

Uma Interface Humano-Computador baseada em detecção e rastreamento de objetos

Cláudio dos Santos Fernandes
Curso de Ciência da Computação
Visão Computacional

21 de abril de 2010

Resumo

A proposta deste trabalho é utilizar um algoritmo de rastreamento de padrões em vídeos para implementar uma interface humano-computador. Tal sistema viabilizaria a substituição do mouse por movimentos livres da mão do usuário. Um modo de usabilidade como este seria bastante útil em aplicações de edição gráfica, por exemplo, nas quais a utilização de um mouse nem sempre oferece um nível adequado de precisão.

1 Pesquisas relacionadas

O artigo (PAVLOVIC; SHARMA; HUANG, 1997) discute as diversas técnicas empregadas por pesquisadores durante o processo de detecção, extração e rastreamento de padrões em imagens, bem como os maiores desafios da área e cita técnicas que visam contorná-los. Nele, é desenvolvido o conceito de sistema global de interpretação gestual baseado em visão computacional. De forma geral, estes sistemas requerem que um modelo matemático de gestos seja inicialmente estabelecido, como esquematizado na Figura 1.

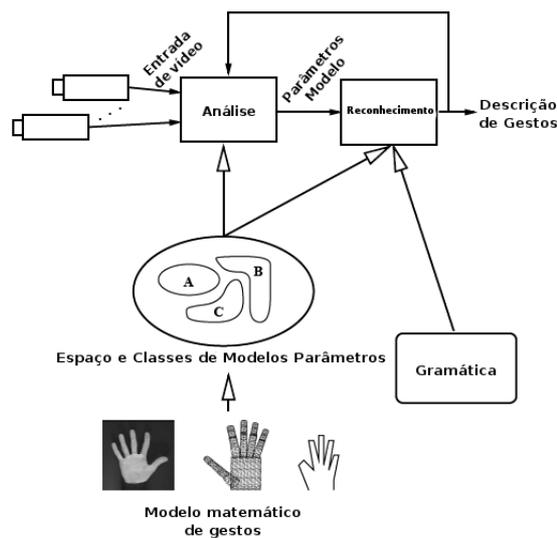


Figura 1: Diagrama do sistema de interpretação gestual definido em Pavlovic, Sharma e Huang (1997). Obtida e traduzido deste mesmo artigo.

Este artigo foca na questão do uso de técnicas de visão computacional aplicadas a problemas da área de Interfaces Humano-Computador. Por isso, seu desenvolvimento parte da definição de conceitos básicos, como gestos, a partir do ponto de vista desta área. Por exemplo, a seguinte proposição é feita acerca do conceito de gesto para ambientes IHC:

Em um ambiente IHC o seguinte conceito de regras determina a segmentação temporal de gestos:

1. O intervalo gestual consiste de três fases: preparação, movimento e retração.
2. A posição da mão durante o movimento segue um caminho classificável no espaço de parâmetros.
3. Gestos são confinados em um volume espacial especificado (chamado workspace).
4. Movimentos repetitivos da mão são gestos.
5. Gestos manipulativos tem tamanhos de intervalo gestual maiores do que gestos comunicativos.

Uma boa parte dos trabalhos realizados na área se baseiam na decomposição do corpo do usuário em objetos articulados. Esta abordagem tem dois grandes obstáculos: a grande quantidade de articulações a serem analisadas, e as dificuldades de obtê-las por meio de técnicas de visão computacional. Em técnicas de mapeamento tridimensional, elas são contornadas com o uso de premissas relacionadas às dependências entre juntas e seus limites em termos de ângulo máximo de rotação da articulação. Nas abordagens baseadas em modelos bidimensionais, a solução adotada é modelar gestos por meio da relação entre a aparência de qualquer gesto à aparência de um grupo de gestos pré-definidos em templates.

O processo de localização e extração das mãos em uma imagem é uma tarefa necessária e bastante complexa. É comum que algumas restrições sejam aplicadas para a simplificação deste processo. Tais restrições geralmente recaem sobre a escolha do background, do usuário e do sistema de captura de imagens. A primeira geralmente significa o uso de um background uniforme, enquanto a segunda se dá por meio do uso de luvas de cor e textura uniforme ou com marcadores em pontos específicos da mão. A abordagem de Kurata et al. () é inovadora em alguns aspectos. Foram utilizadas informações relacionadas às cores para detectar e rastrear a mão do usuário - ao invés de cores pré-definidas, os modelos de cor foram construídos dinamicamente. As premissas para detecção utilizadas foram: O usuário deve iniciar o processo de detecção apontando seu dedo para um “círculo guia”, o que parametrizaria o processo de identificação de sua mão. A partir daí, o rastreamento é feito por meio do algoritmo *Mean Shift*.

Uma abordagem similar foi adotada por Bottino e Laurentini (2007). Nesta pesquisa, foram utilizados os algoritmos de *Mean Shift* para detecção da mão, e *CAMShift* para seu rastreamento. Foi utilizado um marcador referencial na cena analisada. A silhueta da mão foi inicialmente extraída da imagem. Um modelo 3D foi então adaptado à imagem segmentada com o objetivo de reconstruir a postura da mão e a posição do marcador. O reconhecimento da postura da mão foi baseado no casamento da projeção de modelos 3D (composição de elipsóides) em planos bidimensionais e a silhueta da mão, o qual foi feito por meio do algoritmo *ICondensation*.

Referências

- BOTTINO, A.; LAURENTINI, A. How to make a simple and robust 3d hand tracking device using a single camera. In: *ICCOMP'07: Proceedings of the 11th WSEAS International Conference on Computers*. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2007. p. 414–419. ISBN 978-960-8457-95-9.
- KURATA, T. et al. *The Hand Mouse: GMM Hand-color Classification and Mean Shift Tracking*.
- PAVLOVIC, V. I.; SHARMA, R.; HUANG, T. S. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 19, n. 7, p. 677–695, 1997. ISSN 0162-8828.