

SHIVA: Realidade Aumentada Aplicada à infoinclusão de Portadores de Necessidades Especiais

Danilo Cunha¹, Flávio Souza¹, Marcelo Silva¹, Ulysses Potiguara¹, Alessandra Baganha¹

¹Curso de Bacharelado em Ciência da Computação - Área de Ciências Exatas e Tecnologia - Centro Universitário do Pará – Av. Gov. José Malcher, 1963 – CEP 66.060-230 – São Brás – Belém – Pará – Brasil.

{danilocunha85,ramonhell,marcelochaar,ulyssespotiguara}@gmail.com,
alessandra.natasha@cesupa.br

Abstract. *Based on the provision of input devices and data for Special Needs (NAP's) and the difficulty encountered by the class in using computers as well as the popularity of webcams on the market with its low cost of acquisition and handling, we developed a solution based the capture of positions supervised by an ordinary camera in order to increase interaction between users, especially PNE's, through tools based on the principles of interaction of Augmented Reality.*

Resumo. *Fundamentada na oferta de dispositivos de entrada de dados para Portadores de Necessidades Especiais (PNE's) e na dificuldade encontrada por essa categoria em utilizar computadores, assim como a popularidade das webcams no mercado – com seu baixo custo de aquisição e manuseio – foi desenvolvida uma solução baseada na captura supervisionada de posições por uma câmera comum com o objetivo de aumentar a interação dos usuários, principalmente PNE's, utilizando ainda princípios de interação da Realidade Aumentada.*

1. Motivação

O sistema de interação foi desenvolvido para a plataforma Windows®¹ e criado a partir do ambiente integrado de desenvolvimento Delphi®², sendo uma continuidade do trabalho proposto por [Cunha *et al.*, 2010]. Desde 1974, [Monteiro, 2007] surgimento do caráter pessoal atribuído aos então chamados microcomputadores, a microeletrônica passou a receber investimentos cada vez maiores que culminaram nas atuais tecnologias portáteis, citam-se: *palms*, *laptop*, celulares, até que a portabilidade começou a perder espaço, cedendo lugar para a mobilidade. Na definição de computação ubíqua, prevista por Mark Weiser já em 1991 [Coulouris, 2001], a mobilidade tecnológica, associada aos sistemas distribuídos, encontrou apoio e alta possibilidade de crescimento na Internet móvel, desta maneira inaugurando a era da computação em todo lugar.

Termos da atualidade, e frutos do avanço tecnológico de ponta, estão ainda, neste mesmo sentido, a realidade virtual, a realidade aumentada e a nanotecnologia. Entretanto, perante a todos os programas desenvolvidos e novos paradigmas propostos nas diversas subáreas do meio tecnológico, a forma de acesso usual aos sistemas

¹Windows® é marca registrada Microsoft Corp.

²Delphi® é marca registrada Inprise Corp.

computacionais permanece com a classificação apresentada desde 1974, quando os dispositivos para entrada de dados, como o teclado e o *mouse* eram (como ainda são) os meios mais utilizados para estabelecer a interface homem-máquina.

O objetivo desta pesquisa é apresentar um protótipo desenvolvido para possibilitar a interação virtual, por meio de uma câmera de vídeo comum de computador, propondo a adaptação do protótipo às necessidades do usuário. Deste modo, propõe-se um benefício para determinados tipos de portadores de necessidades especiais, visto que um *layout* diferenciado de teclado poderá ser adaptado pelo próprio usuário.

Ivan Laptev e Tony Lindeberg [Laptev e Lindeberg, 2000] conseguiram um sistema de reconhecimento de vários estados da mão baseado em uma hierarquia de imagens de multi escalas. Através de uma câmera o usuário interagiu com um programa que distinguia vários gestos da mão para efetuar diversas funções. Stiefelhagen [Stiefelhagen, 1997], com o auxílio de redes neurais, criou um sistema de rastreamento ocular em tempo real usando uma câmera que ficava localizada na parte inferior do monitor e com isso o usuário controlava um ponteiro na tela.

