



Interação Natural com **imagem** **Projetada**: proposta alternativa ao uso de sensores reais

NATHANAEL OLIVEIRA VASCONCELOS

THIAGO DA SILVA ALMEIDA

ALCIDES XAVIER BENICASA



Resumo

A interação entre humanos e computadores é um campo de pesquisa que tem sido bastante investigada nos últimos anos. O uso de diferentes tipos de sensores, como por exemplo, sensores de movimento, toque, etc, tem proporcionado grandes avanços nesta área. Entretanto, de acordo com a especificidade de cada sensor, o custo para o desenvolvimento de determinadas aplicações torna-se muitas vezes inviável. Sendo assim, este trabalho possui como principal objetivo a proposta de um modelo como alternativa ao uso de sensores reais, neste caso, sensores de toque para interação homem-máquina, para isso foi desenvolvida uma aplicação que torna possível a interação entre o usuário e a imagem projetada por um projetor de imagem (datashow).

Palavras-chave

Interação humano-computador; imagem projetada.



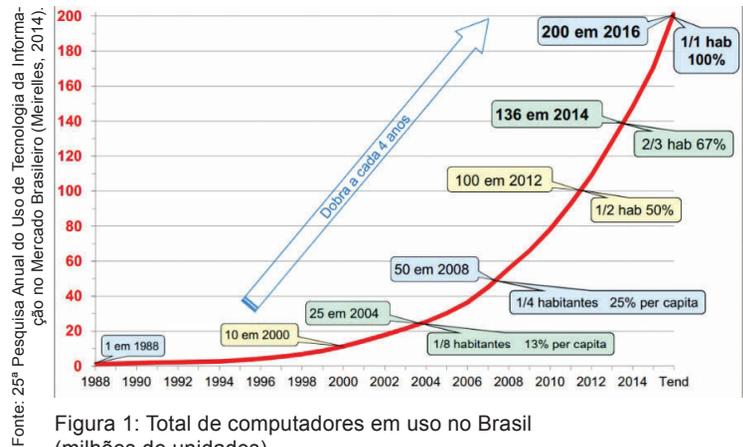
O QUE É INTERAÇÃO NATURAL?

Para Hewett et al. (2009), a Interação Homem-Computador (IHC) é a disciplina que se preocupa com o planejamento, avaliação e implementação de sistemas computacionais interativos para o uso humano. Uma das formas de interação mais conhecida atualmente para interagir com os computadores é o mouse. Entretanto, nos últimos anos, a evolução tecnológica permitiu a criação de vários dispositivos, os quais permitiram o surgimento de uma nova modalidade de interação, disponível em aplicações como telas sensíveis ao toque, reconhecimento de comandos de voz, reconhecimento de gestos, entre outras. Esse novo paradigma é conhecido como Interfaces Naturais de Usuário (INU).

Uma interface INU exige apenas que o usuário seja capaz de interagir com o ambiente por meio de interações previamente já conhecidas pelo mesmo. Esse tipo de interface exige aprendizagem, porém é facilitada, pois não exige que o usuário seja apresentado a um novo dispositivo.

PORQUE É VIÁVEL INVESTIR EM INTERAÇÃO HOMEM-COMPUTADOR?

Nos últimos anos os computadores têm se tornados dispositivos comuns ao cotidiano das pessoas. De acordo com Meirelles (2014), coordenador da 25ª Pesquisa Anual do Uso de Tecnologia da Informação no Mercado Brasileiro, divulgada em abril de 2014 pela Fundação Getúlio Vargas, existem 136 milhões de computadores em uso no Brasil, uma densidade de 67% per capita ou 2 computadores para cada 3 habitantes, ainda conforme a pesquisa, estima-se que em 2016, o Brasil deve ter um computador em uso para cada habitante, ou 200 milhões de máquinas.



Fonte: 25ª Pesquisa Anual do Uso de Tecnologia da Informação no Mercado Brasileiro (Meirelles, 2014).
Figura 1: Total de computadores em uso no Brasil (milhões de unidades)

Exemplos de interação e qual a utilizada neste trabalho

Com o objetivo de aproximar e facilitar a interação entre o homem e a máquina, atualmente, diversos *consoles*, computadores e *smartphones* apresentam funcionalidades de interatividade baseadas em comandos por voz, gestos, ou ainda, pela captura de movimentos do globo ocular. Neste trabalho é proposta uma aplicação em que é possível interagir com a máquina através do uso de gestos. Esta forma de interação está em ascensão, sendo uma forma de interação natural, ou seja, através de ações naturalmente utilizadas pelas pessoas para se comunicarem.

Quais os mecanismos utilizados para construir dispositivos baseados em gestos?

