2
Trabalhos Relacionados

Este capítulo apresenta os algoritmos já existentes que são utilizados nesta dissertação para obter pontos homólogos entre duas imagens de um par estéreo. Pode-se classificar essas técnicas em duas categorias: os métodos baseados em área, apresentados na primeira parte, e os métodos baseados em feições, na segunda.

2.1
Métodos baseados em área

Os primeiros dos métodos para obter mapas de correspondência entre duas imagens são baseados em correspondências de áreas. De um modo geral, as duas vistas de um par estéreo de imagens aéreas ou de satélite são muito parecidas porque mostram as mesmas áreas, e têm poucas distorções uma em relação à outra. Faz então sentido comparar blocos de pixels diretamente extraídos das imagens por meio de correlação, já que eles apresentam similaridades. Esses métodos são os mais utilizados porque a localização dos pontos obtidos é mais exata, e também porque utilizados em conjunto com uma estratégia de crescimento de região, o algoritmo fornece uma nuvem densa de pontos homólogos. Entretanto, eles encontram dificuldades em regiões com mudanças abruptas de elevação e necessitam de uma boa solução inicial, que é normalmente fornecida por um operador humano.

2.1.1
Correlação por Mínimos Quadrados

2.1.1.1
Princípio da Correlação por Mínimos Quadrados

Sejam g_1 e g_2 duas imagens da mesma cena. Os métodos baseados em área tentam localizar a vizinhança centrada nas posições (x_1, y_1) e (x_2, y_2), respectivamente em g_1 e g_2, no qual a relação

$$g_2(x_2, y_2) = \alpha + \beta \cdot g_1(x_1, y_1) \tag{2.1}$$

seja válida no contexto dos mínimos quadrados para algum valor de α e $\beta \in \mathbb{R}$.

Este modelo é invariante quanto ao brilho (α) e ao contraste (β) entre as imagens. A correlação cruzada normalizada admite que (x_1, y_1) e (x_2, y_2), estão relacionados por uma mera translação, como apresentado na coluna esquerda da tabela 2.1. Entretanto, este modelo não se aplica nos casos em que a relação espacial das imagens do objeto não é a de uma simples translação.

Tabela 2.1: Comparação entre Correlação Cruzada Normalizada e Correlação por Mínimos Quadrados

Correlação Cruzada Normalizada	Correlação por Mínimos Quadrados
$\begin{cases} x_2 = x_1 + a \\ y_2 = y_1 + b \end{cases}$	$\begin{cases} x_2 = a + bx_1 + cy_1 \\ y_2 = d + ex_1 + fy_1 \end{cases}$

Um aperfeiçoamento no método de correlação cruzada normalizada é a correlação por mínimos quadrados, que leva em consideração as diferenças geométricas entre as imagens devidas à posição dos sensores, as quais são modeladas através de uma transformação afim aplicada à segunda imagem, como mostra a coluna da direita da tabela 2.1. Além disso, a correlação por mínimos quadrados leva em consideração as mudanças lineares de tons de cinza. Os oito parâmetros são, então, determinados através de ajustamento.

2.1.1.2
Ajustamento da Correlação por Mínimos Quadrados

Por causa de efeitos aleatórios (ruído, mudanças de vista…) entre as duas imagens, raramente se encontram blocos correspondentes $g_1(x_1, y_1)$ (na imagem de referência) e $g_2(x_2, y_2)$ (na imagem de procura) para os quais a relação (2.1) seja rigorosamente verificada. Deve-se adicionar um termo de ruído :

$$g_2(x_2, y_2) = \alpha + \beta . g_1(x_1, y_1) - e(x_1, y_1) \tag{2.2}$$

onde α, β, e $e(x_1, y_1) \in \mathbb{R}$ estão associados ao contraste, ao brilho, e ao ruído relativo entre as imagens. Estima-se a correspondência final entre os blocos minimizando esse ruído $e(x_1, y_1)$ (a função de custo da otimização).

Há 6 incógnitas (*a,b,c,d,e,f*) no modelo da tabela 2.1 que representa as deformações afins dos blocos. Deve-se fornecer ao algoritmo um ponto de partida para a busca na segunda imagem para determinar estas incógnitas. Assim, a otimização começa em uma posição inicial (x_2^0, y_2^0) na segunda imagem, denotada $g_{20} = g_2(x_2^0, y_2^0)$. A equação (2.2) pode ser linearizada aplicando-se a fórmula de Taylor, limitando-se aos termos de primeira ordem, e separando o erro à esquerda:

$$e(x_1, y_1) = \alpha + \beta . g_1(x_1, y_1) - g_{20} - \left.\frac{\partial g_2}{\partial x_2}\right|_{(x_2^0, y_2^0)} dx_2 - \left.\frac{\partial g_2}{\partial y_2}\right|_{(x_2^0, y_2^0)} dy_2 \tag{2.3}$$

