

# CALIBRAGEM VISUAL DE MANIPULADORES ROBÓTICOS COM USO DE INVARIANTES SIFT

MIGUEL A.G. PINTO, MARCO A. MEGGIOLARO, RICARDO TANSCHKEIT

*Dept. de Engenharia Elétrica e Engenharia Mecânica, PUC-Rio*  
*Rua Marquês de São Vicente, 225, Gávea, 22450-200, Rio de Janeiro, RJ*  
*E-mails: miguel.angelo@technomages.com.br, meggi@puc-rio.br,*  
*ricardo@ele.puc-rio.br*

**Abstract**— In this work, techniques are developed to allow a camera system attached to the end-effector of a manipulator (eye-in-hand) to identify its positioning errors with respect to the task reference frame. The identification is performed with the aid of the SIFT technique to correlate multiple images captured by the end-effector camera. Once calibrated within the task frame, it is possible to move the manipulator with safety and precision, as needed for automation of complex tasks. The approach is validated through simulations for a six-degree of freedom manipulator, used in underwater interventions in oil exploring platforms.

**Keywords**— Robotics; Calibration; Computer Vision; Control; Pattern Recognition; Stereoscopia.

**Resumo**— Neste trabalho, técnicas são desenvolvidas para permitir que câmeras ligadas a um sistema de visão computacional detectem automaticamente diferenças entre a posição real de um manipulador e sua posição desejada, localizando-o em seu volume de trabalho. A identificação é feita com o auxílio da técnica SIFT para correlacionar múltiplas imagens capturadas por uma câmera posicionada na extremidade do manipulador. Estas diferenças de posição são então enviadas para os controladores, para que estes corrijam a trajetória pré-programada. Os comandos de movimento do manipulador poderiam ser então pré-programados *off-line* sem a necessidade de ativar o robô. Isto permitiria maior velocidade em sua validação e na resolução de problemas. Simulações comprovam a eficiência da metodologia proposta, aplicadas a uma situação real de um manipulador submarino de seis graus de liberdade, usado em intervenções submarinas em plataformas de petróleo.

**Palavras-chave**— Robótica; Calibragem; Visão Computacional; Controle; Reconhecimento de Padrões; Estereoscopia.

## 1 Introdução

Muitas tarefas executadas por manipuladores robóticos requerem que estes reconheçam sua área de trabalho e tomem decisões com base neste reconhecimento. Um exemplo relevante é o caso de manipuladores robóticos utilizados em intervenções submarinas. Tais manipuladores operam em profundidades perigosas ao ser humano e devem executar tarefas em um ambiente que muda a cada vez em que o robô é transportado ao seu volume de trabalho. As dificuldades envolvidas no processo de posicionamento de um manipulador em ambientes completamente mutáveis é um problema complexo e pouco tratado na literatura, motivando fortemente este estudo.

Para efetuar o posicionamento do manipulador, este trabalho propõe uma técnica baseada em visão computacional. O uso de um sistema de câmeras é potencialmente rápido, automatizado e não-invasivo ao volume de trabalho. Há dois tipos de configuração para sistemas de medição baseados em imagens (Zhuang e Roth 1996). O primeiro é fixar câmeras no ambiente próximo ao da tarefa, para que possam visualizar tanto o ambiente quanto uma referência de calibragem presa na extremidade do robô. O segundo tipo de configuração consiste em montar uma câmera ou um par de câmeras na extremidade do manipulador (*eye-in-hand*). A configuração de câmeras móveis presas na extremidade do manipulador resolve o conflito entre alta precisão e grande volume de trabalho visualizável, e será considerada neste estudo.

O objetivo deste trabalho é desenvolver procedimentos para que um manipulador robótico possa se auto-localizar em relação ao ambiente no qual irá operar utilizando apenas um par de câmeras em sua extremidade. Para tanto, o manipulador precisa capturar pares de imagens obtidas em diferentes configurações de suas juntas, e automaticamente identificar e associar pontos de referência (pontos-chave) comuns aos pares. Através dos pontos correlacionados nessas imagens, é possível auto-calibrar o robô através de triangulação, possibilitando assim que ele possa trabalhar autonomamente mesmo em um ambiente com incertezas. O uso de câmeras para posicionamento do manipulador passa por três estágios: reconhecimento de padrões, acompanhamento de padrões, e recuperação da cinemática da câmera.

