Interfaces baseadas em técnicas de visão computacional

Carlos Hitoshi Morimoto
Departamento de Ciência da Computação - IME/USP
hitoshi@ime.usp.br

1 Resumo

Técnicas avançadas de interação homem-máquina são necessárias para aprimorar os tipos de interfaces de computadores mais comuns. Nessa palestra serão apresentadas algumas técnicas de visão computacional de tempo real para a estimação do olhar, e como utilizar essa informação para a construção de novos tipos de interfaces. Em particular apresentaremos MAGIC (Manual And Gaze Input Cascaded) Pointing, como um exemplo mais natural e eficiente de selecionar elementos na tela de um computador.

2 Introdução

Um dos objetivos da área de interação homem-computador (human-computer interaction - HCI) é estudar novas interfaces que permitam uma interação mais eficiente, ou seja, que produza canais de comunicação entre o usuário e o computador com altas taxas de transmissão de dados (banda larga). Por exemplo, a apresentação de informações visuais através de uma janela gráfica permite que o usuário receba uma grande quantidade de informações, mas infelizmente os dispositivos convencionais de entrada de dados (teclado e mouse) são bastante limitados e requerem a aquisição de algumas habilidades motoras para a sua eficiente utilização, sendo algumas dessas habilidades não triviais (principalmente em jogos).

Uma das sugestões mais comuns para melhorar as interfaces tradicionais do tipo WIMP (Windows, Icons, Menus, and Pointing) seria torná-las perceptivas, através da inclusão de novos modos de interação e técnicas computacionais, como processamento de fala natural e visão computacional. Esses novos modos permitiriam a comunicação direta por fala e gestos, além de expressões faciais, olhares, sons particulares, etc. Essas interfaces seriam mais naturais pois não requerem que o usuário adquira uma nova habilidade, se baseando em habilidades que a maior parte da população já domina.

Porém as dificuldades técnicas relacionadas ao desenvolvimento de sistemas de percepção como visão ou entendimento de fala tornaram o desenvolvimento desse tipo de interfaces, as interfaces perceptuais, inviáveis comercialmente, devido também, em parte, aos seus elevados custos computacionais. Porém, esses custos estão atualmente se viabilizando, e com a constante miniaturização e a crescente disceminação de dispositivos computacionais pessoais que apresentam multiplas funções como PDAs (Personal Digital Assitants) e telefones celulares, dispositivos que não suportam o uso de teclados e mouses convencionais, a pressão para o desenvolvimento de novos modos de interação com computadores vem aumentado.

Para facilitar as atividades do usuário, é nat-

ural tentar automatizar o maior número de tarefas possível, usando, por exemplo, a tecnologia de agentes autônomos [8]. Tais agentes requerem algum tipo de capacidade sensorial para serem capazes de interagir com o usuário. Esse paradigma de interação é conhecido como interfaces perceptuais (PUIs - Perceptual User Interfaces) [13].

3 Agentes Autônomos Baseados em Visão Computacional

Diversos exemplos de agentes autônomos perceptivos podem ser encontrados nas áreas de interação homem-computador e de agentes autônomos, principalmente aqueles com um sistema de percepção [4, 5, 10, 11, 12].

O estudo de comportamentos complexos de entidades biológicas através de agentes autônomos virtuais foram realizados em [10, 11]. Por exemplo, Terzopolous *et al.*[11] demonstram diversos comportamentos de peixes como agrupamento em cardumes, predadores e presas, acasalamento, etc, em ambientes complexos mas completamente virtuais, sem interação com um usuário humano. Seqüências bastante realistas de animação em vídeo usando computação gráfica foram criadas para visualizar esses comportamentos.

Tosa [12] se utiliza de uma rede neural para modelar um bebê artificial que reage emocionalmente aos sons feitos por um usuário olhando para dentro do berço do bebê. Um melhor exemplo de um sistema utilizando agentes autônomos com uma interface baseada em visão computacional é o sistema ALIVE [5], que demonstra diversos aspectos sobre o uso de agentes para aplicações de entretenimento interativo. O sistema é baseado em um sistema de visão computacional que permite a usuários humanos interagirem com um ambiente virtual povoado pelos agentes. O sistema de visão do sistema ALIVE usa uma única câmera para determinar a posição tri-dimensional (3D) da cabeça, mãos e outras partes salientes do corpo. A detecção do usuário é feita através de um algoritmo de segmentação por subtração de figura/fundo, e assume um fundo fixo conhecido. Essas limitações impostas por ALIVE tornam difícil a aplicação desse sistema de visão em computadores pessoais. Para poder ser aplicado em ambientes menos restritos, o sistema que desenvolvemos utiliza algumas propriedades geométricas e fisiológicas dos olhos para facilitar a detecção de faces, que são então rastreadas em tempo real. A próxima seção descreve esse sistema com maiores detalhes.

