, todas as formas de base de dados conhecidas nunca são suficientes para

avaliar qualquer tipo de rastreamento de câmera, pois a cena ou os tipos de marcadores filmados não cobrem todos os tipos de rastreamento, mesmo porque constantemente estão surgindo novas técnicas para modelar tanto padrões [3] como algoritmos de visão computacional criados para detectar esses novos padrões [4].

Recentemente, Zimmerman et al [6] apresentaram um banco de dados com três seqüências de imagens e seus correspondentes movimentos de câmera. Todos os dados gerados foram feitos de forma manual pelos autores usando alguns marcadores em forma de cruz. Esses marcadores foram colados em torno de texturas ou em objetos da cena que foram marcados manualmente na imagem. Essas seqüências têm cerca de 12000 imagens no total. Essa estratégia é a mesma utilizada na nova metodologia proposta. A diferença é que nós fornecemos uma forma de marcar os pontos de referência na imagem de forma rápida e com a possibilidade de erro na identificação reduzida. Essa redução é decorrente do fato do usuário marcar os pontos em apenas poucos quadros.



**Figura 1:** Sistema mecânico de geração de dados de referência da FARO. Fonte: [1]

## 3. Ambiente de teste

Nessa sessão será descrito como montar o ambiente de teste utilizado neste artigo. Os requisitos para nosso método de teste são: um tipo especial de marcador, iluminação ambiente sem mudanças bruscas e suficientemente constante, que permita encontrar nossos marcadores na imagem com alto grau de correteude. Além disso, é preciso medir a posição dos marcadores no ambiente. Então será descrito um método de medição simples e um outro método de alta precisão. Posteriormente será explicado como organizar o processo de testes.



**Figura 2:** Sistema óptico da VICON para geração de dados de referência. Fonte: VICON

### 3.1. Marcadores

Nossos marcadores são esferas de isopor brancas coladas em um pedaço de cartolina preta, como mostrado na figura 3.

O centro da esfera branca é o nosso ponto de referência, dessa forma é possível calcular esse ponto de qualquer posição e perspectiva da câmera em que a esfera possa ser visualizada. O contraste entre a cor do marcador branco e a base de cartolina preta ajuda a reconhecer e ressaltar quadro a quadro os nossos marcadores dentro da cena, sendo facilmente distinguidos na imagem. A captura do centro da esfera como ponto de referência se baseia na seguinte característica: a esfera vista na imagem a partir de qualquer perspectiva vira um círculo cujo centro é sempre um ponto que se somado ao raio da esfera resulta no centro da esfera.

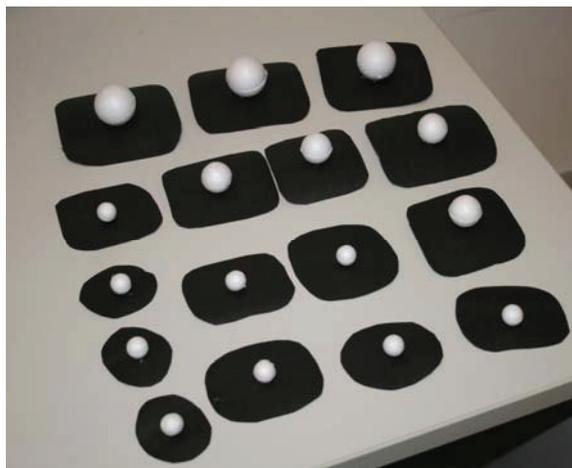
O pedaço de cartolina preta que é colado na esfera também permite uma segmentação mais correta da esfera, pois na maioria dos ambientes as paredes ou os móveis são de cor clara podendo atrapalhar o reconhecimento de nossos marcadores.

Os tamanhos da esfera e da cartolina dependem da distância em que a câmera está dos marcadores. De forma geral, espera-se que a esfera cubra pelo menos 4x4 pixels na imagem e que a cartolina preta seja suficiente para olhar a esfera lateralmente e ainda assim formar uma borda preta ao redor da esfera.

