3
Transformação SIFT (Scale Invariant Feature Transform)

Este capítulo apresenta as seguintes seções:

- 3.1 – Uma Introdução Sobre Descritores Locais: A técnica SIFT ("Scale Invariant Feature Transform") é utilizada para se construir descritores. Nesta seção é apresentado o que são descritores e um panorama do uso de descritores em diversos problemas de visão computacional;
- 3.2 – Descrição do SIFT: Esta seção apresenta a técnica SIFT, apresenta trabalhos recentes que utilizam da técnica e, então, descreve sua teoria e como é aplicada. Esta seção é dividida em:
    - 3.2.1 – Introdução;
    - 3.2.2 – Detecção de Extremos;
    - 3.2.3 – Localização Exata de Pontos Chave;
    - 3.2.4 – Atribuição da Orientação dos Descritores;
    - 3.2.5 – Construção do Descritor Local;
    - 3.2.6 – Encontrando os Pontos em Comum;

3.1.
Uma Introdução Sobre Descritores Locais

Correspondência de imagens é fundamental em diversos problemas de visão computacional como reconhecimento de objetos, reconhecimento de cenas, montagem automática de mosaicos, obtenção da estrutura 3D de múltiplas imagens, correspondência estéreo e perseguição de movimentos. Uma abordagem para se trabalhar com correspondência de imagens é se usar descritores locais para se representar uma imagem. Descritores são vetores de características de uma imagem ou de determinadas regiões de uma imagem e podem ser usados para se comparar regiões em imagens diferentes. Este vetor de características é normalmente formado por descritores locais ou globais. Descritores locais computados em pontos de interesse provaram ser bem sucedidos em aplicações

como correspondência e reconhecimento de imagens [25, 31]. Descritores são distintos, robustos à oclusão e não requerem segmentação.

Existem diversas técnicas para se descrever regiões locais em uma imagem [31]. O mais simples descritor é um vetor com as intensidades dos *pixels* da imagem. A medida de correlação cruzada pode ser então usada para computar a similaridade entre duas regiões como apresentado na seção 2.6.2. Porém, a alta dimensionalidade de tal descritor aumenta a complexidade computacional da comparação. Então, esta técnica é principalmente usada para se encontrar correspondências ponto a ponto entre duas imagens. A vizinhança de um ponto também pode ser escalada de modo a reduzir sua dimensão. Outro descritor simples é a distribuição de intensidades de uma região representada por seu histograma.

Trabalhos recentes têm se concentrado em fazer descritores invariáveis a transformações nas imagens. Mikolajczyk e Schmid [30] propuseram um detector de pontos de interesse invariável a transformações afins através da combinação de um detector invariável à escala e da técnica "*second moment of Harris corners"* [32]. Ling e Jacobs [33] propuseram um sistema para a construção de descritores de intensidade locais invariáveis a deformações em geral. Em [34], Gotze, Drue e Hartmann apresentam descritores baseados na Transformada de Fourier-Mellin de regiões locais. Lowe [25-29] propôs uma maneira rápida e eficiente de computar características invariáveis a transformações em escala, que medem a distribuição do gradiente em regiões detectadas invariáveis à escala.

A seção a seguir apresentará os descritores SIFT e detalhará seu uso.

3.2. Descrição da técnica SIFT

3.2.1. Introdução

SIFT ("*Scale Invariant Feature Transform"*) é uma técnica de processamento de imagens que permite a detecção e extração de descritores locais, razoavelmente invariáveis a mudanças de iluminação, ruído de imagem, rotação, escala e pequenas mudanças de perspectiva. Estes descritores podem ser

utilizados para se fazer a correspondência de diferentes visões de um objeto ou cena.

Descritores obtidos com a técnica SIFT são altamente distintos, ou seja, um determinado ponto pode ser corretamente encontrado com alta probabilidade em um banco de dados extenso com descritores para diversas imagens.

Um aspecto importante da técnica SIFT é a geração de um número grande de descritores que conseguem cobrir densamente uma imagem quanto a escalas e localizações. A quantidade de descritores é particularmente importante para o reconhecimento de objeto, onde a capacidade de se encontrar pequenos objetos em ambientes desordenados requer ao menos 3 pontos encontrados em comum para uma identificação confiável.

