Localização baseada em odometria visual

André Toshio Nogueira Nishitani

Localização baseada em odometria visual

André Toshio Nogueira Nishitani

Orientador: **Prof. Dr. Denis Fernando Wolf**

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional. *EXEMPLAR DE DEFESA*

USP – São Carlos
Maio de 2015

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

N722l

Nishitani, André Toshio
Localização baseada em odometria visual / André Toshio Nishitani; orientador Denis Wolf. -- São Carlos, 2015.
79 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2015.

1. Odometria visual. 2. Localização. 3. Visão computacional. 4. Alinhamento de imagem. I. Wolf, Denis, orient. II. Título.

Resumo

O problema da localização cosiste em estimar a posição de um robô com relação a algum referencial externo e é parte essencial de sistemas de navegação de robôs e veículos autônomos. A localização baseada em odometria visual destaca-se em relação a odometria de *encoders* na obtenção da rotação e direção do movimento do robô. Esse tipo de abordagem é também uma escolha atrativa para sistemas de controle de veículos autônomos em ambientes urbanos, onde a informação visual é necessária para a extração de informações semânticas de placas, semáforos e outras sinalizações. Neste contexto estre trabalho propõe o desenvolvimento de um sistema de odometria visual utilizando informação visual de uma câmera monocular baseado em reconstrução 3D para estimar o posicionamento do veículo. O problema da escala absoluta, inerente ao uso de câmeras monoculares, é resolvido utilizando um conhecimento prévio da relação métrica entre os pontos da imagem e pontos do mundo em um mesmo plano.

Abstract

The localization problem consists of estimating the position of the robot with regards to some external reference and it is an essential part of robots and autonomous vehicles navigation systems. Localization based on visual odometry, compared to encoder based odometry, stands out at the estimation of rotation and direction of the movement. This kind of approach is an interesting choice for vehicle control systems in urban environment, where the visual information is mandatory for the extraction of semantic information contained in the street signs and marks. In this context this project propose the development of a visual odometry system based on structure from motion using visual information acquired from a monocular camera to estimate the vehicle pose. The absolute scale problem, inherent with the use of monocular cameras, is achieved using som previous known information regarding the metric relation between image points and points lying on a same world plane.

Sumário

Lista de Figuras

Capítulo

1

Introdução

1.1 Contextualização

Segundo a World Health Organization (2013), todos os anos cerca 1,3 milhões de pessoas morrem devido a acidentes de trânsito, e entre 20 a 50 milhões de pessoas sofrem acidentes não fatais, que muitas vezes incorrem em debilidade ou invalidez do acidentado. No Brasil segundo dados do Departamento Nacional de Infraestrutura de Transporte - DNIT (2013) de 2011, cerca de 200 mil acidentes ocorreram nas estradas federais, sendo que mais de 7 mil desses acidentes foram fatais. No relatório técnico "Dados de boletins de ocorrência" (Tani et al., 2008) desenvolvido em parceria pelo Departamento Nacional de Infraestrutura de Transporte - DNIT (2013) e pelo Laboratório de Transporte e Logística - LabTrans (2013) da Universidade Federal de Santa Catarina, foram coletadas informações de boletins de ocorrências do período entre 2005 a 2007 coletados pela Polícia Rodoviária Federal de Santa Catarina. As informações coletadas foram classificadas e relacionadas de maneira a facilitar visualização dos dados e permitir a identificação de maneira simples quais rodovias são críticas em relação aos acidentes de trânsito.

A Tabela 1.1 foi montada com valores retirados do relatório técnico (Tani et al., 2008). É relacionado o número de acidentes cujo principal fator contribuinte foi a falta de atenção do condutor com o número total de acidentes na rodovia nos anos de 2005, 2006 e 2007. Os dados apontam que nas rodovias onde foram levantados os dados, a falta de atenção do condutor foi o principal fator contribuinte para acidente.

Rodovia	2005			2006			2007		
	Total	**Nº Acid.**	**%**	**Total**	**Nº Acid.**	**%**	**Total**	**Nº Acid.**	**%**
BR 101	5 928	3 043	51,30	5 927	3 260	55,00	6 346	3 391	53,40
BR 116	601	309	51,41	649	297	45,76	704	326	46,73
BR 153	214	95	44,39	239	99	41,42	225	66	29,33
BR 158	60	30	50,00	55	27	49,09	55	21	38,18
BR 163	92	43	46,74	105	49	46,67	116	36	31,03
BR 280	1 155	695	60,17	1 313	898	68,39	1 289	791	61,39
BR 282	2 220	1 075	48,42	2 286	942	41,21	2 667	1 027	38,51
BR 470	2 457	1 205	49,04	2 398	1 319	55,00	2 990	1 536	51,37

Tabela 1.1: A tabela é separada entre os anos em que se coletou os dados. Para cada ano as colunas representam respectivamente: quantidade total de acidentes na rodovia, quantidade de acidentes causados pela falta de atenção e o respectivo porcentual.

Os dados da tabela expressam um grande problema dos veículos que é a falta de atenção por parte do condutor. Outros fatores, como o sono e o desrespeito às regras da rodovia, também aparecem como grandes contribuintes para acidentes na rodovia, correspondendo entre 10% e 25% dos acidentes nas rodovias observadas. Assim como a falta de atenção esses fatores são atribuídos ao condutor. O uso sistemas de auxílio a direção ou sistemas de direção autônomos é uma forma de reduzir fatalidades no trânsito. Veículos autônomos também podem gerar uma séria de outros benefícios como o aumento da eficiência do trânsito em grandes cidades e a facilidade de transporte para pessoas com limitações ou que não se sintam aptas a dirigir.

As primeiras referências a pesquisas relacionadas veículos autônomos datam dos anos 1980 (Thorpe et al., 1988; Pomerleau, 1989), mas a área obteve um grande salto em relação ao estado da arte através das competições DARPA *Grand Challenge* (Buehler et al., 2007) e DARPA *Urban Challenge* (Buehler et al., 2009). As competições são, provavelmente, os eventos que melhor expressam o quanto é investido em pesquisa para desenvolvimento de veículos autônomos. Criadas pela agência DARPA (*Defense Advanced Research Projects Agency*), essas competições visam o desenvolvimento de novas tecnologias para uso em veículos militares. Apesar do desenvolvimento voltado para fins militares, muitas das tecnologias desenvolvidas nos desafios têm sido utilizadas para fins não militares.

O desafio proposto aos participantes do DARPA *Grand Challenge* foi atravessar um trajeto de aproximadamente 240 km, realizando a tarefas com um sistema de navegação autônoma. O trajeto passava por tipos variados de terrenos e os sistemas deveriam ser capazes de navegar por eles evitando obstáculos que pudessem obstruir seu caminho. Na primeira edição do evento em 2004 nenhum veículo foi capaz de terminar a prova, sendo que o melhor colocado conseguiu completar somente 11.78 km do percurso. No ano

seguinte dos 24 finalistas (dentre quase 200 competidores), somente um não conseguiu ultrapassar a distância de 11.78 km, enquanto cinco veículos conseguiram completar o percurso.

Na Europa, o The European Robot Trial - ELROB (2013) é um evento para demonstração e comparação do estado da arte da robótica européia. Alguns temas relacionados ao eventos são "navegação autônomoa" e "transporte em comboio". O The European Robot Trial - ELROB (2013) ao contrário do DARPA *Grand Challenge* e DARPA *Urban Challenge* não é uma competição, portanto não há um vencedor para as provas propostas. O evento é realizado anualmente, sendo que a cada ano a temática das competições alterna entre civil e militar. Também na Europa, em maio de 2011, ocorreu o primeiro *Grand Cooperative Driving Challenge* (Ploeg et al., 2012). O *Grand Cooperative Driving Challenge* é uma competição que visa impulsionar a pesquisa de sistemas de direção cooperativas, onde o veículo se comunica com outros veículos e com o ambiente com o fim de evitar acidentes e melhorar o fluxo de trânsito. Na primeira edição do evento o desafio proposto foi manter um grupo de carros em comboio, com uma distâncias curtas entre os carros evitando colisões. O desafio sugerido visa solucionar o problema de como aumentar o número de veículos suportados por uma via e baixar o consumo de combustível, sem ter um aumento no número de acidentes.

1.2 Motivação

Para solucionar os desafios propostos pelas competições e eventos mencionados e mesmo para outras tarefas, um veículo autônomo depende de um sistema de navegação precisos e robusto, capaz de tratar com eficiência a natureza imprevisível do trânsito em ruas e rodovias. O sistema precisa conhecer sua localização em relação à rua, aos outros veículos e aos demais elementos (*e.g.* pedestres, sinalizações) que compõe a rua ou qualquer outro ambiente em que o veículo esteja atuando.

Leonard e Durrant-Whyte (1991) definem a navegação com três perguntas: "onde estou?", "onde vou?" e "como chego lá?". As perguntas traduzem os seguintes elementos da navegação respectivamente: a localização do agente, a definição da meta e o planejamento do trajeto à meta.

A localização é parte essencial da navegação autônoma. O problema de localização consiste em estimar a posição do robô com relação a algum referencial externo. Pode-se classificar o problema de localização em dois tipos: global e local. O problema de localização global consiste em estimar a posição do robô independente de qualquer informação sobre sua posição inicial. Já no problema de localização local, dada uma aproximação

da posição inicial do robô uma nova localização para o robô é estimada de maneira incremental utilizando-se o deslocamento obtido com auxílio de sensores e atualizando a localização anterior.

Diferentes sensores podem auxiliar na tarefa de estimar o deslocamento de um robô. Alguns exemplos são IMUs (unidade de medida inercial, ou do inglês *Inertial Measurement Unit*), *encoders* para contagem de rotações da roda, GPS (sistema de posicionamento global, ou do inglês *Global Positioning System*), dentre outras possibilidades. É possível implementar sistemas de odometria para um robô com uma variedade de sensores, inclusive com a combinação de de dois ou mais sensores. Porém cada um desses sensores apresenta vantagens e desvantagens próprias. Sistemas de odometria baseados em IMUs e *encoders*, por exemplo, apresentam acumulo de erros à medida que o robô se move por conta de derrapagens ou calibração imprecisa dos sensores, tornando-os sistemas extremamente imprecisos a longo prazo. Outro exemplo é o GPS que permite a obtenção de uma localização global, mas que em geral apresenta imprecisão local, podendo comprometer a navegação do veículo, além de sofrer com a perda de sinal, período em que o sistema ficaria "às cegas", sem nenhuma informação de localização.

Em geral, sistemas de odometria apresentam erros em suas estimativas para a localização. Sistemas baseados em sensores visuais não estão isentos deste problema. Porém o uso diferentes sensores pode melhorar a qualidade da localização, uma vez que as vantagens de um sensor pode suprir as limitações de outro. Assim um sistema baseado em odometria visual é uma opção importante para o apoio e complementação de outros sensores utilizados para obtenção da localização.

A odometria visual é o processo de estimar a posição da câmera baseado na informação recuperada de uma sequência de imagens. Apesar de ser um sensor com grande riqueza de informação, técnicas de odometria visual são pouco usadas pelo o fato de muitas delas necessitarem de um alto poder computacional ou não atingirem requisitos de tempo real ou mesmo pela alta complexidade dos algoritmos envolvidos. Com o aumento da capacidade computacional e surgimento de novas tecnologias, além do desenvolvimento de novas técnicas de processamento de imagem mais eficientes, métodos de localização baseados em odometria visual tem ganhado espaço dentro da comunidade científica.

Os sistemas baseados em odometria visual destacam-se na obtenção de rotação e direção do movimento. Esse tipo de sistema é uma escolha óbvia para sistemas de controle veicular em ambiente urbano uma vez que a aquisição de imagens do ambiente é uma necessidade por conta da informação semântica contida em placas e outras sinalizações.

1.3 Objetivo

O objetivo principal dessa dissertação é o desenvolvimento de um sistema de localização baseado em visão que possa ser utilizado em um veículo autônomo. O sistema recupera o trajeto do veículo a partir de uma sequência de imagens utilizando métodos baseados em reconstrução 3D (*structure from motion*). A odometria é composta da posição e da orientação do veículo totalizando seis graus de liberdade, três ângulos para a orientação e três coordenadas para a posição.

A informação de deslocamento obtida através de uma câmera monocular é relativa. A menos que haja alguma informação métrica a respeito dos dados extraídos pela câmera, não é possível se obter a escala real do deslocamento da câmera e consequentemente do veículo. Então além da estimação dos ângulos de rotação e da direção do movimento do veículo, objetiva-se determinar a escala real do movimento realizado pelo veículo.

A premissa é que o sistema aqui apresentado trabalhará em conjunto com outros sistemas de odometria, complementando e proporcionando maior precisão ao sistema de localização do veículo. Apesar de se buscar o funcionamento conjunto deste sistema com outros sistemas de odometria espera-se que ele seja capaz de fornecer, sem auxílio dos outros sistemas, informação de qualidade à respeito da odometria do veículo.

1.4 Organização da Dissertação

Esta dissertação é divida em seis Capítulos. O Capítulo 2 traz uma revisão geral dos sistemas de odometria baseados em visão. Nos Capítulos 3, 4, 5 e 6 são apresentadas as teoria e técnicas utilizadas no desenvolvimento do trabalho. O Capítulo 3 descreve a calibração de uma câmera perspectiva, além da descrição e estimação de estruturas que relacionam pontos do mundo com a imagem. O Capítulo 4 descreve métodos para estimação do alinhamento de imagem e como esses métodos podem ser utilizados para a estimação da odometria de um veículo. Os Capítulos 5 e 6 apresentam uma forma alternativa para estimar a odometria de um veículo utilizando elementos distinguíveis em imagens diferentes e a relação geométrica desses elementos na imagem para encontrar o deslocamento real do veículo. Por fim, o Capítulo 7 compara o uso dos métodos independentes e o uso combinado apresentando também os resultados obtidos com o método desenvolvido. No Capítulo 8 são discutidos os pontos forte e fracos do método desenvolvido com uma breve avaliação dos mesmos.

Capítulo

2

Trabalhos Relacionados

Por muito tempo, técnicas para odometria visual foram pouco utilizadas devido à falta de capacidade computacional ou métodos eficientes para lidar com o processamento e extração de informação da imagem. O surgimento de novas técnicas para o processamento de imagem e novas tecnologias, mais eficientes e de menor custo, vem mudando este cenário.

A maior parte dos trabalhos relacionados a estimação do deslocamento de um robô por meio da visão utiliza câmeras estéreo por permitir, de maneira relativamente simples, a obtenção de informações sobre a profundidade de pontos na imagem. O trabalho de Moravec (1980) é tido como um dos primeiros trabalhos com câmera estéreo sendo utilizadas para estimação da odometria de um agente móvel e é referência para muitos outros. Neste trabalho é apresentado um sistema de odometria para um *rover*, um veículo desenvolvido para exploração espacial. Utilizando uma câmera em uma estrutura deslizante que lhe permitia a obtenção de imagens com uma distância entre as imagens conhecida era possível estimar a estrutura 3D da cena e comparar a estrutura entre duas cenas para obter o movimento do *rover*. Este trabalho também apresentou um dos primeiros detectores de cantos conhecidos, e que será descrito no Capítulo 5.

Em trabalho mais recente, Nistér et al. (2006) apresenta um algoritmo baseado em reconstrução 3D para estimação da odometria visual de um veículo. Nesse trabalho são apresentadas versões do algoritmo tanto para câmeras estéreo quanto para câmeras monoculares. A ideia básica para ambos os algoritmos consiste em triangular as características

encontradas em um par de imagens, rastrear essas características nas imagens seguintes e estimar a posição da câmera em relação aos pontos triangulados utilizando o "Algoritmo dos Três Pontos" (Haralick et al., 1994). Porém o algoritmo monocular precisa de imagens de dois instantes diferentes para a triangulação dos pontos, enquanto a câmera estéreo utiliza o par de imagens de um mesmo instante, além de obter a escala absoluta dos pontos. Um ponto a se destacar no trabalho apresentado por Nistér et al. (2006) é o aprimoramento da técnica RANSAC, propondo o *preemptive* RANSAC (Nistér, 2003) para remover valores destoantes ao movimento do veículo e uma implementação eficiente para o "Algoritmo dos Cinco Pontos" (Nistér, 2004) que codifica a geometria entre duas câmeras. Trabalhos baseados em (Nistér et al., 2006) são apresentados por Agrawal e Konolige (2006, 2007), o último difere do primeiro pelo uso da técnica *bundle adjustment* para minimizar o erro em uma janela de frames.

O RANSAC proposto por Fischler e Bolles (1981) é um dos métodos mais utilizados para a remoção de características que atrapalham a estimação da odometria. A quantidade mínima de parâmetros para se obter uma solução possível influencia o desempenho do RANSAC. Considerando o caso de movimento irrestrito, ou seja, movimento com 6 graus de liberdade, os solucionadores mínimos são o "Algoritmos dos Cinco Pontos" (Nistér, 2003) para câmeras calibradas e o "Algoritmo dos Seis Pontos" (Stewenius et al., 2005) para câmeras genéricas. Uma abordagem comum para aumentar a eficiência do algoritmo visa a redução do grau de liberdade do problem de estimação do movimento.

Fraundorfer et al. (2010) propõe a redução do problema à três graus de liberdade, impondo a condição de se conhecer dois ângulos da câmera. Essa condição pode ser alcançada utilizando uma unidade inercial, capaz de fornecer essa informação. A redução a três graus de liberdades permite que o problema seja solucionado com apenas três pontos correspondentes entre as imagens. Zhu et al. (2012) utiliza um método híbrido, onde impõe-se a restrição de movimento planar e reduz a dois o número necessário de correpondências para solucionar o problema de estimar o movimento da câmera, mas utiliza o algoritmo de Haralick et al. (1994) para obter uma reconstrução mais precisa. Scaramuzza et al. (2009b) utiliza a ideia de movimento circular e reduz o problema a um grau de liberdade correspondendo ao ângulo da rotação do veículo. Neste trabalho Scaramuzza et al. (2009b) propõe o uso do método RANSAC, mas também propõe o uso de um método de votação de histograma, que necessitaria menos interações, para solucionar a solução que melhor comporta o conjunto de dados. Civera et al. (2010) utilizam a predição de estado do filtro de kalman extendido permitindo reduzir o problema a um grau de liberdade. Em trabalho mais recente Jiang et al. (2012) propõe uma solução para o problema da odometria utilizando três pontos, destacando que o modelo proposto por

Scaramuzza et al. (2009b) falha na presença de muita variação na variação dos ãngulos de *pitch* e *roll*. Howard (2008) realiza a estimativa da odometria utilizando uma alternativa para métodos baseados em RANSAC para remoção de *outliers*, a técnica é baseada em "clique"e é apresentada por Hirschmuller et al. (2002).

Métodos de odometria visual envolvendo câmeras monoculares frequentemente relevam o problema da ambiguidade da escala do movimento e encontram soluções em função da escala. Kitt et al. (2011) apresentam um método de odometria visual baseado em (Nistér et al., 2006), onde estabelem algumas suposições em relação ao movimento do veículo e à posição de montagem da câmera para que pontos da rua são utilizados para atualizar a estimativa da escala do movimento. Em (Scaramuzza et al., 2009a) o trabalho (Scaramuzza et al., 2009b) é extendido para estimar a escala absoluta do movimento a partir da distância horizontal entre a câmera e o eixo traseiro do veículo.

Além de técnicas baseadas em reconstrução 3D a odometria pode ser obtida com técnicas como o fluxo óptico Barron e Thacker (2005). Campbell e Sukthankar (2004) descrevem um exemplo de algoritmo que usa uma técnica baseada em fluxo óptico para determinar a odometria da câmera. Mais recentemente Lategahn et al. (2012) propuseram um método baseado em câmera omnidirecional onde a estimação de movimento é feita baseada nas relações dos raios de projeção de um ponto. Também trabalhando com uma câmera omnidirecional Lim et al. (2010) utilizam os chamados pontos antipodais para estimar o deslocamento. Com o uso dos pontos antipodais os autores reduzem o número mínimo de pontos para solucionar o problema para quatro e também alcançam o desacoplamento de rotação e translação na solução do problema.

A câmera omnidirecional é um tipo especial de câmera monocular que possui um ângulo de visão de 360°. Scaramuzza et al. (2009b) fazem uso de uma câmera omnidirecional, destacando que métodos utilizado em câmeras omnidirecionais podem, em geral, ser adaptados de câmeras monoculares simples, além de fazer a comparação entre diferentes métodos de extração e pareamento de características da imagem. Scaramuzza e Siegwart (2008) apresentam métodos de cálculo de odometria visual com uma câmera omnidirecional e utilizando o método SIFT (Scale Invariant Feature Transformation) (Lowe, 1999) para extração de características da imagem. Os trabalhos (Tardif et al., 2008) e (Corke et al., 2004) são outros exemplos de odometria visual utilizando câmeras omnidirecionais. O primeiro apresenta uma metodologia semelhante à apresentada em (Nistér et al., 2006) com algumas otimizações, enquanto o segundo apresenta método de estimação do movimento da câmera baseados em fluxo óptico (Barron e Thacker, 2005) e nas restrições epipolares (Ma et al., 2003).

Capítulo

3

Geometria e Calibração da Câmera Monocular

Este Capítulo trata da formação da imagem em uma câmera monocular e a relação da imagem com o mundo. A Seção 3.1 descreve a geometria da projeção de uma cena em uma imagem. As Seções 3.2 e 3.3 descrevem os parâmetros internos e externos da câmera, respectivamente. A Seção 3.4 descreve brevemente a matriz de projeção e aborda o caso em que os pontos no mundo estão todo em um mesmo plano. Por fim a Seção 3.5 apresenta um metodo para estimar a matriz de homografia entre um plano no mundo e a imagem. A teoria apresentada nesse Capítulo é baseada em (Hartley e Zisserman, 2004; Trucco e Verri, 1998)

3.1 Formação da Imagem

A formação de uma imagem em uma câmera ocorre com a entrada de feixes de luz através de uma abertura na câmera e a projeção desses feixes em uma tela, também chamada de **plano de imagem**.

Em uma câmera real, um ponto no mundo reflete diversos feixes de luz. Se todos os feixes refletidos por esse ponto convergirem para um mesmo ponto no plano de imagem, então é dito que a imagem está focada. O modelo de projeção perspectiva é uma simplificação da câmera real.

O modelo de projeção perspectiva é apresentado na Figura 3.1. O centro de projeção $\mathbf{C}$ é a origem do sistema de coordenadas da câmera e também o centro da câmera. O eixo-z

do sistema de coordenadas da câmera é chamado eixo-principal. O plano $z = f$ é o plano de imagem e a intersecção do plano de imagem com o eixo-principal é chamado ponto principal. Considere $\mathbf{X} = [X, Y, Z]^T$ as coordenadas de um ponto no mundo referentes ao sistema de coordenadas da câmera. A intersecção do plano de imagem com o segmento de reta ligando $\mathbf{X}$ e $\mathbf{C}$ é a projeção de $\mathbf{X}$ e é referênciada como $\mathbf{x}$. Por semelhança de triângulos observa-se que $\mathbf{x} = [f\frac{X}{Z}, f\frac{Y}{Z}, f]$ em relação à câmera. Como a última coordenada de $\mathbf{x}$ será sempre f, ela será desconsiderada nas equações daqui em diante.

