

**UNIVERSIDADE DE SÃO PAULO**

**ESCOLA POLITÉCNICA**

**Departamento de Engenharia de Sistemas Eletrônicos**

**SIDNEI ALVES DE ARAÚJO**

**CASAMENTO DE PADRÕES EM IMAGENS DIGITAIS LIVRE DE  
SEGMENTAÇÃO E INVARIANTE SOB TRANSFORMAÇÕES DE  
SIMILARIDADE**

Tese apresentada à Escola Politécnica da  
Universidade de São Paulo para obtenção  
do título de Doutor em Engenharia.



São Paulo  
2009

**UNIVERSIDADE DE SÃO PAULO**

**ESCOLA POLITÉCNICA**

**Departamento de Engenharia de Sistemas Eletrônicos**

**SIDNEI ALVES DE ARAÚJO**

**CASAMENTO DE PADRÕES EM IMAGENS DIGITAIS LIVRE DE  
SEGMENTAÇÃO E INVARIANTE SOB TRANSFORMAÇÕES DE  
SIMILARIDADE**

Tese apresentada à Escola Politécnica da  
Universidade de São Paulo para obtenção  
do título de Doutor em Engenharia.

Área de Concentração:  
Sistemas Eletrônicos

Orientador: Prof. Livre-Docente  
Hae Yong Kim



São Paulo  
2009

**Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.**

**São Paulo, 18 de novembro de 2009.**

**Assinatura do autor**



**Assinatura do orientador**

## **FICHA CATALOGRÁFICA**

**Araújo, Sidnei Alves de**

**Casamento de padrões em imagens digitais livre de segmentação e invariante sob transformações de similaridade / S.A. de Araújo. -- ed.rev. -- São Paulo, 2009.**

**142 p.**

**Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.**

**1. Processamento digital de imagens 2. Visão computacional  
3. Reconhecimento de padrões I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II. t.**

## DEDICATÓRIA

À minha esposa Célia, por todo amor, companheirismo, incentivo e apoio em todos os momentos.

## AGRADECIMENTOS

Primeiramente, agradeço a Deus por iluminar meu caminho e me dar forças, principalmente nos momentos mais difíceis.

Ao Prof. Dr. Hae Yong Kim, meu orientador, pela receptividade, paciência e por todo o apoio e incentivo ao longo do desenvolvimento deste trabalho.

À minha esposa Célia, meus amados filhos Nicole e Henry e enteadas Giovanna e Pietra pelos diversos momentos de suas vidas sacrificados com minha ausência.

Aos meus pais, Nelcina e João, pelo incentivo a continuidade dos estudos até a chegada a este doutorado, orientação e amor durante toda a vida.

Às minhas irmãs Inês, Marinez, Andréia, Valéria e Simone pelo apoio e incentivo.

Aos meus queridos amigos Dr. André Felipe H. Librantz e Ms. Cristina Koyama pela amizade, compreensão e pela preciosa ajuda nas revisões de textos e idéias que foram importantes para o refinamento deste trabalho.

À amiga Ms. Rosane Beatriz Oliveira Severo pelo incentivo e apoio, essenciais para meu ingresso no programa de doutorado.

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

Aos amigos da Universidade Nove de Julho, em especial ao Dr. Marcos Alberto Bussab e ao Dr. Cleber Gustavo Dias.

Aos amigos Arildo, Assis, Edite, Valmir, Ronaldo, Roni, Betinho, Rafael, Ivanildo, Paulo, Tita, Junior e Toni pelos momentos de descontração.

Aos amigos Dr. José Fusari e Dra. Mariazinha Fusari (in memoriam), que sempre me incentivaram a dar continuidade nos estudos.

Aos membros da banca pelas importantes contribuições para a finalização deste trabalho.

A todos que por ventura eu tenha esquecido de mencionar, mas nem por isso são menos importantes!

"A educação faz um povo fácil de ser liderado, mas difícil de ser dirigido; fácil de ser governado, mas impossível de ser escravizado."

Henry Peter

## RESUMO

Reconhecimento de padrões em imagens é um problema clássico da área de visão computacional e consiste em detectar um padrão ou objeto de referência (*template*) em uma imagem digital. A maioria dos métodos para esta finalidade propostos na literatura simplifica as imagens por meio de operações como binarização, segmentação e detecção de bordas ou pontos de contorno, para em seguida extrair um conjunto de atributos descritores. O problema é que esta simplificação pode descartar informações importantes para descrição dos padrões, fazendo diminuir a robustez do processo de detecção. Um método eficiente deve ter a habilidade de identificar um padrão sujeito a algumas transformações geométricas como rotação, escalonamento, translação, cisalhamento e, no caso de métodos para imagens coloridas, deve ainda tratar do problema da constância da cor. Além disso, o conjunto de atributos que descrevem um padrão deve ser pequeno o suficiente para viabilizar o desenvolvimento de aplicações práticas como um sistema de visão robótica ou um sistema de vigilância. Estes são alguns dos motivos que justificam os esforços empreendidos nos inúmeros trabalhos desta natureza encontrados na literatura. Neste trabalho é proposto um método de casamento de padrões em imagens digitais, denominado Ciratefi (*Circular, Radial and Template-Matching Filter*), livre de segmentação e invariante sob transformações de similaridade, brilho e contraste. O Ciratefi consiste de três etapas de filtragem que sucessivamente descartam pontos na imagem analisada que não correspondem ao padrão procurado. Também foram propostas duas extensões do Ciratefi, uma que utiliza operadores morfológicos na extração dos atributos descritores, denominada Ciratefi Morfológico e outra para imagens coloridas chamada de color Ciratefi. Foram realizados vários experimentos com o intuito de comparar o desempenho do método proposto com dois dos principais métodos encontrados na literatura. Os resultados experimentais mostram que o desempenho do Ciratefi é superior ao desempenho dos métodos empregados na análise comparativa.

Palavras-chave: Visão computacional. Casamento de padrões. Reconhecimento de objetos. Transformações de similaridade. Invariância de cores.

## ABSTRACT

Pattern recognition in images is a classical problem in computer vision. It consists in detecting some reference pattern or template in a digital image. Most of the existing pattern recognition techniques usually apply simplifications like binarization, segmentation, interest points or edges detection before extracting features from images. Unfortunately, these simplification operations can discard rich grayscale information used to describe the patterns, decreasing the robustness of the detection process. An efficient method should be able to identify a pattern subject to some geometric transformations such as translation, scale, rotation, shearing and, in the case of color images, should deal with the color constancy problem. In addition, the set of features that describe a pattern should be sufficiently small to make feasible practical applications such as robot vision or surveillance system. These are some of the reasons that justify the effort for development of many works of this nature found in the literature. In this work we propose a segmentation-free template matching method named Ciratefi (*Circular, Radial and Template-Matching Filter*) that is invariant to rotation, scale, translation, brightness and contrast. Ciratefi consists of three cascaded filters that successively exclude pixels that have no chance of matching the template from further processing. Also we propose two extensions of Ciratefi, one using the mathematical morphology approach to extract the descriptors named Morphological Ciratefi and another to deal with color images named Color Ciratefi. We conducted various experiments aiming to compare the performance of the proposed method with two other methods found in the literature. The experimental results show that Ciratefi outperforms the methods used in the comparison analysis.

Keywords: Computer vision. Template Matching. Object recognition. Similarity transformations. Color invariance.

## LISTA DE ILUSTRAÇÕES

Figura 2.1 – Exemplo de casamento de padrões em imagens.....	41
Figura 2.2 – Etapas de um sistema de visão computacional típico. ....	42
Figura 2.3 – Processo de reconhecimento de padrões em imagens digitais.....	47
Figura 2.4 – Matriz de confusão ou tabela de contingência. ....	49
Figura 3.1 – Correspondência de pontos, feita por SIFT, entre duas imagens. ....	51
Figura 3.2 – Esquema de funcionamento do processo de detecção dos pontos candidatos.....	53
Figura 3.3 – Resultado do processo de detecção do padrão sapo por SIFT. A instância com contraste invertido não foi detectada.....	54
Figura 3.4 – Resultado do processo de detecção do padrão sapo por EasyMatch...	55
Figura 3.5 – Tela do EasyMatch para leitura da imagem $T$ e extração de seus atributos.....	56
Figura 3.6 – Tela do EasyMatch para leitura da imagem $A$ e inserção dos parâmetros utilizados pelo algoritmo de casamento de padrões.....	57
Figura 3.7 – Resultados numéricos do processo de detecção do padrão sapo por EasyMatch.....	57
Figura 4.1 – Processo de detecção do padrão sapo pelo algoritmo de força bruta...	62
Figura 4.2 – Filtragem por Cifi.....	66
Figura 4.3 – Filtragem por Rafi.....	68
Figura 4.4 – Resultado da detecção por Ciratefi. Os objetos detectados estão marcados com um “X” vermelho. ....	69
Figura 4.5 – Esquema de funcionamento das três etapas de filtragem do Ciratefi. ..	70
Figura 4.6 – Matriz multi-escala $\mathbf{C}_T$ com 6 linhas (escalas) e 13 colunas (círculos)..	71
Figura 4.7 – Esquema de funcionamento das três etapas de filtragem do Ciratefi morfológico.....	78

Figura 4.8 – Imagens da mesma cena sob diferentes condições de iluminação.....	80
Figura 4.9 – Imagens do mesmo objeto adquiridas sob diferentes ângulos de iluminação. ....	80
Figura 4.10 – Representação gráfica do espaço de cores CIELAB. ....	85
Figura 5.1 – Padrões a serem procurados nas imagens do conjunto I. Cada imagem tem 51×51 pixels. ....	90
Figura 5.2 – Resultados da detecção das instâncias dos 5 padrões por Ciratefi. ....	91
Figura 5.3 – Resultados da detecção do símbolo do McDonald’s por Ciratefi. ....	93
Figura 5.4 – Resultado da detecção do prédio em forma de “H”, em uma imagem de sensoriamento remoto, por Ciratefi. ....	94
Figura 5.5 – Os 12 Padrões a serem detectados nas 30 imagens do conjunto IV. Cada padrão tem 71×71 pixels. ....	95
Figura 5.6 – Resultados da detecção dos padrões em imagens do conjunto IV.....	96
Figura 5.7 – Imagens resultantes da primeira filtragem com os pixels candidatos em magenta .....	103
Figura 5.8 – Imagens resultantes da segunda filtragem com os pixels candidatos em magenta .....	104
Figura 5.9 – Resultado do processo de detecção do padrão cachorro por SIFT. ...	107
Figura 5.10 – Pontos chave extraídos por SIFT de imagens dos conjuntos I e III. ....	110
Figura 5.11 – Resultados da detecção de padrões por SIFT em imagens dos conjuntos II e III. ....	112
Figura 5.12 – Resultados da detecção de padrões por EasyMatch em imagens dos conjuntos I, II e III. ....	114
Figura 5.13 – Exemplos de imagens da base utilizada nos experimentos com color Ciratefi. ....	119
Figura 5.14 – Gráfico ilustrando a taxa de acertos do color Ciratefi para cada um dos subconjuntos de imagens, considerando diferentes pares de $\delta$ e $\eta$ .....	121
Figura 5.15 – Quantidade de pontos chave extraídos por color SIFT de imagens do conjunto bike, gradualmente afetadas por borramento. ....	123

Figura 5.16 – Quantidade de pontos chaves extraídos por color SIFT de imagens do conjunto Ubc, gradualmente afetadas por compressão JPEG..... 124

Figura 5.17 – Resultados da detecção de 10 padrões em uma imagem do conjunto graffiti. .... 125

## LISTA DE TABELAS

Tabela 2.1. Agrupamento das transformações geométricas .....	38
Tabela 5.1. Resultados do experimento com Ciratefi envolvendo as imagens do conjunto I.....	91
Tabela 5.2. Resultados do experimento com Ciratefi envolvendo as imagens do conjunto IV. ....	97
Tabela 5.3. Resumo dos experimentos realizados com Ciratefi.....	97
Tabela 5.4. Sensibilidade do Ciratefi com relação ao número de círculos $l$ .....	98
Tabela 5.5. Sensibilidade do Ciratefi com relação ao número de linhas radiais $m$ ....	99
Tabela 5.6. Sensibilidade do Ciratefi com relação aos limiares $t_1$ , $t_2$ e $t_3$ .....	100
Tabela 5.7. Sensibilidade do Ciratefi com relação aos limiares $t_\beta$ e $t_\gamma$ . ....	101
Tabela 5.8. Resultados do Ciratefi morfológico na detecção do padrão sapo, nas imagens do conjunto I e os resultados obtidos por Ciratefi na mesma tarefa. ....	106
Tabela 5.9. Quantidade média de pixels candidatos na primeira e na segunda fase gerados pelos métodos Ciratefi morfológico e Ciratefi. ....	106
Tabela 5.10. Resultados do SIFT no experimento com as imagens do conjunto I e os resultados obtidos por Ciratefi na mesma tarefa. ....	108
Tabela 5.11. Resultados do SIFT nos experimentos com as imagens dos conjuntos II e III e os resultados obtidos por Ciratefi nas mesmas tarefas.....	109
Tabela 5.12. Resultados do SIFT no experimento com as imagens do conjunto IV e os resultados obtidos por Ciratefi na mesma tarefa. ....	111
Tabela 5.13. Resumo dos experimentos realizados com SIFT. ....	111
Tabela 5.14. Resumo dos experimentos realizados com EasyMatch. ....	113
Tabela 5.15. Taxa de acertos dos métodos Ciratefi, SIFT e EasyMatch.....	115
Tabela 5.16. <i>Ranking</i> dos métodos com base na taxa de acertos. ....	116
Tabela 5.17. <i>Ranking</i> dos métodos com base no tempo de processamento .....	117

Tabela 5.18. Resultados do color Ciratefi nos experimentos com as imagens do subconjunto leuven variando $\delta$ e $\eta$ e <i>fixando-se</i> os demais parâmetros. ....	120
Tabela 5.19. Resultados do color Ciratefi nos experimentos com todas as imagens da base variando $\delta$ e $\eta$ e <i>fixando-se</i> os demais parâmetros. ....	121
Tabela 5.20. Resultados dos algoritmos color SIFT e color Ciratefi nos experimentos com todas as imagens da base. ....	122
Tabela 5.21. Desempenho do color SITF considerando imagens gradualmente afetadas por borramento e por compressão JPEG. ....	124

## LISTA DE ABREVIATURAS E SIGLAS

ALOI	<i>Amsterdam Library of Object Images</i> (Base de imagens com distorções fotométricas)
ART	<i>Angular Radial Transform</i> (transformada angular-radial). Método para reconhecimento de objetos baseado em região.
C/C++	Linguagens de programação. C é uma linguagem estruturada enquanto C++ é orientada a objetos.
CIE	<i>Commission Internationale d'Eclairage</i> .
CIELAB	Sistema de cores perceptualmente uniforme desenvolvido pela CIE. Também conhecido como $L^*a^*b^*$ .
Cifi	<i>Circular sampling filter</i> (filtro de amostragem circular).
Ciratefi	<i>Circular, Radial and Template-Matching Filter</i> . Método para reconhecimento de padrões invariante sob transformações de similaridade brilho e contraste.
Cis	<i>Circular sampling</i> (amostragem circular).
CisCorr	Correlação da amostragem circular.
CisPS	Provável escala associada a cada pixel candidato da primeira fase.
CMYK	<i>Cyan, Magenta, Yellow and Black</i> (Sistema de cores padrão para impressoras).
Cor	Coeficiente de correlação.
Corr	Coeficiente de correlação invariante a alterações de brilho e contraste.
CSS	<i>Curvature Scale Space</i> (espaço-escala da curvatura). Método para reconhecimento de objetos baseado no contorno.
dE	Distância Euclidiana.
Dev-C++	Ambiente integrado de desenvolvimento de aplicações escritas em C/C++.

dH	Distância de Hausdorff.
dMi	Distância de Minkowsky.
DoG	<i>Difference of Gaussian</i> (diferença de duas imagens idênticas geradas pela convolução gaussiana com diferentes fatores de escala).
EasyMatch	Ferramenta que incorpora um método para reconhecimento de padrões em imagens em níveis de cinzas e coloridas
FAST	<i>Features from Accelerated Segment Test</i> .
FPGA	<i>Field Programmable Gate Array</i> .
GLOH	<i>Gradient Location and Orientation Histogram</i> (Método para detecção de pontos chaves em imagens).
HSI	<i>Hue/Saturation/Intensity</i> (Sistema de cores para emulação da percepção visual humana).
HSV	<i>Hue/Saturation/value</i> (Sistema de cores orientado ao usuário).
IMG	Biblioteca para processamento de imagens e visão computacional de autoria do Prof. Dr. Hae Yong Kim do Depto de Eng. de Sistemas Eletrônicos da Escola Politécnica da Universidade de São Paulo.
ISO	<i>International Organization for Standardization</i> (Organização Internacional para Padronização).
JPEG	<i>Joint Photographic Experts Group</i> (Formato de codificação de imagens).
L*u*v*	Sistema de cores perceptualmente uniforme.
Matlab	<i>MATrix LABoratory</i> (software de alto desempenho para cálculo numérico e operações com matrizes).
MPEG	<i>Moving Picture Experts Group</i> (grupo de trabalho da ISO responsável pelo desenvolvimento de padrões para a compressão de áudio e vídeo digitais).
OpenCV	<i>Open Computer Vision</i> (Biblioteca com rotinas para processamento de imagens e visão computacional, desenvolvida pela Intel).

Open eVision	Biblioteca contendo ferramentas de visão computacional, de propriedade da empresa EureSys.
ProEikon	Biblioteca para processamento de imagens e visão computacional de autoria do Prof. Dr. Hae Yong Kim do Depto de Eng. de Sistemas Eletrônicos da Escola Politécnica da Universidade de São Paulo.
Rafi	<i>Radial sampling filter</i> (filtro de amostragem radial).
Ras	<i>Radial sampling</i> (amostragem radial).
RasAng	Provável ângulo de rotação associado a cada pixel candidato da segunda fase.
RasCorr	Correlação da amostragem radial.
RGB	<i>Red, Green and Blue</i> (Sistema de cores padrão para monitores).
ROC	<i>Receiver Operating Characteristic</i> (Métrica amplamente utilizada na avaliação do desempenho de classificadores).
SIFT	<i>Scale-Invariant Feature Transform</i> . Método que permite fazer correspondência de padrões entre duas imagens com base em pontos chaves.
Sim	Medida de similaridade para avaliar a semelhança perceptual entre dois vetores de cores no espaço CIELAB.
SURF	<i>Speeded Up Robust Features</i> (Método para detecção de pontos chaves em imagens).
SUSAN	<i>Smallest Univalve Segment Assimilating Nucleus</i> . (Algoritmo para detecção de cantos e linhas).
Tefi	<i>Template-matching filter</i> (Filtro de casamento de padrões).
YIQ	Sistema de cores empregado na transmissão de imagens de televisão padrão NTSC ( <i>National Television System Committee</i> ).
YUV	Sistema de cores empregado na transmissão de imagens de televisão padrão PAL ( <i>Phase Alternate Lines</i> ).
YCbCr	Sistema de cores empregado pelos padrões JPEG e MPEG
XYZ	Sistema de cores primárias desenvolvido pela CIE.

## LISTA DE SÍMBOLOS

$f(x,y)$	Função bidimensional que define matematicamente uma imagem
$x'$	Valor da coordenada $x$ após transformação geométrica
$y'$	Valor da coordenada $y$ após transformação geométrica
$d_x$	Fator de translação na direção $x$
$d_y$	Fator de translação na direção $y$
$S_x$	Fator de escalonamento em relação ao eixo $x$
$S_y$	Fator de escalonamento em relação ao eixo $y$
$i, j, k$	Índice de vetores
$N_v$	Número de elementos dos vetores $\mathbf{x}$ e $\mathbf{y}$
$\theta$	Ângulo
$A$	Imagem a ser analisada (na qual se faz a busca do padrão)
$T$	Imagem contendo o padrão ou objeto a ser detectado ( <i>template</i> )
$I$	Imagem em níveis de cinza
$\mathbf{x}$	Vetor de atributos extraídos de $T$
$\mathbf{y}$	Vetor de atributos extraídos de uma sub-imagem de $A$
$\bar{\mathbf{x}}$	Média do vetor $\mathbf{x}$
$\bar{\mathbf{y}}$	Média do vetor $\mathbf{y}$
$\rho$	Parâmetro que modifica a distância de Minkowski
$\beta$	Correção de contraste
$\gamma$	Correção de brilho
$\boldsymbol{\varepsilon}$	Vetor de erros residuais
$\mathbf{1}$	Vetor composto por 1's
$\tilde{\mathbf{x}}$	Vetor $\mathbf{x}$ corrigido pela sua média

$\tilde{\mathbf{y}}$	Vetor $\mathbf{y}$ corrigido pela sua média
$r_{xy}$	Coeficiente de correlação modificado (invariante a alterações de brilho e contraste)
$t_{\beta}$	Limiar de correção de contraste
$t_{\gamma}$	Limiar de correção de brilho
$n$	Número de fatores de escala
$m$	Número de ângulos (Número de linhas radiais)
$\alpha_i$	i-ésimo ângulo
$s_i$	i-ésimo fator de escala
$\sigma$	Desvio padrão do filtro gaussiano
$t_f$	Limiar de correlação empregado pelo método de casamento de padrões pela força bruta
$\pi$	Valor $\pi=3,141592$
$r$	Raio (distância do centro a um ponto qualquer da circunferência)
$P$	Número de pixels ao longo de um círculo de raio $r$
$T_i$	Amostras de $T$ escalonadas
$l$	Número de círculos utilizados na filtragem Cifi
$r_i$	i-ésimo raio
$C_T$	Matriz multi-escala de atributos extraídos do padrão pela filtragem Cifi
$x_0, y_0$	Coordenadas que denotam o pixel central de $T$
$C_A$	Matriz 3D de atributos extraídos da imagem analisada pela filtragem Cifi
$t_1$	Limiar utilizado por Cifi para definir os candidatos da primeira fase
$\alpha$	Ângulo de inclinação da linha radial
$\lambda$	Comprimento da linha radial
$Q$	Número de pixels ao longo de uma linha radial com inclinação $\alpha$ e comprimento $\lambda$

$R_T$	Vetor de atributos extraídos do padrão pela filtragem Rafi
$R_A$	Matriz de atributos extraídos da imagem analisada pela filtragem Rafi
$t_2$	Limiar utilizado por Rafi para definir os candidatos da segunda fase
$t_3$	Limiar utilizado por Tefi para definir os pontos de casamento
$O$	Notação O (empregada em análise de complexidade de algoritmos)
$N$	Número de pixels da imagem analisada $A$
$M$	Número de pixels da imagem $T$
$N_1$	Número de pixels candidatos da primeira fase
$N_2$	Número de pixels candidatos da segunda fase
$ \cdot $	Valor absoluto de um número
$B$	Elemento estruturante
$B_r$	Elemento estruturante circular com raio $r$
$\oplus$	Operação morfológica de dilatação
$\ominus$	Operação morfológica de erosão
$q_1$	Número de pixels candidatos a serem promovidos da primeira para segunda etapa de filtragem
$q_2$	Número de pixels candidatos a serem promovidos da segunda para terceira etapa de filtragem
$q_3$	Número de pixels candidatos da terceira fase que representam posições de casamento
$d_1$	Distância entre os pixels candidatos da primeira fase
$d_2$	Distância entre os pixels candidatos da segunda fase
$d_3$	Distância entre os pixels candidatos da terceira fase
$S_c$	Similaridade de cromaticidade
$S_L$	Similaridade de luminosidade
$\delta$	Peso atribuído à similaridade de cromaticidade
$\eta$	Peso atribuído à similaridade de luminosidade

# SUMÁRIO

## LISTA DE ILUSTRAÇÕES

## LISTA DE TABELAS

## LISTA DE ABREVIATURAS E SIGLAS

## LISTA DE SÍMBOLOS

<b>1 INTRODUÇÃO .....</b>	<b>23</b>
1.1 CONSIDERAÇÕES INICIAIS .....	23
1.2 REVISÃO DA LITERATURA .....	25
1.3 OBJETIVOS .....	31
1.4 MOTIVAÇÃO E JUSTIFICATIVA .....	31
1.5 ABORDAGEM PROPOSTA .....	32
1.6 CONTRIBUIÇÕES ORIGINAIS .....	33
1.7 PUBLICAÇÕES ASSOCIADAS A ESTA TESE .....	34
1.8 ORGANIZAÇÃO DA TESE .....	35
<b>2 RECONHECIMENTO DE PADRÕES EM VISÃO COMPUTACIONAL .....</b>	<b>36</b>
2.1 INTRODUÇÃO .....	36
2.2 IMAGENS DIGITAIS .....	37
2.3 TRANSFORMAÇÕES GEOMÉTRICAS .....	37
2.3.1 <i>Transformação de translação</i> .....	38
2.3.2 <i>Transformação de escala</i> .....	39
2.3.3 <i>Transformação de rotação</i> .....	39
2.4 RECONHECIMENTO DE PADRÕES .....	40
2.5 VISÃO COMPUTACIONAL .....	41
2.6 REPRESENTAÇÃO E DESCRIÇÃO DE OBJETOS EM IMAGENS .....	43
2.6.1 <i>Descritores de contorno</i> .....	44
2.6.2 <i>Descritores de região</i> .....	45
2.6.3 <i>Descritores locais e globais</i> .....	46
2.6.4 <i>Medidas de similaridade empregadas pelos algoritmos de reconhecimento de padrões em imagens digitais</i> .....	46

2.7 AVALIAÇÃO DO DESEMPENHO DE ALGORITMOS DE RECONHECIMENTO DE OBJETOS .....	49
<b>3 METODOS UTILIZADOS EM ANÁLISES COMPARATIVAS .....</b>	<b>51</b>
3.1 SCALE-INVARIANT FEATURE TRANSFORM (SIFT).....	51
3.1.1 <i>Estágios do SIFT</i> .....	52
3.1.2 <i>Correspondência de pontos comuns entre duas imagens</i> .....	53
3.2 EASYMATCH.....	55
<b>4 CASAMENTO DE PADRÕES LIVRE DE SEGMENTAÇÃO E INVARIANTE SOB TRANSFORMAÇÕES DE SIMILARIDADE, BRILHO E CONTRASTE.....</b>	<b>58</b>
4.1 INTRODUÇÃO .....	58
4.2 MEDIDA DE SIMILARIDADE ROBUSTA A VARIAÇÕES DE BRILHO E CONTRASTE .....	59
4.3 CASAMENTO DE PADRÕES PELO MÉTODO DA FORÇA BRUTA INVARIANTE SOB TRANSFORMAÇÕES DE SIMILARIDADE, BRILHO E CONTRASTE.....	61
4.4 CIRCULAR, RADIAL AND TEMPLATE-MATCHING FILTER (CIRATEFI) .....	63
4.4.1 <i>Circular sampling filter (Cifi)</i> .....	63
4.4.2 <i>Radial sampling filter (Rafi)</i> .....	66
4.4.3 <i>Template Matching filter (Tefi)</i> .....	69
4.4.4 <i>Abordagem multi-escala</i> .....	71
4.4.5 <i>Uma breve análise da complexidade computacional do Ciratefi</i> .....	72
4.5 CIRATEFI MORFOLÓGICO .....	74
4.5.1 <i>Definição das operações de dilatação e erosão</i> .....	75
4.5.2 <i>Primeira etapa de filtragem</i> .....	75
4.5.3 <i>Segunda etapa de filtragem</i> .....	76
4.5.4 <i>Terceira etapa de filtragem</i> .....	77
4.6 COLOR CIRATEFI .....	79
4.6.1 <i>Introdução</i> .....	79
4.6.2 <i>Descrição do algoritmo Color Ciratefi</i> .....	83
4.6.3 <i>Espaço de cor e medida de similaridade utilizados</i> .....	84
4.6.3.1 <i>O espaço de cor CIELAB</i> .....	84
4.6.3.2 <i>Medida de similaridade proposta</i> .....	86
<b>5 RESULTADOS EXPERIMENTAIS .....</b>	<b>89</b>

5.1 RESULTADOS EXPERIMENTAIS COM CIRATEFI.....	89
5.1.1 <i>Experimentos com as imagens do conjunto I</i> .....	90
5.1.2 <i>Experimentos com as imagens do conjunto II</i> .....	92
5.1.3 <i>Experimentos com as imagens do conjunto III</i> .....	93
5.1.4 <i>Experimentos com as imagens do conjunto IV</i> .....	95
5.1.5 <i>Análise da sensibilidade do Ciratefi com relação aos parâmetros de controle</i> .....	98
5.1.6 <i>Proposta para otimização do Ciratefi</i> .....	102
5.2 RESULTADOS EXPERIMENTAIS COM CIRATEFI MORFOLÓGICO .....	106
5.3 ANÁLISE COMPARATIVA SIFT X CIRATEFI.....	107
5.4 ANÁLISE COMPARATIVA EASYMATCH X CIRATEFI .....	113
5.5 RESUMO DAS COMPARAÇÕES ENTRE CIRATEFI, SIFT E EASYMATCH.....	115
5.7 RESULTADOS EXPERIMENTAIS COM COLOR CIRATEFI.....	118
<b>6 CONCLUSÕES .....</b>	<b>127</b>
6.1 CONTINUIDADE DO TRABALHO .....	129
<b>REFERÊNCIAS.....</b>	<b>130</b>
<b>APÊNDICE A – PUBLICAÇÕES DO AUTOR DURANTE O PERÍODO DO DOUTORADO .....</b>	<b>142</b>

# 1 INTRODUÇÃO

Neste capítulo faz-se uma exposição do tema e motivação desta tese. Primeiramente, é feita uma revisão da literatura. Em seguida, são apresentados os objetivos, a abordagem proposta, as contribuições do trabalho, as publicações a ele associadas e, finalmente, a estrutura geral da tese.

## 1.1 CONSIDERAÇÕES INICIAIS

O casamento de padrões em imagens consiste na verificação da presença de um determinado padrão ou objeto de referência (*template*) em uma imagem digital. A solução deste problema é fundamental para diversas tarefas em visão computacional como registro de imagens, reconhecimento de objetos, detecção de movimento, análise de imagens médicas, entre outras. O termo padrão pode caracterizar qualquer coisa que possa ser definida em termos quantitativos, inclusive os objetos os quais podem ser caracterizados de acordo com sua forma, cor, dimensão e textura, além de outros atributos. Como na maioria dos casos um padrão representa a descrição simbólica ou numérica de um objeto, é comum o uso dos termos reconhecimento de objetos e detecção de objetos em imagens em substituição ao termo reconhecimento de padrões em imagens.

Um dos principais desafios na área de visão computacional é a criação de sistemas computacionais que simulem as capacidades humanas na realização de funções relacionadas à análise de imagens (HUTCHINSON; HAGER; CORKE, 1996). De acordo com Gonzalez e Woods (2002), um sistema de visão computacional deve ser dotado de algumas características como habilidade para extrair apenas informações importantes a partir de uma cena repleta de informações entre as quais muitas irrelevantes, habilidade para fazer inferências a partir de informações incompletas e, principalmente, conseguir identificar um objeto ou padrão em uma imagem com a maior independência possível em relação a fatores como mudanças de posição, tamanho, orientação e variações no ponto de vista (transformações geométricas). Esta idéia é compartilhada por muitos outros autores entre os quais podemos citar

Loncaric (1998), Zhang e Lu (2004) Long, Zhang e Feng (2005), Choksuriwong, Laurent e Emile (2005) e Veltkamp e Latecki (2006). Além disso, um sistema de visão eficiente deve ser capaz de extrair um conjunto de atributos que descreva com precisão um determinado padrão ou objeto e que seja pequeno o suficiente para reduzir o tempo de processamento e viabilizar a construção de aplicações que possam ser utilizadas na prática tais como sistemas de visão para robôs industriais e veículos autônomos, detecção de eventos em sistemas de vigilância, leitura automatizada de placas de veículos e recuperação de informações em bancos de dados por meio de conteúdo visual.

Um sistema de visão computacional típico envolve as etapas de aquisição de imagens, pré-processamento, segmentação, extração de características e reconhecimento (LONCARIC, 1998; GONZALEZ; WOODS, 2002; LONG; ZHANG; FENG, 2005). Assim, a maioria dos métodos de reconhecimento de padrões em imagens propostos na literatura, primeiramente simplifica as imagens usando operações como binarização, segmentação e detecção de bordas e pontos de contorno, para em seguida extrair o conjunto de atributos que descreve o padrão. Para constatar a existência do padrão procurado na imagem analisada normalmente utiliza-se alguma medida de similaridade para comparar os dois conjuntos de atributos, um extraído do padrão e outro da região de interesse na imagem analisada. O problema encontrado nesses métodos é que as operações para simplificar as imagens podem descartar informações importantes para descrição do padrão, tornando o processo de detecção sujeito a erros.

Uma solução óbvia e precisa para o reconhecimento de objetos sujeitos a transformações geométricas é o método da força bruta (descrito na seção 4.3) que consiste em fazer uma série de operações de casamento de padrões entre a imagem analisada e a imagem do objeto, considerando todas as posições possíveis que o padrão possa aparecer na imagem analisada e todas as transformações que possa estar sujeito dentro de um intervalo estabelecido (TSAI; TSAI, 2002). Claramente, esta solução é inviável já que demanda muito tempo de processamento. Nos experimentos realizados neste trabalho, constatou-se que o método de casamento de padrões proposto, denominado Ciratefi (*Circular, Radial and Template-Matching Filter*), foi capaz de obter resultados similares aos resultados do método de força bruta, com tempo de processamento cerca de 400 vezes menor.

## 1.2 REVISÃO DA LITERATURA

Nos últimos anos, o interesse em aplicações de visão computacional tem sido crescente e os avanços conseguidos têm possibilitado o desenvolvimento de inúmeras aplicações práticas em diferentes áreas. Entretanto, observa-se que, se por um lado tem havido uma ampliação ao acesso a câmeras de vídeo de baixo custo, melhorias na qualidade das imagens e o surgimento de computadores com maior capacidade de processamento que possibilitam abordar problemas mais complexos, por outro lado, exige-se cada vez mais das aplicações relativas ao reconhecimento de padrões em imagens, sendo as principais exigências a invariância sob transformações de similaridade e afim, robustez à oclusão parcial, resistência a alterações de brilho e contraste e insensibilidade a ruídos. O desenvolvimento de técnicas que visam atender a estas exigências representa um grande desafio para a comunidade científica interessada no assunto e muitos esforços têm sido empreendidos em pesquisas nesta área. Como consequência, muitos métodos têm sido reportados na literatura utilizando diferentes abordagens, as quais diferem entre si basicamente pela forma como descreve os objetos (a partir de características extraídas de sua região ou do seu contorno) e pelo subconjunto das exigências que atende. Embora haja várias maneiras de classificar as inúmeras técnicas de descrição de objetos encontradas na literatura, região e contorno normalmente são as abordagens utilizadas para esta finalidade (ZHANG; LU, 2004).

Duas técnicas amplamente conhecidas e utilizadas na extração de características invariantes são os momentos de Hu (1962) e os momentos de Zernike (KIM; KIM, 2000; FLUSSER; SUK, 2006; XIN; PAWLAK; LIAO, 2006). Tanto os momentos de Hu quanto os momentos de Zernike são atributos globais extraídos da região do objeto. Estas técnicas são geralmente empregadas em imagens binárias e invariantes apenas a rotação. Para se obter invariância à escala, a área do objeto deve ser normalizada. Para isso, os momentos de Hu e os momentos de Zernike são obtidos, respectivamente, a partir de momentos centrais normalizados e a partir de um círculo de raio unitário para o qual o objeto é mapeado. Nas últimas décadas, muitas outras técnicas com base nos momentos invariantes de Hu e de Zernike têm sido desenvolvidas e algumas delas podem ser encontradas nos trabalhos de Teh (1988), Khotanzad e Hong (1990), Kim e Yuan (1994), Choi e Kim (2002), Li et al.

(2004), Hse e Newton (2004) e Flusser e Suk (2006). Embora amplamente utilizados, os momentos de Hu apresentam sensibilidade ao ruído e não discriminam bem os objetos simétricos (FLUSSER; SUK, 2006).

Outros dois métodos para descrição de objetos muito utilizados em aplicações de visão computacional são o *Curvature Scale Space (CSS)* e o ART (*Angular Radial Transform*). O CSS, descrito em Mokhtarian e Mackworth (1992), Mokhtarian; Abbasi e Kittler (1996), Zhong e Liao (2004), é baseado no espaço-escala da curvatura do contorno da forma e vem sendo amplamente utilizado em aplicações de visão computacional. Dada a sua robustez, CSS foi adotado pelo MPEG-7 como um dos métodos padrão para descrição de formas (BOBER, 2001; SALEMBIER, 2002; VELTKAMP; LATECKI, 2006). O ART, também adotado pelo MPEG-7, é um descritor baseado em região capaz de descrever formas planas simples e complexas por meio da transformada de uma função complexa (BOBER, 2001; HWANG; KIM, 2006; VELTKAMP; LATECKI, 2006).

Algumas técnicas de reconhecimento de padrões invariantes sob transformação de similaridade encontradas na literatura como as descritas por Goshtasby (1985), Taza e Suen (1989), Torres-Mendez et al. (2000) e Dionisio (2005), foram desenvolvidas com o propósito de detectar padrões binários em imagens binárias.

Torres-Mendez et al. (2000) propuseram um método que utiliza o algoritmo de aprendizagem K-Vizinhos mais próximos, para reconhecimento de objetos a partir de suas características topológicas, invariante a transformações de similaridade.

Goshtasby (1985) desenvolveu um método de reconhecimento de padrões invariante sob transformações de similaridade conhecido por "*Shape Matrix*", no qual a descrição da forma do objeto é feita usando uma matriz, denominada *shape matrix*, cujas coordenadas de linhas e colunas são dadas, respectivamente, por linhas radiais e círculos concêntricos que cobrem o objeto e são originados no seu centro de gravidade. Os valores dos pixels que estão na intersecção das linhas radiais com os círculos concêntricos são utilizados para compor a matriz. Alguns anos depois Taza e Suen (1989) propuseram uma melhoria no método *Shape Matrix* atribuindo pesos aos valores da matriz e levando em conta outras características do objeto além dos valores dos pixels.