A obtenção de descritores SIFT é feita através das seguintes etapas:

- Detecção de extremos: Nesta primeira etapa é feita procura para todas escalas e localizações de uma imagem. Isto é feito utilizando-se a diferença de filtros gaussianos de modo a se identificar pontos de interesse invariáveis à escala e rotação. A detecção de extremos é descrita na seção 3.2.2;
- Localização de pontos chave: Para cada localização em que foi detectado um extremo, um modelo detalhado é ajustado de modo a se determinar localização e escala. Pontos chaves, ou pontos de interesse, são então selecionados baseando-se em medidas de estabilidade. A localização dos pontos chaves é apresentada na seção 3.2.3;
- Definição de orientação: É definida a orientação de cada ponto chave através dos gradientes locais da imagem. Toda operação a partir de então será feita com relação a dados da imagem transformados em relação à orientação, escala e localização de cada ponto chave. Desta maneira se obtém invariância a estas transformações. A atribuição da orientação de cada descritor é vista na seção 3.2.4;
- Descritor dos pontos chaves: Nesta etapa é feita a construção dos descritores ao se medir Gradientes locais em uma região vizinha a cada ponto de interesse. Estas medidas são então transformadas para uma representação que permite níveis significativos de distorção e

mudança na iluminação. A construção dos descritores é apresentada em 3.2.5;

Em tarefas de comparação de imagens e reconhecimento, descritores SIFT são extraídos das imagens para então poderem ser comparados.

Na próxima seção será descrito o primeiro passo da obtenção de descritores SIFT.

3.2.2. Detecção de Extremos

A primeira etapa da técnica SIFT é detectar extremos (máximos e mínimos) em uma pirâmide da imagem convoluída com a função Diferença de Gaussiana. Pontos chave correspondem a estes extremos para diferentes escalas. Esta etapa está descrita na presente seção.

A convolução de uma função *f(x,y)* com uma função *h(x,y)* é dada por:

$$f(x,y) * h(x,y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) h(x-m, y-n) \tag{99}$$

Onde *x* varia de 1 a M e *y* varia de 1 a N.

Um filtro Gaussiano passa baixa é dado pela convolução de uma imagem *I* com a função *G*:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{100}$$

Onde:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \tag{101}$$

Perceba que este filtro é variável à escala através do parâmetro σ.

A função DoG (*"Difference of Gaussian"*) é dada pela diferença de imagens filtradas em escalas próximas separadas por uma constante *k*. A função DoG é definida por:

$$\text{DoG} = G(x,y,k\sigma) - G(x,y,\sigma) \tag{102}$$

O resultado de fazer a convolução de uma imagem com o filtro DoG é dado por:

$$\begin{aligned} D(x,y,\sigma) &= \left(G(x,y,k\sigma) - G(x,y,\sigma)\right) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \tag{103}$$

ou seja, é a diferença entre imagens borradas por um filtro gaussiano em escalas σ e $k\sigma$. Este filtro consegue detectar variações de intensidade na imagem, tais como contornos. Perceba que variando-se σ, é possível encontrar descritores para variações em diferentes escalas espaciais.

Nas Figura 3-1 e Figura 3-2 podem-se ver alguns exemplos de aplicação dos filtros descritos.

Figura 3-1: Imagem após filtro gaussiano com σ igual a 1.6, 2.4 e 3.2

Figura 3-2: Filtro DoG para imagens apresentadas na Figura 3-1

Um modo eficiente de se construir $D(x,y,\sigma)$ é apresentado na Figura 3-3. Deseja-se construir s intervalos, onde cada intervalo representa uma imagem filtrada por DoG intervalar entre duas outras. Para se construir s intervalos serão

necessárias s+3 imagens na pilha apresentada pelas imagens superiores da Figura 3-3. A imagem inicial é convoluída progressivamente com funções gaussianas para produzir mais $s+2$ imagens separadas por um fator constante k. A imagem é inicialmente filtrada por filtro Gaussiano com escala σ. A partir de então, são geradas imagens que são progressivamente convoluídas. Cada nova imagem é filtrada com escala k vezes a escala utilizada anteriormente. Para cada duas imagens, pode-se produzir a diferença de Gaussianas D através da subtração de duas imagens consecutivas na pilha de imagens L.

Para exemplificar, imagine que se deseja gerar apenas um intervalo. Serão necessárias, então, quatro imagens na pilha superior (a pilha das imagens filtradas com Gaussianas). Estas imagens serão:

- A imagem original;
- A imagem original filtrada por Gaussiana com escala σ;
- Outras duas imagens filtradas com escalas multiplicadas por k: kσ e k2σ;

Lowe considera em [25] que é necessário fazer a convolução da imagem até 2σ para ser possível a construção de descritores invariáveis à escala. Portanto, para se gerar s intervalos é definido:

$$k = 2^{1/s} \tag{104}$$

Desta maneira, teremos s intervalos produzidas por DoG, sendo que o primeiro é dado por $D(x,y,\sigma)$ e a ultima imagem da pilha de DoG dada por $D(x,y,2\sigma)$.