No presente estudo, será mostrado que, a partir da inspiração obtida no primeiro trabalho [Cunha, Silva, & Souza, 2008] e visando estendê-lo aos PNE's, bem como, consciente da carência de recursos financeiros para a obtenção e manipulação de uma câmera de alta resolução (maior quantidade de *pixel* por *frame* capturado), foi pesquisado o uso de contraste com tonalidade ajustável em relação a uma cor base para entrada de dados através de uma *webcam* e um teclado virtual, podendo este ser alterado para atender necessidades específicas dos diferentes tipos de usuários.

Dada uma superfície plana com o teclado pré-configurado do programa impresso, usuários com limitações, permanentes ou temporárias, nos membros superiores conseguem interagir com o computador, simulando assim a coordenação motora fina humana ao utilizar um teclado convencional. Basta que o usuário consiga produzir um contraste na área da tecla desejada, podendo esse contraste ser feito por meio de uma caneta *laser*.

Outro fator motivacional para esse trabalho baseou-se nos dados do IBGE nos quais, de um total de 34.580.721 brasileiros portadores de necessidades especiais, por volta de 14,5% da população nacional, segundo constam os levantamentos do Censo 2000, 22,96% sofrem de limitações por causa de problemas motores e 4,09% sofrem dificuldades físicas.

Segundo o Programa de ação mundial para as pessoas com deficiência, Resolução 37/52 da ONU, 1982, em virtude de deficiências mentais, físicas ou sensoriais, há no mundo mais de 500 milhões de pessoas com deficiência. A cifra estimada daquele ano vê-se confirmada pelos resultados de pesquisas referentes a diversos segmentos da população e pela observação de peritos. Na maior parte dos países, pelo menos uma em cada dez pessoas tem uma deficiência e a presença dessa deficiência repercute em pelo menos 25% de toda a população. Estima-se também que, no mínimo 350 milhões de pessoas com deficiência vivam em áreas de difícil acesso e que não dispõem dos serviços necessários para ajudá-las a superar as suas necessidades especiais. Grande parcela das pessoas com deficiência está exposta a barreiras físicas, culturais e sociais que constituem obstáculos à sua vida, mesmo quando dispõem de ajuda para a sua reabilitação e inclusão.

2. Módulos do Sistema

O sistema proposto no presente trabalho foi organizado em cinco módulos, com funções específicas, bem como diferenciadas com as finalidades futuras de reaproveitamento dos módulos em eventuais evoluções e atualizações do sistema, como prevêm as boas práticas da Engenharia de Software. São os 5 módulos: captura, processamento, ação, interface gráfica e teclados pré-definidos, conforme descritos a seguir.

2.1. Módulo de Captura

Responsável pela conexão entre o programa desenvolvido e a câmera que capta imagens em tempo real. No programa é utilizado um componente [MAsterC, 2006] que se encarrega dentre outras funções, principalmente em detectar a câmera, ou câmeras, e realizar as devidas operações para o uso dela. Cada *frame* obtido pela câmera é convertido para o formato *bitmap* que é repassado em seguida para o módulo de processamento.

É importante ressaltar que essa conversão é fundamental para o tratamento e manipulação dos *pixels* da imagem nos módulos posteriores a esse.

2.2. Módulo de Processamento

É o módulo encarregado de realizar modificações na imagem, ou seja, é nesse módulo que a imagem é tratada. O processamento da imagem consiste na captura dos contrastes, em preto e branco, e na conseqüente transformação da imagem em monocromática baseado na tolerância definida pelo usuário; inversão horizontal da imagem; inversão das cores preto e branco para captura em ambientes escuros; captura da cor secundária e o cálculo de sua posição em relação ao teclado virtual, feito por meio de um algoritmo que opera nas posições cartesianas X e Y.

2.3. Módulo de Ação

Nesse módulo, o programa recebe a variável que contém o estado das teclas e então aciona o evento correspondente a cada uma. Caso a tecla se encontre “pressionada”, simulando o movimento da coordenação motora fina, o contraste na área da tecla deverá ser computacionalmente equivalente a pressão de um dedo sobre um tecla do teclado convencional. Este módulo também é encarregado de efetuar a repetição do evento em um determinado intervalo de tempo, tal como aconteceria se determinada tecla, do teclado real, estivesse pressionada.