O trabalho proposto por Dionisio (2005) aborda dois descritores de formas, um invariante sob transformações de similaridade e outro invariante sob transformações afim. No primeiro, as características são extraídas da curvatura do contorno do objeto enquanto no segundo as características são baseadas em áreas de triângulos formados a partir de três pontos pertencentes ao contorno do objeto. Os descritores desenvolvidos por Dionisio (2005) foram testados no reconhecimento supervisionado de gestos estáticos de mão, mas os resultados obtidos não foram comparados com outros métodos, como por exemplo o CSS (*Curvature Scale Space*).

Normalmente, os métodos para reconhecimento de padrões em imagens binárias funcionam da seguinte maneira: dada uma imagem em níveis de cinza, o primeiro passo é converter esta imagem em uma imagem binária. Os próximos passos são: separar os objetos do fundo da imagem, extrair um conjunto de características invariantes que descreve cada um dos objetos e, finalmente, realizar o processo de reconhecimento. Infelizmente, em muitos casos práticos, as imagens em níveis de cinza não podem ser convertidas em imagens binárias, tornando a aplicação destas técnicas inviável.

Todas as técnicas anteriormente mencionadas, com exceção de CSS e do primeiro descritor apresentado em Dionisio (2005), extraem características a partir da região do objeto. Diversas outras técnicas, entre elas *generalized Hough transform* (BALLARD, 1981), *geometric hashing* (LAM DAN; WOLFSON, 1988; WOLFSON; RIGOUTSOS, 1997), *shape context* (MORI; BELONGIE; MALIK, 2001; BELONGIE; MALIK; PUZICHA, 2002; MORI; BELONGIE; MALIK, 2005) e *SUSAN (Smallest Univalve Segment Assimilating Nucleus)* (SMITH, 1996; SMITH; BRADY, 1997), objetivam a extração de características invariantes a partir do contorno do objeto.

Ballard (1981) propôs a transformada generalizada de Hough (*generalized Hough transform*) para reconhecimento de objetos em imagens em níveis de cinza utilizando informações direcionais extraídas de suas bordas. A desvantagem deste algoritmo é a grande quantidade de memória utilizada e o alto custo computacional exigido. Recentemente, Ulrich, Steger e Baumgartner (2003) apresentaram uma metodologia para melhoria do algoritmo proposto em Ballard (1981), na qual o processo de detecção é iniciado com a utilização de imagens com baixa resolução, sendo repetido até atingir uma resolução aceitável. A transformada de Hough foi inicialmente introduzida por Hough (1962) para detectar padrões complexos de

pontos em imagens binárias. Na última década o algoritmo se tornou muito popular para detecção de formas parametrizáveis como linhas, círculos e elipses.

Lamdan e Wolfson (1988) desenvolveram um método para indexação e reconhecimento de objetos denominado *geometric hashing*, invariante sob transformações afim e capaz de reconhecer objetos parcialmente oclusos. Este método propõe o uso de uma tabela de *hashing* para indexação dos objetos de acordo com características geométricas extraídas de um conjunto de pontos de interesse. Posteriormente, Wolfson e Rigoutsos (1997) agregaram a abordagem Bayesiana ao método *geometric hashing*, para determinar a probabilidade da distribuição dos índices sobre o espaço de invariantes geométricos, visando otimizar o processo de indexação na tabela de *hashing*.

Nos trabalhos de Mori, Belongie e Malik (2001), Belongie, Malik e Puzicha (2002) e Mori, Belongie e Malik (2005), um algoritmo simples e robusto para detecção de correspondências entre dois objetos a partir de características do contorno é apresentado. O processo de extração de características utiliza um conceito denominado *Shape Context*, o qual se baseia na distribuição de um conjunto de pontos do contorno do objeto com relação a um determinado ponto do mesmo contorno. A correspondência entre dois objetos é determinada pela proximidade de seus "*Shape Contexts*". De acordo com os autores, além da invariância a transformações de similaridade, o método é robusto a pequenas mudanças do ponto de vista da câmera e oclusões parciais. Recentemente, Ghosh e Petkov (2005) propuseram um algoritmo de aprendizagem que tem como base o *shape context* e é capaz de reconhecer objetos mesmo quando estes têm seus contornos incompletos, o que trouxe melhorias para robustez do método *shape context*, no que tange o problema da oclusão parcial.

Em Smith e Brady (1997) um detector de bordas e cantos para imagens em níveis de cinza foi proposto. O SUSAN, embora seja limitado a detectar cantos e linhas, tem sido utilizado como extrator de características ou como parte de outros métodos de reconhecimento de padrões em imagens, como por exemplo o método proposto por Mao, Zhang e Wang (2005) para detecção de textos. Além disso, extensões visando melhorias do método também têm sido propostas como é o caso do FAST (*Features from Accelerated Segment Test*), uma eficiente implementação do SUSAN que utiliza árvores de decisão (ROSTEN; DRUMMOND, 2005). Outro algoritmo para

detecção de cantos largamente utilizado em tarefas de visão computacional é o detector de Harris proposto por Harris e Stephens (1988).

Técnicas capazes de detectar objetos em imagens, sem a necessidade de segmentação prévia e que sejam invariantes sob transformações de similaridade e resistentes a alterações de brilho e contraste, como é o caso do método proposto neste trabalho, também podem ser encontradas na literatura (ULLAH; KANEKO, 2004; MARIMON; EBRAHIMI, 2007; TSAI; TSAI, 2002; LOWE, 1999; LOWE, 2004; EURESYS, 2007). Para estas técnicas o desafio é ainda maior dada a dificuldade de estimar a escala de um objeto e extrair características invariantes, sem determinar previamente seu contorno.

Ullah e Kaneko (2004) propuseram um método de casamento de padrões invariante à rotação para imagens em níveis de cinza denominado "*Orientation Code Histograms*". A idéia principal é a utilização de histogramas de direção de gradiente para descrição dos padrões. Tais histogramas permitem que o método seja também invariante a alterações de brilho. Marimon e Ebrahimi (2007) propuseram uma forma diferente para obtenção do histograma de direção de gradiente, que além de melhorar a invariância a rotação, tornou o método mais rápido. Contudo, dois problemas podem ser destacados nestes métodos: um deles é que dois padrões diferentes podem produzir histogramas semelhantes e o outro é que um padrão representado por uma região pequena na imagem tem pouco impacto na formação do histograma.

Tsai e Tsai (2002) apresentaram um método de casamento de padrões invariante à rotação para imagens coloridas, no qual as características que descrevem um padrão são extraídas considerando os espaços de cores HSI e CIELAB. Eles fizeram experimentos com o método considerando a detecção de objetos em imagens adquiridas sob diferentes condições de iluminação. A principal desvantagem deste método é a falta de invariância a escala e a possibilidade de se obter um grande número de falsos positivos.

Dois eficientes métodos que têm sido utilizados em tarefas de casamento de padrões em imagens em níveis de cinza são o SIFT (*Scale-Invariant Feature Transform*) (LOWE, 1999, 2004, 2005) e o EasyMatch (EURESYS, 2007). O SIFT utiliza como base a distribuição do gradiente de pequenas regiões em diferentes escalas para determinar um conjunto de pontos-chave (*keypoints*) que permitem

fazer correspondência de padrões entre duas imagens. Já o EasyMatch, leva em consideração um conjunto de características locais para detecção de padrões em uma imagem. Ambos os métodos não necessitam de segmentação prévia das imagens e são invariantes sob transformações de similaridade, alterações de brilho moderadas, pequenas mudanças do ponto de vista da câmera e oclusão parcial. O sucesso do SIFT tem motivado o desenvolvimento de outros métodos detectores de pontos-chaves como é o caso do PCA-SIFT (*Principal Component Analysis-based SIFT*) proposto por Ke e Sukthankar (2004), o GLOH (*Gradient Location and Orientation Histogram*) desenvolvido por Mikolajczyk e Schmid (2005) e o SURF (*Speeded Up Robust Features*) proposto por Bay, Tuytelaars e Van Gool (2006).

Dentre todos os métodos anteriormente citados, SIFT e EasyMatch foram os escolhidos para realização de análises comparativas com o método proposto (Ciratefi) por se tratar de duas das soluções mais completas e robustas da atualidade e também pela disponibilidade de implementações que facilitaram suas utilizações. Entretanto, cabe salientar que tanto SIFT quanto EasyMatch funciona de maneira diferente do método proposto neste trabalho. A principal diferença é que SIFT e EasyMatch extraem características locais enquanto no método proposto as características são extraídas de forma global.

A literatura existente a respeito dos métodos de casamento de padrões e reconhecimento de objetos envolve inúmeras outras técnicas e métodos. Entre eles, podemos citar aqueles que utilizam a transformada de Fourier para extração de características tais como *Circular Harmonic Expansion* (HSU; AESENAULT; APRIL, 1982), *Normalized Cross-Correlation* (SORGI; DANILIDS, 2004) e *Circular Short Time Fourier Transform* (FORNASIER; TONIOLO, 2005), os descritores que empregam a geometria fractal (TAO; LOERGER; TANG, 2001; TORRES; FALCÃO; COSTA, 2004; YU et al., 2005), os descritores que utilizam operadores da morfologia matemática, os quais são explorados na seção 4.5, onde é apresentado o Ciratefi Morfológico e também os métodos para reconhecimento de padrões em imagens coloridas que são tratados na seção 4.6, na qual é descrito o color Ciratefi.

### 1.3 OBJETIVOS

O objetivo desta tese é contribuir na solução do problema de reconhecimento de padrões em imagens digitais. Para isso, propõe-se um método de casamento de padrões, denominado Ciratefi (*Circular, Radial and Template-Matching Filter*), que é capaz de detectar padrões ou objetos sem a necessidade de simplificação das imagens (segmentação/binarização). Além disso, Ciratefi é invariante sob transformações de similaridade (rotação, escalonamento uniforme e translação), brilho e contraste. Também foram propostas duas extensões do Ciratefi, uma que utiliza operadores morfológicos na extração dos atributos descritores, denominada Ciratefi Morfológico e outra para imagens coloridas que chamamos de color Ciratefi.

Alguns objetivos específicos do presente trabalho são:

- Descrever as principais abordagens empregadas no reconhecimento de padrões em imagens;
- Propor descritores invariantes sob transformação de similaridade, brilho e contraste;
- Avaliar a sensibilidade do Ciratefi com relação à escolha dos parâmetros que regulam seu funcionamento;
- Desenvolver as extensões do Ciratefi (Ciratefi Morfológico e color Ciratefi);
- Testar os métodos propostos em imagens contendo diferentes tipos de padrões, texturas, fundos, condições de iluminação, etc.;
- Comparar os resultados obtidos pelo Ciratefi e suas extensões com os resultados obtidos por alguns dos principais métodos encontrados na literatura;

### 1.4 MOTIVAÇÃO E JUSTIFICATIVA

As pesquisas em visão computacional têm sido motivadas, principalmente, pelos estudos acerca da cognição visual humana. Com base nestes estudos, diversos métodos têm sido propostos na literatura com o intuito de propiciar ao computador a

capacidade de emular a visão humana e tomar decisões baseadas em entradas visuais. Um dos principais obstáculos para criação de sistemas de visão computacional é que estes deveriam ser dotados de um conjunto de habilidades que são inerentes à inteligência humana para realização de tarefas relativas à interpretação do conteúdo de imagens. Entre estas habilidades podemos citar a capacidade de extrair informações importantes descartando tudo aquilo que é irrelevante, a capacidade de fazer inferências a partir de informações incompletas e a capacidade de aprender e generalizar o conhecimento para ser aplicado em situações novas e diferentes daquelas que foram aprendidas. Estas exigências fazem com que as aplicações no campo da visão computacional sejam projetadas e implementadas para ambientes operacionais limitados.

Diante deste cenário, os esforços na área de visão computacional, mais precisamente voltados para os métodos para reconhecimento padrões em imagens digitais, são importantes porque fortalecem e encorajam o desenvolvimento de soluções para problemas práticos como sistemas de visão para robôs e veículos autônomos, detecção de eventos em sistemas de vigilância e sistemas de inspeção automatizados. Além disso, o método proposto neste trabalho torna-se interessante porque: a) demonstrou ser robusto; b) funciona de maneira diferente da maioria dos métodos encontrados na literatura que utilizam operações para simplificar as imagens e por consequência descartam informações importantes para descrição dos objetos, tornando a detecção sujeita a erros; c) pode ser facilmente adaptado para tratar de imagens coloridas, o que aumenta a sua completude; d) dada a sua simplicidade, pode ter sua velocidade acelerada em milhares de vezes via implementação em hardware, viabilizando o desenvolvimento de sistemas de visão computacional de alto desempenho e de baixo custo.

## 1.5 ABORDAGEM PROPOSTA

O trabalho descrito nesta tese situa-se na área de visão computacional, na sub-área denominada processamento de nível médio ou intermediário, a qual é descrita na seção 2.5. Especificamente, tratamos da representação, descrição e reconhecimento de padrões em imagens digitais em níveis de cinzas e coloridas.

A abordagem proposta baseia-se em métodos restritos ao domínio espacial para realização de tarefas relacionadas à interpretação de imagens com base em características globais extraídas da imagem de referência e da imagem analisada. Para isso não há a necessidade segmentação prévia das imagens e as instâncias do padrão procurado podem estar sujeitas a transformações de similaridade (translação, rotação e escala uniforme), brilho e contraste. No caso do Ciratefi para imagens coloridas (color Ciratefi), levou-se em conta também o problema da constância da cor.

Os algoritmos propostos neste trabalho foram implementados em linguagem C/C++, utilizando os compiladores Dev-C++ e as bibliotecas para Processamento de Imagens e Visão Computacional IMG (KIM, 2004) e ProEikon (KIM, 2006). Foi utilizada também a biblioteca OpenCV (INTEL, 2000) para compilação de alguns algoritmos pertinentes aos métodos utilizados nos estudos comparativos.

## 1.6 CONTRIBUIÇÕES ORIGINAIS

Esta tese contém as seguintes contribuições principais:

- Foi proposta uma maneira de extração de atributos invariantes a rotação e translação para identificar padrões em imagens (KIM; ARAUJO, 2007);
- Foi proposta uma maneira de tornar o conjunto de atributos invariante a mudança de escala uniforme, tornando-o assim invariante a transformações de similaridade (rotação, escala uniforme e translação) (KIM; ARAUJO, 2007);
- Propusemos uma medida de similaridade, baseada no coeficiente de correlação, para avaliar a semelhança entre vetores de atributos extraídos de imagens em níveis de cinzas, robusta a variações de brilho e contraste (KIM; ARAUJO, 2007);
- Foi proposto um método de casamento de padrões para imagens em níveis de cinzas que não requer segmentação prévia, denominado Ciratefi (KIM; ARAUJO, 2007);

- Foi proposta uma extensão do Ciratefi, também para imagens em níveis de cinzas, baseada em descritores morfológicos, na qual o conjunto de atributos é obtido por meio de operações de dilatação e erosão (ARAUJO; KIM, 2007);
- Foi proposta uma extensão do Ciratefi para imagens coloridas levando-se em conta o problema da constância da cor. Também propusemos uma medida de similaridade para avaliar a semelhança perceptual entre vetores de atributos extraídos de imagens coloridas (artigo a ser submetido);
- Foram realizados vários experimentos comparando os métodos propostos com dois dos principais métodos encontrados na literatura (SIFT e EasyMatch).

## 1.7 PUBLICAÇÕES ASSOCIADAS A ESTA TESE

A esta tese estão associadas as seguintes publicações:

- No artigo (KIM; ARAÚJO, 2007) foi proposto um método de reconhecimento de padrões em imagens em níveis de cinzas invariante sob transformações de similaridade, brilho e contraste denominado Ciratefi. Ainda neste trabalho, foi avaliada a sensibilidade do método com relação aos parâmetros que regulam seu funcionamento. Os resultados experimentais mostraram a robustez do método no reconhecimento de objetos sujeitos a transformações de similaridade;
- No artigo (ARAUJO; KIM, 2007) é apresentada uma extensão do Ciratefi que extrai características dos padrões por meio do uso de operadores morfológicos.

Vale ressaltar que o artigo (KIM; ARAÚJO, 2007) recebeu o prêmio de melhor artigo em análise de imagens no Pacific-Rim Symposium on Image and Video Technology (PSIVT), 2007.

## 1.8 ORGANIZAÇÃO DA TESE

Esta tese está organizada da seguinte maneira:

- No capítulo 2 são apresentadas algumas definições, conceitos e informações básicas sobre imagem digital, transformações geométricas e de similaridade, reconhecimento de padrões, visão computacional e descritores de objetos. Em seguida, apresentam-se algumas das principais medidas de similaridade e por fim é feita uma breve explanação sobre métricas comumente utilizadas na avaliação de desempenho de algoritmos de reconhecimento de objetos em imagens;
- No capítulo 3 é feita uma breve descrição dos métodos *Scale-Invariant Feature Transform* (SIFT) e EasyMatch, os quais são utilizados em análises comparativas com os métodos propostos.
- No capítulo 4 descreve-se a técnica de casamento de padrões pela força bruta. Em seguida, descreve-se o método Ciratefi, explicando detalhadamente sua formulação, sua estrutura e seu funcionamento. Na seqüência, apresenta-se o Ciratefi morfológico e por fim descreve-se o color Ciratefi. Também são descritas neste capítulo as medidas de similaridade empregadas pelos métodos propostos.
- No capítulo 5 são apresentados os resultados experimentais obtidos com Ciratefi, Ciratefi Morfológico e color Ciratefi, bem como as análises comparativas com os métodos SIFT e EasyMatch. Também é feita uma análise da sensibilidade do Ciratefi com relação à escolha dos parâmetros que regulam seu funcionamento e por fim é apresentada uma proposta para sua otimização;
- No capítulo 6 são apresentadas as conclusões e algumas sugestões para a continuidade do trabalho.

## 2 RECONHECIMENTO DE PADRÕES EM VISÃO COMPUTACIONAL

Neste capítulo, são definidos alguns termos importantes para a compreensão deste trabalho como imagem digital, transformações geométricas e de similaridade, reconhecimento de padrões, visão computacional, descritores de objetos, entre outros. Em seguida, são tratadas as duas abordagens para descrição de objetos (região e contorno), algumas das principais medidas de similaridade empregadas pelos algoritmos de reconhecimento de objetos e por fim são apresentadas algumas métricas comumente utilizadas na avaliação de desempenho de tais algoritmos.

### 2.1 INTRODUÇÃO

Embora haja muitas pesquisas em visão computacional com intuito de simular os mecanismos envolvidos no processo de visão humana e de alguns animais, sabe-se que há muita coisa a ser desenvolvida ou aperfeiçoada, já que alguns aspectos relativos ao funcionamento da visão biológica ainda não são totalmente compreendidos (ROORDA, 2002; FULTON, 2006).

A habilidade humana de processar e interpretar conteúdos visuais motiva o desenvolvimento de técnicas e dispositivos com intuito de estender estas habilidades para máquinas (HUTCHINSON; HAGER; CORKE, 1996; LONCARIC, 1998; GONZALEZ; WOODS, 2002). Entretanto, a necessidade de hardwares mais específicos que sejam capazes de fornecer respostas em tempo real, aliada a um conjunto de habilidades que um sistema de visão computacional deveria possuir para solucionar problemas encontrados no dia-a-dia, por exemplo a automatização de processos de controle de qualidade, identificação e classificação de produtos, e exploração de ambientes diversos, torna as tarefas relativas à visão computacional mais difíceis, deixando uma lacuna para novos estudos. Atualmente o que existe em visão computacional são algoritmos para resolver problemas bem definidos. Estes algoritmos em geral são bastante especializados e raramente podem ser generalizados para outras aplicações. Diante disso, o desenvolvimento de métodos baseados em aprendizagem está se tornando cada vez mais comum.

## 2.2 IMAGENS DIGITAIS

Uma imagem pode ser definida como a representação visual de um objeto. Do ponto de vista matemático, uma imagem é definida como uma função bidimensional  $f(x,y)$ , com  $x, y \in \mathbb{R}^2$ , onde  $x$  e  $y$  são as coordenadas espaciais que identificam um ponto e o valor  $f$  em qualquer ponto  $(x, y)$  denota a intensidade do nível de cinza da imagem naquele ponto. Uma imagem digital é uma imagem  $f(x,y)$  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; CONCI; AZEVEDO; LETA, 2008).

Em imagens binárias os valores dos pixels podem assumir os valores 0 e 1. Para imagens em níveis de cinzas normalmente os valores dos pixels variam de 0 a 255 e, no caso de imagens coloridas, cada pixel é representado por um conjunto de três ou quatro valores dependendo do sistema de cores empregado. Na maioria dos casos, utilizam-se três valores (tri-estímulos) a partir dos quais é possível identificar uma cor em um sistema de coordenadas que representa o espaço de cores

## 2.3 TRANSFORMAÇÕES GEOMÉTRICAS

Transformações geométricas, no contexto de processamento de imagens digitais, podem ser definidas como operações que modificam a posição dos pixels no espaço da imagem, alterando algumas características dos objetos gráficos como posição, orientação, tamanho ou forma. Todas as transformações geométricas podem ser representadas por meio de equações algébricas ou matriciais. Esta segunda forma de representação é extensamente utilizada já que as matrizes são mais fáceis de usar e entender (CONCI; AZEVEDO; LETA, 2008).

Algumas das principais transformações geométricas são: translação, rotação, escalonamento, cisalhamento e reflexão. De acordo com Dionísio (2005), estas

transformações podem ser divididas em três grupos: Euclidiano, Similaridade e Afim, como mostrado na Tabela 2.1.

Tabela 2.1. Agrupamento das transformações geométricas

Transformação Geométrica	Grupo		
	Euclidiano	Similaridade	Afim
Translação	✓	✓	✓
Rotação	✓	✓	✓
Escala uniforme	✗	✓	✓
Escala não uniforme	✗	✗	✓
Cisalhamento	✗	✗	✓
Reflexão	✗	✗	✓

As transformações euclidianas preservam distâncias e ângulos entre os pontos de um objeto e também as linhas paralelas. As transformações de similaridade (translação, rotação e escala uniforme), sob as quais *Ciratefi* é invariante e que são descritas a seguir, preservam ângulos e linhas paralelas. Já as transformações do grupo afim preservam as linhas paralelas.

### 2.3.1 Transformação de translação

Transladar um ponto significa deslocá-lo adicionando quantidades às suas coordenadas. Dessa forma, para movimentar o ponto  $(x, y)$   $d_x$  unidades em relação ao eixo  $x$  e  $d_y$  unidades em relação ao eixo  $y$ , basta fazer  $x' = x + d_x$  e  $y' = y + d_y$ .

O modelo de translação no plano é representado pela seguinte equação matricial:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d_x \\ d_y \end{bmatrix} \quad (2.1)$$

### 2.3.2 Transformação de escala

Escalonar um objeto significa alterar as dimensões de escala deste objeto. Para isso, multiplica-se as coordenadas de cada ponto que representa o objeto por fatores de escala, ou seja:  $x' = x \cdot S_x$  e  $y' = y \cdot S_y$  onde  $S_x$  e  $S_y$  são os fatores de escala. O modelo de escalonamento no plano é representado pela equação matricial mostrada a seguir.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.2)$$

No escalonamento, quando  $S_x = S_y$  tem-se escala uniforme ou isotrópica. Naturalmente, quando  $S_x \neq S_y$  diz-se que o escalonamento é não uniforme. Se  $S_* = 1$  não ocorre alteração na coordenada; Se  $S_* > 1$ , então há uma ampliação de coordenada, e se  $0 < S_* < 1$ , então ocorre uma compressão de coordenada.

### 2.3.3 Transformação de rotação

Rotacionar um ponto significa girá-lo considerando um ângulo  $\theta$ . Assim, dado um ponto  $(x, y)$  e um ângulo  $\theta$ , para girar este ponto em torno da origem basta fazer  $x' = x \cdot \cos(\theta) - y \cdot \sin(\theta)$  e  $y' = x \cdot \sin(\theta) + y \cdot \cos(\theta)$ . A rotação 2D na forma matricial é obtida pela seguinte equação:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.3)$$

Considerando uma imagem digital, se  $\theta > 0$  a rotação é feita no sentido horário e se  $\theta < 0$  no sentido anti-horário.

## 2.4 RECONHECIMENTO DE PADRÕES

O reconhecimento de padrões é uma tarefa que os seres humanos e os animais realizam com eficiência desde os primeiros dias de vida (RUSSEL; NORVIG, 1995). Um padrão pode ser visto como uma entidade a partir da qual é possível extrair algum tipo de característica, seja ela simbólica ou numérica. Um mecanismo de reconhecimento de padrões tem como objetivo identificar um padrão a partir das suas características. A criação de mecanismos automáticos que imita habilidades dos seres humanos e dos animais em tarefas de reconhecimento de padrões é ainda um sonho e, ao mesmo tempo, uma motivação para realização de intensas pesquisas, as quais têm crescido proporcionalmente ao aumento das exigências associadas às aplicações relacionadas ao tema. Diversas áreas do conhecimento tais como biologia, psicologia, medicina, engenharia, inteligência artificial e visão computacional possuem problemas relacionados ao reconhecimento de padrões (JAIN; DUIN; MAO, 2000). A importância do reconhecimento de padrões em visão computacional está relacionada com a idéia de simulação de processamentos complexos de percepção visual em máquinas (CONCI; AZEVEDO; LETA, 2008).

O desenvolvimento de aplicações como reconhecimento de comandos de voz, identificação de pessoas por meio de suas digitais ou face, análise de seqüência de DNA (*Deoxyribonucleic Acid*), previsão de comportamentos em mercados financeiros e reconhecimento de caracteres impressos e manuscritos têm sido possível graças ao avanço e disponibilidade de recursos computacionais.

Um sistema de reconhecimento de padrões normalmente envolve algumas etapas como: extração de características, elaboração de um conjunto descritor do padrão e um classificador para decidir se um padrão pertence a uma determinada classe. Algumas abordagens empregadas em reconhecimento de padrões são: casamento de padrões, estatística, sintática e redes neurais (JAIN; DUIN; MAO, 2000). A escolha de uma abordagem não é uma tarefa fácil, sendo muitas vezes, dependente do domínio do problema.

O casamento de padrões é uma das primeiras e mais simples abordagens para reconhecer padrões e consiste em determinar a similaridade entre duas entidades do mesmo tipo. Em visão computacional, o casamento de padrões envolve duas

imagens: a imagem está sendo analisada e a imagem que contém o padrão de referência (*template*). Deste ponto em diante chamaremos estas imagens de  $A$  e  $T$ , respectivamente. O processo consiste em escolher um ou mais pontos de interesse em  $A$ , extrair sub-imagens em torno destes pontos e compará-las com  $T$  de acordo com alguma medida de similaridade. Um exemplo deste processo é ilustrado na Figura 2.1.

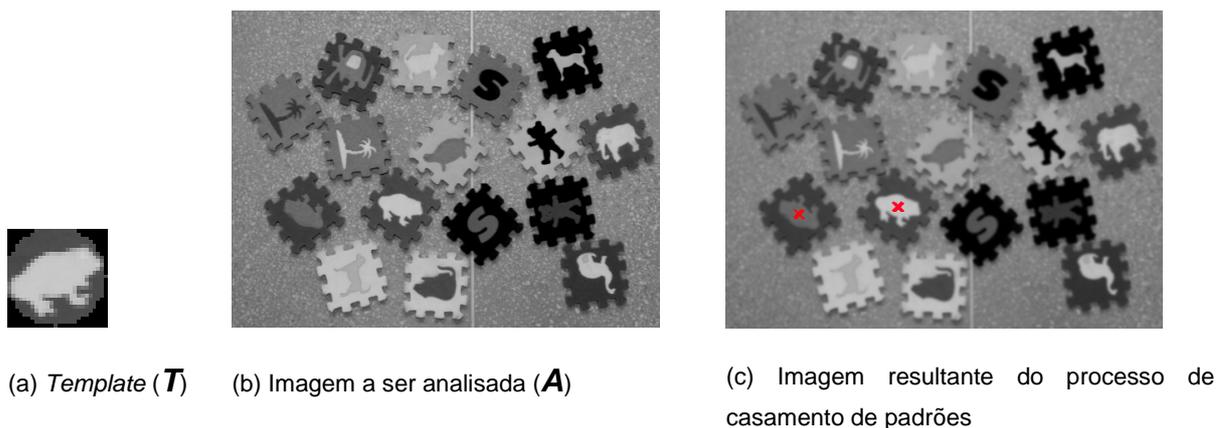


Figura 2.1 – Exemplo de casamento de padrões em imagens.

O exemplo ilustrado na Figura 2.1 mostra casamentos de padrões onde instâncias do padrão procurado aparecem na imagem analisada com diferentes rotações e escalas e também com contraste invertido.

## 2.5 VISÃO COMPUTACIONAL

Visão computacional pode ser definida como a sub-área do processamento de imagens que estuda o desenvolvimento de métodos e técnicas que possibilitam um sistema computacional interpretar imagens (CONCI; AZEVEDO; LETA, 2008). Na área de visão computacional, os métodos de reconhecimento de padrões têm como objetivo detectar padrões em imagens digitais a partir de exemplos fornecidos, conhecidos na literatura como modelos, máscaras ou *templates*. É comum, além da detecção, o método prover a classificação dos padrões em um número de classes ou categorias. Em resumo, um sistema de visão computacional deve dotar uma máquina com algumas capacidades do sistema visual, como a habilidade para descrever uma cena contida em uma imagem digital, a partir do reconhecimento dos

padrões nela contidos, independente de cor, posição, tamanho, orientação ou variações no ponto de vista. Além disso, o conjunto de atributos que descreve um padrão deve ser adequado para viabilizar a construção de aplicações práticas.

Um sistema de visão computacional típico (Figura 2.2) envolve as seguintes etapas: aquisição de imagens, pré-processamento, segmentação, representação e descrição (extração de características) e reconhecimento (LONCARIC, 1998; GONZALEZ; WOODS, 2002; LONG; ZHANG; FENG, 2005) e pode ser classificado, quanto ao grau de abstração, em três diferentes níveis: baixo, médio, alto. À medida que se aumenta o nível de abstração, ocorre uma redução progressiva da quantidade de informações manipuladas. No baixo nível um conjunto de dados composto por valores associados aos pixels da imagem original (por exemplo, intensidade de brilho) é obtido. Ao baixo nível estão associadas as etapas de aquisição e pré-processamento. No nível médio, ao qual estão associadas as etapas de segmentação, representação e descrição e reconhecimento, o conjunto de valores dos pixels é convertido em conjuntos de características que permitem descrever os objetos contidos na imagem. No processamento de alto nível é feita a interpretação do conteúdo da imagem a partir dos conjuntos de características.

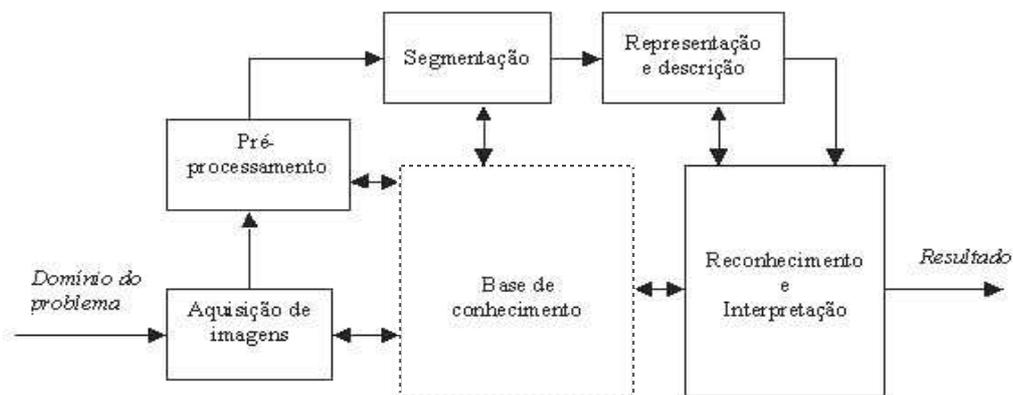


Figura 2.2 – Etapas de um sistema de visão computacional típico.

Extraído de (GONZALEZ, 2002).

Na primeira etapa do sistema de visão, adquire-se a imagem digital. Em seguida, é feito um pré-processamento, ou seja, uma melhoria da imagem visando aumentar as chances de sucesso das etapas seguintes. O pré-processamento normalmente envolve algumas operações como melhoria de contraste, remoção de ruídos e

correção de foco. A próxima etapa trata da segmentação que tem como finalidade isolar o conjunto de objetos ou padrões de interesse. Cabe ressaltar que processos de segmentação automática estão entre as tarefas mais difíceis em processamento de imagens e estão diretamente relacionados à garantia da solução do problema. Daí a importância de métodos capazes de reconhecer padrões em imagens sem a necessidade de segmentação prévia. Na etapa de representação e descrição, procede-se a extração de atributos que caracterizam os padrões. No caso do padrão representar a descrição de um objeto, a extração pode ser feita a partir da região ou do contorno do objeto. Para ambos os casos, o conjunto de atributos extraídos deve ser robusto a variações de tamanho, orientação e translação do objeto (GONZALEZ; WOODS, 2002). Ademais, em algumas aplicações é interessante combinar os descritores de contorno e região. A última etapa trata do reconhecimento e interpretação. O reconhecimento pode ser visto como o processo que atribui um rótulo a um padrão, com base no conjunto de atributos que o descreve enquanto a interpretação está relacionada com a atribuição de significado a um conjunto de padrões reconhecidos. A base de conhecimento é composta por informações relativas ao domínio do problema e sua finalidade é facilitar a execução das etapas.

Uma vez que o método proposto neste trabalho é livre de segmentação, exploramos com mais detalhes apenas as técnicas de representação e descrição de objetos e as principais medidas de similaridade utilizadas no processo de reconhecimento.

## 2.6 REPRESENTAÇÃO E DESCRIÇÃO DE OBJETOS EM IMAGENS

Os objetos podem ser descritos a partir de características como forma, cor, dimensão e textura. A forma de um objeto é uma das características visuais mais importantes para descrevê-lo e por isso é muito utilizada na interpretação do conteúdo de imagens digitais (ZHANG; LU, 2004; DIONISIO, 2005). É também o atributo que traz a maior dificuldade, principalmente pela necessidade de se conhecer o tamanho dos objetos contidos na imagem. Conforme mencionado, o reconhecimento de objetos é um problema clássico da área de visão computacional e tem como objetivo detectar e categorizar objetos em imagens digitais. Para que um objeto contido em uma imagem possa ser reconhecido, normalmente as etapas

ilustradas na Figura 2.2 precisam ser executadas. Os descritores de objetos podem ser classificados de acordo com diversos critérios (LONCARIC, 1998). Contudo, a maioria dos autores considera duas abordagens básicas: descritores de contorno e descritores de região (ZHANG; LU, 2004).

### 2.6.1 Descritores de contorno

Descritores de contorno descrevem um objeto extraíndo-se características do seu contorno, entre as quais perímetro, circularidade, convexidade, excentricidade e corda máxima (maior distância entre dois pontos do contorno fechado). Para isso, métodos de detecção de bordas como o filtro passa-alta são necessários. As bordas podem ser caracterizadas por uma mudança brusca do nível de cinza de duas regiões. Tais descritores são adequados quando o interesse está concentrado nas características próprias do contorno tais como cantos ou pontos de inflexão. Alguns dos principais descritores desta categoria encontrados na literatura são:

- Descritores baseados em transformadas: descrevem um objeto usando um conjunto de coeficientes calculados por uma transformada, por exemplo Fourier ou *Wavelets*. Alguns métodos que utilizam transformadas de Fourier podem ser encontrados em (HSU; AESENAULT; APRIL, 1982; SORGI; DANILIDS, 2004; FORNASIER; TONIOLO, 2005).
- Descritores baseados em contornos normalizados: descrevem o objeto com base no seu contorno normalizado. A normalização dos contornos em um conjunto de objetos que estão sendo analisados é feita para garantir que todos os contornos normalizados tenham a mesma quantidade de pontos equidistantes.
- Descritores baseados em ângulos de curvatura: descrevem um objeto por meio de um conjunto de ângulos da curvatura do seu contorno.
- Descritores baseados em pontos de inflexão: descrevem um objeto com base em um conjunto de pontos de inflexão extraídos do contorno do objeto. Um bom exemplo deste tipo de descritor é o *Curvature Scale Space* (CSS).

- Descritores baseados em saliências do contorno: descrevem um objeto com base nos pontos de saliência do contorno, os quais estão associados às extremidades do esqueleto de um objeto. O descritor baseado em saliências, proposto por Torres e Falcão (2007), combina as regiões de influência e a localização pontos de mais alta curvatura no contorno.

### 2.6.2 Descritores de região

Descritores de região descrevem um objeto simples ou complexo a partir do conjunto de pixels que compõem sua região, ou seja, que estão na parte interna do seu contorno. Nos descritores de região, normalmente o interesse se concentra em propriedades internas da forma tais como textura, área, perímetro, centro de gravidade, valores de máximo e mínimo e média dos níveis de cinzas dos pixels (CONCI; AZEVEDO; LETA, 2008). Alguns dos principais descritores de região encontrados na literatura são:

- Descritores baseados em momentos: descrevem um objeto utilizando valores obtidos por meio do cálculo de momentos estatísticos e constituem a grande maioria dos descritores baseados em momentos. Os mais conhecidos são os momentos de Hu e de Zernike (HU, 1962; FLUSSER; SUK, 2006).
- Descritores geométricos: descrevem um objeto por meio de propriedades geométricas extraídas de sua região como área, perímetro, compactação e projeções de altura e largura. Um exemplo é o descritor proposto por Dionisio (2005).
- Descritores baseados em transformadas: descrevem um objeto usando um conjunto de coeficientes calculados por uma transformada a partir de sua região. Alguns exemplos de transformadas são: Fourier, *Wavelets* e angular-radial (BOBER, 2001; HWANG; KIM, 2006; VELTKAMP; LATECKI, 2006).

