

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
DEPARTAMENTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

FÁBIO BATISTA SOUSA DOS SANTOS

**MANIPULAÇÃO INTERATIVA DE IMAGENS MÉDICAS EM 3D USANDO
KINECT E TÉCNICAS DE VISÃO COMPUTACIONAL**

São Luís

2012

FÁBIO BATISTA SOUSA DOS SANTOS

**MANIPULAÇÃO INTERATIVA DE IMAGENS MÉDICAS EM 3D USANDO
KINECT E TÉCNICAS DE VISÃO COMPUTACIONAL**

Monografia apresentada ao Curso de Ciências da Computação da Universidade Federal do Maranhão, para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Aristófanês Corrêa Silva

São Luís

2012

Santos, Fábio Batista Sousa dos.

Manipulação interativa de imagens médicas em 3D usando kinect e técnicas de visão computacional / Fábio Batista Sousa dos Santos. – 2012.

51 f.

Impresso por computador (Fotocópia).

Orientador: Aristófanês Corrêa Silva.

Monografia (graduação) – Universidade Federal do Maranhão, Curso de Ciência da Computação, 2012.

1. Interface homem – máquina. 2. Kinect. 3. Gestos naturais. 4. Gestos – Reconhecimento. 5. Imagens médicas em 3D. I. Título.

CDU 004.5

FÁBIO BATISTA SOUSA DOS SANTOS

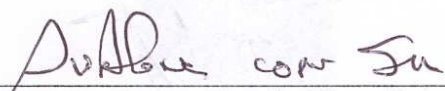
**MANIPULAÇÃO INTERATIVA DE IMAGENS MÉDICAS EM 3D USANDO
KINECT E TÉCNICAS DE VISÃO COMPUTACIONAL**

Monografia apresentada ao Curso de Ciências da Computação da Universidade Federal do Maranhão, para obtenção do grau de Bacharel em Ciência da Computação.

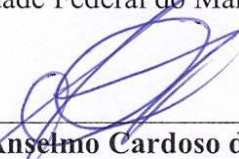
Orientador: Prof. Dr. Aristófanês Corrêa Silva

Aprovada em 31/07/2012

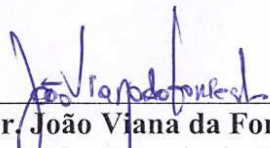
BANCA EXAMINADORA



Prof. Dr. Aristófanês Corrêa Silva (Orientador)
Doutor em Informática
Universidade Federal do Maranhão



Prof. Dr. Anselmo Cardoso de Paiva
Doutor em Informática
Universidade Federal do Maranhão



Prof. Dr. João Viana da Fonseca Neto
Doutor em Engenharia Elétrica
Universidade Federal do Maranhão

A Deus, Rei e Criador de todas as coisa.

A Meus Pais, Batista e Zinha, pelo apoio e dedicação.

As Minhas irmãs Flávia e Fabiana.

A Meus Amigos e Minha namorada Monally, pela ajuda e companheirismo.

A todos que tornaram este sonho realidade.

AGRADECIMENTOS

A Deus, pelo dom da vida e por se fazer presente em todos os momentos da minha vida, e principalmente ao longo deste curso, fazendo com que não desistisse do meu sonho.

A meus pais, João Batista e Maria das Neves, que são meus exemplos de vida, pelo amor, e pelo apoio ao longo da minha vida.

As minhas irmãs, Flávia e Fabiana, por acreditarem em meu trabalho.

A minha Namorada Monally pela compreensão em minha ausência durante estes últimos períodos.

A meu orientador, Aristófanês Corrêa Silva, pela paciência, compreensão, dedicação e pela sinceridade que demonstrou em nossas conversas.

A todos os Amigos do curso de Ciência da Computação não só os que entraram comigo, mas a todos que me ajudaram em momentos de dificuldades durante o curso.

Aos professores do Departamento de Informática pelos conhecimentos passados e pelo profissionalismo.

Enfim, a todos que contribuíram direta ou indiretamente para conclusão deste curso.

“A mente que se abre a uma nova ideia jamais
volta ao seu tamanho original.”

(Albert Einstein)

RESUMO

Sistemas de interface homem-máquina têm sido objetos de pesquisa a vários anos. As pesquisas buscam melhores maneiras de traduzir comandos do usuário para o sistema. Neste contexto, gestos naturais podem ser utilizados como comando para interfaces em sistema computacionais, entretanto, deve-se desenvolver uma maneira de rastrear a mão do usuário e identificar os gestos utilizados como comando para tornar as ações adequadas. Este trabalho propõe uma metodologia de técnicas de manipulação de imagens em 3D, usando técnicas de visão computacional e o Kinect, um dispositivo de reconhecimento de gestos e voz desenvolvido pela Microsoft para o console Xbox, por meio uma câmera e do sensor de profundidade do Kinect, identificar a mão do usuário, rastrear sua posição e identificar gestos de comando produzidos pela mesma. O desempenho do sistema se mostrou eficiente no reconhecimento dos gestos e na manipulação (rotação, escala, translação e animação) das imagens médicas através dos gestos pré-definidos.

Palavras-chave: Interface homem-máquina. Kinect. Gestos naturais. Reconhecimento de gestos. Imagens médicas em 3D.

ABSTRACT

Systems of man-machine interface have been objects of research for several years. Researches seek better ways to translate user commands to the system. In this context, natural gestures can be used as command interfaces for computer system, however, must develop a way to track the user's hand and identify the gestures used as a command to make the appropriate actions. This work proposes a methodology for the manipulation techniques of 3D images using computer vision techniques and Kinect, a device for voice and gesture recognition developed by Microsoft for the Xbox console through a camera and depth sensor Kinect, identify the user's hand, track their position and identify gestures produced by the same command. System performance was efficient in the recognition of gestures and manipulation (rotation, scaling, translation and animation) of medical images through pre-defined gestures.

Keywords: Man-machine interface. Kinect. Natural gestures. Gesture recognition. Medical imaging 3D.

LISTA DE FIGURAS

Figura 2.1 - Microsoft Xbox Kinect.....	19
Figura 2.2 - Abstração das camadas do framework OpenNI.....	21
Figura 2.3 - <i>Hand point detected</i>	22
Figura 2.4 - <i>Full body analysis</i>	23
Figura 2.5 - <i>Wave detected</i>	23
Figura 2.6 - A) Reconstrução tridimensional B) Axial C) Coronal.....	27
Figura 3.1 - Tela do programa MITO.....	32
Figura 3.2 – Arquivo XML de configuração.....	35
Figura 3.3 - Sequência de atividade para interação.	39
Figura 3.4 - Movimento para realizar a animação.....	41
Figura 3.5 - Movimento para realizar o zoom.	41
Figura 3.6 - Movimento para realizar a translação.....	42
Figura 3.7 - Movimento para realizar a rotação.	43
Figura 4.1 - Usando o software com o Kinect (calibração).....	45
Figura 4.2 - Interagindo com o sistema (rotação).....	45
Figura 4.3 - Interagindo com o sistema (zoom).....	46
Figura 4.4 - Interagindo com o sistema (translação).	46
Figura 4.5 - Interagindo com o sistema (animação).	46

LISTA DE SIGLAS

MITO	-	<i>Medical Imaging Toolkit.</i>
OPENNI	-	<i>Open Natural Interaction.</i>
SDK	-	<i>Software Development Kit.</i>
API	-	<i>Application Programming Interface.</i>
IHM	-	Interação Homem-Máquina.
DICOM	-	<i>Digital Imaging and Communications in Medicine.</i>
RM	-	Ressonância Magnética.
TC	-	Tomografia Computadorizada.
PACS	-	<i>Picture Archiving and Communication System.</i>
VTK	-	<i>Visualization Toolkit</i>
ITK	-	<i>Insight Segmentation and Registration Toolkit</i>
OpenGL	-	<i>Open Graphic Library</i>
XML	-	<i>eXtensible Markup Language</i>

SÚMARIO

1. INTRODUÇÃO	14
1.1 MOTIVAÇÃO	15
1.2 JUSTIFICATIVA	16
1.3 OBJETIVOS	17
1.4 ORGANIZAÇÃO DO TRABALHO.....	17
2. CONCEITOS BÁSICOS E TECNOLOGIAS.....	18
2.1 TECNOLOGIAS	18
2.1.1 Kinect.....	18
2.1.2 OpenNI	19
2.1.3 NITE	24
2.1.4 MITO	24
2.2 AQUISIÇÃO DE IMAGENS EM 3D	25
2.2.1 DICOM.....	25
2.2.2 Ressonância Magnética	26
2.2.3 Tomografia Computadorizada.....	27
2.3 INTERAÇÃO NATURAL	28
2.3.1 Reconhecimento de gestos.....	29
3. METODOLOGIA PROPOSTA.....	31
3.1 AQUISIÇÃO DAS IMAGENS	31
3.2 INTERFACE DO USUÁRIO	32
3.3 OPENNI.....	33
3.3.1 Integrando OpenNI ao MITO	33
3.3.2 Comunicação com o sensor Kinect.....	34
3.4 NITE	35
3.4.1 Integrando NITE ao MITO	35
3.4.2 Reconhecimento de gestos.....	36
3.5 CONTROLE POR GESTOS	38
3.6 TÉCNICAS DE INTERAÇÃO	40
3.6.1 Animação	40
3.6.2 Zoom	41
3.6.3 Translação	42

3.6.4	Rotação	42
4.	EXPERIMENTO E RESULTADOS.....	44
5.	CONCLUSÃO	48
	REFERÊNCIAS	50

1. INTRODUÇÃO

A interação homem-máquina vem sendo um assunto pesquisado desde o início da computação, sendo que diferentes aplicações dependem de diferentes tipos de interação entre o usuário e algum dispositivo. Mesmo que grande parte das operações com dispositivos computacionais possa ser realizada por meio de teclado e *mouse*, diferentes tipos de interações podem vir a aumentar consideravelmente a produtividade do usuário em determinada tarefa (CHU; FILHO, 2011).

