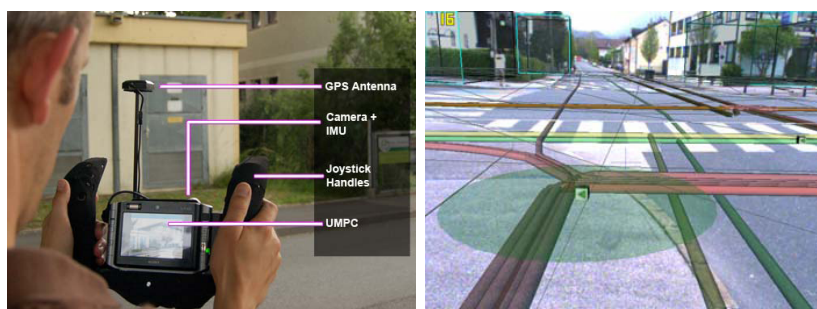


2

Trabalhos Relacionados

Ao se pensar em encontrar uma posição no espaço 3D, uma ideia seria utilizar o sistema de posicionamento global (GPS). No entanto, o GPS comum tem alguns inconvenientes, como por exemplo, não funcionar dentro de edificações e ter um erro de posicionamento de cerca de 10 a 15 metros. Em [30] é apresentado um sistema para localização precisa de dutos embaixo das ruas, onde se utiliza além de um GPS diferencial (DGPS), sensores inerciais e câmeras (figura 2.1). O GPS diferencial apresenta uma maior precisão em relação ao GPS comum, chegando a alcançar um erro entre 1 e 3 metros. Mesmo assim, o DGPS não foi suficiente para localizar a câmera em relação aos dutos, sendo necessária para isso a adição de sensores inerciais e algoritmos de visão computacional. O trabalho[19] utiliza o mesmo tipo de solução para fazer realidade aumentada em ambientes ao ar livre.



2.1(a): Equipamento para visualização
2.1(b): Imagem visualizada pelo usuário.

Figura 2.1: Sistema de localização de dutos sob o chão. Fonte: [30]

Sistemas de rastreamento mecânicos e por luz infravermelha foram descartados nessa dissertação por serem muito caros e de amplitude demasiadamente limitada, como explicado na seção 1. No caso do infravermelho temos o agravante dele não permitir nenhuma fonte de luz infravermelha externa, como uma lâmpada de luz quente ou até mesmo uma janela por onde possa entrar luz natural. O rastreamento infravermelho vem sendo usado para pesquisa de sistemas de localização no âmbito industrial enquanto não há uma solução com a mesma precisão e facilidade de uso [32, 29, 35].

As soluções baseadas em visão computacional sem iluminação infravermelha são as opções realmente viáveis a fim de atingir nossos objetivos. Existem vários sistemas de localização usando visão. Cada um deles tem especificações diferentes quanto ao tipo de preparação do ambiente necessário, a amplitude de rastreamento e a precisão do rastreamento. Na tese de doutorado de Klein[23], ele faz um resumo bastante extenso sobre os todos os principais trabalhos de rastreamento visual. Abaixo será feito um resumo de alguns dos sistemas mais fortemente relacionados a essa dissertação:

Wagner and Schmalstieg[38] apresentam um sistema de localização dentro de um edifício espalhando marcadores fiduciais (figura 2.2). No início da utilização, o usuário seleciona no PDA(Personal digital assistant - Computador de mão) para qual sala ele deseja ir e então aponta o PDA com uma câmera acoplada para um marcador. Esses marcadores ficam distribuídos por todas as paredes em média a cada metro. A imagem capturada pelo PDA é processada no próprio PDA ou enviada para um servidor. O resultado desse processamento é a posição 3D do PDA em relação ao edifício e a direção para a sala requisitada. Então o sistema apresenta uma seta indicando para onde o usuário deve seguir para conseguir chegar ao destino desejado. O custo de processamento desse sistema é sua principal qualidade. O processamento de um único marcador na cena é barato, pois são marcadores fiduciais. Os marcadores usados nesse trabalho são similares a um código de barra 2D, codificando em si um número. Com esse número é fácil consultar uma base de dados para saber qual a posição 3D do marcador e então derivar a posição do PDA. Em contrapartida, a interferência no ambiente e a inexistência de rastreamento do usuário entre dois marcadores inviabilizam o seu uso para localizar objetos pequenos. Além de na maioria dos lugares ser impossível colocar tantos marcadores como é necessário nesse sistema.