O modelo da tabela 2.1 deve então ser diferenciado :

$$\begin{cases} dx_2 = da + db.x_1 + dc.y_1 \\ dy_2 = dd + de.x_1 + df.y_1 \end{cases} \tag{2.4}$$

E obtém-se então, com a notação simplificada $g_{2x0} = \left.\frac{\partial g_2}{\partial x_2}\right|_{(x_2^0, y_2^0)}$ e $g_{2y0} = \left.\frac{\partial g_2}{\partial y_2}\right|_{(x_2^0, y_2^0)}$:

$$e(x_1,y_1)=\begin{bmatrix}1 & g_1(x_1,y_1) & -g_{2x0} & -x_1g_{2x0} & -y_1g_{2x0} & -g_{2y0} & -x_1g_{2y0} & -y_1g_{2y0}\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\\ da\\ db\\ dc\\ dd\\ de\\ df\end{bmatrix}-g_{20}$$

(2.5)

A equação acima também pode ser formulada na notação tradicional de uma estimação por mínimos quadrados com os parâmetros reunidos no vetor $A=\begin{bmatrix}1 & g_1(x_1,y_1) & -g_{2x0} & -x_1g_{2x0} & -y_1g_{2x0} & -g_{2y0} & -x_1g_{2y0} & -y_1g_{2y0}\end{bmatrix}$ e as incógnitas reunidas no vetor $x=\begin{bmatrix}\alpha,\beta,da,db,dc,dd,de,df\end{bmatrix}^T$:

$$e=A.x-g_{20} \tag{2.6}$$

É possível resolver a estimação com os métodos clássicos para obter os 8 parâmetros da transformação. Deve-se minimizar a norma $\|A.x-g_{20}\|_2$. Não são necessárias muitas iterações, desde que os valores iniciais sejam bons, principalmente a posição inicial na imagem de busca (x_2^0,y_2^0), como explicado pelo Gruen [1].

2.1.1.3
Vantagens da Correlação por Mínimos Quadrados

A primeira, e a mais evidente, vantagem de utilizar uma transformação afim entre os blocos é poder compensar a distância sensor-imagem. Enquanto a correlação tradicional falha quando os objetos aparecem em tamanhos diferentes nas imagens, a versão por mínimos quadrados resolve o problema para pequenas variações de escala,

Uma segunda vantagem é ilustrada na figura 2.1, onde são apresentadas dois sensores (nos aviões ou satélites) capturando imagens de dois tipos de superfícies. A primeira à esquerda é plana e horizontal. Um objeto nessa superfície terá o mesmo tamanho em ambas as imagens (os planos dessas imagens são representados entre as setas e com a legenda). Ao contrário, à direita, o

mesmo objeto no solo inclinado terá uma imagem muito maior na imagem do sensor 2, cujo eixo ótico fica quase perpendicular ao solo, do que na foto do sensor 1, cujo eixo ótico é muito mais inclinado. No final, essa mudança no ponto de vista 3D resulta em uma transformação afim dos objetos nas imagens. Se a deformação do solo for regular e se a inclinação for moderada, a correlação por mínimos quadrados também poderá resolver esse problema de ponto de vista.

