

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Informática

Investigação de Técnicas para Extração de
Características e Indexação usando Redes GHSOM
Aplicadas à Recuperação de Imagens por Conteúdo

Saulo de Tarso Oliveira Rodrigues

Dissertação submetida à Coordenação do Curso de Pós-Graduação
em Ciência da Computação da Universidade Federal de Campina
Grande - Campus I como parte dos requisitos necessários para ob-
tenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Modelos Computacionais e Cognitivos

Herman Martins Gomes (Orientador)

Campina Grande, Paraíba, Brasil

©Saulo de Tarso Oliveira Rodrigues, Agosto de 2008

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

R696i Rodrigues, Saulo de Tarso de Oliveira.

2008 Investigação de técnicas para extração de características e indexação usando redes GHSOM aplicadas à recuperação de imagens por conteúdo / Saulo de Tarso de Oliveira Rodrigues. - Campina Grande, 2008.
172f. : il. Col.

Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Orientador: Prof. Dr. Herman Martins Gomes.

Referências.

1. Indexação de Imagens 2. Redes Neurais 3. GHSOM 4. Extração de Características 5. CBIR I. Título

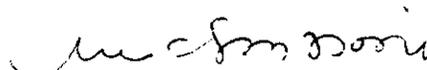
CDU - 004.8(043)

**“INVESTIGAÇÃO DE TÉCNICAS PARA EXTRAÇÃO DE
CARACTERÍSTICAS E INDEXAÇÃO USANDO REDES GHSOM
APLICADAS À RECUPERAÇÃO DE IMAGENS POR
CONTEÚDO”**

SAULO DE TARSO OLIVEIRA RODRIGUES

DISSERTAÇÃO APROVADA EM 26.08.2008


PROF. HERMAN MARTINS GOMES, Ph.D
Orientador


PROF. MARCUS COSTA SAMPAIO, Dr
Examinador


PROF. RICARDO DA SILVA TORRES, Dr.
Examinador

CAMPINA GRANDE – PB

*Dedico este trabalho aos meus pais Eliel e Ana,
minha irmã Eliana, minha namorada Luciana
e a minha tia Socorro(in memoriam),
que me ensinou a lutar sempre pela vida.*

Agradecimentos

Gostaria de primeiramente agradecer a Deus por me proporcionar paciência, perseverança, saúde e coragem para a realização desse trabalho.

Aos meus pais Eliel Nunes e Ana Estrela, meus tutores da vida, que me ensinaram tudo que sei sobre moralidade, respeito e acima de tudo educação. Painho e Mainha vocês são meus exemplos, os meus modelos. Queria também agradecer a minha irmã Eliana que sempre foi companheira.

A minha namorada Luciana Figueirêdo, que se mostrou amorosa e paciente durante minha ausência e por trazer um pouco mais de sentido a minha existência.

A minha família que esteve comigo em todos os momentos me apoiando em todas as decisões. A minha avó querida Chicolina, que sempre orou por meu sucesso e sempre estava presente com carinho e atenção e a minha querida tia Socorro Estrela (*in memoriam*) que é um exemplo na minha vida de luta e amor. Queria também agradecer a hospitalidade e disposição do meu tio Berlindo Estrela e sua esposa Maria Auxiliadora que me acolheram em Campina Grande.

Ao professor Herman Martins Gomes, pela orientação, acompanhamento constante além de seu senso de humor, que foi fundamental em todos os momentos para realização desse trabalho.

Queria agradecer também a todos os meus irmãozinhos de laboratório que contribuíram de alguma forma com esse trabalho, desde uma ajuda na escrita ou implementação como também com conversas e lanches no final da tarde (Bruno Alexandre, Fernando Henrique, Paulo de Tarso, Vinícius Porto, Francisco Fabian, Camilla Falconi, Mikaela Maia e Rafael Figueiredo) que se tornaram uma família ao longo desses dois anos de dissertação.

Ao meu fiel escudeiro (Assistente Geral de Assuntos do Mestrado) Tiago Albuquerque que batalhou em parceria comigo para tornar esse trabalho uma realidade.

Agradecer também a todos os amigos conquistados em minha estadia em Campina Grande ao longo de dois anos, Bruno de Brito (irmãozinho), Severino Neto, Paulo Ditarso, Álvaro (Degas), Celso Brennand, Yuri Lacerda, Halley Freitas, Danillo Cesar, Ramon, Batista Marques, Alex Serres (o frances), Georgina, Lucyanno Holanda, Fabiana (Bia), Ana Clara (Clarinha) e Mirna Maia (que não me deixava escrever).

Ao pessoal da COPIN, Aninha e Vera, por sempre estarem de bom humor e prontas a servir.

Ao pessoal da Cantina, Dona Inês, Romildo, Tibério e Germano, que sempre estavam prontos para ouvir e conversar nos momentos de relaxamento.

A toda turma da Quarta Insana, que fizeram parte das revisões, platéia e banca examinadora durante todo meu trabalho, além de oferecerem os melhores jantares das quartas-feiras, são eles: Marcos Muccini (Marquinhos), o sempre bem humorado e cozinheiro; Guilherme Germoglio (Guiga), o cara mais sincero da turma, conhecido por muitos como bruto; Gilson Junior (Negola), um cara de coração enorme e o mais maluco que eu conheço; Felipe Cardoso (Farinha), prestativo e sempre conhecido por seu estilo playboy de ser; Flávio Santos (Baratão), sempre feliz e maluco beleza; e Fernando Henrique (Nando T), o cara mais inteligente e carismático da turma. Amigos que ficaram para sempre.

Resumo

Esta dissertação propõe um sistema de recuperação de imagens por conteúdo com aplicação na *World Wide Web*. Foi realizada uma revisão bibliográfica das principais técnicas comumente utilizadas na extração de características, indexação e recuperação de imagens. O principal objetivo da dissertação foi avaliar a combinação de uma seleção de métodos de extração de características, baseado em cor, forma e textura, utilizando um método de indexação baseados em Mapas Auto-Organizáveis Construtivos Hierárquicos (GHSOM) a fim de obter uma maior acurácia na recuperação. Como resultado prático desta dissertação, foi desenvolvido um sistema de recuperação de imagens baseado em conteúdo, o qual foi testado através de experimentos objetivos e subjetivos, de forma a permitir uma melhor avaliação das combinações de métodos implementados. Foram realizados dois experimentos principais, um primeiro experimento objetivo utilizou como base de teste um conjunto imagens rotuladas e teve como meta a realização de um estudo comparativo entre métodos simples e combinados. Um segundo experimento subjetivo considerou a opinião de usuários na avaliação subjetiva da precisão dos diferentes métodos implementados. Foram calculadas medidas estatísticas a partir dos dados obtidos nestes experimentos, tendo sido encontrados resultados promissores para as combinações das características extraídas.

Abstract

This dissertation proposes a content-based image retrieval system with application on the World Wide Web. A literature review has been prepared, including common techniques for feature extraction, indexing, and image retrieval. The main goal of this dissertation was to evaluate the combination of several methods for feature extraction (based on color, shape, and texture), utilizing a method of indexing based on Growing Hierarchical Self-Organizing Maps (GHSOM) with the purpose of obtaining higher accuracy. As a practical result, a system for content-based image retrieval has been developed and validated through objective and subjective experiments, in order to allow a better evaluation of the combination of implemented methods. The objective experiment used as test base a set of labeled images and had as its main goal the development of a comparative study among standalone and combined methods. On the other hand, the subjective experiment involved testing all the system and required a group of testers, who performed image retrieval experiments on unlabeled images. The purpose was to subjectively evaluate the precision of the several implemented methods. Some statistical measures have been calculated from the outputs of these experiments and promising results have been found for combinations of the features.

Conteúdo

1	Introdução	1
1.1	Motivação	1
1.2	Descrição do Problema	2
1.3	Objetivos e Relevância	3
1.4	Estrutura da Dissertação	4
2	Revisão Bibliográfica	6
2.1	Extração de Características	7
2.1.1	Cor	7
2.1.2	Forma	10
2.1.3	Textura	12
2.2	Estruturas de Indexação	18
2.3	Combinação de Classificadores	28
2.4	Exemplos de Sistemas	30
2.4.1	QBIC	30
2.4.2	NETRA	31
2.4.3	PicSOM	31
2.4.4	VisualSEEk	32
2.4.5	WebSEEk	32
2.4.6	Sistema de French <i>et al.</i>	33
2.4.7	Sistema de Barcellos <i>et al.</i>	33
2.4.8	Sistema de Prasad <i>et al.</i>	34
2.4.9	Sistema de Edvarden	34
2.5	Análise Comparativa	35
2.6	Considerações Finais	35

3	Projeto e Implementação	38
3.1	Arquitetura do Sistema	38
3.2	Módulos do Sistema	44
3.3	Modos de Operação	48
3.4	Considerações Finais	49
4	Estudo Experimental	51
4.1	Experimento 1 - Busca pelos Melhores Parâmetros dos Classificadores GHSOM	51
4.2	Experimento 2 - Combinação de Classificadores	64
4.3	Experimento 3 - Teste com a Interface <i>Web</i> os Usuários potenciais	74
4.3.1	Experimento com Usuários - Imagens Aleatórias	76
4.3.2	Experimento com Usuários - Conjunto de Imagens Fixas	78
4.3.3	Experimento com Usuários - Única Imagem	83
4.4	Considerações Finais	86
5	Conclusão	88
5.1	Resumo da Dissertação	88
5.2	Contribuições	89
5.3	Trabalhos Futuros	90
A	Robô de Busca (<i>Web Crawler</i>)	97
B	Banco de Imagens	99
C	Respostas da Recuperação	117
C.1	Resultado de Classificação - Conjunto de Imagens Fixas	117
C.2	Resultado de Classificação - Única Imagem	137
D	Roteiro do Experimento com os Usuários	156

Lista de Figuras

2.1	Histograma para 256 níveis de cinza.	8
2.2	Amostras de texturas.	13
2.3	Conjunto de vizinhos igualmente espaçados na superfície circular para diferentes P e R	14
2.4	Rotação buscando o menor número inteiro entre os bits.	15
2.5	Transformada wavelet. Inicia-se com a imagem original e (a) após a transformação, obtêm-se imagens em outras escalas: (b) LL, (c) HL, (d) LH e (e) HH.	17
2.6	Estrutura de um Mapa Auto-Organizável.	22
2.7	Inserção de neurônios.	25
2.8	Exemplo do crescimento de um GHSOM.	26
2.9	Arquiteturas para combinação de classificadores.	29
3.1	Arquitetura do Sistema de Recuperação de Imagens Baseada em Conteúdo.	39
3.2	Telas do Sistema de Recuperação de Imagens Baseada em Conteúdo, em que MB indica o fluxo do Modo de Busca e MC o Modo de Consulta.	40
3.3	Tela dos resultados da recuperação.	41
3.4	Módulo de extração de características.	44
3.5	Módulo de classificação.	45
3.6	Combinação de classificações	46
3.7	Modo de Busca.	48
3.8	Modo de Consulta.	49
4.1	Avaliação do parâmetro τ_2 utilizando o extrator RGB com 128 <i>bins</i> por canal.	54
4.2	Avaliação do parâmetro τ_2 utilizando o extrator RGB com 64 <i>bins</i> por canal.	54
4.3	Avaliação do parâmetro τ_2 utilizando o extrator RGB com 32 <i>bins</i> por canal.	55

4.4	Avaliação do parâmetro τ_2 utilizando o extrator RGB com 16 <i>bins</i> por canal.	55
4.5	Avaliação do parâmetro τ_2 utilizando o extrator HSV com 128 <i>bins</i> por canal.	56
4.6	Avaliação do parâmetro τ_2 utilizando o extrator HSV com 64 <i>bins</i> por canal.	57
4.7	Avaliação do parâmetro τ_2 utilizando o extrator HSV com 32 <i>bins</i> por canal.	57
4.8	Avaliação do parâmetro τ_2 utilizando o extrator HSV com 16 <i>bins</i> por canal.	58
4.9	Avaliação do parâmetro τ_2 utilizando o extrator YCbCr com 128 <i>bins</i> por canal.	58
4.10	Avaliação do parâmetro τ_2 utilizando o extrator YCbCr com 64 <i>bins</i> por canal.	59
4.11	Avaliação do parâmetro τ_2 utilizando o extrator YCbCr com 32 <i>bins</i> por canal.	59
4.12	Avaliação do parâmetro τ_2 utilizando o extrator YCbCr com 16 <i>bins</i> por canal.	60
4.13	Avaliação do parâmetro τ_2 utilizando o extrator Wavelets.	61
4.14	Avaliação do parâmetro τ_2 utilizando o extrator LBP.	61
4.15	Avaliação do parâmetro τ_2 utilizando o extrator Momentos Invariantes de Hu.	62
4.16	Precisões médias para combinação de classificadores dois a dois.	66
4.17	Precisão média para combinação de classificadores <i>versus</i> classificadores simples.	66
4.18	Cobertura para combinação de classificadores <i>versus</i> classificadores simples. .	67
4.19	Precisões médias para a combinação de classificadores três a três.	68
4.20	Precisões médias para a combinação de classificadores três a três.	68
4.21	Precisões médias para a combinação de classificadores quatro a quatro.	69
4.22	Coberturas médias para a combinação de classificadores quatro a quatro. . .	69
4.23	Precisões médias para a combinação de classificadores cinco a cinco.	70
4.24	Coberturas médias para a combinação de classificadores cinco a cinco.	71
4.25	Precisões e coberturas médias para a combinação de classificadores seis a seis.	71
4.26	Precisões médias para comparar os classificadores e suas combinações.	72
4.27	Coberturas médias para comparar os classificadores e suas combinações. . . .	73
4.28	Tela de configurações avançadas do sistema de recuperação.	76
4.29	Tela de resultados recuperados do sistema de recuperação.	77
4.30	Precisões médias para todos os usuário.	77
4.31	Tela inicial dos novos testes com os usuários.	79
4.32	Tela de resultados da interface de testes com os usuários.	80
4.33	Imagens utilizadas nos testes com os usuários.	81
4.34	Precisões para cada classificador utilizado.	81
4.35	Valores de Precisão - Média das Precisões para cada usuário.	82
4.36	Gráfico com as precisões médias para os classificadores sem os usuários incomuns.	83

4.37	Imagem padrão selecionada para todos os testes com os usuários.	84
4.38	Gráfico com as precisões para cada classificador utilizado.	84
4.39	Valores de $ \text{Precisão} - \text{Média das Precisões} $ para cada usuário.	85
4.40	Gráfico com as precisões médias para os classificadores sem os usuários incomuns.	86
A.1	Arquitetura do Robô de Busca.	98
B.1	Grupo de imagens Aviões.	101
B.2	Grupo de imagens Vegetação.	102
B.3	Grupo de imagens Aves.	103
B.4	Grupo de imagens Garrafas.	104
B.5	Grupo de imagens Camelos.	105
B.6	Grupo de imagens Carros.	106
B.7	Grupo de imagens Faces.	107
B.8	Grupo de imagens Flores.	108
B.9	Grupo de imagens Folhas.	109
B.10	Grupo de imagens Futebol.	110
B.11	Grupo de imagens Guitarras.	111
B.12	Grupo de imagens Casas.	112
B.13	Grupo de imagens Motocicletas.	113
B.14	Grupo de imagens Outono.	114
B.15	Grupo de imagens Planetas.	115
B.16	Grupo de imagens Pôr-do-sol.	116
C.1	Resultado da classificação da imagem 1 para RGB com 16 <i>bins</i>	118
C.2	Resultado da classificação da imagem 2 para RGB com 32 <i>bins</i>	119
C.3	Resultado da classificação da imagem 3 para RGB com 64 <i>bins</i>	120
C.4	Resultado da classificação da imagem 4 para HSV com 16 <i>bins</i>	121
C.5	Resultado da classificação da imagem 5 para HSV com 32 <i>bins</i>	122
C.6	Resultado da classificação da imagem 6 para HSV com 64 <i>bins</i>	123
C.7	Resultado da classificação da imagem 7 para YCbCr com 16 <i>bins</i>	124
C.8	Resultado da classificação da imagem 8 para YCbCr com 32 <i>bins</i>	125
C.9	Resultado da classificação da imagem 9 para YCbCr com 64 <i>bins</i>	126
C.10	Resultado da classificação da imagem 10 para LBP.	127

C.11 Resultado da classificação da imagem 11 para <i>Waveletes</i>	128
C.12 Resultado da classificação da imagem 12 para Momentos Invariantes de Hu. .	129
C.13 Resultado da classificação da imagem 13 para combinação de cor.	130
C.14 Resultado da classificação da imagem 14 para combinação de textura.	131
C.15 Resultado da classificação da imagem 15 para combinação de cor e textura. .	132
C.16 Resultado da classificação da imagem 16 para combinação de cor e forma. . .	133
C.17 Resultado da classificação da imagem 17 para combinação de textura e forma.	134
C.18 Resultado da classificação da imagem 18 para combinação cor, textura e forma.	135
C.19 Resultado da classificação da imagem 19 para a melhor combinação.	136
C.20 Resultado da classificação da única imagem para RGB com 16 <i>bins</i>	137
C.21 Resultado da classificação da única imagem para RGB com 32 <i>bins</i>	138
C.22 Resultado da classificação da única imagem para RGB com 64 <i>bins</i>	139
C.23 Resultado da classificação da única imagem para HSV com 16 <i>bins</i>	140
C.24 Resultado da classificação da única imagem para HSV com 32 <i>bins</i>	141
C.25 Resultado da classificação da única imagem para HSV com 64 <i>bins</i>	142
C.26 Resultado da classificação da única imagem para YCbCr com 16 <i>bins</i>	143
C.27 Resultado da classificação da única imagem para YCbCr com 32 <i>bins</i>	144
C.28 Resultado da classificação da única imagem para YCbCr com 64 <i>bins</i>	145
C.29 Resultado da classificação da única imagem para LBP.	146
C.30 Resultado da classificação da única imagem para <i>Waveletes</i>	147
C.31 Resultado da classificação da única imagem para Momentos Invariantes de Hu.	148
C.32 Resultado da classificação da única imagem para combinação de cor.	149
C.33 Resultado da classificação da única imagem para combinação de textura. . . .	150
C.34 Resultado da classificação da única imagem para combinação de cor e textura.	151
C.35 Resultado da classificação da única imagem para combinação de cor e forma.	152
C.36 Resultado da classificação da única imagem para combinação de textura e forma.	153
C.37 Resultado da classificação da única imagem para combinação cor, textura e forma.	154
C.38 Resultado da classificação da única imagem para a melhor combinação. . . .	155

Lista de Tabelas

2.1	Resumo dos sistemas apresentados.	37
4.1	Parâmetros de τ_2 selecionados para cada extrator.	63
4.2	Valores de precisão e cobertura para algumas classes.	63
4.3	Valores de τ_2 selecionados para cada rede GHSOM.	75

Capítulo 1

Introdução

Nesta dissertação são investigadas técnicas de extração de características (baseadas em cor, textura e forma) e combinação de classificadores neurais do tipo GHSOM (*Growing Hierarchical Self-Organizing Maps*), com aplicação ao problema de recuperação de imagens da Internet a partir de conteúdo. Os propósitos deste capítulo são apresentar a motivação para o desenvolvimento de Sistemas de Recuperação de Imagens Baseada em Conteúdo (SRIBCs), as principais limitações das soluções atuais e os principais objetivos desta pesquisa.

1.1 Motivação

Um dos propósitos da área da Visão Computacional é o desenvolvimento de teorias e métodos que objetivam a extração automática de informações úteis a partir de imagens.

Nos últimos anos, tem-se visto um aumento nas coleções de imagens digitais. A cada dia, um grande número de organizações vem coletando e armazenando imagens, como também as adicionando massivamente à *World Wide Web* (Web). Porém, ao mesmo tempo em que esse imenso volume de imagens é armazenado, existe dificuldade de acesso, uma vez que geralmente as imagens não são organizadas de forma a permitir buscas e recuperação. Por outro lado, hoje em dia, a recuperação de imagens, provenientes tanto da Internet quanto de bases de dados privados, é de fundamental importância para diferentes grupos de usuários. Entre esses grupos, podemos citar jornalistas, policiais, cientistas, médicos, *designers*, e outros. Por exemplo, considere um jornalista que está fazendo uma reportagem sobre raças de cães e deseja ilustrar sua matéria com imagens de diferentes raças, através de uma imagem de exemplo um sistema recuperaria imagens com diferentes tipos de cães.

No início dos anos 90 foram identificados dois principais problemas na recuperação de imagens. O primeiro problema está no esforço para se fazer as anotações manuais referentes a cada imagem [RHC97], descritores textuais utilizados para a recuperação. Já o segundo problema se refere à dificuldade para extrair informações das imagens, já que as mesmas possuem um rico conteúdo [SWS⁺00]. Nesse contexto, surgiu a idéia de recuperação de imagens por conteúdo (*Content-Based Image Retrieval* - CBIR), que se caracteriza pela indexação e recuperação de imagens utilizando suas próprias características visuais, como cor, textura e forma. Desde então, as comunidades de Reconhecimento de Padrões, Visão Computacional e afins, vêm se dedicando a desenvolver SRIBCs. Alguns desses sistemas, citados por Huang [RHC97], são: QBIC, PicSOM, VisualSEEk, WebSEEk, Netra, MARS, ART MUSEUM e CAETHML.

Um cenário mais amplo de utilização de um SRIBC seria justamente o de buscas por imagens na Internet efetuadas por um usuário comum: a partir de uma imagem de seu interesse, o usuário poderia solicitar ao sistema imagens que possuem algum tipo de similaridade com a que foi submetida ao sistema. Como exemplificado anteriormente o sistema pode ser aplicado a diversos domínios. Um aplicação desse porte, que possuísse uma execução em tempo real no ambiente da Internet, seria de grande valor para grandes empresas do domínio de buscas na Internet, como Google© e Yahoo!© [DJLW08].

1.2 Descrição do Problema

Mesmo com todos os avanços em sistemas de recuperação de imagens por conteúdo na tentativa de aprimorar as buscas em grandes bancos de imagens, os resultados obtidos ainda não são totalmente satisfatórios [TFa06]. Existem áreas com algumas deficiências no contexto de busca por conteúdo, como a Internet.

A maioria das aplicações investigadas na literatura atual possui resultados de recuperação inferiores a 50% de precisão média, isso para um conjunto de tamanho médio de 1000 imagens [FMWX03; BOLG05]. Os recuperadores que obtiveram os melhores resultados possuíam as seguintes características: classificavam as imagens em um número muito pequeno de classes (por exemplo, florestas, prédios, montanhas, desertos, etc.) ou classificam um número muito pequeno de imagens. Como exemplos, podem-se citar um banco de imagens que possui apenas 500 imagens diferentes, ou um banco de imagens que possui muitas imagens porém poucos rótulos associados às mesmas [PGB01; Edv06].