### 2.6.3 Descritores locais e globais

Os descritores locais consideram apenas os pontos de interesse na imagem e a partir de cada um destes pontos é extraído um vetor de atributos. Problemas normalmente associados aos descritores locais são a sensibilidade à presença de ruídos e a quantidade de informações para descrição de um objeto (DIONISIO, 2005). Apesar disto, este tipo de descritor favorece a construção de métodos robustos à oclusão parcial e livres de segmentação (TUYTELAARS e MIKOLAJCZYK, 2008). Alguns exemplos de descritores locais são os invariantes geométricos (LAMDAN; WOLFSON, 1988; WOLFSON; RIGOUTSOS, 1997), cantos (SMITH, 1996; SMITH e BRADY, 1997) e curvatura do contorno da forma (MOKHTARIAN; MACKWORTH, 1992; MOKHTARIAN; ABBASI; KITTLER, 1996). Os descritores globais levam em consideração toda a região ou todo o contorno do objeto para extração do conjunto de atributos. Se por um lado os descritores globais têm a vantagem de ser mais tolerantes à presença de ruídos e normalmente gerar conjuntos de atributos menores, por outro lado, são incapazes de diferenciar pequenas alterações na imagem. Alguns exemplos de descritores globais são os momentos invariantes descritos em HU (1962) e Flusser e Suk (2006), *shape matrix* (GOSHTASBY, 1985; TAZA; SUEN, 1989) e a transformada angular-radial (BOBER, 2001; HWANG; KIM, 2006; VELTKAMP; LATECKI, 2006).

### 2.6.4 Medidas de similaridade empregadas pelos algoritmos de reconhecimento de padrões em imagens digitais

O processo de reconhecimento de padrões em imagens digitais, representado na Figura 2.3, normalmente é realizado da seguinte maneira: um mecanismo descritor faz extração de um conjunto atributos da imagem que contém o objeto ou padrão de referência ( $T$ ) e um ou mais conjuntos de atributos da imagem que está sendo analisada ( $A$ ). A detecção do padrão é feita verificando-se a semelhança entre o vetor de atributos extraídos de  $T$  e os vetores extraídos de  $A$ . O objetivo de uma medida de similaridade é exprimir esta semelhança. Se o grau de similaridade entre

dois vetores for acima de um limiar definido pelo usuário, diz-se que há um casamento entre  $T$  e a sub-imagem de  $A$ , a partir da qual foi extraído o vetor de atributos. Daí o termo casamento de padrão ou *template matching*.

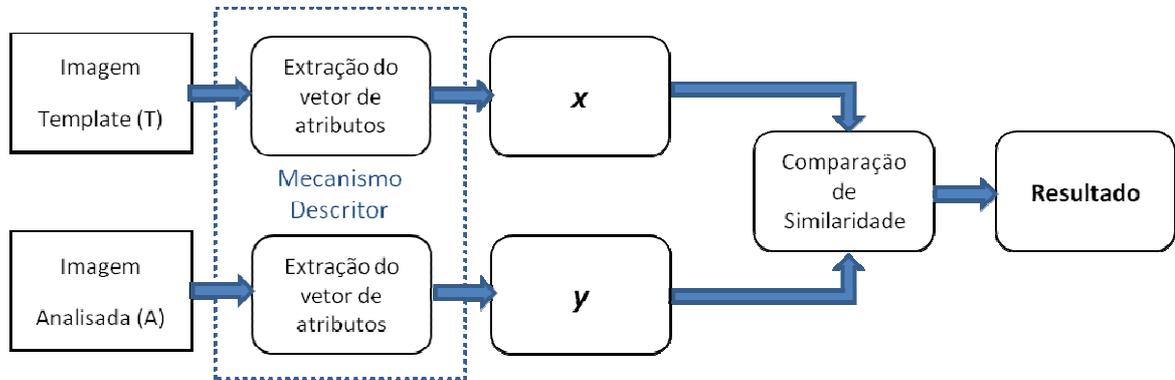


Figura 2.3 – Processo de reconhecimento de padrões em imagens digitais.

Nas últimas décadas, muitas medidas de similaridade têm sido desenvolvidas (LONG; ZHANG; FENG, 2005), sendo as mais utilizadas a distância Euclidiana (dE), distância de Minkowsky (dMi), distância de Hausdorff (dH) e correlação (Cor). Considerando  $\mathbf{x}$  o vetor de atributos extraídos de  $T$  e  $\mathbf{y}$  o vetor de atributos extraídos de uma sub-imagem de  $A$ , então estas medidas de similaridade são dadas por:

Distância Euclidiana:

$$dE(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{N_v} (x_i - y_i)^2} \quad (2.4)$$

onde  $N_v$  é o número de elementos dos vetores  $\mathbf{x}$  e  $\mathbf{y}$  (LONG; ZHANG; FENG, 2005).

Distância de Minkowsky:

$$dMi(\mathbf{x}, \mathbf{y}) = \sqrt[\rho]{\sum_{i=1}^{N_v} |x_i - y_i|^\rho} \quad (2.5)$$

onde  $\rho$  é um parâmetro que modifica a distância de Minkowski resultando em outras distâncias conhecidas. Por exemplo, quando  $\rho=1$  tem-se a distância

city-block ou distância de Manhattan e quando  $\rho=2$  tem-se a distância Euclidiana (LONG; ZHANG; FENG, 2005).

Distância de Hausdorff:

$$dH(\mathbf{x}, \mathbf{y}) = \max\{h(\mathbf{x}, \mathbf{y}), h(\mathbf{y}, \mathbf{x})\} \quad (2.6)$$

onde  $h(\mathbf{x}, \mathbf{y}) = \max_{a \in \mathbf{x}} \min_{b \in \mathbf{y}} d(a, b)$  e  $d(a, b)$  é a distância entre os elementos dos vetores  $\mathbf{x}$  e  $\mathbf{y}$  (ZHANG; LU, 2004). Normalmente se utiliza a distância Euclidiana. A função  $h(\mathbf{x}, \mathbf{y})$  calcula, para cada elemento do vetor  $\mathbf{x}$  (representado por  $a$ ), a sua distância para cada elemento do vetor  $\mathbf{y}$  (representado por  $b$ ), armazenando a menor distância. Finalizado o cálculo de todas as distâncias, a maior entre todas elas é retornada pela função.

Correlação:

$$Cor(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{N_v} (x_i - \bar{\mathbf{x}})(y_i - \bar{\mathbf{y}})}{\sqrt{\sum_{i=1}^{N_v} (x_i - \bar{\mathbf{x}})^2} \cdot \sqrt{\sum_{i=1}^{N_v} (y_i - \bar{\mathbf{y}})^2}} \quad (2.7)$$

$$\text{onde } \bar{\mathbf{x}} = \frac{1}{N_v} \sum_{i=1}^{N_v} x_i \text{ e } \bar{\mathbf{y}} = \frac{1}{N_v} \sum_{i=1}^{N_v} y_i$$

Devido à simplicidade do cálculo, as distâncias Euclidiana e de Minkowsky têm sido amplamente utilizadas no contexto de reconhecimento de padrões em imagens (LONG; ZHANG; FENG, 2005). Entretanto, elas são sensíveis à mudança de escala dos valores que compõem os vetores de atributos. Isto significa que a distância verificada entre dois vetores de atributos, por exemplo  $\mathbf{x}=\{1,2,3,4,5\}$  e  $\mathbf{y}=\{10,20,30,40,50\}$ , representando duas imagens semelhantes mas com diferentes intensidades de brilho sinalizará que se trata de duas imagens distintas quando na verdade não são. Uma maneira de contornar este problema é fazer a normalização dos vetores efetuando-se, para cada elemento do vetor, a subtração da média e a divisão do resultado pelo desvio-padrão. Este problema não acontece com o coeficiente de correlação uma vez que, embora em escalas diferentes, os vetores

estão correlacionados. Além disso, o coeficiente de correlação permite o casamento de instâncias do objeto com contraste invertido como ilustrado na Figura 2.1.

A distância de Hausdorff tem sido empregada como um método de casamento de padrões baseado em correspondência e também como medida de similaridade entre padrões (LONG; ZHANG; FENG, 2005; VELTKAMP; LATECKI, 2006). Uma das vantagens da distância de Hausdorff é que ela permite medir a similaridade entre vetores de diferentes tamanhos, o que permite o casamento parcial de padrões. Sua principal desvantagem é a sensibilidade ao ruído, o que tem motivado diversos estudos propondo modificações em sua formulação (ZHANG; LU, 2004).

## 2.7 AVALIAÇÃO DO DESEMPENHO DE ALGORITMOS DE RECONHECIMENTO DE OBJETOS

A avaliação do desempenho de algoritmos de reconhecimento de padrões/objetos em imagens é um assunto que ainda desperta o interesse de pesquisadores em visão computacional. Um dos principais problemas acerca deste tema é a falta de bases de imagens que permitam a comparação de diferentes algoritmos (WOLF; JOLION, 2006). Apesar disso, diferentes medidas para esta finalidade podem ser encontradas na literatura. As mais utilizadas são as taxas de acertos (*recall*) e erros, além de curvas *recall-precision* e curvas ROC (*Receiver Operating Characteristic*) (FAWCETT, 2005). Estas e outras medidas, algumas das quais descritas pelas equações 2.8 a 2.11, são calculadas a partir da matriz de confusão ou tabela de contingência (FAWCETT, 2005) mostrada na Figura 2.4.

		Existe uma instância do objeto na imagem	
		Positivo (SIM)	Negativo (Não)
O algoritmo detectou uma instância do objeto na imagem	Positivo (SIM)	Verdadeiro Positivo (VP) Detecção correta	Falso Positivo (FP) Erro na detecção
	Negativo (Não)	Falso Negativo (FN) Erro na detecção	Verdadeiro Negativo (VN)

Figura 2.4 – Matriz de confusão ou tabela de contingência.

As categorias VP, FP, FN e VN são definidas como segue:

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

A seguir, são descritas pelas equações 2.8 a 2.11 algumas métricas de avaliação de desempenho com base nas categorias VP, FP, FN e VN.

$$\text{Taxa de acertos (hit rate ou recall) = sensibilidade} = \frac{VP}{VP + FN} \times 100 \quad (2.8)$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \times 100 \quad (2.9)$$

$$\text{Precisão} = \frac{VP}{VP + FP} \times 100 \quad (2.10)$$

$$\text{Acurácia} = \frac{VP + VN}{VP + FP + FN + VN} \times 100 \quad (2.11)$$

A taxa de acertos pode ser definida como a razão entre o número de detecções efetuadas corretamente pelo algoritmo (verdadeiros positivos) e o total de instâncias do padrão que aparecem nas imagens analisadas (casamentos possíveis). A especificidade corresponde à fração dos casos negativos corretamente identificados. A precisão é a razão entre o número de detecções corretas e o total de instâncias detectadas e a acurácia corresponde à fração dos casos (positivos ou negativos) corretamente identificados. Neste trabalho optou-se pelo uso da taxa de acertos para a avaliação de desempenho dos algoritmos.

### 3 METODOS UTILIZADOS EM ANÁLISES COMPARATIVAS

Neste capítulo é feita uma breve descrição dos métodos *Scale-Invariant Feature Transform* (SIFT) e EasyMatch, os quais são utilizados em análises comparativas com os métodos propostos. O fato de serem duas das soluções mais completas encontradas na literatura e disporem de implementações que facilitaram suas utilizações motivou a escolha destes métodos.

#### 3.1 SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

O SIFT, originalmente proposto por Lowe (1999), é um método de extração de atributos locais que permite fazer correspondência de pontos entre imagens, como ilustrado na Figura 3.1, e tem sido amplamente empregado em tarefas relativas à detecção de padrões em imagens. O conjunto de atributos locais extraídos por SIFT é invariante a transformações de similaridade, insensível à presença de ruídos e parcialmente robusto a alterações de brilho e contraste e mudanças de ponto de vista. Além disso, a forma como extrai os atributos locais, facilita o reconhecimento de um objeto parcialmente ocluso. A eficiência e robustez do SIFT têm motivado sua utilização em aplicações embarcadas na área da robótica (BONATO, 2007) e até mesmo em celulares (KNIGHT, 2007).

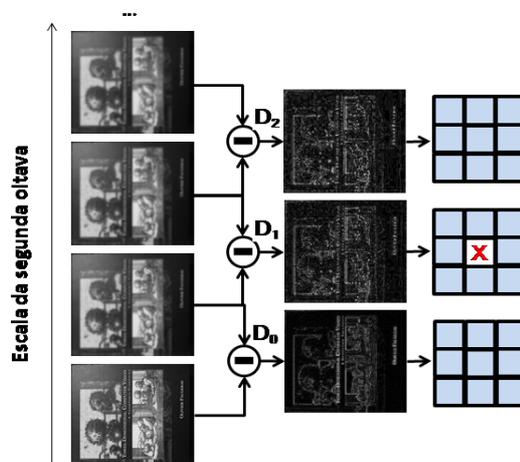


Figura 3.1 – Correspondência de pontos, feita por SIFT, entre duas imagens.

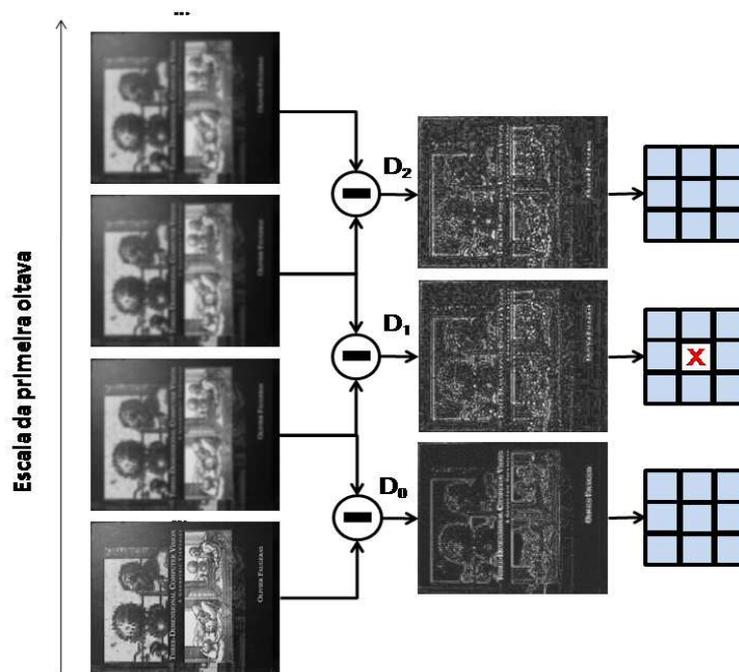
De acordo com Lowe (2004), o SIFT é composto por 4 estágios: detecção de pontos candidatos, seleção e localização dos pontos chaves (*keypoints*), definição da orientação dos pontos chaves e descritores dos pontos chaves.

### 3.1.1 Estágios do SIFT

No primeiro estágio é feita a detecção dos pontos candidatos. Um ponto é considerado candidato se possuir valores máximos ou mínimos em relação à sua vizinhança, definida por uma janela  $3 \times 3$  no mesmo espaço de escala e em escalas adjacentes separadas por uma constante multiplicativa  $k$ . Isto é feito com base na diferença de duas imagens idênticas geradas pela convolução gaussiana com diferentes fatores de escala  $\sigma$ , que por sua vez, é realizada por uma função denominada DoG (*Difference of Gaussian*). No algoritmo, além do número de escalas, deve-se também definir o número de oitavas. Uma oitava compreende um conjunto de imagens em uma determinada resolução processadas com diferentes parâmetros  $\sigma$ , formando uma espécie de pirâmide. A Figura 3.2 ilustra um exemplo do processo de detecção de pontos candidatos onde se utiliza duas oitavas com quatro escalas em cada oitava. A resolução das imagens da primeira oitava (Figura 3.2b) é  $260 \times 220$  pixels e da segunda oitava (Figura 3.2a)  $130 \times 110$  pixels. Normalmente, em uma oitava, a resolução é a metade da resolução da oitava anterior.



(a) Imagens da segunda oitava



(b) Imagens da primeira oitava

Figura 3.2 – Esquema de funcionamento do processo de detecção dos pontos candidatos.

Cada ponto  $D_i(x,y,\sigma)$  é comparado com seus 8 vizinhos na mesma escala e com seus 18 vizinhos de escalas adjacentes de uma mesma oitava e, se ele for um extremo (possuir valor máximo ou mínimo), então é um candidato a ponto chave.

Nos próximos dois estágios, faz-se um refinamento dos pontos candidatos com intuito de aceitar ou rejeitar cada um deles como ponto chave, define-se a localização exata e a escala de cada ponto chave aceito e calcula-se a orientação dominante para cada ponto chave para se conseguir invariância a rotação. O quarto estágio consiste na computação do descritor que representa a região relativa a cada ponto chave. Cada descritor é formado por um vetor de 128 bytes e é obtido a partir de histogramas de direção de gradiente .

### 3.1.2 Correspondência de pontos comuns entre duas imagens

Na tarefa de reconhecimento de padrões ou objetos, deve-se encontrar correspondência entre pontos chaves da imagem do padrão  $T$  e da imagem

analisada  $A$ , como ilustra a Figura 3.3. Assim, primeiro se aplica SIFT nas duas imagens para geração dos descritores (*keypoints*) e depois é feito o casamento dos pontos. Segundo Lowe (2004), considera-se a detecção do padrão  $T$  em  $A$  quando há pelo menos 3 pontos de correspondência entre as imagens  $T$  e  $A$ .

Tomando um ponto chave de  $T$ , para se encontrar pontos candidatos em  $A$ , determina-se o vizinho mais próximo “*nearest neighbour*” entre um ponto chave de  $T$  e todos os possíveis pontos chaves candidatos de  $A$ . A *determinação* do vizinho mais próximo é feita com base na distância Euclidiana dos vetores (Equação 2.4). Ainda segundo Lowe (2004), esta técnica pode ter um custo computacional alto quando se busca pontos correspondentes em grandes bancos de descritores, mas é rápida quando se trata de duas imagens. Além disso, técnicas como árvores de pesquisa para acelerar o processamento têm sido muito utilizadas.

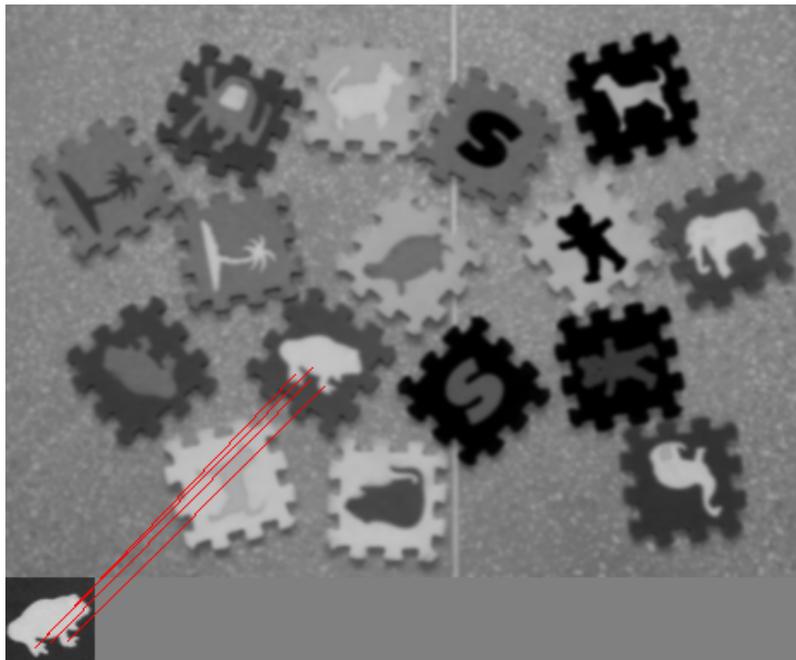


Figura 3.3 – Resultado do processo de detecção do padrão sapo por SIFT. A instância com contraste invertido não foi detectada.

Maiores detalhes acerca do método SIFT podem ser encontrados em Lowe (1999) e Lowe (2004).

### 3.2 EASYMATCH

O EasyMatch é uma ferramenta que incorpora um método de reconhecimento de padrões em imagens em níveis de cinza e coloridas e faz parte de uma biblioteca contendo outras ferramentas de visão computacional denominada Open eVision™, de propriedade da empresa EureSys (EURESYS, 2007). Por se tratar de um software comercial e de código fechado, as únicas informações sobre ele são aquelas disponíveis no site da empresa proprietária.

O método de reconhecimento de padrões incorporado ao EasyMatch baseia-se em atributos locais para detecção de objetos, não necessita de segmentação prévia das imagens, é invariante sob transformações de similaridade e robusto a presença de ruídos, oclusão parcial e alterações de brilho e contraste, permitindo localizar instâncias do padrão procurado com contraste invertido (ou negativo). A medida de similaridade utilizada pelo método é o coeficiente de correlação. Dada a sua robustez e eficiência, é comum encontrar trabalhos na literatura que propõem o uso do EasyMatch como parte de aplicações para funcionar em tempo real. Alguns exemplos são os trabalhos de Leemans e Destain (2004), Fanti e Maggiolo (2004) e Luqiao et al. (2007). A Figura 3.4 ilustra uma imagem processada por EasyMatch onde o objetivo foi localizar o padrão sapo utilizado em exemplos anteriores.

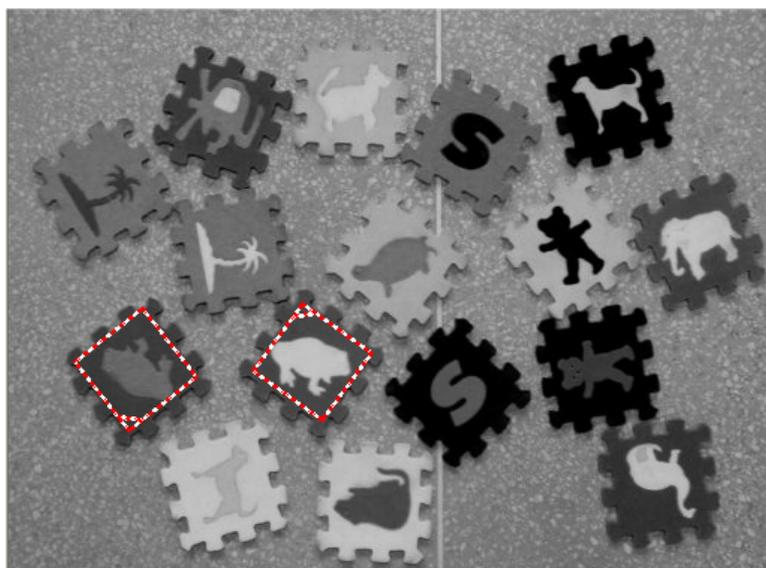


Figura 3.4 – Resultado do processo de detecção do padrão sapo por EasyMatch. Os retângulos formados por linhas pontilhadas indicam as posições onde foram detectadas instâncias do padrão.

O software EasyMatch (versão de demonstração), cujas telas são ilustradas nas Figuras 3.5 a 3.7, é parametrizável e apresenta dados numéricos relativos ao resultado do processo de detecção (Figura 3.7), como o coeficiente de correlação obtido, a localização, a escala e o ângulo de cada instância do objeto que é detectada na imagem.

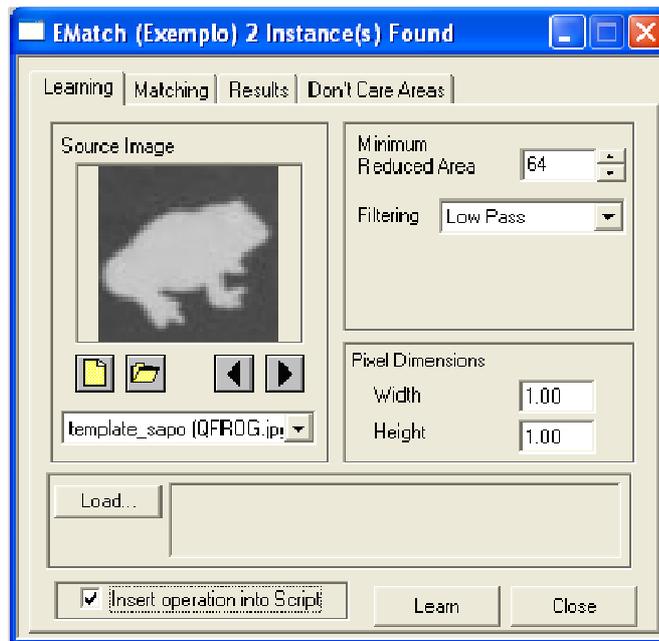


Figura 3.5 – Tela do EasyMatch para leitura da imagem  $T$  e extração de seus atributos.

A operação de EasyMatch é relativamente simples. Primeiro, seleciona-se a aba “*Learning*” (Figura 3.5) para efetuar a leitura da imagem do padrão procurado  $T$  (no exemplo acima o padrão é a imagem do sapo) e em seguida, pressiona-se o botão “*Learn*” para extração dos atributos do padrão. Este processo leva menos de 1s.

O próximo passo consiste em efetuar a leitura da imagem a ser analisada  $A$  e inserir os parâmetros utilizados pelo algoritmo de reconhecimento de padrões. Isto é feito selecionando-se a aba “*Matching*” ilustrada na Figura 3.6. Entre os parâmetros requeridos estão o intervalo angular e de escala, a quantidade máxima de instâncias de  $T$  em  $A$ , informação sobre o tipo de contraste (normal, invertido ou ambos) e o modo como é calculada a correlação.

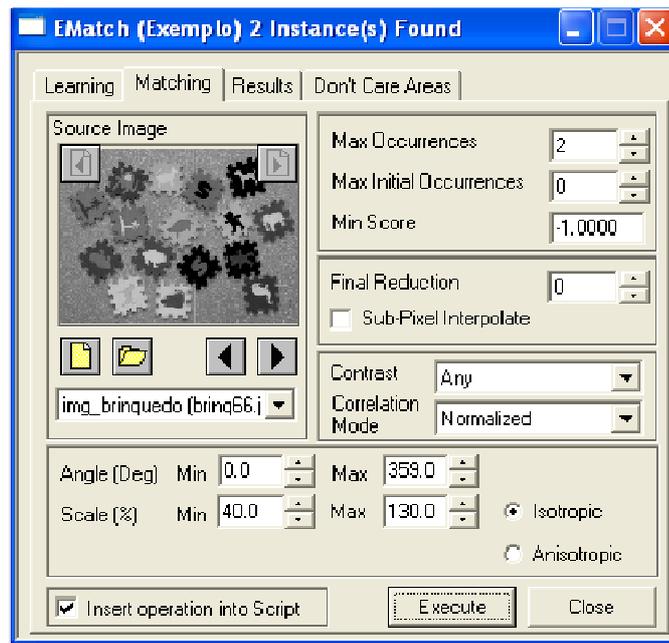


Figura 3.6 – Tela do EasyMatch para leitura da imagem *A* e inserção dos parâmetros utilizados pelo algoritmo de casamento de padrões.

A execução do algoritmo, feita por meio do botão “*Execute*”, é extremamente rápida (cerca de 1 minuto). Caso seja encontrada alguma instância de *T* em *A*, retângulos pontilhados são desenhados (ver Figura 3.4), indicando a localização, o ângulo e a escala de cada instância. Os resultados numéricos podem ser visualizados selecionando-se a aba “*Results*”, como mostra a Figura 3.7.

Score	Center X	Center Y	Angle (Deg)	Scale X (%)	Scale Y (%)
0.9622	203.00	238.50	34.893	111.05	111.05
0.9734	85.00	243.50	232.622	111.05	111.05

Figura 3.7 – Resultados numéricos do processo de detecção do padrão sapo por EasyMatch.

## **4 CASAMENTO DE PADRÕES LIVRE DE SEGMENTAÇÃO E INVARIANTE SOB TRANSFORMAÇÕES DE SIMILARIDADE, BRILHO E CONTRASTE**

Neste capítulo concentram-se as principais contribuições desta tese. Após uma breve introdução, é apresentada a medida de similaridade, denominada Corr, proposta para avaliar a semelhança entre vetores de atributos extraídos de imagens em níveis de cinzas e robusta a variações de brilho e contraste. Em seguida, é descrito o funcionamento do algoritmo de casamento de padrões pelo processo de força bruta. Este algoritmo não é uma contribuição direta deste trabalho, mas emprega Corr e é utilizado como uma das etapas de filtragem nos métodos propostos. Na seqüência, descreve-se o método Ciratefi, proposto para casamento de padrões em imagens em níveis de cinzas e uma extensão deste, baseada em operadores morfológicos, denominada Ciratefi morfológico. A seguir, é feita uma breve revisão da literatura acerca dos mecanismos empregados no reconhecimento de padrões em imagens coloridas e por fim é apresentado o color Ciratefi, bem como o espaço de cores utilizado e a medida de similaridade Sim, proposta para avaliar a semelhança perceptual entre vetores de cores.

### **4.1 INTRODUÇÃO**

Os algoritmos de casamento de padrões em imagens visam solucionar o problema de encontrar correspondências entre representações de um mesmo padrão em imagens distintas, ou ainda, determinar a posição de um padrão em uma imagem (CHOI; KIM, 2002). Assim, dado um conjunto de padrões representando objetos do mundo real e uma imagem de uma cena, por meio de operações de casamento de padrões, é possível interpretar a cena contida na imagem com base nos padrões.

Considerando a diversidade e a importância dos problemas existentes, diversos métodos de casamento de padrões em imagem têm sido implementados. Estes métodos podem ser caracterizados pelo tipo de representação utilizada para

descrever os padrões e as imagens no processo de casamento, sendo a mais comum a representação pictorial na qual se leva em conta os valores dos pixels da imagem. Outra forma de representação é a simbólica, na qual se empregam elementos que representam os objetos e suas interrelações.

## 4.2 MEDIDA DE SIMILARIDADE ROBUSTA A VARIAÇÕES DE BRILHO E CONTRASTE

Técnicas empregadas no reconhecimento de objetos em imagens, como por exemplo o casamento de padrões, utiliza alguma medida de similaridade para avaliar a semelhança entre o vetor de atributos extraídos da imagem  $T$  e os vetores extraídos da imagem  $A$ . Algumas medidas como distância Euclidiana, distância de Minkowsky, distância de Hausdorff e coeficiente de correlação têm sido bastante utilizadas. Neste trabalho, foi proposta uma modificação no coeficiente de correlação visando torná-lo robusto a variações de brilho e contraste. Para isto, seguiu-se o seguinte raciocínio: seja  $\mathbf{x}$  o vetor de atributos extraídos de  $T$ , que deve ser correlacionado com  $\mathbf{y}$  que é o vetor de atributos extraídos de uma sub-imagem de  $A$  então, a correção de brilho e contraste pode ser escrita como um problema de mínimos quadrados como mostrado na equação 4.1:

$$\mathbf{y} = \beta\mathbf{x} + \gamma\mathbf{1} + \boldsymbol{\varepsilon} \quad (4.1)$$

onde  $\mathbf{1}$  é um vetor identidade,  $\boldsymbol{\varepsilon}$  é o vetor de erros residuais,  $\beta$  é a correção de contraste e  $\gamma$  a correção de brilho. O problema, que pode ser computado rapidamente, consiste em encontrar o par de parâmetros  $\beta$  e  $\gamma$  que minimiza  $\boldsymbol{\varepsilon}^2$ .

Sejam  $\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$  e  $\tilde{\mathbf{y}} = \mathbf{y} - \bar{\mathbf{y}}$  os vetores  $\mathbf{x}$  e  $\mathbf{y}$  corrigidos pelas suas médias, onde  $\bar{\mathbf{x}}$  e  $\bar{\mathbf{y}}$  são, respectivamente, as médias de  $\mathbf{x}$  e  $\mathbf{y}$ , então:

$$\beta = \frac{\tilde{\mathbf{x}}\tilde{\mathbf{y}}}{\tilde{\mathbf{x}}^2}, \gamma = \bar{\mathbf{y}} - \beta\bar{\mathbf{x}} \text{ e } \boldsymbol{\varepsilon} = \tilde{\mathbf{y}} - \beta\tilde{\mathbf{x}}. \quad (4.2)$$

e o coeficiente de correlação modificado  $r_{xy}$  pode ser computado por:

$$r_{xy} = \frac{\bar{\mathbf{x}}\bar{\mathbf{y}}}{\|\bar{\mathbf{x}}\|\|\bar{\mathbf{y}}\|} = \frac{\beta\bar{\mathbf{x}}^2}{\|\bar{\mathbf{x}}\|\|\bar{\mathbf{y}}\|}. \quad (4.3)$$

Como pode ser visto na equação 4.2, caso os vetores  $\mathbf{x}$  e  $\mathbf{y}$  sejam idênticos e não representem regiões com níveis de cinzas constantes, então  $\beta=1$ ,  $\gamma=0$  e  $\epsilon^2=0$ , indicando que correções de contraste e brilho são desnecessárias.

A correlação  $r_{xy}$  assume o valor zero nos seguintes casos:

- a) Se a correção de brilho ou contraste requerida for muito alta porque, neste caso, considera-se que  $T$  e a sub-imagem de  $A$  provavelmente são muito diferentes;
- b) Se  $|\beta| \leq t_\beta$  ou  $\frac{1}{t_\beta} \leq |\beta|$ , onde  $0 < t_\beta \leq 1$ , é o limiar de correção de contraste escolhido. Por exemplo, se  $t_\beta=0,5$  significa que regiões de  $A$  com contraste menor que meio ou maior que duas vezes o contraste de  $T$  são consideradas não-correlacionadas. Isto também evita operações de divisão por zero, nos casos onde regiões de  $A$  com níveis de cinzas são constantes já que o valor de  $\|\bar{\mathbf{y}}\|$  é zero;
- c) Se  $|\gamma| > t_\gamma$ , onde  $0 < t_\gamma \leq 1$  é limiar de correção de brilho escolhido, pois assume-se que os níveis de cinzas das imagens são números reais no intervalo  $[0, 1]$ .

De acordo com o que foi exposto, definiu-se o coeficiente de correlação (Corr) que leva em conta correções de brilho e contraste como:

$$\text{Corr}(\mathbf{x}, \mathbf{y}) = \begin{cases} 0, & \text{se } |\beta| \leq t_\beta, \frac{1}{t_\beta} \leq |\beta| \text{ ou } |\gamma| > t_\gamma \\ r_{xy}, & \text{caso contrário} \end{cases} \quad (4.4)$$

Dependendo da aplicação, pode-se usar o valor absoluto de  $\text{Corr}(\mathbf{x}, \mathbf{y})$  permitindo o casamento de instâncias negativas (com contraste invertido) de  $T$  ou o valor de  $\text{Corr}(\mathbf{x}, \mathbf{y})$  com sinal que não permite o casamento de instâncias negativas. Em todos os experimentos realizados com o Ciratefi utilizou-se o valor absoluto de  $\text{Corr}(\mathbf{x}, \mathbf{y})$ .

### 4.3 CASAMENTO DE PADRÕES PELO MÉTODO DA FORÇA BRUTA INVARIANTE SOB TRANSFORMAÇÕES DE SIMILARIDADE, BRILHO E CONTRASTE

O casamento de padrões invariante sob transformações de similaridade pela força bruta (também chamado de casamento de padrões convencional) é uma solução óbvia e precisa para o reconhecimento de objetos em imagens e consiste em fazer uma série de operações de casamento de padrões entre  $T$  e sub-imagens de  $A$ . Para isso, consideram-se todos os pixels de  $A$  como sendo pontos de interesse e todas as escalas e rotações que  $T$  possa estar sujeito dentro de um intervalo discreto estabelecido (TSAI; TSAI, 2002). Claramente, esta solução é inviável para aplicações que possam ser utilizadas na prática já que demanda muito tempo de processamento. Para melhor ilustrar o problema, considere um experimento utilizando esta técnica no qual o objetivo foi detectar o padrão sapo ( $T$ ) (Figura 4.1a) na imagem  $A$  (Figura 4.1b), contendo amostras de  $T$  rotacionadas, em diferentes escalas e com diferentes brilhos e contrastes. Para isso, considerou-se  $T$  rotacionado em  $m=36$  diferentes ângulos ( $\alpha_0=0$ ,  $\alpha_1=10$ , ...,  $\alpha_{35}=350$ ) e escalonado por  $n=6$  diferentes fatores de escala ( $s_0=0,6$ ,  $s_1=0,7$ , ... ,  $s_5=1,1$ ), totalizando 216 amostras de  $T$  (Figura 4.1c). Para evitar possibilidade de erros no processo de casamento,  $A$  e  $T$  foram suavizadas por um filtro passa-baixas gaussiano com  $\sigma=1$ .

O casamento pela força bruta é feito da seguinte maneira: para cada pixel  $(x,y)$  de  $A$  extrai-se uma sub-imagem em torno de  $(x,y)$  e compara-se com cada uma das 216 amostras de  $T$ . Se o maior valor absoluto de Corr (equação 4.4) for acima de um limiar  $t_f$ , considera-se que uma instância de  $T$  foi detectada na posição  $(x,y)$  de  $A$ .

Além de detectar corretamente as ocorrências do padrão  $T$  na imagem  $A$ , como pode ser visto na Figura 4.1d, o algoritmo de casamento de padrões pela força bruta é capaz de fornecer com precisão o fator de escala e o ângulo de rotação para cada ponto de casamento. O grande problema é o tempo de processamento. O algoritmo levou 9.173 segundos ou aproximadamente 2 horas e meia para localização dos alvos utilizando um microcomputador Pentium4-2.8GHz. A realização desta mesma tarefa pelo Ciratefi, descrito a seguir, levou apenas 22 segundos.

