

# CONIC·SEMESP

## 13º Congresso Nacional de Iniciação Científica

Anais do Conic-Semesp. Volume 1, 2013 - Faculdade Anhanguera de Campinas - Unidade 3. ISSN 2357-8904

**TÍTULO:** ESTUDO DO USO DO KINECT PARA INTERPRETAÇÃO DE GESTOS VISANDO LIBRAS

**CATEGORIA:** CONCLUÍDO

**ÁREA:** ENGENHARIAS E TECNOLOGIAS

**SUBÁREA:** COMPUTAÇÃO E INFORMÁTICA

**INSTITUIÇÃO:** FACULDADE ANHANGUERA DE BELO HORIZONTE

**AUTOR(ES):** OTAVIANO MARTINS MONTEIRO, LUCAS LEITE ANTUNES

**ORIENTADOR(ES):** SANDRO RENATO DIAS

**COLABORADOR(ES):** CAMILA FREITAS DE MIRANDA, JONATHAN TIAGO LÉO SANAS, LEANDRO MAGELA DA SILVA

Realização:



Apoio:



## **ESTUDO DO USO DO KINECT PARA INTERPRETAÇÃO DE GESTOS VISANDO LIBRAS**

### **1. RESUMO**

As pessoas da comunidade surda brasileira utilizam LIBRAS (Linguagem Brasileira de Sinais) para se comunicarem. Mas infelizmente, esta que é a primeira língua das pessoas com deficiência auditiva, não é compreendida pela maioria da população brasileira. A dificuldade de entendimento deve-se ao fato dessa comunicação ser expressa por sinais que muitas vezes são complexos. Devido às dificuldades de comunicação entre pessoas surdas e ouvintes, torna-se viável realizar um estudo sobre soluções de hardwares que possam minimizar este problema. Dentre estas opções, existe o Microsoft Kinect, que originalmente surgiu para o console XBOX 360, possibilitando uma grande interatividade entre o jogador e o jogo. Ele fornece recursos para captura e interpretação de movimentos em tempo real, conseguindo ainda entender o modelo completo de um esqueleto humano, inclusive as suas articulações. O estudo sobre este equipamento feito até então neste artigo, demonstra a sua arquitetura, o seu funcionamento, os pontos em que seus recursos são úteis para a interpretação de gestos visando Libras e as possíveis soluções para a sua principal deficiência que é a captação dos movimentos feitos pelos dedos.

### **2. INTRODUÇÃO**

LIBRAS é uma língua que possui uma própria estrutura gramatical. Ao contrário do que muitos leigos pensam, ela não consiste apenas em mímicas e gestos soltos utilizados pelo surdo para facilitar a comunicação. (AJA, 2012).

Sendo a primeira língua aprendida por pessoas com deficiência auditiva, lamentavelmente, grande parte das pessoas fora da comunidade surda não a compreendem. Esta realidade dificulta a inclusão do surdo em vários ambientes.

Segundo Felipe (2007), pelo fato dos surdos não terem facilidades em se comunicarem com pessoas ouvintes, eles acabam optando por se relacionarem com outras pessoas que possuem essa deficiência, podendo ser por identidade ou domínio da língua de sinais.

Com os avanços da área da Tecnologia da Informação, existem vários recursos que podem ser estudados para que futuramente possa ser desenvolvida uma solução que tenha a capacidade de traduzir de LIBRAS para o português. Para a realização de tal tarefa é necessário analisar bem estes recursos disponíveis, pois como cita Siola (2010), a linguagem LIBRAS apresenta vários sinais complexos e para realizar a captação correta deles é necessário atentar para alguns detalhes como: movimento das mãos em relação ao corpo, movimento das pontas dos dedos e até expressões faciais. Dentre as ferramentas que podem ser úteis para a captação destes movimentos citados, o Microsoft Kinect possui destaque.