### 3.2. Medição dos marcadores

A medição dos marcadores em geral é uma tarefa muito sensível a erro. Abaixo vamos descrever duas formas de medição. A primeira é simples e barata de se fazer, pois usa equipamentos comuns como réguas e

nível. No entanto tem uma imprecisão maior e não é aplicável a cenas em que os marcadores não sejam coplanares. A segunda é altamente precisa e mais geral, usando um equipamento de topografia que mede pontos 3D. Em compensação esse segundo método é caro por usar um equipamento especializado e de difícil acesso.



**Figura 3:** Marcadores utilizados para esse trabalho.



**Figura 4:** Modo de medição simples dos marcadores simples, à esquerda e a estação total à direita.

#### 3.2.1 Método Simples

O método simples usa equipamentos de marcenaria e construção. Uma régua de 1 metro, uma trena, um nível e fita crepe, como mostrado na figura 4. Com esses instrumentos é possível espalhar marcadores ao longo de paredes consecutivas e ter uma precisão razoável na medição dos marcadores.

O método se baseia em colocar o eixo Y do sistema de coordenada paralelo ao eixo gravitacional para possibilitar encontrar o eixo X usando o nível. Então começamos marcando um ponto na parede que será a origem do sistema de coordenada. Em seguida, marcamos pontos na mesma altura(Y) que a origem do sistema de meio em meio metro em X usando a régua e o nível. Depois unimos esses pontos com fita crepe. Essa fita será o eixo X. Usando a régua e o nível, medimos a distância do marcador até a fita, que é a coordenada Y do marcador. Por fim medimos com a trena a distância do ponto de intersecção da régua com a fita até o ponto de origem, essa distância é a coordenada X do marcador.

Para continuar a medição em paredes consecutivas, elas precisam ser perpendiculares, assim é só trocar a coordenada X por Z. ou vice-versa, dependendo de qual coordenada estava variando na parede anterior.

### 3.2.2 Método de Alta Precisão

O método de alta precisão usa um equipamento chamado estação total (figura 4). Esse instrumento foi desenvolvido para área de topografia. Ele pode ser considerado um avanço do teodolito digital. O teodolito digital é um equipamento que é capaz de medir a rotação no eixo gravitacional e no eixo perpendicular a ele. A estação total é esse mesmo equipamento somado a um medidor de distâncias a laser. Com essas duas rotações e a distância até o ponto de interesse ele é capaz de calcular a coordenada 3D de um ponto num sistema de coordenada previamente informado.

Para fazer a medição basta apontar o equipamento para o centro da esfera que ele informará a posição 3D do ponto. Subtraindo essa posição da posição da estação total, que também é fornecida pelo equipamento, temos o vetor de intersecção na esfera. Então para encontrar o centro da esfera basta adicionar o valor do raio da esfera na direção desse vetor.

### 3.3. Processo de teste

O processo de teste é um resumo do passo a passo para usar nossa metodologia de testes. Os passos não precisam ser executados exatamente na ordem que são descritos. No entanto, a ordem que propomos é mais eficiente segundo nossos casos de teste.

O primeiro passo é montar a cena de teste e verificar se o algoritmo de rastreamento em análise funciona nessa cena. O segundo passo é escolher o tamanho das esferas para os marcadores e posicioná-los na cena de teste. O terceiro passo é gravar o vídeo de teste que será usado tanto pelo algoritmo de

rastreamento em análise quanto pelo nosso algoritmo de acompanhamento de marcadores. O quarto passo é testar se o algoritmo de rastreamento funciona nesse vídeo gravado, se as esferas têm um tamanho de pelo menos 4x4 pixels na imagem e se há um mínimo de esferas por quadro do vídeo, caso algum desses testes falhe volte para o segundo passo.

