

## 3

### Visão Computacional

#### 3.1.

##### Introdução

A calibração do manipulador TA-40 no seco permite que se tenha uma boa precisão na posição e orientação da extremidade do manipulador em relação à sua base, corrigindo os erros estáticos intrínsecos ao sistema. Em várias aplicações de manipuladores este nível de calibragem é suficiente, pois a base está sempre fixa na mesma posição. É o caso de um manipulador fixado no chão e que manipula objetos, os quais não mudam sensivelmente de posição.

A aplicação desta dissertação possui um desafio a mais para a calibração. O manipulador estará fixado sobre um robô submarino, que o levará até sua área de trabalho. Assim, dificilmente a base do manipulador estará no mesmo local a cada vez em que ele for levado para sua área de trabalho, pois o braço mecânico provavelmente não irá se atracar duas vezes na mesma exata posição no fundo do oceano, além de não haver garantias de que a base não se movimentará junto com o robô uma vez atracada. Logo, um método para se obter uma medida da posição da base do manipulador em relação à sua área de trabalho se faz essencial. Uma maneira consiste no uso de duas câmeras conectadas à garra do manipulador. Essa escolha é apropriada em virtude da redução do custo de câmeras na última década e pela existência de diversos algoritmos que permitirão um posicionamento com bastante precisão da base e dos objetos que serão manipulados.

Neste capítulo, será inicialmente feita uma breve descrição do modelo matemático da câmera e de sua calibração. Também será abordado um método para encontrar pontos-chave na imagem obtida por uma das câmeras e como correlacioná-las com pontos-chave encontrados nas imagens obtidas da outra câmera, procedimento este necessário para a obtenção da posição dos objetos no espaço tridimensional.

### 3.2.

#### Modelagem da Câmera

##### 3.2.1.

##### Modelo Pinhole

Um dos modelos de câmera mais difundidos é o *pinhole*, representado esquematicamente na Figura 9. Segundo este modelo, um raio de luz proveniente de um ponto  $M$  no espaço atravessa um orifício  $C$  e incide em um plano no ponto  $m$ . Dá-se o nome de plano de imagem ao plano de incidência dos raios, enquanto que o orifício  $C$  é chamado de centro da câmera. Um parâmetro intrínseco da câmera, chamado distância focal, pode ser medido pela distância do plano de imagem ao centro da câmera, e é geralmente representado por  $f$ .

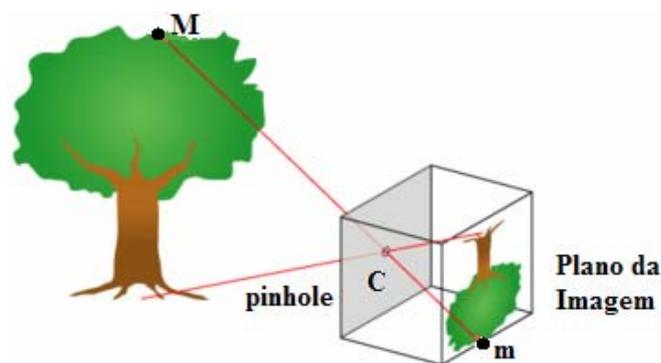


Figura 9: Representação do modelo pinhole.

A Figura 10 representa a geometria do modelo pinhole, tendo o quadro da câmera no espaço 3D com sua origem no orifício  $C$ . Também é representado o plano da imagem normalizado em um sistema de coordenadas em duas dimensões com origem no ponto  $\hat{C}$ . Por semelhança de triângulos, é possível demonstrar que o ponto  $P=(C_x, C_y, z_0)$  é mapeado no plano da imagem normalizado pelo ponto  $\hat{p}=(\hat{u}, \hat{v})=(fC_x/z_0, fC_y/z_0)$ . É possível visualizar os pontos  $P$  e  $\hat{p}$  em coordenadas homogêneas, pela seguinte equação:

$$\begin{cases} \hat{u} = \frac{C_x}{Z} \\ \hat{v} = \frac{C_y}{Z} \end{cases} \Leftrightarrow \hat{p} = \begin{bmatrix} \hat{u} \\ \hat{v} \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} Id & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} {}^c P \\ 1 \end{bmatrix} \quad \text{Eq. 3.2.1}$$

$$\text{Sendo } z = \frac{f}{z_0}.$$

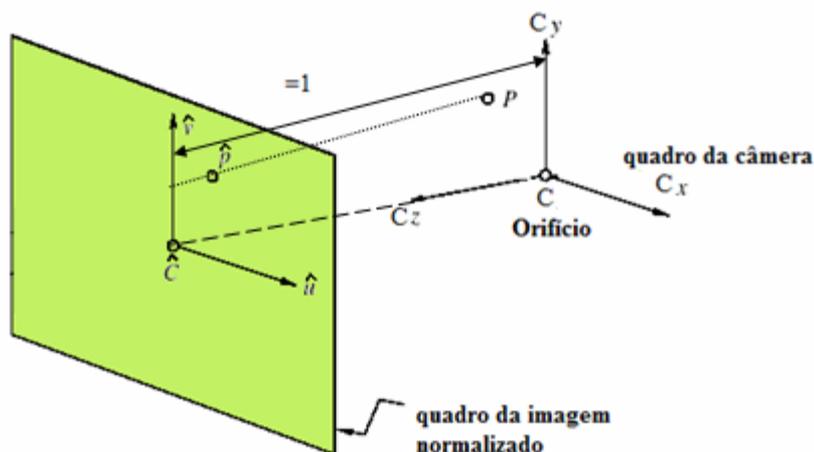


Figura 10: Geometria do modelo pinhole.