Mesmo possuindo alto grau de complexidade, gestos com as mãos são movimentos bastante naturais aos seres humanos, e possuem alta precisão. Desta forma, gestos podem ser utilizados como interface em uma grande gama de aplicações, possibilitando executar tarefas complexas através de comandos gestuais bastante intuitivos ao usuário (CHU; FILHO, 2011).

Devido a grande complexidade dos movimentos manuais humanos, sistemas capazes de reconhecer gestos em grau satisfatório de detalhamento podem ter alguns problemas, existem dificuldades principalmente na variação de gestos de pessoa para pessoa, e como determinar os pontos iniciais e finais dos gestos. Para essa ambiguidade existe diversos mapeamentos e conceitos como está especificado em (MITRA; ACHARYA, 2007). Atualmente não existe técnica de controle de gesto eficaz como relatado em (SHAN, 2010).

Com o lançamento do Kinect, um dispositivo de reconhecimento de gestos e voz desenvolvido pela Microsoft para o console Xbox, as oportunidades de desenvolvimento de tecnologias e ferramentas de interface natural se multiplicaram, não só no mundo dos *games* como também criou incontáveis oportunidades pelo mundo a fora e com certeza uma das que mais impressionou foi a quantidade de oportunidades que surgiram no campo da medicina. As aplicações com o Kinect na área médica tem sido um avanço considerável. Hoje os médicos podem explorar dados de imagens como tomografia computadorizada e ressonância magnética de forma interativa acessando um grande conjunto de dados por diferentes tipos de interface de usuário sem utilizar *mouse* ou teclado.

Como está relatado em (STEAKLEY, 2011) médicos em um hospital *Sunnybrook* de Toronto estão apostando na tecnologia Kinect para economizar tempo e evitar a contaminação na sala de cirurgia. Uma equipe do hospital começou a usar o Kinect Xbox, para virtualmente manipular imagens médicas durante a cirurgia. Os médicos usam gestos com as mãos para fazer *zoom in* e *zoom out* ou selecionar determinadas imagens sem sair da mesa de operações.

Normalmente cirurgiões tem que deixar o campo estéril em torno do paciente para visualizar as imagens, como ressonância magnética ou tomografia computadorizada em um computador próximo. Eles, então, tem que passar por uma limpeza, antes de retornar para a área para se certificar de que eles não trazem as bactérias que podem prejudicar o paciente. Pode demorar até 20 minutos para limpar as mãos, unhas e braços cada vez que um médico consulta uma imagem, disse o Dr. Calvin Law, que ajudou a integrar a tecnologia Kinect na sala de cirurgia (STEAKLEY, 2011). As interrupções por várias vezes, podem causar mais de uma hora de atraso ao longo de uma cirurgia, ao eliminar os atrasos, o hospital poderia economizar tempo suficiente para operar mais pacientes.

Há crescentes interesses por aplicações que utilizam controle por gestos na medicina usando tecnologia Kinect, principalmente para serem aplicados em manipulação de imagens 3D em sala de cirurgias. Pois além da manipulação de imagens sem qualquer auxílio do *mouse* ou teclado é eficiente e fácil de usar.

Desta forma o presente trabalho abordará o desenvolvimento de uma metodologia para interação de imagens médicas em 3D usando Kinect e técnicas de visão computacional, com auxílio de bibliotecas *open source* tais como OpenNI (OPENNI, 2011) e NITE (PRIMESENSE, 2011) que auxiliam na identificação e reconhecimentos de gestos.

1.1 MOTIVAÇÃO

A motivação deste trabalho foi o interesse em aplicações na área médica e novas tecnologias, em particular o Kinect que está fazendo sucesso não só no mundo dos *games*, agora está sendo usado na área médica e tem impressionado, a eficiência de uma interface de computador totalmente baseada em gestos pode reduzir o tempo em salas de cirurgias e consequentemente um maior número de pacientes atendidos diariamente. A seguir será mostrado algumas aplicações usando Kinect:

- Diagnosticar distúrbios¹ em crianças: O Kinect esta sendo usado para medir uma variedade de sintomas do transtorno² em crianças, criando uma forma mais

¹ Anomalia funcional de um órgão ou de um sistema.

² Perturbação mental.

objetiva de avaliá-los, a fim de detectar problemas como autismo³, transtorno de déficit de atenção⁴ e transtorno obsessivo compulsivo⁵ (MLAZARUS, 2011);

- Aumentar a eficácia das interações médico-paciente: O Kinect esta sendo usado para detectar sinais não-verbais. Isso pode aumentar consideravelmente as habilidades dos médicos para detectar as necessidades do paciente via linguagem corporal e expressões faciais (MLAZARUS, 2011);
- Manipular imagens 3D de Tomografia computadorizada: O Kinect esta sendo usado para criação de aplicações que permite visualizar e explorar imagens 3D de tomografia computadorizada e para dar aos cirurgiões a capacidade de manipular imagens sem usar *mouse* (MLAZARUS, 2011).

1.2 JUSTIFICATIVA

Com crescentes custos de *hardware* médicos, houve necessidade de apresentar uma oportunidade não só para diminuir os custos, mas para melhorar o atendimento ao paciente. Apesar da grande disponibilidade de recursos e necessidade de médicos tem em explorar dados de imagens 3D em salas de cirurgias, bem poucos sistemas computacionais têm sido implantados . Outro problema frequente em salas de cirurgias, é a higienização dos equipamentos e até mesmo o ambiente e os profissionais. O problema é que médicos tem que coletar informações do paciente, visualizar imagens em um computador próximo e até mesmo pisar fora da área esterilizada, a cada uma dessas etapas o medico deve passar por limpeza para certificar que não trazem bactérias como mostra (STEAKLEY, 2011).

A implantação desta metodologia usando a tecnologia Kinect pode ajudar os médicos a manipular imagens sem deixar a sala de operação, e sem ter contato a dispositivos como teclado e *mouse* o que reduz o processo de operação, já que o medico não passará por esterilização ao ter contato a interface de visualização de imagens dos pacientes.

³ Síndrome comportamental com causas múltiplas, decorrente de um distúrbio de desenvolvimento.

⁴ Caracteriza principalmente por desatenção crônica, impulsividade, dificuldade de concentração e problemas de memória.

⁵ É um transtorno de ansiedade caracterizado por pensamentos exagerados no qual o indivíduo tem comportamentos considerados estranhos para a sociedade.

1.3 OBJETIVOS

Apresentar uma metodologia computacional para manipular imagens médicas 3D, mais especificamente tomografia computadorizada e ressonância magnética, usando gestos com as mãos. Serão utilizadas a tecnologia Xbox Kinect como dispositivo de entrada e técnicas de visão computacional para realizar este objetivo.

A metodologia proposta disponibilizará recursos de manipulação de imagens em 3D, como rotação, translação, escala e animação, para que as imagens de tomografias computadorizadas e ressonância magnéticas possam ser melhor analisadas.

1.4 ORGANIZAÇÃO DO TRABALHO

Aqui serão descritos, resumidamente, os principais tópicos abordados pelos diversos capítulos deste texto, fazendo um breve resumo do conteúdo de cada capítulo, para que o leitor possa melhor compreender a estrutura, e facilitar como se dirigir aos itens de mais interesse.

No Capítulo 2, aborda a descrição das tecnologias utilizadas, introduz alguns dos conceitos importantes para a compreensão deste texto.

No Capítulo 3, descreve a metodologia proposta, técnica de reconhecimentos de gestos com auxílio a bibliotecas c++, tais como OpenNI e NITE, e detalhes da interface de interação do *Medical Imaging Toolkit* (MITO).

No Capítulo 4, mostra os experimentos da metodologia desenvolvida e os resultados encontrados.

No Capítulo 5, são feitas as considerações finais do trabalho realizado.

2. CONCEITOS BÁSICOS E TECNOLOGIAS

Neste capítulo serão apresentados alguns conceitos introdutórios que servem como base para a compreensão deste trabalho. A visão dos tópicos abordados aqui tem o objetivo de oferecer uma noção geral sobre o assunto.

2.1 TECNOLOGIAS

Aqui será descrito as principais tecnologias utilizadas, tais como o dispositivo de entrada para manipulação de imagens o Kinect, principais bibliotecas com suas características e a estrutura do *Medical Imaging Toolkit* (MITO).

2.1.1 Kinect

O Kinect é um dispositivo de entrada desenvolvido pela Microsoft para uso no Xbox 360, um videogame, que usa o sensor Kinect para que o usuário possa interagir de forma natural, através de gestos corporais e comandos de voz, com a finalidade de entretenimento. A Figura 2.1 mostra os componentes do sensor Kinect (MICROSOFT, 2011). Seus recursos são:

- *RGB camera*: Obtém imagens coloridas numa resolução de 640 x 480, com uma taxa de atualização de 30 quadros por segundo. Cada pixel é representado por 32 bits, sendo que apenas 24 deles são usados, com cada 8 bits determinando o valor de um componente cor RGB (*Red, Green, Blue*);
- *3D depth sensors*: Possui um emissor de luz infra-vermelha e um sensor CMOS (*Complimentary Metal-Oxide Semiconductor*) que funcionam juntos para identificar a profundidade dos objetos independentemente das condições de iluminação. O sensor compara a luz refletida com um padrão gravado no *firmware*. Dessa maneira, é calculado um mapa de profundidade com resolução de 640 x 480, a uma velocidade de também de 30 quadros por segundo;

- *Motorized tilt*: Rotaciona o aparelho em alguns graus na vertical para acompanhar um usuário;
- *Multi-array microphone*: São 4 uniformemente distribuídos na parte inferior, que permitem um processo de som mais avançado, em geral para melhorar sua qualidade, através de redução de ruído do ambiente e cancelamento de eco.

