

## INTERFACES NATURAIS DE USUÁRIO PARA UMA APLICAÇÃO DE TREINAMENTO INDUSTRIAL UTILIZANDO A CÂMERA REALSENSE

Dr. Eduardo Manuel de Freitas Jorge<sup>1</sup>, Msc. Fábio Britto de Carvalho Almeida<sup>1</sup> e Yago Souza Venancio Lima<sup>1</sup>

<sup>1</sup>CIMATEC/SENAI

E-mails: [eduardo.jorge@fiab.org.br](mailto:eduardo.jorge@fiab.org.br), [falmeida@fiab.org.br](mailto:falmeida@fiab.org.br), [yagosvl@hotmail.com](mailto:yagosvl@hotmail.com)

### ABSTRACT

*This article pretends to specify a training application for various industrial contexts through the prototyping technique using concepts of Natural User Interface and RealSense sensor to enhance the user's immersion. In industries that have high turnover of staff or danger sectors such initiatives are crucial to keep the technical quality and cut costs.*

**Keywords:** RealSense, Human Computer Interface, Natural Interface, Industrial Training

### RESUMO

*Este artigo pretende especificar uma aplicação de treinamento para diversos contextos industriais através da técnica de prototipação utilizando conceitos de Interface Natural de Usuário e o sensor RealSense de modo a potencializar a imersão do usuário. Em indústrias que tem alta rotatividade de funcionário ou setores de periculosidade essas iniciativas são fundamentais para a manter a qualidade técnica e cortar gastos.*

**Palavras-chaves:** RealSense, Interface Humano-Computador, Interfaces Naturais, Treinamento Industrial

### 1. INTRODUÇÃO

Entre os vários marcos da história da computação, o surgimento das interfaces gráficas, em meados da década de 80, foi um importante marco [1]. Com isso, houve uma grande expansão dos computadores nos ambientes domésticos, popularizando-os em todo o mundo. À medida que o número de usuários foi aumentando, juntamente com o avanço tecnológico, novas formas de interação foram surgindo e se aprimorando. Um desses avanços significativos foi o surgimento do *touchscreen*, que revolucionou a interação humano-computador, através do

toque dos dedos. Novos conceitos começaram a surgir como NUI (Interfaces Naturais de Usuário) e Computação Perceptiva. Ambos conceitos defendem a criação de experiências nas quais os usuários possam ter a sensação de estarem interligados aos sistemas computacionais através de meios naturais de comunicação, como gestos e/ou vozes [1]. Diante deste contexto, e com às necessidades crescentes dos usuários de se comunicarem de uma forma mais natural com os computadores, foram desenvolvidas várias tecnologias, entre elas câmeras como o Kinect (2010), e mais recentemente a Intel RealSense 3D (2012). Estas câmeras trazem consigo um conjunto de possibilidades em interações de interfaces naturais. O Kinect possui funcionalidades como reconhecimento de voz e de gestos[2]. Apesar de ter sido projetado inicialmente para proporcionar uma melhor interação para os usuários do console Xbox da Microsoft, sua utilização para outros tipos de aplicações vêm gerando várias oportunidades em outros domínios de negócios[2]. Além do Kinect, a Intel também desenvolveu um sensor, a Intel RealSense 3D que, foi criada já visando a sua instalação em diversos aparelhos que vão de computadores, *All in one*, notebooks, Dois em Um, tablets e smartphones. Os ambientes são tão diversos que a finalidade da RealSense se torna quase uma questão criativa. Por conta disso, a Intel vem fomentando a utilização de sua câmera, através de competições entre desenvolvedores, com o intuito de criar um novo nicho de aplicações que proporcionem novas experiências aos usuários[3]. A câmera possui funcionalidades como: rastreamento dos dedos e das mãos, reconhecimento facial, rastreamento do rosto, reconhecimento de voz, realidade aumentada e etc.

O desafio agora é se apropriar das funções desses dispositivos para criar ambientes que respondam melhor os usuários, se beneficiando ao máximo dessas tecnologias.