### 3.2.2.

#### Parâmetros Intrínsecos

Na Equação 3.2.1, supõe-se que a origem do sistema de coordenadas do plano da imagem é a origem da imagem, porém dependendo do programa empregado, a imagem que é digitalmente expressa como uma matriz, pode ter sua origem em uma posição diferente da do plano da imagem. O modelo pinhole também considera escalas iguais para os eixos x e y, o que não é sempre verdade para câmeras CCDs usadas na prática. A Figura 11 mostra a diferença entre o modelo pinhole e um modelo mais realista, onde os eixos x e y têm escalas distintas e as câmeras possuem cisalhamento de pixels.

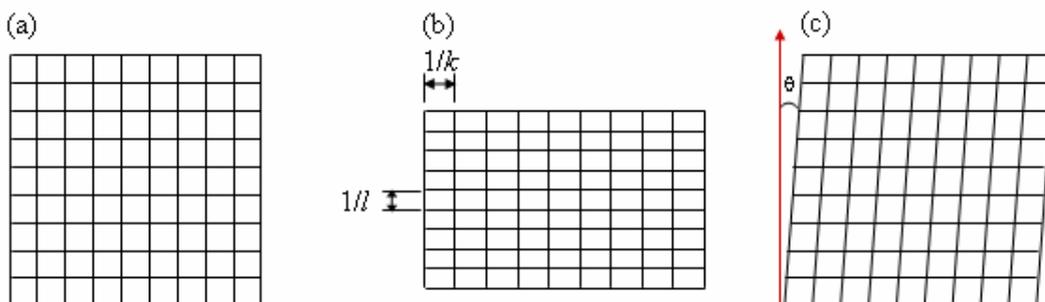


Figura 11: Modificações ao modelo pinhole com base na escala e cisalhamento dos pixels das câmeras. Figura (a) mostra os pixels segundo o modelo pinhole, figura (b) mostra um sistema mais realista onde escalas dos eixos x e y se diferenciam, figura (c) mostra o problema de cisalhamento

Além disto, a lente pode possuir distorções que precisam ser modeladas. Para se adequar a essa realidade, a Equação 3.2.1 deve ser alterada, criando-se na modelagem o quadro da retina física da imagem, após o quadro da imagem normalizada, conforme visto na Figura 12.

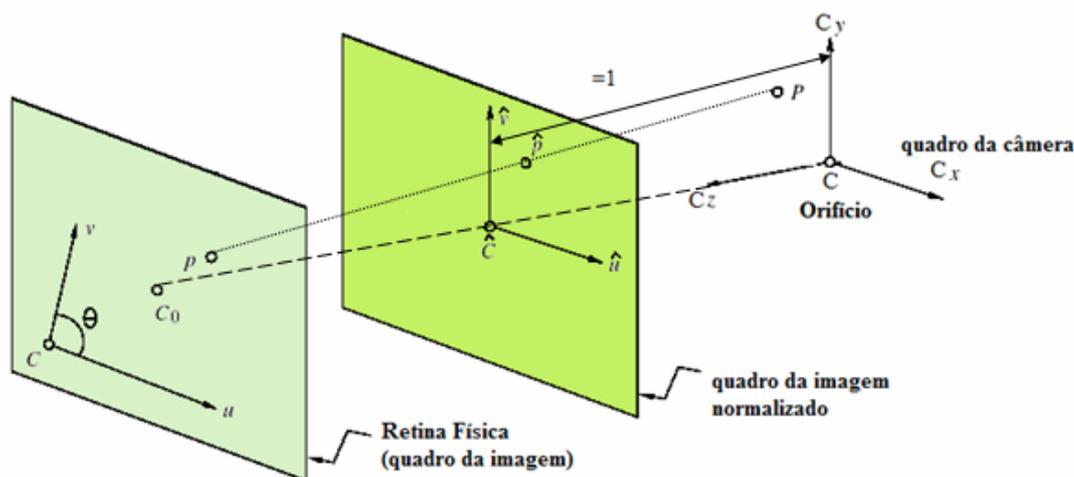


Figura 12: Modelo pinhole modificado.

Considerando-se que os pixels da câmera podem ter escalas diferentes nos eixos  $x$  e  $y$ , com  $1/k$  e  $1/l$  sendo as dimensões horizontal e vertical do pixel, tem-se:

$$\begin{cases} u = k f \hat{u} \\ v = l f \hat{v} \end{cases} \quad \text{Eq. 3.2.2}$$

Além disto, considerando-se que a origem do quadro da imagem não se encontra no centro do quadro, mas sim deslocado em  $(u_0, v_0)$ :

$$\begin{cases} u = \alpha \hat{u} + u_0 \\ v = \beta \hat{v} + v_0 \end{cases} \quad \text{Eq. 3.2.3}$$

, onde  $\alpha = kf$  e  $\beta = lf$

Por fim, considera-se que a matriz de sensores pode se encontrar distorcida de um ângulo  $\theta$ :

$$\begin{cases} u = \alpha \hat{u} - \alpha \cot \theta \hat{v} + u_0 \\ v = \beta \hat{v} + v_0 \end{cases} \quad \text{Eq. 3.2.4}$$