Para melhor entendimento, perceba que na Figura 3-3, a primeira imagem acima à esquerda é $I(x,y)$ e a última imagem acima é $L(x,y,k2\sigma)$. A primeira imagem abaixo é $D(x,y,0)$ e a última é $D(x,y,2\sigma)$. O intervalo é dado por $D(x,y,\sigma)$.

O processo apresentado gera o que é chamado de uma oitava. Este processo é repetido para um número desejado de oitavas. Cada oitava representa um conjunto de imagens L e D para a imagem reescalada com diferentes amostragens.

Isto funciona da seguinte forma: quando uma oitava tiver sido processada, a imagem Gaussiana que possui 2σ (corresponde à penúltima imagem da pilha superior na Figura 3-3) é re-amostrada para a metade de seu tamanho. Esta será a

primeira imagem da próxima oitava. Cada oitava produz o mesmo número de intervalos.

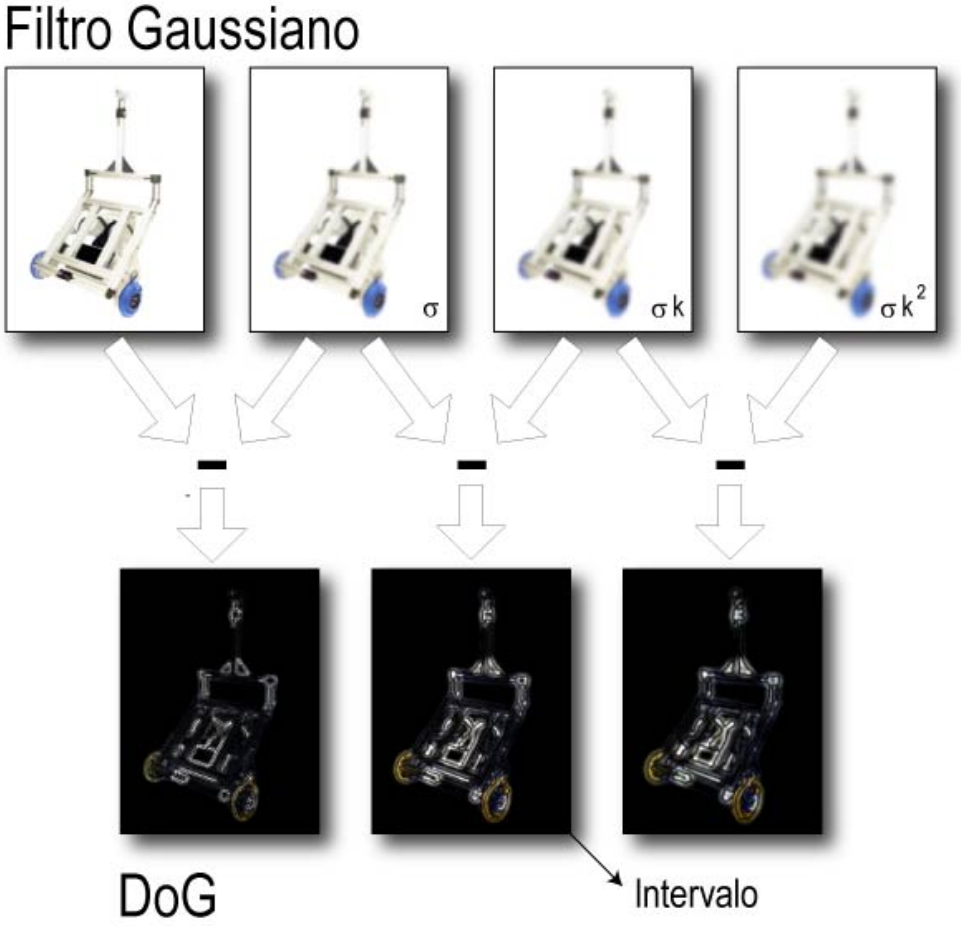

Figura 3-3: Construção das Diferenças de Gaussiana

A geração de oitavas está exemplificada na Figura 3-4.

A partir de agora será feita a detecção de extremos em cada intervalo de cada oitava. Os extremos são dados por valores locais de máximo ou mínimo para cada $D(x, y, \sigma)$ que corresponda a um intervalo. Cada ponto é comparado aos seus oito vizinhos na imagem atual, mais seus nove vizinhos na escala superior e nove vizinhos na escala inferior.

As escalas superior e inferior são correspondentes às imagens vizinhas em uma mesma oitava para a pilha de imagens DoG. Não confunda escala superior e inferior com oitavas onde a amostragem das imagens gera imagens em escalas diferentes. Quando se diz escala superior e inferior aqui, está se fazendo referência à σ.

O procedimento de detecção está exemplificado na Figura 3-5. No exemplo, o ponto marcado como *X* é comparado com seus vizinhos marcados como *O*. As 3 imagens DoG apresentadas são 3 intervalos vizinhos em uma pilha.