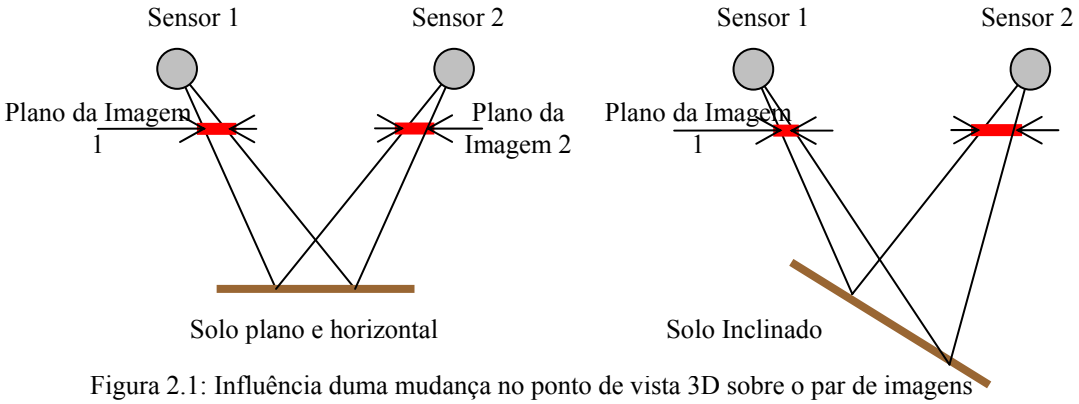

Figura 2.1: Influência duma mudança no ponto de vista 3D sobre o par de imagens

2.1.2
Crescimento de Região

2.1.2.1
Princípio do crescimento de região

O procedimento de crescimento de região [11] começa a partir de um par de pontos homólogos, chamados de sementes, que são normalmente identificadas por um operador humano. Um método baseado em área é então aplicado para determinar com maior exatidão a localização do ponto homólogo na segunda imagem. Dependendo do grau de similaridade (limiar na correlação), os pontos homólogos encontrados são mantidos ou descartados. Se um ponto é mantido, quatro novos pares de pontos são gerados a uma distância de *d* pixels acima, abaixo, à esquerda e à direta, a partir do último ponto localizado, onde *d* é um parâmetro pré-determinado que define o espaçamento entre a localização atual e os novos pontos. Esses pontos homólogos tornam-se agora novas sementes e suas posições também são refinadas por um método baseado em área. O procedimento é recursivamente repetido, espalhando novas sementes em ambas as imagens, assim fornecendo um conjunto denso de pontos homólogos.

2.1.2.2
Exemplos de crescimentos

A figura 2.2 apresenta um primeiro exemplo do resultado produzido utilizando-se a correspondência por mínimos quadrados com crescimento de região, partindo-se de uma semente medida manualmente (grande X branco nas duas imagens do par estéreo). Os pontos encontrados são assinalados pelas pequenas cruzes cinza. À esquerda, foram utilizados blocos de tamanho 10 × 10 pixels para o cálculo das correlações, deslocados de 3 em 3 pixels tanto para as colunas como para as linhas. À direta, a grade não é mais constante já que a posição dos pontos é ajustada pela correlação por mínimos quadrados. A cobertura obtida nessas imagens ficou limitada, porque o limiar de correlação abaixo do qual os pontos homólogos são descartados foi escolhido bastante alto (0.7 aqui). Esse limiar depende de cada imagem e tem que ser ajustado manualmente. A figura 2.3 apresenta imagens da mesma região geográfica, com as mesmas sementes, onde o crescimento de região se espalhou mais graças a um limiar de correlação mais baixo. Contudo, ainda sobram algumas regiões não atingidas.

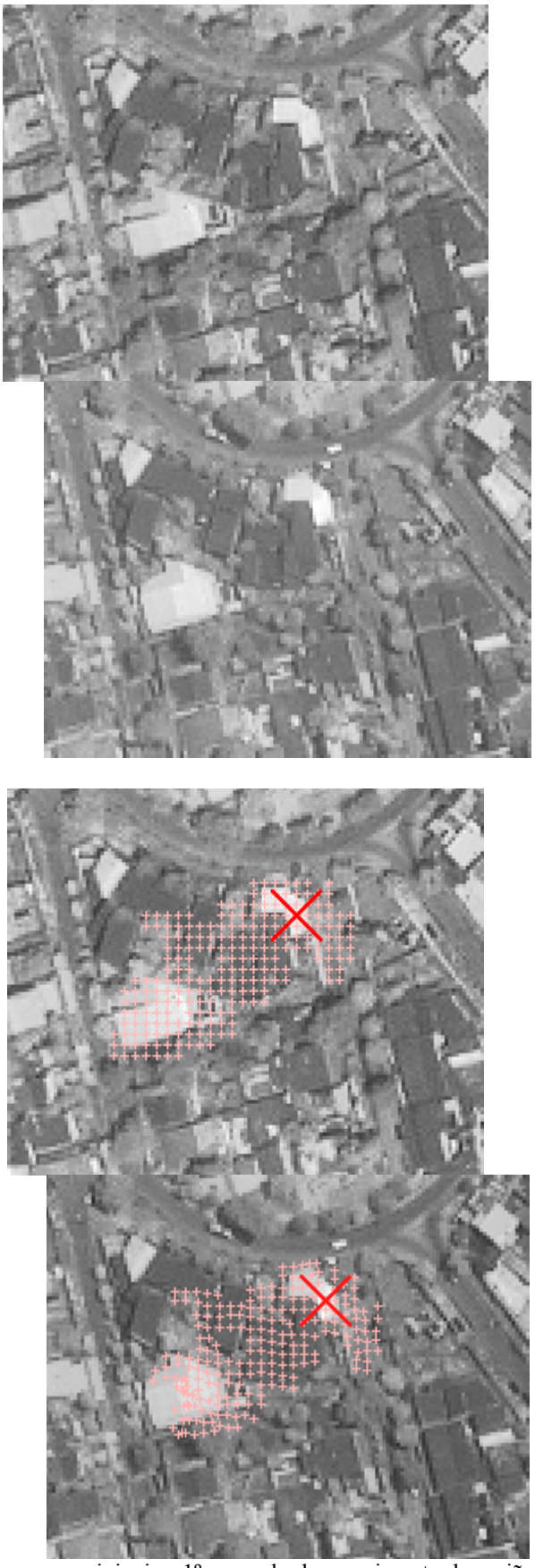

Figura 2.2: Par de imagens originais e 1º exemplo do crescimento de região com cobertura limitada

