1
Introdução

A visão é provavelmente o sentido humano mais utilizado, principalmente na interação com o ambiente. Sendo assim, a ideia de alimentar um computador de informações provenientes de uma imagem vem instigando a comunidade científica há muito tempo. Esse campo de pesquisa é chamado de *visão computacional*(VC). Em paralelo a esse campo de estudo, as técnicas de *realidade virtual*(RV) fornecem meios para substituir as informações sensoriais humanas por informações geradas por computador, normalmente imagens e sons, a fim de transportar a imaginação do usuário para uma realidade totalmente artificial. Nesse trabalho usamos técnicas de visão computacional e realidade virtual para criar uma *realidade aumentada*(RA). O conceito de realidade aumentada mais aceito hoje em dia é o de Azuma [2], no qual RA é: combinação da realidade com elementos virtuais, interativa em tempo real e registrada em três dimensões. Sendo que, o registro é o alinhamento entre o mundo real e o virtual.

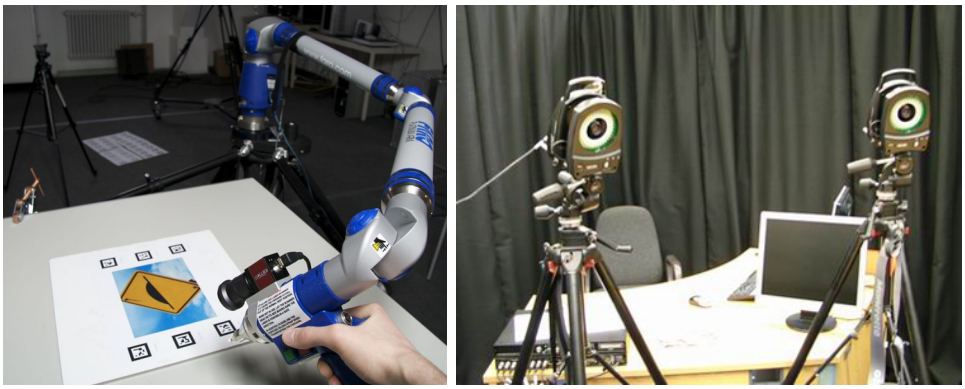
A realidade aumentada de forma geral depende fortemente da capacidade de estimar a posição da câmera dentro do ambiente, pois é através dessa estimativa que será possível escolher e apresentar as informações virtuais corretamente. Essa posição pode ser estimada por diversas técnicas. Entre elas, existem equipamentos mecânicos como, por exemplo, o braço articulado da FARO (figura 1.1(a)) e também sistemas óticos baseados em luz infravermelha como, por exemplo, os sistemas de captura de movimento da Vicon (figura 1.1(b)). Esses dois tipos de equipamento são muito precisos, chegando à precisão milimétrica. No entanto, eles possuem uma limitação de amplitude de movimento e preços na faixa das dezenas de milhares de dólares, o que os torna inviáveis para o público em geral.

Para viabilizar o acesso à realidade aumentada, a comunidade científica foca seus esforços nas técnicas puramente óticas, pois as câmeras de computador, comumente chamadas de webcams, estão amplamente difundidas pelos usuários de computador. Da mesma forma, as câmeras de celulares, também bastante disseminadas, podem ser usadas para realidade aumentada. Sendo assim, para fazer uma aplicação de RA que seja acessível às pessoas comuns

PUC-Rio - Certificação Digital Nº 0721334/CA

hoje em dia, ela precisa ser baseada em visão computacional.

Esta dissertação apresenta um algoritmo para estimação de posição de câmera baseado em visão computacional no intuito de ser viável financeiramente e ao mesmo tempo atender aos requisitos quanto ao tipo de ambiente, amplitude de movimentação de câmera e precisão. Na próxima seção será determinado o conjunto de aplicações em que é esperado que o algoritmo funcione a fim de definir esses requisitos. Em seguida será dada uma ideia geral do algoritmo proposto.