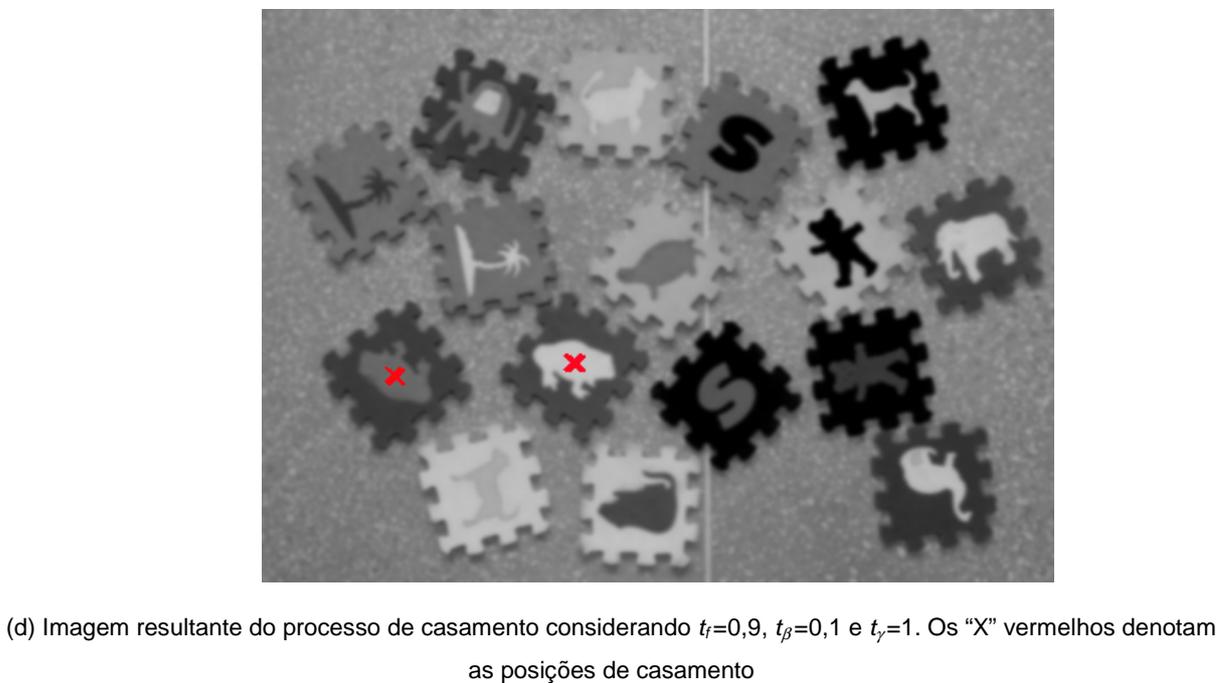
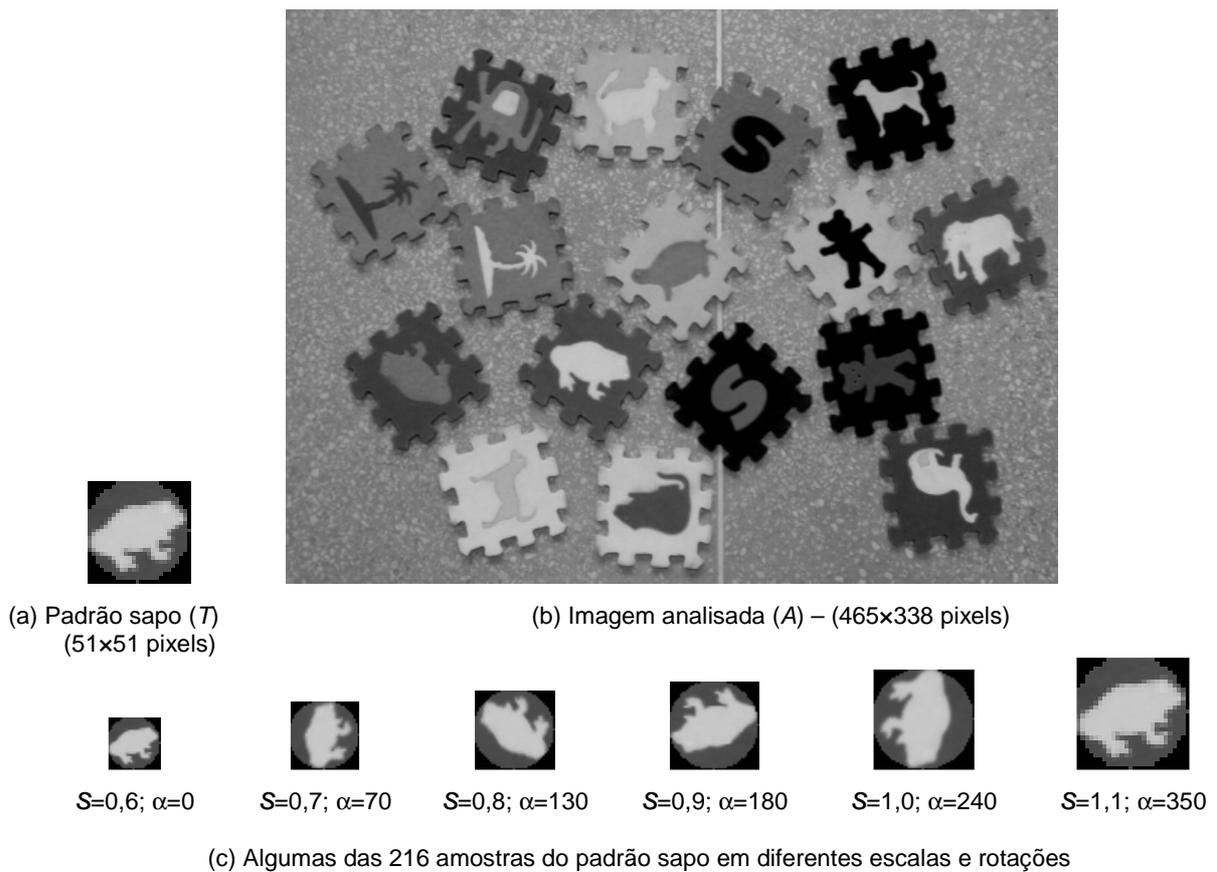


Figura 4.1 – Processo de detecção do padrão sapo pelo algoritmo de força bruta.

Cabe ressaltar que o valor de  $t_f$  foi escolhido de forma empírica e os valores de  $t_\beta$  e  $t_\gamma$  foram determinados para permitir o máximo de correção de contraste e brilho, respectivamente.

#### 4.4 CIRCULAR, RADIAL AND TEMPLATE-MATCHING FILTER (CIRATEFI)

O Ciratefi consiste de um algoritmo de casamento de padrões em imagens, a partir de características globais e invariantes extraídas de suas regiões. Mais especificamente, trata-se de um encadeamento de três etapas de filtragem que sucessivamente descartam pontos na imagem analisada ( $A$ ) que não correspondem ao *padrão* ( $T$ ). Na primeira etapa de filtragem, denominada *Circular sampling filter* (Cifi), a extração dos atributos é feita por meio de projeções circulares. Na segunda etapa, projeções radiais são utilizadas na tarefa de extração. Denominamos a segunda etapa de *Radial sampling filter* (Rafi). Na terceira etapa de filtragem, restando poucos pontos candidatos a serem avaliados na imagem  $A$ , emprega-se a técnica convencional de casamento de padrões (casamento de padrões pela força bruta). Desta forma, dadas duas imagens em níveis de cinza  $A$  e  $T$ , a finalidade do algoritmo é encontrar todas as ocorrências de  $T$  em  $A$ , considerando que instâncias de  $T$  podem estar contidas em  $A$  em diferentes posições, rotações e escalas e com diferentes intensidades de brilho e contraste. Para medir a similaridade entre  $T$  e as sub-imagens extraídas de  $A$  utiliza-se o coeficiente de correlação  $Corr$  (equação 4.4).

É importante ressaltar que muitos trabalhos encontrados na literatura ao longo das três últimas décadas, entre os quais (HSU; AESENAULT; APRIL, 1982; GOSHTASBY, 1985; TAZA; SUEN, 1989; YUEN; FENG; TANG, 1998; TAO; TANG, 1999; CHANG; HORNAK, 2000 e TSAI; TSAI, 2002), descrevem métodos que utilizam projeções circulares e/ou radiais. Contudo, seus funcionamentos e objetivos são completamente diferentes do método proposto neste trabalho.

##### 4.4.1 Circular sampling filter (Cifi)

A primeira etapa de filtragem, denominada Cifi, utiliza projeções das imagens  $A$  e  $T$  sobre um conjunto de círculos (Figura 4.2a) para detectar os pontos (pixels) candidatos em  $A$ . Para cada um destes pontos estima-se a provável escala. Temos assim os pixels candidatos da primeira fase. Como podemos ver na subseção 5.1.5,

a escolha correta do número de círculos não é essencial para o algoritmo dado que os pixels candidatos serão novamente filtrados nas duas etapas seguintes. Apesar disso, uma escolha adequada pode melhorar o desempenho do algoritmo.

Seja  $I$  uma imagem em níveis de cinza, a amostragem circular  $Cis_I(x, y, r)$  pode ser definida como a média dos níveis de cinza dos pixels de  $I$  situados a uma distância  $r$  do pixel localizado na posição  $(x, y)$ :

$$Cis_I(x, y, r) = \frac{1}{2\pi r} \int_0^{2\pi} I(x + r \cos \theta, y + r \sin \theta) d\theta \quad (4.5)$$

Na prática, utiliza-se uma representação discreta de  $Cis_I(x, y, r)$  que é dada por:

$$Cis_I(x, y, r) = \frac{1}{P} \sum_{\theta=0}^{P-1} I\left(x + r \cos \frac{2\pi\theta}{P}, y + r \sin \frac{2\pi\theta}{P}\right) \quad (4.6)$$

Onde  $P$  é o número de pixels ao longo do círculo de raio  $r$ .

Algoritmos para desenhar círculos como os descritos em (BRESENHAM, 1977) podem ser usados para encontrar de forma eficiente todos os pixels ao longo de um círculo de raio  $r$ . Estes algoritmos descartam o uso das funções de seno e cosseno que demandam alto custo computacional.

Dada a imagem  $T$  e um conjunto de  $n$  fatores de escala (em nosso exemplo utilizamos  $s_0=0,6, s_1=0,7, \dots, s_5=1,1$ ) a imagem  $T$  é escalonada por cada fator  $s_i$ , obtendo os padrões escalonados  $T_0, T_1, \dots, T_{n-1}$ , conforme descrito adiante na seção 4.4.4. Depois disso, cada  $T_i$  é circularmente amostrado por um conjunto  $l$  de raios predefinidos (em nosso exemplo utilizamos  $l=13$ , e  $r_0=0$  pixels,  $r_1=2$  pixels, ...,  $r_{12}=24$  pixels), produzindo a matriz multi-escala de características invariantes à rotação  $C_T$  com  $n$  linhas (escalas) e  $l$  colunas (raios):

$$C_T[i, k] = Cis_{T_i}(x_0, y_0, r_k), \quad 0 \leq i < n \text{ e } 0 \leq k < l \quad (4.7)$$

onde  $(x_0, y_0)$  é o pixel central de  $T$ ,  $i$  e  $k$  denotam, respectivamente, a linha e a coluna da matriz  $C_T$ .

A extração de atributos da imagem a ser analisada  $A$ , nesta primeira etapa de filtragem, é feita considerando todos os seus pixels. Com base na imagem  $A$ , uma imagem 3D  $C_A[x,y,k]$  é construída como segue:

$$C_A[x,y,k] = \text{Cis}_A(x,y,r_k), \quad 0 \leq k < l \text{ e } (x,y) \in \text{domínio}(A) \quad (4.8)$$

O filtro Cifi utiliza as matrizes  $C_T$  e  $C_A$  e os limiares (*thresholds*) de contraste  $t_\beta$  e de brilho  $t_\gamma$  para detectar a correlação  $\text{CisCorr}$  da amostragem circular na melhor escala de casamento, para cada pixel  $(x,y)$ .

$$\text{CisCorr}_{A,T}(x,y) = \text{MAX}_{i=0}^{n-1} [ |\text{Corr}(C_T[i], C_A[x,y])| ] \quad (4.9)$$

onde  $n$  é o número de fatores de escala e  $\text{Corr}$  é o coeficiente de correlação modificado, descrito na equação 4.4.

O pixel  $(x,y)$  é classificado como um candidato da primeira fase se  $\text{CisCorr}_{A,T}(x,y) \geq t_1$  para um determinado limiar  $t_1$  (em nosso exemplo, definimos  $t_1=0,95$ ). Conforme pode ser visto na subseção 5.1.5, a escolha do parâmetro  $t_1$  não é crítica, contanto que ele seja pequeno o suficiente para não descartar pontos com reais chances de casamento. A provável escala  $\text{CisPS}$  associada a cada pixel candidato  $(x,y)$  da primeira fase é dada por:

$$\text{CisPS}_{A,T}(x,y) = \text{ARGMAX}_{i=0}^{n-1} [ |\text{Corr}(C_T[i], C_A[x,y])| ] \quad (4.10)$$

onde  $\text{ARGMAX}$  é um operador que retorna o índice  $i$  relacionado à escala de  $T$  que maximiza a correlação  $\text{Corr}$ .

A geração de  $C_A[x,y,k]$  a partir da imagem ilustrada na Figura 4.1b (465x338 pixels) levou 2,5 segundos e a computação de  $\text{CisCorr}_{A,T}(x,y)$  para todos os pixels da mesma imagem levou 4,5 segundos. Neste exemplo, as demais operações consumiram menos de 1 segundo.

A Figura 4.2a ilustra as projeções circulares sobre a imagem  $T$ , a Figura 4.2b ilustra uma representação gráfica do conjunto de atributos extraídos a partir da

amostragem circular e a Figura 4.2c mostra a imagem resultante da filtragem por Cifi, na qual os pixels candidatos da primeira fase aparecem na cor magenta.

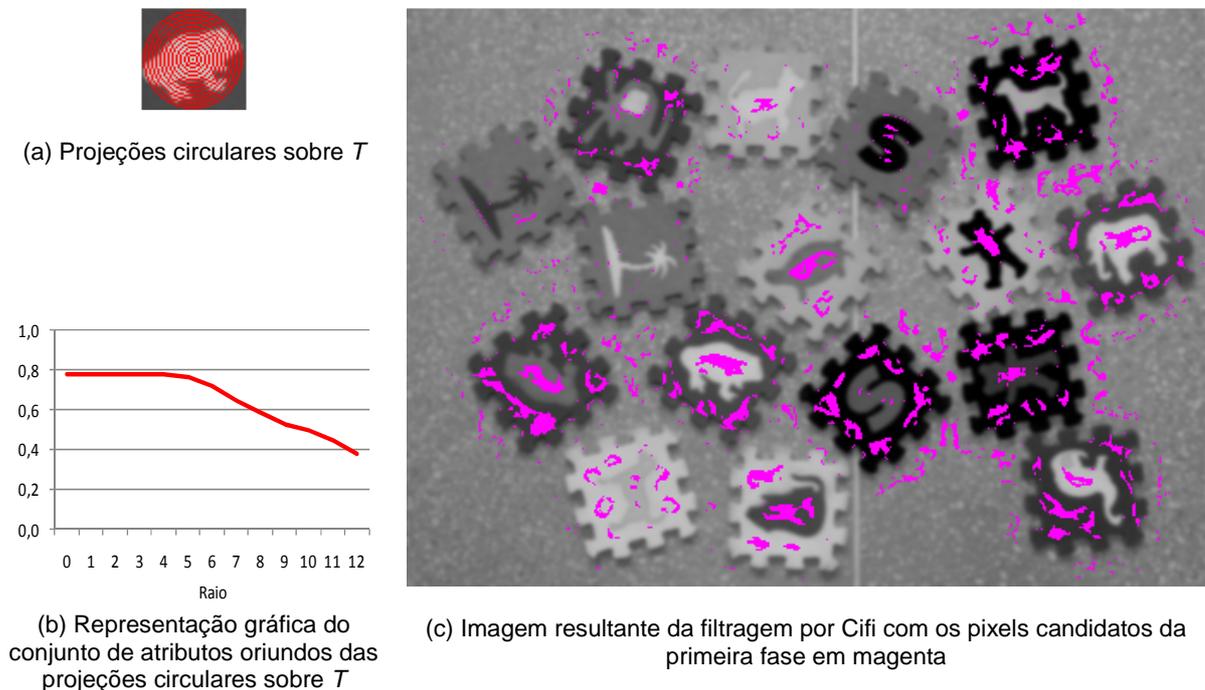


Figura 4.2 – Filtragem por Cifi.

#### 4.4.2 Radial sampling filter (Rafi)

Rafi é a segunda etapa de filtragem e utiliza projeções das imagens  $A$  e  $T$  sobre um conjunto de linhas radiais (Figura 4.3a) para reavaliar os pixels candidatos da primeira fase. Nesta etapa, uma parte dos pixels candidatos são promovidos para compor o conjunto de candidatos da segunda fase e a outra parte é descartada. Rafi também associa o provável ângulo de rotação a cada pixel candidato da segunda fase. O conjunto de inclinações das linhas radiais precisa ser igual ao conjunto de ângulos  $m$  escolhido (em nosso exemplo utilizamos  $m=36$ ,  $\alpha_0=0$ ,  $\alpha_1=10$ , ...,  $\alpha_{35}=350$ ). A escolha de  $m$  não é crítica, contanto que não seja atribuído um valor muito pequeno para não descartar pixels com chances reais de casamento (conforme descrito na subseção 5.1.5).

Dada uma imagem  $I$ , a amostragem radial  $Ras_{\lambda}^{\alpha}(x,y)$  pode ser definida como a média dos níveis de cinza dos pixels de  $I$  localizados sobre linhas radiais com um vértice em  $(x,y)$ , comprimento  $\lambda$  e inclinação  $\alpha$ .

$$\text{Ras}_I^\lambda(x, y, \alpha) = \frac{1}{\lambda} \int_0^\lambda I(x + t \cos \alpha, y + t \sin \alpha) dt \quad (4.11)$$

A exemplo de Cis, na prática, uma representação discreta de Ras é utilizada:

$$\text{Ras}_I^\lambda(x, y, \alpha) = \frac{1}{Q} \sum_{t=0}^{r_{l-1}} I(x + t \cos \alpha, y + t \sin \alpha) \quad (4.12)$$

onde  $Q$  é o número de pixels ao longo da linha radial com inclinação  $\alpha$  e comprimento  $\lambda=r_{l-1}$  (raio do maior círculo utilizado na filtragem por Cifi). Também neste caso, algoritmos eficientes para desenhar as linhas como os descritos em (BRESENHAM, 1965) podem ser utilizados para encontrar os pixels que estão ao longo de uma linha radial.

Dada uma imagem  $T$  e o conjunto de  $m$  ângulos de inclinação  $(\alpha_0, \alpha_1, \dots, \alpha_{m-1})$ ,  $T$  é radialmente amostrada produzindo o vetor  $R_T$  com  $m$  atributos:

$$R_T[j] = \text{Ras}_T^{r_{l-1}}(x_0, y_0, \alpha_j), \quad 0 \leq j < m \quad (4.13)$$

onde  $(x_0, y_0)$  é o pixel central de  $T$ .

Para cada pixel candidato  $(x, y)$  da primeira fase,  $A$  é radialmente amostrada considerando a provável escala  $i = \text{CisPS}_{A,T}(x, y)$ . O maior raio  $r_{l-1}$  redimensionado com base na provável escala  $s_i$ , resulta em  $\lambda = s_i r_{l-1}$ . Assim, uma matriz  $R_A$  de atributos é gerada:

$$R_A[x, y, j] = \text{Ras}_A^{s_i r_{l-1}}(x, y, \alpha_j), \quad 0 \leq j < m \text{ and } (x, y) \in \text{cand\_1}^a \text{\_fase}(A) \quad (4.14)$$

Na seqüência, para cada pixel candidato  $(x, y)$  da primeira fase é computada a correlação  $\text{RasCorr}$  para determinar o melhor ângulo de casamento. Para isso, são utilizados os vetores  $R_A[x, y]$ ,  $R_T$  e os limiares de brilho e contraste  $t_\gamma$  e  $t_\beta$ .

$$\text{RasCorr}_{A,T}(x, y) = \text{MAX}_{j=0}^{m-1} \left[ \left| \text{Corr}(R_A[x, y], \text{cshift}_j(R_T)) \right| \right], \quad (x, y) \in \text{cand\_1}^a \text{\_fase}(A) \quad (4.15)$$

onde  $\text{cshift}_j(R_T)$  significa um deslocamento circular de  $j$  posições do vetor  $R_T$ .

O pixel candidato  $(x,y)$  da primeira fase é promovido para segunda fase se  $\text{RasCorr}_{A,T}(x,y) \geq t_2$  para um dado limiar  $t_2$  (em nosso exemplo,  $t_2=0,9$ ). Como no caso de  $t_1$ , a escolha do parâmetro  $t_2$  não é crítica, contanto que ele seja pequeno o suficiente para não descartar pontos com reais chances de casamento.

O provável ângulo de rotação para cada pixel candidato da segunda fase é o ângulo de melhor casamento, dado por  $\text{RasAng}$  como segue:

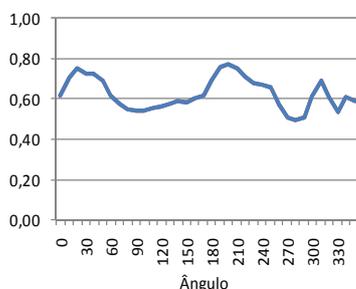
$$\text{RasAng}_{A,T}(x,y) = \underset{j=0}{\text{ARGMAX}}^{m-1} \left[ \left| \text{Corr}(R_A[x,y], \text{cshift}_j(R_T)) \right| \right] \quad (4.16)$$

onde  $\text{ARGMAX}$  é um operador que retorna o deslocamento  $j$  relacionado ao ângulo de rotação que maximiza  $\text{Corr}$ .

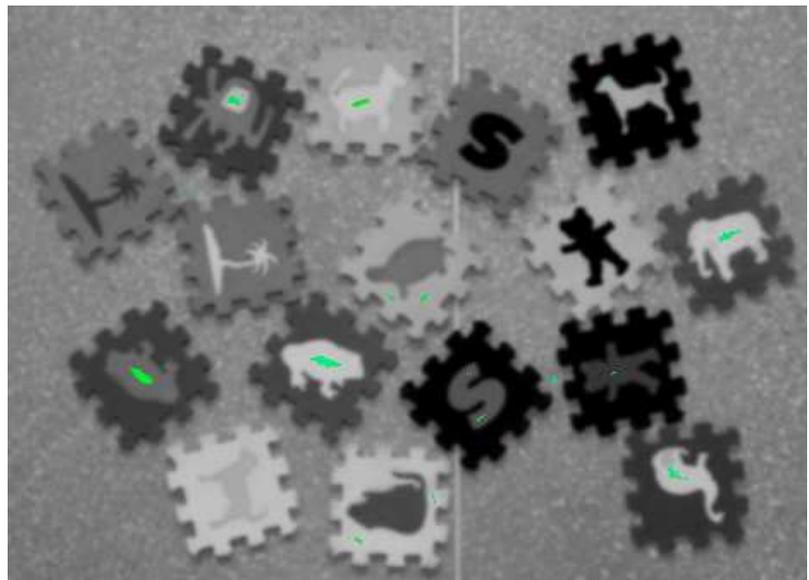
A computação de  $\text{RasCorr}_{A,T}(x,y)$  para todos os pixels  $(x,y)$  da imagem  $A$  utilizada como exemplo levou 13 segundos. As outras operações envolvidas em na filtragem Rafi consumiram menos de 1 segundo. A Figura 4.3a ilustra as linhas radiais sobre  $T$  em azul, a Figura 4.3b ilustra uma representação gráfica do conjunto de atributos extraídos a partir da amostragem radial e a Figura 4.3c ilustra a imagem resultante da filtragem por Rafi, na qual os pixels na cor verde são os candidatos da segunda fase.



(a) Projeções radiais sobre  $T$



(b) Representação gráfica do conjunto de atributos oriundos das projeções radiais sobre  $T$



(c) Imagem resultante da filtragem por Rafi com os pixels candidatos da segunda fase em verde

Figura 4.3 – Filtragem por Rafi.

#### 4.4.3 Template Matching filter (Tefi)

A terceira etapa de filtragem, denominada Tefi, emprega o algoritmo de casamento de padrões convencional invariante a alterações de brilho e contraste, descrito na seção 4.3, aplicado apenas nos pixels candidatos da segunda fase, usando a provável escala e o provável ângulo de inclinação determinados, respectivamente, por Cifi e Rafi. Assim, se  $(x,y)$  é um pixel da segunda fase com provável escala  $i = \text{CisPS}_{A,T}(x,y)$  e provável ângulo  $j = \text{RasAng}_{A,T}(x,y)$ , então Tefi calcula  $\text{Corr}$  (equação 4.4) entre  $T$  na escala  $s_i$ , ângulo  $\alpha_j$  e a sub-imagem de  $A$  em torno do pixel  $(x,y)$ . Se o valor absoluto da correlação for acima de um determinado limiar  $t_3$ , considera-se que uma instância de  $T$  foi detectada em  $A$ , com o centro da instância de  $T$  em  $(x,y)$ . A Figura 4.4 ilustra o resultado final do processo de detecção por Ciratefi. Neste exemplo, Tefi levou 1 segundo para ser computado.

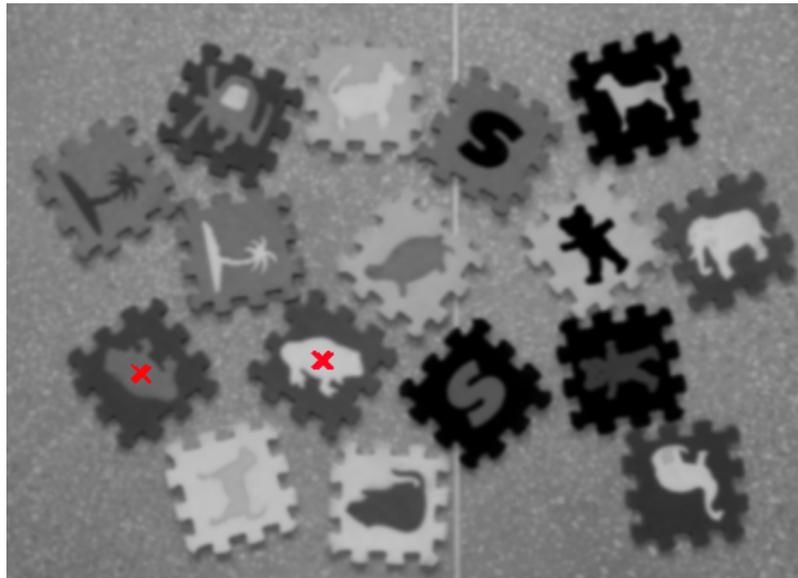


Figura 4.4 – Resultado da detecção por Ciratefi. Os objetos detectados estão marcados com um “X” vermelho.

Adotando o mesmo limiar usado no algoritmo de força bruta, ou seja  $t_3 = t_f$ , o resultado de Ciratefi normalmente é igual ou muito similar ao resultado do algoritmo de força bruta. Uma possibilidade de melhorar a acurácia do Ciratefi é testar casamentos considerando um conjunto de escalas próximas de  $i$  (por exemplo,  $i-1$ ,  $i$ ,  $i+1$ ) e um conjunto de ângulos próximos de  $j$  (por exemplo,  $j-1$ ,  $j$ ,  $j+1$ ) ou computar as adições e subtrações em módulo  $m$ .

O funcionamento do Ciratefi envolvendo as três etapas de filtragem é ilustrado na Figura 4.5.

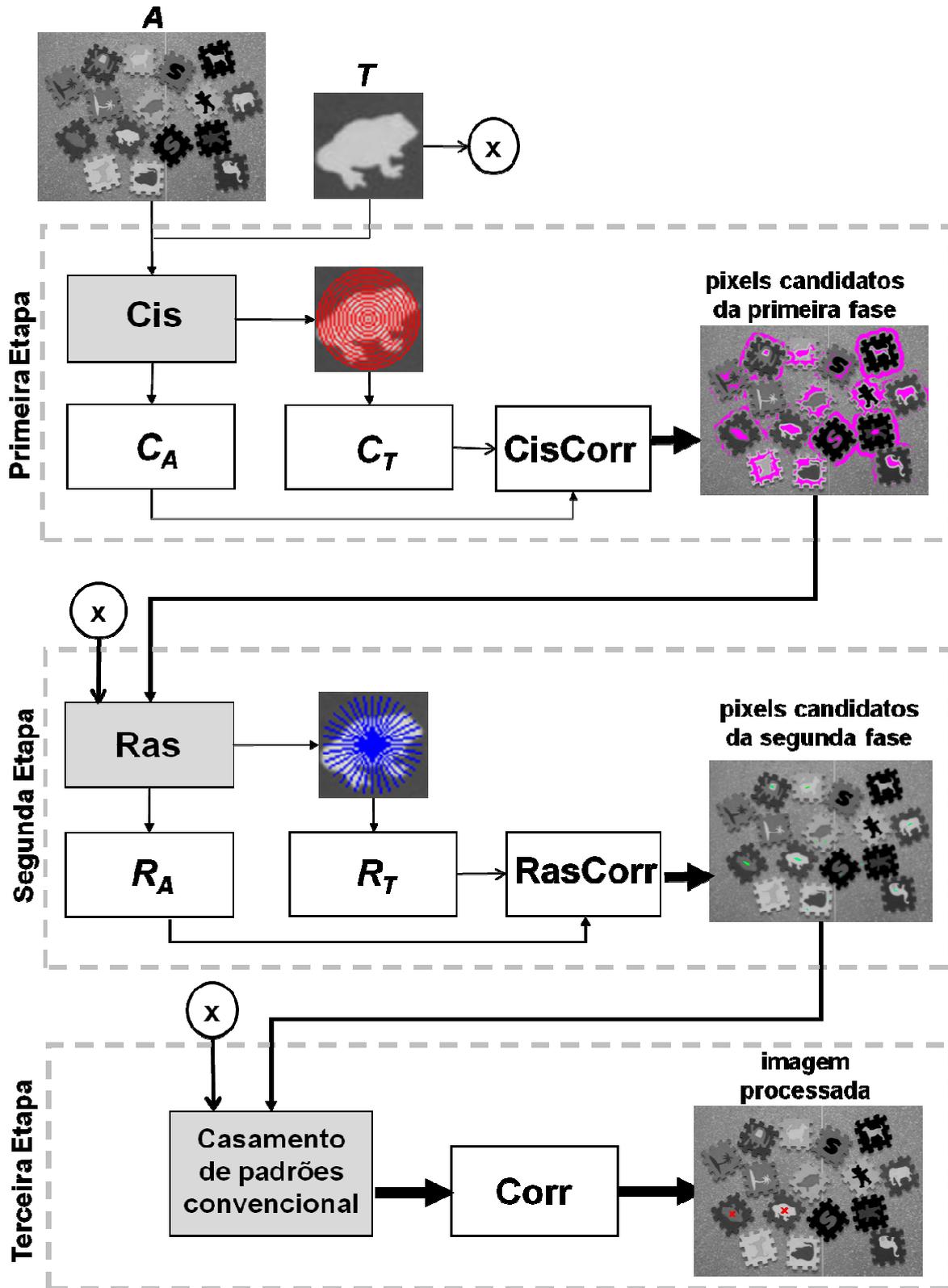


Figura 4.5 – Esquema de funcionamento das três etapas de filtragem do Ciratefi.

#### 4.4.4 Abordagem multi-escala

Ciratefi não é intrinsecamente invariante à escala. Contudo, o método obtém invariância sob transformações de similaridade (rotação, translação e escala uniforme) porque Cifi associa a provável escala a cada pixel candidato da primeira fase, depois de fazer a comparação com os vetores de atributos dos padrões escalonados  $T_0, T_1, \dots, T_{n-1}$  (considerando os  $n$  fatores de escala). Então, a filtragem Rafi é usada para eliminar os falsos positivos. E, se ainda permanecer algum caso de falso positivo, será descartado na terceira etapa de filtragem. Por exemplo, se considerarmos que as escalas das instâncias de  $T$  contidas em  $A$  variam de 60% a 110% do tamanho de  $T$ , escalona-se  $T$  com várias escalas no intervalo especificado (por exemplo: 60%, 70%, 80%, 90%, 100% e 110%), mantendo a mesma distância entre os círculos concêntricos para cada  $T_i$ . Desta maneira é obtida a matriz multi-escala ( $C_T$ ) de características invariantes a rotação com 6 escalas como ilustrado na Figura 4.6. Cabe salientar que em todas as operações de escalonamento e rotação de imagens descritas neste trabalho, empregou-se a técnica de interpolação bilinear para realização do processo de reamostragem.

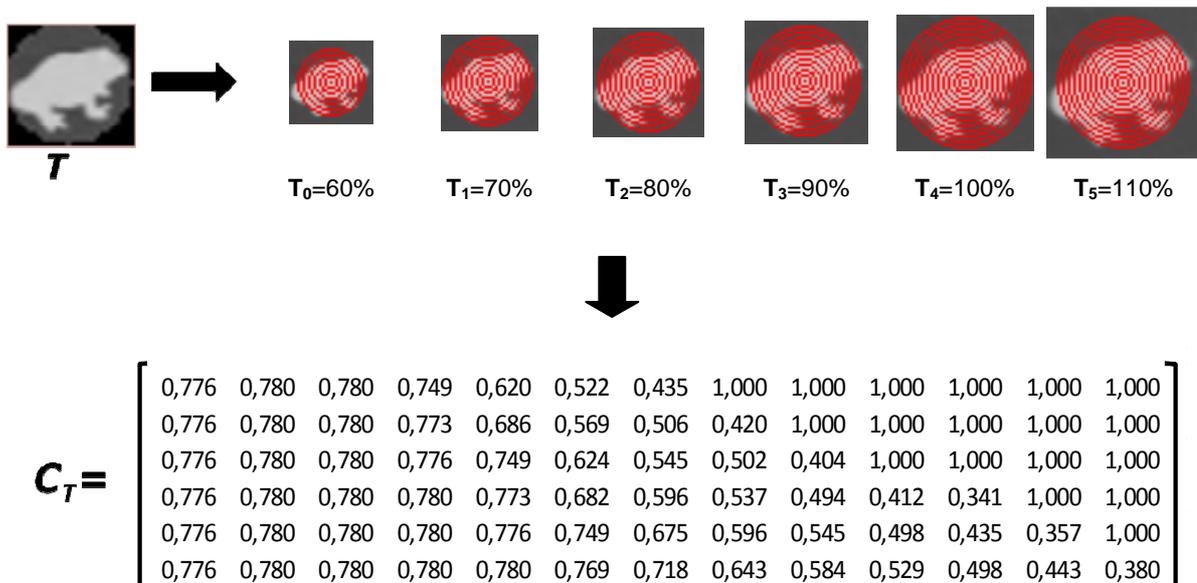


Figura 4.6 – Matriz multi-escala  $C_T$  com 6 linhas (escalas) e 13 colunas (círculos).

Uma maneira de tornar Ciratefi realmente invariante à escala é utilizar uma idéia de pirâmide semelhante à utilizada por SIFT (Figura 3.2). No caso do Ciratefi, a pirâmide deve ser formada de matrizes  $C_A$  computadas a partir de imagens  $A$  em diferentes resoluções divididas por oitavas, sendo que em cada oitava a resolução da imagem é a metade da resolução da imagem da oitava anterior. Então, bastaria construir uma única matriz  $C_T$  com fatores de escala variando de 0,5 a 1,0 e calcular a correlação CisCorr entre a matriz  $C_T$  e as matrizes  $C_A$  da pirâmide, considerando um determinado número de oitavas. Dessa forma, o fator de escala associado a cada pixel candidato oriundo da correlação de  $C_T$  com  $C_A$  da primeira oitava varia de 0,5 a 1,0; o fator de escala associado a cada pixel candidato oriundo da correlação de  $C_T$  com  $C_A$  da segunda oitava varia de 1,0 a 2,0, e assim sucessivamente.

#### 4.4.5 Uma breve análise da complexidade computacional do Ciratefi

A importância da análise de complexidade de um dado algoritmo está relacionada à possibilidade de estimação da quantidade de recursos (tempo e memória) necessários para executá-lo, visto que o fato de haver um algoritmo para resolver um determinado problema não implica, necessariamente, que ele possa ser resolvido na prática (LEWIS; PAPADIMITRIOU, 2000). Em complexidade de algoritmos a notação  $O$  é amplamente utilizada para comparar o crescimento assintótico de duas funções e por este motivo é a notação utilizada nesta seção.

Uma das dificuldades para analisar a complexidade computacional do Ciratefi é que ele depende de muitos fatores imprevisíveis como, por exemplo, o número de pixels candidatos da primeira e da segunda fase. Assim, considerando que  $N$  é o número de pixels da imagem analisada  $A$  e  $M$  o número de pixels da imagem  $T$ , são necessárias as seguintes assunções e aproximações:

- Assume-se que o número de escalas  $n$ , o número de ângulos  $m$ , e o número de círculos  $l$  são todos  $O(\sqrt{M})$ .
- Todas as operações que não dependem de  $N$  são ignoradas, dado que normalmente  $N$  é muito maior que  $M$ ;

O algoritmo Ciratefi tem quatro operações que dependem de  $N$ :

- A geração da imagem 3D  $C_A[x,y,k]$  tem custo computacional  $O(NM)$ , considerando que quase todos os pixels do domínio de  $T$  são levados em conta para cada pixel  $(x,y)$  de  $A$  que é processado.
- A computação de CisCorr para todos os pixels de  $A$  tem custo  $O(Nn)$ , ou aproximadamente  $O(NM)$ .
- A computação de  $R_A[x,y,j]$  e RasCorr para todos os pixels candidatos da primeira fase tem custo  $O(N_1m\sqrt{M})$  e  $O(N_1m^2)$ , respectivamente.  $N_1$  é o número de pixels candidatos da primeira fase.  $O(N_1m\sqrt{M} + N_1m^2)$  pode ser aproximado para  $O(NM)$ .
- A computação de Tefi tem custo  $O(N_2M)$ , onde  $N_2$  é o número de pixels candidatos da segunda fase, e  $O(N_2M) \leq O(NM)$ .