Desta forma, tais limitações são o ponto de partida desta dissertação que tem como objetivos estudar técnicas de extração de características e desenvolver novos métodos de classificação mais robustos, melhorando a precisão e a cobertura, utilizando combinação de classificadores, em um sistema de busca por conteúdo da *World Wide Web*.

1.3 Objetivos e Relevância

O objetivo geral desta dissertação é avaliar a combinação de uma seleção de métodos de extração de características, baseados em cor, forma e textura, utilizando indexadores baseados em Mapas Auto-Organizáveis Construtivos Hierárquicos (GHSOM) a fim de obter uma maior acurácia na recuperação. A seleção do método de indexação GHSOM se deu devido às suas vantagens inerentes, como treinamento não supervisionado, a criação de uma estrutura de árvore de indexação e de agrupamentos dinâmicos. Os sistemas atuais geralmente são aplicados a um domínio específico ou com uma quantidade determinada de imagens. Com base neste fato, o sistema proposto nesta dissertação será aplicado a um domínio de imagens genéricas e que será constantemente atualizada.

Apesar do grande número de aplicações existentes na literatura atual, poucos são aplicados a *World Wide Web*. Além disso, a utilização de uma técnica de combinação de classificadores é pouco abordada para esse tipo de aplicação. Nesse sentido, pretendeu-se atingir os seguintes objetivos específicos:

- investigar técnicas simples e mais utilizadas em SRIBC na extração de características de imagens, de forma que possam ser utilizadas em um sistema de recuperação de imagens na Internet;
- investigar a estrutura e funcionamento da GHSOM, de forma a aumentar o desempenho do sistema na indexação de imagens;
- investigar a aplicação de combinação de classificadores no problema de recuperação de imagens por conteúdo;
- desenvolver um SRIBC que integre todos os itens acima e aplicá-lo em um problema de recuperação de imagens na Internet.

1.4 Estrutura da Dissertação

Esta dissertação é dividida em cinco capítulos.

No Capítulo 2, são descritos os principais pontos do estado da arte sobre sistemas de recuperação de imagens baseada em conteúdo, também são apresentados alguns sistemas dispostos na literatura e suas funções. Além disso, são destacadas as principais características dos SRIBC comuns, os conceitos fundamentais e trabalhos relacionados para as técnicas mais utilizadas na extração de características de imagens, indexação e combinação de classificadores. Por fim, é apresentada uma tabela que resume as características de cada sistema pesquisado, mostrando suas formas de extrações, método de consulta, estrutura de indexação e medida de similaridade.

No Capítulo 3, são apresentadas de forma detalhada uma descrição do projeto e implementação de um SRIBC. Além disso, é apresentada a arquitetura geral do sistema, como também os detalhes dos módulos internos, de extração, classificação, combinação e recuperação de imagens. Também são descritos os dois modos de funcionamento do sistema: modo de busca, que utiliza um robô de busca para localizar e armazenar informações sobre imagens da Internet; e o modo de consulta, que possibilita ao usuário final, utilizar o sistema através de uma interface Web para a recuperação de imagens.

No Capítulo 4, são relatados os principais experimentos envolvendo diferentes tipos de extratores de características e tipos de combinação de classificadores aplicados ao problema de recuperação de imagens com base em conteúdo. Os experimentos foram subdivididos em três etapas: a primeira teve como objetivo encontrar os melhores parâmetros para o treinamento de cada uma das redes GHSOM, utilizadas na classificação; a segunda, que faz uso dos melhores parâmetros determinados pelo experimento anterior, teve como objetivo um estudo comparativo entre classificações simples e combinadas; finalmente, a terceira etapa envolveu testar o sistema como um todo, sendo necessário um conjunto de testadores, os quais realizaram sessões de recuperação de imagens não rotuladas com vistas a uma avaliação subjetiva da precisão dos diferentes métodos implementados.

No Capítulo 5, apresentam-se as conclusões obtidas a partir deste estudo, as principais contribuições e os trabalhos futuros que podem ser derivados a partir do exposto ao longo desta dissertação.

Por fim, o Apêndice A traz conceitos sobre a implementação e a configuração do Robô de Busca que foi utilizado para buscar e armazenar informações sobre imagens da Internet.

O Apêndice B contém amostras de imagens utilizadas no experimento 2 descrito no Capítulo 4.

Capítulo 2

Revisão Bibliográfica

Nas últimas décadas, o rápido desenvolvimento de tecnologias da informação e comunicação, bem como o advento da *Web*, tem acelerado o crescimento de conteúdos digitais, como por exemplo, coleções de imagens. A partir da necessidade de organização e acesso a essas coleções, alguns sistemas [Dow93; HMR96; MM99; NB94; APS96; SC97] foram criados fazendo uso da idéia de recuperação de imagens por conteúdo. A intenção foi a de reduzir as dificuldades existentes até então, considerando que os métodos tradicionais para organização dessas coleções requerem anotações manuais no tocante ao conteúdo das imagens. A expressão “recuperação de imagens por conteúdo” pode ser vista como um processo de encontrar uma imagem ou grupo de imagens de interesse em uma coleção, em que as imagens são indexadas a partir de suas próprias características visuais, como cor, forma e textura. Entretanto, há também abordagens que utilizam outros atributos para recuperação por conteúdo, como alguns metadados associados às imagens (por exemplo, recuperar imagens que foram adquiridas numa determinada faixa de datas).

Deste modo, nas subseções a seguir são apresentadas as etapas fundamentais para o funcionamento dos SRIBCs (Sistemas de Recuperação de Imagens Baseados em Conteúdo). São destacadas as etapas de extração de características, trabalhos relacionados a indexação e abordagens propostas para recuperação. Em adição as etapas tradicionais, foram abordadas também técnicas de combinação de classificadores em sistemas de recuperação de imagens por conteúdo. A maioria dos SRIBCs possuem as seguintes características em comum [Sub98]:

- Extração de características: etapa responsável pela extração de características visuais das imagens, que são geralmente cor, forma e textura;
- durante a construção do sistema, as características extraídas podem ser organizados

em algum tipo de Estrutura de Dados Multidimensional (EDM), com o objetivo de reduzir o espaço de busca. Esta etapa é conhecida como Indexação e é equivalente à etapa de Aprendizagem em outros sistemas de Visão Computacional.

- a etapa de Reconhecimento, geralmente utiliza a distância Euclidiana como medida de dissimilaridade.

2.1 Extração de Características

Nesta seção são apresentadas as principais características e técnicas de extração e representação utilizadas nos SRIBCs. Além disso, são apresentados trabalhos relacionados a cada forma de extração e suas técnicas utilizadas.

2.1.1 Cor

Nas duas próximas subseções, inicialmente são descritos conceitos fundamentais e técnicas utilizadas pelo sistema desenvolvido nesta dissertação, em seguida, é apresentada uma revisão bibliográfica contendo trabalhos relacionados.

Fundamentos e técnicas utilizadas

Uma das características mais utilizadas na recuperação de imagens por conteúdo é a cor, a qual é um descritor poderoso que freqüentemente simplifica a identificação do objeto e a extração de uma cena [GW02]. Em uma imagem, as cores podem ser representadas em diferentes espaços de cor, desta forma, cada espaço de cor fornece um conjunto diferente de características para uma mesma imagem. O espaço de cor pode ser visto como uma representação matemática para cores, através de tuplas de números, formada por três ou quatro elementos, onde cada elemento representa um canal.

O espaço de cor RGB que é a abreviatura do sistema de cores aditivas formado por Vermelho (*Red*), Verde (*Green*) e Azul (*Blue*). O modelo de cores RGB é baseado na teoria de visão colorida tricromática, e no triângulo de cores de Maxwell. O modelo RGB é utilizado como padrão para apresentação de cores na Internet [GW02]. Uma cor no modelo de cores RGB pode ser descrita pela indicação da quantidade de vermelho, verde e azul.

Outro espaço de cor utilizado é o HSV, abreviatura para o sistema de cores formadas pelas componentes Tonalidade (*Hue*), Saturação (*Saturation*) e Valor (*Value*). O espaço de

cor é definido de acordo com seus três parâmetros:

- Tonalidade: identifica o tipo da cor, abrangendo todas as cores do espectro, desde o vermelho até o violeta, mais o magenta.
- Saturação: também chamada de pureza - quanto maior esse valor, mais acinzentada parecerá a cor.
- Valor, ou brilho: define a intensidade do brilho da cor.

YCbCr é também um espaço de cor bastante utilizado em sistemas de vídeo e fotografia digital. Y é o componente de luminância em escala de cinza e Cb e Cr são as componentes de crominância azul e vermelho, respectivamente. O sistema YCbCr é usado em alguns algoritmos de compressão de imagens (como o do formato JPEG) na medida em que permitem que seja descartada uma parte das informações de Cb e Cr, tornando os arquivos de imagens menores mas sem grandes perdas visuais.

As cores freqüentemente são representadas por meio de histogramas. Em imagens coloridas, esta representação pode ser realizada calculando-se o histograma para cada um de seus componentes. Assim, uma imagem no espaço de cor RGB, possuirá três histogramas, uma para cada componente. Na Figura 2.1 é apresentada uma imagem em níveis de cinza e seu respectivo histograma em forma de gráfico de barras. Outra forma de representação do modelo RGB baseia-se em um sistema de coordenadas cartesianas, sendo que o subespaço de cores de interesse é o cubo, no qual os valores de RGB estão nos três cantos do cubo.

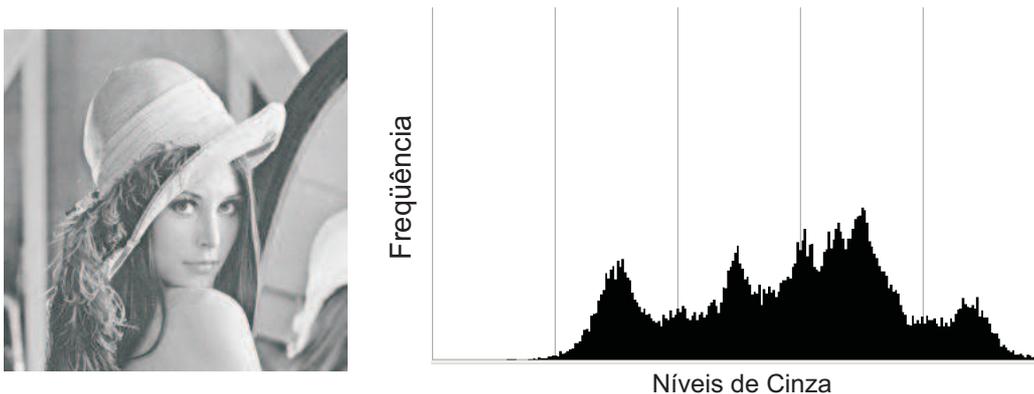


Figura 2.1: Histograma para 256 níveis de cinza.

Histograma de cor é um dos descritores mais comumente utilizado em recuperação de imagens [GW02]. O algoritmo de extração do histograma de cor pode ser dividido em três passos principais: partição do espaço de cor em níveis de intensidade, associação de cada

nível a um *bin* do histograma, contagem do número de pixels para cada nível da imagem e armazenamento desta contagem no correspondente *bin* do histograma. Os descritores que utilizam Histograma de Cor são invariantes à translação e rotação [TFa06].

Trabalhos Relacionados

Além do histograma de cor, muitas outras representações de características de cores são aplicadas na recuperação de imagens, como o Momento da Cor [SO95] e o Conjunto de Cor [SC95].

Segundo Stricker e Orengo [SO95] a aproximação por Momento da Cor diz que toda distribuição de cor pode ser caracterizada por seus momentos. Momento é uma técnica robusta para decomposição de uma imagem de forma arbitrária em um conjunto finito de características invariantes quanto à escala, posição e orientação. Em termos práticos, a utilização de Momento para o reconhecimento de imagens requer a seleção de um subconjunto de valores de momento que contenha informações suficientes para caracterizar unicamente cada imagem. Usualmente, a média (primeira ordem), variância (segunda) e *skewness* (terceira) são usadas no Momento da Cor para formar o vetor de característica. Estes momentos são definidos, respectivamente, como $E_i = (1/N) \sum_{j=1}^N p_{ij}$, $\sigma_i = \sqrt{(1/N) \sum_{j=1}^N (p_{ij} - E_i)^2}$ e $s_i = \sqrt[3]{(1/N) \sum_{j=1}^N (p_{ij} - E_i)^3}$, onde p_{ij} é o valor do i -ésimo componente de cor de um pixel j da imagem e N é o número de pixels da imagem.

Para facilitar a busca em coleções de imagens de grande escala, Stricker e Orengo [SC95] propuseram o uso do conceito de Conjunto de Cor como uma aproximação para o histograma de cor. Inicialmente o espaço de cor é mapeado em um espaço perceptivamente uniforme e então se quantiza o espaço em uma estrutura de tamanho \mathbf{M} . O Conjunto de Cor é definido como uma seleção das cores do espaço de cor quantizado. Uma das vantagens está na forma de armazenamento dos vetores de características, que são alocados em uma estrutura do tipo árvore binária, a qual permite buscas muito mais rápidas.

Apesar de serem ótimas representações para as características de cor, as técnicas citadas acima possuem como desvantagem o tempo de processamento da etapa de extração de características. Um estudo mais detalhado sobre descritores de cor aplicados a recuperação na Web é apresentado por Penatti et al. [PT08]. Como o sistema proposto nesta dissertação tem a Web como cenário de aplicação final, verifica-se que, para permitir tempos de acesso aceitáveis aos usuários, é necessário que tenha um tempo de processamento rápido. Portanto, no tocante a características de cor foram escolhidas para implementação do sistema apenas

técnicas que utilizam histogramas em diferentes espaços de cores.

2.1.2 Forma

Na subseção a seguir são apresentados fundamentos a respeito da extração de características de forma, conceitos básicos e o detalhamento das técnicas utilizadas no desenvolvimento do sistema desta dissertação.

Fundamentos

Em reconhecimento de padrões e áreas afins, a forma é a característica mais importante para identificar e distinguir objetos [ZL04]. Para sistemas de recuperação de imagens por conteúdo, a técnica de extração de características de forma requer um pré-processamento das imagens, através da segmentação das mesmas em regiões de interesse.

Logo após a identificação das formas dos objetos que compõem uma imagem, é necessário descrevê-las com base nas propriedades invariantes a escala, rotação e translação.

Em geral, as representações de forma podem ser divididas em duas categorias: baseadas em contorno e baseadas em regiões inteiras. Esta classificação leva em conta se as características da forma obtida foram de apenas um contorno ou foram de toda uma região. Esta subdivisão leva em consideração se a forma é representada como um todo ou por segmentos/seqções. Uma técnica bastante utilizada é a dos Momentos Invariantes de Hu [RHC97].

Momentos Invariantes de Hu

Momentos têm se apresentado como uma técnica robusta para decomposição de uma imagem de forma arbitrária em um conjunto finito de características, chamados de momentos invariantes.

Os momentos fornecem descrições globais com propriedades invariantes e com vantagens de serem uma descrição compacta que é robusta em relação a ruídos. Desta forma, a técnica dos Momentos tem provado ser popular e bem sucedida em muitas aplicações.

Hu propôs sete momentos invariantes à orientação, escala e translação que podem ser utilizados como descritores de forma [Hu62]. Estes descritores são calculados por meio de expressões simples, tendo como base a seguinte equação [NA02]:

$$m_{pq} = \sum_x \sum_y x^p y^q I(x, y) \quad (2.1)$$

m_{pq} é o momento bi-dimensional da imagem $I(x, y)$, de ordem $p + q$, sendo p e q números naturais.

A invariância à translação é obtida a partir do momento central, dado por:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})_p (y - \bar{y})_q I(x, y) \quad (2.2)$$

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (2.3)$$

em que \bar{x} e \bar{y} são os centros de massa (centróides) da imagem. A invariância à escala é adquirida através da seguinte equação:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (2.4)$$

em que $\gamma = \frac{p+q}{2} + 1 \quad \forall p + q \geq 2$

Por último, os momentos de Hu são obtidos a partir de sete expressões invariantes à rotação:

$$M1 = \eta_{20} + \eta_{02} \quad (2.5)$$

$$M2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (2.6)$$

$$M3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (2.7)$$

$$M4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (2.8)$$

$$\begin{aligned} M5 = & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) + ((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2) \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \end{aligned} \quad (2.9)$$

$$M6 = (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (2.10)$$

$$\begin{aligned} M7 = & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) \\ & + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})(3(\eta_{12} + \eta_{30})^2 - (\eta_{21} + \eta_{03})^2) \end{aligned} \quad (2.11)$$

Trabalhos relacionados

Além da técnica de Momentos Invariantes de Hu, existem outros trabalhos que utilizam diferentes técnicas para descrever as características de forma. Os Descritores de Fourier [Pra91] são uma delas, os quais consideram apenas os contornos dos objetos e tem como vantagem a invariância à orientação, escala e translação. Um problema com os descritores é que sua computação é lenta, desta forma não são de interesse quando a utilização estiver atrelada à produção de resultados em tempo real. Outra técnica é a Programação Dinâmica (DP - *Dynamic Programming*), que utiliza casamento de formas, onde as formas são aproximadas como seqüências de segmentos côncavos e convexos, proposta por Petrakis [PDM02]. Outra técnica bastante promissora é a de escala tensor [AMTF07], que é uma representação que unifica a representação local de estruturas de espessura, orientação e anisotropia. Apesar dos bons resultados apresentados pela técnica, os testes realizados não utilizaram imagens de conteúdo complexo.

2.1.3 Textura

A subseção a seguir apresenta conceitos fundamentais sobre textura e descreve as técnicas utilizadas na aplicação desenvolvida nesta dissertação. Em seguida são apresentados trabalhos relacionados que utilizam diferentes métodos de extração de característica de textura.

Fundamentos

Não existe uma definição clara para textura, entretanto autores concordam em defini-la como propriedades da imagem que são caracterizadas pela existência de primitivas básicas, cuja distribuição espacial criam alguns padrões visuais definidos em termos de granularidade, direcionalidade e repetitividade [TJ93].

Textura refere-se a um padrão visual que tem propriedades de homogeneidade que não resultam da presença apenas de uma simples cor ou intensidade. Tais padrões podem ser resultado de propriedades físicas da superfície do objeto ou da reflexão da luz tal como a cor da superfície. Para a visão humana, o reconhecimento de textura é um processo fácil, o que não acontece para processos automáticos (máquinas) em que esta tarefa exige técnicas computacionais complexas.

Em SRIBCs estas características são muito úteis, já que a maioria dos objetos (por exemplo, ver Figura 2.2) exibem texturas.

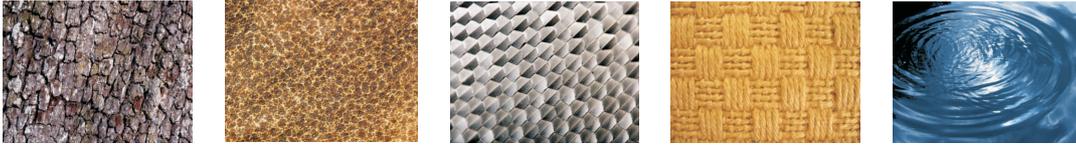


Figura 2.2: Amostras de texturas.

A análise das texturas tem como objetivo estabelecer o relacionamento de vizinhança dos elementos da textura e seu posicionamento em relação aos demais (conectividade), o número de elementos por unidade espacial (densidade) e a sua regularidade (homogeneidade) [CT03].

Existem diferentes métodos de extração e representação de texturas, os quais podem ser classificados em estáticos, geométricos e de processamento de sinais [TJ93].

- *Métodos estáticos*: a textura é definida como distribuições espaciais de tons de cinza;
- *Métodos geométricos*: se caracterizam como sendo compostos de “elementos da textura” ou primitivas;
- *Métodos de processamento de sinais*: utilizam a análise da frequência da imagem para classificar a textura.

A seguir são descritas duas técnicas utilizadas na extração de texturas: LBP (*Local Binary Patterns*) e as *Wavelets*.

LBP - *Local Binary Patterns*

A classificação de texturas utilizando *Local Binary Patterns*, invariante a rotação, é teoricamente simples e bastante eficiente. O método LBP é baseado no reconhecimento de padrões binários, a extração da textura da imagem é realizada através de um operador e organizadas em um histograma de ocorrência desses padrões [OPM02].

O operador de textura sugerido permite a detecção dos padrões binários em uma vizinhança circular em qualquer espaço angular e em qualquer resolução. O operador utilizado é formado por um conjunto de P vizinhos simétricos e em um círculo de raio R , desta forma obtemos $LBP_{P,R}$. O parâmetro P regula a quantização de espaços angulares enquanto R regula a resolução do operador no espaço.

O histograma de ocorrência discreta dos padrões computados sobre uma imagem também mostra-se uma característica de textura poderosa. Ao computar o histograma de ocorrência, as visões estruturais e estatísticas são combinadas de forma eficiente, o operador LBP detecta

as microestruturas (linhas, cantos, pontos, superfícies, etc.), enquanto o histograma estima sua distribuição.

O operador LBP, invariante a rotação e níveis de cinza, é definido através de uma textura T em uma vizinhança de uma imagem monocromática como a distribuição de níveis de cinza em P pixels, com $P > 1$: $T = t(g_c, g_0, \dots, g_{p-1})$, em que o nível de cinza g_c corresponde ao valor de cinza do pixel central, e os valores $g_p (p = 0, \dots, P - 1)$ correspondem aos valores de cinza dos P pixels (vizinhos) igualmente espaçados em um círculo de raio $R (R > 0)$

Se a posição do pixel central g_c for $(0, 0)$, então as coordenadas de g_p serão dados por $(-R \text{sen}(2\pi p/P), R \text{cos}(2\pi p/P))$, como ilustra a Figura 2.3. O valor do tom de cinza dos vizinhos que não estiverem centrados nos pixels são calculados através de interpolação.

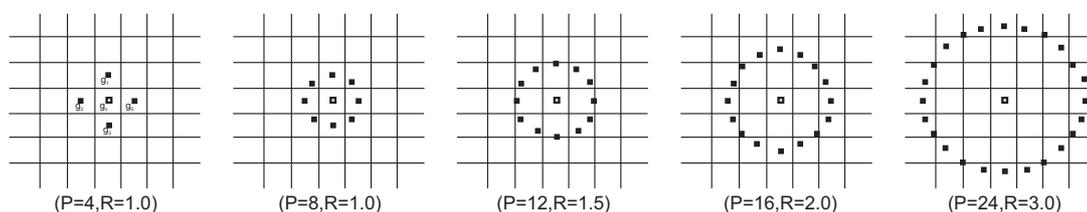


Figura 2.3: Conjunto de vizinhos igualmente espaçados na superfície circular para diferentes P e R .

