

GUILLERMO ÁNGEL PÉREZ LÓPEZ

**AFORAPRO: RECONHECIMENTO DE OBJETOS INVARIANTE SOB
TRANSFORMAÇÕES AFINS**

Dissertação apresentada à Escola Politécnica
da Universidade de São Paulo para obtenção
do Título de Mestre em Engenharia Elétrica.



São Paulo

2011

GUILLERMO ÁNGEL PÉREZ LÓPEZ

**AFORAPRO: RECONHECIMENTO DE OBJETOS INVARIANTE SOB
TRANSFORMAÇÕES AFINS**

Dissertação apresentada à Escola Politécnica
da Universidade de São Paulo para obtenção
do Título de Mestre em Engenharia Elétrica.

Área de Concentração:
Sistemas Eletrônicos

Orientador: Prof. Dr.
Hae Yong Kim

São Paulo

2011

*A meu Deus, aquele
que apareceu na minha vida induzido talvez pelos livros sagrados,
mas que agora não precisa daquelas histórias. Porque manifesta
sua presença constantemente ao meu lado, me fazendo sentir mais
forte a cada dia. Obrigado meu Deus!*

DEDICO

AGRADECIMENTOS

Ao Prof. Dr. Hae Yong Kim, meu orientador, pelas suas importantes orientações e pelo apoio para meu ingresso no programa de mestrado.

À minha família, especialmente aos meus pais e minha sobrinha Laura, por seu apoio e sua compreensão, incondicionais.

A meu amigo e colega Ricardo Hitoshi e sua família, pela convivência agradável, pelos conhecimentos compartilhados e momentos de descontração.

A minha linda Wysllenny Nascimento, por participar de meus sonhos, por seu apoio incondicional e por ter sempre um sorriso para me dar.

Aos colegas do Laboratório de Processamento de Sinais da Escola Politécnica da Universidade de São Paulo pela convivência enriquecedora e pelos conhecimentos compartilhados.

Ao Dr. Guoshen Yu, da Ecole Polytechnique Palaiseau, France, por suas sugestões.

Ao Centro Residencial da Universidade de São Paulo – CRUSP, pela moradia proporcionada durante quase todo o mestrado, em especial à assistente social Luisa por sua compreensão e apoio.

Aos meus amigos do CRUSP, em especial aos meus colegas de apartamento pelo ambiente familiar gerado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, pela bolsa de estudos que me foi concedida.

Aos membros da banca, Prof. Dr. J. A. Quintanilha e Prof. Dr. Sidnei Alves, pelas importantes contribuições para a finalização deste trabalho.

RESUMO

Reconhecimento de objetos é uma aplicação básica da área de processamento de imagens e visão computacional. O procedimento comum do reconhecimento consiste em achar ocorrências de uma imagem modelo Q numa outra imagem a ser analisada A . Conseqüentemente, se as imagens apresentarem mudanças no ponto de vista da câmera o algoritmo normalmente falha. A invariância a pontos de vista é uma qualidade que permite reconhecer um objeto, mesmo que este apresente distorções resultantes de uma transformação em perspectiva causada pela mudança do ponto de vista. Uma abordagem baseada na simulação de pontos de vista, chamada ASIFT, tem sido recentemente proposta no entorno desta problemática. O ASIFT é invariante a pontos de vista, no entanto falha na presença de padrões repetitivos e baixo contraste. O objetivo de nosso trabalho é utilizar uma variante da técnica de simulação de pontos de vista em combinação com a técnica de extração dos coeficientes de Fourier de projeções radiais e circulares (FORAPRO), para propor um algoritmo invariante a pontos de vista, e robusto a padrões repetitivos e baixo contraste. De maneira geral, a nossa proposta resume-se nas seguintes fases: (a) Distorcemos a imagem Q , variando os parâmetros de inclinação e rotação da câmera, para gerar alguns modelos $Q \rightarrow Q_1, Q_2, \dots, Q_n$ e conseguir a invariância a deformações em perspectiva, (b) utilizamos cada Q_i como modelo a ser procurado na imagem A , para escolher o Q_i que melhor case, (c) realizamos o casamento de padrões. As duas últimas fases do processo baseiam-se em características invariantes por rotação, escala, brilho e contraste extraídas pelos coeficientes de Fourier. Nossa proposta, que chamamos AFORAPRO, foi testada com 350 imagens que continham diversidade nos requerimentos, e demonstrou ser invariante a pontos de vista e ter ótimo desempenho na presença de padrões repetitivos e baixo contraste.

Palavras chaves: Reconhecimento de objetos. Casamento de padrões. ASIFT. Distorção de imagens. Coeficientes de Fourier. Invariância afim. Simulação de ponto de vista. Padrões repetitivos. Mudança de contraste. FORAPRO.

ABSTRACT

Object recognition is a basic application from the domain of image processing and computer vision. The common process recognition consists of finding occurrences of an image query Q in another image to be analyzed A . Consequently, if the images changes viewpoint in the camera it will normally result in the algorithm failure. The invariance viewpoints are qualities that permit recognition of an object, even if this present distortion resultant of a transformation of perspective is caused by the change in viewpoint. An approach based on viewpoint simulation, called ASIFT, has recently been proposed surrounding this issue. The ASIFT algorithm is invariant viewpoints; however there are flaws in the presence of repetitive patterns and low contrast. The objective of our work is to use a variant of this technique of viewpoint simulating, in combination with the technique of extraction of the Coefficients of Fourier Projections Radials and Circulars (FORAPRO), and to propose an algorithm of invariant viewpoints and robust repetitive patterns and low contrast. In general, our proposal summarizes the following stages: (a) We distort the image Q , varying the parameters of inclination and rotation of the camera, to produce some models $Q \rightarrow Q_1, Q_2, \dots, Q_n$ and achieve perspective invariance deformation, (b) use Q_i as the model to be search in the image, to choose the Q_i that match best, (c) realize the template matching. The two last stages of process are based on invariant features by images rotation, scale, brightness and contrast extracted by Fourier coefficients. Our approach, that we call AFORAPRO, was tested with 350 images that contained diversity in applications, and demonstrated to have invariant viewpoints, and to have excellent performance in the presence of patterns repetitive and low contrast.

Keywords: Objects recognition. Template-matching. ASIFT. Images distortion. Fourier coefficients. Affine invariant. Viewpoint simulation. Repetitive patterns. Changes contrast. FORAPRO.

LISTA DE FIGURAS

	Pag.
Figura 2.1	<i>Elementos básicos no reconhecimento de objetos.</i> 24
Figura 2.2	<i>Exemplo do reconhecimento de objetos.</i> 25
Figura 2.3	<i>Exemplificação dos casos possíveis casos de reconhecimento.</i> 29
Figura 2.4	<i>Detecção de cantos utilizando o detector de Harris.</i> 31
Figura 2.5	<i>Extremas (pontos chaves) achadas utilizando DoG.</i> 34
Figura 2.6	<i>Descritor de regiões características ao redor de um ponto chave.</i> 35
Figura 2.7	<i>Desempenho do SIFT.</i> 37
Figura 2.8	<i>Projeções utilizadas pelo FORAPRO.</i> 38
Figura 2.9	<i>Núcleos radiais e circulares utilizados pelo FORAPRO.</i> 39
Figura 2.10	<i>Teste do desempenho de FORAPRO.</i> 44
Figura 2.11	<i>Vizinhança envolvida na construção de invariantes a mudanças de brilho e contraste, e presença de padrões repetitivos.</i> 46
Figura 2.12	<i>Robustez a padrões repetitivos e baixo contraste.</i> 46
Figura 3.1	<i>Transformações afins.</i> 49
Figura 3.2	<i>Variação do contraste em função da intensidade.</i> 50
Figura 3.3	<i>Fatores que influenciam as mudanças de contraste.</i> 51
Figura 3.4	<i>Padrões repetitivos.</i> 52
Figura 3.5	<i>Mudança da visual induzida pela variação do ponto de vista da câmera.</i> 53
Figura 3.6	<i>Uma transformação em perspectiva.</i> 54
Figura 3.7	<i>Processo de normalização de regiões.</i> 55
Figura 3.8	<i>Parâmetros geométricos envolvidos na captura da imagem.</i> 57
Figura 3.9	<i>Modelo de distribuição de pontos de vista.</i> 59
Figura 3.10	<i>Exemplo da técnica de simulação de pontos de vista.</i> 60
Figura 3.11	<i>Demonstração de desempenho do MSER.</i> 62
Figura 3.12	<i>Detecção de regiões invariantes.</i> 65
Figura 3.13	<i>Demonstração de desempenho do ASIFT.</i> 66
Figura 3.14	<i>Exemplo de falhas do ASIFT na presença de padrões repetitivos.</i> 68

Figura 3.15	<i>ASIFT falha com grandes mudanças de contraste.</i>	68
Figura 4.1	<i>Imagens de entrada ao processo de reconhecimento.</i>	71
Figura 4.2	<i>Imagens sintéticas Q_i correspondentes à simulação de distorções. . .</i>	72
Figura 4.3	<i>Imagens envolvidas no casamento final.</i>	73
Figura 4.4	<i>Testes realizados com imagens do banco de dados do ASIFT.</i>	73
Figura 4.5	<i>Testes realizados com objetos de formas simples e baixo contraste. . .</i>	74
Figura 4.6	<i>Testes com imagens que apresentam mudanças de contraste.</i>	76
Figura 4.7	<i>Testes comparativos, usando imagens que possuem mudanças de contraste.</i>	76
Figura 4.8	<i>Desempenho do AFORAPRO usando imagens que apresentam padrões repetitivos.</i>	77
Figura 4.9	<i>Teste comparativo, usando imagens que apresentam padrões repetitivos.</i>	77
Figura 5.1	<i>Imagens Q do Conjunto 1.</i>	81
Figura 5.2	<i>Imagens Q do Conjunto 2.</i>	82
Figura 5.3	<i>Imagens Q do Conjunto 3.</i>	83
Figura 5.4	<i>Imagens Q do Conjunto 4.</i>	83
Figura 5.5	<i>Imagens Q do Conjunto 5.</i>	84
Figura 5.6	<i>Resultado do FORAPRO com objetos do Conjunto 1 “extensão”.</i>	85
Figura 5.7	<i>Resultados do AFORAPRO com objeto do Conjunto 1 “extensão”.</i>	85
Figura 5.8	<i>Resultado do ASIFT com o objeto do Conjunto 1 “extensão”.</i>	88
Figura 5.9	<i>Resultado com objeto do Conjunto 2 “extintor”.</i>	89
Figura 5.10	<i>Resultados com imagem do Conjunto 3 “grafite”.</i>	92
Figura 5.11	<i>Resultados com imagem do Conjunto 3 “Adam”.</i>	92
Figura 5.12	<i>Resultados com imagem do Conjunto 3 “Lepoint”.</i>	93
Figura 5.13	<i>Resultados com alguns objetos do Conjunto 4.</i>	94
Figura 5.14	<i>Resultados com alguns objetos do Conjunto 5.</i>	97

LISTA DE GRÁFICOS

	Pag.
Gráfico 5.1 <i>Taxa de acertos do FORAPRO e AFORAPRO na invariância a pontos de vista com imagens do Conjunto 1.</i>	87
Gráfico 5.2 <i>Taxa de Acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 1.</i>	89
Gráfico 5.3 <i>Taxa de acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 2.</i>	91
Gráfico 5.4 <i>Taxa de acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 3.</i>	94
Gráfico 5.5 <i>Taxa de acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 4.</i>	96
Gráfico 5.6 <i>Taxa de acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 5.</i>	98
Gráfico 5.7 <i>Taxa de acertos do ASIFT e AFORAPRO para cada um dos conjuntos de imagens utilizados nos experimentos.</i>	99

LISTA DE TABELAS

	Pag.
Tabela 2.1 <i>Tabela de contingência para avaliação de desempenho.</i>	28
Tabela 4.1 <i>Tempo de execução na geração de imagens sintéticas (AFORAPRO). .</i>	78
Tabela 4.2 <i>Tempo de execução na seleção do Q_i para o casamento.</i>	79
Tabela 4.3 <i>Tempo de execução na etapa do reconhecimento final.</i>	80
Tabela 5.1 <i>Resultado da invariância a pontos de vista relacionando o FORAPRO e AFORAPRO.</i>	86
Tabela 5.2 <i>Resultado da invariância a pontos de vista com imagens do Conjunto 1.</i>	88
Tabela 5.3 <i>Respostas do ASIFT e AFORAPRO com as imagens do Conjunto 2. . .</i>	90
Tabela 5.4 <i>Respostas do ASIFT e AFORAPRO com as imagens do Conjunto 3. . .</i>	93
Tabela 5.5 <i>Respostas do ASIFT e AFORAPRO com as imagens do Conjunto 4. . .</i>	95
Tabela 5.6 <i>Respostas do ASIFT e AFORAPRO com as imagens do Conjunto 5. . .</i>	97

LISTA DE ABREVIATURAS E SIGLAS

AFORAPRO	<i>Affine Fourier coefficients of radial projections – Extensão do FORAPRO que o torna invariante afim.</i>
ASIFT	<i>Affine Scale-Invariant Feature Transform (Affine-SIFT) – Extensão do SIFT que o torna invariante afim.</i>
FORAPRO	<i>Fourier coefficients of radial projections – Técnica de reconhecimento de objetos baseada na extração dos coeficientes de Fourier.</i>
IMG	<i>Biblioteca para processamento de imagens e visão computacional de autoria do Prof. Dr. Hae Yong Kim.</i>
MSER	<i>Maximally Stable Extremal Region – Técnica de casamento de regiões características.</i>
OpenCV	<i>Open Computer Vision - Biblioteca com rotinas para processamento de imagens e visão computacional, desenvolvida pela Intel.</i>
ProEikon	<i>Biblioteca para processamento de imagens e visão computacional de autoria do Prof. Dr. Hae Yong Kim.</i>
SIFT	<i>Scale-Invariant Feature Transform – Técnica de casamento de padrões baseada em extração de pontos chaves.</i>
SURF	<i>Speeded Up Robust Features - Método para detecção de pontos chaves em imagens.</i>

SUMÁRIO

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

	Pág.
1 INTRODUÇÃO	
1.1 Considerações iniciais.	16
1.2 Objetivos.	20
1.3 Motivação e justificativas.	21
1.4 Publicação associada a esta dissertação.	21
1.5 Organização.	22
2 RECONHECIMENTO DE OBJETOS EM IMAGENS	
2.1 Introdução.	23
2.2 Reconhecimento de objetos.	25
2.3 Avaliação de desempenho.	28
2.4 Revisão da literatura.	29
2.5 Descrição dos algoritmos SIFT e FORAPRO.	33
2.5.1 Algoritmo SIFT.	33

2.5.1.1	Detecção de pontos candidatos (detector <i>DoG</i>).	33
2.5.1.2	Descritor de regiões do SIFT.	35
2.5.1.3	Reconhecimento de objetos com SIFT.	36
2.5.2	Algoritmo FORAPRO.	37
2.5.2.1	Coeficientes de Fourier de projeções radiais (CFPR).	38
2.5.2.2	Orientação canônica.	40
2.5.2.3	Vetor de magnitudes radiais.	40
2.5.2.4	Vetor de ângulos radiais.	41
2.5.2.5	Invariância a rotação (vetor de características circulares).	42
2.5.2.6	Combinação das características.	43
2.5.2.7	Transformada de Hough.	43
2.5.2.8	Invariância a escala.	44
2.5.3	Análise comparativo do SIFT e FORAPRO.	45

3 CARACTERÍSTICAS INVARIANTES E TRANSFORMAÇÃO AFIM

3.1	Introdução.	47
3.2	Transformações afins.	47
3.3	Invariância a mudanças de contraste.	50
3.4	Robustez a padrões repetitivos.	51
3.5	Invariância a pontos de vista.	53
3.5.1	Normalização de regiões.	55
3.5.2	Simulação de pontos de vista.	57
3.6	Estado da arte da invariância a pontos de vista	60

3.6.1	MSER (<i>maximally stable extremal regions</i>)	61
3.6.2	Hessian-affine e Harris-affine.	63
3.6.3	Algoritmo ASIFT (<i>Affine- scale invariant feature transform</i>)	66
3.6.3.1	Fraqueza do ASIFT a padrões repetitivos.	67
3.6.3.2	Fraqueza do ASIFT a grandes mudanças de contraste.	68

4 AFORAPRO: FORAPRO INVARIANTE AFIM

4.1	Introdução.	69
4.2	Algoritmo AFORAPRO.	69
4.3	Invariância a pontos de vista.	71
4.4	Robustez a padrões repetitivos e mudanças de contraste.	75
4.5	Análise de complexidade.	78
4.5.1	Geração de imagens sintéticas.	78
4.5.2	Seleção da imagem sintética para o casamento.	79
4.5.3	Etapa de reconhecimento.	79

5 RESULTADOS EXPERIMENTAIS

5.1	Introdução.	81
5.2	Experimentos de invariância a pontos de vista	85
5.2.1	AFORAPRO e FORAPRO.	85
5.2.2	AFORAPRO e ASIFT.	87

5.2.2.1 Experimentos com imagens do Conjunto 1.	87
5.2.2.2 Experimentos com imagens do Conjunto 2.	89
5.2.2.3 Experimentos com imagens do Conjunto 3.	91
5.3 Experimentos de invariância à mudanças de contraste.	94
5.4 Experimentos de robustez a padrões repetitivos.	96
5.5 Resumo dos resultados.	98
5.5.1 AFORAPRO e FORAPRO.	99
5.5.2 AFORAPRO e ASIFT.	99
6 CONCLUSÕES	
6.1 Comentários finais.	101
6.2 Trabalho futuro.	102
REFERÊNCIAS.	103

Capítulo 1.

INTRODUÇÃO

Neste capítulo, apresentamos uma exposição do tema e motivação do trabalho. Iniciamos com uma revisão da literatura, apresentamos os objetivos, a abordagem da proposta, a publicação associada a esta dissertação, e finalizamos com a organização do documento.

1.1 Considerações Iniciais

O reconhecimento de objetos inicia com duas imagens (normalmente em nível de cinza), uma é denominada modelo (Q de “*query*”) e a outra imagem de procura (A de “*analyze*”). O objetivo é achar as instâncias de Q presentes em A . As imagens são tratadas como uma matriz numérica onde cada posição é representada por um nível de cinza (intensidade) e os padrões ou objetos são regiões de A que apresentam um alto grau de semelhança com Q . A técnica consiste em extrair as características do modelo Q e achá-las na imagem A . Algumas das dificuldades que se apresentam são: (1) as imagens podem aparecer em diferentes rotações e escalas; (2) as instâncias de Q em A podem sofrer efeitos fotométricos que incluem mudança de brilho e contraste; (3) o conjunto ótico da câmera pode borrar as imagens (desajustes no foco); (4) as imagens podem sofrer deformação em perspectiva devido à mudança na posição do fotógrafo (mudança do ponto de vista). Uma solução, a estes tipos de dificuldades, é usar algoritmos invariantes a essas deformações.

Nos últimos anos, os avanços alcançados com a utilização das técnicas baseadas em características locais invariantes (e.g. LOWE, 2004; MIKOLAJCZIK; SCHMID, 2004; RABIN, et. al., 2009; SHIN; TJAHJADI, 2010) têm gerado grandes avanços no reconhecimento de objetos. Estas técnicas utilizam dois passos estratégicos para descrever o conteúdo da imagem: (a) detecção de “pontos de interesse” que são aqueles pixels facilmente localizáveis espacialmente mesmo que a imagem tenha sido

afetada por uma transformação geométrica e/ou fotométrica, (b) derivação do vetor de características também chamado “descritor de regiões”. Cada vetor de características representa a informação da imagem disponibilizada na vizinhança ao redor do “ponto de interesse”. O reconhecimento do objeto pode ser então executado comparando os descritores de regiões e sua localização de configuração espacial, pertencente tanto à imagem Q quanto à imagem A . O sucesso destas técnicas se deve principalmente ao fato de não precisar da etapa da segmentação.

O objetivo deste trabalho é obter um algoritmo para reconhecimento de objetos robusto à variação de ponto de vista ou deformação em perspectiva. Em geral existem duas técnicas propostas para conseguir a invariância afim: a normalização e a simulação de pontos de vista.

- A normalização é um processo onde os pixels de uma região da imagem são padronizados, que pode ser um parâmetro de um pixel (e.g. escala e orientação do gradiente), e o objetivo é conseguir que os parâmetros extraídos daquela região sejam aproximadamente invariantes a uma transformação geométrica específica (TREIBER, M., 2010). A normalização é mais utilizada que a simulação, no entanto, os algoritmos baseados na normalização não conseguem a completa invariância a transformações afins, pois os parâmetros de escala e inclinação dos eixos da câmera não podem ser completamente padronizados (MOREL; YU, 2009a). Na literatura temos várias propostas de algoritmos que usam a técnica da normalização de parâmetros, algumas delas são:

Uma análise multi-escala da morfologia afim para extrair cantos, proposto em (ALVAREZ; MORALES, 1997). Os autores deste trabalho assumem que a continuidade de um canto é dada por um trecho de linha formado por pontos e o declive desta linha identifica o ângulo do canto, deste modo é possível recuperar a localização inicial e orientação do canto para modelar a invariância afim. Porém, os pontos de interesse em imagens de cenas reais não podem ser aproximados por cantos perfeitos.

Um método para achar características afins em pequenas regiões de interesse foi proposto por LINDEBERG e GARDING (1997), no qual os “pontos de interesse” são extraídos usando os valores máximos do segundo momento da matriz numa representação espaço-escala utilizando o detector de Harris (HARRIS; STEPHENS, 1988). A proposta de Lindeberg e Garding foi complementada por Baumberg (2000) que utilizou um detector de regiões de forma elíptica para melhorar a normalização. No entanto, os “pontos de interesse” e suas regiões associadas não conseguem a invariância na presença de transformações afins significativas.

Um algoritmo baseado em “pontos de interesse” invariantes que utiliza o detector de Harris-Laplace foi proposto por SCHAFFALITZKY e ZISSERMAN (2001). Nele se faz uma normalização da transformada afim procurando correspondências em texturas de pequenas regiões previamente segmentadas.

Um algoritmo baseado em momentos invariantes derivados de imagens binárias foi proposto por FLUSSER e SUK (1993). Uma falha presente neste tipo de proposta é a alta sensibilidade ao ruído. As duas propostas de MIKOLAJCZYK e SCHMID (2002; 2004), também baseadas em momentos invariantes, aproximam a região característica utilizando uma elipse que posteriormente é normalizada para um círculo. Nestas propostas, utiliza-se o detector Harris-affine e o Hessian-affine respectivamente.

Um algoritmo que, além da invariância afim, também apresenta robustez a oclusões parciais, baseado na transformada de Hough e complementado com a extração de contornos usando informação local, foi apresentado por ECABERT e THIRAN (2004). Esta proposta demanda pouca capacidade de memória e tempo de execução, além de considerável tolerância com respeito a deformações locais.

Um detector de regiões invariantes baseado em linhas de nível MSER (*Maximally stable extremal region*) foi apresentado por MATAS, et. al. (2004). O MSER baseasse numa sequência de linearizações de intensidade, as regiões extremas são componentes conexos de um conjunto de máximo ou mínimo nível de intensidade. O MSER procura a invariância a pontos de vista normalizando os parâmetros da

transformada afim e está definido pelas propriedades extremas de uma função de intensidade numa região e suas regiões próximas.

Recentemente, foi proposto um algoritmo que estende o conceito de regiões invariantes do MSER, denominado “Clique descritor” onde as características são agrupadas em pequenos conjuntos de regiões e co-relacionadas por um mínimo número de ocorrências (SHIN; TJAHJADI, 2010), o Clique descritor utiliza o MSER como detector de formas para combiná-lo com o descritor do SIFT e fazer melhor uso da influência da vizinhança entre regiões invariantes, uma das vantagens desta fusão é a robustez a oclusões parciais. Porém a técnica do MSER utiliza componentes de alto contraste na construção das regiões invariantes, o que constitui uma deficiência dado que as imagens de cenas naturais contêm poucas daquelas componentes.

Todos estes métodos realizam a normalização do casamento local, região, ou linha de nível que tem uma transformação afim, normaliza também a transformação no objeto padrão. Deste modo, eles conseguem eliminar o efeito da transformação. Porém, nenhum dos algoritmos reportados na literatura e baseados na normalização consegue completa robustez a transformações afins (MOREL; YU, 2009a).

- A simulação de pontos de vista é uma técnica proposta no ASIFT (MOREL; YU, 2009a), consiste na utilização da transformada afim como ferramenta para simular os pontos de vista nas imagens. Por meio desta técnica, consegue-se simular os parâmetros de inclinação da câmera para alcançar completa robustez à invariância afim. No ASIFT, esta técnica é proposta como uma extensão do SIFT (LOWE, 2004). O SIFT utiliza a normalização para conseguir invariância por translação e rotação, e uma técnica de simulação de escalas para a invariância a escala. A robustez obtida pelo ASIFT o destaca como um dos melhores algoritmos invariantes afins no estado da arte, porém apresenta dificuldades para o reconhecimento na presença de padrões repetitivos e de imagens com baixo contraste. Isto se deve ao fato de que os atributos característicos de seu “descritor de regiões”, baseado no histograma da direção dominante do gradiente centralizado nos “pontos chaves”, tende a se deteriorar na presença destes dois fatores.

Na procura da construção de um algoritmo invariante afim e robusto na presença de padrões repetitivos e baixo contraste, nossa pesquisa foi orientada na combinação de duas técnicas:

1. A simulação de pontos de vista utilizando as transformadas afins, para modelar as deformações induzidas pelos efeitos das perspectivas, proposta por Morel e You (2009a) e denominada ASIFT.
2. Os coeficientes de Fourier de projeções radiais e circulares para extrair características invariantes por rotação, escalamento, brilho e contraste, proposta por (KIM, H. Y., 2010a) e denominada FORAPRO.

O FORAPRO tem demonstrado experimentalmente que consegue superar as dificuldades que apresenta ASIFT nas aplicações do reconhecimento de objetos onde as imagens possuem padrões repetitivos e/ou grande mudança de contraste. Neste trabalho, utilizamos este fato para propor uma extensão invariante afim do FORAPRO, que denominamos de AFORAPRO.

Os algoritmos foram implementados em linguagem C/C++, utilizando os compiladores Dev-C++ e as bibliotecas para Processamento de Imagens e Visão Computacional IMG (KIM, H. Y., 2010b) e ProEikon (KIM, H. Y., 2009). Foi utilizada também a biblioteca OpenCV (INTEL, 2009) para compilação e otimização de algumas funções.

1.2 Objetivos

Explorar a potencialidade dos coeficientes de Fourier de projeções radiais e circulares na construção de um algoritmo para o reconhecimento de objetos invariante sob transformações afins, induzidas pelas deformações em perspectivas causadas pela mudança de pontos de vista da câmera, e robusto à presença de padrões repetitivos e mudanças de contraste.

O trabalho também tem os seguintes objetivos específicos:

- Descrever as principais abordagens empregadas no reconhecimento de objetos invariante afim.
- Avaliar o método proposto AFORAPRO em diversos conjuntos de imagens, que contenham variação de pontos de vista, padrões repetitivos e grande mudança de contraste.
- Comparar os desempenhos de AFORAPRO e ASIFT e analisar os resultados.

1.3 Motivação e justificativas

Nos últimos anos, uma grande quantidade de algoritmos de reconhecimento de objetos tem sido proposta, pois o reconhecimento de objetos tem-se constituído numa ferramenta de utilização geral para uma variedade de aplicações. Cada aplicação tem seus próprios requerimentos e restrições. Consequentemente, não é possível a construção de um algoritmo para atingir todos os requerimentos, precisando-se então de novas propostas para cobrir a demanda e ampliar o panorama das aplicações.

Dos algoritmos de força bruta, que procuravam um objeto presente em duas imagens fazendo comparações diretas e limitando fortemente o processo de reconhecimento, para algoritmos que utilizam invariantes, a área de processamento de imagens digitais tem conseguido uma ampliação do campo das aplicações práticas e da utilização estratégica de ferramentas matemáticas. Especificamente, nosso trabalho explora ferramentas como os coeficientes radiais da transformada de Fourier e a utilização das características da geometria afim para construir modelos que simulam os pontos de vista.

1.4 Publicação associada a esta dissertação

A seguinte publicação está associada a esta dissertação:

- No artigo (KIM, H. Y.; PÉREZ, G. A, 2010) foi proposto um método de reconhecimento de padrões invariante a transformações afins denominado AFORAPRO. Neste trabalho, foi avaliada a resposta do algoritmo para imagens com

padrões repetitivos e imagens com baixo contraste, apresentando comparações preliminares com o ASIFT. Este artigo foi escolhido como um dos melhores do VI Workshop de Visão Computacional (VI WVC) 2010.

1.5 Organização

Esta dissertação é composta de seis capítulos e está organizada da seguinte maneira:

- Reconhecimento de Objetos: No capítulo 2 se apresenta uma visão geral da temática, com uma revisão da literatura e algumas definições, conceitos e informações sobre os métodos e as ferramentas teóricas adotadas na proposta. Ao final apresentam-se uma descrição e análise comparativa dos algoritmos SIFT e FORAPRO.
- Invariância Afim: No capítulo 3 se faz uma abordagem teórica sobre a invariância afim, robustez a padrões repetitivos e mudanças de contraste. Descreve-se a técnica de “simulação de pontos de vista”, descreve-se os quatro métodos invariantes afins mais relevantes na literatura, entre os quais se destaca o ASIFT.
- Algoritmo proposto: No capítulo 4 se faz a descrição da estrutura do algoritmo proposto AFORAPRO. Também são apresentados alguns testes demonstrativos do desempenho e uma análise do tempo de execução para cada etapa.
- Resultados: No capítulo 5 são apresentados os resultados dos experimentos realizados para avaliar o desempenho do AFORAPRO. Primeiramente são testadas as melhoras do algoritmo por meio de uma comparação com FORAPRO e depois se utiliza o ASIFT como referência para testar a invariância a pontos de vista, robustez a mudanças de contraste e padrões repetitivos.
- Conclusões: No capítulo 6 são apresentadas as discussões finais e as possibilidades para dar continuidade nos trabalhos futuros.

Capítulo 2.

RECONHECIMENTO DE OBJETOS EM IMAGENS

Neste capítulo, apresentamos uma visão geral da temática abordada em nosso trabalho. Realizamos uma revisão da literatura, apresentamos algumas definições básicas como o reconhecimento de objetos e o conceito de avaliação de desempenho utilizado em nossa proposta, e ao final descrevemos e fazemos uma análise comparativa dos algoritmos SIFT e FORAPRO.

2.1 Introdução

Matematicamente, a imagem digital é definida como uma função bidimensional $I(x, y)$, com $x, y \in R^2$, onde x e y são as coordenadas espaciais que identificam um ponto, e o valor $I(x, y)$ denota a intensidade do nível de cinza da imagem naquele ponto. Uma imagem digital é uma imagem $I(x, y)$ discretizada tanto em coordenadas espaciais (amostragem) quanto na intensidade dos níveis de cinzas (quantização). Assim, uma imagem digital pode ser representada através de uma matriz na qual os índices de linhas e de colunas identificam um elemento da imagem (pixel) e o valor do elemento identifica o nível de cinza (GONZALEZ; WOODS, 2002).

De modo geral, uma imagem digital está constituída por um fundo e um ou vários objetos que ficam sobre ele, assim os objetos são definidos como regiões constitutivas que se diferenciam do fundo (e.g. uma letra ou caractere, uma forma ou silueta, um automóvel, um prédio, etc.) (THEODORIDIS; KOUTROUMBAS, 2009).

O reconhecimento de objetos envolve três níveis de processamento: filtragem inicial, extração de características e classificação. A filtragem da entrada tem o objetivo de eliminar informação desnecessária fazendo com que a entrada apresente apenas dados relevantes para o processo. A extração de características consiste na análise dos dados de entrada a fim de extrair e derivar informações úteis. O estágio final do reconhecimento é a classificação, onde através da análise das características o objeto é declarado como pertencente a uma determinada categoria (TREIBER, 2010).

O uso da técnica de “casamento de padrões” (*template matching*) baseada na extração de características locais, no contexto do reconhecimento, tem gerado um interessante progresso em termos de robustez, eficiência e qualidade dos resultados. Nesta técnica, um padrão pode ser visto como uma entidade (pixel e sua vizinhança) a partir da qual é possível extrair algum tipo de característica, que seja simbólica ou numérica (SZELISKI, 2009). O casamento de padrões envolve duas imagens (Q e A), o processo consiste em escolher alguns padrões (pontos de interesse) nas duas imagens e compará-los utilizando alguma medida de similaridade. Na Figura a seguir são apresentados os elementos básicos do casamento de padrões.