Matricialmente, pode-se definir uma matriz  $K$  dos parâmetros intrínsecos da câmera, onde:

$$\mathbf{p} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & -\alpha \cot \theta & u_0 \\ 0 & \beta / \sin \theta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{v} \\ 1 \end{pmatrix} = \mathbf{K} \hat{\mathbf{p}} \quad \text{Eq. 3.2.5}$$

A matriz de parâmetros intrínsecos  $K$  transforma as coordenadas no quadro da imagem normalizado para o quadro da retina física da imagem. Por fim, pode-se utilizar esta matriz para mapear os pontos no quadro da câmera para o quadro da retina física da imagem por meio de matrizes homogêneas:

$$\mathbf{p} = \mathbf{K} \frac{1}{z} (\text{Id } \mathbf{0}) {}^c\mathbf{P} = \frac{1}{z} (\mathbf{K} \mathbf{0}) {}^c\mathbf{P} \quad \text{Eq. 3.2.6}$$

Na Equação 3.2.6, o ponto  ${}^c\mathbf{P}$  representa o ponto  $P$  em coordenadas homogêneas no sistema de coordenadas da câmera.

### 3.2.3.

#### Parâmetros Extrínsecos

Em geral, é mais conveniente expressar os pontos no espaço em termos de um ponto de referência no mundo, sendo o mundo definido como o quadro do espaço objeto. As coordenadas do mundo estão relacionadas com as coordenadas do centro da câmera por uma rotação e uma translação. Assim, se  ${}^w\mathbf{O}$  é o centro do sistemas de coordenadas do mundo e  ${}^c\mathbf{O}$  o centro do sistema de coordenadas da câmera, pode-se mapear a localização de um ponto  $P$  representado no quadro no mundo para o quadro da câmera. Este processo está representado na Figura 13.

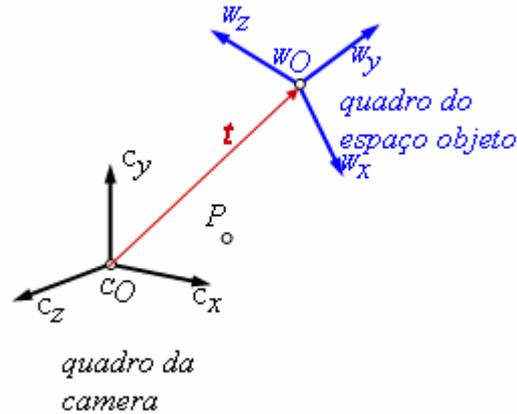


Figura 13: Transformação de coordenadas do mundo para a câmera.

Essa mudança de coordenadas pode ser descrita como uma transformação de matrizes, conforme a Equação 3.2.6:

$${}^c\mathbf{P} = \begin{pmatrix} {}^c\mathbf{R}_w & {}^c\mathbf{O}_w \\ \mathbf{0}^T & 1 \end{pmatrix} {}^w\mathbf{P} = \begin{pmatrix} {}^c\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} {}^w\mathbf{P} \quad \text{Eq. 3.2.7}$$

Os seis parâmetros das matrizes de rotação e translação são chamados de parâmetros extrínsecos da câmera, e mapeiam a posição de um ponto no sistema de coordenadas do mundo até a posição no sistema de coordenadas da câmera. Combinando as Equações 3.2.6 e 3.2.7, obtém-se um mapeamento completo de um ponto no quadro do mundo para o quadro da imagem.

$$\mathbf{p} = \frac{1}{z} (\mathbf{K} \mathbf{0}) \begin{pmatrix} {}^c\mathbf{R}_w & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} {}^w\mathbf{P} = \frac{1}{z} \mathbf{K} \begin{pmatrix} {}^c\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} {}^w\mathbf{P} \quad \text{Eq. 3.2.8}$$

A matriz que representa o mapeamento é denominada Matriz de Projeção Perspectiva (M):

$$\mathbf{M} = \mathbf{K} \begin{pmatrix} {}^c\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \rightarrow \mathbf{p} = \frac{1}{z} \mathbf{M} {}^w\mathbf{P} \quad \text{Eq. 3.2.9}$$

### 3.3.

#### Calibração de Câmera

Como visto na seção 5.2, um ponto no espaço 3D pode ser mapeado em um ponto no plano da imagem através da Equação (3.2.9). Contudo, a matriz de projeção da câmera ( $M$ ) deve ser conhecida. Essa matriz é uma característica particular de cada câmera e depende do processo de fabricação. Além disso, essa matriz pode ser alterada durante a execução normal do sistema por uma mudança na distância focal da câmera, e pode também variar de acordo com as condições do ambiente (variações mecânicas, térmicas). Por isso, a única forma de determinar a matriz de projeção corretamente é através da calibração da câmera.

A calibração da câmera é um procedimento distinto da calibração do manipulador. Enquanto que, no segundo caso, procura-se a matriz de erros de cada elo do manipulador, no caso da calibração de câmeras procura-se obter os parâmetros intrínsecos e extrínsecos da câmera. Como na aplicação serão utilizadas duas câmeras acopladas ao manipulador, a calibração tem como objetivo saber a posição de uma câmera em relação à outra.