Uma das formas é utilizando mecanismos de rastreamento 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.). Outra forma de interação através de gestos é baseada na visão computacional, que tem como objetivo a interpretação automática de cenas complexas através da captura de imagens com câmeras normais, para isso, são



utilizadas técnicas de reconhecimento de padrões nas imagens capturadas, ao invés de dispositivos de rastreamento, além de algumas restrições do ambiente, como por exemplo, o fundo da cena, cores dos objetos a serem reconhecidos, e condições de iluminação, construindo assim, ambientes bem controlados, de maneira a facilitar operabilidade do sistema.

Qual o mecanismo utilizado neste trabalho e quais as vantagens em relação aos outros?

Segundo Aggarwal e Cai (1999), a relevância da comunicação corporal é percebida desde as primeiras investigações da área de IHC, setores de pesquisa se dedicam a estudos na área de rastreamento do corpo humano desde o início da década de 1980. Desde então, os métodos que se propõem a tal tarefa passaram a abranger diferentes formas de perceber o corpo do usuário, como é o caso das luvas de dados (Baudel e Beaudouin-Lafon, 1993; Systems), que simplificam a identificação de gestos.

Nos dias atuais, com o maior acesso à dispositivos de captura de imagens e computadores com capacidades suficientes de processamento, é possível o desenvolvimento da IHC a partir de gestos com mãos livres, baseando-se somente na interpretação da cena através de visão computacional (Freeman et al., 1998; Mistry and Maes, 2009).

É fato que a utilização de dispositivos tais como: luvas de dados, sensores eletromagnéticos, entre outros, simplificam a identificação de gestos, mas por outro lado aumentam o custo do sistema e, em alguns casos, requerem calibração, tornando assim uma tecnologia ainda não muito acessível à maioria da população. Ao contrário desses dispositivos, interfaces baseadas em visão computacional oferecem diversas vantagens, consistindo numa interação natural entre humanos e computadores, sem a necessidade de instalações especiais, nem da utilização de qualquer dispositivo (mecânico, óptico ou magnético) que o usuário deva vestir ou manipular, possibilitando assim, uma maior liberdade e facilidade no uso do sistema.

Sendo assim, este trabalho possui como principal objetivo propor um modelo que, utilizando técnicas de Inteligência Artificial e Processamento de Imagem, torne possível a detecção, reconhecimento da ação e interação do usuário com o computador, a partir de informações providas por uma câmera de vídeo, de modo que não seja necessária a utilização de sensores especiais. Para validação do modelo proposto, foi desenvolvida uma aplicação que possibilita a interação intuitiva entre o usuário e a imagem projetada por um projetor de imagem (*datashow*).

Como podem ser observados na Fig. 2, os dispositivos necessários para a utilização do aplicativo proposto são bastante comuns, sendo os seguintes: computador, *datashow* e câmera, que pode ser a própria câmera embutida no computador.

Fonte: ilustração Gilmar Marcel Oliveira dos Santos



Figura 2: Layout do aplicativo proposto.

Qual o processo necessário para a identificação da interação?

Para permitir a interação do usuário com a imagem projetada foi necessária a inserção de algumas informações adicionais ao slide, ou imagem projetada, de forma que, intuitivamente, saiba-se onde e como a interação deva ocorrer. Para isto, foi criada uma máscara, como pode ser visto na Fig. 2, que é incorporada à lateral direita da imagem projetada, composta, neste trabalho, por quatro funções básicas: anterior, início, próximo e final.

Para a identificação das posições das op-



ções, inicialmente foi inserida uma máscara de reconhecimento, que destaca as regiões em vermelho (para facilitar a localização). Com testes verificou-se que ao capturar a foto da projeção perde-se qualidade, interferindo no resultado da identificação das opções, por esse motivo, é necessário um pré-processamento, em que são destacados os pontos vermelhos (regiões das opções).

Ao capturar a imagem inicial e concluir o pré-processamento, é feito o tratamento e obtenção de informações relevantes da cena que possam identificar os alvos (opções), a etapa seguinte é a segmentação (as regiões das opções são isoladas) e, por fim, a identificação de interação. As etapas desse processo são apresentadas na Fig 3.

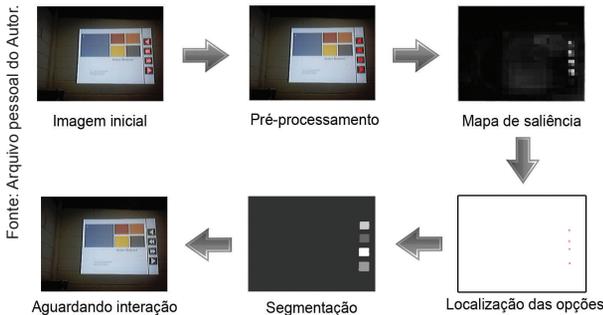


Figura 3: Processo resumido para a identificação das opções.

