CENTRO UNIVERSITÁRIO DA FEI FERNANDO CARUSO OLIVIO

# UM MODELO BAYESIANO COM DIVERGÊNCIA DE KULLBACK-LEIBLER ESTENDIDA PARA RECONHECIMENTO DE OBJETOS 3D BASEADOS EM MÚLTIPLAS VISÕES

São Bernardo do Campo, SP 2009

#### FERNANDO CARUSO OLIVIO

#### UM MODELO BAYESIANO COM DIVERGÊNCIA DE KULLBACK-LEIBLER ESTENDIDA PARA RECONHECIMENTO DE OBJETOS 3D BASEADOS EM MÚLTIPLAS VISÕES

Dissertação de Mestrado apresentada ao Centro Universitário da FEI para obtenção do título de Mestre em Engenharia Elétrica, orientado pelo Prof. Dr. Paulo Sérgio Silva Rodrigues

São Bernardo do Campo, SP 2009

Olivio, Fernando Caruso

Um modelo bayesiano com divergência de Kullback-Leibler estendida para reconhecimento de objetos 3d baseados em múltiplas visões / Fernando Caruso Olivio. - São Bernardo do Campo, 2009. 78 f. : il.

Dissertação (Mestrado) - Centro Universitário da FEI.

Orientador: Prof. Dr. Paulo Sérgio Silva Rodrigues

1. Entropia não-extensiva. 2. Reconhecimento tridimensional. 3. Redes Bayesianas. I. Rodrigues, Paulo Sérgio Silva, orient., II. Título.

CDU 681.3.02



Centro Universitário da FEI

### APRESENTAÇÃO DE DISSERTAÇÃO ATA DA BANCA JULGADORA

#### Programa de Mestrado de Engenharia Elétrica

Aluno: Fernando Caruso Olivio

Matrícula: 1071042

Título do Trabalho: Um modelo bayesiano com divergência de Kullback-Leibler estendida para reconhecimento de objetos 3D baseados em múltiplas visões.

Área de Concentração: Inteligência Artificial Aplicada à Automação

Orientador: Prof. Dr. Paulo Sérgio Silva Rodrigues

Data da realização da defesa: 16 / dezembro / 2009

A Banca Julgadora abaixo-assinada atribuiu ao candidato o seguinte:

APROVADO 💹

REPROVADO

São Bernardo do Campo, 16/12/2009

MEMBROS DA BANCA JULGADORA	
Prof. Dr. Paulo Sérgio Silva Rodrigues	
Ass.: als foligue	
Prof. <sup>a</sup> Dr. <sup>a</sup> Rosângela Barreto Biasi Gin	
Ass.: Lof-	
Prof. Dr. Flávius Portella Ribas Martins	
Ass.: The factule UP	

#### VERSÃO FINAL DA DISSERTAÇÃO

ENDOSSO DO ORIENTADOR APÓS A INCLUSÃO DAS RECOMENDAÇÕES DA BANCA EXAMINADORA Aprovação do Coordenador do Programa de Pós-graduação

marulo An

Prof. Dr. Marcelo Antonio Pavanello

A Deus, a todos que de alguma maneira acreditaram em mim e, principalmente, àqueles que me apoiaram. Dedico este trabalho à minha família e, em especial, à minha avó Maria, à minha irmã Adriana, à minha mãe Lídia e também à minha esposa Christianne.

#### AGRADECIMENTOS

Agradeço à Deus por ter me dado a oportunidade de viver este momento tão especial e por ter sido cercado de pessoas maravilhosas, que me ajudaram e me fizeram acreditar. Dou graças por vivenciar as experiências juntamente com os novos amigos que fiz nestes três anos de mestrado: Danilo Eduardo, Danilo Nunes, Fussek, Giuliano, Haddad, Horst, Leão, Rodolpho Freire, Thiago, Valquiria e Vanessa, entre outros. Sinto-me honrado por ter sido orientado por uma pessoa tão autentica e leal quanto o Prof. Paulo Sérgio Rodrigues, que mostrou o quão longe o trabalho árduo e a dedicação podem levar.

Gostaria de agradecer a todos os amigos de trabalho que me auxiliaram, por vezes trabalhando em dobro, principalmente nos dias em que tive de escolher entre os estudos e a profissão.

Finalmente, agradeço à minha família, meu pai e principalmente minha mãe, que soube educarme, mostrando os caminhos por toda vida, através do incentivo, do suporte, da dedicação e do ânimo para lutar por meus sonhos, mesmo nos momentos difíceis desta jornada. A obra aqui apresentada não seria possível sem o amor, por isso, lembro-me enternecidamente de minha esposa Christianne, que se privou de minha companhia mas, como uma companheira para toda a vida, sempre mostrou uma enorme confiança em meus esforços e no meu potencial, algo tangível apenas por quem ama.

#### **RESUMO**

Este trabalho apresenta um modelo Bayesiano que combina as características de cor, forma e textura para o reconhecimento de objetos em três dimensões. As interações espaciais ou temporais de longo alcance dessas características permitem modelar a probabilidade de se observar em cada dessas evidências nos objetos com a distância de Kullback-Liebler estendida, que é um conceito recentemente proposto na mecânica estatística. O modelo Bayesiano proposto pode ser usado em diversas aplicações, mas enfatizamos o trabalho cooperativo de diversos observadores para executar a tarefa de reconhecimento tridimensional. Os experimentos com uma base de dados de informações a priori sugerem que o modelo atinge o seu melhor desempenho a partir da inclusão do terceiro observador, indicando resultados promissores.

Palavras-chave: Reconhecimento tridimensional. Redes Bayesianas. Entropia não-extensíva.

#### ABSTRACT

This paper presents a Bayesian model that combines the color, shape and texture features for the recognition of objects in three dimensions. The long range spatial and temporal interactions of these features allow to model the probability to observe in each of these evidences the objects with the extended Kullback-Liebler distance, which is a recently proposed concept in statistical mechanics. The Bayesian model proposed can be used in several applications but we emphasize the cooperative work of several observers in order to perform the task of three dimensional recognition. Experiments with a large database of a priori information suggest that the model reaches its best performance with the inclusion of at most a third observer, showing promising results.

Key words: Three-dimensional recognition. Bayesian Networks. Non-extensive entropy.

# Lista de Figuras

3.1 3.2 3.3 3.4	Grafo de representação de uma RB	25 35 40 40
3.5	Resultado da aquisição de geometria e volume de objetos reais	41
4.1	Modelo de RB Genérico.	42
4.2	Modelo de classificador para análise de características de cor	44
4.3	Modelo de classificador para análise de características de forma.	44
4.4	Modelo de classificador para analise de características de textura	45
4.5	Modelo de classificador para analise conjunta de características de cor, forma e textura.	45
4.0	múltiplas visões	45
4.7	Proposta de arquitetura para um sistema de recuperação de informações baseado em múltiplas visões.	47
5.1	Amostra da base de dados Columbia. Nesta amostra, apresenta-se 6 classes distintas	
	na horizontal, sendo cada uma registrada em 5 ângulos diferentes	51
5.2	Gráfico da precisão e revocação do objeto assimétrico (referente à segunda linha da	
	Figura 5.1)	52
5.3	Gráfico da precisão e revocação do objeto simétrico (referente à sexta linha da	
5 1	Figura 5.1). $\ldots$	52
5.4	Representação grafica da Precisão e Revocação de diversas classes com parametro	51
55	entropico $q = 0.1$	54 54
5.6	Gráfico da precisão e revocação representando um aumento na precisão pela adição de novos observadores. Dados extraídos a partir das análises da classe de objetos	54
	68 (Figura 5.7).	55
5.7	Quatro vistas do objeto grampeador da classe 68. Apresentado nos ângulos 80°, 75°, 195° e 315° (da esquerda para a direita).	55
5.8	Gráfico da precisão e revocação representando um aumento na precisão pela adição	
	de novos observadores. Dados extraídos a partir da análise da classe de objetos 14	
	(Figura 5.9)	56
5.9	Quatro vistas do objeto gato da classe 14. Objeto apresentado nos ângulos 90°,	
- 10	$265^{\circ}$ , $175^{\circ}$ e $330^{\circ}$ (da esquerda para a direita)	56
5.10	Gráfico da precisão e revocação representando uma redução na precisão pela adição de novos observadores. Dados extraídos a partir de análises da classe de objetos 84	
	(Figura 5.11)	57
5.11	Quatro vistas do objeto pacote amareio da classe 84. Objeto apresentado nos ângu- los $235^{\circ}$ , $70^{\circ}$ , $105^{\circ}$ o $140^{\circ}$ (de osquerdo paro o direito)	57
5 1 2	$105 \ 253$ , $70$ , $193$ e 140 (ua esquerua para a difella)	51
J.12	$140^{\circ}$ (da esquerda para a direita)	58

5.13	Quatro vistas do objeto da classe 1 representando os ângulos 235°, 70°, 195° e 140°	~0
- 1 4	(da esquerda para a direita)	59
5.14	Quatro vistas do objeto da classe 3 representando os ângulos 235°, 70°, 195° e 140°	50
	(da esquerda para a direita)	59
5.15	Quatro vistas do objeto da classe 6 representando os ângulos 235°, 70°, 195° e 140°	
	(da esquerda para a direita).	59
5.16	Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro	
_	entrópico $q = 0.1$	60
5.17	Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro	
	entrópico $q = 0.2$ .	61
5.18	Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro	
	entrópico $q = 0.3.$	62
5.19	Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro	
	entrópico $q = 0.4$	62
5.20	Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro	
	entrópico $q = 0.5$	63
5.21	Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro	
	entrópico $q = 0.6$	63
5.22	Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro	
	entrópico $q = 0.7$	64
5.23	Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro	
	entrópico $q = 0.8$	65
5.24	Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro	
	entrópico $q = 0.9$	65
5.25	Vistas do objeto da classe 13, que apresenta vistas com grandes variações nas ca-	
	racterísticas de cor, forma e textura entre si	66
5.26	Vistas do objeto da classe 15, que apresenta vistas com grandes variações nas ca-	
	racterísticas de cor, forma e textura entre si	66
5.27	Vistas do objeto da classe 25, que possuem pouca variação nas características de	
	cor, forma e textura entre as vistas	67
5.28	Objetos da classe 35, que possuem pouca variação nas características de cor, forma	
	e textura entre as vistas	67
5.29	Gráfico da média obtida para todas as classes da base de dados	68
5.30	Gráfico que representa o resultado do experimento utilizando o parâmetro entrópico	
	q = 0.1 no <i>matching</i> das características de cor	70
5.31	Gráfico que representa o resultado do experimento utilizando o parâmetro entrópico	
	q = 0.1 no <i>matching</i> das características de forma	71
5.32	Gráfico que representa o resultado do experimento utilizando o parâmetro entrópico	
	q = 0, 1 no <i>matching</i> das características de textura	72

# Lista de Tabelas

3.1	Tabela comparativa entre a entropia extensiva e não-extensiva	31
5.1	Valores ordenados da busca por objetos da classe 84	58
5.2	Médias dos valores das curvas contidas no gráfico da Figura 5.29.	68
5.3	Tabela com os valores ordenados dos cálculos das média das precisões com parâmetro $q$ variando entre 0.1 à 0.9	69
5.4	Tabela com os valores dos desvios padrão das precisões médias com parâmetro não- extensivo $q$ variando entre 0.1 e 0.9.	69
5.5	Percentuais de <i>matchings</i> para cor, forma e textura, exemplificando classes nas quais a forma foi a característica mais relevante na RIBC.	72
5.6	Percentuais de " <i>matchings</i> " para cor, forma e textura, exemplificando classes nas quais a textura foi a característica mais relevante na RIBC.	72

# Lista de Símbolos e Abreviaturas

2D	—	bidimensional, que possui duas dimensões.
3D	—	tridimensional, que possui três dimensões.
q	—	Parâmetro entrópico.
RCO		Reconhecimento Cooperativo de Objetos.
RB		Redes Bayesianas.
SQL		Structured Query Language, linguagem utilizada em bancos de dados.
HTML	—	Hyper Text Markup Language.
SVM		Support Vector Machine.
NB		naïve-bayes.
SNB	—	semi-naïve-bayes.
RBE	—	Redes Bayesianas Entrópicas.
PCA		Principal Component Analysis.
Н		Hipótese.
TPC		Tabela de probabilidade conjunta.
S		Entropia.
$S_q$		Entropia de Tsallis.
$\dot{P(i)}$		Probabilidade de <i>i</i> .
D(A B)		Divergência entre duas distribuições $A \in B$ .
$D_{KL}(A B)$	—	Divergência de Kullback-Leibler entre duas distribuições A e B.
$D_{KL_q}(A B)$		Divergência de Kullback-Leibler estendida entre duas distribuições A e B.
$D_{KL_e}(A B)$		Divergência de Kullback-Leibler estendida na forma simétrica entre
		duas distribuições $A \in B$ .
RGB		Red, Green, Blue, sistema de representação de cores
		formado pela combinação das cores vermelho, verde e azul.
CMY		Cyan, Magenta, Yellow sistema de representação de cores
		formado pela combinação das cores ciano, magenta e amarelo.
HSV		Hue, Saturation, Value sistema de representação de cores
		formado pela combinação de matiz, saturação e valor.
HSV162		Discretização do espaço HSV em apenas 162 valores possíveis.
RA		Realidade Aumentada.
$K_c$		Representação das características de cor no classificador.
$K_f$		Representação das características de forma no classificador.
$K_t$		Representação das características de textura no classificador.
IE		Integrador de Evidências.
RIBC		Recuperação de Imagens Baseado em Conteúdo.
PR		Método da Precisão x Revocação.

# Sumário

Ll	ISTA I	DE FIGURAS	8
Ll	ISTA I	DE TABELAS	10
Ll	ISTA 1	DE SÍMBOLOS E ABREVIATURAS	11
1	INT	ſRODUÇÃO	14
	1.1	Objetivo	15
	1.2	Contribuições da Dissertação	16
2	TR	ABALHOS RELACIONADOS	17
	2.1	Redes Bayesianas (RB)	17
	2.2	Entropia e Divergência de Kullback-Leibler	20
	2.3	Reconhecimento Cooperativo de Objetos	22
3	CO	NCEITOS FUNDAMENTAIS	24
	3.1	Teoria de Bayes	24
	3.2	Redes Bayesianas	24
	3.3	Entropia	26
	3.4	Teoria de informação e entropia	27
		3.4.1 Entropia de Shannon segundo uma abordagem física	29
	3.5	Entropia Não-Extensiva	30
	3.6	Medidas de Distância	32
		3.6.1 Medidas Clássicas	32
		3.6.2 Modelo Vetorial	32
		3.6.3 Divergência de Kullback-Leibler	32
		3.6.4 Divergência de Kullback-Leibler estendida	33
	3.7	Características para reconhecimento de objetos	34
		3.7.1 Descritores de Cor	34
		3.7.2 Descritores de Forma	36
		3.7.3 Descritores de Textura	36
	3.8	Aplicações do modelo proposto	39
		3.8.1 Realidade Aumentada	39

		3.8.2	Análise de cenas e vídeos	. 40
		3.8.3	Scanner 3D	. 41
4	PR	OPOS	ГА	42
	4.1	Model	lo Bayesiano para Recuperação de Informação	. 42
	4.2	Estrut	ura Proposta	. 44
	4.3	Medid	la de Similaridade Proposta	. 47
5	EX	PERIN	MENTOS	49
	5.1	Metod	lologia de testes	. 49
	5.2	Base c	le dados utilizada	. 50
	5.3	Hardw	vare Utilizado	. 52
	5.4	Precisã	io x Revocação (PR)	. 53
	5.5	Desem	penho classe x observador	. 53
	5.6	Relaçã	o Observadores x Objetos x Parâmetro entrópico	. 59
	5.7	Melho	or parâmetro entrópico q	. 67
	5.8	Casam	nento por Cor, Forma e Textura	. 69
6	CO	NCLU	JSÕES	73

# Capítulo 1

# INTRODUÇÃO

Os sistemas digitais têm sido criados para diversos fins, espalhando-se por inúmeros ramos de atividades, sejam eles militares, industriais, médicos ou de entretenimento. Porém, existem tarefas que são executadas quase que exclusivamente por humanos, como é o caso do reconhecimento visual de objetos. Reconhecer objetos em cenas ou individualmente é uma tarefa inerentemente humana, que requer alto grau de cognição e interpretação de imagens. Para que esta capacidade de reconhecimento seja aplicada a sistemas de visão computacional, passa-se por diversas etapas, entre elas aquelas que envolvem a visão de baixo nível (eliminação de ruídos, distorção e transformações geométricas); visão de médio nível (segmentação e descrição) e visão de alto nível (reconhecimento e classificação).

O reconhecimento de objetos é um dos problemas mais antigos em Visão Computacional, onde diversas propostas de modelagem tem sido avaliadas há décadas. O reconhecimento tridimensional de objetos é uma tarefa que envolve a captura de várias perspectivas com o uso de uma ou mais câmeras. Eventualmente, essa tarefa pode ser realizada de maneira cooperativa por vários usuários.

O reconhecimento de objetos cooperativo apresenta diversos problemas que tornam sua modelagem complexa. Estes problemas são notados quando se verifica que, além do reconhecimento em si, também é necessária a troca de informações entre os usuários, a correlação adequada entre as diversas vistas e a correspondência entre os pontos da cena.