Existem diversos métodos de calibração de câmeras presentes na literatura, sendo um processo bem testado, com diversos programas já implementados utilizando algoritmos diferenciados. A base dos algoritmos é a resolução de um sistema equações formado por diversos pontos coletados pelas câmeras, tendo como incógnitas os doze componentes da matriz de projeção  $M$ . A solução pode ser feita de forma linear ou não-linear. Não entra no escopo desta dissertação ir a fundo nos algoritmos, porém podem ser encontradas informações nas referências (Forsyth, D. A., pp.43-59, 2003).

É possível fazer o reconhecimento das posições de pontos no plano do mundo e relacioná-los facilmente com pontos no plano da câmera utilizando-se plataformas de calibragem como as mostradas na Figura 14. Conhecendo a distância dos lados dos quadrados, obtém-se facilmente a posição de cada ponto em relação ao outro.

Os resultados da calibragem levam a uma matriz de projeção que considera a origem do plano do mundo no centróide dos pontos observados na plataforma de calibração, conforme pode ser visto na Figura 15. Assim, a matriz

M considera uma translação e uma rotação em relação a essa origem, que pode não ser conveniente na prática.



Figura 14: Plataforma de Calibração

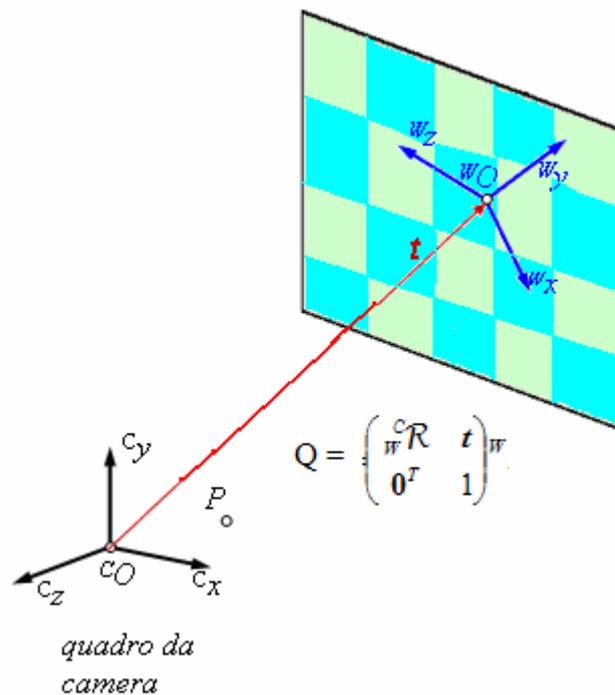


Figura 15: Transformação do quadro do mundo para o quadro da câmera

O manipulador robótico possuirá duas câmeras robóticas acopladas a sua extremidade, apresentando uma translação e rotação entre si. Esta translação e rotação serão futuramente importantes para posicionar os objetos vistos pelas câmeras e posicionar a base do manipulador em relação a pontos captados do ambiente de trabalho pelas câmeras. A calibração de câmeras será revista no capítulo 4 onde a aplicação das técnicas de visão computacional vai ser apresentada.

### 3.4.

#### Emparelhamento de Pontos

Um dos principais problemas de visão computacional é encontrar semelhanças em imagens. Tal problema ocorre em reconhecimento de objetos e cenas, resolução de estruturas 3D em múltiplas imagens, correspondência estéreo e busca de movimento. Na calibração da câmera, faz-se uso de uma plataforma de calibração que facilita o processo, mas no fundo do oceano o problema de emparelhar pontos em duas imagens distintas é bem mais complexo. A Figura 16 mostra como um mesmo objeto pode ser visto de formas diferentes por câmeras posicionadas de forma distinta.

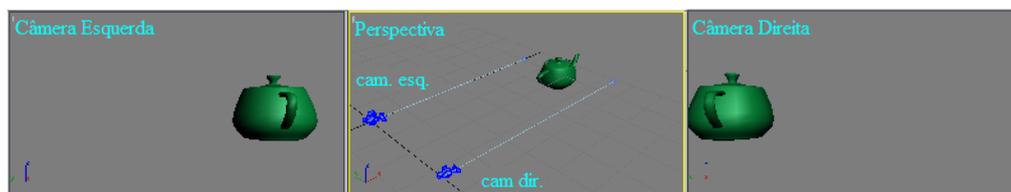


Figura 16: Visão de um mesmo objeto em câmeras distintas deslocadas por uma distância fixa.

Na aplicação, existirá um par de câmeras estéreo, vendo a mesma imagem a não ser por uma diferença de posição entre elas. Elas estarão sujeitas a ruídos na imagem causadas por um ambiente onde há pouca iluminação e distorções por refração. Para que seja possível obter a posição do manipulador em relação ao seu ambiente, será preciso encontrar pontos em ambas as imagens que sejam os mesmos no mundo real.

Nesta seção será apresentado um algoritmo para extrair pontos em uma imagem que possuem diversas propriedades que os habilitam a serem descobertos em outras imagens de um mesmo objeto ou cena. O método que será apresentado obtém, para cada ponto encontrado, propriedades robustas a variações na imagem, tais quais rotação, mudança de escala, iluminação e ponto de vista da câmera. Além disto, as propriedades obtidas dos pontos por esse método tem elevada robustez a ruído. O método é capaz de distinguir pontos diferentes a partir destas propriedades, com alta probabilidade de acerto. Essa metodologia é denominada Transformada de Propriedades Invariantes à Escala (“Scale Invariant Feature

Transform”, SIFT), pois transforma dados da imagem em coordenadas invariantes à escala e a outras características locais, como luminosidade e rotação.