O primeiro estágio dedica-se a encontrar padrões nas imagens que sejam robustos a rotação, translação, mudança de escala e ruído. Diversos métodos são utilizados para este fim, sendo que o mais robusto é a identificação de cantos nas imagens. Este método, porém, dificulta o estágio seguinte, no qual se procura relacionar os mesmos pontos na outra imagem do par de câmeras. É comum utilizar correlação para efetuar este rastreamento, porém esta técnica não é robusta a rotação ou a mudanças de escala, o que torna ambos os métodos insatisfatórios.

Para identificação dos pontos e posterior rastreamento, será utilizada a Transformada de Propriedades Invariantes a Escala (SIFT - *Scale Invariant Feature Transform*) desenvolvido por David G. Lowe (Brown e Lowe 2002, Lowe 2004), que descobre padrões em imagens e associa a cada ponto um con-

junto de parâmetros para que estes possam ser identificados em imagens subseqüentes. Este método é robusto a rotação, translação, escala e ruído (Ait-Aider et al. 2002).

A última etapa da calibragem visa a encontrar a distância dos pontos capturados com relação à câmera, bem como obter a posição e orientação da câmera em relação ao mundo. Levando em conta que a câmera está instalada na extremidade de um manipulador, obtêm-se a posição e a orientação da extremidade em relação à base do robô, bem como as distâncias dos pontos vistos à extremidade. O algoritmo utilizado para tanto se chama triangulação.

Após a identificação da posição do manipulador através de câmeras, tarefas de teleoperação automática ou semi-automática são factíveis. Além disso, o conhecimento da posição absoluta do manipulador e do ambiente de trabalho ao seu redor permitiria gerar um sistema de visualização 3-D virtual que reflita as configurações do sistema em tempo real. A visualização virtual permite ao teleoperador observar áreas obscurecidas utilizando câmeras virtuais (Meggiolaro et al. 1999). No entanto, a utilização dos parâmetros identificados em um sistema automático encontram-se fora do escopo deste trabalho.

## 2 Emparelhamento de Pontos

Um dos principais problemas de visão computacional é encontrar semelhanças em imagens. Tal problema ocorre em reconhecimento de objetos e cenas, resolução de estruturas 3D em múltiplas imagens, correspondência estéreo e busca de movimento (Forstyn e Ponce 2003).

Para que seja possível obter a posição de um manipulador em relação ao seu ambiente, será necessário encontrar pontos correspondentes em ambas as imagens capturadas pelo par de câmeras, mesmo na presença de ruídos e variações na escala e rotação.

O método que será apresentado obtém pontos com propriedades robustas a variações na imagem, tais quais rotação, mudança de escala, iluminação e ponto de vista da câmera. O método é capaz de distinguir pontos diferentes a partir destas propriedades, com alta probabilidade de acerto. Essa metodologia é denominada Transformada de Propriedades Invariantes à Escala (“*Scale Invariant Feature Transform*”, SIFT), pois transforma dados da imagem em coordenadas invariantes à escala e a outras características locais, como luminosidade e rotação.

A primeira etapa do algoritmo é buscar pontos que sejam invariantes a mudanças de escala da imagem, possibilitando a detecção de pontos com a câmera próxima ou distante do objeto de interesse. Tal objetivo é alcançado procurando-se os pontos em todas as possíveis escalas, por meio de uma função chamada de espaço de escala, que, no caso, é a função gaussiana. Assim, o espaço de escalas  $L(x,y,\sigma)$  é formado pela filtragem da imagem  $I(x,y)$  por funções gaussianas  $G(x,y,\sigma)$ , conforme:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (1)$$

onde \* simboliza a convolução nas coordenadas  $x$  e  $y$  da imagem. A função  $G(x,y,\sigma)$  é expressa por:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (2)$$