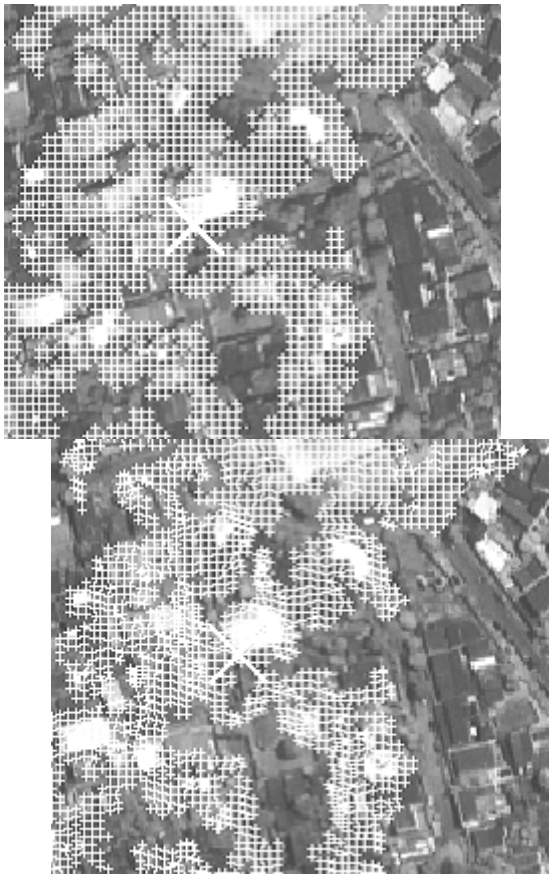

Figura 2.3: Segundo exemplo do crescimento de região com uma cobertura maior

2.1.3
Limitações dos métodos baseados em área

Mesmo se forem utilizadas as configurações mais favoráveis para a correlação por mínimos quadrados junto com o crescimento de região, ainda sobram algumas áreas das imagens onde o algoritmo não consegue encontrar pontos homólogos. É possível classificar essas falhas em duas categorias: por falta de detalhes na imagem, ou por causa de uma má inicialização.

2.1.3.1
Falha por falta de detalhes na imagem

Como se pode ver nas figuras 2.2 e 2.3, o algoritmo não conseguiu atravessar regiões onde os detalhes das imagens foram insuficientes para fornecer uma boa correlação. Encontram-se, por exemplo, problemas em florestas, na água

(mar e lagoas), em regiões com muita sombra ou em objetos imageados de um ponto de vista muito diferente. Na maioria desses casos, nem mesmo o olho humano consegue localizar os pontos homólogos, não havendo alternativa senão contornar essas áreas ou fornecer novas sementes em outra posição para a partir dali se retomar o processo de crescimento.

2.1.3.2
Falha por causa de uma má inicialização

Na figura 2.4 observa-se outra limitação da correlação por mínimos quadrados com o crescimento de região. Note-se que nenhum ponto foi encontrado ao redor dos prédios altos. De um modo geral, o crescimento de região pára sobre regiões onde há grande variação de altura, e, portanto, grandes diferenças nas imagens causadas por oclusão e pelas diferentes vistas de fachadas. Como ilustrado na figura 2.5, um prédio alto não mostra as mesmas fachadas nas duas imagens de um par estéreo, mesmo o topo aparecendo em ambas as imagens, o ângulo com o qual suas arestas são visualizadas é bem diferente de um lado para o outro, dificultando a correspondência.