Um aspecto importante desta metodologia é que ela gera um grande número de características que cobrem densamente a imagem sobre uma ampla gama de escalas e localizações. A quantidade de propriedades é particularmente importante para o reconhecimento de objetos, pois garante menor possibilidade de erros quando se faz a busca do mesmo objeto na nuvem de pontos de outras imagens. A seguir, será explicado de forma resumida cada estágio do algoritmo.

### 3.4.1.

#### Detecção de Extremos no Espaço de Escalas

A primeira etapa do algoritmo é buscar pontos que sejam invariantes a mudanças de escala da imagem, possibilitando a detecção de pontos com a câmera próxima ou distante do objeto de interesse. Tal objetivo é obtido procurando os pontos em todas as possíveis escalas, utilizando-se de uma função chamada de **espaço de escala**, que no caso é a função **gaussiana**. Assim, o espaço de escalas  $L(x,y,\sigma)$  é formado pela filtragem da imagem  $I(x,y)$  por funções gaussianas  $G(x,y,\sigma)$ , conforme visto abaixo:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad \text{Eq. 3.4.1}$$

Onde \* simboliza a convolução em x e y. A função gaussiana  $G(x,y,\sigma)$  é expressa por:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad \text{Eq. 3.4.2}$$

A eficiência da busca por pontos-chaves é aumentada substituindo-se a função  $L(x,y,\sigma)$  por uma função de **Diferença-do-Gaussiano**, formada pela subtração de dois espaços de escala, separadas por um fator multiplicativo k, conforme mostrado na Equação 3.4.3. Esta função se aproxima da função **Laplaciano do Gaussiano** normalizado em escala  $\sigma^2 \Delta^2 G$ , que permite

(Lindeberg 1994) a possui pontos invariantes à escala em seus máximos e mínimos.

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad \text{Eq. 3.4.3}$$

A Figura 17 mostra uma imagem submetida à funções  $L(x, y, \sigma)$  e  $D(x, y, \sigma)$ , com  $\sigma=2$ .



Figura 17: Imagem (à esquerda) submetida a gaussiana (no meio) e à diferença-do-gaussiano (à direita).

A Figura 18 demonstra o algoritmo para criação da diferença do gaussiano:

1- A imagem inicial sofre convoluções incrementais com Gaussianos para produzir imagens separadas por um fator de escala  $k$  no espaço de escala (empilhados na coluna esquerda)

2- Cada octava (grupo de cinco imagens que sofrem seguidas convoluções) do espaço de escala é dividida por um número inteiro,  $s$ , de forma que  $k = 2^{1/s}$ , produzindo assim  $s+3$  imagens na octava de forma que a detecção de extremos cubra toda octava.

3- Imagens em escalas adjacentes são subtraídas para produzir as imagens da diferença do gaussiano mostradas na direita.

4- Uma vez processada a octava, é reduzida a amostragem da imagem (*downsample*) com duas vezes o valor inicial de  $\sigma$  (duas imagens no topo da pilha), voltando-se ao passo 1.

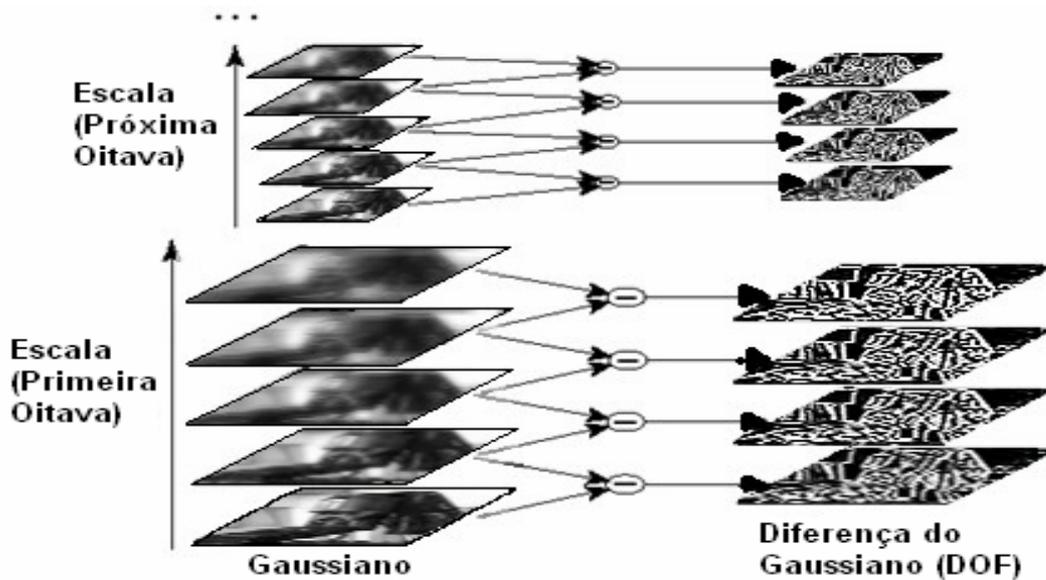


Figura 18: Representação do procedimento de obtenção de diferentes diferenças de gaussianas para diversas octavas da imagem.