A questão da troca de informações conjuntamente com o problema do reconhecimento de objetos pode ser visto, por exemplo, em ambientes virtuais colaborativos. Nestes sistemas cada observador deve reconhecer a cena ou o objeto nela contido individualmente e comunicar aos demais quais as suas inferências e o estado atual. Por outro lado, a importância da análise da correlação ou correspondência entre as diversas vistas de um objeto é amplamente estudada para a construção de aplicações como *scanners* tridimensionais (ou reconstrução tridimensional).

Desta forma, pode-se destacar que a existência de um ou mais observadores, em aplicações como as citadas, feitas para finalidades diferentes, dependem do mesmo requisito, que é o reconhe-

cimento de objetos em imagens bidimensionais. Este é um problema estudado há décadas, envolvendo análise de uma ou mais características, tais como: cor, forma ou textura. Estas informações, tanto individualmente para cada observador, quanto de maneira global para um sistema cooperativo, podem tanto levar a uma acertada inferência (ou precisão no reconhecimento), quanto podem gerar informações ruidosas ou regiões espúrias. Isto significa que não há relação entre o aumento da quantidade de informações e a precisão do reconhecimento, não importando se o espaço analisado é bidimensional ou tridimensional. Logo, percebe-se a necessidade do emprego de modelos que sejam capazes de adequadamente inferir sobre um objeto utilizando informações advindas de uma única vista (ponto de observação) ou de diversas vistas (visão cooperativa).

Por outro lado, estudos recentes têm demonstrado que a relação entre as características de uma imagem (seja em cenas ou em objetos individuais) são altamente correlacionadas, envolvendo interações espaciais e temporais de longo alcance. As interações físicas de longo alcance tanto no espaço quanto no tempo são também estudadas na mecânica estatística, que recentemente tem sido empregada nas áreas de Visão Computacional e Processamento de Imagens através da estatística de Tsallis, chamada q-entropia (RODRIGUES, 2003) (ESQUEF, 2002).

O trabalho proposto nesta dissertação de mestrado apresenta um modelo Bayesiano para o reconhecimento de objetos em três dimensões a partir de diversas vistas capturadas por múltiplos observadores. O modelo utilizado considera o grau de crença de um objeto ser observado como a divergência de Kullback-Leibler estendida para estatística de Tsallis entre as distribuições de cor, forma e textura de um modelo *a priori* observado.

Os experimentos sugerem que o sistema composto por características de cor, forma e textura pode ser considerado como um sistema físico não-extensivo (TSALLIS, 1988), sendo possível aplicá-lo a sistemas de recuperação de imagens com base no conteúdo, reconstrução 3D e sistemas de Realidade Aumentada.

Esta monografia está dividida da seguinte maneira: no Capitulo 2 são apresentados trabalhos relacionados; no Capitulo 3 descritos os principais conceitos que serão utilizados no decorrer do trabalho; no Capitulo 4 é apresentada a proposta de pesquisa; no Capitulo 5 são apresentados os experimentos executados para a validação do método; finalmente, no Capitulo 6 apresenta-se as conclusões e discussões geradas pelo estudo.

# 1.1 Objetivo

Propor um modelo bayesiano utilizando a divergência de Kullback-Leibler estendida para a construção de um sistema de reconhecimento cooperativo de objetos 3D.

# 1.2 Contribuições da Dissertação

- Estudo de um modelo Bayesiano para reconhecimento de objetos 3D a partir de informações 2D;
- Uso da distância de Kullback-Leibler estendida para modelar o grau de crença da causalidade entre dois eventos.

# Capítulo 2

#### **TRABALHOS RELACIONADOS**

Este capítulo tem como objetivo organizar e apresentar trabalhos relacionados às técnicas que serão utilizadas nesta dissertação.

A Seção 2.1 cita trabalhos que contribuem e utilizam métodos e modelos bayesianos semelhantes aos utilizados nesta dissertação; a Seção 2.2 refere-se a trabalhos que utilizam entropia e, em especial, de medidas de similaridade com foco nas aplicações da divergência de Kullback-Leibler; finalmente, na Seção 2.3 são destacados trabalhos que se embasam no Reconhecimento Cooperativo de Objetos (RCO) para recuperação de informações visuais em imagens.

# 2.1 Redes Bayesianas (RB)

Na literatura encontram-se diversas abordagens e aplicações para os modelos RB. Uma das mais comuns é a recuperação de conteúdos para casamento de padrões. De maneira geral, as técnicas que usam as RBs hoje aplicadas a diversas áreas da ciência, têm origem em teorias e formalismos matemáticos provenientes da estatística, como será visto na Seção 3.1.

Um dos trabalhos de maior importância é apresentado por (RIBEIRO-NETO;MUNTZ, 1996a), onde é proposto um modelo de rede de crença para a recuperação de informações de textos em grandes bases de dados. Conceitualmente, as chaves alvo das buscas formuladas pelo usuário, bem como cada documento pesquisado, são representados vetorialmente. Então, aplica-se à RB o cálculo de similaridade entre os diversos vetores, através do método de distância vetorial, apresentado na Seção 3.6.2. Os resultados obtidos mostraram que o modelo proposto pode ser visto como uma alternativa a outros métodos de inferência. Além disto, ainda existem vantagens conceituais, pois o método é intuitivo e abrangente a diversas aplicações. Também é fato que a abrangência da técnica deve-se à capacidade da mesma "moldar-se" a diversos métodos de "pontuação" (*ranking*) ou medidas de similaridade entre os objetos ou coleções sobre as quais fazem inferencia.

Pela flexibilidade e abrangência apresentadas em (RIBEIRO-NETO; MUNTZ, 1996a), o mesmo

método foi adaptado para que pudesse ser revisitado em outros artigos de mesma autoria, como em (RIBEIRO-NETO;MUNTZ, 1996b), onde é feita uma adaptação para o uso em bancos de dados SQL. A flexibilidade é tema explicitamente citado em (CRISTO et al., 2003), no qual é proposto um modelo para recuperação de informação em páginas da Web, demonstrando, também, que a capacidade de agregação de informações provenientes de diversas fontes de dados (contextos) pode trazer ganhos de desempenho na recuperação de diversos tipos de informação.

O modelo proposto em (RIBEIRO-NETO;MUNTZ, 1996a) foi novamente utilizado e adaptado em (COELHO et al., 2004), porém, para o uso em recuperação de imagem com base em seu conteúdo. Este trabalho faz uso dos métodos de busca de imagens na Web até então implementados. Até aquele momento, as técnicas de buscas, quase em seu todo, eram macro-associadas e baseavam-se no uso de palavras-chave relacionadas às imagens ("meta-dados") como informação *a priori* em seus modelos. Entretanto, nos conteúdos disponíveis na internet, as imagens são fracamente relacionadas com estes tipos de informação. Assim, atestou-se a inadequação desta abordagem e propôs-se uma solução que relacionava tanto a informação das imagens, quanto as informações do conteúdo HTML do documento no qual a imagem estava inserida. Esta estratégia de recuperação de informação reforçou os dados propostos em (CRISTO et al., 2003), aumentando a precisão dos resultados entre 50% e 60% em relação aos métodos regulares de extração de dados.

Uma variação deste modelo foi proposta no trabalho (RODRIGUES, 2003) e aplicada em (RO-DRIGUES et al., 2005), introduzindo uma técnica baseada em RB para casamento de padrões, na recuperação de informações visuais em imagens e na inicialização de rastreamento de objetos em cenas reais para Realidade Aumentada. A abordagem adotada inferia simultaneamente sobre três contextos principais de informações em uma imagem, sendo eles: a cor, a textura e a forma. Como resultado, os pesquisadores foram capazes de extrair e utilizar as características necessárias para fazer as correspondências entre os pontos 2D e 3D, embora ainda fossem necessárias algumas melhorias no desempenho da aplicação, conforme citado pelos autores, principalmente em relação ao alinhamento dos objetos 3D gerados por computador com a cena real. Contudo, os trabalhos de (RODRIGUES, 2003) e (RODRIGUES et al., 2005), posteriormente, inspiraram (SILVA, 2006) a aplicar uma abordagem semelhante em sua pesquisa.

Como se nota, a abordagem bayesiana é amplamente utilizada na recuperação de informações, assim como no artigo (OSWALD;LEV, 2001), onde é proposto o uso de múltiplos observadores, espacialmente espalhados para o reconhecimento de objetos, introduzindo o conceito de Reconhecimento Cooperativo de Objetos (RCO) (Seção 2.3). Este método executa a integração das hipóteses individuais geradas a partir de cada ponto de vista, sobre os quais são aplicados testes estatísticos e redes bayesianas para o *matching* (casamento) entre os objetos. Os resultados experimentais apre-

sentados indicam um aumento de robustez no processo, quando aplicada a RCO em relação aos métodos "clássicos" de reconhecimento individual.

A abordagem bayesiana com o uso de múltiplos observadores ou pontos de vista, também é vista no artigo (TANG et al., 2005), no qual foi proposta uma solução de convergência garantida para reconstrução 3D com uso de múltiplas câmeras. Contudo, diferentemente de (OSWALD;LEV, 2001), os observadores não tinham as hipóteses iniciais advindas de um modelo previamente inserido em uma base de dados. Então, para reconstruir um objeto alvo, são analisadas as correspondências entre as linhas 2D, informadas pelas múltiplas visões em um método de projeção ortogonal, chamado Projeção Reconstrutiva. Porém, os resultados apresentados, embora muito robustos e hábeis a lidar com ruídos, mostraram-se lentos para uso em sistemas de reconstrução em tempo real.

Variações nos métodos bayesianos podem ser utilizadas para melhorar o desempenho e diminuir a complexidade dos modelos, como visto no trabalho (JIN et al., 2004), onde foi gerada uma proposta para um sistema utilizando modelos bayesianos em conjunto com técnicas de segmentação e clusterização (agrupamento) com restrições semânticas. Esta abordagem possibilitou a classificação de um grande número de imagens automaticamente. Nos resultados, foram apresentados dados que justificaram o uso das RBs em relação a outras técnicas, como por exemplo, o *Support Vector Machine* (SVM) probabilístico. Outro ponto importante são os resultados da comparação entre os modelos bayesianos *naïve-bayes* (NB) e *semi-naïve-bayes* (SNB). Nesta comparação, foi mostrado que o modelo SNB tem melhor performance (desempenho), mas requer um poder computacional muito maior que a abordagem NB. Isto se deve ao maior número de análises feitas no cluster pela SNB, em comparação com NB.

As RBs também mostraram-se adequadas em (THORNTON et al., 2007), no qual é descrito um procedimento para o reconhecimento de padrões em casos onde existam deformações não-lineares, que ocorrem, por exemplo, na análise de íris ou escrita manual. O objetivo principal da pesquisa era extrair a transformação mais aproximada, "possivelmente não-linear", entre duas imagens para definir se elas participavam de uma mesma classe. Os resultados mostraram que existem vantagens em utilizar um modelo bayesiano para a geração de um *matching* mais preciso, tolerante a distorções e deformações. Por fim, mostrou-se que a técnica não despende alto custo computacional, o que permitiria, inclusive, seu uso em sistemas em tempo real.

A utilidade das RBs pode ser vista em outras sub-áreas do processamento de imagens, como a segmentação. Conforme descrito em (YONG-LI et al., 2007), para segmentar a cristalização em sedimentos urinários, células sanguíneas e outras informações em imagens médicas, um classificador bayesiano foi aplicado em conjunto com morfologia matemática, erosão e dilatação, resultando no aumento da precisão de todo o processo, mesmo envolvendo amostras de imagens médicas ruidosas. Ainda como agregador de desempenho, podemos citar (LI et al., 2003), onde, visando a extração das características dos objetos principais da cena (*foreground*) e também do plano de fundo (*background*), foi apresentado um modelo de classificador bayesiano capaz de separar objetos em imagens em movimento (como galhos de árvores balançando, superfícies de líquidos, etc).

Na presente proposta de mestrado, será utilizado o modelo bayesiano proposto por (RODRIGUES, 2003) e (RODRIGUES et al., 2005) para relacionar características de cor, forma e textura de um objeto. No entanto, o diferencial principal da técnica aqui apresentada é que as probabilidades a *priori* do modelo serão calculadas com base na divergência de Kullback-Leibler estendida. Assim, introduz-se o nome de Redes Bayesianas Entrópicas, ou, Entropia Bayesiana(RBE).

# 2.2 Entropia e Divergência de Kullback-Leibler

Desde os estudos de Rudolf Julius Emanuel Clausius no Séc. XIX, a entropia tem despertado o interesse de diversos ramos da Ciência. É fato que ela encontra cada vez mais aplicações fora da Mecânica Estatística. Uma das áreas onde seu emprego é realmente vasto é a Ciência da Computação, em especial, no processamento de imagens digitais e visão computacional.

As formulações de Boltzmann-Gibbs foram ao encontro das necessidades da medida quantitativa de informação proposta por Shannon em (SHANNON, 1948). Este trabalho fundamentou e introduziu o uso da entropia para sistemas de informação e Ciência da Computação, gerando uma nova área de pesquisa, cujo nome é Teoria da Informação. Em seus estudos, Shannon propôs um modelo para sistemas de comunicação utilizando a entropia como ferramenta fundamental para estimar a quantidade máxima de informação que poderia ser transmitida por um determinado canal.

O livro (GONZALEZ;WOODS, 2003) sintetiza os conceitos criados por Shannon, bem como, faz um paralelo entre as proposições contidas em (SHANNON, 1948) e seus possíveis usos para análise de imagens, como por exemplo, nos casos onde é necessária a compressão e segmentação.

Alguns dos conceitos demonstrados em (SHANNON, 1948) foram utilizados nas pesquisas realizadas por (KULLBACK;LEIBLER, 1951), com a finalidade de gerar uma forma de medir o que os próprios autores denominaram como "distância" ou "divergência" entre duas populações estatísticas. Este trabalho foi um marco, pois, de certo modo, popularizou ainda mais os conceitos da entropia, sendo que seu uso é amplamente encontrado e citado na literatura. Em nosso trabalho, este método, conhecido como Divergência de Kullback-Leibler é o discriminante fundamental entre distribuições de mesmo contexto, conforme será discutido na Seção 3.6.3.

A divergência de Kullback-Leibler é encontrada em vários trabalhos, como é o caso de (VAS-CONCELOS et al., 2004), onde esta abordagem foi utilizada em conjunto com outros métodos estatísticos para que fosse realizado o reconhecimento visual de imagens. As pesquisas dispostas nesta publicação indicaram que o melhor desempenho, dentro de todos os métodos descriminantes analisados, foi a do classificador baseado em Kullback-Leibler.

Embora as formulações de Boltzmann-Gibbs e, por conseqüência, Shannon, tenham obtido relativo sucesso em diversas aplicações, ainda existem alguns fenômenos que esta técnica é incapaz de modelar. Então, por volta da segunda metade da década de 1980, Constantino Tsallis, em seu trabalho (TSALLIS, 1988), introduziu um modelo estendido dos conceitos de entropia, fundamentando o que viria a ser conhecido como Entropia de Tsallis ou Entropia Não-Extensiva. De maneira geral, os modelos anteriores à proposição de Tsallis não eram aptos a modelar sistemas, como por exemplo, os tratados em (BECK, 2002) e (TARUYA;SAKAGAMI, 2002).

A Entropia de Tsallis também foi um dos focos das pesquisas realizadas em (TAVARES, 2003). Através de uma análise, do ponto de vista matemático, são expostas as propriedades e axiomas dos diversos métodos relativos à entropia, em suas diferentes variações e aplicações. Este trabalho teve grande importância em nossas pesquisas, pois, por seu conteúdo estritamente matemático, nos permitiu fundamentar e sedimentar os conceitos relativos à entropia e sua aplicação na Divergência de Kullback-Leibler.

Uma comparação entre os métodos extensivos de Shannon (SHANNON, 1948) e não-extensivos de Tsallis (TSALLIS, 1988) é feita em (MARTIN et al., 2004) para o registro de imagens. Para isso, foi proposta uma contraposição entre os métodos Kullback-Leibler em sua forma clássica e não-extensíva, sendo que esta última pode ser implementada conforme descrito nos trabalhos (ESQUEF, 2002) e (BARÃO, 2003). A pesquisa também utilizou as duas formas de divergência em conjunto com uma perturbação estocástica. Os resultados demonstraram uma redução significativa do número de iterações e ganhos na precisão do registro pela utilização da abordagem não-extensiva.

De fato, conforme apresentado nas pesquisas realizadas em (ESQUEF, 2002), existe um fator de não-extensividade em imagens. Intrinsecamente, esta relação torna a técnica adequada à aplicação em reconhecimento de padrões, o que justifica melhor desempenho nos processos baseados na abordagem não-extensiva. O autor analisou diversos métodos de reconhecimento de padrões e segmentação, de forma que os resultados que mostraram melhor desempenho foram aqueles onde empregou-se a entropia não-extensiva de Tsallis.

Neste trabalho, os conceitos relacionados à entropia, nas formas clássica e não-extensíva, estão estritamente ligados ao casamento de padrões. Nesta monografia, a divergência de Kullback-Leibler estendida será utilizada na execução do processo de casamento de características como a cor, a forma e a textura de imagens.