A eficiência da busca por pontos-chaves é aumentada substituindo-se a função  $L(x,y,\sigma)$  por uma função de Diferença-do-Gaussiano, formada pela subtração de dois espaços de escala, separadas por um fator multiplicativo  $k$ , conforme mostrado na Eq. (3) e nas Figuras 1 e 2. Esta função se aproxima da função Laplaciano do Gaussiano normalizado em escala  $\sigma^2 \Delta^2 G$ , que possui pontos invariantes à escala em seus máximos e mínimos (Lindeberg 1994).

$$D(x,y,\sigma) = [G(x,y,k\sigma) - G(x,y,\sigma)] * I(x,y) \quad (3)$$



Figura 1 - Imagem (à esquerda) submetida à função gaussiana (ao centro) e à Diferença-do-Gaussiano (à direita), com  $\sigma=2$ .

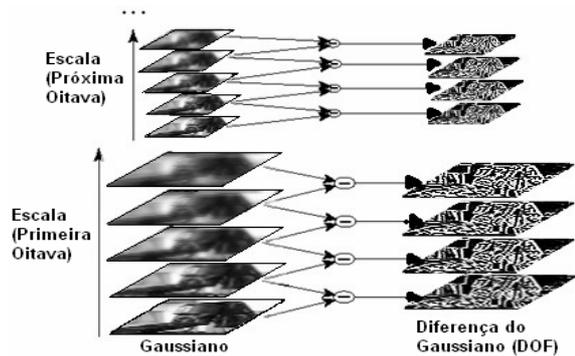


Figura 2 - Procedimento para obtenção de diferentes diferenças de gaussianas em oitavas da imagem.

Os pontos-chave são localizados nos máximos e mínimos locais de cada imagem  $D(x,y,\sigma)$ , que podem ser obtidos comparando-se a intensidade de cada ponto com as intensidades de seus seis pontos vizinhos na sua escala, e com os nove pontos vizinhos nas escalas acima e abaixo, como visto na Figura 3. Procuram-se pontos com valores de intensidade maiores ou menores que todos os seus vizinhos.

Para se obter invariância à rotação, atribui-se uma orientação baseada nas características locais da imagem. Para tanto calculam-se, para cada amostragem da imagem  $L(x,y)$ , na escala do ponto-chave, a magnitude  $m(x,y)$  e a orientação  $\theta(x,y)$  do gradiente usando-se diferenças de *pixels*.

Na região vizinha ao ponto-chave, forma-se um histograma de orientações composto pelas orientações dos pontos amostrados. Esse histograma possui 36 pontos, cobrindo os 360 graus de orientação.

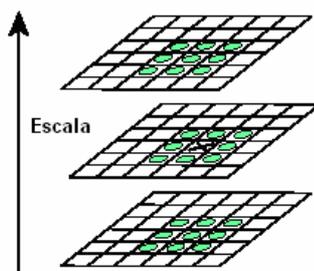


Figura 3 – Detecção de máximos e mínimos locais em uma mesma imagem e em outras escalas.

O pico deste histograma é utilizado para definir a sua orientação. No caso de múltiplos picos de elevada amplitude, o ponto-chave receberá múltiplas orientações, tornando-o ainda mais estável para futura identificação. Ao final, uma parábola é usada para interpolar os três valores do histograma mais próximos do pico, de forma a se ter uma melhor precisão de sua posição.

O próximo passo será atribuir a esses pontos-chave descritores invariantes a mudanças de iluminação e ponto de vista 3D, além de torná-los bem distinguíveis.

As magnitudes e orientações dos gradientes são então amostradas ao redor da localização do ponto-chave, utilizando a sua escala para escolher o nível de suavização Gaussiano para a imagem.

Para que os descritores tenham invariância à rotação, as coordenadas e as orientações dos descritores são rodadas em relação à orientação do ponto-chave. Este procedimento está mostrado pelas pequenas setas em cada amostra da localização no lado esquerdo da Figura 4.

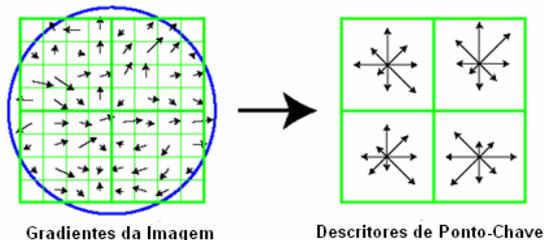


Figura 4 – Computação de descritores de pontos-chaves.