2.4. Interface Gráfica

A interface do modelo apresenta, de forma intuitiva e fundamentando-se em conceitos da Web 2.0 com o *layout* voltado para o conteúdo, a interação por meio de funções configuráveis do programa e a opção de retorno do módulo de processamento, conforme mostrado na figura 1, que demonstra a solução (ambiente gráfico) do teclado virtual para um usuário sem necessidades especiais.

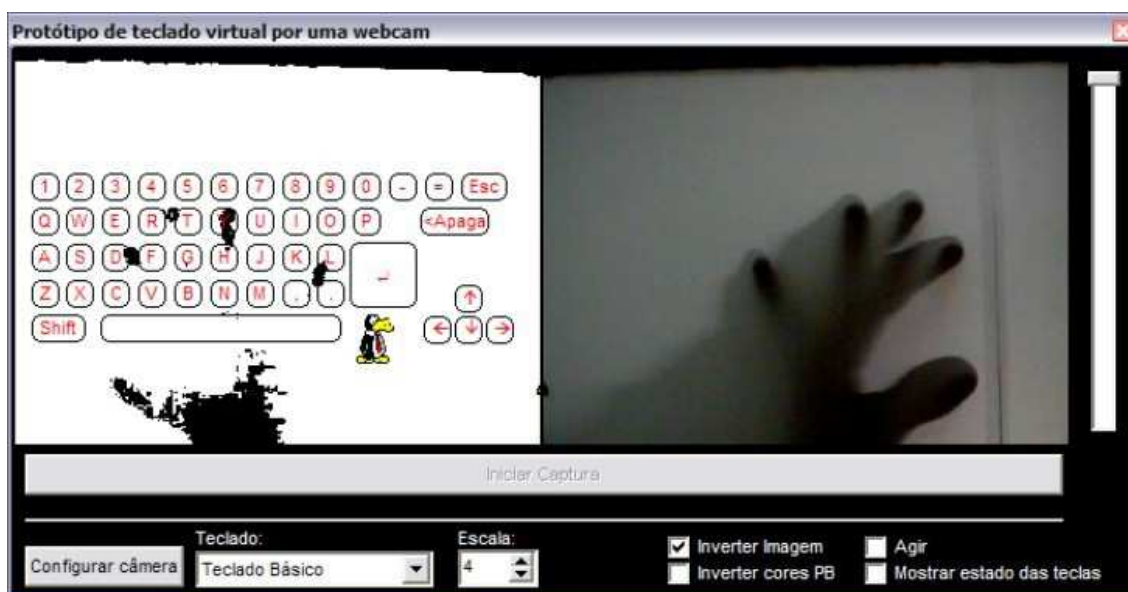


Figura 1. Ambiente Gráfico

2.5. Teclados Pré-Definidos

Este módulo contém as classes de teclados desenvolvidos e compilados junto ao protótipo, personalizáveis via edição do código-fonte. Este é o módulo que prevê futuras personalizações do sistema para adaptação em novos ambientes, como alterações na iluminação. Os tipos de teclados que foram desenvolvidos para este projeto foram: Teclado Básico (vide figura 1, no item 2.4 Interface Gráfica), Calculadora e Apresentação.

Como citado anteriormente na apresentação deste item 2, a organização por módulos, além de permitir futuras evoluções do sistema, também facilitou o desenvolvimento e a organização do código-fonte, permitindo que a manutenção seja feita, se necessário especificamente no módulo que se pretenda editar ou acrescentar algo, sem necessidade de alterações em todo o código.

O código-fonte apresenta-se organizado em um total de três classes (TPilha, TTeclado e Classe Principal), o que igualmente permitiu um maior compartilhamento do código entre os autores. Desta forma, foi possível torná-lo mais enxuto e agradável para entender “onde” e “como” cada módulo tem condições de atuar para contribuir com o sistema, facilitando e permitindo maior socialidade dos problemas pertinentes as implementações módulo a módulo.