### 2.3 Reconhecimento Cooperativo de Objetos

No trabalho de (OSWALD;LEV, 2001), como já citado na Seção 2.1, foi introduzido o método de Reconhecimento Cooperativo de Objetos (RCO), através do uso de múltiplas visões ou pontos de vista de um mesmo objeto. O princípio conceitual deste trabalho baseia-se na idéia proposta em (OSWALD;LEV, 2001, apud (WEISS;RAY, 2001)) de que sempre há perdas de informações, como volume ou profundidade, quando se tenta utilizar projeções de um objeto 3D em uma imagem 2D. Assim, a inferência de diversas faces de um mesmo objeto pode apresentar resultados mais assertivos e acurados. No mesmo trabalho de (OSWALD;LEV, 2001) propõe-se que o maior número de observadores pode complementar e aumentar a eficiência em detectar evidências para identificação de um objeto. Então, nesta dissertação modelou-se uma rede de crença bayesiana, capaz de integrar e propagar estatísticamente as múltiplas hipóteses observadas, onde a inferência individual de cada ponto de observação influência o modelo como um todo. Contudo, é possível que informações incorretas obtidas nas análises iniciais da rede se propaguem por toda a inferência, terminando o processo de reconhecimento erroneamente. Isto ocorre porque em uma RB os valores obtidos no início de um processo de inferência influenciam o resultado como um todo. Assim, caso uma camada superior da RB avalie incorretamente uma imagem, o princípio da causalidade fará com que este erro seja propagado por toda a rede. Para minimizar esses efeitos, foi proposta uma equação que previne erros de interpretações, minimizando as disparidades causadas por evidências isoladas. Os resultados apresentados mostraram que o aumento no número de observadores causa não só o aumento de precisão, como também melhora no desempenho e velocidade da busca. Este trabalho relaciona-se ao trabalho de (OSWALD;LEV, 2001) pelo fato de ter introduzido o conceito do RCO, que é um dos pontos focais estudados, indicando o aumento do desempenho em relação ao reconhecimento baseado em um único ponto de vista.

A idéia de que a análise baseada em múltiplas visões pode ser benéfica para um sistema de reconhecimento de objetos é explorada também em (FARSHIDI et al., 2009). Nesse trabalho, foi apresentado um método para reconhecimento 3D através do uso de duas câmeras móveis, sendo estas dispostas na lateral e logo acima do objeto alvo fixo em uma mesa. A proposta para a extração das características dos objetos-alvo contou com uma versão modificada do cálculo do *Principal Component Analysis* (PCA) para criar uma medida vetorial de representação dos objetos. A modificação do PCA foi justificada pelo fato de que esta técnica avalia as informações globalmente o que a torna sensível a ruídos estruturados e oclusões. Este trabalho ainda utilizou-se de modelos previamente preparados para a execução do casamento de características. As inferências sobre os objetos foram extraídas por uma análise bayesiana a partir de movimentações controladas das

câmeras. Isto foi feito correlacionando-se as imagens observadas com os modelos, baseando-se no posicionamento atual e prévio das câmeras. Por outro lado, sendo a movimentação controlada das câmeras um fator chave neste estudo, a descrição da posição atual, bem como a escolha dos posicionamentos futuros é imprescindível. Assim, para que adequadamente fossem selecionados os movimentos das câmeras que melhor ajudassem a inferir conhecimento sobre o objeto observado utilizou-se a entropia para o cálculo da informação mútua entre a posição atual e a posição futura destes sensores. Como resultado desse trabalho mostrou-se a eficiência no método de recuperação de informações baseada em múltiplas visões. Esse método, que empregou modelos estatísticos, foi capaz de aumentar a robustez e a precisão do reconhecimento mesmo em imagens com ruído estruturado ou oclusão. Entretanto, essa abordagem difere da proposta apresentada por esta monografia. Enquanto em (FARSHIDI et al., 2009) é utilizado um aparato que permite avaliar as diversas vistas de um mesmo objeto pela variação dos parâmetros (movimentação) da câmera, aqui, são empregadas vistas previamente armazenadas para a inferência dos objetos.

Em (NIAN et al., 2007) novamente é exposta a idéia de que é difícil recuperar informações de geometria de objetos 3D apenas com o uso de um único ângulo de visão. Desta forma, foi proposta a recuperação de informações a partir de vídeos 2D contendo múltiplas vistas de um mesmo objeto 3D. Conforme exposto pelo trabalho, os vídeos gerados continham uma quantidade grande de informações, o que não ajudaria no processo de criação de um modelo adequado para o reconhecimento 3D e ainda aumentaria a demanda por esforço computacional. Desta forma, foi proposto um préprocessamento nas imagens, de forma que, imagens muito semelhantes fossem fundidas, gerando modelos que possibilitassem a redução do tamanho da base de dados. Uma vez finalizada a fase de criação de modelos a partir das imagens dos vídeos (cujos autores chamaram de clusterização) pode-se então executar o processamento das imagens. Foram extraídas características de borda das imagens supondo-se que estas e as áreas contidas dentro delas sejam as regiões de interesse para a recuperação de informações. Contudo, para que o descritor escolhido se tornasse invariável à rotação, aplicou-se uma transformação log-polar. Por fim, para o casamento utilizou-se um método probabilístico baseado em modelos bayesianos, onde a probabilidade a posteriori define se o objeto observado é um objeto da base de dados dada a comparação feita entre ele e diversas vistas disponíveis. A abordagem apresentada em (NIAN et al., 2007) mostrou-se computacionalmente eficiente no uso de múltiplas visões para o reconhecimento de objetos 3D. Contudo, a utilização exclusiva das características de forma é uma desvantagem em relação a modelos de reconhecimento capazes de utilizar simultaneamente características de cor, de forma e de textura, pois exclui informações importantes para o processo de casamento dos objetos.

# Capítulo 3

#### **CONCEITOS FUNDAMENTAIS**

### **3.1** Teoria de Bayes

As teorias desenvolvidas pelo matemático e reverendo Thomas Bayes no século XVIII propunham o corolário da probabilidade conjunta (ou total), com o qual é possível calcular a probabilidade da ocorrência de uma hipótese(H), dado o conhecimento sobre o acontecimento de um evento(E). Estas relações foram publicadas apenas após a morte de Bayes (GREEN et al., 2001) e ficaram conhecidas como a lei, regra ou teorema de Bayes, que, conforme apresentado em (JENSEN, 2001), são definidas pela seguinte Equação:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)},$$
(3.1)

onde se lê "probabilidade de uma hipótese H dado que um evento E ocorreu com certo grau de certeza".

Uma vez que a teoria bayesiana permite relacionar as causalidades entre dados probabilísticos, é possível complementá-la, para que se possa quantificar a intensidade dos relacionamentos entre diversos eventos. Estes eventos podem estar direta ou indiretamente relacionados, sendo que estas relações podem ser modeladas através de redes bayesianas, apresentadas na próxima seção.

# **3.2 Redes Bayesianas**

A teoria de Bayes é a base para o estudo das Redes Bayesianas (RB), que são usadas principalmente para o estudo de problemas que não podem ser resolvidos apenas com a relação de uma única evidência e uma única hipótese. A idéia da RB é criar representações causais, modelos próximos da realidade, intuitivamente ligados ao raciocínio lógico utilizado por seres humanos para resolver problemas complexos e com grande gama de hipóteses. Esta similaridade entre os modelos, a realidade e a forma humana de inferência é um fator vantajoso da abordagem, pois existe maior facilidade na extração e aproveitamento dos conhecimentos de um especialista.

Formalmente, uma RB é uma representação gráfica das intensidades das correlações entre eventos probabilísticos. Utilizando o conceito de grafos, uma rede é formada por nós, os quais representam eventos e arcos direcionados, que representam suas dependências. Este conjunto (nós-arcos) deve gerar um grafo direcionado, acíclico e finito (RODRIGUES, 2003). Por outro lado, a representação numérica das intensidades das dependências entre eventos, é definida por tabelas de probabilidade conjunta (TPC). Estas contêm os valores das probabilidades relacionadas às ocorrências de cada um dos eventos. Um exemplo de RB é apresentado pela Figura 3.1.



Figura 3.1: Grafo de representação de uma RB.

O grafo da Figura 3.1 representa a distribuição conjunta do conjunto de variáveis formado por  $\{E_0, E_1, E_2, E_3, E_4, E_5, E_6\}$ , com valores  $\{e_0, e_1, e_2, e_3, e_4, e_5, e_6\}$ . Para esta representação específica, temos  $E_0$  como nó pai, ou raiz, o qual tem como filhos  $E_1$  e  $E_2$ . A probabilidade de  $P(E_0)$ , do nó  $E_0$  é definida como *probabilidade a priori*, pois sua ocorrência é o evento conhecido deste modelo. Agora, seguindo a orientação do grafo, encontramos os eventos  $E_1$  e  $E_2$ , os quais são independentes entre si, porém, suas probabilidades dependem de  $E_0$ . Ainda, analisando  $E_1$  e  $E_2$ , o primeiro é pai da folha (nós sem filhos)  $E_3$  e do nó  $E_4$  e o segundo é pai de  $E_4$  e da folha  $E_5$ . Isto implica que  $P(E_4)$  depende tanto da  $P(E_1)$ , quanto de  $P(E_2)$ , sendo os valores dos outros nós dependentes apenas de seus respectivos pais. Sucessivamente dentro do encadeamento do grafo, encontraremos a folha  $E_6$ , filha e dependente do nó  $E_4$ . Assim, devido às relações entre os eventos, o cálculo da distribuição conjunta de probabilidades do modelo é definida pela Equação (3.2).

$$P(E_0, E_1, E_2, E_3, E_4, E_5, E_6) = P(E_0)P(E_2|E_0)P(E_1|E_0)P(E_3|E_1)$$

$$P(E_4|E_1, E_2)P(E_5|E_2)P(E_6|E_4)$$
(3.2)

Neste trabalho, as RB serão utilizadas para modelar as probabilidades conjuntas entre as evidências de cor, forma e textura de objetos observados.

# 3.3 Entropia

Por volta da metade do Século XIX, em plena revolução industrial, o físico e matemático alemão Rudolf Julius Emanuel Clausius (1822-1888), idealizou uma equação para quantificar as perdas inerentes à transformação de uma forma de energia em outra. O objetivo de sua análise era calcular a transformação da energia térmica em mecânica, relacionada à quantidade de trabalho que uma máquina a vapor era capaz de produzir. Este estudo estabeleceu a segunda lei da termodinâmica e o conceito de entropia.

Segundo (ÇENGEL, 1997), entropia é um conceito e não uma grandeza, cujo interesse dá-se pela observância da variação de seus valores ao longo do tempo, que é válida apenas para análise de um processo definido. Desta forma, embora a entropia tenha sido formalmente definida, trata-se de um conceito intuitivo, que deve ser avaliado conforme o contexto de sua aplicação. Por exemplo, na termodinâmica, como visto no início desta seção, a entropia é um processo utilizado para medir o trabalho de um sistema para transformação de energia; na física, é usada para medir a desordem de uma distribuição e na Teoria da Informação, ela é utilizada como medida quantitativa de informação transmitida em um canal, conforme será visto na Seção 3.4. Nesta dissertação estamos interessados no conceito de entropia como medida de informação.

Do ponto de vista termodinâmico, a variação da entropia pode ser definida pela Equação (3.3), que, com adequações ao contexto da aplicação, é largamente aplicada em diversos ramos do conhecimento, abrangendo desde a física até a sociologia.

$$\Delta S = \frac{\Delta Q}{T},\tag{3.3}$$

onde  $\triangle S$  representa a variação da entropia em relação a variação de calor  $\triangle Q$  em um sistema com temperatura constante T.

Como descrito em (ÇENGEL, 1997), a entropia pode ser entendida como uma medida da desordem ou aleatoriedade em nível molecular. Conceitualmente, a afirmação que relaciona a análise em termos de moléculas à entropia é creditada a Ludwig Boltzmann, que, em 1877, sugeriu a entropia em estados microscópicos e macroscópicos. Os eventos microscópicos tratam da análise de microestados em nível molecular do sistema. Por outro lado, a entropia dos estados macroscópicos é diretamente relacionada à energia interna, pressão e temperatura, ditos parâmetros termodinâmicos. Estas observações fundamentaram a Equação (3.4)

$$S \propto \log \Omega$$
 (3.4)

onde, segundo (TAVARES, 2003), Boltzmann observou que, em um sistema fechado, existe uma proporção direta entre a Entropia S e volume ocupado pelo estado macroscópico  $\Omega$ . Posteriormente, a Equação (3.4) passou a ser descrita como a Equação (3.5), onde surge a constante k, ou constante de Boltzmann. Embora a Equação (3.5) seja atribuída a Boltzmann, a mesma foi publicada apenas 1906, uma ano após sua morte.

$$S = k \ln W \tag{3.5}$$

Formalmente, na Equação (3.5),  $k \in \mathbb{R}^+$  e W é o número de estados microscópicos do sistema em relação ao macroestado analisado.

Baseado nos estudos de Boltzmann, Willard Gibbs criou uma forma mais generalizada da entropia para analisar a movimentação, segundo as leis da mecânica, de corpos com complexidade arbitrária, resultando na Equação (3.6).

$$S = -k\sum_{i=0}^{W} p_i \ln p_i \tag{3.6}$$

Assim, seguindo as formulações (3.4) e (3.5), na Equação (3.6), apresentada por Gibbs, W é o número total de microestados e  $p_i$  é a probabilidade do sistema estar no estado  $\omega_i$ .

Estes conceitos ficaram por quase um século e meio restritos à área termodinâmica, porém, em (SHANNON, 1948), Claude Shannon propôs um modelo para medição quantitativa e probabilística da informação, baseado em entropia. Apresentaremos a entropia do ponto de vista da teoria de Shannon na Seção 3.4, a seguir.

## 3.4 Teoria de informação e entropia

Assim como Rudolf Clausius queria melhorar o desempenho das máquinas à vapor, Claude E. Shannon, motivado pela criação de novos meios de comunicação, como o telefone e o telegráfo, percebeu que seria necessário entender as leis que regem estes sistemas e encontrar uma forma de maximizar a capacidade do envio de informação e reduzir os problemas ligados ao ruído e às características físicas específicas dos meios ou canais de comunicação. Então, em 1948, Shannon publicou um trabalho intitulado "A Mathematical Theory of Communication", no qual objetivava solucionar questões relativas à capacidade de transmissão de informações pelo uso de métodos estatísticos com base na análise das mensagens enviadas (SHANNON, 1948).

De forma geral, Shannon também criou um modelo para um sistema de comunicação e relacionou seus componentes de maneira matemática e probabilística, propondo uma forma quantitativa para medir o conteúdo de informação fornecido por uma mensagem. Esta medição é baseada na probabilidade da ocorrência da mensagem selecionada. Contudo, existe uma razão inversa entre a probabilidade da ocorrência da mensagem e a quantidade de informação nela contida. Assim, quanto maior a probabilidade da ocorrência de uma mensagem, menor a quantidade de informação própria que ela carrega, sendo o inverso também verdadeiro.

Intuitivamente, a entropia está relacionada ao grau de desordem em um sistema fechado. Assim, em uma fonte de informação, quanto maior a equiprobabilidade da ocorrência de mensagens, mais desordenada estará a transmissão e, conseqüentemente, maior será sua entropia. Verificando este fato, John von Neumann sugeriu a Shannon o uso da mesma função matemática da termodinâmica definida por Boltzman e Gibbs (Seção 3.3).

Sendo assim, Shannon em (SHANNON, 1948) propôs o seguinte formalismo para o cálculo da entropia (quantidade de informação de uma fonte): Seja  $P = \{p_1, p_2, ..., p_n\}$  uma distribuição conjunta de probabilidade de um sistema físico, onde  $p_i$  é a probabilidade do sistema estar no estado i e o número de estados possíveis é n. Logo, a entropia do sistema é definida como:

$$S = -k \sum_{i=0}^{n} p_i \ln p_i,$$
(3.7)

onde k é uma constante positiva.

A formulação de Shannon foi um marco para a Teoria da Informação e influenciou muitas outras teorias, principalmente na Ciência da Computação, onde é, desde então, usada em diversas aplicações, tais como: casamento de padrões, transmissão de dados em redes, processamento de sinais e visão computacional.

Nesta dissertação, introduz-se o uso da entropia para o cálculo da probabilidade de um objeto possuindo características de cor, forma e textura, ser encontrado em uma cena real, em uma aplicação de recuperação de informações visuais do ponto de vista bayesiano. Sendo assim, reitera-se o termo Entropia Bayesiana.

