COOPERAÇÃO ENTRE HOMENS E ROBÔS BASEADA EM RECONHECIMENTO DE GESTOS

Flávio Garcia Pereira*, Norbert Schmitz†, Raquel Frizera Vassallo*, Karsten Berns†

* *Universidade Federal do Espírito Santo*
Vitória – Brasil

† *Technische Universität Kaiserslautern*
Kaiserslautern – Alemanha

Emails: flaviog@ele.ufes.br, nschmitz@informatik.uni-kl.de, raquel@ele.ufes.br, berns@informatik.uni-kl.de

Abstract— The development of robots for interaction with human beings is a challenging research topic. The inter-human communication is very complex and offers a variety of interaction possibilities. Although speech is often seen as the primary channel of information, psychologists claim that 60% of the information is transferred non-verbally. Besides body pose, mimics and others gestures like pointing or hand waving are commonly used. In this paper the gesture detection and control system of the humanoid robot ROMAN and the mobile robot Pioneer 3-AT are presented using a predefined dialog situation. The whole information flow from gesture detection till the reaction of the robots is presented in detail.

Keywords— Human-Robot Interaction, Gesture Recognition

Resumo— O desenvolvimento de robôs para a interação com seres humanos é uma linha de pesquisa desafiadora. A comunicação entre humanos é muito complexa e proporciona várias maneiras de interação. Embora a fala seja frequentemente vista como a principal via de comunicação, alguns psicólogos afirmam que 60% da informação é passada de forma não-verbal. Além da pose do corpo, mímicas e outros gestos como apontar ou o ato de balançar a mão são bastante usados. Neste artigo a detecção de gestos e sistemas de controle do robô humanóide ROMAN e do robô móvel Pioneer 3-AT são mostrados usando um diálogo pré-definido. O processo completo da detecção de gestos até a reação dos robôs é mostrada em detalhes.

Palavras-chave— Interação Homem-Robô, Reconhecimento de Gestos

1 Introdução

Teclado e *mouse* ainda são as interfaces de comunicação mais utilizadas entre homem e sistemas de computadores, não importando se este sistema é um *desktop*, um *laptop* ou um robô. Entretanto, há um grande interesse no desenvolvimento de diferentes tipos de interfaces tais como reconhecimento de voz, caracteres, detecção e reconhecimento de gestos, emoções, devido ao fato destas maneiras de interação trazerem maior naturalidade na comunicação homem-computador. Além disso, a maioria das pessoas sentem-se mais à vontade se puderem interagir com robôs da mesma maneira que se comunicam com outros seres humanos. O interessante é que alguns psicólogos afirmam que mais de 60% da interação entre os seres humanos é não verbal (Hall, 1990). Esta é uma das motivações do trabalho apresentado neste artigo.

O principal objetivo do presente trabalho é estabelecer uma interface de comunicação entre um ser humano e os robôs ROMAN (humanóide) e Pioneer 3-AT (móvel) através da detecção e reconhecimento de gestos. Os robôs devem reconhecer alguns gestos manuais feitos pelo ser humano e realizar alguma ação baseada no gesto detectado.

Existem várias técnicas de reconhecimento de gestos baseadas na forma da mão ou no movimento dos braços e mãos do ser humano. Gestos podem ser vistos como uma interação não verbal e podem ser desde um simples sinal definido pela forma das mãos, ações tais como apontar para um objeto a um movimento complexo feito para expressar idéias ou sentimentos permitindo a comunicação entre pessoas.

Assim sendo, para se reconhecer gestos, é necessário encontrar uma maneira de fazer com que os computadores sejam capazes de detectar configurações dinâmicas ou estáticas das mãos, braços e até mesmo de outras partes do corpo humano. Alguns métodos fazem o uso de dispositivos mecânicos tais como luvas (Sturman and Zeltzer, 1994) para estimar a posição das mãos e os ângulos dos braços. Estes métodos têm a desvantagem de o usuário ter que vestir uma luva desconfortável, a qual possui muitos cabos de conexão com o computador restringindo a área de trabalho e limitando os movimentos do usuário. O uso de visão computacional é uma das melhores opções para solucionar esse problema. Com ela é possível detectar e rastrear mãos e braços.