Uma janela de suavização gaussiana de  $\sigma$  igual à metade da largura da janela do descritor é usada para dar pesos à magnitude do gradiente em cada ponto amostrado, conforme visto à esquerda da Figura 4. Esse gaussiano evita mudanças súbitas do descritor com pequenas mudanças na posição da janela, e também reduz a ênfase nos gradientes longe do centro do descritor, que são mais afetados por erros.

Uma vez feita a suavização dos gradientes, criam-se histogramas de orientação sobre regiões de amostragem  $4 \times 4$ . O lado direito da Figura 4 mostra os quatro histogramas formados pela amostragem mostrada à esquerda. Na figura estão representadas oito direções para cada histograma, e o tamanho de cada seta representa a magnitude referente à respectiva entrada do histograma.

Para evitar efeitos de borda, no qual o descritor abruptamente muda enquanto a amostra se desloca suavemente indo de um histograma para outro ou de uma orientação para outra, uma interpolação trilinear é usada para distribuir o valor de cada amostra de gradiente para a sua posição no histograma adjacente.

O descritor de cada ponto-chave identificado é formado por um vetor contendo os valores de todas as orientações dos histogramas, correspondentes aos tamanhos das setas no lado direito da Figura 4. Para correlacionar pontos em imagens distintas, basta então procurar os pares com descritores similares, como descrito a seguir.

### 3 Relacionamento de Pontos-Chave em Imagens Distintas

Uma vez aplicado o algoritmo SIFT sobre duas imagens distintas do mesmo objeto, um conjunto de pontos-chaves em cada imagem e seus respectivos descritores será obtido. De forma a relacionar os pontos de forma robusta, faz-se uso da transformada Hough (Ballard e Brown 1982) para agrupar os pontos a partir de suas propriedades (localização, orientação e descritores) em um espaço de transformação 2D, e em seguida utilizar o algoritmo RANSAC (*Random Sample Consensus*) para melhorar as estimativas (Fischler e Bolles 1981).

A transformada de Hough é um método usualmente utilizado para detectar linhas ou curvas especificadas por uma função paramétrica. Ela cria uma matriz acumuladora n-dimensional onde acumulam-se “votos” para os parâmetros corretos das linhas e curvas encontradas na imagem. O uso da transformada de Hough para relacionar pontos é semelhante. Cada ponto-chave especifica quatro parâmetros: dois deles para localização 2D, e mais dois para escala e orientação. Assim, cria-se uma matriz acumuladora (espaço de transformada 2D) em translações ( $t_1$ ,  $t_2$ ), log escala ( $\log s$ ) e rotação ( $\theta$ ). Discretizam-se os tamanhos dos eixos de um oitavo do tamanho da imagem para translação, um oitavo de escala para escala, e um oitavo de radianos para rotação. A transformada de Hough identifica então agrupamentos de propriedades de pontos-chaves usando cada propriedade de um ponto de uma imagem para votar, dentre todos os outros pontos da outra, aquele que é mais consistente.

Os resultados da transformada Hough são refinados por meio do algoritmo RANSAC. Este método tem a vantagem de ser altamente insensível a pontos anormalmente distantes da maioria dos pontos amostrados, mas falhará se este grupo de pontos anormalmente distantes for muito grande. Por isso, a transformada de Hough é empregada inicialmente, de forma a agrupar as propriedades.

#### 4 Reconstrução

O objetivo final do uso de câmeras na calibragem do manipulador robótico é obter as posições de pontos identificados nas imagens em suas coordenadas reais no mundo. Tendo-se esta posição dos pontos, é possível descobrir a posição do manipulador em relação ao seu volume de trabalho para que ele possa trabalhar automaticamente. Este processo chama-se reconstrução. Dentre diversos algoritmos de reconstrução presentes na literatura, o escolhido foi o algoritmo de Triangulação. Outros métodos como Estrutura do Movimento e Geometria Epipolar foram testados, porém geraram resultados com menor precisão (Pinto 2006).