#### 3.4.1 Entropia de Shannon segundo uma abordagem física

É fato que o formalismo proposto por Shannon, embora utilizado principalmente na Teoria da Comunicação, está intrinsecamente ligado ao conceito físico e à abordagem tomada para a medição do nível de caos em um sistema. Então, podemos exemplificar a Equação (3.7) tomando como exemplo um sistema fechado, com 36 estados possíveis, os quais têm chances equiprováveis de possuir uma das moléculas de um gás qualquer que esteja encerrado no sistema. Desta maneira, entende-se que, pela aplicação dos conceitos de Boltzmann e Gibbs, existem 36 microestados possíveis para o sistema. A quantidade de micro-estados é representada pela variável W, o que confere W = 36. Isto, dado a equiprobabilidade dos estados, permite afirmar que o sistema encontra-se em máxima entropia. Assim, utilizando uma simples equação estatística, é possível verificar que as chances de uma molécula *i* estar em um estado qualquer é de  $\frac{1}{36}$ . Esta probabilidade é representada por P(i) e é definida por:

$$P(i) = \frac{n_i}{W},\tag{3.8}$$

onde  $n_i$  é a quantidade de moléculas em um estado e W é o número total de estados. Aplicando-se a idéia de equiprobabilidade em 36 microestados à Equação (3.7), obtem-se:

$$p_1 = p_2 = p_3 = \dots = p_{36} = \frac{1}{36}$$
 (3.9)

$$S = -[(p_1 \ln p_1) + (p_2 \ln p_2) + (p_3 \ln p_3) + \dots + (p_{36} \ln p_{36})]$$

$$S = -[(\frac{1}{36} \ln \frac{1}{36}) + (\frac{1}{36} \ln \frac{1}{36}) + (\frac{1}{36} \ln \frac{1}{36}) + \dots + (\frac{1}{36} \ln \frac{1}{36})]$$

$$S = -(36 \cdot \frac{1}{36} \ln \frac{1}{36})$$

$$S = -(\ln \frac{1}{36})$$

$$S = -\ln 36^{-1}$$

$$S = \ln 36 > 0,$$
(3.10)

o que mostra que, no estado de equiprobabilidade, a entropia máxima (maior quantidade de caos e desordem no sistema) pode ser calculada como o logaritmo do número total de estados possíveis (Equação (3.11)).

$$S = \ln W \tag{3.11}$$

Por outro lado, no caso de alta concentração de moléculas em um único estado e baixa concentração nos demais, diz-se que o sistema encontra-se em desequilíbrio. Assim, para os estados com baixa concentração tem-se:

$$p_1 = p_2 = p_3 = \ldots = p_W \simeq 0$$
 (3.12)

e para o estado de alta concentração têm-se  $p_j \simeq 1$ . Sendo assim, aplicando a Equação (3.8) à Equação (3.7) proposta por Shannon em (SHANNON, 1948) e assumindo-se que  $0 \ln 0 = 0$ , chegase à:

$$S = -[(0 \ln 0) + (0 \ln 0) + (0 \ln 0) + \dots + (1 \ln 1)]$$

$$S = -0 + 0 + 0 + \dots + 0$$

$$S = -0 + 0 + 0 + \dots + 0$$

$$S = 0$$
(3.13)

Nesse caso, todos os termos de S são iguais a zero, o que indica que, em caso de alta concetração, a entropia é próxima de zero. O que corresponde ao sistema com entropia mínima, ou, maior organização (menor caos).

### 3.5 Entropia Não-Extensiva

Embora a Entropia de Boltzmann/Gibbs/Shannon e seus conceitos tenham extrapolado com sucesso as áreas de interesse da termodinâmica, percebeu-se que as formulações e os modelos propostos por Boltzmann e Gibbs não explicavam adequadamente alguns fenômenos físicos. Estes problemas ocorrem em sistemas nos quais não é respeitado o Princípio da Aditividade (Equação 3.14), o qual prevê que a soma das entropias de n subsistemas independentes, seja igual a entropia total do sistema que os contém. Assim, supondo duas distribuições de probabilidade A e B; para os sistemas clássicos, o princípio da adtividade nos diz que:

$$S(A \oplus B) = S_A + S_B, \tag{3.14}$$

onde  $S_A$  representa a entropia do sistema A e  $S_B$  a entropia do sistema B.

Visando criar um método adequado a equacionar problemas envolvendo sistemas não-extensivos, na metade da década de 1980, Constantino Tsallis, em seu trabalho (TSALLIS, 1988), propôs uma maneira de generalizar a entropia clássica de Boltzman-Gibbs e Shannon, conforme citam (TAVARES, 2003) e (ESQUEF, 2002). As formulações sobre a mecânica estatística introduzidas por Tsallis ficaram conhecidas como "Estatística de Tsallis" ou "Entropia de Tsallis", sendo definidas por:

$$S_q = k \frac{1 - \sum_{i=1}^{W} p_i^q}{q - 1},\tag{3.15}$$

onde k é uma constante positiva,  $q \in \mathbb{R}$  é conhecido como *parâmetro entrópico* e W representa o número total de microestados do sistema. Contudo, a Equação (3.15) pode ser reduzida à formulação original da entropia apresentada por Shannon (Equação 3.6), no limite para  $q \to 1$ .

Em relação à aditividade e seu comportamento em sistemas não-extensivos, Tsallis introduziu o conceito "Pseudo-Aditividade", iniciando o uso do parâmetro q para o cálculo da entropia total do sistema, ficando esta representada por:

$$S(A \oplus B) = S_A + S_B + (1 - q) \cdot (S_A + S_B).$$
(3.16)

Note-se, que esta equação pode ser reduzida à Equação (3.14) no caso limite em que  $q \rightarrow 1$ .

Segundo (TSALLIS, 1988), os sistemas não-extensivos partilham as seguintes características:

- Longo Alcance Espacial
- Longo Alcance Temporal
- Comportamento Fractal nas fronteiras
- Pseudo aditividade

A Tabela 3.1 mostra um paralelo entre a teoria de Tsallis, apresentada nesta seção e as teorias de Boltzmann/Gibbs/Shannon, mostradas na Seção 3.4.

Extensivo (Shannon)	Não-Extensivo (Tsallis)
$S = -k \sum_{i=0}^{n} p_i \ln p_i$	$S_q = k \frac{1 - \sum_{i=1}^{\omega} p_i^q}{q - 1}$
Aditividade	Pseudo-Aditividade
$S(A \oplus B) = S_A + S_B$	$S(A \oplus B) = S_A + S_B + (1 - q) \cdot (S_A + S_B)$

Tabela 3.1: Tabela comparativa entre a entropia extensiva e não-extensiva

Neste trabalho utilizaremos a entropia não-extensíva para calcular o grau de crença de um objeto  $O_i$  ser igual a um objeto  $O_j$ , dentro da medida da divêrgencia de Kullback-Leibler, como será explicado na seção seguinte.

## 3.6 Medidas de Distância

Nesta seção reunimos algumas das abordagens mais utilizadas para medidas de distância. Ao utilizar o termo distância, se faz referência a medidas de similaridade ou divergência entre distribuições  $A = \{a_1, a_2, a_3, \ldots, a_n\}$  e  $B = \{b_1, b_2, b_3, \ldots, b_n\}$  de tamanho n ou de n objetos observados.

#### 3.6.1 Medidas Clássicas

É comum encontrarmos a utilização de medidas de distância baseadas na comparação entre duas distribuições de tamanho n. Uma das técnicas mais conhecidas para esta aplicação é a distância de Minkowiski (Equação (3.17)) que, ao utilizar-se de um parâmetro de ajuste m, pode tanto calcular a distância de Manhattan, quanto a convencional distância Euclidiana.

$$D_m = \left[\sum_{i=1}^n |A(i) - B(i)|^m\right]^{\frac{1}{m}},$$
(3.17)

onde, m é o parâmetro de ajuste; se m = 2, então a Equação (3.17) calculará a distância Eclidiana; se m = 1, a Equação (3.17) resultará na distância de Manhattan.

#### 3.6.2 Modelo Vetorial

As diferenças entre duas distribuições também podem ser calculadas através do modelo vetorial, que compreenda um espaço *n*-dimensional. Segundo (RODRIGUES, 2003), a similaridade entre os vetores pode ser calculada pelo cosseno do ângulo interno por eles delimitado. Assim, temos a Equação:

$$D(A|B) = \frac{\sum_{i=1}^{n} a_i \times b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} + \sqrt{\sum_{i=1}^{n} b_i^2}}$$
(3.18)

#### 3.6.3 Divergência de Kullback-Leibler

Em 1951, Solomon Kullback (1907-1994) e Richard A. Leibler (1914-2003) apresentaram um novo conceito para medir as divergências entre duas distribuições probabilísticas aleatórias denominada Entropia Relativa, representada pela Equação (3.19), a seguir:

$$D(A|B) = \sum_{i=1}^{n} a_i \overset{\cdot}{\log} \frac{a_i}{b_i},$$
(3.19)

onde  $A \in B$  representam as distribuições de probabilidades discretas, variando os estados de 1 até n.

Por derivar das formulações de Boltzmann e Gibbs, a entropia relativa deve apresentar sempre resultados positivos, sendo estes iguais a zero apenas se a = b, ou ainda, segundo (ESQUEF, 2002) e (CABELLA et al., 2008), se  $a \to 0$ , o que convenciona D(0|B) = 0. Por outro lado, utilizando uma convenção, se  $b \to 0$ , então D(A|0) = 0.

A Entropia Relativa é conhecida, entre outras denominações, como "distância de Kullback-Leibler", porém, o conceito de "distância", embora amplamente difundido, não é aplicável, pois, diferentemente do entendimento convencional de distância, a divergência entre duas distribuições é assimétrica, sendo D(A|B) diferente de D(B|A).

Sendo assim, foram propostas diversas formas para transformar a entropia relativa em uma métrica que contorne a questão da assimetria, conforme mostrado por exemplo em trabalhos como (JOHNSON;SINAMOVIC, 2001). Porém, utilizaremos a abordagem apresentada em (ESQUEF, 2002), definida pela Equação (3.20) na forma:

$$D_{KL}(A|B) = D(A|B) + D(B|A),$$
(3.20)

onde  $D_{KL} \in \mathbb{R}^+$ , D(A|B) é a divergência entre a distribuição A para a distribuição B e D(B|A)representa a divergência entre a distribuição B para a distribuição A.

#### 3.6.4 Divergência de Kullback-Leibler estendida

Analogamente às proposições apresentadas na Seção 3.5, a não-extensividade é aplicável à divergência de Kullback-Leibler. Desta forma, é necessário o emprego do parâmetro não-extensivo q, conforme apresentado em (ESQUEF, 2002), o que resulta na Equação (3.21)

$$D_{KL_q}(A|B) = \sum_i \frac{a_i^q}{1-q} \cdot (a_i^{1-q} - b_i^{1-q}).$$
(3.21)

Este método é particularmente útil para aplicações que envolvam casamento de padrões. Desta forma, seguindo a abordagem probabilística apresentada em (BARÃO, 2003), a entropia relativa permite quantificar o grau de certeza sobre a ocorrência de distribuição A, dada a verificação da ocorrência de uma outra distribuição B, arbitrária.

Sendo assim, nesta monografia será utilizada a divergência de Kullback-Leibler na forma estendida apresentada por (ESQUEF, 2002) para a distância entre duas distribuições independentes  $A = \{a_1, a_2, a_3, \dots, a_n\}$  e  $B = \{b_1, b_2, b_3, \dots, b_n\}$  da seguinte forma:

$$D_{KL_q}(A|B) = \sum_{i} \frac{a_i^q}{1-q} \cdot (a_i^{1-q} - b_i^{1-q})$$
(3.22)

e

$$D_{KL_e}(A|B) = D(A|B) + D(B|A)$$
(3.23)

# 3.7 Características para reconhecimento de objetos

Para que seja possível identificar um objeto, é necessário que primeiramente se conheçam as características que o definem. Sendo a alvo deste estudo o reconhecimento de objetos em imagens, deve-se então, enumerar descritores capazes de diferenciar e quantificar informações visuais em imagens bidimensionais. Em trabalhos como (RODRIGUES, 2003) e (SILVA, 2006) são apontadas a cor, a forma e a textura como três características básicas aplicáveis ao reconhecimento e processamento de imagens.

#### 3.7.1 Descritores de Cor

O uso da cor em sistemas de processamento de imagens dá-se pelo fato de que este é um descritor muito eficiente e que pode simplificar processos de extração e identificação de objetos em cena. Essa afirmação, apresentada em (GONZALEZ;WOODS, 2003), ainda é reforçada pela importância da cor no processo de identificação de objetos pelos seres humanos. Com base nessas idéias, é razoável assumir que a cor também seja um descritor de objetos importante para sistemas de recuperação de informações visuais com base no conteúdo. Entretanto, existem várias maneiras de se modelar a natureza da cor em sistemas de processamento de imagens. Dentre as modelagens clássicas encontram-se, por exemplo, os padrões RGB e CMY (GONZALEZ;WOODS, 2003). Porém, estes sistemas de representação possuem uma gama imensa de cores e tons, o que pode acarretar consumo de tempo computacional excessivo.

Por outro lado, pequenas variações entre as cores dos objetos comparados também podem tornar ineficiente o processo de reconhecimento. Com base nestes dois problemas, na literatura é indicado o uso de quantizações, que são reduções feitas no espaço de cores, conforme o apresentado em (RODRIGUES, 2003) e (SILVA, 2006). Desta forma para este trabalho optou-se pelo uso descritores que utilizam o padrão HSV-162.

A sigla HSV significa *Hue*, *Saturation* e *Value*, sendo o digrama deste espaço de cores visto na Figura 3.2.



Figura 3.2: Diagrama do espaço de cores HSV - Adaptado de (MATHWORKS, 2009).

Na Figura 3.2 pode-se observar as variações das três componentes que definem as cores do espaço HSV. A variação de matizes (*Hue*) está entre 0° e 360° no sentido anti-horário. Por outro lado, a quantidade de branco (*Saturation*) varia entre 0 e 1 (do centro para as bordas do diagrama), sendo que o valor 0 representa a máxima quantidade de branco (saturação) em uma cor. Complementando, o componente *Value* mede a quantidade de luz na cor, variando entre 0 e 1 (do fundo para o topo do diagrama), sendo que 0 representa total ausência de luz (preto).

O espaço HSV-162 utilizado em (RODRIGUES, 2003) é uma discretização do HSV em dezoito tons de cores (H = 18), três níveis de saturação (S = 3) e três níveis de iluminação(V = 3), perfazendo-se  $162 = 18 \times 3 \times 3$ , variações de cores possíveis.

Nos experimentos realizados neste trabalho o espaço de cores HSV-162 foi utilizado para descrever as informações de cores do objeto. A justificativa de sua utilização deu-se por sua grande capacidade de representação de informações de cores a partir de um histograma baseado em um conjunto de apenas 162 valores.

Assim, assume-se que o conjunto de cores de um objeto é descrito como:

$$C_j = c_1, c_2, \dots, c_n, (3.24)$$

onde  $c_i$  é a probabilidade de ser observada a cor *i* no objeto *j*, como restrição de  $\sum_{c_i} = 1.0$ , para n = 162 neste trabalho.
### 3.7.2 Descritores de Forma

A forma de um objeto pode ser capturada automaticamente por sistemas de reconhecimento visual de várias maneiras. Tradicionalmente, este tipo de informação é relacionado aos componentes de alta frequência da cena, sendo portanto descritos por detectores de mapas de borda como Canny (usado em (OSWALD;LEV, 2001)), Sobel, Prewitt ou Roberts (GONZALEZ;WOODS, 2003) (NIAN et al., 2007).

Semanticamente falando, as bordas de um objeto podem ser descritas por seu contorno mais externo. No entanto, para a análise de imagens naturais, este conceito pode não ser válido pois, dependendo da complexidade do objeto, a análise do contorno mais externo pode confundir-se com o próprio mapa de bordas, dificultando sua descrição. No caso de objetos específicos em uma cena, mesmo com *background* homogêneo, os contornos de regiões internas ainda são um forte descritor de informação. Por este motivo, para descrever as formas dos objetos, neste trabalho utilizou-se um descritor de bordas baseado na extração da orientação do gradiente a partir de imagens em tons de cinza.

Formalmente, segundo (GONZALEZ;WOODS, 2003), a análise feita pelo operador utilizado aqui tem como resultado o ângulo  $\alpha(x, y)$  da direção do gradiente  $\nabla f$  na posição (x, y), sendo representado pela Equação (3.25):

$$\alpha(x,y) = \tan^{-1}\left(\frac{G_y}{G_x}\right) \tag{3.25}$$

discretizando os valores de  $\alpha(x, y)$  em N direções difrentes.

Nos experimentos realizados neste trabalho, os valores  $\alpha(x, y)$  obtidos pelo cálculo dos gradientes foram discretizados em 18 valores possíveis (N = 18), de acordo com o sugerido em (RO-DRIGUES, 2003). Esta sugestão é importante, uma vez que a análise destes dados será feita a partir de um histograma de N entradas.

Assim, assume-se que a borda de um objeto é descrita como:

$$F_j = f_1, f_2, \dots, f_n, (3.26)$$

onde  $f_i$  é a probabilidade de ser observada a direção i no objeto j, como restrição de  $\sum_{f_i} = 1.0$ .

### 3.7.3 Descritores de Textura

Neste trabalho, a textura é usada como um dos três descritores chave para o reconhecimento de objetos. Para a extração de informações sobre as características de textura das imagens estudadas

foi implementado o método clássico da matriz de co-ocorrência, conforme descrito em (GONZA-LEZ;WOODS, 2003). As matrizes de co-ocorrência são calculadas apenas a partir de imagens em tons de cinza (neste trabalho discretizados em 18 valores possíveis). O princípio básico deste descritor é buscar padrões e informações de homogeneidade dentre as variações dos tons de cinza dos *pixels* que formam uma imagem.