4 Sistema de Visão

O sistema de visão que estamos utilizando para detectar faces e rastreá-las é similar ao apresentado em [6]. O sistema usa duas fontes de luz infra-vermelha (IV) para criar imagens de pupilas brilhantes e escuras, como mostra a Figura 1, e uma camêra branco e preto sensível ao comprimento de onda IV utilizada. A pupila brilhante é gerada por uma fonte de luz IV colocada bem próxima ao centro óptico da câmera[1], enquanto a pupila escura é gerada pela segunda fonte de luz IV colocada um pouco distante do centro óptico (as duas fontes de luz nunca estão ligadas ao mesmo tempo) de forma a criar uma imagem de intensidade luminosa semelhante à primeira imagem, mas mostrando a pupila escura. Dispondo dessas duas imagens, a detecção das pupilas se dá através de

[1] o brilho da pupila é causado por um fenômeno semelhante ao ocorrido em fotos tiradas com um flash intenso, que tornam as pupilas vermelhas nas fotos

uma simples subtração entre as imagens, seguida de binarização. As pupilas são agrupadas em pares que correspondem a faces com o uso de algumas regras heurísticas e restrições geométricas. Após a detecção, cada face pode ser idependentemente rastreada.

4.1 Rastreador de olhar

Através de mínimas alterações no hardware do sistema de visão, este pode ser utilizado também como um rastreador de olhar [7]. Um rastreador de olhar é um dispositivo que permite estimar o local ou objeto que se encontra sob observação pelo usuário.

A única alteração de hardware que necessitamos é o uso de uma lente de maior magnificação, para observar mais claramente os movimentos da pupila. Uma vez que a pupila é detectada utilizando o método descrito anteriormente, o seu centro (CP) é calculado como sendo o centro de massa dos pixels pertencentes à região segmentada. Além do centro da pupila CP, o brilho produzido pela reflexão das fontes de luz sobre a córnea (veja a Figura 1a e b) também é segmentado e rastreado. Esse ponto brilhante (PB) é relativamente fácil de ser detectado, e é utilizado como ponto de referência para estimar o local sendo observado, após um processo de calibração.

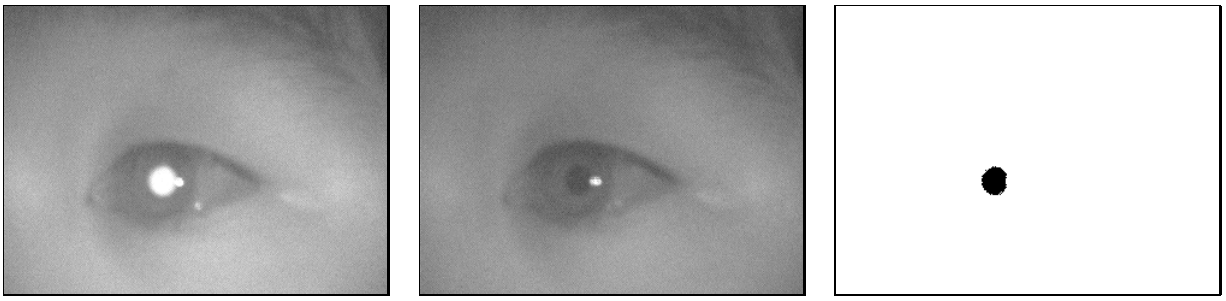
A calibração assume movimentos rígidos mínimos para a face, ou seja, o rosto pouco se move. Sendo assim, o movimento do olho pode ser aproximado como o de rotação de uma esfera, e o ponto de referência PB fixo no espaço (observe que, como PB é apenas a reflexão das fontes IV, sua posição é invariante a rotações do olho). Utilizando PB como referência, e dispondo do centro da pupila CP, podemos calcular um vetor $CP - PB$. Para calcular o mapeamento desse vetor com coordenadas sobre a tela do monitor, o usuário deve primeiramente fixar o seu olhar sobre um dentre 9 posições específicas sobre a o monitor quando solicitado, apertar uma tecla, e mover seu olhar para o próximo ponto. A partir das correspondências entre os vetores $CP - PB$ e os 9 pontos coletados, é possível interpolar todos os demais pontos, usando como aproximação, um polinômio de segundo grau.

5 MAGIC Pointing

Interfaces baseadas em rastreadores de olhar tem o potencial de serem mais naturais que as interfaces padrões pois requerem muito pouco treinamento. Jacob [3] descreve diversas maneiras para se utilizar esses dispositivos para apontar e selecionar objetos em interfaces, além de suas principais dificuldades de implementação, como o de não ativar todos os objetos selecionados (conhecido como o toque de Midas). Seus estudos indicam uma melhora de cerca de 30% no tempo para seleção por tempo de fixação do olhar quando comparado com um mouse. Outros experimentos e aplicações são apresentados em [2, 1].

Quando o tempo de fixação do olhar é utilizado para seleção, o usuário deve ajustar seu comportamento para evitar olhar para um objeto por longos períodos. Caso contrário, o objeto pode ser acidentalmente acionado. Algumas alternativas sugeridas a esse método é o uso de dispositivos mecânicos como uma chave, botão ou pedal.