Figura 2.4: Crescimento de região falhando em topos de prédios altos

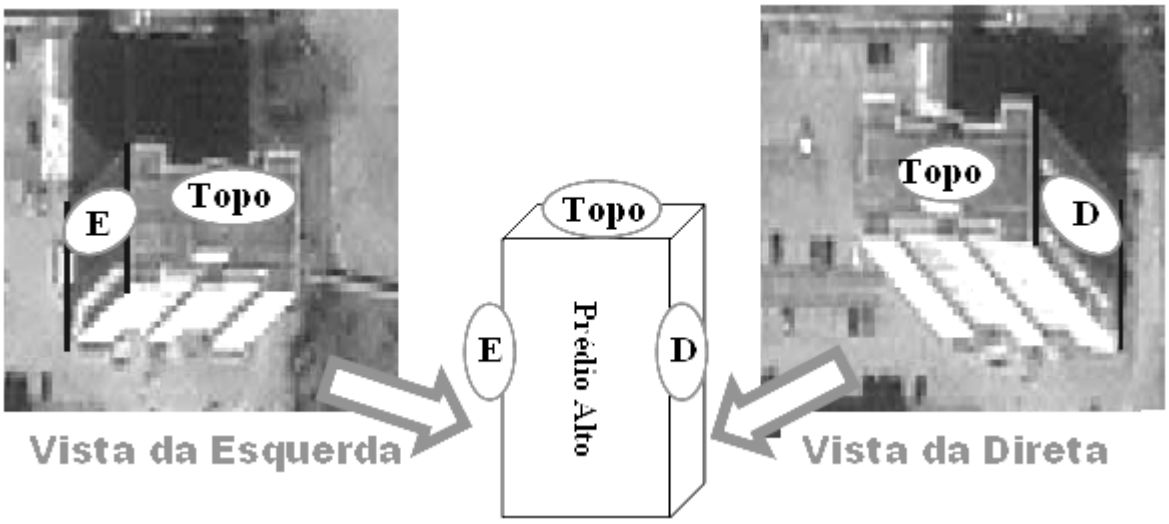

Figura 2.5: Efeito de uma mudança do ponto de vista 3D na imagem de um prédio alto

2.2
Métodos baseados em feições – SIFT

O segundo tipo de método para obter mapas de correspondência entre duas imagens é baseado em correspondências de feições. Selecionam-se poucos pontos característicos das duas imagens antes de procurar, para cada um da primeira imagem, o seu correspondente no banco de pontos da segunda imagem. É um método eficiente devido ao número limitado de candidatos, e a serem esses pontos mais salientes na imagem o que facilita a tarefa de distingui-los. Pode-se conseguir uma robustez melhor para oclusões, mudanças do ponto de vista 3D e outras alterações das imagens. As etapas principais de uma correspondência de feições são ilustradas na figura 2.6.

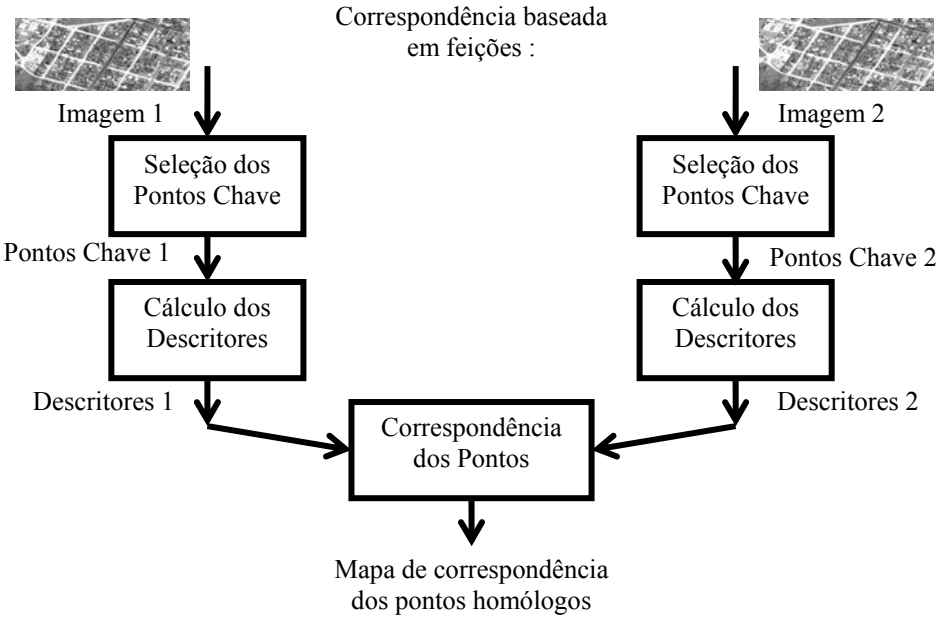

Figura 2.6: Etapas de um método de correspondência baseado em feições

Um método baseado em feições, proposto pelo Lowe em [2], conhecido pelo acrônimo SIFT (*Scale Invariant Feature Transform*), ou transformação de feições invariantes à escala, tem sido empregado com sucesso na área de robótica. Ele é invariante quanto à escala e rotação, e parcialmente quanto à iluminação e ao ponto de vista. A seguir serão apresentados os três passos principais que compõem este método.

