

UNIVERSIDADE DO VALE DO RIO DOS SINOS  
CIÊNCIAS EXATAS E TECNOLÓGICAS  
CURSO DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

**Utilização da Informação de Cor na  
Obtenção de Descritores Locais para  
o Registro de Imagens**

por

RAFAEL GUTERRES JEFFMAN

Dissertação submetida a avaliação  
como requisito parcial para a obtenção do grau de  
Mestre em Computação Aplicada

Prof Dr. Fernando S. Osório  
Orientador

São Leopoldo, agosto de 2005

**CIP — CATALOGAÇÃO NA PUBLICAÇÃO**

Jeffman, Rafael Guterres

Utilização da Informação de Cor na Obtenção de Descritores Locais para o Registro de Imagens / por Rafael Guterres Jeffman. — São Leopoldo: Ciências Exatas e Tecnológicas da UNISINOS, 2005.

61 f.: il.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos. Ciências Exatas e Tecnológicas. Curso de Pós-Graduação em Computação Aplicada São Leopoldo, BR-RS, 2005. Orientador: Osório, Fernando S..

I. Osório, Fernando S.. II. Título.

UNIVERSIDADE DO VALE DO RIO DOS SINOS

Reitor: Dr. Aloysio Bohnen

Diretora da Unidade de Pós-Graduação e Pesquisa: Prof<sup>a</sup>. Dr<sup>a</sup>. Ione Bentz

Coordenador do PIPCA: Prof. Dr. Arthur Tórgo Gómez

# Agradecimentos

Sempre é difícil agradecer ao diverso número de pessoas que auxiliam durante um trabalho longo como a realização de um mestrado, por isso começo agradecendo a todos aqueles que de uma forma ou outra participaram e me ajudaram durante este tempo.

Uma das pessoas que mais fico em dívida agora é o meu orientador, já de longa data, desde a iniciação científica, passando pelo trabalho de conclusão de graduação e agora no mestrado, Prof. Dr. Fernando Santos Osório. Sem sombra de dúvida, este trabalho não teria chegado ao fim não fosse seu apoio e dedicação.

Agradeço também todo o apoio e compreensão da coordenação do curso, com destaque para o Prof. Dr. Arthur Gomez, e a secretária do curso, Rejane Weissheimer. Muito obrigado por tudo.

A minha família, pelo apoio e ajuda mesmo quando nem eu acreditava em mais nada. A minha esposa Taís e a minha filha Gabriela e a minha avó Edy, meu amor eterno.

Por fim, agradeço a Universidade do Vale do Rio dos Sinos pelo apoio financeiro na forma de bolsa de estudos, sem o qual não seria possível a realização deste mestrado.

# Sumário

<b>Lista de Figuras</b>	<b>6</b>
<b>Lista de Tabelas</b>	<b>8</b>
<b>Resumo</b>	<b>9</b>
<b>Abstract</b>	<b>10</b>
<b>1 Introdução</b>	<b>11</b>
1.1 Objetivos . . . . .	14
<b>2 Percepção de Cores</b>	<b>18</b>
2.1 Espaços de Cores . . . . .	19
2.1.1 CIE XYZ . . . . .	20
2.1.2 CIE RGB . . . . .	20
2.1.3 Os modelos HSI . . . . .	21
2.1.4 CIE $L^*a^*b^*$ . . . . .	22
2.2 Temperatura de Cor . . . . .	23
2.3 Utilização da Cromaticidade no Registro de Imagens . . . . .	24
<b>3 Registro de Imagens</b>	<b>26</b>
3.1 Extração de Características . . . . .	26
3.2 Descritores Globais . . . . .	27
3.2.1 Histograma de Cores . . . . .	27
3.2.2 Métodos Baseados em Correlação . . . . .	28
3.3 Descritores Locais . . . . .	30
3.3.1 Pontos de Interesse . . . . .	30
3.3.2 <i>Harris Corner Detection</i> . . . . .	30
3.3.3 <i>Scale Invariant Feature Transform</i> (SIFT) . . . . .	31
3.4 Comparação de características . . . . .	37
3.5 Estimativa da transformação do modelo . . . . .	37
3.6 Transformação da Imagem . . . . .	38
3.7 Transformações geométricas entre duas imagens . . . . .	39
3.7.1 Transformações <i>Affine</i> . . . . .	39
3.8 Iluminação e Cor . . . . .	40
3.9 Utilização da Cor em Descritores Locais . . . . .	40

<b>4</b>	<b>Obtenção de descritores locais utilizando a informação de cor</b>	<b>41</b>
4.1	Base de Dados de Imagens . . . . .	42
4.2	Número de pontos obtidos . . . . .	43
4.3	Robustez dos pontos obtidos com a informação de cor . . . . .	45
4.3.1	Transformações <i>Affine</i> e Perspectiva . . . . .	45
4.3.2	Posição da Fonte de Luz . . . . .	47
4.3.3	Temperatura de Cor . . . . .	48
4.4	Problemas com a presença de ruído nas imagens . . . . .	50
<b>5</b>	<b>Conclusão e Perspectivas Futuras</b>	<b>53</b>
5.1	Perspectivas Futuras . . . . .	54
<b>A</b>	<b>Objetos Utilizados nos Experimentos</b>	<b>56</b>
	<b>Bibliografia</b>	<b>57</b>

# Lista de Figuras

FIGURA 1.1 – Indexação e recuperação de imagens, por similaridade. . . .	12
FIGURA 1.2 – Reconhecimento de objetos em imagens complexas. . . . .	12
FIGURA 1.3 – Sistema de Navegação Visual. . . . .	13
FIGURA 1.4 – Exemplo de fotografia da iris.[CSBR, 2005] . . . . .	13
FIGURA 1.5 – Geração automática de panoramas. . . . .	16
FIGURA 1.6 – Exemplo de aplicação de SIFT. . . . .	17
FIGURA 2.1 – Curvas de resposta espectral do olho humano de acordo com o comprimento de onda (nm). [Gomes and Velho, 1994] . . . . .	18
FIGURA 2.2 – Cubo RGB . . . . .	21
FIGURA 2.3 – Cone HSV . . . . .	22
FIGURA 2.4 – Exemplo de objeto com variação da temperatura de cor da fonte de luz. [Geusebroek et al., 2005] . . . . .	24
FIGURA 3.1 – Histogramas de um mesmo objeto alterados pela inclusão de artefatos. . . . .	28
FIGURA 3.2 – Resultado da correlação de uma janela com uma imagem. [Righes, 2004] . . . . .	29
FIGURA 3.3 – Pontos de interesse detectados com a utilização do <i>Harris Corner Detector</i> . . . . .	31
FIGURA 3.4 – Pirâmide para obtenção das diferenças de Gaussianas. . . .	32
FIGURA 3.5 – Detecção de máximos e mínimos das diferenças de Gaussi- anas. . . . .	33
FIGURA 3.6 – Navegação Visual com a utilização de descritores invarian- tes à escala. . . . .	36
FIGURA 3.7 – Obtenção dos descritores SIFT a partir dos histogramas de orientação. Neste exemplo são utilizadas regiões 2x2 e 8 <i>bins</i> para os histogramas. . . . .	37
FIGURA 3.8 – Exemplos de diversas funções de mapeamento. No sen- tido horário similaridade, <i>affine</i> , perspectiva e transformação elástica. [Zitova and Flusser, 2003] . . . . .	38
FIGURA 3.9 – Transformações <i>affine</i> . . . . .	39
FIGURA 4.1 – Pontos obtidos com a intensidade de luz e os canais V e H.	44
FIGURA 4.2 – Exemplo de objeto com variação da posição da fonte de luz.	47
FIGURA 4.3 – Comparação entre imagens com problema de calibragem da temperatura de cor e sua versão em escala de cinzas. . . . .	48
FIGURA 4.4 – Ruído presente no canal H. . . . .	50
FIGURA 4.5 – Baixo contraste apresentado no canal $b^*$ . . . . .	51

FIGURA 4.6 – Ruído apresentado no canal $b^*$ , após a equalização do histograma da imagem para aumentar o contraste. . . . .	52
FIGURA A.1 – Imagens dos objetos utilizados nos experimentos. . . . .	56

# Lista de Tabelas

TABELA 2.1 – Valores de temperatura de cor para fontes de luz comuns. . . . .	24
TABELA 4.1 – Média e desvio padrão do número total de descritores complementares obtidos entre os canais. . . . .	43
TABELA 4.2 – Média e desvio padrão do número total de descritores obtidos em cada canal e percentual médio de pontos complementares. . . . .	44
TABELA 4.3 – Média e desvio padrão do percentual de descritores similares encontrados na comparação entre uma imagem de um objeto sem alteração a mesma imagem rotacionada $30^\circ$ . . . . .	45
TABELA 4.4 – Média e desvio padrão do percentual de descritores similares encontrados na comparação entre uma imagem de um objeto sem alteração a mesma imagem com escala um terço menor. . . . .	45
TABELA 4.5 – Média e desvio padrão do percentual de descritores similares encontrados na comparação entre uma imagem de um objeto sem alteração a mesma imagem rotacionada $30^\circ$ e com escala um terço menor. . . . .	46
TABELA 4.6 – Média e desvio padrão do percentual de descritores encontrados na comparação entre uma imagem de um objeto sem alteração a mesma imagem com o objeto em um ângulo de $15^\circ$ em relação a câmera. . . . .	46
TABELA 4.7 – Percentual médio do número de descritores semelhantes encontrados entre duas imagens, em relação a transformações nas imagens. . . . .	46
TABELA 4.8 – Média e desvio padrão do percentual de descritores similares encontrados na comparação entre uma imagem de um objeto sem alteração e as imagens do mesmo objeto com alteração na posição da fonte de luz. . . . .	47
TABELA 4.9 – Média e desvio padrão do percentual de descritores encontrados na comparação entre uma imagem de um objeto sem alteração e as imagens do mesmo objeto com alteração na temperatura de cor da fonte de luz. . . . .	49
TABELA 4.10 – Variação na comparação de descritores obtidos através da informação de cor frente a alteração da temperatura de cor. . . . .	49

# Resumo

Uma imagem oferece uma grande quantidade de informação. O olho humano é responsável por aproximadamente 80% dos dados que um ser humano recebe do ambiente. Sistemas de visão computacional tentam dar a máquinas a capacidade de obter informações a partir de imagens. No entanto, esta informação não está prontamente disponível, pois uma imagem capturada por um dispositivo sofreu diversas transformações, incluindo, mas não limitadas a, transformações *affine*, transformações de projeção, transformações de cor e luz.

O registro de imagens tem se mostrado muito importante em sistemas de visão computacional, e consiste em encontrar uma função de mapeamento entre os pontos de duas imagens. Entre as técnicas de registro de imagens, o registro baseado em características locais, ou descritores locais, tem recebido muita atenção ultimamente.

Apesar do sistema de visão humano ser baseado em cores, a maioria dos sistemas de registro de imagens baseados em descritores locais trabalha com imagens em escalas de cinza onde apenas a intensidade de luz é levada em consideração.

A pesquisa desenvolvida para este trabalho está focada no estudo de técnicas do estado da arte relacionadas a obtenção de descritores locais baseados em características invariantes à escala, onde este trabalho busca dar sua contribuição nesta área através de um estudo sobre a utilização e o impacto do informação de cor no registro de imagens. Como conclusão deste trabalho, baseado em experimentos práticos, será demonstrado que a utilização da informação de cor na obtenção de descritores locais permite complementar o processo de registro de imagens com um aumento no número de pontos obtidos. Será demonstrado também a necessidade da aplicação de correções às imagens quando da utilização da cor para a obtenção de descritores locais.

**Palavras-chave:** Registro de Imagens, Processamento de Imagens Coloridas, Descritores Locais.

**TITLE:** “USE OF COLOR INFORMATION TO OBTAIN LOCAL DESCRIPTORS FOR IMAGE REGISTRATION”

## Abstract

An image offers a huge amount of information. The human eye is responsible for nearly 80% of the input data humans receive from the environment. Computer vision systems try to give machines the ability to gather information from images. Although, this information is not readily available, since a image captured by a device has gone through several transforms, including, but not limited to affine transforms, projection transforms, light and color transforms.

Image registration has been shown to be very important in Computer vision systems, consists in finding a mapping function between the points of two images. Among image registration techniques, the registration based on local features, or local descriptors, have recently gain a lot of attention lately.

Although the human visual system works based on color, most of the image registration systems based on local descriptors works with gray scale images where only the light intensity is taken into consideration.

The research developed for this work is focused on obtaining local descriptors based on invariant scale features using color information. It will be shown that color information can be used for such task and may improve the results of a image registration system.

**Keywords:** Image Registration, Color Image Processing, Local Descriptors.

# Capítulo 1

## Introdução

Sistemas de visão computacional têm o objetivo de recuperar informações a partir de imagens. Como as imagens são projeções bi-dimensionais de um mundo tri-dimensional, a informação não está diretamente disponível. Para recuperar as informações contidas em uma imagem, pode ser necessária a inversão das alterações sofridas pelos objetos existentes nessa imagem. Entre as diversas aplicações de sistemas de visão computacional, destacam-se a recuperação de imagens e vídeos em bancos de dados multimídia [Felipe et al., 2003, Ravela et al., 1996, Bender, 2003], reconhecimento e localização de objetos [Rothganger et al., 2003, Lowe, 1999], navegação visual de robôs [Lowe et al., 2002, Jones et al., 1997, Matsumoto et al., 1996], reconstrução de modelos tri-dimensionais de objetos a partir de imagens [Rothganger et al., 2003], sensoriamento remoto [Chalermwat and El-ghazawi, 1999], busca de diferenças entre imagens obtidas em momentos diferentes e/ou sob condições diferentes [Hu et al., 2004], análise de imagens médicas [Makela et al., 2002, Hill et al., 2001], geração automática de panoramas [Brown and Lowe, 2003].

Com o aumento significativo de bases de dados formadas por imagens e vídeos, e a necessidade de indexação dessas bases de dados para consultas mais eficientes, são utilizados sistemas de visão computacional para indexar e recuperar imagens com relação ao seu conteúdo. Diversos atributos podem ser utilizados para essa tarefa, sendo cor e textura, dois atributos bastante utilizados. A Figura 1.1 [Bender, 2003] mostra um exemplo de consulta a um sistema de recuperação de imagens baseado em conteúdo – CBIR (*Content Based Image Retrieval*). Sistemas de recuperação e indexação de imagens podem utilizar tanto descritores globais da imagem [Bender, 2003, Faloutsos et al., 1994], como descritores locais [Gouet and Boujemaa, 2001].

No reconhecimento de objetos em imagens complexas, o objetivo é encontrar objetos em imagens que contém outros elementos onde o objeto alvo pode estar parcialmente ocluso, em uma orientação diferente da esperada, em outra escala de tamanho, ou ainda deformado devido a projeção 3D da cena. A Figura 1.2 [Lowe, 2004] mostra o resultado de uma aplicação de reconhecimento de objetos. À esquerda da figura, encontram-se as imagens de referência para o reconhecimento. Ao centro a cena em que os objetos devem ser localizados. À direita, o resultado da localização dos objetos. Os retângulos maiores representam a borda da imagem de referência de acordo com os parâmetros obtidos durante o registro de imagem. Os retângulos menores representam os pontos a partir dos quais os objetos foram

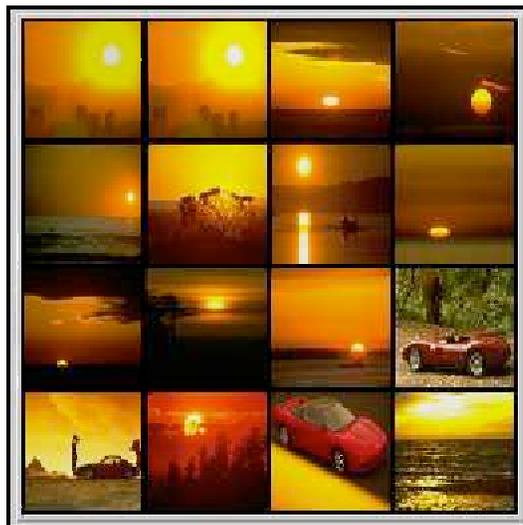


FIGURA 1.1 – Indexação e recuperação de imagens, por similaridade.

reconhecidos. [Lowe, 2004]



FIGURA 1.2 – Reconhecimento de objetos em imagens complexas.