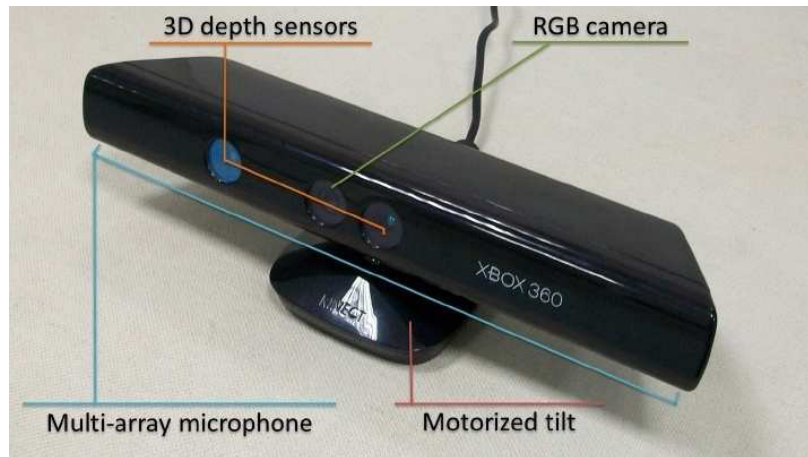


Figura 2.1 - Microsoft Xbox Kinect.
Fonte: (GALLO; PLACITELLI; CIAMPI, 2011).

A grande vantagem do Kinect foi conseguir calcular com precisão de centímetros a distância de objetos e pessoas na cena dentro de um intervalo de 60cm a 5m em tempo real, com *hardware* de baixo custo. Como se usa um conector USB e uma fonte de energia, foi possível adaptá-lo para usar em computadores.

Com estes recursos o Kinect está gerando grande inovação no mundo da medicina, com o lançamento do *Software Development Kit* (SDK) Kinect da Microsoft várias aplicações na área médica tem sido implantados (STEAKLEY, 2011).

2.1.2 OpenNI

OpenNI ou *Open Natural Interaction* é uma organização criada em Novembro de 2010 com o objetivo de melhorar a compatibilidade e interoperabilidade de dispositivos de interface natural, aplicações e *middleware*⁶. O OpenNI lançada por uma empresa israelense PrimeSense responsável pela tecnologia de sensoriamento 3D (*Light Coding*) é usada no

⁶ É um programa de computador que faz a mediação entre software e demais aplicações.

hardware do Kinect (OPENNI, 2011). Também foi utilizada como forma de "*hackear*" o Microsoft Kinect para ser utilizado em aplicações em máquinas (computador) além do Xbox 360, que era seu principal e único alvo originalmente.

O OpenNI é um *framework* multiplataforma com suporte a múltiplas linguagens de programação que define *Application Programming Interface* (API) para a escrita de aplicações as quais utilizam interação natural, ou seja, fazem uso de gestos para a interação com sistemas computacionais. Seu principal propósito é construir uma API padrão para comunicação com sensores de vídeo e áudio, e com *middleware* que realize percepção de visão e audição (OPENNI, 2011).

Uma característica importante, consiste no fato de que as APIs fornecidas pelo OpenNI permitem ao desenvolvedor realizar o escaneamento de cenas tridimensionais do mundo real, através dos tipos de dados gerados a partir da entrada do sensor (o Kinect), podendo obter a representação do corpo, a localização de uma mão ou até uma matriz de *pixels* representando um mapa de profundidades da cena. Além disso, todas essas operações, podem ser realizadas de forma independente do sensor a ser utilizado e até mesmo do *middleware*, facilitando muito a reutilização do código.

A Figura 2.2 mostra uma visualização em três camadas do conceito utilizado pelo OpenNI.

- ***Application***: Representa o *software* que implementa aplicações que fazem uso de gestos e voz para a interação;
- ***OpenNI Interfaces***: Representa o OpenNI, provendo interfaces de comunicação com os sensores e os componentes de *middleware*, que analisam dados dos sensores;
- ***Hardware Device(sensor)***: Representa o dispositivo de *hardware* que captura dados de áudio e vídeo da cena.

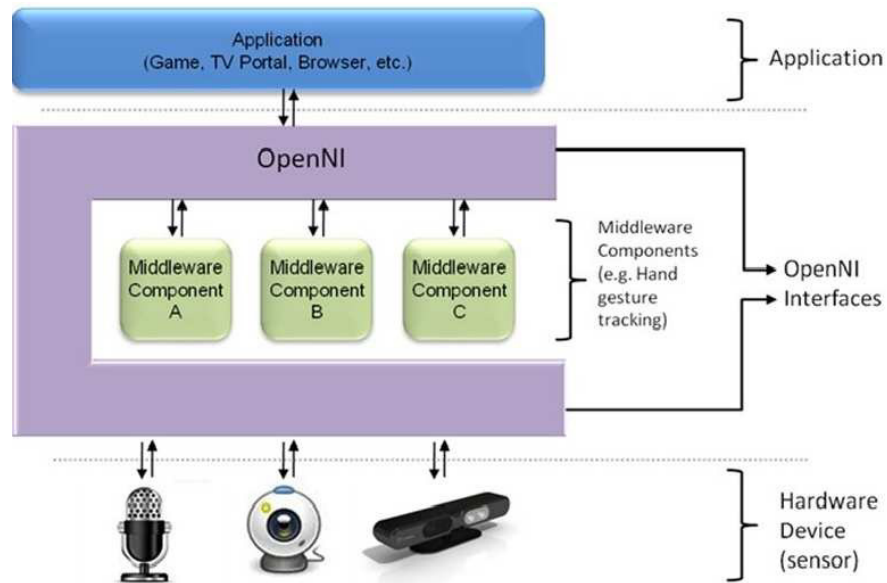


Figura 2.2 - Abstração das camadas do framework OpenNI.

Fonte: <http://openni.org/>.

2.1.2.1 Componentes OpenNI

Alguns componentes OpenNI é de fundamental importância para aplicação neste trabalho. Será apresentado nesta seção uma descrição dos principais componentes utilizados pelo OpenNI para implementar funcionalidades abordadas no presente trabalho.

- *Análise Completa do Corpo (Full body analysis)*: Um componente que processa os dados e gera informação do organismo relacionado, tipicamente estrutura de dados que descreve articulações, orientação, centro de massa, e assim por diante;
- *Análise de pontos da mão (Hand point analysis)*: Um componente de *software* que processa os dados e gera a localização do ponto de centro da mão;
- *Detecção de gesto (Gesture detection)*: Um componente que identifica gestos predefinidos (por exemplo, *wave*) e alerta o aplicativo;
- *Análise de cena (Scene Analyzer)*: Um componente que analisa a imagem da cena, a fim de produzir informação tais como: separar o que faz parte da cena e o fundo, as coordenadas do plano, identificação individual das figuras na cena.

O OpenNI define Produção de Nós ou *Production Nodes*, como sendo um conjunto de componentes que têm um papel produtivo no processo de criação de dados necessários para aplicações baseadas em interação natural. Os nós ou conexões são os principais pontos do corpo como ombro, cotovelos, mãos, pés, quadril e cabeça esses pontos são identificados pelo componente *Full body analysis*. Cada *Production Nodes* encapsula a funcionalidade que se relaciona com os tipos de dados específicos. Esses nós são elementos fundamentais na interface OpenNI, no entanto apenas define a linguagem, a lógica de geração de novos dados e funcionalidades deve ser implementada pelos módulos que se conectam ao OpenNI.

Por exemplo, há um *Production Nodes* que representa a funcionalidade de geração do dados do ponto da mão. A lógica para geração de outros tipo de dados deve ser feito por componentes externo. O *middleware Natural Interaction (NITE)* por exemplo que está ligado ao OpenNI, implementa novos gestos tendo como dado a localização da mão.

A Figura 2.3 exemplifica uma aplicação para acompanhar o movimento da mão da pessoa. A localização da mão do usuário, que pode ser tanto o centro da palma da mão, como também as pontas dos dedos.

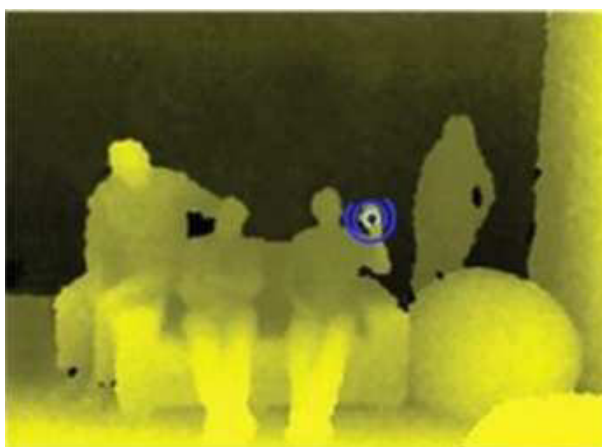


Figura 2.3 - *Hand point detected*.
Fonte: <http://openni.org>.

A Figura 2.4 identifica a localização, orientação e conexões para descrever o corpo na cena. O OpenNI define os principais pontos do corpo na cena como o centro da cabeça, ombro esquerdo e direito, cotovelo esquerdo e direito, mão esquerda e direita, centro da coluna, quadril esquerdo e direito, joelho esquerdo e direito, pé esquerdo e direito.

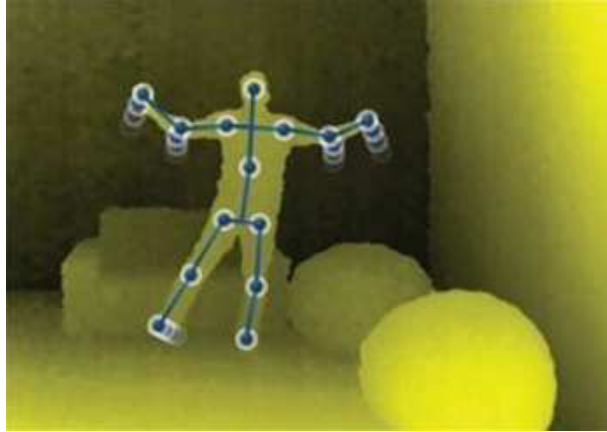


Figura 2.4 - *Full body analysis*.
Fonte: <http://openni.org/>.

Com a definição destes nós e com auxílio de componentes externos e bibliotecas serão a base para a metodologia proposta.

A Figura 2.5 mostra quando o gesto *wave* é detectado ou reconhecido. O gesto *wave* é definido na biblioteca OpenNI, e é chamado quando o usuário realiza mudanças de direção com a mão indo da esquerda para direita e vice-versa em direção ao sensor Kinect.



