

Detecção de Alvos Terrestres em Imagens Redimensionadas, Rotacionadas e Parcialmente Encobertas

Ronald Annoni Junior

Instituto Tecnológico de Aeronáutica, Pça Mal. Eduardo Gomes, 50, 12228-900, São José dos Campos, SP

Resumo — Um dos principais requisitos de sistemas ATD/R (*Automatic Target Detection/Recognition*) é a robustez de seu algoritmo de localização de alvos em imagens. Em situações reais, as imagens podem apresentar-se em escalas e orientações variadas, e até mesmo parcialmente encobertas por diversas razões. Este artigo apresenta os resultados de testes realizados utilizando o método SIFT (*Scale-Invariant Feature Transform*) para avaliar o seu desempenho na localização de alvos terrestres em imagens redimensionadas, rotacionadas e parcialmente encobertas.

Palavras-chaves — armamento aéreo, detecção automática de alvos, SIFT.

I. INTRODUÇÃO

Algoritmos ATD/R (*Automatic Target Detection / Recognition*) são utilizados para se esquadriar uma imagem e se determinar a localização de possíveis alvos.

Apesar de ser algo trivial para humanos, que em apenas um vislumbre são capazes de identificar vários objetos em uma cena, essa habilidade tem se mostrado extremamente complexa para ser reproduzida artificialmente.

O desenvolvimento de soluções para esse problema esbarra em diversas dificuldades de ordem prática tais como a variedade de tipos de alvos, variações de distância, de ângulos de visão, de iluminação, de condições geográficas e atmosféricas e até mesmo o encobrimento parcial ou a camuflagem desses alvos.

Na literatura científica encontramos uma ampla gama de abordagens para a detecção de objetos em imagens. Referência [1] enumera várias dessas técnicas e as classifica em quatro grupos: 1. classificação supervisionada, 2. segmentação, 3. modelagem de segundo plano e, 4. detecção de pontos de interesse.

Dentro deste último grupo destaca-se a técnica conhecida como SIFT [2,3] (*Scale-Invariant Feature Transform*) que em testes de desempenho realizados em [4] demonstrou maior robustez em comparação às técnicas congêneres sobre um conjunto de transformações aplicadas às imagens que incluiu rotação, mudança de escala, mudança de perspectiva, enevoamento (*blur*) da imagem, compressão JPEG e variações de iluminação.

Neste trabalho vamos apresentar os resultados de um conjunto de testes realizados para avaliar o desempenho do método SIFT para a detecção de alvos terrestres a partir de imagens em diferentes orientações e escalas e, ainda, parcialmente encobertos, procurando aproximar as condições de teste das condições reais que poderiam ser enfrentadas em campo.

II. CONCEITOS BÁSICOS

SIFT (*Scale-Invariant Feature Transform*), tal como descrito em [2], é um método que permite a extração de descritores locais de imagens que podem ser utilizados para realizar a correspondência entre diferentes visões de um objeto ou de uma cena. Descritores locais são vetores de características de uma imagem ou de determinadas regiões de uma imagem e possuem as qualidades de serem distintivos, resistentes a encobrimentos e não necessitarem de segmentação. Os descritores locais extraídos com o método SIFT são invariantes a rotações e a mudanças de escala, e também são razoavelmente capazes de fornecer a correspondência entre imagens com pequenas diferenças de perspectiva, iluminação e adição de ruído. A geração dos vetores de características SIFT é realizado em quatro estágios:

1) Determinação dos extremos do espaço de escalas

Em primeiro lugar, um espaço de escalas $D(x, y, \sigma)$ é obtido pela convolução da imagem $I(x, y)$ com uma diferença de filtros gaussianos $G(x, y, \sigma)$ em diferentes escalas σ separados por um fator multiplicativo k :

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (1)$$

Onde:

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

Os pontos candidatos a extremos de $D(x, y, \sigma)$ são então obtidos pela comparação de cada ponto de amostra com seus 26 vizinhos no espaço-escala (Fig. 1).