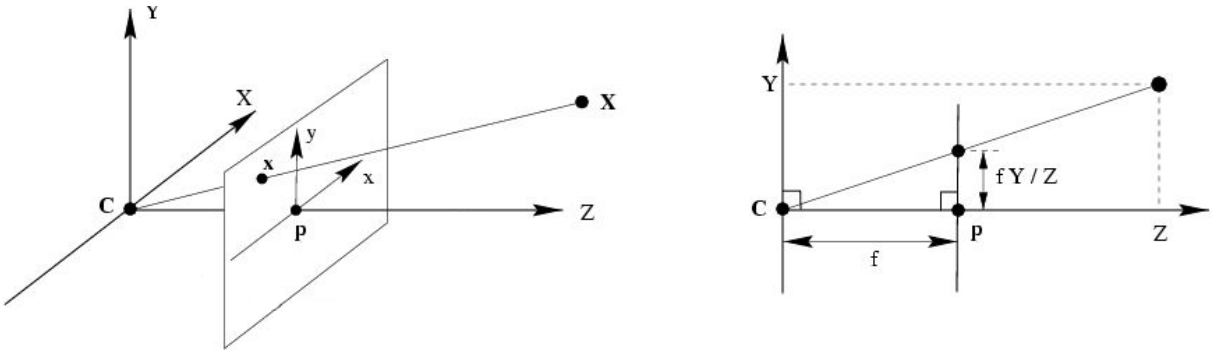


Figura 3.1: Esquema do modelo de projeção perspectiva. Fonte:Hartley e Zisserman (2004)

Reescrevendo a relação entre $\mathbf{X}$ e $\mathbf{x}$ utilizando coordenadas homogêneas na forma matricial

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{3.1}$$

Note que a última coordenada w é a escala da coordenada homogênea e não a distância focal f, que como já foi dito, é desconsiderada daqui em diante. Daqui em diante as coordenadas homogêneas serão representadas por $\mathbf{x}$ e $\mathbf{X}$.

3.2 Matriz de Calibração e Parâmetros Intrínsecos

A Equação 3.1 descreve o mapeamenteo de um ponto no mundo $\mathbf{X}$ no plano de imagem $\mathbf{x}$. Note que essa relação é dada no sistema de coordenadas da câmera, que utiliza alguma unidade de distância própria (*e.g.* metro). Porém, para aplicações de visão computacional é interessante conhecer a representação da projeção no sistema de coordenadas da imagem, ou seja, conhecer a Equação 3.1 em *pixels*.

A Equação 3.1 considera que o ponto principal p é a origem do sistema de coordenadas no plano da imagem, mas para aplicações reais essa suposição nem sempre é verdadeira.

O deslocamento pelas coordenadas $p = [p_x, p_y]$ corrige esse problema. Outra diferença nos sistemas de coordenadas é a unidade de medida de distâcia. A unidade da câmera pode estar em metros, centímetros ou alguma outra, enquanto a da imagem, em geral, é dada em *pixels*. Os fatores s_x e s_y dão o tamanho do pixel na unidade de medida do sistema de coordenadas da câmera.

A relação entre um ponto no sistema de coordenadas da câmera com um ponto no sistema de coordenadas da imagem é dado então por

$$\begin{aligned} x &= -(x_{im} - p_x)s_x \\ y &= -(y_{im} - p_y)s_y \end{aligned}. \tag{3.2}$$

Substituindo a Equação 3.2 na Equação 3.1

$$\begin{bmatrix} x_{im} \\ y_{im} \\ w \end{bmatrix} = \begin{bmatrix} -\frac{f}{s_x} & 0 & p_x & 0 \\ 0 & -\frac{f}{s_y} & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3.3}$$

Reescrevendo a Equação 3.3 como

$$\begin{bmatrix} x_{im} \\ y_{im} \\ w \end{bmatrix} = \begin{bmatrix} -\frac{f}{s_x} & 0 & p_x \\ 0 & -\frac{f}{s_y} & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{3.4}$$

obtém-se a matriz

$$K = \begin{bmatrix} -\frac{f}{s_x} & 0 & p_x \\ 0 & -\frac{f}{s_y} & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

chamada de matriz de calibração ou matriz de parâmetros intrínsecos, sendo (f, p_x, p_y, s_x, s_y) são chamados parâmetros intrínsecos por serem as características internas da câmera. Também fazem parte dos parâmetros intrínsecos os coeficientes de distorção. Porém neste trabalho os coeficientes de distorção serão desconsiderados, significando que as imagens não apresentam nenhum tipo de distorção causada pela lente da câmera. Se a matriz de calibração da câmera é conhecida e aplicada a imagem, diz-se que a câmera está calibrada.

3.3 Parâmetros Extrínsecos

Como mencionado anteriormente a Equação 3.1 mapeia pontos do mundo no plano de imagem em relação ao sistema de coordenadas da câmera. Na Seção 3.2 foi introduzida a matriz de calibração que tranforma pontos no plano de imagem, na referência do sistema de coordenadas da câmera para o sistema de coordenadas da imagem. Nesta Seção a mesma ideia será aplicada para pontos do mundo.

Em geral os pontos do mundo são descritos em relação a um sistema de coordenadas global. A relação entre os dois sistemas é dada por uma tranformação de corpo rígido do tipo

$$\mathbf{X} = R\mathbf{X}_w + \mathbf{T}, \tag{3.5}$$

onde $\mathbf{X}_w$ são as coordenadas do ponto $\mathbf{X}$ em relação ao sistema de coordenadas global. A matriz $R \in SO(3)$ é a rotação que alinha o sistema de coordenadas global com o sistema de coordenadas câmera e $\mathbf{T} \in \mathbb{R}^3$ é o vetor de translação entre os dois sistemas de coordenadas. Os parâmetros de R e $\mathbf{T}$ são chamados de parâmetros extrínsecos e a matriz

$$\left[R|\mathbf{T}\right] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}$$

é a matriz de parâmetros extrínsecos. A Figura 3.2 mostra a relação entre os dois sistemas de coordenadas.

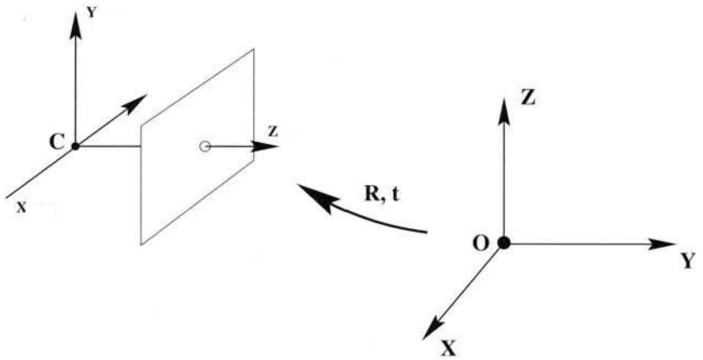


Figura 3.2: Parâmetros que definem a posição e orientação do sistema de coordenadas da câmera com um sistema de coordenadas global. Fonte:Hartley e Zisserman (2004)

Substituindo a Equação 3.5 na Equação 3.1

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = K[R|\mathbf{T}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \tag{3.6}$$

onde $[X_w, Y_w, Z_w, 1]$ são as coordenadas do ponto $\mathbf{X}$ no sistema de coordenadas global.

3.4 Matriz de Projeção e Homografia

Na Seção anterior foram introduzidas as matrizes K e $[R|\mathbf{T}]$ que descrevem respectivamente as características internas e externas da câmera. Tem-se então as informações necessárias para montar a seguinte relação

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3.7}$$

A matriz $P = K[R|\mathbf{T}]$ é chamada matriz de projeção e define a relação que mapeia um ponto $\mathbf{X}$ no mundo em um ponto $\mathbf{x}$ no plano de imagem.

Um caso especial dessa relação ocorre quando alguma informação sobre X é conhecida. Supondo que somente os pontos no chão serão mapeados no plano de imagem, então $\mathbf{X} = [X, Y, 0, 1]$, ou seja, todos se situam no plano $Z = 0$. A Figura 3.3 exemplifica essa situação. Considere a representação $P = \begin{bmatrix} \mathbf{p_1} & \mathbf{p_2} & \mathbf{p_3} & \mathbf{p_4} \end{bmatrix}$, onde $\mathbf{p_i}$, $i = 1, \ldots, 4$, são as

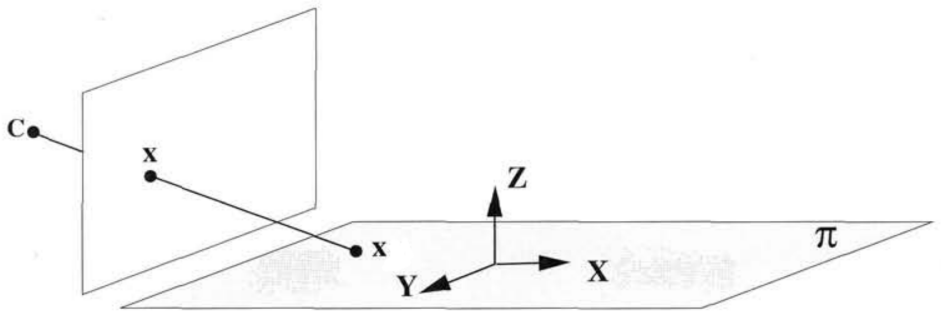


Figura 3.3: O ponto $\mathbf{X}$ se situa no plano $Z = 0$ é projetado no plano de imagem no ponto $\mathbf{x}$. Fonte:Hartley e Zisserman (2004)

colunas de P, então a Equação 3.7 pode ser reescrita como

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \mathbf{p_1} & \mathbf{p_2} & \mathbf{p_3} & \mathbf{p_4} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{p_1} & \mathbf{p_2} & \mathbf{p_4} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{3.8}$$

A matriz $H = \begin{bmatrix} \mathbf{p_1} & \mathbf{p_2} & \mathbf{p_3} \end{bmatrix}$ é a matriz de homografia que projeta os pontos do plano $Z = 0$ no plano de imagem. Note que é possível realizar a operação inversa e projetar pontos do plano de imagem no plano $Z = 0$. A matriz de homografia que projeta pontos do plano de imagem no plano $Z = 0$ é a matriz inversa $H' = H^{-1}$.

3.5 Estimação da Matriz de Homografia

Suponha $\mathbf{x}_i = [x_i, y_i, w_i]^T$ um conjunto de pontos no plano π_1 e $\mathbf{X}_i = [X_i, Y_i, W_i]^T$ um conjunto de pontos (sendo a coordenada $Z = 0$) no plano π_2. Suponha também uma matriz de homografia $H \in \mathbb{R}^{3\times3}$ que projeta o ponto do plano π_1 no plano π_2 através da relação

$$\begin{bmatrix} x_i \\ y_i \\ w_i \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ W_i \end{bmatrix}. \tag{3.9}$$

Note que ambos os pontos na Equação 3.9 são homogêneos e portanto os dois lados da Equação são iguais a menos de um fator escalar. Note também que a Equação 3.9 pode ser expressada como $\mathbf{x}_i \times H\mathbf{X}_i = \mathbf{0}$. Desenvolvendo essa Equação obtém a seguinte relação

$$\mathbf{x}_i \times H\mathbf{X}_i = \begin{bmatrix} \mathbf{0}^T & -w_i\mathbf{X}_i^T & y_i\mathbf{X}_i^T \\ w_i\mathbf{X}_i^T & \mathbf{0}^T & -x_i\mathbf{X}_i^T \\ -y_i\mathbf{X}_i^T & x_i\mathbf{X}_i^T & \mathbf{0}^T \end{bmatrix} \mathbf{h} = \mathbf{0}, \tag{3.10}$$

onde

$$\mathbf{h} = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 \end{bmatrix}^T.$$

Seja a matriz $A = [a_1, a_2, \ldots, a_n]^T$, onde

$$a_i = \begin{bmatrix} \mathbf{0}^T & -w_i\mathbf{X}_i^T & y_i\mathbf{X}_i^T \\ w_i\mathbf{X}_i^T & \mathbf{0}^T & -x_i\mathbf{X}_i^T \end{bmatrix},$$

então pode-se montar o sistema linear $A\mathbf{h} = \mathbf{0}$. A matriz de homografia $H \in \mathbb{R}^{3\times3}$ é deficiente em um grau de liberdade, portanto pode-se encontrar a solução para $\mathbf{h}$ com oito equações ou mais em A. Desse modo, é necessário um valor mínimo para n de quatro pontos, sendo que não podem haver três colineares, para determinar a matriz de homografia H. Note que somente as duas primeiras linhas da matriz encontrada na Equação 3.10 são utilizadas. Isso ocorre pois a terceira linha é uma combinação linear das duas anteriores e portanto não adiciona nenhuma informação nova ao sistema.

Antes do cálculo da homografia é sugerida a normalização dos pontos. Uma normalização é realizada para os pontos do plano de imagem e outra normalização é aplicada ao plano $Z = 0$. Essa normalização consiste em uma translação e uma escala aplicadas ao conjunto de pontos de maneira que o centróide do conjunto seja a origem do sistema de coordenadas e a média das distâncias dos pontos a origem seja $\sqrt{2}$. Essa normalização visa diminuir o efeito da diferença entre a magnitude da escala para as demais coordenadas no vetor homogêneo em casos de ruído na imagem.

Considere S e S' as matrizes que normalizam os conjuntos de pontos $\mathbf{x}_i$ e $\mathbf{X}_i$, respectivamente. Considere também $\widetilde{\mathbf{x}_i} = S\mathbf{x}_i$ e $\widetilde{\mathbf{X}}_i = S'\mathbf{X}_i$ os conjuntos de pontos normalizados. A matriz A é montada como descrito anteriormente, mas utilizando os pontos normalizados.

Para $n = 4$, sendo que não há três pontos colineares, $A\mathbf{h} = \mathbf{0}$ tem solução exata. Porém se $n > 4$, então a solução pode ser achada para minimizar uma função de erro. Decompondo a matriz A em valores singulares, o vetor singular unitário correspondente ao menor valor singular é a solução para $\mathbf{h}$. Se for utilizado o método SVD (Singular Value Decomposition) (Golub e Kahan, 1965), onde $A = UDV^T$ e D é a matriz diagonal com os valores singulares de A em ordem decrescente, então a solução $\|\mathbf{h}\| = 1$ corresponde à última coluna da matriz V.

Encontrada a solução $\|\mathbf{h}\| = 1$, então $\mathbf{h}$ pode ser reescrita na forma matricial. Note que a forma matricial será referente às coordenadas normalizadas então será chamada $\widetilde{H}$. Como

$$S\widetilde{\mathbf{x}_i} = HS'\widetilde{\mathbf{X}}_i,$$

então

$$H = S'^{-1}\widetilde{H}S.$$

A matriz H é então a homografia que relaciona dois planos. Com essa matriz é possível mapear pontos da imagem em um plano do mundo e vice-versa.

Capítulo

4

Odometria Visual Direta

Um par de imagens de uma sequência ou de uma mesma cena estão relacionadas através de uma transformação. O processo de identificar a transformação é chamado de alinhamento de imagem ou registro de imagem. Os métodos de alinhamento de imagem podem ser separados em métodos diretos e métodos baseados em características (Irani e Anandan, 1999). Os métodos diretos utilizam informações diretamente disponíveis na imagem (e.g. intensidade) para estimar a transformação de alinhamento do par. Os métodos baseados em características utilizam informações de mais alto nível como cantos ou bordas para realizar o alinhamento.

Existem muitas discussões acerca das vantagens e desvantagens de ambos os métodos Szeliski (2006); Irani e Anandan (1999). De maneira geral os métodos diretos apresentam pior desempenho em situações onde o deslocamento entre as duas imagens é grande. O uso de estratégias hierárquicas de múltiplas resoluções da imagem melhora a performance dos métodos diretos, mas ainda assim métodos baseados em características são mais indicados para alinhamento de imagens com grande deslocamento. Uma das vantagens dos métodos diretos de alinhamento é a utilização completa da informação contida nas imagens, uma vez que o alinhamento é feito comparando a diferença de intensidade de todos os pixels das imagens. Essa diferença torna os métodos diretos mais eficientes que os métodos baseados em características (Capítulo 6) em cenários com pouca textura ou com padrões muito repetitivos, onde torna-se difícil a identificação de caracaterísticas.

Entre os métodos de registro de imagem, aqueles baseados em gradientes descendentes são provavelmente os mais populares e mais utilizados. Os métodos baseados em gradiente descendente foram introduzidos por Lucas e Kanade (1981) e diversas variações foram propostas ao longo dos anos. Em seu trabalho, Lucas e Kanade propõe o método de alinhamento em uma dimensão visando solucionar o problema de casamento de pontos em um sistema estéreo. Além de considerar o problema em uma dimensão, também são considerados os casos de maior dimensionalidade.

4.1 Método Direto Baseado em Gradiente Descendente

Os métodos diretos estimam o movimento da câmera e a estrutura associadas a um par de imagens utilizando as informações mesuradas diretamente dos pontos da imagem. Os métodos de gradiente descendente vis am encontrar os parâmetros da função que alinha as duas imagens minimizando a função de custo.

Em (Baker e Matthews, 2004), os autores propõe um modelo para classifiação de métodos de alinhamento de imagem basedos em gradientes descendentes. Os métodos são classificados em métodos por adição (Lucas e Kanade, 1981) e por composição (Shum e Szeliski, 2000) com relação a atualização do parâmetro de alinhamento. Além disso, Baker e Matthews também propõe o método inverso para ambos os métodos. Alguns métodos como o método Minimização Eficiente de Segunda Ordem (Benhimane e Malis, 2004) não se adequavam à classfiicação proposta. Mégret et al. (2008) apresentam uma estrutura mais genérica que a apresentada em (Baker e Matthews, 2004), para classificação dos métodos de alinhamento. A seguir serão apresentados alguns desses métodos que são relevantes para a compreensão deste trabalho.

4.1.1 Aplicação de Transformação

A aplicação $W(\boldsymbol{\mu}, \mathbf{x})$ aplica uma transformação $\mathbf{T}(\boldsymbol{\mu}) \in \mathbb{R}^{n \times n}$ na coordenada homogênea $\mathbf{x} \in \mathbb{R}^n$.

$$W(\boldsymbol{\mu}, \mathbf{x}) = \mathbf{T}(\boldsymbol{\mu})\,\mathbf{x}$$

Por exemplo, seja $\mathbf{x} = [x, y, 1]^T \in \mathbb{R}^3$ o vetor homogêneo representando as coordenadas de um ponto no espaço de uma imagem e $\boldsymbol{\mu} = [\mu_1, \mu_2]^T \in \mathbb{R}^2$ o vetor contendo os parâmetros que descrevem a transformação de translação simples, então a aplicação da transformação

é dada por

$$W(\boldsymbol{\mu}, \mathbf{x}) = \mathbf{T}(\boldsymbol{\mu})\,\mathbf{x} = \begin{bmatrix} 1 & 0 & \mu_1 \\ 0 & 1 & \mu_2 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x + \mu_1 \\ y + \mu_2 \end{bmatrix}.$$

Define-se as seguintes propriedades para a apliação $W(\cdot, \cdot)$

$$W(\boldsymbol{\mu}, W(\boldsymbol{\beta}, \mathbf{x})) = \mathbf{T}(\boldsymbol{\mu})\,\mathbf{T}(\boldsymbol{\beta})\,\mathbf{x} = W(\boldsymbol{\mu} \circ \boldsymbol{\beta}, \mathbf{x}) \tag{4.1}$$

$$W(\mathbf{0}, \mathbf{x}) = \mathbf{x} \tag{4.2}$$

4.1.2 Método Aditivo

O método proposto por Lucas e Kanade (1981) é a primeira referência de métodos de alinhamento de imagem. O método apresentado pelos autores minimiza a diferença entre duas imagens calculando iterativamente os parâmetros de alinhamento calculando passos de primeira ordem. Os parâmetros de alinhamento descrevem a transformação que alinha as imagens.

Sejam $\mathcal{I}_c(\mathbf{x})$ e $\mathcal{I}_r(\mathbf{x})$ a imagem corrente no instante t e a imagem referência no instante $t-1$, respectivamente, o método proposto por Lucas e Kanade busca alinhar a imagem referência com a imagem corrente transformada por $W(\boldsymbol{\mu}, \mathbf{x})$ que minimiza a soma do quadrado das diferenças entre as coordenadas das imagens:

$$\sum_{\mathbf{x}} \left[\mathcal{I}_c(W(\boldsymbol{\mu}, \mathbf{x})) - \mathcal{I}_r(\mathbf{x})\right]^2. \tag{4.3}$$

A atualização entre as iterações do parâmetro $\boldsymbol{\mu}$ é feita incrementando o vetor $\boldsymbol{\mu}$ ao vetor $\boldsymbol{\mu}$ original

$$\boldsymbol{\mu}^{k+1} \leftarrow \boldsymbol{\mu}^k + \boldsymbol{\mu}. \tag{4.4}$$

Pelo fato do vetor de parâmetros ser atualizado por incremento, o método proposto por Lucas e Kanade é denominado método aditivo direto. Aplicando o critério de atualização descrito na Equação 4.4 à Equação 4.3, a minimização da diferença entre as imagens pode ser escrita como

$$E(\boldsymbol{\mu}) = \sum_{\mathbf{x}} \left[\mathcal{I}_c(W\left(\boldsymbol{\mu}^k + \boldsymbol{\mu}, \mathbf{x}\right)) - \mathcal{I}_r(\mathbf{x})\right]^2, \tag{4.5}$$

onde $E(\boldsymbol{\mu})$ é o erro residual da diferença das imagens para todo o $\mathbf{x}$ no espaço da imagem $\mathcal{I}_r$ e $\boldsymbol{\mu}$ é o incremento calculado que aproxima $E(\boldsymbol{\mu})$ do mínimo a cada iteração. O termo $\mathcal{I}_c(W\left(\boldsymbol{\mu}^k + \boldsymbol{\mu}, \mathbf{x}\right))$ é linearizado utilizando a expansão de Taylor de primeira ordem,

resultando em

$$E(\boldsymbol{\mu}) \approx \sum_{\mathbf{x}} \left[\mathcal{I}_c(W\left(\boldsymbol{\mu}^k, \mathbf{x}\right)) + \nabla \mathcal{I}_c \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}} \boldsymbol{\mu} - \mathcal{I}_r(\mathbf{x}) \right]^2 . \tag{4.6}$$

O gradiente $\nabla \mathcal{I}_c$ é o gradiente da imagem I_c posteriormente transformado por $W(\boldsymbol{\mu}, \mathbf{x})$. O jacobiano $\frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}$ é dependete do modelo de transformação escolhido, e.g. para o modelo de translação o resultado o gradiente seria

$$\frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}} = \begin{pmatrix} \frac{\partial \mathbf{W}_x}{\partial \boldsymbol{\mu}_1} & \frac{\partial \mathbf{W}_x}{\partial \boldsymbol{\mu}_2} \\ \frac{\partial \mathbf{W}_y}{\partial \boldsymbol{\mu}_1} & \frac{\partial \mathbf{W}_y}{\partial \boldsymbol{\mu}_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Observe que

$$\frac{\partial E(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = 0,$$

é condição necessária para a minimização da Equação 4.6, onde

$$\begin{aligned} \frac{\partial E(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} &\approx \frac{\partial}{\partial \boldsymbol{\mu}} \sum_{\mathbf{x}} \left[\mathcal{I}_c(W\left(\boldsymbol{\mu}^k, \mathbf{x}\right)) + \nabla \mathcal{I}_c \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}} \boldsymbol{\mu} - \mathcal{I}_r(\mathbf{x}) \right]^2 \\ &= 2 \sum_{\mathbf{x}} \left[\nabla \mathcal{I}_c \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}} \right]^T \left[\mathcal{I}_c(W\left(\boldsymbol{\mu}^k, \mathbf{x}\right)) + \nabla \mathcal{I}_c \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}} \boldsymbol{\mu} - \mathcal{I}_r(\mathbf{x}) \right] . \end{aligned} \tag{4.7}$$

Então, resolvendo a Equação 4.7 para $\boldsymbol{\mu}$ obtém-se

$$\boldsymbol{\mu} = \mathbf{H}^{-1} \sum_{\mathbf{x}} \left[\nabla \mathcal{I}_c \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}} \right]^T \left[\mathcal{I}_c(W(\boldsymbol{\mu}, \mathbf{x})) - \mathcal{I}_r(\mathbf{x}) \right] \tag{4.8}$$

onde $\mathbf{H}$ é o Hessiano

$$\mathbf{H} = \sum_{\mathbf{x}} \left[\nabla \mathcal{I}_c \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}} \right]^T \left[\nabla \mathcal{I}_c \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}} \right]$$

Encontrado o valor de $\boldsymbol{\mu}$ para a iteração, se o valor for maior que um limite, o parâmetro $\boldsymbol{\mu}^{k+1}$ é atualizado utilizando 4.4 e uma nova iteração é repetida.