Para obter algum valor de textura da imagem subtrai-se o valor do pixel central dos valores de cada pixel de sua vizinhança, conforme indicado pela Equação 2.12, assim certo contexto local é criado.

$$T = t(g_0 - g_c, g_1 - g_c, \dots, g_{p-1} - g_c) \quad (2.12)$$

O operador acima é sensível a variações na escala de cinza, então para se obter invariância de textura independente de alterações na escala de cinza de uma imagem consideramos apenas os sinais dos resultados ao invés de seu valor exato, como na Equação 2.13:

$$T = t(s(g_0 - g_c), s(g_1 - g_c), \dots, s(g_{p-1} - g_c)) \quad (2.13)$$

em que,

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases} \quad (2.14)$$

Assim os valores obtidos de cada vizinhança em uma imagem qualquer tornam-se confiáveis mas que não dependem de variações que se apliquem a toda a imagem. Finalmente, atribuindo a potência 2^p para cada sinal $s(x)$, obtém-se um valor único que identifica cada estrutura da textura obtendo o operador:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \tag{2.15}$$

O operador $LBP_{P,R}$ produz 2^P valores diferentes correspondentes aos 2^P diferentes padrões que podem ser formados pelos P pixels vizinhos, quando uma imagem sofre uma rotação os P pixels irão orbitar no perímetro do círculo ao redor de g_c , como g_0 sempre é associado ao pixel à direita de g_c uma rotação dos valores naturalmente (mas não necessariamente) irá resultar em uma mudança no valor resultante do operador $LBP_{P,R}$. Para remover este efeito sobre a rotação basta assinalar um identificador único para cada padrão. Definido através da seguinte equação:

$$LBP_{P,R}^i = \min \{ROR(LBP_{P,R}, i) \mid i = 0, 1, \dots, P - 1\}; \tag{2.16}$$

onde $ROR(x, i)$ faz uma rotação nos bits do numero x , i vezes, isso significa achar o número que tenha a maior quantidade de zeros nos bits mais significativos, na Figura 2.4 é apresentada a forma de seleção do padrão.

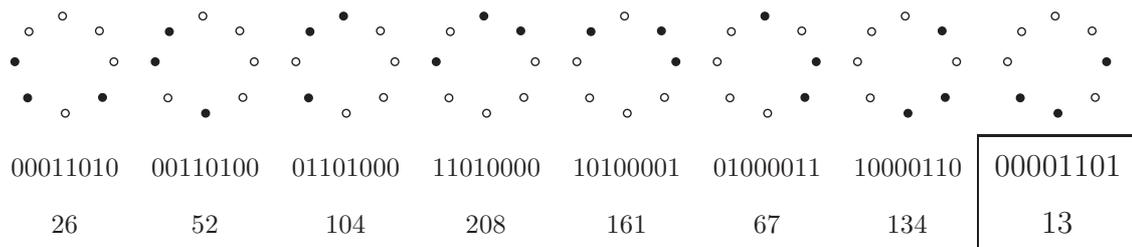


Figura 2.4: Rotação buscando o menor número inteiro entre os bits.

Transformações Wavelets (WT)

Transformadas matemáticas são utilizadas para extrair características do sinal que não são normalmente visualizadas no sinal puro. No processo, toma-se o sinal inicial no domínio do tempo, aplicam-se as transformadas, sendo capturadas informações importantes, principalmente o espectro da frequência - o qual corresponde ao sinal processado.

As Wavelets foram desenvolvidas com base na Matemática, Física Quântica e na Estatística [Add02]. As transformadas decompõem o sinal em diferentes componentes de frequência, desta forma, permitindo estudar cada componente com uma resolução relacionada à sua escala [NA02]. Possuem grande vantagem em relação à transformada de Fourier principalmente quando há descontinuidades do sinal.

Diferentemente da Transformada de Fourier, as transformadas Wavelets têm um conjunto infinito de funções base; enquanto as de Fourier têm apenas as funções seno e cosseno.

As transformadas Wavelets fornecem uma representação tempo x frequência, ou seja, com elas é possível determinar quais são e onde estão localizados os componentes de frequência. Tal representação é muito útil, visto que freqüentemente é necessário conhecer a ocorrência de um componente espectral num determinado instante. Nesta abordagem, é utilizado método de multi-resolução. Os algoritmos que implementam essas transformadas processam dados em diferentes escalas e resoluções, permitindo que sejam mostrados tanto global como os detalhes das imagens. O algoritmo processa a imagem da seguinte forma:

1. Sinal de entrada é recebido;
2. Sinal de entrada é filtrado: são utilizados o filtro passa-alta e o filtro passa-baixa; esses filtros dividirão o sinal em dois sinais menores, um de frequências maiores que um limiar e outro de frequências menores.

A aplicação dos filtros passa alta e passa baixa é realizada a partir de uma iteração nas linhas e colunas da matriz que representa a imagem. A cada iteração, são geradas três sub-imagens (os coeficientes Wavelets), extraídos em diferentes orientações - cada uma responsável por um ângulo na variação de intensidade. Na Figura 2.5 pode-se observar o conceito da transformada, aplicada em uma imagem.

A imagem é processada pela transformação e quatro novas imagens são produzidas pela técnica descrita acima:

1. LL - Imagem obtida a partir da passagem do filtro de passa-baixa nas linhas e colunas da matriz da imagem;
2. LH - Imagem obtida a partir da passagem do filtro de passa-baixa nas linhas e passa-alta nas colunas da matriz da imagem;
3. HL - Imagem obtida a partir da passagem do filtro de passa-alta nas linhas e passa-baixa nas colunas da matriz da imagem;



Figura 2.5: Transformada wavelet. Inicia-se com a imagem original e (a) após a transformação, obtêm-se imagens em outras escalas: (b) LL, (c) HL, (d) LH e (e) HH.

4. HH - Imagem obtida a partir da passagem do filtro de passa-alta nas linhas e colunas da matriz da imagem.

As Wavelets geralmente abordadas são as de *Haar* e *Daubechies* [WWFXW98]. Os filtros passa-baixa e passa-alta de *Haar* e *Daubechies* podem ser observados nas Equações 2.17 e 2.18 respectivamente.

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \quad (2.17)$$

$$\begin{bmatrix} \frac{(\sqrt{3}+1)}{4\sqrt{2}} & \frac{\sqrt{3}(\sqrt{3}+1)}{4\sqrt{2}} & \frac{(\sqrt{3}-1)}{4\sqrt{2}} & \frac{\sqrt{3}(\sqrt{3}-1)}{4\sqrt{2}} \\ -\frac{\sqrt{3}(\sqrt{3}-1)}{4\sqrt{2}} & -\frac{(\sqrt{3}-1)}{4\sqrt{2}} & \frac{\sqrt{3}(\sqrt{3}+1)}{4\sqrt{2}} & -\frac{(\sqrt{3}+1)}{4\sqrt{2}} \end{bmatrix} \quad (2.18)$$

Trabalhos relacionados

Dos métodos estáticos existentes na literatura o LBP (*Local Binary Patterns*) se destaca, já que trata-se de um método de classificação de textura invariante a rotação. O trabalho de Takala *et al.* [TAP05] apresenta uma técnica de extração de textura, baseada em blocos, para um SRIBC em que o LBP foi utilizado como descritor das imagens.

No trabalho proposto por Ma e Manjunath [MM99], as texturas são representadas por um vetor de características contendo a média normalizada e o desvio padrão da série de

transformadas Wavelets de Gabor. Essas características são utilizadas em SRIBC chamado NETRA, além das texturas o sistema também utiliza outros métodos de extração.

Outro SRIBC o QBIC [NB94], também utilizada extração de características baseada em textura. O métodos de extração utiliza versões modificadas de *coarseness* (fineza), *contrast* (contraste) e *directionality* (direcionalidade) proposta por Tamura [TMY78]. Métodos que são baseados na percepção visual humana.

2.2 Estruturas de Indexação

Esta seção está organizada em duas subseções: na primeira são apresentados conceitos e a técnica de indexação utilizada no sistema desenvolvido nesta dissertação. Na segunda subseção são apresentados trabalhos relacionados e outras técnicas utilizadas na indexação.

Fundamentos

Depois de extraídas as características de uma imagem é necessário representar essas informações de forma que seja possível uma recuperação. Em SRIBC, essas características são comumente representadas por meio de vetores de características e estruturas de dados multidimensionais.

Os vetores de características são a chave principal, utilizada para decidir se duas imagens são ou não similares.

Devido ao fato de grande parte das técnicas de extração de características de imagens produzirem vetores de grande dimensão (tipicamente da ordem de 10^2 [RHC97]), os sistemas de recuperação de imagens por conteúdo precisam ser escaláveis quando operando com grandes coleções de imagens, e eficientes (rápidos) na busca de imagens baseando-se nas características extraídas.

Para desenvolver um recuperador de imagem baseado em conteúdo verdadeiramente escalável para uma grande quantidade de imagens é necessário explorar técnicas multidimensionais eficientes de indexação. Algumas comunidades de pesquisa contribuíram com essa área, como a Geometria Computacional, Gerenciamento de Banco de Dados e Reconhecimento de Padrões.

Entre os métodos existentes podemos destacar os SOMs (Self-Organizing Maps) [Koh82], que são um tipo de redes neurais passível de utilização como estrutura de indexação de dados multidimensionais. Esse tipo de rede neural é utilizada como ferramenta na construção de

estruturas de árvore de indexação em sistemas de recuperação de imagens. As vantagens da utilização do SOM estão na habilidade do aprendizado não-supervisionado, na criação de grupos dinâmicos e capacidade de suporte a medidas de similaridades arbitrárias [RHC97]. Outra técnica de destaque na literatura, trata-se de um aprimoramento do algoritmo tradicional do SOM, chamado de (*Growing Hierarchical Self-Organizing Maps*), devido a sua capacidade de adaptação na fase de treinamento, esse algoritmo além das propriedades básicas de um SOM, possui aprimoramentos, como a capacidade de inserção de novos neurônios em regiões com grandes diferenças entre os padrões e a subdivisão dos padrões em hierarquia, isto é, a busca por padrões nos mapas é reduzida [RMD02]. A seguir são apresentados alguns conceitos sobre redes neurais e um maior detalhamento sobre redes do tipo SOM.

Introdução às Redes Neurais Artificiais

O desenvolvimento das redes neurais artificiais (RNA) foi motivado pelo funcionamento do cérebro humano, o qual tem a capacidade de processar informações de uma forma inteiramente diferente de um computador convencional. O cérebro humano é um sistema de processamento de informação muito complexo, não-linear e paralelo [RN03].

Segundo Haykin [HE01], uma rede neural é um processador paralelamente distribuído que formado por unidade de processamento simples, possui uma habilidade de armazenar conhecimento e torná-lo disponível para o uso.

A unidade fundamental de processamento de uma Rede Neural Artificial é o Neurônio Artificial. Seu funcionamento é inspirado no do neurônio biológico, tendo implementação computacional normalmente eficiente.

Um neurônio artificial é uma entidade que recebe do ambiente externo, através de suas entradas, impulsos ou estímulos (x_i) que são combinados pelo somatório dos produtos de todos os impulsos com o peso de sua respectiva entrada (w_i) de modo a se determinar um sinal de entrada ($i = \sum_{i=1}^n x_i w_i$) para o neurônio [HE01].

A entrada é submetida a uma função que determina o sinal de saída do neurônio, que é denominada função de ativação. Esta função compara o sinal de entrada (i) do neurônio com um limiar pré-estabelecido (t) e gera como saída o valor 1 nos casos em que $i > t$, ou 0 nos casos contrários, conforme a Equação 2.19.

$$f(i) = \begin{cases} 1, & i > t \\ 0, & i \leq t. \end{cases} \quad (2.19)$$

Em uma rede neural pode existir uma ou múltiplas camadas de neurônios. Por exemplo, numa rede que possua três camadas, a primeira camada é a de entrada, onde as unidades (neurônios) recebem os padrões, a segunda camada é a intermediária (escondida), onde é realizado o processamento e extração de padrões. Por último a terceira camada, a camada de saída que apresenta o resultado final.

As redes neurais podem ser estruturadas em dois tipos:

- Feedforward: estrutura em que a saída de um neurônio de uma determinada camada é utilizada como entrada para os neurônios da camada seguinte.
- Feedback: estrutura em que a saída de um neurônio é utilizada como entrada para neurônios da mesma camada, ou de camadas anteriores.

Assim como na biologia, as redes neurais artificiais possuem também mecanismos de aprendizado, ou seja, a partir de um conjunto de dados utilizados para treinamento, ela passa a ser capaz de tratar dados imprecisos, nebulosos ou ruidosos sem qualquer dificuldade de resposta, além de ser capaz de fazer generalizações.

Em geral, o aprendizado de uma rede neural artificial se dá por meio de ajustes nos seus pesos sinápticos conforme os dados de treinamento vão sendo apresentados à mesma. Há também formas alternativas de aprendizado, como a alteração na topologia da rede, criando-se e removendo-se sinapses conforme seja conveniente para a mesma.

Os dois principais paradigmas de aprendizados seguidos pelas redes neurais artificiais são: supervisionado e o não-supervisionado.

O aprendizado supervisionado ou com professor, similar ao papel de um professor lecionando a um aluno: o desempenho do mesmo é acompanhado, desta forma, na medida em que o aluno erra o professor corrige para que ele passe a acertar. Nas redes neurais o aprendizado supervisionado ocorre quando se fornece um conjunto de dados de treinamento previamente rotulado (o qual é utilizado como guia), para treinar a rede e torná-la capaz de reproduzir o mapeamento entrada-saída do conjunto de treinamento.

Já a aprendizado não-supervisionado (auto-organizável) consiste em fazer com que a rede neural por si só ajuste os pesos de seus neurônios com base na experiência adquirida após ter sido apresentada aos dados que formam o conjunto de treinamento. Este tipo de aprendizado é utilizado pelas redes neurais que se dedicam a tarefas de reconhecimento de padrão por *clustering*, como por exemplo os Mapas Auto-Organizáveis (*Self-Organizing Maps* - SOM).

SOM - Self-Organizing Maps

Os SOM são um tipo de rede neural que utiliza aprendizagem não-supervisionada, o objetivo desse tipo de aprendizagem é descobrir padrões significativos ou características discriminantes nos dados de entrada sem a intervenção de um professor. Este processo consiste em modificar repetidamente os pesos sinápticos de todas as conexões do sistema em resposta a padrões de entrada e de acordo com regras pré-determinadas, até se desenvolver uma configuração final [HE01].

As redes neurais do tipo SOM também conhecidas como Mapas Auto-Organizáveis, são grades neurais baseadas na aprendizagem competitiva, isto é, os neurônios de saída da grade competem entre si para serem ativados ou disparados (Figura 2.6), após apresentado um padrão de entrada a rede, com o resultado que apenas um neurônio de saída está ligado ou

ativo em um instante de tempo.

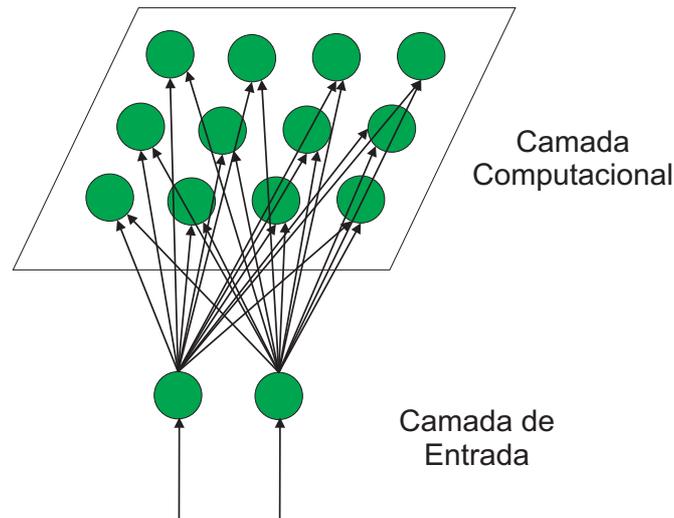


Figura 2.6: Estrutura de um Mapa Auto-Organizável.

O principal objetivo do mapa auto-organizável é transformar um padrão de sinal incidente em seus neurônios em uma maneira topologicamente ordenada [HE01]. O algoritmo básico de treinamento dos mapas auto-organizáveis é subdividido em três processos essenciais:

1. **Competição:** na fase competitiva, os neurônios de saída competem entre si, segundo algum critério, para encontrar um único vencedor para um dado instante.
2. **Cooperação:** na fase de cooperação é definida uma vizinhança topológica de neurônios excitados, para o neurônio vencedor.
3. **Adaptação Sináptica:** na fase adaptativa, os pesos do neurônio vencedor e de sua vizinhança topológica são ajustados.

Resumo do Algoritmo SOM

Segundo Haykin [HE01], o principal objetivo do algoritmo SOM é transformar um padrão de entrada com dimensão arbitrária em um mapa discreto uni ou bidimensional e realizar esta transformação adaptativamente de uma maneira ordenada. Na execução do algoritmo (ver Algoritmo 1) há três passos básicos envolvidos após a inicialização: amostragem, casamento por similaridade e atualização. Estes três passos são repetidos até que a formação do mapa de características esteja completa.

Algoritmo 1 Algoritmo SOM**1. Inicialização:**

Os valores dos pesos são inicializados aleatoriamente, $w_j(0)$, desde que os valores $w_j(0)$ sejam distintos para $j = 1, 2, \dots, l$, onde l representa a quantidade de neurônios do mapa.

2. Amostragem:

Apresente uma amostra x do espaço de entrada ao mapa de neurônios M .

3. Casamento por Similaridade:

Buscar o neurônio vencedor $i(x)$ no tempo n usando os critérios de menor distância Euclidiana:

$$i(x) = \arg \min \|x(n) - w_j\|, \quad j = 1, 2, \dots, l \quad (2.20)$$

4. Atualização:

Ajustar os pesos para o neurônio vencedor i e para sua vizinhança topológica $h_{j,i(x)}(n)$

$$w_j(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x(n) - w_j(n)) \quad (2.21)$$

onde $\eta(n)$ é o parâmetro da taxa de aprendizagem. Ambos $h_{j,i(x)}(n)$ e $\eta(n)$ são variados dinamicamente durante a aprendizagem para obter melhores resultados.

5. Continuação:

Repita a partir do passo 2 até atingir o número de épocas de treinamento k

GHSOM - Growing Hierarchical Self-Organizing Maps

Um problema encontrado na utilização da rede tipo SOM ocorre quando o número de neurônios é muito grande. Neste caso, a busca pelo neurônio vencedor exige muito tempo computacional e a utilização desse tipo de redes neurais em alguns sistemas torna-se muitas vezes inaplicáveis.

Uma solução para a redução desse tempo computacional exigido em grandes mapas é uma reestruturação dos neurônios em uma organização hierárquica, mais especificamente numa estrutura de árvore. Um exemplo de uma rede neural que utiliza essa estratégia é a GHSOM (*Growing Hierarchical Self-Organizing Maps*)[RMD02], que é formada por um conjunto de redes SOMs, onde a profundidade e as dimensões dessas redes crescem automaticamente durante a fase de treinamento.

O início do treinamento de uma GHSOM é similar ao algoritmo de uma SOM tradicional: para cada iteração λ , os neurônios i que possuírem o maior erro de quantização q_i serão identificados.

Após a identificação, uma nova linha ou coluna de neurônios será inserida entre o neurônio com maior q_i e o seu vizinho imediato mais dissimilar.

O cálculo do erro de quantização de um neurônio é realizado através da soma das distâncias Euclidianas entre seu vetor de peso sináptico W_i e os vetores de entrada X_k , representado pela Equação 2.22:

$$q_i = \sum_{j=0}^l \|W_j - X_j\| \quad (2.22)$$

A Figura 2.7 ilustra um exemplo de inserção de coluna de neurônios numa rede GHSOM, em que $N1$ representa o *error unit* e $N2$ representa o vizinho mais dissimilar. O peso atribuído aos novos neurônios inseridos é calculado por meio da média entre os pesos do neurônio $N1$ e o neurônio $N2$.

O crescimento do mapa continua até que a média dos erros de quantização dos neurônios seja menor que o limiar MQE_m , conforme apresentado na Equação 2.23:

$$MQE_m < \tau_m \cdot q_i \quad (2.23)$$

em que o parâmetro τ_m representa uma fração do erro de quantização q_i do neurônio i que

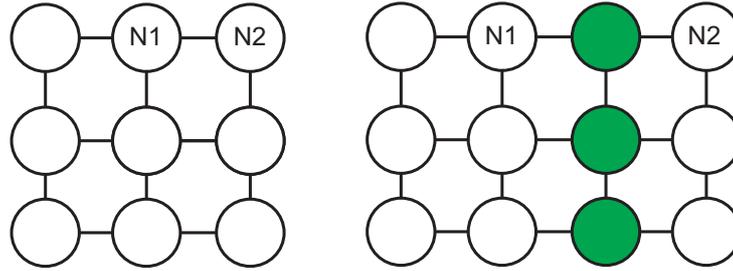


Figura 2.7: Inserção de neurônios.

deu origem ao mapa.

Logo que o processo de crescimento da SOM anterior está finalizado, os neurônios que possuem erro de quantização maiores que o outro limiar mqe_i serão expandidos (ver Equação 2.24), dando origem a outro mapa SOM, como no exemplo da Figura 2.8.

$$mqe_i > \tau_u \cdot q_0 \quad (2.24)$$

em que o parâmetro τ_u representa uma fração do erro de quantização inicial q_0 , da camada 0.

Desta forma, padrões mapeados em neurônios que não conseguiram reduzir seus erros de quantização terão a oportunidade de se auto-organizar em uma nova SOM, em um nível mais profundo de hierarquia (nos filhos desse neurônio).

O processo de formação da rede GHSOM é encerrado quando as unidades não mais requerem maiores expansões, ou seja, quando todos os neurônios do GHSOM possuírem erros de quantização menores que τ_u . Note-se que este processo de formação não conduz necessariamente a uma hierarquia equilibrada, ou seja, uma hierarquia com profundidade igual em cada ramo, na Figura 2.8 é apresentado um modelo de rede GHSOM.