Os métodos baseados em visão estão focados no reconhecimento da forma da mão (gestos estáticos) ou na interpretação de gestos dinâmicos. O reconhecimento de gestos estáticos é feito através da identificação da forma da mão, silhuetas, contornos 2D ou modelos 3D (Chang et al., 2008; Stenger et al., 2001; Wu et al., 2001), enquanto os métodos que consideram gestos dinâ-

micos estão preocupados com a análise dos movimentos (Quek et al., 2000; Yang et al., 2006). Existem, ainda, trabalhos como o de (Waldherr et al., 2000) que trabalha tanto com gestos dinâmicos quanto com gestos estáticos.

A abordagem apresentada neste trabalho tem por objetivo criar uma interação com um robô e controlar alguns de seus movimentos. Além disso, pretende-se melhorar a interação e comunicação com o robô permitindo o uso de gestos e diálogos. Em (Chang et al., 2008), a forma da mão é usada para controlar a navegação de um robô utilizando gestos que indicam comandos como "parar", "mover-se", "para frente", etc. O sistema possui quatro módulos: detecção da mão usando segmentação da cor da pele, rastreamento da mão, reconhecimento da forma da mão e o controlador do robô. No trabalho de (Zhang et al., 2002) uma integração entre olhares e gestos é usada para instruir um robô em uma tarefa de montagem. Em (Sugiyama et al., 2006) um modelo de atenção para um robô humanóide é definido usando gestos e fala.

Este trabalho está focado em um método baseado em aparência para reconhecimento de gestos que ajudarão a melhorar a interação com um robô permitindo o uso de gestos e diálogos. A abordagem aqui apresentada também inclui um módulo de rastreamento da mão, assim o usuário pode fazer diferentes sequências de gestos enquanto o robô se mantém olhando para a mão (no caso do robô humanóide), sem ter que processar a imagem inteira para detectá-la e classificar um novo gesto. Os resultados iniciais são satisfatórios e para trabalhos futuros pretende-se obter informações temporais para classificar gestos mais complexos.

Este artigo apresenta, inicialmente na Seção 2, um método para detecção da cor da pele, a qual é um pré-processamento para a detecção de gestos tratada na Seção 3. A Seção 4 apresenta o algoritmo para o rastreamento da mão, enquanto as Seções 5 e 6, mostram, respectivamente, os resultados experimentais obtidos, as conclusões e trabalhos futuros.

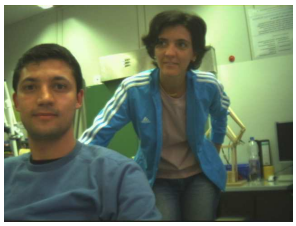
2 Detecção de Cor da Pele

Uma importante característica para a localização de faces, detecção e reconhecimento de gestos e identificação de pessoas, é a detecção da cor da pele. Esta detecção é normalmente uma etapa de pré-processamento, após a qual, apenas as regiões que contêm cor da pele são processadas.

Dentre as diversas maneiras de se realizar a segmentação da cor da pele (Kakumanu et al., 2007), é utilizado neste artigo, o modelo de cores RGB em conjunto com uma combinação de 16 Gaussianas para se determinar a probabilidade de cada *pixel* ser pele. Esta probabilidade é determinada por

$$P(x) = \sum_{i=1}^{N} \omega_i \frac{1}{\sqrt{(2\pi)^3 |\Phi_i|}} e^A, \qquad (1)$$

com $A = -\frac{1}{2}(x-\mu_i)^T \Phi_i^{-1}(x-\mu_i)$. Onde N é o número de Gaussianas, x é um vetor contendo os valores RGB para cada *pixel* da imagem, μ_i e Φ_i são, respectivamente, o vetor média e a matriz diagonal de covariância. O valor ω_i é a contribuição de cada Gaussiana. Os parâmetros da Gaussiana (μ_i, Φ_i e ω_i) são baseados no conjunto de treinamento mostrado em (Jones and Rehg, 2002). A Figura 1 (a) mostra uma imagem RGB enquanto a Figura 1 (b) mostra a imagem de saída do algoritmo de detecção de cor da pele. *Pixels* mais claros tem uma maior probabilidade de ser pele.



(a)

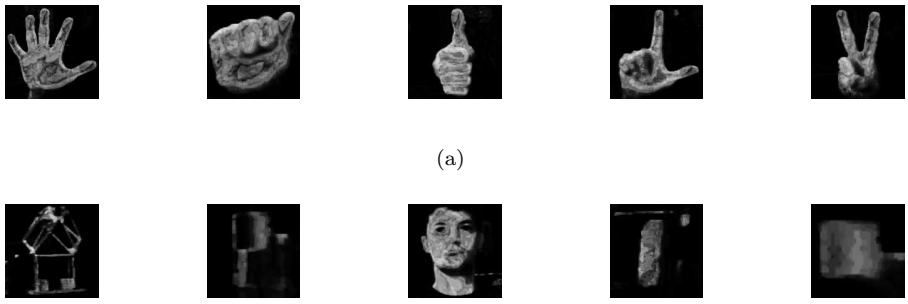
(b)

Figura 1: (a) Imagem RGB. (b) Imagem em tons de cinza após a detecção de pele. Um alto brilho indica uma alta probabilidade.