Os pontos-chave são localizados nos máximos e mínimos locais de cada imagem  $D(x,y,\sigma)$ , que podem ser obtidos comparando-se a intensidade de cada ponto com as intensidades de seus seis pontos vizinhos na sua escala, e com os nove pontos vizinhos na escala acima e abaixo, como visto na Figura 19. Procuram-se pontos com valores de intensidade maiores ou menores que todos os seus vizinhos. É importante ressaltar que é possível encontrar um ponto máximo ou mínimo em uma imagem após ter sua amostragem reduzida, e portanto a sua posição nesta imagem reduzida não é equivalente à sua posição na imagem original, que possui mais pixels, sendo necessária uma interpolação para descobrir a posição na imagem original.

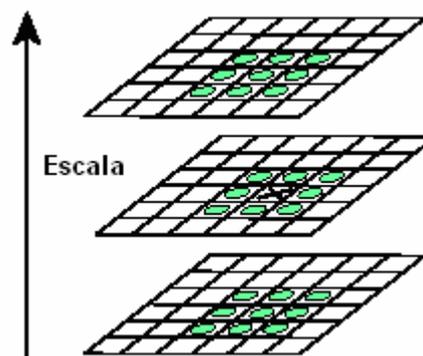


Figura 19: Máximo e mínimo das imagens de Diferença-do-Gaussiano são detectados comparando o pixel (marcado por X) com seus 26 vizinhos em 3x3 regions nas escalas corrente e adjacentes (marcados por círculos).

### 3.4.2.

#### Localização Precisa de Pontos-Chaves

Os pontos encontrados na seção anterior devem ser filtrados para retirar aqueles que possuem pouco contraste (sendo sensíveis a ruído) e aqueles próximos de extremidades. A eliminação de pontos sensíveis a ruídos é obtida utilizando uma expansão de Taylor sobre a função Diferença-do-Gaussiano deslocada de modo a ter sua origem no ponto de amostragem (Brown, M. and Lowe):

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} \quad \text{Eq. 3.4.4}$$

onde  $D$  e suas derivadas são calculadas no ponto de amostragem e  $x=(x,y,\sigma)^T$  é o deslocamento deste ponto. A localização do extremo,  $\hat{x}$ , é dada por:

$$\hat{x} = - \frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \quad \text{Eq. 3.4.5}$$

O valor da função no extremo,  $D(\hat{x})$ , é útil para rejeição de extremos instáveis com baixo contraste. Pode-se obter este resultado substituindo a Equação 3.4.5 na Equação 3.4.4, descartando-se valores muito baixos de  $D(\hat{x})$  :

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x} \quad \text{Eq. 3.4.6}$$

A eliminação de pontos-chaves próximos da extremidade é feita usando-se uma matriz Hessiana 2x2,  $H$ , na localização e escala do ponto-chave na função  $D$ . A Hessiana representa uma segunda derivada, permitindo mensurar o tamanho das curvaturas de  $D$  em seus autovalores. A curvatura próxima a pontos cujo gradiente é paralelo a extremidades possui baixa amplitude, mas em pontos com gradiente perpendicular a essas extremidades a amplitude é elevada.

$$H = \begin{bmatrix} D_{XX} & D_{XY} \\ D_{XY} & D_{YY} \end{bmatrix} \quad \text{Eq. 3.4.7}$$

Sendo  $\alpha$  o autovalor com maior magnitude e  $\beta$  o de menor magnitude, pode-se computar a soma e o produto destes autovalores:

$$\begin{aligned} Tr(H) &= D_{XX} + D_{YY} = \alpha + \beta \\ Det(H) &= D_{XX}D_{YY} - D_{XY}^2 = \alpha\beta \end{aligned} \quad \text{Eq. 3.4.8}$$

Caso o determinante seja negativo, as curvaturas têm sinais diferentes e o ponto é descartado, não sendo um extremo. Sendo  $r$  a razão entre o maior e o menor autovalor,  $\alpha=r\beta$ :

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \quad \text{Eq. 3.4.9}$$

A equação 3.4.9 depende apenas da razão entre os autovalores. A quantidade  $(r+1)^2/r$  oferece uma medida de quanto os autovalores são distintos. Eliminam-se pontos próximos a extremidades descartando-se pontos abaixo de um determinado *threshold*:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \quad \text{Eq. 3.4.10}$$

Assim, eliminam-se pontos-chave que não são estáveis, apesar de estarem próximos de extremidades.

### 3.4.3.

#### Atribuição de Orientação

A invariância ao ruído e à escala já são um passo importante para os pontos-chave, porém eles devem ser invariantes à rotação da câmera. Para se obter essa invariância, atribui-se uma orientação baseada nas características locais da imagem.

Para tanto, calcula-se para cada amostragem da imagem,  $L(x,y)$ , na escala do ponto-chave, a magnitude,  $m(x,y)$  e a orientação  $\theta(x,y)$  do gradiente usando-se diferenças de pixels.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad \text{Eq. 3.4.11}$$

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right) \quad \text{Eq. 3.4.12}$$

Na região vizinha ao ponto-chave, forma-se um histograma de orientações composto pelas orientações dos pontos amostrados. Esse histograma possui 36 pontos, cobrindo os 360 graus de orientação, conforme visto na figura 20.

O pico deste histograma é utilizado para definir a sua orientação. No caso de múltiplos picos de elevada amplitude, o ponto-chave receberá múltiplas orientações, tornando-o ainda mais estável para futura identificação. Ao final, uma parábola é usada para interpolar os três valores do histograma mais próximos do pico, de forma a se ter uma melhor precisão de sua posição.