Segundo a Microsoft (2010), o Kinect é um dispositivo que foi desenvolvido inicialmente para o console XBOX 360 e possui a capacidade de proporcionar aos jogadores um divertimento de forma interativa. Com este equipamento, é possível que o jogador realize movimentos e estes sejam captados e reproduzidos dentro do jogo.

Recentemente, este dispositivo passou a ser utilizado por diversas comunidades para realização de pesquisas científicas.

De acordo com Andersson (2011), o Kinect pode entender o modelo completo de um esqueleto humano, inclusive as articulações. Ele faz uso da realidade virtual, conjunto de câmeras e sistema de projeção. Dessa forma, é possível captar informações muito próximas da realidade e fazer a análise em tempo real dos movimentos feitos.

Segundo Paula (2011), outra característica importante do Kinect é o tratamento de profundidade, permitindo interpretar se um dado obtido pertence ao corpo humano ou não. Este equipamento pode diferenciar até seis pessoas que estiverem à sua frente, e reconhecer as informações do esqueleto de até dois jogadores.

Apesar de fornecer diversos recursos quanto à captação de movimentos e mapeamento de partes do corpo humano, a maior dificuldade em usar o aparelho para reconhecimento de sinais de Libras deve-se à sua limitação em reconhecer os dedos, que são fundamentais para a comunicação com Libras. Contudo, este obstáculo pode ser solucionado com a ajuda de outros frameworks e assim, o Microsoft Kinect pode ser considerado útil na captação e interpretação dos sinais feitos em LIBRAS.

### **3. OBJETIVOS**

O objetivo geral consiste em estudar o dispositivo kinect, visando a sua capacidade de mapear os sinais de Libras. Para este fim, foi feita uma análise de sua arquitetura e verificado seus recursos referentes à captação de movimentos, no qual foram examinadas as suas vantagens e deficiências quanto ao reconhecimento de gestos feitos principalmente pelas mãos.

Após o término deste estudo geral sobre o kinect, foi iniciado um estudo mais específico, aonde já foi entendido que o dispositivo é eficiente em captar movimentos, mas deixa a desejar quando a captação é referente aos dedos. Dessa forma, o objetivo atual consiste em estudar tecnologias em termos de softwares como OpenNI e OpenCV que possam complementar o aparelho para suprir esta deficiência.

### **4. METODOLOGIA**

A metodologia de trabalho consistiu inicialmente em estudar e documentar a arquitetura do dispositivo Kinect, de modo a entender sua arquitetura e seus componentes. Após esta etapa, foi estudado como é feita a captação de movimentos pelo aparelho, observando em até qual ponto este dispositivo pode ser útil para o reconhecimento dos sinais feitos em LIBRAS. Foi verificado o seu entendimento de movimentos feitos pelas mãos (principalmente ponta dos dedos), corpo, expressões faciais, entre outros detalhes que são imprescindíveis para o entendimento desta linguagem. Também foi observado quando é necessário complementar o Kinect com algum outro framework.

### **5. DESENVOLVIMENTO**

De acordo com Libras (2011), a língua brasileira de sinais (LIBRAS) é usada pela maioria das pessoas da comunidade surda brasileira, sendo reconhecida por lei como uma língua oficial do país. Ela é composta por níveis lingüísticos bem estruturados como: morfologia, fonologia, sintaxe e semântica. A comunicação feita

por esta língua gestual consiste em sinais que são obtidos através de movimentos feitos principalmente pelas mãos e dedos, incluindo outras partes do corpo e expressões faciais.

A comunicação entre uma pessoa com deficiência auditiva que conheça Libras, e uma pessoa ouvinte que não tenha o conhecimento desta língua, pode ser uma tarefa complicada, pois esta língua gestual possui sinais que são complexos e difíceis de serem compreendidos por um leigo. Devido aos problemas encontrados nesta comunicação, pode ser válido realizar um estudo sobre os hardwares existentes que possuam capacidade de captação de movimentos e que possibilite adaptações para posteriormente reconhecer sinais de Libras e traduzi-los para o português. Um dispositivo que possui destaque quanto a esta característica é o Microsoft Kinect.