Os sistemas de navegação visual usualmente adquirem imagens de uma única câmera monocromática [Matsumoto et al., 1999], mas podem também ser adaptados para usar visão estéreo [Jones et al., 1997], câmeras omnidirecionais [Matsumoto et al., 1999], e câmeras com aquisição de imagens coloridas [Righes, 2004]. O objetivo destes sistemas é buscar imitar um comportamento típico dos seres humanos: uma pessoa é capaz de ser conduzida por um caminho, armazenar em sua memória o caminho percorrido, e em um momento posterior, realizar de modo autônomo novamente este mesmo caminho, baseando-se em suas lembranças da paisagem e de alguns pontos de referências memorizados. Estes sistemas de navegação visual usualmente são compostos de uma base de imagens, ou seja, uma seqüência de imagens capturadas em intervalos regulares, que descreve o caminho a ser percorrido. Esta seqüência pode ser "anotada", incluindo a associação de ações a serem realizadas quando uma certa posição for alcançada [Jung et al., 2005] (Figura 1.3).

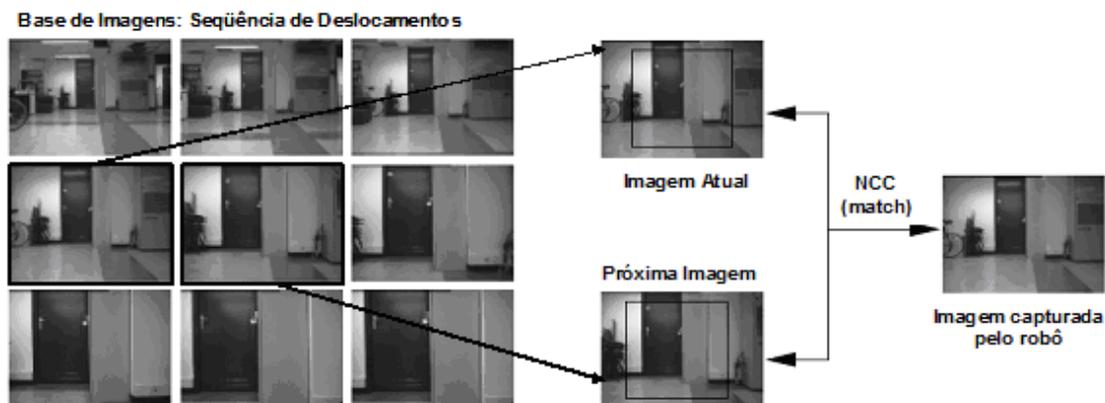


FIGURA 1.3 – Sistema de Navegação Visual.

Sistemas de biometria baseados na análise de imagens da íris humana fazem uso do registro de imagens a fim de identificar similaridades entre os artefatos encontrados em uma imagem capturada da íris e uma imagem candidata [Daugman, 1993, Vatsa et al., 2004]. Como pode ser visto na Figura 1.4, as imagens possuem uma série de artefatos que as caracterizam. Usualmente as características são extraídas a partir de imagens monocromáticas [Daugman, 1993], porém na imagem de exemplo, é possível notar artefatos coloridos. Muitos trabalhos utilizam como base de dados de referência a base monocromática CASIA do CSBR [CSBR, 2005].

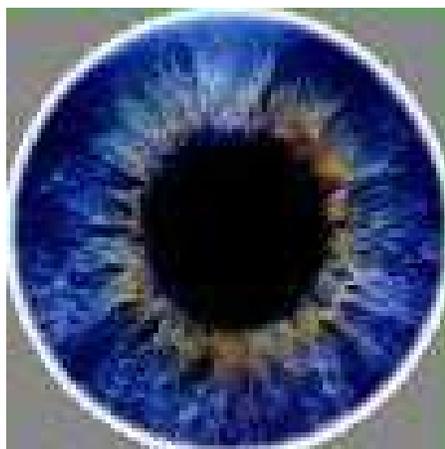


FIGURA 1.4 – Exemplo de fotografia da íris.[CSBR, 2005]

A geração de panoramas a partir de imagens pode ser automatizada utilizando sistemas de registro de imagem. Neste tipo de aplicação, é necessário encontrar o mapeamento entre diversas imagens, considerando possíveis diferenças de iluminação, distância focal, translação, rotação, entre outras. Descritores locais foram utilizados neste tipo de aplicação com grande sucesso [Brown and Lowe, 2003]. A Figura 1.5 mostra um exemplo de geração automática de panoramas.

Em todas estas aplicações, nota-se a necessidade de comparação de características semelhantes entre imagens. Normalmente, a simples comparação dos *pixels* de duas imagens não é suficiente para que se possa obter um resultado satisfatório, principalmente devido às transformações que podem ocorrer em uma imagem. Algumas dessas alterações são detalhadas na seção 3.7.

O registro de imagens consiste em encontrar um mapeamento entre pontos de uma imagem com pontos de outra. Na literatura encontram-se diversas técnicas que permitem este mapeamento ([Zitova and Flusser, 2003, Lowe, 2004, Rothganger et al., 2003]), mas no entanto este ainda é um tema de pesquisa atual, desafiador, constituído-se de uma área onde diversos grupos tem focado seus esforços no aperfeiçoamento e implementação de novas técnicas e algoritmos. O registro de imagens tem se mostrado importante em aplicações onde a informação final das tarefas de análise de imagens depende da combinação de diversas fontes de dados. Estes sistemas se caracterizam, principalmente, com relação a forma como são extraídas as informações de uma imagem. Com relação a esse aspecto, os sistemas de registro de imagens podem ser classificados como sistemas que utilizam características *globais* ou *locais*. Entre os sistemas que utilizam características globais encontram-se, por exemplo, os baseados em histogramas [Bender, 2003] ou em correlação [Righes, 2004]. Sistemas que utilizam características locais como Valores de Cinza Invariantes [Schmid and Mohr, 1997] ou a Transformada Local Independente de Escala (SIFT) [Lowe, 2004] têm tido sucesso em diversas aplicações atualmente. Este último pode ser considerado com um exemplo do estado-da-arte nesta área de pesquisa. A Figura 1.6 mostra um exemplo de aplicação de SIFT, mostrando os pontos similares encontrados entre duas imagens que sofreram transformações *affine*.

Os sistemas baseados em informações globais, apesar do grande sucesso obtido na sua aplicação, possuem diversas limitações, como a dificuldade em detectar objetos parcialmente oclusos ou com formas e tamanhos arbitrários, e que procuram ser sanadas pelos sistemas que utilizam informações locais. No entanto, os sistemas que utilizam informações locais, na sua grande maioria, desprezam a informação da cor, sendo utilizadas imagens em escala de cinza.

Como a cor tem um papel importante na visão humana, sendo o sistema visual humano todo baseado na cor, excetuando-se situações extremas, foi realizado um estudo sobre a utilização da informação de cor na obtenção de pontos de interesse. também buscamos avaliar o impacto na utilização desta informação na tarefa de registro de imagens. Para isso, nessa dissertação, será investigada uma extensão do algoritmo SIFT [Lowe, 2004], onde a informação de cor faz parte das características obtidas e é utilizada durante a fase de comparação dos pontos de interesse identificados na imagem.

## 1.1 Objetivos

Ao converter uma imagem colorida em sua equivalente em tons de cinza, existe perda de parte da informação disponível. A informação perdida é a informação da cromaticidade das cores. Uma imagem em tons de cinza representa, normalmente, a intensidade de luz existente na imagem. Apesar deste efeito, a maioria dos sistemas de registro de imagens baseados em características locais utilizam apenas imagens em

escala de cinza e não imagens coloridas. Um dos objetivos deste trabalho é verificar o quanto da informação da cromaticidade é relevante para aplicações de registro de imagens por características locais, contribuindo desta forma para o desenvolvimento da pesquisa de novas técnicas.

Neste trabalho também será feita uma revisão do estado da arte em registro de imagens, com ênfase nos sistemas de registro de imagens por descritores locais, com destaque especial a transformada de características invariantes à escala (SIFT - *Scale Invariant Feature Transform*). Esta técnica será estudada e estendida mediante a utilização da informação de cor na obtenção dos descritores locais. Também faz parte deste trabalho um estudo sobre a percepção de cor e como ela pode ser aplicada em sistemas de registro de imagem.

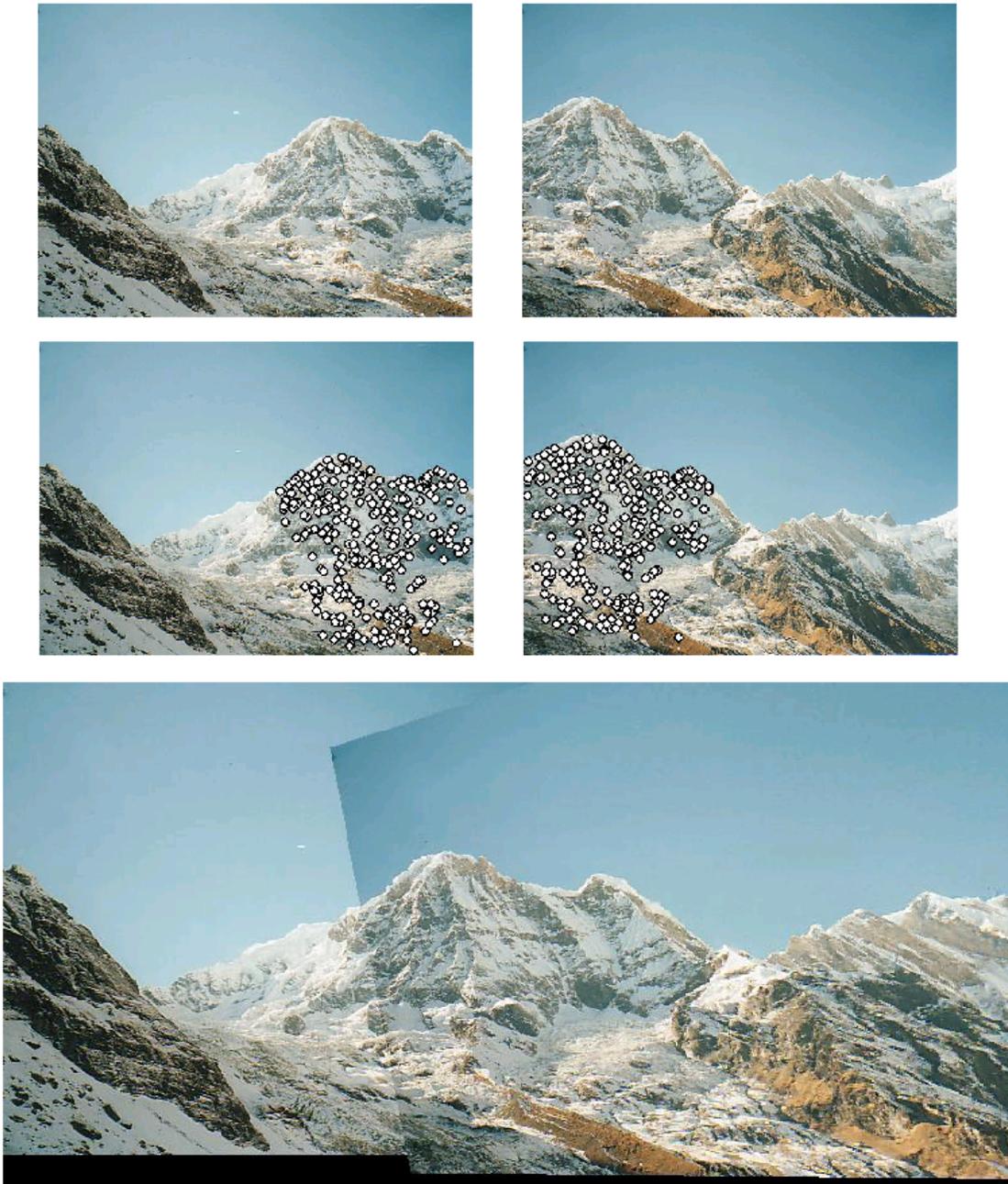


FIGURA 1.5 – Geração automática de panoramas.



FIGURA 1.6 – Exemplo de aplicação de SIFT.

## Capítulo 2

# Percepção de Cores

Cor é uma sensação criada em resposta a um estímulo do sistema visual humano à radiação eletromagnética (luz). A cor é o resultado da percepção da luz na região visível do espectro electromagnético, com comprimento de onda de  $400\text{nm}$  até  $700\text{nm}$ , incidente na retina do olho humano.

A retina humana possui dois tipos de células foto-receptoras, cones e bastonetes. Os bastonetes são importantes apenas em condições de pouca luminosidade, por exemplo na visão noturna, e não tem um papel importante na percepção de imagens [Lammens, 1994, Johnson and Fairchild, 1999]. Existem três tipos de cones, que são células foto-receptoras responsáveis pela percepção da cor, cada um com uma curva de resposta espectral diferente ao estímulo da radiação incidente [Wyszecki and Stiles, 1982]. Com base em experimentos, observou-se que o olho humano possui uma resposta espectral com um ponto máximo nas faixas vermelho, verde e azul do espectro visível [Padgham and Saunders, 1975]. A Figura 2.1 mostra um esboço das três curvas de resposta espectral do olho humano.

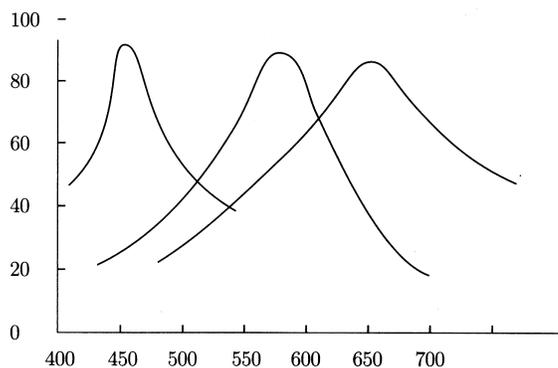


FIGURA 2.1 – Curvas de resposta espectral do olho humano de acordo com o comprimento de onda (nm). [Gomes and Velho, 1994]

Como existem exatamente três tipos diferentes de células cones foto-receptoras, três componentes numéricos são necessários e suficientes para descrever uma cor. Desta forma, uma cor pode ser especificada por um vetor tridimensional. O conjunto de todas as cores formam um espaço vetorial chamado de *espaço de cores* ou *modelo de cores*. Os três componentes de uma cor podem ser definidos de diversas formas, dando origem a vários espaços de cores.

## 2.1 Espaços de Cores

Para utilizar a cor em aplicações como multimídia, processamento de imagens, e visão computacional, é necessário definir um método para representar a cor. Para esse fim foram desenvolvidos diversos *espaços de cores*. Os espaços de cores permitem especificar, ordenar, manipular e visualizar as cores de forma racional [Plataniotis and Venetsanopoulos, 2000]. Diferentes espaços de cores oferecem diferentes abordagens e soluções para o problema em questão. O processo de seleção da melhor representação de cores envolve saber como os sinais de cores são gerados e quais as informações e propriedades deste sinal são importantes. Apesar de impor certas limitações na percepção e representação das cores, os espaços de cores são muito úteis para definir, discriminar, comparar e identificar categorias de cores.

Podemos classificar os espaços de cores de duas maneiras diferentes, pela forma como são definidos os componentes que descrevem a cor, ou pelo tipo de aplicação ao qual se destina o modelo.

Com relação a forma como são definidos os componentes que descrevem a cor, podemos classificar os espaços de cores como:

- **Baseados em Colorimetria:** modelos baseados em medições físicas do espectro.
- **Psicofísicos:** baseados na percepção humana da cor. Estes modelos são baseados em critérios subjetivos de observação e percepção.
- **Fisiológicos:** baseados nas três cores primárias dos três tipos de *cones* da retina. O espaço de cores RGB utilizado em computadores é talvez o modelo fisiológico mais conhecido.
- **Modelos de Cores Oponentes:** baseados em experimentos de percepção utilizando cores primárias oponentes como os pares Amarelo-Azul e Vermelho-Verde.

Os espaços de cores podem também ser classificados de acordo com o tipo de aplicação a que se destinam:

- **Modelos Orientados a Dispositivos:** associados ao sinal de entrada, ao processamento e ao sinal de saída de dispositivos.
- **Modelos Orientado ao Usuário:** utilizados em interfaces Homem-Máquinas. Estes modelos utilizam atributos de cores relacionados a forma como o ser humano as identifica.
- **Modelos Independentes de Dispositivos:** utilizados para definir cores independentemente do tipo de dispositivo ou aplicação. Estes modelos são muito importante em aplicações onde as cores serão comparadas ou exibidas em dispositivos diferentes daqueles utilizados para a sua obtenção.

A seguir serão descritos alguns espaços de cores, que direta ou indiretamente foram utilizados neste trabalho. A descrição destes espaços de cores foi baseada nas referências [Gonzalez and Woods, 1992] e [Plataniotis and Venetsanopoulos, 2000].

### 2.1.1 CIE XYZ

Em 1931, a *Comission Internationale de L'Eclairage* (CIE) adotou um padrão de *curvas* de cores para um observador hipotético padrão. Estas curvas de cores especificam como uma distribuição do espectro de força de um *estímulo externo* (luz visível incidindo no olho) pode ser transformada em um conjunto de três números que especificam aquela cor.