2.2.1
Detecção dos pontos chave do SIFT

2.2.1.1
Invariância à escala

A invariância quanto à escala é obtida através da construção de uma pirâmide de imagens, conforme mostra a figura 2.7. Koenderink (1984) e Lindeberg (1994) mostraram que a melhor função conservando as propriedades da representação escala-espaço é a função Gaussiana. Portanto, novas imagens são geradas através da suavização da imagem original por um filtro gaussiano de desvio padrão σ variável $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right)$, criando assim uma

oitava de imagens, com diversos níveis, cada um desses níveis correspondendo a uma escala, ou seja a um desvio padrão do filtro gaussiano diferente. Uma dessas imagens é reduzida no tamanho gerando uma nova imagem com a metade da resolução da imagem original. O sucessivo processo de suavização, seguido pela redução de tamanho da imagem pode ser repetido várias vezes, gerando assim novas oitavas. As imagens da pirâmide possuem diferentes escalas e, desta forma, diferentes níveis de detalhes.

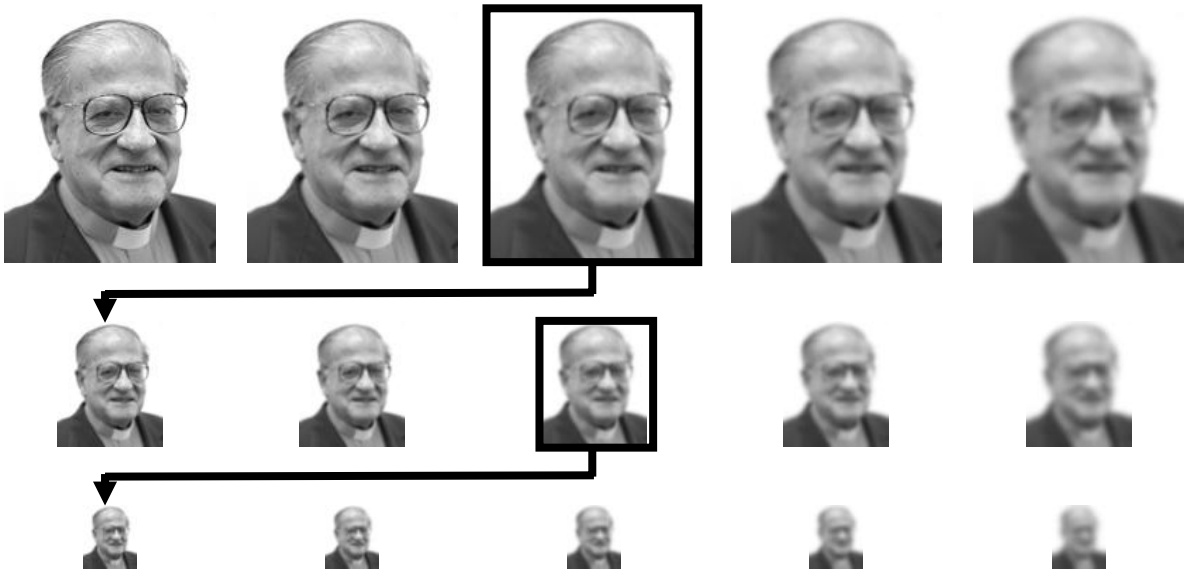

Figura 2.7: Pirâmide de imagens filtradas pelas gaussianas com 3 oitavas e 5 níveis cada.

2.2.1.2
Seleção dos pontos característicos do SIFT

Pontos chave são selecionados através desta pirâmide. Vários métodos foram testados por Mikolajczyk e Schmid ([3] e [5]) para se encontrar os melhores pontos chave. Mostraram que o LoG (Laplaciano de Gaussianas) normalizado na escala é uma das feições mais estáveis quando comparadas a outras como o gradiente, o Hessiano ou os cantos de Harris. Lowe propõe em [2] uma maneira muito eficiente de se aproximação do LoG : para cada oitava, calcula-se a diferença entre imagens em níveis adjacentes. Esta operação produz a pirâmide de Diferença de Gaussiana (DoG), conforme ilustra a figura 2.8. Os valores extremos da pirâmide DoG computados no espaço e na escala são selecionados como pontos chave.

Para detectar esses extremos locais na pirâmide DoG, todos os pontos são comparados com seus 8 vizinhos na sua própria imagem, seus 9 vizinhos na escala acima (nível acima da mesma escala), e seus outros 9 abaixo (nível abaixo da mesma escala, vide Figura 2.9). As três imagens utilizadas para a detecção dos extremas estão extraídas da mesma oitava e têm então a mesma resolução. Um ponto é selecionado somente se seu valor for maior ou menor do que o de todos esses 26 vizinhos.