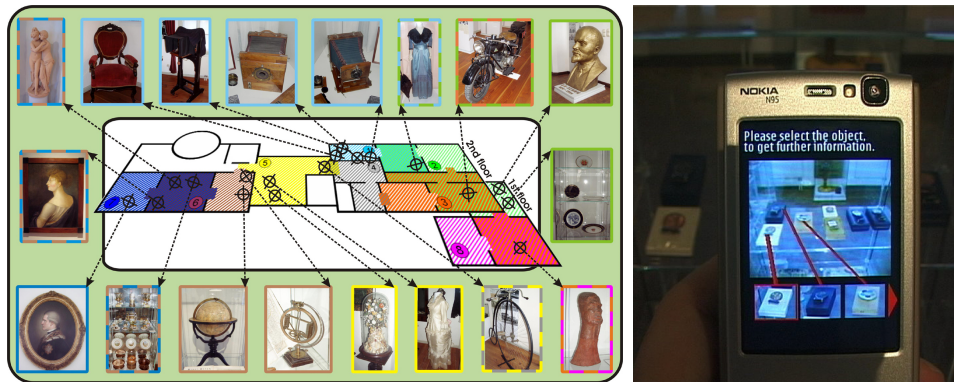
2.2(a): Sistema PDA e marcadores para localização



2.2(b): Aplicação de RA sendo apresentada ao usuário

Figura 2.2: Sistema de RA de Wagner e Schmalstieg. Fonte: [38]

Föckler et al[18, 9] apresentam um sistema que é capaz de auxiliar o visitante de um museu(figura 2.3(a)) na forma de um guia virtual. O sistema mostra informações em texto e em vídeo na tela de um celular após o usuário apontar a câmera do celular para o objeto do museu, como mostrado na figura 2.3(b). O sistema funciona por reconhecimento de padrões e RFID. Na fase de preparação do ambiente, fotos de cada objeto são usados para treinar uma rede neural que será utilizada para reconhecer os objetos posteriormente. Os Emissores de RFID são usados para definir regiões do museu e, com isso o software de reconhecimento é capaz de eliminar da lista de candidatos os objetos de salas que não estão próximas a um determinado emissor, melhorando o desempenho e a precisão. A melhor qualidade desse sistema é que ele não é invasivo ao ambiente, uma vez que os marcadores são os próprios objetos do museu. Esse tipo de marcador recebe o nome de marcador natural. Os emissores de RFID(Radio-Frequency IDentification) também são pequenos, do tamanho de duas pilhas AA. No teste feito pelo autor, em um museu de dois andares com 13 salas e 155 objetos, foram necessários apenas 8 emissores. O problema desse sistema é que ele foi projetado para dar informações quando o usuário está em frente ao objeto. O máximo que essa abordagem pode fornecer é uma posição grosseira do usuário no momento em que ele se encontra em frente a um objeto, pois é apenas um sistema de identificação de objetos, não um rastreador.



2.3(a): Mapa dos objetos dentro do museu passíveis de serem reconhecidos pelo sistema 2.3(b): Aplicação de RA sendo apresentada ao usuário

Figura 2.3: Sistema de RA Phone Guide. Fonte: [18]

Davison [14] apresenta um sistema de SLAM (Simultaneous Localisation and Mapping). Sistemas de SLAM são capazes de, a partir de uma posição conhecida, se mover no espaço e manter sua posição conhecida ao mesmo tempo em que geram um mapa do ambiente ao redor. Esses sistemas funcionam de