1.1(a): Braço articulado da FARO. Fonte: [25]

1.1(b): Sistema Vicon. Fonte: Laboratorio de Robótica de Briston

Figura 1.1: Sistemas de rastreamento

1.1 Motivação e Objetivo

A motivação desse trabalho é o desafio de localizar um objeto dentro de um ambiente muito grande. Alguns exemplos desse problema são: localizar uma determinada peça dentro de um estoque, onde são armazenadas milhares de diferentes peças; conseguir encontrar um determinado mantimento dentro de um supermercado ou um determinado livro em uma biblioteca[27]. A biblioteca do Trinity College Dublin com 200.000 livros na mesma sala(figura 1.2(a)) e um dos estoques do site de venda Amazon(figura 1.2(b)) do tamanho de 10 campos de futebol são ótimos exemplos de como localizar um objeto específico pode ser muito díficil. Inclusive já existe uma área da Realidade Aumentada estudando como apresentar as informações de forma ao usuário consiga rapidamente recolher(pick) objetos dentro de estoques. Essa área é intitulada Pick-by-Vision [32]

Então, sabendo que todo objeto ocupa um lugar no espaço, ao invés de procurar pelo objeto pode-se procurar pela sua localização no espaço. Sendo assim, o objetivo desse trabalho é *localizar uma posição em um espaço*

PUC-Rio - Certificação Digital Nº 0721334/CA

1.2(a): Câmara principal da biblioteca do Trinity College Dublin. Fonte: Flickr

1.2(b): Estoque do site de vendas eletrônica Amazon. Fonte: The Guardian, fotógrafo Gareth Phillips

Figura 1.2: Exemplos de ambientes onde é dificil localizar um objeto

tridimensional (3D) previamente conhecido, como por exemplo uma biblioteca ou museu específicos.

Para um usuário localizar uma posição 3D específica usando realidade aumentada são necessárias duas informações. A primeira, conhecer essa posição 3D em coordenadas absolutas, enquanto a segunda é conhecer a posição absoluta do usuário no espaço e a direção para onde ele está olhando. Com esses dois dados é possível traçar uma rota que leve o usuário à posição requerida. Como os objetos são considerados estáticos, então as suas posições em coordenadas absolutas podem ser calculadas em pré-processamento. O mesmo não acontece com a posição e orientação do usuário, pois este está em movimento. Sendo assim é necessário rastrear a sua posição absoluta no espaço 3D e orientação. Como neste trabalho vamos utilizar métodos de visão computacional para resolver o problema e viabilizar a realidade aumentada, a posição do usuário é dada pela posição de uma câmera acoplada a ele.

1.2 Abordagem

Para estimar a posição da câmera acoplada ao usuário em coordenadas absolutas propomos uma técnica de rastreamento híbrida baseada em Visual SLAM[14](Simultaneous Localisation and Mapping) e marcadores, tanto os naturais quanto os fiduciais. Os marcadores fiduciais são posicionados em lugares em que não façam oclusão ou atrapalhem a estética do ambiente. Eles são usados para inicializar o sistema de rastreamento. Como marcadores fiduciais do tipo código de barra 2D fornecem um número identificador para o marcador e a posição relativa da câmera ao marcador, o algoritmo proposto

PUC-Rio - Certificação Digital Nº 0721334/CA

usa esse número para consultar uma base de dados e descobrir a posição absoluta do marcador e consecutivamente a posição da câmera. A detecção de marcadores naturais é mais difícil e cara, pois é necessário comparar a imagem da câmera atual com cada um dos possíveis marcadores presentes no ambiente. Essa comparação com um grande número de marcadores naturais é inviável em tempo interativo, então nós usamos a última posição conhecida da câmera para inferir um pequeno grupo de marcadores possíveis, assim viabilizando a detecção de marcadores naturais em tempo interativo(cerca de 3 detecções por segundo). Quando o marcador é identificado, acontece o mesmo processo dos marcadores fiduciais resultando na posição da câmera em coordenadas absolutas. Com esses marcadores conseguimos descobrir a posição da câmera apenas nos momentos em que a câmera esteja olhando para um marcador. Para não ser necessário ter marcadores em todos os lugares do ambiente, propomos usar um algoritmo de Visual SLAM para estimar a posição da câmera nos momentos em que não haja marcadores disponíveis. Chamamos esta abordagem de Local SLAM, visto que, enquanto os algoritmos de SLAM armazenam todos os pontos mapeados, o algoritmo de SLAM proposto só armazena dados de mapeamento localmente. Os detalhes sobre os algoritmos citados serão mais bem detalhados na seção de trabalhos relacionados.

1.3 Organização

Esse trabalho é organizado da seguinte forma. No capítulo 2 são apresentados os trabalhos relacionados. Em seguida,no capítulo 3, o algoritmo Local SLAM é explicado. No capítulo 4 são explicadas a implementação, as condições de testes e os testes propriamente ditos. Por fim, as conclusões e trabalhos futuros são expostos.

PUC-Rio - Certificação Digital Nº 0721334/CA