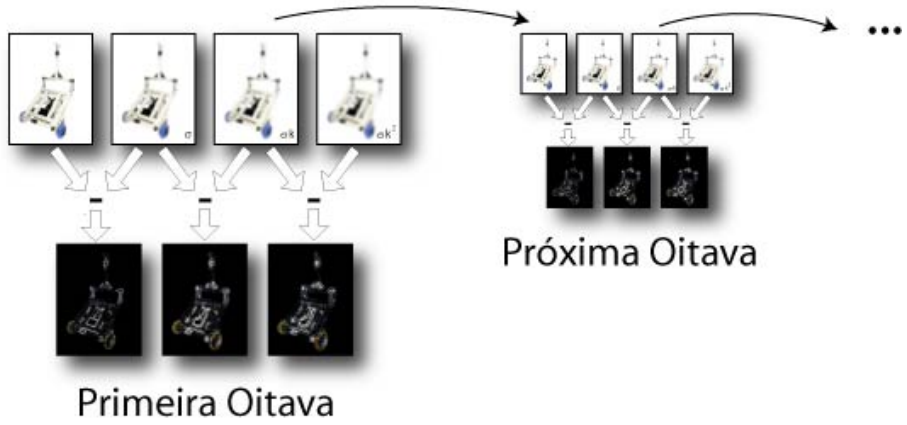

Figura 3-4: Formação das Oitavas

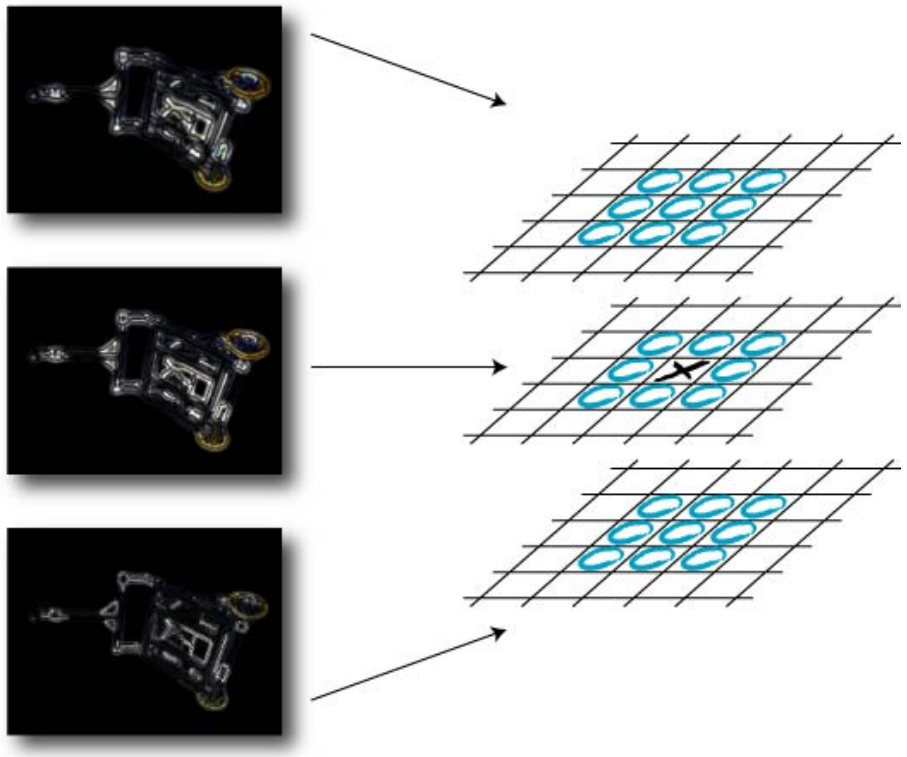

Figura 3-5: Detecção de Máximos e Mínimos

A próxima etapa é definir a localização exata dos pontos chave e fazer o descarte de pontos chave instáveis. Isto será visto na próxima seção.

3.2.3.
Localização Exata de Pontos Chave

Todos os pontos detectados como extremos são possíveis pontos chave. Deseja-se agora calcular a localização e escala Gaussiana detalhadas de cada um destes pontos. Recalcular a localização e escala interpoladas dos pontos de máximo traz melhoria para à técnica. A localização dos pontos chaves como será apresentada é especialmente importante para as últimas oitavas, porque o espaçamento amostral destas representa grandes distâncias na imagem base.

O método consiste em enquadrar uma função quadrática 3D do ponto de amostragem local de modo a determinar uma localização interpolada do máximo.