Um domínio que poderia se beneficiar de interações mais naturais são as aplicações de treinamento. Existem seguimentos empresariais cuja rotatividade de empregados é muito alta. O treinamento desses empregados recém ingressos pode ser custoso pelo volume de pessoas que precisa ser treinado, pelo tempo que esse treinamento pode exigir ou pela alocação de funcionários mais experientes para supervisionar o treino. [4] enfatiza tanto as vantagens cognitivas quanto as econômicas de treinamento simulado, mas cria um ambiente cuja interação se dá via mouse e teclado.

Nesse contexto, o objetivo desse projeto é especificar uma aplicação de treinamento para diversos contextos industriais através da técnica de prototipação utilizando conceitos NUI atendidos pela câmera RealSense de modo a potencializar a imersão do usuário. Para isso, será apresentado na seção 3 um protótipo de baixa fidelidade de interface atendendo aos requisitos de uma aplicação de treinamento industrial e com as formas de interação *Hover Select*, *Swipe* e *Move* [3]. Além disso, apresentam-se os algoritmos e técnicas computacionais que compõem o estado da técnica desta solução computacional. Para a realização desta pesquisa a metodologia adotada foi baseada fortemente na prototipagem rápida e iterações cíclicas e evolucionárias.

Com uma base de dados é possível apresentar diversas das ferramentas de trabalho para serem manipuladas por gestos já mapeados pelo sensor RealSense. Além disso, distribuindo essas

ferramentas em contextos bem discriminados, teremos aprendizagem mais consolidada e de uma maneira mais interessante do que uma simples lida no manual.

Com o surgimento de novas tecnologias, ferramentas, processos frente a diversos outros que já existem, os funcionários de uma indústria, independente do setor, precisam conhecer uma enorme gama de equipamentos. Além disso, nos casos de novas contratações ou mudanças de equipamento ou processos é necessário um desprendimento de tempo e gente para treinar aqueles funcionários novatos. Este trabalho permitirá treiná-los de uma maneira imersiva, sem grandes gastos a curto prazo uma vez que as câmeras RealSense serão embutidas nos mais diversos dispositivos laptops, Dois em Um, tablets e derivados.

Durante a pesquisa, foram estudados os conceitos e boas práticas de Interfaces Homem-Computador, explicadas na seção 2 e 2.1, que nortearam o presente estudo. Em seguida, um estudo sobre os sensores que auxiliassem nessas boas práticas. O sensor RealSense será bem detalhado na seção 2.2. Finalmente, a seção 3 apresenta a aplicação, suas telas de protótipo, os gestos que serão empregados em cada momento da aplicação. Por último, a seção 4 traz a conclusão do trabalho.

## **2. INTERFACE HUMANO COMPUTADOR: INTERAÇÕES NATURAIS ATRAVÉS DE CÂMERAS**

Segundo [5], o termo interface é comumente utilizado para denominar aquilo que interliga dois sistemas. A Interação Homem-Computador (IHC) é a área da computação que investiga o design, avalia e implementa interfaces para que seres humanos possam interagir com sistemas computacionais de maneira eficiente e intuitiva [6]. Já segundo a Computer Science and Engineering o objetivo da IHC é o desenvolvimento e implantação de interfaces “amigáveis”, *user-friendly*, permitindo ao utilizador realizar as suas tarefas com segurança, eficácia, eficiência e satisfação, ou seja, originando interfaces com elevado grau de usabilidade.

Devido aos constantes avanços tecnológicos, as interfaces vêm sofrendo significativas mudanças ao longo dos anos, buscando sempre se adequar às novas demandas dos usuários [7]. Segundo [8] pode-se definir três tipos principais de interfaces Homem-Computador: Linhas de Comando (CLI), Interfaces Gráficas (GUI) e Interfaces Naturais (NUI). A CLI caracteriza-se por possuir um modelo de interação baseado em uma série de transações de pedidos e respostas, expressos em um vocabulário especializado. São interfaces muito ágeis, porém pouco intuitivas. A GUI surgiu com a necessidade da manipulação de objetos na tela. É o tipo de interface mais utilizado até os dias de hoje, principalmente pelos seus conceitos de janelas, ícones, ponteiros e menus (WIMP). Já a NUI caracteriza-se por interfaces voltadas para a captação e interpretação de comandos naturais.