O método da triangulação é puramente geométrico. Tendo-se um par de câmeras calibradas e um conjunto de pontos  $p$  na primeira imagem corretamente relacionado a pontos  $p'$  na segunda imagem, e sendo a posição das câmeras  $O$  e  $O'$  respectivamente, encontra-se a posição do ponto no espaço 3D na interseção dos raios  $R = Op$  e  $R' = O'p'$ , conforme visto na Figura 5.

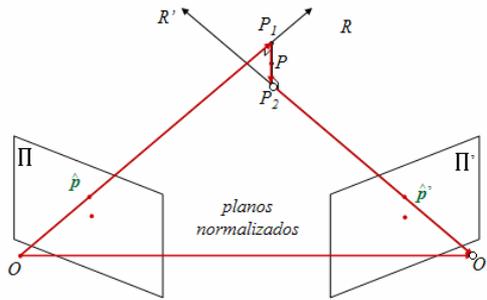


Figura 5 – Método da triangulação.

Na prática, esta interseção não ocorre devido a erros de calibragem de câmera e de localização dos pontos. Para solucionar este problema, calcula-se o ponto médio do segmento que mais aproxima os dois raios  $R$  e  $R'$ .

$$\overline{OP_1} + \overline{P_1P_2} + \overline{P_2O'} = \overline{OO'} \quad (4)$$

Expandindo a Eq. (4) com a posição relativa das câmeras:

$$z\hat{p} + c(\hat{p} \times R\hat{p}') + z'R\hat{p}' = t \quad (5)$$

onde  $R$  é a matriz de rotação entre as câmeras e  $t$  é o vetor de translação. Substituindo na Eq. (5) os parâmetros intrínsecos ( $K$  e  $K'$ ), tem-se:

$$zK^{-1}p + c(K^{-1}p \times RK'^{-1}p) + z'RK'^{-1}\hat{p}' = t \quad (6)$$

Resolvendo o sistema da Eq. (6), encontram-se os parâmetros  $z$  e  $c$ . O ponto  $P$  é obtido pela equação:

$$P = zK^{-1}p + c(K^{-1}p \times RK'^{-1}p) / 2 \quad (7)$$

O algoritmo de triangulação utilizado é confiável quando os pontos utilizados são bem correlacionados. O algoritmo SIFT é adequado para a obtenção de pontos estáveis e, em conjunto com RANSAC e a

transformada de Hough, é uma ferramenta robusta para correlacionar estes pontos.

Com as técnicas apresentadas acima, é possível identificar a posição de um par de câmeras em relação ao ambiente. Quando as câmeras estão montadas na extremidade de um manipulador, essas informações podem ser combinadas às dos sensores das juntas do robô e das equações da cinemática direta para identificar a posição de sua base.

#### 5 Auto-Localização Visual do Manipulador

Matrizes de transformação homogêneas são normalmente empregadas para localizar a posição da extremidade de um manipulador em relação à sua base. Este tipo de matriz é também utilizado para transformar um ponto físico no espaço tridimensional visto por uma câmera, presa na extremidade do robô, em um ponto no quadro da imagem. A matriz que faz esta transformação se chama Matriz de Projeção Perspectiva, e acumula parâmetros intrínsecos e extrínsecos da câmera.

É possível, a partir das imagens do ambiente, obter matrizes de transformação homogêneas que definam a posição de pontos neste ambiente em relação a um par de câmeras, e em seguida utilizar as matrizes obtidas na cinemática direta para obter a posição destes pontos em relação à base do manipulador.

Tendo-se duas câmeras calibradas acopladas à extremidade do manipulador em uma posição fixa, obtêm-se imagens da área de trabalho com ambas. A partir do uso do algoritmo SIFT, as imagens passam a ser convertidas em conjuntos de pontos  $p_L = (\rho_1, \rho_2, \rho_3, \dots, \rho_n)$  na câmera montada à esquerda e  $p_R = (\rho'_1, \rho'_2, \rho'_3, \dots, \rho'_m)$  na câmera direita, onde  $\rho_i = (u, v)$  são coordenadas na imagem. Por transformada Hough e RANSAC, são encontrados pontos em  $p_L$  que se relacionem a pontos em  $p_R$ .