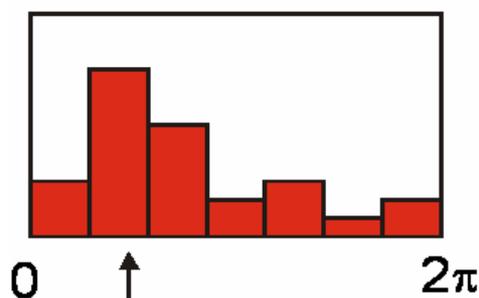


Figura 20: Histograma de orientação do ponto-chave.

#### 3.4.4.

#### Descritor Local da Imagem

As seções anteriores apresentaram procedimentos para posicionamento de pontos-chaves, e atribuir-lhes escala, resistência a ruído e orientações, permitindo a imposição de um sistema de coordenadas 2D na qual descrever uma região da imagem localmente, e prover invariância a esses parâmetros. O próximo passo será de atribuir a esses pontos-chave descritores invariantes a mudanças de iluminação e ponto de vista 3D, além de torná-los bem distinguíveis.

As magnitudes e orientações dos gradientes são amostradas ao redor da localização do ponto-chave, utilizando a escala do ponto-chave para escolher o nível de suavização Gaussiano para a imagem.

Para que os descritores tenham invariância a rotação, as coordenadas e as orientações dos descritores são rotacionadas relativo a orientação do ponto-chave. Este procedimento está mostrado pelas pequenas setas em cada amostra da localização no lado esquerdo da Figura 21.

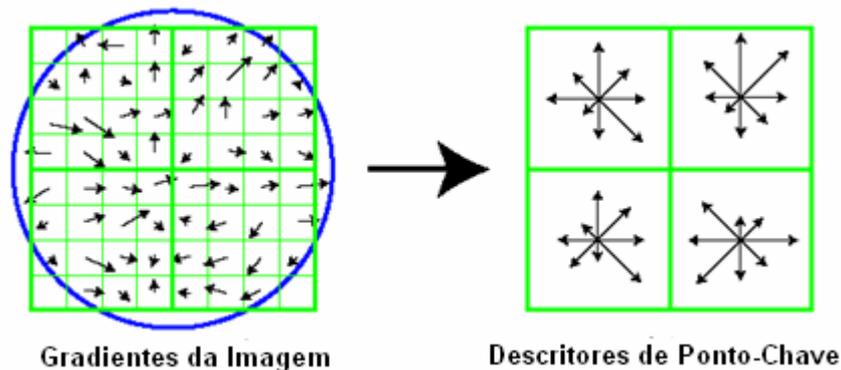


Figura 21: Ilustração da computação de descritores de pontos-chaves

Uma janela de suavização gaussiana de  $\sigma$  igual à metade da largura da janela do descritor é usada para dar pesos à magnitude do gradiente em cada ponto amostrado, conforme visto na esquerda da Figura 18. Esse gaussiano evita mudanças súbitas do descritor com pequenas mudanças na posição da janela e também reduz a ênfase nos gradientes longe do centro do descritor, que são mais afetados por erros.

Uma vez feita a suavização dos gradientes, criam-se histogramas de orientação sobre 4x4 regiões de amostragem. O lado direito da Figura 18 mostra os quatro histogramas formados pela amostragem da esquerda. Na figura estão representadas oito direções para cada histograma e o tamanho de cada seta representa a magnitude referente a respectiva entrada do histograma.

Para evitar efeitos de borda na qual o descritor abruptamente muda enquanto a amostra se desloca suavemente indo de um histograma para outro ou de uma orientação para outra, uma interpolação trilinear é usada para distribuir o valor de cada amostra de gradiente para a sua posição no histograma adjacente.

O descritor é formado de um vetor contendo os valores de todas as orientações dos histogramas, correspondentes aos tamanhos das setas no lado direito da Figura 18.

### 3.4.5.

#### **Invariância à Luminosidade**

Duas imagens de um mesmo objeto podem possuir variações de luminosidade que modifiquem os descritores obtidos sensivelmente. Devido a este motivo deve-se fazer um tratamento dos descritores para torná-los menos sensíveis a essas variações.

Os descritores são invariantes a mudanças homogêneas de brilho da imagem, uma vez que esta variação representa uma adição de todos os pixels da imagem por uma constante e os descritores são calculados por diferenças de pixels. Quanto a mudanças homogêneas de contraste, representada pela multiplicação de todos os pixels por uma constante, é corrigida normalizando os descritores.

Variações não-lineares, causadas quer por saturação das câmeras, quer por efeito de iluminação de superfícies tridimensionais em diferentes orientações, podem provocar elevada influência sobre as magnitudes dos descritores, mas pouco influenciam na orientação. Reduz-se este efeito impondo um valor máximo às magnitudes. Assim sendo, reduz-se a influência da magnitude na correspondência dos pontos em imagens distintas e aumenta-se a importância dos descritores de orientação.

### 3.4.6.

#### **Relacionamento de Pontos-Chaves em Imagens Distintas**

Uma vez aplicado o algoritmo SIFT sobre duas imagens distintas do mesmo objeto, o algoritmo retornará um conjunto de pontos-chaves em cada imagem e seus respectivos descritores. Se faz necessário relacionar os pontos de forma robusta. Diversos métodos possíveis são buscar pontos com menor distância euclidiana, distância de *Mahalanobis*, máxima probabilidade, entre outros. O procedimento escolhido consiste no uso de transformada Hough (Ballard D. , 1982) para agrupar os pontos a partir de suas propriedades (localização, orientação e descritores) em um espaço de transformação 2D, e em seguida utilizar o algoritmo RANSAC (Random Sample Consensus) (Fischler M. A., 1981) para melhorar as estimativas.