Figura 2.5 - *Wave detected*.
Fonte: http://openni.org.

Com o estudo das características do OpenNI, pôde-se concluir que ela atenderia as necessidades do projeto proposto e que, portanto, será aplicado no reconhecimento de gestos do usuário, tendo o Kinect como dispositivo de entrada. A biblioteca OpenNI dará suporte à utilização do *middleware* NITE, da PrimeSense.

2.1.3 NITE

A utilização do dispositivo Kinect, demandou a adoção de ferramentas de desenvolvimento que tornassem possível a utilização de seus recursos. O *middleware Natural Interaction NITE*, da PrimeSense, tem como função prover uma interface de alto nível para a comunicação com este dispositivo, e ainda tem como grande vantagem a possibilidade de sua utilização com outros dispositivos que possuam recursos similares, sem a necessidade de grandes modificações na aplicação (PRIMESENSE, 2011).

O *middleware NITE* inclui métodos para processamento de dados de vídeo, inclusive com o recurso do mapeamento da profundidade da cena, contendo uma infra estrutura de algoritmos capaz de realizar identificação de pessoas, características de objetos em geral e reconhecimento de gestos, permitindo ao desenvolvedor criar aplicação controladas por gestos, encapsulando diversos algoritmos de processamento de imagens que são utilizados largamente em aplicações de visão computacional. Além disso por fazer parte da arquitetura de desenvolvimento de aplicações da PrimeSense, o *middleware* possui alta eficiência em termo de performance, robustez e taxa de detecção (PRIMESENSE, 2011).

Em adição a uma metodologia orientada a eventos, o NITE possui a capacidade de capturar gestos pré-definidos. Desta forma, juntamente com o dispositivo Kinect, optou-se pela utilização do OpenNI e do *middleware NITE*, devido sua maturidade e os recursos já disponíveis.

As características descritas para o *middleware* em questão são consideradas de suma importância para a metodologia proposta, com recursos para uma aplicação com boas taxas de reconhecimento de gestos e recursos avançados. No projeto proposto será usado gestos pré-definidos pelos NITE, na qual cada gesto reconhecido executa determinada ação, um exemplo e a *gesture circle control*, ou seja, o quando o usuário faz um ciclo completo em direção ao Kinect esse gesto é reconhecido e ativa a ação de rotacionar a imagem.

2.1.4 MITO

O *software MITO* ou *Medical Imaging Toolkit* (ALFANO et al., 2009) é um visualizador de imagens médicas gratuitos, e de código aberto, totalmente compatível com o

padrão *Digital Imaging and Communications in Medicine* (DICOM). O MITO foi escrito em c++ e implementado com bibliotecas *open-source* e multiplataforma, que foram integrados e estendidos para suportar processamento de imagens médicas, funcionalidades de visualização e interação. As principais bibliotecas usadas no visualizador de imagens MITO são: *Open Graphic Library* (OpenGL) - para exibição em 2D e 3D rapidamente; *Visualization Toolkit* (VTK) - para renderização em 3D ; *Insight Segmentation and Registration Toolkit* (ITK) para a segmentação de imagens e *Wxwidgets* para construção da interface gráfica do usuário.

O MITO será o interface principal, adicionando biblioteca e código em c++, estendendo para que seja utilizado o sensor Kinect como dispositivo de entrada.

2.2 AQUISIÇÃO DE IMAGENS EM 3D

A aquisição da imagens consiste em adquirir uma imagem por meio de um sensor e digitalizá-la. Podem ser divididas em invasivas e não invasivas, de acordo com a forma como são obtidas. Os métodos invasivos caracterizam-se pela introdução de um instrumento no interior do corpo humano, de forma a obter as imagens pretendidas. Nesta categoria incluem-se as angiografias e as imagens de medicina nuclear. Nos métodos não invasivos incluem-se os raios X, ultra-sonografia, tomografia computadorizada e ressonância magnética. Neste trabalho serão abordadas as características gerais de dois métodos não invasivos, são imagens de ressonância magnética e tomografia computadorizada e usando o padrão DICOM.

2.2.1 DICOM

O DICOM foi criado com a finalidade de se padronizar as imagens diagnósticas , como tomografias, ressonâncias magnéticas, radiografias, ultrassonografias, etc. O padrão DICOM é uma série de regras que permite que imagens médicas e informações associadas sejam trocadas entre equipamentos de imagem, computadores e hospitais. O padrão estabelece uma linguagem comum entre os equipamentos de marcas diferentes, que geralmente não são compatíveis, e entre equipamentos de imagem e computadores, estejam esses em hospitais, clínicas ou laboratórios (JUNIOR; YAMASHITA, 2001).

As dificuldades de se criar um padrão estão no seu planejamento e desenvolvimento, o que envolve os interesses das empresas. A padronização só será eficiente se os usuários forem beneficiados com as mudanças feitas e se for amplamente utilizado. Grande parte do tempo é perdida nas negociações. Atualmente o DICOM é gerido por um comitê composto por praticamente todos os grandes fabricantes de equipamentos para imagem diagnóstica e, por grandes instituições médico científicas em todo o mundo.

A visualização de forma padronizada possibilita melhor definição na interpretação médica, pois só uso de texto não é suficiente para certas situações, sendo que através das imagens médicos abstraem informações muitas vezes não correlacionadas nos laudos textuais.

2.2.2 Ressonância Magnética

A ressonância magnética, antigamente chamada de Ressonância Nuclear Magnética, é um método de obtenção de imagens computadorizadas que não utiliza radiação ionizante.

Esta técnica se divide em três etapas: alinhamento, excitação e detecção de radiofrequência. O alinhamento apresenta o comportamento dos átomos quando submetidos a um campo magnético.

Na excitação cada núcleo de hidrogênio "vibra" numa determinada frequência proporcional ao campo magnético aplicado. O aparelho emite uma onda eletromagnética na mesma frequência. Neste momento acontece uma troca de energia da onda emitida pelo equipamento para os átomos de hidrogênio, este fenômeno é conhecido como ressonância.

Na detecção de radiofrequência, os núcleos de hidrogênio que receberam a energia tornaram-se instáveis. Ao retornar ao estado normal, eles emitem ondas na mesma frequência. Então o equipamento detecta essas ondas e determina a posição no espaço e a intensidade da energia (JUNIOR; YAMASHITA, 2001).

A Figura 2.6 mostra um caso específico de imagens de Ressonância Magnética (RM), como são capazes de demonstrar diferentes estruturas no cérebro.

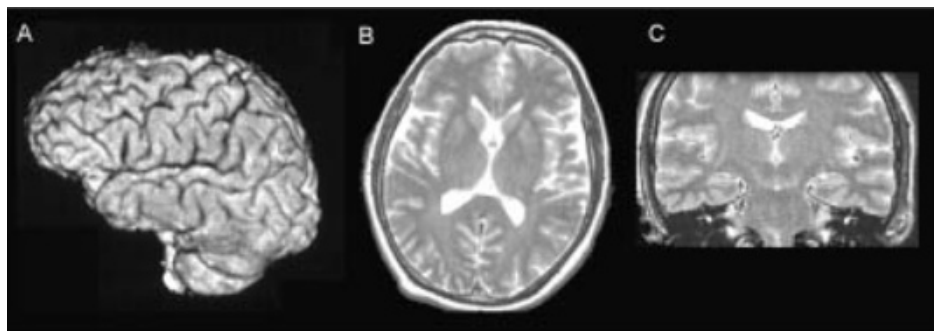


Figura 2.6 - A) Reconstrução tridimensional B) Axial C) Coronal.
Fonte: <http://www.scielo.br>.

Este tipo de aquisição tem uma capacidade de demonstrar aos médicos diferentes estruturas do local examinado bem como facilidade em demonstrar mínimas alterações na maioria das doenças. As alterações morfológicas são mais facilmente avaliadas do que na Tomografia computadorizada (TC), e com maior sensibilidade para doenças desmielinizantes⁷ e processos infiltrativos⁸.

2.2.3 Tomografia Computadorizada

A tomografia computadorizada trata-se de um método radiológico que permite a obtenção de uma imagem de determinado segmento do corpo humano com a finalidade de visualização e estudo de suas estruturas anatômicas. A palavra tomografia, deriva da palavra grega "Tomos", que significa corte ou fatia, e "grafos", que significa uma imagem ou gráfico (Monteiro, 2005).

O aparelho consiste em uma fonte de raios-X que é acionada ao mesmo tempo em que realiza um movimento circular ao redor do corpo do paciente, emitindo um feixe de raios-X em forma de leque. No lado oposto a essa fonte, está localizada uma série de detectores que transformam a radiação em um sinal elétrico que é convertido em imagem digital. Dessa forma, as imagens correspondem a fatias do local examinado. A intensidade (brilho) reflete a absorção dos raios-X e pode ser medida em uma escala (JUNIOR; YAMASHITA, 2001).

⁷ É qualquer doença do sistema nervoso na qual a bainha de mielina dos neurônios é danificada.

⁸ Infiltrativos quer dizer que ele já atingiu tecidos vizinhos e tem grandes chances de metástase (espalhar).

Com a evolução tecnológica, é possível adquirir imagens rapidamente através de técnicas de helicoidal⁹ e multidetectores¹⁰. Podem ser feita aquisição de até 2000 imagens por exame em tempos de 0.5 segundos (Dutra, 2009). Apesar dos avanços, ainda é limitada a capacidade de diferenciar entre substância branca e cinzentas notadamente na região do cerebelo¹¹ e núcleos da base¹². A grande deficiência é vista nas doenças desmielinizantes ou em algumas lesões neoplásicas infiltrativos e em transtornos psiquiátricos¹³

Vimos que o a ressonância magnética e tomografias computadorizadas são métodos de aquisição de imagem, que diferem, basicamente, pelo tipo de sinal utilizado na formação das imagens. A tomografia computadorizada possui um emissor de raio-X, que emite radiação ionizante, enquanto a ressonância magnética utiliza um intenso campo magnético acoplado a um emissor de radiofrequência.