### **2.1 Interface Natural**

As novas formas de interação modificaram a forma de pensar interfaces digitais e abriram espaço para discutir novos paradigmas, como o de Interfaces Naturais de Usuários (NUI)

afirmam [1]. Segundo [8] o futuro da IHC está na utilização das interações naturais. De maneira discreta, os computadores estarão por toda a parte, relógios, carteiras, óculos, sendo operados naturalmente, chegando então a real integração entre ser humano e computador [8].

As NUI são baseadas em captura e reconhecimento de gestos através de algoritmos matemáticos, que recebem as informações dos sensores, as interpreta e as converte para um formato reconhecido pelos computadores [1]. Segundo [9], um gesto pode ser considerado como qualquer movimento físico detectado através de sensores por um sistema digital, ao qual poderá responder sem o auxílio de mecanismos tradicionais, como mouses ou canetas específicas. Ainda de acordo com [9], as interfaces gestuais podem ser classificadas em: Touchscreen - Interface onde o usuário pode se comunicar através do toque na tela com os dedos; Interações de Forma Livre - Permitem gestos livres sem estar em contato com nenhuma superfície, ou seja o corpo humano é o dispositivo de entrada.

Neste artigo utiliza-se os gestos livres como fonte de entrada de dados. Embora a tecnologia *Touchscreen* já esteja estabelecida a alguns anos, e tenha aberto várias oportunidades na IHC, são as interações de forma livre, através das novas tecnologias perceptivas como o Kinect e a Intel RealSense 3D, que prometem alavancar este tipo de interfaces a um novo nível. As tecnologias perceptivas nos permitem elaborar interações cada vez mais humanas e intuitivas, aumentando o grau de imersão dos usuários e consequentemente de usabilidade e satisfação dos mesmos.

## **2.2 Tecnologias Perceptivas**

As tecnologias perceptivas possuem um conjunto de hardwares e softwares que lhes possibilitam as seguintes percepções: reconhecimento de gestos, reconhecimento facial, reconhecimento de voz e reconhecimento de toque. Com base nestas percepções os computadores têm acesso a uma detecção sensorial mais detalhada dos humanos, o que possibilita a criação de vários tipos de aplicações voltadas para as Interfaces de Interação Natural. Para esta pesquisa, analisa-se o dispositivo Intel RealSense 3D.

### **2.2.1 Intel RealSense 3D**

O RealSense não é nenhum equipamento tecnológico específico, e sim o nome dado pela Intel a um grupo de avanços em hardware e software em computação de percepção. O primeiro produto desse grupo é uma câmera 3D que permite a interação com seus dispositivos de uma forma mais natural [10]. Durante o ano de 2014, empresas como Acer, Asus, Dell, Fujitsu, HP, Lenovo e NEC começaram a integrar a nova câmera da Intel RealSense 3D em seus laptops. As novas câmeras serão integradas dentro de notebooks, tablets e dispositivos de dois-em-um, e provavelmente residirá no topo, onde fica a webcam atualmente [11].

A Intel vem realizando parcerias com empresas como a Microsoft, a Google e a 3DSystems Assistente Nuance Dragão Digital, para a utilização de recursos como: o Voip do Skype, o Street View e o reconhecedor de voz Dragon Naturally Speaking [10] [11]. Segundo a Intel, a primeira câmera Intel RealSense 3D conta com o melhor sensor de profundidade do

segmento. O aparelho possui a habilidade de detectar os movimentos dos dedos, o que habilita o reconhecimento altamente preciso de gestos e expressões faciais para compreender o movimento e as emoções. A câmera pode compreender o que está no fundo e à frente para permitir o controle, melhorar a realidade aumentada interativa, simplificar a varredura de itens em 3 dimensões e mais [12]. A Intel RealSense 3D possui uma câmera colorida RGB FullHD 1080p e dois microfones integrados (intel1). Ver figura 1.