O sexto passo é usar o aplicativo que implementa o nosso algoritmo de acompanhamento de marcadores para marcar os pontos de entrada e saída. Esses pontos serão mais bem explicados na próxima sessão. Com esses pontos e o vídeo, o algoritmo é capaz de gerar uma lista de marcadores para cada quadro do vídeo contendo a posição na imagem de cada um dos marcadores visíveis. O sétimo passo é fazer a medição da posição 3D dos marcadores. Essa é, sem dúvida, a tarefa mais demorada de se fazer, então só deve ser realizada quando todas as outras tiverem ocorrido com sucesso e não seja necessária nenhuma modificação na posição dos marcadores.

## 4. Algoritmo de Acompanhamento dos Marcadores

O algoritmo de acompanhamento dos marcadores é o responsável por descobrir a posição do marcador no frame corrente a partir da posição do marcador no frame anterior. O algoritmo proposto usa um algoritmo de fluxo óptico para fazer isso. Além disso, usamos uma estratégia para segmentar os marcadores e extrair o centro do círculo que representa a esfera na imagem.

### 4.1 Visão Geral

Será descrito, primeiramente, o processo geral do algoritmo e depois daremos detalhes sobre cada um dos sub-algoritmos.

Para o nosso algoritmo, a primeira etapa é o usuário marcar no vídeo os pontos de entrada e saída dos marcadores. Um ponto de entrada é a posição em que um marcador está no quadro corrente. É necessário marcar um ponto de entrada toda vez que no quadro anterior o marcador em questão não esteja visível. Um ponto de entrada é composto pela coordenada X e Y do marcador no quadro K e pelo ID do marcador ( $Pts_{entrada}(K)=\{X,Y,K,ID\}$ ). Já um ponto de saída é o quadro em que um marcador desapareceu, sendo que no quadro anterior ele estava presente ( $Pts_{saida}(K)=\{K,ID\}$ ).

Todavia, é muito difícil marcar todos os pontos de entrada e saída do vídeo olhando imagem por imagem. Então para isso o processo de marcação interage com o processo de acompanhamento dos marcadores, que

será explicado abaixo. Basicamente, o usuário marca os pontos de entrada no quadro 1 e aciona o acompanhador dos marcadores que vai atualizando a posição dos marcadores ao mesmo tempo em que avança o vídeo. Quando o usuário chegar a um quadro em que entrou um novo marcador, ele pára o avanço do vídeo e marca o ponto na imagem. Em seguida liga o avanço do vídeo novamente. O mesmo é feito para os pontos de saída. Quando o usuário chegar ao final do vídeo todos os pontos de entrada e saída vão estar marcados. Após esse procedimento basta executar o acompanhamento do vídeo desde o quadro 1 até o final salvando a lista de pontos (posições dos marcadores na imagem) a cada quadro, como mostrado na figura 5.

O processo de acompanhamento inicia com a lista de pontos vazia no quadro  $i=1$ . Então ele entra num ciclo que o primeiro passo é adicionar os pontos de entrada que foram marcados no quadro( $i$ ). Caso já exista um ponto para o marcador daquele ID, ele é substituído. Já os pontos de saída são usados para eliminar da lista de pontos aqueles com os mesmos IDs.

Em seguida, cada ponto da lista tem suas coordenadas corrigidas usando um algoritmo para encontrar o centro de um círculo. Abaixo este algoritmo será explicado com mais detalhes.

Após a correção, os pontos são salvos para o usuário usar como pontos de referência. Se o quadro atual não for o último, ele calcula o fluxo óptico, que informa a posição dos pontos do quadro atual no próximo quadro. Com essas novas posições os pontos retornam ao início do ciclo.

#### 4.2 Correção dos Pontos

A correção dos pontos é feita, primeiramente, binarizando a imagem. Todos os pixels na imagem em tons de cinza abaixo de um limiar viram pretos e acima brancos. Depois identificamos na imagem todos os grupos de pixels unidos com mais de 4 pixels e menos de 100 pixels. Cada grupo recebe o nome de Blob. Então procuramos o blob mais perto de cada marcador e com no máximo 8 pixels de distância do ponto atual. Caso exista esse blob, calculamos qual o maior círculo que cabe dentro desse blob e o centro desse círculo passa a ser o ponto corrigido.