4.1.3 Método por Composição

O método aditivo encontra o valor de $\boldsymbol{\mu}$ que minimiza a Equação 4.5 incrementando em cada iteração o valor $\boldsymbol{\mu}$, como descrito na Equação 4.4, em direação ao mínimo. O método por composição por sua vez resolve o problema de minimizar o erro do alinhamento de imagens calculando em cada iteração a tranformação que aproxima o erro do mínimo.

A função de custo do método por composição é escrita como

$$E(\boldsymbol{\mu}) = \sum_{\mathbf{x}} \left[\mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k, W\left(\boldsymbol{\mu}, \mathbf{x}\right)\right)\right) - \mathcal{I}_r\left(\mathbf{x}\right)\right]^2, \tag{4.9}$$

Como dito anteriormente, em cada iteração é calculada a transformação $W\left(\boldsymbol{\mu}^k, \mathbf{x}\right)$ que aproxima o erro do mínimo. No passo de atualização da tranformação é realizada a composição da estimativa inicial com o passo que aproxima do mínimo como descrito na Equação 4.10 a seguir

$$W\left(\boldsymbol{\mu}^{k+1}, \mathbf{x}\right) \leftarrow W\left(\boldsymbol{\mu}^k, \mathbf{x}\right) \circ W\left(\boldsymbol{\mu}, \mathbf{x}\right). \tag{4.10}$$

De forma semelante ao método aditivo realiza-se a aproximação de $\mathcal{I}_c(W\left(\boldsymbol{\mu}, \mathbf{x}\right))$ por sua série de Taylor em $\boldsymbol{\mu} = \mathbf{0}$, obtém-se

$$E\left(\boldsymbol{\mu}\right) \approx \sum_{\mathbf{x}} \left[\mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k, W\left(\mathbf{0}, \mathbf{x}\right)\right)\right) + \nabla\mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k, W\left(\mathbf{0}, \mathbf{x}\right)\right)\right)\frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}\boldsymbol{\mu} - \mathcal{I}_r\left(\mathbf{x}\right)\right]^2.$$

Utilizando as propriedades das Equações 4.1 e 4.2 temos que

$$\mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k, W\left(\mathbf{0}, \mathbf{x}\right)\right)\right) = \mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k, \mathbf{x}\right)\right) = \mathcal{I}_w\left(\mathbf{x}\right).$$

A imagem $\mathcal{I}_w\left(\mathbf{x}\right)$ é calculada no ínicio da iteração para calcular o vetor de erro e pode ser reutilizada. Note também que no método por composição a derivada $\frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}$ é avaliada em $(\mathbf{0}, \mathbf{x})$, enquanto no método aditivo a derivada é avaliada em $(\boldsymbol{\mu}^k, \mathbf{x})$, significando que a derivada $\frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}$ pode ser calculada uma única vez antes das iterações.

Resolvendo o problema da minimização do erro de alinhamento da Equação 4.9 para $\boldsymbol{\mu}$, obtém-se

$$\boldsymbol{\mu} = \mathbf{H}^{-1} \sum_{\mathbf{x}} \left[\nabla\mathcal{I}_w \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}\right]^T [\mathcal{I}_w(\mathbf{x}) - \mathcal{I}_r(\mathbf{x})], \tag{4.11}$$

com Hessiano

$$\mathbf{H} = \sum_{\mathbf{x}} \left[\nabla\mathcal{I}_w \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}\right]^T \left[\nabla\mathcal{I}_w \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}\right]. \tag{4.12}$$

4.1.4 Método por Composição Inversa

Baker e Matthews (2004) propõe uma formulação diferente para o problema de alinhamento de imagem. Semelhante ao método por composição, a formulação proposta por

Baker e Matthews utiliza a seguinte função para cálculo do erro

$$E(\boldsymbol{\mu}) = \sum_{\mathbf{x}} \left[\mathcal{I}_r(W(\boldsymbol{\mu}, \mathbf{x})) - \mathcal{I}_c(W\left(\boldsymbol{\mu}^k, \mathbf{x}\right))\right]^2. \tag{4.13}$$

Diferente das Equações 4.5 e 4.9 a Equação 4.13 calcula a tranformação sobre a imagem referência $\mathcal{I}_r$ que minimiza o erro de alinhamento.

A imagem corrente $\mathcal{I}_c$ ainda é recalculada utilizando transformações, mas a aproximação é feita de $\mathcal{I}_r$ para $\mathcal{I}_c$. O passo de atualização da transformação então fica

$$W\left(\boldsymbol{\mu}^{k+1}, \mathbf{x}\right) \leftarrow W\left(\boldsymbol{\mu}^k, \mathbf{x}\right) \circ W(\boldsymbol{\mu}, \mathbf{x})^{-1}. \tag{4.14}$$

Como pode-se notar do passo de atualização um dos requisitos necessários para o uso do método é que as transformações possam ser invertidas. Felizmente esse requisito é satisfeito pela maioria dos tipos de transformações utilizadas nos problemas de alinhamento.

A solução do problema de minimização do erro de alinhamento da Equação 4.13 é semelhante aos dos outros dois métodos. Inicialmente aproxima-se $\mathcal{I}_r(W(\boldsymbol{\mu}, \mathbf{x}))$ por sua série de Taylor de primeira ordem

$$E \approx \sum_{\mathbf{x}} \left[\mathcal{I}_r(W(\mathbf{0}, \mathbf{x})) + \nabla\mathcal{I}_r \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}\boldsymbol{\mu} - \mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k, \mathbf{x}\right)\right)\right]^2. \tag{4.15}$$

Resolvemos o problema de minimização para $\boldsymbol{\mu}$, semelhante a 4.7 obtém-se

$$\boldsymbol{\mu} = \mathbf{H}^{-1} \sum_{\mathbf{x}} \left[\nabla\mathcal{I}_r \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}\right]^T [\mathcal{I}_w(\mathbf{x}) - \mathcal{I}_r(\mathbf{x})], \tag{4.16}$$

com Hessiano

$$\mathbf{H} = \sum_{\mathbf{x}} \left[\nabla\mathcal{I}_r \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}\right]^T \left[\nabla\mathcal{I}_r \frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}\right]. \tag{4.17}$$

Note que assim como no método por composição de Shum e Szeliski a derivada $\frac{\partial \mathbf{W}}{\partial \boldsymbol{\mu}}$ é calculada em $(\mathbf{0}, \mathbf{x})$, podendo assim ser préviamente calculada às iterações. A diferença está no gradiente da imagem, enquanto o método de Shum e Szeliski utiliza o gradiente $\nabla\mathcal{I}_c$ que precisa ser calculado a cada nova iteração, o método proposto por Baker e Matthews utiliza o gradiente $\nabla\mathcal{I}_r$ que também pode ser calculado antes das iterações. Isso significa que o calculo Hessiano, que tem um custo de processamento alto, pode ser realizado uma única vez antes do início das iterações.

4.1.5 Minimização Eficiente de Segunda Ordem

Os métodos por adição, composição e composição inversa apresentados nas Subseções anteriores utilizam o método de Gauss-Newton para resolver o problema de minimização de mínimos quadrados descrito pelas funções de custo. O método de Gauss-Newton é uma aproximação do método Newton-Raphson. Outros métodos de minimização são descritos e comparados em Baker e Matthews (2004). O uso de aproximações para solução do problema de minimização ocorre devido ao custo elevado da matriz Hessiana e porque podem ocorrer problemas na convergência se a matriz não for positiva.

Apesar dos problemas da matriz Hessiana, o método Newton-Raphson apresenta uma característica desejável que é a taxa de convergência alta. Um método para resolução do problema de minimização com alta taxa de convergência e sem o custo do cálculo da matriz Hessiana é apresentado em Malis (2004); Benhimane e Malis (2004). O método é denominado Minimização Eficiente de Segunda Ordem (ESM, do inglês *Efficient Second-Order Minimization*), pois utiliza a segunda ordem da série de Taylor para encontrar o minimizador local de cada iteração, semelhante ao método Newton-Raphson.

Podemos escrever o erro de alinhamento na forma geral de um problema de minimização em relação ao vetor de parâmetros $\boldsymbol{\mu}$ como

$$E\left(\boldsymbol{\mu}\right) = \sum_{\mathbf{x}} \frac{1}{2} f\left(\boldsymbol{\mu}, \mathbf{x}\right)^2, \tag{4.18}$$

onde

$$f\left(\boldsymbol{\mu}\right) = \mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k \circ \boldsymbol{\mu}, \mathbf{x}\right)\right) - \mathcal{I}_r\left(\mathbf{x}\right) \tag{4.19}$$

é a função a ser minimizada. Expandindo a função $f\left(\boldsymbol{\mu}\right)$ por sua série de Taylor em segunda ordem em torno de $\boldsymbol{\mu} = \mathbf{0}$, obtém-se

$$f\left(\boldsymbol{\mu}\right) = f\left(\mathbf{0}\right) + \mathbf{J}_c\left(\mathbf{0}\right)\boldsymbol{\mu} + \frac{1}{2}\mathbf{M}\left(\mathbf{0}, \boldsymbol{\mu}\right)\boldsymbol{\mu} + \mathbb{O}(3), \tag{4.20}$$

onde $\mathbb{O}(3)$ são os termos de terceira ordem e maiores,

$$\mathbf{J}_c\left(\mathbf{0}\right) = \left.\frac{\partial f\left(\boldsymbol{\mu}\right)}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \left.\frac{\partial \mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k \circ \boldsymbol{\mu}, \mathbf{x}\right)\right)}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\mathbf{0}}$$

e

$$\mathbf{M}\left(\mathbf{0}, \boldsymbol{\mu}\right) = \left.\frac{\partial \mathbf{J}_c\left(\boldsymbol{\mu}\right)}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\mathbf{0}} \boldsymbol{\mu}.$$

Expandindo agora $\mathbf{J}_c(\boldsymbol{\mu})$ em séries de Taylor em primeira ordem em torno de $(\mathbf{0})$

$$\mathbf{J}_c(\boldsymbol{\mu}) = \mathbf{J}_c(\mathbf{0}) + M(\mathbf{0}, \boldsymbol{\mu}) + \mathbb{O}(2), \tag{4.21}$$

onde $\mathbb{O}(2)$ são os termos de segunda ordem ou maiores. Substituindo a Equação 4.21 na Equação 4.20 e descartando os termos de ordem maior, obtém-se

$$f(\boldsymbol{\mu}) \approx f(\mathbf{0}) + \left(\frac{\mathbf{J}_c(\mathbf{0}) + \mathbf{J}_c(\boldsymbol{\mu})}{2}\right)\boldsymbol{\mu}. \tag{4.22}$$

Com esta substituição fica-se livre da necessidade de calular a matriz Hessiana. Note porém que o jacobiano $\mathbf{J}_c(\boldsymbol{\mu})$ é definido em função do vetor $\boldsymbol{\mu}$, justamente a variável que deseja-se encontrar. Assumimos aqui que para um valor pequeno de $\boldsymbol{\mu}$ temos que

$$\mathbf{J}_c(\boldsymbol{\mu})\,\boldsymbol{\mu} = \mathbf{J}_r(\mathbf{0})\,\boldsymbol{\mu}, \tag{4.23}$$

onde

$$\mathbf{J}_r(\mathbf{0}) = \left.\frac{\partial \mathcal{I}_r(W(\boldsymbol{\mu}, \mathbf{x}))}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\mathbf{0}}.$$

Essa suposição é satisfeita, atendidas algumas condições que serão apresentadas na Seção seguinte.

Novamente, busca-se o valor de mínimo resolvendo $\frac{\partial E(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = \mathbf{0}$ para $\boldsymbol{\mu}$ e obtém-se

$$\boldsymbol{\mu} = -\left(\frac{\mathbf{J}_c(\mathbf{0}) + \mathbf{J}_r(\mathbf{0})}{2}\right)^{+} f(\mathbf{0}), \tag{4.24}$$

onde $(\cdot)^{+}$ é o operador de pseudo-inversa para matrizes.

A atualização da transformação entre as iterações é realizado por composição, semelhante aos métodos de Shum e Szeliski; Baker e Matthews

$$W\left(\boldsymbol{\mu}^{k+1}, \mathbf{x}\right) \leftarrow W\left(\boldsymbol{\mu}^{k}, \mathbf{x}\right) \circ W(\boldsymbol{\mu}, \mathbf{x}). \tag{4.25}$$

4.2 Estimação de Odometria com Escala

Os métodos diretos apresentados na Seção 4.1 buscam estimar os parâmetros que descrevem um modelo de transformação e que minimizam a diferença do alinhamento entre duas imagens. As tranformações representadas pelo modelo paramétrico, desde transformações simples como a descrito em Lucas e Kanade (1981), onde o deslocamento

da imagem ocorrem em uma dimensão, até transformações mais complexas como as de corpo rígido.

Nas próximas Subseções será apresentado um modelo de transformação que se adequa ao movimento de veículos em áreas urbanas e a derivação dos jacobianos utilizando o método ESM.

4.2.1 Modelo Paramétrico

Considere o deslocamento de um veículo entre os instantes $t-1$ e t. Chamamos de estado referência r, o instante $t-1$ e estado corrente c o instante t. Em um cenário urbano é plausível assumir que o veículo se desloca em uma superfície localmente plana. Pode-se então descrever o deslocamento do veículo V entre o estado referêcia e o estado corrente é descrito pela transformação

$$^{V_c}\mathbf{T}_{V_r}(\boldsymbol{\mu}) \in SE(2), \tag{4.26}$$

onde V_r e V_c são respectivamente, os estados referência e corrente do veículo e $\boldsymbol{\mu}$ é o vetor que parametriza a transformação $\mathbf{T}$.

Suponha uma câmera C rígidamente fixada ao veículo V, a transformação da câmera em relação ao veículo é descrita como

$$^{V}\mathbf{T}_{C} = {^{C}\mathbf{T}_{V}}^{-1}, \tag{4.27}$$

e a transformação entre a câmera do estado r para o estado c é descrita como

$$^{C_c}\mathbf{T}_{C_r} = {^{C}\mathbf{T}_{V}}\,{^{V_c}\mathbf{T}_{V_r}}(\boldsymbol{\mu})\,{^{V}\mathbf{T}_{C}}. \tag{4.28}$$

Note que a notação de estado foi descartada da transformação entre câmera e veículo, pois a mesma é constante uma vez que considera-se uma câmera rigidamente ligada ao veículo.

A Equação 4.28 descreve a relação entre o deslocamento da câmera e o deslocamento do veículo. Como visto no Capítulo 3, a imagem se relaciona com pontos no sistema de coordenadas da câmera através da matriz de projeção, ou para o caso específicos de regiões planas, através de homografias. A forma da homografia que relaciona planos de duas imagens é dada por

$$\mathbf{H} = \mathbf{K}\left(\mathbf{R} - \frac{\mathbf{t}\hat{\mathbf{n}}^T}{d}\right)\mathbf{K}^{-1} = \mathbf{K}\ \mathbf{T}\left(\ \mathbf{I}\ \middle|\ -\tfrac{\hat{\mathbf{n}}}{d}\ \right)^T\mathbf{K}^{-1}, \tag{4.29}$$

onde $\mathbf{T} = \left(\mathbf{R} \,\middle|\, \mathbf{t} \right)$, $\mathbf{I} \in \mathbb{R}^{3\times3}$ é a matriz identidade, $\hat{\mathbf{n}}$ é o vetor normal no sistema de coordenadas de C_c e d é a distância da câmera ao plano utilizado para relacionar as imagens, ambos relativos à superfície da rua e $\mathbf{K} \in \mathbb{R}^{3\times3}$ é a matriz de parâmetros intrínsecos da câmera. Substituindo a Equação 4.28 em

$$ {}^{C_c}\mathbf{H}_{C_r}(\boldsymbol{\mu}) = \mathbf{K}\ {}^{C}\mathbf{T}_{V}\,{}^{V_c}\mathbf{T}_{V_r}(\boldsymbol{\mu})\,{}^{V}\mathbf{T}_{C} \left(\mathbf{I} \,\middle|\, -\tfrac{\hat{\mathbf{n}}}{d} \right)^{T} \mathbf{K}^{-1}. \tag{4.30} $$

A função descrita pela Equação 4.30 relaciona as duas imagens $\mathcal{I}_r$ e $\mathcal{I}_c$. Utiliza-se a função $\boldsymbol{\pi}$ para a desomogenização das coordenadas coordenadas da imagem. Temos então que a função de transformação entre as imagens é dada como

$$ W(\boldsymbol{\mu}, \mathbf{x}) = \boldsymbol{\pi}\left(\mathbf{H}(\boldsymbol{\mu}), \mathbf{x}\right). \tag{4.31} $$

Sendo

$$ \mathbf{H}(\boldsymbol{\mu}) = \begin{bmatrix} h_{11}(\boldsymbol{\mu}) & h_{12}(\boldsymbol{\mu}) & h_{13}(\boldsymbol{\mu}) \\ h_{21}(\boldsymbol{\mu}) & h_{22}(\boldsymbol{\mu}) & h_{23}(\boldsymbol{\mu}) \\ h_{31}(\boldsymbol{\mu}) & h_{32}(\boldsymbol{\mu}) & h_{33}(\boldsymbol{\mu}) \end{bmatrix}, $$

então

$$ \boldsymbol{\pi}\left(\mathbf{H}(\boldsymbol{\mu}), \mathbf{x}\right) = \begin{bmatrix} \frac{h_{11}(\boldsymbol{\mu})x + h_{12}(\boldsymbol{\mu})y + h_{13}(\boldsymbol{\mu})}{h_{31}(\boldsymbol{\mu})x + h_{32}(\boldsymbol{\mu})y + h_{33}(\boldsymbol{\mu})} \\ \frac{h_{21}(\boldsymbol{\mu})x + h_{22}(\boldsymbol{\mu})y + h_{23}(\boldsymbol{\mu})}{h_{31}(\boldsymbol{\mu})x + h_{32}(\boldsymbol{\mu})y + h_{33}(\boldsymbol{\mu})} \\ 1 \end{bmatrix}. $$

As mesmas propriedades de composição e elemento neutro da transformação $W(\cdot,\cdot)$ são válidas para $\boldsymbol{\pi}(\cdot,\cdot)$. Note que as referências aos sistemas de coordenadas foram omitidos e serão omitidos daqui em diante, ficando a cargo do leitor a interpretação dos mesmos.

4.2.2 Parametrização do Movimento Através da Algebra Lie

No método ESM, apresentado na Seção 4.1, fazemos a suposição 4.23 para valores pequenos de $\boldsymbol{\mu}$. Nesta Seção é apresentada a parametrização do problema através dos grupos Lie, de forma que a suposição seja válida. Maiores detalhes a respeito dos grupos Lie podem ser encontrados em Varadarajan (1984); Hall (2003).

Um grupo Lie G é uma variedade cujas operações de grupo

$$ (g,h) \in G \times G \to gh \in G $$
$$ g \in G \to g^{-1} \in G $$

são aplicações diferenciáveis. O grupo linear geral é um grupo Lie definido como

$$GL(n) = \left\{ \mathbf{A} \in \mathbb{R}^{n \times n} | \det \mathbf{A} \neq 0 \right\}$$

com a seguinte propriedade:

Propriedade 4.1. *Dado G um subgrupo de $GL(n)$, seja $\mathbf{A}_m$ uma sequência de matrizes em G, se $\mathbf{A}_m$ convergir para alguma matriz $\mathbf{A}$, então $\mathbf{A} \in G$ ou $\mathbf{A}$ não é inversível.*

Assumimos anteriormente que o deslocamento do veículo descreve um movimento planar sobre a superfície da via. Tal movimento pode ser descrito pela transformação

$$T(\boldsymbol{\mu}) = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta & x \\ -\operatorname{sen}\theta & \cos\theta & y \\ 0 & 0 & 1 \end{bmatrix}, \tag{4.32}$$

onde $\boldsymbol{\mu} = [x, y, \theta]^T$. A transformação $T(\cdot)$ pertence ao grupo $SE(2)$, que é um subgrupo de $GL(3, \mathbb{R})$ e também um grupo Lie.

Existe para todo grupo Lie G um espaço vetorial $\mathfrak{g}$ tangente à identidade de G denominado álgebra Lie. Esse espaço vetorial é composto pelos vetores tangentes aos caminhos diferenciáveis que passam pela identidade. Um caminho $\mathbf{A}(t) \in G$ é dito diferenciável se a derivada $\mathbf{A}'(t)$ existe e $\mathbf{A}(0)$ é a identidade de G para todo t pertencente a um intervalo real. Seja então

$$\mathbf{A}(t) = \begin{bmatrix} \cos\theta(t) & \operatorname{sen}\theta(t) & x(t) \\ -\operatorname{sen}\theta(t) & \cos\theta(t) & y(t) \\ 0 & 0 & 1 \end{bmatrix}$$

um caminho diferenciável em $SE(2)$, com $\mathbf{A}(0) = \mathbf{I}$. Derivando a matriz $\mathbf{A}(t)$ em $t = 0$, obtemos

$$\left.\frac{d\mathbf{A}(t)}{dt}\right|_{t=0} = \begin{bmatrix} 0 & -\dot{\theta}(0) & \dot{x}(0) \\ \dot{\theta}(0) & 0 & \dot{y}(0) \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -\mu_3 & \mu_1 \\ \mu_3 & 0 & \mu_2 \\ 0 & 0 & 1 \end{bmatrix}.$$

A álgebra Lie $\mathfrak{se}(2)$ relacionada ao grupo Lie $SE(2)$ é então

$$\mathfrak{se}(2) = \left\{ \mathbf{A} \in \mathbb{R}^{3\times 3} | \mathbf{A} = \left[\begin{smallmatrix} 0 & -\mu_3 & \mu_1 \\ \mu_3 & 0 & \mu_2 \\ 0 & 0 & 1 \end{smallmatrix} \right]; \mu_1, \mu_2, \mu_3 \in \mathbb{R} \right\}.$$

A base de $\mathfrak{se}(2)$ é compostas pelas matrizes

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{A}_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

que são chamadas geradores da álgebra Lie $\mathfrak{se}(2)$ e os elementos da álgebra Lie $\mathfrak{se}(2)$ são gerados como

$$\mathbf{A}(\boldsymbol{\mu}) = \sum_{i=0}^{3} \mu_i \mathbf{A}_i \tag{4.33}$$

A relação entre o grupo Lie G e a álgebra Lie $\mathfrak{g}$ associada é dada pelo mapa exponencial. O mapa exponencial não fornece uma associação de um para um da álgebra para o grupo, mas localmente na vizinhança da identidade o mapa é bijetivo. O mapa exponencial de $\mathfrak{se}(2)$ para $SE(2)$ é dado como

$$e^{\mathbf{A}(\boldsymbol{\mu})} = \mathbf{I} + \frac{\operatorname{sen}\mu_3}{\mu_3}\mathbf{A} + \frac{(1-\cos\mu_3)}{{\mu_3}^2}\mathbf{A}^2 = \begin{bmatrix} \cos\mu_3 & -\operatorname{sen}\mu_3 & \frac{\mu_1 \operatorname{sen}\mu_3 + \mu_2 \cos\mu_3 - \mu_2}{\mu_3} \\ \operatorname{sen}\mu_3 & -\cos\mu_3 & \frac{\mu_2 \operatorname{sen}\mu_3 - \mu_1 \cos\mu_3 - \mu_1}{\mu_3} \\ 0 & 0 & 1 \end{bmatrix}.$$

Para o caso de μ_3 pequeno, o limite de $\mu_3 \to 0$ resulta em

$$e^{\mathbf{A}(\boldsymbol{\mu})} = \begin{bmatrix} 1 & 0 & \mu_1 \\ 0 & 1 & \mu_2 \\ 0 & 0 & 1 \end{bmatrix}.$$

4.2.3 Prova da Suposição de Igualdade no Método ESM

A suposição descrita na Equação 4.23 é ponto central no desenvolvimento do método. A seguir será mostrado que a suposição da Equação 4.23 é válida, se a função $W(\boldsymbol{\mu}, \mathbf{x})$ for uma aplicação de um elemento do grupo Lie em um ponto. A prova foi adaptada de Authesserres (2010); Comport et al. (2010). Abordagem semelhante é apresentada em Mei (2007) e outra abordagem é apresentada em Benhimane (2006).