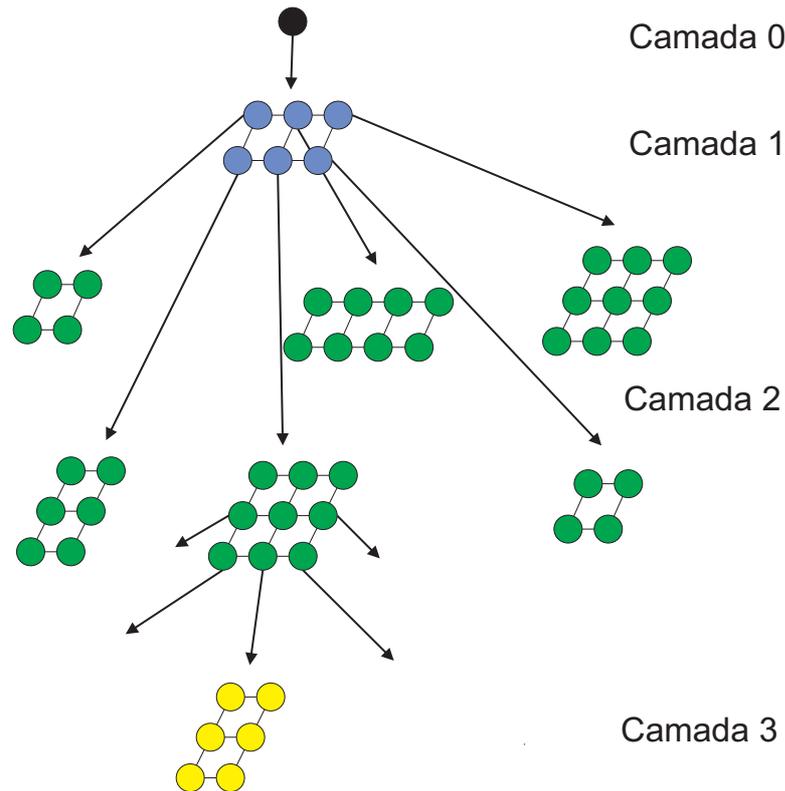


Figura 2.8: Exemplo do crescimento de um GHSOM.

A representação da execução do algoritmo GHSOM é apresentado no Algoritmo 2

Trabalhos relacionados

Além das técnicas baseadas em redes neurais, outra técnica popular de indexação de dados multidimensionais é a R^* -tree. O trabalho proposto por Kulkarni e Joshi [KJ02] apresenta uma técnica de recuperação de imagens baseada na similaridade onde a estrutura de indexação utilizada foi R^* -tree. Mesmo possuindo bons resultados, as estruturas baseadas na R^* -tree não são escaláveis para dimensões maiores que 20.

Outra técnica bastante referenciada é a de *Clustering*, que é amplamente utilizada em Reconhecimento de Padrões e é promissora em resolver o problema da alta dimensão na indexação. O trabalho proposto por Chen *et al.* [CWK03] apresenta um SRIBC que utiliza a técnica de *Clustering* para indexação das imagens, são apresentadas algumas vantagens, como: suporte a estruturas dinâmicas, capacidade de manipular dados de alta dimensão e o suporte a similaridade não Euclidiana.

Algoritmo 2 Algoritmo GHSOM**1. Inicialização:**

Os valores dos pesos são inicializados aleatoriamente, $w_j(0)$, desde que os valores $w_j(0)$ sejam distintos para $j = 1, 2, \dots, l$, onde l representa a quantidade de neurônios do mapa.

2. Amostragem:

Apresente uma amostra x do espaço de entrada ao mapa de neurônios M .

3. Cálculo do erro de quantização:

Identificar os neurônios i que possuem o maior erro de quantização q_i :

$$q_i = \sum_{j=0}^l \|W_j - X_j\| \quad (2.25)$$

em que W_j representa o vetor de pesos sinápticos do neurônio j e X_i representa o vetor de pesos de entrada.

4. Inserção de Neurônios:

Após a identificação, uma nova linha ou coluna de neurônios será inserida entre o neurônio com maior q_i e o seu vizinho imediato mais dissimilar.

5. Continuação do Crescimento do Mapa:

Repita a partir do passo 3 até que a média dos erros de quantização dos neurônios seja menor que o limiar MQE_m :

$$MQE_m < \tau_m \cdot q_i \quad (2.26)$$

em que o parâmetro τ_m representa uma fração do erro de quantização q_i do neurônio i que deu origem ao mapa.

6. Inserção de Mapas:

Os neurônios que possuírem erro de quantização maiores que o limiar mqe_i serão expandidos, dando origem a outro mapa.

$$mqe_i > \tau_u \cdot q_0 \quad (2.27)$$

em que o parâmetro τ_u representa uma fração do erro de quantização inicial q_0 , da camada 0.

7. Continuação:

Repita a parti do passo 3 até que todos os neurônios possuam erros de quantização menores que τ_u .

2.3 Combinação de Classificadores

Com a finalidade de alcançar um melhor resultado de classificação para uma dada tarefa, foi observado que a combinação de classificadores como um método para diminuir a probabilidade de erro na classificação é de grande valia. De um modo geral, a vantagem em usar um combinador é que se pode melhorar o desempenho do sistema fazendo com que as deficiências de um classificador sejam suprimidas pelo bom desempenho de outros [KH98]. No domínio da recuperação de imagens por conteúdo, a combinação de classificadores pode auxiliar numa maior acurácia no processo de recuperação das imagens.

As principais vantagens de empregar uma estratégia de combinação, segundo Jain *et al.*[JDM00] são:

- Habilidade de lidar com predições de classificadores baseados em diferentes espaços de características. Uma pessoa, por exemplo, poderia ser identificada por sua voz, imagem da face, assinatura, etc. Ou uma imagem poderia ser classificada, por meio de diferentes características, como cor, forma e textura.
- Habilidade de lidar com tipos diferentes de classificadores baseados no mesmo espaço de características. Diferentes classificadores possuem desempenhos locais distintos, isto é, em dados intervalos do espaço de características, os classificadores têm desempenhos diferentes. Desta forma, o bom desempenho local de um classificador pode compensar as deficiências locais de outros.
- Habilidade de lidar com classificadores homogêneos baseados no mesmo espaço de características. Alguns classificadores, como redes neurais, mesmo quando treinados com os mesmos dados realizam predições distintas em função da aleatoriedade no processo de inicialização, redundando nas vantagens mencionadas no item anterior.

A evidência experimental em diversos domínios de aplicação demonstra que a combinação de classificadores pode ser uma boa opção para melhorar a eficácia de sistemas de reconhecimento. A idéia é não confiar em uma decisão realizada por uma simples classificação. Em vez disso, todo o projeto, ou um subconjunto dos resultados serão usados para produzir uma decisão pela combinação das opiniões individuais derivando a uma decisão em consenso. A principal razão para combinar classificadores é a melhoria da acurácia.

Geralmente um esquema de combinação é composto por um conjunto de classificadores organizados em uma arquitetura e uma regra de combinação. Em relação à arquitetura, es-

tes esquemas são classificados como lineares, paralelos e hierárquicos conforme ilustrado na Figura 2.9. Na arquitetura do tipo linear, os classificadores são invocados em série, isto é, a saída de um classificador fornece uma classificação que é sucessivamente refinada por classificadores mais especializados em série. Na abordagem em paralelo, as saídas dos diversos classificadores são combinadas simultaneamente para realização de uma predição única pelo combinador. Na abordagem hierárquica, ocorre uma combinação das arquiteturas apresentadas anteriormente, de forma que as predições dos classificadores são combinadas tanto em série como em paralelo [Mat04].

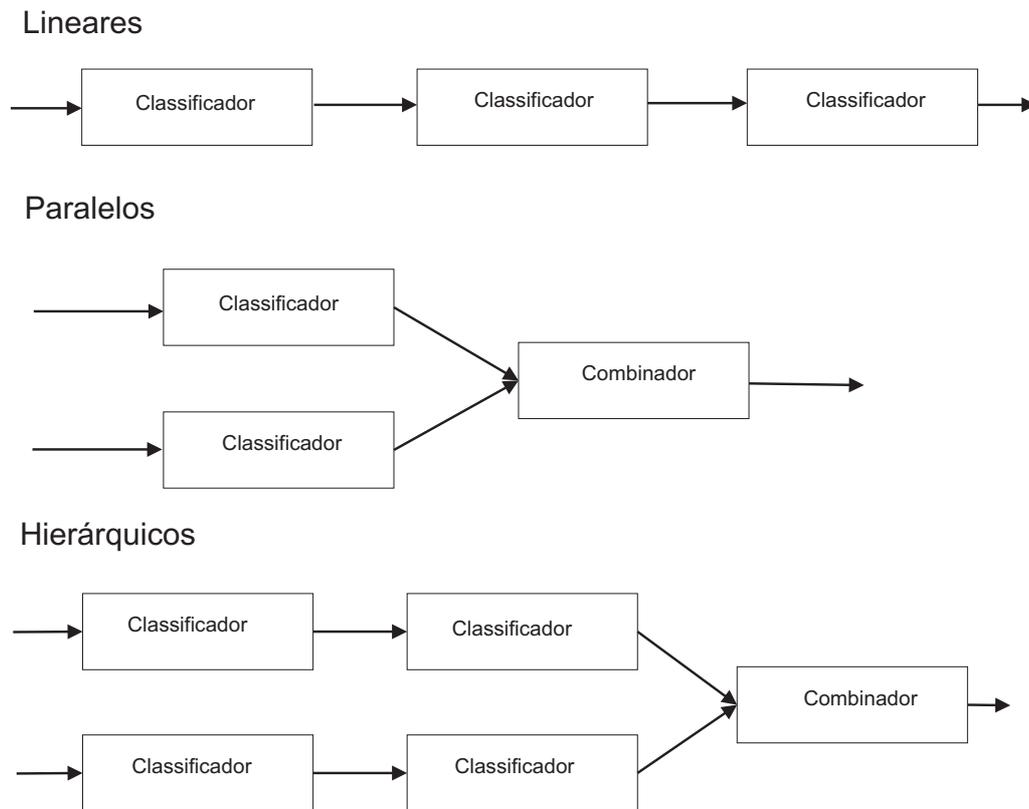


Figura 2.9: Arquiteturas para combinação de classificadores.

Com relação à regra de combinação segundo Kittler [KHDM98], os combinadores são classificados como baseados em regras fixas ou estáticas. Regras de combinação fixas realizam um procedimento pré-determinado. A forma como as entradas são combinadas para gerar uma saída única é estabelecida inicialmente, isto é, o conhecimento requerido para realizar a combinação não é adquirido através de um processo de treinamento [Mat04]. Segundo Xu *et al.* [XKS92] a combinação de classificadores é medida por três tipos de níveis de informação:

- Nível abstrato - são classificadores que produzem apenas como saída, rótulos predefi-

nidos pelo padrão de entrada. Ex: classificador simples.

- Nível posto - são classificadores que organizam os rótulos de saída de acordo com votos (pesos). Esse tipo de classificação possibilita apresentar resultados de classificação ordenados, por exemplo: apresentando uma lista das saídas de maior probabilidade até as de menor probabilidade de pertinência a uma dada classe. Ex. classificadores baseado em redes neurais.
- Nível de mediação - são classificadores que fornecem uma probabilidade condicional para cada rótulo da saída. Ex. classificador Bayesiano.

As regras de combinação mais utilizadas na literatura são a regra da média, regra do produto, regra da soma e voto majoritário [Mat04].

2.4 Exemplos de Sistemas

Nesta seção são apresentados sistemas pesquisados durante a revisão bibliográfica. Inicia-se com uma breve descrição desses sistemas, os tipos de características que são utilizadas, como são executadas as consultas, quais os resultados de recuperação obtidos e os tipos de estruturas de indexação utilizados. Em seguida é apresentada uma tabela que sumariza toda a revisão mostrando as principais características de cada sistema.

2.4.1 QBIC

O sistema QBIC (*Query By Image Content*) [NB94] foi desenvolvido pela IBM. O QBIC utiliza métodos de extração de características baseados em cor, forma e textura. Para as características referentes a cor, o sistema utiliza histogramas de cor em diferentes representações de espaço de cor. A característica de forma é extraída através de um conjunto de momentos invariantes, que representa a forma em relação a área, circularidade e orientação. Para textura, o QBIC utiliza versões modificadas de características de *coarseness* (fineza), *contrast* (contraste) e *directionality* (direcionalidade) propostas por Tamura [TMY78].

As consultas do QBIC podem ser realizadas de três formas distintas: através de imagens de consulta, em que uma imagem exemplo é submetida ao sistema ou através de um esboço feito pelo usuário, descrevendo os principais elementos da imagem, ou também através da seleção de padrões de cores e texturas. A busca por imagens similares no banco de dados é

feita através da distância Euclidiana e o método de indexação utilizado é baseado na estrutura R^* -tree. Durante a pesquisa não foram encontradas informações ou referências sobre o banco de imagens utilizado na recuperação do sistema.

2.4.2 NETRA

Desenvolvido pelo Departamento de Engenharia Elétrica e da Computação da Universidade da Califórnia, o sistema NETRA [MM99] utiliza um banco de imagens que são segmentadas em regiões de cores homogêneas, a partir dessas regiões é que são extraídas as características de cor, forma e textura que serão utilizadas pelo sistema. Na extração por meio de características de cor, utiliza-se uma representação quantizada do espaço de cor RGB em 256 cores. Já as texturas são representadas por um vetor de características contendo a média normalizada e o desvio padrão da série de transformadas Wavelet de Gabor. Por fim, a característica de forma é representada pela aplicação da transformada rápida de Fourier.

A consulta ao sistema é realizada a partir de imagens fixas, isto quer dizer, são utilizadas 2500 imagens da coleção de fotografias Corel [Dat99], as quais são organizadas em 25 grupos, cada qual contendo 100 imagens. A consulta pode ser realizada por meio da seleção de uma região específica da imagem ou por pela imagem inteira. A busca por imagens similares no banco de dados é feita usando-se a distância Euclidiana. O método de indexação utilizado é baseado na estrutura SS -tree.

2.4.3 PicSOM

O sistema foi desenvolvido pelo Laboratório de Ciências da Computação e Informação da Universidade de Tecnologia de Helsinki. No PicSOM [LKLO00] as características são computadas com a subdivisão da imagem em uma região central e quatro regiões nas laterais, a média dos canais R, G e B das regiões fornece um vetor de característica de dimensão 15, em seguida, são aplicados os momentos de cor para cada canal fornecendo assim um vetor de característica de dimensão 45 que será utilizado como entrada em uma rede SOM. Além disso, é aplicado um detector de bordas (Sobel) para a identificação das formas. Por fim, a imagem é normalizada e binarizada para a aplicação da transformada de Fourier. Todos os dados extraídos são utilizados na identificação da imagem.

A consulta no PicSOM é realizada em um processo iterativo que considera o *feedback* do usuário. A consulta é iniciada com a seleção de uma imagem representativa fixa e seu

resultado é refinado no processo de *feedback*. A busca de similaridade é baseada na distância Euclidiana nos Mapas Auto-Organizáveis (SOM), em que cada conjunto de imagens similares é representada por um neurônio. A estrutura de indexação utilizada é um vetor de características que serve para o treinamento da rede SOM. O sistema PicSOM utiliza um banco de imagens fixo e estático, mas durante a pesquisa não foram encontrados referências a esse banco.

2.4.4 VisualSEEk

Este sistema foi desenvolvido pelo Laboratório de Imagens e Televisão Avançada da Universidade de Columbia. No VisualSEEk [SC97] cada imagem é automaticamente decomposta em regiões de cores dominantes durante a criação do banco de dados. A consulta consiste em encontrar imagens que possuam o maior número de regiões similares. Para a representação das cores é selecionado um conjunto de cores do espaço HSV, de forma que as regiões da imagem são mapeadas utilizando estes conjuntos.

A consulta no VisualSEEk é realizada a partir de um esboço criado na interface pelo usuário, onde é configurado o número de regiões da imagem do esboço e definido as cores de cada região. A procura por imagens similares ocorre por casamento de padrões por região, utilizando a distância Euclidiana. A imagem que possuir o maior número de regiões similares é considerada mais próxima a imagem de consulta. Na indexação é utilizada uma estrutura baseada na *R-tree*. Não foram encontradas referências durante a pesquisa do banco de imagens utilizada na recuperação das imagens deste sistema.

2.4.5 WebSEEk

O WebSEEk [Smi97] foi desenvolvido pelo Laboratório de Imagens e Televisão Avançada da Universidade de Columbia. O sistema possui uma recuperação baseada em texto e cor e tem como objetivo catalogar coleções de imagens e vídeos da Internet. Da mesma forma que o sistema VisualSEEk, o WebSEEk utiliza uma representação de cores na extração de características, em que as cores das regiões, nas quais a imagem é decomposta, são mapeadas em um conjunto de cores do espaço HSV.

Na consulta, o usuário inicialmente escolhe um tema disponível a partir de um catálogo ou insere algum novo tópico. Os resultados obtidos com a consulta podem ser utilizados para uma consulta a todo o catálogo organizado por cor ou a uma lista de resultados com

cores semelhantes à consulta, organizado de forma decrescente. Além disso, o usuário tem a possibilidade de modificar manualmente o histograma de cor de uma imagem na tentativa de um refinamento no resultado da recuperação. A similaridade entre a consulta e o catálogo de imagens/vídeos é realizada a partir do cálculo da distância entre os histogramas de cor, em que é criada uma matriz de similaridade de cor que contém as cores mais significantes. Para todo o banco de imagens são computados os histogramas de cor que são indexados individualmente.

2.4.6 Sistema de French *et al.*

O sistema de French *et al.* [FMWX03] foi desenvolvido pelo Departamento de Ciências da Computação da Universidade de Virgínia.

A principal objetivo do sistema é criar diferentes representações de cor para uma imagem e utilizar algumas dessas representações ou todas durante o processo de recuperação, de forma a tentar aumentar a acurácia. Uma imagem é representada em quatro formatos de cor diferentes, (C+) que utiliza as cores originais, (B+) que utiliza imagem no formato preto e branco, (C-) que utiliza o negativo da imagem com as cores originais e (B-) que utiliza o negativo da imagem em preto e branco.

A consulta é baseada em imagens de exemplo. O usuário submete uma imagem para ser consultada. A procura por similaridade entre a imagem de consulta e o banco ocorre a partir da comparação de histogramas das imagens em suas diferentes formas de representação. Nos experimentos foram utilizadas 3400 imagens da coleção Corel [Dat99], que são subdivididas em 34 grupos de 100 imagens. Na indexação são criados quatro índices correspondentes, um para cada forma de representação da imagem. Por meio da combinação dos resultados da recuperação para cada formato, o sistema obteve uma precisão média de 22%.

2.4.7 Sistema de Barcellos *et al.*

O sistema de Barcellos *et al.* [BOLG05] foi desenvolvido na Universidade de São Paulo. Este sistema utiliza autocorrelogramas de cor para a recuperação de imagens por conteúdo. As características são extraídas baseadas em cor, para o espaço HSV quantizado.

O sistema utiliza consultas a partir de exemplos de imagens, em que os usuários submetem uma imagem ao sistema, sendo a procura por imagens similares no banco guiada pela distância Euclidiana entre os autocorrelogramas, que são representações gráficas da covariância entre

duas variáveis estatísticas, das imagens. As imagens são indexadas no banco através dos autocorrelogramas calculados para cada imagem.

Nos experimentos foram utilizadas 248 imagens obtidas da Internet, que foram divididas em 62 grupos. Com o testes foi obtida uma taxa de 75% de precisão média, para um banco de imagens muito pequeno.

2.4.8 Sistema de Prasad *et al.*

O sistema proposto por Prasad *et al.* [PGB01] tem como foco uma técnica de indexação baseada em cor e forma.

Na extração de características de cor é utilizada uma técnica de mapeamento das cores dominantes de uma região, em que são utilizadas 25 cores de um espaço RGB quantizado. Para a característica de forma é utilizada uma técnica de mapeamento de regiões utilizando grades de tamanho fixo, as características internas a cada espaço da grade são extraídas referente ao maior eixo, menor eixo e centro de gravidade. Na indexação das características é utilizada uma estrutura de *hash table*. A consulta no sistema pode ser realizada baseada em imagem exemplo ou a partir de padrões de cor. A similaridade entre a imagem de consulta e as imagens do banco é obtida a partir do cálculo das diferença entre as regiões das imagens. Caso o somatório dessa diferença seja menor que um determinado limiar, a imagem é considerada similar.

Os experimentos utilizaram um banco com 560 imagens divididas em 4 grupos distintos (bandeiras, frutas, flores e objetos). Foi obtida nos testes uma precisão média maior que 50%, da mesma forma que no sistema anterior, para um banco de imagens relativamente pequeno e com poucos grupos de imagens.

2.4.9 Sistema de Edvarden

Desenvolvido na Universidade de Ciência e Tecnologia da Noruega, o sistema de Edvarden [Edv06] tem como principal função explorar a utilização de métodos de recuperação em combinação com algoritmo genético na classificação.

No sistema, as características extraídas das imagens se baseiam em cor, tendo sido utilizados histogramas no espaço HSV (com quantidades de *bins* diferentes para cada canal, 18, 3 e 3 para H, S e V respectivamente). Para as medidas de similaridade são utilizados três métodos: distância Euclidiana, interseção de histogramas e distância quadrática. Na classi-

ficação são utilizados os vetores de características da imagem de consulta e os pesos gerados pelo algoritmo genético. No final do processo, é determinada qual classe que a imagem de consulta pertence.

Para os experimentos do sistema, foi utilizado um banco de imagens com aproximadamente 900 imagens subdivididas em 5 grupos distintos. Apesar dos testes comprovarem uma precisão média acima de 50% para todas as medidas de similaridade.

2.5 Análise Comparativa

Nesta seção é realizada uma análise comparativa contendo as principais funções e características dos sistemas apresentados nas subseções anteriores. A Tabela 2.1 sintetiza os sistemas abordados.

Observando os dados apresentados na Tabela 2.1 é possível afirmar que grande parte dos sistemas abordados na revisão bibliográfica utilizam como forma de consulta imagens-exemplos, alguns casos como os sistemas VisualSEEk e o WebSEEk utilizam esboços criados pelos usuários e apenas o QBIC utiliza mais de uma forma de consulta, utilizando imagens-exemplos, esboços dos usuários e padrões de cor e textura.

Outra questão importante diz a respeito à extração de características. A grande maioria dos sistemas utiliza histogramas de cor como método de extração, possivelmente devido a sua simplicidade e eficácia, mas também por se destacar de outros métodos citados como Wavelet de Gabor, Transformada de Fourier e Momentos de Cor. Uma grande parte dos sistemas utiliza como medida de similaridade a distância Euclidiana e a distância entre pontos do histogramas, apenas o sistema de Edvarden utilizou mais de uma medida de similaridade.

Entre os métodos de indexação foi observada uma divergência entre os sistemas, possivelmente devido a diferentes prioridades no desenvolvimento dos sistemas, como por exemplo, busca melhor desempenho, melhor acurácia ou tempo de resposta ao usuário.

Com base nessas informações algumas técnicas foram selecionadas para o desenvolvimento do sistema proposto nesta dissertação.

