



**Data Science
Academy**

www.datascienceacademy.com.br

Introdução à Inteligência Artificial

Reconstrução a Partir de Múltiplas Visões



A estereopsia binocular funciona porque, para cada ponto, temos quatro medidas restringindo três graus de liberdade desconhecidos. As quatro medidas são as posições (x, y) do ponto em cada visão, e os graus desconhecidos de liberdade são os valores das coordenadas (x, y, z) do ponto na cena.

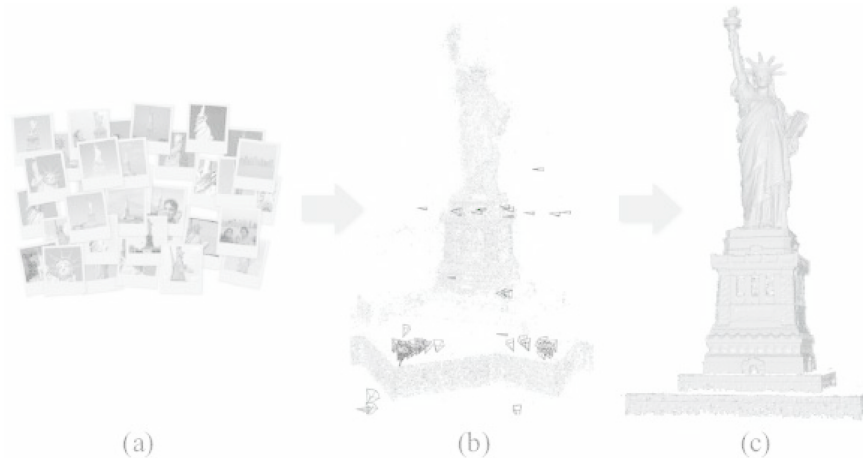
Esse argumento geral sugere, corretamente, que existem restrições geométricas que impedem a maioria dos pares de pontos de serem correspondências aceitáveis. Muitas imagens de um conjunto de pontos deveriam revelar as suas posições de forma inequívoca.

Nem sempre é preciso uma segunda imagem para obter uma segunda visão de um conjunto de pontos. Se acreditarmos que o conjunto de pontos original vem de um objeto familiar 3-D rígido, poderemos ter um modelo de objeto disponível como fonte de informação. Se esse modelo de objeto consiste em um conjunto de pontos 3-D ou um conjunto de imagens do objeto, e se pudermos estabelecer as correspondências dos pontos, poderemos determinar os parâmetros da câmera que produziram os pontos na imagem original. Essa é uma informação muito poderosa.

Poderíamos usá-la para avaliar a nossa hipótese inicial de que os pontos vêm de um modelo de objeto. Fazemos isso usando alguns pontos para determinar os parâmetros da câmera, em seguida projetando pontos de modelo nessa câmera e verificando para constatar se há pontos de imagem nas proximidades.

Esboçamos aqui uma tecnologia que agora está muito bem desenvolvida. A tecnologia pode ser generalizada para lidar com visões que não são ortográficas; para lidar com pontos que são observados em apenas algumas visões; para lidar com propriedades da câmera desconhecidas, como o comprimento focal; para explorar várias buscas sofisticadas para correspondências apropriadas; e para fazer a reconstrução a partir de grande número de pontos e de visões. Se a localização dos pontos nas imagens for conhecida com alguma precisão, e as direções de visão forem razoáveis, pode-se obter uma câmera de precisão muito alta e informações do ponto. Algumas aplicações são:

- Construção de modelo: por exemplo, pode-se construir um sistema de modelagem que toma uma sequência de vídeo e produz uma estrutura entrelaçada de polígonos texturizados tridimensional, muito detalhada, para utilização em aplicações de computação gráfica e de realidade virtual. Modelos como esse podem agora ser construídos a partir de conjuntos de imagens aparentemente nada promissores. Por exemplo, a figura abaixo mostra um modelo da Estátua da Liberdade construída a partir de imagens encontradas na Internet.



Essa figura descreve um sistema construído por Michael Goesele e colegas da Universidade de Washington, TU Darmstadt, e Microsoft Research. De uma coleção de imagens de um monumento tomada por uma grande comunidade de usuários e postadas na Internet (a), seu sistema pode determinar as direções de visualização dessas imagens, mostradas por pirâmides pretas pequenas em (b) e por uma reconstrução abrangente 3-D em (c).

- Casamento de movimentos: para colocar personagens de computação gráfica em vídeo real, precisamos saber como a câmera se moveu para o vídeo real, para que possamos representar o personagem corretamente.
- Reconstrução de caminho: robôs móveis precisam saber onde estavam. Se estiverem se movendo em um mundo de objetos rígidos, executar a reconstrução e manter a informação da câmera é uma maneira de obter o caminho.