Retornando à Equação 4.23 temos

$$\mathbf{J}_c(\boldsymbol{\mu})\,\boldsymbol{\mu} = \left.\frac{\partial \mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k \circ \boldsymbol{\mu}, \mathbf{x}\right)\right)}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}} \boldsymbol{\mu},$$

onde é realizada a seguinte substituição de variáveis $\boldsymbol{\mu} = \boldsymbol{\mu} + \boldsymbol{\omega}$, para se obter

$$\mathbf{J}_c(\boldsymbol{\mu})\,\boldsymbol{\mu} = \left.\frac{\partial \mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k \circ (\boldsymbol{\mu}+\boldsymbol{\omega}), \mathbf{x}\right)\right)}{\partial \boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=0} \frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\mu}}\boldsymbol{\mu}.$$

Devido a parametrização em através da álgebra Lie, pode-se utilizar a seguinte propriedade

$$W(\boldsymbol{\mu}+\boldsymbol{\omega}, \mathbf{x}) = \mathbf{T}(\boldsymbol{\mu}+\boldsymbol{\omega})\,\mathbf{x} = \mathbf{T}(\boldsymbol{\mu})\,\mathbf{T}(\boldsymbol{\omega})\,\mathbf{x} = W(\boldsymbol{\mu}\circ\boldsymbol{\omega}, \mathbf{x})$$

e portanto

$$\mathbf{J}_c(\boldsymbol{\mu})\,\boldsymbol{\mu} = \left.\frac{\partial \mathcal{I}_c\left(W\left(\boldsymbol{\mu}^k \circ \boldsymbol{\mu} \circ \boldsymbol{\omega}, \mathbf{x}\right)\right)}{\partial \boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=0} \boldsymbol{\mu}.$$

Assume-se aqui que o parâmetro $\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}^k \circ \boldsymbol{\mu}$, é a solução exata e portanto temos que

$$\mathbf{J}_c(\boldsymbol{\mu})\,\boldsymbol{\mu} = \left.\frac{\partial \mathcal{I}_r(W(\boldsymbol{\omega}, \mathbf{x}))}{\partial \boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=0} \boldsymbol{\mu} = \mathbf{J}_r(\mathbf{0})\,\boldsymbol{\mu}.$$

4.2.4 Derivação da Função de Custo

Nesta Subseção será apresentada a derivação do método ESM utilizando o modelo de movimento descrito na Subseção 4.2.1. Substituindo a Equação 4.31 na Equação 4.19, obtém-se

$$f(\boldsymbol{\mu}) = \mathcal{I}_c\left(\boldsymbol{\pi}\left(\mathbf{H}\left(\boldsymbol{\mu}^k\right), \boldsymbol{\pi}\left(\mathbf{H}(\boldsymbol{\mu}), \mathbf{x}\right)\right)\right) - \mathcal{I}_r(\mathbf{x}) \tag{4.34}$$

e derivando a Equação 4.34 obtém-se

$$\mathbf{J}_c(\boldsymbol{\mu}) = \left.\frac{\partial \mathcal{I}_c\left(\boldsymbol{\pi}\left(\mathbf{H}\left(\boldsymbol{\mu}^k\right), \mathbf{a}\right)\right)}{\partial \mathbf{a}}\right|_{\mathbf{a}=\boldsymbol{\pi}(\mathbf{H}(\boldsymbol{\mu}),\mathbf{x})} \left.\frac{\partial \boldsymbol{\pi}(\mathbf{b}, \mathbf{x})}{\partial \mathbf{b}}\right|_{\mathbf{b}=\mathbf{H}(\boldsymbol{\mu})} \frac{\partial \mathbf{H}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}, \tag{4.35}$$

onde

$$\frac{\partial \mathbf{H}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = \mathbf{K}^C\mathbf{T}_V \frac{\partial \mathbf{T}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}{}^V\mathbf{T}_C \left(\begin{array}{c|c} \mathbf{I} & -\frac{\hat{\mathbf{n}}}{d} \end{array}\right)^T \mathbf{K}^{-1}. \tag{4.36}$$

Utiliza-se a algebra Lie de $SE(2)$ para parametrizar localmente $\frac{\partial \mathbf{T}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$

$$\mathbf{T}(\boldsymbol{\mu}) = \exp\left(\sum_{i=1}^{3} \mu_i \mathbf{A}_i\right),$$

onde $\mathbf{A}_i$ são os geradores da algebra Lie associada a $SE(2)$. A atualização da transformação é então dada como

$$\mathbf{T}\left(\boldsymbol{\mu}^{k+1}\right) \leftarrow \mathbf{T}\left(\boldsymbol{\mu}^{k}\right) \exp\left(\sum_{i=1}^{3} \mu_i \mathbf{A}_i\right).$$

Retornando aos Jacobiano da Equação 4.24 e utilizando a Equação 4.35, aplicas-e a regra da cadeia para encontrar $\mathbf{J}_c\left(\mathbf{0}\right)$ como

$$\mathbf{J}_c\left(\mathbf{0}\right) = \left.\frac{\partial \mathcal{I}_c\left(\boldsymbol{\pi}\left(\mathbf{H}\left(\boldsymbol{\mu}^k \circ \boldsymbol{\mu}\right)\mathbf{x}\right)\right)}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \mathbf{J}_{\mathcal{I}_c}\mathbf{J}_{\boldsymbol{\pi}}\mathbf{J}_{\mathbf{H}}$$

e

$$\mathbf{J}_r\left(\mathbf{0}\right) = \left.\frac{\partial \mathcal{I}_r\left(\boldsymbol{\pi}\left(\mathbf{H}\left(\boldsymbol{\mu}\right)\mathbf{x}\right)\right)}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \mathbf{J}_{\mathcal{I}_r}\mathbf{J}_{\boldsymbol{\pi}}\mathbf{J}_{\mathbf{H}}. \tag{4.37}$$

Pode-se então reescrever a Equação 4.24 como

$$\boldsymbol{\mu} = -\left(\left(\frac{\mathbf{J}_{\mathcal{I}_c} + \mathbf{J}_{\mathcal{I}_r}}{2}\right)\mathbf{J}_{\boldsymbol{\pi}}\mathbf{J}_{\mathbf{H}}\right)^{+} f(\mathbf{0}), \tag{4.38}$$

O Jacobiano $\mathbf{J}_{\mathcal{I}_c}$ é cálculado como

$$\mathbf{J}_{\mathcal{I}_c} = \left.\frac{\partial \mathcal{I}_c\left(\boldsymbol{\pi}\left(\mathbf{H}\left(\boldsymbol{\mu}^k\right),\mathbf{a}\right)\right)}{\partial \mathbf{a}}\right|_{\mathbf{a}=\boldsymbol{\pi}(\mathbf{H}(\mathbf{0}),\mathbf{x})}$$

que é a derivada da imagem da iteração atual, enquanto

$$\mathbf{J}_{\mathcal{I}_r} = \left.\frac{\partial \mathcal{I}_r\left(\mathbf{a}\right)}{\partial \mathbf{a}}\right|_{\mathbf{a}=\boldsymbol{\pi}(\mathbf{H}(\mathbf{0}),\mathbf{x})}$$

é a derivada da imagem referência. O Jacobiano $\mathbf{J}_{\mathcal{I}_r}$ precisa ser calculado toda iteração, umas vez que depende de $\boldsymbol{\mu}^k$, mas o Jacobiano $\mathbf{J}_{\mathcal{I}_r}$ pode ser calculado antes das iterações. As derivadas das imagens podem ser calculadas utilizando, por exemplo, o operador de Söbel.

O segundo termo, comum a ambos os Jacobianos, é a derivada da função de desomogenização da Equação 4.31. A derivada da função resulta em

$$\mathbf{J}_{\boldsymbol{\pi}} = \begin{bmatrix} \mathbf{x}^T & \mathbf{0}^T & -x\mathbf{x}^T \\ \mathbf{0}^T & \mathbf{x}^T & -y\mathbf{x}^T \\ \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T \end{bmatrix}, \tag{4.39}$$

onde $\mathbf{x} = [u, v, 1]^T$.

O Jacobiano $\mathbf{J_H}$ é constante apesar do termo $\frac{\partial \mathbf{T}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$. Pode-se reescrever o termo como

$$\frac{\partial \mathbf{T}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = \begin{bmatrix} \vec{\mathbf{A}}_1 & \vec{\mathbf{A}}_2 & \vec{\mathbf{A}}_3 \end{bmatrix} \tag{4.40}$$

onde $\vec{\mathbf{A}}_i$ é a forma vetorizada da base $\mathbf{A}_i, i = \{1, 2, 3\}$ para a álgebra Lie $\mathfrak{se}(2)$. Tem-se então que

$$\mathbf{J_H} = \mathbf{K}^C\mathbf{T}_V \begin{bmatrix} \vec{\mathbf{A}}_1 & \vec{\mathbf{A}}_2 & \vec{\mathbf{A}}_3 \end{bmatrix} {}^V\mathbf{T}_C \left(\begin{array}{c|c} \mathbf{I} & -\frac{\hat{\mathbf{n}}}{d} \end{array} \right)^T \mathbf{K}^{-1}$$

4.3 Estimação da Escala

Como descrito nos anteriormente considera-se para o problema de odometria, o cenário de um veículo trafegando em uma via localmente plana. Nestas condições, a imagem da via em duas cenas diferentes induz uma homografia, a mesma associada à transformção induzida pelo deslocamento da câmera entre duas imagens. A Equação da homografia é descrita em 4.29. Conhecendo então o vetor normal ao plano e a distância do plano à câmera pode-se estimar com escala absoluta a tranformação entre a câmera e consequentemente a transformação de qualquer corpo rigidamente ligada a ela.

Dada as suposições anteriores, pode-se utilizar regiões da via identificadas em imagnes diferentes para induzir a homografia necessária para estimar a transformação. Utilizando a via temos que a distância d da câmera ao plano é a altura em que a câmera foi instalada e o vetor normal é dado por

$$\hat{\mathbf{n}} = \mathbf{R}_{roll}\mathbf{R}_{pitch}\mathbf{n}_z,$$

onde $\mathbf{R}_{roll}$ e $\mathbf{R}_{pitch}$ são respectivamente as matrizes de rotação e inclinação da câmera e $\mathbf{n}_z$ é o vetor unitário na direção do eixo Z no sistema de coordenadas do mundo.

Calculada a transformação, pode-se extrair a escala absoluta $\alpha = \|\mathbf{t}\|$. Uma vez estimada, a escala pode ser aplicada à translação da transformação estimada por outros métodos menos restritivos. Nas Seções seguintes será feita a apresentação de um desses métodos.

Capítulo

5

Características e Descritores da Imagem

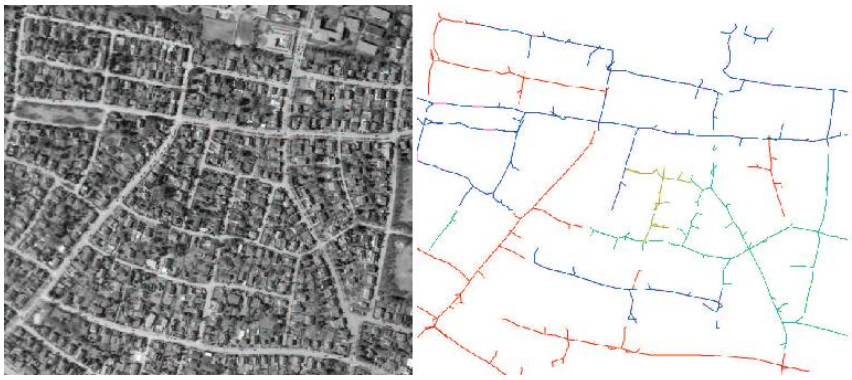
Métodos de odometria visual baseados em reconstrução 3D necessitam de referências em uma cena que possam ser identificadas em imagens de diversos pontos de vista. Essas referências são chamadas **características**. Neste Capítulo serão abordados métodos para a detecção das características de uma imagem (Seção 5.1) e extração de uma estrutura que descreva a região em que se encontra a característica (Seção 5.2), permitindo a identificação da características em imagens diferentes.

5.1 Detectores de Características

Como o próprio nome sugere, características são pontos ou regiões representativas da imagem. Uma característica pode ser uma região de uma cor específica, um determinado padrão na imagem, um ponto onde ocorre a variação de cores ou outra informação que possa ser extraídas da imagem.

O uso de características pode ser identificado em diversas aplicações (Figura 5.1) como a identificação de rodovias com base em mapas aéreos utilizando segmentação de bordas, reconhecimento de faces, reconstrução de modelos 3D, entre outras. No caso da odometria visual baseada em reconstrução 3D, são utilizadas características pontuais como valores mínimos e máximos de funções como gradiente da imagem e a diferença de gaussianas.

Para a extração de características de uma imagem são usados os *detectores* de características. Os detectores utilizam alguma função e métrica para identificar e selecionar as



(a) Detecção de Ruas

(b) Detecção de Faces

(c) Reconstrução 3D

Figura 5.1: Exemplo de aplicações que utilizam características.

cararacterísticas da imagem. Os detector apresentam algumas propriedades referentes às características identificadas e ao seu desempenho descritas por Tuytelaars e Mikolajczyk (2007). As propriedades interessantes para este trabalho são: repetibilidade, diferenciabilidade, quantidade e eficiência.

A **repetibilidade** garante que uma grande porcentagem das características visíveis a duas imagens serão identificadas em ambas as imagens. Caso as imagens sejam tiradas de pontos de vista muito diferentes, como no caso da identificação de objetos em um ambiente, a **invariância** se torna uma propriedade da característica importante para a repetibilidade. A invariância a uma transformação garante que a característica será identificada igualmente com ou sem a tranformação.

A **diferenciabilidade** de uma característica esta mais ligada aos descritores, que serão apresentados mais adiante no capítulo, mas espera-se que o detector gere informação sobre a característica de uma imagem suficiente para identificá-la em outra imagem da mesma cena e diferenciá-la das demais características. Idealmente cada característica em uma imagem deve ser unicamente identificável tanto na mesma, quanto em outras imagens.

A **quantidade** de características identificadas na imagem influencia a qualidade do resultado da aplicação que usa o detector. Em geral, quanto maior a quantidade de características identificadas na imagem, melhor o resultado da aplicação. Em alguns

casos a aplicação exige uma quantidade mínima de características necessárias para a execução, como no caso da odometria visual que necessita de pelo menos cinco pontos correlacionados em duas imagens diferentes (Kruppa, 1913; Nistér, 2004) para a estimação da odometria. O aumento da quantidade de características, em geral, melhora a qualidade do resultado final da aplicação, mas acarreta em um custo maior para o procesamento da informação. Portanto a quantidade de características na imagem deve ser suficiente para obter bons resultados da aplicação sem que haja diminuição no desempenho.

A **eficiência** está relacionada ao tempo que o algoritmo de detecção gasta para identificar as características de uma imagem. Muitas aplicações de visão computacional não apresentam requisito exigentes quanto ao tempo de processamento, como no caso de construção de vistas panorâmicas, e podem utilizar detectores mais lentos, mas que forneçam características mais precisas ou diferenciáveis. Por outro lado aplicações de tempo real, como a odometria visual no caso de veículos autônomos, necessitam de detectores eficientes, muitas vezes penalizando a qualidade do resultado final.

Nas Subseções 5.1.1 e 5.1.2 serão apresentados dois detectores muito utilizados em sistemas de odometria visual. Os detectores de Harris e de Shi-Tomasi utilizam o gradiente da imagem para determinar a formação de cantos e são descritos primeiro. O segundo detector utiliza diferença de gaussianas para obter características invariantes à escala e rotação.

5.1.1 Detector de Cantos

Um canto em uma imagem é o ponto onde há uma grande variação da intensidade dos *pixels* em duas direções dominantes. Moravec (1980) desenvolveu um sistema robótico autônomo utilizando uma câmera em um sistema estéreo e propôs em seu trabalho um detector de características. As caraceterísticas identificadas no trabalho de Moravec (1980) são chamadas cantos.

O detector de Moravec calcula a variação de intensidade de uma imagem em tons de cinza. A variação é detectada dentro de uma janela reduzida da imagem em torno de um ponto no centro dessa janela. A detecção da variação é realizada em quatro direções principais $(u, v) = \{(1, 0); (1, 1); (0, 1); (-1, 1)\}$. Utilizando uma janela de tamanho $2w+1$ a seguinte equação representa a variação de intensidade na direção (u, v)

$$E_{u,v}(x, y) = \sum_{i=-w}^{w} \sum_{j=-w}^{w} (I_{x+i,y+j} - I_{x+i+u,y+j+v})^2, \tag{5.1}$$

onde $I_{x,y}$ é a intensidade do *pixel* na posição (x, y). Se o menor valor de $E_{u,v}(x, y)$ obtido para as quatro direções estiver abaixo de um limiar, o centro da janela é uma característica interessante da imagem e é chamada de canto. O detector de Moravec é um detector de cantos simples e apresenta alguns problemas. Diversos detectores foram propostos ao longo dos anos para solucionar os problemas do detector de Moravec e também aumentar sua eficiência.

Segundo Harris e Stephens (1988) o detector de Moravec apresenta três problemas principais e propõe um novo detector para solucionar esses problemas. O primeiro problema abordado por Harris e Stephen é a anisotropia da resposta. Como o detector de Moravec verifica a variação de intensidade somente em um conjunto discreto de direções, pequenas rotações geram resposta completamente diferentes como exemplificado na figura 5.2. Para cobrir todas as direções o detector de Harris expande o termo da variação

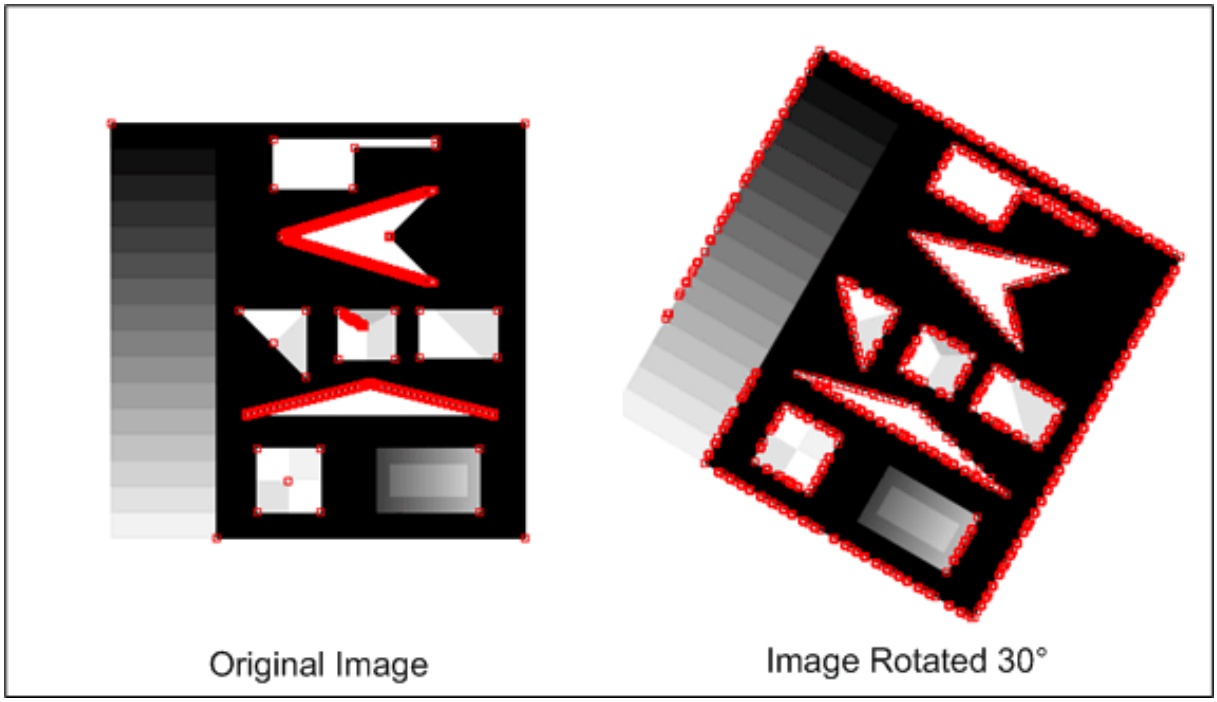


Figura 5.2: Exemplo de anisotropia utilizando o detector de Moravec. A imagem da direita está rotacionada de 30° em relação à da esquerda. Fonte: Parks e Gravel (2013)

utillizando séries de Taylor da seguinte forma

$$I_{x+u,y+v} = I_{x,y} + \frac{\partial I_{x,y}}{\partial x}u + \frac{\partial I_{x,y}}{\partial y}v. \tag{5.2}$$

A Equação 5.2 é substituída na Equação 5.1 obtendo

$$E_{u,v}(x,y) = \sum_{i=-w}^{w}\sum_{j=-w}^{w} G(i,j)\left(\frac{\partial I_{x+i,y+j}}{\partial x+i}u + \frac{\partial I_{x+i,y+j}}{\partial y+j}v\right)^2. \tag{5.3}$$

A matriz $G(i,j)$ é uma janela circular proposta por Harris e Stephens (1988) para suavização do ruído da imagem e é melhor explicada mais adiante na Subseção. Expandindo o termo quadrático na Equação 5.3 e aplicando as devidas distribuições obtém-se

$$E_{u,v}(x,y) = A(x,y)u^2 + 2C(x,y)uv + B(x,y)v^2, \tag{5.4}$$

onde

$$\begin{aligned} A(x,y) &= \sum_{i=-w}^{w}\sum_{j=-w}^{w}\left(\frac{\partial I_{x+i,y+j}}{\partial x}\right)^2 \\ B(x,y) &= \sum_{i=-w}^{w}\sum_{j=-w}^{w}\left(\frac{\partial I_{x+i,y+j}}{\partial y}\right)^2 \\ C(x,y) &= \sum_{i=-w}^{w}\sum_{j=-w}^{w}(\frac{\partial I_{x+i,y+j}}{\partial x})(\frac{\partial I_{x+i,y+j}}{\partial y}). \end{aligned} \tag{5.5}$$

Essa abordagem permite que todas as direções sejam avaliadas quanto à variação da intensidade.

O segundo problema citado por Harris e Stephens (1988) é a sensibilidade a ruídos do operados de Moravec. Em (Moravec, 1980) a janela utilizada é binária e quadrada, não tendo nenhuma suavização do ruído, já em (Harris e Stephens, 1988) os autores sugerem o uso de uma janela circular gaussiana

$$G(u,v) = e^{\frac{-(u^2+v^2)}{2\sigma^2}}$$

para a suavização da imagem.