3. Algoritmo de captura da imagem e posição

Como cada dispositivo de captura de imagem pode possuir distintos meios de resolução e ainda devido às possíveis variações de luminescência dos ambientes de uso do sistema, foi disponibilizado um artifício que permite ao usuário adaptar o programa às mais diversas qualidades de *webcams* e claridades de ambientes. Para tal, o usuário pode ajustar uma porcentagem de determinada cor primária, onde qualquer outra cor, fora deste limiar, será considerada como cor secundária e portanto desconsiderada, ou seja, o usuário define um ponto e os valores abaixo deste são cortados.

Deste modo, apenas duas cores distintas são captadas pelo programa, sendo que a

primeira é a cor neutra, que não dispara eventos, e a segunda o programa considera como ponteiro que, se detectada sobre a área de qualquer tecla, o evento associado é acionado, como se fosse pressionada a tecla correspondente no teclado, vide detalhes na figura 2 com o ambiente gráfico e o ambiente físico, além do usuário que adaptou a caneta *laser* ao pé, devido à impossibilidade de coordená-la com outra parte do corpo.



Figura 2. Usuário interagindo com o ambiente gráfico e o físico

Tratada a forma de acionamento da tecla, a partir do *layout* do teclado foi feito então com que o programa diferenciasse contrastes nesta captura com uma certa tolerância definida, sendo que apenas o preto e o branco seriam utilizados. É definida uma cor pela captura para servir de fundo e então o algoritmo, através da tolerância, determina o que é branco ou preto, sendo que cores mais próximas ao fundo são consideradas brancas. Detalhes na figura 1, anteriormente mostrada, mostram que na imagem processada, a cor de fundo definida foi o branco, e o preto está acionando as funções das teclas.

Com o teclado mapeado virtualmente pelo programa e a imagem capturada pela câmera processada de forma monocromática, é realizada a varredura detectando as áreas escuras. Escaneando as posições das teclas e suas respectivas áreas, são confrontadas essas informações com a posição cartesiana dos pontos escuros obtidos anteriormente e incrementa-se 1 (unidade de pixel) a cada ponto escuro encontrado numa variável. Se após terminar a varredura, a percentagem de pontos escuros na área da tecla atingir dada percentagem, o protótipo dispara o evento correspondente da tecla. A figura 3 demonstra, no detalhe, a invasão do *laser* na área de uma tecla do *layout* impresso, a caneta *laser* e a *webcam*.



Figura 3. Detecção da luz *laser* invadindo a área de uma tecla

4. O Teclado Virtual

Diferentemente do dispositivo feito com uma borda de compensado fino e papel vegetal proposto por Cunha *et. al* (2010), este trabalho capta a imagem diretamente do *layout* impresso via câmera que fica direcionada de frente para a folha, enquanto isso o teclado é mapeado virtualmente pelo protótipo, como pode ser observado na Figura 2, mostrada anteriormente no item 3.

O feixe de luz laser é projetado no papel e capturado pela câmera enquanto que o algoritmo converte-o em informações como “pressionado” e “liberado”, para cada tecla mapeada na posição onde a sombra foi projetada. Então, dependendo da ação, um evento pré-configurado e associado à área é disparado.

Com isso, é possível a criação de novas teclas com funções especiais, definidas pelo usuário, podendo ser reajustado de acordo com suas necessidades, como por exemplo, automatizar o endereço residencial ou o correio eletrônico. O teclado virtual também permitirá a total modificação de seu *layout*, ajustando-se tanto ao usuário comum quanto, e principalmente, ao usuário especial.

Na Figura 1, já mostrada no sub-item 2.4, tem-se um dos teclados pré-configurados e o algoritmo de captura da imagem trabalhando na imagem original, de forma divergente do proposto em [Cunha *et al.*, 2010] no qual a captura é feita linha a linha de pixel e não mais pixel a pixel, o que possibilita a interação em tempo real, diminuindo consideravelmente o tempo de resposta. A câmera fica localizada em frente ao protótipo e o usuário pode interagir projetando o *laser* sobre o papel, do lado oposto. O espelhamento da imagem, neste caso, é necessário para que não haja confusão nos movimentos realizados em frente à câmera, já que, posicionada em frente ao protótipo, capta uma imagem invertida em relação ao usuário.