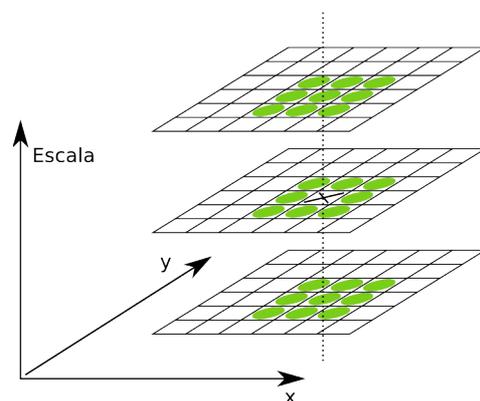


Fig. 1. Comparação de um ponto com seus vizinhos

2) Localização precisa dos pontos de interesse

Uma vez encontrados os pontos candidatos é realizado um ajuste detalhado sobre esses extremos para se obter a sua localização precisa, escala e curvaturas principais.

Para isso utiliza-se uma expansão de Taylor de $D(x, y, \sigma)$ transladada de forma que a origem da expansão coincida com o ponto:

$$D(x) = D + \frac{\partial D}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \quad (3)$$

A localização do extremo é obtida igualando-se a derivada a zero, resultando em:

$$\hat{x} = -\frac{\partial^2 D}{\partial x^2}^{-1} \frac{\partial D}{\partial x} \quad (4)$$

Substituindo (4) em (3) temos o valor da função no extremo, que é útil para se rejeitar extremos instáveis e com baixo contraste:

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D}{\partial x}^T \hat{x} \quad (5)$$

Para garantir estabilidade não é suficiente rejeitar pontos com baixo contraste mas também é necessário eliminar os picos de resposta ao longo das arestas. Para isso é preciso verificar a razão r entre as curvaturas principais por meio da expressão:

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} < \frac{(r+1)^2}{r} \quad (6)$$

Onde H é a matriz Hessiana calculada na localização e na escala do ponto dada por:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (7)$$

3) Atribuição de orientação aos descritores

A invariância com relação à rotação da imagem é alcançada atribuindo-se uma orientação consistente a cada ponto chave baseada em propriedades locais da imagem, de forma que os descritores podem ser representados relativamente a essas orientações. A escala do ponto chave é utilizada para selecionar a imagem convolvida com a Gaussiana que possui a escala mais próxima. Para cada amostra dessa imagem $L(x, y)$, na escala selecionada, calcula-se a magnitude $m(x, y)$ e a orientação $\theta(x, y)$ do gradiente:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (8)$$

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x+1, y) - L(x-1, y)}{L(x, y+1) - L(x, y-1)} \right) \quad (9)$$

Um histograma de orientações é formado a partir das orientações dos gradientes dos pontos dentro de uma região em torno do ponto chave. Os picos desse histograma correspondem às orientações dominantes neste ponto. Localizações com mais de um pico possuirão mais de um ponto chave.

4) Representação dos descritores locais da imagem

A Fig. 2 demonstra como um descritor de ponto chave é criado inicialmente pelo cálculo da magnitude e da orientação do gradiente em cada ponto de amostra da imagem em uma região em torno da localização do ponto chave. Essas amostras são ponderadas por uma Gaussiana (região circular da figura) e então são acumuladas em histogramas de orientações reduzindo o seu conteúdo para uma sub-região com o comprimento de cada seta correspondendo à soma das magnitudes próximas àquela direção dentro da sub-região.