O terceiro problema citado por Harris e Stephens (1988) também refere-se à sensibilidade a ruídos e a quantidade excessiva de falsos positivos. O detector de Moravec utiliza a menor resposta de $E_{u,v}(x,y)$ entre todas as direções para decidir se o ponto avaliado é ou não um canto. O problema dessa abordagem é que ruídos podem acarretar em descontinuidades das bordas, onde o valor mínimo passa a ser grande e caracteriza um canto. Para resolver ess problema o detector de Harris utiliza a variação entre os valores de $E_{u,v}(x,y)$ conforme muda-se a direção (u,v).

Harris e Stephens (1988) reescrevem então a Equação 5.4 na seguinte forma matricial

$$E_{u,v}(x,y) = \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix} \tag{5.6}$$

onde

$$M = \begin{bmatrix} A(x,y) & C(x,y) \\ C(x,y) & B(x,y) \end{bmatrix} \tag{5.7}$$

é a matriz do segundo momento da imagem. Os autovetores e autovalores de M resumem a distribuição do gradiente da imagem.

Os autovalores α e β da matriz M são proporcionais à curvatura principal da função de autocorrelação da imagem em torno do ponto (x, y). Essa proporção permite as seguintes observações:

- Se $\alpha = \beta \approx 0$, então a região em torno do ponto analisado apresenta intensidade constante.
- Se $\alpha > 0, \beta = 0$, então a região em torno do ponto analisado é uma borda.
- Se $\alpha = \beta > \tau > 0$, onde τ é um valor de limiar, então o ponto analisado é um canto.

Harris e Stephens (1988) definem, além da classificação de regiões em cantos e bordas, uma medida para a qualidade do canto ou borda. Utilizando as observações anteriores é definida a seguinte medida

$$R = \alpha\beta - k(\alpha + \beta)^2. \tag{5.8}$$

A constante k é utilizada para controlar a sensibilidade do detector. Note que os valores $\alpha\beta$ e $\alpha + \beta$ podem ser encontrados facilmente utilizando as propriedades dos autovalores

$$\begin{aligned} det(M) &= \alpha\beta \\ tr(M) &= \alpha + \beta \end{aligned} \tag{5.9}$$

Substituindo o determinante e o traço da matriz M na Equação 5.8

$$R = A(x,y)B(x,y) - C^2(x,y) - k(A(x,y) + B(x,y))^2. \tag{5.10}$$

O valor k serve como um ajuste da sensibilidade do detector. Valores maiores de k aumentam a sensibilidade do detector à cantos e bordas, mas ao mesmo tempo aumentam a suscetibilidade à detecção de ruídos.

Shi e Tomasi (1994) mostraram mais tarde que bons resultados são obtidos da observação $min(\alpha, \beta) > \tau$, ou seja, um ponto de interesse é um canto desde que o menor dos autovalores de M seja maior que um limiar τ. O detector de Harris utilizando o critério apontado por Shi e Tomasi (1994) é conhecido como detector de detector de Shi-Tomasi e apresenta resultados melhores que os obtidos por Harris e Stephens (1988) com essa simples modificação.

Os detectores de cantos de Harris e de Shi-Tomasi são eficientes em relação à técnicas mais robustas como a diferença de guassianas, porém apresentam problemas. Esses detectores não são invariantes à rotação e à escala, mas apresentam alta repetibilidade das características. Além da alta repetibilidade as características são encontradas em região que aumentam a diferenciabilidade, facilitando a correspondência da característica entre imagens.

5.1.2 Detector SIFT

A diferença de gaussianas (daqui em diante será referenciada como DoG, do inglês *Difference of Gaussians*) foi utilizada por Lowe (1999, 2004) para identificar características invariantes à escala e rotação e parcialmente invariantes à variação na luminosidade e tranformações afins. O método desenvolvido por Lowe (1999, 2004) é chamado SIFT (do inglês *Scale Invariant Feature Tranform*) e é composto por um detector e um descritor de caracaterísticas. O método SIFT é composto por um detector de características baseado em DoG e um descritor. Nesta Seção será apresentado o detector de característica, enquanto na próximo Seção será abordado o descritor do método SIFT.

Baseado nos trabalhos de Koenderink (1984) e Lindeberg (1994), Lowe (1999) utiliza a função gaussiana sobre uma imagem com escalas diferentes para gerar o espaço de escalas da imagem.

A função

$$L(x, y, \sigma) = G(x, y, \sigma) * I_{x,y} \tag{5.11}$$

define o espaço de escalas da imagem $I_{x,y}$. Na Equação 5.11, o operador $*$ define a convolução da função gaussiana

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

A aplicação de uma gaussiana sobre a imagem causa efeito de desfoque que deixa a imagem com aspecto embaçado ou borrada.

Como dito anteriormente o SIFT utiliza a DoG para obtenção de pontos de interesse no espaço de escalas. A DoG é uma aproximação para o laplaciano da gaussiana $G(x, y, \sigma)$ com a escala normalizada, necessário para a invariância da escala segundo Lindeberg (1994). A função da DoG é dada por

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I_{x,y} \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{5.12}$$

O algoritmo descrito por Lowe (1999, 2004) inicia pela obtenção do espaço de escalas. O espaço de escalas é dividido em oitavas. Cada oitava é composta pelas imagens borradas utilizando o filtro gaussiano. A escala aplicada às imagens é crescente.

$$\begin{aligned} I_0 &= G(x, y, \sigma) * I_{x,y} \\ I_i &= G(x, y, \sigma) * I_{i-1} \end{aligned}. \tag{5.13}$$

A Equação 5.13 representa o crescimento da escala dentro da oitava. Uma vez calculadas as escalas da oitava, a imagem cujo valor de σ é duas vezes o valor do σ inicial na oitava é selecionada para ser reamostrada. A reamostragem seleciona um *pixel* a cada duas linhas ou colunas e gera uma imagem com metade do tamanho da amostra original que será utilizada para gerar uma nova oitava. Para evitar perda de informação a imagem original é redimensionada para o dobro do seu tamanho utilizando interpolação linear e a imagem redimensionada é utilizada para gerar a primeira oitava.

Uma vez calculadas as escalas de uma oitava, as DoGs $D(x, y, \sigma)$ são estimadas para valores de escala adjacentes. A Figura 5.3 apresenta o esquema para se encontrar o espaço de escalas.

Calculadas as DoGs, deseja-se obter os máximos e mínimos locais em todas as escalas de $D(x, y, \sigma)$. Para isso avalia-se cada ponto da primeira oitava com seus 26 vizinhos, 8 no mesmo nível e 9 nos níveis superiores e inferiores de escala. A Figura 5.4 apresenta essa comparação. Se o ponto for extremo nessa oitava, sua posição é estimada na oitava seguinte e o processo é repetido para esse ponto. As informações referentes à escala e oitava alcançadas são guardadas e farão parte da característica.

A implementação apresentada por Lowe (2004) difere da sua primeira implementação (Lowe, 1999). Em seu trabalho mais recente, uma vez obtidos os pontos de extremo, Lowe (2004) refina a posição do ponto de extremo para obter uma posição em *subpixels*. É utilizada uma expansão de Taylor em torno do ponto de extremo

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x} \tag{5.14}$$

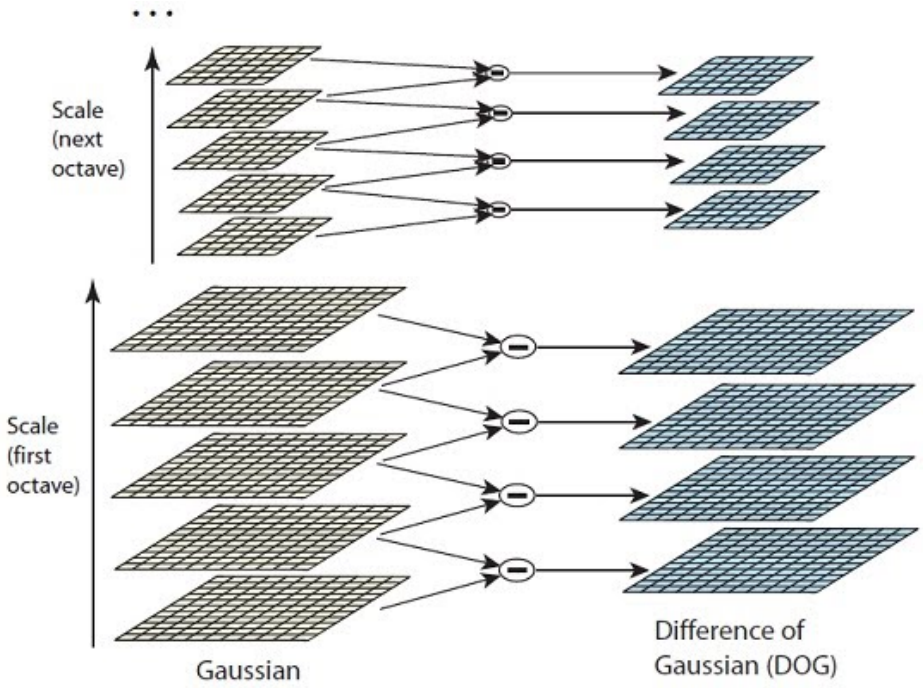


Figura 5.3: A pirâmide de gaussianas da figura apresenta dois níveis, ou seja, apresenta duas oitavas. Do lado esquerdo estão as imagens borradas com o filtro gaussiano e no lado direito estão as DoGs. Fonte:Lowe (2004)

onde $\mathbf{x}$ é o deslocamento em torno do ponto de extremo. O ponto extremo $\hat{\mathbf{x}}$ é determinado derivando a Equação 5.14 em relação a $\mathbf{x}$ e igualando a zero. O resultado é dado por

$$\hat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial^2 \mathbf{x}^2}\frac{\partial D}{\partial \mathbf{x}}. \tag{5.15}$$

Se houver variação maior que 0.5 em qualquer uma das direções o ponto muda de *pixel* e o valor de extremo é interpolado no novo ponto. Lowe (2004) também propõe utilizar o valor do ponto de extremo refinado para remover extremos com baixo contraste. Essa operação é feita substituindo a Equação 5.15 na Equação 5.14, resultando em

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2}\frac{\partial D^T}{\partial \mathbf{x}}\hat{\mathbf{x}}.$$

Outro filtro proposto por Lowe (2004) usa uma ideia semelhante a de Harris e Stephens (1988) para remover pontos de borda. A matriz hessiana

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

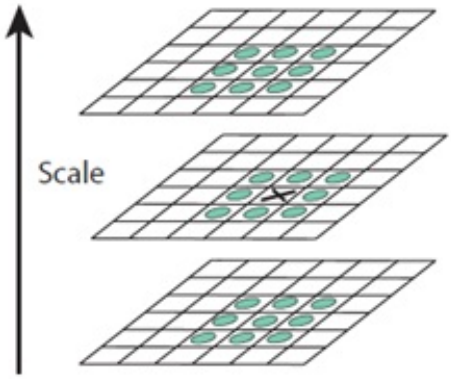


Figura 5.4: O *pixel* maracado com um x é avaliado em relação aos seus vizinhos no mesmo nível e nos níveis superior e inferior. Fonte:Lowe (2004)

descreve a curvatura principal da imagem em torno do ponto de extremo. Analisando a relação entre os autovalores de H Lowe (2004) avalia que se o determinate de H é negativo o ponto é descartado. Assim como Harris e Stephens (1988) uma avaliação é feita sobre a relação dos autovalores, como mostra a Equação a seguir

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r},$$

onde $\alpha = r\beta$. Um valor extremo será descartado caso a relação

$$\frac{Tr(H)^2}{Det(H)} < \frac{(\tau+1)^2}{\tau},$$

para um limiar τ a ser definido. Em seu trabalho, Lowe (2004) define $\tau = 10$.

Até este ponto foram obtidas localização e escala dos pontos de extremos e foram removidos extremos com baixo contraste e em bordas. Resta obter uma orientação para o ponto. Para isso escolhe-se em cada oitava a imagem gaussiana L cuja escala mais se aproxima da escala do ponto de extremo. Para cada uma das imagens $L(x, y)$ (note que o valor σ não aparece pois é definido como o mesmo do ponto de extremo) a magnitude $m(x, y)$ e a orientação $\theta(x, y)$ são calculadas usando

$$\begin{aligned} m(x,y) &= \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \\ \theta(x,y) &= tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))) \end{aligned}.$$

Calcula-se as orientações em torno do ponto de extremo e um histograma é montado com 36 orientações possíveis para a característica, cada um respondendo por 10° dos 360°

possíveis para $\theta(x, y)$. Cada uma das orientações $\theta(x, y)$ é pesada utilizando a magnitude $m(x, y)$ e uma janela gaussiana com σ sendo 1.5 vezes a escala do ponto de extremo antes de ser adicionada à orientação da característica.

Picos no histograma das orientações em torno do ponto de extremo correspondem a direções dominantes do gradiente local. O maior pico do histograma é identificado e uma característica é gerada com localização, escala e orientação. Picos no histograma que sejam máximos locais e cuja magnitude seja maior que 80% da do pico máximo também geram características com a mesma localização e escala, mas com a orientação diferente. Por fim, para cada pico que gerou uma característica, uma parábola é traçada pelo pico e os dois valores do histograma adjacentes a ele. Para se obter maior precisão o pico máximo é então tomado como o máximo da parábola gerada pela interpolação dos três picos.

Nesta Subseção é apresentada a característica utilizada no método SIFT. A característica obtida é invariante à escala e à rotação, e robusto à varições na luminosidade e tranformações afins. O detector apresenta quantidade, repetibilidade e diferenciabilidade das características. Mas é mais caro computacionalmente que o detector de Harris. Existem detectores baseados no SIFT que apresentam melhor desempenho computacional e não tem grandes perdas na robustez.

Na próxima Seção serão apresentados descritores que podem ser utilizados com esta e outras características. O descritor proposto por Lowe (1999, 2004) é apresentado na Subseção 5.2.2.

5.2 Descritores

Como mencionado no início da Seção 5.1, espera-se que uma boa característica tenha como propriedade a diferenciabilidade. A diferenciabilidade permite que a característica seja identificada (idealmente) de maneira única. Porém se for utilizado somente o ponto de interesse onde localiza-se a característica, a diferenciação torna-se difícil, uma vez que existe pouca informação para gerar um identificador para aquele ponto.

Os descritores surgem como uma forma de gerar uma identificação para a característica. Eles utilizam informação do ponto onde se encontra a característica e de sua vizinhança para determinar uma assinatura para o local mais distinta possível. Idealmente os descritores devem ser únicos e invariantes a tranformações, mas a invariância para algumas transformações é mais difícil (se mesmo possível) que a invariância para outras.

5.2.1 Recorte

Um descritor baseado em recorte utiliza a região em torno da característica de maneira bruta para identificar a característica. O recorte é uma maneira simples e rápida de se identificar uma característica, mas não apresenta as propriedades desejadas de invariância a nenhuma tranformação.

Por ser um descritor tão simples, os método para identificação de uma característica em imagens diferentes normalmente não retornam um único correspondente, retornando correspondências com características diferentes. Além do problema das múltiplas correspondências, por não ser invariante a diversas transformações, variações na luminosidade, deslocamento da câmera, entre outras mudanças na imagem, acarretam em uma grande chance de falha em encontrar uma característica correspondente em imagens diferentes, mesmo que a característica esteja visível nas imagens.

Apesar de não ser um descritor robusto, a simplicidade e velocidade para gerar e se comparar o recorte tornam atrante seu uso em aplicações que necessitam de resposta rápida e podem abrir mão da invariância. Uma aplicação que utiliza imagens de um vídeo, onde há pouca variação entre imagens sequenciais, ou uma aplicação em que é possível estimar a posição da característica nas demais imagens podem fazer uso desse descritor sem grandes prejuízos no resultado da aplicação.

No sistema de odometria visual apresentado por Nistér et al. (2006) as características identificadas pelo detector de Harris são correlacionadas com as características da imagem seguinte. Um recorte da imagem em torno da característica é utilizado para a comparação com as características da imagem seguinte. Como o deslocamento entre duas imagens é pequeno Nistér et al. (2006) impõe uma distância máxima para a busca da característica na imagem seguinte. Para a comparação dos recortes foi utilizado o método de correlação cruzada normalizada.

Em (Kitt et al., 2011) é utilizado o descritor baseado em recorte, mas não usam para identificar uma característica pontual na imagem. Kitt et al. (2011) utilizam cantos como característica para estimar o movimento da câmera, mas em uma superfície uniforme como a rua a obteção de características se torna uma tarefa difícil. Em vez de buscar características na rua, Kitt et al. (2011) definem um recorte próximo à câmera, onde a imagem processada apresenta melhor resolução. Esse recorte além de próximo à câmera está dentro de uma região maior de interesse. O recorte é então comparado com a região de interesse da imagem anterior utilizando o método da soma da diferença absoluta.

A Figura 5.5a mostra em branco a região de interesse. A Figura 5.5b mostra em branco o recorte perto da câmera dentro da região de interesse. Observe que é aplicada

uma transformação na região de interesse da Figura 5.5a para se obter a região de interesse na Figura 5.5b.

(a) Região de interesse **(b)** Recorte

Figura 5.5: Exemplo onde o descritor baseado em recorte é utilizado. A Figura (b) apresenta o recorte utilizado na comparação. A região de interesse onde se localiza o recorte é apresentada na Figura (a). Fonte:Kitt et al. (2011)

5.2.2 Descritor SIFT

O descritor a seguir é parte do método apresentado na Subseção 5.1.2 e é utilizado junto com as características encontradas na mesma Subseção. O descritor é baseado em um modelo biológico de visão proposto por Edelman et al. (1997).

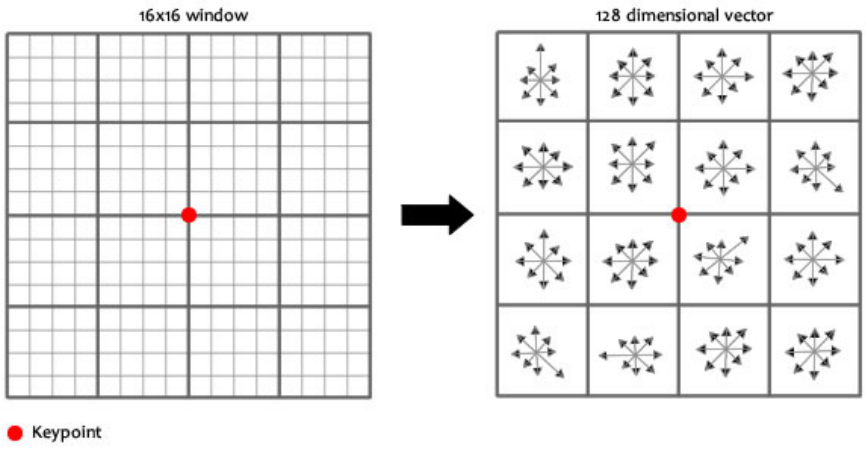


Figura 5.6: À esquerda a janela 16×16 composta por janelas 4×4. À direita o a janela 16×16 com as orientações geradas pelo histograma.

Para cada característica um descritor é gerado utilizando a imagem na mesma escala e na mesma oitava da característica. Uma janela de tamanho 16×16 *pixels* é definida com a característica no centro como mostrado na Figura 5.6. Para cada *pixel* da janela é calculada a orientação e magnitude como feito na Subseção 5.1.2. Subtraindo a orientação

da característica das demais orientações, alinha-se a janela à característica. A orientação da janela relativa à característica garante ao descritor a invariância a rotação.

A janela 16×16 é dividida por janelas menores de tamanho 4×4. A divisão é mostrada na Figura 5.6 à esquerda. Em cada uma das janelas 4×4 as orientações são separadas em um histograma. O histograma é separado em oito conjuntos, cada conjunto associado a um intervalo de 45°. Cada gradiente adicionado ao histograma é pesado em relação à sua magnitude, dando um peso maior aos gradientes com meior magnitude. O gradiente também recebe um peso em relação à sua distância da característica. Esse peso é dado por uma função gaussiana com centro na característica e com σ igual à metade da largura da janela, no caso igual a 8.

As oito orientações, uma para cada janela 4×4, dão origem a um vetor chamado *feature vector*. Para se ter maior robustez a luminosidade Lowe (2004) propõe a normalização do vetor, uma limiarização dos valores que forem maiores que 0.2 e uma nova normalização do vetor.

Capítulo

6

Odometria Visual Baseada em Reconstrução

Neste Capítulo são tratados os principais conceitos da geometria epipolar, essencial para o desenvolvimento de um sistema de odometria visual baseado em reconstrução 3D. Também são abordadas técnicas utilizadas para estimação dos elementos que compões a geometria de duas câmeras. São descritos também métodos para obtenção da odometria baseado na geometria epipolar.

A Seção 6.1 descreve brevemente da obtenção da odometria de um agente móvel baseado em uma câmera monocular. A Seção 6.2 descreve a relação existente entre duas câmera que dá origem à geometria epipolar. As Seções 6.3, 6.4 e 6.5 apresenta métodos necessários para a obtenção e validação das estruturas da geometria epipolar. A Seção 6.3 descreve o "Algoritmo dos Oito Pontos" que estima a matriz fundamental a partir de um conjunto de pontos correspondentes entre as imagens. A Seção 6.4 descreve como é possível obter a tranformação entre o posicionamento das câmeras. Na Seção 6.5 é apresentado um método para a reconstrução 3D de um ponto dada a matriz de projeção das câmeras. E por fim a Seção 6.6 descreve o método RANSAC utilizado como ferramenta de apoio aos métodos de estimação baseados em correspondências.

6.1 Formulação do Problema

Suponha um agente móvel vagando em um ambiente. Suponha também uma câmera monocular fixa a esse agente e que essa câmera captura imagens do ambiente em intervalos

de tempo discretos k. Como a câmera esta fixa ao agente, a odometria estimada para um dos corpos pode ser aproximada para a do outro corpo. Portanto encontrar a odometria do agente pode ser tratado como o problema de encontrar a odometria da câmera.

A partir de um par de imagens de uma mesma cena é possível recuperar a transformação de corpo rígido $g(\cdot) \in SE(3)$. A transformação $g(\cdot)$ é composta pela rotação $R \in SO(3)$ e pela translação $\mathbf{T} \in \mathbb{R}^3$ que move a câmera do posicionamento na primeira imagem para o posicionamento na segunda imagem. O posicionamento de uma câmera é a posição e orientação do sistema de coordenadas da câmera em relação a algum outro sistema de coordenadas.

Considere que a câmera tenha um posicionamento inicial conhecido. A transformação $g(\cdot)$ entre duas imagens é estimada a cada intante k. Cada nova transformação é utilizada para atualizar a odometria do agente. Então o problema pode ser resumido em obter a tranformação $g(\cdot)$ entre duas imagens e concatenar cada uma das tranformações para toda a sequência de imagens.

A geometria epipolar descreve a relação entre duas imagens e permite estimar a transformação que relaciona o posiconamento da câmera em cada imagem. Na Seção 6.2 será descrita a geometria epipolar e como recuperar a matriz de rotação R e o vetor de translação t que compões a tranformação de corpo rígido.

6.2 Geometria Epipolar

Considere duas câmeras calibradas, como apresentado no Capítulo 3, com origem em $\mathbf{C}$ e $\mathbf{C}'$. Tomando o centro de projeção $\mathbf{C}$ como referência, o posicionamento da segunda câmera é dado pela transformação de corpo rígido $g = (R, \mathbf{T}) \in SE(3)$, onde $R \in SO(3)$ é uma matriz de rotação e $\mathbf{T} \in \mathbb{R}^3$ é um vetor de translação. Se um ponto no mundo tem coordenadas $\mathbf{X} \in \mathbb{R}^3$ em relação à $\mathbf{C}$ e $\mathbf{X}' \in \mathbb{R}^3$ em relação à $\mathbf{C}'$, então a relação entre os pontos $\mathbf{X}$ e $\mathbf{X}'$ é dada por uma tranformação de corpo rígido da seguinte maneira

$$\mathbf{X}' = R\mathbf{X} + \mathbf{T} \tag{6.1}$$

O desenho das relações descritas anteriormente é apresentado na Figura 6.1.