3 Reconhecimento de Gestos

Após a segmentação da cor da pele, são gerados alguns *blobs* contendo cor da pele. O algoritmo de detecção de gestos implementado neste artigo processa apenas os *blobs* com área maior que 200 *pixels*. O algoritmo detector de gestos aqui implementado utiliza a técnica de Análise de Componentes Principais (PCA - do inglês *Principal Component Analysis*) para criar uma base de autovetores e classificar os *blobs* que são encontrados.

Para construir a base de autovetores foram selecionados cinco gestos de diferentes pessoas e algumas regiões contendo cor da pele que não representam nenhum gesto. Os cinco gestos (Mão Aberta, Mão Fechada, Positivo, L e V) usados



(a)

(b)

Figura 2: (a) Os cinco gestos utilizados no conjunto de treinamento. (b) Regiões de cor da pele que não representam nenhum gesto.

para gerar as matrizes utilizadas pelo algoritmo PCA são mostrados na Figura 2 (a) e os falsos exemplos podem ser vistos na Figura 2 (b).

O conjunto de treinamento possui 90 imagens. Quinze imagens para cada um dos 5 gestos e 15 que não representam nada. A base gerada pela técnica de PCA possui 22 autovetores. Este número foi determinado baseado na energia dos autovetores, a qual é calculada com base nos autovalores associados a cada autovetor segundo

$$E(n) = \frac{\sum_{j=1}^{n} autovalor(j)}{\sum_{k=1}^{T} autovalor(k)}, \qquad (2)$$

onde n é o número de autovetores que serão usados para calcular a energia dos autovetores e T é o número total de autovalores. Este é o número mínimo de autovetores para que a energia do sistema seja maior que 80% da energia total dos autovetores. Um gráfico contendo as informações sobre este processo é mostrado na Figura 3. Neste gráfico pode ser visto que a energia dos autovetores cresce muito rápido devido ao fato desta energia ser calculada usando os autovalores, e os primeiros autovalores são grandes comparados com os últimos.

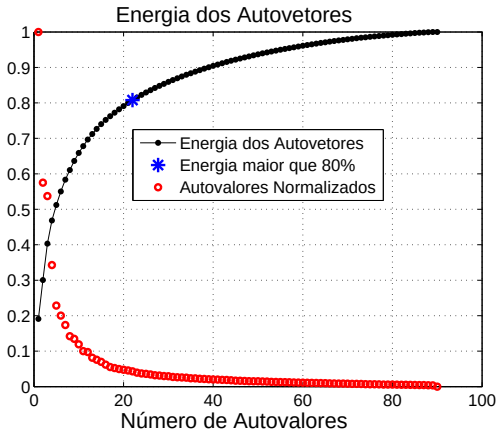


Figura 3: Seleção do número de vetores que irão compor a base de autovetores.

Após a segmentação da cor da pele, gera-se um vetor para cada *blob* detectado. Este vetor é projetado na base de autovetores para, desta forma, determinar o gesto que mais se assemelha com o *blob* em questão.

O algoritmo de reconhecimento de gestos implementado neste artigo é capaz de classificar corretamente gestos feitos diante de uma câmera. Assim, com as informações fornecidas pelo algoritmo de reconhecimento de gestos, é possível que um humano interaja com um robô usando um conjunto específico de gestos. O ser humano pode, por exemplo, parar o robô com um simples gesto.

Para comprovar a funcionalidade do algoritmo de reconhecimento de gestos, foram realizados testes com quatro diferentes pessoas. Colocou-se o algoritmo de reconhecimento de gestos para funcionar e, enquanto isso, cada pessoa apresentou, várias vezes, os cinco gestos que o algoritmo é capaz de detectar. Foram realizados 500 gestos em frente à câmera. Os resultados obtidos durante este teste podem ser vistos na Tabela 1.

Analisando a primeira linha da Tabela 1 pode-se notar que o gesto "Mão Aberta" foi corretamente classificado em 92% das amostras e em 8% não foi reconhecido.