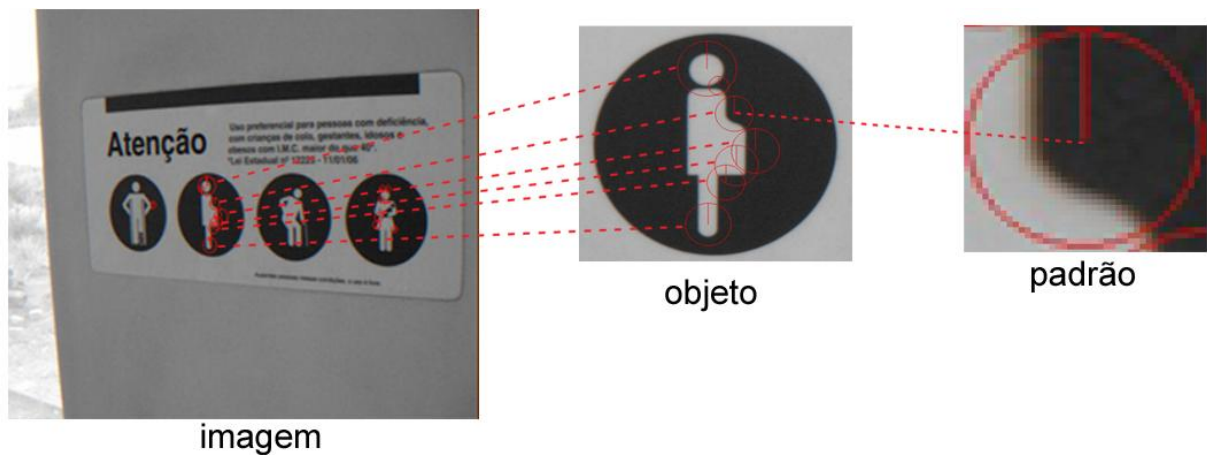


Figura 2.1 Elementos básicos no reconhecimento de objetos. As imagens apresentam o casamento de seis padrões, as linhas vermelhas entre a imagem e o objeto indicam a correspondência dos padrões. No complemento a região de um dos padrões é ampliada, o pixel central e a área dentro do círculo vermelho indicam a região característica do casamento.

O casamento de padrões, além do reconhecimento de objetos também possui outras aplicações (e.g. restauração e registro de imagens). De modo geral, os algoritmos destinados ao casamento de padrões (e.g. ASIFT) costumam usar as linhas de correspondência para indicar a posição dos padrões casados, já os algoritmos destinados exclusivamente ao reconhecimento de objetos (e.g. AFORAPRO) utilizam um rótulo sobre o objeto reconhecido, como será apresentado a seguir.

2.2 Reconhecimento de objetos

Entre as ferramentas ou recursos fundamentais da detecção de objetos no interior de uma imagem está o casamento de padrões (*template matching*), no qual uma réplica do objeto de interesse Q é comparada para cada objeto desconhecido na imagem A . Se o casamento entre o objeto desconhecido e Q satisfizer certos requerimentos, o objeto desconhecido é rotulado para indicar a presença de uma instancia de Q dentro de A . Na Figura 2.2, apresenta-se um exemplo típico dos resultados de um reconhecimento de objetos.

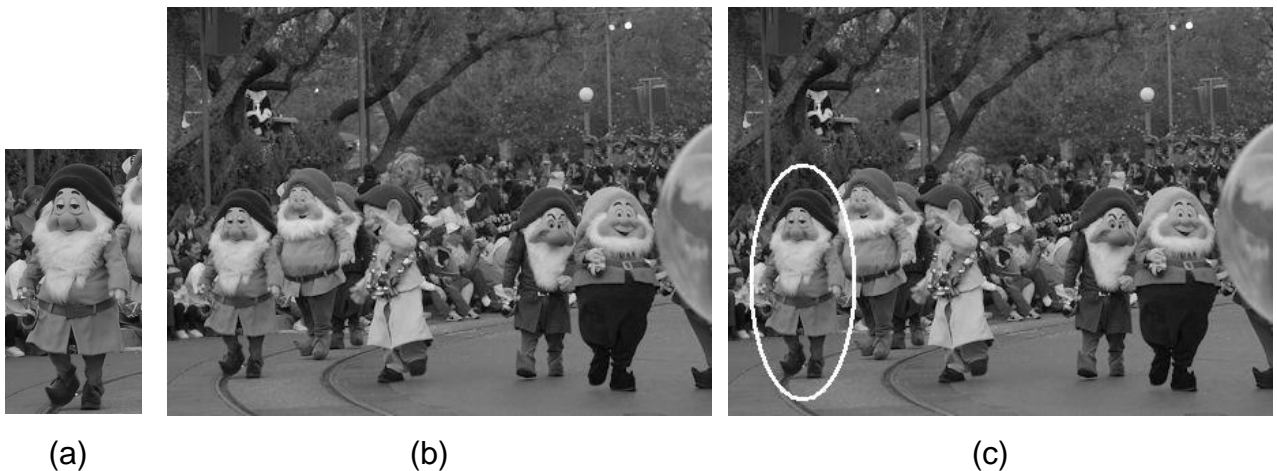


Figura 2.2 Exemplo do reconhecimento de objetos. (a) Imagem Q , (b) Imagem A , (c) resposta, rótulo branco sobre instância do objeto reconhecido.

O nível de dificuldade no reconhecimento de objetos aumenta na medida em que as imagens apresentem variabilidades que podem ser causadas devido a (TREIBER, 2010):

1. Ponto de vista; as imagens podem ser capturadas a partir de diferentes pontos de vista (direções e distâncias), este efeito pode induzir transformações em perspectivas que deformam, transladam, rotacionam e/ou mudam a escala dos objetos dentro da imagem.
2. Efeitos fotométricos; incluem as posições e distribuição das fontes de luz na cena, sua amplitude, o efeito de iluminação mútua entre objetos, a distribuição de sombras e reflexão nas superfícies dos objetos físicos. Este tipo de efeito tende a mudar drasticamente a distribuição de brilho e contraste nas imagens.

3. Disposição dos objetos; em cenas naturais os objetos raramente ficam isolados, eles normalmente se encontram parcialmente oclusos sobre algum fundo ou com outros objetos.
4. Mudança de forma; muitos objetos físicos podem manter sua identidade enquanto mudam de forma (propriedades elásticas).

O reconhecimento dos objetos é mais fácil quando se tem uma vista armazenada que coincide com o ponto de vista do objeto a ser reconhecido (TARR; BULTHOFF, 1995). Poderia se conseguir uma solução ao problema do reconhecimento de objetos se fosse armazenado um número suficiente de diferentes vistas associadas com cada objeto, para depois comparar a imagem da vista atual com todas as vistas armazenadas na memória. Existem algumas propostas que usam esta idéia, conhecidas como memórias associativas, que fazem a aproximação diretamente ao reconhecimento. Estes mecanismos, normalmente baseados em redes neurais, podem armazenar grande quantidade de padrões (Q_1, Q_2, \dots, Q_3) , e depois, atendendo a um padrão de entrada A , tentam encontrar o padrão Q_i que seja o mais similar a A (RIPLEY, B., 1996). No entanto, a simples comparação ou comparação direta entre duas imagens é insuficiente por duas razões: o espaço de todas as possíveis opções é grande demais e é possível que a imagem a ser reconhecida não possua suficiente similaridade com as imagens da memória.

A criação de uma base de dados com todas as vistas que pode adotar um objeto é algo complexo. Porém, os modelos matemáticos podem fornecer alguns tipos de soluções ou alternativas para superar este problema como, por exemplo, a utilização dos invariantes de uma configuração geométrica, que são funções, cujos valores não variam mesmo que o objeto tenha sofrido certa transformação. Entre as categorias mais destacadas no estudo de invariantes temos os momentos centrais proposto por HU (1962) e os momentos de Zernike (TEAGUE, 1980). Eles apresentam teorias baseadas em momentos aplicáveis a imagens binárias e sua álgebra de invariantes. Embora estas técnicas sejam somente invariantes a rotação, elas são amplamente

reconhecidas pela contribuição no complemento e desenvolvimento de outras técnicas como:

- Os descritores baseados em momentos invariantes (VAN GOOL, et. al. 1996), onde os momentos são calculados pelas derivadas da intensidade da imagem $I_d(x, y)$ com relação à direção d ;
- Numa outra proposta (FLUSSER; SUK, 1993), os autores demonstram que o valor de uma combinação específica de momentos invariantes em imagens binárias pode ficar constante mesmo que a região, da qual este foi derivado, tenha sofrido uma transformação afim, evidenciando que o uso dos momentos é uma alternativa para obter descritores invariantes a pontos de vista. Porém, a quantidade de características invariantes afins que se consegue desta forma é muito pequena.

Nos métodos baseados em correspondências locais, os objetos são representados por meio de um conjunto de características locais, que são extraídas das duas imagens envolvidas no processo. Depois de organizar e ordenar as características se procede ao casamento de padrões levando em conta o número de correspondências. As vantagens principais da técnica da correspondência de características locais são:

- A representação local está baseada na aparência, por isso não precisa da extração de primitivas geométricas (e.g. linhas, eixos, círculos, retângulos).
- Não requer segmentação prévia.
- A técnica se torna robusta a oclusões parciais e a mudança de fundos, pois o reconhecimento não precisa do casamento de todas as características locais.
- As variações complexas na aparência dos objetos causadas pela variação de pontos de vista e pelas condições na iluminação, podem se aproximar usando a modelagem por meio de transformações a uma escala local.

2.3 Avaliação de desempenho

Diferentes medidas que permitem a comparação de algoritmos para o reconhecimento de objetos podem ser encontradas na literatura. As mais utilizadas são as taxas de acertos e erros, as curvas de precisão, e as curvas ROC (*receiver operating characteristic*). Estas medidas são calculadas a partir da matriz de confusão ou tabela de contingência apresentada a seguir (FAWCETT, 2006):

Tabela 2.1 Tabela de contingência para avaliação de desempenho.

		<i>O objeto Q está presente na imagem A</i>	
		<i>Sim</i>	<i>Não</i>
<i>O objeto Q foi detectado na imagem A</i>	<i>Sim</i>	VP	FP
	<i>Não</i>	FN	VN

Onde:

- VP: Verdadeiro Positivo, uma instância do objeto procurado está presente na imagem analisada e é detectada pelo algoritmo.
- FP: Falso Positivo, uma instância do objeto procurado não está presente na imagem analisada e é detectada pelo algoritmo.
- FN: Falso Negativo, uma instância do objeto procurado está presente na imagem analisada e não é detectada pelo algoritmo.
- VN: Verdadeiro Negativo, uma instância do objeto procurado não está presente na imagem analisada e não é detectada pelo algoritmo.

A métrica de avaliação adotada em nosso trabalho foi a taxa de acertos, apresentada pela seguinte equação:

$$taxa\ de\ acertos = \frac{VP}{VP + FN} * 100 \quad (2.1)$$

Nossa avaliação está baseada no VP e FN, e estes dois fatores estão representados na Figura 2.3. Neste exemplo apresentam-se a imagem modelo (Q) (Figura 2.3a) e três casos de reconhecimento, o rótulo vermelho indica a área detectada. Na Figura 2.3b apresenta-se um VP, a instância de Q é detectada com sucesso. Nas imagens complementares a instância de Q não é reconhecida com sucesso: na Figura 2.3c o rótulo não especifica a localização do objeto procurado (FN) e na Figura 2.3d é rotulado um objeto errado (FN).

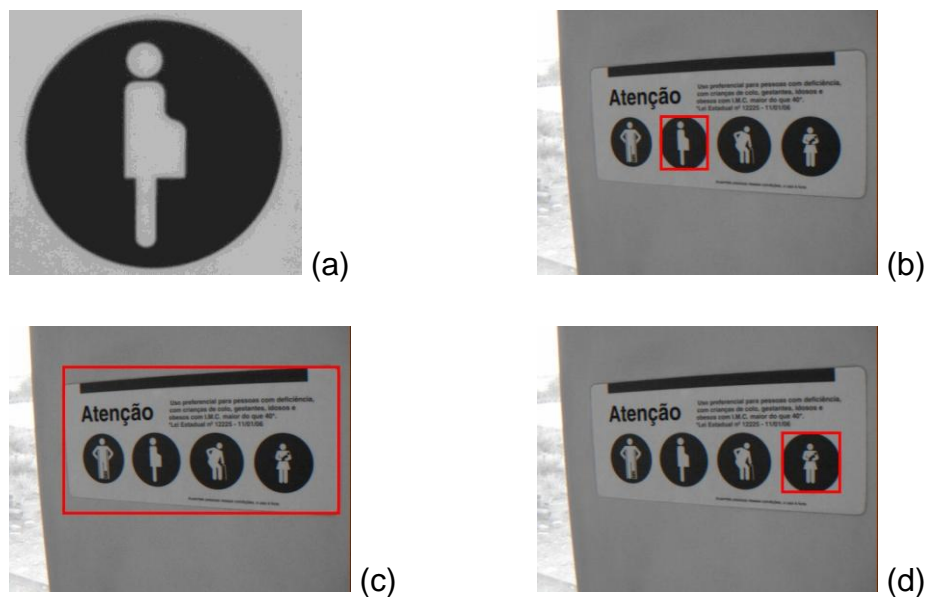


Figura 2.3 Exemplificação dos possíveis casos de reconhecimentos, (a) imagem Q , (b) VP, a instância de Q aparece e é detectada com sucesso em A , (c-d) FN, a instância de Q aparece, mas não é detectada em A .

Em nosso trabalho optamos pelo uso da taxa de acertos, pois assumimos que a imagem Q deve aparecer uma e só uma única vez na imagem A (então FP e VN nunca podem acontecer).

2.4 Revisão da literatura

Um modelo de reconhecimento de objetos é escolhido dependendo da robustez de suas características e da sua adaptação a aplicações específicas. Porém, cada área de

aplicação tem suas próprias exigências. Por exemplo, na robótica móvel a preferência é que o tempo de execução seja o mais reduzido possível mesmo que se perca um pouco na robustez do desempenho. O contrário acontece com as aplicações médicas, onde se pode ser flexível com o tempo de execução para obter um ótimo desempenho do algoritmo.

Um dos primeiros passos no desenvolvimento do reconhecimento de objetos foi o detector proposto por MORAVEC (1977), que definiu pela primeira vez o conceito de “pontos de interesse” como regiões representativas da imagem que podem ser usados para procurar regiões repetitivas em quadros consecutivos. Moravec propôs medir a variação de intensidade utilizando uma pequena janela (tipicamente, 3x3 ou 5x5 pixels), para percorrer a imagem pixel a pixel, usando as oito direções principais do pixel central (horizontal, vertical, e as quatro diagonais). Este tipo de operador pode ser classificado de acordo com a forma com que é modelada a informação da imagem, deste jeito temos dois grupos: (a) aqueles que procuram ou calculam a curvatura principal ao redor de cada ponto: (BEAUDET, 1978; KITCHEN; ROSENFELD, 1982), e (b) aqueles que utilizam a distribuição do gradiente ao redor do ponto: (HARRIS; STEPHENS, 1988; FORTNER, 1994; SHI; TOMASI, 1994), que são os mais comumente usados.

Um detector popularmente conhecido como “Detector de Harris” foi proposto por HARRIS e STEPHENS (1988), este detector apresenta forte invariância à rotação, variação de iluminação, e ruído na imagem. O grande incremento dado por Harris e Stephens, ao método de Moravec, foi a utilização de uma máscara de operador gaussiano tornando o detector menos sensível a ruídos. Onde o equivalente das pequenas variações $[u, v]$ do operador de Moravec é substituído por diferenciais $[\Delta x, \Delta y]$. Na Figura 2.4 apresenta-se um exemplo da resposta do detector de Harris, os pontos brancos na imagem são os pontos indicados como cantos pelo detector. Note que alguns pontos ficam sobre os eixos e nem todos os cantos foram detectados, a otimização deste detector depende da aplicação específica.

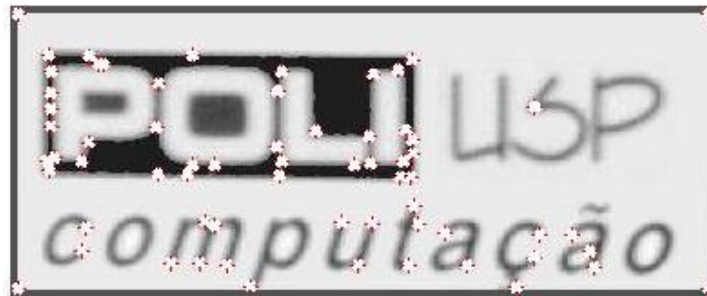


Figura 2.4 Detecção de cantos utilizando o detector de Harris. Aplicação implementada nos programas C/C++ utilizando a biblioteca ProEikon (KIM, H. Y., 2010c)

O detector de Harris possui muitas versões complementares e tem sido a base para uma grande quantidade de trabalhos na área de reconhecimento de objetos, alguns destes trabalhos são abordados no capítulo seguinte. O próprio Harris, (HARRIS, 1992), propõe uma modificação que melhora a eficiência do detector para os rastreamentos de movimento e recuperação de estruturas tridimensionais.

Em (SCHMID; MOHR, 1997) foi apresentada uma proposta que amplia mais a utilidade do detector de Harris. Eles mostram que o casamento de características locais invariantes pode ser estendido para o problema do reconhecimento de objetos no qual uma característica é compatível frente a uma grande base de dados, permitindo casamentos na presença de arbitrarias mudanças de orientação. Entre muitos outros métodos, que também conseguem invariância a rotação, temos: a proposta de TSAI e CHAING (2002) que utiliza a decomposição da transformada Wavelet; ULLAH e KANEKOI (2004) que usam a orientação de códigos de cadeia onde a idéia principal é a utilização de histogramas de direção de gradiente para descrição dos padrões; LIN, CHEN e WEI (2006) que utilizam uma combinação da transformada de Fourier e métodos paramétricos.

A invariância a escala foi introduzida por TANIMOTO (1975, 1980), WITKIN (1983), KOENDERINK (1984), e LINDEBERG (1990, 1994), a ferramenta que eles utilizaram foi chamada de “espaço de escala”. A teoria dos espaços de escala aparece pela necessidade de incorporar a noção de escala de uma forma explícita na representação da imagem, mediante a geração de uma família de imagens derivadas. Trabalhos como os de MIKOLAJCZYK, et. al. (2002) e BAUMBERG (2000) utilizarem o espaço de

escala para conseguir invariância a escala, nestas propostas é utilizado o detector de Harris em diferentes níveis do espaço de escala discretizada, para detectar pontos chaves.

Em LOWE (1999, 2001, 2004) são procurados máximos e mínimos locais num espaço de escala construído com *DoG* (*Difference-of-Gaussian*). O trabalho de Lowe denominado SIFT (*Scale Invariant Feature Transform*) é abordado na seção 2.5.

Em KLINGER (1971) foi utilizada uma pirâmide multi-resolução que consiste de um grupo de representações com sucessivos alisamentos da imagem original, organizados de forma que o número de pixels decresce com um fator constante de uma sub-imagem para outra. Alguns trabalhos se fundamentam nesta técnica e tem sido importantes em determinadas aplicações como no caso das transformadas Wavelets (MALLAT, 1989). A principal diferença entre multi-escala e multi-resolução é precisamente a variação sistemática do número de pixels que ocorre na representação multi-resolução, isso faz com que a resolução da imagem decresça gradualmente dependendo de um fator que normalmente é dois ou quatro. Alguns dos métodos que utilizam a técnica Wavelets para obter a invariância à escala são: o método denominado SURF (*Speeded up robust features*) (BAY; TUYTELAARS; GOOL, 2006) que utiliza a Haar Wavelet, e o método proposto por JU HAN e KAI-KUANG (2007) que utilizam a Gabor Wavelets.

Já os algoritmos FORAPRO (KIM, H., 2010a) e o CiRaTeFi (KIM, H.; ARAÚJO, 2007) que não são intrinsecamente invariantes a escala, utilizam uma estrutura de várias escalas num intervalo específico construindo uma matriz multi-escala de características invariantes por rotação onde se considera estar contido o parâmetro de escala procurado.

MATAS et. al. (2002) introduziram o MSER, o qual foi avaliado no trabalho de MIKOLAJCZYK (2005) como o melhor detector no momento em comparação para outros detectores baseados na extração de pontos chaves e invariantes a transformações afins, este método é abordado no Capítulo 3.

Recentemente, MOREL e YU (2009a) introduziram um método completamente invariante afim para comparação de imagens, denominado ASIFT (*Affine-SIFT*), onde o

destaque é procurar as correspondências dos pontos chaves em imagens que simulam as distorções de pontos de vista. O ASIFT será abordado no Capítulo 3. Recentemente, GAO et.al. (2010) apresentam um método que combina a técnica de simulação de pontos de vista com uma técnica de extração de características estáveis, onde o objetivo é acelerar o processo de reconhecimento diminuindo o número de vistas simuladas, este método baseia-se no modelado do objeto utilizando as características da cor.

2.5 Descrição dos algoritmos SIFT e FORAPRO

Os algoritmos SIFT e FORAPRO constituem a base do ASIFT e AFORAPRO, e portanto é com base nesses dois algoritmos que devemos iniciar a análise da nossa proposta.

2.5.1 Algoritmo SIFT (*Scale Invariant Feature Transform*) (LOWE, 1999, 2001, 2004)

O SIFT é um método que permite fazer correspondência de pontos entre imagens, e tem sido amplamente empregado em tarefas relativas à detecção de padrões. O conjunto de atributos locais extraídos pelo SIFT é invariante a transformações de rotação, translação e escalamento, é insensível a presença de ruídos e parcialmente robusto a alterações de brilho, contraste, mudanças de ponto de vista e oclusões parciais.

2.5.1.1 Detecção de pontos candidatos (detector *DoG*)

A ideia principal desta etapa é aumentar a dimensão da imagem (x, y) para aproveitar as estruturas através de escalas distintas, assim uma imagem gera um espaço de escala tridimensional onde a terceira dimensão é a escala (x, y, σ) .

O passo de uma escala fina para uma grossa simplifica o processo porque elimina informação que não faz diferença. A função que satisfaz a representação da imagem em diferentes escalas é o kernel Gaussiano (Equação 2.3). A representação de uma

imagem $I_s(x, y, \sigma)$ numa escala específica s pode ser calculada pela convolução da imagem original $I(x, y)$ com $G(x, y, \sigma)$, assim:

$$I_s(x, y) = G(x, y, \sigma) * I(x, y) \quad (2.2)$$

Onde

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2.3)$$

A escolha de σ define a escala s . Para localizar os pontos de interesse (chamados por Lowe como “keypoint”), são extraídos os máximos e mínimos locais no espaço de escala. Para este fim, são calculadas as diferenças entre imagens de escalas próximas utilizando a convolução da imagem com a função de Diferenças Gaussianas (*DoG*) (Equação 2.4), onde os parâmetros σ das Gaussianas diferem por algum fator constante k (tipicamente no intervalo de 1,1 a 1,4).

$$D(x, y, \sigma) = DoG * I(x, y) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (2.4)$$

O espaço de escala pode ser explorado variando o valor de σ . Na Figura a seguir é apresentado um exemplo da extração de máximos (vermelhos) e mínimos (verdes) utilizando *DoG*.



Figura 2.5 Extremas (pontos chaves) achadas utilizando *DoG*. Aplicação implementada utilizando a biblioteca *ProEikon* (KIM, H. Y., 2010c) e *SIFT* disponível em LOWE (2010).

Lowe 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.

2.5.1.2 Descritor de regiões do SIFT

O objetivo desta etapa é fazer com que cada “ponto chave” possua um vetor de descrição $(x, y, orientação, \sigma)$. O conceito deste descritor é motivado pela visão biológica, pela observação de que certos neurônios no córtex visual primário respondem a uma particular orientação do gradiente.

Cada região ao redor de um ponto chave é dividida em sub-regiões retangulares de tamanho 3×3 . Em seguida, os gradientes da intensidade são calculados e suas orientações são acumuladas num histograma para cada sub-região separadamente (ver Figura 2.6). Cada ponto ganha um peso dependendo da magnitude do gradiente e da distância ao centro da região.

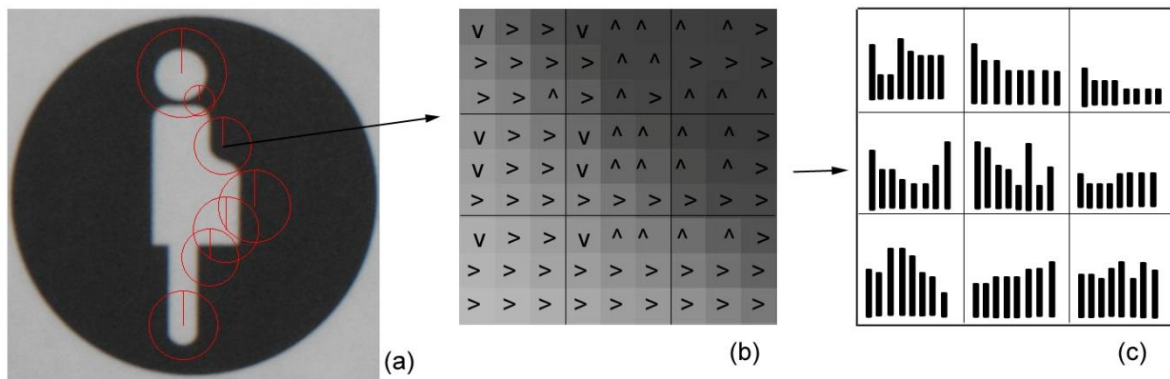


Figura 2.6 Descritor de regiões características ao redor de um ponto chave. (a) Pontos chaves extraídos pelas DoG e filtrados, (b) direção de intensidade do gradiente em sub-regiões, (c) histogramas de acumulação.

Os picos na orientação do histograma correspondem a direções dominantes para os gradientes locais. O maior pico no histograma e aqueles acima de 80% de seu valor são usados para se definir a orientação de cada ponto chave. No caso de se ter múltiplos picos de magnitudes similares, são criados diferentes pontos chaves na mesma localização, mas com diferentes orientações. Para se definir com maior precisão a orientação, uma parábola é interpolada entre os três (3) valores do histograma próximos de cada pico, e então é interpolada a posição do pico.

2.5.1.3 Reconhecimento de objetos com SIFT

Para uma aplicação no reconhecimento de objetos, o SIFT apresenta cinco fases que podem ser sumarizadas assim: (a) detecção de pontos de interesse procurando extremas no espaço escala da imagem construído pelo detector DoG , (b) extrair a localização dos pontos chaves (x, y, σ) , para refinar o espaço escala, (c) determinar a orientação dominante para cada ponto chave $(x, y, orientação, \sigma)$, (d) construção do descritor. Estas quatro fases asseguram a invariância do SIFT com respeito a rotação, translação e escala.

(e) A quinta e última fase consiste em casamento de padrões: A correspondência entre duas imagens depende dos pontos em comum encontrados entre elas. Os descritores são extraídos das duas imagens e considera-se que o descritor d_Q extraído da imagem Q casa-se com o descritor d_A extraído da imagem A , se a distância euclidiana entre d_Q e d_A é pelo menos 60% menor que a distância entre d_Q e qualquer outro descritor da imagem A . A complexidade desta tarefa depende se o objeto está-se procurando numa base de dados ou está-se procurando numa cena simples. Para achar o objeto numa cena, o SIFT utiliza a técnica do “vizinho mais próximo”. O casamento de vários descritores pode ser agrupado para detectar um objeto, usando a transformada de Hough generalizado (BALLARD, 1981).

Na Figura 2.7 apresenta-se o resultado do reconhecimento de objetos utilizando o SIFT, onde as linhas vermelhas indicam o casamento dos pontos detectados nas imagens. Note que na Figura 2.7b o objeto tem uma pequena mudança no ponto de vista, mas o casamento é realizado com sucesso.

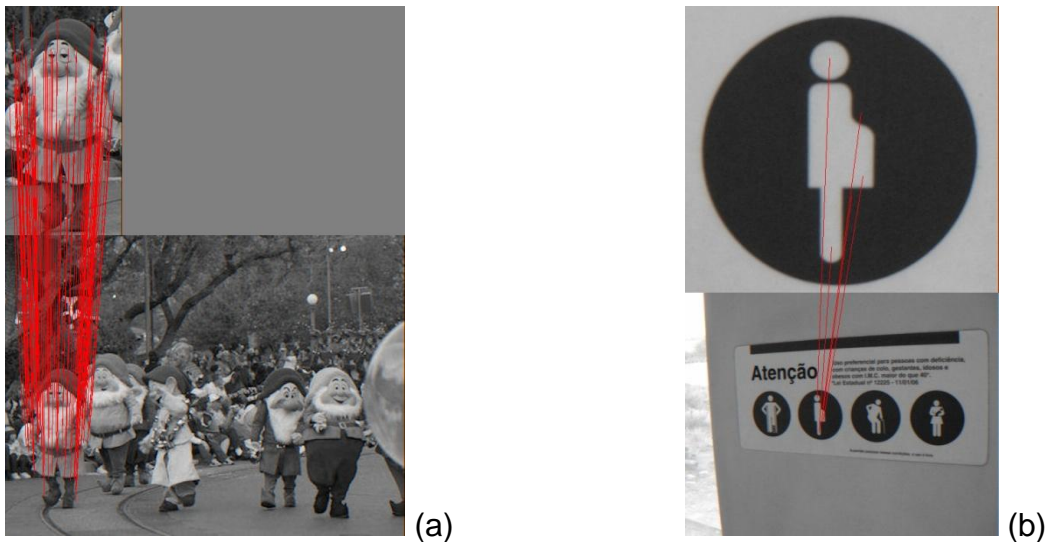


Figura 2.7 Desempenho do SIFT. Aplicação implementada utilizando a biblioteca ProEikon (KIM, H., 2010c) e SIFT disponível em LOWE (2010). (a) palhaço, (b) sinalização.

Várias extensões do SIFT já foram apresentadas; entre elas pode-se destacar a proposta por MIKOLAJCZYK e SCHMID (2005), eles também fazem comparação de vários algoritmos e demonstrarem o excelente desempenho do algoritmo SIFT.

2.5.2 Algoritmo FORAPRO (KIM, H., 2010a)

FORAPRO (*Fourier Coefficients of Radial Projections*) é um algoritmo de reconhecimento de padrões intrinsecamente invariante por rotação, brilho e contraste. Ele é invariante à mudança de escala dentro de um intervalo especificado. FORAPRO calcula os coeficientes de Fourier de projeções radiais usando FFT (*Fast Fourier Transform*). A partir desses coeficientes são calculadas, para cada pixel, a “orientação canônica” e as características invariantes por rotação, brilho e contraste.

Para buscar uma imagem Q em A , são extraídas várias sub-imagens circulares $T_i \subset Q$, calculando as suas características invariantes por rotação e procurando em A os pixels com características semelhantes. Essas correspondências são agrupadas através da transformada de Hough, resultando na detecção final. Para obter a invariância à escala, redimensiona-se Q para várias escalas e procura-se cada instância de Q na imagem A . Este processo é rápido, pois as características da

imagem A só precisam ser calculadas uma única vez. A seguir se faz uma descrição das características do FORAPRO.

2.5.2.1 Coeficientes de Fourier de projeções radiais (CFPR)

Dada uma imagem A em níveis de cinza, a projeção radial $R_\alpha^\lambda(A(x, y))$ é definida como o nível de cinza médio dos pixels da imagem A localizados sobre a linha radial com um dos vértices em (x, y) , onde λ é o comprimento e α a inclinação, assim:

$$R_\alpha^\lambda(A(x, y)) = \frac{1}{\lambda} \int_0^\lambda A(x + t \cos(\alpha) + t \sin(\alpha)) dt \quad (2.5)$$

O vetor de M projeções radiais discretas no pixel $A(x, y)$ com raio λ pode ser obtido variando o ângulo α , assim:

$$R_\alpha^\lambda(A(x, y)) = R_{2\pi m/M}^\lambda(A(x, y)) \quad \text{onde } 0 \leq m < M \quad (2.6)$$

A Figura 2.8 mostra um esquema das projeções radiais e circulares traçadas sobre um pixel central.

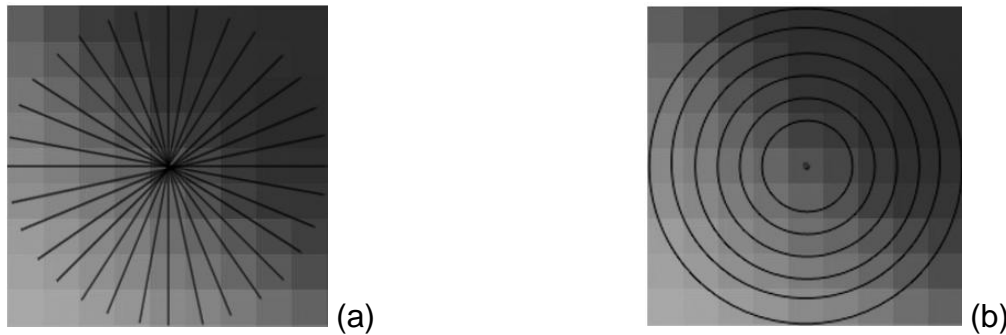


Figura 2.8 Projeções utilizadas pelo FORAPRO, (a) Projeções radiais com $M = 36$, (b) projeções circulares.