De acordo com Ferreira (2013), o dispositivo Kinect possui versões diferentes. Os aparelhos que são vendidos juntamente com o console XBOX necessitam de um cabo especial de alimentação quando usado juntamente com um computador, enquanto que o *KINECT FOR WINDOWS* já possui este cabo integrado. Além desta vantagem, ele também possui alguns recursos que os outros dispositivos não possuem. Ele consegue captar pessoas que estão sentadas e permite configurações avançadas das câmeras, como ajustes de luminosidade e exposição. Ambas as versões podem ser integradas com computadores pessoais que possuam diferentes tipos de sistemas operacionais, entre eles Windows (versões 7 e 8) e algumas versões do Linux. É possível desenvolver aplicações trabalhando em conjunto com qualquer uma dessas duas versões.

Segundo Da Silva e Sanas (2012) e Torgoes (2011), os principais componentes da arquitetura do kinect são: um conjunto de microfones que permitem captar uma voz próxima e diferenciá-la de outros ruídos, um projetor de luz infravermelho, um motor que permite girar, alterando o ângulo de visão do dispositivo além de possuir duas câmeras, sendo uma com um sensor do tipo RGB e uma outra câmera específica para captura de sinais infravermelhos. Como pode ser visto na figura 1.



Figura 1: Arquitetura do Kinect

Fonte: MICROSOFT RESEARCH, 2011. In: Da Silva e Sanas, 2012

Conforme cita Paula (2011), ao desenvolver uma aplicação utilizando o Kinect, o dispositivo devolve ao programador três tipos de informações, sendo elas o fluxo de imagem (*Image Stream*), informações de profundidade (*Depth Stream*) e o fluxo de áudio (*Audio Stream*).

- O *Image Stream* pode ser obtido tanto da câmera RGB quanto da câmera infravermelha. Cada pixel representa uma cor e sua resolução pode ser de 640 x 480 pixels, em 30 frames por segundo ou em 1280 x 1024 pixels com até 15 frames por segundo.
- O *Depth Stream* permite que cada pixel indique a distância do objeto em relação ao aparelho e pode também retornar se este pixel pertence ou não ao objeto detectado. O Kinect consegue detectar até 2000 níveis de sensibilidade e reconhece objetos que estiverem à sua frente a uma distância de 1,2 a 3,5 metros.
- O *Audio Stream* possui um conjunto com quatro microfones que anula ruídos e ecos. Permite também o reconhecimento da fala em inglês.

De acordo com Teixeira Et.al (2011), o sensor de profundidade permite a captação de imagens de fundo, este fato é importante, pois assim é possível obter o rastreamento de mãos, detecção de colisões entre outras finalidades.

Realizando um teste prático com o dispositivo Kinect, utilizamos o seu sensor de profundidade e juntamente com o framework OpenNi, conseguimos realizar a gravação de vídeos. Esta imagem abaixo é referente a uma dessas gravações.





Figura 2: Captação do Sensor de Profundidade do Kinect

Fonte: Autores deste trabalho.

Conforme cita Paula (2011), o sensor citado acima é um dos principais destaques do Kinect, pois ele obtém com muita precisão a distância de cada pixel de uma imagem em relação ao sensor. Esta distância é calculada através da projeção de um padrão infravermelho e a deformação neste padrão é medida, realizando assim o cálculo da distância. Este sensor, ainda possui a capacidade de devolver a informação de profundidade com um bom tratamento, informando a cada pixel se este pertence ou não ao corpo de uma pessoa, assim permitindo a distinção de até seis pessoas que estiverem à sua frente e ainda reconhecer o esqueleto de até duas delas.