O sistema de especificação de cores do CIE é baseado na descrição da cor através de um componente de *luminosidade*  $Y$  e dois componentes adicionais  $X$  e  $Z$ . O modelo CIE XYZ permite descrever qualquer cor, é independente de dispositivo e muito útil em aplicações onde se deseja manter a representação de cores entre diferentes dispositivos.

O espaço de cores CIE XYZ não é perceptualmente uniforme, logo, não é apropriado para manipulações quantitativas envolvendo a percepção de cores e raramente é utilizado em aplicações de visão computacional.

No entanto, o espaço de cores CIE XYZ possui um papel significativo, porém não direto, no processamento de imagens, uma vez que todos os outros espaços de cores poder ser obtidos a partir do CIE XYZ através de transformações matemáticas. Esta propriedade permite a conversão de representações de cores entre diferentes espaços de cores, quando for possível representar determinada cor nos dois espaços de cores em questão.

### 2.1.2 CIE RGB

O espaço de cores mais utilizado em processamento de imagens, computação gráfica e sistemas multimídia é o CIE RGB. Neste espaço de cores, cada cor é definida a partir das cores primárias Vermelho (***R***ed), Verde (***G***reen) e Azul (***B***lue). O sistema RGB pode ser visto como formado por um cubo, onde cada cor é representada pelas coordenadas  $R$ ,  $G$  e  $B$  (Figura 2.2).

Duas variações importantes do espaço de cores RGB são o RGB linear e o RGB não-linear. Os valores do RGB linear são a representação física da luz emitida por um objeto. Este espaço de cores é perceptualmente *não-uniforme* e não é indicado para análise numérica dos atributos relativos a percepção de cores, porém apresenta a vantagem de ser independente de dispositivo, sendo por isso mesmo utilizado em sistemas de gerenciamento de cores. O espaço de cores RGB *não-linear* é perceptualmente mais uniforme que o RGB linear, sendo utilizado em aplicações de processamento de imagens, como filtros e nos padrões JPEG e MPEG, porém, é dependente de dispositivo. Uma importante característica comum a estes espaços de cores é a alta correlação existente entre os componentes que definem as cores. O espaço de cores RGB não-linear não é adequadamente uniforme e não deve ser utilizado quando se deseja acuidade perceptual, para estes casos são mais indicados os espaços de cores CIE  $L^*u^*v^*$  e  $L^*a^*b^*$ .

A conversão do espaço RGB linear para o espaço XYZ, considerando valores no intervalo  $[0, 1]$ , pode ser obtida com uma transformação de matrizes, conforme a Equação 2.1.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.1)$$

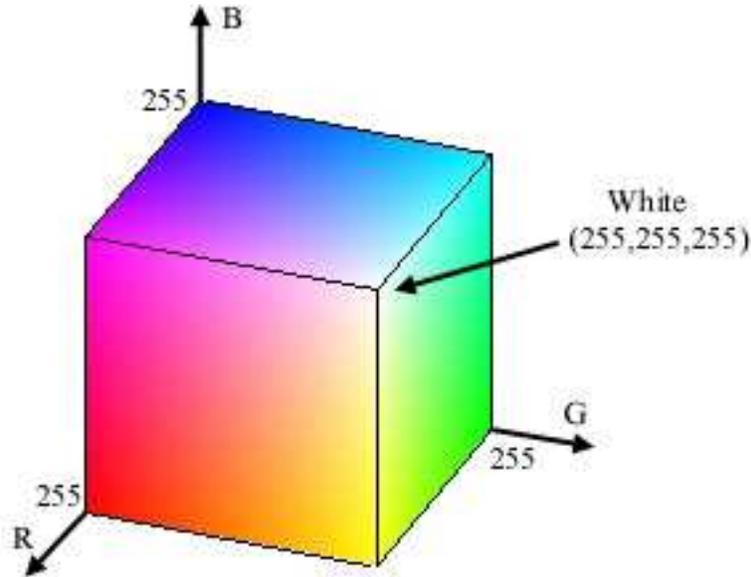


FIGURA 2.2 – Cubo RGB

Em um sistema de aquisição de imagem é utilizado o espaço de cores RGB *não-linear*, obtido através da conversão do sistema RGB linear por meio de uma transformação não-linear conhecida como correção *gamma*. Esta conversão é feita de acordo com as Equações 2.2, 2.3, 2.4), onde  $\gamma C$  é conhecido com o "fator gama", que varia de acordo com o dispositivo (por esse motivo esse espaço de cores é dependente de dispositivo). Para câmeras de vídeo, o valor de  $\gamma C$  mais comumente utilizado é  $\frac{1}{0.45}$ , ou aproximadamente 2.22,

$$R' = \begin{cases} 4.5R, & \text{se } R \leq 0.018 \\ 1.099R^{\frac{1}{\gamma C}} - 0.099 & \text{caso contrario} \end{cases} \quad (2.2)$$

$$G' = \begin{cases} 4.5G, & \text{se } G \leq 0.018 \\ 1.099G^{\frac{1}{\gamma C}} - 0.099 & \text{caso contrario} \end{cases} \quad (2.3)$$

$$B' = \begin{cases} 4.5B, & \text{se } B \leq 0.018 \\ 1.099B^{\frac{1}{\gamma C}} - 0.099 & \text{caso contrario} \end{cases} \quad (2.4)$$

### 2.1.3 Os modelos HSI

Os espaços de cores baseados no CIE RGB são convenientes para especificar cores em dispositivos. No entanto, apesar de computacionalmente eficientes, esses modelos não são indicados para utilização em interfaces com usuários. O sistema de visão do ser humano utiliza atributos como luminosidade, matiz e saturação na tarefa de percepção de imagens. Os modelos cor da família HSI (*Hue, Saturation and Intensity* – Matiz, Saturação e Intensidade) são mais intuitivos para o ser humano. Também apresentam características interessantes como a fraca correlação entre os componentes e componentes independentes para intensidade e matiz. O espaço de cor HSV pertence a este grupo. O sistema de coordenadas HSV pode ser

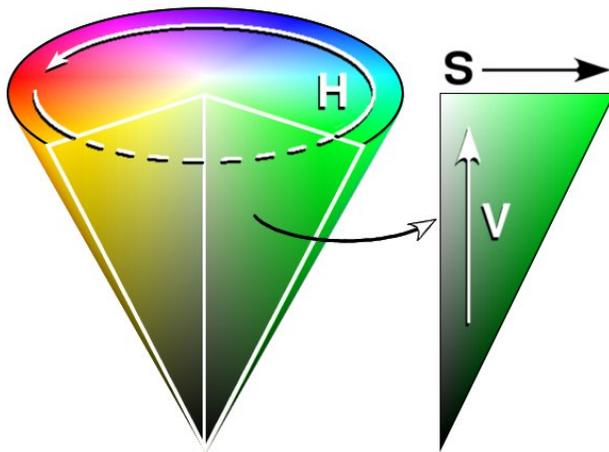


FIGURA 2.3 – Cone HSV

representado por um cone, como na Figura 2.3.

A partir de uma cor no modelo RGB representada por um inteiro de 8 bits (intervalo  $[0; 255]$ ), a conversão para HSV é dada pelas Eq. 2.5, 2.6, e 2.7.

$$H_1 = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right\}$$

$$H = \begin{cases} H_1 & \text{se } B \leq G \\ 360^\circ - H_1 & \text{se } B > G \end{cases} \quad (2.5)$$

$$S = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)} \quad (2.6)$$

$$V = \frac{\max(R, G, B)}{255} \quad (2.7)$$

#### 2.1.4 CIE $L^*a^*b^*$

Para a tarefa de percepção de cores, é muito importante possuir um sistema altamente preciso para defini-las. No processamento de imagens coloridas é interessante utilizar um sistema onde a distribuição das cores, com relação a sua percepção, seja uniforme. Como no sistema visual humano a percepção de cores não ocorre de forma linear, é necessária uma transformação não-linear para obter tal espaço de cores. Dois espaços de cores perceptualmente uniformes foram padronizados pelo CIE, o CIE  $L^*u^*v^*$  e o CIE  $L^*a^*b^*$ . Ambos são perceptualmente uniformes e possuem uma estimativa muito boa da diferença entre cores a partir de suas distâncias (num espaço vetorial formado pelos componentes dos espaços de cores). Estes dois modelos se diferenciam apenas pela forma como são obtidos.

A conversão de um espaço de cor para o CIE  $L^*a^*b^*$  (assim como para o CIE  $L^*u^*v^*$ , que não será abordado neste trabalho), é baseada no espaço de cor CIE XYZ e em uma cor branca de referência. Este "branco referência" é baseado no sistema RGB linear e representado, nesse sistema, pela cor  $[1,1,1]$  (considerando-se o

intervalo  $[0; 1]$ ) convertida para o espaço XYZ (Eq. 2.8). Existem outras alternativas para o "branco referência" utilizado, como os definidos pela Federal Communications Commission (FCC) ou pela European Broadcasting Union (EBU).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (2.8)$$

O componente  $L^*$  representa a luminosidade, sendo obtido pela Eq. 2.9, onde  $Y_n$  é o valor de  $Y$  para o branco de referência.

$$L^* = \begin{cases} 116 \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & \text{se } \frac{Y}{Y_n} > 0.008856 \\ 903.3 \left(\frac{Y}{Y_n}\right) & \text{caso contrario} \end{cases} \quad (2.9)$$

Na prática, podemos assumir que  $Y/Y_n > 0,01$ , sendo  $L^*$  obtido pela Eq. 2.10

$$L^* = 116 \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 \quad (2.10)$$

Os componentes  $a^*$  e  $b^*$  representam a cromaticidade da cor, sendo o eixo  $a^*$  referente ao eixo Vermelho-Verde, e  $b^*$  o eixo Amarelo-Azul. Estes componentes são obtidos de acordo com as Eq. 2.11 e 2.12, onde  $X_n$ ,  $Y_n$  e  $Z_n$  são os valores dos componentes XYZ para o branco de referência.

$$a^* = 500 \left[ \left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} \right] \quad (2.11)$$

$$b^* = 200 \left[ \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}} \right] \quad (2.12)$$

Espaços de cores perceptualmente uniformes são úteis em sistemas onde é importante uma reprodução perceptualmente exata de cores (reprodução de imagens digitalizadas), ajuste de cores frente a alterações na iluminação (navegação visual), e segmentação de cenas "naturais" utilizando técnicas como histogramas ou clusteração.

## 2.2 Temperatura de Cor

Outro conceito importante na realização deste trabalho é a temperatura de cor. Expressa em graus Kelvin, a temperatura de cor corresponde a cor irradiada por um corpo preto aquecido àquela temperatura. A Tabela 2.1 resume algumas faixas de luz e as suas respectivas temperaturas [Poynton, 2005].

Sistemas de captura de imagem são mais sensíveis que o olho humano quanto a variação na temperatura de cor. Nestes sistemas de aquisição da imagem é necessário modificar a percepção da temperatura de cor por meio de ajustes no equipamento de captura para reduzir estes efeitos. Por exemplo, a utilização de filmes ou filtros específicos para lâmpadas de Tungstênio em máquinas fotográficas analógicas ou a calibragem de branco em câmeras de vídeo ou máquina fotográficas digitais.

Com a variação da temperatura de cor a imagem pode adquirir tons avermelhados, quando a temperatura de cor da cena for mais baixa que a utilizada pelo

Temperatura	Fonte de luz
1500 K	Luz de vela.
2680 K	Lâmpada incandescente de 40 W.
3000 K	Lâmpada incandescente de 200 W.
3200 K	Nascer/pôr do sol.
3400 K	Lâmpada de Tungstênio.
3400 K	1 hora do nascer/pôr do sol.
4500-5000 K	Lâmpada de Xeônio.
5500 K	Dia de sol ao meio-dia.
5500-5600 K	<i>Flash</i> Eletrônico.
6500-7500 K	Céu nublado.
9000-12000 K	Céu claro.

TABELA 2.1 – Valores de temperatura de cor para fontes de luz comuns.

sistema de aquisição ou tons azulados, em caso contrário. A Figura 2.4 mostra o resultado da variação da temperatura de cor para um objeto.

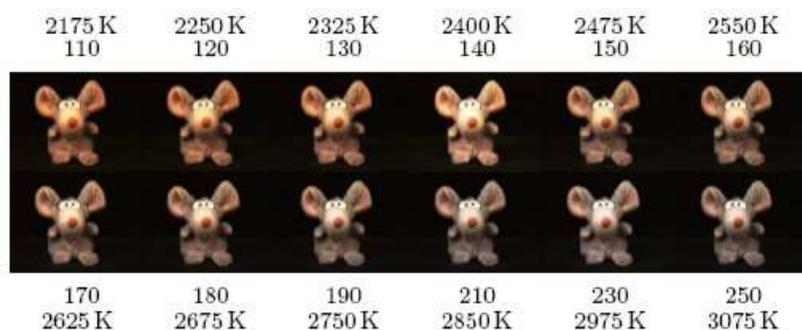


FIGURA 2.4 – Exemplo de objeto com variação da temperatura de cor da fonte de luz. [Geusebroek et al., 2005]

### 2.3 Utilização da Cromaticidade no Registro de Imagens

Nos métodos de registro de imagens por descritores locais descritos no Capítulo 3, nota-se a utilização quase exclusiva de imagens em escalas de cinza. O uso da cor se restringe a sistemas que utilizam técnicas de extração de características globais, como histogramas de cor [Bender, 2003, Faloutsos et al., 1994, Nastar et al., 1998].

O mapeamento de uma imagem colorida em uma imagem em tons de cinza reduz o volume de dados a ser tratado pelo sistema de registro de imagens. No entanto, ao realizar este mapeamento, parte da informação é perdida. Encontrase indícios na literatura [Marcel and Bengio, 2002] que a utilização da cor pode melhorar a classificação de imagens em um domínio específico. Ainda é uma questão em aberto se esta informação pode ser relevante para o processo do registro de imagens em geral.

Este estudo pretende estender um algoritmo de registro de imagens por descritores locais, de modo que este algoritmo inclua a informação de cromaticidade além da informação de intensidade de luz (*gray scale*). Atualmente, a maioria dos algoritmos para registro de imagens utilizam apenas a informação da intensidade de luz. Outro objetivo deste estudo é avaliar o quanto a informação de cromaticidade pode contribuir para melhorar o processo do registro de imagem.

# Capítulo 3

## Registro de Imagens

O registro de imagens consiste em encontrar um mapeamento entre pontos de uma imagem com pontos de outra. Como as imagens podem ser obtidas em resoluções diferentes, com pontos de vista diferentes, em instantes de tempo distintos, não é possível apenas sobrepor as imagens. Em alguns casos, até mesmo os sensores utilizados para captar as imagens são diferentes. No registro de imagens geradas em alguns diagnósticos médicos, são usadas, em conjunto, a Tomografia Computadorizada (CT), que mostra claramente as estruturas do corpo, e a Tomografia por Emissão de Pósitron (PET), que pode detectar atividades metabólicas específicas. O registro destas imagens pode ser utilizado para, por exemplo, identificar estruturas em que estão atividades metabólicas alteradas (como no caso de tumores). Neste caso, o registro consiste em encontrar um mapeamento capaz de associar as imagens dos dois sensores [Brown, 1992].

Em geral, o registro de imagens consiste de quatro etapas:

- *Extração de características* (Seção 3.1): seleção de características que representam a imagem. Essas características devem permitir um alinhamento entre imagens diferentes, ou seja, um ponto em uma imagem com uma determinada característica deve ser mapeada em uma outra imagem em um ponto com a mesma característica.
- *Comparação de características* (Seção 3.4): consiste em encontrar a correspondência entre os pontos da imagem sendo analisada com relação a uma imagem de referência.
- *Estimativa das transformações do modelo* (Seção 3.5): encontrar uma função que mapeie os pontos correspondentes em duas imagens de forma a obter um modelo das transformações ocorridas.
- *Transformação da imagem* (Seção 3.6): a imagem sendo analisada é transformada de acordo com o modelo obtido na etapa anterior.

### 3.1 Extração de Características

O objetivo da extração de características é obter uma transformação da imagem do domínio espacial para um domínio mais apropriado para a realização do

registro, resultando num vetor de características que descrevem a imagem. O domínio mais apropriado para a realização do registro depende do problema a ser tratado. A extração de características é importante para o registro de imagens por diminuir o espaço de procura (que passa a ser o vetor de características e não a imagem original) e obter uma descrição da estrutura intrínseca da imagem.

A transformação utilizada para obter o vetor de características da imagem, deve gerar características distintas e que devem ser ao menos parcialmente invariantes a transformações da projeção 3D, translação, rotação e escala, variações na luminosidade e oclusão parcial de objetos. Algumas alterações podem ser minimizadas ou removidas durante o pré-processamento da imagem, como o ruído; outras, como a iluminação, com a escolha de uma representação apropriada. Nos últimos anos, diversos trabalhos têm se preocupado com a extração de características parcialmente invariantes a diversas distorções.