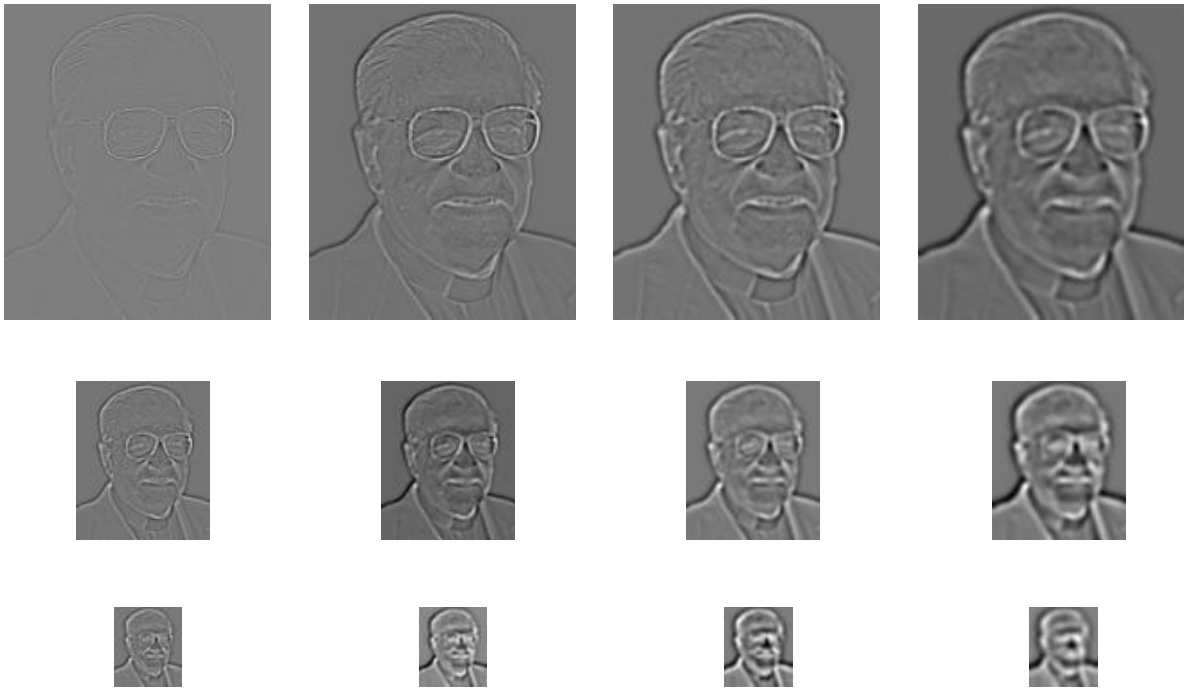

Figura 2.8: Pirâmide de diferenças de gaussianas das imagens (DoG).

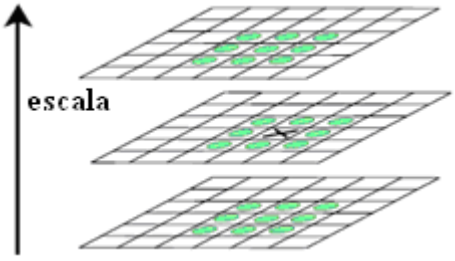

Figura 2.9: Seleção dos extremos na pirâmide DoG entre os 26 vizinhos.

2.2.1.3
Refinamento da localização dos pontos chave

Alguns dos extremos selecionados nas pirâmides de DoG das duas imagens se situam em áreas da imagem com menos informações, e são então descartados. Eliminam-se primeiro os pontos chave em vizinhanças de baixo contraste, que são mais sensíveis ao ruído e são, portanto, mais sensíveis a erros na correspondência. Esses pontos são eliminados se os seus valores na pirâmide DoG ficam abaixo de um limiar (um dos parâmetros cuja influência será estudada).

Eliminam-se também pontos chave situados sobre arestas das imagens. De fato, muitos extremos locais na pirâmide de DoG são localizados na vizinhança de arestas (linhas retas) onde o sinal muda em uma direção só, mas esses pontos são menos estáveis porque as arestas contêm poucas informações, suas localizações não podem ser definidas com muita exatidão. Assim, para cada um dos pontos selecionados, a matriz μ de segundo momento é calculada na escala σ da pirâmide e no espaço (*x, y*) onde ele é localizado :

$$\mu(x,y,\sigma) = \sigma^2 \begin{bmatrix} D_x^2(x,y,\sigma) & D_y D_x(x,y,\sigma) \\ D_y D_x(x,y,\sigma) & D_y^2(x,y,\sigma) \end{bmatrix} \tag{2.7}$$