forma incremental. A partir de uma posição conhecida, ele mapeia novos pontos que estejam no campo de visão. Com esses novos pontos, que abrangem uma área maior que a inicial, a câmera é deslocada. Nesse novo ponto de vista novamente existem pontos conhecidos e pontos desconhecidos que são mapeados e assim recursivamente. A implementação de Davison foi revolucionária, pois foi a primeira a atingir tempo real de processamento em um computador de mesa comum e a utilizar imagens provenientes de uma única câmera como dado de entrada. Esse sistema rastreia os pontos característicos usando a correlação cruzada e um pequeno retalho 2D ao redor dele. O sistema é inicializado a partir de um retângulo preto sobre fundo branco. Esse objeto geométrico fornece os 4 pontos com posições conhecidas, sendo que o centro do sistema de coordenadas é o centro do retângulo. Dando as coordenadas iniciais desses quatro pontos no sistema métrico, o mapeamento também será métrico. Em seguida o sistema procura pontos na imagem que se sobressaíam segundo o detector de Shi-Tomasi[33]. Ele escolhe o ponto mais nítido e tenta adicioná-lo no mapa. Essa adição é bem sucedida caso a câmera se movimente o suficiente para calcular a profundidade do ponto em relação à câmera por triangulação estéreo. O estado do sistema é armazenado em um filtro de Kalman estendido. Esse filtro é o responsável por armazenar a posição da câmera, a velocidade da câmera, as posições das características e a incerteza associada a cada variável do estado. O tempo real desse sistema é atingido graças ao uso de uma técnica chamada busca ativa(Active Search). Normalmente, a busca é feita de forma passiva, ou seja, o pareamento é feito da seguinte maneira: os pontos característicos no quadro atual são procurados no quadro anterior. Essa busca é feita delimitando, no quadro anterior, uma área de busca de tamanho fixo ao redor da posição do ponto no quadro atual. Caso no quadro anterior tenha, dentro da área de busca, um ponto com características semelhantes, então eles são considerados correspondentes. Já na técnica de busca ativa, o ponto do quadro anterior, que no nosso caso pertence ao mapa e está armazenado no filtro de Kalman, é buscado no quadro atual. Além disso, a busca é inteligente, pois a área de busca é definida pelo preditor do filtro de Kalman e pela incerteza associada a cada ponto característico. Quanto maior a incerteza, maior a área de busca. Como o movimento dos pontos característicos que são candidatos a entrar no mapa não pode ser descrito por incertezas gaussianas, pois sua profundidade em relação a câmera não é conhecida, Davison usa um filtro de partícula para prever a posição desses pontos no próximo quadro. Apenas quando esse filtro converge para um comportamento parecido com o da função densidade de probabilidade(pdf) de uma gaussiana da profundidade, em outras palavras, quando o filtro de partícula apresenta um comportamento parecido

com o de uma gaussiana, o ponto candidato passa a pertencer ao mapa. Esse é o ponto no qual a profundidade é suficientemente conhecida para ser inserida no mapa. Caso essa convergência não aconteça em um número determinado de quadros, o ponto característico é descartado da lista de candidatos.

As qualidades desse sistema são muitas. Em princípio, qualquer posição 3D, se posicionada em relação ao retângulo preto que inicializa o sistema, pode ser alcançada. Isso significa que nenhum tipo de modificação seria necessária no ambiente além de fixar um único marcador em uma parede. Mas esse sistema também tem problemas. O tamanho do estado do filtro de Kalman cresce quadraticamente em relação ao número de pontos característicos no mapa, o que, segundo o próprio autor, deixa de ser computável em tempo real na faixa de 400 pontos. Eade[1] propôs uma estratégia usando um grafo espacial que indica para cada ponto quais outros pontos precisam ter suas covariâncias atualizadas. Então esse problema pode ser significativamente reduzido. Um grave problema do sistema de Davison é o erro acumulado no mapa. Como os novos pontos são calculados a partir de pontos que foram anteriormente estimados, por menor que seja o erro em cada medida, para pontos distantes dos quatro pontos que inicializaram existe um erro acumulado potencialmente grande. Infelizmente, nesse sistema, é quase impossível fazer uma volta fechada, em torno do usuário segurando a câmera, em que ao voltar à posição de inicialização do sistema ele consiga identificar os pontos iniciais como sendo, no mapa, aqueles que foram iniciados na primeira volta.