As características geradas devem ser repetíveis, ou seja, para objetos ou imagens semelhantes as características geradas devem ser semelhantes. Os métodos de comparação de características dependem do tipo de características obtidas e de como o vetor de características pode ser avaliado.

A forma como são obtidos os vetores de características da imagem (descritores) pode ser utilizada como classificação de sistemas de registro de imagens. Estes descritores são divididos em *descritores globais*, obtidos a partir da análise de toda a imagem, ou *descritores locais*, obtidos a partir de uma operação em um ponto da imagem levando em consideração os pontos vizinhos.

## 3.2 Descritores Globais

Os descritores globais não levam em consideração variações locais. Este tipo de descritor pode obter bons resultados na indexação de imagens em bancos de dados, mas são incapazes de diferenciar pequenas variações locais que podem ser importantes em outras aplicações como o registro de imagens médicas [Shyu et al., 1998], distinguir entre objetos diferentes com características globais semelhantes, como por exemplo distinguir um campo de flores laranjas de um tigre [Belongie et al., 1998], ou identificar objetos parcialmente ocultos.

### 3.2.1 Histograma de Cores

Um histograma de cores aproxima a distribuição de probabilidade das cores presentes em uma imagem. A caracterização de uma imagem ou região de uma imagem por um histograma de cores, tem como principal vantagem ser parcialmente invariante a alterações de perspectiva, rotação e translação. Outra grande vantagem é o baixo custo computacional para obter-se o histograma de cores. No entanto, perde-se as informações espaciais com relação a distribuição de cores na imagem [Sebe and Lew, 2000]. Objetos parcialmente ocultos podem ter sua distribuição de cores alterada, já que a parte oculta do objeto não será considerada na geração do histograma. Para minimizar este efeito, podem ser utilizados auto-correlogramas de cor [Huang et al., 1997].

Histogramas de cores são utilizados com sucesso em sistemas de recuperação e classificação de imagens como o sistema QBIC [Faloutsos et al., 1994] utiliza características globais como histograma de cores e características baseadas em texturas,

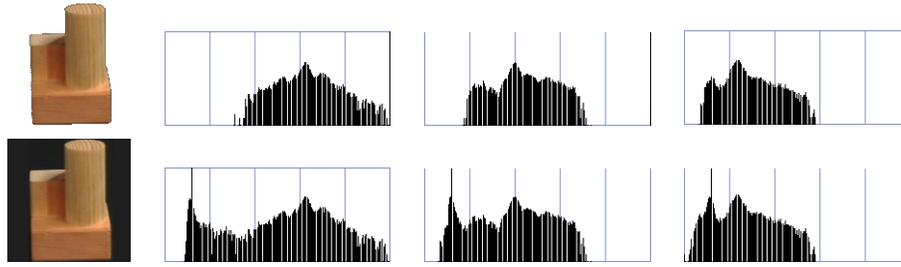


FIGURA 3.1 – Histogramas de um mesmo objeto alterados pela inclusão de artefatos.

e a sistema IMatch [Bender, 2003] utiliza histograma de cores como descritor para um sistema de classificação baseado em técnicas de aprendizagem de máquinas.

O registro de imagens baseado unicamente nas cores que compõem a imagem pode resultar na seleção de imagens com conteúdo diferente. Isto faz com que a utilização de histogramas não seja uma alternativa tão eficaz para registro de imagens onde objetos e estruturas devem ser localizados. A utilização de outras características, como forma e textura, podem auxiliar o registro com a utilização de histogramas [Belongie et al., 1998, Stricker and Orengo, 1995, Bender, 2003]. A Figura 3.1 mostra a alteração causada pela inclusão de um "artefato" na avaliação do histograma. Neste caso a alteração da cor de fundo da imagem foi utilizada para obter este efeito, o que poderia ter sido evitado com uma melhor definição dos limites do objeto, porém, este mesmo efeito ocorre quando da oclusão parcial do objeto.

A seleção do espaço de cores a ser utilizado também pode ter grande influência no histograma de cores. No caso do espaço de cores RGB [Foley et al., 1990], a informação da luminosidade está distribuída entre os três canais e diferenças de iluminação podem influenciar no histograma. Como alternativa podem ser utilizados espaços de cores que separam a cromaticidade da intensidade de iluminação, permitindo uma relativa invariância em relação a luminosidade. Entre os espaços de cores que separam a cromaticidade da luminosidade estão os espaços de cores HSI [Gonzalez and Woods, 1992], HSV [Foley et al., 1990] e CIE-L\*a\*b\* [Wyszecki and Styles, 1982].

### 3.2.2 Métodos Baseados em Correlação

Os métodos baseados em correlação, também conhecidos como métodos de área [Zitova and Flusser, 2003], não possuem explicitamente uma fase de extração de características, sendo um método de comparação baseado em padrões (*templates*). Nos métodos baseados em correlação, uma imagem, ou parte de uma imagem (janela), é utilizada como padrão a ser procurado, sem uma maior preocupação quanto ao conteúdo da imagem. A Figura 3.2 mostra o resultado da correlação de uma janela com uma imagem. Em destaque, a janela utilizada na correlação. (a) e (b) mostram o resultado da correlação de duas imagens candidatas com a janela utilizada, (c) é a imagem selecionada. [Righes, 2004]

A utilização de uma janela retangular para a correlação acaba por limitar estes métodos ao registro de imagens onde a única alteração local permitida é a trans-

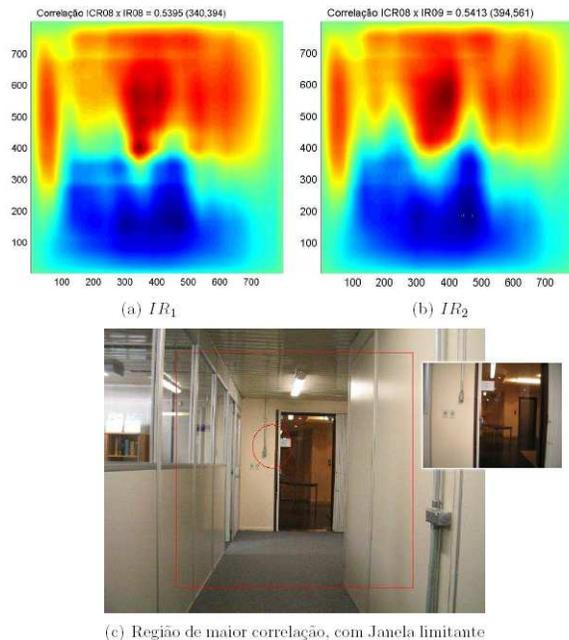


FIGURA 3.2 – Resultado da correlação de uma janela com uma imagem.  
[Righes, 2004]

lação. Outros formatos de janelas foram propostos, porém, a utilização de janelas com formas geométricas simples acaba não contemplando outras transformações geométricas mais complexas, por exemplo, alterações na perspectiva.

Outro problema é a utilização de uma janela que contenha apenas elementos com alterações suaves, sem detalhes proeminentes (contrastes de cor, contornos destacados). Neste caso, a correlação poderá ser alta com outras regiões sem saliências, mesmo que essas regiões não representem a região procurada.

Nos métodos tradicionais, como Correlação Cruzada e seus métodos derivados, Correlação Cruzada Normalizada [Crowley and Martin, 1995] e Correlação Cruzada Normalizada Rápida [Lewis, 1995]), a comparação entre as imagens é feita baseada na intensidade de cada pixel. Por este motivo, estes métodos são muito sensíveis a variações de iluminação e ruído.

Diversas técnicas foram desenvolvidas para minimizar os problemas inerentes aos métodos baseados em correlação, como por exemplo, a utilização de janelas com geometrias complexas para tratar outras transformações, e não apenas a translação [Berthilsson, 1998], assim como métodos menos sensíveis a variações de intensidade e com registro de objetos parcialmente oclusos [Kaneko et al., 2002].

Em geral, os métodos de registro de imagens baseados em correlação são computacionalmente custosos. Mesmo com essas desvantagens, estes métodos ainda são muito utilizados para o registro de imagens, principalmente devido a facilidade de implementação em hardware, sendo alternativas interessantes para aplicações em tempo-real nestes casos.

### 3.3 Descritores Locais

Outra forma de caracterizar uma imagem, é extraindo características em um único ponto da imagem, por meio de descritores locais. Ao contrário das técnicas globais de caracterização da imagem, os descritores locais tratam das características em um ponto específico da imagem. Para cada ponto, é calculado um vetor de características que descreve as estruturas presentes naquele ponto.

Descritores locais têm sido utilizados em aplicações de registro de imagens principalmente por serem robustos frente a oclusão parcial de objetos e imagens com muitos objetos. A dificuldade na utilização destes descritores está em obtê-los de forma que não sejam afetados por transformações nas imagens. Diversas soluções foram propostas para solucionar este problema recentemente [Lowe, 1999, Schmid, 2001, Rothganger et al., 2003, K.Mikolajczyk and C.Schmid, 2001, Mikolajczyk and Schmid, 2002]. Estas soluções determinam uma série de características a partir das quais são obtidos os descritores locais. Uma técnica que se mostrou particularmente útil são os pontos de interesse.

#### 3.3.1 Pontos de Interesse

Em uma imagem existem muitos pontos que não fornecem informações relevantes para o registro de imagens. Ao se extrair descritores locais de uma imagem, existe a necessidade de selecionar de quais pontos devem ser extraídos os descritores. Extrair características de cada ponto da imagem geraria muita informação, e grande parte desta informação seria irrelevante, servindo apenas para dificultar o registro. Para reduzir o volume de informações utilizam-se pontos de interesse na imagem. Os pontos de interesse se caracterizam por fornecerem um alto grau de informação e serem robustos à oclusão parcial de objetos.

Pontos de interesse são ocorrências em uma imagem onde as características locais sofrem alterações em duas dimensões [Mikolajczyk and Schmid, 2002], ou seja dado um ponto em uma imagem e uma escala, existe uma alteração em duas dimensões do sinal (há a ocorrência de um *corner*). Para uma eficiente utilização de descritores locais no registro de imagens, é importante que os descritores obtidos a partir de pontos de interesse sejam repetíveis e discriminantes sob alterações na imagem, ou seja, que os pontos detectados em uma imagem possam ser obtidos em uma outra imagem mesmo na presença de transformações em relação a imagem original, e que as características determinadas por esse ponto sejam semelhantes para as duas imagens.

Existem vários métodos para determinar pontos de interesse, entre os quais, detecção de ângulos (*Harris Corner Detection*) [Harris and Stephens, 1988, Schmid and Mohr, 1997], que é invariante à rotação da imagem, diferença de Gaussianas [Lowe, 2004] e *Laplacian* [K.Mikolajczyk and C.Schmid, 2001], os quais são invariantes a escala.

#### 3.3.2 *Harris Corner Detection*

A idéia básica deste detector de ângulos [Harris and Stephens, 1988] é utilizar a auto-correlação para determinar os locais onde ocorrem alterações no sinal em duas dimensões. A matriz relacionada com a função de auto-correlação, levando

em consideração as primeiras derivadas do sinal em uma janela, é obtida através da Equação 3.1, onde  $\otimes$  denota convolução.

$$e^{-\frac{x^2+y^2}{2\sigma^2}} \otimes \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (3.1)$$

Os autovalores da matriz obtida pela Equação 3.1 são as principais curvaturas da função de auto-correlação. Dois valores significativos indicam a presença de um ponto de interesse. Os pontos de interesse obtidos com esta técnica são invariantes a translações e rotações. A Figura 3.3 mostra pontos de interesse detectados com a utilização do *Harris Corner Detector*. A imagem foi rotacionada 155° e 92% dos pontos de interesse ainda foram detectados [Schmid and Mohr, 1997].

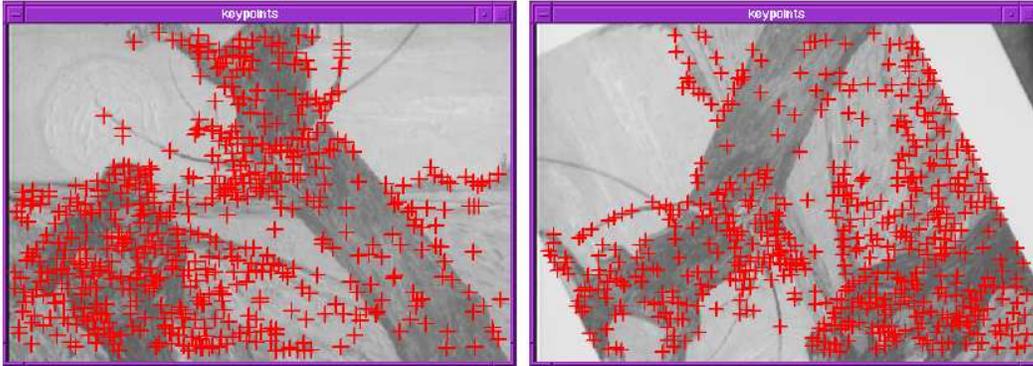


FIGURA 3.3 – Pontos de interesse detectados com a utilização do *Harris Corner Detector*.

### 3.3.3 *Scale Invariant Feature Transform*(SIFT)

Para obter pontos de interesse invariantes a alterações na escala, pode ser feita uma procura por características estáveis em todas as escalas possíveis utilizando uma função contínua de escala conhecida por *espaço de escala* [Witkin, 1986]. O espaço de escala de uma imagem pode ser definido pela convolução de uma escala variável Gaussiana,  $G(x, y, \sigma)$ , com uma imagem de entrada  $I(x, y)$ , como pode ser visto na Equação 3.2 [Lowe, 2004].

$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y) \quad (3.2)$$

Para detectar pontos de interesse no espaço de escala, pode ser utilizada a convolução da diferença de Gaussianas com a imagem de entrada, que pode ser calculada através da diferença de duas escalas próximas, separadas por um fator constante  $k$  (Equação 3.3) [Lowe, 1999].

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) \otimes I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3.3)$$

Para obter os pontos candidatos a ponto de interesse, é construída uma pirâmide  $D$ , variando  $\sigma$ . É feita a convolução da imagem original com valores incrementais de  $k$ . Divide-se cada oitava do espaço escala (obtidas dobrando  $\sigma$ ), por um

número inteiro de intervalos  $s$ , de modo que  $k = 2^{\frac{1}{s}}$ . Devem ser produzidas  $s + 3$  imagens para cada oitava desta forma. Imagens adjacentes são subtraídas, para criar a diferença de Gaussianas. Assim que uma oitava do espaço escala é processada, uma nova etapa é realizada com a imagem obtida com  $\sigma$  com o dobro do valor inicial, tomando cada segundo pixel das linhas e colunas. O processo de obtenção da pirâmide pode ser visto na Figura 3.4, onde à esquerda estão as imagens do espaço escala e à direita as diferenças de Gaussianas [Lowe, 2004].

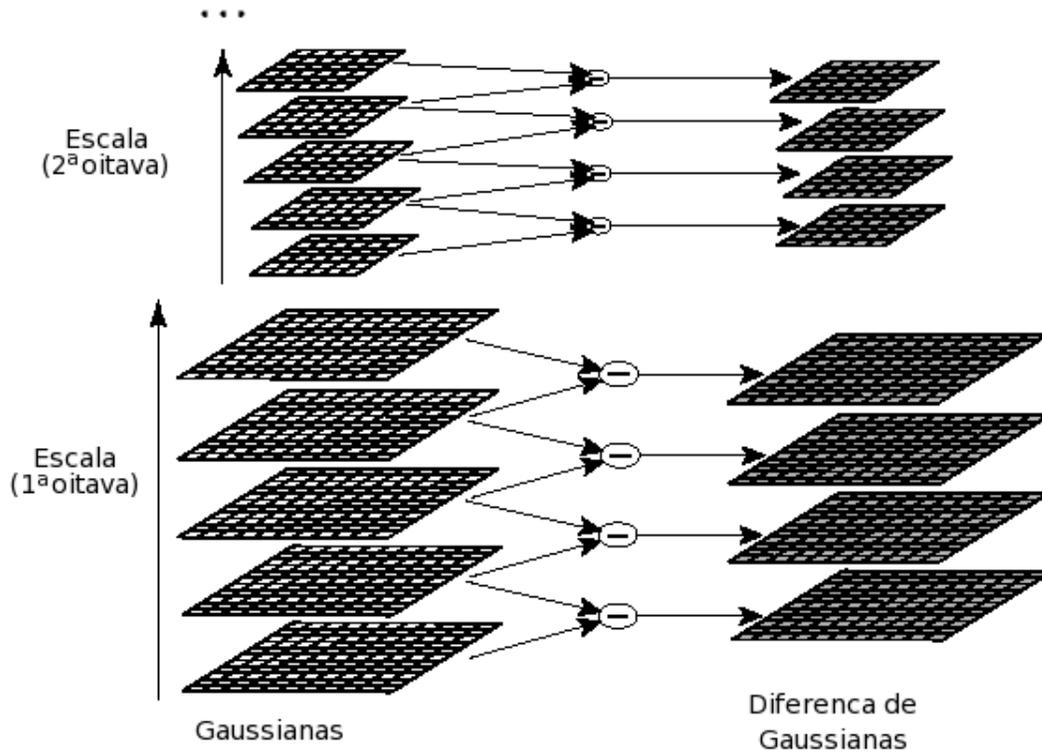


FIGURA 3.4 – Pirâmide para obtenção das diferenças de Gaussianas.