2.6 Considerações Finais

Neste capítulo foram abordadas as principais características dos sistemas de recuperação de imagens por conteúdo, os quais se caracterizam por indexar imagens a partir de suas pró-

prias características visuais, como cor, forma e textura. Foram descritas algumas das técnicas utilizadas em SRIBCs, como o Histogramas de Cor, que apesar de simples são invariantes a translação e rotação, de forma a serem destaque na literatura quando se tratam de características baseadas em cor. Nos descritores de Forma destacam-se os Momentos Invariantes de Hu, por serem bem sucedidos em muitas aplicações, além de incorporarem invariâncias a escala, rotação e translação. Já nos descritores de textura, destacamos o LBP (*Local Binary Patterns*), que são descritores invariantes a rotação e a variações na intensidade da imagem. As Wavelets também obtiveram destaque devido aos bons resultados referenciados na literatura.

Quanto a estruturas de indexação de dados multidimensionais, foram abordados alguns métodos tradicionais que possuem problemas com dimensionalidade. Desta forma, destacamos a utilização de redes neurais, mais especificamente GHSOM (Growing Hierarchical Self-Organizing Maps) devido a sua capacidade de operar sobre dados de alta dimensionalidade. Também foram apresentados métodos para combinação de classificadores em função dos benefícios que esta estratégia vem trazendo a sistemas de classificação em geral.

Todas as técnicas selecionadas foram analisadas levando em consideração o tipo de aplicação a ser desenvolvida na dissertação, tempo de resposta ao usuário, tipo de classificação e acesso às informações, o que poderá conduzir a uma eventual diferença de desempenho de pesquisa e velocidade.

Por último foram analisados alguns sistemas de RIBC encontrados na literatura. Foram observadas suas principais características: formas de extração, indexação e recuperação e banco de imagens utilizados, além dos resultados obtidos em termos de precisão e de cobertura (quando eram disponibilizados).

Tabela 2.1: Resumo dos sistemas apresentados.

Sistema	Consulta	Forma de Extração	Medida de similaridade	Indexação
QBIC	Imagem Exemplo, esboço e por padrões de cor e textura	Histograma de Cor, Texturas segundo Tamura e Momentos invariantes	Distância Euclidiana	R^* -tree
NETRA	Imagens Pré-definidas	Histograma de cor quantizados, Wavelet de Gabor e Transformada rápida de Fourier	Distância Euclidiana	SS -tree
PicSOM	Imagens Pré-definidas com refinamento interativo	Histogramas de Cor, Momentos de Cor e Transformada de Fourier	Distância Euclidiana	SOM
VisualSEEk	Esboço	Histograma de cor quantizado	Distância Euclidiana	R -tree
WebSEEk	Esboço e texto	Histograma de cor quantizado	Distância entre histogramas	Dados da Extração
Sistema de French <i>et al.</i>	Imagem exemplo	Histograma de cor (para as diferentes representações da imagem de consulta)	Distância entre histogramas	Dados extraídos de cara representação
Sistema de Barcellos <i>et al.</i>	Imagem exemplo	Histograma de cor quantizado	Distância Euclidiana	Dados da Extração
Sistema de Presad <i>et al.</i>	Imagem exemplo	Histograma de cor quantizado, Forma (Maior Eixo, Menor Eixo e Centro de gravidade)	Diferença entre as regiões extraídas, utilizando um limiar	dado não informado
Sistema de Edvardsen	Imagem exemplo	Histograma de cor quantizado	Distância Euclidiana, Interseção de histograma e distância quadrática	dados extraídos

Capítulo 3

Projeto e Implementação

O propósito deste capítulo é apresentar uma descrição detalhada do projeto e da implementação de um Sistema de Recuperação de Imagens Baseada em Conteúdo, proposto nesta dissertação, em que foram combinados diversos tipos de extratores de características por meio de redes neurais do tipo GHSOM (*Growing Hierarchical Self Organizing Map*). A rede neural foi utilizada como mecanismo de indexação das imagens, tendo sido desenvolvido um módulo dedicado à combinação de múltiplos classificadores utilizando cada uma das características distintas. Inicialmente, são descritos em linhas gerais os módulos componentes do sistema proposto. Em seguida, a macro arquitetura e os modos de operação são detalhados.

3.1 Arquitetura do Sistema

A macro arquitetura do sistema pode ser observada na Figura 3.1. Os principais módulos apresentados e discutidos são: interface com usuário, robô de busca, extração de características, classificadores e combinação de classificadores.

A seguir todos os módulos do sistema serão melhor detalhados.

Interface com usuário

Com o intuito de aumentar a interatividade com o usuário foi desenvolvida uma interface *Web* simples, por meio da qual o usuário tem a possibilidade de submeter imagens de consulta ao sistema. As imagens podem ser submetidas a partir do *upload* direto da imagem da máquina local do usuário para o sistema. A interface foi desenvolvida em HTML e utilizando *scripts* de PHP para interpretação de entrada e saída de cada módulo do sistema.

A interface permite ao usuário uma consulta direta, usando a da tela principal, (Figura

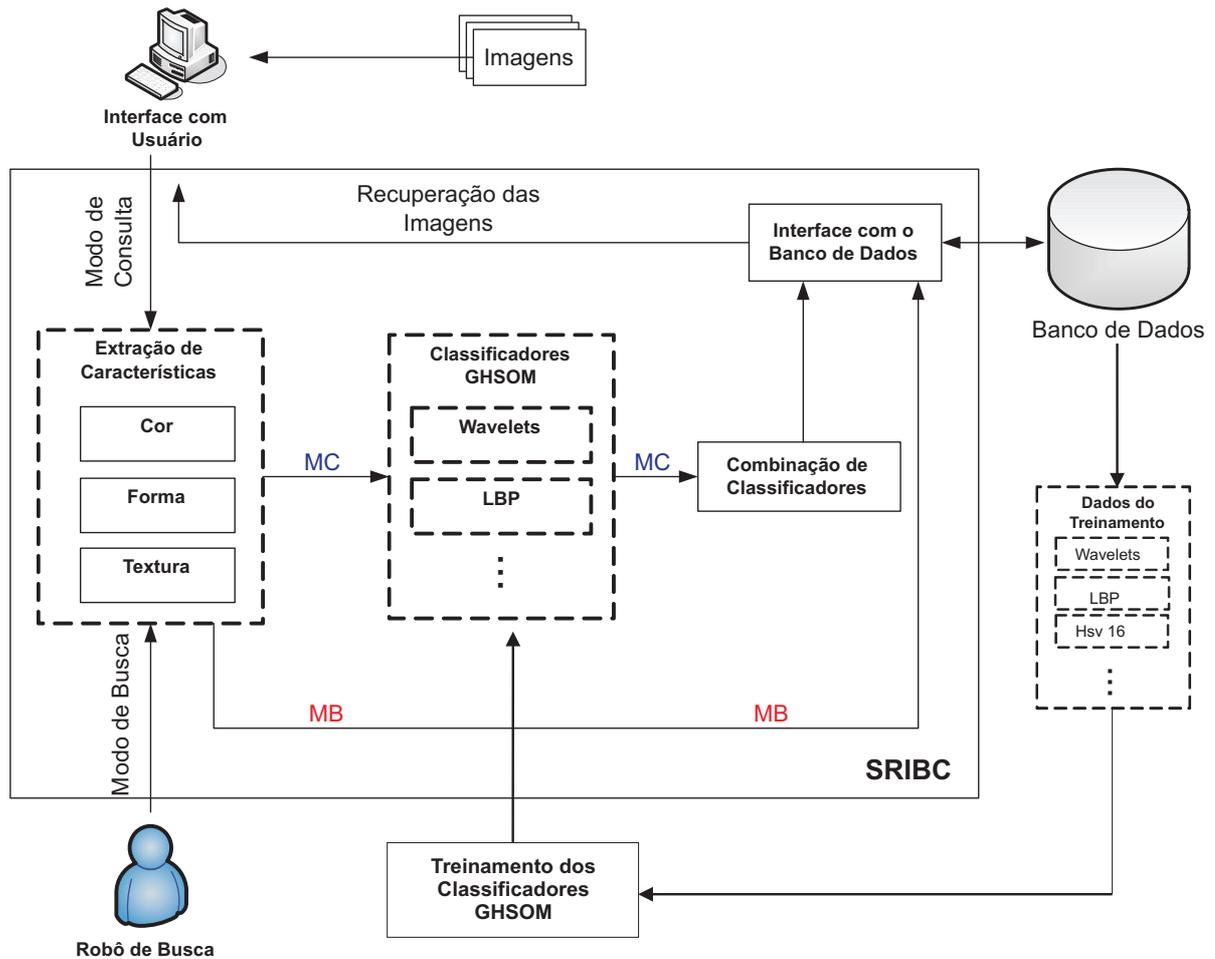
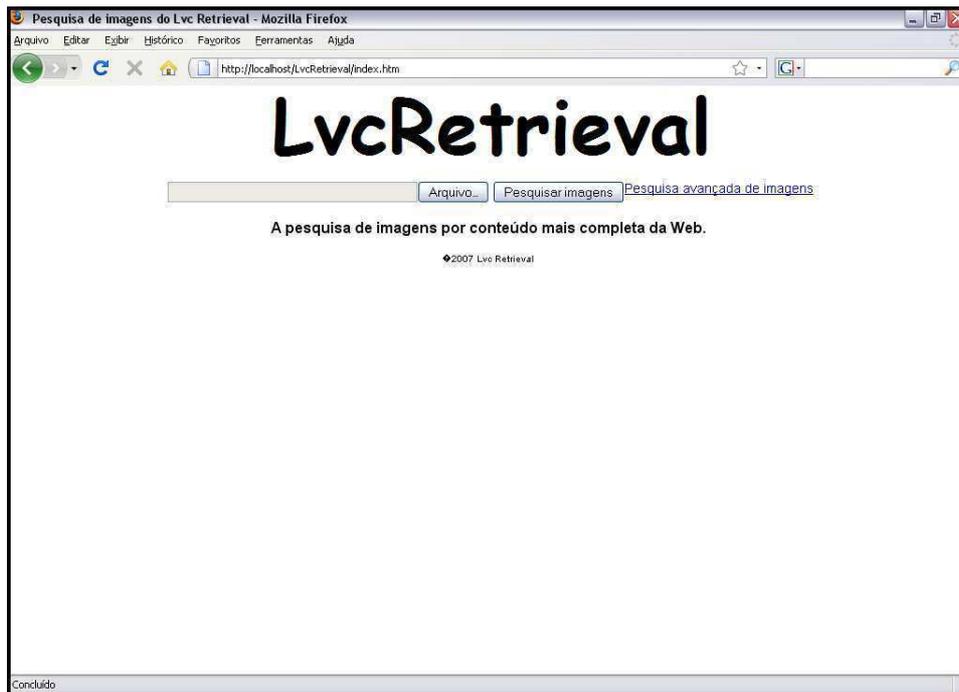


Figura 3.1: Arquitetura do Sistema de Recuperação de Imagens Baseada em Conteúdo.

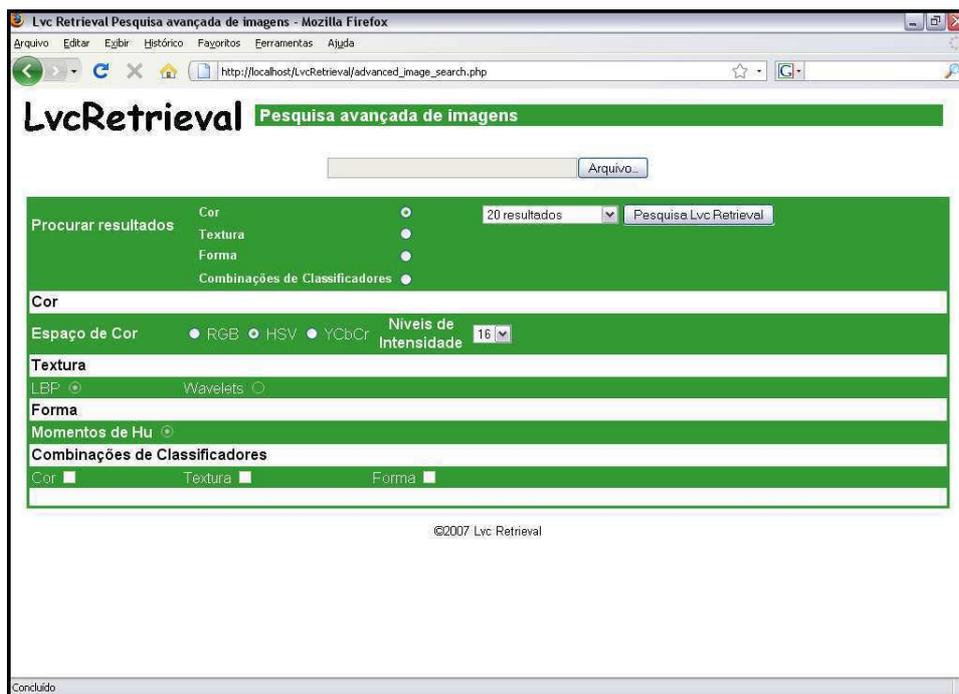
3.2(a)), em que são utilizados os extratores com melhores desempenhos combinados. Outra possibilidade é a consulta avançada (Figura 3.2(b)), em que o usuário poderá informar que tipo de extrator deseja utilizar com sua imagem de consulta e definir os valores dos parâmetros como o número de *bins* para os espaços de cores (RGB, HSV e YCbCr).

A interface avançada também permite ao usuário poderá definir que tipo de classificação utilizará: simples ou combinada. Na classificação simples, apenas uma característica é utilizada para recuperação das imagens. Por outro lado, a classificação combinada utiliza mais de uma característica extraída, em que, de uma maneira simples, cada resultado de um classificador é combinado com os demais. Na classificação combinada, o usuário poderá selecionar quais características deseja combinar (Cor, Forma e Textura).

Por último, o módulo de interface exibe uma tela de resultado, contendo a imagem de consulta e todas as imagens mais similares conforme as opções escolhidas, em ordem decrescente, da esquerda para a direita de cima para baixo (conforme ilustrado na Figura 3.3).



(a) Tela Principal.



(b) Tela Avançada.

Figura 3.2: Telas do Sistema de Recuperação de Imagens Baseada em Conteúdo, em que MB indica o fluxo do Modo de Busca e MC o Modo de Consulta.

O usuário ainda poderá colaborar com a avaliação da precisão do sistema ao indicar se as imagens recuperadas são similares segundo o seu critério pessoal, usando o campo *checkbox* abaixo de cada imagem recuperada.

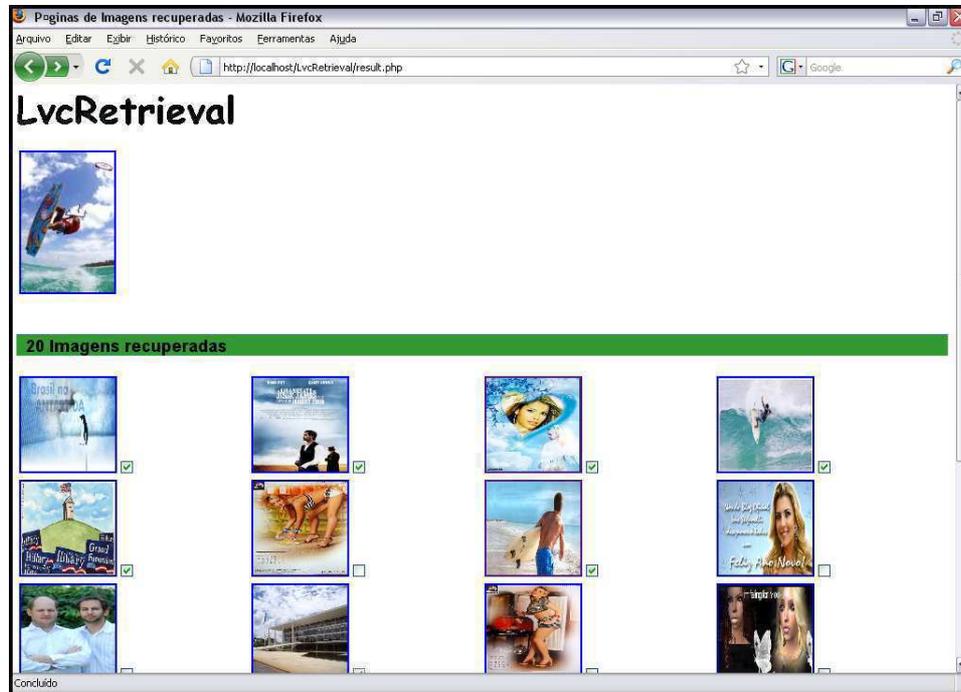


Figura 3.3: Tela dos resultados da recuperação.

Robô de Busca

Este módulo (também conhecido como *Web Crawler* ou *Web Spider*) é responsável pela navegação, busca e indexação automática de imagens disponíveis na *World Wide Web*.

O *Web Crawler* é responsável por procurar imagens que são utilizadas como resposta na recuperação. Inicialmente, é necessária a inserção de uma semente (URL), que indicará o ponto inicial das buscas do *Crawler*, de forma que todos os links de imagens que possuírem alguma ligação com a semente serão adicionados a uma tabela no banco de dados, que corresponde às possíveis imagens que serão utilizadas nos extratores de características.

Definiu-se que todas as imagens submetidas aos extratores de características, obedecendo às regras estabelecidas na configuração do *Crawler*, deveriam possuir tamanho mínimo de 150 x 150 pixels e razão entre altura e largura entre 3 e 1/3. Desta forma, evita-se a captura de imagens muito pequenas ou imagens de elementos de *design* (barras, botões, laterais de sites, rodapés, etc.). Também, foi imposta a restrição de que todas as imagens devem estar no formato **JPEG** (*Joint Photographic Experts Group*), já que esse formato é amplamente utilizado na Internet. Mais detalhes podem ser encontrados no Apêndice A.

Todas as imagens que satisfazem as condições acima podem ser processadas pelos demais módulos do sistema.

Extração de Características

Este módulo é responsável por extrair características das imagens submetidas ao sistema ou de imagens encontradas pelo robô de busca. A extração de características realizada para imagens submetidas pelo usuário é configurada pelo próprio usuário, o que significa que o usuário irá informar quais extratores devem ser utilizados ou qual combinação deve ser aplicada. Desta forma, pode-se optar por utilizar apenas um tipo de característica na consulta ou um conjunto de características combinadas.

Para as imagens encontradas pelo do robô de busca, o processo de extração de características ocorre de maneira diferente. Todas as imagens encontradas têm suas características extraídas utilizando-se de todos os métodos. As características de cor, por exemplo, são extraídas utilizando diferentes espaços de cores (RGB, HSV e YCbCr) e diferentes números de *bins* por componente (128, 64, 32 e 16). Por outro lado, características de forma são obtidas a partir dos Momentos Invariantes de Hu. Finalmente, características de textura, são extraídas através dos métodos das Wavelets e LBP.

Ao final do processo de extração, todas as características são inseridas em um banco de dados, como também a URL em que a imagem se encontra, de forma a serem utilizadas posteriormente no treinamento das redes neurais GHSOM.

Treinamento dos Classificadores

Este módulo é responsável pelo treinamento das redes GHSOM, utilizando as características das imagens que foram extraídas e armazenadas no banco de dados. Para simplificar o processo, foi decidido que esse treinamento fosse realizado de forma individual, ou seja, para cada classificador é utilizado apenas um tipo de característica específica (por exemplo, um classificador para Wavelets, outro para LBP).

Para a realização desse treinamento, é necessária a criação de um arquivo de entrada contendo todos os elementos da característica selecionada e as URLs das imagens, como pode ser observado na Figura 3.1. Esse arquivo é utilizado como entrada em uma biblioteca, que foi desenvolvida na Universidade de Tecnologia de Viena (<http://www.ifs.tuwien.ac.at/andi/ghsom/>), que dará origem a uma rede GHSOM treinada. Como resultado do treinamento, funções da biblioteca permitem a criação de um arquivo de saída contendo os grupos de imagens formados, os pesos associados a cada um dos neurônios que representam esses grupos, bem como os níveis (*layers*) em que cada grupo se encontra na estrutura da rede.

Após a realização dessa etapa, que periodicamente é executada para atualização das imagens e dos mapas, os próximos módulos poderão ser executados.

Classificadores

O módulo de classificação é responsável por encontrar as imagens mais similares à imagem de consulta, com base no módulo de extração de características. Para isso, cada característica é submetida a seu classificador GHSOM específico, resultando ao final em uma lista dos neurônios vencedores para cada nível (*layer*) da rede.

Para encontrar os neurônios vencedores, é necessário o cálculo do erro de quantização (*QE* - *Quantization Error*), isto é, buscam-se neurônios que possuam pesos com o menor erro de quantização em relação à imagem de consulta. O cálculo do *QE* é realizado de acordo com a Equação 3.1, a seguir:

$$qe_i = \sum_{x_j \in C_i} \|m_i - x_j\| \quad (3.1)$$

em que i e j representam as dimensionalidades dos pesos dos neurônios do mapa m_i e do vetor de características da imagem de consulta x_j , respectivamente.

De acordo com o tipo de configuração utilizada, mais de um classificador pode ser acionado com diferentes características extraídas e os resultados dessas classificações poderão ser combinadas pelo módulo de combinação.

Combinação de Classificadores

O módulo de combinação de classificadores funcionará sempre que mais de um classificador for acionado em uma consulta, isto quer dizer que o usuário, de forma explícita, irá selecionar e configurar o módulo de combinação.

A regra de encadeamento utilizada na combinação de classificadores é uma modificação da técnica de voto majoritário, como descrito na Seção 3.2. Após a combinação dos resultados dos classificadores, as imagens são ordenadas de forma decrescente de acordo com suas similaridades com a imagem de consulta. O resultado final é enviado ao usuário em uma tela de resultado final contendo a imagem de consulta e as imagens classificadas como mais similares.

3.2 Módulos do Sistema

A exemplo dos modelos clássicos de Sistemas de Recuperação de Imagens Baseada em Conteúdo [Dow93; HMR96; MM99; NB94; APS96; SC97], o sistema proposto possui um módulo de extração de características e um módulo de classificação (indexação). Adicionalmente, foi desenvolvido um módulo responsável pela combinação de classificadores com vistas à obtenção de melhores resultados de classificação.

Módulo de Extração de Características

O módulo de extração de características é responsável por computar vetores de cor, forma e textura a partir das imagens de entrada, esse módulo está ilustrado na Figura 3.4. Na fase de tratamento, todas as imagens submetidas ao sistema têm seus níveis de intensidade normalizados no intervalo $[0, 1]$. A equação 3.2 é usada no processo de normalização. Dessa forma é possível comparar imagens com diferentes números de *pixels*.