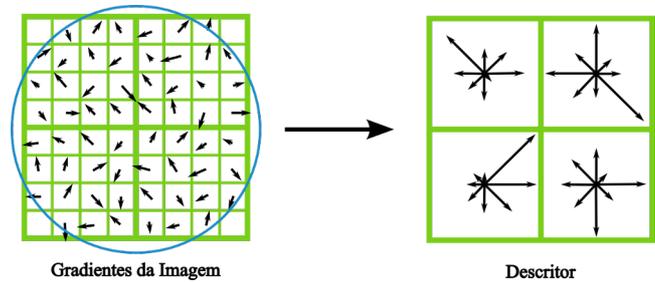


Fig. 2. Construção do Descritor

O descritor então é representado por um vetor em que cada componente corresponde a uma direção em cada sub-região dos histogramas. Por fim, para reduzir os possíveis efeitos de variações de iluminação das imagens, este descritor é normalizado de forma que os componentes com valores acima de determinado limiar sejam ajustados para esse limiar, evitando que direções com magnitude muito grande dominem a representação do descritor. Com isso conclui-se o processo de geração dos descritores SIFT para uma imagem.

Aplicando esse processo em duas imagens distintas geramos dois conjuntos de descritores a partir dos quais podemos determinar os pontos de correspondência dessas imagens encontrando-se, para cada descritor da primeira, o descritor da segunda que estiver à menor distância Euclidiana e eliminando deste resultado aqueles cuja razão entre as distâncias entre o primeiro e o segundo mais próximo seja maior que determinado valor de corte.

III. APLICAÇÃO DO MÉTODO SIFT E RESULTADOS

Nos testes apresentados neste artigo, uma implementação de David Lowe do algoritmo SIFT [5] foi utilizada para a geração dos descritores. *Scripts* MATLAB® foram utilizados para ler os descritores gerados, determinar e contar as correspondências e desenhar uma linha entre as imagens ligando cada correspondência encontrada.

A partir de uma imagem base foi extraída a imagem de uma edificação que foi utilizada como alvo a ser localizado de volta na imagem base após esta passar pelas transformações de escala, de rotação e de adição de ruído para simular um encobrimento natural.

A Fig. 3 mostra o alvo selecionado e a sua imagem sobreposta pelos descritores encontrados pelo método SIFT.

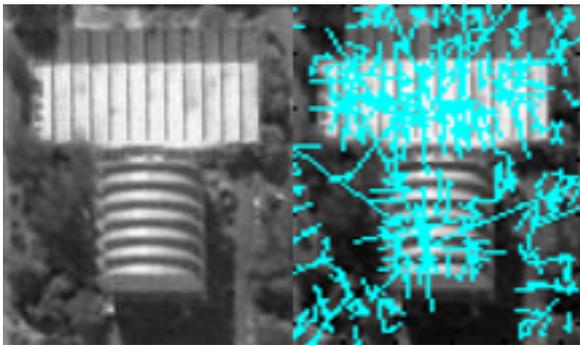


Fig. 3. Alvo selecionado para teste com 241 descritores

Como referência, foi feita a correspondência do alvo com a imagem de origem sem qualquer modificação. Nessas condições a imagem de origem gerou 5.556 descritores, que comparados com os 241 do alvo forneceram 204 correspondências corretas e 5 incorretas (Fig. 4).