Figura 1: Intel RealSense 3D – (Intel)

Esta câmera engloba várias categorias. Na categoria mãos, ela é capaz de rastrear as mãos e os dedos e reconhecer gestos estáticos e dinâmicos, tudo através do reconhecimento de 22 pontos de cada mão [3].

Na categoria face, ela é capaz de identificação facial, rastreamento da cabeça e do rosto, detecção de pontos de referência, detecção de emoção, tudo através da identificação de 78 pontos no rosto [3].

Na figura 2, pode-se observar ainda a tolerância a certos movimentos de rotação nos três eixos do rosto.

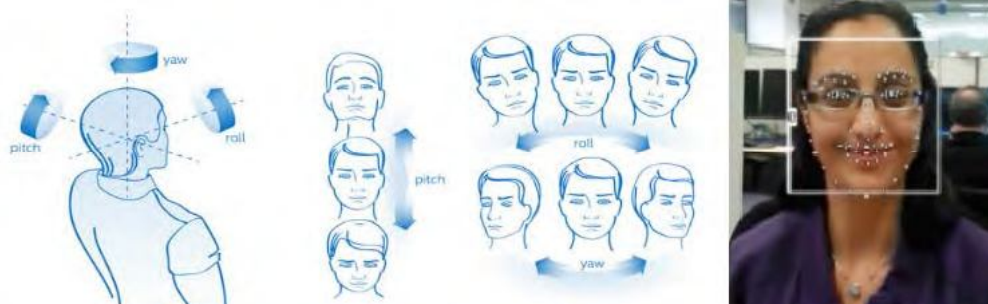


Figura 2: Rastreamento de cabeça - (Intel)

Ainda que o usuário incline (pitch) a cabeça para cima e para baixo, rotacione (roll) o rosto para direita ou para a esquerda, ou gire (yaw) o pescoço o algoritmo de reconhecimento facial ainda é capaz de reconhecer o rosto.

Na categoria voz, ela é capaz de reconhecer e executar comandos via voz, de converter texto em voz e voz em texto. Na categoria ambiente, ela é capaz de proporcionar realidade aumentada, percepção de cenas, modelar objetos 3D e imprimir objetos 3D (intel1).

Com todos estes recursos a Intel RealSense 3D possibilitará o desenvolvimento de diversas aplicações em várias áreas como jogos, colaboração imersiva, entretenimento educacional, aprimoramento de chamadas de videoconferências e etc.

### 3. APLICAÇÃO DE TREINAMENTO INDUSTRIAL USANDO NUI

Nessa pesquisa será apresentado a modelagem de um sistema que funcionará como repositório de contextos e atividades industriais. Para cada contexto, existirão um grupo de objetos associados relevantes ao contexto. Por exemplo, em se tratando do contexto Segurança do Trabalho, objetos de proteção como capacetes, luvas e óculos serão relevantes enquanto que ferramentas de trabalho como chaves de fenda, alicates não. Mas para o contexto de Manutenção de Circuitos, a lista de objetos de interesse citada anteriormente, poderia se inverter.

Então, para começar o treinamento, o primeiro passo seria escolher o contexto a ser praticado. A seguir, a figura 3 traz o protótipo de baixa fidelidade da tela de contextos.