Os pontos de interesse nesse método são obtidos nos máximos e mínimos das diferenças de Gaussianas. Esses máximos e mínimos são obtidos comparando cada ponto com os seus oito vizinhos na mesma escala e os nove vizinhos uma escala acima e outra abaixo (Figura 3.5) [Lowe, 2004]. Os pontos com as maiores e as menores diferenças são selecionados como *pontos candidatos* para as próximas etapas do processamento.

Após determinar os pontos candidatos através das máximas e mínimas da diferença das Gaussianas, é necessário verificar a estabilidade do ponto candidato, avaliando a sua localização, escala e razão das curvaturas principais. Estas informações permitem eliminar pontos que possuem baixo contraste, e por esse motivo são muito sensíveis a ruídos, e pontos localizados em arestas, os quais não são estáveis e também sensíveis a ruído.

Para determinar com precisão a localização dos pontos de interesse, é utilizado o ajuste dos pontos candidatos a uma função quadrática 3D [Brown and Lowe, 2002]. A função é obtida pela utilização da expansão de Taylor da função do espaço escala  $D(x, y, \sigma)$ , transladada de forma que a origem seja a

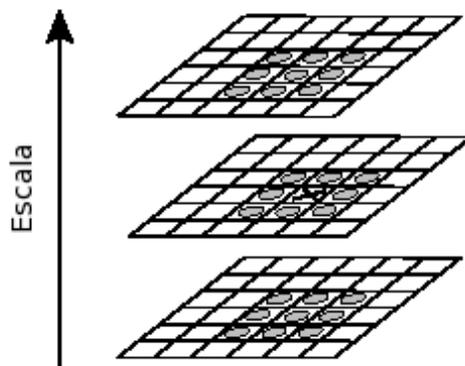


FIGURA 3.5 – Detecção de máximos e mínimos das diferenças de Gaussianas.

posição do ponto candidato (Equação 3.4), onde  $D$  e suas derivadas são computados em relação ao ponto candidato e  $\mathbf{x} = (x, y, \sigma)^T$  é a distância deste ponto.

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} \quad (3.4)$$

A localização do *extremum*,  $\hat{\mathbf{x}}$ , é determinada tomando-se a derivada da Equação 3.4 em relação a  $x$  e igualando-a a zero (Equação 3.5).

$$\hat{\mathbf{x}} = - \frac{\partial^2 D^{-1} \partial D}{\partial \mathbf{x}^2 \partial \mathbf{x}} \quad (3.5)$$

O *extremum*  $\hat{\mathbf{x}}$  é distância em relação ao ponto candidato que se encontra o máximo ou mínimo da diferença de Gaussiana, com precisão de *sub-pixel*. Se o offset  $\hat{\mathbf{x}}$  for maior que 0.5 em qualquer dimensão, o *extremum* está mais próximo de outro ponto, esse ponto é então utilizado para localização do ponto de interesse. O offset final  $\hat{\mathbf{x}}$  é então adicionado ao ponto amostrado, para ter uma estimativa interpolada da localização do *extremum* [Lowe, 2004].

O valor da função no *extremum*,  $D(\hat{\mathbf{x}})$  (Equação 3.6), é utilizado para rejeitar pontos onde, apesar de serem máximos ou mínimos da diferença de Gaussianas, possuem baixo contraste, e portanto são considerados instáveis. Valores menores que um determinado linear para  $|D(\hat{\mathbf{x}})|$  podem ser descartados. Neste trabalho, considerando intensidades na imagem original no intervalo  $[0; 1]$ , foi utilizado o limiar 0,03, como sugerido em [Lowe, 2004].

$$D(\hat{\mathbf{x}}) = \frac{1}{2} \frac{\partial D^T}{\partial \mathbf{x}} \hat{\mathbf{x}} \quad (3.6)$$

As bordas de objetos existentes em uma imagem também mostram-se instáveis, para determinar pontos de interesse. Intuitivamente, esta afirmação parte do princípio que qualquer ponto desta borda terá um alto contraste, porém é difícil determinar um ponto repetível na extensão da borda. Estes pontos, se utilizados, seriam sensíveis ao ruído.

Para eliminar estes pontos, as curvaturas principais podem ser calculadas por uma matriz Hessiana 2x2,  $\mathbf{H}$ , na posição e escala do ponto de interesse. As derivadas podem ser estimadas pelas diferenças entre os pontos vizinhos.

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

Seja  $\alpha$  o autovalor de  $\mathbf{H}$  com maior magnitude e  $\beta$ , o autovalor de menor magnitude, podemos calcular a soma dos autovalores e o seu produto, através do traço o do determinante de  $\mathbf{H}$ .

$$Tr(\mathbf{H}) = D_{xx} + D_{yy} \quad (3.7)$$

$$Det(\mathbf{H}) = D_{xx}D_{yy} - D_{xy}^2 \quad (3.8)$$

Caso o determinante seja negativo, o ponto de interesse é rejeitado por não ser um *extremum*. Seja  $r$  a razão entre o autovalor de maior magnitude e o autovalor de menor magnitude, tal que  $\alpha = r\beta$ , temos

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

que depende apenas da razão dos autovalores e não dos seus valores individuais. Desta forma podemos verificar se um ponto de interesse deve ser rejeitado fixando um valor arbitrário para  $r$  e verificar se as curvaturas principais estão abaixo de um limiar pela Equação 3.9.

$$\frac{Tr(\mathbf{H})^2}{Det(\mathbf{H})} < \frac{(r + 1)^2}{r} \quad (3.9)$$

Após determinar a localização e a escala onde se encontra o ponto de interesse, é necessário determinar a orientação do ponto de interesse. Ao associar uma orientação consistente para cada ponto de interesse, o descritor local pode ser representado relativo a essa orientação, obtendo assim, invariância quanto a rotação da imagem.

Para definir a orientação do vetor seleciona-se a imagem  $L$  com a escala mais próxima da escala ao qual pertence o ponto de interesse. Dessa forma a invariância à escala é garantida. Para cada amostra desta imagem, são calculada a magnitude  $m$  e a orientação  $\theta$ , utilizando a diferença dos *pixels*, conforme as Equações 3.10 e 3.11.

$$m = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \quad (3.10)$$

$$\theta = \tan^{-1} \left( \frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \right) \quad (3.11)$$

Um histograma de orientações, com 36 *bins* cobrindo os 360° dos possíveis ângulos, é formado com as orientações do gradiente dos pontos próximos ao ponto de interesse. Cada amostra adicionada ao histograma é ponderada pela sua magnitude e por uma janela Gaussiana com  $\sigma$  1.5 vezes maior que a escala do ponto de interesse.

Picos no histograma correspondem a direções dominantes no gradiente local. O maior pico do histograma e os picos com até 80% deste pico maior são utilizados para

obter um descritor local com a orientação de cada um dos picos selecionados. Podem existir pontos de interesse que acabem gerando vários descritores com orientações diferentes.

Com a localização, escala e orientação do ponto de interesse, tem-se um conjunto de parâmetros repetíveis que descrevem uma região da imagem. Este ponto de interesse também é invariante aos parâmetros que o definem. O próximo passo é obter um vetor de características que seja altamente discriminante e tão invariante quanto possível a outras alterações, por exemplo, iluminação e perspectiva.

Para obter os descritores, para cada ponto de interesse são computadas as magnitudes e orientações dos gradientes próximos ao ponto de interesse. A invariância em relação à rotação é obtida rotacionando-se as coordenadas e as orientações em relação à orientação do ponto de interesse. Uma janela Gaussiana com  $\sigma$  igual à metade do tamanho do descritor é utilizada como "peso" para as magnitudes de cada ponto. O objetivo dessa janela é minimizar alterações abruptas no descritor e dar menos ênfase aos pontos distantes do ponto de interesse. São criados histograma de orientações, como os descritos anteriormente, em regiões de  $4 \times 4$  pontos. O descritor é então criado com as magnitudes das entradas dos histogramas de orientação. Para atingir um certo grau de invariância parcial a alterações na iluminação, o vetor de características é normalizado.

Cada descritor obtido é uma matriz quadrada com as magnitudes do histograma de orientações. O tamanho do descritor é dado por  $rn^2$ , onde  $r$  é o número de entradas no histograma, e  $n$  a ordem da matriz. Quanto maior o tamanho do descritor, melhor será a discriminação obtida, no entanto, o descritor será mais sensível a distorções nas formas e à oclusão parcial.

O maior problema na obtenção destes descritores é o custo computacional da convolução da imagem para gerar as diferenças de Gaussianas nas várias escalas. Comparados isoladamente, o custo computacional para a obtenção destes descritores é bem superior ao necessário para obter o histograma de uma imagem, por exemplo. No entanto, a utilização de histogramas para o reconhecimento de um objeto parcialmente ocluso ou em cenas complexas, não é tão simples e pode levar a um alto custo computacional devido principalmente à complexidade da comparação das características.

A obtenção de descritores utilizando SIFT foi aplicada com sucesso em diversas aplicações onde se utiliza o registro de imagens, como na obtenção de panoramas a partir de um conjunto de imagens [Brown and Lowe, 2003], como pode ser visto na Figura 1.5 (pág. 16); na localização de objetos em cenas complexas [Lowe, 2004], como mostra a Figura 1.2 à página 12; e na tarefa de localização e mapeamento para navegação visual de robôs [Lowe et al., 2002], como visto na Figura 3.6, onde à esquerda vemos os descritores detectados em uma imagem, ao centro a diferença de posição entre os descritores encontrados em duas imagens "estéreo" (as linhas verticais representam a disparidade vertical e as horizontais a disparidade horizontal), e à direita a diferença da posição dos descritores entre duas imagens após a rotação da câmera.

Na obtenção de descritores locais utilizando SIFT, alguns parâmetros podem ser alterados de forma a ajustar o resultado do algoritmo à aplicação a ser desenvolvida. Entre estes parâmetros, os mais importantes são o número de *bins* dos histogramas de orientação, o número de histogramas para cada região da imagem e o limite de rejeição do *extremum*.

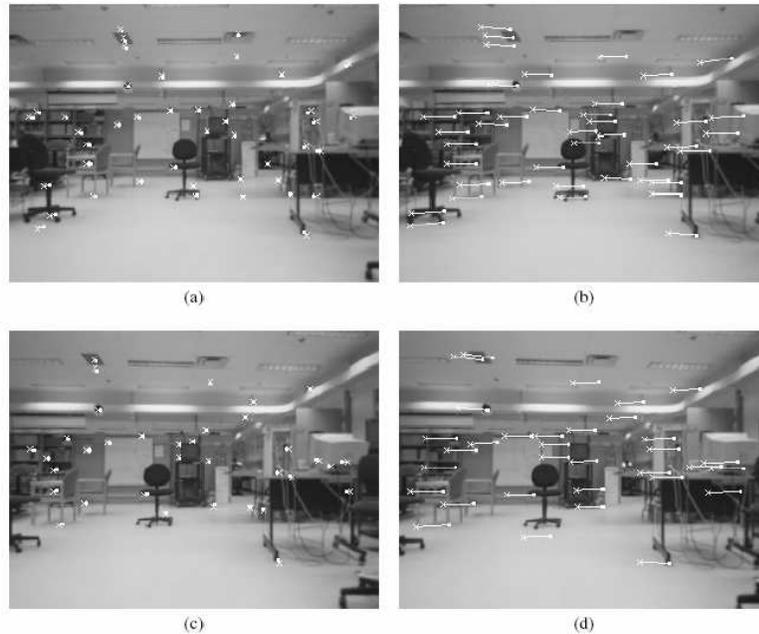


FIGURA 3.6 – Navegação Visual com a utilização de descritores invariantes à escala.

A variação do limite de rejeição do *extremum* faz com que pontos com maior ou menor contraste sejam aceitos como pontos de interesse para a obtenção dos descritores. Limites mais baixos permitem a obtenção de um maior número de descritores. Entretanto, pontos com contraste mais baixo podem ser menos estáveis. Neste trabalho, foi utilizado o valor 0.03 como limite de rejeição do *extremum* ([Lowe, 2004]).

Os parâmetros mais importantes no entanto, são o número de *bins* dos histogramas de orientação e o número de histogramas para cada região da imagem. A variação destes parâmetros altera não somente a qualidade dos descritores obtidos, mas também o tamanho do vetor de características dos descritores.

Neste trabalho foram usados vetores de características de 128 dimensões, sendo obtidos através de histogramas de 8 *bins* obtidos a partir de regiões de tamanho 4x4. Cada região é formada por regiões retangulares de 4x4 pontos. A figura 3.7 exemplifica a obtenção de um descritor em regiões 2x2.

A variação do tamanho das regiões ou do número de *bins* altera o tamanho do vetor de características. Neste trabalho foram utilizados vetores formados por regiões de tamanho 4x4 e histogramas de 8 *bins*, resultando em vetores de 128 posições (4x4x8).

A alteração do número de *bins* dos histogramas, aumenta a especificidade do descritor, fazendo com que o descritor defina melhor um ponto. No entanto, essa alteração faz com que seja mais difícil encontrar pontos semelhantes, uma vez que pequenas variações locais terão mais impacto no descritor.

A alteração do tamanho das regiões também afeta a precisão com que um descritor define um ponto. Para regiões maiores, uma maior área terá influência no ponto, fazendo com que o contexto onde o ponto está inserido tenha mais ou

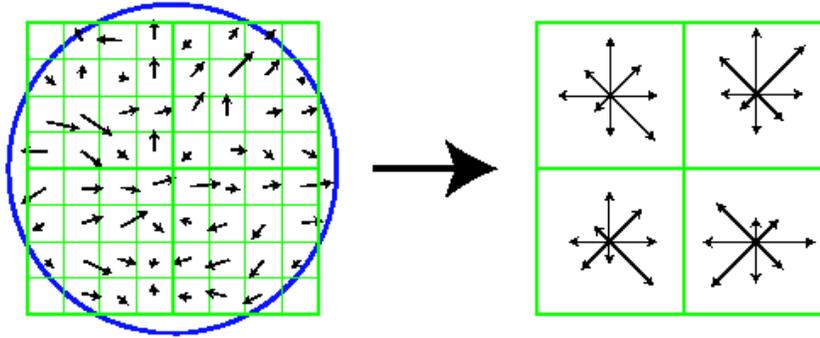


FIGURA 3.7 – Obtenção dos descritores SIFT a partir dos histogramas de orientação. Neste exemplo são utilizadas regiões  $2 \times 2$  e 8 *bins* para os histogramas.

menos relevância. Regiões pequenas tendem a ser mais facilmente observadas em outras imagens, mesmo que não tenham relação com a original, regiões maiores são mais difíceis de serem encontradas em outras imagens. No entanto, regiões maiores tem mais probabilidade de serem parcialmente oclusas, o que ocorre com menos frequência em regiões menores.

Sempre que possível, é interessante identificar, experimentalmente, qual a melhor relação entre a especificidade dos descritores e a aplicação na qual serão utilizados.

### 3.4 Comparação de características

Após a extração das características das imagens de referência e capturada, é realizada a comparação destas características. Nesta etapa procura-se obter correspondências entre as imagens sendo registradas. Alguns métodos, paralelamente à comparação de características, estimam os parâmetros das funções de mapeamento nesta fase, unindo as segunda e terceira etapas do registro de imagens.

Os métodos utilizados para a comparação de características dependem do tipo de características obtidas, globais ou locais. Entre os métodos de comparação de características globais, encontram-se os métodos baseados em correlação, os métodos de Fourier e os métodos de informação mútua. Entre os métodos de comparação de características locais, destacam-se métodos que exploram relações espaciais e descritores invariantes.

### 3.5 Estimativa da transformação do modelo

O objetivo desta etapa do registro de imagens é encontrar os parâmetros da transformação entre a imagem de referência e a imagem capturada, de forma que seja possível sobrepor a segunda à primeira [Zitova and Flusser, 2003].