Uma questão mais fundamental é se o uso do olhar é mais adequado para fins de apontamento e seleção, ou seja, se um dispositivo de apontamento baseado em rastreadores de olhar substituirão o mouse como o dispositivo padrão. As pessoas estão acostumadas a usar seus olhos para exploração do ambiente (en-



(a) (b) (c)

Figura 1: (a) Imagem de uma pupila brilhante e (b) pupila escura. Note o brilho perto do canto da pupila gerado pela reflexão da fonte de luz infra-vermelha. (c) Diferença entre as imagens de pupila escura e brilhante, após um processo de binarização.

trada sensorial) e não para manipulação (atuar sobre o ambiente). Assim mais estudos devem ser conduzidos para testar essa hipótese. Outro problema com a tecnologia atual é a falta de precisão, não sendo possível manipular objetos muito pequenos.

MAGIC (Manual And Gaze Input Cascaded) Pointing é uma forma elegante apresentada por Zhai *et al.*[14] de se utilizar um rastreador de olhar para fins de apontamento e seleção. A idéia básica é mover o cursor para o local sendo observado apenas quando o usuário demonstrar a intenção de fazê-lo, ou seja, quando ele toca ou movimenta o mouse. Se o cursor estiver originalmente distante da posição desejada, ele pode ser imediatamente transportado para lá, dentro das limitações do rastreador, e o mouse pode então ser utilizado apenas para realizar movimentos finos de correção.

Mesmo que o olhar se mostre inadequado para as funções de apontamento e seleção em geral, nós propomos outras possíveis aplicações, como deteção de contato visual para diferenciar entre dispositivos que recebam um comando vocal, ou realizar o "cache"de informações em hipertextos, perto da região sendo lida, ou contar o número de vezes que um certo objeto é observado, etc. Esses e outros projetos são descritos em [9].

6 Conclusão

Nesse artigo foram apresentados algumas possibilidades de se utilizar sistemas de visão computacional como novos modos de interação homem-computador. Foram introduzidos dois sistemas de visão, o primeiro com um campo visual largo apropriado para rastrear faces, e outro com um campo visual bem restrito, apropriado para rastrear olhares. A partir desses sistemas estamos desenvolvendo e testando novas formas de interação, buscando sempre interfaces com maior usabilidade, que facilitem a utilização de dispositivos computacionais em geral, aumentando a produtividade e o nível de satisfação das pessoas, sem exigir um longo período de treinamento.

Referências

[1] A. Glenstrup and T. Engell-Nielsen. Eye controlled media: Present and future state. Master's thesis, University of Copenhagen DIKU (Institute of Comput-

er Science), Universitetsparken 1 DK-2100 Denmark, June 1995.

[2] T.E. Hutchinson, K.P. White Jr., K.C. Reichert, and L.A. Frey. Human-computer interaction using eye-gaze input. *IEEE Transactions on Systems, Man, and Cybernetics*, 19:1527–1533, Nov/Dec 1989.

[3] R.J.K. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems*, 9(3):152–169, April 1991.

[4] Pattie Maes. Agents that reduce work and information overload. *Communications of the ACM*, 37(7):31–40, July 1994.

[5] Pattie Maes. Artificial life meets entertainment: lifelike autonomous agents. *Communications of the ACM*, 38(11):108–114, November 1995.

[6] C.H. Morimoto and M. Flickner. Real-time multiple face detection using active illumination. In *Proc. of the 3rd Int. Conf. on Automatic Face and Gesture Recognition*, Grenoble, France, March 2000.

[7] C.H. Morimoto, D. Koons, A. Amir, and M. Flickner. Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, 18(4):331–336, March 2000.

[8] H.S. Nwana and D.T. Ndumu. An introduction to agent technology. In H.S. Nwana and N. Azarmi, editors, *Software agents and soft computing*, pages 3–26, 10662 Los Vaqueros Circle, P.O. Box 3014 Los Alamitos, CA 90720-1314, 1997. Springer Verlag.

[9] IBM Almaden Research Center: BlueEyes Project. URL: http:// www.almaden.ibm.com/ cs/ blueeyes.

[10] Craig Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Computer Graphics: Proceedings of ACM SIGGRAPH 87*, volume 21, July 1987.

[11] D. Terzopoulos. Artificial fishes with autonomous locomotion, perception, behavior and learning, in a physical world. In P. Maes and R. Brooks, editors, *Proc. of the Artificial Life IV Workshop*. MIT Press, 1994.

[12] N. Tosa. Neurobaby. In *ACM SIGGRAPH-93 Visual Proceedings, Tomorrow's Realities*, pages 212–213, 1993.

[13] M. Turk. Moving from guis to puis. Technical Report MSR-TR-98-69, Microsoft Research, 1998.

[14] S. Zhai, C.H. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *Proc. ACM SIGCHI - Human Factors in Computing Systems Conference*, pages 246–253, Pittsburgh, PA, May 1999.