5. Considerações Finais

No início do projeto, a pesquisa por trabalho que se assemelhassem a este nos levou ao encontro dos trabalhos de [Laptev e Lindeberg, 2000] e [Stiefelhagen, 1997]. Ao estudar os prós e contras de cada uma das pesquisas citadas, resolveu-se optar por um meio termo dentre estes. Já que uma câmera de alta definição como a que foi utilizada por Stiefelhagen (1997), não seria adequada no contexto da pesquisa apresentada neste artigo. Ao mesmo tempo, Laptev e Lindeberg conseguiram reconhecer vários estados das mãos baseando-se em hierarquia de imagens, entretanto essa técnica normalmente

não se adequaria ao contexto da presente pesquisa por exigir mais recursos computacionais.

Sendo assim, foram estudadas distintas formas de interação homem-máquina e encontrou-se na Realidade Aumentada – RA e no Processamento Digital de Imagem – PDI uma possível solução para a inclusão de portadores de necessidades especiais.

A RA é uma linha de pesquisa da grande área da ciência da computação que prevê a integração de elementos virtuais com o mundo real através do processamento *real-time* de imagens capturadas por câmeras ou vídeos [Azuma, 1997]. Utilizando parte das técnicas estudadas em RA, como por exemplo: combinação de elementos virtuais com o ambiente real; interação e processamento em tempo real possibilitando adequar o sistema às necessidades dos usuários. Porém, para isto, foi deixada de fora da pesquisa, bem como do desenvolvimento, a técnica de criação de objetos em três dimensões, assim evitando o aumento do custo computacional [Gonzalez, 2000].

A pesquisa seguiu estes limites entre os dois trabalhos [Azuma, 1997] e [Gonzalez, 2000], o que permite o início de uma nova pesquisa para estudar-se novas formas de interação homem-máquina (exemplo: interação por voz), processamento digital imagem em tempo real (exemplo: algoritmos de segmentação de imagem, reconhecimento de padrão em imagem, clusterização de imagem) e inclusão.

5.1. Relatos de Experiências Relevantes

A experiência mais relevante para esta pesquisa foi o momento em que iniciou-se a fase de testes do *software*, encontrou-se algumas dificuldades para ter acesso aos hospitais públicos de Belém – PA, tanto na escala municipal quanto na estadual. Devido a isso, partiu-se para uma clínica onde as crianças passavam apenas por atividades de lazer, fisioterapia e educacionais.

Os testes foram considerados satisfatórios pois, os problemas encontrados para a utilização do sistema eram relacionados com os pré-requisitos, exemplo: a caneta *laser* ficava com a bateria fraca, o *layout* do teclado impresso era pequeno demais e o feixe *laser* acionava mais de uma tecla ao mesmo tempo. Problemas estes que foram contornados sem a edição do código-fonte e sendo assim, permitindo maior tempo dedicado aos testes com os usuários.

5.2. Estudos de Caso

Os usuários demonstraram facilidade em interagir com o sistema, devido ao fato de poderem fazer uso de *layout* de teclados impressos de tamanhos variados, além da simplicidade destes, o que possibilita o aumento ou a diminuição da área de cada tecla, dependendo da habilidade que o usuário já possua ou tenha adquirido utilizando o sistema por algum tempo.

Quanto maior o tempo dedicado dos usuários ao utilizar e conhecer o sistema, mais à vontade para questionar o tamanho das teclas e a proximidade entre elas. Tais tipos de questionamentos foram solucionados imprimindo *layouts* de outros tamanhos. Notou-se também a dificuldade na confecção de textos pois, os usuários só conseguem coordenar uma caneta *laser* de cada vez. Para automatizar esse uso do teclado, poderia ser aplicado o uso do modelo de teclado T9 (Nuance Communications) e a sua tecnologia preditiva de texto, encontrada em vários modelos de telefones celulares.