O vetor de projeções radiais $R_m^\lambda[A(x, y)]$ caracteriza a vizinhança de raio λ de $A(x, y)$. Se A rotaciona este vetor desloca-se circularmente e o k -ésimo CFPR pode-se calcular assim:

$$r[k] = \sum_{m=0}^{M-1} R_m^\lambda[A(x, y)] \exp(-j2\pi k m/M), \quad 0 \leq k < M \quad (2.7)$$

Os CFPR podem ser calculados diretamente fazendo convolução de A com núcleos apropriados k , sem a necessidade de se calcular explicitamente as projeções radiais.

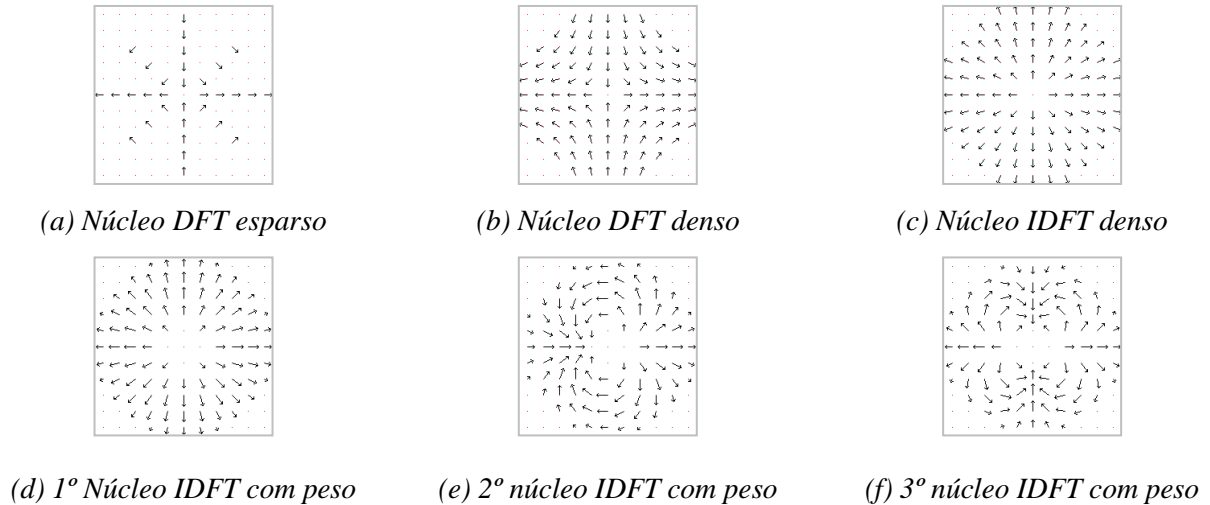


Figura 2.9. Núcleos radiais e circulares utilizados pelo FORAPRO. DFT (Discrete Fourier Transformation), IDFT (Inverse Discrete Fourier Transformation).

A Figura 2.9a mostra o núcleo DFT esparso K (com $M = 8$ ângulos) tal que a convolução $A * \check{K}$ resulta no primeiro coeficiente de Fourier das oito projeções radiais, onde $\check{K}(x, y) = K(-x, -y)$.

$$(A * \check{K})(x, y) = \sum_p \sum_q A(p, q) K(p - x, q - y) \quad (2.8)$$

A convolução $A * \check{K}$ pode ser calculada através de multiplicações no domínio da frequência:

$$A * \check{K} \Leftrightarrow A\check{K} \quad (2.9)$$

Onde A e \check{K} são respectivamente as transformadas de Fourier discretas de A e \check{K} .

O núcleo esparso da Figura 2.9a não leva em conta a maioria dos pixels externos e assim não gera características robustas. Para superar este problema, “núcleo DFT denso” da Figura 2.9b pode ser usado. Ele preenche todos os pixels do núcleo, exceto o elemento central. Os elementos não-nulos deste kernel são definidos:

$$\mathfrak{R}_k[x, y] = \exp(-jk\angle(x + yj)) \quad (2.10)$$

Onde k é a ordem do coeficiente de Fourier e $\angle(\cdot)$ é o ângulo do número complexo. O filtro linear usando este núcleo não tem um significado intuitivo. Usando a transformada discreta de Fourier inversa (IDFT), o resultado da convolução adquire um significado: torna-se análogo ao gradiente. A Figura 2.9c mostra o núcleo obtido usando IDFT e $k = 1$. Para tornar o núcleo mais “estável”, isto é, para fazer o resultado da convolução menos sensível a perturbações como translação ou rotação por sub-pixel, são atribuídos pesos menores aos pixels nas regiões externa e central, resultando em núcleo com peso apresentado na Figura 2.9d. O núcleo radial resultante é:

$$\mathfrak{R}_k[x, y] = \sqrt{r(\lambda - r)} \exp(-jk\angle(x + jy)) \quad (2.11)$$

Onde $r = \sqrt{x^2 + y^2}$ e λ é o raio do núcleo. Os núcleos usados para obter coeficientes IDFT para $k = 2$ e $k = 3$ estão mostrados nas Figuras 2.9e e Figura 2.9f. É chamado a convolução de $A(x, y)$ com a reflexão dupla do k -ésimo núcleo radial de “ k -ésimo coeficiente radial” e é denotado por $r_k(A(x, y))$ ou simplesmente r_k . Também são chamados $\angle r_k$ e $|r_k|$, respectivamente, “ k -ésimo ângulo radial” e “ k -ésima magnitude radial”.

2.5.2.2 Orientação canônica

A característica que discrimina a rotação local ou a “orientação canônica” é o primeiro ângulo radial $\angle r_1$. A orientação canônica $\angle r_1(A(x, y))$ indica a direção local de $A(x, y)$ na vizinhança de raio λ . Se $A(x, y)$ rotaciona θ radianos, o vetor de projeções radiais $R_m^\lambda[A(x, y)]$ desloca circularmente θ radianos e conseqüentemente o primeiro ângulo radial $\angle r_1(A(x, y))$ rotaciona pelo mesmo ângulo. Uma mudança de brilho/contraste não altera a orientação canônica.

2.5.2.3 Vetor de Magnitudes Radiais

As magnitudes radiais são invariantes por rotação, pois se $A(x, y)$ rotaciona então o vetor de projeções radiais $R_m^\lambda[A(x, y)]$ desloca circularmente, e um deslocamento circular

não muda as magnitudes dos coeficientes IDFT (somente os seus ângulos são alterados). As magnitudes radiais $|r_k|$, $k \geq 1$, também são invariantes a brilho, pois a mudança de brilho somente afeta o coeficiente r_0 . Finalmente, as razões entre magnitudes radiais são invariantes a contraste, pois uma alteração no contraste multiplica pelo mesmo fator todos os coeficientes radiais. FORAPRO utiliza o vetor de magnitudes radiais v_{rm} que leva em conta magnitudes de todos os coeficientes radiais até o grau K :

$$v_{rm} = \upsilon[|r_1|, |r_2|, \dots, |r_k|] \quad (2.12)$$

Onde υ significa L^1 -versor que consiste em dividir cada elemento do vetor pelo seu comprimento $L^1 = |r_1| + |r_2| + \dots + |r_k|$. Este vetor é invariante por rotação, brilho e contraste. A função distância Λ entre dois v_{rm} 's está limitada ao intervalo $[0,1]$ e é definida como:

$$\Lambda(v_{rm}(A(x, y)), v_{rm}(T(x_0, y_0))) = \frac{1}{2} \|v_{rm}(A(x, y)), v_{rm}(T(x_0, y_0))\|_1 \quad (2.13)$$

2.5.2.4 Vetor de ângulos radiais

Se $A(x, y)$ rotaciona θ radianos, então o vetor de projeções radiais $R_m^\lambda[A(x, y)]$ desloca circularmente θ radianos e o k -ésimo coeficiente radial $r_k(A(x, y))$ é multiplicado por $e^{jk\theta}$. Além disso, a mudança de brilho/contraste não afeta o $\angle r_k$. Assim, a diferença entre $\angle r_k$ e $k \angle r_1$ é invariante por rotação e brilho/contraste. Denominamos este valor de diferença entre ângulos radiais k e 1:

$$dra_k = \text{mod}(\angle r_k - k \angle r_1, 2\pi), \quad k \geq 2 \quad (2.14)$$

Esta característica é calculada no módulo 2π . Todas as diferenças de ângulos radiais até ordem K são empacotadas numa estrutura denominada vetor de ângulos radiais v_{ra} :

$$v_{ra} = [dra_2, dra_3, \dots, dra_k] \quad (2.15)$$

Este vetor é invariante por rotação, brilho e contraste. A distância Λ entre dois v_{ra} 's é definida como:

$$\begin{aligned} \Lambda[v_{ra}(A(x, y), v_{ra}(T(x_0, y_0)))] &= (w_2/\pi w_t)\Phi[dra_2(A(x, y), dra_2(T(x_0, y_0)))] + \\ & (w_3/\pi w_t)\Phi[dra_3(A(x, y), dra_3(T(x_0, y_0)))] + \dots + \\ & (w_k/\pi w_t)\Phi[dra_k(A(x, y), dra_k(T(x_0, y_0)))] \end{aligned} \quad (2.16)$$

Onde $w_k = 1/k$ com $(2 \leq k \leq K)$, $w_t = w_2 + w_3 + \dots + w_k$ e Φ é a diferença entre dois ângulos: $\Phi(a, b) = \min(\text{mod}(a - b, 2\pi - \text{mod}(a - b, 2\pi)))$ esta função distância está limitada ao intervalo $[0, 1]$.

2.5.2.5 Invariância a rotação (vetor de características circulares)

Nesta etapa utilizam-se características derivadas das projeções circulares junto com as derivadas de projeções radiais, o núcleo circular utilizado é:

$$\mathfrak{S}_l[x, y] = \begin{cases} \frac{1}{2\pi r} \exp\left(\frac{jlr}{\lambda}\right), & \text{se } r > 0 \\ 0,73, & \text{se } r = 0 \end{cases} \quad (2.17)$$

Onde $r = \sqrt{x^2 + y^2}$ e λ é o raio do núcleo. O peso $1/2\pi r$ é o inverso do perímetro do círculo onde o pixel está situado. O peso para $r = 0$ é definido como 0,73 para distribuir uniformemente os ângulos da imagem complexa resultante da convolução.

A convolução de $A(x, y)$ com a dupla reflexão do primeiro núcleo circular é denotada $c_l(A(x, y))$ ou simplesmente c_l . O “vetor de características circulares” abaixo leva em conta os componentes reais e imaginários de todos os coeficientes circulares até o grau L : $v_{cf} = v[re(c_1), im(c_1), re(c_2), im(c_2), \dots, re(c_L), im(c_L)]$, onde v significa L^1 -versor, e “re” e “im” são respectivamente as partes, real e imaginária, do número complexo. Este vetor é invariante por rotação, brilho e contraste. A distância Λ entre dois v_{cf} 's está limitada ao intervalo $[0, 1]$ e é definida como:

$$\Lambda(v_{cf}(A(x, y), v_{cf}(T(x_0, y_0)))) = \frac{1}{2} \|v_{cf}(A(x, y)) - v_{cf}(T(x_0, y_0))\|_1 \quad (2.18)$$

2.5.2.6 Combinação das características

Nas subseções anteriores, obtiveram-se três classes de características invariantes por rotação, brilho e contraste usando K radial e L coeficientes circulares e empacotando-as em três vetores: v_{rm} , v_{ra} , e v_{cf} . Esses três vetores serão agrupados numa outra estrutura denominada “vetor de características”:

$$v_f = (v_{rm}, v_{ra}, v_{cf}) \quad (2.19)$$

A função distância Λ entre dois vetores de características é definida como uma média ponderada da distância dos três vetores constituintes e expressada assim:

$$\Lambda[v_f(A(x, y), v_f(T(x_0, y_0)))] = (w_m/\pi w_t) \Lambda[v_{rm}(A(x, y), v_{rm}(T(x_0, y_0)))] + (w_a/w_t) \Lambda[v_{ra}(A(x, y), v_{ra}(T(x_0, y_0)))] + (w_c/w_t) \Lambda[v_{cf}(A(x, y), v_{cf}(T(x_0, y_0)))] \quad (2.20)$$

Onde $w_m = w_a = K - 1$, $w_c = 2L - 1$ e $w_t = w_m + w_a + w_c$.

2.5.2.7 Transformada de Hough

Para detectar Q em A com robustez, propõe-se localizar N sub-imagens circulares $T_1, \dots, T_N \subset Q$ em A . Existe a possibilidade de que a detecção de uma única imagem gere erro. Porém, se muitas sub-imagens concordam em apontar um pixel como o ponto de casamento, a probabilidade de erro é minimizada.

O algoritmo usando a transformada de Hough generalizada se torna: dada uma imagem Q , extrair N sub-imagens “apropriadas”. Para cada sub-imagem circular T_i , é gerada a imagem de distâncias D_i , onde cada pixel $D_i(x, y)$ é a distância entre as características de $A(x, y)$ e T_i :

$$D_i(x, y) = \Lambda(v_f(A(x, y), v_f(T_i(x_0, y_0)))) \quad (2.21)$$

Localiza-se o conjunto C_i de pixels candidatos ao casamento selecionando n_c pixels com as menores distâncias. A partir dos pixels candidatos e das suas orientações canônicas, é possível calcular o pixel central de Q em A . Incrementa-se a matriz de

acumulação da transformada de Hough. Os casamentos são aqueles pixels da matriz de acumulação com os maiores valores.

2.5.2.8 Invariância a escala

FORAPRO não é intrinsecamente invariante à escala, a invariância a escala é obtida para um intervalo especificado. A imagem Q é redimensionada ou escalonada (variação do tamanho) considerando um conjunto de fatores de escala num intervalo especificado t e obtendo Q_1, Q_2, \dots, Q_i . O casamento é realizado procurando, por comparação direta, cada imagem re-escalada Q_i na imagem A , executando o processo descrito para cada comparação.

Na Figura 2.10 apresenta-se o resultado do reconhecimento de objetos utilizando o FORAPRO. As imagens utilizadas são as mesmas com as quais foi testado o SIFT, Na Figura 2.10b, o objeto apresenta uma pequena perspectiva e mesmo assim os dois algoritmos o reconhecem com sucesso.

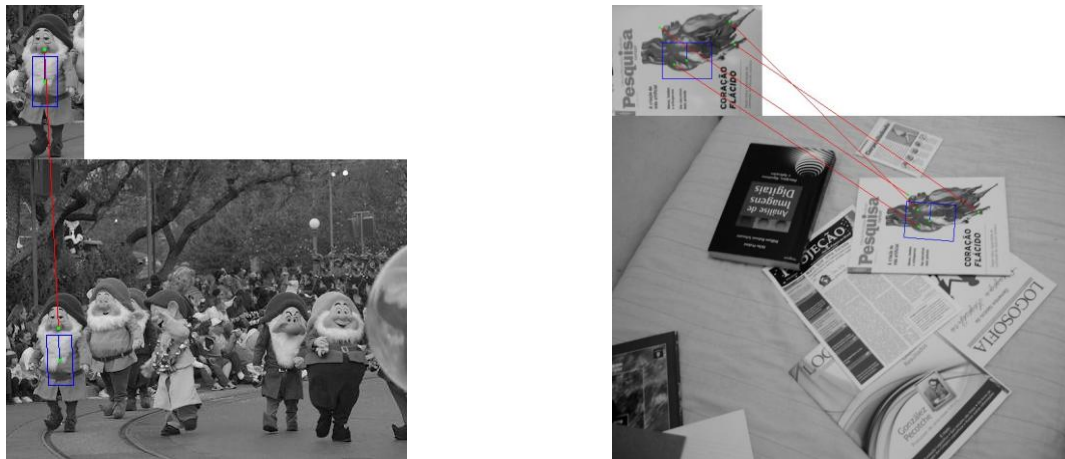


Figura 2.10 Teste do desempenho de FORAPRO (imagens redimensionadas pelo FORAPRO)

A resposta do FORAPRO apresenta um pequeno retângulo ao redor do centro do objeto para complementar as linhas do casamento de pontos. As linhas vermelhas são utilizadas para fazer uma ligação (casamento) nos pontos detectados, no FORAPRO o

número de casamentos é normalmente menor do que no SIFT pois este número depende de uma variável escolhida pelo usuário.

2.5.3 Análise comparativa do SIFT e FORAPRO

Estes dois algoritmos são baseados na extração de características locais para achar correspondências entre os objetos das duas imagens, os dois utilizam “pontos de interesse”, mas suas técnicas são muito diferentes:

- SIFT é invariante por translação, rotação, completamente invariante a escala e apresenta pouca robustez a mudanças de brilho e contraste, e padrões repetitivos. FORAPRO é invariante por translação, rotação, invariante a escala dentro de um intervalo especificado e robusto a mudanças de brilho e contraste, e padrões repetitivos.
- SIFT detecta “pontos chaves” invariantes a escala, depois acha a direção do gradiente de cada ponto chave para torná-los invariantes a rotação e mudanças de iluminação, finalmente constrói descritores para realizar o casamento. FORAPRO detecta características invariantes por rotação e mudanças de iluminação para todos os pixels da imagem, constrói os descritores baseado em “vetores de características”, redimensiona a imagem Q para varias escalas num intervalo especificado obtendo a invariância a escala, e finalmente realizar o casamento procurando cada imagem Q redimensionada em A .
- SIFT utiliza a distribuição do gradiente de pequenas regiões em diferentes escalas para determinar um conjunto de pontos-chave (*keypoints*) que permitem realizar o casamento. A base deste processo são as mudanças extremas da intensidade, de preto-branco ou branco-preto (ver Figura 2.11a), por isso os pontos chaves são afetados pela presença de padrões repetitivos e baixo contraste. FORAPRO não tem estas dificuldades porque em seu processo de extração de características invariantes envolve o valor médio de várias trajetórias circulares e radiais sobre a vizinhança do pixel (ver Figura 2.11b).

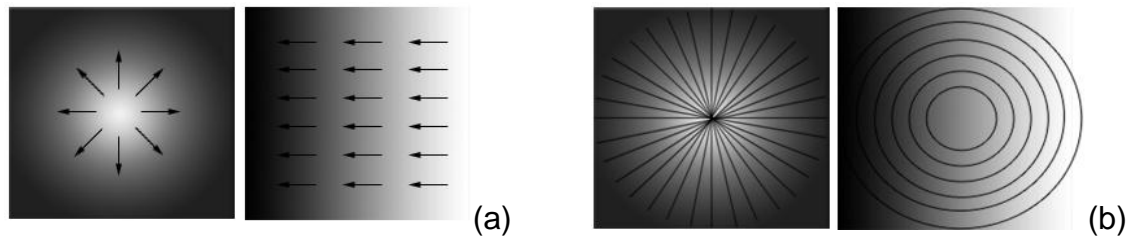


Figura 2.11 *Vizinhança envolvida na construção de invariantes a mudanças de brilho e contraste, e presença de padrões repetitivos. (a) SIFT, orientação do gradiente, (b) FORAPRO, vetor soma de projeções radiais e circulares.*

- No teste de desempenho apresentado na Figura 2.12 percebe-se como o SIFT falha na presença de padrões repetitivos (Figura 2.12a) e baixo contraste (Figura 2.12b), já o FORAPRO nas mesmas condições reconhece o objeto com sucesso.

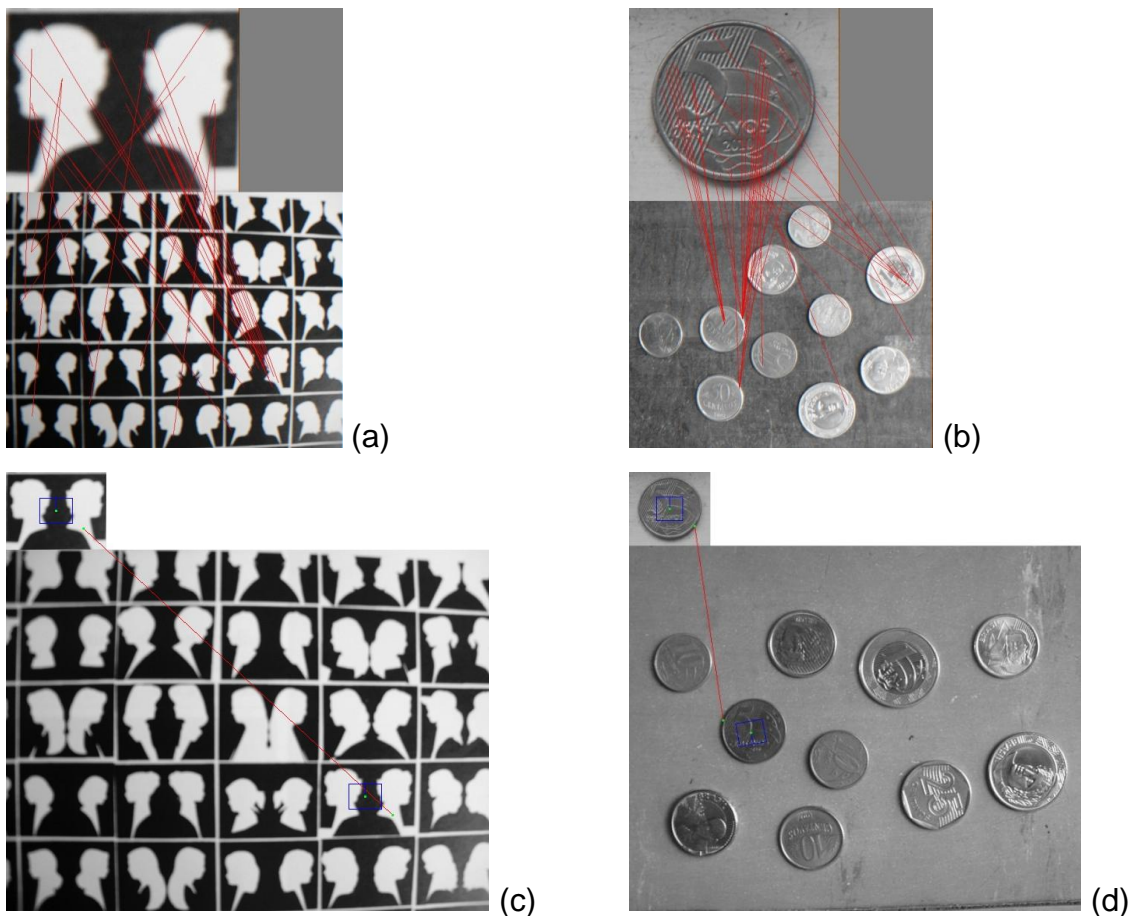


Figura 2.12 *Robustez a padrões repetitivos e baixo contraste. (a-b) SIFT falha, (c-d) FORAPRO reconhece. O tamanho das imagens A e Q foram 650x600 e 450x400. Os algoritmos redimensionam o tamanho das imagens na saída. Aplicação implementada utilizando a biblioteca ProEikon (KIM, H. Y., 2010c) e SIFT disponível por LOWE (2010).*

Capítulo 3.

CARACTERÍSTICAS INVARIANTES E TRANSFORMAÇÃO AFIM

Neste capítulo apresentamos a definição dos conceitos de transformações afins, padrões repetitivos, baixo contraste e pontos de vista. Também, apresentamos as técnicas de simulação de pontos de vista e normalização de regiões. Na sequência descreveremos quatro dos métodos mais relevantes na literatura (MASER, Harris-affine, Hessian-affine e ASIFT).

3.1 Introdução

O objetivo deste capítulo é apresentar as definições adotadas para os conceitos fundamentais de nosso trabalho e abordar o estado da arte do reconhecimento de objetos invariante a pontos de vista.

3.2 Transformações afins

Uma transformação afim de uma imagem $I_1(x_1, y_1)$ é um mapeamento que leva essa imagem a outra imagem $I_2(x_2, y_2)$ seguindo a seguinte equação:

$$I_2(x_2, y_2) = A * I_1(x_1, y_1) + T \quad (3.1)$$

Onde $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ é uma matriz inversível e $T \in R^2$ é o vetor de translação. As transformações afins podem ser decompostas:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & -\text{sen} \phi \\ \text{sen} \phi & \cos \phi \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \quad (3.2)$$

Onde S_x e S_y são fatores de escala para os eixos x e y respectivamente, k é um fator de não ortogonalidade entre eixos, ϕ é o ângulo de rotação sobre o plano, e e f são fatores de translação. Desenvolvendo esta equação obtemos:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} S_x (\cos \phi + k \cdot \text{sen} \phi) & S_x (-\text{sen} \phi + k \cdot \cos \phi) \\ S_y \text{sen} \phi & S_y \cos \phi \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \quad (3.3)$$

Que também pode ser expressa assim (SZELISKI, 2009):

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \quad (3.4)$$

Onde

$$\begin{aligned} a &= S_x (\cos \phi + k \cdot \text{sen} \phi) & b &= S_x (-\text{sen} \phi + k \cdot \cos \phi) \\ c &= S_y \text{sen} \phi & d &= S_y \cos \phi \end{aligned} \quad (3.5)$$

Resumindo, a transformação afim de uma imagem $I(x, y)$ é uma transformação linear com seis graus de liberdade (2-escalamento, 2-rotação e 2-translação) e pode ser descrita como:

$$I(x, y) \rightarrow I(ax + by + e, cx + dy + f) \quad (3.6)$$

Uma transformação afim no espaço euclidiano possui as seguintes propriedades:

- Preserva a relação de colinearidade entre pontos, isto é, três pontos colineares continuam sendo colineares após a transformação. Como consequência desta propriedade, a transformação afim preserva o paralelismo das retas.
- Preserva a razão das distâncias ao longo de uma linha, isto é, para pontos colineares distintos p_1, p_2 e p_3 , a razão $|p_2 - p_1|/|p_3 - p_2|$ é preservada.

As transformações afins incluem rotação, mudança de escala uniforme e não uniforme, assim como suas combinações. A Figura 3.1 apresenta exemplos de transformações afins obtidas utilizando a Equação 3.4.

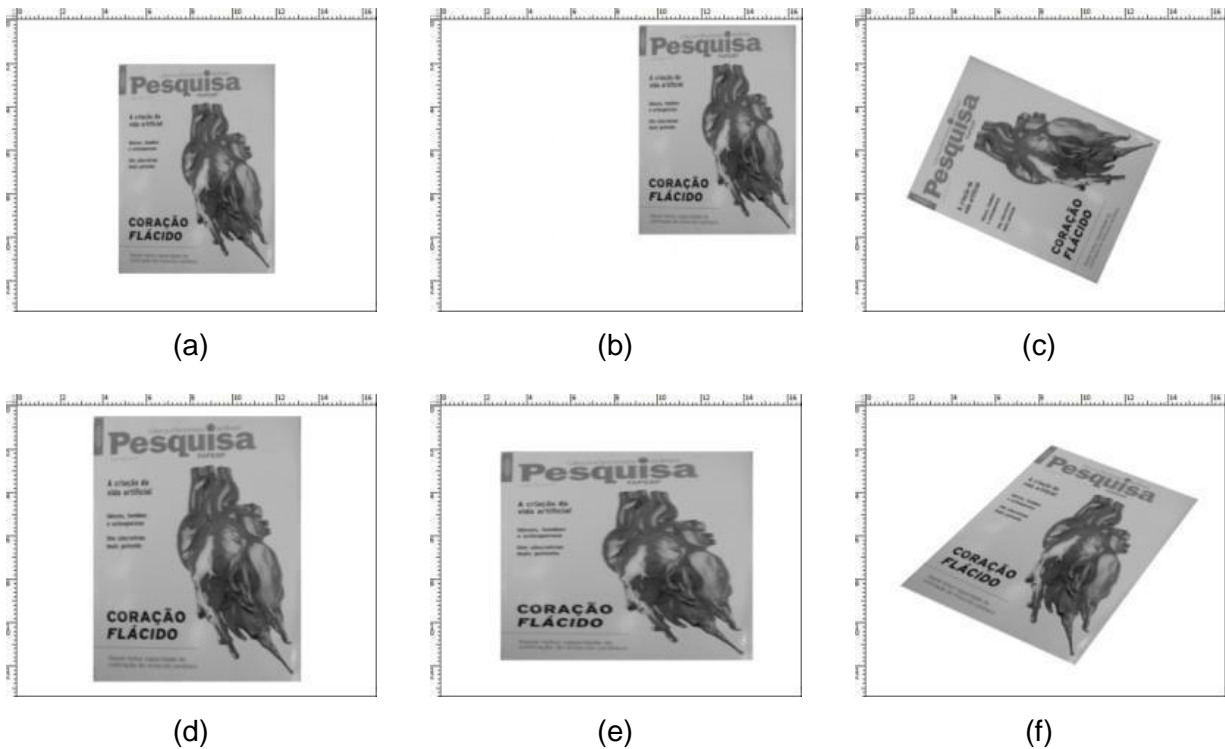


Figura 3.1 Transformações afins; a) imagem fonte, b) translação, c) rotação, d) mudança uniforme de escala, e) mudança não uniforme de escala, e f) combinação das anteriores.

Uma transformação afim é qualquer transformação que preserva a colinearidade e relações de distâncias (por exemplo, o ponto médio de um segmento de linha continua sendo o ponto médio após a transformação). Nesse sentido, afim indica uma classe especial de transformações projetivas onde os objetos conservam algumas de suas propriedades do espaço afim para o plano real da imagem. Uma transformação afim é também chamada de afinidade.

Geometricamente a transformação afim em R^2 é um caso especial de transformação em perspectiva que transforma um plano da imagem em outro plano equivalente. A principal diferença entre uma transformação em perspectiva e uma transformação afim são os efeitos não-lineares, que causam a distorção das formas de uma perspectiva. Assim uma afinidade igualmente transforma círculos em elipses e retângulos em paralelogramos, mas preserva a relação de colinearidade entre pontos e a razão das distâncias ao longo de uma linha (TREIBER, 2010).

3.3 Invariância a mudanças de contraste

Uma forma comum de definir o contraste numa imagem é por meio do *rms* (*root-mean-square*), expressado matematicamente assim:

$$rms = \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2} \quad (3.7)$$

Onde x_i é o valor normalizado ($0 \leq x_i \leq 1$) do nível de cinza no pixel, \bar{x} é a média normalizada de todos os níveis de cinza na imagem e n o número de pixels da imagem (FREDERIC; JEAN-MICHEL; ROBERT, 2010; ELI, 1990; GONZÁLEZ; WOODS, 2004). Uma definição do contraste local faz referência à mudança de contraste em regiões específicas e inclui os conceitos de alto e baixo contraste modelados na Figura 3.2.

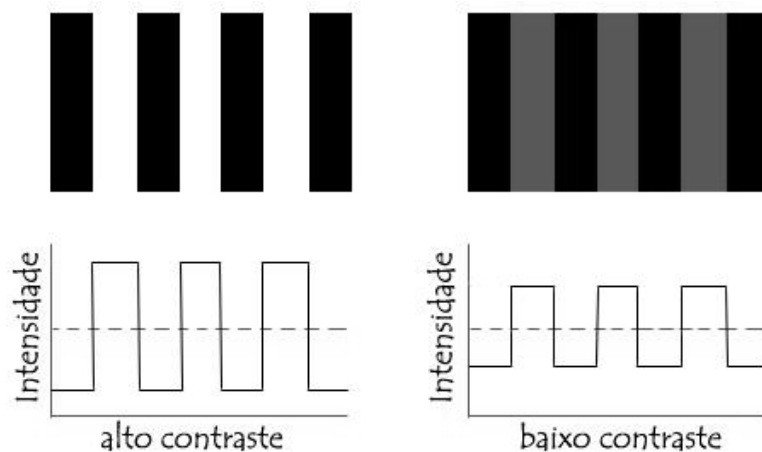


Figura 3.2 *Varição do contraste em função da intensidade.*

A variação do contraste nas imagens normalmente acontece por fatores tais como a variação da iluminação, mudança do fundo e efeitos de sombras. Na Figura 3.3 são ilustradas imagens com mudanças de contraste em decorrência dos fatores mencionados. Dada a impossibilidade para manter um controle constante sobre estes fatores a invariância a mudanças de contraste adquire importância no reconhecimento de objetos.

