

1 Introdução

Muitos usuários tentam imaginar por que os computadores ainda permanecem “cegos” apesar do grande avanço no poder de processamento. Mas como pode-se esperar “inteligência” de máquinas que não estão cientes do que acontece a seu redor, se unicamente podem sentir sinais do teclado e movimentos do mouse?

Recentemente, o preço das câmeras digitais de vídeo tem diminuído drasticamente. Por menos de 200 reais, cada computador pode ser equipado com um novo e útil sentido: a visão. Como a Visão Computacional pode ser utilizada para permitir uma interação mais veloz ou simplesmente mais intuitiva? Este trabalho tentará responder tais perguntas com um exemplo de dispositivo de interação baseado em gestos da mão.

1.1. Motivação

Por muito tempo, a pesquisa em Interfaces Humano-Computador (IHC) tem sido restrita à utilização do monitor, do teclado e do mouse. Recentemente este paradigma tem mudado. Técnicas como Visão Computacional, Reconhecimento de Voz e Dispositivos de Rastreamento (*Tracking*) têm permitido uma interação mais intuitiva e natural entre homem e máquina. Mesmo com estas novas tecnologias, existe ainda um enorme abismo entre o mundo dos humanos e o mundo das máquinas. A maioria dos computadores existentes não está ciente de nós ou de nossos sentimentos a menos que os expressemos explicitamente.

Os dispositivos de interação baseados em gestos são tradicionalmente construídos utilizando diferentes dispositivos de rastreamento (*Tracking*) óptico, magnético ou mecânico ligados ao computador e/ou colocados no corpo do usuário. Muitos desses sistemas utilizam equipamentos sofisticados (dispositivos de rastreamento, luvas, câmeras especiais, etc.) que os tornam difíceis de serem comercializados em massa devido ao alto custo. Além disso, muitas vezes, devido a sua operabilidade, requerem espaços e ambientes especiais.

Alguns sistemas que não utilizam dispositivos de rastreamento, os baseados em Visão, utilizam técnicas de reconhecimento de padrões nas imagens capturadas. Para conseguir isto, os sistemas utilizam algumas restrições do ambiente, do fundo da cena, da cor dos objetos a serem reconhecidos, das condições de iluminação e das características das câmeras e constroem ambientes bem controlados, de maneira a facilitar as fases iniciais do reconhecimento em benefício da operabilidade do sistema. Alguns desses sistemas são apresentados no Capítulo 2 da presente dissertação.

Uma interação natural entre humanos e computadores não deveria nem requerer instalações especiais nem envolver qualquer dispositivo (mecânico, óptico ou magnético) que a pessoa devesse vestir ou manipular. Assim como temos a habilidade de estarmos cientes do ambiente à nossa volta através de nossos sentidos, a princípio o computador deveria ser capaz de imitar esta habilidade usando interfaces que emulem os sentidos, o que o tornaria mais fácil e mais intuitivo de usar. Conseqüentemente, novas linguagens e interfaces não convencionais precisam ser construídas e devem ser capazes de traduzir para o computador os gestos, sentimentos, expressões, linguagem e demais meios de comunicação utilizados na interação natural entre humanos.

1.2. Objetivos

Este trabalho estará focado na interação Humano-Computador baseado em gestos de uma mão limpa. “Limpa” neste contexto significa que nenhum dispositivo (mecânico, magnético ou óptico) é colocado na mão para poder interagir com o computador.

Além disso, o foco também estará centrado em diminuir ou eliminar as restrições especiais impostas para a interação. Ambientes especiais e com condições de iluminação controladas, cor de fundo preestabelecida e cores específicas dos objetos, a princípio, são descartados. Em termos de custo são levados em conta dispositivos (câmeras) de prateleira, facilmente acessíveis.

Considerando as características que devem ser levadas em conta, o objetivo principal deste trabalho é mostrar que através do uso de diferentes técnicas de Visão Computacional combinado com a utilização de dispositivos de baixo custo como simples câmeras *Web*, é possível construir um mecanismo de interação Humano-Computador sem considerar grandes restrições de ambiente,

iluminação, cor dos objetos, etc. de forma que a interação com o computador seja feita da maneira mais natural possível. Para isto construímos uma interface baseada em gestos da mão na qual os gestos e as posições dos dedos são diretamente utilizados para interagir com as aplicações.

O escopo desta pesquisa estará limitado ao reconhecimento de alguns gestos da mão (identificação de qual gesto dentre os gestos predefinidos está sendo mostrado pelo usuário) e à detecção e rastreamento dos dedos (localização da posição de certos dedos em imagens de vídeo). Às vezes são realizados alguns ajustes baseados nas características dos ambientes, mas tais ajustes são pequenos e feitos numa fase de instalação de forma a não comprometer a usabilidade do sistema. Para avaliar o desempenho do sistema de interação baseado em gestos da mão e por sua vez avaliar o mecanismo de interação Humano-Computador sem restrições significativas, desenvolvemos algumas aplicações para demonstrar o sistema em funcionamento.

1.3. Estrutura do Documento

O capítulo 2 está dividido em duas partes. Na primeira parte é feita uma revisão de alguns conceitos básicos, importantes para o melhor entendimento e contextualização do mecanismo de interação proposto. São definidos alguns requisitos funcionais e não-funcionais dos sistemas de interação em tempo real. Finalmente são expostas algumas vantagens dos sistemas de interação baseados em Visão que procuram justificar a sua utilização. Na segunda parte é apresentado um conjunto de trabalhos de pesquisa relacionados com o tema de estudo deste trabalho.

No capítulo 3 é apresentada a técnica de Subtração de Fundo, utilizada para segmentar o objeto de interesse do fundo da cena. Nesse capítulo é feito um estudo e avaliação dos diferentes algoritmos propostos para este tema. O objetivo principal, além de facilitar a detecção do objeto, é diminuir ou eliminar as restrições do ambiente, de iluminação, de cor, etc. presentes em outros sistemas. Julgamos este capítulo importante, em linhas gerais, por ser a base sobre a qual foram implementadas as idéias desenvolvidas nos capítulos subsequentes.

No capítulo 4 é apresentado um mecanismo para a detecção da silhueta da mão. Nesse capítulo, em uma primeira etapa é feita uma melhoria da segmentação, e depois é apresentada uma abordagem para a obtenção das

silhuetas. Os algoritmos implementados nesta etapa objetivam detectar, coletar e armazenar os pontos da imagem que conformam a silhueta da mão.

No capítulo 5 é apresentado o processo de reconhecimento de gestos da mão. Além dos gestos, alguns parâmetros, tais como a posição e orientação dos dedos, também são calculados. O processo de reconhecimento é feito tendo como referência o número de dedos detectados na mão e um conjunto predefinido de gestos a serem reconhecidos.

No capítulo 6 são apresentadas algumas aplicações do mecanismo de interação Humano-Computador utilizando gestos da mão. Os gestos da mão e a posição dos dedos são diretamente utilizados para interagir com as aplicações.

Finalmente no capítulo 7 temos as considerações finais sobre o trabalho de pesquisa desenvolvido e os trabalhos futuros que podem resultar a partir deste.