Os pontos correlacionados são utilizados na triangulação para se obter a posição dos mesmos no espaço tridimensional,  $P^C$ , em relação à posição de uma das câmeras. A posição da câmera em relação à base do manipulador é conhecida, uma vez que se conhecem a cinemática direta do manipulador, os ângulos das juntas, e a posição da câmera em relação à extremidade, determinada por uma matriz de transformação homogênea  $A_C^n$  entre o sistema de coordenadas  $n$  e a câmera escolhida no par. A Eq. (8) transforma as posições dos pontos em relação à câmera  $P^C$  para sua posição em relação à base do manipulador  $P^O$ , obtendo assim a posição e orientação da base em relação ao volume de trabalho.

$$P^O = A_n^0 A_C^n P^C = A_1^0 A_2^1 \dots A_i^{i-1} \dots A_n^{n-1} A_C^n P^C \quad (8)$$

Por fim, após localizar a base do manipulador em relação ao ambiente de trabalho, pode-se movimentar a sua extremidade para qualquer dos pontos

desejados utilizando a sua cinemática inversa. A utilização deste algoritmo permite ao manipulador posicionar-se em qualquer ambiente sem a intervenção humana. Ao operador será oferecida apenas a tarefa de escolher posições da imagem onde ele deseja que a ferramenta esteja, e o algoritmo calculará automaticamente os ângulos das juntas necessários para alcançar o objetivo.

## 6 Resultados

Nesta seção, simulações são conduzidas para testar a eficiência da metodologia proposta para identificar a posição da base do manipulador em relação aos objetos vistos pela câmera. Em um cenário de intervenções submarinas, por exemplo, as técnicas apresentadas permitiriam que um manipulador conseguisse identificar a posição do veículo submarino (ROV) ao qual sua base está acoplada em relação ao seu volume de trabalho.

Primeiramente foi criado um conjunto de pontos  $P_R$  (na verdade, pequenas esferas na simulação) a distâncias conhecidas (consideradas as “reais”) da câmera esquerda de um par de câmeras fixo no espaço. O algoritmo foi testado para dez distâncias distintas aos pontos, variando entre 50mm e 2000mm no eixo Z da Figura 6. Para cada distância, foram criados  $n$  pontos com valores nos eixos X e Y variando aleatoriamente, mas sempre possíveis de serem visualizados por ambas as câmeras. As câmeras simuladas, com resolução de 320 x 240 pixels, foram afastadas entre si de 50mm no eixo X, e dispostas paralelamente, vide a Figura 6.

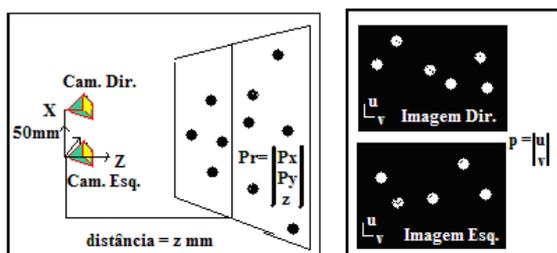


Figura 6 – Simulação para distância fixa de 50mm entre câmeras.

Uma vez determinados os pontos no plano da imagem, eles são utilizados no processo de triangulação. São utilizados os mesmos parâmetros intrínsecos e extrínsecos para obtenção de um conjunto de pontos  $P_P$  estimados por triangulação no espaço tridimensional em relação à câmera esquerda. Uma medida de erro  $P_E$  para cada distância  $d$  é obtida calculando-se a distância Euclideana entre os vetores  $P_R$  e  $P_P$ , para cada ponto, e em seguida calculando-se o erro médio quadrático:

$$P_E(d) = \sqrt{\frac{1}{n} \sum_{i=1}^n [P_R(i, d) - P_P(i, d)]^2} \quad (9)$$

A Figura 7 mostra os erros médios quadráticos em diversas distâncias de acordo com o eixo. Percebe-se que o erro aumenta com a distância das esferas à câmera, como esperado.