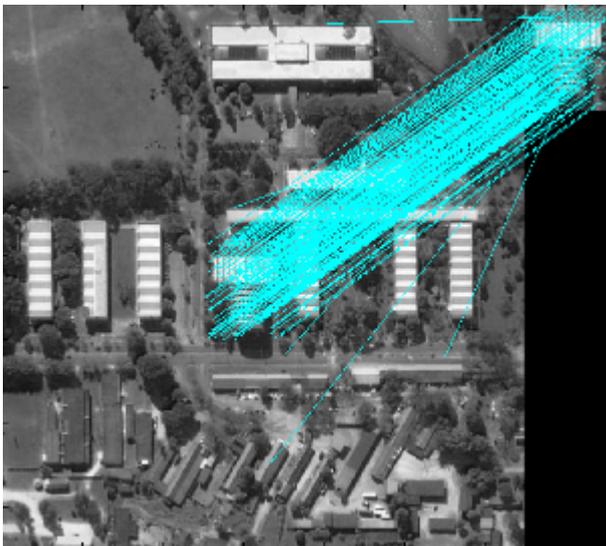


Fig. 4. Correspondência do alvo com a imagem original

Em seguida, com a ferramenta GIMP [6] operamos diversas modificações na imagem base original a fim de que pudéssemos testar a capacidade do método de localizar o alvo diante de uma imagem com orientação diversa, com baixo contraste, com ruído, em escala diferente e até mesmo com o alvo parcialmente encoberto.

1) Rotação da imagem

Testamos a detecção do alvo em imagens rotacionadas de 15° em 15° a partir da original.

Nas imagens rotacionadas o número de correspondências encontrado foi menor do que o obtido com a imagem na orientação original, mas ainda alto, com número de correspondências variando entre 127 e 181 (Fig. 5), quase todas dentro do alvo (Fig. 6).

Deste resultado ficou clara a invariância dos descritores SIFT com relação à rotação pela sua capacidade de distinguir um grande número de pontos do alvo na imagem transformada com quase a totalidade de acertos.