Para cada ponto analisado é utilizada uma expansão de Taylor da função $D(x,y,\sigma)$ transladada de modo que a origem desta expansão esteja localizada no ponto:

$$D(\bar{x}) = D + \frac{\partial D^T}{\partial \bar{x}}\bar{x} + \frac{1}{2}\bar{x}^T \frac{\partial^2 D}{\partial \bar{x}^2}\bar{x}\ldots \tag{105}$$

Onde:

$$D = D(x,y,\sigma) \tag{106}$$

$$D(\bar{x}) = D(x+x',y+y',\sigma+\sigma') \tag{107}$$

Esta equação deve ser entendida da seguinte maneira, D e suas derivadas são avaliados a partir do ponto analisado e $\bar{x} = (x',y',\sigma')^T$ é o *offset* em relação a este ponto. Ou seja, D é o valor da função D(x,y,σ) no ponto avaliado, $\bar{x}$ é o *offset* em relação a este ponto e $D(\bar{x})$ é a aproximação do valor de D(x,y,σ) interpolado para um ponto transladado com *offset* $\bar{x}$.

Os coeficientes quadráticos são computados aproximando-se as derivadas através das diferenças entre *pixels* das imagens já filtradas.

A localização *sub-pixel* / sub-escala do ponto de interesse é dada pelo extremo da função apresentada na eq. (105). Esta localização, $\hat{x}$, é determinada ao se fazer a derivada segunda da eq. (105) com relação a *x* e igualando o resultado a zero. Isto é feito como a seguir:

$$\frac{\partial D(\hat{x})}{\partial \bar{x}} = \frac{\partial D^T}{\partial \bar{x}} + \frac{\partial D}{\partial \bar{x}^2}\hat{x} = 0 \tag{108}$$

Perceba que esta derivada usa a expansão de Taylor até $\frac{\partial D}{\partial \bar{x}}$. Tem-se então a posição do extremo dada por:

$$\hat{x} = -\frac{\partial^2 D^{T-1}}{\partial \bar{x}^2} \frac{\partial D}{\partial \bar{x}} \tag{109}$$

O resultado é um sistema linear 3x3 que pode ser resolvido com custo mínimo. Caso $\hat{x}$ seja maior que 0.5 em alguma dimensão, isto significa que o extremo se aproxima mais de outro ponto. Neste caso, o ponto é re-alocado e a interpolação é realizada para este novo ponto. O *offset* $\hat{x}$ final é adicionado à localização do ponto analisado para se chegar à interpolação estimada da localização do extremo.

A localização estimada deverá ser usada a partir de então nos procedimentos que seguirão.

O valor da função no extremo, $D(\hat{x})$, é utilizado para se rejeitar extremos instáveis com baixo contraste. Substituindo-se a eq. (109) na eq. (105) obtemos:

$$D(\hat{x}) = D + \frac{1}{2} \frac{\delta D^T}{\delta \bar{x}} \hat{x} \tag{110}$$

É aconselhável por Lowe que se rejeite valores de $|D(\hat{x})|$ inferiores a um determinado valor. Em [25] é aconselhado trabalhar-se com o valor 0.03 (assumindo-se que os *pixels* da imagem estejam entre [0,1]).

Aqui não é refinada a posição como apresentado, porém são descartados valores de $|D(x,y,\sigma)|$ inferiores a determinado limiar.

Alem do procedimento apresentado para se descartar pontos, Lowe ainda aponta que a função DoG possui resposta forte ao longo de arestas, mesmo que a localização ao longo da borda seja mal determinada. Isto faz com que estes pontos sejam instáveis para ruído em até pequenas quantias.

Um pico mal definido em DoG terá grande curvatura principal ao longo da borda, porém pequena curvatura em sua direção perpendicular. As curvaturas principais podem ser computadas através da matriz Hessiana 2x2, *H*, computada na localização e escala do ponto:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{xx} \end{bmatrix} \tag{111}$$

Onde:

$$D_{ab} = \delta_{ab} D = \frac{\partial\left(\frac{\partial D(\hat{x})}{\partial b}\right)}{\partial a} \tag{112}$$

Onde D_{xy} é a derivada de $D(x,y,\sigma)$ na localização e escala dados, em relação a x, e então a y; D_{xx} é a derivada segunda em relação a x; D_{yy} é a derivada segunda em relação a y;

As derivadas são estimadas através das diferenças entre pontos vizinhos à localização e escala definidos, podendo ser aproximada por:

$$D_{xx} = D(x+1,y,\sigma) - 2D(x,y,\sigma) + D(x-1,y,\sigma) \tag{113}$$

$$D_{yy} = D(x,y+1,\sigma) - 2D(x,y,\sigma) + D(x,y-1,\sigma) \tag{114}$$

$$D_{xy} = \begin{pmatrix} D(x-1,y+1,\sigma) - D(x+1,y+1,\sigma) \\ +D(x+1,y-1,\sigma) - D(x-1,y-1,\sigma) \end{pmatrix} \Big/ 4 \tag{115}$$