Formalmente, segundo descrito por (GONZALEZ;WOODS, 2003), sendo A uma matriz de tamanho  $k \times k$  contendo elementos  $a_{ij}$  e sendo P um operador de posição; cada elemento  $a_{ij}$  deve conter o número de vezes em que um *pixel* com um tom de cinza  $z_i$  ocorre na posição indicada por P em relação ao ponto com nível de cinza  $z_j$ . Exemplificando, supondo uma imagem de tamanho  $4 \times 4$  com apenas 3 tons de cinza (0, 1, 2) caracterizada pela matriz com os seguintes valores:

$$A = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 2 & 1 & 2 & 2 \\ 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Utilizando o operador de posição P com a regra "um *pixel* à direita", obtém-se a seguinte matriz de co-ocorrência C:

$$C = \left[ \begin{array}{rrr} 1 & 2 & 0 \\ 3 & 1 & 2 \\ 0 & 2 & 1 \end{array} \right]$$

É importante observar que o operador P pode ser alterado, gerando um variedade de novos descritores para uma mesma textura. Isso permite gerar análises diversas sobre a relação espacial dos tons de cinza na imagem. Isto é possível pois, pode-se empregar um operador P, com a premissa de verificar, por exemplo, pares de pixels que apresentem a correlação "um *pixel* abaixo". Nesta pesquisa, cada imagem analisada gerou um conjunto de oito matrizes de co-ocorrência. Cada uma destas matrizes foi gerada pela aplicação de um operador P diferente, usando as seguintes regras:

- 1. um pixel acima.
- 2. um *pixel* à direita e um *pixel* acima.
- 3. um *pixel* à direita.
- 4. um *pixel* à direita e um *pixel* abaixo.

- 5. um pixel abaixo.
- 6. um *pixel* abaixo e um *pixel* à esquerda.
- 7. um pixel à esquerda.
- 8. um *pixel* à esquerda e um *pixel* acima.

Sendo gerada a matriz de co-ocorrência C, é possivel caracterizar o seu conteúdo segundo abordagens estatísticas. Para isso, podem ser empregados diversos métodos, como os apresentados em (GONZALEZ;WOODS, 2003). Neste trabalho foram empregados os seguintes descritores:

- 1. Descritor de Energia ou Uniformidade (Equação (3.27));
- 2. Máxima probabilidade (Equação (3.28));
- 3. Descritor de Homogeneidade (Equação (3.29)).

$$\sum_{i=1}^{K} \sum_{j=1}^{K} c_{ij}^2 \tag{3.27}$$

$$\max_{i,j}(c_{ij}) \tag{3.28}$$

$$\sum_{i=1}^{K} \sum_{j=1}^{K} \frac{C_{ij}}{1+|i-j||}$$
(3.29)

O descritor de energia calculado através da Equação (3.27), tem valores que variam de 0 até 1, onde o valor 1 representa uma imagem sem nenhuma alteração na textura (GONZALEZ;WOODS, 2003). Por outro lado, o descritor de máxima probabilidade, apresentado pela Equação 3.28 mede a resposta mais forte dos valores contidos pela matriz C (GONZALEZ;WOODS, 2003). Da mesma forma, o descritor de homogeneidade representado pela Equação 3.29 mede a proximidade espacial entre os elementos  $c_{ij}$  da matriz de co-ocorrência C, assumindo valores que podem variar entre 0 e 1 (GONZALEZ;WOODS, 2003).

Complementando, neste trabalho os descritores de textura extraídos serão dispostos de maneira que gerem um conjunto de valores que serão utilizados como um histograma de textura na seguinte

forma:

$$T_j = t_1, t_2, \dots, t_n, (3.30)$$

onde  $t_i$  é um conjunto de valores formados pelos resultados das equações (3.27), (3.28) e (3.29) de cada uma das oito matrizes de co-ocorrência, representando a probabilidade de ser obsevado padrão de textura *i* no objeto *j*.

## 3.8 Aplicações do modelo proposto

Nas áreas de Visão Computacional (VC) e de processamento de imagens existe uma vasta gama de aplicações dependentes de recuperação de informações visuais e reconhecimento de objetos naturais ou artificiais em cena. Assim sendo, os resultados obtidos por este trabalho podem ser integrados às inúmeras aplicações como, por exemplo, realidade aumentada e análise de cenas e vídeos. Por estes motivos nas seções seguintes são relacionadas algumas possíveis aplicações para o modelo proposto neste tabalho. Esta relação é importante à medida que por meio dela se podem vislumbrar possíveis estensões ao tema e à pesquisa aqui sugeridos.

#### 3.8.1 Realidade Aumentada

Uma das possíveis aplicações do modelo aqui proposto é em Realidade Aumentada (RA). O objetivo de aplicações de RA é adicionar informação a objetos ou localidades do mundo real, incorporando contextos e descrições a objetos reais (EDUCASE, 2005), possibilitando uma maior e mais completa experiência de um usuário em relação ao assunto, localidade, ou objeto observado e simulado por um sistema.

De maneira sucinta, sistemas intitulados de RA devem ser capazes de mesclar o mundo real e o mundo virtual, permitindo a interação em tempo real do usuário com os objetos sintéticos gerados pela simulação. A capacidade de inserir corretamente um objeto sintético em uma cena natural é conhecida como Registro 3D (AZUMA, 1997).

Dentre as características básicas que devem atender às aplicações de RA, o processo conhecido como registro é o mais importante. Executá-lo de forma eficiente depende do reconhecimento dos marcadores (Figura 3.3) que são objetos previamente inseridos na cena e servem de referência para a aplicação RA inserir, de maneira adequada, novas informações nas imagens naturais já existentes (Figura 3.4). Com base nestas idéias e sendo a RA dependente do reconhecimento de objetos em cena para o seu fucionamento, abre-se uma perspectiva de uso das pesquisas aqui realizadas no processo de registro. Isto porque as técnicas propostas aqui podem ser facilmente empregadas na

fase de reconhecimento dos marcadores. Assim, é possível que se utilizem tanto objetos artificiais, quanto objetos naturais. Isto pode ser feito extraindo-se previamente as características de cor, forma e textura dos objetos usados como marcadores, utilizando estas informações no reconhecimento de objetos em cena através do uso de múltiplos observadores.



Figura 3.3: Marcador utilizado em aplicações de RA - Adaptado de (BILLINGHURST et al., 2002).



Figura 3.4: Mescla entre uma imagem real e um objeto sintético através da RA utilizando a biblioteca ARToolKit (KATO, 2009).

### 3.8.2 Análise de cenas e vídeos

A recuperação de informações visuais baseada em conteúdo pode ser aplicada não só a imagens estáticas, mas também a vídeos contendo cenas reais ou sintéticas. Aplicações, como a feita em (NIAN et al., 2007) se utilizam do processamento de vídeos para aquisição do modelo e reconhecimento de objetos. Baseado nos conceitos apresentados, o método proposto nesta dissertação pode facilmente ser adaptado ao reconhecimento de objetos em cenas sejam eles 3D ou 2D, naturais ou sintéticos. Isto porque, admitindo-se que um vídeo é uma seqüência de imagens, a busca por informações baseada em modelos é possível através da repetição contínua dos processos tomados para a análise de imagens estáticas. Desta forma, o uso de múltiplos observadores poderia ser aplicado em três contextos: busca por cenas, busca por objetos em cenas e busca por múltiplos objetos em cena. A busca por cenas é de aplicação direta; neste contexto, pode-se utilizar um ou mais *"frames*"(quadros) do vídeo como objeto alvo da busca. Por outro lado, a busca por um único objeto em cena pode ser feita informando-se o objeto alvo para múltiplos observadores; estes, por sua vez analisariam a mesma imagem retornando suas inferências sobre a existência do objeto. Neste ponto, seria possível, com alterações adequadas no sistema, avaliar as inferências dos observadores em relação aos objetos e retornar, não só a inferência conjunta, mas também o posicionamento do objeto na cena, assim como o que poderia ser feito na Realidade Aumentada. Por ultimo, a busca por múltiplos objetos em cena seria possível levando-se em consideração os mesmos pontos levantados para a busca por um único objeto, só que distribuindo modelos de objetos distintos para cada observador. A busca por múltiplos objetos é interessante também, à medida em que é possível a criação de um sistema de recuperação de informações contextualizadas baseadas não só em uma, mas em multiplas características de um mesmo objeto, de forma que o modelo ao qual se confronta seja uma parte específica e não todo o objeto. Desta forma, seria possível projetar sistemas com finalidades semelhantes às apresentadas em (KIM;KWEON, 2008a) e (KIM;KWEON, 2008b), porém implementados de maneira completamente distintas.

### 3.8.3 Scanner 3D

O uso de múltiplas visões aplicadas a esquemas de aquisição de informações de geometria e de volume a partir de objetos tridimensionais é direta. Um cenário seria o uso de várias câmeras em perspectivas fixas ou poucas câmeras em perspectivas móveis focalizando um objeto 3D fixo (similarmente ao que foi feito em (FARSHIDI et al., 2009)). Em seguida, as diversas imagens capturadas deveriam passar por um processo de registro ou correspondência entre os pontos, possívelmente utilizando abordagens como a apresentada em (TANG et al., 2005).

Um exemplo da utilidade da implementação de sistemas de extração de características tridimensionais de objetos é o trabalho (WASSERMAN, 1999). Na pesquisa, os objetos cujas características são extraídas, são obras de arte. Isso é feito visando estudos detalhados das peças em questão. Ilustrativamente, o resultado desta aquisição de informações pode ser visto na Figura 3.5, onde, à esquerda, temos a imagem da obra original e, na parte superior direita, o resultado da extração da geometria, do volume, da cor e da textura da obra de arte.



Figura 3.5: Resultado da aquisição de informações de geometria e volume a partir de objetos reais – Adaptado de (WASSERMAN, 1999).

# **Capítulo 4**

## PROPOSTA

## 4.1 Modelo Bayesiano para Recuperação de Informação

Como citado anteriormente na Seção 2.1, em (RIBEIRO-NETO;MUNTZ, 1996a) foi proposto um modelo para recuperação de informações textuais baseado em Redes de Crença Bayesianas. Este modelo foi posteriormente adaptado em (RODRIGUES, 2003), que introduziu o seu uso para recuperação de informações visuais em imagens digitais, baseando-se na avaliação de características de cor, forma e textura.

Neste trabalho será utilizado o modelo bayesiano conforme o apresentado em (RODRIGUES, 2003) e (SILVA, 2006), o qual pode ser visto de forma esquematizada na Figura 4.1. Este modelo consiste em uma RB de duas camadas. Destas, a camada superior (K) representa as probabilidades *a priori* e a inferior (O) as probabilidades *a posteriori*. Todos os elementos da camada K ligam-se a todos elementos da camada O através de arcos direcionados dos nós  $k_n$  (integrantes da camada K) para os nós  $O_m$  (integrantes da camada O). Os arcos indicam a probabilidade da ocorrência de um Objeto  $O_j$ , dado que foi observado o termo  $k_i$ . O modelo de (RODRIGUES, 2003) e (SILVA, 2006) propõe que os termos  $k_i$  se refiram à características como cor, forma e textura.



Figura 4.1: Modelo de RB Genérico - Adaptado de (RODRIGUES, 2003).

Dado que possuímos um modelo bayesiano definido, a probabilidade da ocorrência de um objeto

 $O_i$ , dado que observemos um objeto modelo K é formalizada pela Equação (4.1)

$$P(O_j|k_1, k_2, \dots, k_n) = P(O_j|K),$$
(4.1)

onde  $O_j$  é um objeto que queremos encontrar dada a observação das características  $k_1, k_2, ..., k_n$  de um objeto modelo representado por K.

Como visto nos trabalhos de (RODRIGUES, 2003) ,(RODRIGUES et al., 2004), (RODRIGUES et al., 2005) e (SILVA, 2006), a Equação (4.1) pode ser modelada por uma cláusula OU. Desta forma, o modelo representado pela Figura 4.1 e formalizado pela Equação (4.2), pode ser estendido para que outros contextos possam ser validados. Assim, conforme simplificações apresentadas em (RIBEIRO-NETO;MUNTZ, 1996a) e (COELHO et al., 2004), é possível adaptarmos o modelo genérico para que se converta em novos classificadores, capazes de calcular as probabilidades de  $O_j$  dadas as características de Cor ( $K_c$ ), Forma ( $K_f$ ) e Textura ( $K_t$ ), como demonstrado em (RO-DRIGUES, 2003) e (SILVA, 2006), resultando na Equação (4.2)

$$P(O_j|K) = 1 - \left[ (1 - P(O_j|K_c)) \times (1 - P(O_j|K_f)) \times (1 - P(O_j|K_t)) \right]$$
(4.2)

De forma geral, a Equação (4.2) representa a união entre diversas evidências, cada uma representada por um termo  $(1 - P(O_j|k_i))$ . Este conjunto de evidências é passível de inserções ou remoções de novos elementos, sem que isso acarrete mudanças na estrutura do modelo. Exemplificando: supondo que, para a análise de um objeto seja necessária a remoção<sup>1</sup> das características de forma (representadas pelo termo  $(1 - P(O_j|K_f))$ ). Logo, a Equação (4.2) será reescrita na forma da Equação (4.3).

$$P(O_j|K) = 1 - \left[ (1 - P(O_j|K_c)) \times (1 - P(O_j|K_t)) \right]$$
(4.3)

Similarmente, definindo a probabilidade de  $K_f$  como nula ( $P(O_j|K_f) = 0$ ), a representação desta evidência dentro da Equação genérica (4.2) fica na forma:

$$P(O_{i}|K) = 1 - \left[ (1 - P(O_{i}|K_{c})) \times (1 - 0) \times (1 - P(O_{i}|K_{t})) \right], \tag{4.4}$$

o que não altera a estrutura da modelagem e permite que os outros termos continuem a ser avaliados.

O sistema genérico formalizado pela Equação (4.2) pode ser modelado como a união de três modelos equivalentes ao apresentado na Figura 4.1. Desta forma, temos uma RB para as caracte-

<sup>&</sup>lt;sup>1</sup>Demonstraremos apenas a remoção de três evidências, pois, de maneira intuitiva, podemos verificar que a adição de novos termos nos leva à mesma consideração.

rísticas de cor, representada pela Figura 4.2, uma para características de forma, representada pela Figura 4.3 e uma para as características de textura, representada pela Figura 4.4. A RB proposta em (RODRIGUES, 2003), que engloba os classificadores de cor, forma e textura é esquematizada na Figura 4.5 e será utilizada para modelar o *matching* entre duas imagens (visão de um observador, conforme será discutido na Seção 4.2) a partir de um ponto de vista de um objeto.



Figura 4.2: Modelo de classificador para análise de características de cor - adaptado de (RODRIGUES, 2003).



Figura 4.3: Modelo de classificador para análise de características de forma – adaptado de (RODRIGUES, 2003).

# 4.2 Estrutura Proposta

Neste trabalho é proposto o emprego do modelo representado pela Equação (4.2) em um sistema de reconhecimento cooperativo de objetos. Na Figura 4.6, é apresentada a arquitetura para a qual o modelo aqui desenvolvido se aplica.

A Figura 4.6 mostra que arquitetura proposta se baseia na aquisição de imagens de um objeto A, a partir de n pontos de vista. Estas imagens são processadas pelos *Observadores*  $Z_i$  e em seguida são



Figura 4.4: Modelo de classificador para análise de características de textura - adaptado de (RODRIGUES, 2003).



Figura 4.5: Modelo de classificador para análise conjunta de características de cor, forma e textura – adaptado de (RODRIGUES, 2003).



Figura 4.6: Proposta de arquitetura para um sistema de recuperação de informações baseado em múltiplas visões.

enviadas para um ponto central de processamento de informações, o qual é chamado de *Integrador de Evidências* (IE)

O IE tem como atribuições principais a integração das informações enviadas pelos observadores, integrando as evidências no processo de *Fusão* (OSWALD;LEV, 2001). Por sua vez, os observadores, têm a incumbência de estimar e retornar a probabilidade das visões capturadas pertencerem

a um dos objetos contidos em sua base de dados (Seção 5.2).

O processo completo do reconhecimento funciona nas seguintes etapas:

- 1. Captura de visões do objeto A;
- 2. Cálculo das evidências individuais de cada Observador;
- 3. Envio das evidências dos observadores ao IE;
- 4. Fusão.

No item 1 é executado um processo que consiste em posicionar n câmeras de forma adequada ao redor de um objeto A, de forma que seja possível capturar imagens em diversos pontos de vista, gerando o que é chamado de *visões do objeto*;

No item 2 é feito o processamento das visões capturadas em 1 para os n observadores. Nesta etapa também, cada um dos observadores executa o processo de avaliação das visões, através do uso do classificador bayesiano representado pela Equação (4.2) e ilustrado pela Figura 4.5. Os observadores utilizam a divergência de Kullback-Leibler estendida, a partir dos procedimentos descritos na Seção 4.3, para efetuar o *matching* das características de cor, forma e textura do objeto A com as características dos modelos  $O_j$  contidos na base de dados;

O item 3 consiste no envio das evidências calculadas pelos observadores ao IE. Ao fim de cada análise, os observadores devem enviar as probabilidades, em termos percentuais, de volta ao IE.

Finalmente, no item 4 é feita a *Fusão*, que assim como apresentado no trabalho (OSWALD;LEV, 2001), consiste na integração das evidências aferidas pelos observadores. Esta integração é executada aplicando-se a média aritmética às probabilidades definidas pelos observadores. Os resultados da probabilidade de A (objeto observado) ser o objeto modelo  $O_j$ , obtidos pelos n observadores (z), são "fundidos", gerando-se um único percentual, baseado em todas as evidências observadas, da seguinte maneira:

$$P(O_j|z_1, z_2, \dots, z_n) = \frac{1}{n} \sum_{i=1}^n P_{z_i},$$
(4.5)

onde os termos  $z_1, z_2, \ldots, z_n$  representam os resultados das inferências dos observadores em relação às visões do objeto analisado de acordo com os procedimentos da Seção 4.1.