Conseqüentemente, a complexidade do Ciratefi é  $O(NM)$ . Já o algoritmo de força bruta descrito na seção 4.3 realiza  $n \times m$  casamentos de padrões para cada pixel de  $A$ . Considerando que cada casamento faz  $O(M)$  operações, a complexidade deste algoritmo é  $O(NnmM)$ , ou aproximadamente  $O(NM^2)$ . Isto faz uma grande diferença e justifica Ciratefi ter sido, nos experimentos realizados, aproximadamente 400 vezes mais rápido que o algoritmo de força bruta.

## 4.5 CIRATEFI MORFOLÓGICO

A morfologia matemática baseia-se na teoria dos conjuntos e em operações lógicas para extrair características geométricas e topológicas de uma imagem por meio da utilização de padrões de formatos pré-definidos denominados elementos estruturantes. As operações elementares da morfologia matemática são a dilatação e a erosão, as quais compõem a base do mecanismo descritor do Ciratefi morfológico. No caso de imagens em níveis de cinza, estas operações são usadas para computar os níveis de cinza máximos e mínimos de uma região da imagem, de acordo com o elemento estruturante.

Dadas as suas características, a morfologia matemática tem sido amplamente utilizada para descrição de padrões em imagens e também em outras tarefas como realce, segmentação, filtragem e esqueletonização (BANON, 1995; YU; WANG, 2005; JALBA; WILKINSON; ROERDINK, 2006). Em se tratando de descrição de padrões, algumas aplicações em diferentes áreas podem ser encontradas na literatura. Entre elas podemos citar análise de imagens médicas (YU-QIAN et al., 2005; NAEGEL et al., 2007; MEYENHOFER et al., 2007), reconhecimento de padrões em imagens de sensoriamento remoto (PINA, BARATA e BANDEIRA, 2006), reconhecimento de caracteres impressos e manuscritos (HA; ZIMMERMANN; BUNKE, 1998; KIM, 1999) e extração de textos em imagens digitais (RETORNAZ; MARCOTEGUI, 2007; WU; HSIEH; CHEN, 2008). O sucesso do emprego da morfologia matemática nestas e tantas outras aplicações motivou a proposta de uma extensão do Ciratefi utilizando a abordagem morfológica, denominada Ciratefi morfológico.

O Ciratefi morfológico preserva as mesmas características do Ciratefi, ou seja, é um método de casamento de padrões para imagens em níveis de cinza, invariante a transformações de similaridade, não requer segmentação prévia e também é composto por três etapas de filtragem. A diferença entre os dois métodos é que no Ciratefi morfológico as duas primeiras etapas consistem em filtros baseados em operações de dilatação e erosão. Todo o aparato restante, incluindo os parâmetros de controle, a medida de similaridade utilizada e o uso da técnica de casamento de padrões convencional na terceira etapa de filtragem são mantidos inalterados.

#### 4.5.1 Definição das operações de dilatação e erosão

As operações de dilatação e erosão de uma imagem em níveis de cinza  $I$ , pelo elemento estruturante  $B$ , com domínio  $D_B$  podem ser definidas, respectivamente, por:

$$(I \oplus B)(s, t) = \max\{I(s + x, t + y)\}, (x, y) \in D_B \quad (4.17)$$

$$(I \ominus B)(s, t) = \min\{I(s + x, t + y)\}, (x, y) \in D_B \quad (4.18)$$

Note que, diferentemente do usual, a operação de dilatação não inclui a reflexão do elemento estruturante.

#### 4.5.2 Primeira etapa de filtragem

Na primeira etapa, operações morfológicas de dilatação e erosão por elementos estruturantes circulares planos, cujos raios são determinados pelo conjunto  $l = \{r_0, r_1, \dots, r_{l-1}\}$ , são utilizadas na extração de características das imagens  $A$  e  $T$ . Isto significa que tanto a imagem 3D  $C_A$  quanto a matriz multi-escala  $C_T$ , utilizadas para determinar os pixels da imagem  $A$  que serão classificados como candidatos da primeira fase, são construídas com base nos operadores morfológicos. Desta maneira, dada uma imagem a ser analisada  $A$  e um conjunto de raios  $l$ , a imagem 3D  $C_A[x, y, k]$  é obtida como segue:

$$C_A[x, y, k] = \begin{cases} (A \oplus B_{r_{k/2}})(x, y), & \text{se } k \text{ é par} \\ (A \ominus B_{r_{(k-1)/2}})(x, y), & \text{se } k \text{ é ímpar} \end{cases} \quad (4.19)$$

onde  $0 \leq k < 2l$  (Considera-se  $2l$  uma vez que, para cada raio, são computados dois valores resultantes das operações de erosão e dilatação) e  $B_r$  é o elemento

estruturante circular com raio  $r$ . Nos experimentos realizados neste trabalho utilizou-se  $l=13$ , e  $r_0=0$  pixels,  $r_1=2$  pixels, ...,  $r_{12}=24$  pixels.

Dada a *imagem*  $T$  e o conjunto de  $n$  fatores de escala  $\{s_0, s_1, \dots, s_{n-1}\}$ ,  $T$  é escalonada por cada fator  $s_i$ , obtendo-se os padrões escalonados  $T_0, T_1, \dots, T_{n-1}$  e na seqüência, a matriz  $C_T$  com  $n$  linhas (escalas) e  $2l$  colunas (raios) é construída da seguinte forma:

$$C_T[i, k] = \begin{cases} (T_i \oplus B_{r_{k/2}})(x_0, y_0), & \text{se } k \text{ é par} \\ (T_i \ominus B_{r_{(k-1)/2}})(x_0, y_0), & \text{se } k \text{ é ímpar} \end{cases} \quad (4.20)$$

onde  $(x_0, y_0)$  é o pixel central de  $T$ ,  $0 \leq i < n$  e  $0 \leq k < 2l$

Uma vez construídas as matrizes  $C_A$  e  $C_T$  e definidos os limiares  $t_\beta$ ,  $t_\gamma$  e  $t_1$ , computam-se os pixels candidatos da primeira fase e a provável escala associada a cada um deles por meio das equações 4.9 e 4.10 (CisCorr e CisPS). O pixel  $(x, y)$  é classificado como um candidato da primeira fase quando  $\text{CisCorr}_{A,T}(x, y) \geq t_1$  para um determinado limiar  $t_1$ .

#### 4.5.3 Segunda etapa de filtragem

A segunda etapa, assim como ocorre no Ciratefi, é responsável por promover parte dos pixels candidatos da primeira para a segunda fase, descartando aqueles que não têm chances de casamento com o padrão de referência, além de associar o provável ângulo de rotação a cada pixel candidato da segunda fase. Para isto, são realizadas operações de dilatação e erosão utilizando elementos estruturantes planos (preenchidos com valor constante em todo o suporte) no formato de linhas radiais. Este processo é feito da seguinte maneira: para cada pixel candidato  $(x, y)$  da primeira fase, a matriz  $R_A$  de atributos é computada de acordo com a equação a seguir:

$$R_A[x, y, j] = \begin{cases} (A \oplus B_{\alpha_{j/2}}^\lambda)(x, y), & \text{se } j \text{ é par} \\ (A \ominus B_{\alpha_{(j-1)/2}}^\lambda)(x, y), & \text{se } j \text{ é ímpar} \end{cases} \quad (4.21)$$

onde  $\lambda = s_i r_{l-1}$  é maior raio  $r_{l-1}$  redimensionado com base na provável escala  $s_i$ , e  $0 \leq j < 2m$  ( $m$  é multiplicado por 2 uma vez que para linha radial são computados dois valores resultantes das operações de erosão e dilatação).

O vetor de atributos  $R_T$  com  $2m$  elementos é construído a partir da imagem  $T$ , considerando o conjunto de inclinações ( $\alpha_0=0, \alpha_1=10, \dots, \alpha_{m-1}=350$ ), como segue:

$$R_T[j] = \begin{cases} (T \oplus B_{\alpha_{j/2}}^{r_{l-1}})(x_0, y_0), & \text{se } j \text{ é par} \\ (T \ominus B_{\alpha_{(j-1)/2}}^{r_{l-1}})(x_0, y_0), & \text{se } j \text{ é ímpar} \end{cases} \quad (4.22)$$

onde  $(x_0, y_0)$  é o pixel central de  $T$ ,  $B_{\alpha}^{r_{l-1}}$  é o elemento estruturante no formato de linha radial com inclinação  $\alpha$  e comprimento  $r_{l-1}$  (raio do maior elemento estruturante circular utilizado na etapa anterior), e  $0 \leq j < 2m$ .

A partir dos vetores  $R_A[x,y]$  e  $R_T$ , e os limiares  $t_\beta$ ,  $t_\gamma$  e  $t_2$ , definem-se os pixels candidatos da segunda fase e a provável escala associada a cada um deles utilizando-se as equações 4.15 e 4.16 (RasCorr e RasAng), com a variável  $j$  no intervalo  $[0, 2m-1]$ . O pixel candidato  $(x,y)$  da primeira fase é promovido para candidato da segunda fase se  $RasCorr_{A,T}(x,y) \geq t_2$  para um determinado limiar  $t_2$ .

#### 4.5.4 Terceira etapa de filtragem

Nesta etapa, utilizando as informações de provável escala e provável ângulo de inclinação determinados nas etapas anteriores, os pixels candidatos da segunda fase são filtrados usando a técnica de casamento de padrões convencional. Se o valor absoluto da correlação entre  $T$  e uma sub-imagem de  $A$  centrada na posição  $(x,y)$  for maior que um determinado limiar  $t_3$ , considera-se que  $T$  foi detectado em  $A$ . O funcionamento do Ciratefi morfológico, envolvendo as três etapas de filtragem é ilustrado na Figura 4.7.

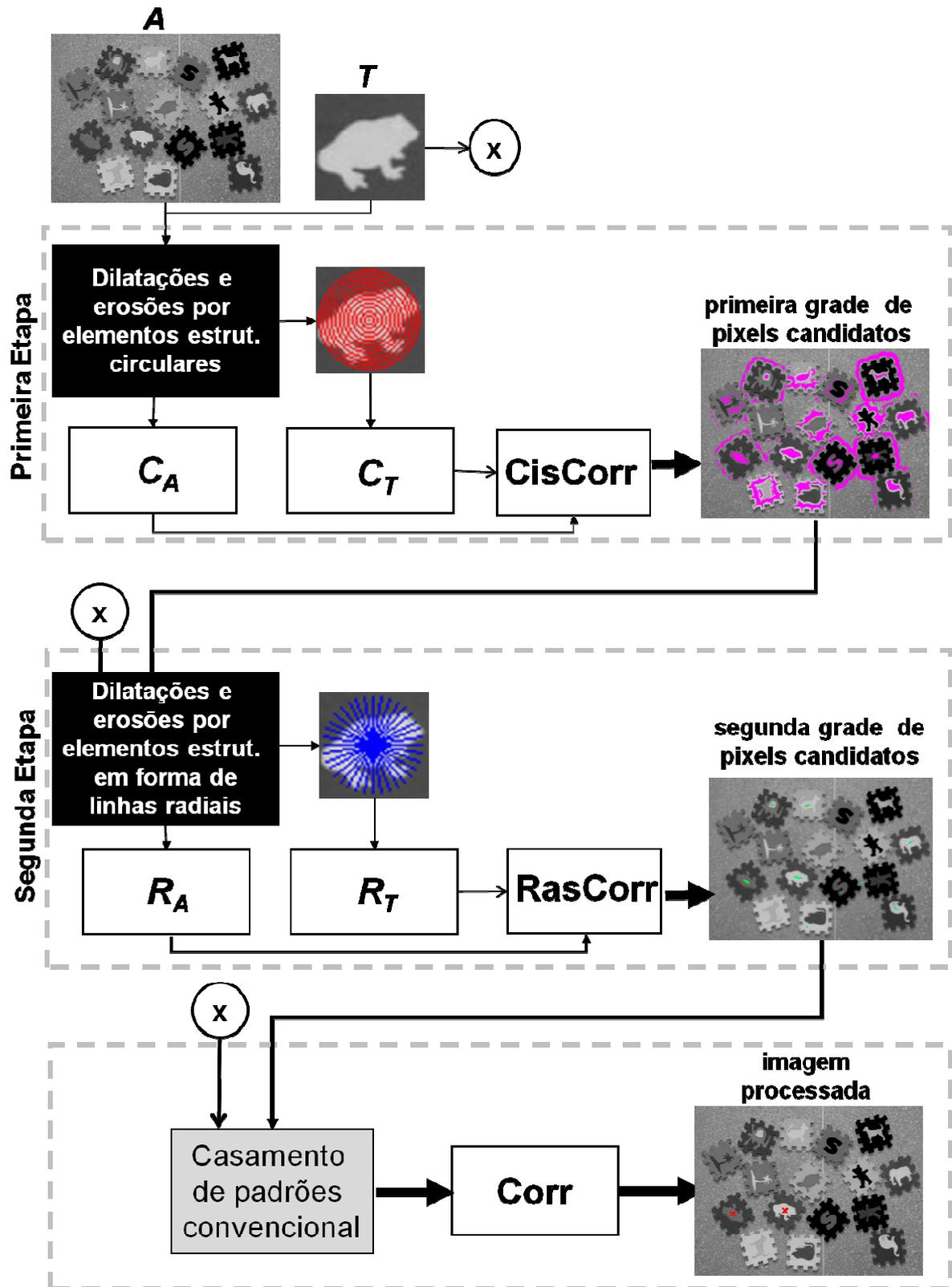


Figura 4.7 – Esquema de funcionamento das três etapas de filtragem do Ciratefi morfológico.

## 4.6 COLOR CIRATEFI

### 4.6.1 Introdução

A cor é um atributo muito importante em tarefas de reconhecimento de objetos. Isto porque usando apenas a informação de cor é possível, em muitos casos, distinguir um objeto dos demais contidos em uma imagem. A maioria das técnicas de casamento de padrões existentes foi desenvolvida para tratar apenas imagens em níveis de cinzas. Elas não levam em conta o poder da informação de cor e, por consequência, não podem ser empregadas em muitas aplicações práticas (TSAI; TSAI, 2002). Um exemplo que pode ser citado é a inspeção visual automática de alguns produtos industrializados.

Uma dificuldade que surge no reconhecimento de objetos em imagens coloridas é a constância da cor. Isto se refere ao fato das cores das superfícies dos objetos permanecerem constantes mesmo quando são vistas sob diferentes condições de iluminação. O problema é que a cor não é uma propriedade intrínseca dos objetos (MOERLAND; JURIE, 2005; SCHAEFER, 2006). Ela depende da composição da fonte luminosa, das propriedades de refletância do material que compõe o objeto e também da cor do ambiente em que o objeto se encontra. Assim, mudanças nas condições de iluminação (Figura 4.8) podem alterar as cores dos objetos afetando negativamente o desempenho dos algoritmos de reconhecimento de padrões que utilizam informação de cor (MOERLAND; JURIE, 2005). Efeitos locais de iluminação e sombreamento, como ilustrado na Figura 4.9, aumentam a dificuldade do problema descrito.

O sistema de visão humano tem a capacidade de extrair aquilo que é invariante sob mudanças de iluminação. Desta forma, mesmo que as condições de iluminação sejam alteradas, nosso cérebro é capaz de reconhecer certos padrões constantes permitindo o reconhecimento adequado dos objetos.

A constância da cor é ainda um problema em aberto e tem sido um assunto de grande importância em pesquisas na área de visão computacional nas duas últimas décadas e, por consequência, diversos mecanismos para tratar constância da cor

incluindo modelos de cores invariantes e medidas de distância entre cores têm sido propostas (CHONG; GORTLER; ZICKLER, 2008; BASCLE; BERNIER; LEMAIRE, 2006; PARK; LEE; LEE, 2008).

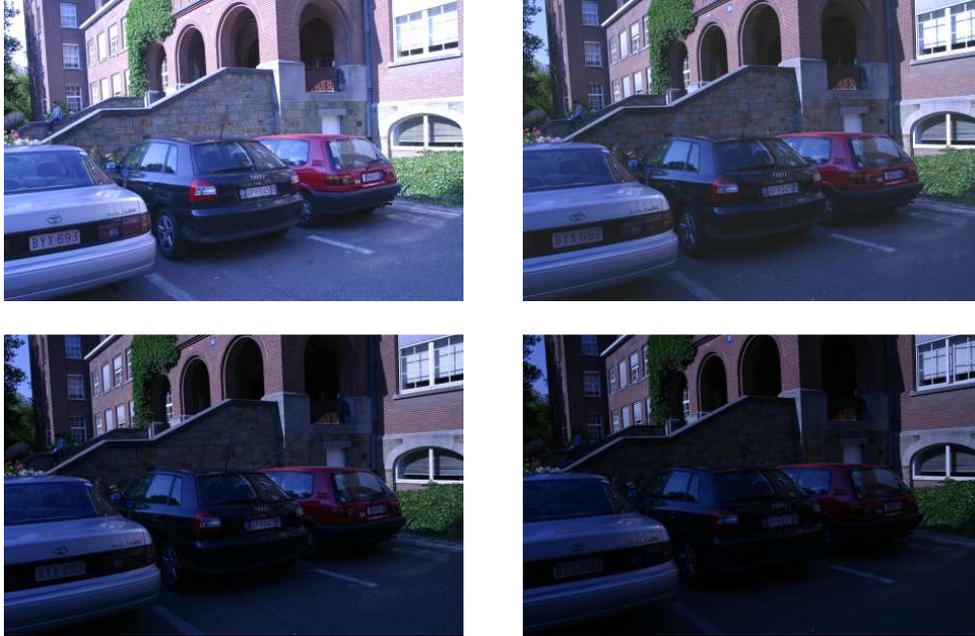


Figura 4.8 – Imagens da mesma cena sob diferentes condições de iluminação.

Extraído de (MIKOLAJCZYK; SCHMID, 2005).



Figura 4.9 – Imagens do mesmo objeto adquiridas sob diferentes ângulos de iluminação.

Extraído de (GEUSEBROEK; BURGHOUTS; SMEULDERS, 2005).

Dois algoritmos simples e bastante conhecidos para tratar o problema da constância da cor são o “*grey world*” e o “*white patch*” (WEIJER; GEVERS; GIJSENIJ, 2007; GIJSENIJ; GEVERS, 2007). No primeiro deles assume-se que, dada uma imagem com um número de variações de cor suficiente, o valor médio das componentes de cor da imagem é cinza. Isto implica que qualquer desvio da cor média distante do cinza, é causado pela cor da fonte luminosa. Desta forma, para estimar a cor da fonte luminosa basta determinar a media das componentes de cor da imagem. No segundo supõe-se que a resposta máxima em uma imagem é causada por uma

refletância perfeita. Assim, determinando a resposta máxima dos três canais de cores de todos os pixels da imagem, a cor da fonte luminosa pode ser estimada. A principal vantagem destes dois algoritmos é que eles não precisam de nenhum tipo de treinamento e não requerem muito esforço computacional para serem utilizados (GIJSENIJ; GEVERS, 2007). Apesar disso, o fato de funcionarem sob circunstâncias limitadas, torna suas utilizações em aplicações práticas comprometidas.

Swain e Ballard (1991) propuseram um método pioneiro chamado “*colour indexing*” no qual se compara a similaridade entre o histograma de atributos de cor extraído do objeto procurado e os histogramas extraídos de objetos conhecidos e armazenados em uma base. O método *colour indexing* apresenta invariância à rotação, translação e escalonamento, mas é severamente afetado por mudanças nas condições de iluminação (FUNT; BARNARD; MARTIN, 1998). Para amenizar este problema Funt e Finlayson (1995) propuseram um conjunto de descritores invariantes sob mudanças de iluminação. Entretanto, devido ao fato de tais descritores dependerem da geometria do objeto, o desempenho do processo de reconhecimento de objetos pode ser severamente afetado.

Gevers e Smeulders (1999) analisaram e avaliaram vários modelos de cores empregados em tarefas de reconhecimento de objetos considerando diferentes condições de iluminação. Eles também propuseram três novos modelos de cores derivados do sistema RGB, cada qual dotado de propriedades invariantes específicas, os quais têm sido largamente empregados em tarefas de reconhecimento de objetos.

Um importante estudo envolvendo esquemas de cores invariantes foi apresentado por Geusebroek et al. (2001). Neste estudo os autores propuseram um conjunto de características de cor invariantes sob mudanças de iluminação, sombra e ruídos, o qual, mais tarde, foi embutido no conhecido método SIFT (*Scale Invariant Feature Transform*) (LOWE, 1999) fazendo surgir descritores com a finalidade de tratar o problema da invariância de cor (ABDEL-HAKIM; FARAG, 2006; BURGHOUS; GEUSEBROEK, 2008). Atualmente, muitas das abordagens visando a invariância de cores propostas na literatura têm sido empregadas no SIFT uma vez que este tem provado ser um dos mais eficientes e populares métodos de extração de características em imagens. Em decorrência disto muitas extensões do SIFT para extração de características em imagens coloridas têm surgido nos últimos anos.

Entre elas podemos citar: CSIFT (ABDEL-HAKIM; FARAG, 2006); HSV-SIFT, Hue-SIFT, OpponentSIFT, W-SIFT, rgSIFT e *Transformed color SIFT* (VAN DE SANDE; GEVERS; SNOEK, 2008); SIFT-CCH (ANCUTI; BEKAERT, 2007); W-color-SIFT, H-color-SIFT e C-color-SIFT (BURGHOUTS; GEUSEBROEK, 2008). Estas três últimas são, na verdade, variações do método denominado color SIFT, proposto pelos mesmos autores e que emprega as características invariantes desenvolvidas por Geusebroek et al. (2001).

Há muitos outros métodos propostos na literatura. Alguns deles, por exemplo os descritos em (MOERLAND; JURIE, 2002; BASCLE; BERNIER; LEMAIRE, 2006; GEVERS; SMEULDERS, 1999) requerem algum tipo de informação prévia da imagem a ser analisada e como consequência têm seu uso comprometido no que tange as aplicações práticas.

A despeito da quantidade de estudos envolvendo constância de cor, alguns deles têm questionado a eficácia dos métodos propostos para este fim. Funt et al. (1998) investigaram a efetividade de vários métodos empregados em tarefas de reconhecimento de objetos. Entre eles, os algoritmos *grey world*, *white patch* e o *colour indexing*. Neste estudo eles concluíram que todos os métodos investigados não são confiáveis para serem empregados no reconhecimento de objetos em imagens coloridas. Mais recentemente, Schaefer (2006) investigou a utilidade dos invariantes de cor em tarefas de recuperação de imagens incluindo os algoritmos *grey world*, *colour indexing* e os três modelos de cor propostos por Gevers e Smeulders (1999). A conclusão foi que os métodos investigados nem sempre são úteis para o propósito de recuperação de imagens e também que, para melhorar o desempenho de tais métodos, alguma informação a respeito do domínio da aplicação é necessária.

Algoritmos para reconhecimento de padrões em imagens coloridas geralmente utilizam histogramas baseados em atributos de cor ignorando a informação espacial da imagem (FUNT; CARDEI, 2000). Por este motivo, tais algoritmos têm sido mais adequados para aplicações envolvendo recuperação de imagens em bases (*image retrieval*) a partir de características de cor ou aplicações considerando o reconhecimento de objetos em imagens com fundo constante (TSAI; TSAI, 2002). Um problema que normalmente surge em decorrência de não se levar em conta a

informação espacial é que duas imagens com composição de cores parecida podem ser consideradas similares mesmo sendo de contextos diferentes.

Tsai e Tsai (2002) apresentaram um algoritmo de casamento de padrões nomeado “*color ring-projection*” para detecção de padrões em imagens coloridas com fundos complexos. Eles compararam o desempenho do algoritmo proposto usando oito diferentes características de cor derivadas dos espaços de cores HSI e CIELAB. Eles também fizeram experimentos considerando a detecção de padrões em imagens adquiridas sob diferentes condições de iluminação. A principal desvantagem do *color ring-projection* é a falta de invariância à escala e a possibilidade de se obter um grande número de falsos positivos.

Da mesma forma que existem inúmeras abordagens para tratamento do problema da constância da cor, também existem muitas abordagens para mensuração da distância entre cores, tema que tem recebido muita atenção dos pesquisadores da área de visão computacional. Duas medidas de distância amplamente utilizadas neste contexto são a distância Euclidiana e o erro angular (GIJSENIJ; GEVERS; LUCASSEN, 2008).

#### 4.6.2 Descrição do algoritmo Color Ciratefi

O método color Ciratefi, proposto para reconhecimento de padrões em imagens coloridas e descrito nesta seção é uma extensão do Ciratefi e, portanto, herda as suas principais características, ou seja, trata-se de um método de casamento de padrões composto por três etapas de filtragem em cascata, é livre de segmentação e invariante sob translação, rotação e escala uniforme. Além disso, como pode ser visto na seção 5.7, ele apresentou bom desempenho no caso de imagens adquiridas sob diferentes condições de iluminação como as ilustradas na figura 4.8.

Dadas duas imagens coloridas  $A$  e  $T$  seu objetivo é encontrar as instâncias de  $T$  em  $A$ . Estas instâncias podem ser afetadas por transformações geométricas (rotação, escala, translação e pequenas mudanças de ponto de vista) e fotométricas (variações uniformes nas condições de luminosidade).

As principais diferenças entre os métodos Ciratefi e color Ciratefi são as seguintes:

- a) Como as imagens  $A$  e  $T$  são coloridas, cada elemento dos vetores de atributos  $\mathbf{x}$  e  $\mathbf{y}$  é composto por três valores que identificam uma cor no espaço CIELAB;
- b) Para as três etapas de filtragem, a medida de similaridade **Corr** (equação 4.4) é substituída pela medida de similaridade **Sim** apresentada na equação 4.30;

#### 4.6.3 Espaço de cor e medida de similaridade utilizados

Um espaço ou sistema de cores pode ser definido como um modelo matemático para formalizar a descrição de cores. Esta descrição é feita, na maioria dos casos, através de um conjunto de três valores por meio dos quais é possível identificar uma cor em um sistema de coordenadas. Entre os espaços de cores mais utilizados no processamento de imagens coloridas estão o RGB (padrão para monitores), CMYK (padrão para impressoras), YIQ, YUV e YCbCr (empregados na transmissão de imagens de televisão e de vídeo), HSV/HSI (para emulação da percepção visual humana), XYZ (conjunto padrão das cores primárias),  $L^*u^*v^*$  e  $L^*a^*b^*$  (perceptualmente uniforme) (GEVERS AND SMEULDERS, 1999; KERMINEN; GABBOUJ, 2000; ANCUTI; BEKAERT, 2007). Apesar da grande variedade de espaços de cores existentes, muitos são derivados do RGB que é o mais popular entre todos eles (GEVERS; SMEULDERS, 1999).

A seguir são descritos o sistema de cor CIELAB ( $L^*a^*b^*$ ) e a medida de similaridade empregados no color Ciratefi. Cabe ressaltar que eles foram escolhidos levando-se em conta fatores como a simplicidade e a correlação com a percepção humana das cores (RUBNER; GUIBAS; TOMASI, 1997; GIJSENIJ; GEVERS; LUCASSEN, 2008).

##### 4.6.3.1 O espaço de cor CIELAB

Embora o espaço de cores RGB seja muito popular e amplamente utilizado para representar cores, ele não é perceptualmente uniforme (TSAI; TSAI, 2002). Em um espaço de cor perceptualmente uniforme quando se altera o valor de um dos parâmetros que descrevem uma cor não se altera os outros dois perceptualmente.

Além disso, duas cores igualmente distantes no espaço de cor são percebidas como igualmente distantes pelo sistema visual humano. Assim, utilizando critérios baseados na percepção da cor pelo olho humano, a Comissão Internacional de Iluminação, (CIE - Commission Internationale d'Eclairage) desenvolveu, em 1976, o sistema de cores denominado CIELAB também conhecido como CIE  $L^*a^*b^*$  ou simplesmente Lab. Este espaço de cores é tido como um dos mais completos sistemas usados para descrever as cores visíveis pelo olho humano.

No espaço CIELAB, cuja representação gráfica é ilustrada na Figura 4.10, uma cor qualquer é representada por três valores ou tri-estímulos ( $L^*$ ,  $a^*$  e  $b^*$ ). O primeiro valor ( $L^*$ ) representa a luminosidade e os outros dois ( $a^*$  e  $b^*$ ) representam a cromaticidade. O valor de  $L^*$  varia de 0 (preto) a 100 (branco). O segundo valor ( $a^*$ ) indica a posição entre as cores oponentes verde e vermelho (valores negativos de  $a^*$  indicam a cor verde enquanto valores positivos indicam a cor vermelho). O terceiro valor ( $b^*$ ) indica a posição entre as cores oponentes azul e amarelo (valores negativos de  $b^*$  indicam a cor azul enquanto valores positivos indicam a cor amarelo).

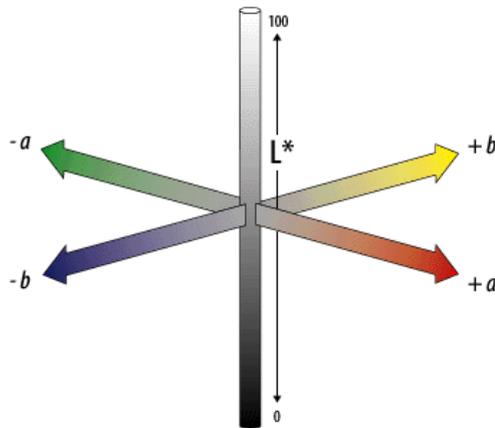


Figura 4.10 – Representação gráfica do espaço de cores CIELAB. Extraído de (ADOBE, 2009).

A conversão do espaço de cores RGB para o espaço CIELAB é feita usando uma transformação intermediária para o espaço XYZ, conforme mostrado na equação 4.23 (GIJSENIJ; GEVERS; LUCASSEN, 2008). Os demais passos envolvidos na conversão são descritos pelas equações 4.24 a 4.27 a seguir.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,4125 & 0,3576 & 0,1804 \\ 0,2127 & 0,7152 & 0,0722 \\ 0,0193 & 0,1192 & 0,9502 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4.23)$$

$$L^* = 116 f\left(\frac{Y}{Y_0}\right) - 16 \quad (4.24)$$

$$a^* = 500 \left[ f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right) \right] \quad (4.25)$$

$$b^* = 200 \left[ f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right) \right] \quad (4.26)$$

$$f(t) = \begin{cases} t^{1/3}, & t > 0,008856 \\ 7,787t + 16/116, & \text{caso contrário} \end{cases} \quad (4.27)$$

onde:  $X_0, Y_0, Z_0$  é o conjunto de valores que representam o branco de referência. Um conjunto padrão definido pelo CIE e amplamente utilizado é o D65 cujos valores são: 0,313, 0,329 e 1,0. O espaço XYZ é um padrão também estabelecido pela CIE para descrever as cores visíveis através de 3 coeficientes (X, Y e Z) que representam as porcentagens das cores primárias (vermelho, verde e azul) existentes em uma cor.

#### 4.6.3.2 *Medida de similaridade proposta*

A questão da similaridade envolvendo imagens coloridas tem apresentado alguma complexidade, principalmente devido à questão da constância da cor, e por isso tem sido tema de diversas pesquisas em visão computacional. Muitos algoritmos para avaliação de similaridade que levam em conta o problema da constância de cor usam apenas as componentes de cromaticidade ignorando a componente de luminosidade (FUNT; BARNARD; MARTIN, 1998; GIJSENIJ; GEVERS; LUCASSEN, 2008). Apesar disso, é sabido que o canal de luminosidade carrega parte da

informação acerca da imagem e, portanto, não pode ser desprezado. A distância Euclidiana ou outra medida derivada dela é tipicamente utilizada como medida de similaridade em espaços de cores perceptualmente uniforme como é o caso do CIELAB (KERMINEEN; GABBOUJ, 2000; GIJSENIJ; GEVERS; LUCASSEN, 2008; MOERLAND; JURIE, 2005; CHONG; GORTLER; ZICKLER, 2008).

A medida de similaridade proposta (*Sim*) é uma composição ponderada das similaridades de cromaticidade e luminosidade e procura avaliar a semelhança perceptual entre vetores de cores  $\mathbf{x}$  e  $\mathbf{y}$  no espaço CIELAB, de forma invariante à mudanças uniformes nas condições de luminosidade. No caso do color Ciratefi, os elementos dos dois vetores são obtidos calculando as médias das cores em círculos e linhas radiais. Sejam  $\{x_1, x_2, \dots, x_n\}$  e  $\{y_1, y_2, \dots, y_n\}$  os elementos dos vetores  $\mathbf{x}$  e  $\mathbf{y}$ , respectivamente. Cada elemento  $x_i$  e  $y_i$  possui as componentes L,  $a^*$  e  $b^*$  e são denotados, respectivamente, por  $x_{iL}, x_{ia}, x_{ib}$  e  $y_{iL}, y_{ia}, y_{ib}$ . Para similaridade de cromaticidade ( $S_c$ ) empregou-se a distância Euclidiana das componentes  $a^*$  e  $b^*$ :

$$S_c(\mathbf{x}, \mathbf{y}) = 1 - \sqrt{\frac{\sum_{i=1}^{N_v} (x_{ia} - y_{ia})^2 + (x_{ib} - y_{ib})^2}{2 \cdot 200^2 \cdot N_v}} \quad (4.28)$$

onde  $N_v$  é o número de elementos dos vetores  $\mathbf{x}$  e  $\mathbf{y}$  e 200 é a maior diferença nas componentes  $a^*$  e  $b^*$  (variam de -100 a 100).

A distância Euclidiana denota a diferença de cromaticidade entre os vetores  $\mathbf{x}$  e  $\mathbf{y}$ . Para obter a medida de similaridade, a distância é subtraída de um. Para similaridade de luminosidade ( $S_L$ ) foi empregado o coeficiente de correlação (equação 4.29), uma vez que é invariante ao brilho e contraste.

$$S_L(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{N_v} (x_{iL} - \bar{\mathbf{x}})(y_{iL} - \bar{\mathbf{y}})}{\sqrt{\sum_{i=1}^{N_v} (x_{iL} - \bar{\mathbf{x}})^2} \sqrt{\sum_{i=1}^{N_v} (y_{iL} - \bar{\mathbf{y}})^2}} \quad (4.29)$$

onde  $\bar{\mathbf{x}}$  e  $\bar{\mathbf{y}}$  são, respectivamente, as médias de  $\mathbf{x}$  e  $\mathbf{y}$ . A medida de similaridade proposta é:

$$\text{Sim}(X, Y) = [S_C(X, Y)]^\delta \cdot [S_L(X, Y)]^\eta \quad (4.30)$$

onde  $\delta$  e  $\eta$  são os pesos atribuídos às similaridades da cromaticidade e da luminosidade.

Como pode ser observado na equação 4.30, a importância de cada termo depende do peso que lhe for atribuído. Isso propicia aumentar ou diminuir, de acordo com a necessidade, a importância dada para as informações de cromaticidade e de luminosidade, de forma independente. Por exemplo, se  $\delta=0$  considera-se apenas a correlação dos níveis de cinzas das imagens. Por outro lado, se  $\eta=0$  apenas a informação de cromaticidade é levada em conta na avaliação da similaridade. O fato de Sim ser resultante do produto das similaridades de cromaticidade ( $S_C$ ) e luminosidade ( $S_L$ ) força a necessidade de haver compromisso entre estes dois termos. Assim, caso um dos termos seja nulo, Sim também será nula.

## 5 RESULTADOS EXPERIMENTAIS

Este capítulo é dedicado à descrição dos experimentos com os métodos propostos e às análises comparativas com os métodos SIFT e Easymatch. Primeiro, são apresentados os resultados obtidos pelo Ciratefi envolvendo 4 conjuntos imagens com diferentes características e uma análise da sua sensibilidade com relação aos parâmetros de controle. Depois, são apresentados alguns resultados obtidos com o Ciratefi morfológico. Em seguida, comparam-se os resultados de Ciratefi com os resultados obtidos pelos métodos SIFT e EasyMatch, considerando os mesmos conjuntos de imagens. Na seqüência, é feito um resumo das análises comparativas entre os métodos Ciratefi, SIFT e Easymatch e apresenta-se uma proposta para otimização do Ciratefi. Por último, comparam-se os resultados de color Ciratefi com os obtidos por color SIFT.