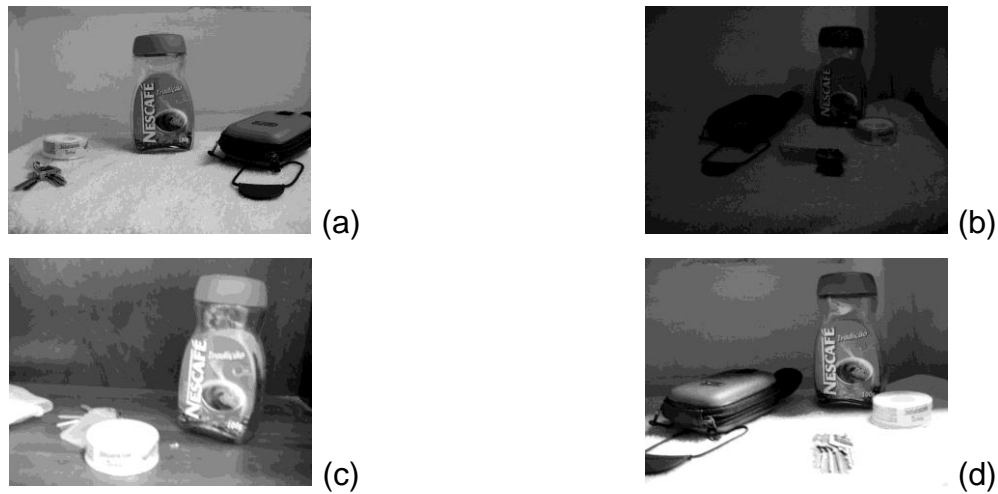


Figura 3.3 Fatores que influenciam as mudanças de contraste: (a) Imagem com alto contraste, (b) escassa iluminação, (c) mudança de fundo, (d) efeito de sombras.

Definição 1:

Sejam duas imagens, imagem Q e imagem A . Nós consideramos que, um algoritmo de reconhecimento de objetos é invariante a contraste se consegue achar instâncias de Q em A independentemente do contraste da instância.

3.4 Robustez a padrões repetitivos

Uma imagem de cenas reais pode conter muitos elementos que se repetem (e.g. janelas na fachada, azulejos no chão ou árvores no parque), estes elementos ou padrões repetitivos representam um problema no processo de casamento de padrões porque computacionalmente eles são tratados como elementos independentes e, dado que eles são individualmente indistinguíveis, incrementam o número de casamentos errados (DOUBEK, et. al., 2010; WENDEL; DONOSER; BISCHOF, 2010). Um caso típico de padrões repetitivos são as *self-similarities*, definidas nos casos onde o formato do objeto é aproximadamente similar à estrutura do fundo ou então similar ao formato de outros objetos presentes na cena, e.g. na Figura 3.4.

Entre as técnicas que têm buscado a robustez a padrões repetitivos temos o trabalho de BALLESTER e GONZÁLEZ (2004). Sua técnica é válida para imagens com abundantes texturas, mas não para estruturas simples. No trabalho de PODBREZNIK e POTOCHNIK (2010), os autores propõem uma combinação do ASIFT com uma etapa de

segmentação para extrair regiões invariantes afins que apresentam robustez a padrões repetitivos visando a implementação de um algoritmo para reconhecer objetos em imagens de estruturas e construções (e.g. janelas, paredes e fachadas).

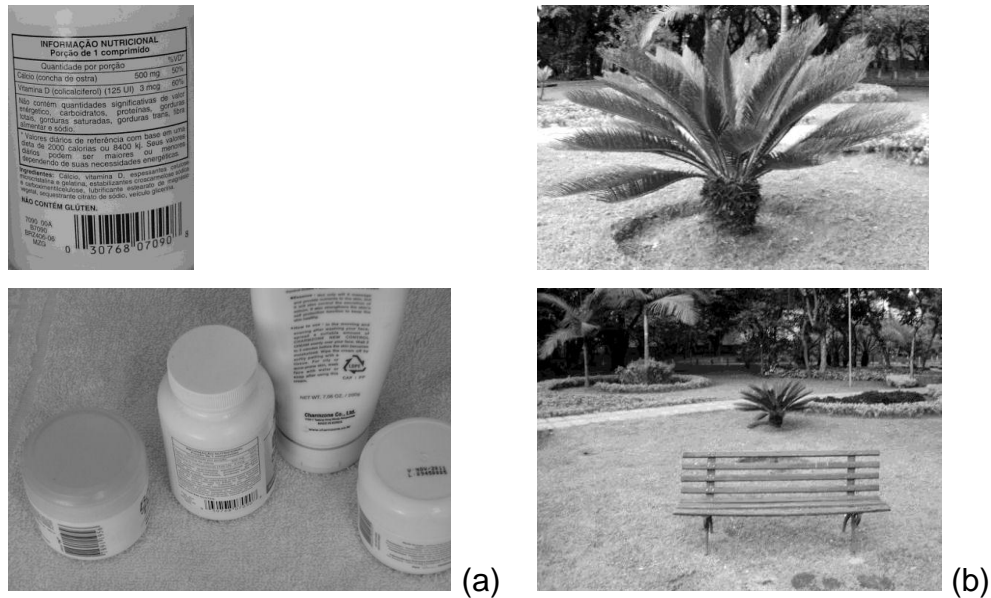


Figura 3.4 Padrões repetitivos: (a) as letras da imagem Q aparecem repetidamente nos outros objetos, (b) a estrutura do formato do objeto procurado se repete no fundo da imagem.

Nem sempre os padrões repetitivos representam problemas, alguns trabalhos os utilizam como uma ferramenta para facilitar a detecção e construção de descritores: (SHECHTMAN; IRANI, 2007) apresentam uma proposta para o casamento de dados visualmente complexos usando as *self-similarities*, eles correlacionam uma área central com o ponto de interesse que envolve uma ampla região e usam valores de máxima correlação centrados numa função de logaritmo polar binária como descritores, demonstrando que a forma é um fator importante no casamento de padrões.

Definição 2:

Sejam duas imagens, imagem modelo Q e imagem a analisar A , onde n instâncias de Q aparecem em A . Dizemos que A possui padrões de Q repetidos se algumas sub-partes de Q aparecerem em A repetindo-se no complemento da imagem (fora das n instâncias). Consideramos que um algoritmo de reconhecimento de objetos é robusto

a padrões repetitivos se consegue achar as n instâncias de Q em A , mesmo que A possua padrões de Q repetidos.

3.5 Invariância a pontos de vista

O ponto de vista faz referência à posição espacial do fotógrafo com respeito à cena, na hora da captura da imagem ou aos ângulos de inclinação e rotação da câmera respeito ao quadro fotografado (PRATT, 2007). Como descrito na seção 2.1, uma imagem digital é uma projeção bidimensional de uma cena tridimensional do mundo real. Desta forma, um objeto pode ser fotografado de múltiplos pontos de vista e a possibilidade de duas imagens capturarem o mesmo ponto de vista é baixa, devido a diversos fatores que podem condicionar a captura (e.g. distância objeto-câmera e orientação de posicionamento da câmera). Um efeito da variação de ponto de vista pode ser observado na Figura 3.5, onde são apresentados dois pontos de vista do mesmo terreno (vista frontal e vista lateral). Note como o panorama visual muda de um ponto de vista para o outro.

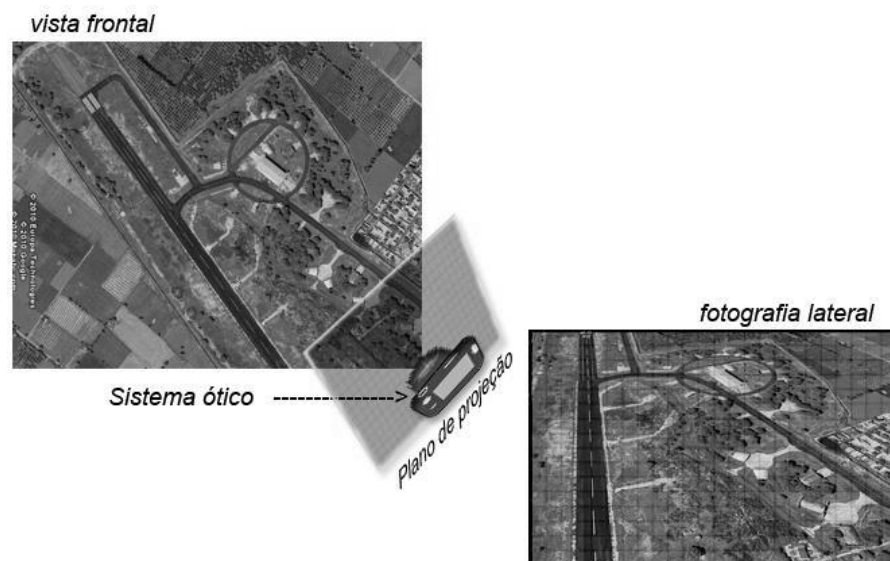


Figura 3.5 Mudança da visual induzida pela variação do ponto de vista da câmera.

Como pode se observar na Figura 3.5, a mudança do ponto de vista da câmera introduz uma deformação nos objetos. Esta deformação pode dificultar o reconhecimento do objeto procurado.

Definição 3:

Sejam duas imagens, imagem modelo Q e imagem a analisar A . Consideramos que um algoritmo de reconhecimento de objetos é invariante a pontos de vista se consegue achar instâncias de Q que aparecem com diferentes pontos de vista na imagem A .

Algumas propostas para a construção de algoritmos invariantes a pontos de vista baseiam-se no seguinte postulado:

A aparente deformação da vista plana de um objeto, induzida pela mudança no ponto de vista, é uma transformação em perspectiva que pode ser localmente modelada por uma transformação afim em cada região da imagem (TUYTELAARS; GOOL, 2000; HSIAO, et. al., 2007; WU; YANG, 2008; MOREL; YU, 2009a).

A Figura 3.6 apresenta uma deformação em perspectiva, induzida pelo ponto de vista adotado pela câmera, onde os retângulos tornam-se trapezóides, mas localmente pode-se interpretar como uma transformação afim levando em conta que cada ladrilho é aproximadamente um paralelogramo. Observe que o trapezóide preto grande no chão era originalmente um retângulo, mas tornou-se trapezoidal devido à transformação em perspectiva.

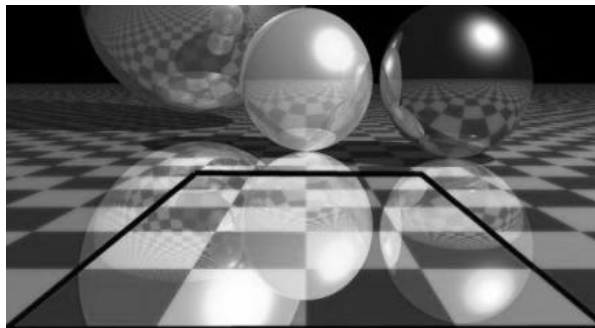


Figura 3.6 *Uma transformação em perspectiva pode ser localmente aproximada por uma transformação afim. Cada ladrilho deformado é aproximadamente um paralelogramo, embora o retângulo preto deformado seja claramente trapezoidal.*

Assim pode-se procurar instâncias da imagem Q que aparecem com deformações em perspectiva na imagem A , utilizando sub-imagens Q_i modeladas com transformações

afins. Na literatura são duas as técnicas destacadas na procura de invariância a pontos de vista, a técnica de normalização de regiões e a de simulação de pontos de vista.

3.5.1 Normalização de regiões

A técnica mais utilizada para construir algoritmos invariantes afins é a normalização de regiões, que pode ser vista como a introdução de um sistema adicional de referência na imagem. A normalização dos parâmetros da translação ou da rotação apresenta boa viabilidade, praticamente o processo se reduz a transladar ou rotacionar de volta um ponto (x, y) para outro ponto especificado, levando a conta o centro do objeto ou uma direção principal em conjunto com o sistema adicional de referência. A normalização dos parâmetros da escala ou da inclinação é muito mais complexa, no entanto existem alguns métodos que exploram este tipo de normalização. Na Figura 3.7 apresenta-se um exemplo da técnica de normalização utilizada para obter a invariância afim pelo método do Hessian-affine (MIKOLAJCZYK; SCHMID, 2002) o qual é abordado mais adiante.

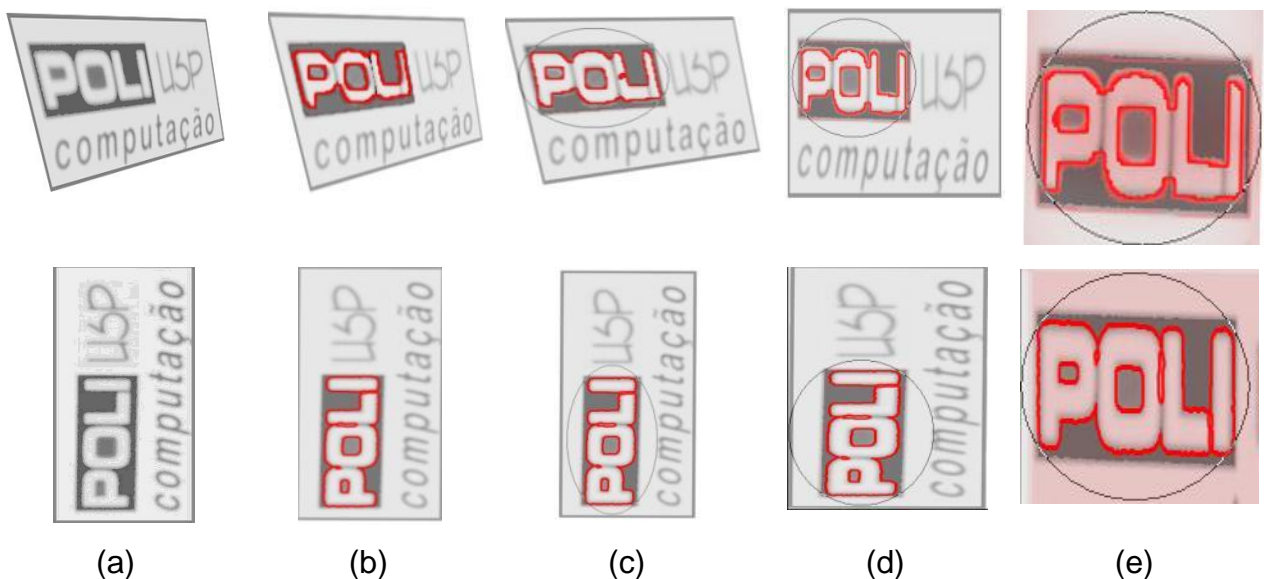


Figura 3.7 Processo de normalização de regiões. Apresentam-se duas instâncias do mesmo objeto diferenciadas pela mudança no ponto de vista: a) imagens fonte, b) detecção da região do objeto, c) traço elíptico na área detectada, d) normalização geométrica para o círculo, e) normalização fotométrica e ajuste da rotação. Aplicação disponível em MIKOLAJCZYK (2010a).

Alguns métodos que utilizam a normalização para procurar a invariância a pontos de vista são:

- Uma abordagem para recuperar a transformação, utilizando estruturas agrupadas de vértices e ângulos modeladas na imagem (THOMPSON; MUNDY, 1987);
- O reconhecimento da forma plana do objeto por alinhamento de pontos triplos (HUTTENLOCHER; ULLMAN, 1987), esta proposta requer um alto custo computacional devido à complexidade do algoritmo.
- Alguns métodos (e.g. CYGANSKI; et. al., 1987) utilizam características globais para procurar a invariância, mas o estes método precisam de toda a região do objeto para o correto reconhecimento, ou seja, não reconhecem objetos parcialmente oclusos.
- Algumas propostas (e.g. TAUBIN; COOPER, 1992; RAHTU; SALO, 2005) baseadas em características globais utilizam os momentos invariantes para procurar a invariância afim, mas apresentam alta sensibilidade ao ruído.
- Numa análise multi-escala para extração de cantos utilizando morfologia matemática afim (ALVAREZ; MORALES, 1997; DERICHE; GIRAUDON, 1993), o canto é representado por um extremo local do operador diferencial com resposta da segunda derivada na direção ortogonal do gradiente, assume-se que a evolução de um canto é dada por uma série linear de pontos e que a inclinação desta linha identifica o ângulo do canto, isto permite calcular na imagem a localização inicial e a orientação do canto para construir um modelo invariante afim, o grande problema deste método apresenta-se nas imagens de cenas naturais as quais não podem ser aproximadas por um modelo de canto perfeito.
- Outro método baseado na extração de cantos e eixos (TUYTELAARS; GOOL, 1999) reduz a região de busca utilizando o detector de pontos de Harris, cada ponto movimenta-se ao longo de seus dois eixos mais próximos formando um paralelogramo, depois uma função baseada em intensidades e um detector de eixos são utilizados para processar a região do paralelogramo e extrair um modelo invariante afim, este método é robusto com imagens sintéticas, mas não para imagens de cenas naturais.
- Existem algumas propostas baseadas em elementos de textura que procuram regiões invariantes afins (e.g. RUBNER; TOMASI, 1999; SCHAFFALITZKY;

ZISSERMAN, 2001). Primeiro, extrai-se uma pequena região afim de uma textura na imagem utilizando para isso um detector multi-escala. A seguir, a região é normalizada utilizando um histograma de distribuição de intensidades para conseguir o descritor invariante afim. Na sequência, os descritores são distribuídos na forma de uma assinatura contendo um grupo central e valores relativos. Por fim, as assinaturas são comparadas utilizando algoritmos de classificação por distâncias. Estes métodos são altamente sensíveis às mudanças de iluminação. Eles são ótimos para a restauração e classificação de imagens, mas não para reconhecimento de objetos.

Uma limitação de todos os métodos baseados em normalização (MOREL; YU, 2009a) é a não-comutação entre o borrimento ótico e a transformação afim. O processo de normalização não elimina exatamente a deformação, porque $A^{-1}G_1AI_0 \neq G_1I_0$. As diferenças são proporcionais ao tamanho do filtro de borrimento A^{-1} utilizado, desse modo quando as deformações incluem grandes escalamentos ou grandes inclinações na imagem e deve-se usar um grande filtro de borrimento para compensar a transformação afim (A), o processo de normalização apresenta falhas no reconhecimento (I_0 imagem fonte, G_1 fator de correção gaussiano).

3.5.2 Simulação de pontos de vista (MOREL; YU, 2009a)

Esta técnica baseia-se na utilização dos parâmetros de orientação da câmera presentes na interpretação geométrica do sistema plano do objeto e câmera fotográfica:

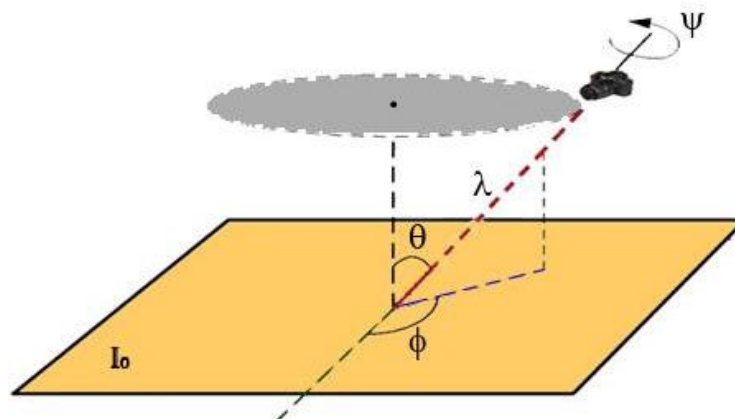


Figura 3.8 Parâmetros geométricos envolvidos na captura da imagem.

A Figura 3.8 descreve a interpretação geométrica de uma decomposição afim, onde o plano I_0 é a vista frontal do objeto físico, ψ é o ângulo de rotação da câmera sobre seu eixo ótico, λ é o parâmetro de escala (distância câmera-ponto central do plano), ϕ é o ângulo de rotação da câmera sobre o plano da imagem, e θ é o ângulo de inclinação medido entre a normal do plano da imagem e o eixo ótico da câmera. Assume-se que a vista frontal está determinada pelos valores; $\lambda = 1, \theta = 0, \phi = \psi = 0$.

Este modelo transforma o núcleo da Equação 3.1 (a matriz A) assim:

$$A = H_\lambda R_1(\psi) T_t R_2(\phi) = \lambda \begin{bmatrix} \cos \psi & -\text{sen} \psi \\ \text{sen} \psi & \cos \psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & -\text{sen} \phi \\ \text{sen} \phi & \cos \phi \end{bmatrix} \quad (3.8)$$

Onde H_λ é a matriz de escala, com $\lambda > 0$, os R_i são as rotações, $\phi \in [0, \pi)$ que é suficiente para cobrir todas as transformações, e T_t é a mudança de inclinação com $t = 1/\cos \theta$. O parâmetro t é chamado de tilt (grau de inclinação de uma vista para outra) e é utilizado com duas definições: tilt absoluto, que é a diferença de inclinação de uma imagem para sua vista frontal, e tilt de transição, que é a medida de inclinação entre dois pontos de vista de uma imagem. Em nosso trabalho consideramos somente o tilt absoluto.

A simulação de pontos de vista utiliza os valores de θ e ϕ para simular todas as possíveis distorções causadas pela mudança de orientação do eixo ótico da câmera. A escolha correta do intervalo de valores destes dois parâmetros garante que um pequeno número de imagens simuladas seja o suficiente para ter correspondência com todas as distorções presentes nas imagens capturadas.

Os valores de θ associados ao tilt t , seguem uma série geométrica $1, a, a^2, \dots, a^n$ para $a > 1$. Com valores de $a = \sqrt{2}$ e $n \geq 5$ escolhidos experimentalmente por Morel e Yu. Com o valor de $t_{\max} \approx 4\sqrt{2}$, correspondente a um ângulo $\theta_{\max} = \arccos(1/t_{\max}) \approx 80^\circ$. Os valores de ϕ para cada tilt seguem uma série aritmética $0, b/t, \dots, kb/t$ onde $b \approx 72^\circ$ e $kb/t < 180^\circ$.

A Figura 3.9 apresenta o esquema de distribuição dos diferentes pontos de vista que são simulados matematicamente, onde cada ponto é semelhante a uma posição da câmera no momento da captura. Na Figura 3.9a utiliza-se uma semi-esfera para esquematizar a posição de três valores de $t = 2, 2\sqrt{2}, 4$, cada um destes valores introduz uma linha por ângulo θ e um ponto por ângulo ϕ . Por exemplo, no valor de $t = 2$ se faz um semicírculo com $t = 1/\cos \theta \rightarrow \theta = 60^\circ$ e posicionam-se nesta linha cinco pontos que correspondem aos valores de $\phi = 0, 36^\circ, 72^\circ, 108^\circ, 144^\circ$.

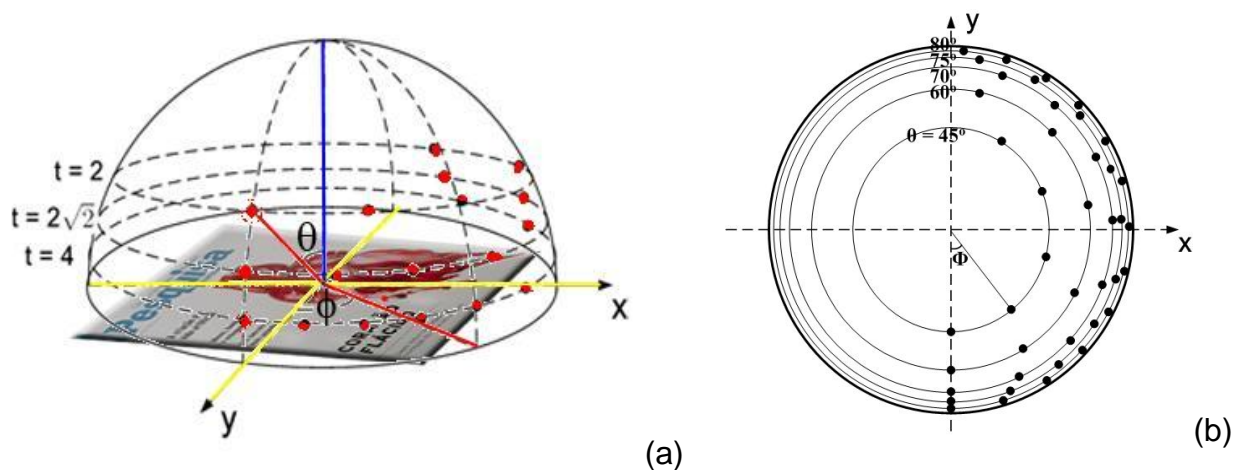


Figura 3.9 Modelo de distribuição de pontos de vista, calculados pelos valores de θ e ϕ , (a) três valores do tilt sobre suas respectivas linhas, (b) vista esférica superior com 41 pontos correspondentes ao intervalo de tilt [1,32]. Imagens adaptadas de (MOREL; YU, 2009a).

Na Figura 3.9b se modela a posição dos pontos de simulação distribuídos numa semi-esfera e vistos desde a parte superior. Note que são poucos (41), mas eles são suficientes para proporcionar cobertura às possíveis variações de pontos de vista, mantendo uma correta distribuição dos intervalos de tolerância.

Um exemplo da técnica de simulação de pontos de vista é apresentado na Figura 3.10, onde o objetivo é simular o ponto de vista que apresenta a instância da imagem Q contida na imagem A . Este exemplo foi executado com a técnica de “simulação de pontos de vista” implementada no algoritmo proposto. Forem simulados 27 pontos de vista $\left(t = \sqrt{2}^4\right)$ dos quais são apresentados só seis (Figura 3.10c a Figura 3.10h).

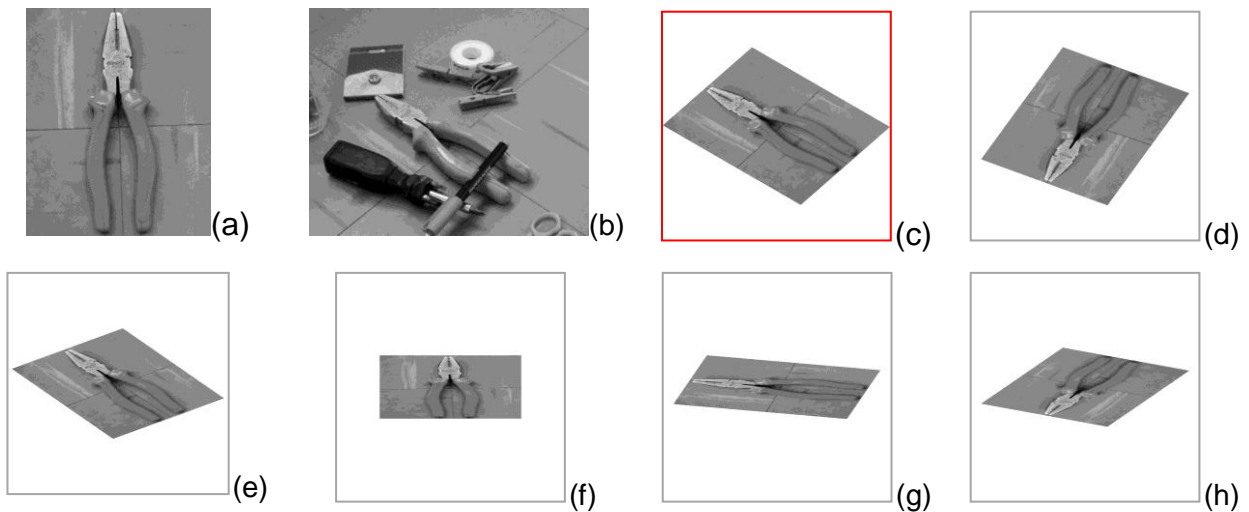


Figura 3.10 Exemplo da técnica de simulação de pontos de vista. (a) Imagem Q , (b) Imagem A , (c) Q_i simulado que melhor corresponde com o ponto de vista da instância de Q contida na imagem A , (d-h) outros Q_i gerados na simulação.

O algoritmo do ASIFT (MOREL; YU, 2009a) utiliza esta técnica e reporta uma ampla superioridade sobre seus concorrentes, conseguindo uma invariância a transição de tilt de $(4\sqrt{2})^2 = 32$, que é um valor bem superior aos valores equivalentes para outros métodos (MSER, Harris-affine e Hessian-affine) reportados na literatura. O ASIFT, que é descrito mais adiante, combina a técnica de simulação de pontos de vista com o algoritmo SIFT. Utiliza o SIFT para obter a invariância a 4 dos 6 parâmetros da transformada afim (2-translação, 1-rotação e 1-escalamento) e complementa com os 2 parâmetros proporcionados pela simulação de pontos de vista (inclinação e rotação da câmera) para obter completa invariância afim.

3.6 Estado da arte da invariância a pontos de vista

A seguir são descritos quatro métodos que procuram a invariantes afins. Três que utilizam a normalização e um que utiliza a simulação de pontos de vista (ASIFT). Visto que uma transformação afim possui seis graus de liberdade, deve-se procurar a invariância para cada um destes parâmetros.

3.6.1 MSER (*maximally stable extremal regions*) (MATAS, et. al., 2002, 2004)

Trata-se de um detector de regiões invariantes que simplifica o casamento, baseado numa sequência de limiarizações da imagem. As regiões extremas são componentes conexos de um conjunto de máximo ou mínimo nível de intensidade. O MSER procura a invariância a pontos de vista normalizando os seis parâmetros da transformada afim e é definido pelas propriedades extremas de uma função de intensidade numa região e suas regiões próximas: seja $I(p)$ uma função que retorna os valores de intensidade, onde p é um ponto na imagem, de um conjunto $I = \{0,1,2,\dots,255\}$. Uma região de máxima intensidade R_{\max} é definida por:

$$R_{\max} = \{p \mid I(p) > I(q) \text{ onde } \forall p \in R, \forall q \in \partial R\}, \quad (3.9)$$

onde R representa uma região na imagem e ∂R são suas regiões próximas. As regiões de mínima intensidade R_{\min} são definidas pela condição oposta $I(p) < I(q)$. Na prática, as regiões extremas R_m são obtidas por limiarização da imagem de forma que uma R_m é dividida em múltiplas regiões ou unida a outras regiões em relação a variados valores de limiar, isto produz uma sequência de R_m circundadas (umas dentro das outras), que por sua vez são utilizadas para determinar as regiões invariantes que serão aquelas que satisfaçam à condição de estabilidade de:

$$d(\varepsilon_i) = \frac{|\varepsilon_{i+\Delta}| - |\varepsilon_{i-\Delta}|}{|\varepsilon_i|}, \quad (3.10)$$

Onde $|\cdot|$ denota o número de elementos num grupo e Δ é um pequeno incremento. Para uma sequência de R_m circundadas, $\varepsilon_1 \subset \varepsilon_2 \subset \dots \subset \varepsilon_k$, a i -ésima extrema região ε_i , onde $1 < i < k$, é selecionada como uma região máxima estável quando ε_i é um mínimo local da Equação 3.10. Um resultado da detecção do MSER inclui regiões binarizadas com áreas que não mudam significativamente enquanto o valor do limiar apresenta pequenas alterações.

A matriz de covariância do MSER determina uma região invariante de forma elíptica. A máxima estável ε_i é representada por uma matriz 2×2 ,

$C_i = |\varepsilon_i|^{-1} \sum_{p \in \varepsilon_i} (p - m_i)(p - m_i)^T$, onde m_i é a posição média de ε_i , e a isotropia é determinada pela razão de dois autovalores de C_i . Os autovalores de C_i e m_i definem uma janela de referência local nas regiões invariantes, a qual é usada para procurar a vizinhança local.

A normalização do MSER é um processo para transformar regiões invariantes elípticas, de diferente orientação e escala, para sub-imagens $N_p \times N_p$, na procura de um casamento robusto. De acordo com SHIN e TJAHDADI (2010) a matriz de covariância pode ser decomposta para: $C_i = U \text{diag}(\lambda_1, \lambda_2) U^T$, onde $UU^T = I$ e $\text{diag}(\lambda_1, \dots, \lambda_i)$ é a diagonal da matriz quadrada. Para fazer C_i isotrópico é preciso a transformação;

$$\bar{C}_i = \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}\right) U^T C_i U \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}\right), \quad (3.11)$$

onde \bar{C}_i é a matriz de covariância normalizada. Assim um ponto normalizado \bar{p} é obtido usando $\bar{p}_i = s \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}) U^T p$, onde s é o fator de escala e p é um ponto pertencente a uma região invariante elíptica. Um teste do desempenho do MSER é apresentado na Figura 3.11:



(a)



(b)

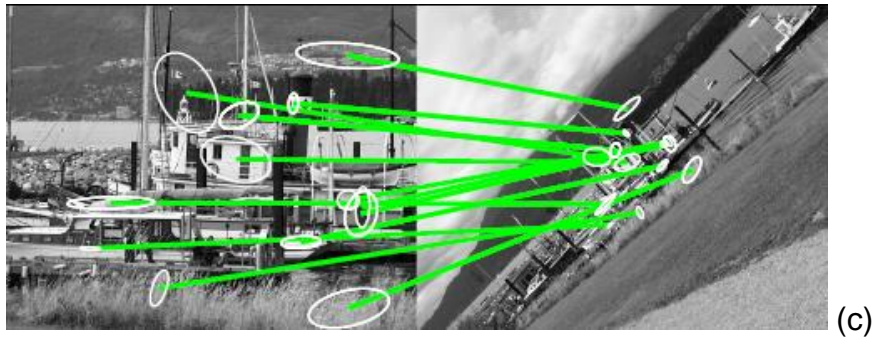


Figura 3.11 Demonstração de desempenho do MSER; (a) imagem Q ; (b) imagem A com mudança de escala por fator $\lambda = 2,5$ e rotação de $\phi = 45^0$; (c) reconhecimento do MSER, as elipses brancas indicam regiões estáveis e as linhas verdes mostram a correspondência ou casamento das regiões. Imagens obtidas de (FORSSËN; LOWE, 2007).