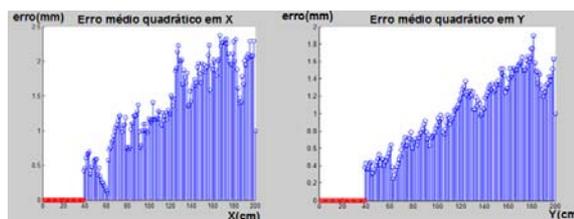


Figura 7 – Simulação de triangulação em função da distância de objetos à câmera.

Para implementar as simulações das câmeras montadas na extremidade de um robô, foi criada uma réplica em ambiente virtual de um manipulador robótico típico utilizado em intervenções submarinas, o TA-40 (Slingsby, 1998), vide Figura 8, e adicionadas duas câmeras virtuais calibradas em posições fixas de sua extremidade.



Figura 8 – Modelo do manipulador TA-40 e miniatura utilizada como “mestre” na tele-operação (Slingsby, 1998).

O manipulador TA-40 possui 6 juntas rotativas com ângulos representados pelas variáveis  $\theta_1$  a  $\theta_6$ . Seus parâmetros de Denavit-Hartenberg podem ser vistos na Tabela 1. Ele pode alcançar uma distância máxima de cerca de dois metros da base. Por esse motivo, nessa simulação pretende-se testar a capacidade do algoritmo de triangulação de calcular posições de objetos em imagens cuja distância até a câmera seja de até dois metros, dentro dos limites de operação do manipulador.

Tabela 1 –Parâmetros de DH do manipulador TA-40.

Elo $i$	$a_i$ (mm)	$d_i$ (mm)	$\alpha_i$ ( $^\circ$ )	$\theta_i$
1	115	0	90 $^\circ$	$\theta_1$
2	753	0	0 $^\circ$	$\theta_2$
3	188	0	90 $^\circ$	$\theta_3$
4	0	747	-90 $^\circ$	$\theta_4$
5	0	0	90 $^\circ$	$\theta_5$
6	0	360	0	$\theta_6$

Foram então inseridas esferas de cores diferenciadas no ambiente virtual, com posições fixas. O sistema de coordenadas do mundo é fixado na base do manipulador, em relação ao qual pretende-se identificar as posições reais das esferas. Em seguida, são obtidas imagens das esferas em diversas configurações do TA-40, longe de suas configurações de singularidade. A Figura 9 mostra um exemplo de configuração do manipulador. Os quadros à direita mostram a vista frontal e superior do manipulador,

juntamente com as esferas e uma representação dos ângulos de visão das duas câmeras. Nos quadros à esquerda, encontram-se as imagens que seriam capturadas pelas câmeras esquerda e direita.

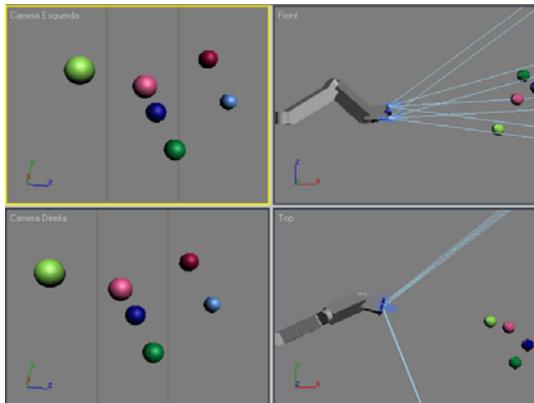


Figura 9 – Vistas da simulação de calibragem, com manipulador configurado com  $\theta_1=26^\circ$ ,  $\theta_2=30^\circ$ ,  $\theta_3=21^\circ$ ,  $\theta_4=47^\circ$ ,  $\theta_5=60^\circ$ ,  $\theta_6=-43^\circ$ .

Os algoritmos apresentados nesse trabalho são então aplicados a esses pares de imagens para calcular, por triangulação, as posições das esferas em relação à câmera esquerda. A Eq. (8) é então utilizada para obter essas posições em relação à base do manipulador, completando o processo de auto-localização do TA-40. É calculado o erro médio quadrático da posição real de cada esfera em relação à calculada.