O problema a ser resolvido é encontrar o tipo da função de mapeamento e os seus parâmetros (Figura 3.8). O tipo de função de mapeamento deve corresponder à deformação geométrica esperada na imagem capturada, ao método de aquisição da imagem e a exatidão requerida do registro. Em situações especiais, por exemplo,

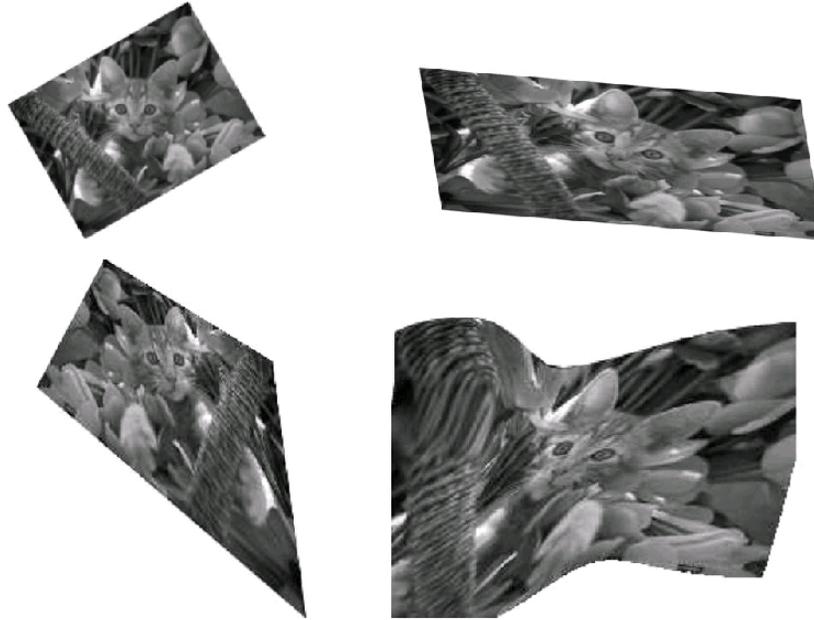


FIGURA 3.8 – Exemplos de diversas funções de mapeamento. No sentido horário similaridade, *affine*, perspectiva e transformação elástica.  
[Zitova and Flusser, 2003]

quando a deformação geométrica é parcialmente conhecida, pode ser feita uma pré-correção baseada no inverso da deformação esperada.

Neste trabalho, o registro de imagens será efetuado levando em consideração apenas as funções de mapeamento de similaridade, *affine* e perspectiva.

### 3.6 Transformação da Imagem

As funções de mapeamento obtidas na etapa anterior são utilizadas para transformar a imagem capturada e dessa forma, registrar a imagem. A transformação pode ser realizada com os métodos diretos e reverso. O método direto consiste em transformar diretamente cada ponto da imagem capturada utilizando os parâmetros do modelo estimado. Esta abordagem no entanto é de difícil implementação e sujeita a defeitos na imagem final (como "buracos" e sobreposição). Normalmente é utilizado o método reverso, onde a imagem resultante do registro é obtida utilizando as coordenadas do pixel alvo (no mesmo sistema de coordenadas da imagem de referência) e o inverso da função de mapeamento encontrada. Após é feita a interpolação na imagem capturada, onde a interpolação bilinear oferece, provavelmente, a melhor relação entre custo computacional e precisão. Caso seja necessário um grande aumento da imagem capturada, é aconselhável o uso de interpolação cúbica para obter melhores resultados.

## 3.7 Transformações geométricas entre duas imagens

Este trabalho estará focado no uso de imagens 2D e não na obtenção de modelos 3D ou no processamento de imagens multi-espectrais. No entanto, imagens 2D sofrem diversas transformações influenciadas pelo espaço 3D, como alterações na geometria aparente dos objetos, por exemplo, devido à alterações na perspectiva. Além da transformação perspectiva, são comuns também alterações na escala, translação e rotação entre as imagens de referência e capturada.

### 3.7.1 Transformações *Affine*

As transformações *affine* são suficientes para alinhar imagens de uma cena obtida sob o mesmo ângulo, mas de diferentes posições. São transformações lineares as operações cartesianas de escala, translação e rotação [Brown, 1992]. Este tipo de transformação é globalmente rígida, ou seja, a relação geométrica entre os pontos não são alteradas entre as imagens. Tipicamente, esta transformação possui 4 parâmetros,  $t_x$ ,  $t_y$  (translação em  $x$  e  $y$ ),  $s$  (escala) e  $\theta$  (ângulo de rotação), os quais mapeiam um ponto  $(x_1, y_1)$  em uma imagem, para o ponto  $(x_2, y_2)$  em outra, conforme a Equação 3.12.

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + s \cdot \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \quad (3.12)$$

Na sua forma geral, a transformação *affine* pode ser utilizada para considerar também a obliquidade [Van Wie and Stein, 1977]. Sem a adição do termo de translação, a transformação torna-se linear. A Figura 3.9 mostra as transformações *affine*.

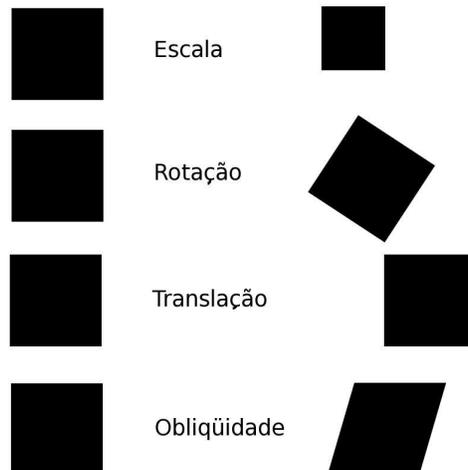


FIGURA 3.9 – Transformações *affine*.

### 3.8 Iluminação e Cor

Outra fonte de alterações significativas da imagem é a fonte de luz. A fonte de luz de uma imagem pode alterar sua intensidade ou sua posição relativa aos objetos na cena. Diferenças na fonte de luz podem alterar a percepção de um objeto, por exemplo, gerando sombras e criando falsos contornos. Outra alteração que pode ser verificada é nas cores dos objetos. Como a cor de um objeto é definida pela forma como a luz é refletida neste objeto, alterações na fonte de luz podem causar alterações na percepção dos objetos.

### 3.9 Utilização da Cor em Descritores Locais

Sistemas que utilizam descritores globais têm utilizado a informação de cor com relativo sucesso [Bender, 2003, Faloutsos et al., 1994, Righes, 2004]. Por outro lado, a maioria dos sistemas de registro de imagem que utilizam descritores locais para caracterização da imagem trabalham apenas com imagens em escalas de cinza [Schmid and Mohr, 1997, Lowe, 2004]. Considerando que os sistemas com descritores locais são, normalmente, baseados em pontos de interesse, e que esse pontos de interesse encontram-se em locais onde existem variações nas duas dimensões da imagem, ao inserir a informação de cor estariam sendo analisadas não apenas variações na luminosidade, mas também variações na cromaticidade existente na cena.

Tendo como base o algoritmo SIFT, foi realizado um estudo para avaliar o impacto da cromaticidade como informação básica para a obtenção de pontos de interesse e, conseqüentemente, descritores locais baseados nas alterações da cromaticidade.

A informação de cor pode ser importante também, como uma informação de mais alto nível, por exemplo, em um processo de decisão. Será feita também uma análise de como a cor pode ser utilizada como uma informação de alto nível, sendo utilizada como mais um atributo de um descritor de imagem.

## Capítulo 4

# Obtenção de descritores locais utilizando a informação de cor

A informação da intensidade de luz é a mais utilizada em sistemas de visão computacional, principalmente nos sistemas de registro de imagens baseados em descritores locais. Nestes sistemas, os descritores são obtidos em pontos onde existem alterações em duas dimensões, por serem pontos que contém informações mais relevantes para o processo de registro (ou "casamento") de imagens. Nota-se que o que se procura são "cantos" (*corners*) na imagem, e a intensidade de luz é uma característica que permite a identificação de bordas (e por consequência dos cantos) com boa precisão. No entanto, o sistema de visão do ser humano é baseado na cor, e esta pode ser uma informação importante. Os experimentos descritos a seguir procuram avaliar o quanto a informação de cor pode ser útil na obtenção de descritores locais utilizando o algoritmo SIFT.

O primeiro experimento procura avaliar se os descritores obtidos mediante a utilização exclusiva da informação de cor (canal H do HSV) são complementares àqueles obtidos pelos canais de intensidade de luz (imagem em escala de cinza e canal V do HSV) e se são gerados em número significativo. É importante obter uma quantidade razoável de descritores a partir de uma imagem para diminuir a probabilidade de que os pontos fiquem oclusos quando da comparação com outra imagem. A verificação quanto aos pontos serem complementares, ou seja diferentes dos pontos obtidos a partir da intensidade de luz, é importante no caso destes pontos serem utilizados em conjunto, aumentando o número total de pontos de interesse obtidos a partir de uma única imagem. Neste experimento, esta verificação é feita pela distância, em *pixels*, da posição dos descritores obtidos no canal H em relação a posição dos descritores no canal V e na imagem monocromática. Se a distância entre os descritores for maior que um limiar pré-definido, os pontos são considerados complementares.

Os descritores obtidos com SIFT são parcialmente invariantes as transformações *affine* [Lowe, 2004]. Para verificar se o mesmo ocorre com os descritores obtidos a partir da informação de cor, foram executados experimentos com imagens rotacionadas, escaladas e com alteração no ângulo de visão do objeto.

Para verificar a robustez dos pontos obtidos utilizado-se exclusivamente a informação de cor, foram utilizadas as bases de dados com alteração da posição da fonte de luz e com alteração na temperatura de cor da fonte de luz. A variação da posição da fonte de luz acaba gerando falsos contornos nos objetos, e a variação da

temperatura de cor, alterações na percepção da cor dos objetos. Nestes experimentos, os descritores obtidos em uma imagem foram comparados com os descritores obtidos na imagem "r0" do mesmo objeto (imagem sem alteração, ver Seção 4.1). A comparação dos descritores é feita como descrito em [Lowe, 2004].

Resumindo, foram realizados experimentos para verificar as seguintes características em relação a obtenção de descritores utilizando a informação de cor:

- Quantidade de pontos complementares;
- Robustez em relação a transformação *affine*;
- Robustez em relação a posição da fonte de luz;
- Robustez em relação a variações da temperatura de luz.
- Utilização do espaço de cores CIE  $L^*a^*b^*$ .

## 4.1 Base de Dados de Imagens

Nos experimentos realizados, foi utilizada a Biblioteca de Imagens de Objetos de Amsterdam (ALOI - *Amsterdam Library of Object Images*) [Geusebroek et al., 2005]. Esta base de dados é composta de 1000 (mil) objetos e tem algumas características interessantes como a variação da posição da fonte de luz sem alteração da posição da câmera nem do objeto, e também alteração da temperatura de cor. Somente uma parte dos objetos da base de dados foram utilizados, dada a extensa lista de possíveis elementos disponíveis, que não pode ser toda tratada por limitações de tempo e do alto custo computacional envolvido. No Apêndice A, são exibidos todos os objetos utilizados nos experimentos.

Os objetos da base de dados são numerados de 1 a 999, perfazendo um total de 1000 objetos. Foram utilizados, para estes experimentos, os objetos de índice 110 à 139, incluindo as seguintes variações:

- Sem variações: foi utilizada uma imagem de cada objeto sem variação do ponto de vista do objeto, da temperatura de cor ou da posição da fonte de luz. Na base de dados, estes objetos são representados pelo código "r0".
- Variação do ponto de vista: foi utilizado uma imagem de cada objeto com variação do ponto de vista e sem nenhuma outra alteração. O objeto, neste caso, está rotacionado  $15^\circ$  em relação ao objeto sem variações.
- Variação da posição da fonte de luz: cada objeto é iluminado por uma de cinco fontes de luz, direcionadas ao objeto em ângulos diferentes. Além da variação do ângulo de luz em relação ao objeto, é feita uma variação da luz em relação à câmera de  $15^\circ$  e  $30^\circ$ , perfazendo um total de 15 alterações na posição da fonte luz, como visto na Figura 4.2.
- Variação da temperatura de cor: cada objeto é iluminado por uma fonte de luz frontal com variação apenas da temperatura de cor, com o sistema de captura de imagem calibrado para uma temperatura de cor de 3075K e a temperatura de cor variando entre 2175K (tons avermelhados) a 3075K (luz branca) com

Canal	Média de pontos	Desvio Padrão ( $\sigma$ )
<i>Grayscale</i>	166,50	28,50
V ( <i>value</i> )	175,43	30,43
H ( <i>hue</i> )	94,57	14,61

TABELA 4.1 – Média e desvio padrão do número total de descritores complementares obtidos entre os canais.

variação de 75K entre as imagens. São utilizadas 12 imagens para cada objeto, como visto na Figura 2.4, na página 24.

- Transformações *affine*: em adição às alterações presentes na base de dados, foram adicionadas três alterações às imagens sem variações. A partir de uma imagem sem alterações, foram obtidas outras três, uma por rotação de 30 °, uma por redução da escala para um terço do tamanho original, e uma pela aplicação das duas alterações anteriores. Estas imagens foram geradas a partir de objetos da base de dados, mas não estão presentes na base de dados original.

## 4.2 Número de pontos obtidos

Quando o registro de imagens é realizado utilizando-se descritores locais, é muito importante a obtenção de um número expressivo de características para que o registro possa ser realizado com mais precisão. Em aplicações como o reconhecimento de objetos em cenas complexas, esse número de características permite o reconhecimento de objetos parcialmente oclusos, uma vez que dado um número grande de pontos de interesse, a probabilidade de todos pontos estarem oclusos é menor. Nestes experimentos foram utilizados os objetos da base de dados sem alterações.

Para avaliar a quantidade de pontos de interesse encontrados apenas utilizando-se a informação de cor, comparou-se o número de pontos de interesse encontrado em uma imagem composta pela intensidade de luz (chamadas daqui por diante de *Grayscale*), que pode ser obtida pela conversão de uma imagem RGB pela Eq. 4.1 [BT.709, 1990], com imagens obtidas pela separação dos canais H e V do espaço de cores HSV, sendo que o canal H (*hue*) representa a informação de cor propriamente dita e o canal V (*value*) aproxima a intensidade de luz. Para este experimento foram utilizados os objetos da base de dados sem alterações.

$$G = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B \quad (4.1)$$

Em um primeiro experimento foi levada em consideração apenas a quantidade dos descritores obtidos para cada imagem. Os resultados estão sumarizados na Tabela 4.2, onde pode ser vista a média do número de descritores obtidos sobre as imagens e o desvio padrão dessa média.

Um dado importante é a disposição dos pontos na imagem. Se os descritores dos diversos canais forem obtidos em posições diferentes, a utilização da informação de cor será complementar à informação da intensidade de luz.

Para verificar a diferença de posição entre os descritores, foi comparada a distância de cada descritor de um canal ao descritor mais próximo em outro canal. Se

Canais	Média de pontos	Desvio Padrão ( $\sigma$ )	%
H $\times$ <i>Grayscale</i>	59,4	10,35	21.12%
H $\times$ V	61,9	11,09	18.62%
V $\times$ <i>Grayscale</i>	101,0	21,76	44.70%

TABELA 4.2 – Média e desvio padrão do número total de descritores obtidos em cada canal e percentual médio de pontos complementares.

a distância entre os descritores for menor que 3 *pixels*, os descritores são considerados "próximos". Descritores próximos pouco contribuem para o registro de imagem quando levamos em consideração a oclusão parcial de objetos. A Tabela 4.1 mostra a média e o desvio padrão do número de pontos com distância menor que o limiar estabelecido, e o percentual médio de pontos complementares obtidos.

Os resultados obtidos mostram que a utilização da informação de cor permite a obtenção de um número significativo de descritores, porém apenas parte destes descritores são complementares aos obtidos pelos canais com informação da intensidade de luz. A Figura 4.1 mostra uma comparação visual entre os pontos obtidos (na parte de baixo da imagem) com a intensidade de luz (à direita) e os canais V (ao centro) e H (à direita) do espaço de cores HSV. Na parte superior da imagem, as imagens das quais foram obtidos os pontos.

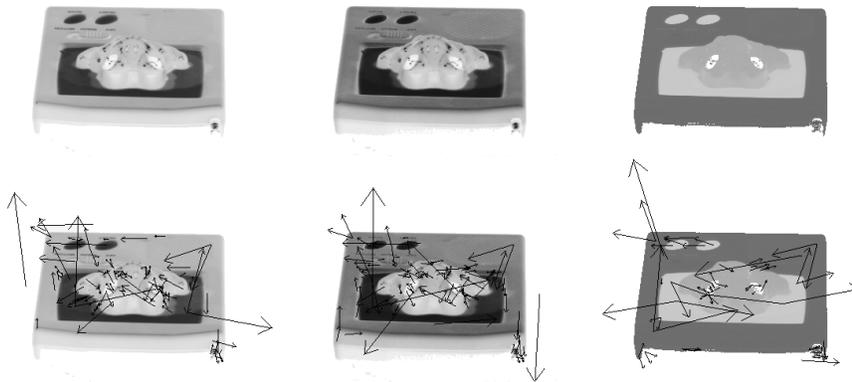


FIGURA 4.1 – Pontos obtidos com a intensidade de luz e os canais V e H.