Se os parâmetros instrínsecos das câmeras forem conhecidos, então $\mathbf{X}$ pode ser substituído pela projeção $\lambda\mathbf{x}$ em coordenadas homogêneas, onde $\mathbf{x} \in \mathbb{R}^3$ é a coordenada da projeção e $\lambda \in \mathbb{R}^+$ é um fator desconhecido de escala. Substituindo $\mathbf{X}$ e $\mathbf{X}'$ por suas projeções na Equação 6.1 tem-se:

$$\lambda'\mathbf{x}' = R\lambda\mathbf{x} + \mathbf{T} \tag{6.2}$$

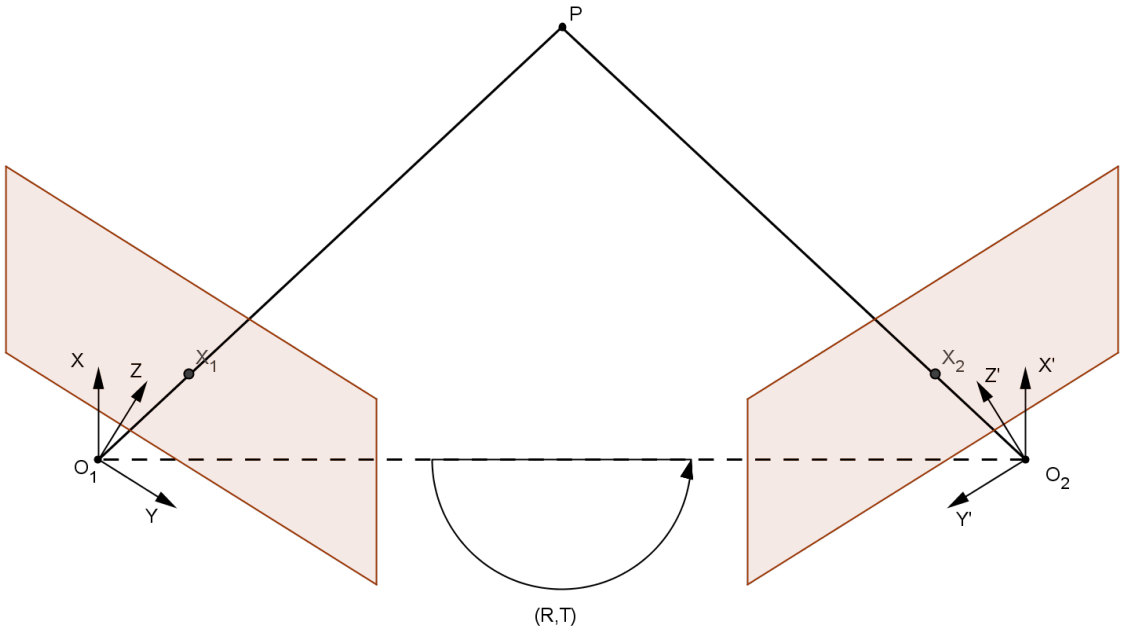


Figura 6.1: Representação da geometria epipolar. Duas câmera $\mathbf{C}$ e $\mathbf{C}'$ projetam o ponto $\mathbf{X}$ em seus planos de imagem. As projeções são respectivamente $\mathbf{x}$ e $\mathbf{x}'$.

Multiplicando os dois lados da Equação 6.2 por $\widehat{T}$ obtém-se a Equação 6.2 fica:

$$\lambda' \widehat{\mathbf{T}} \mathbf{x}' = \widehat{\mathbf{T}} R \lambda \mathbf{x}, \tag{6.3}$$

onde

$$\widehat{\mathbf{T}} = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$$

é uma matriz anti-simétrica e a multiplicação por $\widehat{\mathbf{T}}$ equivale ao produto cruzado do vetor $\mathbf{T}$ por outro vetor.

Pré-multiplicando novamente, desta vez por $\mathbf{x}'^T$ obtemos a seguinte Equação

$$\lambda' \mathbf{x}'^T \widehat{\mathbf{T}} \mathbf{x}' = \mathbf{x}'^T \widehat{\mathbf{T}} R \lambda \mathbf{x} \tag{6.4}$$

O termo $\mathbf{x}'^T \widehat{\mathbf{T}} \mathbf{x}'$ é um produto triplo e pode ser interpretado como o volume do paralelepípedo formado pelos três vetores que compõe o produto. Como todos os vetores do produto triplo $\mathbf{x}'^T \widehat{\mathbf{T}} \mathbf{x}'$ estão no mesmo plano o resultado do produto é 0. Utilizando esse resultado na Equação 6.4 obtém-se

$$\mathbf{x}'^T \widehat{\mathbf{T}} R \lambda \mathbf{x} = 0 \tag{6.5}$$

O valor λ é sempre maior que 0 para pontos na imagem. Dessa última informação obtemos a Equação 6.6 que é chamada de restrição epipolar.

$$\mathbf{x}'^{T}\widehat{\mathbf{T}}R\mathbf{x} = 0 \tag{6.6}$$

Na Equação 6.6 o produto cruzado define a *matriz essencial*

$$E = \widehat{\mathbf{T}}R \in \mathbb{R}^{3\times 3}.$$

A matriz essencial foi primeiramente apresentada por Longuet-Higgins (1981). Como visto na Equação 6.2, a matriz essencial são codificados a orientação e a posição relativa entre as câmeras da geometria epipolar. O Teorema seguinte, proposto por Ma et al. (2003), é uma versão mais forte do teorema proposto por Huang e Faugeras (1989) e define uma característica da matriz essencial importante para o desenvolvimento dos métodos para encontrar a matriz essencial e para recuperar a rotação e translação relativa.

Teorema 6.1. *Uma matriz não nula $E \in \mathbb{R}^{3\times 3}$ é uma matriz essencial se e somente se E tem uma decomposição em valores singulares da forma $E = U\Sigma V^T$, com:*

$$\Sigma = \begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

onde $\sigma \in \mathbb{R}_+$ e $U, V \in SO(3)$

O método de decomposição em valores singulares, ou SVD (do inglês *Singular Value Decomposition*), é descrito por Golub e Kahan (1965). O Teorema 6.1 diz que uma condição necessária e suficiente para uma matriz ser essencial é a de que a matriz possua dois valores singulares iguais e um valor singular nulo.

Nesta Seção é apresentada a derivação algébrica da restrição epipolar, também é possível obter a derivação geométrica. A derivação geométrica, assim como a derivação algébrica podem ser encontradas nos livros Hartley e Zisserman (2004) e Ma et al. (2003).

6.3 Algoritmo dos Oito Pontos

Como visto na Seção anterior, a matriz essencial codifica a posição e a orientação relativa entre as câmeras que compõe a geometria epipolar. Deseja-se recuperar essa informação separadamente, mas primeiro precisa-se obter a matriz essencial. Uma solução

linear para encontrar a matriz essencial é apresentada por Longuet-Higgins (1981) e é mais detalhadamente explicada em (Hartley e Zisserman, 2004) e (Ma et al., 2003).

Sejam $(\mathbf{x}_i, \mathbf{x'}_i)$ pares de caracaterísticas correspondentes nas câmera $\mathbf{C}$ e $\mathbf{C'}$ e seja $\mathbf{x}_i \otimes \mathbf{x'}_i$ o produto de Kronecker entre o par i. A matriz $M \in \mathbb{R}^{m\times 9}$ é formada da seguinte maneira:

$$M = \begin{bmatrix} \mathbf{x}_1 \otimes \mathbf{x'}_1 \\ \mathbf{x}_2 \otimes \mathbf{x'}_2 \\ \vdots \\ \mathbf{x}_m \otimes \mathbf{x'}_m \end{bmatrix}.$$

Agora, seja também

$$\mathbf{e} = [E_{11}, E_{12}, E_{13}, E_{21}, E_{22}, E_{23}, E_{31}, E_{32}, E_{33}]^T \in \mathbb{R}^9$$

o vetor formado pelos elementos da matriz essencial, onde E_{ij} é o elemento da linha i e coluna j. Observe que se pode reescrever a Equação 6.6 como:

$$M\mathbf{e} = 0. \tag{6.7}$$

Caso a matrix M tenha exatamente 8 graus de liberdade, então a solução é única. Já se a matriz M tiver 9 graus de liberdade, então pode-se encontrar o vetor solução $\mathbf{e}$ que minimiza o erro por mínimos quadrados. Utiliza-se então o método SVD (Golub e Kahan, 1965) para decompor a matriz M. A solução que minimiza o erro corresponde à coluna $\mathbf{v} \in \mathbb{R}^9$ de $V \in \mathbb{R}^{9\times 9}$, associado ao menor valor singular de M, com $\text{SVD}(M) = UDV^T$ e $D \in \mathbb{R}^{m\times 9}$ a matriz diagonal contendo os valores singulares de M.

A matriz $E' \in \mathbb{R}^{3\times 3}$ é formada pelos elementos do vetor $\mathbf{v}$

$$E' = \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix}$$

e é a solução da Equação 6.7. Mesmo solucionando a Equação 6.7, E' pode não satisfazer a restrição do Teorema 6.1 de que dois valores singulares devem ter o mesmo valor e um deve ser nulo. Ma et al. (2003) definem um teorema para forçar a condição do Teorema 6.1 projetando a matriz E' no espaço das matrizes essenciais.

Teorema 6.2. *Seja $E' \in \mathbb{R}^{3\times 3}$ uma matriz com decomposição em valores singulares da forma $SVD(E') = Udiag\{\lambda_1, \lambda_2, \lambda_3\}V^T$, onde $diag\{\lambda_1, \lambda_2, \lambda_3\}$ é a matriz diagonal $\mathbb{R}^{3\times 3}$, $U, V \in SO(3)$ e $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Então a matriz essencial E que minimiza o erro*

$||E - E'||_f^2$ é dado por $E = Udiag\{\sigma, \sigma, 0\}V^T$, com $\sigma = \frac{\lambda_1+\lambda_2}{2}$. O índice $_f$ no erro indica que é utilizada a norma de Frobenius.

O Teorema 6.2 encontra a matriz E dentro do espaço das matrizes essenciais que tem a menor distância para a matriz E'.

Obtida a matriz essencial, Longuet-Higgins (1981) ainda propõe a normalização da magnitude do vetor de translação para $\|\mathbf{T}\| = 1$. Utilizando o método de decomposição em valores singulares em $E' = U\Sigma V^T$, a normalização $\|\mathbf{T}\| = 1$ equivale à substituição da matriz $\Sigma = diag\{\sigma, \sigma, 0\}$ pela matriz $\Sigma' = diag\{1, 1, 0\}$.

O algoritmo dos oito pontos, como foi apresentado é muito sensível a ruídos e não é um algoritmo considerado aplicável em muitas aplicações reais. Porém Hartley (1997) propõe a normalização dos pontos antes da estimação da matriz essencial como maneira de diminuir o efeito do ruído na imagem, melhorando seu desempenho e permitindo o uso do mesmo em aplicações reais. A normalização sugerida por Hartley (1997), a mesma aplicada na homografia, é uma translação da imagem de maneira que o centróide das características esteja na origem e o valor quadrático médio das características para o centro seja igual a $\sqrt{2}$.

6.4 Matriz Essencial: Rotação e Translação

Na Seção 6.3 é apresentado o método dos oito pontos, utilizado para se estimar a matriz essencial. O objetivo de se estimar a matriz essencial é que nela estão codificados rotação e translação relativa entre as câmeras da geometria epipolar.

Como foi visto na Seção 6.2, uma matriz essencial pode ser decomposta como $E = U\Sigma V^T$, onde $U, V \in SO(3)$ e $\Sigma = diag\{\sigma, \sigma, 0\}$, é a matriz diagonal contendo os valores singulares de E, como mostrado no Teorema 6.1. Relembrando que $E = \widehat{\mathbf{T}}R$, observa-se que os valores singulares de $\widehat{\mathbf{T}}$ coincidem com os de E, uma vez que as duas matriz diferem apenas por uma rotação. Assim sendo, segundo o teorema espectral aplicado ao caso da matriz antissimétrica, pode-se decompor $\widehat{\mathbf{T}}$ em

$$\widehat{\mathbf{T}} = Q\Sigma W Q^T, \tag{6.8}$$

onde $Q \in \mathbb{R}^{3\times3}$ é uma matriz ortogonal e

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

é a matriz de rotação em torno do eixo-z pelo ângulo de 90°.

Substituindo a Equação 6.8 na Equação 6.2 obtém-se a seguinte Equação

$$E = Q\Sigma W Q^T R. \tag{6.9}$$

Comparando a Equação 6.9 com a decomposição $E = U\Sigma V^T$ observa-se que a matriz Q equivale à matriz U da decomposição. Também podemos encontrar a Equação para R

$$R = UW^T V^T, \tag{6.10}$$

além da Equação 6.8 agora com o valor de U no lugar de Q

$$\widehat{T} = U\Sigma W U^T. \tag{6.11}$$

A configuração $(R, \mathbf{T})$ formada pelas Equações 6.10 e 6.11 é apenas uma das quatro configurações possíveis à partir da matriz E. Ma et al. (2003) mostram que cada matriz essencial E gera o chamado *twiste pair* (aqui chamado de par reverso). Se uma configuração possível é dada por

$$\begin{aligned} R_1 &= UW^T V \\ \widehat{\mathbf{T}}_1 &= U\Sigma W U^T \end{aligned},$$

então o seu par reverso é dado por

$$\begin{aligned} R_2 &= UWV \\ \widehat{\mathbf{T}}_2 &= U\Sigma W^T U^T \end{aligned}.$$

Note que o par reverso é a inversão de $\mathbf{T}_1$ e a rotação de R_1 em torno da linha que liga os centros ópticos das câmeras (*baseline*). Na Figura 6.2 os pares de configurações (a),(d) e (b),(c) são pares reversos.

Tem-se agora dois pares de configurações possíveis entre as câmeras. As duas configurações restantes surgem do fato de que a matriz essencial, assim como a translação, é recuperada a menos de um fator escalar e um sinal. Considerando os pares gerados pela ambiguidade do sinal da matriz essencial e os encontrados anteriormente temos que as

possíveis configurações entre as câmeras são

$$\begin{aligned}
&(R_1 = UW^TV \ , \ \widehat{\mathbf{T}}_1 = U\Sigma WU^T \ ),\\
&(R_2 = UWV \ \ , \ \widehat{\mathbf{T}}_2 = U\Sigma W^TU^T),\\
&(R_3 = UW^TV \ , \ \widehat{\mathbf{T}}_3 = U\Sigma W^TU^T),\\
&(R_4 = UWV \ \ , \ \widehat{\mathbf{T}}_4 = U\Sigma WU^T \ ).
\end{aligned}$$

Note que o vetor $\mathbf{T}$ compõe o espaço nulo de $\widehat{\mathbf{T}}$, portanto $\widehat{\mathbf{T}}\mathbf{T} = \mathbf{0}$. Sendo assim

$$\widehat{\mathbf{T}}\mathbf{T} = U\Sigma WU^T\mathbf{T} = \mathbf{0}. \tag{6.12}$$

Da Equação 6.12 obtém-se que $\mathbf{T} = U(0,0,1)^T = \mathbf{u_3}$, onde $\mathbf{u_3}$ é o vetor correspondendo à terceira coluna da matriz U, assim a translação pode ser obtida sem a necessidade das multiplicações de matrizes. As possíveis configurações podem ser reescritas comos

$$\begin{aligned}
&(R_1 = UW^TV \ , \ \mathbf{T}_1 = \ \ u_3),\\
&(R_2 = UWV \ \ , \ \mathbf{T}_2 = -u_3),\\
&(R_3 = UW^TV \ , \ \mathbf{T}_3 = -u_3),\\
&(R_4 = UWV \ \ , \ \mathbf{T}_4 = \ \ u_3).
\end{aligned}$$

A Figura 6.2 mostra todas as possibilidades de configurações para $(R, \mathbf{T})$

A escolha da configuração correta é feita obtendo a strutura 3D de um ponto da imagem. Os pontos reconstruídos com as configurações incorretas terão profundidade negativa em relação a uma ou as duas câmeras. A configuração (a) na Figura 6.2 corresponde à configuração correta. O método de recontrução do ponto 3D será apresentado na Seção 6.5.

Seja $\mathbf{X} = [X, Y, Z, W]^T$ um ponto 3D, $\mathbf{x} = \lambda[x, y, 1]^T$ a projeção do ponto e $P = [M|\mathbf{p_4}]$ a matriz de projeção tal que $\lambda\mathbf{x} = P\mathbf{X}$, então a profundidade do ponto $\mathbf{X}$ em relação à matriz de projeção P é dada por

$$depth(\mathbf{X}; P) = \frac{sign(det(M))\lambda}{W\|\mathbf{m^3}\|},$$

onde $\mathbf{m_3}$ é a terceira linha da matriz $M = KR$.

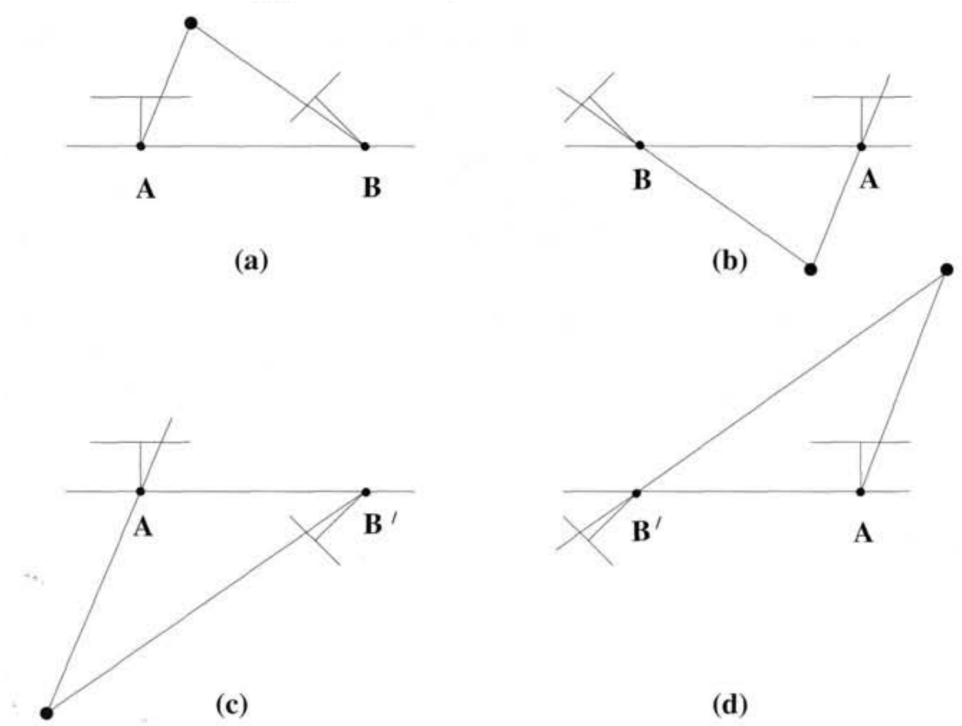


Figura 6.2: Todas as posíveis configurações $(R, \mathbf{T})$ dada uma matriz essencial E.

6.5 Recontrução 3D de um Ponto

Se conhecida a posição relativa entre duas imagens, é possível estimar a estrutura 3D de um par de característica correpondentes. O processo de reconstrução 3D de um ponto também é conhecido como triangulação.

Hartley e Zisserman (2004) abordam de maneira semelhante ao utilizado para encontrar a matriz de homografia, a triangulação linear do ponto. Na homografia a relação entre um conjunto de pontos e suas projeções davam origem a um sistema linear que solucionado resulta em uma homografia entre o plano dos pontos e a imagem. Na triangulação, porém, a projeção é conhecida, então relacionam-se as matrizes de projeção de duas câmeras e as projeções de um mesmo ponto para gerar um sistema linear que solucionado resulta na estutura 3D relativa do ponto.

Sendo então

$$\begin{aligned} \mathbf{x} &= P\mathbf{X} \\ \mathbf{x}' &= P'\mathbf{X} \end{aligned} \tag{6.13}$$

as projeções de um mesmo ponto X em imagens diferentes. As matrizes P e P' são as matrizes de projeção referentes a primeira e a segunda câmera respectivamente. Assume-se que P é a origem do sistema, portanto $P = K[I|\mathbf{0}]$ e $P' = K[R|\mathbf{t}]$.

As Equações 6.13 são homogêneas e portanto $\lambda\mathbf{x} = P\mathbf{X}$. Removendo tem-se que $\mathbf{x} \times (P\mathbf{X}) = \mathbf{0}$. O produto cruzado gera a seguinte Equação

$$\begin{bmatrix} x(\mathbf{p}_3^T\mathbf{X}) - (\mathbf{p}_1^T\mathbf{X}) \\ y(\mathbf{p}_3^T\mathbf{X}) - (\mathbf{p}_2^T\mathbf{X}) \\ x(\mathbf{p}_2^T\mathbf{X}) - y(\mathbf{p}_1^T\mathbf{X}) \end{bmatrix} = \mathbf{0}, \tag{6.14}$$

onde $\mathbf{p^i}$ é a linha da i da matriz e projeção P. Note que a última linha na Equação 6.15 é uma combinação linear das linhas anteriores. Então as Equações 6.13 podem ser combinadas como

$$A\mathbf{X} = \begin{bmatrix} x(\mathbf{p^{3}}^T) - \mathbf{p^{1}}^T \\ y(\mathbf{p^{3}}^T) - \mathbf{p^{2}}^T \\ x'(\mathbf{p'^{3}}^T) - \mathbf{p'^{1}}^T \\ y'(\mathbf{p'^{3}}^T) - \mathbf{p'^{2}}^T \end{bmatrix} \mathbf{X} = \mathbf{0}, \tag{6.15}$$

gerando um sistema de quatro Equações para as quatro variáveis desconhecidas de $\mathbf{X}$.

O método para solucionar o problema é semelhante ao do problema da homografia. Assim como no caso da Seção 3.5 a solução para o sistema linear pode ser encontrada utilizando o método SVD (Golub e Kahan, 1965). O vetor singular de tamanho unitário associado ao menor valor singular de A é a solução do problema que minimiza o erro por mínimos quadrados.

6.6 Remoção de Outliers

Durante o processo de identificação de características correspondentes, pode ocorrer casamentos errados entre as características de duas imagens. Também podem ser identificadas na imagem características que se movimentem independetes da câmera dando origem correspondências que não foram geradas pelo movimento da câmera. Este caso é muito comum em ambientes urbanos onde existem pedestres e carros se movimentando livremente pela via e calçada. Estas correspondências inválidas, ou *outliers*, influenciam a estimação do movimento do agente e podem levar à estimação incorreta da sua posição se não tratadas. Uma das abordagens mais comuns para tratar esse tipo de problema são métodos baseados no paradigma RANSAC.

Em 1981, Fischler e Bolles (1981) propuseram o paradigma RANSAC (Random Sample Consensus) para ajustar um modelo à um conjunto de dados contendo uma grande

quantidade de dados inválidos. O paradigma RANSAC encontra os valores dos parâmetros livres de um modelo, ajustando-o ao conjunto de dados e identifica quais elementos desse conjunto desviam do padrão. A figura 6.3 apresenta um exemplo de aplicação de um método baseado em RANSAC. O segmento de reta vermelho é a aproximação linear dos pontos, o segmento de reta azul é a estimação da reta usando o paradigma RANSAC e o segmento de reta púrpura é o modelo correto. Os pontos pretos na imagem apresentam um desvio grande em relação ao padrão, causando um desvio grande se utilizados no cálculo da aproximação linear. A imagem foi gerada à partir de um código fornecido por Jones et al. (2001).