Por fim, percebeu-se que ao movimentar a *webcam*, mesmo que sutilmente, o processo de configuração do ambiente deveria ser retomado. Isso causava certo desconforto nos usuários pois, a configuração do *layout* com a *webcam* exige ajustes finos e precisos. Logo, era feito por outra pessoa, que fazia o papel de usuário de apoio. Sendo assim, uma possível solução seria o uso de redes neurais aplicadas ao reconhecimento de padrões, ou seja, a rede neural artificial reconheceria o padrão e faria a calibragem automaticamente, dispensando a presença do usuário de apoio e a dependência causada ao usuário final.

6. Conclusão

A programação voltada para o desenvolvimento de sistemas robustos com enfoque na interatividade humano-computador deve receber cada vez mais a atenção e cuidados das equipes de análise, projeto, desenvolvimento e manutenção de software nas empresas especializadas.

O sistema desenvolvido foi baseado no princípio de que a interatividade é um dos principais elementos que devem ser levados em consideração ao optar-se por dispositivos de entrada e saída de um sistema computacional.

O modelo prevê a utilização de sistemas de entrada de dados baseados nas populares *webcams* de baixo custo focando nos benefícios das pessoas com necessidades especiais e propondo uma solução flexível e modular em software para uma limitação imposta pelo hardware. É mais conveniente, do ponto de vista financeiro, configurar um sistema a um usuário do que adquirir diferentes hardwares para cada tipo de usuário com algo tipo de necessidade especial.

Como proposta de trabalhos futuros ficam o aperfeiçoamento do protótipo, com o uso de interfaces tangíveis adequadas e especializadas para diversas outras finalidades. Para isso será necessário uma nova pesquisa dentro da área de Computação Natural - ciência que estuda modelos inspirados na natureza para resolver problemas complexos, como por exemplo o reconhecimento de padrão em imagens.

A pesquisa demonstrou que com baixo custo financeiro e poder computacional, é possível o aperfeiçoamento dos meios de interação homem-máquina utilizando a Realidade Aumentada para isto e deixando o usuário à vontade para adaptar/desenvolver a própria ferramenta de trabalho. Devido a possibilidade de adaptação do programa, usuários portadores de necessidades especiais puderam usufruir de maneira não apenas mais confortável, mas também mais eficaz dos computadores e em alguns casos mantiveram seu primeiro contato com a máquina e seus softwares.

Referências

- Shi, J., Malik, J. 2000. Normalized Cuts and Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell.
- Cunha, D. S., Silva, M. C., & Souza, F. R. (Dezembro de 2008). Teclado virtual : uma solução para inclusão digital. Belém, PA, Brasil.
- Azuma, R.T., (1997) “A Survey of Augmented Reality”. In Presence: Teleoperators and Virtual Environments.
- Laptev, I.; Lindeberg, T. (2000). “Tracking of multistate hand models using particle filtering and a hierarchy of multiscale image features”, Stockholm, Suécia.
- Stiefekhagen, R.; Yang, J.; Waibel, A. (1997). “Tracking Eyes and Monitoring Eye Gaze”, Universidade de Karlsruhe, Alemanha; Universidade Carnegie Mellon, EUA.
- Coulouris, G.; Dollimore, J.; Kinderg, T. Distributed systems: concepts and design. 3. ed. Massachusetts: Addison-Wesley, 2001.
- Monteiro, M. A. Introdução à organização de computadores. 4. ed. Rio de Janeiro: LTC, 2002.
- MAsterC (2006). <<http://www.delphifr.com/membresite.aspx?ID=4058>>. Capturado em 27/08/2006.
- Planeta Delphi (2006). <<http://www.delphi.eti.br/>>. Capturado em 01/06/2006.
- Gonzalez, R. C. Processamento De Imagens Digitais. EDGARD BLUCHER. 2000. 1ed.
- Nuance Communications. (s.d.). *Nuance*. Acesso em 21 de 02 de 2011, disponível em Nuance: http://www.nuance.com/news/pressreleases/2007/20070824_tegic.asp
- CUNHA, D. S., SOUZA, F. R. A., SILVA, M. C., POTIGUARA, U. J. C., BAGANHA, A. SHIVA: Dispositivo de Entrada de Dados e Interação Virtual por Webcam para Portadores de Necessidades Especiais In: COMPUTER ON THE BEACH, 2010, Florianópolis.