Uma das principais aplicações da visão é de fornecer informações tanto para a manipulação de objetos — pegá-los, segurá-los, girá-los, e assim por diante — como para a navegação enquanto se evita obstáculos. A capacidade de usar a visão para esses fins está presente nos sistemas visuais mais primitivos do animal. Em muitos casos, o sistema visual é mínimo, no sentido de que extrai do campo de luz disponível apenas a informação que o animal precisa para informar o seu comportamento. Muito provavelmente, os sistemas de visão moderna evoluíram desde o início, organismos primitivos que usavam um ponto fotossensível em uma das extremidades para se orientar na direção da luz (ou para longe dela). As moscas utilizam um sistema de detecção de fluxo ótimo muito simples para aterrissar em paredes. Um estudo clássico, *What the Frog's Eye Tells the Frog's Brain* (Lettvin et al., 1959), observa sobre um sapo que “Ele vai morrer de fome cercado por alimentos se eles não estiverem se movendo. Sua escolha de alimento é determinada apenas pelo tamanho e pelo movimento”.



Vamos considerar um sistema de visão para um veículo automatizado dirigindo em uma autoestrada. As tarefas enfrentadas pelo motorista incluem o seguinte:

1. Controle lateral — assegura que o veículo permaneça seguro dentro de sua pista ou troca de pista suavemente quando necessário.
2. Controle longitudinal — garante que haja uma distância segura do veículo da frente.
3. Esquiva de obstáculos — monitora veículos nas pistas vizinhas e está preparado para manobras evasivas, se um deles decidir mudar de faixa.

O problema para o motorista é gerar direção adequada, aceleração e ações de frenagem para melhor realizar essas tarefas.

Para o controle lateral, é preciso manter uma representação da posição e orientação do carro em relação à pista. Podemos usar algoritmos de detecção de arestas para encontrar as arestas correspondentes aos segmentos marcadores de pista. Podemos, então, ajustar as curvas suaves a esses elementos de aresta. Os parâmetros dessas curvas levam informações sobre a posição lateral do carro, a direção que ele aponta em relação à pista e a curvatura da pista. Essa informação, juntamente com informações sobre a dinâmica do carro, é tudo o que é necessário para o sistema de controle de direção. Se tivermos bons mapas detalhados da estrada, o sistema de visão servirá para confirmar nossa posição (e observar os obstáculos que não estão no mapa).

Para controle longitudinal, é preciso conhecer as distâncias dos veículos da frente. Isso pode ser realizado com estereopsia binocular ou fluxo óptico. Usando essas técnicas, carros controlados por visão podem agora dirigir de forma confiável em estrada de velocidades.

O caso mais geral de robôs móveis navegando em vários ambientes interiores e exteriores também tem sido estudado. Um problema particular, a localização do robô em seu ambiente, agora tem soluções muito boas. Um grupo de Sarnoff desenvolveu um sistema baseado em duas câmeras procurando pontos com características de rastro em 3-D e usa isso para reconstruir a posição do robô em relação ao meio ambiente. Na verdade, existem dois sistemas de câmeras estereoscópicas, uma que olha para a frente e outra que olha para trás — isso dá maior robustez, caso o robô tenha de passar por um trecho sem características devido às sombras escuras, paredes brancas, e assim por diante. É improvável que não haja características na frente ou atrás. Agora, é claro que isso pode acontecer, então é providenciada uma cópia de segurança usando uma unidade de movimento inercial (UMI) um pouco parecida com os mecanismos para a detecção de aceleração que nós humanos temos em nosso ouvido interno. Através da integração da aceleração sentida duas vezes, pode-se acompanhar a mudança de posição. Combinar os dados da visão com os da UMI é um problema de fusão de evidência probabilística e pode ser atacado utilizando técnicas como filtragem de Kalman.



No uso de odometria visual (estimativa da mudança de posição), como em outros problemas de odometria, há o problema de “integração de informação incremental do movimento”, que envolve erros posicionais acumulando-se ao longo do tempo. A solução para isso é usar pontos de referência para fornecer correções da posição absoluta: assim que o robô passar por um local em seu mapa interno, pode ajustar a estimativa de sua posição adequadamente. Com essa técnica foram demonstradas precisões da ordem de centímetros. O exemplo da condução torna um ponto muito claro: para uma tarefa específica, não é preciso recuperar toda a informação que, em princípio, pode ser recuperada de uma imagem. Não é necessário recuperar a forma exata de cada veículo, resolver por forma da textura da superfície da grama adjacente à autoestrada, e assim por diante. Em vez disso, um sistema de visão deve calcular exatamente o que é necessário para realizar a tarefa.

Referências:

Livro: Inteligência Artificial

Autor: Peter Norvig