Os autovalores de H são proporcionais às principais curvaturas de D. Porém, não será necessário computar os autovalores pois o que se busca é a razão entre as curvaturas. Determina-se α, o autovalor com maior magnitude, e β, o de menor. Pode-se, então, calcular a soma dos autovalores pelo traço de H e o produto pelo determinante:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \tag{116}$$

$$Det(H) = D_{xx} D_{yy} - (D_{xy})^2 = \alpha\beta \tag{117}$$

Para o caso em que o determinante é negativo, as curvaturas possuem sinais diferentes e então o ponto é descartado como não sendo um extremo. Sendo r a razão entre o autovalor de maior magnitude e o de menor, de modo que $\alpha = r\beta$, então:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \tag{118}$$

A eq. (118) depende apenas da razão entre os autovalores, independente de seus valores individuais. O valor $(r+1)^2/r$ é mínimo quando os dois autovalores são idênticos e cresce com r. Portanto, para se conferir a razão entre as curvaturas está abaixo de determinado limiar $\tilde{r}$, basta checar:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(\tilde{r}+1)^2}{\tilde{r}} \tag{119}$$

A eq. (119) é altamente eficiente de ser computada. Lowe propõe o uso de $r = 10$.

Após a definição de quais pontos serão usados como pontos chaves ainda serão feitas duas etapas na construção dos descritores. A primeira etapa consiste em estabelecer uma ou mais orientações para cada ponto chave. Então, é feita a construção dos descritores com relação às orientações definidas.

A definição das orientações de cada ponto chave é descrita na próxima seção.

3.2.4. Atribuição da Orientação dos Descritores

Ao se atribuir uma orientação para cada ponto chave, pode-se representar os descritores em relação a esta orientação, conseguindo-se assim invariância quanto à rotação. O método utilizado para se atribuir esta orientação é apresentado como se segue.

A escala Gaussiana σ é utilizada para se escolher a imagem filtrada L, com a escala mais próxima, e de oitava referente ao ponto avaliado. Dessa maneira, todas os cálculos passam a ser feitos com invariância à escala.

Para cada ponto de cada imagem $L(x,y,\sigma)$ intervalar, referente às escalas e oitavas utilizadas, são calculados os gradientes. Magnitude $m(x,y)$ e orientação $\theta(x,y)$ são calculados como se segue:

$$m(x,y) = \sqrt{\begin{pmatrix} (L(x+1,y) - L(x-1,y))^2 \\ +(L(x,y+1) - L(x,y-1))^2 \end{pmatrix}} \tag{120}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{(L(x,y+1) - L(x,y-1))}{(L(x+1,y) - L(x-11,y))}\right) \tag{121}$$

Observe que σ não aparece nas equações. Isto foi feito para simplificar, pois o processamento é feito para cada imagem L. Somente as imagens correspondentes a intervalos precisam ser processadas.

Agora, monta-se um histograma das orientações para *pixels* em uma região ao redor do ponto chave. O histograma é uma função discreta h_θ um determinado número de valores discretos de θ (Lowe sugere 36) cobrindo os 360^o de orientações.

Cada ponto na vizinhança do ponto chave é adicionado ao histograma para até dois θ's discretos mais próximos de sua orientação com uma serie de pesos.

O primeiro peso é dado pela distância entre a orientação e θ discreto normalizada pelas distâncias entre θ's discretos. Este peso é dado por:

$$\alpha = \begin{cases} d/i, d < i \\ 0, d > i \end{cases} \tag{122}$$

Onde *d* é a distância absoluta em graus entre a orientação do ponto e θ discreto, e *i* é o intervalo em graus entre θ's discretos.

Por exemplo, para h_θ com θ dado por 0º,10 º,20 º...350 º, ou seja, com intervalos de 10º, um ponto com orientação de 15 º seria acrescentado em h_{10} e h_{20}. Como a distância da orientação para 10 º e 20 º é de 5 º, o peso utilizado para adicionar este ponto em para h_{10} e h_{20} é dado por 5/10, ou seja, a distância sobre o intervalo entre θ's para h_θ.

O segundo peso é dado pela magnitude *m(x,y)* de cada ponto adicionado a h_θ.

O último peso é dado por uma janela gaussiana circular com σ'com valor 1.5 vezes maior que a escala σ do ponto chave. Esta janela é definida pela função gaussiana:

$$g(\Delta x, \Delta y, \sigma') = \frac{1}{2\pi\sigma'^2} e^{-\left(\Delta x^2 + \Delta y^2\right)/2\sigma'^2} \tag{123}$$

E,

$$\sigma' = \sigma \tag{124}$$

Onde Δx e Δy são as distâncias entre cada ponto verificado e o ponto chave.