Um ponto importante desta arquitetura é que, através de pequenas modificações, conforme visto na Figura 4.7, é possível aplicá-la em sistemas de reconhecimento cooperativo de objetos, onde os pontos de processamento (observadores) podem estar geograficamente espalhados. Isto é factível apenas fazendo-se com que o IE se torne também um servidor central que acumule as funções de executar a fusão e distribuir as imagens do objeto observado entre os observadores. Por outro lado, os observadores não precisariam ser alterados e continuariam a executar o levantamento individual de evidências, enviando-as ao IE.



Figura 4.7: Proposta de arquitetura para um sistema de recuperação de informações baseado em múltiplas visões.

# 4.3 Medida de Similaridade Proposta

Neste trabalho a probabilidade de um Objeto  $O_i$  ocorrer dado que um Objeto  $A_j$  tenha sido observado é dada pela divergência de Kullback-Leibler Estendida (DKLE), vista no Capítulo 3 na Seção 3.6.3, de acordo com as Equações (3.22) e (3.23).

A Equação 3.23 será modelada para diversos tipos de informação. É através dela que serão quantificadas as divergências entre as características dos objetos alvo  $A_j$  e os padrões encontrados  $O_j$ .

Considerando informações de cor, o Objeto  $O_i$  será representado pelo histograma 162, o qual será extraído das imagens através do método proposto no Capitulo 3, Seção 3.7.1.

Na análise de informações de forma, utilizaremos o histograma do mapa de bordas, que será calculado conforme descrito no Capitulo 3, Seção 3.7.2.

Para modelar informações de textura, consideraremos as características de co-ocorrência, conforme definido no Capitulo 3, Seção 3.7.3.

Assim, cada evidência observada em um objeto é modelada de acordo com a Equação (3.23). Finalmente, cada evidência (cor, forma e textura) calculada com a Equação (3.23) é integrada ao modelo Bayesiano proposto pela Equação (4.2). Cada observador, a partir de um ponto de vista, calcula a Equação (4.2) e envia o resultado ao Integrador de Evidências que usa a Equação(4.5) para avaliar conjuntamente as inferências individuais dos observadores.

Os experimentos apresentados no Capítulo 5 mostram o desempenho do modelo proposto.

# Capítulo 5

## **EXPERIMENTOS**

## 5.1 Metodologia de testes

Embora o modelo aqui proposto seja principalmente focado no reconhecimento 3D a partir de várias vistas, seu princípio pode ser aplicado diretamente a sistemas de recuperação de imagens baseado em conteúdo (RIBC). Isto se deve, ao fato de que, tanto aplicações de reconhecimento 3D, quanto RIBC, dependem diretamente da recuperação de informações visuais para o seu funcionamento. Isto fica evidente quando se consideram as várias vistas de um objeto como informações relevantes para sua recuperação através de uma consulta em um sistema RIBC. Isto também é uma premissa em sistemas de reconhecimento 3D, pois, neste caso, os mesmos precisam identificar corretamente os objetos com base nas imagens que recebem a partir de diferentes ângulos. Sendo assim, analogamente, pode-se utilizar metodologias de testes similares às que são aplicadas em RIBC para avaliar sistemas de reconhecimento 3D. Na área de RIBC é comum encontrar avaliações baseadas na razão entre a quantidade de imagens que devem ser retornadas e a precisão com que isto ocorre, o que é conhecido como o Método da Precisão x Revocação (PR) (ver a Seção 5.4).

Neste trabalho, esta técnica foi aplicada para medir a quantidade de vistas reconhecidas de um objeto (Revocação) em relação à precisão com que o sistema fez este reconhecimento (Precisão). Estas medidas foram executadas avaliando-se e experimentando-se diversas hipóteses através da variação de parâmetros, tais como o parâmetro entrópico q, a quantidade de Observadores, o tamanho da base de dados, entre outros.

As hipóteses consideradas neste conjunto de experimentos são as seguintes:

- (a) análise do melhor parâmetro entrópico q superextensívo;
- (b) análise do desempenho, individual e conjunto, das características de cor, forma e textura;
- (c) análise comparativa da precisão utilizando múltiplas visões.

A hipótese (a) trata da execução de uma análise onde o melhor parâmetro entrópico q, deve ser apontado. Esta hipótese, considera apenas a entropia não-extensiva para a execução do matching das características das imagens, uma vez que pode haver grande quantidade de interações de longo alcance entre as diversas características e visões de um mesmo objeto (interações de cor, forma e textura). Assim sendo, assume-se, por hipótese, que um melhor reconhecimento pode ser alcançado com o valor de q < 1, que representa, segundo Tsallis (TSALLIS, 1988), um sistema superextensívo;

Como hipótese (b), considera-se avaliar o desempenho das características de cor, forma e textura, bem como a possibilidade de que alguma delas, isoladamente, seja mais relevante para a recuperação de um objeto;

Finalmente, a hipótese (c) avalia se o uso da recuperação de objetos baseada em múltiplas visões é mais preciso do que a aplicação de um único referencial no processo de recuperação de informações visuais de um objeto.

Este trabalho propõe a avaliação dessas hipóteses através do modelo bayesiano proposto, teoria da informação não-extensiva e modelo de múltiplas visões utilizado.

# 5.2 Base de dados utilizada

Neste trabalho os testes foram conduzidos utilizando-se uma base de imagens da Universidade de Columbia (RODRIGUES, 2003, apud (COLUMBIA UNIVERSITY, 2001)). Esta base possui 100 objetos registrados em 72 ângulos distintos, uniforme e seqüencialmente espaçados de 5 em 5 graus, sendo o ângulo inicial igual a 0 graus e o último igual a 355, totalizando  $100 \times 72 = 7200$  imagens. Uma amostra desta base é encontrada na Figura 5.1. Nela, apresentamos 6 das 100 classes (1 por linha) e 5 das 72 vistas (1 por coluna). Na base de dados aqui amostrada, existem alguns objetos simétricos que facilitam a recuperação e geram um alto valor de precisão e revocação, como é o caso do objeto apresentado na sexta linha da Figura 5.1. Contudo, outros objetos, como o registrado na linha 2 da Figura 5.1, são assimétricos e, assim, sua recuperação depende do ângulo sob o qual é analisado, de forma que sua precisão pode ser seriamente afetada. Os gráficos das Figuras 5.2 e 5.3 mostram com que precisão (eixo y) são recuperadas as 72 vistas (revocação, eixo x), das classes 14 e 35, respectivamente, entre as 7200 imagens da base.

Pode-se notar que a precisão alcançada para a imagem da segunda linha (gato, classe 14) da Figura 5.1, cujo gráfico da precisão e revocação é representado pela Figura 5.2, sofre um decréscimo repentino. Isto indica que, neste momento, as vistas analisadas continham informações muito divergentes das da vista usada como modelo, o que se deve à grande assimetria encontrada no objeto pois, conforme o ângulo observado, os padrões de cor, forma e textura são muito diferentes. Por outro lado, a classe à qual pertence a imagem da sexta linha (prato, classe 35) da Figura 5.1, tem taxas de precisão e revocação iguais a 100%, conforme se pode verificar na Figura 5.3. Esta taxa elevada de precisão e revocação indica que, pelo fato de o objeto ser totalmente simétrico, não importa qual ângulo da imagem seja analisado, a taxa de precisão sempre será alta. Para ambos os experimentos mostrados nas Figuras 5.2 e 5.3, usou-se o parâmetro entrópico q = 0.1 na Equação (3.23), por ter sido aquele que possibilitou os melhores desempenhos segundo experimentos que serão mostrados nas próximas seções.



Figura 5.1: Amostra da base de dados Columbia. Nesta amostra, apresenta-se 6 classes distintas na horizontal, sendo cada uma registrada em 5 ângulos diferentes – Adaptado de (RODRIGUES, 2003, apud (COLUMBIA UNIVERSITY, 2001)).

A principal vantagem de se utilizar a base de dados da Columbia para testar o modelo proposto, é que se trata de uma base "pseudo sintética", pois, apesar de os objetos da base serem reais, capturados com uma camera CCD real, as imagens contam com ângulo de visão, iluminação e foco controlado, pois foram produzidas em laboratório, o que permite um maior controle dos experimentos executados. Também, considerando-se o *background* homogêneo e a qualidade das imagens, não é necessário empregar técnicas de extração de *background*, pois este não é o foco do trabalho.



Figura 5.2: Gráfico da precisão e revocação do objeto assimétrico (referente à segunda linha da Figura 5.1).



Figura 5.3: Gráfico da precisão e revocação do objeto simétrico (referente à sexta linha da Figura 5.1).

# 5.3 Hardware Utilizado

Todos os experimentos foram executados utilizando-se os seguintes equipamentos: Notebook HP Pavillon dv6150br, com processador AMD Turion 64X2 de 1.6GHz, com 1GB de RAM; Desktop Dell Optiplex 330, com processador Intel Core2 Duo E8900 de 3GHz, com 3GB de RAM. Ambos os modelos usados para os testes rodavam sob plataforma Windows XP 32bits.

## 5.4 Precisão x Revocação (PR)

Com o objetivo de quantitativamente avaliar os resultados gerados pelo sistema, foi empregada a técnica da Precisão e Revocação (PR) utilizada em (RODRIGUES, 2003). Esta medida baseiase na relação entre a quantidade de objetos existentes em uma base A e a quantidade de objetos recuperados pelo sistema B, resultando nas seguintes relações

$$\operatorname{precisão} = \frac{|A \bigcap B|}{|B|} = \frac{quantidade \ de \ imagens \ relevantes \ recuperadas}{numero \ de \ imagens \ recuperadas}$$
(5.1)

$$\operatorname{revoca}_{\operatorname{ca}}_{\operatorname{ca}} = \frac{|A \bigcap B|}{|A|} = \frac{quantidade \ de \ imagens \ relevantes \ recuperadas}{numero \ de \ imagens \ relevantes}$$
(5.2)

Ao utilizar o termo "imagens relevantes", é importante que se tenha ciência de que esta tarefa depende do contexto da imagem, bem como da subjetividade humana em inferir qual é o resultado mais apropriado (RODRIGUES, 2003). Neste trabalho, por exemplo, a relevância é indicada pela probabilidade da imagem recuperada fazer parte do mesmo grupo das 72 vistas de um objeto que compõem uma classe.

Uma das maneiras mais comuns de se representar dados de PR é na forma gráfica. Nestas representações, o eixo das abscissas representa os valores da precisão e o das ordenadas os valores da revocação, ambos variando ente 0 e 1, sendo 0 o pior resultado, e 1 o melhor.

Alguns exemplos de gráficos PR podem ser observados na Figura 5.4. A figura citada é composta por um conjunto de quatro gráficos, cada um representando a PR de uma das quatro classes de objetos contidas na Figura 5.5. À medida em que os gráficos são observados, nota-se que cada classe de objetos possui valores muito distintos de desempenho, sendo que alguns, como o objeto da classe 9, possuem valores de PR de aproximadamente 50%. Por outro lado, como o objeto da classe 25, possuem 100% de PR.

## 5.5 Desempenho classe x observador

Uma das hipóteses que motivou este trabalho está ligada à possível razão entre o acréscimo do número de observadores e o aumento da precisão na recuperação de informações. Um exemplo do acréscimo da precisão pela adição de novos observadores pode ser visto nas Figuras 5.6 e 5.8 que representam os gráficos de PR das classes 68 e 14, respectivamente.

No gráfico da Figura 5.6 percebe-se claramente que, à medida em que se adiciona um novo observador, ocorre um incremento no valor de precisão. Isto porque as inferências de cada um dos novos observadores adicionados são melhores ou pelo menos iguais à inferência levantada pelos



(a) PR da classe 1, caixa amarela contida na Figura 5.5.



(b) PR da classe 9, fio dental contido na Figura 5.5.



(c) PR da classe 25, vaso contido na Figura 5.5.



(d) PR da classe 81, xícara contida na Figura 5.5.

Figura 5.4: Representação gráfica da Precisão e Revocação de diversas classes com parâmetro entrópico q = 0.1.



(a) Objeto da classe 1 (caixa amarela).



(b) Objeto da classe 9 (fio dental).



(c) Objeto da classe 25 (vaso).

(d) Objeto da classe 81 (xícara).

Figura 5.5: Amostra dos objetos analisados para a geração dos gráfico contidos na Figura 5.4

observadores já avaliados. Deve-se perceber também que a precisão de recuperação de informações para este objeto não foi muito alta. Porém, manteve-se com poucas oscilações em torno dos 30%.



Figura 5.6: Gráfico da precisão e revocação representando um aumento na precisão pela adição de novos observadores. Dados extraídos a partir das análises da classe de objetos 68 (Figura 5.7).

Por outro lado, a taxa de recuperação de informações apresentada por este objeto sugere que existe um grau de divergência entre as vistas desta classe.

É importante observar que objetos não simétricos ou com vistas muito divergentes podem tornar o processo de recuperação mais difícil, isto porque vistas que compartilham poucas características de cor, forma, e textura entre as demais vistas são menos capazes de descrever o objeto como um todo. Nesse ponto, torna-se importante a inserção de mais observadores (como o feito) o que complementaria o processo de inferência. Assim, para este objeto em especial foram selecionadas aleatóriamente vistas com ângulos iguais a 80°, 75°, 195° e 315° para os observadores 1,2,3 e 4, conforme apresentados na Figura 5.7.



Figura 5.7: Quatro vistas do objeto grampeador da classe 68. Apresentado nos ângulos  $80^{\circ}$ ,  $75^{\circ}$ ,  $195^{\circ}$  e  $315^{\circ}$  (da esquerda para a direita).

O acréscimo na precisão pela adição de novos observadores também é notado nos experimentos realizados com a classe 14. Através do gráfico contido na Figura 5.8, cujos dados foram levantados

sob o emprego de quatro observadores com as vistas 90°, 265°, 175° e 330° (observador 1, 2, 3 e quatro, nesta ordem) é possível observar-se que a classe 14, como um todo, possui valores de recuperação de informações que beiram aos 50%. Contudo, notam-se também quedas abruptas nas curvas da PR a partir de um certo ponto. Os pontos de quedas abruptas no gráfico da PR do objeto 14 com vista de 90°, por exemplo, ocorrem nas faixas entre 0° e 30° e 155° e 195°, o que sugere que as vistas apresentam grandes divergências com o modelo utilizado como probabilidade *a priori*. Desta forma, verifica-se que a queda na precisão acontece porque as vistas usadas como referência não são capazes de descrever o objeto como um todo. De maneira geral, isto ocorre quando se realiza o *matching* em objetos em que há baixa similaridade de cor, forma e textura entre as vistas. Assim, a avaliação conjunta de diversos pontos de vista (Figura 5.9) feita pela inserção de novos observadores diminui os impactos causados pelas informações não recuperadas durante o processo de inferência individual, o que acarreta a melhora global dos índices de PR.



Figura 5.8: Gráfico da precisão e revocação representando um aumento na precisão pela adição de novos observadores. Dados extraídos a partir da análise da classe de objetos 14 (Figura 5.9).



Figura 5.9: Quatro vistas do objeto gato da classe 14. Objeto apresentado nos ângulos 90°, 265°, 175° e 330° (da esquerda para a direita).

De forma contrária ao observado para as classes 68 (Figura 5.6) e 14 (Figura 5.8) a adição de novos observadores nem sempre resulta em incremento na recuperação de objetos. Para algumas

classes analisadas pelo sistema, como, por exemplo, a classe 84 (Figura 5.11), ocorre uma redução na precisão à medida em que se empregam mais observadores. Isto pode ser visto nos resultados representados pelo gráfico contido na Figura 5.10, cujos dados foram extraídos a partir do uso de quatro observadores com as vistas com ângulos de 235°, 70°, 195° e 140°.



Figura 5.10: Gráfico da precisão e revocação representando uma redução na precisão pela adição de novos observadores. Dados extraídos a partir de análises da classe de objetos 84 (Figura 5.11).

No gráfico de desempenho da classe 84 (Figura 5.10) observa-se que ocorre um decréscimo na precisão a partir da inserção do segundo observador. Isto acontece porque os observadores consecutivamente tiveram desempenhos ruins. Assim, as falhas nos processos individuais de inferências incidiram negativamente sobre o resultado geral do reconhecimento.



Figura 5.11: Quatro vistas do objeto pacote amarelo da classe 84. Objeto apresentado nos ângulos 235°, 70°, 195° e 140° (da esquerda para a direita).

Baixos índices de precisão e revocação, como os ocorridos com a classe 84, devem-se muitas vezes à existência de outras classes na base de dados que possuem características muito semelhantes de cor, forma ou textura às do objeto observado. Estas classes dificultam o processo de reconhecimento e acarretam reduções nos valores de precisão. Isto pode ser observado analisando-se a Tabela 5.1 que contém o *ranking* de busca por objetos da classe 84.

Rank	Classe	Rank	Classe	Rank	Classe	Rank	Classe
1	84	19	79	37	79	55	1
2	84	20	79	38	84	56	79
3	84	21	79	39	79	57	3
4	84	22	79	40	79	58	79
5	84	23	84	41	1	59	3
6	79	24	79	42	79	60	3
7	79	25	79	43	84	61	79
8	79	26	79	44	84	62	79
9	79	27	79	45	79	63	3
10	79	28	79	46	79	64	84
11	79	29	79	47	79	65	79
12	79	30	84	48	79	66	1
13	79	31	79	49	79	67	6
14	84	32	79	50	79	68	3
15	84	33	79	51	79	69	3
16	79	34	79	52	1	70	84
17	79	35	79	53	1	71	3
18	79	36	84	54	1	72	1

Tabela 5.1: Valores ordenados da busca por objetos da classe 84

A Tabela 5.1, apresenta os valores retornados pela busca de elementos da classe 84. Nesta tabela, nas colunas *Classe*, estão os números identificadores de cada uma das classes retornadas pela busca. Por outro lado, as colunas *Rank* contêm o índice que apresenta de forma ordenada quais as classes que resultaram no maior valor de *matching* com a classe 84. Idealmente, a Tabela 5.1 deveria conter apenas a classe 84 como resultado da busca. Contudo, são apresentadas também as classes 79 (Figura 5.12), 1 (Figura 5.13), 3 (Figura 5.14) e 6 (Figura 5.15). Isto sugere que estas classes têm uma ou mais características (cor, forma ou textura) em comum com a classe 84, o que reduz o percentual de precisão na busca de informações da classe 84.