O número de pontos complementares obtidos através da utilização do canal H mostra um incremento apenas razoável no número de pontos úteis para o registro de imagens, aliado ao baixo número total de pontos obtidos. A utilização apenas da informação cor, obtida através do canal H, para a obtenção de descritores locais utilizando SIFT não apresenta vantagem em relação a utilização exclusiva dos canais de intensidade de luz.

Canais	Média do Percentual de pontos encontrados	Desvio Padrão ( $\sigma$ )
H	62.95	2.76
V	73.25	2.30
<i>Grayscale</i>	72.19	1.71

TABELA 4.3 – Média e desvio padrão do percentual de descritores similares encontrados na comparação entre uma imagem de um objeto sem alteração a mesma imagem rotacionada 30°.

Canais	Média do Percentual de pontos encontrados	Desvio Padrão ( $\sigma$ )
H	65.69	4.11
V	76.74	2.64
<i>Grayscale</i>	79.87	2.48

TABELA 4.4 – Média e desvio padrão do percentual de descritores similares encontrados na comparação entre uma imagem de um objeto sem alteração a mesma imagem com escala um terço menor.

### 4.3 Robustez dos pontos obtidos com a informação de cor

O número de descritores obtidos com a utilização da informação de cor é importante para o registro de imagens, devido a necessidade de um grande número de pontos quando são utilizadas características locais para esta tarefa. Porém, é necessário verificar a robustez destes descritores frente a alterações na imagem. Os próximos experimentos procuram demonstrar o quanto os descritores obtidos a partir da informação de cor são robustos a transformações *affine*, e alterações na posição da fonte de iluminação e na temperatura de cor.

#### 4.3.1 Transformações *Affine* e Perspectiva

Neste experimento foi utilizada a base de dados com transformações *affine* e as imagens com alteração do ângulo de visão do objeto. Foram utilizadas imagens rotacionadas 30° no sentido horário, imagens com escala um terço menor que a original, imagens com as duas alterações anteriores combinadas, e imagens onde os objetos estão com um ângulo em relação à câmera de 15°

A Tabelas 4.3, 4.4, 4.5 e 4.6 mostram a média do percentual de descritores similares encontrados para cada alteração e o desvio padrão desta média.

Nota-se neste experimento que a utilização do canal de cor se mostra um menos robusto na obtenção de descritores em relação aos canais que representam a luminosidade. Neste experimento, quanto mais pontos com descritores similares, melhor será o registro entre as imagens, mesmo na presença de transformações.

A Tabela 4.7 mostra um resumo dos resultados obtidos neste experimento. Os descritores obtidos apenas com a informação de cor mostram-se um menos robustos a transformações *affine* e de perspectiva que os descritores obtidos a partir da intensidade de luz.

Canais	Média do Percentual de pontos encontrados	Desvio Padrão ( $\sigma$ )
H	58.00	4.50
V	79.37	2.21
<i>Grayscale</i>	77.15	2.27

TABELA 4.5 – Média e desvio padrão do percentual de descritores similares encontrados na comparação entre uma imagem de um objeto sem alteração a mesma imagem rotacionada  $30^\circ$  e com escala um terço menor.

Canais	Média do Percentual de pontos encontrados	Desvio Padrão ( $\sigma$ )
H	43.42	3.73
V	49.46	2.96
<i>Grayscale</i>	51.46	2.47

TABELA 4.6 – Média e desvio padrão do percentual de descritores encontrados na comparação entre uma imagem de um objeto sem alteração a mesma imagem com o objeto em um ângulo de  $15^\circ$  em relação a câmera.

Canais	Rotação	Escala	Escala e Rotação	Perspectiva
<i>Hue</i>	62.95	65.69	58.00	43.42
<i>Value</i>	73.25	76.74	79.73	49.46
<i>Grayscale</i>	72.19	79.87	77.15	51.46

TABELA 4.7 – Percentual médio do número de descritores semelhantes encontrados entre duas imagens, em relação a transformações nas imagens.

Canais	Média do Percentual de pontos encontrados	Desvio Padrão ( $\sigma$ )
H	25.64	0.86
V	6.77	0.93
<i>Grayscale</i>	37.63	0.93

TABELA 4.8 – Média e desvio padrão do percentual de descritores similares encontrados na comparação entre uma imagem de um objeto sem alteração e as imagens do mesmo objeto com alteração na posição da fonte de luz.

### 4.3.2 Posição da Fonte de Luz

Para avaliar os pontos obtidos com a informação de cor em relação a posição da fonte de luz, foi utilizada a base de dados com variação a posição da fonte de luz. Exemplos de imagens para um objeto são apresentados na Figura 4.2.

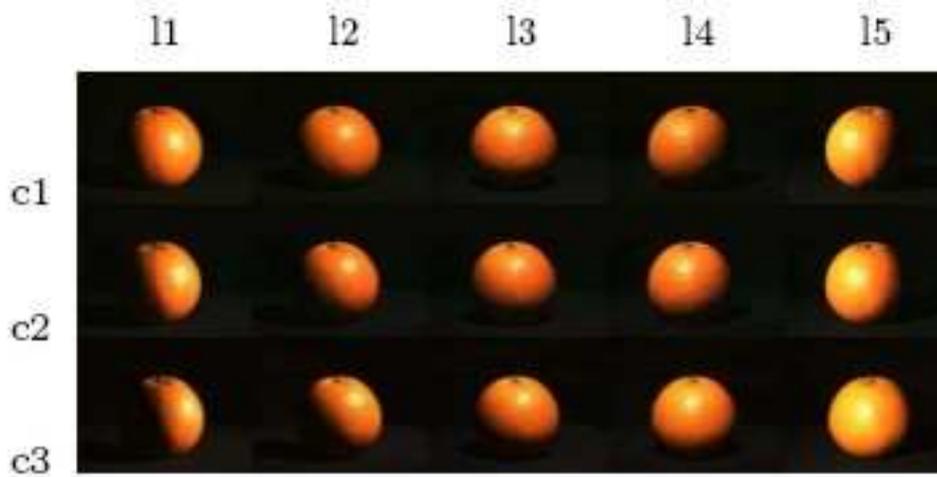


FIGURA 4.2 – Exemplo de objeto com variação da posição da fonte de luz.

Neste experimento, é feita a comparação dos descritores de uma imagem de referência, da base de dados sem alteração ("r0"), com os descritores de todas as imagens do mesmo objeto na base de dados com alteração da posição da fonte de luz. A comparação dos descritores é feita como descrito em [Lowe, 2004]. A Tabela 4.8 mostra a média e o desvio padrão do percentual de descritores detectados com sucesso. Para efeito de comparação são exibidos os resultados com os descritores obtidos com a informação da intensidade de luz.

Nota-se neste experimento uma redução significativa do número de descritores semelhantes encontrados nas imagens com variação da posição da fonte de luz em todos os canais. O canal com a informação de cor teve um desempenho um pouco abaixo dos canais de luminosidade. Assim como observado no experimento das transformações *affine*, os descritores obtidos a partir da utilização apenas da informação de cor contida no canal H, mostraram-se menos robustos em relação aos descritores obtidos a partir da intensidade de luz.

### 4.3.3 Temperatura de Cor

Para ilustrar o problema da temperatura de cor, utilizou-se as imagens da Figura 4.3 onde encontram-se imagens de uma mesma cena, obtidas com uma pequena variação de tempo (aproximadamente 2 minutos) entre cada tomada.

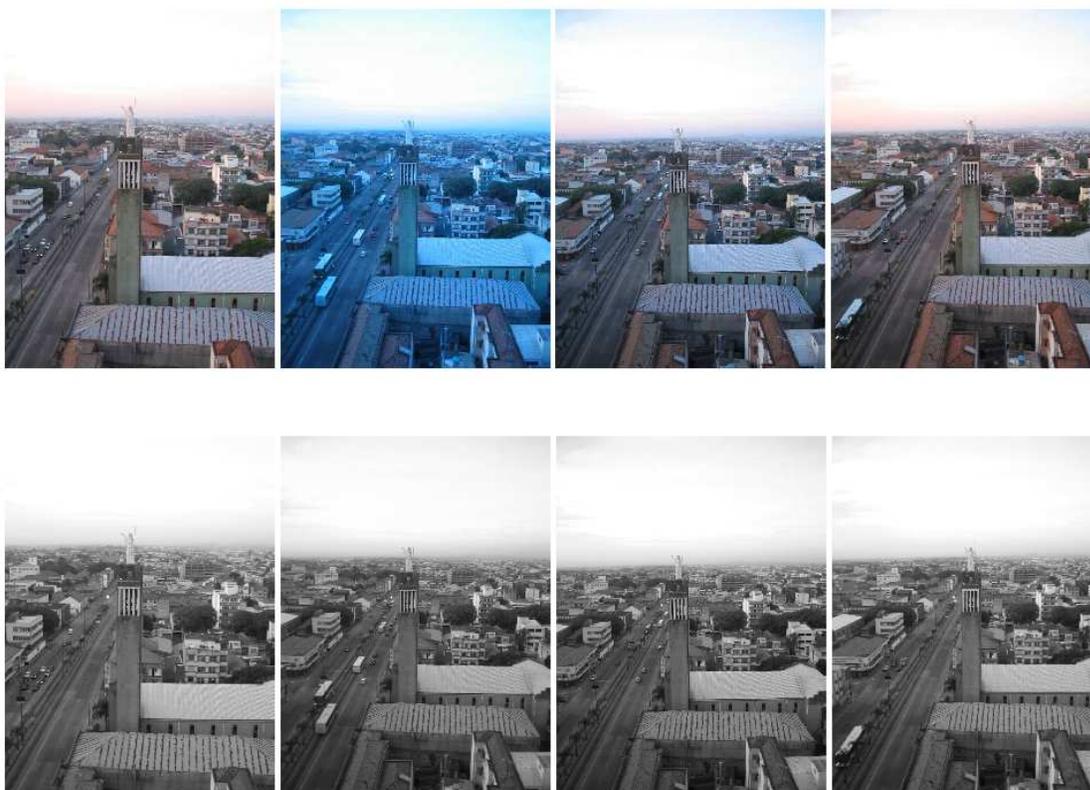


FIGURA 4.3 – Comparação entre imagens com problema de calibragem da temperatura de cor e sua versão em escala de cinzas.

Na parte superior da Figura 4.3 aparecem as versões coloridas das imagens com diferentes configurações de uma câmera digital para o ajuste da temperatura de cor. Nota-se claramente a alteração das cores nas imagens. No entanto, a intensidade de luz (escala de cinza) nas imagens é muito pouco afetada, como pode ser visto nas imagens na parte inferior da Figura 4.3. Estas imagens em escala de cinza foram obtidas a partir das respectivas imagens na parte superior, e nota-se pouca diferença entre elas. Este breve experimento demonstra a necessidade de calibrar corretamente a temperatura de cor em sistemas de visão.

Para verificar a robustez dos descritores obtidos exclusivamente pela informação de cor em relação a alteração da temperatura de cor na iluminação, foi realizado um experimento, variando a temperatura de cor da fonte de luz de 2175K até 3075K. Como as câmeras estão balanceadas para uma iluminação de 3075K, com essa temperatura o efeito que se tem é de luz branca. Temperaturas de cor menores adicionam um tom avermelhado aos objetos. Para cada um dos objetos, são utilizadas 12 imagens, com variação na temperatura de cor de 2175K a 3075, sendo uma variação de 75K para cada imagem. A comparação das imagens será realizada como no experimento anterior. A Tabela 4.9 mostra a média e o desvio padrão do

Canais	Média do Percentual de descritores encontrados	Desvio Padrão ( $\sigma$ )
H	50.91	1.25
V	81.07	0.65
<i>Grayscale</i>	87.70	0.41

TABELA 4.9 – Média e desvio padrão do percentual de descritores encontrados na comparação entre uma imagem de um objeto sem alteração e as imagens do mesmo objeto com alteração na temperatura de cor da fonte de luz.

Temperatura de cor	Número de descritores encontrados	Número de descritores similares	Percentual
2175K	47	9	19.15
2250K	41	19	46.34
2325K	29	15	51.72
2400K	42	9	21.43
2475K	36	12	33.33
2550K	39	18	46.15
2625K	34	16	47.06
2675K	46	29	63.04
2750K	35	28	80.00
2850K	49	37	75.51
2875K	44	28	63.64
3075K	49	40	81.63

TABELA 4.10 – Variação na comparação de descritores obtidos através da informação de cor frente a alteração da temperatura de cor.

percentual de descritores detectados com sucesso. Para efeito de comparação são exibidos os resultados com os descritores obtidos com a informação da intensidade de luz.

Neste experimento, nota-se que ocorre uma grande variação na qualidade dos descritores obtidos apenas com a informação de cor quando na presença de alterações na temperatura de cor. Para reduzir este problema é necessária a utilização de técnicas que permitam calibrar as cores antes da obtenção dos descritores. A Tabela 4.10 mostra para um objeto, a variação no número de descritores obtidos e no número de descritores semelhantes encontrados frente a alteração da temperatura de cor.

Apesar da variação do percentual de pontos similares encontrados não ter se mostrado diretamente relacionada a variação do temperatura de cor, foi observado nos experimentos que os melhores resultados estavam em temperaturas de cor mais próximas da imagem de referência, sendo indicado portanto, a calibragem do sistema de aquisição de imagem, ou o processamento posterior da imagem capturada a fim de diminuir a variação da temperatura de cor quando da obtenção de descritores locais utilizando a informação de cor.

Os descritores obtidos exclusivamente a partir da informação de cor contida no

canal H, mostraram-se muito sensíveis a variação da temperatura de cor, em relação aos canais de intensidade de luz. Para a obtenção de descritores locais, a partir da informação de cor do canal H, utilizando-se SIFT, é recomendável o uso de métodos de calibragem da temperatura de cor, seja no sistema de aquisição da imagem, seja por meio de pós-processamento da imagem obtida.

#### 4.4 Problemas com a presença de ruído nas imagens

Durante a realização dos experimentos, foi observado que as imagens utilizadas, apresentavam um nível de ruído nos canais de cores que dificultou a obtenção de descritores locais utilizando SIFT. O principal efeito deste ruído é a geração de "falsos descritores", ou seja, descritores obtidos a partir de ruído e não do objeto representado na imagem. Outras bases de dados, como por exemplo a COIL-100 ([Nene et al., 1996]), apresentam além do ruído na imagem, *artefatos* gerados, normalmente, por certos algoritmos de compactação. Estes artefatos não foram encontrados nas imagens utilizadas, porém é um ponto a ser avaliado quando do desenvolvimento de aplicações de visão computacional. Aparentemente, os problemas encontrados relativos ao ruído nas imagens, devem-se ao método e ao equipamento de captura das imagens.

A Figura 4.4 mostra o ruído presente no canal H (*Hue*) do espaço de cores HSV. Esta imagem foi obtida aumentando o contraste da imagem com o uso de equalização de histograma. A utilização de técnicas de processamento de imagens para a redução deste ruído, aliado ao aumento do contraste da imagem pode resultar em imagens melhores para a obtenção de um maior número de descritores locais utilizando SIFT, já que o contraste entre regiões da imagem é importante neste caso.



FIGURA 4.4 – Ruído presente no canal H.

Foram realizados alguns experimentos utilizando o espaço de cores CIE  $L^*a^*b^*$ . Durante estes experimentos foi verificado que as imagens resultante dos canais de cores ( $a^*$  e  $b^*$ ) possuía baixo contraste e muito ruído, como pode ser verificado nas

Figuras 4.5 e 4.6, que representam o canal  $b^*$ , sem e com equalização do histograma para aumentar o contraste.



FIGURA 4.5 – Baixo contraste apresentado no canal  $b^*$ .

Em aplicações de sistemas de visão, principalmente, para aquelas baseadas em descritores locais obtidos através de SIFT, ficou clara a importância e o impacto da aplicação de técnicas de processamento de sinal de forma a diminuir o ruído e aumentar o contraste das imagens. Neste trabalho, não foram utilizadas tais técnicas, e como efeito, não foi possível executar experimentos utilizando o espaço de cores CIE  $L^*a^*b^*$ , ou tentar aumentar o número de descritores locais obtidos, por meio do aumento de contraste.