#### 4.3 Fluxo Óptico

Fluxo óptico é uma técnica que calcula para onde cada pixel da tela se deslocou entre dois quadros do vídeo. No nosso caso estamos interessados apenas nos pontos da lista. Sendo assim o gasto computacional é

bastante inferior. Nós usamos o fluxo óptico fornecido pelo OpenCV, que implementa uma versão piramidal do rastreador de pontos Lucas Kanade[5].

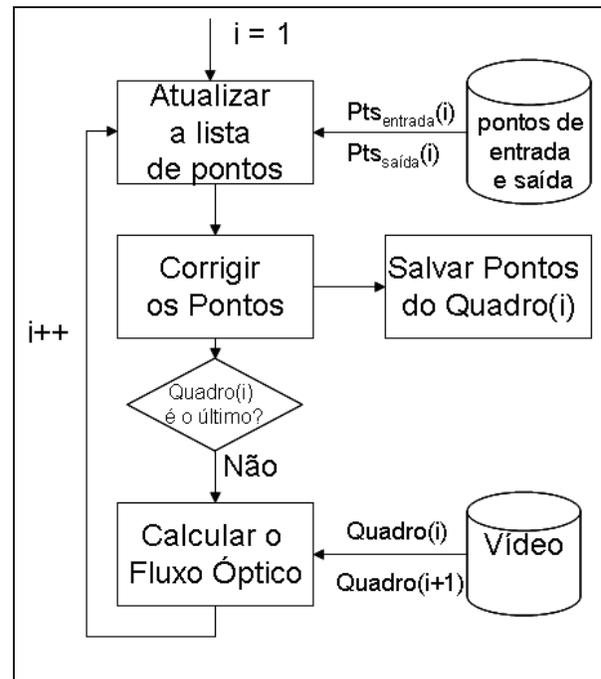


Figura 5: processo de acompanhamento

### 5. Resultados

Para validar nossa metodologia vamos primeiramente comparar o modo de medição simples com o de alta precisão para analisar a faixa de erro na medição no modo simples. Em seguida vamos analisar o número de marcações que o usuário precisa fazer para conseguir o rastreamento em uma cena típica. Por fim vamos fazer um estudo de caso com um marcador do ARToolKitPlus para analisar o nível de erro de reprojeção desse toolkit.

Nossa implementação do algoritmo de acompanhamento dos pontos foi desenvolvido usando a biblioteca OpenCV. Nossa câmera é uma unibrain Fire-I, na resolução de 320x240 a 30fps. Nossa Estação total é uma Sokkia 530RK3 e tem um erro menor que 5 mm.

#### 5.1 Diferença entre os modos de medição

A nossa avaliação sobre os modos de medição foi feito dispondo 8 marcadores ao longo de uma parede. Então usamos os dois métodos de medição para descobrir a posição dos marcadores na parede. Na tabela 1 podemos ver os resultados.

Como podemos ver na tabela, o erro médio foi de 2.22 cm. Sendo assim, quando for usar o método

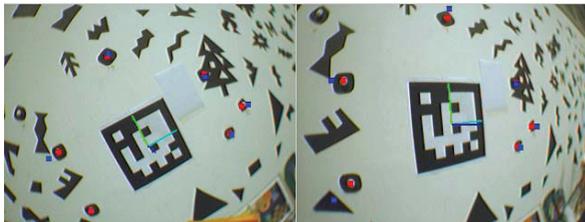
simples de medição é importante usar distâncias maiores até a parede, onde esse erro não é significativo. Por exemplo, se tivesse um marcador de ARToolKit de 5 centímetro próximo a uma esfera branca que foi medida com um erro de 2.2 cm. Esse erro seria inaceitável.