### 5.1 RESULTADOS EXPERIMENTAIS COM CIRATEFI

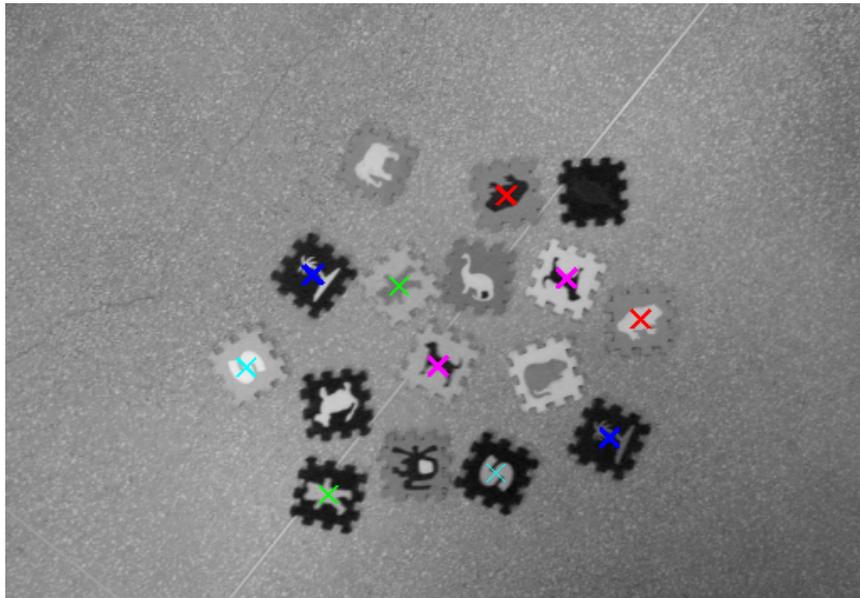
Para avaliação do Ciratefi foram realizados experimentos utilizando 175 imagens, divididas em 4 conjuntos I, II, III e IV. Nestas imagens buscou-se por um total de 1363 padrões. O conjunto I é composto por 70 imagens de brinquedos espalhados pelo chão, todas com tamanho de 512×384 pixels. O conjunto II é composto por 60 imagens, com tamanhos variando de 96×94 pixels até 698×461 pixels, contendo símbolos do McDonald's<sup>®</sup>. O conjunto III é formado por 15 imagens de sensoriamento remoto, todas com o tamanho de 515×412 pixels. Por fim, o conjunto IV é composto por 30 imagens, com tamanhos de 400×400 e 345×469, contendo cartões *Disney*, oriundos de um “Jogo da memória”, sobre fundos complexos. Os padrões foram extraídos de uma única imagem de cada conjunto. Nas 145 imagens, instâncias dos padrões aparecem em diferentes posições, rotações e escalas e também com diferentes níveis de brilho e contraste. Além disso, algumas imagens do conjunto 4 são afetadas por oclusão parcial.

### 5.1.1 Experimentos com as imagens do conjunto I

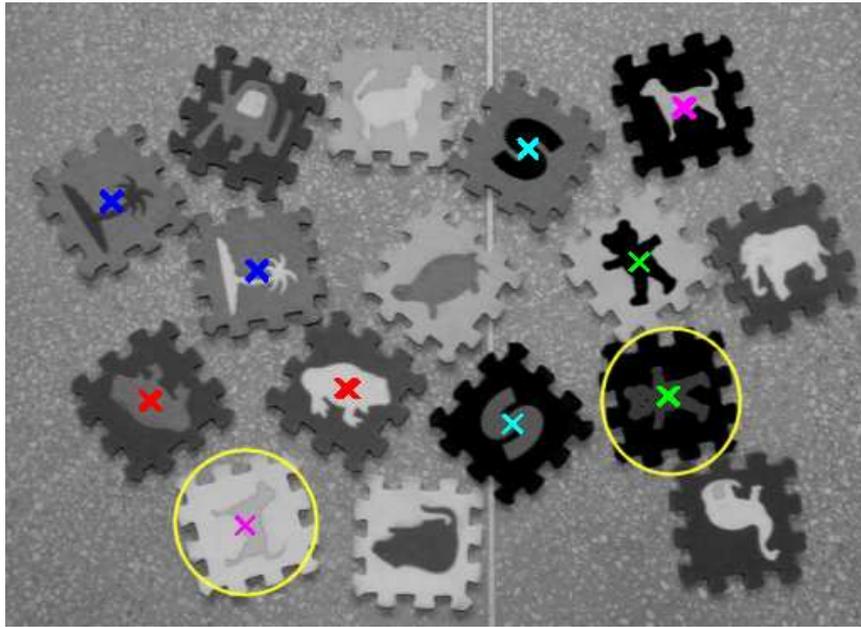
Neste primeiro experimento, o objetivo foi procurar por 5 padrões: sapo, cachorro, palmeira, urso e letra-S (Figura 5.1) nas 70 imagens dos brinquedos. Em cada uma das imagens aparecem 10 instâncias dos padrões procurados (2 instâncias de cada padrão). A Figura 5.2 ilustra duas imagens do conjunto I processadas por Ciratefi onde os padrões detectados estão marcados com “x” em diferentes cores, cada cor indicando uma categoria.



Figura 5.1 – Padrões a serem procurados nas imagens do conjunto I. Cada imagem tem 51×51 pixels.



(a) Resultado da detecção por Ciratefi onde as instâncias dos padrões aparecem em escala reduzida



(b) Instâncias dos padrões pouco visíveis assinaladas por círculos amarelos

Figura 5.2 – Resultados da detecção das instâncias dos 5 padrões por Ciratefi.

Como pode ser visto na Tabela 5.1, todas as 700 instâncias dos padrões foram corretamente detectadas, sem nenhum falso positivo ou falso negativo, mesmo nos casos em que o padrão é pouco visível, por exemplo os padrões cachorro e urso, assinalados por círculos amarelos na imagem da Figura 5.2b.

Tabela 5.1. Resultados do experimento com Ciratefi envolvendo as imagens do conjunto I.

Padrão	Casamentos possíveis	Deteções corretas	Falsos positivos	Falsos negativos
Sapo	140	140	0	0
Urso	140	140	0	0
Palmeira	140	140	0	0
Letra-S	140	140	0	0
Cachorro	140	140	0	0
<b>Total</b>	<b>700</b>	<b>700</b>	<b>0</b>	<b>0</b>

Para este experimento empregaram-se os seguintes parâmetros:  $l=13$ ,  $m=36$  ( $\alpha_0=0$ ,  $\dots$ ,  $\alpha_{35}=350$ ),  $n=10$  ( $s_0=0,4$ ,  $\dots$ ,  $s_9=1,3$ ),  $t_1=0,95$ ,  $t_2=0,9$ ,  $t_3=0,9$ ,  $t_\beta=0,1$  e  $t_\gamma=1$ .

### 5.1.2 Experimentos com as imagens do conjunto II

No segundo experimento o objetivo foi detectar o símbolo do McDonald's® (Figura 5.3a) nas 60 imagens do conjunto II, as quais foram adquiridas de diferentes lugares e objetos. O símbolo procurado aparece 116 vezes nas 60 imagens. As Figuras 5.3b e 5.3c apresentam duas imagens processadas onde cada objeto detectado está marcado com um "x" vermelho.

Este experimento apresentou apenas um caso de falso positivo e 2 casos de falsos negativos. Um dos casos de falsos negativos, assinalado por um círculo amarelo na Figura 5.3c, pode ter sido causado pelo brilho de fundo da região onde se localiza a instância do padrão.



(a) Padrão



(b) Detecções corretas



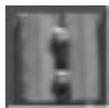
(c) Caso de falso negativo assinalado com um círculo amarelo

Figura 5.3 – Resultados da detecção do símbolo do McDonald's por Ciratefi.

### 5.1.3 Experimentos com as imagens do conjunto III

O experimento com as imagens do conjunto III consistiu na detecção de um prédio em forma de “H” em 15 imagens de sensoriamento remoto extraídas do Google Earth®. O padrão procurado (Figura 5.4a) possui o tamanho de 41×41 pixels e instâncias dele aparece 187 vezes nas 15 imagens.

Neste experimento foram computados 18 casos de falsos positivos e 16 de falsos negativos, causados principalmente pela existência de sombras, oclusões e borrimento nas imagens. A Figura 5.4b mostra um caso de falso negativo causado por oclusão parcial e a Figura 5.4c mostra uma imagem, na qual todos os prédios foram corretamente detectados.



(a) Padrão



(b) Caso de falso negativo assinalado por um círculo amarelo



(c) Prédios corretamente detectados

Figura 5.4 – Resultado da detecção do prédio em forma de “H”, em uma imagem de sensoriamento remoto, por Ciratefi.

#### 5.1.4 Experimentos com as imagens do conjunto IV

Nos experimentos com o conjunto IV procurou-se por 360 instâncias de 12 padrões, ilustrados na Figura 5.5. As imagens deste conjunto foram divididas em 3 subconjuntos (A, B e C) com 10 imagens cada e com as seguintes características: A) Imagens com tamanho 345×469 nas quais as instâncias dos padrões aparecem transladadas e rotacionadas; B) Imagens com tamanho 400×400 onde as instâncias dos padrões aparecem transladadas, rotacionadas e escalonadas; e C) Imagens com tamanho 345×469 nas quais as instâncias dos padrões aparecem transladadas, rotacionadas e com oclusão parcial (até 20% de oclusão). Exemplos de imagens deste conjunto são mostrados na Figura 5.6.

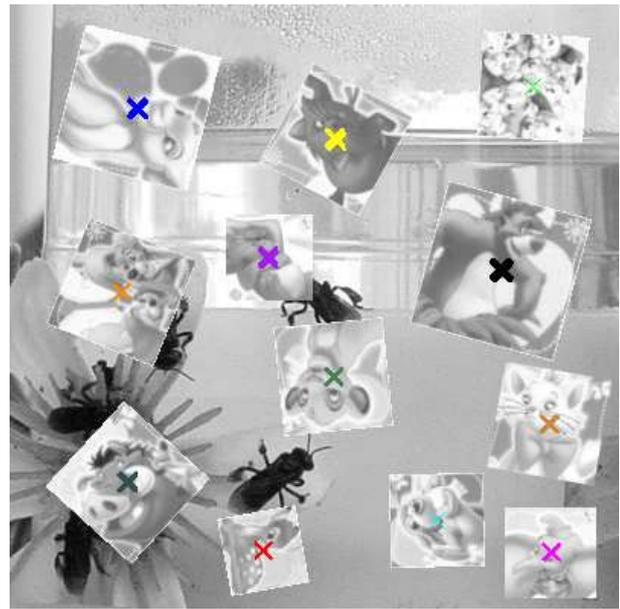


Figura 5.5 – Os 12 Padrões a serem detectados nas 30 imagens do conjunto IV. Cada padrão tem 71×71 pixels.

Na Figura 5.5 é possível observar que os padrões procurados apresentam riqueza de detalhes (texturas/contrastos), ao contrário daqueles do conjunto I (brinquedos espalhados pelo chão). Também é evidente que a utilização de fundos complexos nas imagens e a existência de oclusão parcial aumentam o grau de dificuldade dos métodos no processo de detecção dos padrões.



(a) Resultado da detecção em uma imagem do subconjunto A (translação+rotação)



(b) Resultado da detecção em uma imagem do subconjunto B (translação+rotação+ escala)



(c) Resultado da detecção em uma imagem do subconjunto C (translação+rotação+occlusão). Os casos de falsos negativos estão assinalados por círculos amarelos

Figura 5.6 – Resultados da detecção dos padrões em imagens do conjunto IV.

Como podemos observar na imagem 5.6c, o efeito de oclusão é caracterizado pela alteração de pequenas regiões retangulares da imagem, escolhidas aleatoriamente.

Os resultados do Ciratefi nos experimentos com as imagens do conjunto IV estão sumarizados na Tabela 5.2 a partir da qual pode-se verificar que a maior parte dos casos de falsos positivos e falsos negativos concentrou-se nos experimentos com as imagens do subconjunto C, afetadas por oclusão parcial. Isso já era esperado uma vez que o Ciratefi é um algoritmo baseado em atributos globais extraídos da região do objeto, diferente de outros métodos como o SIFT que extrai atributos locais.

Tabela 5.2. Resultados do experimento com Ciratefi envolvendo as imagens do conjunto IV.

Subconjunto	Casamentos possíveis	Detecções corretas	Falsos positivos	Falsos negativos
<b>A</b> (translação+rotação)	120	120	1	0
<b>B</b> (translação+rotação+escala)	120	118	5	2
<b>C</b> (translação+rotação+occlusão parcial)	120	89	2	31
<b>Total</b>	<b>360</b>	<b>327</b>	<b>8</b>	<b>33</b>

Os resultados dos experimentos realizados com Ciratefi estão sumarizados na Tabela 5.3, a qual permite avaliar seu desempenho. Ainda com base na Tabela 5.3, é possível verificar que a maior parte dos casos de falsos negativos concentrou-se nas imagens do conjunto IV.

Tabela 5.3. Resumo dos experimentos realizados com Ciratefi.

Conjunto	Casamentos possíveis	Detecções corretas	Falsos positivos	Falsos negativos	Taxa de acertos(%)
I – Brinquedos	700	700	0	0	<b>100,0</b>
II – McDonald's®	116	114	1	2	<b>98,3</b>
III – Sensoriamento remoto	187	171	18	16	<b>91,4</b>
IV – Cartões jogo da memória	360	327	8	33	<b>90,8</b>
<b>Total</b>	<b>1.363</b>	<b>1.312</b>	<b>27</b>	<b>51</b>	<b>96,3</b>

Os resultados obtidos pelo Ciratefi envolvendo os 4 conjuntos de imagens são comparados com os resultados obtidos por SIFT e EasyMatch adiante na seção 5.5.

### 5.1.5 Análise da sensibilidade do Ciratefi com relação aos parâmetros de controle

Como vimos, o funcionamento do algoritmo Ciratefi depende da escolha dos seguintes parâmetros: número de ângulos ( $m$ ), número de fatores de escala ( $n$ ), número de círculos ( $l$ ), limiar utilizado por Cifi para definir os candidatos da primeira fase ( $t_1$ ), limiar utilizado por Rafi para definir os candidatos da segunda fase ( $t_2$ ), limiar utilizado por Tefi para definir os pontos de casamento ( $t_3$ ), limiar de correção de contraste ( $t_\beta$ ) e o limiar de correção de brilho ( $t_\gamma$ ). Para avaliar a sensibilidade do método com relação a estes parâmetros, foram realizados experimentos, utilizando as imagens do conjunto I (brinquedos). Em cada experimento, executou-se o Ciratefi para localizar as 140 instâncias do padrão sapo que aparecem nas 70 imagens, variando um ou mais parâmetros e fixando-se os demais. As Tabelas 5.4 a 5.7 apresentam os resultados dos experimentos relativos à variação dos parâmetros. Os parâmetros fixos são aqueles descritos na primeira linha de cada tabela. Em todos os experimentos foram utilizados 10 fatores de escala ( $s_0=0,4, s_1=0,5, \dots, s_5=1,3$ ). O tempo médio de processamento é o tempo total gasto para fazer a detecção das 140 instâncias do padrão sapo dividido pelo total de imagens processadas.

Tabela 5.4. Sensibilidade do Ciratefi com relação ao número de círculos  $l$ .

$m = 36, t_1 = 0,95, t_2 = 0,9, t_3 = 0,9, t_\beta = 0,1, t_\gamma = 1$				
Número de círculos $l$	Total de falsos positivos	Total de falsos negativos	Qtd. Média de pixels candidatos da primeira fase	Tempo médio de processamento
05	0	2	37.316	22,6s
10	0	1	31.211	21,6s
15	0	0	21.447	19,4s
20	0	0	33.532	27,6s
25	0	0	34.297	30,1s

A Tabela 5.4 mostra que o número de círculos não tem grande influência no resultado final. Entretanto, verifica-se que se o número de círculos for muito

pequeno, aumentam-se as chances de erros pela existência de falsos negativos. Por outro lado, se o número de círculos for muito grande, aumenta-se a quantidade de pixels candidatos da primeira fase acarretando em maior tempo de processamento sem, no entanto, apresentar ganhos no processo de detecção. Assim, uma escolha adequada do número de círculos pode melhorar o desempenho do algoritmo porque minimiza a quantidade de pixels candidatos na primeira fase. Por exemplo, a Tabela 5.4 sugere  $l=15$  para o conjunto de imagens I.

Tabela 5.5. Sensibilidade do Ciratefi com relação ao número de linhas radiais  $m$ .

$l = 13, t_1 = 0,95, t_2 = 0,9, t_3 = 0,9, t_\beta = 0,1, t_\gamma = 1$				
Núm. de linhas radiais $m$	Total de falsos positivos	Total de falsos negativos	Qtd. Média de pixels candidatos da segunda fase	Tempo médio de processamento
08	0	91	152	14,3s
15	0	41	22	16,2s
20	0	11	11	17,8s
30	0	0	16	22,2s
40	0	0	17	25,7s

A partir dos resultados mostrados na Tabela 5.5 pode-se inferir que um número pequeno de linhas radiais (por exemplo  $m \leq 20$ ) pode produzir falsos negativos. A razão disto é que pixels com chances reais de casamento são eliminados pela filtragem Rafi. Por outro lado, um número grande de linhas radiais (por exemplo  $m > 30$ ) aumenta o tempo de processamento da filtragem, sem auferir ganhos no processo de detecção. De acordo com os dados apresentados na Tabela 5.5, 30 é o número ideal de linhas radiais para processar o conjunto de imagens I. Esta tabela mostra ainda um dado interessante que é a reduzida quantidade de pixels candidatos da segunda fase, fator que torna possível o uso da técnica de casamento de padrões convencional na terceira etapa de filtragem.

Tabela 5.6. Sensibilidade do Ciratefi com relação aos limiares  $t_1$ ,  $t_2$  e  $t_3$ .

$l = 13, m = 36, t_\beta = 0,1, t_\gamma = 1$							
Limiares $t_1, t_2, t_3$	Total de falsos positivos	Total de falsos negativos	Tempo médio de processa- mento	Limiares $t_1, t_2, t_3$	Total de falsos positivos	Total de falsos negativos	Tempo médio de processa- mento
0,50 0,50 0,50	266.258	0	145,6s	0,90 0,75 0,90	0	0	32,4s
0,50 0,50 0,75	13.159	0	145,8s	0,90 0,90 0,90	0	0	28,2s
0,50 0,50 0,90	0	0	145,9s	0,90 0,90 0,95	0	2	28,4s
0,50 0,50 0,95	0	2	145,7s	0,90 0,95 0,90	0	4	28,2s
0,75 0,75 0,50	41.526	0	39,3s	0,95 0,75 0,75	4.960	0	27,1s
0,75 0,75 0,75	5.950	0	38,8s	0,95 0,75 0,95	0	2	26,9s
0,75 0,75 0,90	0	0	38,9s	0,95 0,90 0,90	0	0	23,5s
0,75 0,90 0,90	0	0	34,7s	0,95 0,95 0,95	0	8	23,5s
0,75 0,95 0,95	0	8	33,9s	0,99 0,90 0,90	0	0	14,9s

Com base na Tabela 5.6 pode-se perceber que a escolha inadequada de  $t_3$  pode produzir falsos negativos ou falsos positivos, e que a escolha de  $t_1$  e  $t_2$  não é crítica desde que seus valores não sejam muito altos ( $t_1 > 0,99$  e  $t_2 > 0,9$ ) a ponto de descartar pixels com chances reais de casamento. Também é possível perceber que o processo de detecção não apresentou erro algum quando  $t_3 = 0,9$ , para qualquer  $t_1 \leq 0,99$  e  $t_2 \leq 0,9$ . No entanto, cabe salientar que valores pequenos para  $t_1$  e  $t_2$  deixam o algoritmo lento.

Tabela 5.7. Sensibilidade do Ciratefi com relação aos limiares  $t_\beta$  e  $t_\gamma$ .

$l = 13, m = 36, t_1 = 0,95, t_2 = 0,9, t_3 = 0,9$							
Limiares $t_\beta, t_\gamma$	Total de falsos positivos	Total de falsos negativos	Tempo médio de processa- mento	Limiares $t_\beta, t_\gamma$	Total de falsos positivos	Total de falsos negativos	Tempo médio de processa- mento
0,10 , 0,10	0	137	13,0s	0,50 , 0,10	0	137	12,6s
0,10 , 0,50	0	65	21,0s	0,50 , 0,50	0	129	13,3s
0,10 , 1,00	0	0	23,7s	0,50 , 1,00	0	64	15,2s
0,20 , 0,50	0	65	20,0s	0,75 , 0,10	0	137	12,3s
0,20 , 0,75	0	0	21,1s	0,75 , 0,50	0	137	12,3s
0,20 , 1,00	0	0	21,7s	0,75 , 1,00	0	132	12,8s
0,25 , 0,10	0	137	12,8s	1,00 , 0,10	0	140	12,1s
0,25 , 0,50	0	67	18,3s	1,00 , 0,50	0	140	12,1s
0,25 , 1,00	0	2	19,8s	1,00 , 1,00	0	140	12,1s

Os resultados apresentados na Tabela 5.7 mostram que valores de  $t_\beta$  maiores que 0,2 ou valores de  $t_\gamma$  menores que 0,75 fazem com que o processo de detecção produza falsos negativos. Contudo, é possível notar que há intervalos de valores para estes parâmetros que não causam erro algum. Considere, por exemplo os intervalos  $0,1 \leq t_\beta \leq 0,2$  e  $0,75 \leq t_\gamma \leq 1,0$ . Veja que, neste caso em particular, mesmo valores muito pequenos de  $t_\beta$  não produzem falsos negativos. Isto ocorreu porque não foram percebidas regiões com níveis de cinzas constantes nas imagens do conjunto I. Contudo, é importante destacar que, na maioria dos casos, valores muito pequenos de  $t_\beta$  podem produzir falsos negativos ou divisão por zero.

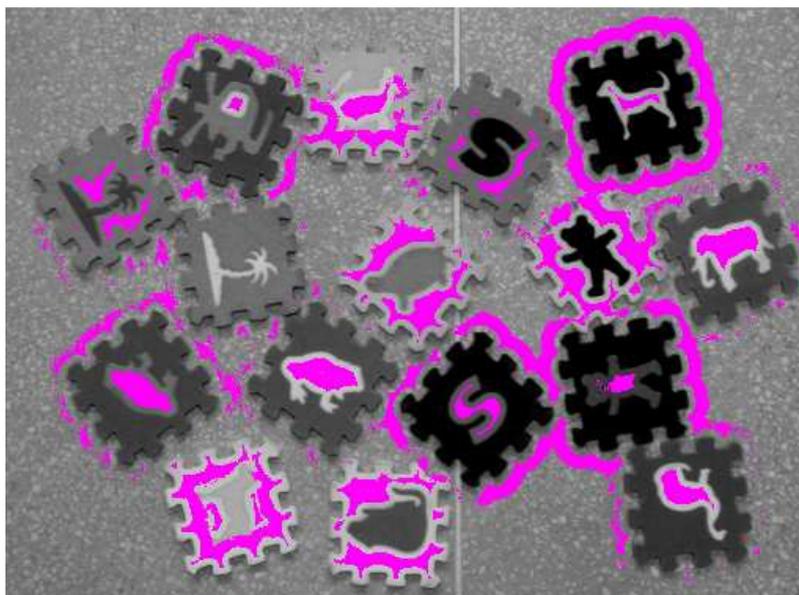
### 5.1.6 Proposta para otimização do Ciratefi

O funcionamento do Ciratefi depende da escolha de vários parâmetros. Conforme descrito na subseção 5.1.5, para alguns deles, por exemplo os limiares de correção de contraste ( $t_b$ ) e de brilho ( $t_l$ ) há grandes intervalos de valores para os quais não são produzidos erros, o que permite maior flexibilidade nas suas escolhas. Outros como o número de círculos ( $l$ ), o número de linhas radiais ( $m$ ) e os limiares  $t_1$  e  $t_2$  não são críticos, mas escolha inadequada destes parâmetros pode produzir falsos positivos ou falsos negativos bem como deixar o algoritmo lento. Já o limiar  $t_3$  está diretamente relacionado com a precisão do algoritmo. Diante disso, fica evidente a importância de mecanismos que permitam otimizar o funcionamento do algoritmo.

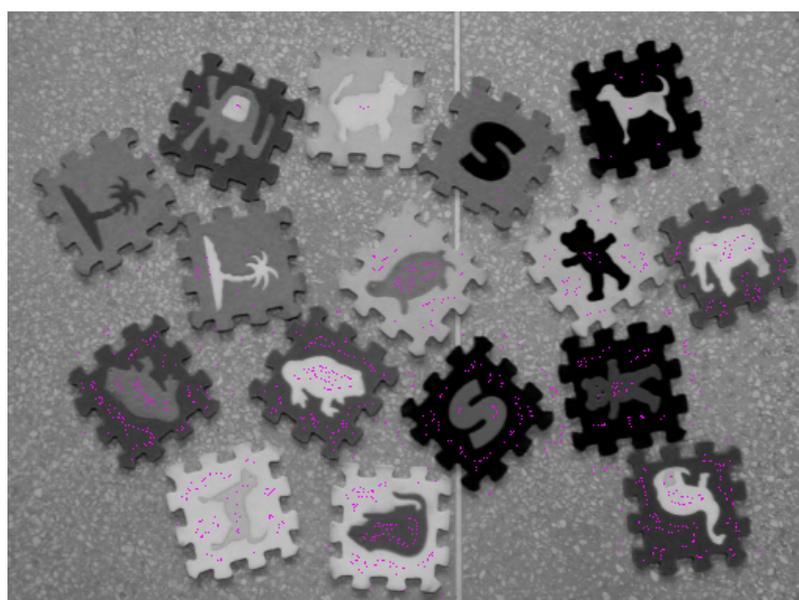
Uma idéia para otimizar as etapas de filtragem do Ciratefi consiste na fixação de valores baixos para  $t_1$ ,  $t_2$  e  $t_3$  (por exemplo 0,5, 0,5 e 0,5) e definir uma quantidade  $q_i$  de pixels candidatos a serem promovidos de uma fase para outra. Cabe ressaltar que tal mecanismo pode ser melhor empregado nos casos onde se sabe, a priori, a quantidade de instâncias do padrão que devem ser detectadas na imagem analisada. Quando não se tem tal informação, pode-se aplicar esta idéia nas duas primeiras etapas de filtragem, deixando a filtragem final a cargo do limiar  $t_3$ . Além disso, pode-se também definir a distância  $d_i$  que os candidatos devem estar separados uns dos outros. Neste caso, em uma região cujo raio depende de  $d_i$ , onde há diversos pixels candidatos, apenas aquele com valor máximo de similaridade é promovido para a fase seguinte.

Considere  $q_1$  a quantidade de candidatos a serem promovidos da primeira para segunda fase,  $q_2$  a quantidade de candidatos a serem promovidos da segunda para terceira fase,  $q_3$  a quantidade de candidatos que restarão na terceira fase (que representarão posições de casamento) e  $d_1$ ,  $d_2$  e  $d_3$  as distâncias entre os candidatos da primeira, segunda e terceira fase, respectivamente. Se fixarmos, por exemplo,  $q_1=1000$ ,  $q_2=100$ ,  $q_3=2$ ,  $d_1=2$ ,  $d_2=2$  e  $d_3=20$ , na busca por instâncias de um padrão em uma imagem, somente os 1000 melhores candidatos passam da primeira para segunda fase, os 100 melhores passam da segunda para terceira e apenas os dois melhores candidatos da terceira fase representarão posições de casamento. Além disso, os candidatos da primeira e segunda fase estarão separados um do

outro por uma distância de 2 pixels e os pontos de casamento estarão separados por pelo menos 20 pixels. Para um melhor entendimento da idéia proposta, considere a detecção do padrão sapo em uma imagem do conjunto I, empregando os seguintes parâmetros de controle:  $l=13$  (0, 2,..., 24),  $m=36$  (0, 10,..., 350),  $n=10$  (0,4, 0,5,..., 1,3),  $t_1=0,5$ ,  $t_2=0,5$ ,  $t_3=0,5$ ,  $t_4=0,1$  e  $t_7=1$  e como parâmetros de otimização:  $q_1=1000$ ,  $q_2=100$ ,  $q_3=2$ ,  $d_1=2$ ,  $d_2=2$  e  $d_3=20$ . As Figuras 5.7 e 5.8 apresentam as imagens resultantes das duas primeiras etapas de filtragem (Cifi e Rafi) com e sem o esquema otimização.

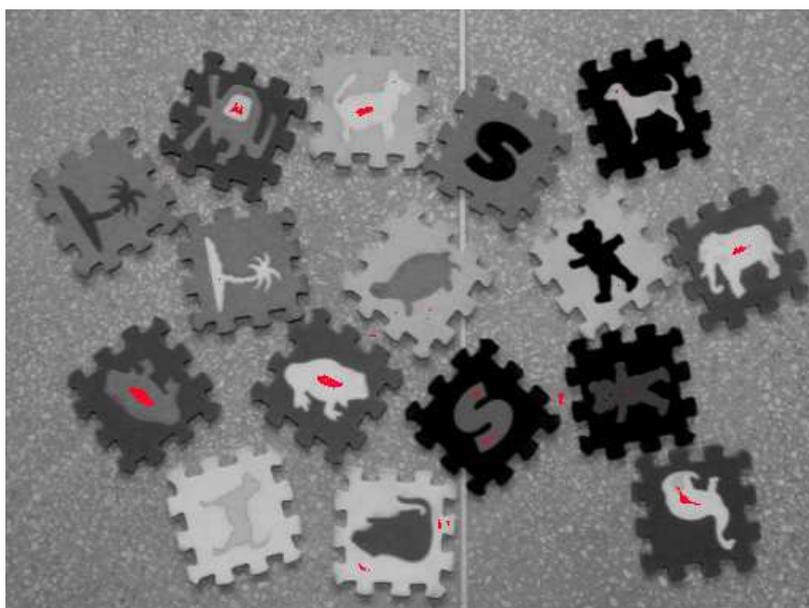


(a) Pixels candidatos da primeira fase sem o esquema de otimização (**20.496 candidatos**)

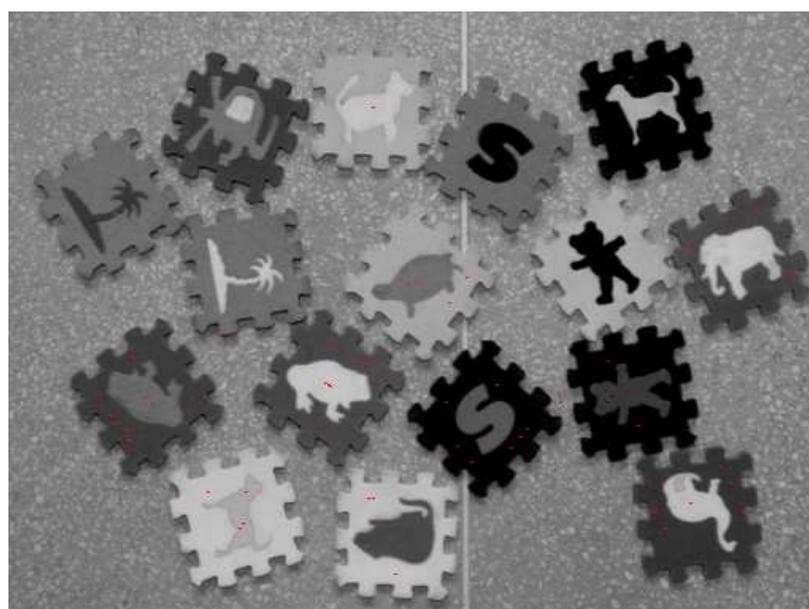


(b) Pixels candidatos da primeira fase com o esquema de otimização (**1.000 candidatos**)

Figura 5.7 – Imagens resultantes da primeira filtragem com os pixels candidatos em magenta



(a) Pixels candidatos da segunda fase sem o esquema de otimização (**394 candidatos**)



(b) Pixels candidatos da segunda fase com o esquema de otimização (**55 candidatos**)

Figura 5.8 – Imagens resultantes da segunda filtragem com os pixels candidatos em magenta

Como pode ser visto nas Figuras 5.7 e 5.8, o esquema de otimização proposto possibilitou uma redução significativa da quantidade de pixels candidatos sem, no entanto, comprometer o processo de detecção. É importante observar na Figura 5.8b que a quantidade de pixels da segunda fase utilizando o esquema de otimização foi menor do que aquela previamente estipulada. O problema é que não foi possível selecionar 100 candidatos separados por 2 pixels de distância, uma vez

que eles se encontravam aglomerados em algumas poucas regiões na imagem antes do processo seletivo. Assim, em uma janela com  $5 \times 5$  pixels candidatos possíveis, apenas aquele com o maior valor de correlação foi selecionado.

Com intuito de fazer uma avaliação mais ampla do esquema de otimização apresentado, refizemos os experimentos com o conjunto de imagens I (brinquedos espalhados pelo chão) utilizando os parâmetros:  $t_1=0,8$ ,  $t_2=0,7$ ,  $t_3=0,7$ ,  $q_1=5000$ ,  $q_2=300$ ,  $q_3=2$ ,  $d_1=2$ ,  $d_2=2$  e  $d_3=20$  e preservando os demais descritos anteriormente nesta seção. Os resultados obtidos foram os mesmos descritos na subseção 5.1.1 (Tabela 5.1), mas com um tempo de processamento muito inferior, em média 11 segundos para detecção de um padrão contra 22 segundos na versão sem o esquema de otimização.

## 5.2 RESULTADOS EXPERIMENTAIS COM CIRATEFI MORFOLÓGICO

O Ciratefi morfológico foi testado apenas com as imagens do conjunto I (brinquedos espalhados pelo chão). O experimento consistiu na aplicação do Ciratefi morfológico para detecção das 140 instâncias do padrão sapo nas 70 imagens (em cada imagem aparecem 2 instâncias do padrão em diferentes rotações, escalas, brilhos e contrastes). Os resultados deste experimento são apresentados na Tabela 5.8, a qual mostra também os resultados obtidos pelo Ciratefi na mesma tarefa.

Tabela 5.8. Resultados do Ciratefi morfológico na detecção do padrão sapo, nas imagens do conjunto I e os resultados obtidos por Ciratefi na mesma tarefa.

Método	Casamentos possíveis	Detecções corretas	Falsos positivos	Falsos negativos	Taxa de acertos(%)
Ciratefi morfológico	140	80	14	60	<b>57,1</b>
Ciratefi	140	140	0	0	<b>100,0</b>

A quantidade de falsos positivos e falsos negativos apresentados na Tabela 5.8 mostra claramente que o Ciratefi morfológico precisa ser melhorado. Além disso, o Ciratefi morfológico gasta, em média, 80 segundos para processar uma imagem do conjunto I enquanto Ciratefi gasta, em média, 22 segundos. O custo de tempo do Ciratefi morfológico pode ser explicado, em parte, pela quantidade média de pixels candidatos da primeira e da segunda fase, como mostrado na Tabela 5.9. Outro fator agravante é que o vetor de atributos extraídos por Ciratefi morfológico é duas vezes maior que o vetor extraído por Ciratefi. Tais fatores justificam a realização de experimentos com o método envolvendo apenas as imagens do conjunto I.

Tabela 5.9. Quantidade média de pixels candidatos na primeira e na segunda fase gerados pelos métodos Ciratefi morfológico e Ciratefi.

Método	Qtd. média de pixels candidatos da 1ª fase	Qtd. média de pixels candidatos da 2ª fase	Tempo médio de processamento
Ciratefi morfológico	43.112	1.899	80s
Ciratefi	31.716	16	22s

### 5.3 ANÁLISE COMPARATIVA SIFT X CIRATEFI

Os experimentos com SIFT envolveram os quatro conjuntos de imagens apresentados na seção 5.1 (imagens de brinquedos espalhados pelo chão, imagens contendo símbolo do McDonald's<sup>®</sup>, imagens de sensoriamento remoto e imagens de cartões do jogo da memória). As implementações disponíveis do método SIFT, incluindo o algoritmo escrito por Lowe (2005) nas linguagens C/C++ e Matlab, foram utilizadas para fazer extração e correspondência de pontos chaves nas imagens como ilustrado na Figura 5.9.

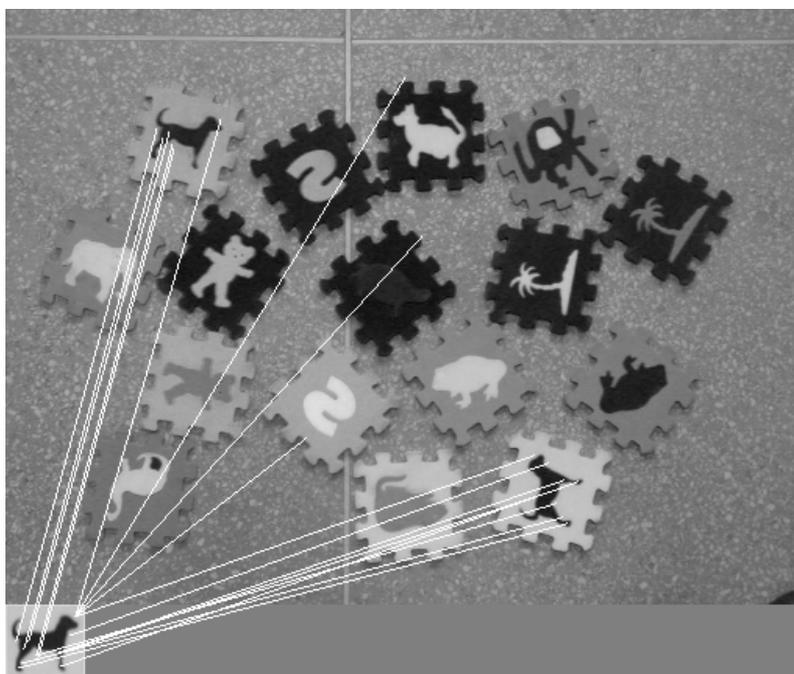


Figura 5.9 – Resultado do processo de detecção do padrão cachorro por SIFT.

No caso ilustrado na figura 5.9 as duas instâncias do padrão cachorro foram corretamente detectadas, sem nenhum falso positivo ou falso negativo. Isto porque, conforme indicado por Lowe (2004), a detecção de um objeto necessita de pelo menos 3 pontos de correspondência.

Um problema encontrado nos experimentos com SIFT foi que a detecção de instâncias dos padrões em uma mesma imagem deixaria as imagens resultantes (do processo de detecção) visualmente poluídas pelas linhas que associam os pontos, tornando difícil a visualização dos resultados. Assim, uma implementação adequada para o processo de casamento dos pontos chaves foi necessária para realização

dos experimentos. Um algoritmo para esta finalidade foi então desenvolvido em linguagem C com o uso das bibliotecas Proeikon (KIM, 2006) e OpenCV (INTEL, 2000), e possibilitou que as imagens resultantes do processo de detecção pelo SIFT fossem iguais às processadas por Ciratefi, o que facilitou a visualização e apuração dos resultados. É importante salientar que o algoritmo de casamento dos pontos chaves desenvolvido não dispensou o algoritmo de extração dos descritores SIFT desenvolvido e parametrizado por Lowe (2005), mas foi acoplado a ele.