2.3 INTERAÇÃO NATURAL

O Início da era computacional foi marcada pela necessidade de se obter um processamento preciso de dados, por este motivo não se deu importância ao processo de Interação Homem-Máquina (IHM). Com a evolução e disseminação dos computadores pessoais a demanda por metodologias específicas para IHM teve um aumento significativo, visto que os usuários passaram a necessitar de interfaces que maximizassem sua produtividade, sendo mais intuitivas e/ou de fácil manipulação (PREECE et al., 1994).

Interação natural (ou *Natural Interaction*) é considerada um conceito, que estuda formas do homem interagir com dispositivos eletrônicos através dos 5 sentidos do ser humano, seja através de gestos, comandos de voz, movimentos e expressões corporais ou detecção e identificação de partes do corpo humano como rosto, mão, polegar, retina, articulações, entre outros. São exemplos de interação natural:

⁹ Tomografia helicoidal ou de quarta geração, além do tubo de raios X e dos detectores girarem, a mesa também é deslocada e a trajetória do feixe de raios X ao redor do corpo é em hélice (ou espiral).

¹⁰ Tem maior quantidade de fileiras de detectores de raio x, ou seja, tem o a obtenção de mais de uma imagem a cada giro do tubo.

¹¹ É a parte do encéfalo responsável pela manutenção do equilíbrio e pelo controle do tônus muscular e dos movimentos voluntários.

¹² São um grupo de núcleos no cérebro interconectados com o córtex cerebral, tálamo e tronco cerebral. Os núcleos da base de mamíferos estão associados a diversas funções: controle motor, cognição, emoções e aprendizado.

¹³ São condições de anormalidade, sofrimento ou comprometimento de ordem psicológica, mental ou cognitiva.

- O reconhecimento de fala, onde os dispositivos recebem instruções através de comandos vocais;
- Gestos com a mão, é o que será aplicado neste trabalho, onde gestos pré-definidos com as mãos são reconhecidos e interpretados para ativar e controlar dispositivos.

As pessoas naturalmente se comunicam através de gestos, expressões, movimentos fala e outros. A interação natural é fazer com que novos sistemas computacionais compreendam essas ações e envolver a pessoa em um diálogo com a máquina, permitindo-lhes interagir naturalmente com sistema sem qualquer dispositivo. A interação é intuitiva e fácil de usar sem qualquer instruções complexas para manusear a interface. O sucesso de um sistema de interação natural depende como esta técnica influencia as pessoas que utilizam. Deve atender as expectativas e até mesmo surpreender o usuário pela facilidade de manuseio com está escrito em (VALLI, 2007).

Para que haja a interação natural é necessário uma forma das entidades se comunicarem, para a metodologia proposta os gestos do usuário serão a forma de se comunicar com o sistema, eles serão adotados para interagir com a interface MITO, e assim manipular às imagens, cada gesto reconhecido pelo sistema aciona determinada ação.

2.3.1 Reconhecimento de gestos

Reconhecimento de gestos é uma tarefa complexa, que envolve análise e modelagem do movimento, reconhecimento de padrões e aprendizado de máquina. Em alguns casos, até estudos linguísticos (TOGORES, 2011). A princípio, eles podem ser formados por qualquer parte do corpo humano, todavia faremos interpretação de apenas gestos realizados com as mãos.

Para o reconhecimento de gesto aplicados no experimentos utilizamos bibliotecas *Open Source* como OpenNI (OPENNI, 2011) e NITE (PRIMESENSE, 2011) e algumas funções pré-definidas. O *middleware* NITE suporta a criação de gestos através de programação em c++ e a utilização de formas de controle pré-definidas, são elas:

- ***Push Detector***: verifica se ocorreu um movimento de aproximação e afastamento do sensor;
- ***Swipe Detector***: apura se ocorre um leve movimento para qualquer direção sucedido de uma parada;
- ***Steady Detector***: é ativado se a mão ficar parada;
- ***Wave Detector***: é ativado se mão faz um movimento de mudança de direção dentro de um limite de tempo. Por padrão, quatro mudanças de direção são necessários para identificar o *wave*;
- ***Circle Detector***: aguarda que o usuário efetue um movimento circular completo com a mão, Sentido horário é considerado o sentido positivo, e anti-horário é considerado o negativo;
- ***SelectableSlider1D***: implementa um controle deslizante unidimensional (x ou y ou z);
- ***SelectableSlider2D***: implementa um controle deslizante bidimensional (x-y).

Através do mecanismo de reconhecimento da posição da mão e dos métodos de controle contidos no *middleware* NITE, foi desenvolvida a metodologia de interação com a Interface MITO utilizada neste trabalho.

3. METODOLOGIA PROPOSTA

Este capítulo contém a descrição da metodologia utilizada para implementação e desenvolvimento do manipulador de imagens médicas em 3D usando Kinect como dispositivo de entrada. Além disso, serão utilizadas uma interface MITO, e bibliotecas *open source* como OpenNI e NITE, para reconhecimento de gestos e interação com o sistema. Aqui será mostrado o procedimento de aquisição das imagens, componentes de interface do usuário, a comunicação com o sensor Kinect, técnicas utilizadas no reconhecimento de gestos e técnicas de manipulação tais como escala, translação, rotação e animação de imagens 3D.

3.1 AQUISIÇÃO DAS IMAGENS

Foram utilizadas imagens do OsiriX DICOM *files*, o OsiriX é um *software* de processamento de imagem dedicada a imagens DICOM. Este *software* possui uma base com imagens de ressonância magnética e tomografia computadorizada no padrão DICOM disponibilizado gratuitamente na internet, e tem o apoio do Instituto Nacional de Saúde dos Estados Unidos.

O banco de imagens OsiriX DICOM *files* tem propósito de pesquisa e ensino bastante utilizada pela comunidade científica. Foram selecionados um conjunto de imagens de cada modalidade (ressonância magnética e tomografia computadorizada). As imagens utilizadas foram:

- A uma imagem de ressonância magnética do cérebro. O processo para aquisição deste tipo de imagem, o paciente deve ficar deitado de costas com os braços ao longo do corpo e um equipamento é instalado ao redor da cabeça. Durante o exame, somente a cabeça vai ser colocada no interior do anel (equipamento que captura a imagem).
- Uma tomografia computadorizada da coluna cervical. No procedimento de aquisição, o paciente é posicionado numa mesa que se desloca para o interior de um anel com cerca de 80 cm de diâmetro onde serão adquiridas as imagens por raios-x.

3.2 INTERFACE DO USUÁRIO

O *software* MITO foi utilizado como interface do usuário para visualizar as imagens, consiste em uma interface intuitiva, serão integrados os módulos OpenNI e NITE com objetivo comunicar com o sensor Kinect e fazer o reconhecimento de gesto. O MITO está inteiramente escrito em linguagem c++ e será estendido para dar suporte as funcionalidades e técnicas de interação. Toda interação com a imagens são mapeados por gestos, que serão executados a uma distância do Kinect entre 1,5 a 3m. Para permitir a transição de estado a interface de controle por gestos, baseia-se no conceito de área de ativação. O sistema verifica continuamente o postura da mão do usuário dentro da área de ativação, esperando uma postura que identifique algum gesto predefinido para que realize a ação desejada pelo usuário. Além disso, o sistema não permite transição de estados indesejáveis, isso porque para mudar a interação com a imagens, o usuário deve manter a mão parada sem variação por aproximadamente 2 segundos , para que o sistema possa reconhecer outros gestos, feitos pelo usuário.

A Figura 3.1 mostra a tela do programa MITO, um exame de Ressonância magnética é exibido.



Figura 3.1 - Tela do programa MITO.
Fonte: O autor.

3.3 OPENNI

O OpenNI será utilizado para realizar a comunicação entre os componentes da aplicação, sendo eles os dispositivos de entrada (Kinect), reconhecimento dos gestos e o *software* MITO. As principais razões que levaram a escolha do OpenNI foram o alto nível de abstração, possibilitando a utilização de diversos componentes e permitindo o foco no resultado desejado, e o fato de ter código aberto e divulgado para múltiplas plataformas permite que a mesma aplicação seja utilizada, mesmo que com ajustes posteriores, em diversos ambientes e ampliando a área de atuação do sistema.

3.3.1 Integrando OpenNI ao MITO

A seguir será mostrado os passos para adicionar o OpenNI ao MITO usando o Visual Studio¹⁴.

- O projeto MITO aberto no Visual Studio, Abra a opção propriedade do projeto e configure como mostra os passos a seguir;
- Na opção *Linker*, sob o item *input*, escolha “*Additional Dependencies*” e adicione "OpenNI.lib";
- Na opção C/C++, sob o item "*general*", escolha "*Additional Include Directories*" e adicione "\$ (OPEN_NI_INCLUDE)". Esta é uma variável de ambiente que aponta para a localização do diretório *include* do OpenNI (local padrão é C:\Arquivos de programas\OpenNI\include);
- Na opção *Linker*, sob o item "*general*", escolha "*Additional Library Directories*" e adicione "\$ (OPEN_NI_LIB)". Esta é uma variável de ambiente que aponta para a localização do diretório *lib* do OpenNI (O local padrão é: C:\Arquivos de programas\OpenNI\Lib).

¹⁴ É um pacote de programas da Microsoft para desenvolvimento de software especialmente dedicado ao .NET Framework e às linguagens Visual Basic (VB), C, C++, C# (C Sharp) e J# (J Sharp).

Após a realização dos passos descritos anteriormente o projeto pode utilizar os componentes do OpenNI basta incluir no código as classes ¹⁵ “XnOpenNI.h” e “XnCppWrapper.h”.

3.3.2 Comunicação com o sensor Kinect

Levando em conta as características do OpenNI, qualquer dispositivo de rastreamento 3D poderia ser conectado à aplicação, portanto, a inicialização deste envolve um fase de determinação de suas capacidades. Para isso, inicialmente, é necessário definir um arquivo *extensible Markup Language* (XML) que determina quais recursos serão utilizados pela aplicação.

O arquivo de configuração é composta por TAG "*Production Nodes*" e uma série de TAGs "*Nodes*", os quais compõe os serviços que serão utilizados na aplicação.