Esse problema é conhecido como 'loop-closing' (fechamento de volta) e é bem discutido na comunidade científica, como pode ser visto nos trabalhos [13, 39, 15]. Todas essas técnicas de volta fechada têm como princípio identificar o momento em que a volta se fechou para corrigir a posição de todos os pontos do mapa e, com isso os erros de acumulação seriam eliminados. No entanto essa ideia tem um problema em sua essência. Enquanto a volta não se fechar os pontos continuam errados, além de a correção dos pontos não acontecer em tempo real. No nosso caso, de localização de posições 3D no espaço, podemos ter uma fase de preparação, apesar de não ser a ideia inicial do SLAM. No entanto, para podermos ter os benefícios da correção do mapa através de uma volta fechada é necessário fazer, na fase de preparação do ambiente, o mapa com quantas voltas fechadas sejam necessárias e, com isso, não usar mais apenas os quatro pontos inicialmente conhecidos, mas todos os que já foram calculados na fase de preparação. Essa ideia funcionaria se considerarmos que o ambiente não sofre nenhuma mudança significativa de iluminação ao longo do tempo. Como isso não é verdade, então a ideia de deixar um mapa completo armazenado não é viável, pois a iluminação vai mudar ao longo do tempo e

o pareamento, pixel a pixel (block-matching), dos retalhos armazenados no estado do filtro de Kalman com os capturados em tempo real com a câmera, não funcionará, já que esse método não é invariante à mudança de iluminação.

O último problema dessa abordagem para a localização de uma posição 3D no espaço é a inicialização. A implementação de Davison considera que o movimento sempre será inicializado com a câmera apontando para o retângulo preto que é o centro do sistema. Esse tipo de requisito torna o algoritmo de rastreamento muito limitado, dado que para qualquer problema em que o rastreamento se perca, é necessário retornar à posição do retângulo preto.

Castle et al [11, 12] propõem uma modificação no sistema de Davison. Eles sugerem adicionar uma base de dados com imagens (texturas) que estão presentes na cena, mas sem necessidade de conhecer a posição das imagens no espaço. Com isso, em paralelo ao processo do SLAM de Davison, é mantido um processo que procura por texturas conhecidas. Quando a textura é localizada, os vértices da caixa envolvente são adicionados ao mapa. Com essa abordagem ele é capaz de reinicializar constantemente os pontos da textura, ajudando na estabilidade do rastreamento. Essa modificação mantém os problemas do sistema de Davison, mas melhora a estabilidade e a tolerância às falhas. Além disso, o sistema pode ser reinicializado por uma das texturas mapeadas. Quanto ao erro acumulado, como as texturas são mapeadas pelo processo de SLAM de Davison, elas também possuem erro de posicionamento. Outro problema apresentado por essa abordagem é que o processo de localização de texturas utilizado, o SIFT [26], é bastante caro computacionalmente. Segundo testes realizados por Castle et al., um banco de dados de apenas 16 objetos já possui cerca de 30.000 pontos chave para serem comparados com os pontos recuperados da imagem da câmera. Assim, podemos inferir que, mesmo com uma organização em árvore, esse banco de dados não pode crescer muito, pois corre o risco de tornar a detecção de texturas incomputável em tempo aceitável, além de agravar o problema da ambiguidade gerado pelo excesso de texturas parecidas.

Klein e Murray [24] apresentaram um sistema chamado de 'Parallel Tracking and Mapping for Small AR Workspaces' (PTAM). Eles argumentam que uma pessoa segurando uma câmera é muito diferente de um robô se movendo. Sendo assim, um filtro de Kalman estendido não é suficiente para modelar o problema. Esse algoritmo separou em dois processos independentes a localização (rastreamento) e o mapeamento. Ambos os processos são modelados com técnicas bastante conhecidas em outras áreas da visão computacional. O mapeamento passa a ser tratado como um problema de Structure-from-Motion (SfM) a partir de quadros chave. O algoritmo escolhe alguns quadros

suficientemente distantes e então alimenta o algoritmo de SfM. Esse algoritmo faz uma otimização local entre os quadros mais próximos, além da otimização global, mais cara, do mapa inteiro. Essa otimização global tem os mesmos efeitos das otimizações utilizadas no fechamento de volta (loop-closing). Com isso o mapa tem uma qualidade bastante superior a dos algoritmos baseados em filtro de Kalman, principalmente pela falta de otimização, como demonstrado nesse trabalho de Klein e Murray. Em contrapartida essa otimização é muito cara. Ela que é a maior responsável pelo nome 'Small AR Workspaces', pois, segundo recomendações do autor, o sistema só funciona bem com um mapa de até cerca de 6.000 pontos e cerca de 150 quadros chave. Com esses valores a otimização já leva cerca de 7 segundos para ser executada. O valor de 6.000 pontos pode parecer muito comparado aos 400 do algoritmo de Davison, mas não é. Enquanto em uma volta completa em torno do usuário, usando o algoritmo de Davison, o mapa tem cerca de 100 pontos, no algoritmo de Klein ele precisa de 11.000 pontos, o que o próprio autor afirma estar fora do escopo de 'pequena área de trabalho'.