4 Rastreamento da Mão

A interação baseada em gestos requer a detecção estática e dinâmica dos gestos. O rastreamento de um gesto pode ser usado para apresentar comandos mais complexos como olhar para um ponto ou girar em uma determinada direção.

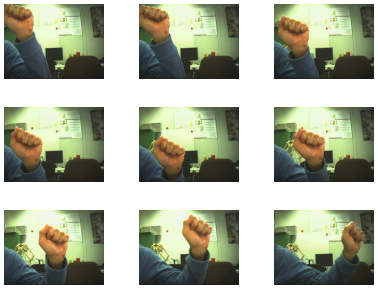
Neste trabalho esse rastreamento é realizado através do algoritmo *Camshift* (Bradski, 1998), o qual é um rastreador que usa a informação do histograma de cor de uma região de interesse. A região inicial do rastreamento é a posição e o tamanho do gesto detectado.

A Figura 4 mostra o rastreamento baseado em cor em um ambiente de teste. Na Figura 4 (a) podem ser vistas algumas posições da mão durante o

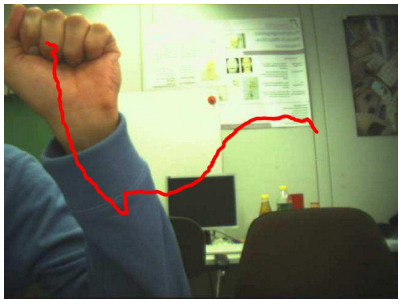
Tabela 1: Resultados do algoritmo de detecção de gestos.

	Mão Aberta	Mão Fechada	Positivo	V	L	Gesto Desconhecido
Mão Aberta	92%	0	0	0	0	8%
Mão Fechada	0	85%	0	0	0	15%
Positivo	0	0	88%	2%	0	10%
V	0	0	5%	90%	0	5%
L	0	4%	0	2%	87%	7%

processo de rastreamento enquanto a Figura 4 (b) mostra a posição inicial da mão e a trajetória descrita pela mesma determinada pelo algoritmo de rastreamento.



(a)



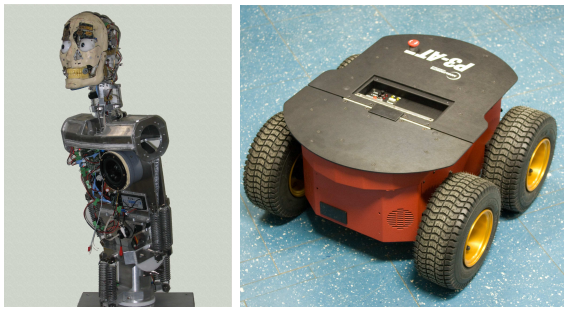
(b)

Figura 4: (a) Imagens da mão durante o processo de rastreamento. (b) Imagem de entrada capturada pela câmera. O movimento da mão é representado pela linha desenhada nesta imagem.

5 Experimentos

Os experimentos são realizados usando um pequeno diálogo com sinais de entrada não verbais. Durante o diálogo, o usuário pode apresentar cada um dos cinco gestos: Mão Aberta, Mão Fechada, Positivo, L e V.

Foram realizados dois tipos distintos de experimentos. Em um deles, faz-se o uso de um robô humanóide e, no outro, um robô móvel é empregado. A Figura 5 mostra em (a) e (b), respectivamente, o robô humanóide ROMAN e o robô móvel Pioneer 3-AT, utilizados nos experimentos aqui apresentados.



(a) (b)

Figura 5: (a) Robô humanóide ROMAN. (b) Robô móvel Pioneer 3-AT.

Com o robô humanóide ROMAN, os gestos permitem que o robô realize as seguintes ações. A Mão Aberta imediatamente para o robô qualquer que seja a ação que ele esteja executando, enquanto o gesto V permite que o robô se mova novamente. Quando o gesto L é detectado o humanóide começa a tocar uma canção e só para quando reconhece uma Mão Fechada ou Mão Aberta. O gesto Positivo inicializa o rastreamento e faz com que o robô olhe para o objeto a ser rastreado (a mão neste caso) e só para quando uma Mão Aberta é reconhecida.

Pode ser visto na Figura 6 o robô humanóide ROMAN em um ambiente de interação com duas pessoas.