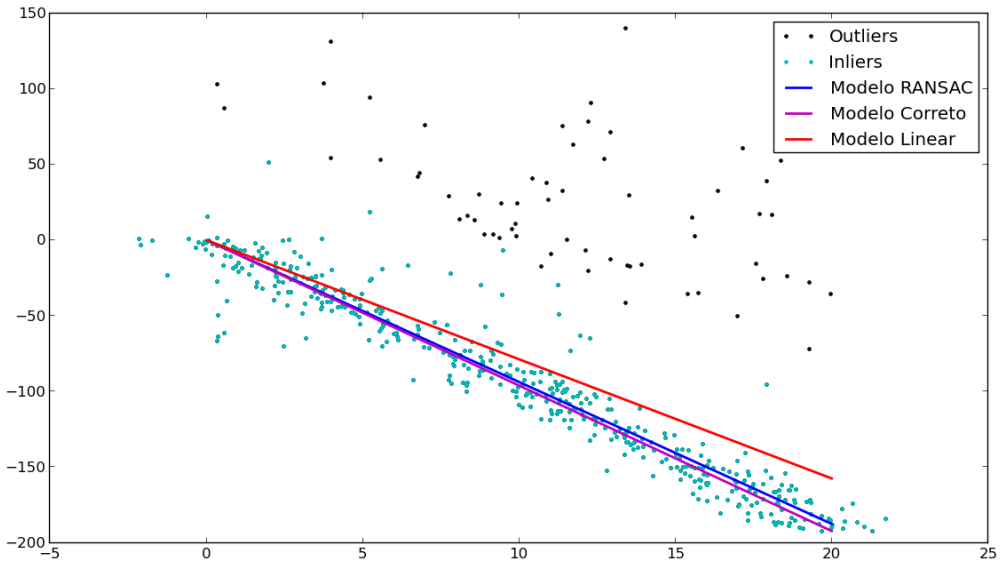


Figura 6.3: Exemplo do uso de um método baseado em RANSAC para ajuste de um modelo à dados que tenham valores com grande desvio. No caso foram gerados dados ruidoso usando como base um modelo linear. O segmento de reta púrpura representa o modelo original, o segmento de reta azul representa o modelo encontrado utilizando um método RANSAC e o segmento de reta vermelho é a aproximação linear do dados. Os pontos pretos são elementos do conjunto que apresentam um desvio muito grande do padrão.

Os métodos baseados em RANSAC apresentam duas etapas principais: geração de hipótese e teste de hipótese. Na etapa de geração de hipótese é escolhida uma amostra aleatória entre o conjunto de dados com a quantidade de informação mínima para determinar os parâmetros do modelo. A partir da amostra é gerada uma hipótese que será usada na etapa de teste de hipótese. Na etapa de teste de hipótese, a hipótese geradas pela amostras é testada verificando quão bem o modelo gerado com a hipótese se ajusta

ao conjunto de dados. Essas duas etapas são repetidas até a obtenção de uma hipótese satisfatória. Essa hipótese servirá como referência para a remoção de *outliers*, elementos cujo desvio em relação padrão da hipótese sejam maiores que um limiar τ.

O teste de hipótese neste caso consiste em encontrar a matriz fundamental que apresente o menor erro de projeção total. O erro de projeção total é a soma das distâncias euclidianas, ou erros de projeção, entre as projeções das características da imagem I_{k-1} às suas correpondências na imagem I_k. As etapas do RANSAC para odometria visual são mostradas na figura 6.4.

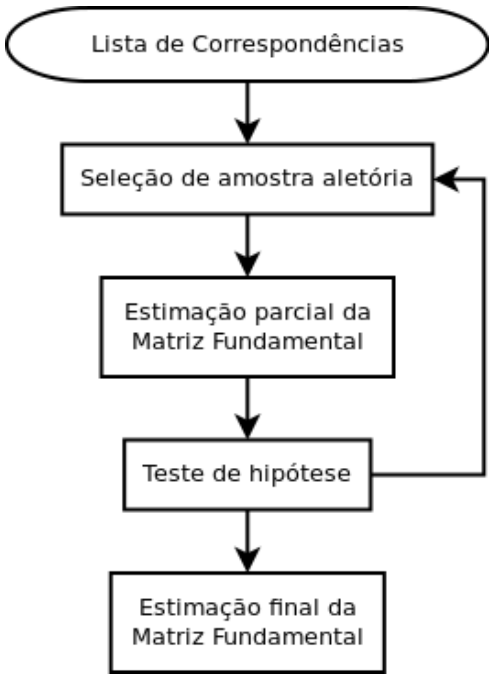


Figura 6.4: Etapas do método baseado em RANSAC aplicado à odometria visual.

O algoritmo realiza o processo de geração e teste de hipótese por um número n máximo de repetições. Se não for encontrada uma hipótese satisfatória dentro de n, a melhor hipótese é utilizada como matriz fundamental final. Para uma hipótese ser considerada satisfatória ela deve conter um número mínimo de *inliers*. Um par de caracaterísticas correpondentes é um *inlier* se seu erro de projeção for menor que um limiar τ, caso contrário o par é um *oulier*.

Capítulo

7

Odometria Visual com Estimação de Escala

Nos Capítulos anteriores foram apresentados dois métodos utilizados para se obter a odometria de um veiculo através de umas sequência de imagens. O primeiro método parte do pressuposto que a região em que o veículo se desloca é localmente plana, permitino descrever o deslocamento do veículo como uma tranformação planar no intervalo de duas imagens. A tranformação planar permite o cálculo da matriz de homografia, utilizada para o método de alinhamento das imagens. O segundo método estima a relação geométrica entre pontos de imagens diferentes e consequentemente a relação geométrica entre o posicionamento relativo da câmera nas imagem.

7.1 Base de Dados Utilizada

Foi utilizada para os experimentos a base de dados disponibilizada no KITTI Benchmark Suite Geiger et al. (2012). O conjunto de dados é composto por vinte e duas sequência, totalizando $39,2$ km de ruas e rodovias registradas pelos sensores. O conjunto de sensores é composto por duas câmeras coloridas, duas câmeras em tons de cinza, um sensor laser de múltiplos feixes e uma unidade inercial com GPS integrado e correção RTK. Neste trabalho foram utilizadas somente as câmeras em tons de cinza e as medidas tomadas do GPS, fornecidas já como poses pelo próprio KITTI Benchmark Suite. Além dos dados fornecidos pelo grupo, também foi utilizado a ferramenta utilizada para geração de erro e desenho das trajetórias.

Foram utilizadas 11 das 22 sequências da base de dados. As 11 sequências escolhidas contém informação do GPS de alta precisão que foi utilizado como trajeto real realizado pelo veículo. Em todas as sequências considerou-se a câmera a uma altura de 165 cm. A orientação das câmeras é fornecida junto com a calibração das câmeras com a própria base de dados. As sequências são gravadas em ruas e rodovias contendo elementos comuns a esses ambientes (e.g. pedestres e carros).

7.2 Configuração dos Experimentos

O problema de estimar o deslocamento aparente de uma câmera entre duas imagens pode ser modelado como um problema de alinhamento, dada uma região planar conhecida na imagem. No caso específico de um veículo, o plano da via pode ser utilizado para atender essa condição. Neste trabalho assume-se que uma porção da via é vísivel à frente do veículo, como visto na Figura 7.1. Essa condição é válida para a maioria dos casos em que o veículo está em movimento.

Figura 7.1: Região de interesse que supõe-se parte da via.

Durante o desenvolvimento do projeto, uma ferramenta foi implementada para estimação do movimento do veículo utilizando o método dos oito pontos descrito na Seção 6.3. Porém, por razões de eficiência e por fornecer uma maior robustez utilizando de métodos de apoio, foi utilizada a implementação do método monocular de Kitt et al. (2011), fornecida pela biblioteca Libviso2 (Geiger, 2015). O método utiliza o detector de cantos de Harris (Harris e Stephens, 1988) em conjunto com um descritor próprio (Geiger et al., 2012).

O método resultante deste projeto integra os dois métodos de forma simples. Primeiro estima-se o deslocamento do veículo utilizando o método direto ESM como descrito na Seção 4.2. Em caso de sucesso de convergência do método a escala é aplicada na odometria

utilizando o método dos oito pontos. Para facilitar a referêcia, daqui em diante o método será denominado **ESM-8p**.

A escala é inicializada com valor 1 e em caso de falha de convergência na etapa do método ESM, a escala anteriormente calculada é reutilizada. Essa hipótese é válida para falhas exporádicas, mas causa um erro grande para longas sequências com falha de convergência. É ideal que exista uma aproximação inicial da tranformação para uma melhor convergência do método ESM. Reutiliza-se então a transformação estimada em uma iteração para alimentar a próxima. Essa suposição é suportada pelo fato da variação de velocidade entre dois frames ser relativamente pequena. Porém este método apresenta a desvantagem de estimativas erradas atrapalharem o desempenho das iterações seguintes.

7.3 Resultado dos Experimentos

Os testes foram realizados aplicando o método desenvolvido nas sequências da base de dados e armazenando as transformações estimadas são armazenadas em um arquivo. Os dados foram analisados com relação à precisão e ao tempo de execução. Também foram analisados alguns dos problemas presentes no método **ESM-8p**.

A Figura 7.2 é a representação dos valores estimados pelo método **ESM-8p** (azul pontilhado) e dos valores do GPS de alta precisão (vermelho contínuo). A geração dos trajetos é feita projetando os pontos estimados no plano correspondente ao plano da via na inicialização do sistema, com o sistema inicializando na origem.

Nas Figuras 7.3 e 7.4 são apresentados respectivamente os percentuais de erro de translação e rotação pela distância percorrida para as 11 sequências avaliadas. Essa é uma representação média, do quanto o algoritmo desvia da posição real. Através da Figura 7.3 observa-se que em deslocamentos de 50m o método desvia-se aproximadamente 7m em média e para deslocamentos de 800m o desvio é de aproximadamente 80m. Assim como no erro da translação o erro da rotação na Figura 7.4 é maior para trechos curtos, mas diminui com o aumento da distância. Os erros do método são relativamente baixos dado a proporção do deslocamento, mesmo não sendo viável estimar a localização apenas através o método **ESM-8p**, é comum o uso de um conjunto de métodos e ferramentas com precisão menor para se obter localização confiável para navegação.

Os testes de performance foram executados em uma máquina Intel Core i5 3.55 GHz. Para o teste de performance, o método foi executado 10 vezes para cada sequência e o tempo foi tomado para execução da iteração completa, do tempo de execução do método ESM e do método dos oito pontos. O melhor e o pior resultado de cada sequência é descartado e é tirada a média dos valores restantes. Na Figura 7.5 estão representados

os tempos médios das iterações do método para as 11 sequências, assim como o tempo médio das etapas dos métodos diretos e por características.

Analisando-se as curvas apresentadas no gráfico da Figura 7.5 nota-se que o tempo de execução do método segue o padrão do método dos oito pontos e também que o algoritmo dos oito pontos consome a maior parte do tempo da iteração. Por outro lado o método ESM ocupa uma parcela pequena do tempo de execução (em torno de 10%). Algumas opções baseadas nessa observações para melhorar o tempo de execução serão discutidas no Capítulo 8.

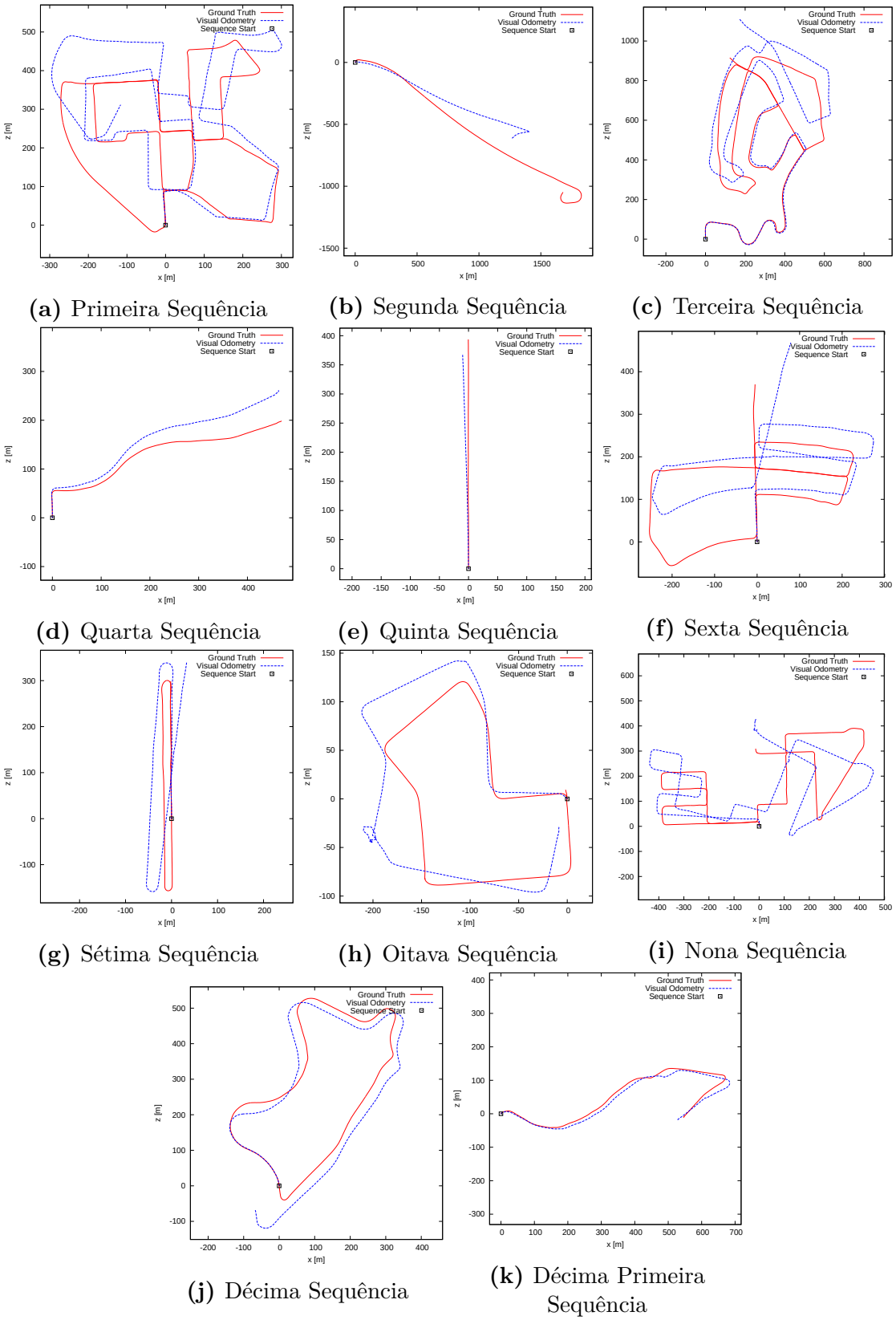
Devido à região visível limitada, os métodos diretos sofrem com o chamado problema da abertura. O problema da abertura ocorre em situações que movimentos diferentes podem dar origem ao mesmo padrão de imagem. A Figura 7.6 é um trecho da primeira sequência onde tal situação ocorre. Note que a diferença das imagens é pequena, o que induz o método a acreditar erroneamente que não ocorreu movimento. Essa situação tende a ter piores resultados que o caso de não convergência, pois o método acredita que o valor encontrado é o correto. A Figura 7.7 apresenta o gráfico do trecho final da primeira sequência, onde o problema da abertura ocorre durante grande parte do trecho de curva. O padrão rua no centro e meio fio no topo direito, se repete diversas vezes.

Outro problema enfrentado pelos métodos diretos é a variação de luminosidade que ocorre frequentemente pela reflexão de luz na via ou pelo ajuste da câmera ao passar por uma região mais escura e depois retorna a uma região bem iluminada. A Figura 7.8 apresenta uma situação onde o reflexo da pista acaba escondendo toda a textura da pista. O gráfico do erro do trecho do qual a Figura 7.8 faz parte é apresentado na Figura 7.9.

A Figura 7.9 traz também outro dado interessante. Após uma região de alta luminosidade, o veículo faz uma curva a direita e a região 2 destaca o erro gerado nesta desta curva. O erro é causado principalmente pelo posicionamento de veículos, muros e outros elementos da cena que se elevam do nível da via. Esse caso é bem comum principalmente em cenários urbanos, onde a via fica próxima a muros de casas e onde a câmera registra durante alguns quadros os veículos estacionados junto às calçadas.

Um último problema comum aos métodos de odometria visual em geral e também presente no método **ESM-8p** é a pertubação na estimação causada por elementos dinâmicos na cena. Tal problema é descrito na Seção 6.6 e o método RANSAC é apresentado como uma forma de amenizar esse problema. Nos testes o método dos oito pontos faz uso do RANSAC para diminuir o efeito dos chamados *outliers*, mas em situações onde corpos com movimento independente ocupam uma região grande da imagem o método RANSAC pode não conseguir evitar essa pertubação. Além do problema do método dos oito pontos o método ESM também é influenciado por elementos com movimento independente.

A Figura 7.10 representa os dois problemas. No caso do algoritmo dos oito pontos o caminhão ocupa a maior parte da imagem o que tende a resultar na maioria dos pontos utilizados no método a ter o mesmo movimento do caminhão. Além do problema causado ao método dos oito pontos, o método ESM também é influenciado. Note que a região utilizada para o alihamento é em grande parte a imagem do caminhão que além de representar um fluxo óptico que não o da câmera, ainda é um elemento perpendicular ao plano esperado (plano da via).



(a) Primeira Sequência **(b)** Segunda Sequência **(c)** Terceira Sequência

(d) Quarta Sequência **(e)** Quinta Sequência **(f)** Sexta Sequência

(g) Sétima Sequência **(h)** Oitava Sequência **(i)** Nona Sequência

(j) Décima Sequência **(k)** Décima Primeira Sequência

Figura 7.2: Odometria obtida utilizando o método direto ESM para estimar o deslocamento entre pares de imagens. Em vermelho o caminho real e em azul tracejado o caminho estimado pelo método de odometria.

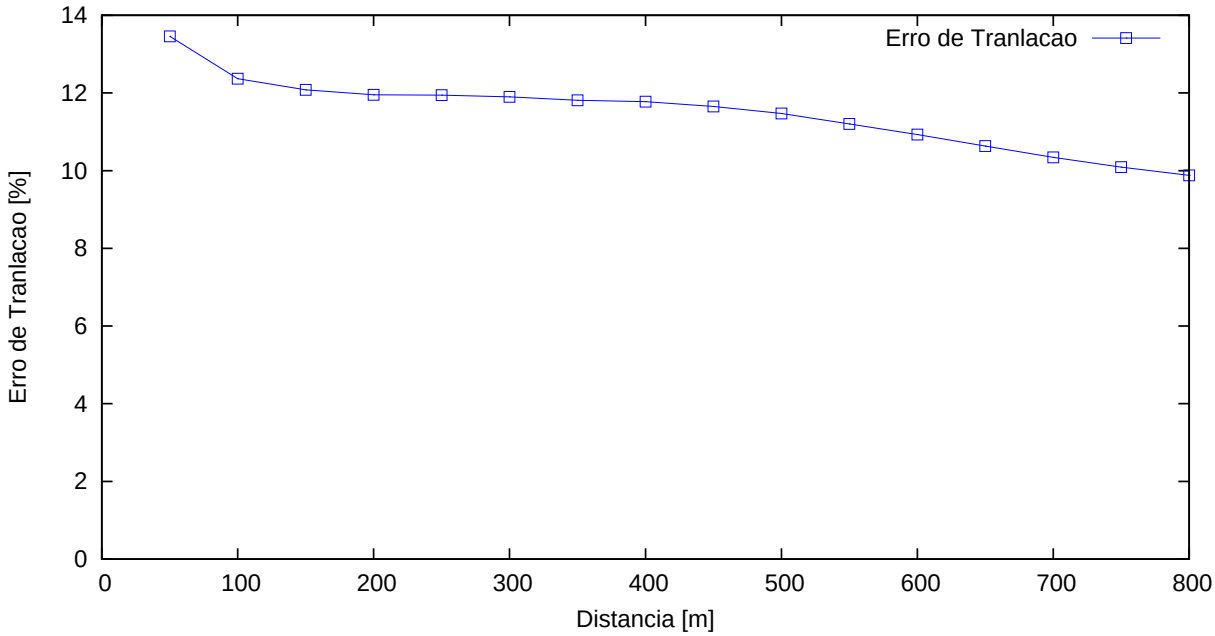


Figura 7.3: Percentual médio dos erros de translação pela distância percorrida

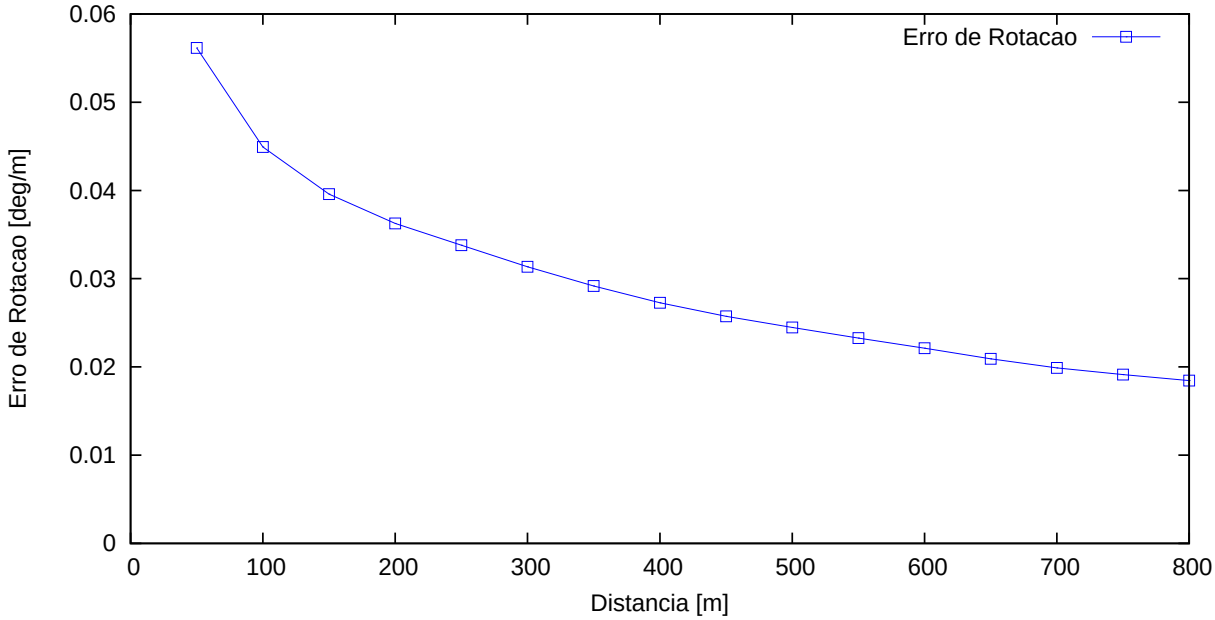


Figura 7.4: Percentual médio erros de rotação pela distância percorrida

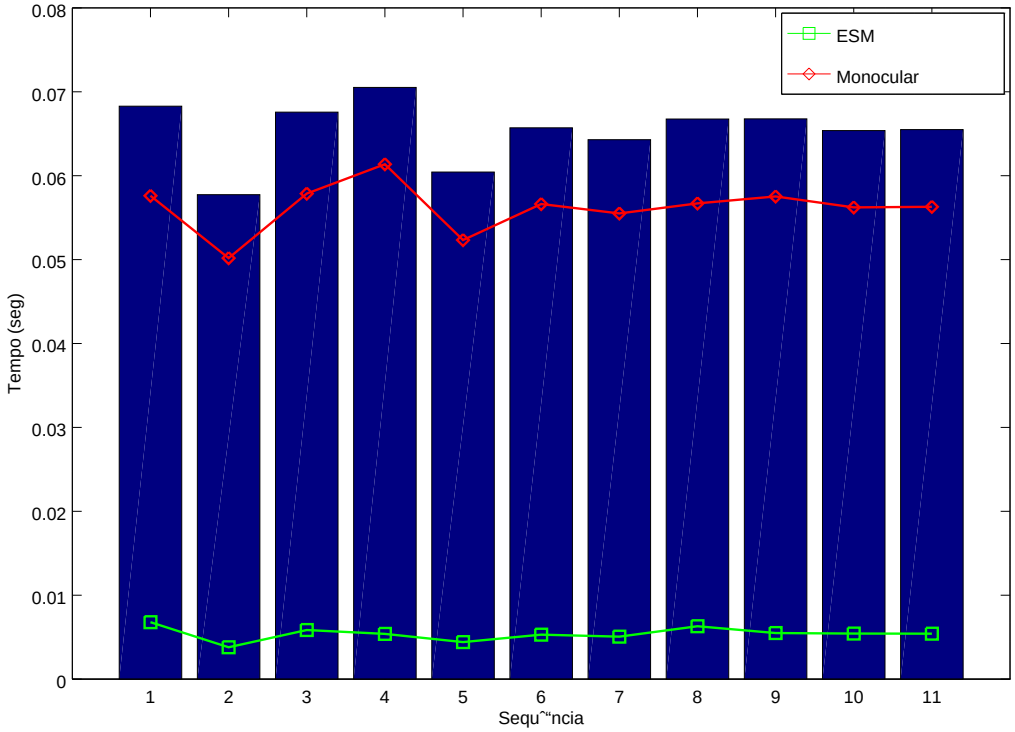


Figura 7.5: Tempo médio de execução por iteração do método **ESM-8p** para cada uma das sequências. A **barra** em azul é o tempo total de execução para uma iteração. A linha verde marcada com **quadrados** representa o tempo médio do método ESM por iteração. A linha vemelha marcada com **dimantes** representa o tempo médio do método dos oito pontos por iteração.