O processo de rastreamento é uma versão modificada do KLT [34]. No KLT o pareamento é feito por busca passiva, como explicado anteriormente. A modificação de Klein é que não são usados os pontos detectados no quadro anterior. Ao invés disso, ele usa a posição da câmera estimada no quadro anterior e projeta cerca de 50 pontos escolhidos aleatoriamente do mapa no plano da imagem. Depois de conseguir identificar no plano 2D qual foi o movimento de alguns desses 50 pontos, ele usa a posição tridimensional dos pontos para estimar a nova posição da câmera. Como Klein cria um mapa denso e utiliza, no rastreamento, cerca de 1.000 pontos reprojatados para realizar a otimização da posição da câmera, ele consegue um rastreamento muito estável. Esse algoritmo pode funcionar como localizador de pontos 3D, mas apenas em regiões muito pequenas, como, por exemplo, a área de uma mesa de trabalho, o que não atende os casos descritos na seção de motivação. Outro problema apresentado é a falta de invariância a mudanças de iluminação, da mesma forma que o algoritmo de Davison. Apesar de ter cerca de mil pontos sendo rastreados, uma mudança na iluminação do ambiente faz com que praticamente todos os pontos sejam perdidos.

Castle et al [10] fez uma modificação em cima do algoritmo PTAM que possibilitou criar múltiplos mapas e navegar entre eles. Com isso ele fez um algoritmo bastante preciso em pequenas ilhas de interesse. Além disso, ele é capaz de indicar onde estão as ilhas ao redor. Ele supõe que o usuário é suficientemente capaz de caminhar até a próxima ilha de interesse. De fato esse sistema atende ao nosso problema de localização de posições 3D, mas

apresenta o mesmo problema de iluminação do PTAM.

2.1

Posicionamento desse trabalho

Neste trabalho apresentamos uma versão modificada do algoritmo de Davison [14]. A primeira modificação é a substituição da inicialização que utiliza um retângulo preto por uma inicialização usando um marcador fiducial aos moldes do algoritmo de Wagner [38], ou seja, para cada marcador fiducial temos sua posição no espaço. Com isso é possível inicializar o sistema de posições diferentes do centro do sistema de coordenadas.

A segunda modificação é a introdução de um detector de texturas naturais em paralelo, análogo à Castle et al [12]. Mas, diferentemente de Castle, vamos mapear a posição exata da textura na fase de preparação do ambiente e, com isso eliminar o erro inerente ao mapa do SLAM de Davison sem otimização. Além disso, como as posições das texturas naturais são conhecidas, podemos prever a posição delas no ambiente e, com isso podemos eliminar da pesquisa de reconhecimento de padrões as texturas que estão fora da visibilidade da câmera.

A terceira modificação é a constante reinicialização. Sabendo que o erro de um algoritmo de SLAM é diretamente proporcional à distância do marcador de inicialização, então vamos reinicializar o sistema toda vez que conseguirmos encontrar um marcador natural ou fiducial com sucesso. Com isso, reduzimos o número de pontos característicos no filtro de Kalman, o que garante bastante amplitude de operação.

Como os marcadores fiduciais são invariantes à mudança de iluminação e existem detectores de textura natural invariantes à mudança de iluminação global, podemos fazer um mapa que só terá que ser alterado caso os marcadores sejam movidos de lugar. Outra qualidade dessa abordagem é que como inicializamos com marcadores fiduciais e, além deles marcadores naturais que estão no campo visível, então temos uma quantidade reduzida de texturas para procurar. Essa ideia é análoga à de Föckler et al[18] no uso do RFID para reduzir o campo de busca e consequentemente o tempo de processamento e as ambiguidades. Além disso, como temos, além dos marcadores fiduciais, os marcadores naturais, podemos usar menos marcadores fiduciais, o que é importante, pois esses muitas vezes provocam a oclusão de partes importantes do ambiente, além de interferir no seu aspecto.