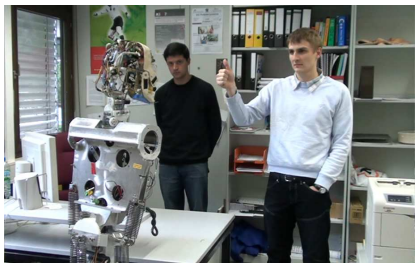


Figura 6: Robô humanóide ROMAN em uma situação típica de interação com seres humanos.

Os testes realizados com o robô móvel simulam uma situação em que uma pessoa chega em

um determinado lugar e o robô age como um recepcionista. Assim que o robô detecta a presença de um ser humano, ele faz uma saudação à pessoa e oferece as opções de onde o ser humano pode ir. Desta maneira, a pessoa pode escolher o lugar desejado apresentando um dos cinco gestos que o robô é capaz de reconhecer.

O robô móvel é dotado de um controlador de posição final (Secchi, 1998; Freire, 2002) e conhece as coordenadas dos lugares que ele pode ajudar o ser humano a chegar. Desta forma, assim que o ser humano apresenta um gesto, e este é identificado, o robô carrega as coordenadas do lugar a que está relacionado com o gesto em questão e inicia o movimento em direção ao lugar desejado. Após chegar ao seu destino, o robô retorna para a posição inicial e aguarda por uma nova pessoa. O fluxograma deste algoritmo pode ser visto na Figura 7.

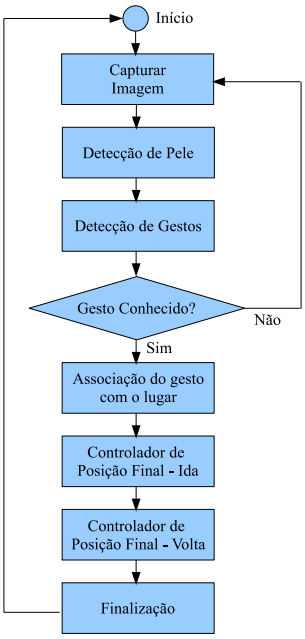


Figura 7: Fluxograma do algoritmo utilizado no robô Pioneer 3-AT.

Os experimentos foram realizados em um ambiente externo e sem nenhum obstáculo. Esta condição foi imposta pois o robô móvel utilizado, o Pioneer 3-AT, não possui nenhum sensor capaz de detectar os obstáculos que podem, eventualmente, aparecer em seu caminho.

Para ilustrar o funcionamento do algoritmo proposto, determinou-se a posição de algumas árvores localizadas em um campo aberto em relação à posição inicial do robô (considerado como "recepção"). A Tabela 2 apresenta as coordenadas destas árvores na área de trabalho do robô em relação à "recepção", assim como os gestos relacionados com cada uma delas.

Tabela 2: Posições relativas das árvores em relação à posição inicial do robô associada com cada gesto.

Gesto	Coordenadas (X; Y) [m]
Mão Aberta	(10; 17)
Mão Fechada	(11; 7)
Positivo	(13; -1)
V	(12.5; -9)
L	(8; 1)

Figura 8: Área de trabalho do robô móvel.

A Figura 8 ilustra o robô móvel em sua área de trabalho enquanto o gráfico apresentado na Figura 9 ilustra o experimento realizado com o robô móvel. Neste experimento, após o usuário humano apresentar o gesto "L" para a câmera, o robô navega até a posição (x, y) = (8m, 1m) de uma maneira que a pessoa que interagiu com ele possa segui-lo. Assim que o robô chega no destino, ele permanece ali por cinco segundos e, após isso, retorna à posição inicial e espera por um novo pedido de ajuda.

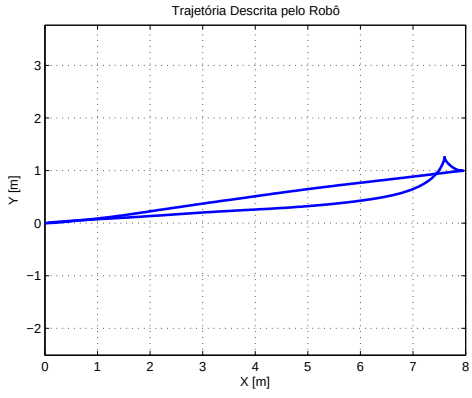


Figura 9: Trajetória descrita pelo robô durante a interação com o ser humano.

Os experimentos realizados foram satisfatórios visto que, assim que uma pessoa faz um dos cinco gestos conhecidos em frente da câmera, o robô se move em direção à posição desejada pelo usuário, conduzindo-o até lá e depois retorna para

a posição inicial, onde se mantém até que receba uma nova requisição de ajuda.