Os *Production Nodes* são definidos como um conjunto de serviços necessários para a construção de aplicações que utilizem interação natural (OPENNI, 2011). São utilizados na implementação os seguintes *Production Nodes*:

- *Depth Generator*: Este nó gera o mapa de profundidade da cena;
- *Hand Point Generator*: Dá suporte à detecção e rastreamento da mão.

Também são definidos os nós "*Gesture*", "*Hands*" , "*User*" e "*Scene*", que são responsáveis por prover a captura de gestos, da posição da mão, do usuário e da cena, respectivamente.

Como mostra a Figura 3.2 os nós foram definidos no arquivo XML de configuração. Este arquivo é necessário carregar no programa, já que é responsável pela inicialização do sensor Kinect.

¹⁵ É uma estrutura que abstrai um conjunto de objetos com características similares. Uma classe define o comportamento de seus objetos através de métodos e os estados possíveis destes objetos através de atributos.

```

1 <OpenNI>
2 <Licenses>
3   <License vendor="PrimeSense" key="0KOIk2JeIBYClPWVnMoRkn5cdY4="/>
4 </Licenses>
5 <Log writeToConsole="true" writeToFile="false">
6   <!-- 0 - Verbose, 1 - Info, 2 - Warning, 3 - Error (default) -->
7   <LogLevel value="3"/>
8   <Masks>
9     <Mask name="ALL" on="false"/>
10  </Masks>
11  <Dumps>
12  </Dumps>
13 </Log>
14 <ProductionNodes>
15   <Node type="Depth">
16     <Configuration>
17       <MapOutputMode xRes="640" yRes="480" FPS="30"/>
18       <Mirror on="true"/>
19     </Configuration>
20   </Node>
21   <Node type="Scene" />
22   <Node type="Gesture" />
23   <Node type="Hands" />
24   <Node type="User" />
25 </ProductionNodes>
26 </OpenNI>

```

Figura 3.2 – Arquivo XML de configuração.
Fonte: O autor.

3.4 NITE

O NITE será utilizado para fazer o reconhecimento e o controle por gestos. O controle utilizado é o gestos das mãos, os dados são adquiridos através do analisador de cena definidos pelo OpenNI na Seção 3.3. Os gestos das mãos são interpretados de forma a gerar uma ação.

3.4.1 Integrando NITE ao MITO

A seguir será mostrado os passos para adicionar o NITE ao MITO usando o Visual Studio.

- O projeto MITO aberto no Visual Studio, abra a opção propriedade do projeto e configure como mostra os passos a seguir;
- Na opção *Linker*, sob o item *input*, escolha “*Additional Dependencies*” e adicione "XnVNite.lib";

- Na opção C/C++, sob o item "*general*", escolha "*Additional Include Directories*" e adicione "\$\$(XN_NITE_INSTALL_PATH)" seguido por `\include` (o local padrão é C:\Arquivos de programas\Sense Prime\NITE\Include);
- Na opção *Linker*, sob o item "*general*", escolha "*Additional Library Directories*" e adicione "\$\$(XN_NITE_INSTALL_PATH)" seguido por `\lib` (O local padrão é: C:\Arquivos de programas\Sense Prime\NITE\Lib).

Após a integração do NITE ao MITO, este terá todos os recursos necessários para o reconhecimento de gesto, atualização da mão do usuário e tratamentos de sessões.

3.4.2 Reconhecimento de gestos

Com a utilização dos recursos de profundidade da cena, reconhecimento de usuário, determinação da posição da mão, oferecidos pelo OpenNI, o reconhecimento de gestos de cada um dos respectivos objetos se torna um canal de comunicação que possibilita o acesso e manipulação dos dados do sensor. Desta forma, os procedimentos para utilização da posição da mão e detecção de usuário são efetuados com auxílio dos métodos contidos nos respectivos objetos.

Reconhecimentos dos gestos e atualização da posição da mão é efetuado através no mecanismo de eventos da classe *XnVSessionManager*. Com o início da sessão é possível monitorar a ocorrência de ações do usuário através de classe relacionadas ao tipo de ação desejada para tratar esses eventos.

Os gestos quando detectados por sua vez tem a função de tratar o evento recebido e efetuar as ações necessárias. Os métodos tem o objetivo de mostrar o tratamento dado quando se tem um evento que determina um gesto predefinido. Observa-se o Quadro 3.1 mostra o mapeamento de gesto (*Push, Swipe, Steady, Wave e Circle*) associados aos métodos correspondentes. Estes métodos efetua a modificação do contexto, ou seja, altera o modo de visualizar as imagens de acordo com a *gesture* ativada e atualização da movimentação da mão.

A seguir será descrito os métodos implementados para ativação das ações Animação, Translação, Rotação e Zoom.

- **SessionProgress:** Este método é chamado assim que a interface do usuário está sendo inicializado. É o primeiro método a ser chamado após a inicialização do sensor. A sessão está sendo processada, ou seja, a espera do movimento(*wave*) que identifica a mão do usuário para fazer a calibração e iniciar o sistema;
- **SessionStart:** Após a calibração e identificação da mão do usuário a sessão é iniciada, a partir deste momento a interface do usuário junto ao sensor Kinect espera a realização de algum movimento que identifique com os gestos implementados. São eles *wave*, *push*, *circle*, *swipe* e *steady*;
- **SessionEnd:** Este método é chamado caso o sistema perca a localização da mão do usuário por algum motivo depois de um certo período de tempo cerca de 10 segundos. Principais motivos de perda de sessão a mão do usuário não está na área de visão do sensor Kinect, ou usuário está muito distante. Caso a sessão seja finalizada o usuário deve iniciar o sistema novamente;
- **OnWaveCB:** Verifica se a mão efetuou trocas de direção, movimentos com a mão da esquerda para a direita e vice-versa, dentro de um espaço de tempo pré-determinado, caso isso ocorra este método é chamado e a *gesture* é ativada(*Gesture = WAVE*), a ação animação de imagens será ativada, e o usuário pode fazer movimentos com a mão que descrevem uma linha horizontal tornando o sistema a avançar uma série de imagens;
- **Push_Pushed:** Verifica se ocorreu um movimento de aproximação e afastamento do sensor, caso isso ocorra este método é chamado e a *gesture* é ativada(*Gesture = PUSH*), o que torna a ação de transladar a imagem ativa, ou seja, deve ser feito movimento em qualquer direção, a imagem segue os movimentos da mão do usuário, acompanha a velocidade e posição das mãos;
- **UpSwipeCB:** Verifica se ocorreu um movimento na vertical, sentido de baixo para cima, o método é chamado e a *gesture* é ativada (*Gesture = SWIPEUP*), o que torna a ação de zoom+ ativa, ou seja, deve ser feito movimento para cima e a imagem será ampliada conforme o movimento da mão do usuário;
- **DownSwipeCB:** O método é chamado quando ocorre um movimento na vertical, sentido de cima para baixo, o método é chamado e a *gesture* é ativada (*Gesture = SWIPEDOWN*), o que torna a ação de zoom- ativa, ou seja, deve ser feito movimento para baixo e a imagem será reduzida conforme o movimento da mão do usuário;

- **CircleCB:** Aguarda que o usuário efetue um movimento circular completo com a mão, ao efetuar este gesto o método é chamado e a *gesture* é ativada (Gesture = CIRCLE) o que torna a ação rotação ativa, Assim o usuário deve fazer movimentos circulares com a mão sentido horário ou anti-horário para rotacionar a imagem;
- **OnPointUpdate:** Este método é chamado para atualizar a posição da mão do usuário, tanto antes de realizar o gesto reconhecido, como depois. Um gesto só pode ser reconhecido se o ponto da localização da mão do usuário é diferente de *null*, ou seja, o posição da mão deve está dentro da área de ativação do sistema;
- **OnSteadyCB:** É ativado se a mão ficar parada, o método é chamado e a *gesture* é ativa (Gesture = STEADY). Esta ação foi implementada para determinar quando terminar a ação, por exemplo, se estiver em uma ação de rotação da imagem, esta ação só será finalizada quando a mão ficar parada por um tempo mínimo de 2 segundos e a *gesture* parada for identificada.

Quadro 3.1 - Mapeamentos de gesto.

Gesto	Método	Ação
Push	Push_Pushed	Translação
Swipe	UpSwipeCB	Zoom
Steady	OnSteadyCB	Parado
Wave	OnWaveCB	Animação
Circle	CircleCB	Rotação

Fonte: O autor.

3.5 CONTROLE POR GESTOS

A metodologia define um fluxo de execução para interação com o interface. A sequência de atividades pode ser visualizadas na Figura 3.3.

Para começar a controlar o dispositivo, deve-se primeiro iniciar a sessão de reconhecimento de gestos. Isto é feito através da realização de um gesto pré-definido através do gesto "wave". as mão deve está aberta em direção do sensor, movê-la várias vezes da esquerda para a direita.

A captura da posição da mão é um pré-requisito para a realização de gestos. Logo, a operação de inicialização da posição deve ser efetuada toda vez que o usuário iniciar o sistema. Caso a posição da mão seja perdida, seja por um oclusão ou pela ausência da mão do usuário no campo de visão do sensor, deve-se repetir o gesto de inicialização. Essa operação de inicialização.

Após o início da sessão, tem-se a atualização da posição da mão e a verificação se algum gesto está sendo realizado.

Se ocorrer algum gesto, este é identificado e caso esteja mapeado alguma ação, e efetua a ação correspondente, e atualiza as modificações necessárias no ambiente até que o gesto seja finalizado. Qualquer gesto pode ser finalizado a qualquer momento basta manter a mão parada por aproximadamente 2 segundos.