onde $D_x(.)$ e $D_y(.)$ são as derivadas espaciais calculadas nas direções *x* e *y*. Essa matriz descreve a distribuição do gradiente da imagem na vizinhança do ponto central (*x, y*), e os seus autovalores representam as duas principais mudanças de sinal ao redor do centro. É possível identificar pontos localizados em arestas retas porque um dos seus autovalores (denotado *A*) é muito maior do que o outro (denotado *B*). Eliminam-se os pontos cuja razão r = *A/B* é superior a um dado limite, um parâmetro do algoritmo. Só que o cálculo de autovalores necessita muita computação. Lowe propôs em [2] um critério mais eficiente, que não envolve o cálculo dos autovalores, mais que aproxima a razão r = A/B :

$$CRIT\_LOWE = \frac{Tr^2(\mu)}{\det(\mu)} = \frac{(A+B)^2}{AB} = \frac{(r+1)^2}{r} \approx r \tag{2.8}$$

Este critério é similar ao critério proposto por Harris e Stephens em [8] para encontrar cantos nas imagens, que também combina o traço e o determinante da matriz de segunda ordem μ:

$$CRIT\_HARRIS = \det(\mu) - \lambda.Tr^2(\mu) \quad (2.9)$$

que envolve um parâmetro adicional (λ). São eliminados os pontos chave cujo critério de Lowe (equação 2.8) é superior a um limiar escolhido pelo operador. Cabe notar que um limiar mais elevado fornece mais pontos.

2.2.2
Descritores dos pontos chave

Uma vez que os pontos chaves são determinados, os seus descritores são calculados da seguinte maneira:

1) Os gradientes da vizinhança em torno de cada ponto chave são calculados no nível em que foram localizados. São representados na figura 2.10 pelas pequenas setas na imagem da esquerda, onde a vizinhança mede 8 × 8 pixels.
2) A vizinhança é dividida em sub-regiões. Aqui são 2 × 2 sub-regiões de 4 × 4 pixels cada uma.
3) Para cada sub-região, calcula-se o histograma de direções dos gradientes. Na composição desses histogramas, os valores acumulados são ponderados pelas respectivas magnitudes dos gradientes e por uma gaussiana centrada no ponto chave, o que dá mais importância ao centro da vizinhança. São utilizadas as direções relativas em vez das absolutas, o que faz que os descritores sejam invariantes quanto à rotação (a referência é a direção do gradiente no próprio ponto chave).
4) A contagem nos histogramas é empilhada, formando um vetor que constitui o descritor daquele ponto chave particular. No exemplo da figura 2.10, há quatro histogramas compreendo 8 direções principais, o que fornece um descritor de 36 coeficientes. Em [2], Lowe sugere o uso de descritores com 128 coeficientes para aplicações de robótica (4 × 4 sub-regiões de 4 × 4 pixels cada uma, com 8 direções nos histogramas).

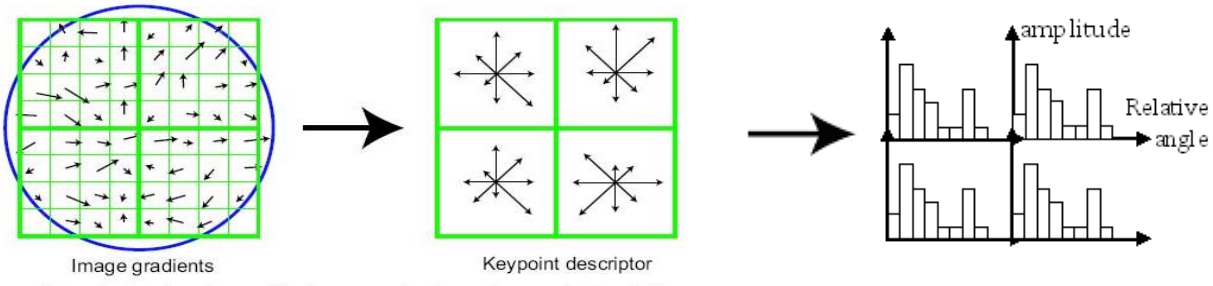

Figura 2.10: O descritor do SIFT

2.2.3
Correspondência

O processo descrito até aqui é aplicado a ambas as imagens do estereograma, produzindo dois conjuntos de pontos chave, onde cada ponto chave é representado pelo seu descritor. O grau de correspondência entre pontos chave de ambos os conjuntos é dado pela distância euclidiana entre seus descritores. Um par de pontos será considerado homólogo se:

- a distância euclidiana entre os descritores for menor que um dado limiar, e,
- a menor distância euclidiana em relação à segunda menor distância for maior que um segundo dado limiar.