Um dos principais pontos deste trabalho está relacionado à identificação do momento no qual uma determinada interação tenha ocorrido, assim como a função de navegação correta a ser executada. A diferença entre os valores dos *pixels* de cada imagem capturada é a principal informação para a identificação da interação. Apesar de parecer uma tarefa trivial, a aplicação proposta neste trabalho, em um ambiente real, pode tornar-se instável, uma vez que duas imagens capturadas em um intervalo de segundos, ou até menos do que um segundo, podem apresentar diferenças sutis nos valores dos *pixels*, geradas por interferências do ambiente, mesmo que imperceptíveis à visão humana. Sendo assim, para que a comparação entre duas imagens, aparentemente iguais, seja realizada de forma correta, foi necessário definir um valor aceitável de diferença entre duas imagens, ou seja, um va-

lor máximo de mudança, tornando possível a diferenciação entre uma interação real e variações naturais do ambiente.

O valor de mudança aceitável foi definido baseado em cálculos estáticos dos valores dos histogramas para um conjunto de imagens capturadas sequencialmente em um intervalo curto de tempo. A medida utilizada compara diferenças entre um conjunto de dados, determinando a dispersão dos valores em relação à média do conjunto.

De forma resumida, após esse cálculo, a apresentação é iniciada em modo interativo e, em intervalos de 400ms, uma imagem da apresentação é capturada, calculando-se em seguida a média dos valores dos *pixels* referente a cada função. A seguir, a média obtida é comparada com a diferença entre a média dos valores dos *pixels* obtidos das imagens iniciais e o valor aceitável de mudança, de forma que, caso o valor da média dos valores dos *pixels* obtidos da imagem atual seja menor, pode-se concluir que houve uma interação, de forma que o comando relacionado à função que tenha apresentado a diferença seja executado. Para o caso de detecção simultânea de interação, nenhuma função deverá ser executada.

A aplicação proposta neste trabalho pode ser considerada como uma interessante alternativa para o desenvolvimento da interação entre o homem e a máquina?

De acordo com os experimentos realizados, fatores como luz alta sobre a apresentação (consequência do reflexo do próprio *datashow* e/ou de lâmpadas), ou luz ambiente alta (causada pelas lâmpadas e/ou grande quantidade de janelas) podem interferir no funcionamento do modelo. Entretanto, a aplicação apresentou desempenho excelente em ambiente ideal, com uma taxa de interação bem sucedida acima de 90% em todos os ambientes testados, de modo que não houveram interpretações errôneas de funções e com um pequeno número de disparo de função na ausência de interação, sendo que, este último, deu-se por consequência de variações do ambiente, como



por exemplo, sombra, reflexo ou outros fatores. O ambiente ideal foi definido como uma projeção realizada em superfície de fundo claro, com iluminação ambiente normal, ou

ainda, sem iluminação no ambiente, contanto que, não exista incidência direta de luz forte sobre a projeção. Exemplos desses ambientes são apresentados na Fig. 4.

Fonte: Arquivo pessoal do Autor.



Figura 4: Exemplos de ambientes ideais.

CONSIDERAÇÕES FINAIS

Como principal contribuição, a aplicação proposta neste trabalho pode ser considerada como uma interessante alternativa para o desenvolvimento da IHC, a qual, através de visão computacional, foi possível dispensar o uso de dispositivos menos acessíveis. Como trabalhos futuros destacam-se o reconhecimento automático das opções, sem a necessidade de uma máscara de treinamento, bem como a melhoria da aplicação para abranger ambientes diferentes do ambiente ideal, como por exemplo, projeções em su-

perfícies escuras ou com incidência de luz forte sobre a projeção. Outra melhoria à vista é que a aplicação permita a interação com a projeção independente do programa de exibição.

AGRADECIMENTOS

Ao programa PIBIC-COPES pela concessão da bolsa de Iniciação Científica e ao Departamento de Sistemas de Informação da Universidade Federal de Sergipe, Campus Universitário Prof. Alberto Carvalho, pelo apoio durante a execução do trabalho.

Referências

- AGGARWAL, J. K.; CAI, Q. **Human motion analysis: A review**. 1999.
- BAUDEL, M; BEAUDOUIN-L, T. Charade. **Communications of the ACM**, 7:28–35, 1993.
- DOYLE, J.; DEAN, T. **Strategic directions in artificial intelligence**. ACM Computing Surveys, 1996.
- FREEMAN, W.; ANDERSON, D.; BEARDSLEY, P. **Computer vision for interactive computer graphics**. IEEE Computer Graphics and Applications, 1998.
- HEWETT, T. et al. Chapter 2: Human-computer interaction. ACM SIGCHI Curricula for Human-Computer Interaction, 2009.
- MEIRELLES, F. S. 25ª pesquisa anual do uso de ti. FGV-EAESP-CIA, 2014.
- MISTRY, P; MAES, P. **Sixthsense: a wearable gestural interface**. Proceedings of SIGGRAPH Asia, Emerging Technologies. Yokohama, Japan, 2009.
- SYSTEMS, C. Cyberglove ii. Disponível em:
<http://www.cyberglovesystems.com> - Acesso em 24 de ago de 2014