O MSER apresenta duas deficiências; (1) procura a invariância afim normalizando os seis parâmetros, (2) utiliza componentes de alto contraste na construção das regiões invariantes, isto constitui uma deficiência dado a escassa presença de componentes de alto contraste em imagens de cenas naturais. No entanto, o MSER apresenta uma limitada invariância a inclinações em condições favoráveis, pequenas mudanças de escala (diferenças menores a 3) e alto contraste nos objetos. Este algoritmo falha na presença de grandes mudanças de inclinação ou variações de escala (MOREL; YU, 2009b; LOWE, 2004).

Na comparação do MSER e SIFT (FORSSËN; LOWE, 2007), concluiu-se que o SIFT teve melhor desempenho para objetos planos (e.g. desenhos de parede), enquanto que o MSER teve melhor desempenho especialmente em cenários que incluem objetos menos triviais (e.g. janelas e paisagens).

3.6.2 Hessian-affine e Harris-affine (MIKOLAJCZYK; SCHMID, 2002, 2004)

O Hessian-affine, assim como o Harris-affine, normaliza os seis parâmetros da transformada afim. A diferença entre eles fica na primeira etapa, onde o Hessian-affine detecta regiões de interesse, enquanto que o Harris-affine detecta cantos. Os máximos locais dos determinantes da matriz hessiana são usados como pontos base, e o processo restante é igual ao do Harris-affine. Os dois métodos realizam o processo de

seleção da escala baseados na Laplaciana e a forma da região elíptica é determinada pelo segundo momento da matriz da intensidade do gradiente.

O Harris-affine utiliza o segundo momento da matriz, também chamado matriz de autocorrelação, para a detecção de características e para descrever estruturas locais na imagem. Esta matriz descreve a distribuição do gradiente na vizinhança local de um ponto:

$$M = \mu(x, y, \sigma_1, \sigma_D) = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} = \sigma_D^2 G(\sigma_1) * \begin{bmatrix} I_x^2(x, y, \sigma_D) & I_x I_y(x, y, \sigma_D) \\ I_x I_y(x, y, \sigma_D) & I_y^2(x, y, \sigma_D) \end{bmatrix} \quad (3.12)$$

Onde

$$I_x^2 = \left(\frac{\partial}{\partial x} \right)^2 \quad I_x I_y = \frac{\partial^2}{\partial x \partial y}$$

E σ_D (escala de diferenciação) é utilizado como o fator de escala do kernel gaussiano utilizado para calcular as derivadas locais da imagem. A média das derivadas é calculada na vizinhança de um ponto, por borramento com um kernel gaussiano G de escala σ_1 (escala de integração). Os autovalores desta matriz descrevem duas principais mudanças de intensidade na vizinhança do ponto. Esta propriedade permite a extração de pontos, para os quais a mudança de intensidade nas direções ortogonais é significativa. Estes pontos são representativos na imagem porque eles apresentam invariância a condições arbitrárias de iluminação. Este é o princípio no qual se baseia o detector de Harris (apresentado no capítulo 2). O Hessian-affine explora uma ideia similar à anterior, baseada na matriz Hessiana:

$$H = H(x, y, \sigma_D) = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} I_{xx}(x, y, \sigma_D) & I_x I_y(x, y, \sigma_D) \\ I_x I_y(x, y, \sigma_D) & I_{yy}(x, y, \sigma_D) \end{bmatrix} \quad (3.13)$$

Onde

$$I_{xx} = \frac{\partial^2}{\partial x^2} \quad I_x I_y = \frac{\partial^2}{\partial x \partial y}$$

A matriz Hessiana utiliza a segunda derivada para obter uma melhor resposta na detecção de regiões e eixos. Um máximo local do determinante indica a presença de uma estrutura local de interesse.

A normalização afim é realizada por um processo iterativo que avalia os parâmetros de uma região elíptica e os normaliza para uma região circular (ver Figura 3.7), procurando

uma região isotrópica invariante sob transformações afins. A cada iteração os parâmetros da região elíptica são avaliados, para minimizar as diferenças entre os autovalores dos momentos de segunda ordem da matriz da região seleccionada. Os autovalores do segundo momento da matriz são também utilizados para avaliar a disposição dos pontos na vizinhança. A transformação afim é determinada por um fator de rotação baseado na normalização da orientação do gradiente.

O processo sequencial da região iterativa (ver Figura 3.7) compreende: (a) detecção da região e seleção da escala, utilizando o espaço de escala, (b) estimação da forma, usando o segundo momento da matriz, (c) normalização da região afim para uma forma circular, (d) voltar à estimação da forma, se os autovalores do segundo momento da matriz para o novo ponto não forem iguais. A Figura 3.12 apresenta o desempenho do Harris-affine e Hessian-affine na detecção de regiões invariantes antes e depois da transformação afim induzida pela mudança de ponto de vista.

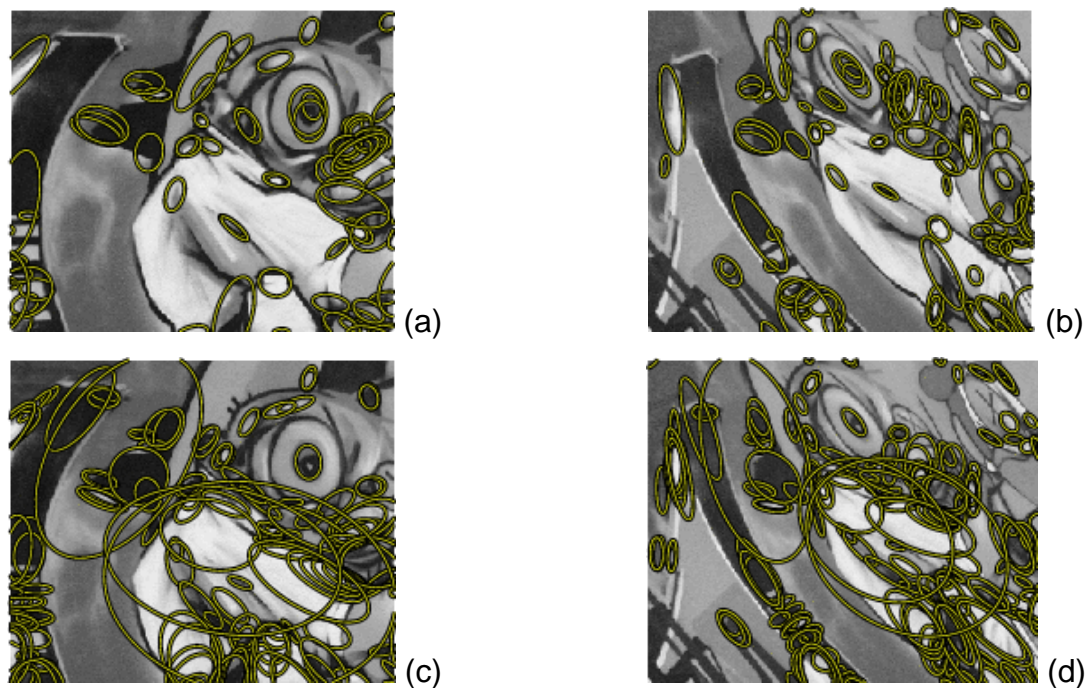


Figura 3.12 Detecção de regiões invariantes, à esquerda ponto de vista frontal e direita fotografia lateral (mudança do ponto de vista), (a-b) Harris-affine, (c-d) Hessian-affine. Note como as regiões são detectadas após de ficar submetidas à transformação afim. Imagens e aplicação disponíveis em (MIKOLAJCZYK, 2010a).

Nos resultados da comparação de seis detectores de regiões invariantes afins (MIKOLAJCZYK; et. al. 2005), onde não se inclui o ASIFT, demonstra-se o destaque no desempenho do MSER e o Hessian-affine. Estes algoritmos apresentam boa resposta invariante afim. No entanto, quando a transformação inclui grandes mudanças de escala (diferenças de escala maiores a 3) estes algoritmos tornam-se não-invariantes-afins, ou seja, nenhum deles é completamente invariante afim (MOREL; YU, 2009a; MIKOLAJCZYK, et. al., 2005; LOWE, 2004).

3.6.3 Algoritmo ASIFT (*Affine- scale invariant feature transform*) (MOREL; YU, 2009a)

Nenhum dos algoritmos reportados na literatura apresenta total invariância a mudanças de pontos de vista quanto o ASIFT, dado que a técnica de normalização utilizada comumente não abriga mudanças extremas. O destaque na proposta do ASIFT é utilização da “simulação de pontos de vista” no lugar da normalização. Morel e Yu demonstram que com esta técnica consegue-se total invariância a mudanças de pontos de vista. ASIFT é proposto como uma extensão do SIFT (descrito no capítulo 2) com o objetivo de obter um algoritmo invariante a pontos de vista. A Figura 3.13 apresenta uma demonstração de desempenho do ASIFT.

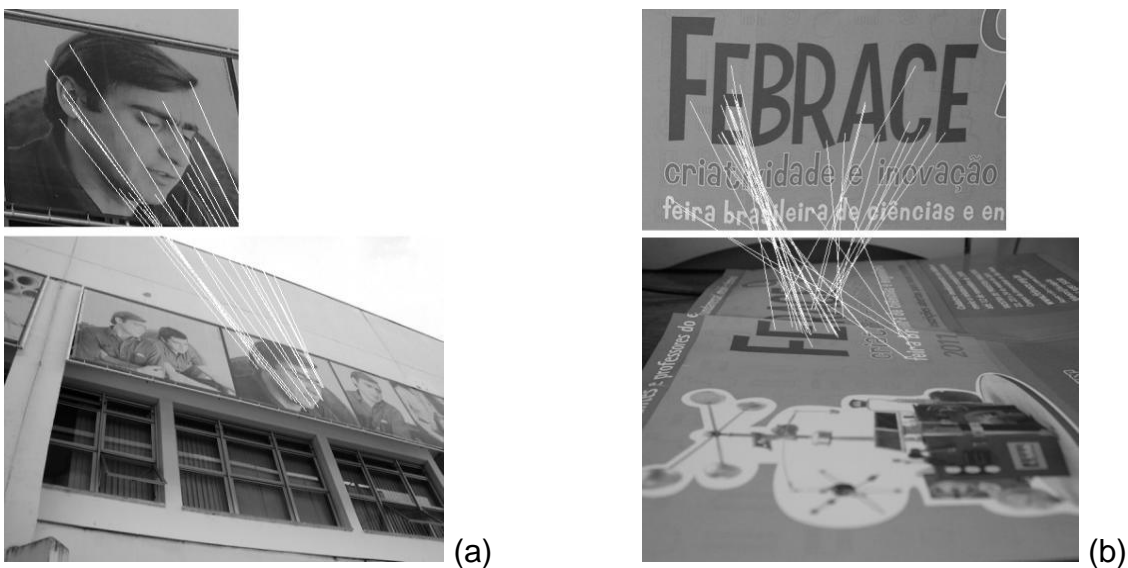


Figura 3.13 Demonstração de desempenho do ASIF, a) reconhecimento de rosto $t = 2$, b) reconhecimento de texto $t = 6$. Aplicação disponível em (MOREL; YU, 2010).

Podemos resumir o algoritmo ASIFT assim:

- Cada imagem (Q e A) é transformada para simular as possíveis distorções (Q_i e A_i) causadas pela mudança de ponto de vista da câmera. As transformações são realizadas mediante um pequeno número de variações dos parâmetros de inclinação θ e rotação ϕ .
- Uma a uma das sub-imagens Q_i são comparadas com as sub-imagens A_i utilizando o algoritmo SIFT e procurando as duas imagens que apresentem melhor correspondência.
- O casamento de padrões é executado e indicado mediante linhas que interconectam os padrões.

A técnica do espaço de escalas, que possibilita a simulação de escalas, faz com que o SIFT apresente completa invariância a mudanças de escala. A combinação deste algoritmo com a técnica de simulação de pontos de vista torna o ASIFT completamente invariante afim (MOREL; YU, 2009a). Porém, ASIFT apresenta dificuldades para o reconhecimento de um objeto na presença de padrões repetitivos e grande mudança de contraste. A seguir são apresentados alguns testes demonstrativos onde se pode verificar este fato.

3.6.3.1 Fraqueza do ASIFT a padrões repetitivos

Quando as imagens apresentam *self-similarities* ou estruturas periódicas o ASIFT falha (MOREL; YU, 2010). Na Figura 3.14 apresentamos uma aplicação demonstrativa do ASIFT onde os casamentos foram errados, devido à similaridade de estrutura entre o objeto e o complemento da cena.

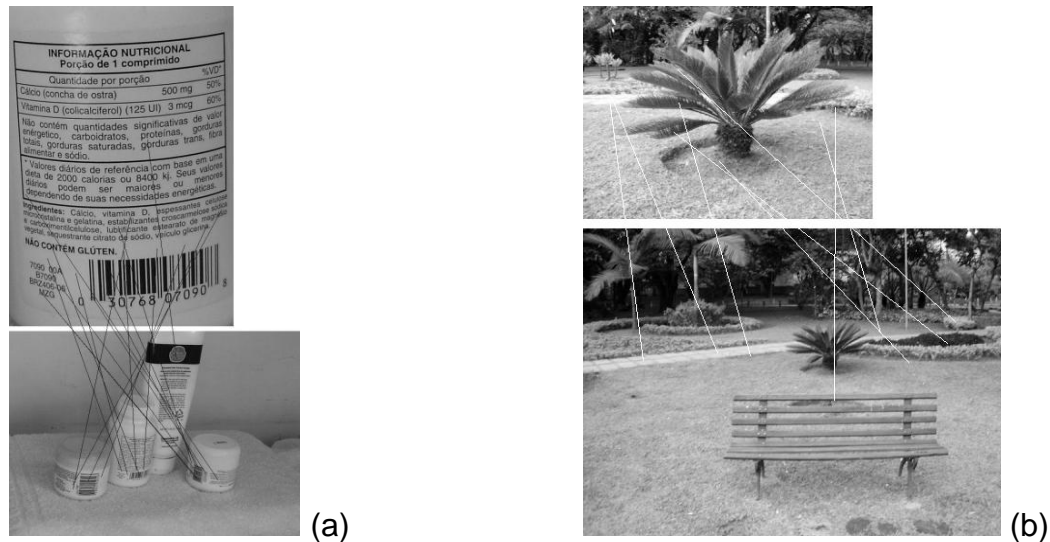


Figura 3.14 Exemplos de falhas do ASIFT na presença de padrões repetitivos. (a) Os padrões de Q (letras) repetem-se nos outros objetos e induzem falsos casamentos, (b) Os padrões de Q repetem-se no fundo da imagem A. Aplicação disponível em MOREL e YU (2010).

3.6.3.2 Fraqueza do ASIFT a grandes mudanças de contraste

Um dos casos típicos onde o ASIFT falha é quando as condições de iluminação são diferentes entre as imagens (MOREL; YU, 2010). Na Figura 3.15 apresentamos uma aplicação demonstrativa, onde o ASIFT falha no reconhecimento, devido à mudança de contraste nas imagens.

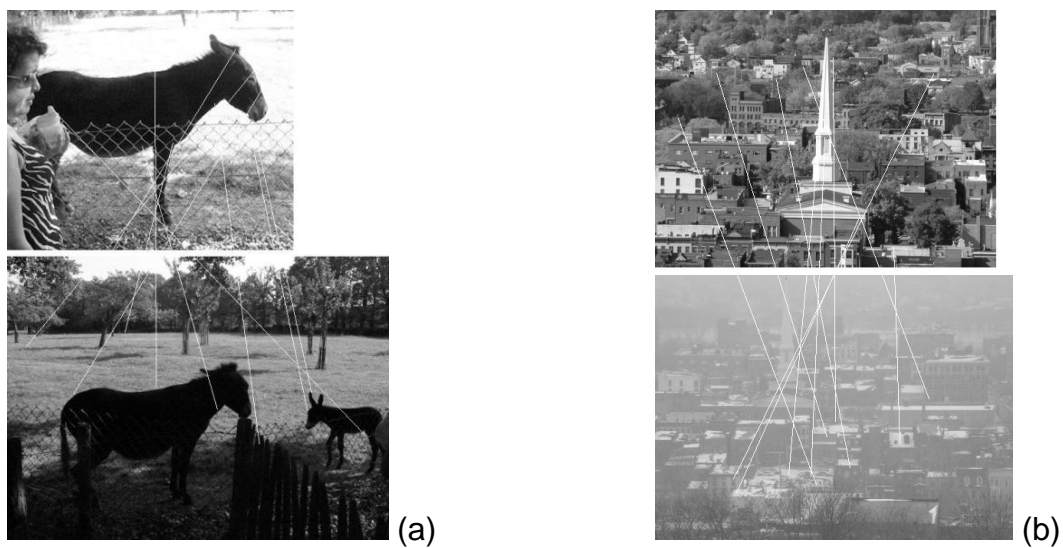


Figura 3.15 ASIFT falha com grandes mudanças de contraste. Todos os casamentos nestas imagens são errados, (a) efeito de sombra, (b) efeito de neblina.

Capítulo 4.

AFORAPRO: FORAPRO INVARIANTE AFIM

AFORAPRO (*Affine-Fourier Coefficients of Radial Projections*) é o nome do algoritmo proposto neste trabalho. É uma extensão que torna o FORAPRO (descrito no capítulo 2) invariante a pontos de vista. Neste capítulo, descrevemos o AFORAPRO e apresentamos alguns testes demonstrativos do desempenho na invariância a pontos de vista, robustez a padrões repetitivos e mudança de contraste. Na sequência, apresentamos a análise de complexidade quanto a recursos de tempo de execução do algoritmo em cada etapa.

4.1 Introdução

O escopo de nossa proposta baseia-se nas diferenças existentes entre o FORAPRO e o SIFT. O primeiro é robusto na presença de padrões repetitivos e grandes mudanças de contraste, devido à utilização dos coeficientes de Fourier na extração de características. O segundo, baseado em texturas locais e extração de pontos chaves é amplamente reconhecido pela sua invariância a escala, mas apresenta dificuldades no reconhecimento de formas simples com baixo contraste. A combinação do FORAPRO com a técnica de simulação de pontos de vista nos permite construir um algoritmo invariante a transformações afim, robusto a padrões repetitivos e a mudanças de contraste.

4.2 Algoritmo AFORAPRO

Dadas duas imagens (Q e A), assumiremos que a imagem Q apresenta um ponto de vista frontal ($\theta = 0, \phi = 0, \lambda = 1$) e a imagem A contém uma instância de Q fotografada desde um ponto de vista aleatório. Então, o processo de reconhecimento do AFORAPRO tem a seguinte sequência: Utiliza-se a imagem Q para gerar um grupo de imagens sintéticas ($Q \rightarrow Q_1, Q_2, \dots, Q_n$), simulando a transição do eixo ótico da câmera numa observação hemisférica (ver Figura 3.9). Cada imagem Q_i é procurada na

imagem A , utilizando um vetor de características extraído pelo FORAPRO. Subsequentemente é realizado o casamento de padrões entre a Q_i selecionada e A .

Os passos do algoritmo proposto são resumidos assim:

1. Transforma-se a imagem Q , simulando as possíveis distorções causadas pela mudança de ponto de vista da câmera para obter as imagens sintéticas ($Q \rightarrow Q_1, Q_2, \dots, Q_n$). Supomos que a imagem Q foi fotografada na posição frontal, mas que pode aparecer na imagem A com distorção. Isto é, diferentemente do algoritmo ASIFT, não consideramos as “inclinações de transição” onde ambas Q e A podem estar inclinadas em direções diferentes. As distorções dependem dos parâmetros de rotação ϕ e inclinação $\theta \rightarrow t = 1/\cos \theta$ (ver Figuras 3.8 e 3.9).
2. Todas as imagens Q_i são procuradas na imagem A , com o objetivo de selecionar a imagem sintética que apresente a simulação mais próxima à distorção do objeto procurado na imagem A . Neste processo o AFORAPRO escolhe a imagem distorcida Q_i que gere o maior valor na matriz de acumulação H da transformada de Hough.
3. Realiza-se o casamento de padrões entre Q_i escolhida e A . Uma vez que o objeto é reconhecido com pelo menos três casamentos de padrões, o AFORAPRO procede a rotular o objeto.

Para dar início à execução do AFORAPRO deve-se definir o número de simulações, este parâmetro assegura que a quantidade das imagens simuladas seja suficiente para cobrir todos os possíveis pontos de vista que a câmera pode adotar. O número de simulações tem dependência direta com a variação dos parâmetros de rotação e inclinação dos eixos da câmera, seguindo a sugestão de Morel e You os valores adotados são $\Delta t = t_{k+1}/t_k = \sqrt{2}$ e $\Delta \phi = 72^\circ/t$, em consequência os intervalos são:

- $t = \{\sqrt{2}^0, \sqrt{2}^1, \sqrt{2}^2, \sqrt{2}^3, \sqrt{2}^4\}$, onde $t = 1/\cos \theta$
- $\phi = \{0, b/t, \dots, kb/t\}$, onde $b = 72^\circ$ e $kb/t < 180^\circ$.

Estes valores asseguram que o número de simulações atinja o intervalo dos possíveis pontos de vista no qual se encontra a vista do objeto na imagem A . O intervalo básico utilizado no AFORAPRO vai até $t = \sqrt{2}^4$ e gera 27 imagens sintéticas Q_i , atingindo até $\theta \approx 76^\circ$, no entanto o AFORAPRO pode ser ajustado para outros valores dependendo das exigências, por exemplo, $t = \sqrt{2}^5$ que gera 41 Q_i e atinge um intervalo de inclinação que vai até $\theta \approx 80^\circ$.

4.3 Invariância a pontos de vista

O AFORAPRO obtém esta característica pela técnica de simulação de pontos de vista adotada do ASIFT. A seguir são apresentados três exemplos demonstrativos do desempenho de AFORAPRO. O objetivo do primeiro exemplo é apresentar as etapas do AFORAPRO e para isso realizamos uma descrição passo a passo. O objetivo do segundo exemplo é comparar a resposta do AFORAPRO com ASIFT, e o objetivo do terceiro é realizar uma demonstração com imagens de formas simples e pouca textura.

Exemplo 4.3.1:

Neste teste apresentamos o reconhecimento de um objeto (mão de sinalização), onde o modelo Q aparece na imagem de procura A com uma distorção em perspectiva. Iniciamos apresentando na Figura 4.1 as imagens envolvidas no processo.

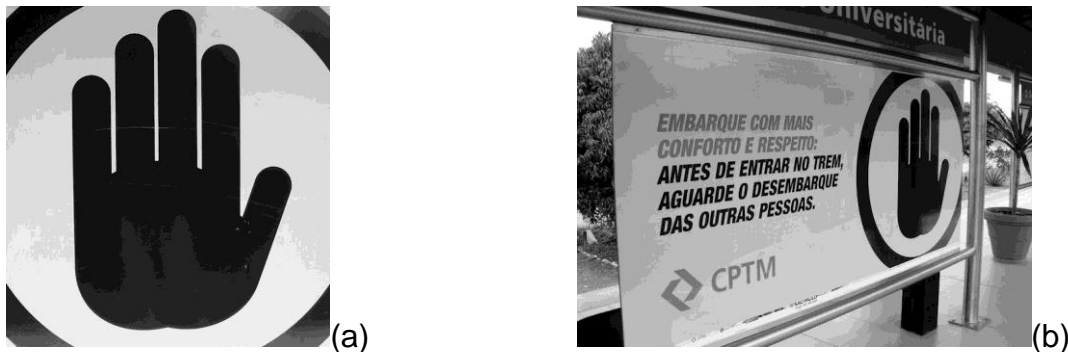


Figura 4.1 *Imagens de entrada ao processo de reconhecimento. (a) imagem Q (tamanho 600x426). (b) imagem A (tamanho 600x700).*

Na sequência, a imagem Q é distorcida para simular os possíveis pontos de vista que poderia ter o objeto. Cada *tilt* é uma amostragem que divide a área da imagem pelo fator t . O número de imagens rotacionadas por cada *tilt* são $(180/72)t = 2.5t$.

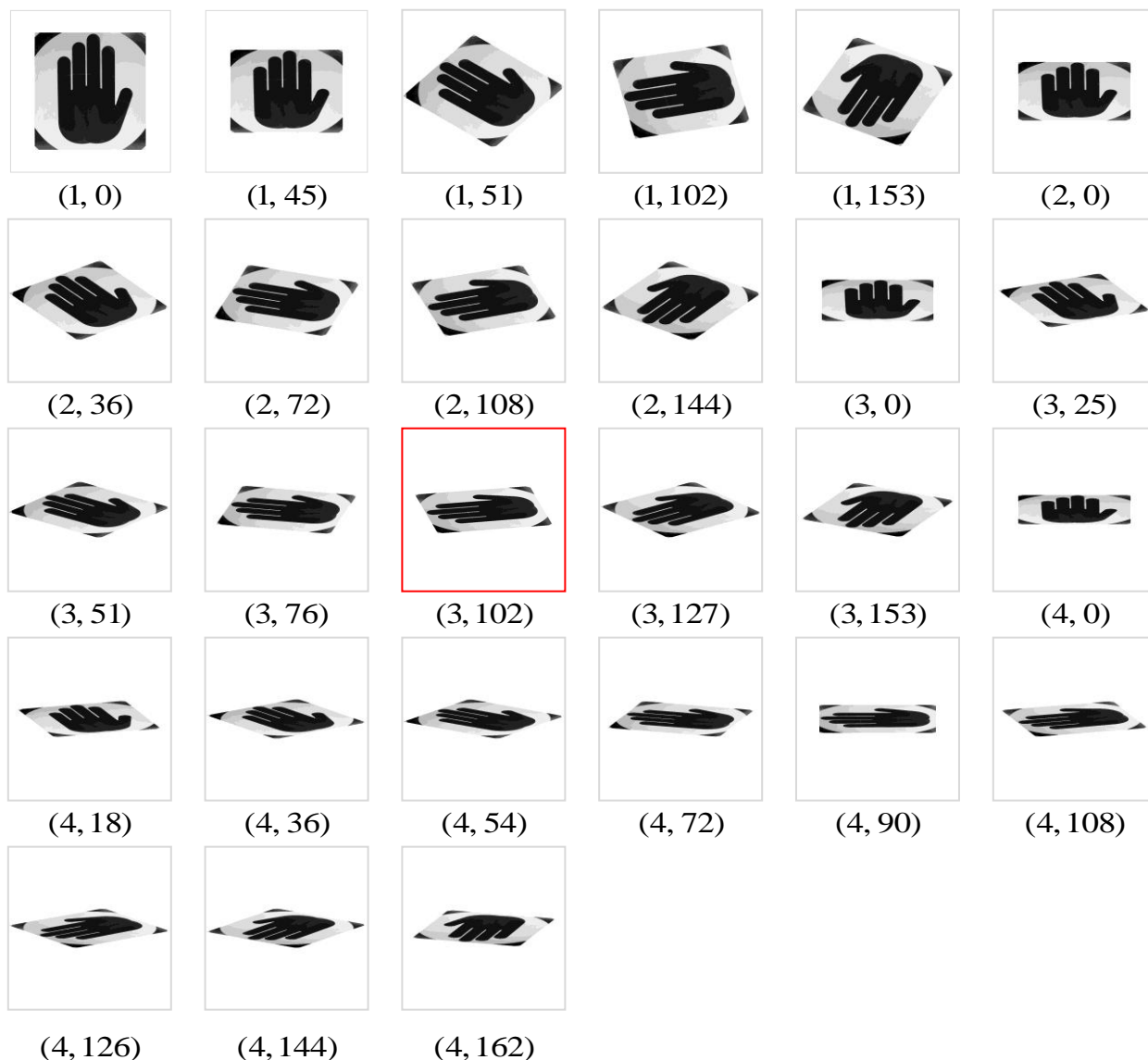


Figura 4.2 Imagens sintéticas Q_i , correspondentes à simulação de distorções com valores variáveis (t, ϕ) . O número de imagens simuladas (27) corresponde ao intervalo mínimo utilizado no AFORAPRO, $t = \{\sqrt{2}^0, \sqrt{2}^1, \sqrt{2}^2, \sqrt{2}^3, \sqrt{2}^4\}$. A imagem de parâmetros $(3, 102)$, rotulada em vermelho, foi selecionada automaticamente para prosseguir no reconhecimento.

A imagem sintética selecionada é aquela que apresente a distorção mais próxima, relacionada com a transformação que possui a instância do objeto procurado e presente na imagem A . A etapa final é o casamento entre a Q_i selecionada e a imagem A . A resposta do AFORAPRO é apresentada na Figura 4.3.



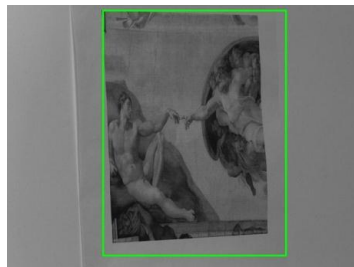
Figura 4.3 *Imagens envolvidas no casamento final. (a) imagem sintética Q_i escolhida (3, 102). (b) resposta do sistema, objeto reconhecido e rotulado de verde na imagem A .*

Exemplo 4.3.2:

Testes realizados com imagens reconhecidas na área do casamento de padrões (Figura 4.4). “The Creation of Adam” é uma das imagens utilizadas na comparação do ASIFT com SIFT, MSER, Harris-affine e Hessian-affine, pela informação fornecida por MOREL e YU (2010) sabemos que: (a) nas Figuras 4.4b-c dos concorrentes do ASIFT apenas o SIFT e MSER conseguiram reconhecer com sucesso, (b) nas Figuras 4.4d-f somente ASIFT teve sucesso. Lembremo-nos que os algoritmos mencionados utilizam as linhas de casamento entre as imagens para indicar a correspondência de padrões e que a quantidade de linhas acertadas determina a qualidade do reconhecimento. Nas aplicações específicas do reconhecimento de objetos é de uso geral a utilização de um rótulo sobre o objeto, como fazemos no AFORAPRO.



(a) $t = 1, \phi = 0, \lambda = 1$



(b) $t = 2.4, \phi = 65, \lambda = 0.9$



(c) $t = 2.4, \phi = -65, \lambda = 0.9$

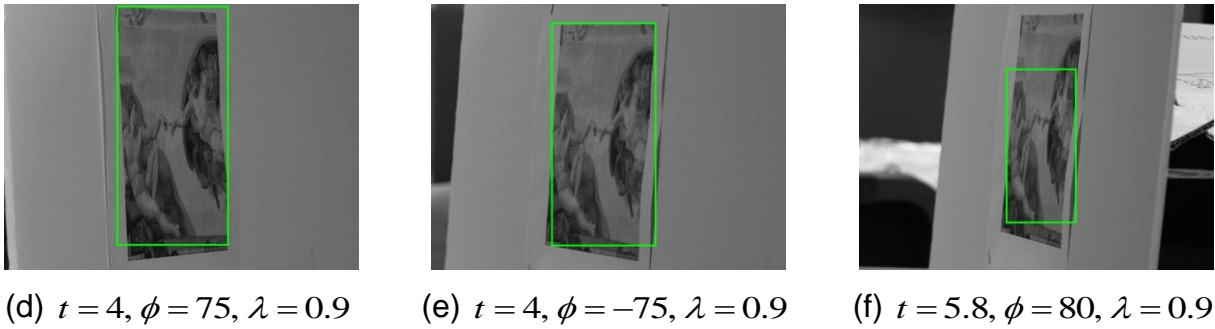


Figura 4.4 Testes realizados com imagens do banco de dados do ASIFT. (a) Vista frontal da imagem modelo Q . (b-f) Reconhecimento bem sucedido pelo AFORAPRO, objetos reconhecidos e rotulado de cor verde sobre a imagem A .

Exemplo 4.3.3:

Reconhecimento de objeto de formas simples e baixo contraste (sinalização de trânsito). Note que o objeto aparece em diferentes escalas e que AFORAPRO é invariante a intervalos específicos de escala, no entanto o intervalo de escala entre Q e a instância do objeto que aparece em A deve ser fornecida para o algoritmo.