O Kinect possui a capacidade de reconhecer o esqueleto humano de acordo com as suas articulações, conforme pode ser visto na figura 2:

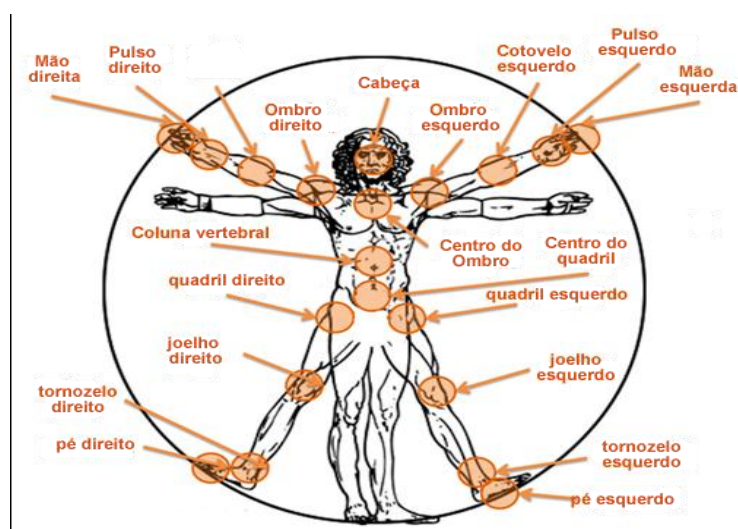


Figura 3: Esqueleto Humano

Fonte: MICROSOFT RESEARCH, 2011. In: Da Silva e Sanas, 2012

Conforme cita Da Silva e Sanas (2012), o Kinect possui a capacidade de reconhecer as articulações e os movimentos das pessoas, graças à realidade virtual, microfones, câmeras e sistema de projeção. Ele capta informações próximas a realidade em tempo real e de uma maneira que compreende todos os movimentos do corpo humano. Porém, o dispositivo não capta de uma maneira precisa os movimentos dos dedos, o que é muito importante na língua de sinais.

Para sobressair a esta dificuldade, Da Silva e Sanas (2012), basearam-se em soluções utilizadas em outros projetos em que os pesquisadores conseguiram através do Kinect obter de uma maneira eficiente a captação de movimentos feitos pelas mãos, incluindo os dedos. Um dos projetos citados é o Vitruvius, que utiliza o Kinect como base de desenvolvimento e através da utilização de redes neurais, conseguiram que o dispositivo possa aprender os movimentos feitos pelas mãos de usuários. Outro projeto citado é o projeto de Reconhecimento de gestos 3D, que também utiliza rede neural e consegue identificar os gestos realizados pelo usuário em tempo real. O reconhecimento dos gestos é feito por um conjunto de técnicas de processamento de imagens e análise de séries temporais, determinando o início e o fim de cada gesto em termos de frames de movimentos. Para conseguir este feito foi montado uma estrutura composta por frameworks como OpenCV e OpenNI, além de middlewares como o NITE. Com a utilização destes recursos combinados com o Kinect é possível captar de uma maneira eficiente os movimentos feitos pelas mãos e dedos.

## **6. RESULTADOS**

Com as informações obtidas até o momento, que vieram através de um estudo sobre a arquitetura e funcionamento do Kinect, incluindo alguns testes práticos com o dispositivo, é possível observar que este equipamento pode ser útil para o desenvolvimento de uma aplicação que traduza Libras para português. Pois este aparelho permite de uma forma eficiente a captação de movimentos, reconhecendo o esqueleto humano e as suas articulações. Este fato é importante, pois alguns movimentos de Libras englobam várias partes do corpo, como pernas e cabeça. Entretanto, uma deficiência do aparelho é a captação dos dedos, que é muito importante ao se tratar de Libras. Para sobressair a esta dificuldade será necessário



a utilização de uma estrutura composta por frameworks e middlewares como OpenNI, OpenCV e Nite, que foram citados no desenvolvimento. É esperado citar até no final deste artigo os possíveis recursos em termos de software que podem trabalhar em conjunto com o Kinect para que ele tenha uma eficiente captação de sinais feitos em Libras.