id	Xs	Xp	$\Delta X$	Ys	Yp	$\Delta Y$	erro
1	38	37.5	0.5	10	10.1	-0.1	0.51
2	14.5	16.5	-2	-6.1	-6.5	0.4	2.04
3	51.4	51.4	0	-40.5	-43.5	3	3.00
4	56	55.7	0.3	45.5	45.4	0.1	0.32
5	64.7	64.6	0.1	-10.5	-13.2	2.7	2.70
6	82	82.4	-0.4	18.3	18.7	-0.4	0.57
7	99.5	99.5	0	-27.5	-29.9	2.4	2.40
8	115.1	112.7	2.4	47	47	0	2.40
						média	2.22

**Tabela 1 :** Valores em cm. O par Xs e Ys são o ponto medido com o modo de medição simples e o par Xp e Yp no modo de alta precisão.  $\Delta X$  e  $\Delta Y$  são as variações nas medidas em cada um dos eixos. Erro é a distancia euclidiana entre os pontos.

### 5.2 Caso de uso

Nós usamos um marcador do ARToolKitPlus para ilustrar o funcionamento da nossa metodologia. Na figura 6 podemos ver o quadro 336 à esquerda e o quadro 641 à direita. Na figura 7, o gráfico do erro de reprojeção para o vídeo de teste. Nesse teste foi necessário marcar apenas 9 pontos.



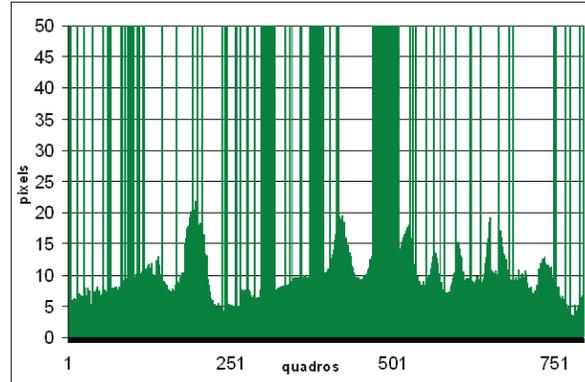
**Figura 6** Vídeo de teste. Os pontos vermelhos são os centros das esferas e os azuis são os pontos 3D reprojctados.

### 6. Conclusões e Trabalhos Futuros

A metodologia se mostrou plenamente capaz de fornecer os dados necessários para o teste de algoritmos de rastreamento de uma forma geral. Os marcadores utilizados são minimamente invasivos na cena e fornecem ótima precisão, sendo assim uma opção para aqueles que estiverem procurando um

método para gerar dados de referência com relativa facilidade.

Como trabalho futuro, seria interessante investigar a possibilidade de detectar os pontos de entrada e de saída automaticamente.



**Figura 7:** Gráfico de erro de reprojeção médio em cada um dos quadros do vídeo. Os erros que atingem o topo do gráfico significa que o ARToolKit não conseguiu calibrar a câmera.

*Agradecimentos:* à Petrobras que é a principal financiadora do Tecgraf/PUC-Rio e à Fabiola Maffra pela ajuda na montagem do caso de teste e revisão do artigo

### 7. Referências

[1] S Lieberknecht, S Benhimane, P Meier, N Navab, "A dataset and evaluation methodology for template-based tracking algorithms," IEEE / ACM ISMAR 2009

[2]. S. Baker, D. Scharstein, J.P Lewis, S Roth, M.J Black. , R. Szeliski: "A database and evaluation methodology for optical flow". ICCV. 2007

[3] M. Loaiza, A. Raposo and M. Gattass "A Novel Optical Tracking Algorithm for Point-Based Projective Invariant Marker Patterns", ISVC 2007

[4] L. Teixeira, M. Loaiza, A. Raposo and M. Gattass "Augmented Reality Using Projective Invariant Patterns", ISVC 2008

[5] J. Bouguet. Pyramidal implementation of the Lucas-Kanade feature tracker: description of the algorithm. Technical report, OpenCV Document, Intel Microprocessor Research Labs, 2000.

[6] K. Zimmerman, J. Matas, and T. Svoboda. Tracking by an optimal sequence of linear predictors. PAMI, 31(4):677–692, April 2009.