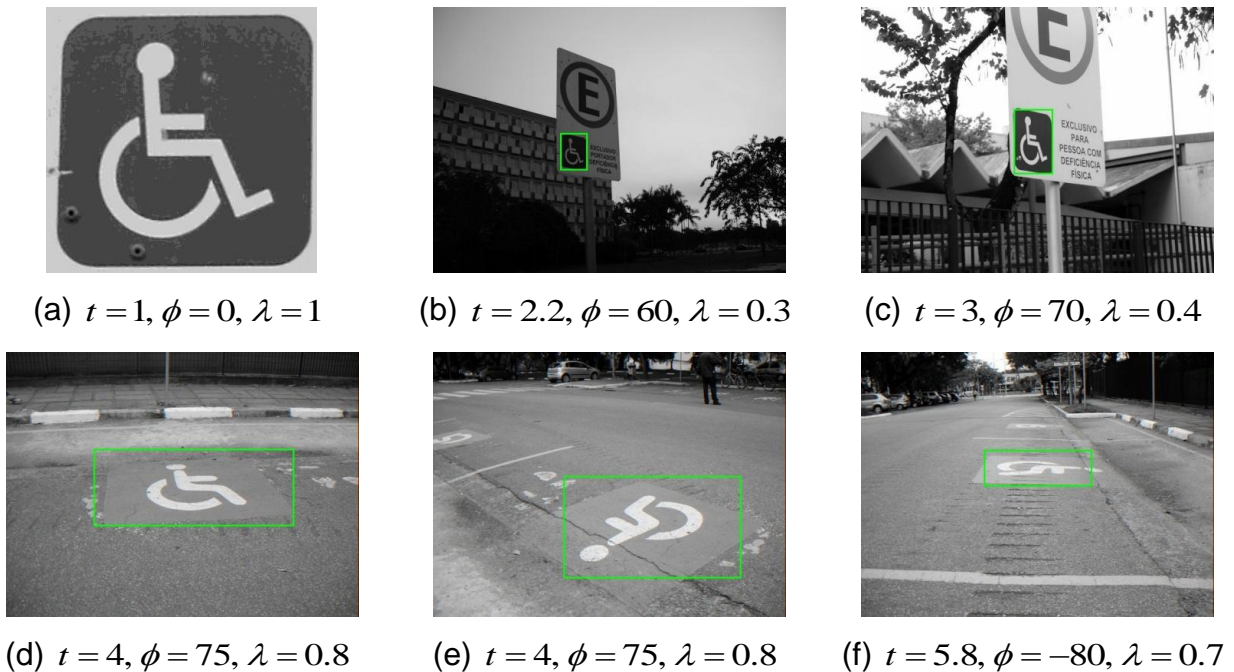


Figura 4.5 Testes realizados com objetos de formas simples e baixo contraste. (a) Imagem Q . (b-f) Reconhecimento bem sucedido pelo AFORAPRO, objetos reconhecidos e rotulados de verde na imagem A .

4.4 Robustez a padrões repetitivos e mudanças de contraste

A robustez a padrões repetitivos e mudanças de contraste é herdada do FORAPRO. Os coeficientes de Fourier de projeções radiais e circulares são equivalentes a um vetor soma de intensidades dos pixels ao longo da trajetória, descrita pelas linhas radiais ou circulares, que reduz a informação bidimensional da vizinhança do pixel para um vetor unidimensional, robusto a variações de intensidade porque o vetor soma elimina os efeitos de todas as mudanças de brilho correspondentes à intensidade média da região avaliada (CHOI; KIM, 2002; KIM, H. Y., 2010a). Resumindo, uma alteração de contraste multiplica todos os coeficientes pelo mesmo fator e não altera a média dos pixels. Deste mesmo jeito a correspondência entre padrões é estabilizada dado que o vetor soma envolve suficiente informação da vizinhança de cada padrão.

O significado físico do vetor soma é equivalente ao coeficiente do primeiro harmônico da projeção radial ou circular. No AFORAPRO o processo para obter os coeficientes é realizado utilizando a propriedade de convolução da transformada de Fourier.

Para testar a robustez a baixo contraste e padrões repetitivos do AFORAPRO foram realizados quatro testes correspondentes a duas classes de provas: na primeira, realizou-se o reconhecimento sobre imagens que apresentam baixo contraste e padrões repetitivos, e na segunda prova foram utilizadas imagens onde ASIFT anteriormente havia falhado. Os resultados são referenciados nos exemplos a seguir.

Exemplo 4.4.1:

Testes do reconhecimento em imagens com mudanças de contraste. As imagens a seguir, além da variação do ponto de vista, apresentam variação do contraste por mudança de fundo, iluminação e efeitos de sombras.



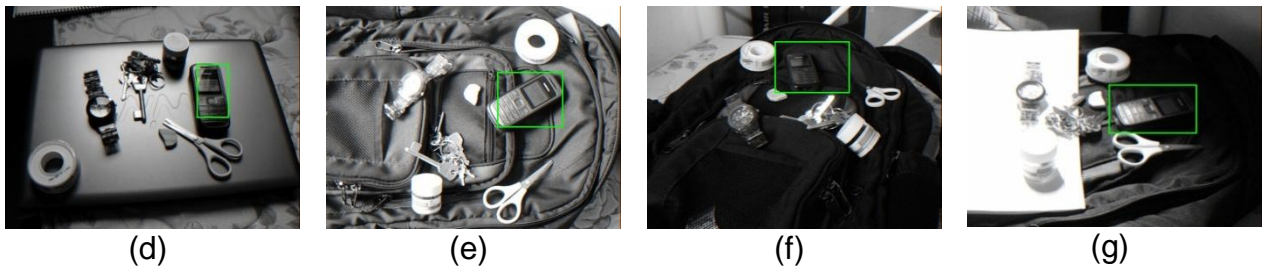


Figura 4.6 Testes com imagens que apresentam mudanças de contraste. (a) Imagem Q com alto contraste. (b-g) Variação do contraste e reconhecimento bem sucedido pelo AFORAPRO.

Exemplo 4.4.2:

Neste teste são utilizadas duas imagens onde o ASIFT tinha falha devido a mudanças de contraste. O AFORAPRO reconhece o objeto com sucesso.



Figura 4.7 Testes comparativos, usando imagens que possuem mudanças de contraste testadas no capítulo anterior (Figura 3.15) onde o ASIFT não conseguiu reconhecer. (a-b) Imagem Q , (c-d) objeto reconhecido e rotulado pelo AFORAPRO, (c) efeito de sombra, (d) efeito de neblina.

Exemplo 4.4.3:

Testes do reconhecimento em imagens com padrões repetitivos. As imagens a seguir, além da variação do ponto de vista, apresentam coincidência nos padrões entre o objeto Q e o complemento da imagem.

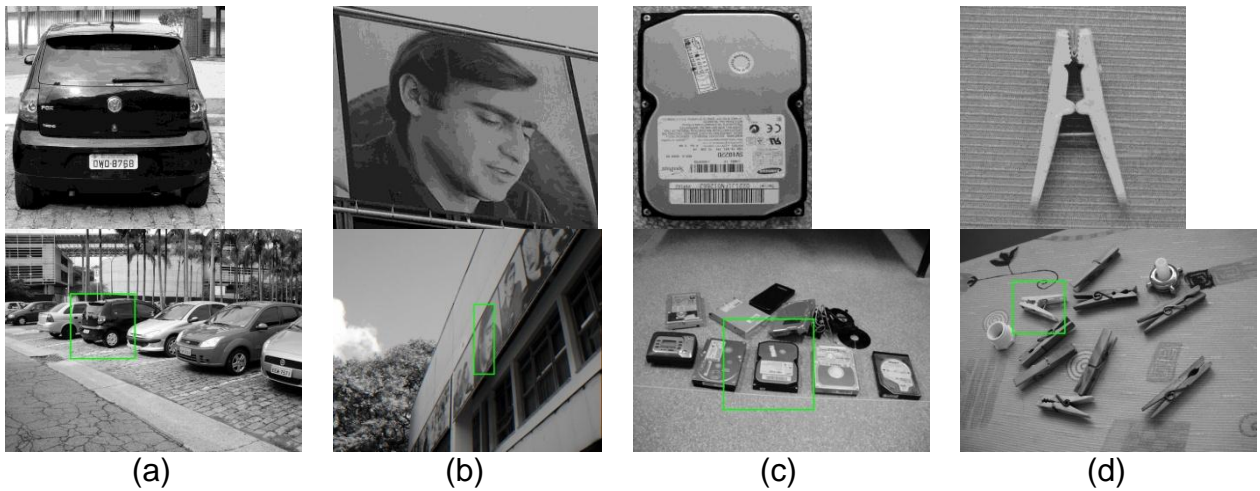


Figura 4.8 Desempenho do AFORAPRO usando imagens que apresentam padrões repetitivos.

Exemplo 4.4.4:

Neste teste são utilizadas duas imagens onde o ASIFT tinha falha devido à presença de padrões repetitivos. O AFORAPRO reconheceu os objetos com sucesso.

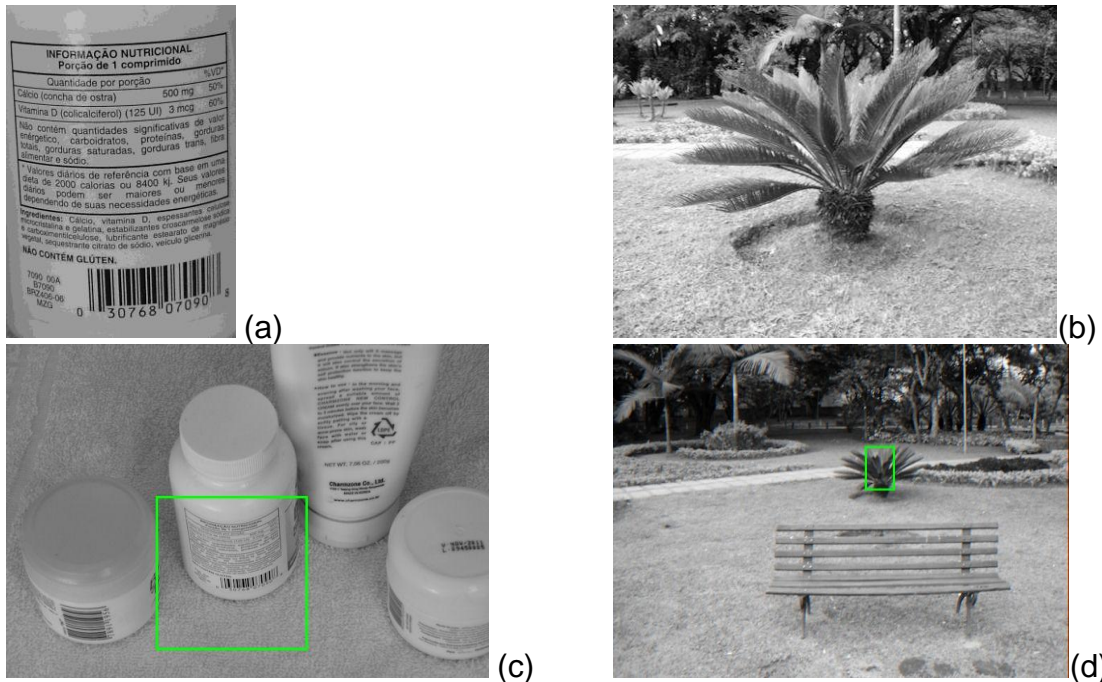


Figura 4.9 Teste comparativo, usando imagens que apresentam padrões repetitivos testadas no capítulo anterior (Figura 3.14) onde o ASIFT não conseguiu reconhecer. (a-b) Imagem Q , (c-d) objeto reconhecido e rotulado pelo AFORAPRO.

4.5 Análise de complexidade

A análise de complexidade está relacionada à quantidade do tempo que o AFORAPRO precisa na execução de cada uma das três etapas; simular as distorções nas imagens sintéticas, selecionar uma imagem sintética para o casamento e realizar o reconhecimento. O número de pixels da imagem é um fator determinante neste tipo de análise, aumentando a complexidade proporcionalmente. Desta maneira o tamanho padrão das imagens utilizadas foi 650×488 pixels. Utilizamos o primeiro exemplo da seção 4.3 como referência para nossa análise.

4.5.1 Geração de imagens sintéticas

Dado que o objetivo do AFORAPRO é especificamente o reconhecimento de objetos, a complexidade na geração de imagens sintéticas é consideravelmente menor do que na técnica original do ASIFT. O reconhecimento de objetos permite pressupor a disponibilidade de uma vista frontal da imagem Q , pelo qual se descarta a simulação de pontos de vista na imagem A , em consequência as distorções são feitas somente em Q . Na tabela 4.1 descreve-se um exemplo representativo do tempo de execução do AFORAPRO, note que o tempo varia proporcionalmente ao número de Q_i . Lembremos que o número básico de Q_i no AFORAPRO é 27, mas pode ser modificado por um parâmetro na entrada.

Tabela 4.1 Tempo de execução na geração de imagens sintéticas mostradas na Figura 4.2. Note que neste exemplo (exemplo 4.3.1) precisa-se apenas de 27 Q_i , os casos com 41 e 61 Q_i são apresentados demonstrativamente.

Número de Imagens sintéticas Q_i	27	41	61
Tempo de execução (segundos)	3,43	5,20	7,46

Não é possível estabelecer uma comparação direta com o tempo empregado no ASIFT, dado que o ASIFT combina a geração de imagens sintéticas com outras funções ao

longo da sua execução, distintamente do AFORAPRO que gera todas as imagens num bloco para depois trabalhar sobre o conjunto.

4.5.2 Seleção da imagem sintética para o casamento

Nesta etapa o objetivo é selecionar a Q_i que melhor corresponda à imagem A . Esta função é realizada pelo FORAPRO, utilizando a transformada de Hough e escolhendo a Q_i que gere o maior valor da matriz de acumulação. O processo fica relativamente lento por AFORAPRO não possuir uma estrutura para a organização dos pixels que apresentam características similares. O tempo gasto no “exemplo 4.3.1” é referenciado na tabela a seguir e a imagem Q_i selecionada foi apresentada na Figura 4.2 com o rótulo vermelho.

Tabela 4.2 Tempo de execução na seleção do Q_i para o casamento.

Número de Imagens sintéticas Q_i	27	41	61
Tempo de execução (minutos)	7,15	11,52	14,08

O controle da complexidade desta etapa poderia ser reduzido num trabalho futuro assim: (a) melhorando o FORAPRO com a implementação de alguma estrutura de dados para acelerar a busca dos pixels com as características similares (por exemplo, usando árvore de decisão, “boosting” ou árvore aleatória), (b) procurando identificar os parâmetros mais adequados (t, ϕ) , da técnica de simulação de pontos de vista, para realizar a amostragem da imagem fonte somente com estes dois valores.

4.5.3 Etapa de reconhecimento

Dada a imagem Q_i são extraídas N sub-imagens circulares $T_1, \dots, T_N \subset Q_i$. Para cada T_i é gerada a imagem de distâncias D_i , onde cada pixel $D_i(x, y)$ é a distância entre as características de $A(x, y)$ e T_i . Na sequência, são selecionados n_c pixels com as menores distâncias, estes pixels são determinados como candidatos e utilizando suas

orientações canônicas procede-se a calcular o pixel central da Q_i em A . Este processo permite incrementar a matriz de acumulação da transformada de Hough e os casamentos são aqueles pixels desta matriz com os maiores valores. Esta etapa também seria beneficiada com a implementação de uma estrutura de dados, para acelerar a busca dos pixels candidatos. Na tabela a seguir é apresentado o tempo gasto em alguns dos exemplos anteriores.

Tabela 4.3 Tempo de execução na etapa do reconhecimento final. Foram referenciadas as imagens apresentadas na Figura 4.1 e Figura 4.9.

Imagens	Figura 4.3	Figura 4.9a-c	Figura 4.9b-d
Tamanhos (Q / A)	600x426 / 600x700	600x350 / 600x700	600x670 / 600x680
Tempo de execução (segundos)	39,13	31,48	28,26

Resumindo, o tempo médio gasto pelo AFORAPRO em todo o processo de reconhecimento de um objeto, para 27 Q_i é $t_{med} \approx 8 \text{ min}$. O tempo médio gasto pelo ASIFT em iguais condições é $t_{med} \approx 2,27 \text{ min}$. A desvantagem do nosso algoritmo fica na segunda etapa. Em um trabalho futuro poderíamos otimizar algumas funções e implementar as sugestões mencionadas anteriormente. Cabe mencionar que o tempo de execução do ASIFT não é constante de uma prova para outra, ou seja, o ASIFT chegou a demorar até $t_{med} \approx 9 \text{ min}$ para algumas imagens que apresentavam estruturas com abundantes detalhes (e.g. revistas, grafites), este caso será retomado no próximo capítulo.

Capítulo 5.

RESULTADOS EXPERIMENTAIS

Neste capítulo descrevemos os experimentos realizados com AFORAPRO e comparamos os resultados obtidos com os de ASIFT e FORAPRO.

5.1 Introdução

Para avaliar o algoritmo proposto, construímos uma database com 350 imagens divididas em cinco conjuntos:

- Conjunto 1: Composto por 100 imagens divididas em 10 grupos, cada grupo contendo um objeto padrão (Q) e nove imagens complementares onde aparece uma instância de Q fotografada desde pontos de vista diferentes e rodeada por outros objetos. O ambiente para adequação das imagens foi uma sala de laboratório e utilizaram-se objetos de uso comum (martelo, alicate, controle de tv, revista, flores, chinelo, pente, colher, óculos e extensão elétrica) apresentados na Figura 5.1:

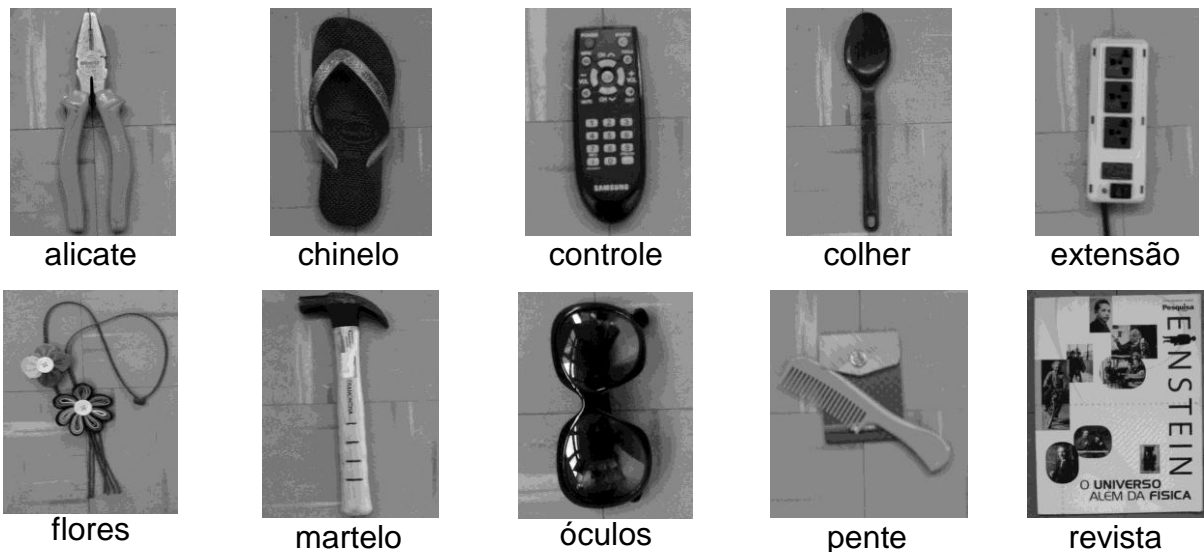


Figura 5.1 Imagens Q do Conjunto 1.

O objeto Q foi fotografado primeiro de maneira isolada e posteriormente fotografado ao lado de outros objetos, tentando recriar uma cena cotidiana.

- Conjunto 2: Composto por 100 imagens de objetos (placas, anúncios, construções, logotipos, grafites e aparelhos) divididas em 20 grupos, cada grupo contém um objeto padrão (Q) e quatro imagens complementares, onde aparece uma instância de Q fotografada desde pontos de vista diferentes e em ambientes naturais.

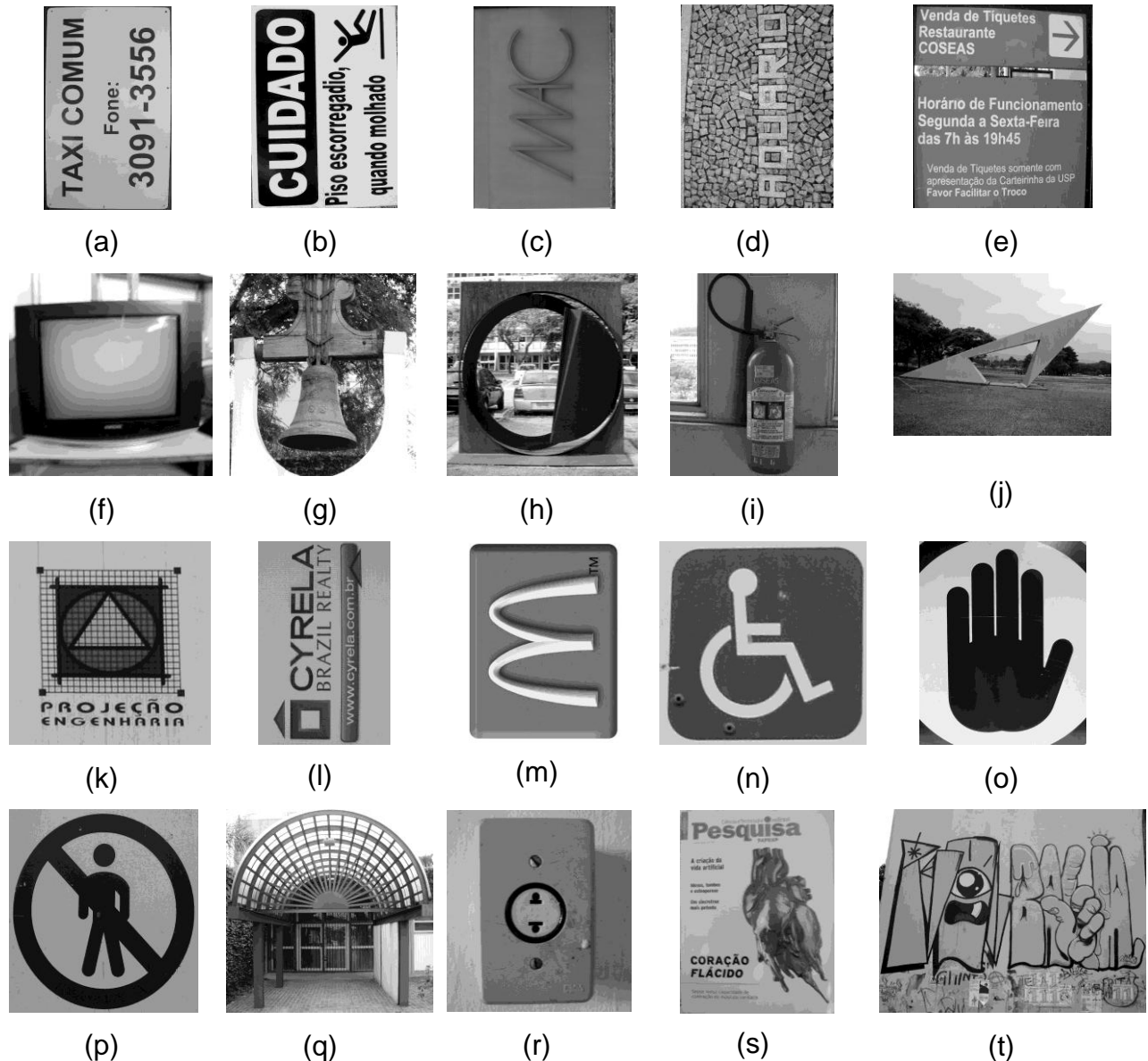


Figura 5.2 Imagens Q do Conjunto 2, (a-e) anúncios, (f-j) objetos “A”, (k-m) logotipos, (n-p) placas, (q-r) objetos “B”, (s) revista e (t) grafite.

- Conjunto 3: Composto por 50 imagens de uso frequente na área de reconhecimento de objetos invariante afim, obtidas das databases de Morel e Yu (2010) e Mikolajczyk (2010b). Estas imagens têm sido usadas para avaliar o ASIFT, Hessian-affine e

Harris-affine, como também têm sido frequentemente usadas como referentes por outros pesquisadores.

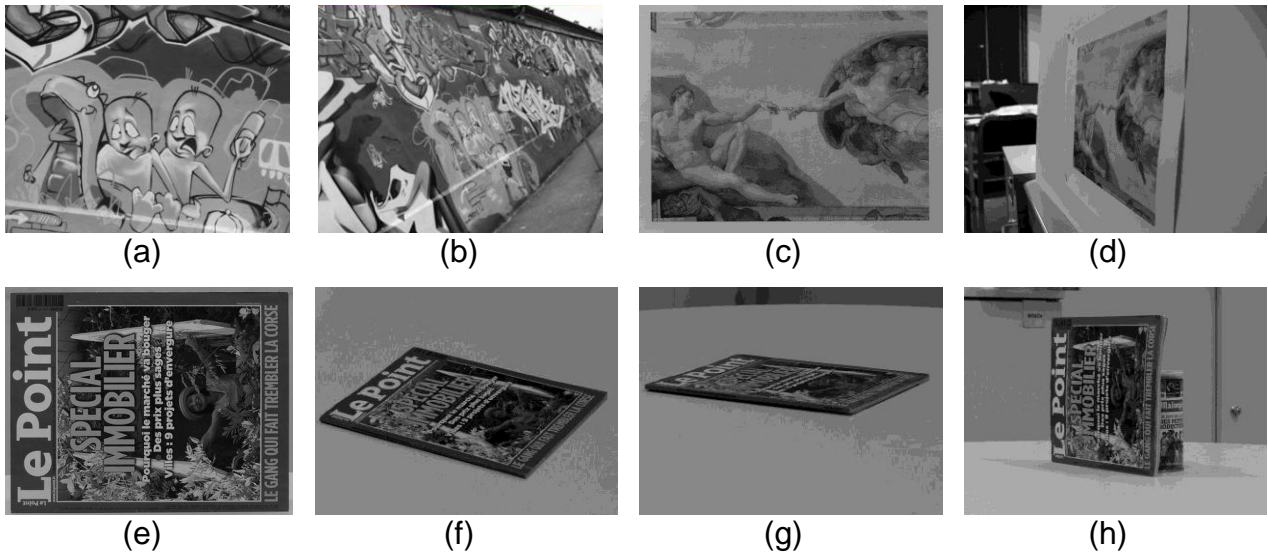


Figura 5.3 Imagens Q do Conjunto 3. (a-b) Grafite database Mikolajzyck, série de nove imagens onde o grafite aparece fotografado com diferentes pontos de vista. (c-h) Database Morel e Yu, (c-d) imagem da criação (Adam), duas séries de nove imagens cada, usando duas escalas e variando a inclinação. (e-h) Revista (Lepoint), 23 imagens distribuídas em três series, (f-g) duas series usando valores θ de (60, 75) e variando o ϕ entre (10, 20, ..., 90), (h) serie de três imagens em posição vertical e com valores θ de (60, 80, 81).

- Conjunto 4: Composto por 50 imagens divididas em 10 grupos, cada grupo contendo um objeto padrão (Q) e quatro imagens complementares onde aparece uma instância de Q fotografada desde pontos de vista diferentes. Uma alteração de contraste foi induzida nas imagens realizando mudanças de fundo, variação de iluminação e efeitos de sombras. O ambiente foi uma sala de laboratório e utilizaram-se objetos de uso comum, apresentados na Figura 5.4.



celular

separador

USB

urso

chave



Figura 5.4 Imagens Q do Conjunto 4.

- Conjunto 5: Composto por 50 imagens divididas em 10 grupos, cada grupo contendo um objeto padrão (Q) e quatro imagens complementares, onde aparece uma instância de Q fotografada desde pontos de vista diferentes. Neste conjunto utilizaram-se imagens de cenas de ambientes naturais que apresentam padrões repetitivos (sinais, anúncios, siluetas, arbustos e aparelhos).

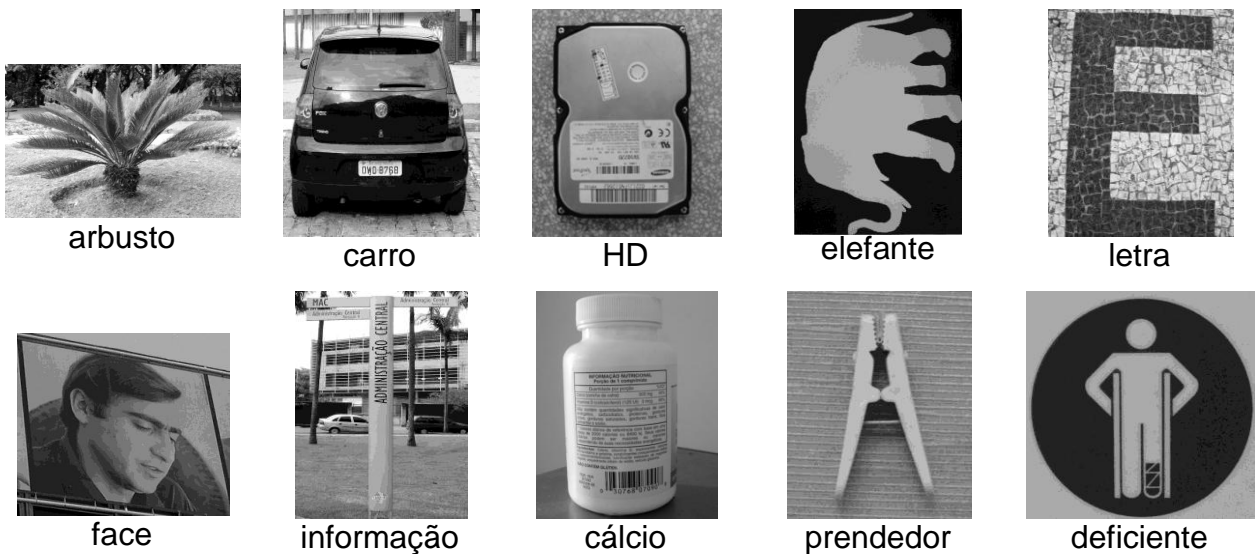


Figura 5.5 Imagens Q do Conjunto 5.

Cada conjunto de imagens possui uma imagem do objeto procurado Q e um número de imagens complementares que contém uma instância de Q . O objetivo é procurar Q nas imagens onde ele aparece. Considera-se o casamento bem sucedido nas seguintes condições: (a) AFORAPRO e FORAPRO, se o traço do retângulo fica no interior do objeto ou se ao menos 50% da área do mesmo fica dentro do objeto procurado, (b) ASIFT, se ao menos 30% das linhas de correspondência marcam casamentos corretos. A média do tamanho das imagens utilizadas nos 5 conjuntos é 650x488 pixels.

Foram realizadas três classes de experimentos, utilizando o AFORAPRO, ASIFT e FORAPRO. O parâmetro de avaliação utilizado foi a taxa de acertos (Equação 2.1). Os resultados demonstram que o AFORAPRO apresenta invariância a pontos de vista, e robustez a padrões repetitivos e mudança de contraste. Também se percebeu superioridade do AFORAPRO sobre o ASIFT no reconhecimento dos objetos.

5.2 Experimentos de invariância a pontos de vista

Estes experimentos testam a capacidade do AFORAPRO no reconhecimento de objetos invariante a pontos de vista. Os objetivos destes experimentos são: (a) Comparando o desempenho de AFORAPRO com o de FORAPRO, (b) Comparando o desempenho de AFORAPRO com o de ASIFT.

5.2.1 AFORAPRO e FORAPRO

Na realização deste experimento utilizaram-se as imagens do Conjunto 1. A Figura 5.6 ilustra os resultados obtidos pelo FORAPRO no reconhecimento de um dos objetos (extensão), com apenas um acerto na imagem com parâmetros (45,0).

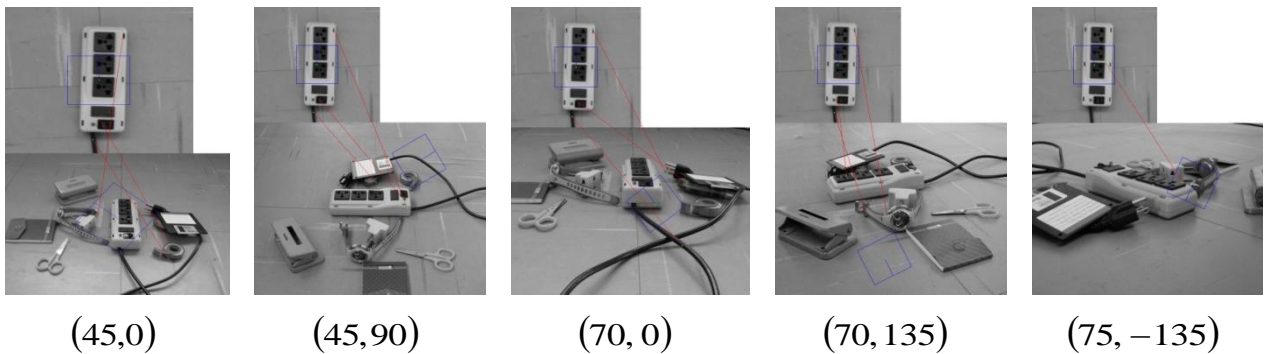
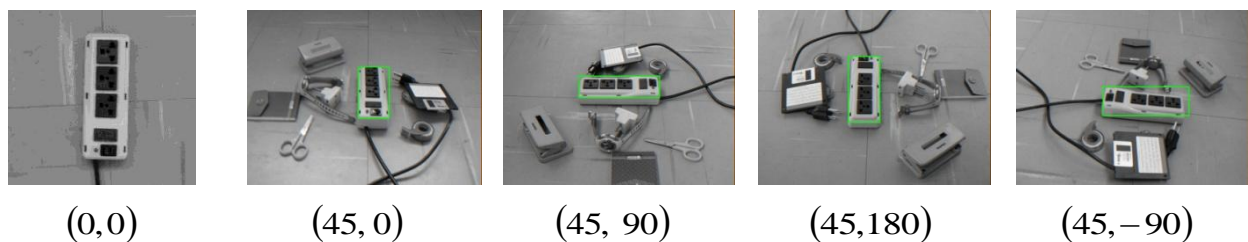


Figura 5.6 Resultado do FORAPRO com objetos do Conjunto 1 “extensão”, (θ, ϕ) .