Por fim, h_θ é atualizado com estes pesos, para cada ponto na vizinhança localizado em *(x,y)*, da seguinte forma:

$$h_\theta' = h_\theta + \alpha \cdot m(x,y) \cdot g(\Delta x, \Delta y, \sigma') \tag{125}$$

Onde h_θ' é a atualização de h_θ.

Perceba que não é necessário fazer a atualização de h_θ para todos os pontos da imagem porque a função $g(\Delta x, \Delta y, \sigma')$ retorna valores muito baixos (aproximadamente zero) para a grande maioria dos pontos.

Picos na orientação do histograma correspondem a direções dominantes para os gradientes locais. O maior pico no histograma e aqueles acima de 80% do

valor do maior pico são usados para se definir a orientação de cada ponto chave. Portanto, para localizações com múltiplos picos de magnitude similar, são criados diferentes pontos chaves na mesma localização, mas com diferentes orientações.

Para se definir com maior precisão à orientação, uma parábola é interpolada entre os 3 valores do histograma próximos de cada pico, e então é interpolada a posição do pico.

Finalmente é possível construir os descritores para os pontos chaves definidos. Isto será abordado na seção a seguir.

3.2.5. Construção do Descritor Local

Até então, para cada oitava, foram escolhidos pontos chaves para localizações, escala σ e orientação definidos. A etapa atual consiste em computar o descritor que represente as regiões relativas aos pontos chaves. Não se esqueça que os procedimentos a seguir serão feitos normalizados em relação à orientação definida na seção anterior para cada ponto chave.

Também serão utilizados os gradientes das imagens *L* já calculados na etapa descrita na seção 3.2.4.

Para cada ponto chave, a construção do descritor é feita através dos seguintes passos:

- Escolhe-se a imagem filtrada L referente à escala σ e oitava relativas ao ponto chave;
- De modo a se conseguir invariância, as coordenadas dos pontos vizinhos ao descritor e das orientações dos gradientes destes pontos são giradas em relação ao ponto chave de acordo com a orientação definida na seção anterior;
- Uma função gaussiana, como apresentada na eq. (123), é utilizada como peso para se ajustar as magnitudes de cada ponto na vizinhança do ponto chave. σ' é escolhido igual a metade da largura da janela em que será calculado o descritor;
- São definidas n x n regiões, com k x k *pixels* cada, ao redor da localização do ponto chave. Geralmente n = k = 4 como exemplificado na Figura 3-6;

- Para cada região, é feito um histograma hθ. para 8 direções, como na Figura 3-7. Este histograma é feito com as magnitudes dos *pixels* pertencentes a cada região. A construção do histograma é similar à apresentada na seção 3.2.4. O peso referente à magnitude de cada *pixel* foi atenuado pela função gaussiana como já ajustado. Perceba que a função gaussiana não é aplicada de modo idêntico ao na seção anterior. Também como feito em 3.2.4, é utilizado um peso α para interpolar a direção relativa no histograma.
- O descritor é então representado pelos histogramas das regiões. A Figura 3-8 exemplifica como fica o descritor para 2 x 2 regiões (n = 2 e k = 4);
- O descritor é representado por um vetor, onde cada valor do vetor é referente a uma das direções de um dos histogramas. Para n e k iguais a 4, o vetor tem tamanho 128.
- Para que o descritor tenha invariância à iluminação, este é normalizado. Após a normalização, todos os valores acima de um determinado limiar são ajustados para este limiar. Isto é feito para que direções com magnitude muito grande não dominem a representação do descritor. Lowe sugere usar limiar 0.2. Por fim, o vetor é normalizado novamente.

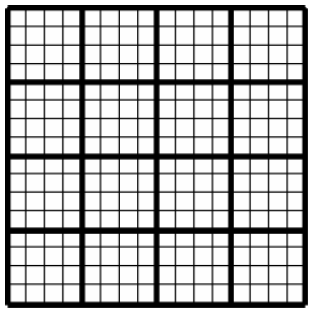

Figura 3-6: Regiões com *n* = 4 e *k* = 4;