Concluindo, apesar da invariância parcial a transformações *affine* e perspectiva e a invariância à escala obtidas com a utilização do algoritmo SIFT para a obtenção de descritores de uma imagem para fins de registro, a utilização de cores nesta tarefa apresenta outros problemas em relação à utilização de imagens em escalas de cinza. Problemas como o baixo contraste e ruído nos canais de cor, e alterações na temperatura de cor devem ser tratados com a utilização de técnicas de processamento de imagens em uma fase de pré-processamento, com o intuito de minimizar os efeitos destas alterações na obtenção de descritores locais, como os verificados durante os experimentos aqui apresentados.



FIGURA 4.6 – Ruído apresentado no canal  $b^*$ , após a equalização do histograma da imagem para aumentar o contraste.

## Capítulo 5

# Conclusão e Perspectivas Futuras

O registro de imagens tem um papel fundamental em sistemas de visão computacional e é a base de diversas aplicações de sistemas de visão computacional. A utilização de descritores locais permite grande flexibilidade para a realização do registro de imagens, e podem ser utilizados em uma larga gama de aplicações.

Neste trabalho, foi realizada uma revisão do estado da arte das técnicas utilizadas para registro de imagens, com foco em técnicas que utilizassem descritores locais. Estas técnicas são normalmente utilizadas com imagens em escalas de cinza representando intensidade de luz.

Buscando estender uma das técnicas de registro de imagens por descritores locais, a transformada de características invariantes a escala (SIFT), foi realizado também um estudo sobre a percepção de cores com o objetivo de adicionar a esta técnica a utilização da informação de cor para a obtenção dos descritores.

Através de experimentos foi verificado que a informação de cor, obtida a partir do canal H (*Hue*) do espaço de cores HSV, pode complementar o número de descritores utilizados para o registro de imagens. No entanto, a utilização da cor está sujeita a outras alterações na imagem que ou não afetam a utilização da intensidade de luz, ou tem um impacto menor. Por exemplo, a mudança de cores devido a alteração da temperatura de cor altera significativamente os descritores obtidos, dificultando o processo de registro. Medidas para balancear a iluminação, ou calibrar as cores, devem ser utilizadas nesse caso.

Este trabalho procurou avaliar a utilização da informação de cor diretamente na obtenção de descritores locais, como uma informação de baixo nível. Os resultados, não mostram claramente um ganho de eficiência no registro de imagem. Com a utilização do canal H (*Hue*) do espaço de cores HSV, verificou-se um aumento no número de pontos complementares para o registro de imagens que apresentam invariância parcial a transformações perspectiva e *affine* e a alterações na posição da fonte de luz. No entanto, estes pontos possuem muita sensibilidade em relação à temperatura de cor e não são tão robustos como os pontos obtidos através de imagens em escala de cinza.

Durante os experimentos foi observado que existe a necessidade de calibragem do sistema de captura de imagem, ou o processamento posterior da imagem para diminuir a variação da temperatura de cor entre a imagem capturada e as imagens de referência. A utilização do canal de cor (H) do espaço de cores HSV mostrou-se muito sensível a estas alterações.

Outro ponto observado foi a dificuldade de se obter descritores quando a ima-

gem apresenta baixo contraste. Este fato ocorreu tanto nos experimentos com o espaço de cores HSV, quanto em experimentos iniciais com o espaço de cores CIE  $L^*a^*b^*$ . A utilização da equalização de histograma para aumentar o contraste foi tentada, porém o aumento de ruído nos canais de cores, H do HSV,  $a^*$  e  $b^*$  do CIE  $L^*a^*b^*$ , foi muito acentuada, impossibilitando seu uso sem a utilização de filtros que reduzissem o ruído. Este aumento de ruído foi notado nos canais com informação de intensidade de luz, porém esse ruído era bem menor que o observado nos canais de cor.

A informação de cor pode ser utilizada também com uma informação de alto nível, aplicada no momento da comparação de características. Neste tipo de aplicação, os descritores SIFT são obtidos em imagens representadas pela intensidade de luz, mas a informação de cor é adicionada ao descritor da mesma maneira que a posição, a orientação e a escala, não sendo utilizada na comparação dos descritores.

Experimentos iniciais mostram um aumento de desempenho quando esta informação é utilizada para reduzir o espaço de busca na comparação de descritores SIFT. Como estes descritores são, normalmente, formados por vetores de características de 128 dimensões e existem muitos descritores em uma imagem (como demonstrado na Tabel 4.2), a seleção dos descritores pela sua cor antes da comparação diminui consideravelmente volume de dados a ser tratado e o tempo de comparação. Este tipo de seleção de descritores pode ser utilizado, por exemplo, em uma aplicação de navegação visual onde artefatos de uma cor específica indicam caminhos e/ou pontos de referência.

## 5.1 Perspectivas Futuras

Acredita-se que este trabalho tenha mostrado que a utilização unicamente da cor, descrita pelo canal H do espaço de cores HSV, na obtenção de descritores locais para o processo de registro de imagens oferece alguns desafios. Devido ao tempo disponível e a pouca bibliografia sobre a utilização da cor na obtenção de registro de imagens, nem todas as possibilidades foram abordadas e poderiam ser verificadas em trabalhos futuros.

É interessante investigar a utilização de outros espaços de cores, na tentativa de melhorar a resposta da cor na obtenção de descritores locais. Em experimentos iniciais com o espaço de cores  $L^*a^*b^*$ , a informação de luminosidade deste canal se mostrou melhor que o canal V do HSV ou a conversão para escala de cinza utilizada neste trabalho. No entanto, a informação de cor dividida entre os canais  $a^*$  e  $b^*$  mostrou-se sujeita a ruído e com contraste muito baixo. A utilização de técnicas de processamento de imagem para filtrar o ruído e aumentar o contraste a fim de melhorar a obtenção de pontos a partir destes canais, seria uma interessante possibilidade de estudo em trabalhos futuros.

Durante o trabalho, foi verificada a dificuldade de obter descritores semelhantes em condições de variação da temperatura de cor. Nestes casos, o espaço de cores HSV pode não ser o mais apropriado, porém o tratamento prévio da imagem corrigindo a alteração das cores pode melhorar consideravelmente a qualidade dos pontos obtidos.

Outra possibilidade é realizar uma investigação mais abrangente na utilização da cor como uma nova informação, de alto nível. Esta informação pode ser utilizada

quando da comparação de descritores, e pode ser interessante para diversos tipos de aplicações. Neste cenário, os descritores locais são obtidos utilizando apenas a informação da intensidade de luz e a informação de cor é uma informação adicional para este descritor. De posse desta informação, apenas os descritores com a mesma cor seriam processados durante o processo de comparação de características, acelerando o processo.

# Apêndice A

## Objetos Utilizados nos Experimentos

A Figura A.1 mostra as imagens dos objetos utilizados nos experimentos realizados, posicionados de frente para a câmera e iluminados por luz branca.



FIGURA A.1 – Imagens dos objetos utilizados nos experimentos.

# Bibliografia

- [Belongie et al., 1998] Belongie, S., Carson, C., Greenspan, H., and Malik, J. (1998). Color- and texture-based image segmentation using the expectation-maximization algorithm and its application to content-based image retrieval. In *IEEE International Conference on Computer Vision*, pages 675–682, Mumbai, India.
- [Bender, 2003] Bender, T. (2003). Classificação e recuperação de imagens por cor utilizando redes neurais artificiais. Master’s thesis, Mestrado em Computação Aplicada - UNISINOS.
- [Berthilsson, 1998] Berthilsson, R. (1998). Affine correlation. In *International Conference on Pattern Recognition*, pages 1458–1460.
- [Brown, 1992] Brown, L. G. (1992). A survey of image registration techniques. *ACM Computing Surveys*, pages 326–376.
- [Brown and Lowe, 2002] Brown, M. and Lowe, D. G. (2002). Invariant features from interest point groups. In *British Machine Vision Conference*, pages 656–665, Cardiff, Wales.
- [Brown and Lowe, 2003] Brown, M. and Lowe, D. G. (2003). Recognising panoramas. In *International Conference on Computer Vision*, pages 1218–25.
- [BT.709, 1990] BT.709, I.-R. R. (1990). *Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange*. International Telecommunication Union - Radiocommunication Sector, 1211 Geneva 20, Switzerland.
- [Chalermwat and El-ghazawi, 1999] Chalermwat, P. and El-ghazawi, T. (1999). Multi-resolution image registration using genetics. In *IEEE International Conference on Image Processing*.
- [Crowley and Martin, 1995] Crowley, J. L. and Martin, J. (1995). Comparison of correlation techniques. In *International Conference on Intelligent Autonomous Systems*, pages 86–93, Karlsruhe.
- [CSBR, 2005] CSBR (2005). The Center for Biometrics and Security Research. <http://www.sinobiometrics.com/main.html>. acessado em maio de 2005.
- [Daugman, 1993] Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11).

- [Faloutsos et al., 1994] Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., and Equitz, W. (1994). Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3(3/4):231–262.
- [Felipe et al., 2003] Felipe, J. C., Traina, A. J. M., , and Jr., C. T. (2003). Retrieval by content of medical images using texture for tissue identification. In *Proc. 16th IEEE Symposium on Computer-based Medical Systems*, pages 26 – 27, New York.
- [Foley et al., 1990] Foley, J. D., van Dam, A., Feiner, S. K., and Hughes, J. F. (1990). *Computer Graphics – Principles and Practice (2nd Ed.)*. Addison-Wesley.
- [Geusebroek et al., 2005] Geusebroek, J. M., Burghouts, G. J., and Smeulders, A. W. M. (2005). The Amsterdam library of object images. *Int. J. Comput. Vision*, 61(1):103–112.
- [Gomes and Velho, 1994] Gomes, J. and Velho, L. (1994). *Computação Gráfica: Imagem*. IMPA/SBM, Rio de Janeiro.
- [Gonzalez and Woods, 1992] Gonzalez, R. C. and Woods, R. E. (1992). *Digital Image Processing*. Addison-Wesley, Reading, MA, USA, 3rd edition.
- [Gouet and Boujemaa, 2001] Gouet, V. and Boujemaa, N. (2001). Object-based queries using color points of interest. In *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'01)*.
- [Harris and Stephens, 1988] Harris, C. J. and Stephens, M. (1988). A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, Manchester.
- [Hill et al., 2001] Hill, D. L. G., Batchelor, P. G., Holden, M., and Hawkes, D. J. (2001). Medical image registration. *Physics in Medicine and Biology*, 46.
- [Hu et al., 2004] Hu, W., Tan, T., Wang, L., and Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. In *IEEE Transactions on Systems, Man and Cybernetics, Part C*, volume 34.
- [Huang et al., 1997] Huang, J., Kumar, S. H., Mithra, M., Zhu, W., and Zabih, R. (1997). Image indexing using color correlogram. In *EEE Comp. Soc. Conf. Comp. Vis. and Patt. Rec.*, pages 762–768.
- [Johnson and Fairchild, 1999] Johnson, G. M. and Fairchild, M. D. (1999). Full-spectral color calculations in realistic image synthesis. *IEEE Computer Graphics and Applications*, 19(4):47–53.
- [Jones et al., 1997] Jones, S. D., Andersen, C. S., and Crowley, J. L. (1997). Appearance based process for visual navigation. In *IEEE Intelligent Robots and Systems Conference*, volume 2.
- [Jung et al., 2005] Jung, C. R., Osório, F. S., Kelber, C. R., and Heinen, F. J. (2005). Computação embarcada: Projeto e implementação de veículos autôno-

- mos inteligentes. In *JAI – Jornada de Atualização em Informática, Congresso do SBC 2005*, São Leopoldo, RS.
- [Kaneko et al., 2002] Kaneko, S., Murase, I., and Igarashi, S. (2002). Robust image registration by increment sign correlation. *Pattern Recognition*, 35(10):2223–2234.
- [K.Mikolajczyk and C.Schmid, 2001] K.Mikolajczyk and C.Schmid (2001). Indexing based on scale invariant interest points. In *International Conference on Computer Vision*, pages 525–531.
- [Lammens, 1994] Lammens, J. M. (1994). *A Computational Model of Color Perception and Color Naming*. PhD thesis, Technical Report 94-26, Department of Computer Science, State University of New York at Buffalo, Buffalo, NY.
- [Lewis, 1995] Lewis, J. P. (1995). Fast template matching. In *Vision Interface*, pages 120–123.
- [Lowe et al., 2002] Lowe, D., Little, J., and Se, S. (2002). Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research*, 21(8).
- [Lowe, 1999] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *IEEE International Conference on Computer Vision*.
- [Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110.
- [Makela et al., 2002] Makela, T., Clarysse, P., Sipila, O., Pauna, N., Pham, Q. C., Katila, T., and Magnin, I. (2002). A review of cardiac image registration methods. *IEEE Transactions on Medical Imaging*, 21(9).
- [Marcel and Bengio, 2002] Marcel, S. and Bengio, S. (2002). Improving face verification using skin color information. In *16th International Conference on Pattern Recognition, ICPR*, volume 2, pages 11–15. IEEE Computer Society Press.
- [Matsumoto et al., 1996] Matsumoto, Y., Inaba, M., and Inoue, H. (1996). Visual navigation using view-sequenced route representation. In *IEEE International Conference on Robotics and Automation*.
- [Matsumoto et al., 1999] Matsumoto, Y., Inaba, M., and Inoue, H. (1999). Visual navigation using omnidirectional view sequence. In *IEEE/RSJ International Conference on Intelligent Robots and System*.
- [Mikolajczyk and Schmid, 2002] Mikolajczyk, K. and Schmid, C. (2002). An affine invariant interest point detector. In *European Conference on Computer Vision*, volume 1, pages 128–142.
- [Nastar et al., 1998] Nastar, C., Mitschke, M., Boujemaa, N., Meilhac, C., Bernard, H., and Mautref, M. (1998). Retrieving images by content: The surfimage system. *Lecture Notes in Computer Science*, 1508:110–??

- [Nene et al., 1996] Nene, S. A., Nayar, S. K., and Murase, H. (1996). Columbia object image library (coil-100). Technical Report CUCS-006-96, Columbia University.
- [Padgham and Saunders, 1975] Padgham, C. A. and Saunders, J. E. (1975). *The Perception of Light and Color*. Academic Press, New York.
- [Plataniotis and Venetsanopoulos, 2000] Plataniotis, K. N. and Venetsanopoulos, A. N. (2000). *Color image processing and applications*. Springer-Verlag New York, Inc., New York, NY, USA.
- [Poynton, 2005] Poynton, C. (Acessado em MAio do 2005). *Frequently Asked Questions about Color*. <http://www.poynton.com/PDFs/ColorFAQ.pdf>.
- [Ravela et al., 1996] Ravela, S., Manmatha, R., and Riseman, E. M. (1996). Image retrieval using scale-space matching. *Lecture Notes in Computer Science*, 1064.
- [Righes, 2004] Righes, E. M. (2004). Processamento de imagens para navegação de robôs autônomos. Trabalho de Conclusão de Curso. Unisinos. Informática.
- [Rothganger et al., 2003] Rothganger, F., Lazebnik, S., Schmid, C., and Ponce, J. (2003). 3D object modeling and recognition using affine-invariant patches and multi-view spatial constraints. *ICCV*, 0(0).
- [Schmid, 2001] Schmid, C. (2001). Constructing models for content-based image retrieval. In *Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, USA.
- [Schmid and Mohr, 1997] Schmid, C. and Mohr, R. (1997). Local greyvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535.
- [Sebe and Lew, 2000] Sebe, N. and Lew, M. S. (2000). Color based retrieval and recognition. In *IEEE International Conference on Multimedia and Expo (I)*, pages 311–314.
- [Shyu et al., 1998] Shyu, C. R., Brodley, C. E., Kak, A. C., Kosaka, A., Aisen, A., and Broderick, L. (1998). Local versus global features for content-based image retrieval. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, Santa Barbara, CA. IEEE Computer Society.
- [Stricker and Orengo, 1995] Stricker, M. A. and Orengo, M. (1995). Similarity of color images. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 381–392.
- [Van Wie and Stein, 1977] Van Wie, P. and Stein, M. (1977). A landsat digital image recognition system. *IEEE Transactions on Geoscience Eletronics*, GE-15(3):130–137.
- [Vatsa et al., 2004] Vatsa, M., Singh, R., and Gupta, P. (2004). Comparision of iris recognition algorithms. In *International Conference on Intelligence Sensing and Information Processing*.

- [Witkin, 1986] Witkin, A. P. (1986). Scale space filtering. In Pentland, A. P., editor, *From Pixels to Predicates: Recent Advances in Computational and Robot Vision*, pages 5–19, Norwood, NJ. Ablex.
- [Wyszecki and Stiles, 1982] Wyszecki, G. and Stiles, W. S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley.
- [Wyszecki and Styles, 1982] Wyszecki, G. and Styles, W. S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, New York, NY.
- [Zitova and Flusser, 2003] Zitova, B. and Flusser, J. (2003). Image registration methods: A survey. In *Second International Workshop on Biomedical Image Registration*.