Figura 7.6: Sequência de quadros (4149 até 4156) com alta similaridade, a despeito do movimento. Essa similaridade induz o problema da abertura.

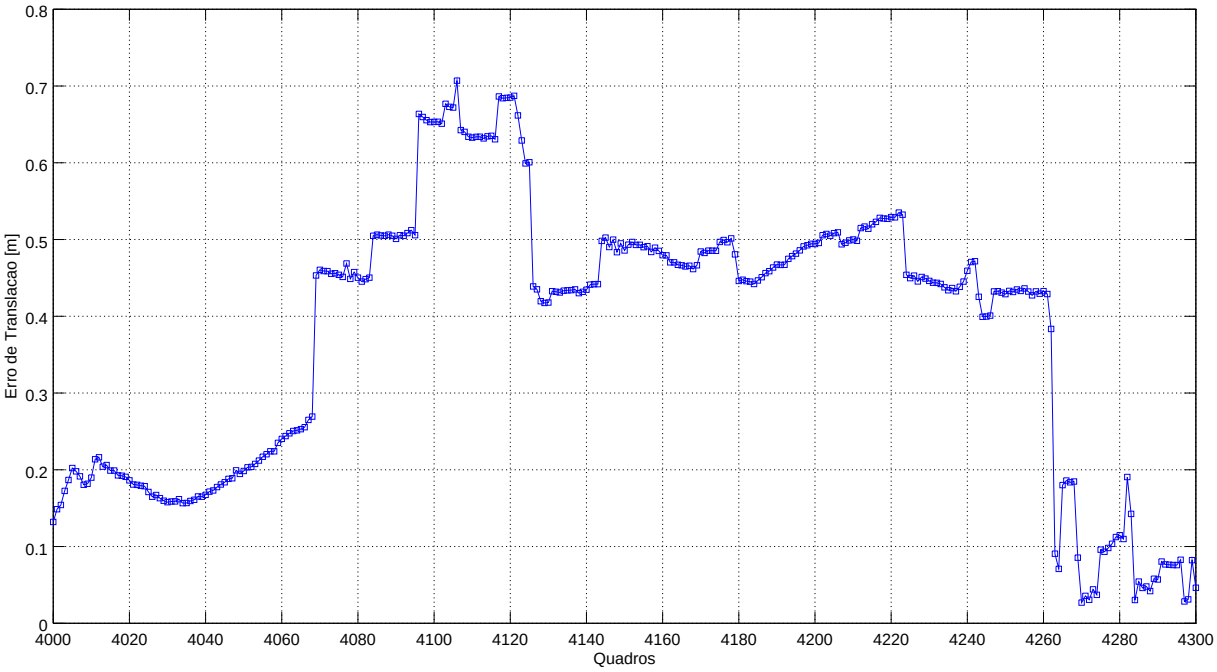


Figura 7.7: Gráfico do erro de translação (em metros) por quadro do trecho final da primeira sequência. O erro é causado pelo problema da abertura, onde a imagem no decorrer da sequência a região visível limitada causa a ilusão de não existir deslocamento.

Figura 7.8: Região da pista completamente com muita reflexão de luz. Fica difícil identificar qualquer textura na pista.

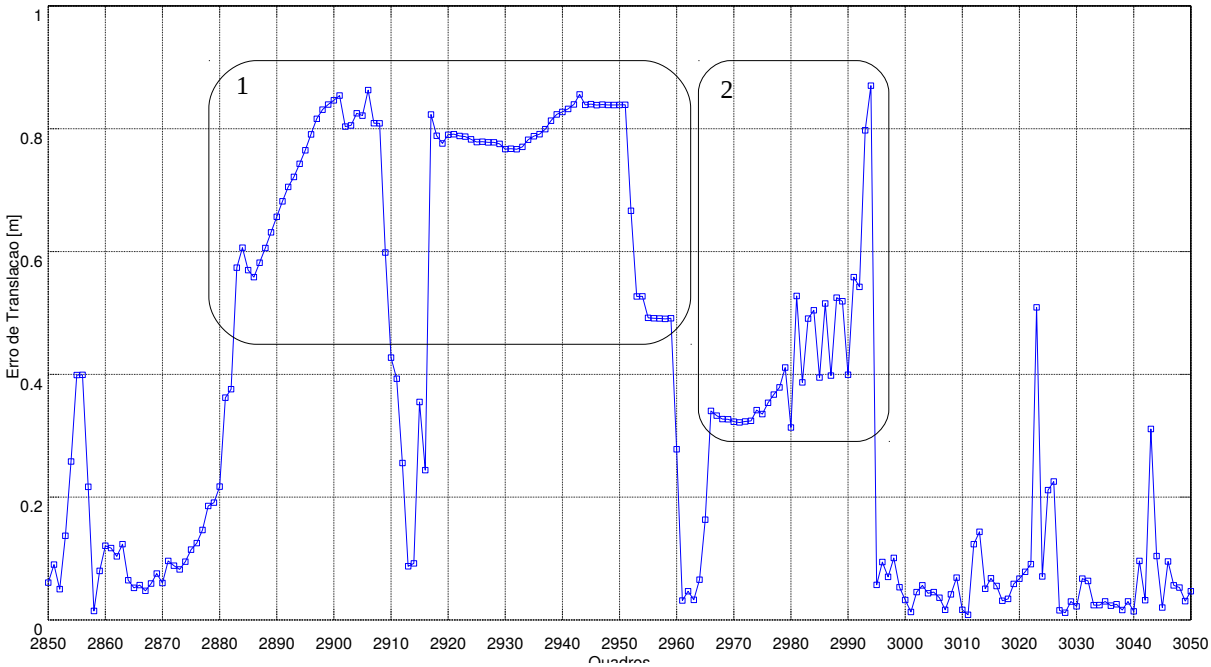


Figura 7.9: Gráfico do erro de translação (em metros) por quadro causado iluminação excessiva da cena e pelo erro de alinhamento em curvas. A região destacada número 1 representa o erro de translação causado pela iluminação excessiva. A região destacada número 2 representa o erro causado pela pelo erro de alinhamento geralmente ocorrido em curvas em regiões urbanas.

Figura 7.10: Exemplo de elemento com movimento independente e dominante na imagem.

Capítulo

8

Conclusão

Nesta dissertação é apresentada uma solução para veículos do problema de localização com estimação da escala real do movimento. Foram descritas duas classes de métodos que solucionam o problema de localização: métodos diretos e métodos baseados em reconstrução. Além da descrição foram apresentados detalhes da implementação de ambos os métodos. A combinação de dois métodos, um direto e um baseado em reconstrução, forma a solução denominada **ESM-8p** capaz de estimar a localização de um veículo em escala real.

Na Seção 8.1 serão feitas algumas ponderações referentes ao método apresentado quanto ao seu desempenho e limitações. Além das ponderações a respeito da aplicação desenvolvida, algumas possibilidades de trabalhos futuros também serão apresentados.

8.1 Discussão

O método apresentado nesta dissertação é capaz de, atendidas as restrições e condições para estimação da escala, continuamente estimar a localização de um veículo. Porém, como apresentado no Capítulo 7, a localização não é tão precisa e não deve ser comparada a sistemas como GPS com correção ou os métodos estado da arte de odometria. O método, porém pode ser integrado com outros métodos menos precisos para formar um sistema que forneça localização confiável para navegação. Exemplos de trabalhos em que métodos

de odometria menos precisos são utilizados como parte de um sistema de localização são vistos em (Lovegrove et al., 2011; Scaramuzza et al., 2009b).

Ainda com relação a precisão do método **ESM-8p**, existem algumas possíveis técnicas e estratégias que podem ser adotadas para se obter resultados mais precisos. A primeira proposta seria a implementação de um métodos direto com mais graus de liberdade. Essa abordagem não só teria efeito sobre a precisão, como também diminuiria as restrições impostas ao método. O próprio método ESM possui variações na modelagem do problema que permitem a estimação de deslocamentos com seis graus de liberdade (Mei et al., 2008; Forster et al., 2014). Estimando o deslocamento com seis graus de liberdade, a aproximação do alinhamento e consequentemente a estimação da odometria seria mais precisa. Outra opção seria o uso de múltiplas regiões sendo observadas simultâneamente (Mei et al., 2008; Ke e Kanade, 2003). A ideia por trás dessa abordagem é que as múltiplas regiões podem ser avaliadas de forma semelhante ao RANSAC, onde a estimação de localização deveria ser coerente com o melhor conjunto de regiões, tornando o algoritmo mais robusto e preciso.

O uso de várias regiões para alinhamento nos métodos diretos induz a ideia de paralelismo. De fato a parelização do método para as diversas regiões é uma opção para melhora de performance. Além da paralelização por regiões, os métodos diretos são altamente paralelizáveis, dado que os cálculos de minimização são realizados para cada ponto da região de interesse. O uso de técnicas de processamento paralelo dentro do próprio método de alinhamento é um possível ponto de otimização.

Observando os resultados da Seção 7.3, nota-se que a melhora da performance do método está muito associada ao método dos oito pontos. Visto que o método ESM apresenta desempenho melhor, quanto ao tempo de execução, mas precisão pior que o método dos oito pontos, uma estratégia de balanceamento da chamada dos métodos pode ser estudada, diminuindo a frequência de chamada do método dos oito pontos.

Referências Bibliográficas

Agrawal, M.; Konolige, K. Real-time localization in outdoor environments using stereo vision and inexpensive gps. In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, IEEE, 2006, p. 1063–1068.

Agrawal, M.; Konolige, K. Rough terrain visual odometry. In: *Proceedings of the International Conference on Advanced Robotics (ICAR)*, 2007.

Authesserres, J.-B. *Alignement d'images paramétrique: proposition d'un formalisme unifié et prise en compte du bruit pour le suivi d'objets.* Tese de Doutoramento, L'Université de Bordeaux, École Doctorale des Sciences Physiques et de L'Ingénieur, 2010.

Baker, S.; Matthews, I. Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*, v. 56, n. 3, p. 221–255, 2004.

Barron, J. L.; Thacker, N. A. Tutorial: Computing 2D and 3D optical flow. *Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester*, , n. 2004, p. 1–12, 2005.

Benhimane, S. *Vers une approche unifiee pour le suivi temps-reel et l'asservissement visuel.* Docteur en sciences - specialite: Informatique temps-reel, automatique et robotique, Ecole Nationale Superieure des Mines de Paris, 2006.

Benhimane, S.; Malis, E. Real-time image-based tracking of planes using efficient second-order minimization. *International Conference on Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ*, v. 1, p. 943–948, 2004.

Buehler, M.; Iagnemma, K.; Singh, S. *The 2005 darpa grand challenge: The great robot race.* 1st ed. Springer Publishing Company, Incorporated, 2007.

Buehler, M.; Iagnemma, K.; Singh, S. *The darpa urban challenge: Autonomous vehicles in city traffic.* 1st ed. Springer Publishing Company, Incorporated, 2009.

Campbell, J.; Sukthankar, R. Techniques for evaluating optical flow for visual odometry in extreme terrain. *Intelligent Robots and*, p. 3704–3711, 2004.

Civera, J.; Grasa, O. G.; Davison, A. J.; Montiel, J. M. M. 1-point ransac for extended kalman filtering: Application to real-time structure from motion and visual odometry. *J. Field Robot.*, v. 27, n. 5, p. 609–631, 2010.

Comport, A.; Malis, E.; Rives, P. Real-time Quadrifocal Visual Odometry. *The International Journal of Robotics Research*, v. 29, p. 245–266, 2010.

Corke, P.; Strelow, D.; Singh, S. Omnidirectional visual odometry for a planetary rover. *Intelligent Robots and Systems,*, v. 4, p. 4007–4012, 2004.

Departamento Nacional de Infraestrutura de Transporte - DNIT Departamento nacional de infraestrutura de transporte - DNIT. http://www.dnit.gov.br/, acessado: 03 de Março, 2013.

Edelman, S.; Intrator, N.; Poggio, T. Complex cells and object recognition, manuscrito não publicado, 1997.

Fischler, M.; Bolles, R. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, v. 24, n. 6, p. 381–395, 1981.

Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast Semi-Direct Monocular Visual Odometry. *Proc. IEEE Intl. Conf. on Robotics ...*, 2014.

Fraundorfer, F.; Tanskanen, P.; Pollefeys, M. A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In: *Proceedings of the 11th European conference on Computer vision: Part IV*, ECCV'10, Berlin, Heidelberg: Springer-Verlag, 2010, p. 269–282 (*ECCV'10*,).

Geiger, A. LIBVISO2: C++ Library for Visual Odometry 2. http://www.cvlibs.net/software/libviso/, acessado: 04 de Abril de 2015, 2015.

Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2012, p. 3354–3361.

GOLUB, G.; KAHAN, W. Calculating the Singular Values and Pseudo-Inverse of a Matrix. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, v. 2, n. 2, p. 205–224, 1965.

HALL, B. C. *Lie Groups, Lie Algebras, and Representations*, v. 222. Springer New York, 2003.

HARALICK, R. M.; LEE, C.-N.; OTTENBERG, K.; NÖLLE, M. Review and analysis of solutions of the three point perspective pose estimation problem. *Int. J. Comput. Vision*, v. 13, n. 3, p. 331–356, 1994.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: *Alvey vision conference*, Manchester, UK, 1988, p. 50.

HARTLEY, R. I. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 19, n. 6, p. 580–593, 1997.

HARTLEY, R. I.; ZISSERMAN, A. *Multiple view geometry in computer vision.* Second ed. Cambridge University Press, ISBN: 0521540518, 2004.

HIRSCHMULLER, H.; INNOCENT, P.; GARIBALDI, J. Fast, unconstrained camera motion estimation from stereo without tracking and robust statistics. In: *Control, Automation, Robotics and Vision, 2002. ICARCV 2002. 7th International Conference on*, IEEE, 2002, p. 1099–1104.

HOWARD, A. Real-time stereo visual odometry for autonomous ground vehicles. In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Ieee, 2008, p. 3946–3952.

HUANG, T. S.; FAUGERAS, O. D. Some properties of the e matrix in two-view motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 11, n. 12, p. 1310–1312, 1989.

IRANI, M.; ANANDAN, P. About direct methods. *ICCV workshop on Vision Algorithms*, p. 267–277, 1999.

JIANG, Y.; CHEN, H.; XIONG, G.; GONG, J.; JIANG, Y. Kinematic constraints in visual odometry of intelligent vehicles. In: *Intelligent Vehicles Symposium (IV), 2012 IEEE*, 2012, p. 1126 –1131.

JONES, E.; OLIPHANT, T.; PETERSON, P.; ET AL. SciPy: Open source scientific tools for Python. `http://www.scipy.org/`, 2001.

Ke, Q.; Kanade, T. Transforming camera geometry to a virtual downward-looking camera: robust ego-motion estimation and ground-layer detection. In: *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, IEEE Comput. Soc, 2003, p. I–390–I–397.

Kitt, B.; Chambers, A.; Lategahn, H.; Singh, S.; Systems, C. Monocular Visual Odometry using a Planar Road Model to Solve Scale Ambiguity. *Measurement And Control*, p. 1–6, 2011.

Koenderink, J. J. The structure of images. *Biological Cybernetics*, v. 50, n. 5, p. 363–370–370, 1984.

Kruppa, E. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. *Sitzungsberichte der Mathematisch Naturwissenschaftlichen Kaiserlichen Akademie der Wissenschaften*, v. 122, p. 1939–1948, 1913.

Laboratório de Transporte e Logística - LabTrans LabTrans. http://www.labtrans.ufsc.br/, acessado: 03 de Março de 2014, 2013.

Lategahn, H.; Geiger, A.; Kitt, B.; Stiller, C. Motion-without-structure: Real-time multipose optimization for accurate visual odometry. In: *Intelligent Vehicles Symposium (IV), 2012 IEEE*, 2012, p. 649 –654.

Leonard, J.; Durrant-Whyte, H. Mobile robot localization by tracking geometric beacons. *Robotics and Automation, IEEE Transactions on*, v. 7, n. 3, p. 376 –382, 1991.

Lim, J.; Barnes, N.; Li, H. Estimating relative camera motion from the antipodal-epipolar constraint. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 32, n. 10, p. 1907 –1914, 2010.

Lindeberg, T. Scale-space theory: A basic tool for analysing structures at different scales. *J. of Applied Statistics*, v. 21(2), p. 224–270, 1994.

Longuet-Higgins, H. C. A computer algorithm for reconstructing a scene from two projections. *Nature*, v. 293, p. 133–135, 1981.

Lovegrove, S.; Davison, A. J.; Ibanez-Guzman, J. Accurate visual odometry from a rear parking camera. In: *2011 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2011, p. 788–793.

Lowe, D. Object recognition from local scale-invariant features. In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, Ieee, 1999, p. 1150–1157.

Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, v. 60, n. 2, p. 91–110, 2004.

Lucas, B. D.; Kanade, T. An iterative image registration technique with an application to stereo vision. *IJCAI*, , n. x, p. 674–679, 1981.

Ma, Y.; Soatto, S.; Kosecka, J.; Sastry, S. S. *An invitation to 3-d vision: From images to geometric models*. SpringerVerlag, 2003.

Malis, E. Improving vision-based control using efficient second-order minimization techniques. In: *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, IEEE, 2004, p. 1843–1848 Vol.2.

Mégret, R.; Authesserre, J. B.; Berthoumieu, Y. The Bi-directional framework for unifying parametric image alignment approaches. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, p. 400–411.

Mei, C. *Laser-augmented omnidirectional vision for 3D localisation and mapping*. Tese de Doutoramento, INRIA Sophia Antipolis, 2007.

Mei, C.; Benhimane, S.; Malis, E.; Rives, P. Efficient Homography-Based Tracking and 3-D Reconstruction for Single-Viewpoint Sensors. *IEEE Transactions on Robotics*, v. 24, n. 6, p. 1352–1364, 2008.

Moravec, H. Obstacle avoidance and navigation in the real world by a seeing robot rover. *tech report CMURITR8003 Robotics Institute Carnegie Mellon University doctoral dissertation Stanford University*, 1980.

Nistér, D. Preemptive RANSAC for live structure and motion estimation. *Machine Vision and Applications*, v. 16, n. 5, p. 321–329, 2003.

Nistér, D. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, v. 26, n. 6, p. 756–77, 2004.

Nistér, D.; Naroditsky, O.; Bergen, J. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, v. 23, n. 1, p. 3–20, 2006.

PARKS, D.; GRAVEL, J.-P. Corner detection - moravec operator. Acessado: 03 de Março, 2013.
Disponível em `http://kiwi.cs.dal.ca/~dparks/CornerDetection/moravec.htm`

PLOEG, J.; SHLADOVER, S.; NIJMEIJER, H.; WOUW, N. Introduction to the special issue on the 2011 grand cooperative driving challenge. *Intelligent Transportation Systems, IEEE Transactions on*, v. 13, n. 3, p. 989 –993, 2012.

POMERLEAU, D. Alvinn: An autonomous land vehicle in a neural network. In: TOURETZKY, D., ed. *Advances in Neural Information Processing Systems 1*, Morgan Kaufmann, 1989.

SCARAMUZZA, D.; FRAUNDORFER, F.; POLLEFEYS, M.; SIEGWART, R. Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints. In: *Computer Vision, 2009 IEEE 12th International Conference on*, 2009a, p. 1413 –1419.

SCARAMUZZA, D.; FRAUNDORFER, F.; SIEGWART, R. Real-time monocular visual odometry for on-road vehicles with 1-point ransac. In: *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, 2009b, p. 4293 –4299.

SCARAMUZZA, D.; SIEGWART, R. Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *Robotics, IEEE Transactions on*, v. 24, n. 5, p. 1015–1026, 2008.

SHI, J.; TOMASI, C. Good features to track. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, IEEE Comput. Soc. Press, 1994, p. 593–600.

SHUM, H.; SZELISKI, R. Construction of panoramic image mosaics with global and local alignment. *International Journal of Computer Vision*, v. 36, n. 2, p. 101–130, 2000.

STEWENIUS, H.; NISTÉR, D.; AL. Solutions to minimal generalized relative pose problems. In: *IN WORKSHOP ON OMNIDIRECTIONAL VISION*, 2005.

SZELISKI, R. Image Alignment and Stitching: A Tutorial. *Foundations and Trends® in Computer Graphics and Vision*, v. 2, n. 1, p. 1–104, 2006.

TANI, V. Z.; OTTO, G. G.; PEÑA, C. C. Dados de boletins de ocorrência. 2008.

Tardif, J.; Pavlidis, Y.; Daniilidis, K. Monocular visual odometry in urban environments using an omnidirectional camera. In: *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, IEEE, 2008, p. 2531–2538.

The European Robot Trial - ELROB The european robot trial - ELROB. http://www.elrob.org/, acessado: 03 de Março de 2013, 2013.

Thorpe, C.; Hebert, M.; Kanade, T.; Shafer, S. Vision and navigation for the Carnegie-Mellon Navlab. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 10, n. 3, p. 362–373, 1988.

Trucco, E.; Verri, A. *Introductory techniques for 3-d computer vision*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

Tuytelaars, T.; Mikolajczyk, K. Local invariant feature detectors: A survey. *Foundations and Trends® in Computer Graphics and Vision*, v. 3, n. 3, p. 177–280, 2007.

Varadarajan, V. S. *Lie Groups, Lie Algebras, and Their Representations*, v. 102. Springer-Verlag New York, xiii+430 p., 1984.

World Health Organization WHO | world health organization. http://www.who.int/en/, acessado: 03 de Março, 2013.
Disponível em `http://www.who.int/mediacentre/factsheets/fs358/en/`

Zhu, M.; Ramalingam, S.; Taguchi, Y.; Garaas, T. W. Monocular visual odometry and dense 3d reconstruction for on-road vehicles. In: Fusiello, A.; Murino, V.; Cucchiara, R., eds. *ECCV Workshops (2)*, Springer, 2012, p. 596–606 (*Lecture Notes in Computer Science*, v.7584).