O experimento com as imagens do conjunto I, da mesma forma como foi feito com Ciratefi, consistiu na detecção das 700 instâncias dos 5 padrões que aparecem nas 70 imagens dos brinquedos. Os resultados deste experimento são apresentados na Tabela 5.10.

Tabela 5.10. Resultados do SIFT no experimento com as imagens do conjunto I e os resultados obtidos por Ciratefi na mesma tarefa.

Padrão	Casamentos possíveis	SIFT			Ciratefi		
		Detecções corretas	Falsos positivos	Falsos negativos	Detecções corretas	Falsos positivos	Falsos negativos
Sapo	140	134	0	6	140	0	0
Urso	140	122	4	18	140	0	0
Palmeira	140	99	0	41	140	0	0
Letra-S	140	77	0	63	140	0	0
Cachorro	140	132	0	8	140	0	0
<b>Total</b>	<b>700</b>	<b>564</b>	<b>4</b>	<b>136</b>	<b>700</b>	<b>0</b>	<b>0</b>

Os dados apresentados na Tabela 5.10 mostram que o Ciratefi obteve melhores resultados do que SIFT. Enquanto SIFT detectou corretamente 564 instâncias dos padrões, apresentando 4 falsos positivos e 136 falsos negativos, Ciratefi detectou todas as 700 instâncias sem nenhum caso de falso positivo ou falso negativo. Entretanto, cabe ressaltar que o processo de detecção por SIFT, envolvendo a extração dos atributos e a execução do algoritmo de casamento de pontos chaves, é cerca de 10 vezes mais rápido do que o processo de detecção por Ciratefi (SIFT

gasta aproximadamente 2 segundos para detectar as instâncias de um padrão em uma imagem do conjunto I).

A Tabela 5.11 apresenta resultados do método SIFT envolvendo os conjuntos de imagens II e III. Ressalta-se que todos os resultados de SIFT foram obtidos considerando pelo menos 3 pontos correspondentes entre as imagens *T* e *A* para confirmar a detecção, como sugerido em Lowe (2004).

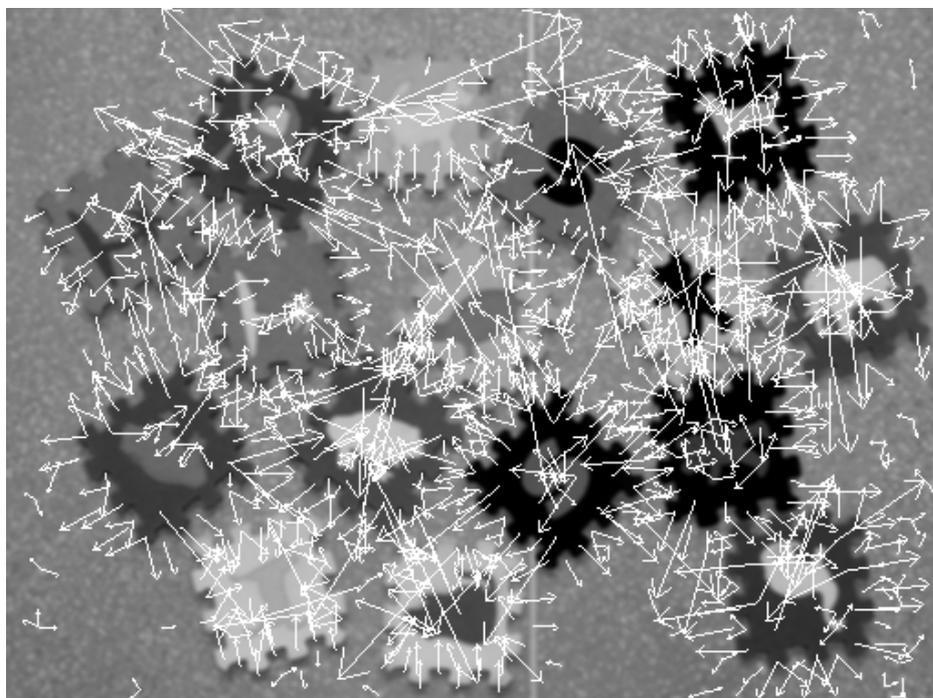
Tabela 5.11. Resultados do SIFT nos experimentos com as imagens dos conjuntos II e III e os resultados obtidos por Ciratefi nas mesmas tarefas.

Conjunto	Casamentos possíveis	SIFT			Ciratefi		
		Detecções corretas	Falsos positivos	Falsos negativos	Detecções corretas	Falsos positivos	Falsos negativos
II McDonald's®	116	54	23	62	114	1	2
III Sensoriamento remoto	187	61	99	126	171	18	16

Como pode ser visto na Tabela 5.11, para os conjuntos de imagens II e III, os resultados obtidos por Ciratefi também foram superiores aos obtidos por SIFT. Por outro lado, nos experimentos com o conjunto de imagens IV SIFT apresentou resultados melhores que os do Ciratefi, conforme mostra a tabela 5.12.

A baixa precisão do SIFT nos conjuntos I, II e III pode estar relacionada com falta ou excesso de detalhes (textura) para caracterização dos padrões contidos nas imagens. Isso fica aparente diante dos seus resultados com o conjunto IV, no qual as imagens possuem mais detalhes que as imagens dos conjuntos I e II, mas não a ponto de apresentar excesso como as imagens do conjunto III, para o qual SIFT apresentou os piores resultados. A falta de detalhes faz com que os pontos chaves fiquem limitados às bordas dos objetos, gerando um alto número de falsos negativos. Já o excesso de detalhes faz SIFT produzir um grande número de falsos pontos chaves, influenciando negativamente o processo de detecção. Exemplos

destas observações podem ser vistos na Figura 5.10, a qual ilustra os pontos chaves extraídos de duas imagens por SIFT.



(a) Pontos chaves extraídos de uma imagem do conjunto I



(b) Pontos chaves extraídos de uma imagem do conjunto III

Figura 5.10 – Pontos chaves extraídos por SIFT de imagens dos conjuntos I e III.

Tabela 5.12. Resultados do SIFT no experimento com as imagens do conjunto IV e os resultados obtidos por Ciratefi na mesma tarefa.

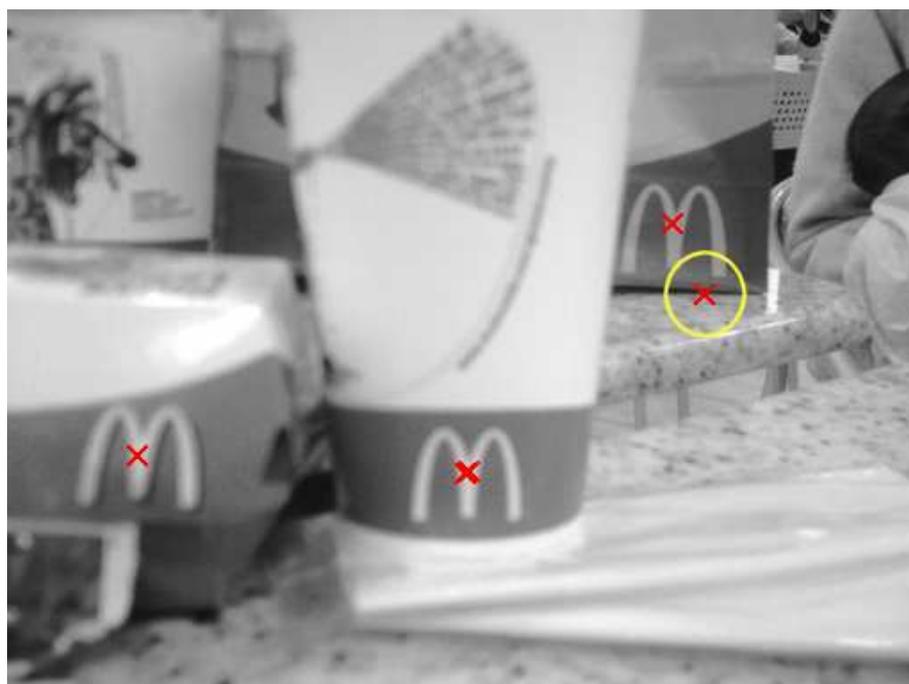
Subconjunto	Casamentos possíveis	SIFT			Ciratefi		
		Detecções corretas	Falsos positivos	Falsos negativos	Detecções corretas	Falsos positivos	Falsos negativos
A	120	120	0	0	120	1	0
B	120	120	2	0	118	5	2
C	120	120	0	0	89	2	31
<b>Total</b>	<b>360</b>	<b>360</b>	<b>2</b>	<b>0</b>	<b>327</b>	<b>8</b>	<b>33</b>

A Tabela 5.13 sumariza os resultados dos experimentos com SIFT, mostrando também os resultados obtidos por Ciratefi. A Figura 5.11 ilustra dois resultados do processo de detecção por SIFT em imagens dos conjuntos II e III.

Tabela 5.13. Resumo dos experimentos realizados com SIFT.

Conjunto	Casamentos possíveis	SIFT				Ciratefi			
		Detec. corretas	Falsos pos.	Falsos neg.	Taxa de acertos (%)	Detec. corretas	Falsos pos.	Falsos neg.	Taxa de acertos (%)
I – Brinquedos	700	564	4	136	<b>80,6</b>	700	0	0	<b>100,0</b>
II – McDonald's®	116	54	23	62	<b>46,6</b>	114	1	2	<b>98,3</b>
III – Sensoriamento remoto	187	61	99	126	<b>32,6</b>	171	18	16	<b>91,4</b>
IV – Cartões jogo da memória	360	360	2	0	<b>100,0</b>	327	8	33	<b>90,8</b>
<b>Total</b>	<b>1.363</b>	<b>1.039</b>	<b>128</b>	<b>324</b>	<b>76,2</b>	<b>1.312</b>	<b>27</b>	<b>51</b>	<b>96,3</b>

Na seção seguinte são apresentados os resultados do método EasyMatch considerando os quatro conjuntos de imagens.



(a) Resultado da detecção por SIFT em uma imagem do conjunto II com três detecções corretas e um caso de falso positivo assinalado por um círculo amarelo



(b) Resultado da detecção, por SIFT, do prédio em forma de "H" em uma imagem do conjunto III com inúmeros casos de falsos positivos

Figura 5.11 – Resultados da detecção de padrões por SIFT em imagens dos conjuntos II e III.

## 5.4 ANÁLISE COMPARATIVA EASYMATCH X CIRATEFI

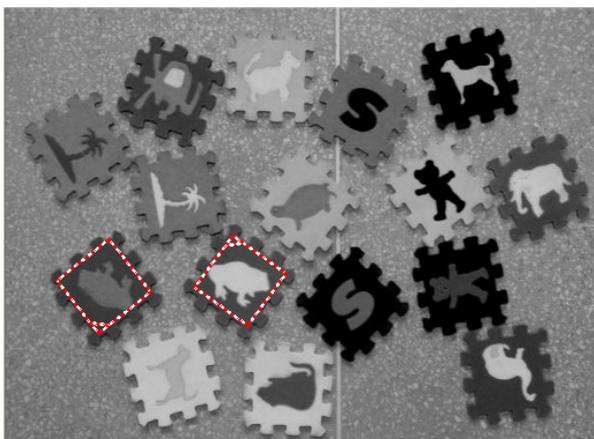
Os experimentos com EasyMatch foram realizados com o uso de uma versão de demonstração do software comercial Open eVision™, apresentada na seção 3.2, e envolveu os conjuntos de imagens I, II, III e IV. No experimento com as imagens do conjunto I (brinquedos), considerou-se apenas a detecção das 140 instâncias do padrão sapo que aparecem nas 70 imagens. Esta redução das amostras foi devida à impossibilidade da realização dos testes de forma automatizada, uma vez que a versão de demonstração do software é limitada neste sentido. Os resultados dos experimentos com o EasyMatch estão sumarizados na Tabela 5.14, a qual mostra também os resultados obtidos por Ciratefi.

Tabela 5.14. Resumo dos experimentos realizados com EasyMatch.

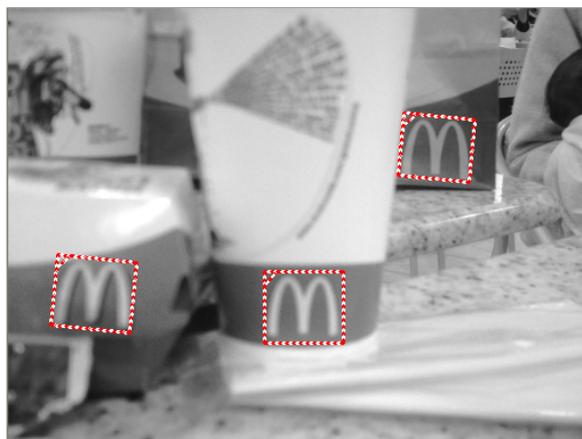
Conjunto	Casamentos possíveis	EasyMatch				Ciratefi			
		Detec. corretas	Falsos pos.	Falsos neg.	Taxa de acertos (%)	Detec. corretas	Falsos pos.	Falsos neg.	Taxa de acertos (%)
I – Brinquedos	140	131	9	9	<b>93,6</b>	140	0	0	<b>100,0</b>
II – McDonald's®	116	101	15	15	<b>87,1</b>	114	1	2	<b>98,3</b>
III – Sensoriamento remoto	187	118	69	69	<b>63,1</b>	171	18	16	<b>91,4</b>
IV – Cartões jogo da memória	360	277	83	83	<b>76,9</b>	327	8	33	<b>90,8</b>
<b>Total</b>	<b>803</b>	<b>627</b>	<b>176</b>	<b>176</b>	<b>78,1</b>	<b>752</b>	<b>27</b>	<b>51</b>	<b>93,6</b>

Os resultados apresentados na Tabela 5.14 mostram que o Ciratefi teve melhor desempenho que EasyMatch nos experimentos com os quatro conjuntos de imagens. Enquanto EasyMatch detectou corretamente 627 instâncias dos padrões, Ciratefi detectou 752 instâncias, apresentando menos casos de falsos positivos e falso negativos. No entanto, os resultados do EasyMatch foram superiores aos resultados do SIFT para as imagens dos conjuntos II e III. Além disso, EasyMatch é mais rápido do que SIFT visto que leva apenas cerca de 1,5 segundos para detectar

as instâncias de um padrão em uma imagem do conjunto I. A Figura 5.12 ilustra três resultados da detecção por EasyMatch em imagens dos conjuntos I, II e III.



(a) detecção perfeita de duas instâncias do padrão sapo



(b) detecção perfeita das instâncias do símbolo do McDonald's®



(c) detecção das instâncias do prédio em forma de "H". Os casos de falsos negativos estão assinalados por círculos amarelos e o caso de falso positivo assinalado por um círculo azul

Figura 5.12 – Resultados da detecção de padrões por EasyMatch em imagens dos conjuntos I, II e III.

## 5.5 RESUMO DAS COMPARAÇÕES ENTRE CIRATEFI, SIFT E EASYMATCH

Nesta seção apresenta-se uma síntese das comparações realizadas entre os métodos Ciratefi, SIFT e EasyMatch com o objetivo de mostrar as qualidades associadas a cada um deles. Cabe ressaltar que Ciratefi morfológico não foi incluído nesta síntese já que poucos experimentos foram realizados com o método em virtude do seu desempenho. Ressalta-se também que nos experimentos com EasyMatch envolvendo as imagens do conjunto I, explorou-se apenas a detecção das 140 instâncias do padrão sapo pelas razões expostas na seção 5.4. Na Tabela 5.15 é apresentada a taxa de acertos dos métodos com base nos experimentos realizados.

Tabela 5.15. Taxa de acertos dos métodos Ciratefi, SIFT e EasyMatch.

Métodos		Conjunto I					Conjunto II	Conjunto III	Conjunto IV	Total	
		sapo	urso	palm.	letra-S	cachor.					Total
Casamentos possíveis		140	140	140	140	140	700	116	187	360	<b>1.363</b>
<b>Ciratefi</b>	Detecções corretas	140	140	140	140	140	700	114	171	327	<b>1.312</b>
	Falsos positivos	0	0	0	0	0	0	1	18	8	<b>27</b>
	Falsos negativos	0	0	0	0	0	0	2	16	33	<b>51</b>
	<b>Taxa de acertos (%)</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>98,3</b>	<b>91,4</b>	<b>90,8</b>	<b>96,3</b>
<b>SIFT</b>	Detecções corretas	134	122	99	77	132	564	54	61	360	<b>1.039</b>
	Falsos positivos	0	4	0	0	0	4	23	99	2	<b>128</b>
	Falsos negativos	6	18	41	63	8	136	62	126	0	<b>324</b>
	<b>Taxa de acertos (%)</b>	<b>95,7</b>	<b>87,1</b>	<b>70,7</b>	<b>55,0</b>	<b>94,3</b>	<b>80,6</b>	<b>46,6</b>	<b>32,6</b>	<b>100,0</b>	<b>76,2</b>
<b>EasyMatch</b>	Detecções corretas	131	-	-	-	-	131	101	118	277	<b>627</b>
	Falsos positivos	9	-	-	-	-	9	15	69	83	<b>176</b>
	Falsos negativos	9	-	-	-	-	9	15	69	83	<b>176</b>
	<b>Taxa de acertos (%)</b>	<b>93,6</b>	-	-	-	-	<b>93,6</b>	<b>87,1</b>	<b>63,1</b>	<b>76,9</b>	<b>78,1</b>

Na Tabela 5.15, para cada método, são apresentados os seguintes dados: quantidade de detecções corretas (verdadeiros positivos), quantidade de falsos positivos, quantidade de falsos negativos e a taxa de acertos ou recall definida na equação 2.8. Com base nos dados apresentados na Tabela 5.15, elaborou-se um *ranking* dos métodos (Tabela 5.16) considerando a taxa de acertos. Na Tabela 5.17 apresenta-se o *ranking* dos métodos com relação ao tempo médio de processamento. Vale lembrar que o tempo médio de processamento foi calculado com base no conjunto I, no qual todas as imagens possuem o mesmo tamanho (512x384 pixels), e consiste no tempo total gasto para fazer a detecção das 140 instâncias do padrão sapo dividido pelo total de imagens processadas (70 imagens).

Tabela 5.16. *Ranking* dos métodos com base na taxa de acertos.

Método	Casamentos possíveis	Total de detecções corretas	Taxa de acertos em %
1. Ciratefi	1.363	1.312	96,3
2. EasyMatch	803	627	78,1
3. SIFT	1.363	1.039	76,2

Os dados apresentados nas Tabelas 5.15 e 5.16 deixam claro que, no geral, a precisão do Ciratefi foi superior à precisão dos métodos SIFT e EasyMatch. A baixa precisão do SIFT talvez possa ser explicada pela dependência de detalhes nas imagens que melhor caracterizem os objetos ou padrões, haja vista que regiões locais com informações de magnitude e gradiente parecidos resultam em descritores semelhantes. Se observarmos, principalmente no conjunto I, as imagens possuem poucos detalhes, fazendo com que os pontos chaves se concentrem apenas nas bordas dos objetos, como pode ser visto nas Figuras 3.3, 5.9 e 5.10a, o que resulta em um grande número de falsos negativos. Já no caso das imagens do conjunto III, onde a riqueza de detalhes é excessiva, há uma confusão no processo de casamento de pontos chaves, ocasionando um grande número de falsos positivos, como pode observado nas Figuras 5.10b e 5.11b.

Ciratefi somente foi superado por SIFT nos experimentos com as imagens do conjunto IV, principalmente do subconjunto C onde as imagens as instâncias dos padrões nas imagens são afetadas por oclusão parcial. Uma sugestão para contornar este problema é fazer busca de partes do padrão, ou seja, dividir o padrão em partes menores e tomar cada uma destas partes como sendo um padrão a ser detectado.

Com relação ao desempenho do EasyMatch, pouco se pode concluir uma vez que não se conhece a metodologia empregada na extração dos atributos descritores.

Tabela 5.17. *Ranking* dos métodos com base no tempo de processamento

Método	Tempo médio de processamento
1. EasyMatch	1,5 segundos
2. SIFT	2,0 segundos
3. Ciratefi	22,0 segundos

Como pode ser visto na Tabela 5.17, a velocidade de processamento dos métodos SIFT e EasyMatch são muito superiores à velocidade do Ciratefi, o que explica a utilização destes métodos em diversas aplicações práticas. Contudo, o Ciratefi, conforme descrito em Nobre e Kim (2009), pode ser acelerado em milhares de vezes, via implementação em hardware (FPGA), enquanto a aceleração do SIFT utilizando o mesmo tipo de implementação chega, no máximo, a uma dezena de vezes devido a complexidade das operações aritméticas e trigonométricas empregadas (BONATO; MARQUES; CONSTANTINIDES, 2008) e também porque não repete as mesmas operações para todos os pixels (NOBRE; KIM, 2009). Em adição, mesmo considerando implementação em software, ainda é possível fazer algumas otimizações no Ciratefi visando diminuir sua dependência dos parâmetros e aumentar sua velocidade. Uma idéia simples com este intuito é apresentada a seguir.

## 5.7 RESULTADOS EXPERIMENTAIS COM COLOR CIRATEFI

A despeito da vasta quantidade de estudos envolvendo o reconhecimento de padrões com base em atributos de cor, não foram encontrados na literatura outros métodos que compartilham as mesmas características e objetivos do color Ciratefi. Assim, para avaliar seu desempenho, comparamos seus resultados com os resultados obtidos pelo método color SIFT (BURGHOUTS; GEUSEBROEK, 2008). Embora o SIFT não tenha como propósito a detecção de objetos em imagens, ele tem sido utilizado com sucesso em muitas aplicações com esta finalidade, incluindo uma aplicação proposta por Lowe (1999).

Os experimentos com color Ciratefi envolveram um conjunto de 42 imagens de cenas reais com diferentes distorções geométricas e fotométricas. A base de imagens utilizada foi proposta por Mikolajczyk e Schmid (2005) e é composta por 48 imagens divididas em 8 subconjuntos rotulados pelos autores como “bark”, “bikes”, “boat”, “wall”, “leuven”, “graffiti”, “trees” e “ubc”. Entretanto, um dos subconjuntos (“boat” – Figura 5.13h) contém apenas imagens em níveis de cinzas e, portanto, não foi utilizado nos experimentos. Cada subconjunto é composto por 6 imagens da mesma cena gradualmente afetadas por distorções geométricas e/ou fotométricas. A Figura 5.13 ilustra exemplos das imagens contidas na base citada. Há seis diferentes distorções: rotação e escala (Figura 5.13a); mudanças na iluminação (Figura 5.13b); mudanças de ponto de vista (Figuras 5.13c e 5.13d); borramento (Figuras 5.13e e 5.13f) e compressão JPEG (Figura 5.13g).

A escolha desta base foi motivada pelo fato dela conter imagens cujas distorções parecem ser mais realistas, do ponto de vista de situações práticas, do que aquelas contidas em bases bastante conhecidas, por exemplo, a base ALOI (*Amsterdam Library of Object Images*) proposta por Geusebroek, Burghouts e Smeulders (2005). Além disso, tal base tem sido utilizada para avaliação do desempenho de algoritmos de reconhecimento de objetos em vários trabalhos recentes. Entre eles podemos citar: Mikolajczyk e Schmid (2005), Mikolajczyk et al. (2006), Bay, Tuytelaars e Van Gool (2006), Marimon e Ebrahimi (2007) e Cui et al. (2009).



Figura 5.13 – Exemplos de imagens da base utilizada nos experimentos com color Ciratefi.

Para realização dos experimentos, primeiramente as imagens foram reduzidas para 50% dos seus tamanhos originais. Na seqüência, extraiu-se 20 padrões (sub-imagens) de uma única imagem de cada um dos 7 subconjuntos para serem procurados nas imagens do mesmo subconjunto. Cada padrão, com tamanho 61x61 pixels, foi extraído considerando posição, ângulo de rotação e escala aleatórios. Para escala considerou-se o intervalo [0,5, 1,0] e para rotação [0, 360].

Em resumo, o experimento consistiu na busca de 840 instâncias dos padrões nas 42 imagens da base (20x6x7). Para cada padrão deveriam ser encontradas 6 instâncias nas 6 imagens do mesmo subconjunto. O desempenho do color Ciratefi foi

comparado com o desempenho do color SIFT (BURGHOUTS; GEUSEBROEK, 2008), o qual utiliza o conjunto de características de cor invariantes proposto por Geusebroek et al. (2001). Visando uma comparação justa, utilizou-se o descritor que apresentou melhor desempenho (C-color-SIFT), de acordo com Burghouts e Geusebroek (2008).

Os seguintes parâmetros foram utilizados nos experimentos com color Ciratefi:  $l=21$ ,  $m=36$  e  $n=9$  (0,3 a 1,1). Dado que o método utiliza o esquema de otimização descrito na subseção 5.1.6, fixou-se  $t_1=0,7$ ,  $t_2=0,5$ ,  $t_3=0,5$  e determinou-se  $q_1=2000$ ,  $q_2=300$ ,  $q_3=1$ ,  $d_1=3$ ,  $d_2=3$ ,  $d_3=1$ . Os pesos utilizados na medida de similaridade (equação 4.30) que apresentaram os melhores resultados foram:  $\delta=0,8$  e  $\eta=0,2$ .

Para se chegar a estes valores, o ponto de partida foi a realização de experimentos considerando apenas o subconjunto leuven cujas imagens são afetadas por mudanças de iluminação, variando gradativamente os valores de  $\delta$  e  $\eta$  e fixando-se os demais parâmetros. Os resultados destes experimentos estão na Tabela 5.18.

Tabela 5.18. Resultados do color Ciratefi nos experimentos com as imagens do subconjunto leuven variando  $\delta$  e  $\eta$  e fixando-se os demais parâmetros.

Parâmetros $\delta$ e $\eta$	Casamentos possíveis	Detecções corretas	Taxa de acertos (em %)
1,0 ; 0,0	120	31	25,8
<b>0,9 ; 0,1</b>	<b>120</b>	<b>84</b>	<b>70,0</b>
<b>0,8 ; 0,2</b>	<b>120</b>	<b>94</b>	<b>78,3</b>
<b>0,7 ; 0,3</b>	<b>120</b>	<b>95</b>	<b>79,2</b>
0,6 ; 0,4	120	83	69,2
0,5 ; 0,5	120	47	39,2

No passo seguinte, selecionou-se os pares de parâmetros que apresentaram os melhores resultados na Tabela 5.18, para empregá-los em experimentos envolvendo todas as imagens da base e então identificar o par  $\delta$  e  $\eta$  que possibilitou a maior taxa de acertos. Os resultados destes experimentos são mostrados no gráfico ilustrado na Figura 5.14 e sumarizados na Tabela 5.19.

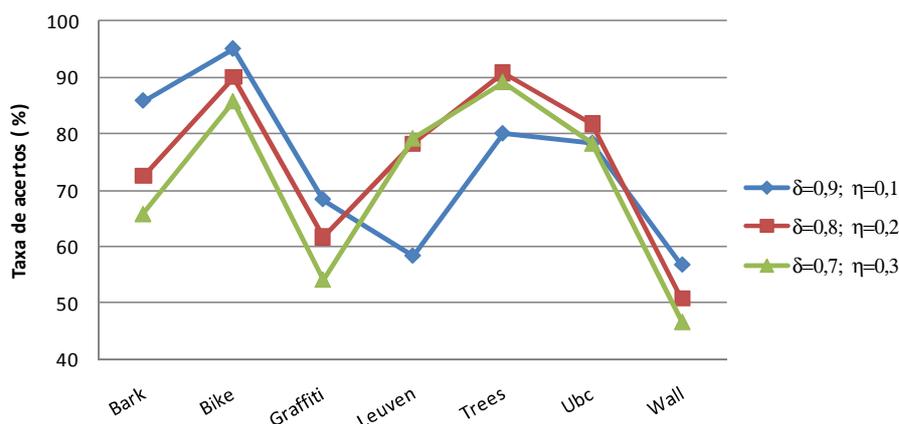


Figura 5.14 – Gráfico ilustrando a taxa de acertos do color Ciratefi para cada um dos subconjuntos de imagens, considerando diferentes pares de  $\delta$  e  $\eta$ .

Tabela 5.19. Resultados do color Ciratefi nos experimentos com todas as imagens da base variando  $\delta$  e  $\eta$  e fixando-se os demais parâmetros.

Parâmetros $\delta$ e $\eta$	Casamentos possíveis	Detecções corretas	Taxa de acertos (em %)
0,9 ; 0,1	840	627	74,6
<b>0,8 ; 0,2</b>	<b>840</b>	<b>631</b>	<b>75,1</b>
0,7 ; 0,3	840	599	71,3

Embora os resultados com as imagens do subconjunto leuven (Tabela 5.18) apontassem o par  $\delta=0,7$  e  $\eta=0,3$  como sendo o melhor, os experimentos com todas as imagens da base apontaram outro par ( $\delta=0,8$  e  $\eta=0,2$ ). A explicação disso pode estar na importância da componente de luminosidade na avaliação de similaridade nos experimentos envolvendo as imagens deste subconjunto, tendo em vista que partes das imagens são muito escuras, como pode ser visto na Figura 5.13b, deixando a informação de cromaticidade com pouco poder discriminativo.

Os resultados dos algoritmos color Ciratefi e color SIFT nos experimentos envolvendo todas as imagens da base utilizada, em termos de taxa de acertos, são mostrados na Tabela 5.20. Cabe salientar os experimentos apresentados nesta seção, a exemplo daqueles realizados com o Ciratefi, foram executados em um microcomputador Pentium 4 2.8GHz.

Tabela 5.20. Resultados dos algoritmos color SIFT e color Ciratefi nos experimentos com todas as imagens da base.

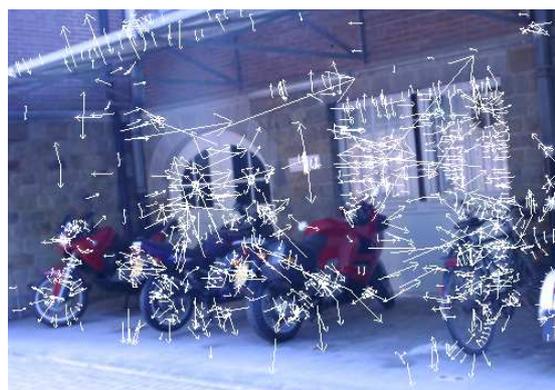
Subconjunto	Distorções contidas nas imagens	Casamentos possíveis	color SIFT			color Ciratefi		
			Detec. corretas	Falsos neg.	Taxa de acertos(%)	Detec. corretas	Falsos neg.	Taxa de acertos(%)
<b>Bark</b>	Rotação/Escala	120	83	37	<b>69,2</b>	87	33	<b>72,5</b>
<b>Bike</b>	Borramento	120	76	44	<b>63,3</b>	108	12	<b>90,0</b>
<b>Graffiti</b>	Dif. pontos de vista	120	48	72	<b>40,0</b>	74	46	<b>61,7</b>
<b>Leuven</b>	Mud.de iluminação	120	76	44	<b>63,3</b>	94	26	<b>78,3</b>
<b>Trees</b>	Borramento	120	94	26	<b>78,3</b>	109	11	<b>90,8</b>
<b>Ubc</b>	Compressão JPEG	120	58	62	<b>48,3</b>	98	22	<b>81,7</b>
<b>Wall</b>	Dif. pontos de vista	120	94	26	<b>78,3</b>	61	59	<b>50,8</b>
<b>Totais</b>		<b>840</b>	<b>529</b>	<b>311</b>	<b>63,0</b>	<b>631</b>	<b>209</b>	<b>75,1</b>

Os resultados sumarizados na tabela 5.20 deixam claro que o desempenho médio do color Ciratefi foi superior ao do color SIFT. Para imagens com mudanças de escala e rotação, mudanças nas condições de iluminação e degradação por borramento e por compressão JPEG color Ciratefi foi superior, sendo as maiores diferenças observadas nas imagens afetadas por borramento e por compressão JPEG (subconjuntos bike e ubc).

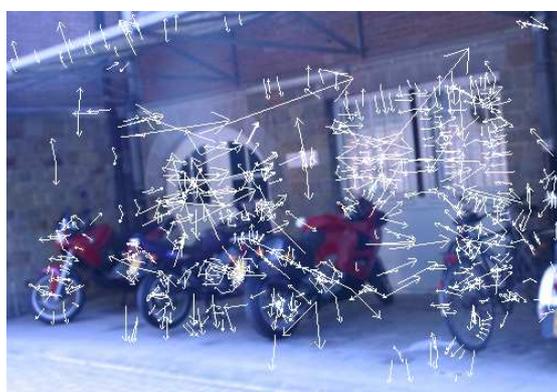
No que diz respeito às imagens afetadas por borramento, o fraco desempenho do color SIFT pode ser explicado pelo baixo contraste das imagens tendo em vista que, quanto mais borrada a imagem, menor foi a quantidade de pontos chaves extraídos por color SIFT, conforme pode ser visto na Figura 5.15.



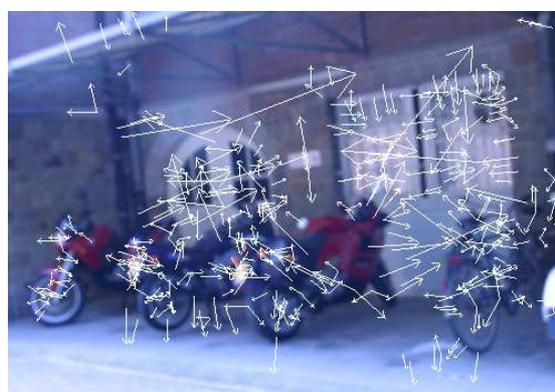
(a) 1.144 pontos chaves extraídos da imagem 1



(b) 1.036 pontos chaves extraídos da imagem 3



(c) 810 pontos chaves extraídos da imagem 4



(d) 439 pontos chaves extraídos da imagem 6

Figura 5.15 – Quantidade de pontos chaves extraídos por color SIFT de imagens do conjunto bike, gradualmente afetadas por borramento.

No caso das imagens degradadas por compressão JPEG, os artefatos introduzidos pelo esquema de compressão induzem o color SIFT a produzir um grande número de falsos pontos chaves influenciando negativamente o processo de casamento dos pontos chaves. Detalhes acerca destas observações podem ser vistos na Figura 5.16.

A Tabela 5.21 mostra a taxa de acertos do color SIFT na busca pelos 20 padrões em cada uma das imagens dos subconjuntos Bike e Ubc. A imagem 1 é a menos afetada (por borramento ou compressão JPEG) enquanto a imagem 6 é a mais afetada. Fica evidente que à medida que aumenta a degradação da imagem, diminui a taxa de acertos.

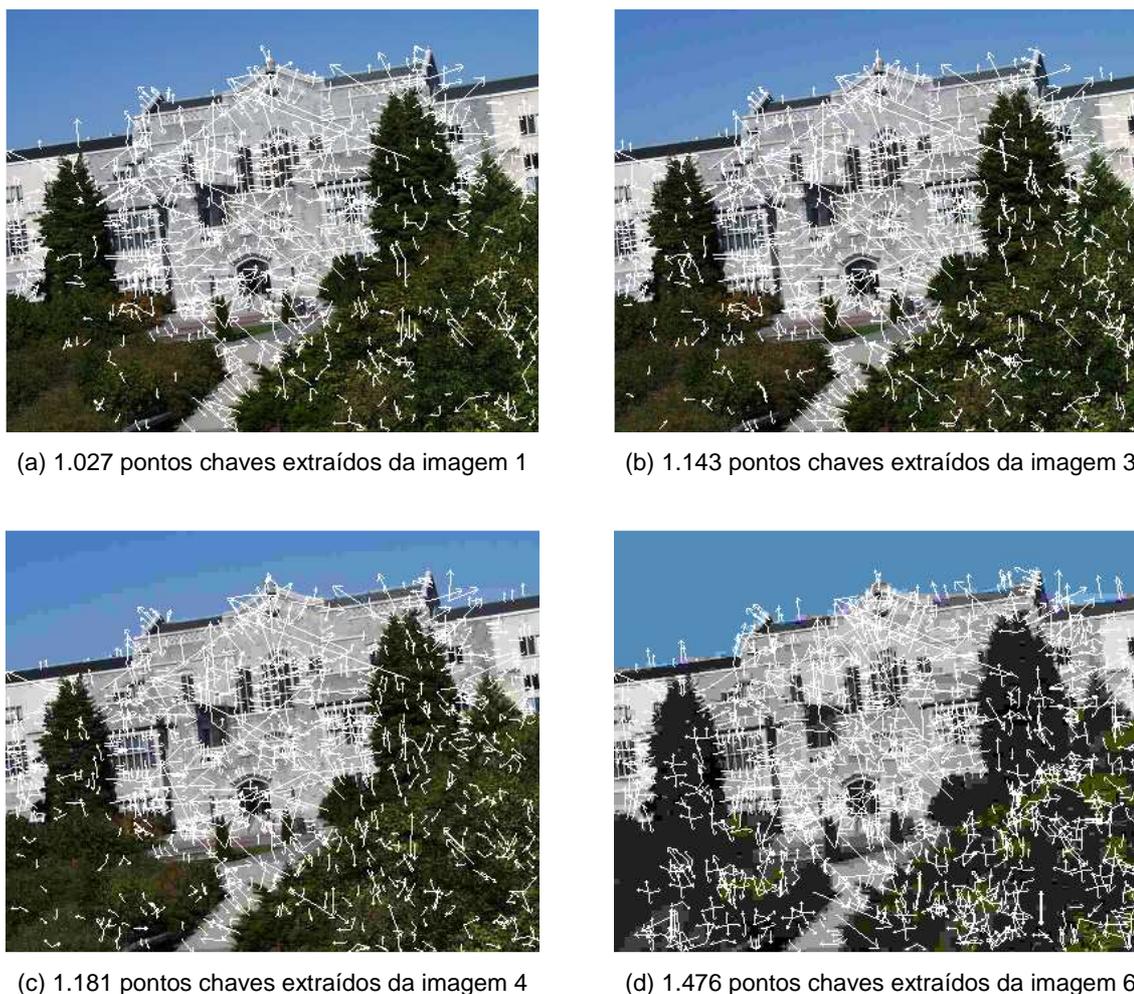


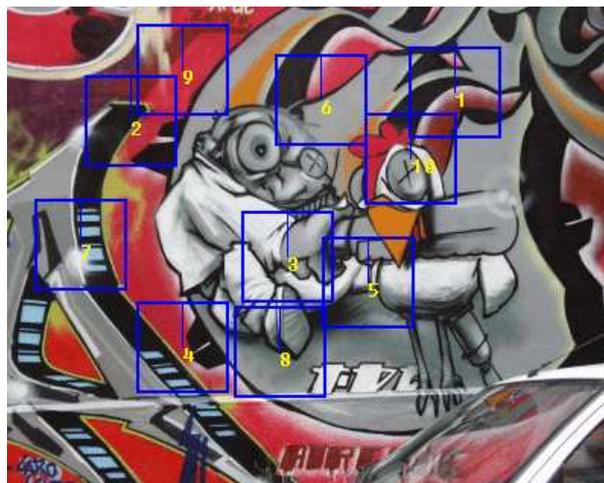
Figura 5.16 – Quantidade de pontos chaves extraídos por color SIFT de imagens do conjunto Ubc, gradualmente afetadas por compressão JPEG.