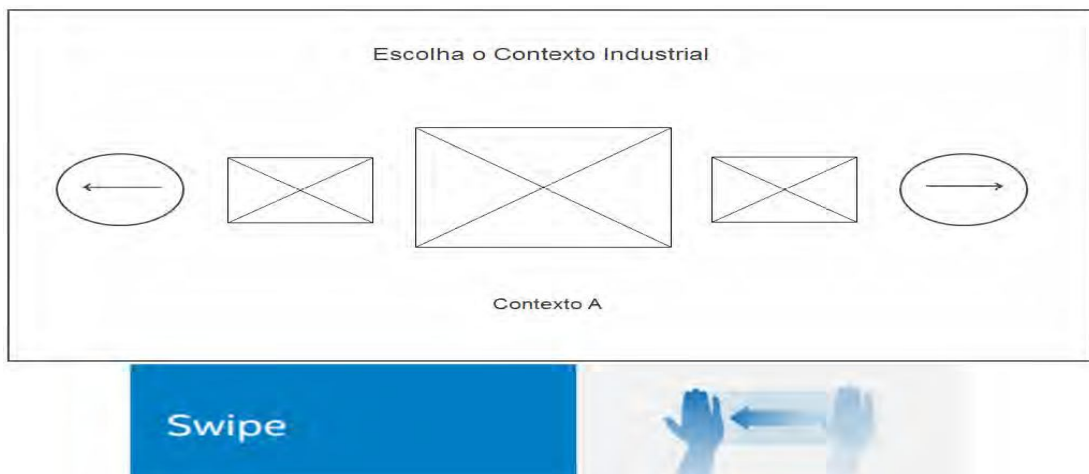


Figura 3: Protótipo do menu da aplicação.

Cada retângulo representa uma imagem que identifica o contexto, seguido pelo nome do contexto abaixo. Bastaria que o usuário percorresse a lista circular com gestos de *Swipe* já programados na câmera e escolhesse o contexto de seu interesse. O *Swipe* é um gesto que o usuário move o braço com a mão aberta, como se estivesse passando as páginas de um livro.

Uma vez escolhido o contexto, o usuário seria levado para uma segunda tela onde a partida se iniciaria de fato. Em um ambiente aberto, vários objetos aleatoriamente seriam apresentados ao usuário que deve capturar e arrastar para os devidos locais na tela, aqueles relevantes ao contexto escolhido antes que o tempo acabe. Novamente, a figura 4 traz o conceito da tela descrita.

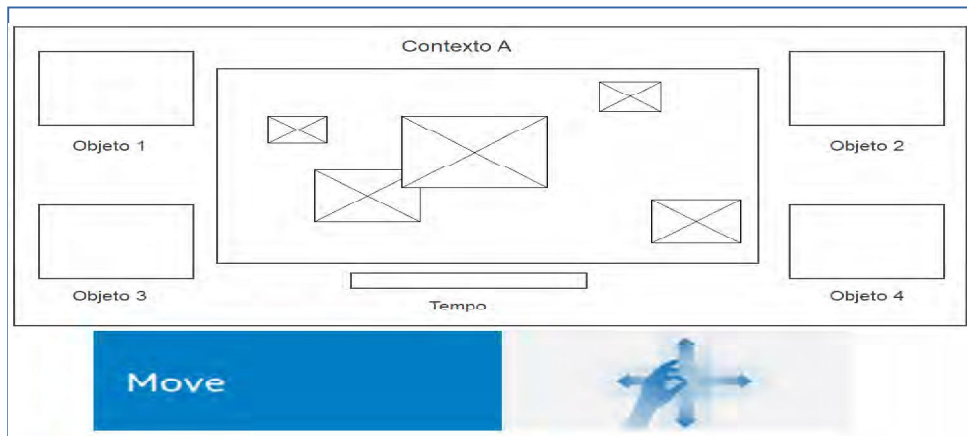


Figura 4: Protótipo da tela da partida.

Essa tela contaria ainda com uma função de alterar a perspectiva dos objetos apresentados a partir do reconhecimento facial do usuário, explicado na seção 2.2.1. Quando o usuário realizar algum movimento com o rosto, o foco da tela seguirá esse movimento para dar diferentes pontos de observação a quem estiver sendo treinado. Esse método foi aplicado para aumentar a inserção do participante.

Ao final do tempo, o usuário será levado a uma tela com os objetos coletados, uma explicação de seus usos ensinando a importância daquele item para o contexto escolhido. A tela descrita está representada na figura 5. Para obter informações dos objetos recuperados, o gesto *Hover Select* será aplicado. Esse gesto consiste em deixar a mão também aberta por um curto espaço de tempo em cima da imagem que for de interesse.