6 Conclusões e Trabalhos Futuros

O desenvolvimento da capacidade de interação não verbal entre robôs e seres humanos tem se tornado muito importante nos último anos. Além da interação baseada em voz, os sinais não verbais podem ser utilizados para melhorar esta comunicação entre robôs e seres humanos. Neste trabalho utilizou-se a detecção de gestos para controlar o robô humanóide ROMAN e o robô móvel Pioneer 3-AT.

A partir da segmentação por cor de pele em uma imagem colorida, a detecção de gestos foi realizada baseada no princípio da PCA. A informação dos gestos é capturada e transferida aos robôs os quais utilizam essa informação para realizar uma ação específica.

Os trabalhos futuros estão focados na integração de gestos estáticos e dinâmicos com sinais verbais e não verbais em um diálogo multimodal.

Além disso, pretende-se também montar um sensor laser sobre o robô móvel Pioneer 3-AT e, assim, introduzir ao controlador de posição final um módulo de desvio tangencial de obstáculos como o apresentado em (Pereira, 2006) aumentando a aplicabilidade em ambientes dinâmicos e com obstáculos. Também tem-se o objetivo de introduzir uma navegação baseada em um mapa do ambiente caso o robô seja usado em um ambiente interno, sem, todavia, desabilitar o controlador de desvio de obstáculos.

Agradecimentos

Os autores gostariam de agradecer à CAPES (Brasil) e ao DAAD (Alemanha), através do projeto de cooperação bi-nacional PROBRAL 282/07, pelo suporte financeiro. Este projeto de cooperação permitiu ao Norbert Schmitz passar um mês em Vitória – ES, no Brasil assim como possibilitou que Flávio Garcia Pereira passasse nove meses em Kaiserslautern, na Alemanha.

Referências

Bradski, G. R. (1998). Computer vision face tracking for use in a perceptual user interface.

Chang, J. S., Kim, E. Y. and Kim, H. J. (2008). Mobile robot control using hand-shape recognition, *Transactions of the Institute of Measurement and Control*, Vol. 30, pp. 143–152.

Freire, E. O. (2002). *Controle de Robôs Móveis por Fusão de Sinais de Controle Usando Filtro de Informação Descentralizado*, PhD thesis, Universidade Federal do Espírito Santo - UFES.

Hall, E. (1990). *The Silent Language*, B and T.

Jones, M. J. and Rehg, J. M. (2002). Statistical color models with application to skin detection, *International Journal of Computer Vision* **46**(1): 81 – 96.

Kakumanu, P., Makrogiannis, S. and Bourbakis, N. (2007). A survey of skin-color modeling and detection methods, *Pattern Recognition* **40**(3): 1106 – 1122.

Pereira, F. G. (2006). *Navegação e desvio de obstáculos usando um robô móvel dotado de sensor de varredura laser.*, Master's thesis, Universidade Federal do Espírito Santo - UFES.

Quek, F., McNeill, D., Bryll, R., Kirbas, C., Arslan, H., McCullough, K., Furuyama, N. and Ansari, R. (2000). Gesture, speech, and gaze cues for discourse segmentation, Vol. 2, pp. 247–254 vol.2.

Secchi, H. A. (1998). *Control de vehículos autoguiados con realimentación sensorial*, Master's thesis, Instituto de Automática de la Universidad de San Juan - INAUT/UNSJ.

Stenger, B., Mendonça, P. R. S. and Cipolla, R. (2001). Model-based 3D tracking of an articulated hand, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2**: 310.

Sturman, D. and Zeltzer, D. (1994). A survey of glove-based input, *Computer Graphics and Applications, IEEE* **14**(1): 30–39.

Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H. and Hagita, N. (2006). Three-layer model for generation and recognition of attention-drawing behavior, pp. 5843–5850.

Waldherr, S., Romero, R. and Thrun, S. (2000). A gesture based interface for human-robot interaction, *Autonomous Robots* **9**(2): 151–173.

Wu, Y., Lin, J. Y. and Huang, T. S. (2001). Capturing natural hand articulation, *IEEE International Conference on Computer Vision* **2**: 426.

Yang, H.-D., Park, A.-Y. and Lee, S.-W. (2006). Human-robot interaction by whole body gesture spotting and recognition, *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, IEEE Computer Society, Washington, DC, USA, pp. 774–777.

Zhang, J., Baier, T. and Hueser, M. (2002). Integration of gaze and gesture detection in nature language instructing of robot in an assembly scenario, pp. 241–246.