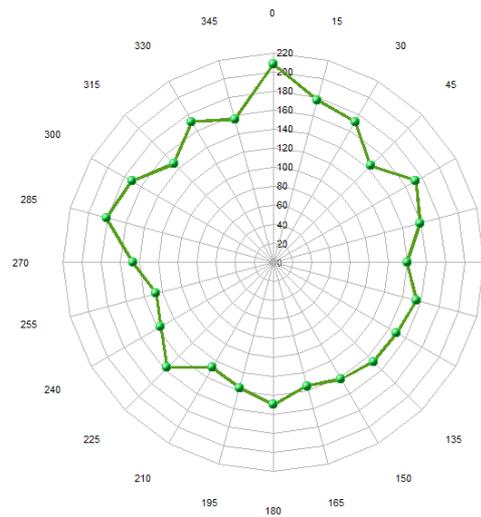


Fig. 5. Número de correspondências em função do ângulo de rotação

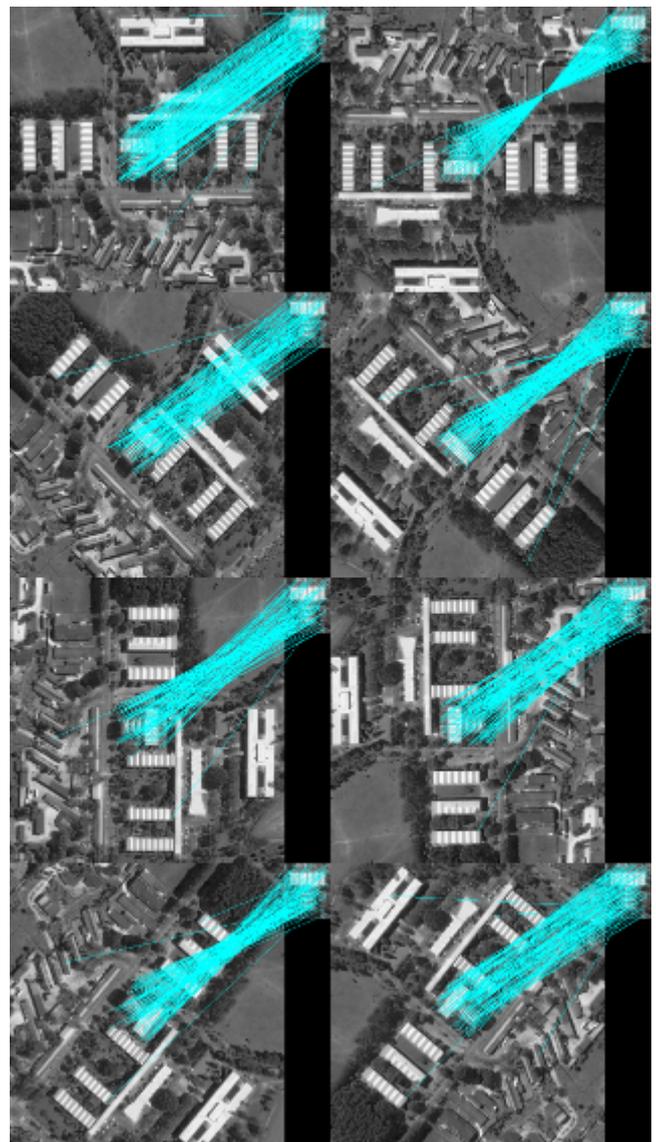


Fig. 6. Sequência com várias rotações da imagem

2) Baixo contraste

Neste teste aplicamos primeiramente uma máscara branca sobre a imagem variando a sua opacidade de 10% até 90% e depois repetimos o mesmo processo aplicando uma máscara preta.

Na Fig. 7 pode-se observar que o número de descritores que o método consegue gerar diminui com a redução do contraste e, conseqüentemente, também diminui o número de correspondências.

Na Tabela I vemos que a 80% nenhuma correspondência foi localizada e a 90% nem mesmo um descritor foi gerado para a imagem. Também é interessante observar que os resultados foram praticamente idênticos para a máscara preta e a branca demonstrando que o método considera apenas valores relativos de intensidade entre os *pixels* e não valores absolutos.

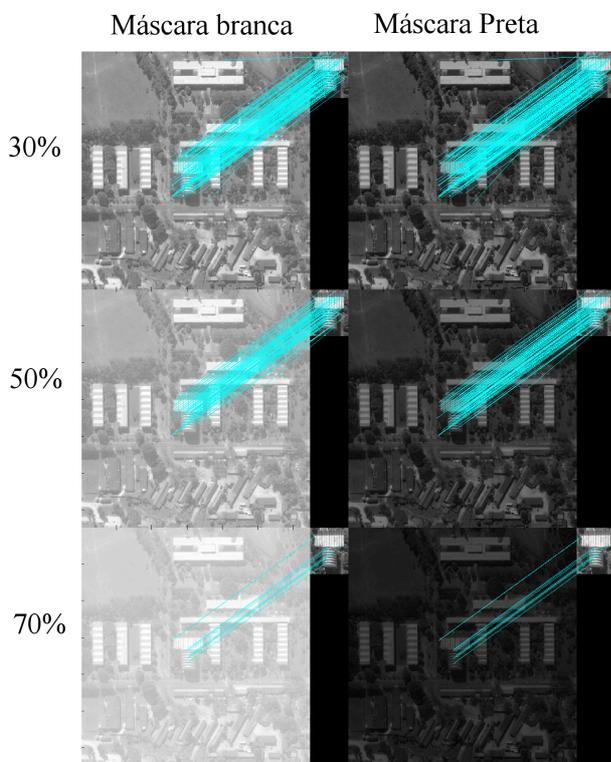


Fig. 7. Efeito da redução do contraste

TABELA I – EFEITOS DA REDUÇÃO DO CONTRASTE

Opacidade	Máscara Branca		Máscara Preta	
	Descritores	Correspondências	Descritores	Correspondências
10,00%	4871	192	4870	192
20,00%	4089	179	4089	180
30,00%	3298	149	3296	150
40,00%	2452	117	2455	117
50,00%	1634	80	1639	82
60,00%	954	44	953	44
70,00%	397	16	400	16
80,00%	40	0	40	0
90,00%	0	0	0	0

Como era de se esperar, o achatamento da distribuição dos valores dos *pixels* provocou a perda de informação da imagem dificultando a obtenção dos descritores e das correspondências. No entanto o fato do método ter suportado até 70% de opacidade da máscara localizando corretamente o alvo sem se confundir com as outras áreas da imagem demonstra robustez do algoritmo.