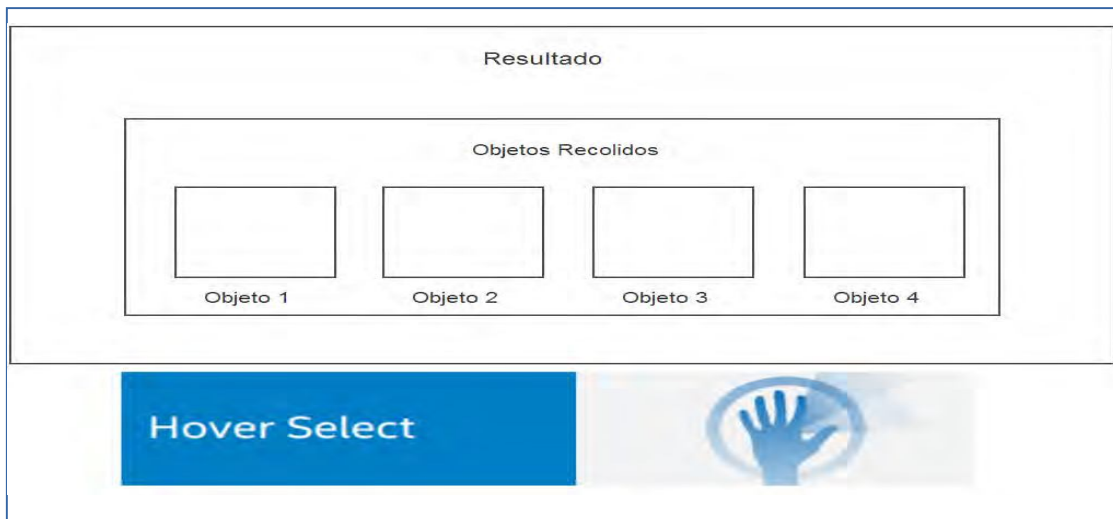


Figura 5: Protótipo da tela de resultados da aplicação.

A escolha da utilização da RealSense traz algumas exigências de hardware específicas. Uma vez que se trata de um produto novo o *Software Development Kit* (SDK) só está disponível para o Sistema Operacional Windows. Além disso, por conta do grande volume de processamento de frames por segundo capturados pela câmera, a comunicação do computador com o dispositivo deve se dar através de um USB 2.0 e a máquina ainda deve possuir um processador de Segunda geração ou superior.

Por outro lado, é possível desenvolver aplicações que utilizem a câmera RealSense em diferentes linguagens. Java (Java Developer Kit 1.7.0\_11 ou superior), Java Script, C# (Microsoft .NET Framework 4.0), C++ e Processing (2.1.2 ou superior) são suportadas. Para o presente projeto, foi escolhido a linguagem C# por afinidade dos desenvolvedores. Para um exemplo do código, ver a figura 6.



```
// Create the SenseManager instance
PXCMSenseManager sm=PXCMSenseManager.CreateInstance();

// Enable face tracking
sm.EnableFace();

// Get a face instance here (or inside the AcquireFrame/ReleaseFrame loop) for configuration
PXCMFaceModule face=sm.QueryFace();
...

// Initialize the pipeline
sm.Init();

// Stream data
while (sm.AcquireFrame(true)>=pxcmStatus.PXCM_STATUS_NO_ERROR) {
    // retrieve the face tracking results
    PXCMFaceModule face2=sm.QueryFace();
    if (sm!=null) {
        ...
    }

    // Resume next frame processing
    sm.ReleaseFrame();
}

// Clean up
sm.Dispose();
```

Figura 5: Exemplo de implementação de rastreamento de face em C# com a RealSense.

#### 4. CONCLUSÃO

O uso de tecnologias perceptivas para criação de interações Naturais é um caminho evolutivo natural assim como as interfaces gráficas avançaram a partir da entrada via teclado. É de fundamental importância o estudo desses dispositivos para que se possa extrair o máximo de suas respostas de maneira harmoniosa seja qual for o ambiente de interação com o computador.