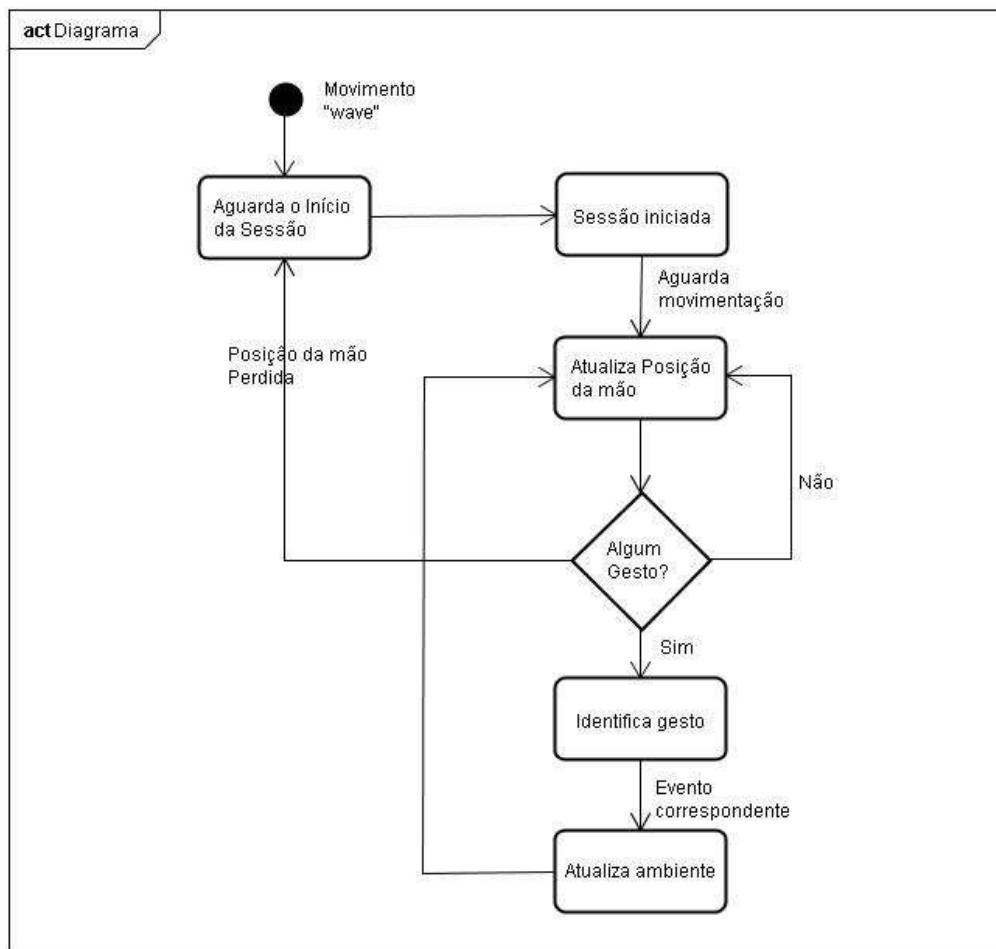


Figura 3.3 - Sequência de atividade para interação.

Fonte: O autor.

Condições para uso da interface:

- Tentar manter a mão que executa o gesto, a uma distância do seu corpo;
- Para obter o gesto de foco o movimento deve ser composto de pelo menos 4 movimentos horizontais (esquerdo-direito ou direito-esquerdo);
- Se tiver dificuldade de obter o gesto de foco, tentar ficar mais perto do sensor, e certificar-se que sua mão está dentro do campo de visão;
- Tentar manter a mão de controle longe de objetos, incluindo a outra mão, cabeça ou corpo. Se outro objeto está perto da mão o ponto de controle pode passar para o outro objeto.

3.6 TÉCNICAS DE INTERAÇÃO

Nesta seção será descrito, as principais funcionalidades do manipulador de imagens médicas 3D bem com forma de utilização e funcionamento das operações.

3.6.1 Animação

Para essa funcionalidade, o usuário com a mão dentro da área de ativação, deve fazer movimentos com a mão indo da esquerda e para a direita e vice-versa descrevendo uma linha horizontal tornando o sistema a avançar uma série de imagens, quanto mais rápido se move mais rápido as imagens deslizam. Desta forma, movendo-se lentamente, o usuário é capaz de selecionar uma única imagem, mesmo em um grande conjunto de dados médicos.

A Figura 3.4 mostra o procedimento para realizar a animação.



Figura 3.4 - Movimento para realizar a animação.
Fonte: O autor.

3.6.2 Zoom

O *zoom* é ativado movendo a mão aberta dentro da área de ativação em direção vertical de baixo para cima e vice-versa descrevendo uma linha vertical, O sistema irá identificar e realizar o *zoom* na imagem de acordo com o movimento da mão, movimento para cima será aplicado o *zoom in* na imagem, movimento para baixo será aplicado *zoom out* na imagem.

A Figura 3.5 mostra o procedimento para realizar o zoom.



Figura 3.5 - Movimento para realizar o zoom.
Fonte: O autor.

3.6.3 Translação

A translação é ativada com a mão dentro da área de ativação aproximando a mão do sensor Kinect, assim que a translação é ativada deve ser feito movimento em qualquer direção, a imagem segue os movimentos da mão do usuário, acompanha a velocidade e posição das mãos do usuário, quando o usuário move as mãos lentamente o movimento de translação é reduzido.

A Figura 3.6 mostra o procedimento para realizar a translação.



Figura 3.6 - Movimento para realizar a translação.
Fonte: O autor.

3.6.4 Rotação

Para entrar neste estado, a mão do usuário dentro da área de ativação fazendo movimento de círculo em direção ao sensor Kinect. Assim que a mão do usuário realiza um círculo completo, sentido horário ou anti-horário o movimento rotação é ativo.

A Figura 3.7 mostra o procedimento para realizar a rotação.



Figura 3.7 - Movimento para realizar a rotação.
Fonte: O autor.

Este capítulo apresentou os principais tópicos para este trabalho como a interface do usuário, os detalhes mais relevantes na implementação, descrição dos métodos implementados e como os métodos foram utilizados.

Abordou-se os aspectos de interação do usuário com a interface gráfica da aplicação, bem com os procedimentos utilizados pela detecção de gestos, comunicação com o sensor Kinect e interação com o usuário.

Através deste mecanismo foi desenvolvida a metodologia para um manipulador de interação com imagens médicas em 3D usando o Kinect como dispositivo de entrada. Com a finalidade de condensar a contribuição deste trabalho, o próximo capítulo trata dos resultados obtidos.

4. EXPERIMENTO E RESULTADOS

Foram realizados testes com a implementação da metodologia proposta, onde foi possível avaliar visualização das imagens TC e RM, interagindo com o usuário usando a tecnologia Kinect. Além disso, foi identificadas condições para que o usuário não tenha problema durante a manipulação com as imagens. Os testes foram realizados usando os seguintes recursos:

- Notebook: Processador Intel Core i5-2430M CPU 2.40 GHz, 4 GB RAM, placa de vídeo integrada:
- Sensor Kinect: Lentes com detecção de cores e profundidade, motor de inclinação para ajuste vertical e com fluxo de dados 30 frames por segundo(FPS).
- Ambiente: Iluminação do ambiente bem distribuída, sem grandes focos de luz e sombra e sem fonte de luz em direção ao sensor Kinect.
- Imagens: Imagens médicas em 3D, padrão DICOM foi selecionado um exame de RM da região do cérebro com 22 fatias e um exame TC com 82 fatias da coluna cervical.
- Drive do Sensor kinect : É instalado ao conectar o kinect no USB do computador;

Após toda implementação, durante os testes o procedimento de reconhecimento de gestos se mostrou eficiente, cada gesto foi identificado como era esperado e realizando a ação correspondente.

A Figura 4.1 mostra a iteração do software com o Kinect, com uma postura ideal para calibração da mão.



Figura 4.1 - Usando o software com o Kinect (calibração).
Fonte: O Autor.

A Figura 4.2 um dos casos de teste de interação com o sistema usando uma imagem de ressonância magnética. O gesto *circle* é identificado e é aplicado a rotação na imagem.

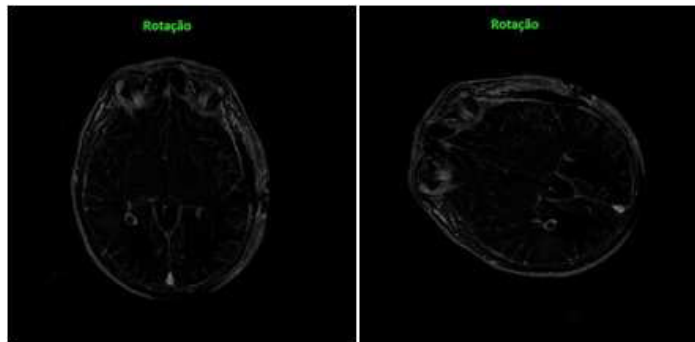


Figura 4.2 - Interagindo com o sistema (rotação).
Fonte: O autor.

A Figura 4.3, mostrar um caso em que o gesto *swipe* foi identificado e será aplicado o *zoom*. Movimento para cima será aplicado o *zoom in* na imagem, movimento para baixo será aplicado *zoom out*.

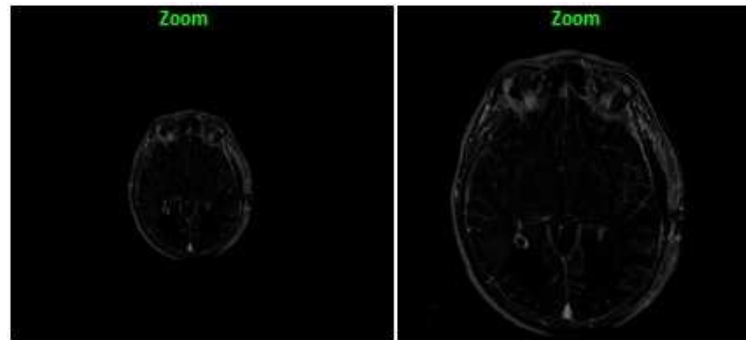


Figura 4.3 - Interagindo com o sistema (zoom).
Fonte: O autor.

A Figura 4.4 mostra o momento em que o gesto *push* é reconhecido e a imagem será transladada de acordo com o movimento da mão do usuário.