3) Ruído

Este caso é uma variação do anterior sendo que desta vez foi aplicada uma máscara de ruído composta de uma distribuição aleatória de valores de *pixels* e de camada de transparência.

Também atribuímos uma opacidade a esta camada variando de 10% a 90% e os resultados podem ser vistos na Fig. 8 e na Tabela II.

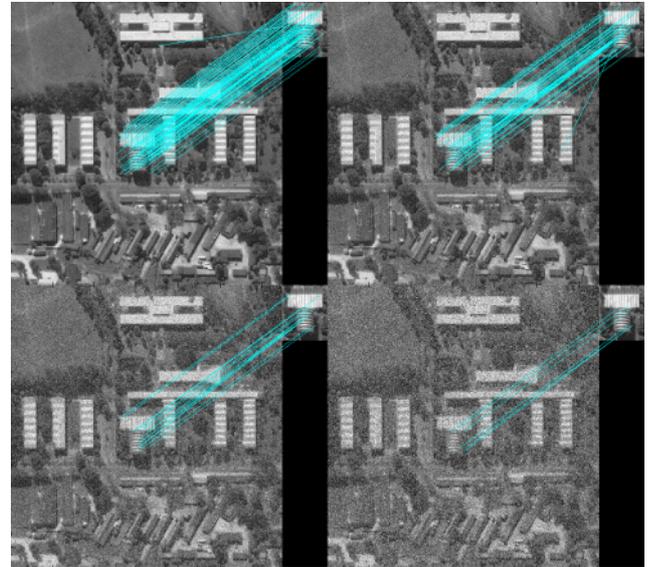


Fig. 8. Máscara de ruído com 20%, 40%, 60% e 80%

TABELA II – EFEITOS DA MÁSCARA DE RUÍDO

Opacidade	Descritores	Correspondências
10,00%	5236	166
20,00%	4847	97
30,00%	4410	79
40,00%	4030	00
50,00%	3779	24
60,00%	3291	17
70,00%	2999	9
80,00%	2700	7
90,00%	2000	0
100,00%	2347	0

Aqui não houve o achatamento dos valores dos *pixels* e, portanto, a queda no número de descritores foi mais lenta.

Por outro lado, a desorganização dos *pixels* promovida pela máscara de ruído fez com que os descritores gerados fossem menos distintivos, causando uma queda mais rápida do número de correspondências encontradas em relação ao caso anterior. Ainda assim, com uma quantidade significativa de ruído na imagem o algoritmo foi capaz de identificar os pontos concordantes com o alvo sem falsas correspondências.

4) Variações de escala

A variação de escala que nos interessa aqui é a redução da imagem, porque neste processo há perda de informação. Como a imagem é reduzida, vários *pixels* tem que ser interpolados em um único *pixel*.

Neste teste tomamos a imagem original e reduzimos as suas dimensões (largura e altura) a 90%, 80%, e assim por diante até 10%. Isso obviamente reduz a quantidade de *pixels* a uma porcentagem ainda menor da quantidade de *pixels* da imagem original. A imagem foi reduzida utilizando interpolação cúbica e os resultados após aplicação do algoritmo SIFT podem ser vistos na Tabela III abaixo:

IV. CONCLUSÕES

TABELA III – RESULTADOS DA VARIÇÃO DE ESCALA

Redução das Dimensões	Pixels Resultantes	Descritores	Correspondências
90%	81%	4933	152
80%	64%	4050	121
70%	49%	3214	85
60%	36%	2279	46
50%	25%	1592	14
40%	16%	1178	11
30%	9%	659	6
20%	4%	337	3
10%	1%	98	0

Mais uma vez ficou clara a robustez do algoritmo ao conseguir identificar o alvo em uma imagem bastante transformada, com apenas 4% do número de pixels da imagem original (Fig. 9).