## 7. CONSIDERAÇÕES FINAIS

O estudo feito até então demonstra a arquitetura e as principais características do dispositivo Kinect que podem ser utilizadas para o reconhecimento de sinais de Libras. Também foi analisado algumas saídas em termos de softwares para complementar o aparelho. O principal ponto a ser aprimorado é o reconhecimento dos movimentos feitos pelos dedos.

A continuação deste estudo é viável, pois existem tecnologias em termos de softwares como OpenNi e OpenCV que podem trabalhar juntos com o Kinect e assim podem possibilitar que o equipamento mapeie os movimentos feitos pelos dedos que são fundamentais para reconhecer Libras.

A continuação deste estudo resultará em uma documentação que poderá ser útil para um futuro desenvolvimento de uma aplicação que contribua significativamente para a comunidade surda.

## 8. FONTES CONSULTADAS

AJA – Associação do Jovem Aprendiz. **O que é LIBRAS.** Disponível em: <<http://www.libras.org.br/libras.php>.> Acesso em 26. nov. 2012.

ANDERSSON, Virginia O.; GRAÑA, Gustavo; ARAÚJO, Ricardo M.(2011) **Investigando o Uso do Kinect para Biometria Através do Caminhar Humano.** Artigo de Pós-Graduação em Computação, Universidade Federal de Pelotas, RS.

COLLET, Lucas Azzi; SILVA (2011), Luciano, **GESTUS: Hardware e Framework de Baixo Custo para Interfaces Gestuais.** Artigo de Graduação em Computação Gráfica, Universidade Presbiteriana Mackenzie, SP.

FELIPE, Tanya A. **LIBRAS em contexto: Curso básico: Livro do estudante**. 8ª Ed. Rio de Janeiro, WalPrint, 2007.

FERREIRA, Fernando Henrique. (2013) Desenvolvendo com o Kinect para o Windows. Disponível em <<http://ferhenriquef.com/2013/04/23/desenvolvendo-com-o-kinect-para-o-windows/>>. Acesso em: 23. Jun. 2013.

LIBRAS. (2011) **O que é Libras ?**. Disponível em: <<http://www.libras.com.br/libras/o-que-e-libras>> . Acesso em 23. Jun.2013.

MICROSOFT, (2010) . **Kinect para Xbox 360 é o nome oficial do dispositivo de jogos sem fios da Microsoft**. Disponível em: <<http://www.microsoft.com/portugal/presspass/Comunicados.aspx?id=322>>. Acesso em: 12 jan. 2013.

PAULA, Bruno Campagnolo de. (2011)**Adaptando e desenvolvendo jogos para uso com o Microsoft Kinect**. Dissertação de graduação em Engenharia de Sistemas.Pontifícia Universidade Católica, Paraná.

SANAS,Jonathan Leo Tiago; DA SILVA, Leandro Magela. (2012) **Levantamento de Dados da Tecnologia de Reconhecimento de Movimentos Utilizando Kinect**. Monografia de Bacharelado no curso de Sistemas de Informação, Faculdade Anhanguera de Belo Horizonte.

SIOLA, Felipe Breve. (2010) **Desenvolvimento de um software para reconhecimento de sinais em LIBRAS através de vídeo**. Monografia para graduação em Ciência da Computação, Universidade Federal do ABC.

TEIXEIRA, João Marcelo; REIS, Bernardo; MACEDO, Samuel; KELNER, Judith. (2012) **Uma avaliação estatística do problema de registro de imagens de profundidade usando o Kinect**, Artigo de Graduação em Tecnologia da Informação, Universidade Federal de Pernambuco.

TOGORES, Tiago Andrade. (2011) **Vitruvius - Um Reconhecedor de Gestos para o Kinect**. Monografia para graduação em Ciência de Computação, Universidade de São Paulo.