Figura 5.12: Quatro vistas do objeto da classe 79 representando os ângulos 235°, 70°, 195° e 140° (da esquerda para a direita).

Os gráficos de PR encontrados nas Figuras 5.6, 5.8 e 5.10 extraídos a partir do uso das vistas mostradas nas Figuras 5.7, 5.9, e 5.11, respectivamente como modelos a *priori* sugerem um comportamento médio, cujo desempenho apresentado pode ser bom ou ruim dependendo da classe de objetos observada. Sendo assim, objetivando uma análise coerente do desempenho do método, é razoável que, neste trabalho, os resultados sejam apresentados levando-se em consideração a média



Figura 5.13: Quatro vistas do objeto da classe 1 representando os ângulos 235°, 70°, 195° e 140° (da esquerda para a direita).



Figura 5.14: Quatro vistas do objeto da classe 3 representando os ângulos 235°, 70°, 195° e 140° (da esquerda para a direita).



Figura 5.15: Quatro vistas do objeto da classe 6 representando os ângulos 235°, 70°, 195° e 140° (da esquerda para a direita).

da precisão para todas as classes e todos observadores, como será mostrado a seguir.

## 5.6 Relação Observadores x Objetos x Parâmetro entrópico

Uma vez que a precisão da recuperação de um objeto da base de dados pode variar de observador para observador de acordo com o ângulo de visão e o parâmetro q da entropia não-extensiva, neste trabalho os experimentos foram organizados variando-se os seguintes parâmetros: número de observadores, número de classes e valor q da Entropia de Tsallis. Assim, os gráficos das Figuras 5.16 a 5.24 representam os resultados do processo de fusão por média aritmética, nos quais foram integradas as evidências de diversos observadores para diferentes valores de q.

Os resultados destes experimentos são apresentados na forma de gráficos, como por exemplo a Figura 5.16, que é o resultado dos experimentos para q = 0.1. No gráfico, o eixo x apresenta a variação da quantidade de observadores, enquanto o eixo y corresponde à média da precisão alcançada. Cada ponto (i, j) deste gráfico representa a precisão j média de 72 buscas de uma mesma classe para i observadores dentre as 7200 imagens da base de dados.

Outro ponto a ser notado é que, à medida em que se varia o tamanho da base de dados, através da inclusão de novas classes espera-se uma oscilação na precisão média global com que os objetos

são recuperados. Sendo assim, com o objetivo de simular bancos de dados de tamanhos distintos, repetimos os experimentos para quantidades variáveis de classes. Desta forma, cada curva da Figura 5.16 corresponde a uma quantidade diferente de classes.

Neste experimento, os gráficos foram organizados dispondo-se os valores de q em ordem crescente. Assim, o primeiro gráfico é o da Figura 5.16, que utiliza o parâmetro q = 0.1. Neste gráfico, nota-se que as séries possuem comportamentos homogêneos, excetuando-se a curva para uma única classe. Isto demonstra que a aplicação do parâmetro de não-extensividade q = 0.1 é capaz de atenuar as discrepâncias em relação à quantidade de classes e número de observadores para grande parte dos objetos da base de dados.



Figura 5.16: Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro entrópico q = 0.1.

A Figura 5.17 representa os valores obtidos quando q = 0.2. Inicialmente, verificamos que, para uma única classe, não é apresentada uma tendência quanto ao incremento do número de observadores, porém, de forma oposta, quando a quantidade de classes é superior a 3, a adição de observadores beneficia a precisão de maneira mais acentuada, até o limite de 3 observadores. Desta forma é possível concluir que, para o parâmetro entrópico q = 0.2, tem-se uma precisão sempre acima dos 50%, e para este valor de q, a melhor faixa de precisão esta situada em 3 ou mais classes na base com 4 ou mais observadores.

Na Figura 5.18, observamos uma acentuada tendência de cerca de 15 pontos percentuais, de aumento de precisão pela adição do segundo observador. Este acréscimo de precisão, embora em menor proporção, é mantido principalmente na curva que representa uma classe, até a faixa onde é inserido o sétimo observador, onde a precisão volta a cair ficando finalmente em torno dos 30%. O mesmo foi notado, embora em menor intensidade, para as curvas de 3 e 10 classes. Por outro



Figura 5.17: Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro entrópico q = 0.2.

lado, as curvas que representam 20, 50, 75 e 100 classes estabilizaram-se com a aplicação do segundo observador. Assim, pelo comportamento das curvas, especificamente para este Gráfico, a melhor precisão deu-se com 10 classes conhecidas e 7 observadores aplicados. Uma outra questão importante notada neste experimento é que ocorreu um grande incremento na precisão pela adição do segundo observador mas, após a aplicação do sétimo observador, quase todas as curvas apresentaram um decréscimo (em maior ou menor grau) na precisão. Isto sugere a existência de uma relação e até mesmo um limiar entre a quantidade de classes conhecidas e o número de observadores.

O comportamento das curvas para o experimento com o parâmetro entrópico q = 0.4 (Figura 5.19) apresentou uma tendência pelo incremento de precisão em função do acréscimo de observadores. Houve novamente para os quatro maiores conjuntos de classes (20,50,75 e 100 classes), uma estabilização mais rápida a partir do quarto observador. Nestas curvas, notaram-se também pequenas oscilações, em média de 0.5% à 1% (de crescimento em geral), completando-se o maior nível de precisão quando houve a adição do  $10^{\circ}$  observador, finalizando as 4 curvas com aproximadamente 75% de precisão. Devemos notar também o expressivo aumento da precisão para a curva que representa "1 classe", que saltou de uma precisão de menos de 10% com um observador, para 57.9% com o uso de 10 de observadores.

O gráfico na Figura 5.20 apresentou para todas as curvas valores muito próximos tanto no emprego de seis, quanto no emprego de dez observadores. Porém, após o sexto observador, em todas as curvas houve uma tendência de queda, o que só foi revertido a partir do nono observador.



Figura 5.18: Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro entrópico q = 0.3.



Figura 5.19: Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro entrópico q = 0.4.

Os resultados da aplicação do parâmetro entrópico q = 0.6, apresentados na Figura 5.21, apresentam um padrão interessante, principalmente em relação às curvas que representam a base de dados de 1, 3, 10 e 20 classes. Nestas curvas, observamos valores de mínimos globais pela adição do terceiro observador, porém, a partir de então, houve apenas acréscimos na precisão, à medida em que novos observadores foram sendo adicionados, sendo que a curva que representa a base



Figura 5.20: Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro entrópico q = 0.5.

com 3 classes chegou à maior precisão dentre as outras 7, com uma precisão de 76.46% para 8 observadores. As demais curvas (50, 75 e 100 classes), como em outros gráficos, alcançaram um comportamento próximo à estabilidade, pois notaram-se algumas ligeiras oscilações positivas, com o uso de apenas dois observadores.



Figura 5.21: Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro entrópico q = 0.6.

Em relação às curvas contidas na Figura 5.22, foi possível observar, novamente, que quando

analisadas as curvas que executaram as médias para quantidades acima de 50 classes, obtiveram-se resultados semelhantes, pelo menos até o limite da base de dados utilizada, que é de 100 classes. Por outro lado, nota-se também que, para uma única classe, o parâmetro entrópico q = 0.7 mostrou-se o melhor. Isto é claro quando aplicado para 3 ou 4 observadores, onde se obtiveram os melhores valores de todo o experimento (para essa quantidade de classes), chegando-se aos 73.79% e 73.20% de precisão, respectivamente.



Figura 5.22: Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro entrópico q = 0.7.

Para o parâmetro entrópico q = 0.8, conforme mostrado na Figura 5.23, as curvas comportaramse de forma homogênea, pois, embora tenham ocorrido algumas pequenas oscilações, em geral, a adição de observadores acarretou um aumento da precisão. A respeito deste comportamento, houve uma única exceção: a curva que representa a simulação para a base de dados com uma classe, que apresentou recuo de 22.20% para 20.28%, quando foi aplicado o terceiro observador. Também, é interessante observar que, para esta quantidade de classes, q = 0.8 mostrou-se o pior valor para o parâmetro entrópico, ficando a precisão média para "1 classe" em torno de 29.43%.

A aplicação do parâmetro não-extensivo q = 0.9 na simulação gerou curvas com um comportamento muito semelhante para as bases de dados com dez ou mais classes, onde foi alcançada a estabilidade na precisão a partir da adição do quarto observador, conforme pode ser visto na Figura 5.24. Por outro lado, a curva que representa a análise com bases de dados constituída por três classes apresentou oscilações na precisão, estabilizando-se apenas com a inserção do quarto observador. Diferentemente do que aconteceu com as outras curvas, a curva relativa à simulação com



Figura 5.23: Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro entrópico q = 0.8.

apenas uma classe apresentou elevado nível de oscilação, sendo que a adição de novos observadores fez com que seus valores de precisão média decrescessem.



Figura 5.24: Gráfico que representa o resultado do experimento da Seção 5.6 usando o parâmetro entrópico q = 0.9.

Após as análises das 63 curvas, onde foram variadas as quantidades de classes, a quantidade de observadores e o parâmetro entrópico, pode-se observar que apenas nas curvas que representam a base de dados contendo uma classe há uma inconstância na precisão. Este comportamento foi

verificado na maioria dos experimentos. Sua ocorrência dá-se porque esta curva representa o cálculo da precisão média do "*matching*" para a primeira classe da base de dados. Então, é natural que, à medida em que são inseridas outras classes na análise, o valor da precisão média aproxime-se dos valores que representam o comportamento médio do sistema. Por esse motivo, é incorreto afirmar que a adição de novas classes ao banco de dados trará melhorias na média da precisão da recuperação de informações. Entretanto, verificou-se que classes de objetos com valores de precisão muito abaixo ou muito acima da média podem causar desvios de precisão no sistema. Contudo, estes desvios podem ser atenuados pela adição de mais classes no processo de análise da precisão global do método. A redução da precisão é causada por classes que possuem vistas com características muito distintas de cor, forma ou textura dependendo do ângulo observado. Classes como a 13 (Figura 5.25) e a 15 (Figura 5.26) representam objetos que possuem baixo percentual de precisão, apresentando grandes divergências entre os ângulos observados. Por outro lado, na base de dados também existem classes simétricas e com pouca variação de cor e textura entre as vistas que ocasionam um aumento da precisão média da recuperação. Exemplos de objetos com estas características podem ser vistos em classes como 25 (Figura 5.27) e a 35 (Figura 5.28).



Figura 5.25: Vistas do objeto da classe 13, que apresenta vistas com grandes variações nas características de cor, forma e textura entre si.



Figura 5.26: Vistas do objeto da classe 15, que apresenta vistas com grandes variações nas características de cor, forma e textura entre si.

Assim, é justificada a importância da análise do comportamento do sistema levando-se em conta diferentes quantidades de objetos da base de dados, pois este procedimento minimiza efeitos indesejados causados por classes com ótimos ou péssimos percentuais de recuperação. A validade deste procedimento pode ser comprovada tendo em vista que, em grande parte dos experimentos,



Figura 5.27: Vistas do objeto da classe 25, que possuem pouca variação nas características de cor, forma e textura entre as vistas.



Figura 5.28: Objetos da classe 35, que possuem pouca variação nas características de cor, forma e textura entre as vistas.

observou-se uma homogeneidade no comportamento das curvas a partir da inserção da décima ou vigésima classe.

Outra característica apresentada por este experimento diz respeito à quantidade ideal de observadores. Nos testes executados, verificou-se a tendência de aumento, seguida pela estabilização na precisão média, devido à adição de novos observadores. Enquanto o incremento na precisão é verificado em quase todas as curvas, o ponto onde ocorre a estabilização varia de acordo com o parâmetro não-extensivo empregado. Contudo, observou-se que um número pequeno de novos observadores três ou quatro - é o suficiente para atingir a precisão média máxima.

# 5.7 Melhor parâmetro entrópico q

Para identificar a existência de um parâmetro entrópico q ideal ou uma aproximação deste, é necessário que se crie uma medida que evidencie globalmente qual o desempenho e o comportamento geral da aplicação. Os gráficos da Seção 5.6 representam estas informações, portanto, é razoável que a obtenção de um valor médio, baseado nas curvas que representam o desempenho médio para todas as 100 classes apresente uma perspectiva válida sobre o desempenho para cada um dos parâmetros. Como método comparativo, pode-se analisar visualmente o comportamento de cada um dos parâmetros, inserindo-os conjuntamente em um mesmo gráfico. Esta análise conjunta das curvas resultou no gráfico contido na Figura 5.29, onde o eixo x representa a quantidade de observadores e o eixo y a precisão média alcançada.

Visualmente, nota-se que a curva que representa o parâmetro q = 0.1 apresenta melhores va-

lores de precisão, principalmente quando se analisa o comportamento do sistema com um ou dois observadores apenas. De maneira geral, o comportamento das curvas apresenta uma tendência de alta mais acentuada até a inserção do terceiro ou quarto observador, a partir de onde ocorrem apenas pequenas oscilações na precisão. Os valores de precisão alcançados pelas curvas encontram-se na Tabela 5.2 e estão dispostos de acordo com as diferentes quantidades de observadores.



Figura 5.29: Gráfico da média obtida para todas as classes da base de dados.

Observ.	q=0.1	q=0.2	q=0.3	q=0.4	q=0.5	q=0.6	q=0.7	q=0.8	q=0.9
1	0.6592	0.6374	0.6213	0.6099	0.6342	0.6098	0.5867	0.6147	0.6387
2	0.7303	0.7109	0.7056	0.6943	0.6865	0.6917	0.6909	0.7182	0.7059
3	0.7371	0.7297	0.7272	0.7229	0.7167	0.7115	0.7183	0.7346	0.7372
4	0.7537	0.7395	0.7379	0.7308	0.7237	0.7220	0.7313	0.7464	0.7546
5	0.7539	0.7448	0.7411	0.7321	0.7314	0.7268	0.7321	0.7577	0.7549
6	0.7573	0.7522	0.7441	0.7416	0.7334	0.7375	0.7333	0.7576	0.7593
7	0.7621	0.7533	0.7414	0.7417	0.7345	0.7393	0.7345	0.7608	0.7630
8	0.7613	0.7535	0.7435	0.7441	0.7348	0.7397	0.7326	0.7623	0.7633
9	0.7635	0.7529	0.7435	0.7471	0.7336	0.7375	0.7357	0.7631	0.7663
10	0.7696	0.7530	0.7421	0.7459	0.7356	0.7409	0.7353	0.7626	0.7672
Média	0.7448	0.7327	0.7248	0.7210	0.7164	0.7157	0.7131	0.7378	0.7411

Tabela 5.2: Médias dos valores das curvas contidas no gráfico da Figura 5.29.

A Tabela 5.2 contém os dados da simulação da Figura 5.29, onde cada coluna representa os dados de desempenho obtidos nas simulações para todas as 100 classes. As linhas, por sua vez, representam a média da precisão alcançada para cada quantidade de observadores. Com o intuito de encontrar o melhor parâmetro não-extensivo, utilizou-se a média do desempenho de cada um dos parâmetros. Os valores encontrados pelos cálculos dessas médias constam na última linha de cada coluna. Conseqüentemente, é possível organizar os valores das médias extraídas em ordem decrescente, de forma a indicar qual parâmetro q possui o melhor valor médio da precisão, conforme feito na Tabela 5.3.

Rank	Parâmetro q	Precisão Média (%)		
1	0.1	74.479		
2	0.9	74.105		
3	0.8	73.781		
4	0.2	73.271		
5	0.3	72.477		
6	0.4	72.103		
7	0.5	71.644		
8	0.6	71.568		
9	0.7	71.308		

Tabela 5.3: Tabela com os valores ordenados dos cálculos das média das precisões com parâmetro q variando entre 0.1 à 0.9.

A Tabela 5.4 apresenta as precisões médias correspondentes para cada valor do parâmetro entrópico usado nos experimentos, juntamente com cada desvio-padrão. Pode-se observar que o valor para q = 0.1, apresentou o melhor desvio-padrão e também a melhor combinação precisão - desviopadrão. Isto indica que o melhor parâmetro a ser utilizado para a base de dados e os descritores empregados é q = 0.1.

Sendo assim, para os demais experimentos apresentados nesse trabalho, sempre será utilizado q = 0.1. Este valor também indica um comportamento não-extensivo (hipótese **a** da Seção 5.1).

q	Precisão Média(%)	Desvio Padrão
0.1	74.479	3.241
0.2	73.271	3.623
0.3	72.477	3.826
0.4	72.103	4.213
0.5	71.644	3.257
0.6	71.568	4.039
0.7	71.308	4.647
0.8	73.781	4.571
0.9	74.105	4.048

Tabela 5.4: Tabela com os valores dos desvios padrão das precisões médias com parâmetro não-extensivo q variando entre 0.1 e 0.9.