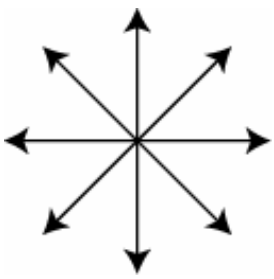

Figura 3-7: Direções do histograma

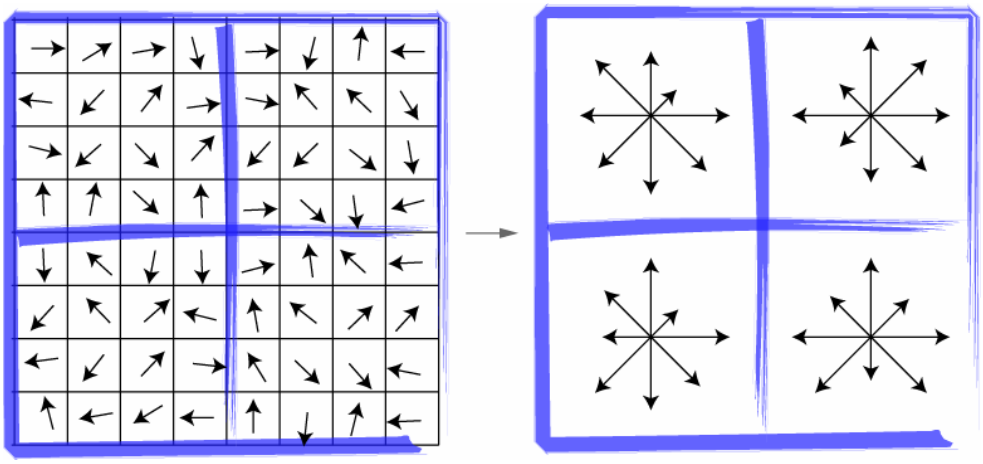

Figura 3-8: Construção do descritor

O descritor está construído. Para cada imagem, são construídos diversos descritores, cada um referente a um ponto chave. Quando se aplica a técnica SIFT em uma imagem, tem-se como resultado, portanto, um conjunto de descritores. Estes descritores podem ser, então, usados para se fazer a correspondência da imagem em outra imagem como será visto na próxima seção.

3.2.6. Encontrando os Pontos em Comum

Para se encontrar a correspondência entre duas imagens, deve-se encontrar pontos em comum entre as duas. Quando se trabalha com a técnica SIFT, pontos de interesse são detectados pelo método e representados em descritores. Tendo-se descritores de duas imagens, a tarefa de se encontrar a correspondência de uma imagem em outra é resumida por se encontrar entre os descritores de uma imagem, os melhores candidatos a serem seus equivalentes na outra imagem. Portanto, dadas duas imagens I_1 e I_2, a tarefa de se encontrar a correspondência de I_1 em I_2 pode ser definida como se segue.

Os descritores são respectivamente definidos por di_1 e dj_2, onde i e j são aos índices para cada um dos descritores de cada imagem e k é o tamanho de cada descritor:

$$di_1 = (m_1 i_1, m_1 i_2, m_1 i_3 \cdots, m_1 i_k) \tag{126}$$

$$dj_2 = (m_2 j_1, m_2 j_2, m_2 j_3 \cdots, m_2 j_k) \tag{127}$$

A magnitude de cada valor dos vetores di_1 e dj_2 é dada por $m_a i_b$, onde a representa a qual imagem se refere o descritor, i é o índice do descritor e b é o índice de cada magnitude dentro do vetor.

A correspondência é feita achando-se os descritores dj_2 que mais se assemelham aos descritores di_1, encontrando-se as falsas equivalências e eliminando-as e por fim, encontrando-se a transformação de I_1 para I_2.

A tarefa de se encontrar o melhor candidato dj_2 para determinado di_1 é feita procurando-se o vizinho mais próximo ou *"nearest neighbor"* de di_1 entre todos os possíveis candidatos, ou seja, para todo o índice j. Quando se procura classificar uma imagem em um extenso banco de dados de descritores para vários objetos, a busca exaustiva de vizinho mais próximo pode ser demorada e para tal existem diversas técnicas para se acelerar a busca. Porém, para o caso de se comparar duas imagens, a busca exaustiva não exige processamento pesado e, portanto, foi a escolhida.

O vizinho mais próximo de di_1 para i dado é definido por dj_2 que possua a menor distância euclideana em relação à di_1. Ou seja, deseja se encontrar j que minimize a função:

$$|di_1 - dj_2| = \sqrt{\begin{pmatrix}(m_1 i_1 - m_2 j_1)^2 + (m_1 i_2 - m_2 j_2)^2 + \ldots \\ +(m_1 i_k - m_2 j_k)^2\end{pmatrix}} \tag{128}$$

Isto é feito para todo i de modo a serem encontrados todos os pares de descritores correspondentes. Perceba que muitos dos pares encontrados correspondem a falsas equivalências, portanto, as correspondências serão refinadas e falsos pares descartados.

O capítulo seguinte aborda métodos de registro e correspondências entre imagens, mostrando como utilizar as correspondências entre pontos encontradas com o método SIFT.