Tabela 5.21. Desempenho do color SIFT considerando imagens gradualmente afetadas por borramento e por compressão JPEG.

Subconjunto	Distorções contidas nas imagens	Taxa de acertos (em %)					
		Imagem 1	Imagem 2	Imagem 3	Imagem 4	Imagem 5	Imagem 6
<b>Bike</b>	Borramento	90,0	80,0	65,0	60,0	50,0	35,0
<b>Ubc</b>	Compressão JPEG	90,0	85,0	60,0	45,0	10,0	00,0

Considerando as imagens adquiridas sob diferentes pontos de vistas (subconjuntos graffiti e wall), o desempenho médio do color SIFT foi superior ao do color Ciratefi. No entanto, como pode ser visto na Tabela 5.20, color Ciratefi apresentou uma

vantagem substancial no que tange as imagens do subconjunto graffiti, indicando que o método foi capaz de tirar proveito da riqueza das cores das imagens. A Figura 5.17 ilustra resultados do color Ciratefi e color SIFT na detecção de 10 padrões (numerados para facilitar a visualização) em uma imagem do conjunto graffiti.



(a) Padrões a serem detectados



(b) Resultado da detecção por color Ciratefi



(c) Resultado da detecção por color SIFT

Figura 5.17 – Resultados da detecção de 10 padrões em uma imagem do conjunto graffiti.

De fato, o algoritmo color Ciratefi não foi desenvolvido para tratar imagens com mudanças muito severas nas condições de iluminação e de mudanças de pontos de vistas. Assim, já era esperado que o método não apresentasse bons resultados, principalmente no caso de imagens adquiridas sob diferentes de pontos de vista. Contudo, color Ciratefi teve melhor desempenho que color SIFT na detecção dos padrões em imagens envolvendo mudanças de iluminação mesmo este último

empregando o importante conjunto de características de cor invariantes proposto por Geusebroek et al. (2001).

Com relação ao tempo de processamento, o algoritmo color SIFT é muito mais rápido que o color Ciratefi. Enquanto color SIFT gastou, em média, 5 segundos para detectar um padrão, color Ciratefi gastou 28 segundos. Contudo, este problema pode ser contornado já que o algoritmo Ciratefi, e conseqüentemente o color Ciratefi, pode ter sua velocidade acelerada em milhares de vezes via implementação em hardware enquanto a aceleração do algoritmo SIFT chega apenas a 10 vezes conforme mencionado na seção 5.5.

Além disso, a robustez do color Ciratefi pode ser aumentada pela incorporação de informações de textura bem como pela busca de partes do padrão visando tratar do problema da oclusão parcial. Neste último caso, bastaria dividir o padrão em partes menores e tomar cada uma delas como sendo um padrão a ser detectado.

## 6 CONCLUSÕES

Neste trabalho foi proposto um algoritmo de casamento de padrões em imagens digitais que é livre de segmentação e invariante sob transformações de similaridade brilho e contraste, denominado Ciratefi. Também foram propostas duas extensões do Ciratefi, uma que utiliza operadores morfológicos na extração dos atributos descritores, denominada Ciratefi Morfológico e outra para imagens coloridas que chamamos de color Ciratefi. Os três métodos são compostos por 3 etapas de filtragem, sendo que a terceira etapa consiste na técnica de casamento de padrões convencional.

Para avaliar o Ciratefi foram realizados experimentos envolvendo um total de 175 imagens divididas em quatro conjuntos, onde se procurou por 1363 instâncias de 19 diferentes padrões. Os experimentos com Ciratefi permitiram avaliar sua eficiência e robustez, mostrando que ele foi capaz de obter resultados similares aos resultados do método de força bruta, com um tempo de processamento de aproximadamente 400 vezes menor. Além disso, utilizou-se os mesmos conjuntos de imagens para comparar os resultados obtidos pelo Ciratefi com os obtidos pelos métodos SIFT e EasyMatch que são duas das soluções mais completas e robustas para reconhecimento de padrões em imagens encontradas na literatura. Esta comparação mostrou a superioridade do Ciratefi no que tange à taxa de acertos (96,3% do Ciratefi contra 78,1% do EasyMatch e 76,2% do SIFT), mas serviu também para mostrar pontos onde Ciratefi precisa ser melhorado, como é o caso do problema da oclusão parcial.

Ao contrário do esperado, o método SIFT apresentou uma baixa taxa de acertos nos experimentos realizados. Uma possível explicação pode ser a dependência de detalhes (texturas) nas imagens que melhor caracterizem os objetos ou padrões, visto que regiões locais com informações de magnitude e gradiente parecidas resultam em descritores semelhantes. Cabe investigar uma maneira de atenuar este problema.

A análise do desempenho do EasyMatch ficou comprometida porque não se teve acesso à metodologia por ele empregada na extração dos atributos descritores, por se tratar de uma aplicação comercial.

Nos experimentos realizados com Ciratefi morfológico foram utilizadas as 70 imagens do conjunto I nas quais 140 instâncias de um único padrão foram procuradas. Estes experimentos mostraram que o método precisa ser melhorado uma vez que conseguiu detectar apenas 57,1% das instâncias do padrão procurado com um custo computacional, em média, quatro vezes maior que o do Ciratefi. O fraco desempenho do Ciratefi morfológico pode ser atribuído à sensibilidade ao ruído já que os vetores de atributos são baseados em valores mínimos e máximos dos níveis de cinzas das imagens. Apesar dos resultados, o método pode abrir uma nova frente de investigação acerca da incorporação de teorias recentes no campo da morfologia matemática como, por exemplo, o operador *hit-or-mis* para imagens em níveis de cinzas proposto por (NAEGEL; PASSAT e RONSE, 2007).

O color Ciratefi teve seu desempenho avaliado em uma tarefa de detecção de 840 padrões em um conjunto de 42 imagens com diferentes transformações geométricas e fotométricas e seus resultados foram comparados com os do método color SIFT. Surpreendentemente, os experimentos demonstraram que o desempenho médio do color Ciratefi (75,1%) foi superior ao do color SIFT (63,0%), mesmo este último empregando um poderoso conjunto de características de cor invariantes.

Sabemos que uma comparação justa entre Ciratefi e SIFT é difícil porque ambos têm funcionamento e objetivos diferentes. O problema é a falta de métodos na literatura que compartilhem as mesmas características do método proposto. Cabe salientar que, embora SIFT não tenha como objetivo principal o reconhecimento de objetos, ele tem sido empregado em muitas aplicações com esta finalidade. Um exemplo é uma aplicação proposta por Lowe (1999).

No que tange ao tempo de processamento, em todos os experimentos descritos neste trabalho, SIFT e EasyMatch foram muito mais rápidos que o Ciratefi. Contudo, este problema pode ser contornado uma vez que o Ciratefi, e por conseqüência suas extensões, podem ter suas velocidades aceleradas em milhares de vezes via implementação em hardware, o que não acontece com o SIFT devido a complexidade das operações aritméticas e trigonométricas empregadas e também pelo fato de não repetir as mesmas operações para todos os pixels. Este fato é importante porque sinaliza a possibilidade de novas frentes de estudo visando o desenvolvimento de sistemas de visão computacional de alto desempenho e de baixo custo. Ressalta-se ainda que a robustez do Ciratefi pode ser aumentada pela

incorporação de informações de textura bem como pela busca de partes do padrão procurado visando tratar do problema da oclusão parcial.

A escolha das 217 imagens utilizadas nos experimentos com os métodos propostos foi feita com intuito de garantir a diversidade, ou seja, imagens com diferentes tipos de padrões, texturas, fundos, condições de iluminação, etc.

Por fim, cabe salientar que os resultados obtidos pelos métodos propostos neste trabalho são muito promissores e incentivam novos estudos visando suas melhorias.

## 6.1 CONTINUIDADE DO TRABALHO

Como mencionado anteriormente, Ciratefi (incluindo suas extensões) teve seu desempenho prejudicado em imagens adquiridas sob diferentes pontos de vistas e em imagens afetadas por oclusão parcial. Assim, fica o desafio de novos estudos visando contornar este problema. Tais estudos devem envolver a utilização de partes do padrão (para o caso da oclusão parcial) bem como a incorporação de algum mecanismo que permita a extração de atributos invariantes à transformação afim (para o caso de mudanças de ponto de vista).

Quanto ao Ciratefi morfológico, cabe uma investigação acerca das teorias recentes no campo da morfologia matemática visando sua melhoria.

Com relação específica ao color Ciratefi, novos estudos acerca dos invariantes de cores serão necessários para torná-lo mais robusto ao problema de constância de cor, principalmente nos casos que envolvem mudança da cor de iluminação e também variações locais de iluminação. Além disso, estudar uma forma de estimar, de forma automática, os pesos  $\delta$  e  $\eta$  que sejam adequados para o processamento de uma determinada imagem se faz necessário.

Por último, cabe investigar uma forma de diminuir a quantidade de falsos pontos chaves extraídos por SIFT em imagens que apresentam riqueza excessiva de detalhes como aquelas do conjunto III (sensoriamento remoto). Cabe ressaltar que muitos estudos neste sentido têm sido apresentados na literatura, mas são raros aqueles que descrevem aplicações de SIFT envolvendo este tipo de imagem.

## REFERÊNCIAS

ABDEL-HAKIM, A. E.; FARAG, A. A. CSIFT: A SIFT Descriptor with Color Invariant Characteristics, In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, New York, USA. **Proceedings...** New York: IEEE Computer Society, 2006, p. 1978-1983.

ADOBE Systems. Technical Guide (Color Models). Disponível em: [http://dba.med.sc.edu/price/irf/Adobe\\_tg/models/main.html](http://dba.med.sc.edu/price/irf/Adobe_tg/models/main.html). Acesso em 23 mai. 2009.

ANCUTI, C.; BEKAERT, P. SIFT-CCH: Increasing the SIFT distinctness by Color Co-occurrence Histograms. In: 5TH INTERNATIONAL SYMPOSIUM ON IMAGE AND SIGNAL PROCESSING AND ANALYSIS, 2007, Istanbul, Turkey. **Proceedings...** Istanbul: IEEE Computer Society, 2007, 130-135.

ARAÚJO, S. A.; KIM, H. Y. Rotation, scale and translation-invariant segmentation-free grayscale shape recognition using mathematical morphology. In: INTERNATIONAL SYMPOSIUM ON MATHEMATICAL MORPHOLOGY, 8., 2007, Rio de Janeiro. **Proceedings...** São José Dos Campos : MCT/INPE, 2007. v. 2. p. 61-62.

BALLARD, D. H. Generalizing the hough transform to detect arbitrary shapes. **Pattern Recognition**, v. 13, n. 2, p. 111-122, 1981.

BANON, G. J. F. Characterization of translation-invariant elementary operators for gray-level morphology. In: NEURAL, MORPHOLOGICAL AND STOCHASTIC METHODS IN IMAGE AND SIGNAL PROCESSING, 1995, San Diego, USA. **Proceedings of SPIE**. San Diego, 1995. v. 2568, p. 68-79.

BASCLE, B.; BERNIER, O.; LEMAIRE, V. Illumination-Invariant Color Image Correction. **Lecture Notes in Computer Science**, Springer-Verlag, v. 4153, p. 359-368, 2006. Apresentado ao INTERNATIONAL WORKSHOP ON INTELLIGENT COMPUTING IN PATTERN ANALYSIS/SYNTHESIS (IWICPAS), Xi'an, China, 2006.

BAY, H.; TUYTELAARS, T.; VAN GOOL, L. SURF: Speeded Up Robust Features. **Lecture Notes in Computer Science**, Springer-Verlag, v. 3951, part 1, p. 404-417. Apresentado ao 9th European Conference on Computer Vision, Graz, Austria, 2006.

BELONGIE, S.; MALIK, J.; PUZICHA, J. Shape matching and object recognition using shape contexts. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 24, p. 509-522, 2002.

BOBER, M. MPEG-7 Visual shape descriptors. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 11, n. 6, p. 716-719, 2001.

BONATO, V. **Proposta de uma arquitetura de hardware em FPGA implementada para SLAM com multi-câmeras aplicada à robótica móvel**. 2008. 168 p. Tese (Doutorado) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, São Paulo, 2008.

BONATO, V.; MARQUES, E.; CONSTANTINIDES, G. A. A Parallel Hardware Architecture for Scale and Rotation Invariant Feature Detection. **IEEE Transactions on Circuits and Systems for Video Technology**. v. 18, n.12, p. 1703-1712, 2008.

BRESENHAM, J. E. Algorithm for computer control of a digital plotter. **IBM Systems Journal**, v. 4, n. 1, p. 25-30, 1965.

BRESENHAM, J. E. A linear algorithm for incremental digital display of circular arcs. **Communications of the ACM**, v. 20 n. 2, p. 100-106, 1977.

BURGHOUTS, G. J.; GEUSEBROEK, J. M. Performance evaluation of local colour invariants. **Computer Vision and Image Understanding**, v. 113, n.1, p. 48-62, 2009.

CHANG, D. H.; HORNAK, J. P. Fingerprint recognition through circular sampling. **The Journal of Imaging Science and Technology**, v. 44, n.6, p. 560-564, 2000.

CHOI, M. S.; KIM W. Y. A novel two stage template matching method for rotation and illumination invariance. **Pattern Recognition**, v. 35, n. 1, p. 119-129, 2002.

CHOKSURIWONG, A.; LAURENT, H.; EMILE, B. Comparison of invariant descriptors for object recognition. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2005, Genova, Italy. **Proceedings...** Genova: IEEE Computer Society, 2005. P. 377-380.

CHONG, H. Y.; GORTLER, S. J.; ZICKLER, T. A perception-based color space for illumination-invariant image processing. **ACM Transactions on Graphics**, V. 27, n. 3 p 1-7, 2008.

CONCI, A.; AZEVEDO, E.; LETA, F. R. **Computação Gráfica: Teoria e Prática**. v.2, Rio de Janeiro: Campus, 2008. 407 p.

CUI, Y. et al. Scale invariant feature transform with irregular orientation histogram binning. **Lecture Notes in Computer Science**, Springer-Verlag. Apresentado ao INTERNATIONAL CONFERENCE ON IMAGE ANALYSIS AND RECOGNITION, Halifax, Canadá, 2009. Não publicado.

DIONISIO, C. R. P.; KIM, H. Y. A supervised shape classification technique invariant under rotation and scaling. In: INTERNATIONAL TELECOMMUNICATIONS SYMPOSIUM, 2002, Natal. **Proceedings...** Natal, 2002. p. 533-537.

DIONISIO, C. R. P. **Características de formas planas invariantes sob transformações de similaridade e afim**. 2005. 93 p. Tese (Doutorado) – Escola Politécnica, Universidade de São Paulo, São Paulo, São Paulo, 2005.

EURESYS. **Open eVision Image Analysis Tools**. 2007. Disponível em: <<http://www.euresys.com/Products/SoftwareDevelopment/MachineVision.asp>>. Acesso entre nov. 2007 e jul. 2008.

FANTI, G.; MAGGIOLO, R. The double superficiality of the frontal image of the Turin Shroud. **Journal of Optics A: Pure and Applied Optics**. v. 6, p. 491-503, 2004.

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**. v. 27, n. 8, p. 861-874, 2006.

FLUSSER, J.; SUK, T. Rotation moment invariants for recognition of symmetric objects. **IEEE Transactions on Image Processing**, v.15, n. 12, p. 3784-3790, 2006.

FORNASIER, M.; TONIOLO, D. Fast, robust and efficient 2D pattern recognition for re-assembling fragmented images. **Pattern Recognition**, v. 38, n. 11, p. 2074-2087, 2005.

FULTON, J. T. **Processes in Biological Vision**. Corona Del Mar, CA, 2006. Disponível em: <<http://4colorvision.com>>. Acesso em: 05 jul. 2008.

FUNT, B. V.; BARNARD, K.; MARTIN, L. Is Machine Colour Constancy Good Enough? **Lecture Notes in Computer Science**, Springer-Verlag, v. 1406, p 445-459. Apresentado ao 5TH EUROPEAN CONFERENCE ON COMPUTER VISION. Freiburg, Germany, 1998.

FUNT, B. V.; CARDEI, V. Computational Uses of Colour. In: DAVIS, S. **Color Perception: Philosophical, Psychological, Artistic and Computational Perspectives**, Vancouver, Oxford University Press. 2000. p. 1-19. Disponível em: <http://www.cs.sfu.ca/~colour/publications/VSCS2000/index.html>. Acesso em: 29 mai. 2009.

FUNT, B. V.; FINLAYSON, G. D. Color Constant Color Indexing, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 17, n. 5, p. 522-529, 1995.

GEUSEBROEK, J. M.; BURGHOUTS, G. J.; SMEULDERS, A. W. M. The Amsterdam library of object images. **International Journal of Computer Vision**, v. 61, n. 1, p. 103-112, 2005.

GEUSEBROEK, J. M. et al. Color invariance. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 23, n. 12, p. 1338-1350, 2001.

GEVERS, T.; SMEULDERS, A. Color-based object recognition, **Pattern Recognition**, v. 32, n. 3, p. 453-464, 1999.

GHOSH, A.; PETKOV, N. Robustness of shape descriptors to incomplete contour representations. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 27, n. 11, p. 1793-1804, 2005.

GIJSENIJ A.; GEVERS, T. Color Constancy using Image Regions. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, San Antonio, Texas, 2007. **Proceedings...** San Antonio: Signal Processing Society, 2007, p. 501-504.

GIJSENIJ, A.; GEVERS, T.; LUCASSEN, M. P. A Perceptual Comparison of Distance Measures for Color Constancy Algorithms. **Lecture Notes in Computer Science**, Springer-Verlag, v. 5302, p. 208-221. Apresentado ao 10th European Conference on Computer Vision, Marseille, France, 2008.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. Massachusetts: Addison-Wesley, 1992, 716 p.

GOSHTASBY, A. Description and discrimination of planar shapes using shape matrices. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 7, n.6, p. 738-743, 1985.

HA, T. M.; ZIMMERMANN, M.; BUNKE, H. Off-line handwritten numeral string recognition by combining segmentation-based and segmentation-free methods. **Pattern Recognition**, v. 31, n. 3, p. 257-272, 1998.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: IV Alvey Vision Conference, 1988, Manchester, UK. **Proceedings...** Manchester, 1988, p. 147-151.

HOUGH, P. V. C. Methods and Means for Recognizing Complex Patterns, U.S. Patent 3069654, 1962.

HSE, H.; NEWTON, A. R. Sketched symbol recognition using Zernike moments. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 17., 2004, Cambridge, UK. **Proceedings...** Cambridge: IEEE CS Press, 2004. v. 1, p. 367-370.

HSU, Y. N.; ARSENAULT, H. H.; APRIL, G. Rotation-invariant digital pattern recognition using circular harmonic expansion. **Applied Optics**, v. 21, n. 22, p. 4012-4015, 1982.

HU, M. K. Visual pattern recognition by moment invariants. **IRE Transactions on Information Theory**, v. 1, n. 8, p. 179-187, 1962.

HUTCHINSON, S.; HAGER, G. D.; CORKE, P. I. A tutorial on visual servo control. **IEEE Transactions on Robotics and Automation**, v. 13, n. 5, p. 651-670, 1996.

HWANG, S. K.; KIM, W. Y. Fast and efficient method for computing ART. **IEEE Transactions on Image Processing**, v. 15, n. 1, p. 112-117, 2006.

INTEL. **OpenCV - Open Source Computer Vision Library**. 2000. Disponível em: <<http://www.intel.com/technology/computing/opencv/>>. Acesso em: 09 out. 2007.

JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical Pattern Recognition: A review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 1, p. 4-37, 2000.

JALBA, A. C; WILKINSON, M. H. F.; ROERDINK, J. B. T. M. Shape representation and recognition through morphological curvature scale spaces. **IEEE Transactions on Image Processing**, v.15, n. 2, p. 331- 341, 2006.

KE, Y.; SUKTHANKAR, R. PCA-SIFT: A more distinctive representation for local image descriptors. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2004, Washington, USA. **Proceedings...** Washington: IEEE Computer Society, 2004, p. 511- 517.

KERMINEN, P.; GABBOUJ, M. The visual goodness evaluation of color-based retrieval processes. In: SIGNAL PROCESSING X–THEORIES AND APPLICATIONS Tampere, Finland, 2000. **Proceedings of EUSIPCO 2000**, Tampere, 2000, p. 2153-2156.

KHOTANZAD, A.; HONGS, Y. H. Invariant image recognition by Zernike moments. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 12, n. 5, p. 489-497, 1990.

KIM, H. Y. **IMG - Rotinas e programas em C++ para processamento de imagens e visão computacional**. São Paulo, 2004. Disponível em: <<http://www.lps.usp.br/~hae/software>>. Acesso entre maio 2004 e set. 2006.

KIM, H. Y. **ProEikon - Rotinas e programas em C++ para processamento de imagens e visão computacional**. São Paulo, 2006. Disponível em: <<http://www.lps.usp.br/~hae/software>>. Acesso entre out. 2006 e fev. 2008.

KIM, H. Y. Segmentation-free printed character recognition by relaxed nearest neighbor learning of windowed operator. In: XII BRAZILIAN SYMPOSIUM ON COMPUTER GRAPHICS AND IMAGE PROCESSING (SIBGRAPI), 12., 1999, Campinas, São Paulo. **Proceedings...** Campinas, 1999. p. 195-204.

KIM, H. Y.; ARAÚJO, S. A. Grayscale Template-Matching Invariant to Rotation, Scale, Translation, Brightness and Contrast. **Lecture Notes in Computer Science**, Springer-Verlag, v. 4872, p. 100-113, 2007. Apresentado ao Second Pacific-Rim Symposium on Image and Video Technology (PSIVT), Chile, 2007.

KIM, W. Y.; KIM, Y. S. A region-based shape descriptor using Zernike moments. **Signal Processing: Image Communication**, v. 16, n. 1, p. 95-102, 2000.

KIM, W. Y.; YUAN, P. A practical pattern recognition system for translation, scale and rotation invariance. In: COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 1994, Seattle, USA. **Proceedings...** Seattle: IEEE Computer Society, 1994. p. 391-396.

KNIGHT, W, **Video search makes phone a 'second pair of eyes'**, 2007. Disponível em: <<http://technology.newscientist.com/article/dn12831-video-search-makes-phone-a-second-pair-of-eyes.html>>. Acesso em: 12 jul. 2008.

LAMDAN, Y.; WOLFSON, H. J. Geometric hashing: a general and efficient model-based recognition scheme. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2., 1988, Florida, USA. **Proceedings...** Flórida, 1988. p. 238-249.

LEEMANS, V.; DESTAIN M. F. A real-time grading method of apples based on features extracted from defects. **Journal of Food Engineering**, v. 61, n. 1, p. 83-89, 2004.

LEWIS, H. R.; PAPADIMITRIOU, C. H. **Elementos de Teoria da Computação**. 2. ed. Porto Alegre: Bookmann, 2000, 344 p.

LI, J. H. et al. Image recognition based on invariant moment in the projection space. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND CYBERNETICS, 3., 2004, Shanghai, China. **Proceedings...** Shanghai: IEEE Computer Society, 2004. v. 6, p. 3606-3610.

LONCARIC, S. A. Survey of Shape Analysis Techniques. **Pattern Recognition**, v. 31, n. 8, p. 983-1001, 1998.

LONG, F.; ZHANG, H.; FENG, D. Fundamentals of Content-based Image Retrieval. In: FENG, D.; SIU, W. C.; ZHANG, H. **Multimedia Information Retrieval and Management: Technological Fundamentals and Applications (Signals and Communication Technology)**, Berlin, 2005. p. 1-26.

LOWE, D. G. Object recognition from local scale-invariant features. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION, 7., 1999, Corfu, Greece. **Proceedings...** Corfu: IEEE Computer Society, 1999. p. 1150-1157.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91-110, 2004.

LOWE, D. G. **Demo Software: SIFT Keypoint Detector**. Vancouver, Canada, 2005. Disponível em <<http://www.cs.ubc.ca/~lowe/keypoints/>>. Acesso em: 09 out. 2007.

LUQIAO, F. et al. A Single-hand and binocular visual system for EOD robot automation and logistics. In: IEEE INTERNATIONAL CONFERENCE ON

AUTOMATION AND LOGISTICS, 2007, Jinan, China. **Proceedings...** Jinan: IEEE Computer Society, 2007, p. 1930-1935.

MAO, W.; ZHANG, T.; WANG, L. Detection of text in images using SUSAN edge detector. **Journal of Harbin Institute of Technology**, v. 12, part 1, p. 34-40, 2005.

MARIMON, D.; EBRAHIMI, T. Efficient Rotation-Discriminative Template Matching, **Lecture Notes in Computer Science**, v. 4756, p. 221-230, 2008. Apresentado ao Iberoamerican Congress on Pattern Recognition (CIARP07), Viña del Mar-Valparaiso-Chile, 2007.

MEYENHOFER, F. et al. Automatic Analysis of microRNA microarray images using mathematical morphology. In: 29<sup>th</sup> ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 29., 2007, Lyon, France. **Proceedings...** Lyon: IEEE Computer Society, 2007. p. 6235-6238.

MIKOLAJCZYK, K. et al. A comparison of affine region detectors. **International Journal of Computer Vision**, v. 65, n. 1-2, p. 43-72, 2005.

MIKOLAJCZYK, K.; SCHMID, C. A performance evaluation of local descriptors, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 10, n. 27, p. 1615-1630, 2005.

MOERLAND, T.; JURIE, F. Learned Color Constancy From Local Correspondences. In: IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA & EXPO, 2005, Amsterdam, Netherlands. **Proceedings...** Amsterdam: IEEE Computer Society, 2005.

MOKHTARIAN, F.; ABBASI, S.; KITTLER, J. Robust and efficient shape indexing through curvature scale space. In: SEVENTH BRITISH MACHINE VISION CONFERENCE (BMVC'96), 7., 1996, Edinburgh, Scotland. **Proceedings...** Edinburgh, 1996. v. 1, p. 53-62.

MOKHTARIAN, F.; MACKWORTH, A. K. A Theory of multi-scale, curvature based shape representation for planar curves. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 14, n. 8, p. 789-805, 1992.

MORI, G.; BELONGIE, S.; MALIK, J. Shape contexts enable efficient retrieval of similar shapes. In: COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2001, Kauai Marriott, Hawaii. **Proceedings...** Kauai Marriott, Hawaii: IEEE Computer Society, 2001. p. 723-730.

MORI, G.; BELONGIE, S.; MALIK, J. Efficient Shape matching using shape contexts. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 27, n. 11, p. 1832-1837, 2005.

NAEGEL, B. et al. Segmentation using vector-attribute filters: methodology and application to dermatological imaging. In: INTERNATIONAL SYMPOSIUM ON MATHEMATICAL MORPHOLOGY (ISMM), 8., 2007, Rio de Janeiro. **Proceedings...** São José dos Campos: INPE, 2007. v. 1, p. 239-250.

NAEGEL, B.; PASSAT, N.; RONSE, C. Grey-level hit-or-miss transforms - Part I: Unified theory. **Pattern Recognition**, v. 40, n. 2, p. 635-647, 2007.

NOBRE, H. P. A.; KIM, H. Y. Automatic VHDL Generation for Solving Rotation and Scale-Invariant Template Matching in FPGA. In: V SOUTHERN PROGRAMMABLE LOGIC CONFERENCE, São Carlos, Brazil, 2009. **Proceedings...** São Carlos: USP, 2009, p. 21-26.

PARK, B. G.; LEE, K. M.; LEE, S. U. Color-based Image Retrieval Using Perceptually Modified Hausdorff Distance. **EURASIP Journal Image and Video Processing**, v. 2008, p. 1-10, 2008.

PINA, P.; BARATA, T.; BANDEIRA, L. P. C. Morphological recognition of the spatial patterns of olive trees. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 18., 2006, Hong Kong. **Proceedings...** Hong Kong: IEEE Computer Society, 2006. v. 4, p. 845-848.

RETORNAZ, T.; MARCOTEGUI, B. Scene text localization based on the ultimate opening. In: INTERNATIONAL SYMPOSIUM ON MATHEMATICAL MORPHOLOGY (ISMM), 8., 2007, Rio de Janeiro. **Proceedings...** São José dos Campos: INPE, 2007. v. 1, p. 177-188.

ROORDA, A. Human Visual System - Image Formation. In: HORNACK, J. P. (Ed.) **Encyclopedia of Imaging Science and Technology**, New York: Wiley & Sons, 2002, p. 539-557. Disponível em: <http://vision.berkeley.edu/roordalab/Pubs/EISTChapterRoorda.pdf>. Acesso em: 21 jun. 2009.

ROSTEN, E.; DRUMMOND, T. Fusing points and lines for high performance tracking. In: IEEE International Conference on Computer Vision, 2005, Beijing, China. **Proceedings...** Beijing: IEEE Computer Society, 2005, p.1508-1511.

RUBNER, Y.; GUIBAS, L.; TOMASI, C. The earth mover's distance, multi-dimensional scaling, and color-based image retrieval, In: ARPA IMAGE UNDERSTANDING WORKSHOP, New Orleans, LA, 1997. **Proceedings...** New Orleans, 1997, p. 661-668.

RUSSEL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. New Jersey: Prentice Hall, 1995. 932 p.

SALEMBIER, P. Overview of the MPEG-7 standard and of future challenges for visual information analysis. **Journal on Applied Signal Processing**, v. 2002, n. 4, p. 343-353, 2002.

SCHAEFER, G. How useful are colour invariants for image retrieval?. **Computational Imaging and Vision**, v. 32, p. 381-386, 2005.

SMITH, S. M. Flexible filter neighbourhood designation. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 13., 1996, Vienna, Austria. **Proceedings...** Vienna: IEEE Computer Society, 1996. v. 1, p. 206-212.

SMITH, S. M.; BRADY, J. M. SUSAN - A new approach to low level image processing. **International Journal of Computer Vision**, v. 23, n. 1, p. 45-78, 1997.

SORGI, L.; DANIILIDIS, K. Normalized Cross-Correlation for Spherical Images. **Lecture Notes in Computer Science**, Springer-Verlag, v. 3022, p. 542-553, 2004. Apresentado ao 8th European Conference on Computer Vision, Prague, Czech Republic, 2004.

SWAIN, M. J.; BALLARD, D. H. Color Indexing. **International Journal of Computer Vision**, v. 7, n.1, p. 11-32, 1991.

TAO, Y.; LOERGER T. R.; TANG, Y. Y. Extraction of rotation invariant signature based on fractal geometry. In: IEEE INTERNATIONAL CONFERENCE IMAGE PROCESSING, 1., 2001, Thessaloniki, Greece. **Proceedings...** Thessaloniki: IEEE Computer Society, 2001. v. 1, p. 1090-1093.

TAO, Y.; TANG, Y. Y. The feature extraction of chinese character based on contour information. In: INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS RECOGNITION, 1999, Bangalore, India. **Proceedings...** Bangalore: IEEE Computer Society, 1999. p. 637-640.

TAZA, A.; SUEN, C. Y. Discrimination of planar shapes using shape matrices. **IEEE Transactions on Systems Man and Cybernetics**, v. 19, n. 5, p. 1281-1289, 1989.

TEH, C. H.; CHIN, R. T. On image analysis by the methods of moments. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 10, n. 4, p. 496-513, 1988.

TORRES, R. da S.; FALCÃO, A. X.; COSTA, L. da F. A graph-based approach for multiscale shape analysis, **Pattern Recognition**, v. 37, n. 6, p. 1163–1174, 2004.

TORRES, R. da S.; FALCÃO, A. X. Contour salience descriptors for effective image retrieval and analysis, **Image and Vision Computing**, v. 25, n. 1, p. 3–13, 2007.

TORRES-MENDEZ, L. A. et al. Translation, rotation and scale-invariant object recognition. **IEEE Transactions Systems, Man and Cybernetics - part C: Applications and Reviews**, v. 30, n. 1, p. 125-130, 2000.

TSAI, D. M.; TSAI, Y. H. Rotation-invariant pattern matching with color ring-projection. **Pattern Recognition**, v. 35, n. 1, p. 131-141, 2002.

TUYTELAARS, T.; MIKOLAJCZYK, K. Local invariant feature detectors: a survey. **Foundations and Trends. Computer Graphics and Vision**, v. 3, n. 3, p. 177-280, 2008.

ULLAH, F.; KANEKO, S. Using orientation codes for rotation-invariant template matching. **Pattern Recognition**, v. 37, n. 2, p. 201-209, 2004.

ULRICH, M.; STEGER, C.; BAUMGARTNER, A. Real-Time Object Recognition using a Modified Generalized Hough Transform. **Pattern Recognition**, v. 36, n. 11, p. 2557-2570, 2003.

VAN DE SANDE, K. E. A.; GEVERS, T.; SNOEK, C. G. M. Evaluation of color descriptors for object and scene recognition. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. 2008, Anchorage, Alaska. **Proceedings...** Anchorage:IEEE Computer Society, 2008. p.1-8.

VELTKAMP, R. C.; LATECKI, L. J. Properties and performance of shape similarity measures. In: INTERNATIONAL CONFERENCE ON DATA SCIENCE AND CLASSIFICATION (IFCS), 2006, Slovenia. **Proceedings...** Slovenia, 2006. p. 1-9.

WEIJER, J.; GEVERS, T.; GIJSENIJ, A. Edge-based Color Constancy. **IEEE Transactions on Image Processing**, v. 16, n. 9, p. 2207-2214, 2007.

WOLF, C.; JOLION, J. M. Object count/area graphs for the evaluation of object detection and segmentation algorithms. **International Journal on Document Analysis and Recognition**, v. 8, n. 4, p. 280-296, 2006.

WOLFSON, H. J.; RIGOUTSOS, I. Geometric hashing: an overview. **IEEE Computational Science & Engineering**, v. 4, n. 4, p. 10-21, 1997.

WU, J. C.; HSIEH, J. W.; CHEN, Y. S. Morphology-based text line extraction. **Machine Vision and Applications**, v. 19, n. 3, p. 195-207, 2008.

XIN, Y.; PAWLAK, M.; LIAO, S. X. Accurate computation of Zernike moments in polar coordinates. **IEEE Transactions on Image Processing**, v. 16, n. 2, p. 581-587, 2007.

YU, L. et al. Coarse iris classification using box-counting to estimate fractal dimensions. **Pattern Recognition**, v. 38, n. 11, p. 1791-1798, 2005.

YU, L.; WANG, R. Shape representation based on mathematical morphology. **Pattern Recognition Letters**, v. 26, n. 9, p. 1354-1362, 2005.

YU-QIAN, Z. et al. Medical images edge detection based on mathematical morphology. In: 27<sup>TH</sup> ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 27., 2005, Shanghai, China. **Proceedings...** Shanghai: IEEE Computer Society, 2005. p. 6492-6495.

YUEN, P. C.; FENG, G. C.; TANG, Y. Y. Printed Chinese character similarity measurement using ring projection and distance transform. **International Journal of Pattern Recognition and Artificial Intelligence**, v. 12, n 2, p. 209-221, 1998.

ZHANG, D.; LU G. Review of shape representation and description techniques. **Pattern Recognition**, v. 37, n.1, p. 1-19, 2004.

ZHONG, B.; LIAO, W. A hybrid method for fast computing the curvature scale space image. In: INTERNATIONAL CONFERENCE ON GEOMETRIC MODELLING AND PROCESSING, 2004, Beijing, China. **Proceedings...** Beijing, 2004. p. 124-130.

## APÊNDICE A – PUBLICAÇÕES DO AUTOR DURANTE O PERÍODO DO DOUTORADO

ARAÚJO, S. A. ; KIM, H. Y. Rotation, scale and translation-invariant segmentation-free grayscale shape recognition using mathematical morphology. In: INTERNATIONAL SYMPOSIUM ON MATHEMATICAL MORPHOLOGY, 8., 2007, Rio de Janeiro, Brazil. **Proceedings...** São José Dos Campos : MCT/INPE, 2007. v. 2. p. 61-62.

ARAÚJO, S. A. ; KIM, H. Y. Meio-Tom Inverso Usando Redes Neurais Artificiais. In: XXII SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, 2005, Campinas, São Paulo. **Anais do XXII Simpósio Brasileiro de Telecomunicações**, Campinas, São Paulo: SBRT, 2005. v. 1, p. 226-231.

KIM, H. Y.; ARAÚJO, S. A. Grayscale Template-Matching Invariant to Rotation, Scale, Translation, Brightness and Contrast. **Lecture Notes in Computer Science**, Springer-Verlag, v. 4872, p. 100-113, 2007. Apresentado ao Second Pacific-Rim Symposium on Image and Video Technology (PSIVT), Chile, 2007.