# 5.8 Casamento por Cor, Forma e Textura

Neste trabalho todas as imagens foram analisadas a partir de três características básicas: a cor, a forma e a textura. Uma das premissas desta pesquisa é avaliar qual destas três características (cor, forma e textura) é mais relevante no processo de recuperação de informações visuais. Com este propósito, foram feitos três testes distintos nos quais cada característica serviu exclusivamente como parâmetro de comparação. Sendo assim, executou-se isoladamente para cada uma das 7200 imagens da base de dados, o *matching* (casamento) de cor, o *matching* de forma e o *matching* de

textura. Cada uma destas três simulações também levou em conta a influência do número de observadores no cálculo da precisão, bem como o uso do parâmetro não-extensivo q = 0.1. Como nos testes anteriores, utilizou-se como medida o método da precisão e revocação, a partir do qual foram gerados três gráficos de precisão, sendo um para cor, um para forma e um para textura cobrindo assim as três características analisadas na monografia.

A Figura 5.30 apresenta o gráfico de desempenho do *matching* utilizando apenas as informações de cor das imagens. Pelo comportamento descrito pela curva, nota-se que para esta característica, a adição de novos observadores é um fator de incremento na precisão, embora este incremento seja aproximadamente 0.5 pontos percentuais por observador. Também observou-se que a precisão alcançada apresentou valores muito satisfatórios, ficando sempre acima dos 90%, o que indica que esta característica tem alta relevância no processo de recuperação de informações na base de dados utilizada sob o modelo proposto.



Figura 5.30: Gráfico que representa o resultado do experimento utilizando o parâmetro entrópico q = 0.1 no *matching* das características de cor.

O comportamento do *matching* baseado apenas nas características de forma é representado pela Figura 5.31. Neste gráfico verificou-se que o processo de recuperação desta característica é pouco influenciado pela adição de novos observadores, isto porque a curva é praticamente estável, apresentando desvio-padrão de 0.43, mantendo um desempenho médio de 41.55% na precisão.

Os valores do *matching* da característica de textura das imagens são descritos pela curva contida no gráfico da Figura 5.32. O comportamento da curva indica que o aumento da quantidade de obser-



Figura 5.31: Gráfico que representa o resultado do experimento utilizando o parâmetro entrópico q = 0.1 no *matching* das características de forma.

vadores influencia muito pouco a precisão da recuperação de informações. Contudo, o desempenho do matching das características de textura apresenta as menores taxas de precisão, com uma média de 25.61% e desvio-padrão 0.72.

Após a análise individual das características de cor, forma e textura, fica evidente que, para os descritores selecionados para esta pesquisa, a maior relevância no processo de recuperação de informações encontra-se na característica de cor em relação à forma e à textura, nesta ordem. Entretanto, mesmo sendo o *matching* por cor o mais eficiente, na média as interações de longo alcance entre os três descritores (cor, forma e textura) impedem o descarte das características de forma e textura no processo de recuperação de informações visuais. Fato é que nenhum dos descritores é capaz de responder adequadamente a todos os tipos de objetos e situações.Para validar esta afirmação, preparou-se um experimento utilizando-se dois objetos da mesma classe, porém com ângulos diversos. O objetivo deste experimento foi avaliar se existia alternância entre relevância das características no processo de matching dos objetos. Assim, para as classes 1, 8 e 11 da base de dados, a forma foi o principal descritor no processo de RIBC. Isto pode ser verificado observando-se a Tabela 5.5 onde são apresentados os valores de "*matching*" para cada uma das três características (cor, forma e textura) básicas estudadas.

Por outro lado, o mesmo acontece com os descritores de textura, que foram mais eficientes, por exemplo, para objetos das classes 9, 45 e 80, cujos percentuais de matching para cada uma das


Figura 5.32: Gráfico que representa o resultado do experimento utilizando o parâmetro entrópico q = 0, 1 no *matching* das características de textura.

Tabela 5.5: Percentuais de *matchings* para cor, forma e textura, exemplificando classes nas quais a forma foi a característica mais relevante na RIBC.

Classe	Cor (%)	Forma (%)	Textura (%)
1	96.85	98.18	84.41
8	95.07	97.21	69.13
11	96.64	99.05	91.58

características constam na Tabela 5.6.

Tabela 5.6: Percentuais de *"matchings"* para cor, forma e textura, exemplificando classes nas quais a textura foi a característica mais relevante na <u>RIBC</u>.

Classe	Cor (%)	Forma (%)	Textura (%)
9	94.75	98.95	99.72
45	89.81	98.89	99.18
80	96.37	96.26	99.81

Uma vez que as características mais influentes na precisão média são variáveis e dependem de muitos fatores, é necessário um modelo que combine cor, forma e textura adequadamente, visando a maximização da precisão no processo de reconhecimento. Isso justifica o uso do modelo Bayesiano proposto, dada a sua capacidade de combinar as três características avaliadas, ressaltando sempre a maior quando conveniente.

## Capítulo 6

## CONCLUSÕES

Neste trabalho foi apresentado uma metodologia para estudo de um modelo bayesiano com base na distância de Kullback-Leibler estendida para um sistema de reconhecimento de objetos a partir de várias vistas (observadores).

Os experimentos foram tomados sob uma base de dados de 100 classes e 7200 imagens. A primeira hipótese apresentada, de que o sistema poderia ser não-extensivo, foi confirmada experimentalmente para q = 0.1, que apresentou melhor desempenho, em termos de precisão e revocação (Tabela 5.3). Isso sugere a forte correlação entre as diversas características utilizadas para reconhecimento de objetos, indicando interações de longo alcance espaciais e temporais.

Os experimentos executados comprovaram que, conforme sugere a segunda hipótese da Seção 5.1 do Capítulo 5, existe uma característica mais relevante no processo de recuperação de informações visuais, que neste trabalho é a cor, seguida pelas características de forma e textura, nesta ordem. Contudo, embora seja comprovada a existência de uma característica dominante no processo de recuperação, como visto, não se deve desconsiderar as outras características pois, dadas as análises feitas, nenhum dos descritores (cor, forma e textura) prevalece de forma absoluta no processo de recuperação de informações para todos os tipos de objetos (ver Tabelas 5.5 e 5.5). Isto reforça a afirmação de que a correlação entre as características é benéfica para o processo de RIBC, uma vez que não é possível prever qual característica é a mais relevante para um determinado objeto. Portanto, o modelo proposto mostra-se, no momento, mais adequado para a combinação das três características. Desta forma, a correlação entre as características justifica plenamente o uso simultâneo da cor, da forma e da textura, para a recuperação de informações, o que foi feito através do modelo bayesiano com o uso do "*matching*" baseado em Kullback-Leibler estendido.

Com relação às múltiplas visões, terceira hipótese levantada no Capítulo 5, todos os experimentos mostram que, na média, o desempenho do reconhecimento do sistema aumenta sensivelmente à medida em que mais de um observador são adicionados ao sistema, apresentando uma estabilidade a partir do terceiro ou quarto observadores, indicando a não necessidade de se incluir mais observadores. Isto se deve a capacidade do sistema de recuperar informações de objetos mesmo em ângulos distintos, assim sendo, para que a adição de novos observadores fosse necessária além de um determinado limite, a base deveria conter amostras dos objetos não apenas rotacionados verticalmente, mas também horizontalmente. Como conseqüência, por ser necessário somente poucas vistas, baseando-se nos dados apresentados, pode-se dizer que o sistema é de baixo custo computacional, pelo menos até o limite de 100 classes.

Aumentando o número de classes, também é verificada uma estabilidade de performance. Então, de acordo com os experimentos, pode-se dizer que o sistema tem seu melhor desempenho quando q = 0.1 com até quatro observadores, para até 100 classes diferentes.

Durante a análise do processo de recuperação verificou-se a existência de classes de objetos como a 13 (Figura 5.25) e a 15 (Figura 5.26), cujos valores das precisões eram abaixo da média. Avaliando-se as características destas classes notou-se que este efeito de redução no valor da precisão é devido a grandes discrepâncias de cor, forma ou textura entre as diferentes vistas do objeto, portanto, é natural que ocorra uma diminuição da precisão, dependendo do ângulo da vista analisado. Por outro lado, as classes como a 25 (Figura 5.27) e a 35 (Figura 5.28) apresentam geometria simétrica, o que favorece o reconhecimento pela característica de forma, e pouca ou nenhuma variação de cor ou textura em todas as 72 vistas analisadas, elevando os percentuais de precisão à 100% na RIBC.

Desta forma, as três principais hipóteses relativas a este trabalho foram experimentadas e conforme dados constantes no Capítulo 5, o que indica que a arquitetura proposta no Capítulo 4 foi capaz de cobrir os requerimentos propostos, mostrando-se viável e aplicável em soluções de problemas relacionados ao reconhecimento de objetos, sendo sua empregabilidade, extensível a sistemas de outras naturezas, como, por exemplo, a reconstrução 3D, Realidade Aumentada e *scanner* 3D.

## **Referências Bibliográficas**

- AZUMA, R. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, v. 6, n. 4, p. 355–385, 1997.
- BARÃO, M. Entropia, Entropia Relativa e Informação Mútua, 2003. Disponível em: <http://www.di.uevora.pt/~mjsb/TI-02/Artigos/cap2.pdf>. Acesso em: 3 dez. 2009.
- BECK, C. Non-extensive estatistical mechanics approuch to fully develop hydrodynamic turbulence. *Chaos, Solutions and Fractals*, v. 13, p. 499–506, 2002.
- BILLINGHURST, M.; CHEOK, A.; PRINCE, S.; KATO, H. Real World Teleconferencing. *IEEE Computer Graphics and Applications*, v. 22, n. 6, p. 11–13, Dezembro 2002.
- CABELLA, B. C. T.; STURZBECHER, M. J.; TEDESCHI, W.; FILHO, O. B.; ARAÚJO, D. B.; NEVES, U. P. C. A Numerical Study of the Kullback-Leibler Distance in Functional Magnetic Resonance Imaging. *Brazilian Journal of Physics*, v. 38, n. 1, Março 2008.
- COELHO, T.; CALADO, P.; SOUZA, L.; RIBEIRO-NETO, B.; MUNTZ, R. Image Retrieval Using Multiple Evidence Ranking. *IEEE Transactions on Knowledge and Data Engineering*, v. 16, p. 408–417, Abril 2004.
- COLUMBIA UNIVERSITY, DEPARTMENT OF COMPUTER SCIENCE. internet, 2001. Disponível em: <a href="http://www.cs.columbia.edu/CAVE/research/softlib/coil-100.html">http://www.cs.columbia.edu/CAVE/research/softlib/coil-100.html</a>.
- CRISTO, M.; CALADO, P.; SILVEIRA, M. L.; SILVA, I.; MUNTZ, R.; RIBEIRO-NETO, B. Bayesian Belief Networks for IR. International Journal of Approximate Reasoning, v. 40, p. 163–179, Novembro 2003.
- EDUCASE. 7 things you should know about Augmented Reality, 2005. Disponível em: <http://net.educause.edu/ir/library/pdf/ELI7007.pdf>. Acesso em: 5 ago. 2008.

- ÇENGEL, Y. A. Introduction to thermodinamycs and heat transfer. II. McGraw-Hill, international ed., 1997. ISBN 0-07-011498-6.
- ESQUEF, I. A. **Técnicas de Entropia em processamento de imagens**. Tese de mestrado, Centro Brasileiro de Pesquisas Científicas, 2002.
- FARSHIDI, F.; SIROUSPOUR, S.; KIRUBARAJAN, T. Robust sequential view planning for object recognition using multiple cameras. *Image and Vision Computing*, v. 27, p. 1072–1082, 2009.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento de Imagens Digitais*. Edgard Blücher, São Paulo, 2003. ISBN 85-212-0264-4.
- GREEN, P.; WOLPERT, R.; RODRIGUEZ, C.; BAYARRI, S.; ZELLNER, A.; EVANS, M. What is Bayesian Analysis?, 2001. Disponível em: <http://www.bayesian.org/index. html>. Acesso em: 5 ago. 2008.
- JENSEN, F. V. *Bayesian networks and decision graphs*. Spring-Verlag, New York, 2001. ISBN 0-387-95259-4.
- JIN, W.; SHI, R.; CHUA, T. A Semi-Naïve Bayesian Method Incorporating Clustering with Pair-wise Constraints for Auto Image Annotation. *Proceedings of the 12th annual ACM international conference on Multimedia*, p. 336–339, Outubro 2004.
- JOHNSON, D. H.; SINAMOVIC, S. Symmetrizing the Kullback-Leibler Distance. Relatório técnico, IEEE Transactions on Information Theory, 2001.
- KATO, H. ARToolKit Home Page. internet, 2009. Disponível em: <http://www.hitl. washington.edu/artoolkit/>. Acesso em: 3 dez 2009.
- KIM, S.; KWEON, I. S. Simultaneous place and object recognition using collaborative context information. *Image and Vision Computing*, v. 27n. 6, p. 824–833, 2008a.
- KIM, S.; KWEON, I. S. Scalable representation for 3D object recognition using feature sharing and view clustering. *Pattern Recognition*, v. 41n. 2, p. 754–773, 2008b.
- KULLBACK, S.; LEIBLER, R. A. On Information and Sufficiency. *The Annals of Mathematic and Statistics*, v. 22, n. 1, p. 79–86, Março 1951.

- LI, L.; HUANG, W.; GU, I. Y. H.; TIAN, Q. Foreground object detection from videos containing complex background. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, p. 2–10, 2003.
- MARTIN, S.; MORISON, G.; NAILON, W.; DURRANI, T. Fast and accurate image registration using Tsallis entropy and simultaneous perturbation stochastic approximation. *ELEC*-*TRONICS LETTERS*, v. 40, n. 10, p. 595–597, Maio 2004.
- MATHWORKS. internet, 2009. Disponível em: <http://www.mathworks.com/access/ helpdesk/help/toolbox/images/f8-20792.%html>. Acesso em: 3 dez. 2009.
- NIAN, R.; JI, G.; ZHAO, W.; FENG, C. Probabilistic 3D object recognition from 2D invariant view sequence based on similarity. *Neurocomputing*, v. 70, p. 785–793, 2007.
- OSWALD, N.; LEV, P. Cooperative Object Recognition. Pattern Recognition Letters, v. 12, n. 22, p. 1273–1282, 2001.
- RIBEIRO-NETO, B.; MUNTZ, R. R. A Belief Network Model for IR. ACM, p. 253–260, Novembro 1996a.
- RIBEIRO-NETO, B.; MUNTZ, R. R. Approximate Answers in Databases Through Bayesian Belief Networks. XVI Int. Conference of the Chilean Computer Science Society, p. 31–42, Novembro 1996b.
- RODRIGUES, P. S. S. Um Modelo Bayesiano Combinando Análise Semântica Latente e Atributos Espaciais para Recuperação de Informação Visual. Tese de doutorado, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Março 2003.
- RODRIGUES, P. S. S.; SILVA, R.L.S; OLIVEIRA, J. C.; GIRALDI, G. Augmented Reality for Scientific Visualization: Bringing DataSets Into the RealWorld. 2004.
- RODRIGUES, P. S. S.; SILVA, R.L.S.; GIRALDI, G.; CUNHA, G. Object Recognition using Bayesian Networks for Augmented Reality Systems. 2005.
- SHANNON, C. E. Mathematical Theory of Communication. *The Bell System Technical Journal*, v. 27, p. 379–423 e 623–656, Julho, Outubro 1948.
- SILVA, R. L. S. Um Modelo de Redes Bayesianas Aplicado a Sistemas de Realidade Aumentada. Tese de doutorado, COPPE/UFRJ, Dezembro 2006.

- TANG, A. W. K.; P., NG T.; HUNG, Y. S.; LEUNG, C. H. **Projective Reconstruction from line-correspondence in multiple uncalibrated images**. *Pattern Recognition*, Junho 2005.
- TARUYA, A.; SAKAGAMI, M. Gravothermal catastrophe and Tsallis' generalized entropy of self-gravitating systems. *Physica A*, v. 307, p. 185–206, 2002.
- TAVARES, A. H. M. P. Aspectos Matemáticos da Entropia. Tese de mestrado, Universidade de Aveiro, 2003.
- THORNTON, J.; SAVVIDES, M.; KUMAR, B. V. K. V. A Bayesian Approach to Deformed Pattern Matching of Iris Images. *IEEE Transations on Pattern Analisys and Machine Inteligence*, v. 29, n. 4, p. 596–606, Abril 2007.
- TSALLIS, C. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, v. 52, p. 479–487, 1988.
- VASCONCELOS, N.; HO, P.; MORENO, P. The Kullback-Leibler Kernel as a Framework for Discriminant and Localized Representations for Visual Recognition. 2004. Disponível em: <a href="http://www.escholarship.org/uc/item/6h9259w1">http://www.escholarship.org/uc/item/6h9259w1</a>. Acesso em: 3 dez. 2009.
- WASSERMAN, J. internet, 1999. Disponível em: <http://www.research.ibm.com/ pieta/>. Acesso em: 3 dez. 2009.
- WEISS, I.; RAY, M. Model-Based Recognition of 3D Objects from Single. *IEEE Transactions* on Pattern Analisys and Machine Intelligence, v. 23n. 2, p. 116–128, 2001.
- YONG-LI, L.; WEI-ZHOU, G.; LING-YAN, Z. The Application of Bayesian Method in Image Segmentation. *IEEE*, 2007.