Para a configuração da Figura 9, o erro de posicionamento da base do manipulador resultou em 12,6mm. Diversas outras configurações aleatórias do manipulador foram testadas, resultando no pior caso em um erro de 29,4mm. Considerando que o alcance máximo do manipulador TA-40 é de cerca de 2 metros, isto significa que a base do manipulador pode ser auto-calibrada com erros relativos da ordem de no máximo  $29,4/2000 = 1,5\%$ .

Estes resultados podem ser ainda melhorados com o uso de câmeras com maior resolução. Além disso, a simulação não pôde utilizar o algoritmo SIFT em virtude da dificuldade de gerar imagens em computação gráfica com textura suficiente para o seu emprego. Experimentos mostram que os pontos SIFT produzem bons resultados em pares de imagens reais. Isto porque objetos reais apresentam imperfeições e características distintas que são mais facilmente correlacionáveis entre imagens. Experimentos em laboratório foram conduzidos com objetos reais afastados de cerca de 1 metro de um par de câmeras de resolução  $640 \times 480 \text{ pixels}$  resultou em erros de calibração da base de menos de 5mm (Augustson 2007). Essa melhora se deveu não apenas à melhor resolução das câmeras, mas também à melhor eficiência do algoritmo SIFT em imagens reais. A calibragem das câmeras reais utilizadas na identificação dos erros também é abordada em (Augustson 2007), incluindo o efeito de distorções radiais e do alinhamento do centro ótico. Estes resultados estão sendo consolidados e serão apresentados em trabalhos subsequentes.

## 7 Conclusões

Foram apresentadas neste trabalho técnicas de calibragem visando o posicionamento de manipuladores em relação ao seu ambiente de trabalho. As técnicas abordadas são aplicáveis a qualquer manipulador robótico serial. Elas utilizam o algoritmo SIFT em conjunto com técnicas tradicionais de visão computacional para reconhecer padrões e estimar a posição tridimensional de objetos no volume de trabalho do manipulador. A junção dos algoritmos de visão computacional à descrição cinemática do robô permite que um manipulador consiga se auto-localizar automaticamente em seu ambiente de trabalho. Com isso, trajetórias pré-programadas podem ser executadas mesmo na presença de incertezas, uma vez que erros no posicionamento da base do robô podem ser identificados automaticamente.

## Referências Bibliográficas

- Ait-Aider, O., Hoppenot, P., Colle, E. (2002). Adaptation of Lowe's Camera Pose Recovery Algorithm to Mobile Robot Self-Localization. *Robotica*, 2002.
- Augustson, T.M. (2007). Calibragem Visual In Situ de Manipuladores Robóticos com Aplicação em Intervenções Submarinas. Dissertação de Mestrado, Dept. Eng. Mecânica, PUC-Rio.
- Ballard, D., Brown, C. (1982). *Computer Vision*. Prentice-Hall.
- Brown, M., Lowe, D.G. (2002). Invariant Features from Interest Point Groups. *Vision Conference*, Cardiff, País de Gales, pp.656-665.
- Fischler, M.A., Bolles, R.C. (1981). Random Sample Consensus; A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM*, Vol. 24, pp. 381-395.
- Forsyth, D.A., Ponce, J. (2003). *Computer Vision, a Modern Approach*. Ed. Prentice Hall.
- Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. University of British Columbia Vancouver, B.C., Canadá.
- Meggiolaro, M.A., Jaffe, P., Iagnemma, K., Dubowsky, S. (1999). A Force-Updated Kinematic Virtual Viewing System With Application to Nuclear Power Plant Maintenance. Proceedings of the *Tenth World Congress on the Theory of Machine and Mechanisms (IFTOMM)*, Finlândia.
- Pinto, M.A.G. (2006). Posicionamento e Calibração de um Manipulador Robótico Submarino com Uso de Visão Computacional. Dissertação de Mestrado, Dept. Eng. Elétrica, PUC-Rio.
- Slingsby (1998). Operation and Maintenance Manual for TA40 Manipulator. Perry Slingsby Syst. Ltd.
- Zhuang, H., Roth, S.Z. (1996). *Camera-Aided Robot Calibration*. Ed. CRC Press.