Figura 3.4: Módulo de extração de características.

$$N_{x,y} = \left(\frac{N_{max} - N_{min}}{O_{max} - O_{min}} \right) (O_{x,y} - O_{min}) + N_{min} \quad (3.2)$$

em que $N_{x,y}$ corresponde ao valor normalizado de intensidade da imagem e $O_{x,y}$ corresponde ao valor original de intensidade; N_{max} e N_{min} correspondem aos novos valores de máximo e mínimo após a normalização (neste trabalho, 1 e 0 foram utilizados, respectivamente); Finalmente, O_{max} e O_{min} correspondem aos valores máximos e mínimo da imagem original (sem normalização).

Em seguida, no módulo de extração de características, diferentes extratores (cor, forma e textura) atuam sobre as imagens normalizadas. No extrator baseado em cores, optou-se pela utilização de uma técnica simples, mas bastante eficaz, que é a de histogramas de cores. Para aplicação da técnica de histograma foram selecionados diferentes espaços de

cores (RGB, HSV e YCbCr) representados com diferentes quantidades de *bins* (128, 64, 32 e 16). Para o extrator baseado em texturas, optou-se pela utilização de duas técnicas: Wavelets, mais especificamente, filtros de Daubechies e Haar [WWFXW98] sobre imagens de entrada convertidas para níveis de cinza por meio da biblioteca OpenCv da Intel; e a técnica LBP (*Local Binary Patterns*) [OPM02], também aplicada sobre imagens convertidas para níveis de cinza. Para o extrator baseado em formas, foi utilizada a técnica dos Momentos Invariantes de Hu [Hu62]. Todas as técnicas selecionadas para extração de características foram mais detalhadas no capítulo anterior.

Módulo de Classificação

O módulo de classificação é responsável por associar imagens similares a uma dada imagem de consulta fornecida como entrada para o sistema. Para executar a classificação, são utilizados classificadores baseados em redes neurais do tipo GHSOM, treinados previamente com características extraídas (histogramas RGB, HSV, YCbCr; Momentos Invariantes de Hu, Wavelets e LBP) de uma base de imagens de treinamento. As vantagens da utilização da GHSOM estão na propriedade de treinamento não-supervisionado, criação de grupos dinâmicos, capacidade de inserção de novos neurônios em regiões com grandes diferenças entre os padrões e a subdivisão dos padrões em hierarquia. Desta forma, o tempo de busca por padrões nos mapas é reduzida. A Figura 3.5 contém uma ilustração do módulo de classificação.

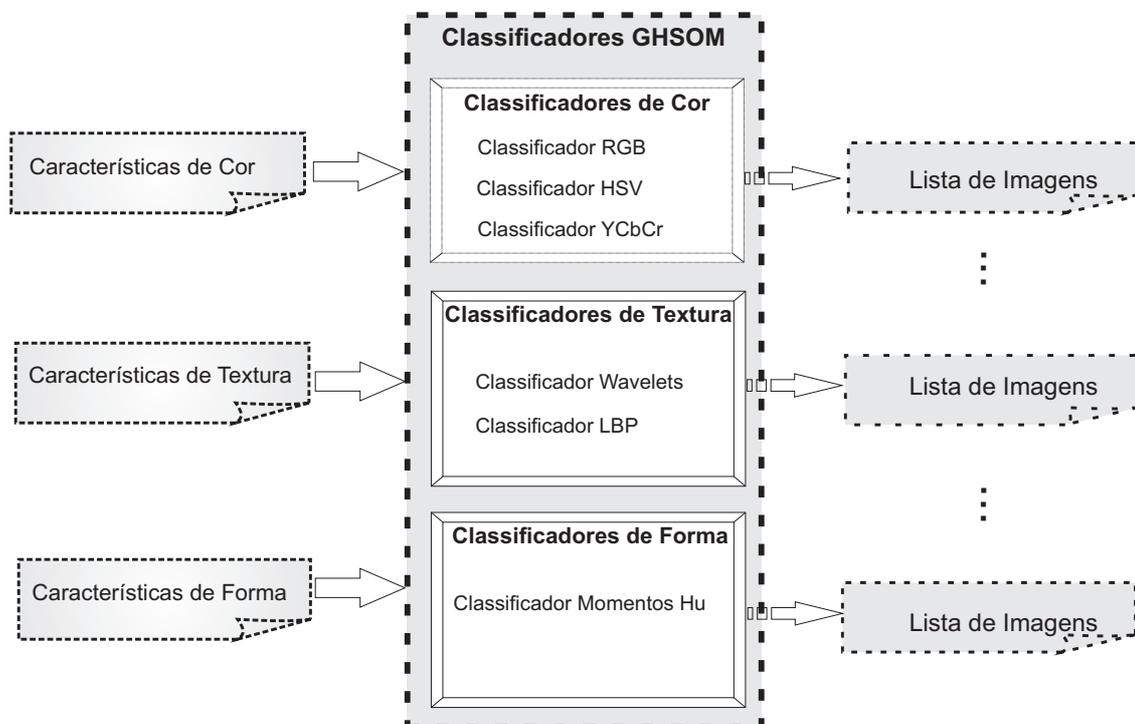


Figura 3.5: Módulo de classificação.

Para o treinamento das redes GHSOM, foi utilizado um banco de dados de imagens com suas características extraídas. Cada tipo de característica alimenta um classificador GMSOM específico, conforme ilustrado na Figura 3.5. Ao final do treinamento, é criado um mapa baseado em uma estrutura de árvore da rede GHSOM, em que imagens similares são mapeadas em neurônios próximos. Para criação do banco de dados de imagens com suas características extraídas, foi utilizado um robô de busca (*Web Crawler*), cuja função é automaticamente seguir links em páginas da *World Wide Web*, encontrar imagens satisfazendo alguns critérios pré-estabelecidos e extrair as características dessas imagens.

No processo de classificação, as características extraídas da imagem de consulta são utilizadas como entradas nas redes GHSOM treinadas, de forma que cada característica serve de entrada a apenas uma rede que foi anteriormente treinada com características idênticas. Por exemplo, para vetores de características utilizando o método LBP, utiliza-se a rede GHSOM treinada com características LBP. Após a propagação das características da imagem de consulta pelas redes correspondentes, os neurônios vencedores de cada camada (*layer*) são produzidos como resposta, e todas as imagens associadas a esses neurônios formam a lista resultante de imagens mais similares à imagem de consulta.

Módulo de Combinação dos Resultados de Classificação

Este módulo é responsável por combinar os resultados de todos os classificadores em uma só lista de imagens, conforme ilustrado na Figura 3.6.

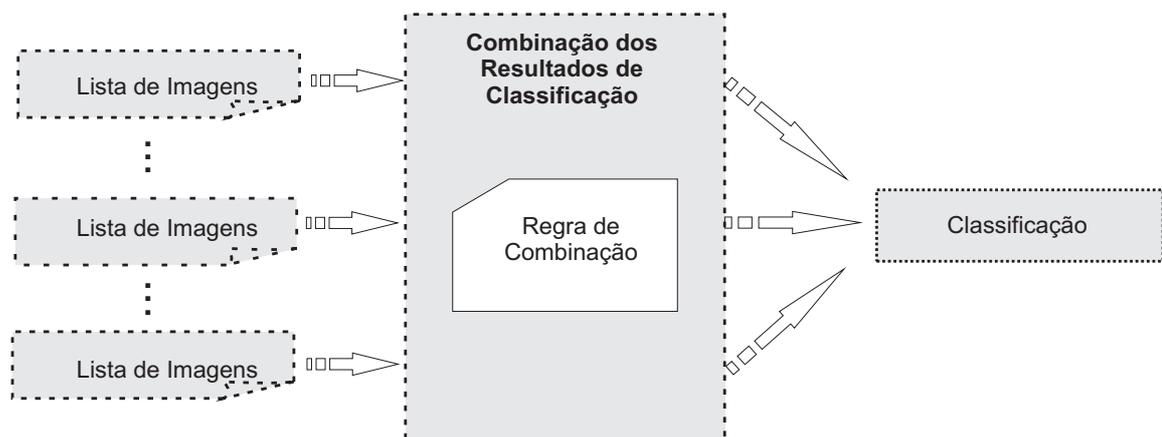


Figura 3.6: Combinação de classificações

A regra de combinação selecionada foi a de voto majoritário, uma estratégia simples e computacionalmente eficiente, consistindo apenas em uma contagem do número de classificadores que retornaram uma imagem específica. Desta forma, a imagem que receber um maior número de votos dos classificadores estará mais bem cotada entre as imagens retornadas pela

pesquisa.

O método de voto majoritário, quando aplicado ao sistema proposto neste trabalho, está sujeito a uma deficiência que diminui sua precisão. Esta deficiência é ocasionada ao se utilizar uma quantidade reduzida de classificadores, desta forma diversas imagens podem receber a mesma votação.

A classificação da imagem de consulta é realizada com base no erro de quantização, de forma que imagens que possuem um valor de erro muito próximo a zero são aquelas com maior grau de similaridade relativo à imagem de consulta, considerando um classificador GHSOM específico.

Portanto, a informação do grau de similaridade entre as imagens pode ser usada em proveito da combinação de classificadores. Para a combinação de classificadores, foi utilizado um método em que cada imagem recuperada recebe um voto conforme uma fórmula que leva em conta todas as redes GHSOM que retornaram aquela imagem como resultado da consulta. O valor do voto é utilizado para ordenar a lista de imagens resultante da combinação. As imagens com os maiores números de votos permanecem na lista resultante e são consideradas como as mais similares à imagem de consulta. Por fim, o valor do voto é utilizado para ordenar as imagens ao final da combinação. As imagens que possuem os maiores votos são classificadas como as mais similares à imagem de consulta.

Para o cálculo do voto de cada imagem, são levados em consideração os erros de quantização produzidos por cada rede GHSOM para a imagem, os níveis (*layers*) em que a imagem foi classificada e a quantidade de níveis de cada rede GHSOM utilizada no processo de recuperação.

O cálculo do voto para a combinação é expresso pela seguinte equação:

$$v_i = \sum_g^G (1 - QE_{g,i}) \cdot \left(\frac{LC_{g,i}}{LT_g} \right) \quad (3.3)$$

em que G representa as redes GHSOM utilizadas no processo de combinação, o valor $QE_{g,i}$ indica o erro de quantização da imagem para uma determinada rede selecionada, o valor $LC_{g,i}$ representa qual o nível da rede em que a imagem foi recuperada e, por fim, LT_g indica o número total de níveis da rede GHSOM selecionada. Ao final do cálculo, um voto com peso igual a v_i é atribuído a uma imagem recuperada.

Esta estratégia reduz drasticamente as chances de empates na votação das imagens mais similares e facilita a ordenação.

3.3 Modos de Operação

Na arquitetura proposta existem dois modos de operação distintos, o modo de busca, o qual utiliza o robô de busca (*Web Crawler*) para obter imagens na Internet. Tais imagens, após passarem por todos os processos de extração, têm suas características inseridas no banco de dados, onde posteriormente são utilizadas no treinamento das redes GHSOM. Toda a trajetória do modo de busca, destacada com setas na cor vermelha, pode ser observada na Figura 3.7. Ao final do processo de extração, apenas as características e a URL da imagem são armazenadas no banco de dados, de forma a reduzir o espaço em disco requisitado.

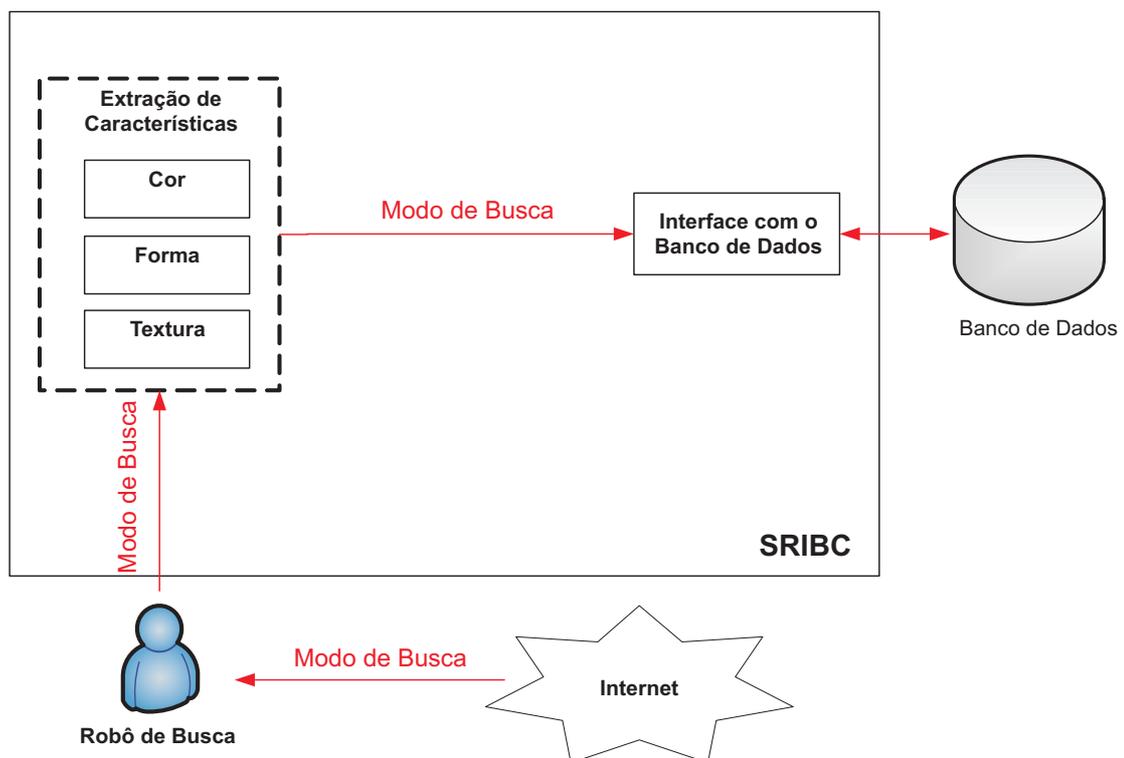


Figura 3.7: Modo de Busca.

O outro modo de operação se trata do modo de consulta, no qual o usuário final a partir da interface *Web* poderá submeter ao sistema uma imagem de consulta, que ao passar por todos os processos de extração, classificação e combinação, como ilustrado na Figura 3.8, retornará um conjunto imagens similares à imagem de consulta.

Todo o caminho percorrido pelo modo de consulta está destacado em cor azul, desde a submissão da imagem até a recuperação retornada ao usuário.

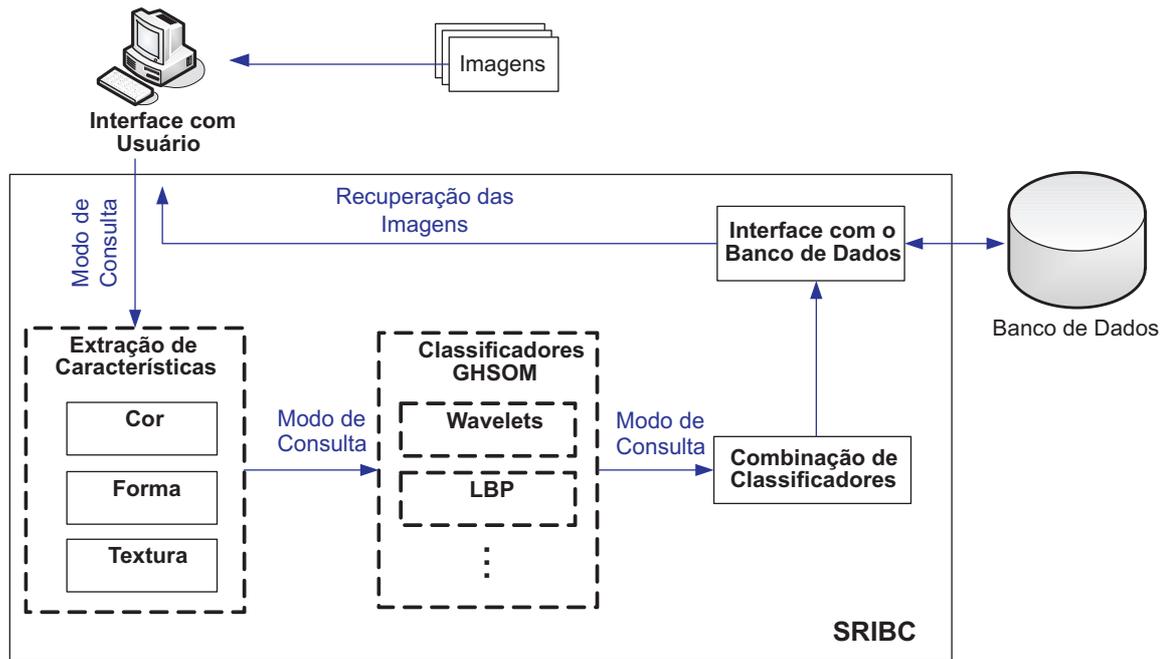


Figura 3.8: Modo de Consulta.

3.4 Considerações Finais

Neste capítulo, foram descritos de forma detalhada os principais módulos do sistema proposto e desenvolvido nesta dissertação, destacando as funções de cada módulo, a sua macro arquitetura, demonstrando as estruturas e configurações, e os seus modos de operação.

Foram apresentados inicialmente os módulos em que o sistema foi subdividido, desde o módulo de extração, destacando os extratores utilizados e suas configurações, passando em seguida pelo módulo de classificação, em que foi apresentado o processo de treinamento dos classificadores e a utilização dos mesmos e por último, foi destacado o módulo de combinação de classificadores, que utiliza como regra de combinação uma técnica modificada do voto majoritário.

Na seqüência, a macro arquitetura do sistema foi apresentada de uma forma detalhada, desde a interface *Web*, passando pelos extratores de características e seus métodos. O robô de busca foi apresentando, destacando sua função de busca de imagens e armazenamentos de características e URL's. O treinamento das redes GHSOM também foi detalhado, assim como a utilização dos classificadores e métodos de busca. Por fim, foi apresentada a combinação de classificadores, discutindo-se como as saídas de múltiplos classificadores podem ser transformadas em uma única classificação.

Por último, foram destacados os modos de operação. O modo de busca, o qual utiliza

o robô de busca como forma de alimentar o banco de dados, e modo de pesquisa, o qual é utilizado pelo usuário do sistema para a recuperação.

No próximo capítulo, são apresentados os principais experimentos realizados no decorrer desta pesquisa, os quais foram fundamentais na seleção dos melhores parâmetros para a classificação e recuperação de imagens. Além disso, é apresentada uma análise do módulo de combinação e sua comparação com o método de classificação simples. Por fim, é apresentado um estudo de caso no qual o sistema proposto é testado por usuários em buscas de imagens na *Web*.

Capítulo 4

Estudo Experimental

Neste capítulo, são apresentados experimentos envolvendo diferentes tipos de extratores de características e tipos de combinação de classificadores aplicados ao problema de recuperação de imagens por conteúdo. O primeiro experimento teve como objetivo encontrar os melhores parâmetros para o treinamento de cada uma das redes GHSOM que utilizam diferentes características extraídas das imagens. Para o segundo experimento, os melhores parâmetros selecionados no primeiro experimento foram utilizados no treinamento das redes GHSOM, as quais serviram de base para um estudo comparativo entre classificações simples e combinadas. Por último, foi realizado um experimento envolvendo todo o sistema e um conjunto de usuários finais, os quais realizaram testes de recuperação de imagens com avaliação subjetiva a respeito da precisão do sistema.

4.1 Experimento 1 - Busca pelos Melhores Parâmetros dos Classificadores GHSOM

O Experimento 1 teve como objetivo a busca pelos melhores parâmetros para o treinamento de cada uma das redes GHSOM que são treinadas com diferentes características extraídas dos banco de imagens rotuladas. O parâmetro buscado durante o experimento foi o τ_2 (Seção 2.2), que é responsável pelo crescimento da profundidade em uma GHSOM. O parâmetro τ_1 foi fixado em 0,1, devido a alguns testes realizados que mostraram que a modificação do valor desse parâmetro não influi diretamente na modificação da rede GHSOM treinada.

A variação do parâmetro τ_2 é necessária para a busca de redes em que os elementos (imagens) sejam distribuídos uniformemente nos grupos (classes). Essa uniformidade dos

grupos é necessária devido à utilização de um banco de imagens rotuladas como conjunto de treinamento, no qual cada grupo (de imagens rotuladas) desse banco possui um mesmo número de imagens vinculadas.

O banco rotulado utilizado como conjunto de treinamento para as redes foi formado por 800 imagens rotuladas, distribuídas em 16 grupos (classes) diferentes, todas as imagens estão em um formato JPEG e com tamanho máximo de 256 x 256 *pixels*. O banco de imagens foi criado a partir da junção de imagens capturadas na Internet com um banco de imagens MRCORID (*Microsoft Research Cambridge Object Recognition Image Database* (<ftp://ftp.research.microsoft.com/pub/download/orid>)). Dos 16 grupos existentes no banco, 5 são pertencentes ao banco da Microsoft e os 11 grupos restantes foram rotulados manualmente. Maiores detalhes podem ser encontrados no Apêndice B.

Os rótulos de cada grupo foram atribuídos de acordo com o conteúdo mais aparente de cada uma das imagens. Desta forma, os rótulos obtidos para os grupos foram: aviões, vegetação, aves, garrafas, camelos, carros, faces, flores, folhas, futebol, guitarras, casas, motocicletas, outono, planetas e pôr-do-sol.

Como conjunto de teste, foi utilizado um banco contendo 128 imagens rotuladas nos mesmos 16 grupos do banco de imagens descrito acima. Como conjunto de treinamento foram utilizadas 800 imagens rotuladas, diferentes das imagens utilizadas para o conjunto de teste.

Os valores de τ_2 avaliados foram os seguintes: 0,5; 0,1; 0,05; 0,01; 0,005; 0,001; 0,0005; 0,0001; 0,00005; 0,00001; 0,000005; 0,000001 e 0,0000001, tais valores foram obtidos por meio de experimentos anteriores com a biblioteca de criação de GHSOM em que foi verificado crescimento da rede, número de grupos criados e o tamanho dos grupos. Desta forma, foram treinadas redes com diferentes parâmetros de τ_2 para cada extrator de característica, totalizando 195 redes GHSOM treinadas ao término da fase de treinamento. Depois de finalizado todo o processo de treinamento, foram submetidas à base de teste para recuperação, os resultados obtidos nesta fase foram utilizados como medida para a seleção dos melhores valores de τ_2 para cada rede.

O critério utilizado para analisar os resultados obtidos após o término da fase de teste foi baseado em três métricas: Precisão (*Precision*), Cobertura (*Recall*) e Medida-F (*F-Measure*) [Ren04], que são freqüentemente utilizadas como medidas de avaliação em sistemas de recuperação de informação.