No caso da aplicação, seu intuito é ser um repositório de contextos industriais simples mas imersivo. A interação com o sistema a partir de gestos foi pensada para torná-la mais atraente e memorável. Se valendo ainda da vantagem de que no futuro a tecnologia RealSense virá embutido em diversos dispositivos, notebooks, tablets, dois em um, e celulares. Como perspectivas futuras esta aplicação será evoluída através de um processo de validação da aplicação, onde serão feitas entrevistas sob a forma de questionários com grupos de funcionários, visando medir a satisfação e a usabilidade dos gestos escolhidos.

#### REFERÊNCIAS

<sup>1</sup> CABREIRA Arthur e MULLING Tobias, 2012. Perspectivas para novas interfaces: Kinect e interações gestuais sob o panorama de interfaces naturais do usuário. IV Congresso Sul Americano de Design de Interação.

<sup>2</sup>MICROSOFT. Kinect [online]. Disponível em: <http://www.xbox.com/en-US/kinect>. [Acessado Outubro 2014].

<sup>3</sup> INTEL. 2012a. Intel RealSense Technology [online]. Disponível em: <https://software.intel.com/en-us/articles/realsense-overview>. [Acessado Outubro 2014].

<sup>4</sup>SOUSA Macos Paulo Alves de, PAMPLONA Alcides Renato da Silva Junior RIBEIRO Manoel Filho e REIS Felipe Vaz. 2008. Treinamento de Montagem e Manutenção em uma Unidade Hidroelétrica de Energia Usando Realidade Virtual Desktop. Revista IEEE América Lantina, 6, Set 2008.

<sup>5</sup>NORMAN Donald A. e DRAPER Stephen W. User centered system design. New Perspectives on Human-Computer Interaction, L. Erlbaum Associates Inc., Hillsdale, NJ, 1986.21.

<sup>6</sup>SANTOS Sérgio Leandro e TEIXEIRA F.G. 2010. Design de uma interface de interação tridimensional com foco na usabilidade e no desempenho gráfico. PgDesing, pages 39-50..

<sup>7</sup>AFONSO Ana Paula, LIMA Jorge Reis e COTA Manuel Perez. Uma avaliação heurística da usabilidade de interfaces web [online]. Disponível em: <http://article.sapub.org/10.5923.j.computer.20120001.04.html#Sec2.1>. [Acessado Outubro 2014].

<sup>8</sup>OLIVEIRA Igor Aguiar. Interface de usuário: A interação homem-computador através dos tempos. Olhar Científico, 1(2):178-184, 2011.

<sup>9</sup>SAFFER Dan. 2008. Designing gestural interfaces: Touchscreens and interactive devices. "O'Reilly Media, Inc."

<sup>10</sup>BARRETT Brian. Realsense: como a intel quer transformar o mundo inteiro em um holodeck [online]. Disponível em: <http://gizmodo.uol.com.br/intel-realsense/>. [Acessado Outubro 2014].

<sup>11</sup>TUDO SOBRE TECNOLOGIA. RealSense da Intel é produtiva e divertida [online]. Disponível em: <http://www.tudosobretecnologia.com.br/2014/01/realsense-da-intel-e-produtiva-e.html> [Acessado Outubro 2014].

<sup>12</sup>DIÁRIO DE PERNAMBUCO. Intel Revela Módulo Integrado de Câmera de Profundidade e Assistente Pessoal de Linguagem Natural [online]. Disponível em: [http://www.diariodepernambuco.com.br/app/noticia/tecnologia/2014/01/08/interna\\_tecnologia\\_483329/intel-revela-modulo-integrado-de-camera-de-profundidade-e-assistente-pessoal-de-linguagem-natural.shtml](http://www.diariodepernambuco.com.br/app/noticia/tecnologia/2014/01/08/interna_tecnologia_483329/intel-revela-modulo-integrado-de-camera-de-profundidade-e-assistente-pessoal-de-linguagem-natural.shtml). [Acessado Outubro 2014].