A transformada de Hough é um método usualmente utilizado para detectar linhas ou curvas especificadas por uma função paramétrica. Ela cria uma matriz acumuladora  $n$ -dimensional onde acumulam-se “votos” para os parâmetros corretos das linhas e curvas encontradas na imagem.

O uso da transformada Hough para relacionar pontos é parecida. Cada ponto-chave especifica quatro parâmetros: localização 2D, escala e orientação. Sendo assim, cria-se uma matriz acumuladora (espaço de transformada 2D) em translações ( $t_1$ ,  $t_2$ ), log escala ( $\log s$ ) e rotação ( $\theta$ ). Discretizam-se os tamanhos dos eixos de com um oitavo do tamanho da imagem para translação, um oitavo de escala para escala e um oitavo de radianos para rotação.

A transformada Hough identifica agrupamentos de propriedades de pontos-chaves usando cada propriedade de um ponto de uma imagem para votar entre todos os outros pontos da outra imagem aquele que é mais consistente.

Os resultados da transformada Hough são refinados usando RANSAC. Este método tem a vantagem de ser altamente insensitivo a pontos anormalmente distantes da maioria dos pontos amostrados, mas falhará se este grupo de pontos anormalmente distantes for muito grande. Por isso é usado Hough antes para agrupar as propriedades.

### 3.5.

#### Reconstrução

O objetivo final do uso de câmeras na calibração do braço mecânico é de obter as posições de pontos identificados nas imagens nas coordenadas reais deles no mundo. Tendo-se esta posição dos pontos, é possível descobrir a posição do manipulador em relação à sua área de trabalho para que ele possa trabalhar automaticamente.

Este processo chama-se reconstrução e dentre diversos algoritmos de reconstrução presentes na literatura, o escolhido foi o algoritmo de Triangulação. Outros métodos como Estrutura do Movimento e Geometria Epipolar foram testados mas trouxeram poucos resultados positivos.

O método da triangulação é puramente geométrico. Tendo-se um par de câmeras calibradas e um conjunto de pontos  $p$  na primeira imagem corretamente relacionado a pontos  $p'$  na segunda imagem, e sendo a posição das câmeras  $O$  e

$O'$  respectivamente, encontra-se a posição do ponto no espaço 3D na interseção dos raios  $R=Op$  e  $R'=O'p'$ , conforme visto na Figura 22. Na prática esta interseção não ocorre devido a erros de calibração e de localização dos pontos. Para solucionar este problema se calcula o ponto médio do segmento que mais aproxima os dois raios  $R$  e  $R'$ .

$$\overrightarrow{OP_1} + \overrightarrow{P_1P_2} + \overrightarrow{P_2O'} = \overrightarrow{OO'} \quad \text{Eq. 3.4.13}$$

Expandindo a Equação 3.4.13 com a posição relativa das câmeras:

$$z\hat{p} + c(\hat{p}xR\hat{p}') + z'R\hat{p}' = t \quad \text{Eq. 3.4.14}$$

Onde  $R$  é a matriz de rotação entre as câmeras e  $t$  é o vetor de translação. Substituindo em 3.4.14 os parâmetros intrínsecos ( $K$  e  $K'$ ):

$$zK^{-1}p + c(K^{-1}pxRK'^{-1}p) + z'RK'^{-1}\hat{p}' = t \quad \text{Eq. 3.4.15}$$

Resolvendo o sistema da Equação 3.4.15 encontra-se os parâmetros  $z$  e  $c$ . O ponto  $P$  é obtido pela equação:

$$P = (zK^{-1}p + c(K^{-1}pxRK'^{-1}p))/2 \quad \text{Eq. 3.4.16}$$

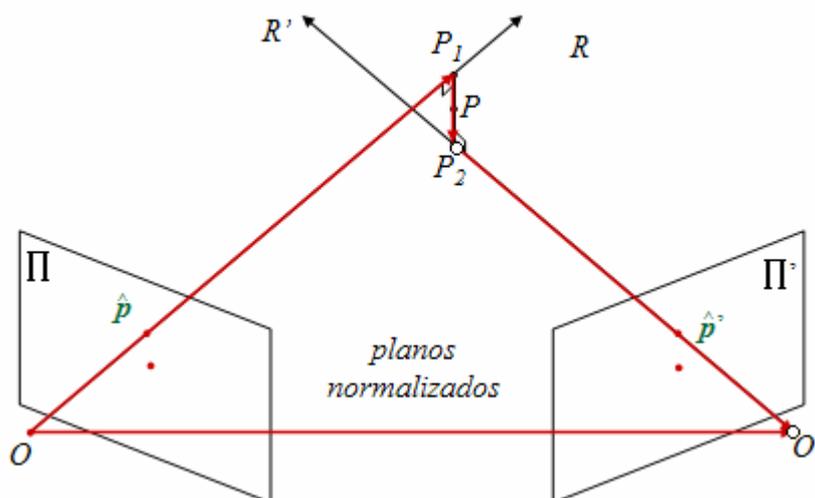


Figura 22: Triangulação