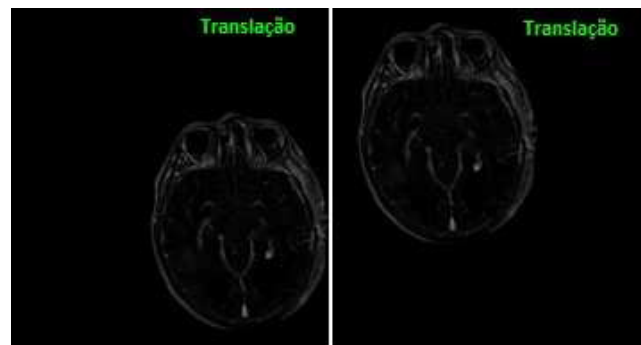


Figura 4.4 - Interagindo com o sistema (translação).
Fonte: O autor.

E por fim, a Figura 4.5 mostra a funcionalidade de animação, será aplicado no momento que o gesto *wave* for reconhecido, o usuário deve movimentar as mão da esquerda para direita ou vice-versa.

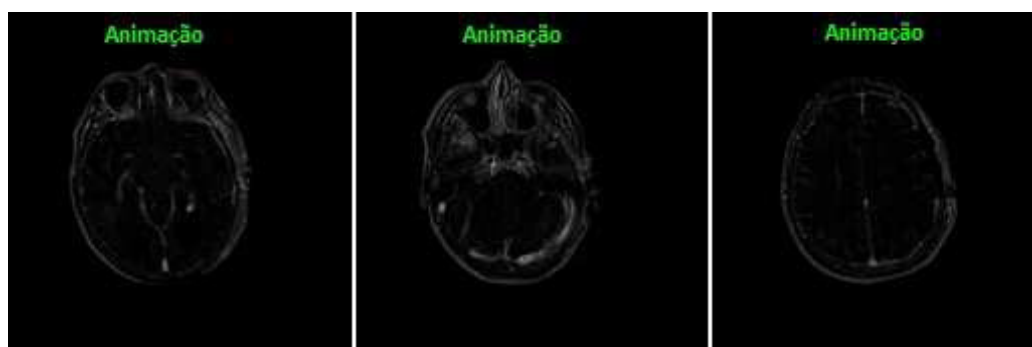


Figura 4.5 - Interagindo com o sistema (animação).
Fonte: O autor.

Além disso foi constatado, durante os testes, a existência de alguns problemas que podem ocorrer durante a iteração. Os problemas encontrados foram:

- A detecção da posição dos braços torna-se menos precisa quando estes estão próximos ao tronco da pessoa detectada ou quando existe sobreposição dos braços;
- Mau desempenho quando a distancia do usuário para o sensor Kinect estiver fora da distancia ideal (entre 1 a 3,5 metros).
- Qualquer tipo de iluminação direcionada diretamente para o Kinect ou a iluminação do ambiente não estiver bem distribuída, há dificuldades para detectar o usuário.
- Em alguns casos movimentos muito rápidos não foram detectados e geraram posições incorretas para o modelo;
- A ocorrência das situações anteriores pode gerar a necessidade de uma nova calibração do usuário ao modelo para que a detecção seja estabilizada.

5. CONCLUSÃO

Este trabalho apresentou o desenvolvimento de uma metodologia para interação de imagens médicas em 3D. Foram utilizadas técnicas de visão computacional e o Kinect juntamente com o *framework* OpenNI e o *middleware* NITE, os quais consistem em módulos de *software* de uso específico para aplicações com interação por gestos.

O desempenho do sistema baseado na implementação desta metodologia é satisfatório. Os gestos são reconhecidos na maioria das vezes quando em condições adequadas (distância em relação ao sensor Kinect e posição da mão dentro da área de ativação do sistema), sendo que um gesto é reconhecido de cada vez. Não há transição de estados indesejáveis, ou seja, o sistema fica bloqueado enquanto uma *gesture* está ativa e não permite ir para outra sem que a mesma seja finalizada.

Além disso, o dispositivo Kinect mostrou nos testes a eficiência na interação com o reconhecimento dos gestos e atualização do ponto de localização da mão, fazendo com que as manipulações de escala, rotação, translação e animação nas imagens, quando ativadas ocorram instantaneamente.

Em determinadas condições o sistema pode ter mau desempenho, caso o usuário esteja fora da distância ideal (entre 1 a 3,5 metros) para o sensor Kinect, ou movendo a mão para fora da área de ativação do sistema, neste caso a localização da mão é perdida e o usuário deve iniciar novamente a calibração da mão com o sensor.

Com o sistema concluído pode-se entender o quão poderosa esta tecnologia Kinect tem mostrado nas aplicações de interação homem-máquina. Afinal, pode-se utilizá-la no dia a dia simplesmente com o uso de um computador sem a utilização de um objeto físico, no caso, o mouse ou teclado, para interação com o usuário. Além disso, tal sistema também pode ser utilizado para manipulação de imagens de tomografia computadorizada e ressonância magnética em salas de cirurgias. Desde que as imagens pertençam ao padrão DICOM.

Com as técnicas apresentadas nos capítulos anteriores obteve-se bons resultados. No entanto, ainda têm-se várias propostas e técnicas a serem implementadas, como trabalhos futuros:

- A substituição do *framework* OpenNI pelo Microsoft Kinect SDK, que em sua versão mais recente não exige calibração dos usuários e possui uma estimação das articulações mais correta;

- Estender a técnicas de manipulação para imagens 3D, já que o MITO versão 2.0, faz a reconstrução da imagem 3D, baseado nas fatias;
- Informar que um gesto não foi reconhecido, para que o usuário ajustar a sua mão em relação ao sensor Kinect;
- Adicionar novas técnicas para manipular as imagens, como contraste, e selecionar região de interesse.

REFERÊNCIAS

Alfano, B. et al. **Medical Imaging TOolkit(MITO)**. 2009. Disponível em: <<http://amico.icar.cnr.it/mito.php>>. Acesso: Fev/2012.

CHU, Gean Francesco Derosso; FILHO Gilberto José Verona. **Interação homem-máquina por meio de gestos manuais naturais(hand flow)**. 2011. Trabalho de Conclusão de Curso. (Graduação em Engenharia de Computação) - Pontifícia Universidade Católica do Paraná. Disponível em: <<http://www.ppgia.pucpr.br/~laplima/ensino/pfec/concluidos/2011/handflow.pdf>>. Acesso: Maio/2012.

Dutra, Dayvison. **Gerações de Aparelhos de Tomografia Computadorizada**. 2009. Disponível em <<http://dicasderadiologia.com.br/site/2009/08/geracoes-de-aparelhos-de-tomografia-computadorizada/>>. Acesso: JUL/2012.

Gallo, Luigi; Placitelli, Alessio Pierluigi; Ciampi, Mario. **Controller-free exploration of medical image data: experiencing the Kinect**. IEEE International Symposium on Computer-Based Medical Systems, CBMS 2011, (Los Alamitos, CA, USA), pp. 1–6, IEEE Press, 2011. doi:10.1109/CBMS.2011.5999138.

JUNIOR, Edson; YAMASHITA Helio. **Aspectos básicos de tomografia computadorizada e ressonância magnética**. Revista Brasileira de Psiquiatria, volume.23 suppl.1 São Paulo. Maio,2001. Disponível em: < <http://www.scielo.br/pdf/rbp/v23s1/5560.pdf> >. Acesso: Maio/2012.

MICROSOFT. Kinect SDK. nov 2011. Disponível em: <<http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/>>. Acesso: Jun/2012.

Mitra, S; Acharya, T. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(3):311–324, 2007.

Mlazarus, **Microsoft Kinect in Medicine**. April 28, 2011. Disponível em: < <http://www.healthgamers.com/2011/using-health-games/microsoft-kinect-in-medicine/>>. Acesso: Fev/2012.

Monteiro, Denyse N. B. **Estudo sobre a Visualização de Imagens obtida por Exames Virtuais**. Dissertação (Mestrado em Computação Visual e Interfaces)- Universidade Federal Fluminense , Niterói, 2005. Disponível em: <

http://www.bdt.d.ndc.uff.br/tde_arquivos/41/TDE-2008-03-06T121916Z-1367/Publico/Dissert-DM.pdf >. Acesso: Jun/2012.

OPENNI. OpenNI User Guide. [S.I.]. Jan 2011. Disponível em: < <http://openni.org/>>. Acesso: Fev/2012.

PREECE et al, 1994) PREECE, j. et al. **Human-Computer Interaction**. : Addison-Wesley, England, 1994.

PRIMESENSE. NITE Control 1.3 - Programmer's. [S.I.]. Jan 2011. Disponível em: <pr.cs.cornell.edu/humanactivities/data/NITE.pdf > . Acesso: Jan/2012.

STEAKLEY, L. Canadian hospital tests Kinect in the operating room, March 2011. Disponível em: < <http://scopeblog.stanford.edu/archives/2011/03/kinect-in-the-operating-room.html>>. Acesso : Dez/2011.

Santos, Eduardo de Souza. **Manipulação de objetos 3D em aplicações de realidade aumentada por meio da movimentação da mão**. Dissertação (Mestre em Ciências) Universidade Federal de Uberlândia. 2011.

Shan, C. Gesture control for consumer electronics. In L. Shao et al., editors, *Multimedia Interaction and Intelligent User Interfaces*, Advances in Pattern Recognition, pages 107–128. Springer London, 2010.

Togores, Tiago Andrade. Vitruvius, **Um Reconhecedor de Gestos para o Kinect**. 2011. Trabalho de Conclusão de Curso. (Bacharelado em Ciência da Computação) - Universidade de São Paulo. Disponível em: <<http://www.linux.ime.usp.br/~togores/mac499/pdf/poster.pdf> >. Acesso: Mai/2012.

Valli, Php Alessandro. *Notes on Natural Interaction*. 2007. Disponível em: <<http://www.idemployee.id.tue.nl/g.w.m.rauterberg/Movies/NotesOnNaturalInteraction.pdf>>. Acesso: Maio/2012.

Valli, Php Alessandro. **The Design of Natural Interaction**. 2008 Revised October 28th, 2006. Disponível em: <<http://research.microsoft.com/en-us/um/cambridge/projects/hci2020/pdf/thedesignofnaturalinteraction.pdf>>. Acesso: Maio/2012.