Precisão (*Precision*)

A métrica de precisão corresponde à quantidade de cada recuperação, ou seja, a estimativa da probabilidade de cada imagem retornada estar correta.

$$Precisão = \frac{\text{quantidadeDeImagensRelevantesRecuperadas}}{\text{quantidadeTotalDeImagensRecuperadas}} \quad (4.1)$$

Cobertura (*Recall*)

Já a métrica de cobertura corresponde à proporção de respostas corretas e é uma estimativa de probabilidade de se obter pelo menos uma imagem relevante (imagens classificadas previamente em grupos).

$$Cobertura = \frac{\text{quantidadeDeImagensRelevantesRecuperadas}}{\text{quantidadeTotalDeImagensRelevantes}} \quad (4.2)$$

Medida-F (*F-Measure*)

Por último, a Medida-F corresponde à média harmônica entre a precisão e a cobertura.

$$Medida - F = \frac{(1 + \beta^2) \cdot (Precisão \cdot Cobertura)}{(\beta^2 \cdot Precisão + Cobertura)} \quad (4.3)$$

onde $\beta = 1$ para que a precisão e a cobertura possuam o mesmo peso.

Para a execução dos experimentos foram utilizados todos os extratores de características (os histogramas de cor HSV com 128 *bins*, HSV com 64 *bins*, HSV com 32 *bins*, HSV com 16 *bins*, RGB com 128 *bins*, RGB com 64 *bins*, RGB com 32 *bins*, RGB com 16 *bins*, YCbCr com 128 *bins*, YCbCr com 64 *bins*, YCbCr com 32 *bins*, YCbCr com 16 *bins*, LBP, Wavelets e Momentos de Hu). Cada característica extraída deu origem a 13 redes GHSOM treinadas com diferentes valores de τ_2 , utilizadas na fase de teste. O melhor parâmetro para cada rede foi selecionado de acordo com as métricas de precisão e cobertura médias e Medida-F.

Para os extratores baseados em Cor, foi utilizada, além de diferentes tipos de espaço de cor (RGB, HSV e YCbCr), a quantização dos níveis de intensidade em diferentes quantidades de *bins*. Desta forma, cada espaço de cor utiliza quatro diferentes redes GHSOM, devido à quantidade de *bins* (128, 64, 32 e 16), para as classificações.

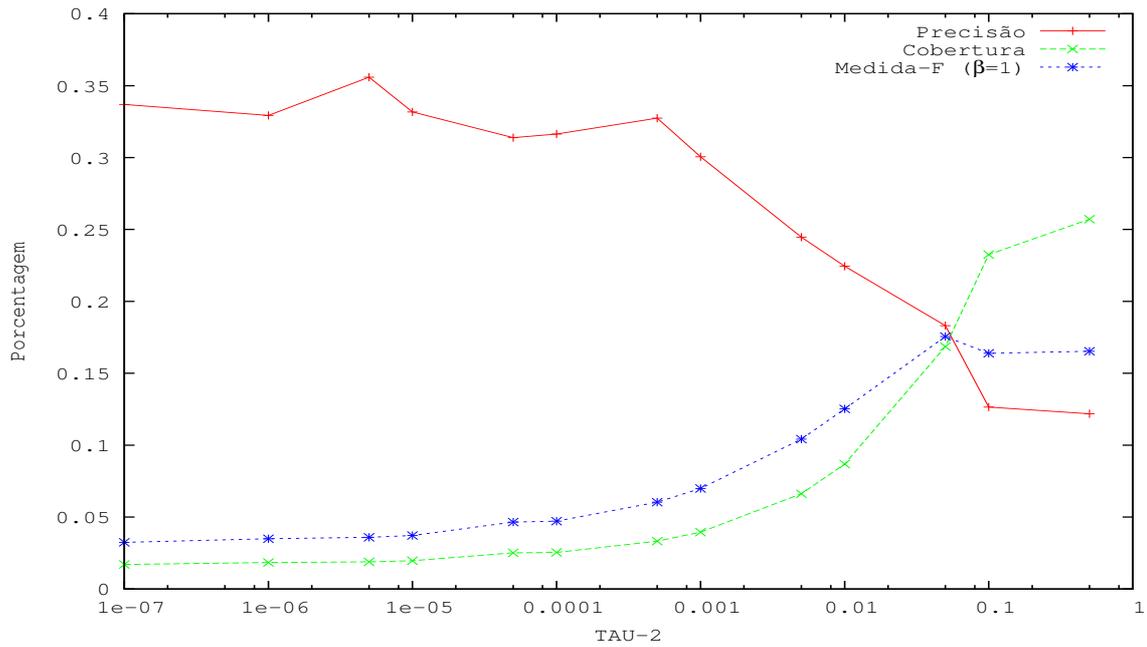


Figura 4.1: Avaliação do parâmetro τ_2 utilizando o extrator RGB com 128 *bins* por canal.

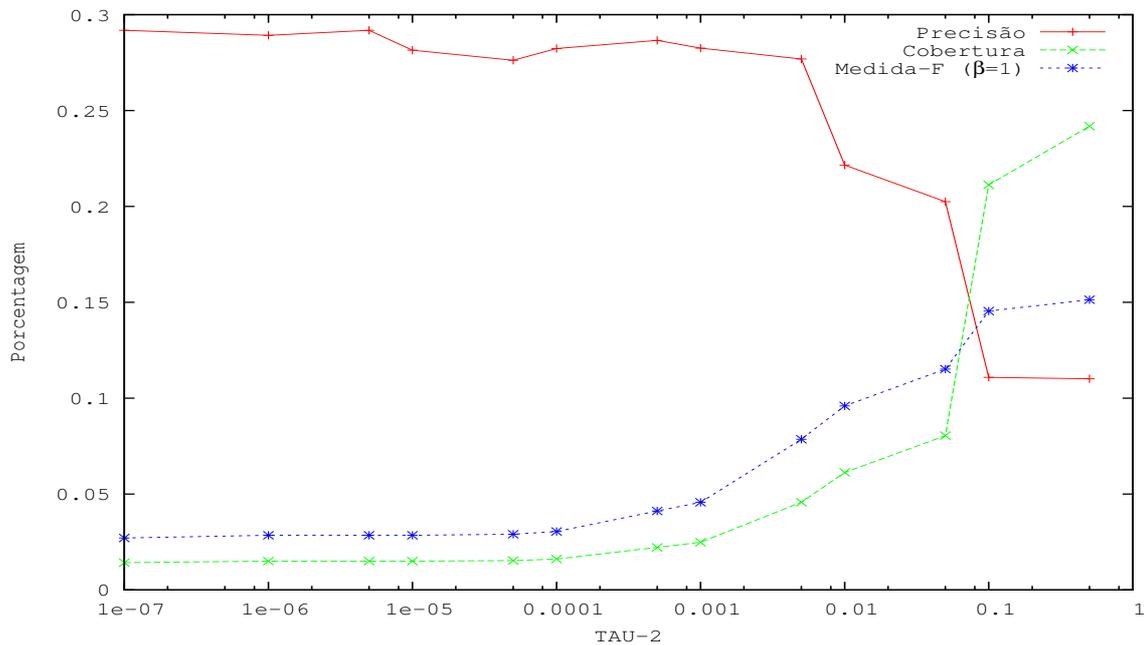


Figura 4.2: Avaliação do parâmetro τ_2 utilizando o extrator RGB com 64 *bins* por canal.

Os resultados das variações do parâmetro τ_2 para o extrator de características baseado no espaço de cor RGB podem ser observado nas Figuras 4.1, 4.2, 4.3 e 4.4.

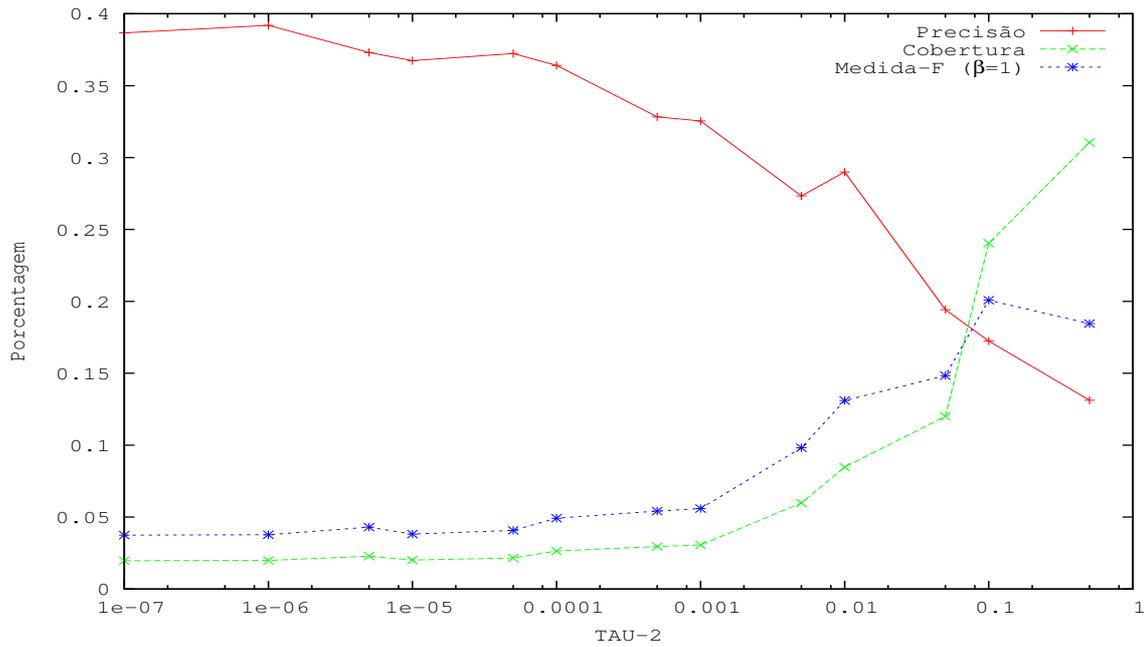


Figura 4.3: Avaliação do parâmetro τ_2 utilizando o extrator RGB com 32 *bins* por canal.

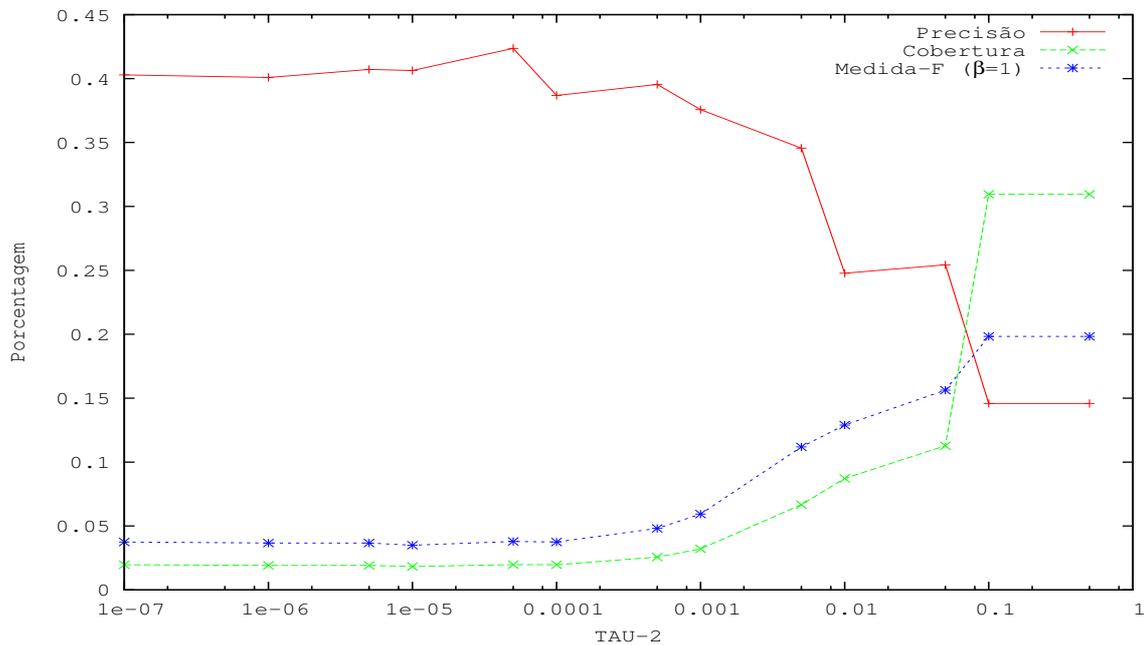


Figura 4.4: Avaliação do parâmetro τ_2 utilizando o extrator RGB com 16 *bins* por canal.

O gráfico na Figura 4.1 exibe os testes para os diferentes valores de τ_2 utilizando RGB com 128 *bins*, o melhor resultado do Medida-F foi obtido quando τ_2 assume o valor 0,05.

Para a rede que utiliza o extrator RGB com 64 *bins*, na Figura 4.2, o melhor valor para a Medida-F foi obtido com o valor 0,5, já os demais extratores RGB com 32 e 16 *bins* obtiveram seu melhor valor de Medida-F quando o valor de τ_2 foi de 0,1. Dos extratores baseados no espaço de cor RGB o que obteve melhor resultado foi o RGB com 32 *bins*, Figura 4.3, com uma Medida-F média de 0,20083396.

Os resultados do experimentos referentes ao espaço de HSV estão presentes nas Figuras 4.5, 4.6, 4.7 e 4.8.

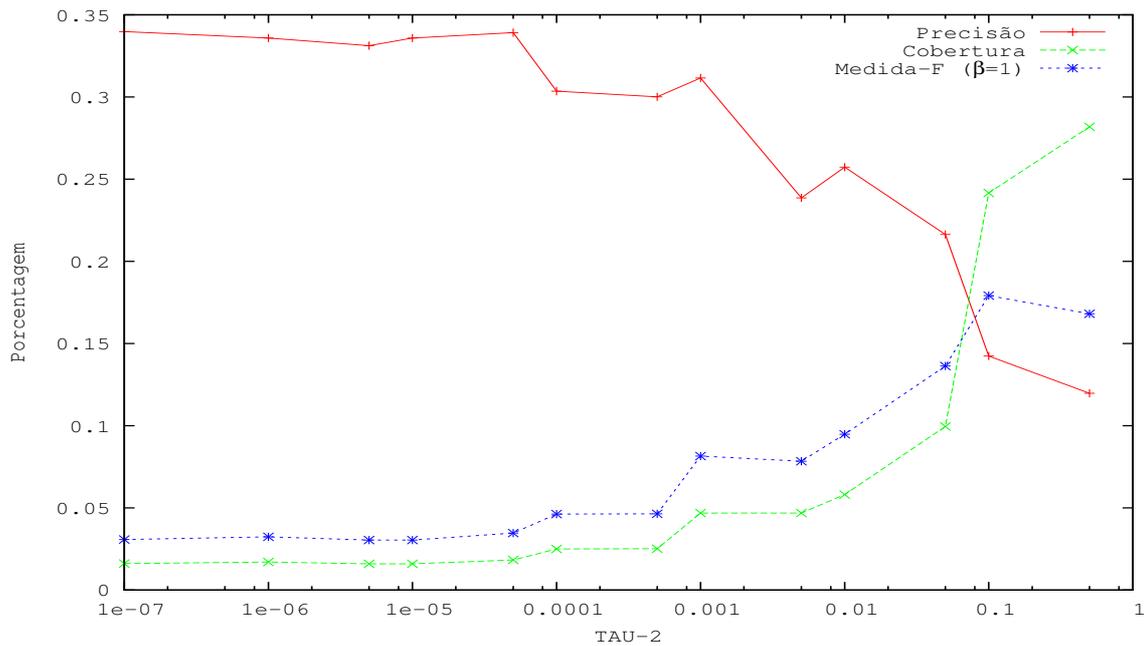


Figura 4.5: Avaliação do parâmetro τ_2 utilizando o extrator HSV com 128 *bins* por canal.

O melhor valor da Medida-F (0,21777563) considerando os diferentes níveis de quantização para o espaço de cor HSV, foi obtido pela rede treinada com 16 *bins*, presente no gráfico da Figura 4.8. O valor τ_2 selecionado para essa rede foi de 0,1.

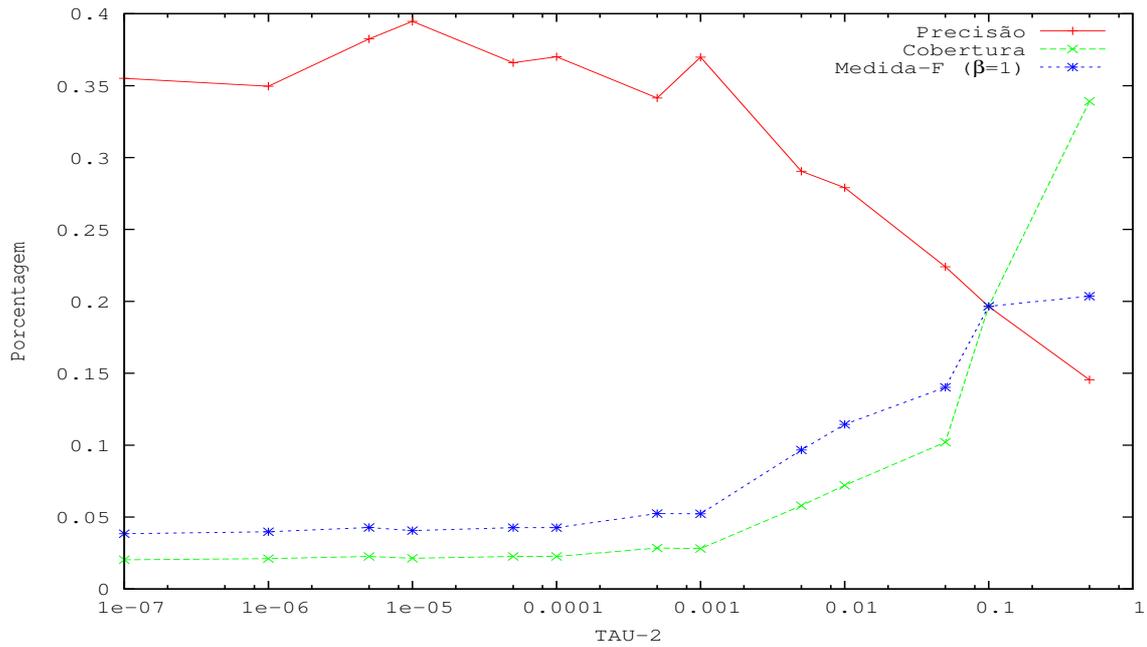


Figura 4.6: Avaliação do parâmetro τ_2 utilizando o extrator HSV com 64 *bins* por canal.

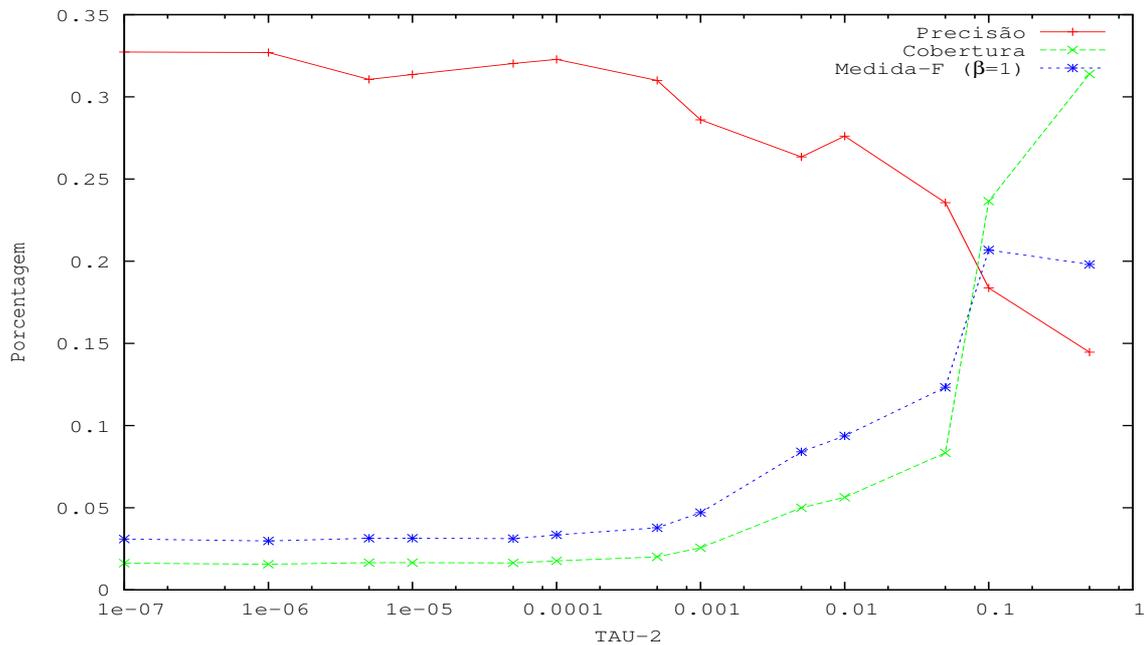


Figura 4.7: Avaliação do parâmetro τ_2 utilizando o extrator HSV com 32 *bins* por canal.

Já para os extratores baseados no espaço de cor YCbCr os melhores valores de τ_2 selecionados para todos os níveis de quantização foram todos iguais a 0,1, como pode ser observado

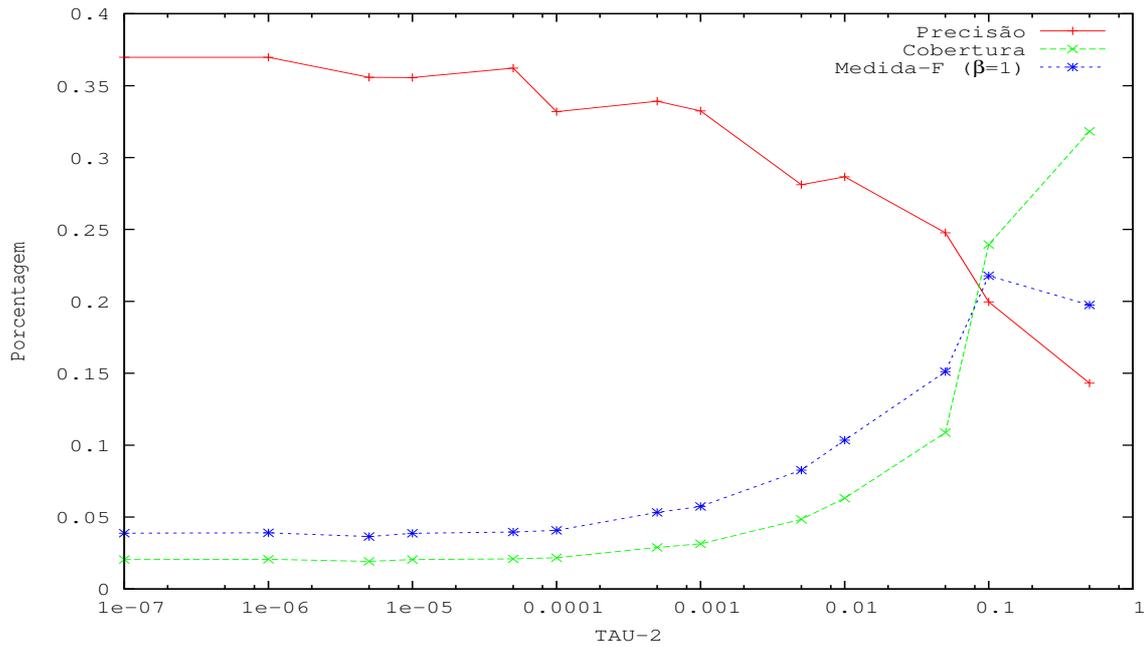


Figura 4.8: Avaliação do parâmetro τ_2 utilizando o extrator HSV com 16 *bins* por canal.

nos gráficos das Figuras 4.9, 4.10, 4.11 e 4.12.