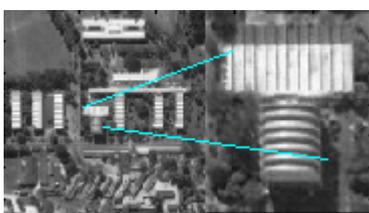


Fig. 9. Localização do alvo em imagem com apenas 4% do número de pixels da imagem original

5) Encobrimento parcial do alvo

Uma das características mais interessantes dos descritores neste caso que estamos investigando se deve ao fato de que eles descrevem características locais dos objetos e não dependem do objeto todo, ou seja, é possível identificar correspondências entre imagens mesmo que uma delas contenha apenas uma parte do objeto que se pretende localizar. Conforme se pode observar pela Fig. 10, isso ficou comprovado em diversos testes em que se simulou uma cobertura de nuvens sobre a área. Essa característica garante confiabilidade de resultado em situações reais em que não se tem controle das condições ambientais, entre outras.

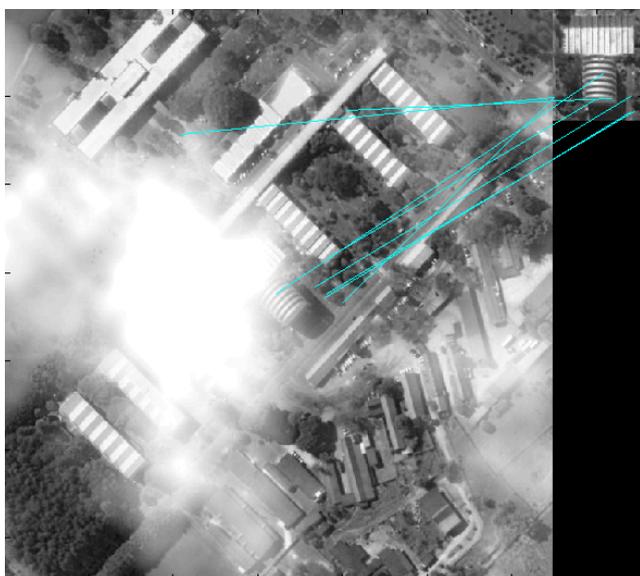


Fig. 10. Alvo parcialmente encoberto foi localizado

Neste trabalho apresentamos os resultados de testes realizados utilizando o método SIFT (*Scale-Invariant Feature Transform*) para avaliar a sua capacidade de localizar um alvo terrestre em imagens em diversas orientações e escalas, submetidas a redução de contraste, aplicação de ruído e até mesmo com encobrimento parcial do alvo.

Ficou demonstrado que, nas condições em que estes testes foram realizados, o algoritmo SIFT é bastante eficaz e robusto, localizando o alvo mesmo em imagens com transformações relativamente expressivas.

Um ponto importante que ainda precisa ser avaliado é a capacidade deste algoritmo de localizar objetos em imagens com mudança de perspectiva, coisa também muito comum em situações reais.

Entretanto, a conclusão geral que se chegou com este trabalho é que o método SIFT se mostrou bastante promissor como alternativa de algoritmo de detecção/reconhecimento de alvos.

REFERÊNCIAS

[1] A. Yilmaz, O. Javed, M. Shah, "Object Tracking: A Survey", ACM Computing Surveys, Vol. 38, No. 4, Article 13, Publication date: December 2006.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

[3] D. G. Lowe, "Object recognition from local scale-invariant features," International Conference on Computer Vision, Corfu, Greece (September 1999), pp. 1150-1157.

[4] K. Mikolajczyk, C. Schmid, "A Performance Evaluation of Local Descriptors", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 10, October 2005, pp. 1615-1630

[5] D. G. Lowe, "SIFT Keypoint Detector", demo program, v. 4, 2005, disponível em <http://www.cs.ubc.ca/~lowe/keypoints/>, acesso em 31/maio/2009.

[6] The GNU Image Manipulation Program, disponível em <http://www.gimp.org/>.