AFORAPRO, em iguais condições, apresentou 100% de acertos:



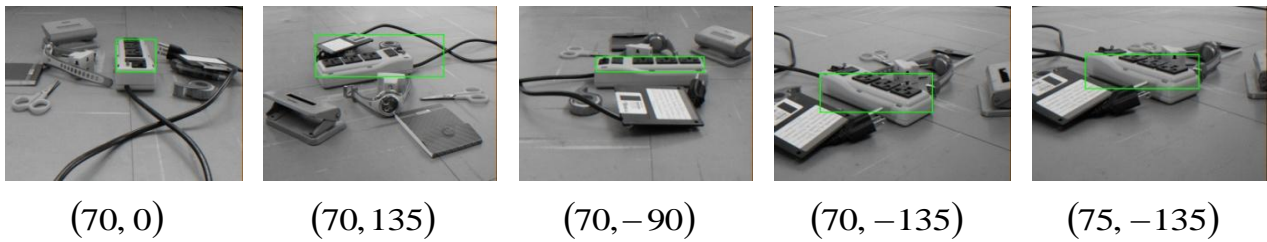


Figura 5.7 Resultados do AFORAPRO com objeto do Conjunto 1 (extensão). Na primeira imagem apresenta-se a vista frontal de Q e nas imagens restantes o resultado do reconhecimento invariante a pontos de vista, (θ, ϕ) .

Os resultados do AFORAPRO e FORAPRO nos experimentos do Conjunto-1 estão sumarizados na Tabela 5.1. Pode-se verificar que AFORAPRO é nitidamente superior a FORAPRO. O FORAPRO apresenta 18,88% na taxa de acertos onde a maior parte destes casos foi para imagens com pequenas variações do θ . O AFORAPRO supera amplamente essa marca, com 100% de acertos.

Tabela 5.1: Resultado da invariância a pontos de vista relacionando o FORAPRO e AFORAPRO.

Objeto (Q)	Reconhecimentos possíveis	FORAPRO		AFORAPRO	
		*VP	**FN	VP	FN
Alicate	9	2	7	9	0
Chinelo	9	1	8	9	0
Controle	9	0	9	9	0
Colher	9	4	5	9	0
Extensão	9	1	8	9	0
Flores	9	2	7	9	0
Martelo	9	2	7	9	0
Óculos	9	3	6	9	0
Pente	9	1	8	9	0
Revista	9	1	8	9	0
Total	90	17	73	90	0
Taxa de acertos (%)		18,88		100,00	

*VP: verdadeiro positivo e **FN: falso negativo (ver tabela 2.1).

O Gráfico 5.1 apresenta a taxa de acertos para cada objeto do Conjunto 1. Como já era previsto, o FORAPRO falha na presença de variações no ponto de vista. Estes resultados atestam que o AFORAPRO estende a aplicabilidade do FORAPRO.

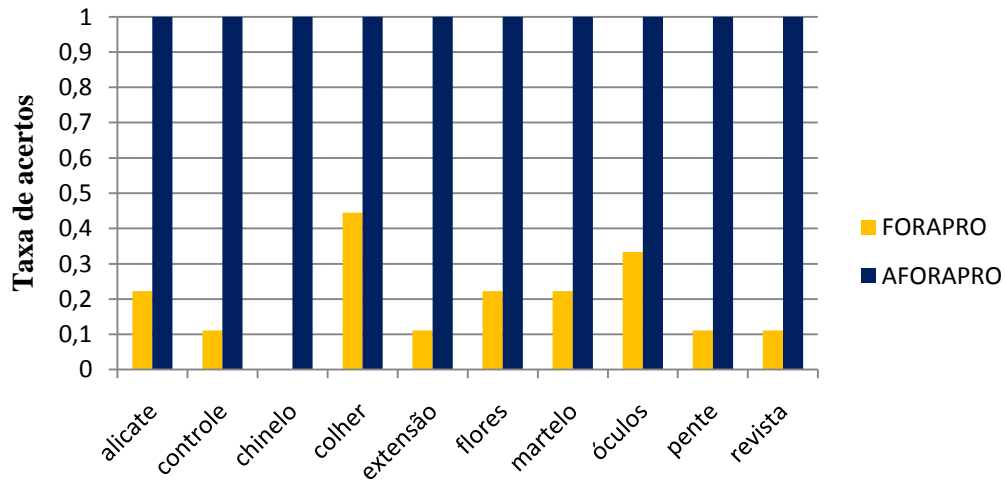


Gráfico 5.1: Taxa de acertos do FORAPRO e AFORAPRO na invariância a pontos de vista com imagens do Conjunto 1.

5.2.2 AFORAPRO e ASIFT

Na realização destes testes utilizaram-se as imagens dos Conjuntos 1, 2 e 3. Também foram utilizadas as duas versões do algoritmo ASIFT disponibilizadas pelos autores (MOREL; YU, 2010) e executadas na plataforma Windows Vista. Inicialmente todos os testes foram executados na versão antiga do ASIFT, posteriormente foi usada a versão nova para corroborar aqueles resultados onde o ASIFT tinha falhado (a nova versão do ASIFT foi disponibilizada em Nov/2010).

5.2.2.1 Experimentos com imagens do Conjunto 1

A Figura 5.8 ilustra os resultados obtidos pelo ASIFT com um dos objetos do Conjunto 1 (extensão). Nos testes realizados para este objeto o ASIFT falhou duas vezes, uma delas é apresentada na Figura 5.8 com valores (70,135) onde todos os casamentos foram errados.

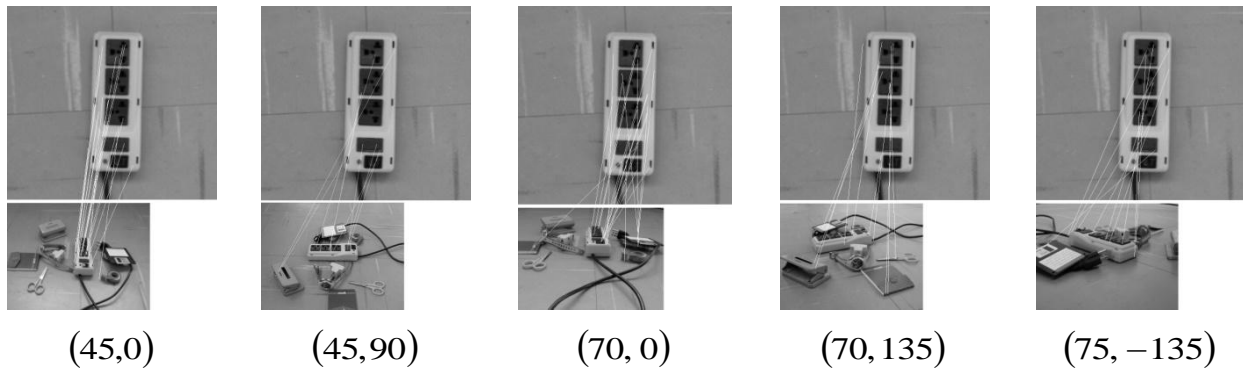


Figura 5.8 Resultado do ASIFT com objeto do Conjunto 1 “extensão”, (θ, ϕ) . O teste com valores $(70,135)$ mostra um caso onde o ASIFT falha no reconhecimento.

Os resultados do AFORAPRO e ASIFT nos experimentos para imagens do Conjunto-1 estão sumarizados na Tabela 5.2. Enquanto ASIFT detectou corretamente apenas 46 dos 90 padrões, AFORAPRO conseguiu reconhecer todas as 90 instâncias dos padrões sem nenhum erro. Porém, o processo de reconhecimento por ASIFT é cerca de quatro vezes mais rápido do que o processo por AFORAPRO.

Tabela 5.2: Resultado da invariância a pontos de vista com imagens do Conjunto 1.

Objeto (Q)	Reconhecimentos possíveis	ASIFT		AFORAPRO	
		VP	FN	VP	FN
Alicate	9	3	6	9	0
Chinelo	9	3	6	9	0
Controle	9	4	5	9	0
Colher	9	4	5	9	0
Extensão	9	7	2	9	0
Flores	9	8	1	9	0
Martelo	9	3	6	9	0
Óculos	9	1	8	9	0
Pente	9	4	5	9	0
Revista	9	9	0	9	0
Total	90	46	44	90	0
Taxa de acertos (%)		51,11		100,00	

O Gráfico 5.2 apresenta o resumo da taxa de acertos para cada objeto do Conjunto 1. Pode-se perceber que AFORAPRO supera ao ASIFT. A resposta do ASIFT é boa para objetos que apresentam abundantes detalhes na estrutura (revista, flores) e tem dificuldades para objetos de formas simples e pouca textura (óculos, martelo).

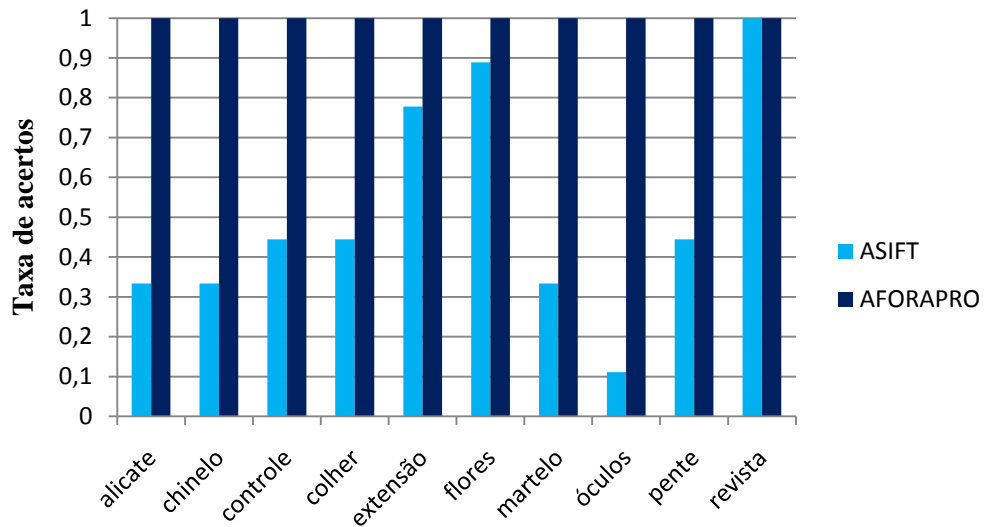
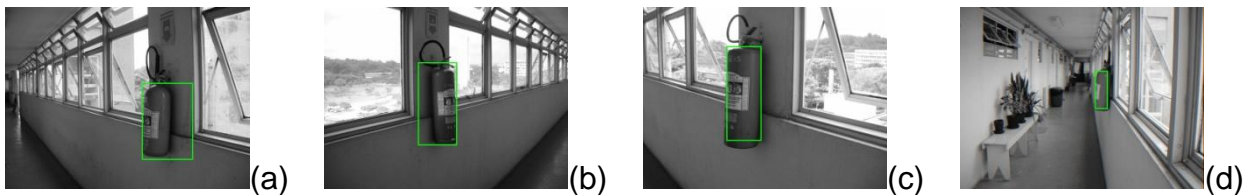


Gráfico 5.2: Taxa de Acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 1.

5.2.2.2 Experimentos com imagens do Conjunto 2

Na Figura 5.9 apresentam-se exemplos da resposta do AFORAPRO e ASIFT no reconhecimento de um extintor de incêndio. O ASIFT só acertou uma vez (Figura 5.9e), enquanto que o AFORAPRO acertou nos quatro casos.



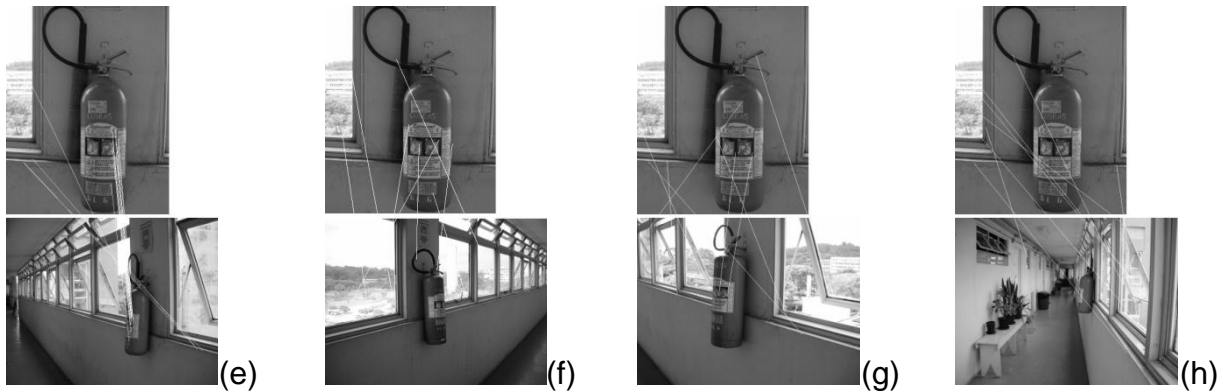


Figura 5.9 Resultado com objeto do Conjunto 2 “extintor”. (a-d) Resposta do AFORAPRO, (e-h) resposta do ASIFT.

A Tabela 5.3 resume os resultados dos experimentos para o Conjunto 2. O AFORAPRO apresenta 5 erros ocasionados principalmente pelas grandes mudanças de escala, enquanto o ASIFT apresenta 40 erros ocasionados provavelmente pela ausência de detalhes ou textura nos objetos. A taxa de acertos do AFORAPRO é claramente superior à do ASIFT.

Tabela 5.3: Respostas do ASIFT e AFORAPRO com as imagens do Conjunto 2.

Objeto (Q)	Reconhecimentos possíveis	ASIFT		AFORAPRO	
		VP	FN	VP	FN
Anúncios	20	11	9	18	2
Objetos A	20	3	17	20	0
Logotipos	12	6	6	10	2
Sinais	12	6	6	11	1
Objetos B	8	6	2	8	0
Revista	4	4	0	4	0
Grafite	4	4	0	4	0
Total	80	40	40	75	5
Taxa de acertos (%)		50,00		93,75	

O Gráfico 5.3 apresenta o resumo da taxa de acertos para cada grupo de objetos do Conjunto-2. Verifica-se que a resposta do ASIFT é boa nos grupos de objetos que

apresentam abundantes detalhes na estrutura (revista, grafite) e ruim para grupos onde os objetos possuem poucos detalhes (logotipos, objetos “A”).

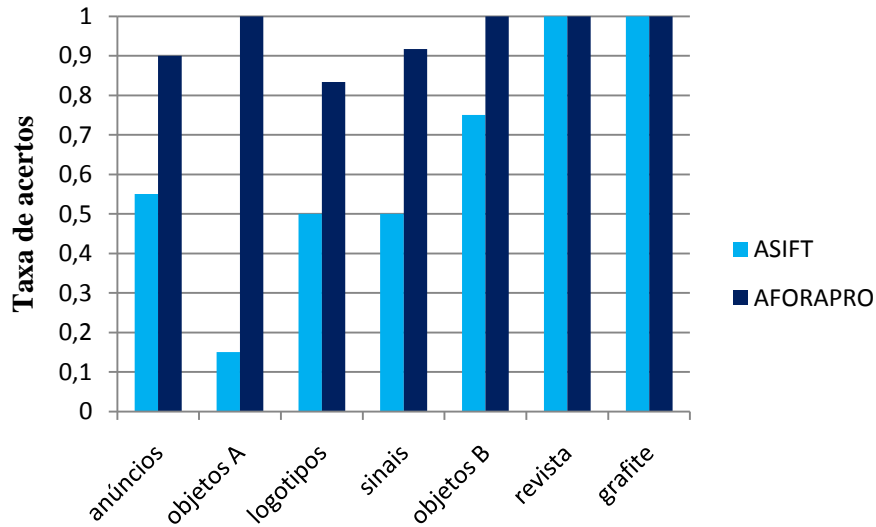
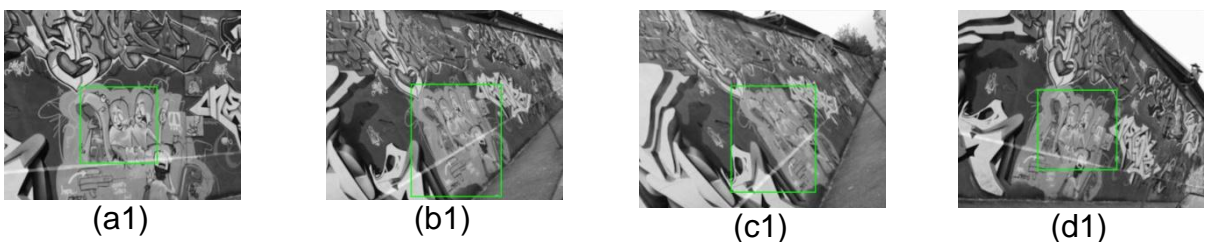


Gráfico 5.3: Taxa de acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 2.

5.2.2.3 Experimentos com imagens do Conjunto 3

Os testes realizados neste experimento, além de permitir comparar o AFORAPRO com o ASIFT, também permitem uma comparação indireta do AFORAPRO com os três outros métodos invariantes a transformações afins (M-SER, Harris-affine, Hessian-affine), pois nos resultados obtidos no trabalho de Morel e You (2009a) utilizando as mesmas imagens conclui-se que nenhum destes métodos consegue achar objetos com uma mudança de ponto de vista $\theta \geq 70^\circ$. A seguir são apresentados os resultados de quatro testes por cada objeto (Grafite, Adam e Lepoin) do Conjunto-3. Na Figura 5.10 apresentam-se os resultados obtidos para o Grafite. Tanto AFORAPRO como ASIFT acertou 100%.



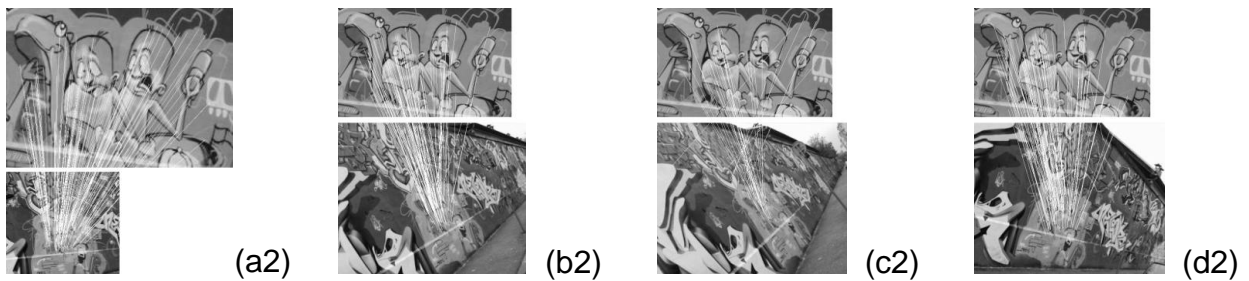


Figura 5.10 Resultados com imagem do Conjunto 3 “grafite”. (a1-d1) Respostas do AFORAPRO, (a2-d2) respostas do ASIFT.

Os resultados para o objeto Adam são apresentados na Figura 5.11. O Adam apresenta duas séries (escalas 1 e 10). Apresentamos dois resultados para cada escala. AFORAPRO supera M-SER, Harris-affine e Hessian-affine para as variações $\theta = 45, 65, 75, 80$. Tanto AFORAPRO como ASIFT acertou 100%.

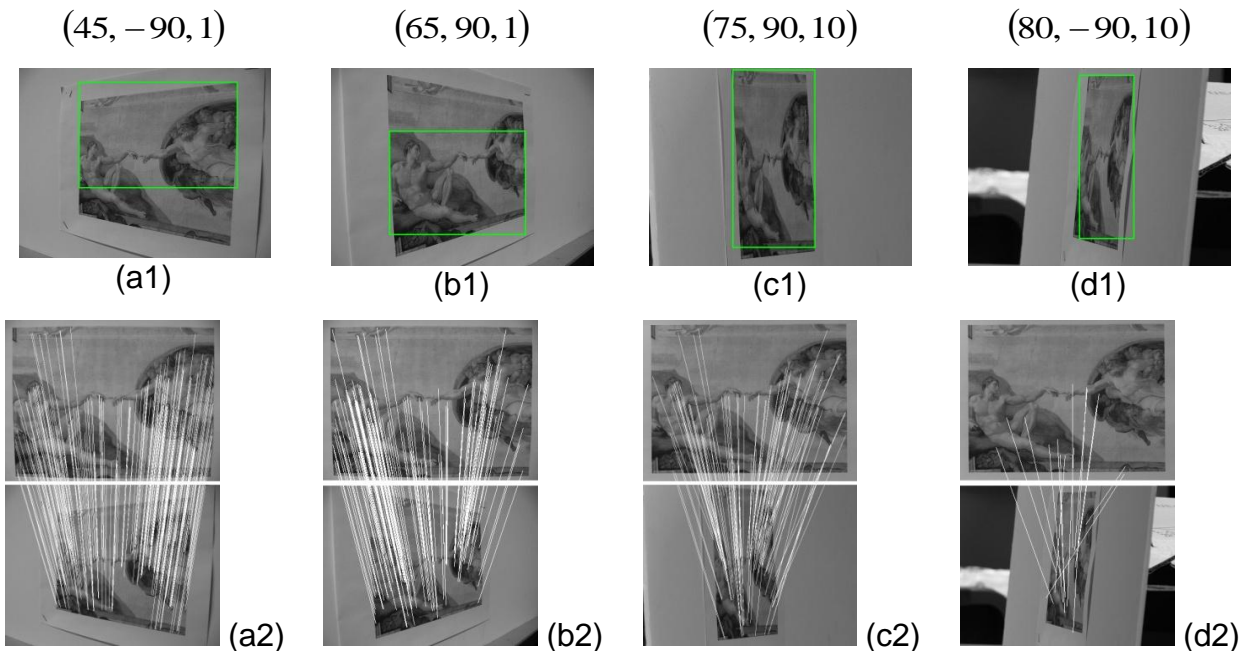


Figura 5.11 Resultados com imagem do Conjunto-3 “Adam”. (a1-d1) Resposta do AFORAPRO, (a2-d2) resposta do ASIFT, (θ, ϕ, r) .

O objeto Lepoint possui três séries, duas em posição horizontal para $\theta = 60$ e $\theta = 75$, e mais uma em posição vertical para $\theta = 60, 80, 81$. O AFORAPRO conseguiu reconhecer todas as instâncias da mesma forma que o ASIFT. A Figura 5.12 apresenta os resultados obtidos para o Lepoint.

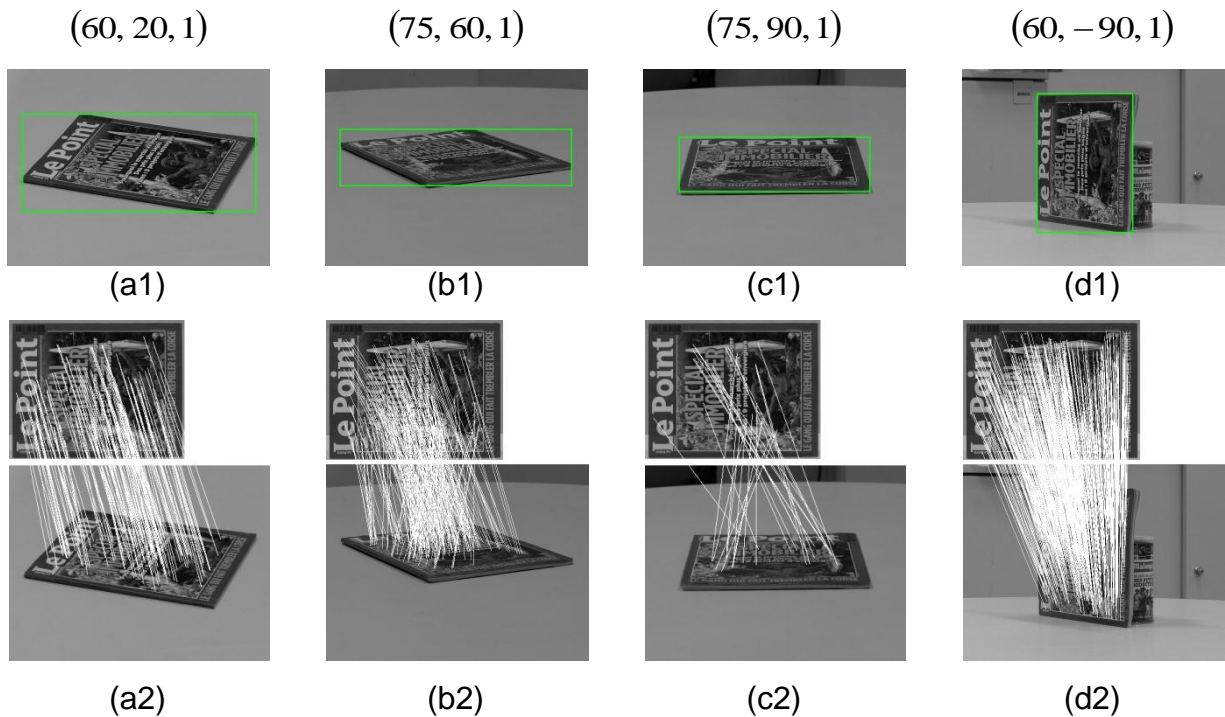


Figura 5.12 Resultados com imagem do Conjunto-3 “LePoint”. (a1-d1) respostas do AFORAPRO, (a2-d2) respostas do ASIFT, (θ, ϕ, r) .

Na Tabela 5.4 resumam-se os resultados obtidos com as imagens do Conjunto-3, onde os dois algoritmos AFORAPRO e ASIFT obtiveram um ótimo desempenho.

Tabela 5.4: Respostas do ASIFT e AFORAPRO com as imagens do Conjunto 3.

Objeto (Q)	Reconhecimentos possíveis	ASIFT		AFORAPRO	
		VP	FN	VP	FN
Adam	16	16	0	16	0
LePoint	23	23	0	23	0
Grafite	8	8	0	8	0
Total	47	47	0	47	0
Taxa de acertos (%)		100,00		100,00	

O Gráfico 5.4 apresenta o resumo da taxa de acertos para os objetos do Conjunto 3. Estes resultados nós permitem verificar a invariância a pontos de vista do AFORAPRO.

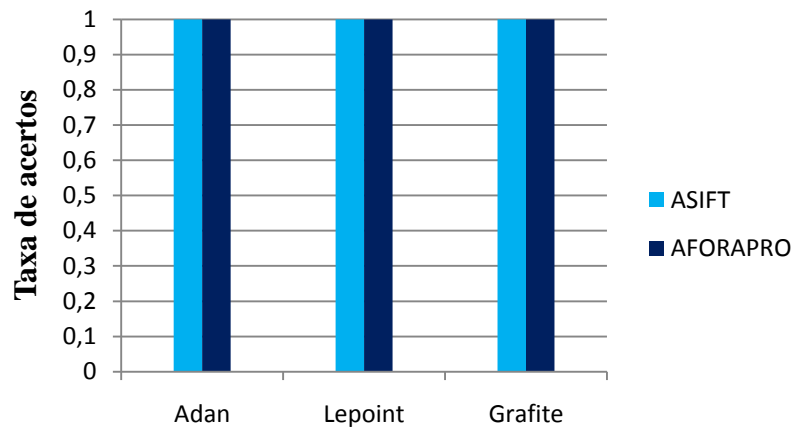
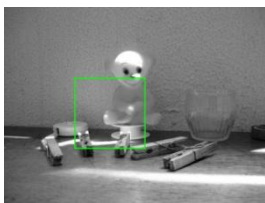


Gráfico 5.4: Taxa de acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 3.

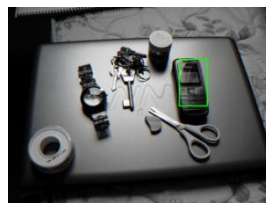
5.3 Experimentos de invariância à mudança de contraste

Os objetivos destes testes foram: (a) Demonstrar que o AFORAPRO é robusto a mudanças de contraste, reconhecendo objetos mesmo em imagens com baixo contraste, (b) Comparar o desempenho de AFORAPRO com o de ASIFT.

Na realização destes experimentos foram utilizadas as imagens do Conjunto-4. Alguns resultados obtidos são apresentados na Figura 5.13. No primeiro caso apresenta-se um teste onde os dois algoritmos AFORAPRO e ASIFT falham (Figura 5.13a e Figura 5.13e), nos dois casos seguintes o AFORAPRO teve sucesso, mas o ASIFT falhou (Figuras 5.13b a Figura 5.13c e Figura 5.13f a Figura 5.13g), e finalmente um teste onde os dois algoritmos tiveram sucesso (Figura 5.13d e Figura 5.13 h).



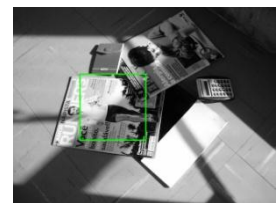
(a)



(b)



(c)



(d)

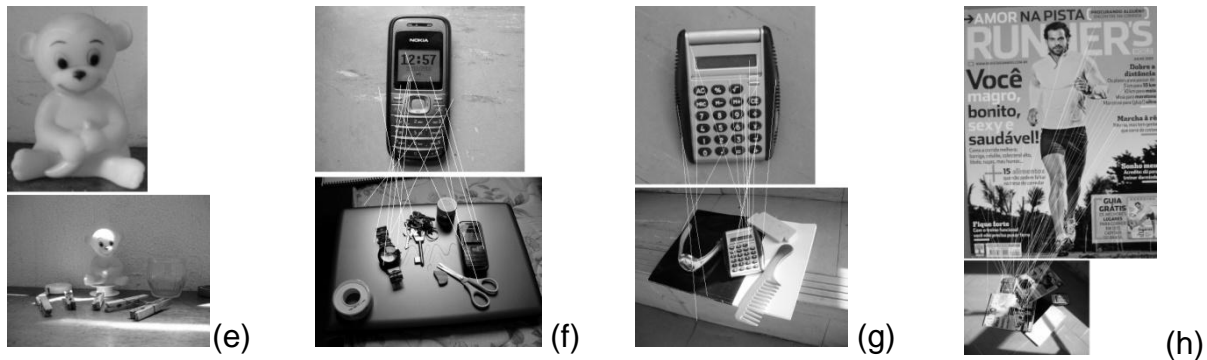


Figura 5.13 Resultados com alguns objetos do Conjunto 4. (a-d) Resposta do AFORAPRO, (e-h) resposta do ASIFT.

O resumo dos resultados obtidos com imagens do Conjunto-4 é apresentado na Tabela 5.5. O ASIFT apresenta 27 erros e foi superado pelo AFORAPRO que apresenta apenas 3 erros. A mudança parcial de iluminação foi a causa de falha nos três casos onde o AFORAPRO errou (e.g. Figura 5.13a). A mudança parcial de iluminação introduz erro na média dos níveis de cinza, usado pelo AFORAPRO no reconhecimento de objetos.

Tabela 5.5: Resposta do ASIFT e AFORAPRO nas imagens do Conjunto 4.

Objeto (Q)	Reconhecimentos possíveis	ASIFT		AFORAPRO	
		VP	FN	VP	FN
Angulômetro	4	2	2	4	0
Calculadora	4	1	3	4	0
Caneca	4	2	2	4	0
Celular	4	0	4	4	0
Separador	4	2	2	3	1
Conector T	4	0	4	3	1
Revista	4	4	0	4	0
Chave	4	1	3	4	0
USB	4	1	3	4	0
Urso	4	0	4	3	1
Total	40	13	27	37	3
Taxa de acertos (%)		32,50		92,25	

O Gráfico 5.6 apresenta a taxa de acertos obtida pelos algoritmos em cada objeto do Conjunto-4. O desempenho do AFORAPRO supera amplamente o ASIFT em objetos com pouca textura. O ASIFT manteve seu ótimo desempenho no reconhecimento de objetos que apresentam abundância de detalhes (e.g. revista).

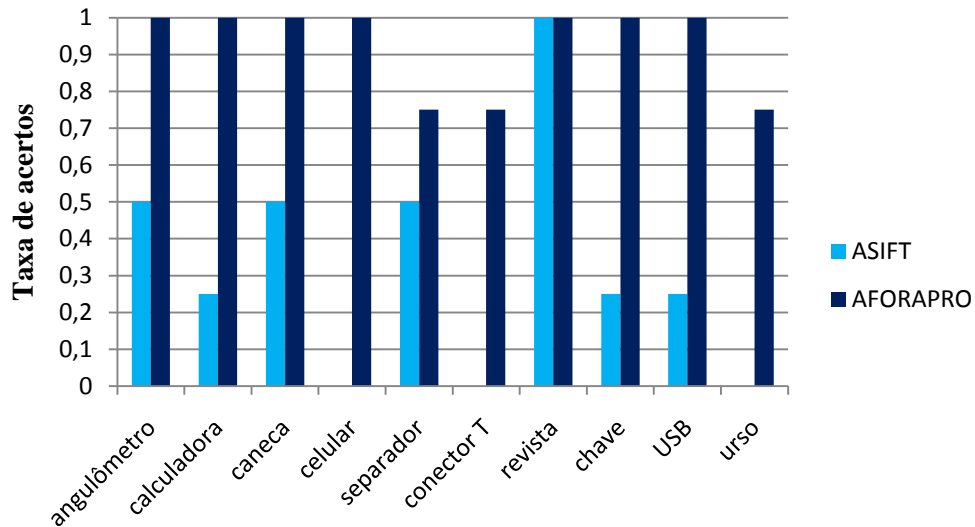


Gráfico 5.5: Taxa de acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 4.

Estes resultados mostram que o AFORAPRO é invariante a pontos de vista e à mudança de contraste, e que seu desempenho supera o ASIFT.

5.4 Experimentos de robustez a padrões repetitivos

Os objetivos destes testes foram: (a) Demonstrar que o AFORAPRO é robusto a padrões repetitivos. (b) Comparar os resultados do AFORAPRO com o ASIFT.

Na realização destes experimentos foram utilizadas as imagens do Conjunto-5. Na Figura 5.14 são apresentados os resultados de quatro testes representativos dos experimentos desta seção. Na sequência, são apresentados os testes onde o AFORAPRO e ASIFT falham (Figura 5.14a e Figura 5.14e), AFORAPRO acerta e ASIFT falha (Figura 5.14b, Figura 5.14c, Figura 5.14f e Figura 5.14g), e os dois acertam (Figura 5.14d e Figura 5.14h).

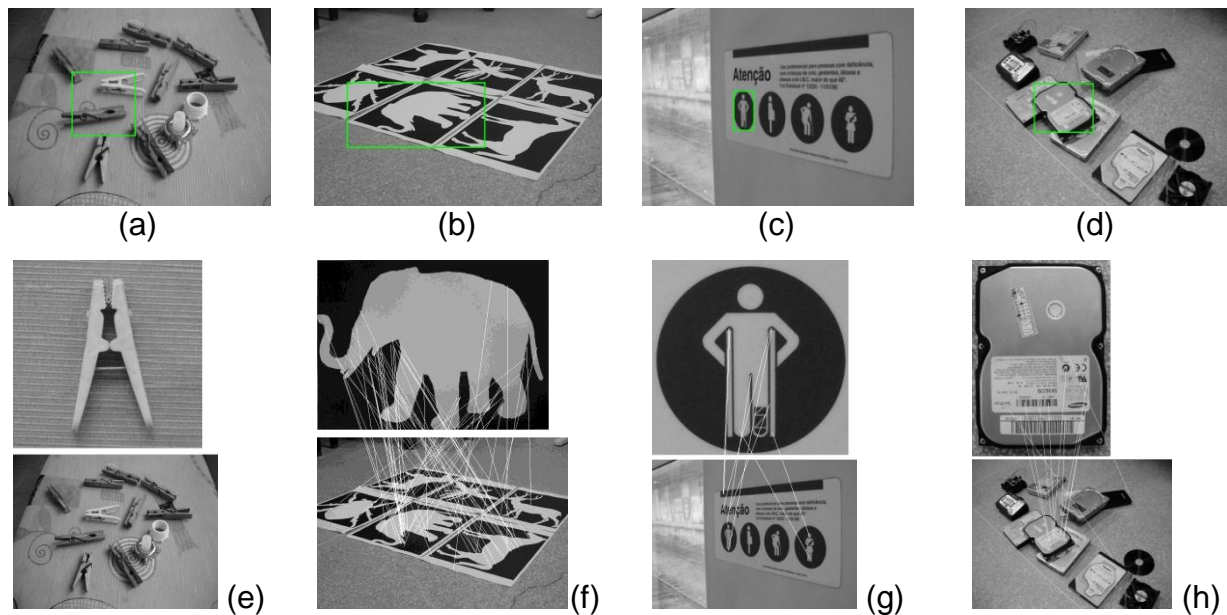


Figura 5.14 Resultados com alguns objetos do Conjunto 5. (a-d) respostas do AFORAPRO, (e-h) respostas do ASIFT.

Os resultados dos experimentos com o Conjunto 5 estão sumarizados na Tabela 5.6.

Tabela 5.6: Respostas do ASIFT e AFORAPRO nas imagens do Conjunto 5.

Objeto (Q)	Reconhecimentos possíveis	ASIFT		AFORAPRO	
		VP	FN	VP	FN
Informação	4	0	4	4	0
Arbusto	4	1	3	4	0
Cálcio	4	3	1	4	0
Carro	4	2	2	3	1
Prendedor	4	0	4	3	1
Elefante	4	0	4	4	0
HD	4	2	2	4	0
Face	4	2	2	4	0
Letra	4	2	2	4	0
Deficiente	4	2	2	4	0
Total	40	14	26	38	2
Taxa de acertos (%)		35,00		95,00	

Estes resultados mostraram que o AFORAPRO teve melhor desempenho que ASIFT. Enquanto ASIFT detectou corretamente 14 instâncias do Q , AFORAPRO detectou 38 instâncias, apresentando menos erros no reconhecimento.

No Gráfico 5.6 é apresentada a taxa de acertos dos métodos ASIFT e AFORAPRO no reconhecimento de cada objeto do Conjunto 5. Há dois casos onde AFORAPRO acerta 100% (informação e elefante) e ASIFT erra todos. Nestes experimentos ASIFT manteve a tendência obtendo seu melhor desempenho no objeto que apresenta mais detalhes na sua estrutura (cálcio).

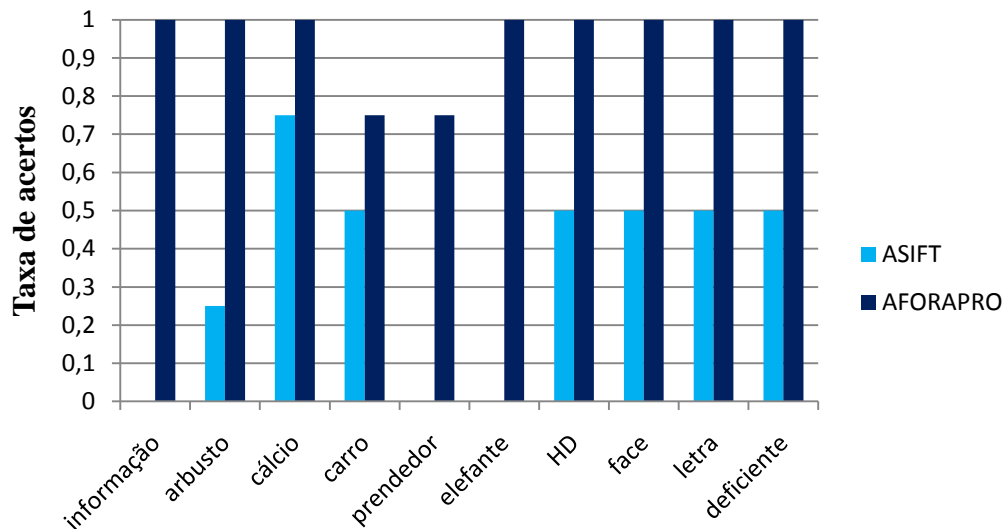


Gráfico 5.6: Taxa de acertos do ASIFT e AFORAPRO nos testes com imagens do Conjunto 5.

Estes resultados mostram que o AFORAPRO é invariante a pontos de vista e padrões repetitivos, e que seu desempenho supera o ASIFT.

5.5 Resumo dos resultados

Foram realizadas três classes de experimentos para testar as qualidades do AFORAPRO frente à invariância a pontos de vista, mudanças de contraste e padrões repetitivos. Nesta seção apresenta-se uma síntese dos experimentos efetuados neste capítulo. O resumo é organizado em duas subseções, uma comparando os resultados de AFORAPRO e FORAPRO e a outras comparando AFORAPRO e ASIFT.

5.5.1 AFORAPRO e FORAPRO

Os dados apresentados na Tabela 5.1 deixam claro que FORAPRO não é invariante a transformações afins e que nossa proposta AFORAPRO consegue melhorar o desempenho deste algoritmo para torná-lo invariante a transformações afins induzidas pela variação de pontos de vista.

5.5.2 AFORAPRO e ASIFT

ASIFT foi utilizado como referência para demonstrar a qualidade do desempenho de AFORAPRO nas três classes de experimento realizados. No Gráfico 5.7 apresenta-se uma síntese da taxa de acertos obtida para cada conjunto de imagens.

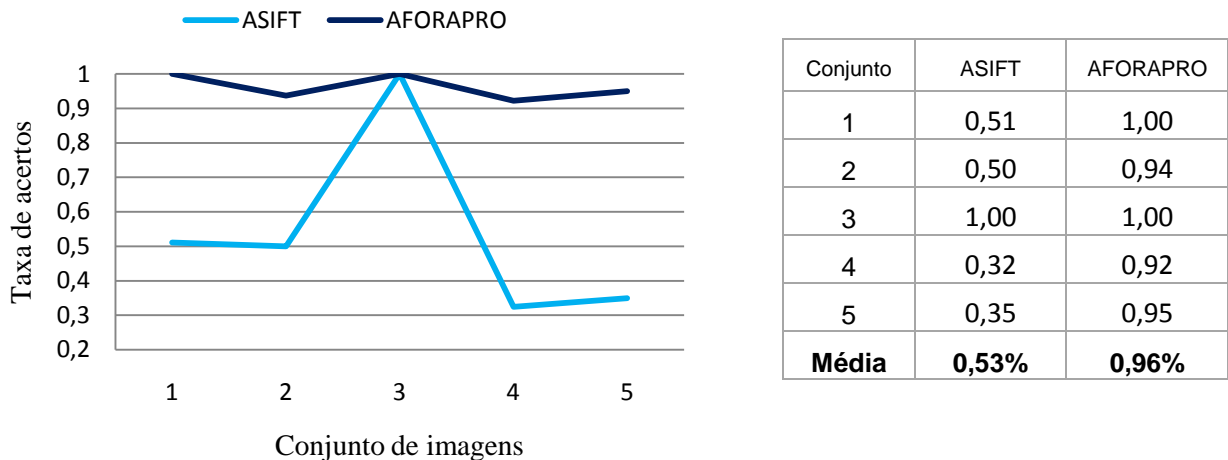


Gráfico 5.7: Taxa de acertos do ASIFT e AFORAPRO para cada um dos conjuntos de imagens utilizados nos experimentos.

O Gráfico 5.7 apresenta um desempenho constante do AFORAPRO que em todos os casos supera a marca dos 0,90%. No entanto ASIFT tem seu pior desempenho nos Conjuntos 4 e 5, que são aqueles que possuem variação de contraste e padrões repetitivos. Estes resultados demonstram que AFORAPRO é invariante a pontos de vista, a grandes mudanças de contraste e a padrões repetitivos. Também podemos verificar que na média da taxa de acertos o AFORAPRO (0,96%) superou amplamente ao ASIFT (0,53%). Note que:

- ASIFT demonstrou em todos os experimentos ser cerca de quatro vezes mais rápido do que AFORAPRO. Os tempos no desempenho do AFORAPRO em cada um dos testes tiveram baixa variação. Na média AFORAPRO gastou 8,52 minutos para cada teste. No entanto, a variação dos tempos no ASIFT foi mais significativa, resolvendo rapidamente aqueles casos onde não apresentou casamentos (0,56 minutos na média) como na Figura 5.14e, mantendo uma média de 2,56 minutos nos casos onde conseguiu algum tipo de reconhecimento e gastando muito mais tempo (9,50 minutos na média) nas imagens que apresentavam quantidade elevada de detalhes na estrutura (e.g. revistas, grafites). A média geral no desempenho do ASIFT foi 2,35 minutos.
- No Gráfico 5.7 verifica-se claramente que o desempenho do ASIFT foi ótimo nas imagens do Conjunto-3, pois aquelas não apresentavam variação do contraste, nem padrões repetitivos. Além disso, as imagens do Conjunto-3 tinham boa quantidade de detalhes e isso possibilita ao ASIFT a extração de pontos chaves. As imagens dos conjuntos restantes possuíam três características específicas: (a) tinham formas simples, (b) tamanho relativamente pequeno, e (c) tinham poucos detalhes ou textura. Todo isso contava a favor de um algoritmo determinado para o reconhecimento de objetos como no caso do AFORAPRO. A baixa taxa de acerto no desempenho do ASIFT sugere então a que ASIFT é um algoritmo com propósitos mais gerais do que o reconhecimento de objetos, como são o reconhecimento de cenas ou a reconstrução de imagens. ASIFT foi escolhido para estes experimentos dado que é o melhor algoritmo proposto na literatura porque na literatura não temos um algoritmo invariante a pontos de vista que seja aplicado especificamente ao reconhecimento de objetos.
- ASIFT demonstrou ser completamente invariante a escala, no entanto AFORAPRO é invariante a escala só num intervalo especificado. Isso quer dizer que o desempenho de AFORAPRO depende do intervalo de escala fornecido pelo usuário e que corresponde ao intervalo de variação entre as escalas dos objetos Q e A .

Capítulo 6.

CONCLUSÕES

Nesta dissertação foi proposto um algoritmo de reconhecimento de objetos invariante a pontos de vista e robusto à mudança de contraste e padrões repetitivos, denominado AFORAPRO. A principal contribuição de nosso trabalho foi estender à aplicabilidade da técnica do FORAPRO (coeficientes de Fourier de projeções radiais e circulares) tornando-lo invariante a mudanças de pontos de vista, e destacando suas vantagens na robustez a padrões repetitivos e mudanças de contraste quando comparado com o ASIFT. Nesta seção, apresentamos os comentários finais e as perspectivas para o trabalho futuro.

6.1 Comentários finais

Nos capítulos 2 e 3 foi apresentada uma descrição evolutiva da área do reconhecimento de objetos e as principais abordagens do reconhecimento de objetos invariante afim.

Foi construída uma base de dados com 350 imagens divididas em cinco grupos, com diversas classes de objetos fotografados em cenários variados, as imagens continham variação de ponto de vista, mudanças de contraste e padrões repetitivos. Com estas imagens foram realizadas três classes de experimentos que nos permitiram avaliar o desempenho do AFORAPRO. Os resultados atestam que AFORAPRO é invariante a pontos de vista, robusto a mudanças de contraste e a padrões repetitivos.

Os resultados do desempenho obtidos com AFORAPRO foram comparados com os resultados dos algoritmos FORAPRO e ASIFT. Cabe mencionar que o ASIFT constitui atualmente o melhor método da literatura para o reconhecimento de padrões invariante a pontos de vista. Na primeira comparação foi demonstrado que AFORAPRO torna ao FORAPRO invariante a pontos de vista. Já na segunda comparação AFORAPRO

demonstrou superioridade (taxa de acerto 0,96%) contra ASIFT (taxa de acerto 0,53%). Em todos os experimentos, o ASIFT foi muito mais rápido do que AFORAPRO. Percebeu-se que o 0,70% do tempo de execução, nas três etapas do AFORAPRO (simular os Q_i , escolher o Q_i apropriado e executar o casamento), é consumido na segunda etapa devido à falta de uma estrutura de dados para indexar os vetores de características.

6.2 Trabalho futuro

O desafio de novas pesquisas visando melhorar o desempenho do AFORAPRO pode ser dividido em três partes:

- Diminuir o tempo de execução: poderíamos estudar três possibilidades: (a) Implementar uma estrutura de dados para acelerar o processo de recuperação das correspondências entre os vetores de características. (b) Analisar a possível diminuição do número de distorções simuladas Q_i . (c) Acelerar o tempo de execução por meio da implementação em Hardware.
- Conseguir completa invariância a escala: como foi mencionado no capítulo anterior, AFORAPRO precisa da pré-definição do intervalo de escala. Este problema poderia ser solucionado mediante a incorporação de uma estrutura piramidal.
- Explorar a utilização das propriedades da cor: muitas aplicações práticas no reconhecimento de objetos poderiam ser bem mais simples com a utilização das propriedades da cor.

REFERÊNCIAS

- ALVAREZ, L.; MORALES, F. Affine morphological multi-scale analysis of corners and multiple junctions. *International Journal of Computer Vision*. vol. 2 (25): pp 95-107. (1997).
- BALLARD, D.H.: Generalizing the Hough transform to detect arbitrary. *Pattern Recognition*. vol. 13 (2): pp. 111-122. (1981).
- BALLESTER, C.; GONZÁLEZ, M.: Affine invariant texture segmentation and shape from texture by variational methods, *Journal of Mathematical Imaging and Vision*. vol. 9 (2): pp. 141–171. (2004).
- BAUMBERG. A. Reliable feature matching across widely separated views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, USA, pp. 774-781. (2000).
- BAY, HERBERT.; TUYTELAARS, T.; GOOL, V.: SURF: Speeded up robust features. In *Proceedings of the ninth European Conference on Computer Vision*, May (2006).
- BEAUDET, P. R. Rotational invariant image operators. In *Proc. IAPR 1978*, pp. 579-583. (1978).
- CHOI, M.; KIM, W.: A novel two stage template matching method for rotation and illumination invariance, *Pattern Recognition*. vol. 35 (1): pp. 119–129. (2002).
- CYGANSKI, R.; ORR, T.; COTT; DODSON, R. Development, implementation, testing, and application of an affine transform invariant curvature function. *Proceedings of the 1st Int. Conf on Computer Vision*, pp. 496-500. London. (1987).
- DERICHE, R.; GIRAUDON, G. A computational approach for corner and vertex detection. *International Journal of Computer Vision*, vol. 10(2): pp. 101-124. (1993).
- DONOSER, M.; BISCHOF, H. Efficient maximally stable extremal region (MSER) tracking. In *Proc. of IEEE CVPR*, pp 553-560. (2006).
- DOUBEK, P.; MATAS, J.; PERDOCH, M.; CHUM, O.: Image Matching and Retrieval by Repetitive Patterns. *ICPR*. pp. 3195-3198. (2010).
- ECABERT, O.; THIRAN, J. Adaptive Hough Transform for the Detection of Natural Shapes Under Weak Affine Transformations. *Pattern Recognition Letters*, vol. 25(12): pp. 1411–1419. (2004).

- ELI, PELI: Contrast in complex images. Optical Society of American. Boston, Massachusetts. vol. 7 (10): pp. 2032-2040. Oct. (1990).
- FAWCETT, TOM. An introduction to ROC analysis. Pattern Recognition Letters. USA. vol. 27 (8): pp. 861-874. (2006).
- FLUSSER, J.; SUK, T. Pattern Recognition by Affine Moment Invariants. Pattern Recognition. vol. 26(1): pp. 167–174. (1993).
- FORSSÈN P.; LOWE, D. Shape descriptors for maximally stable extremal regions. In IEEE ICCV, Rio de Janeiro Brazil. vol. 31 (1): pp. 59-73. (2007).
- FORTNER, W. A framework for low level feature extraction. Lecture notes in computer science In Proc. ECCV94, Sweden. pp. 383-394. May (1994).
- FREDERIC, G.: JEAN-MICHEL, M.: ROBERT, R.: **Contrast invariant image analysis and PDE's**. IPAM GBM Tutorial. (2001).
- GAO, KE; ZHANG, Y.; ZHANG, W.; LIN, S.: Affine stable characteristic based sample expansion for object detection. Xian China. ACM 978-1-4503-0117. CIVR'10. pp. 422-429. (2010).
- GONZÁLEZ, R.; WOODS, R.: **Digital images processing**. Second edition. Prentice Hall, Inc. Upper Saddle River, New Jersey. (2002).
- GONZÁLEZ, R.; WOODS, R.; EDDINS S.: **Digital image processing using Matlab**. Second edition. Prentice Hall, Upper Saddle River, New Jersey. (2004).
- HARRIS, C.: Geometry from visual motion. In: Active Vision, A. Blake and A. Yuille (Eds.), MIT Press. pp. 263-284. (1992).
- HARRIS, C; STEPHENS, M.: A combined corner and edge detector. In Proceedings from the Fourth Alvey Vision Conference. vol. 15: pp. 147–151. (1988).
- HSIAO, J.: CHEN, C.: CHIEN, L.: CHEN, M.: A new approach to image copy detection based on extended feature sets, IEEE Transactions on Image Processing. vol. 16(8): pp. 2069–2079. (2007).
- HU, M. K.: Visual Pattern Recognition by Moment Invariants. IEEE Transactions in Information Theory. vol. 8: pp 179-187, (1962).
- HUTTENLOCHER, D.; ULLMAN, S. Object Recognition Using Alignment. Proc. of the 1st Int. Conf. on Computer Vision: pp. 102-111. London. (1987).

- INTEL. *OpenCV - Open Source Computer Vision Library*. 2009. Disponível em: <<http://software.intel.com/en-us/articles/intel-ipp-support-model-changed-in-opencv-2x/>>. Acesso em junho (2009).
- JU, HAN; KAI-KUANG, M. Rotation-invariant and scale-invariant Gabor feature for texture image retrieval. *Image and Vision Computing*. Nayang Singapore. vol. 25 (9): pp. 1474-1481. (2007).
- KIM, H. Y. *ProEikon - Rotinas e programas em C++ para processamento de imagens e visão computacional*. São Paulo, 2009. Disponível em: <<http://www.lps.usp.br/~hae/software>>. Acesso entre janeiro. 2009 e agosto. 2010c.
- KIM, H. Y.: Forapro Template Matching. Versão disponível para testes. <<http://www.lps.usp.br/~hae/software/forapro/index.html> >. Acesso em 22 nov, (2009).
- KIM, H. Y.: FORAPRO: Rotation-Discriminating Template Matching Based on Fourier Coefficients of Radial Projections with Robustness to Scaling and Partial Occlusion. *Pattern Recognition*. vol. 43 (3): pp. 859-872. (2010a).
- KIM, H. Y.: *IMG - Rotinas e programas em C++ para processamento de imagens e visão computacional*. São Paulo, 2008. Disponível em: <<http://www.lps.usp.br/~hae/software>>. Acesso entre outubro 2008 e agosto 2010b.
- KIM, H. Y.; ARAÚJO, S. A. Grayscale Template-Matching Invariant to Rotation, Scale, Translation, Brightness and Contrast. *Lecture Notes in Computer Science*, Springer-Verlag, vol. 4872: pp. 100-113. (2007).
- KIM, H. Y.; PÉREZ, G. A. Reconhecimento de Padrões Invariante Afim Baseado nos Coeficientes de Fourier de Projeções Radiais. VI Workshop de Visão Computacional, UNESP Campus de Presidente Prudente – Brasil. (2010).
- KITCHEN, L.; ROSENFELD, A. Gray-level corner detection. *Pattern Recognition Letters*, vol. 1: pp. 95-102. (1982).
- KLINGER, A. Pattern and Search Statistics. In Rustagi J.S. (Ed) *Optimizing Methods in Statistics*. Academic Press. New York. (1971).
- KOENDERINK J.J. The structure of images. *Biological Cybernetics*. vol. 50: pp. 363–370. (1984).
- LIN, Y.; CHEN, C.; WEI, C. New method for subpixel image matching with rotation invariance by combining the parametric template method and the ring projection transform process. *Optical Engineering*. vol.45 (6): pp. 067202(1-9). (2006).

- LINDEBERG, T. Scale-space for discrete signals. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12(3): pp. 234–254. (1990).
- LINDEBERG, T. Scale-space theory: A basic tool for analyzing structures at different scales. J. of Applied Statistics, vol. 21(2): pp. 224–270. (1994).
- LINDEBERG, T.; GARDING, J. Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure. Image and Vision Computing, vol. 15(6): pp. 415-434. (1997).
- LOWE, D.: Object recognition from local scale-Invariant features (siftWin32.exe). Versão disponível para testes. <<http://www.cs.ubc.ca/~lowe/keypoints/>>. Acesso em 22 Jan, (2010).
- LOWE, DAVID G.: Local feature view clustering for 3d object recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA, pp. 682-688, Dec (2001).
- LOWE, DAVID G.: Object recognition from local scale-invariant features (SIFT). International Conference on Computer Vision. Corfu, Greece. pp. 1150-1157. Sep (1999).
- LOWE, DAVID: Distinctive image features from scale invariant keypoints, In International Journal on Computer Vision, vol. 60(2): pp. 91-110. (2004).
- MALLAT, S.G. A Theory of Multiresolution Signal Processing: The Wavelet Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. PAMI-11 (6): pp 674-693. (1989).
- MATAS, J.; CHUM, O.; URBAN, M.; PAJDLA, T. MSER: Robust wide baseline stereo from maximally stable extremal regions, Image Vision Compute. vol. 22 (10): pp 761–767. (2004).
- MATAS, J.; CHUM, O.; URBAN, M.; PAJDLA, T. Robust wide baseline stereo from maximally stable extremal regions. In: British Machine Vision Conference. pp. 384–393. (2002).
- MIKOLAJCZYK, C.: Harris-affine e Hessian-affine. <<http://www.robots.ox.ac.uk/~vgg/research/affine/>>. Acesso em 13 agosto, (2010a).
- MIKOLAJCZYK, C.: <http://lear.inrialpes.fr/people/mikolajczyk/Database>. Acesso em: 7 setembro (2010b).

- MIKOLAJCZYK, K.; SCHMID, C.: A performance evaluation of local descriptors. In International Conference on Computer Vision and Pattern Recognition, vol. 2: pp. 257–263. Jun (2003).
- MIKOLAJCZYK, K.; SCHMID, C.: An affine invariant interest point detector. In European Conference on Computer Vision (ECCV). Copenhagen, Denmark. pp. 128-142. (2002).
- MIKOLAJCZYK, K.; SCHMID, C.: Indexing based on scale invariant interest points. *Proc. ICCV*, 1: pp. 525–531. (2001).
- MIKOLAJCZYK, K.; SCHMID, C.: Scale and Affine Invariant Interest Point Detectors. International Journal of Computer Vision, Springer Netherlands. vol. 60 (1): pp. 63-86. Oct (2004).
- MIKOLAJCZYK, K.; TUYTELAARS, T.; SCHMID, C.; ZISSERMAN, A.; MATAS, J.; SCHAFFALITZKY, F.; KADIR, T.; GOOL, L.: A comparison of affine region detector. International journal of computer vision. vol. 65 (1): pp. 43-72. (2005).
- MORAVEC, H. P.: Towards automatic visual obstacle avoidance. In IJCAI. pp. 584-590. (1977).
- MOREL, J. M.; YU, G.: ASIFT: A new framework for fully affine invariant image comparison. In: SIAM Journal on Image Sciences, vol. 2 (2). (2009a).
- MOREL, J.; YU, G.: ASIFT: A Fully Affine Invariant Image Comparison Method. in IEEE Int. Conf. Acoustics, Speech and Signal Processing. pp. 1597-1600. (2009b).
- MOREL, J.M.; YU, G.: SIFT and ASIFT: IPOL Workshop (2009). <<http://www.cmap.polytechnique.fr/~yu/research/ASIFT/>>. Acesso desde: 9 jan-15 dez (2010).
- PODBREZNIK, P.; POTOČNIK, B.: Estimating correspondence between arbitrarily selected points in two widely-separated views. Advanced Engineering Informatics. vol. 24 (3): pp. 367-366. (2010).
- PRATT, WILLIAM: **Digital image processing**. Fourth edition. Wiley-interscience. Los Altos, California USA. ISBN: 978-0-471-76777-0. (2007).
- RABIN, J.; DELON, J.; GOUSSEAU, Y. A statistical approach to the matching of local features. SIAM Journal on Imaging Sciences. vol. 2: pp 931-958. (2009).
- RAHTU, E.; SALO, M.; HEIKKILA. Affine invariant pattern recognition using multi-scale auto-convolution. IEEE Trans. Pattern Analysis and Machine Intelligence. vol. 27(6): pp. 908-918. (2005).

- RIPLEY, B. D.: **Pattern Recognition and Neural Networks**. Cambridge: Cambridge University Press. (1996).
- RUBNER, Y.; TOMASI, C. Texture-Based Image Retrieval Without Segmentation. *Proc. ICCV*. pp. 1018-1024. (1999).
- SCHAFFALITZKY, F.; ZISSERMAN, A. Viewpoint Invariant Texture Matching and Wide Baseline Stereo, *Proc. ICCV*. vol. 2: pp. 636-643. (2001).
- SCHMID, C.; MOHR, R. Local gray value invariants for image retrieval. In: IEEE Trans. on Pattern Analysis and Machine Intelligence. pp. 530-534. (1997).
- SHECHTMAN, E.; IRANI, M.: Matching local Self-Similarities across images and videos. In: Proceedings of CVPR. (2007).
- SHI, J.; TOMASI, C. Good features to track. IEEE conference on Computer Vision and Pattern Recognition. CVPR94, Seattle. June (1994).
- SHIN, D.; TJAHHADI, T. Clique descriptor of affine invariant regions for robust wide baseline image matching. *Pattern Recognition*. vol. 43: pp. 3261-3272. (2010).
- SZELISKI, RICHARD.: **Computer Vision: Algorithms and Applications**, Springer, ISBN: 1848829345. (2009).
- TANIMOTO, S. L. Image data structures, in *Structured Computer Vision*. (S. L. Tanimoto and A. Klinger. Eds.), Academic Press, New York USA. (1980).
- TANIMOTO, S. L.; PAVLIDIS, T. A hierarchical data structure for picture processing. *Computer Graphics and Image Processing*. vol. 4: pp. 104-119. (1975).
- TARR, M.; BULTHOFF H.: Is Human Object Recognition Better Described By Geon-Structural-Descriptions Or By Multiple-Views? *Journal of Experimental Psychology: Human Perception and Performance*. vol. 21(6): pp. 1494-1505. (1995).
- TAUBIN, G.; COOPER, D. Object recognition based on moment invariants. In *Geometric invariance in computer vision*, J L Mundy and A Zisserman(eds.), MIT Press, Cambridge, Mass. pp. 375-497. (1992).
- TEAGUE, M. R. Image Analysis via the General Theory of Moments. *Journal of Optical Society of America*. vol. 70 (8): pp. 920-930. (1980).
- THEODORIDIS, S.; KOUTROUMBAS, K.: **Pattern Recognition**. Fourth edition. San Diego, California, USA. ISBN: 978-1-59749-272-0. (2009).

- THOMPSON, D.; MUNDY, J. Three-Dimensional Model Matching from an Unconstrained Viewpoint. Proc. of the IEEE Int. Conf. on Robotics and Automation, Raleigh, North Carolina. pp. 208-220. (1987).
- TREIBER, MARCO.: **An Introduction to Object Recognition.** Springer, London Dordrecht Heidelberg, New York. (2010).
- TSAI, D.; CHANG, C. Rotation-invariant pattern matching using wavelet decomposition. Pattern Recognition Letters. vol. 23: pp. 191-201. (2002).
- TUYTELAARS, T.; GOOL, L. J. Wide baseline stereo matching based on local, affinely invariant regions, in: Proceedings of the British Machine Vision Conference. pp. 42–56. (2000).
- TUYTELAARS, T.; GOOL, V. Content-based image retrieval based on local affinely invariant regions. In Int. Conf. on Visual Information Systems. pp. 493-500. (1999).
- ULLAH, F.; KANEKOI, S. Using orientation codes for rotation-invariant template matching. Pattern Recognition. vol. 37 (2): pp. 201-209. (2004).
- VAN GOOL; MOONS, T.; UNGUREANU, D. Affine/Photometric Invariants for Planar Intensity Patterns. Proceedings of European Conference on Computer Vision. vol. 1: pp. 642–651. (1996).
- WENDEL, A.; DONOSER, M.; BISCHOF, H.: Unsupervised Facade Segmentation Using Repetitive Patterns. Lecture notes in computer science. Pattern Recognition. vol. 6376: pp. 51-60. (2010).
- WU, W.: YANG, J.: Object fingerprints for content analysis with applications to street landmark localization. ACM Multimedia. (2008).