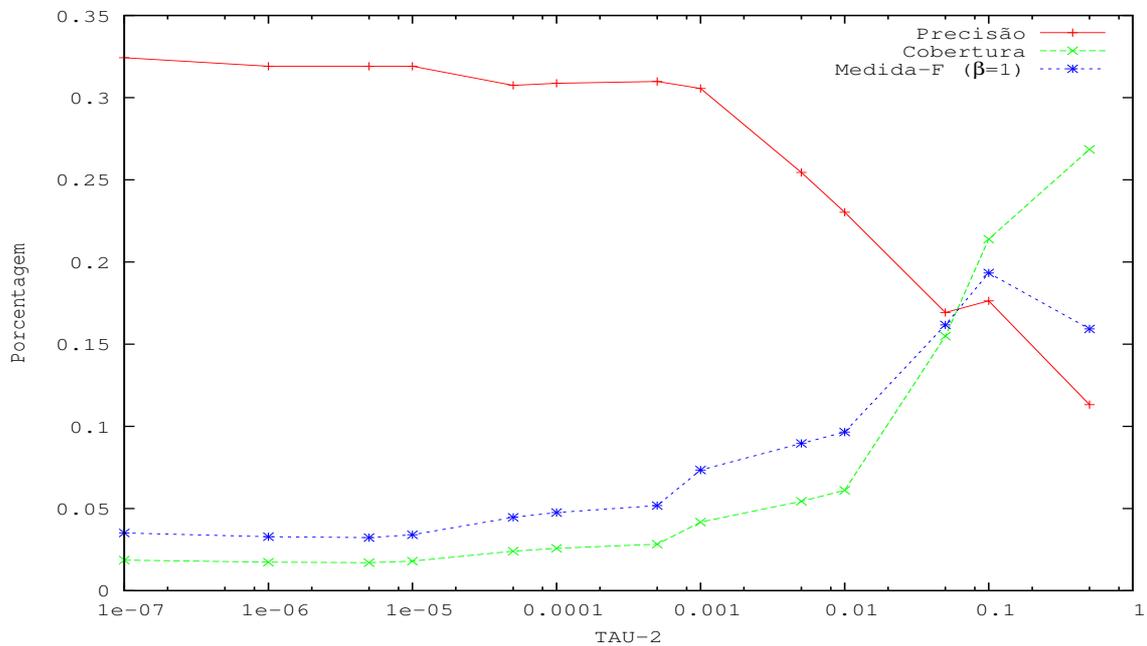


Figura 4.9: Avaliação do parâmetro τ_2 utilizando o extrator YCbCr com 128 *bins* por canal.

A rede GHSOM no espaço YCbCr com o melhor valor da Medida-F (0,217867218) foi

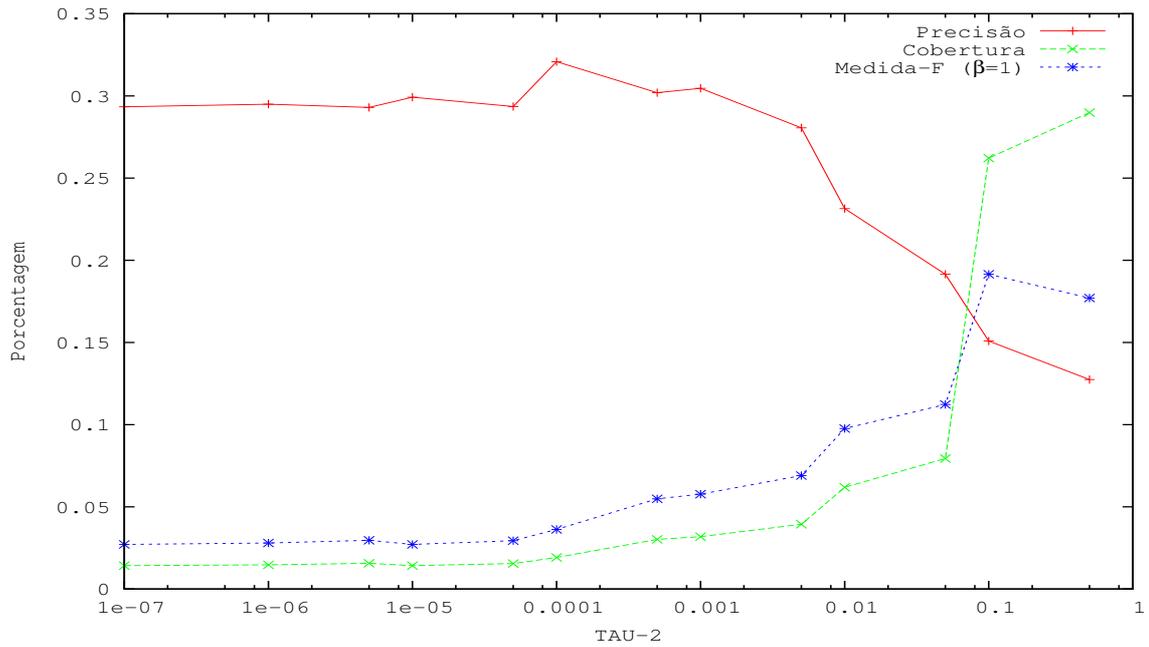


Figura 4.10: Avaliação do parâmetro τ_2 utilizando o extrator YCbCr com 64 *bins* por canal.

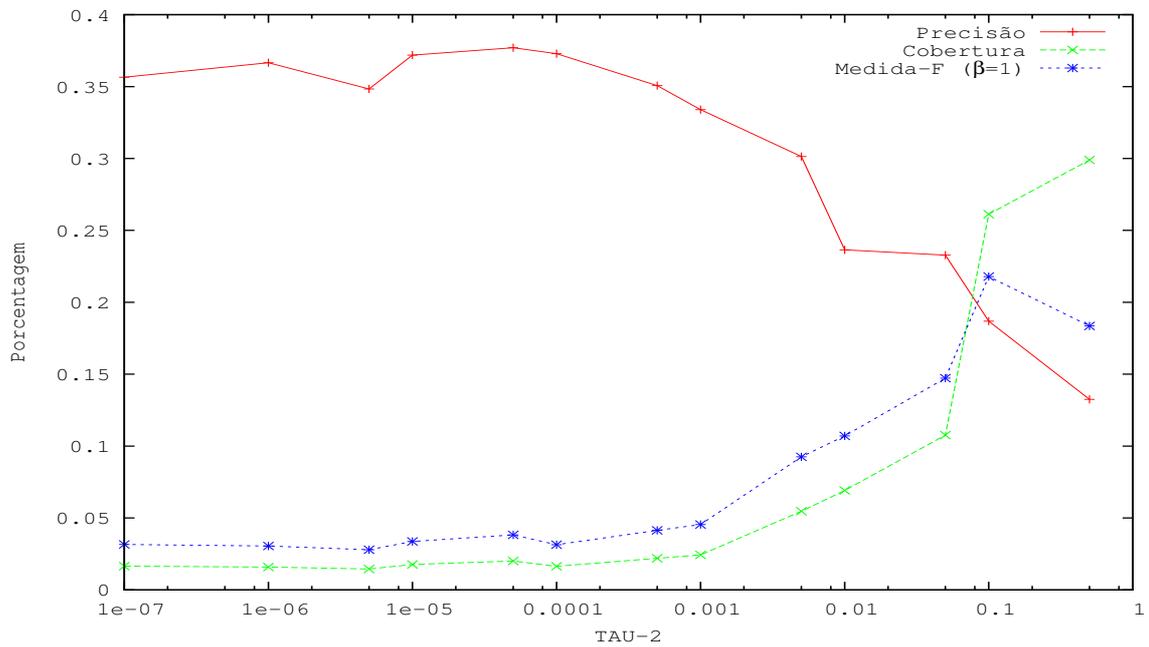


Figura 4.11: Avaliação do parâmetro τ_2 utilizando o extrator YCbCr com 32 *bins* por canal.

a que utilizou um extrator de características com 32 níveis de intensidade (*bins*), conforme apresenta a Figura 4.11.

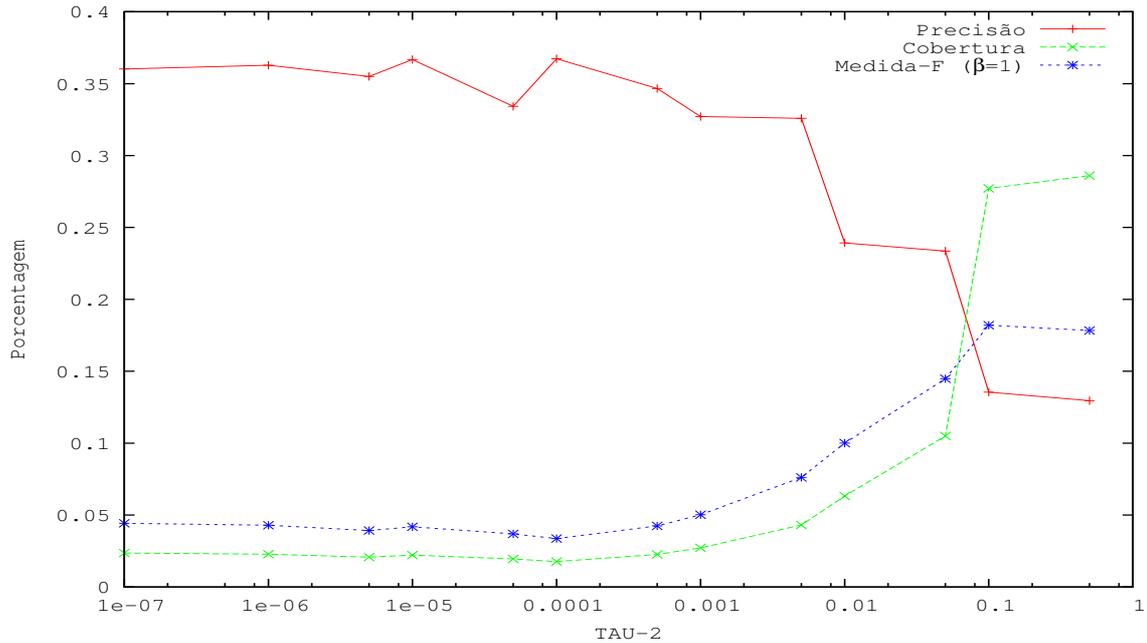


Figura 4.12: Avaliação do parâmetro τ_2 utilizando o extrator YCbCr com 16 *bins* por canal.

Para as características referentes à textura, foram utilizados como extratores dois métodos: as *Wavelets* (Daubechies e Haar) e a técnica LBP (*Local Binary Patterns*). O melhor valor da medida-F (0,195531029) para *Wavelets* foi obtido quando o parâmetro τ_2 possuía o valor de 0,1, como pode ser observado na Figura 4.13.

Para o extrator LBP foi necessário inicialmente definir alguns parâmetros específicos antes da execução dos testes. Foram definidos $P = 4$ (quantidade de *pixels* vizinhos) e $R = 1$ (raio), tais configurações são de grande importância para a velocidade de execução da extração das características e para o número de padrões criados para cada imagem (Seção 2.1.3). O resultado da execução dos testes com o LBP permitiu a seleção de τ_2 com o valor 0,05, obtendo um valor para Medida-F de 0,227855102. O resultado é ilustrado no gráfico da Figura 4.14.

Por último, foram realizados os testes referentes à característica de forma, para isso foi utilizado como técnica de extração os Momentos Invariantes de Hu. O melhor valor para o parâmetro τ_2 foi obtido com valor de 0,1, com uma Medida-F igual a 0,151617916. O resultado pode ser observado no gráfico da Figura 4.15.

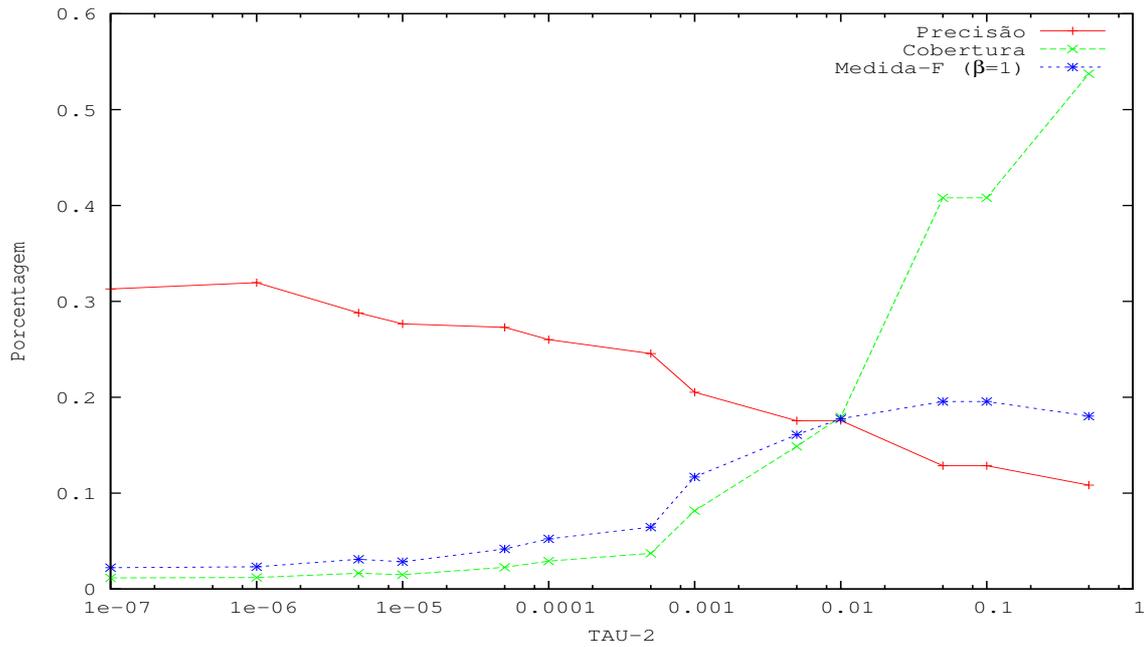


Figura 4.13: Avaliação do parâmetro τ_2 utilizando o extrator Wavelets.

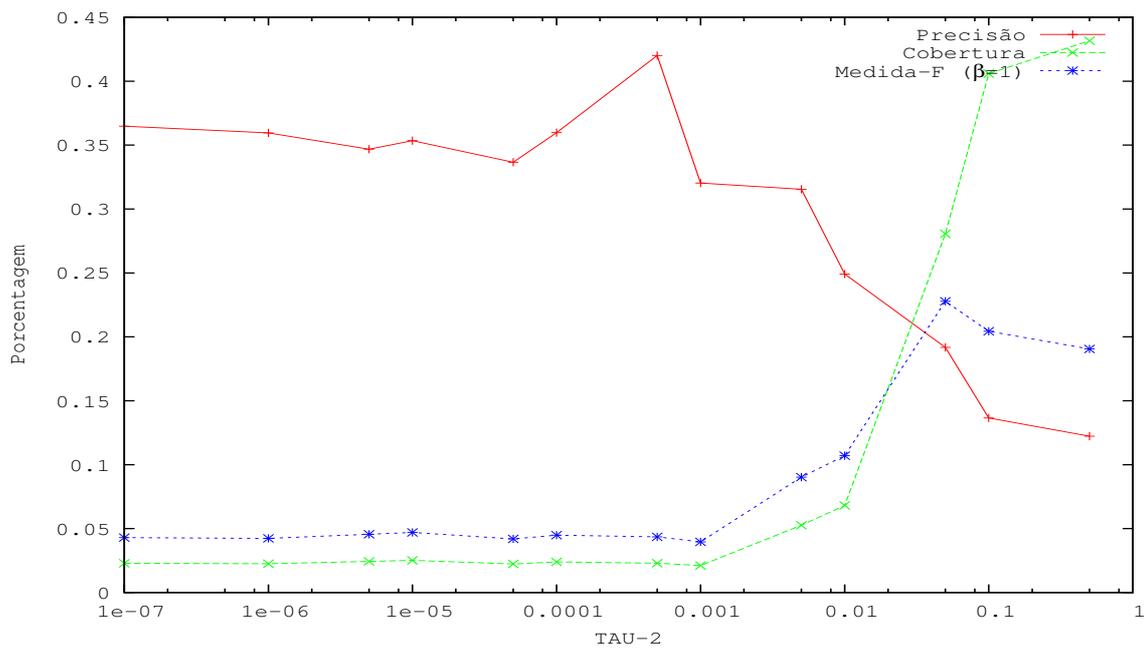


Figura 4.14: Avaliação do parâmetro τ_2 utilizando o extrator LBP.

A Tabela 4.1 sumariza todo o Experimento 1 com os valores de τ_2 selecionados ao longo dos testes, também são apresentadas a precisão e cobertura médias para cada tipo de extração de características.

Dentre os extratores utilizados para o treinamento, o que mais se destacou em relação à

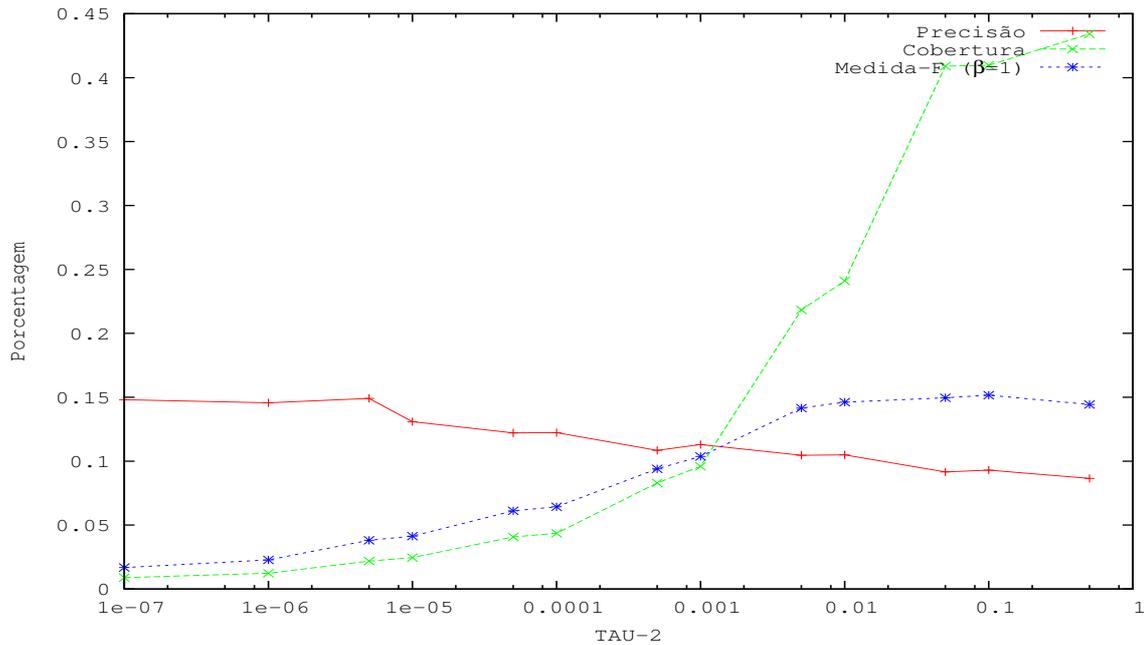


Figura 4.15: Avaliação do parâmetro τ_2 utilizando o extrator Momentos Invariantes de Hu.

métrica medida-F foi o LBP. Em relação à precisão o extrator que se destacou foi baseado no espaço de cor HSV utilizando 16 *bins* (níveis de intensidade) com um valor médio de 19,96%, já para cobertura o extrator que obteve o maior valor foi o baseado em forma, os Momentos Invariantes de Hu, com um valor médio de 40,97%. Os melhores extratores foram indicados pelo tom de cinza escuro na Tabela 4.1.

Foi observado durante a fase de teste das redes GHSOM que algumas imagens foram mapeadas em grupos maiores que outros ou em grupos totalmente diferentes dos seus rótulos originais. Mesmo o conjunto de treinamento sendo composto por grupos com o mesmo número de imagens, este comportamento indica que para alguns extratores certos grupos não são facilmente discriminados. Como exemplo verifica-se que para certos grupos, os elementos mais discriminantes das imagens relacionam-se a cor, e, portanto, foram erroneamente classificados pelos outros classificadores (baseados em forma ou textura).

Com base na observação acima, foi realizado um experimento em que se buscaram grupos que se destacaram em termos de precisão e cobertura. A Tabela 4.2 apresenta esses grupos, com respectivos extratores de características e taxas de precisão e cobertura. Por exemplo, a classe **futebol** por apresentar em suas imagens características mais salientes de cor, é melhor classificada quando são utilizadas redes alimentadas com características de cor (HSV 16,

Tabela 4.1: Parâmetros de τ_2 selecionados para cada extrator.

Extratores	τ_2	Precisão	Cobertura	Medida-F ($\beta = 1$)
LBP	0,05	0,19179	0,28062	0,22785
WAV	0,10	0,12856	0,40812	0,19553
HUMOMENTS	0,10	0,09302	0,40968	0,15161
HSV128	0,10	0,14234	0,24156	0,17913
HSV64	0,50	0,14548	0,33906	0,20360
HSV32	0,10	0,18372	0,23656	0,20682
HSV16	0,10	0,19964	0,23953	0,21777
RGB128	0,05	0,18307	0,16859	0,17553
RGB64	0,50	0,11012	0,24187	0,15133
RGB32	0,10	0,17241	0,24046	0,20083
RGB16	0,10	0,14586	0,30953	0,19828
YCBCR128	0,10	0,17631	0,21390	0,19330
YCBCR64	0,10	0,15086	0,26218	0,19152
YCBCR32	0,10	0,18684	0,26125	0,21786
YCBCR16	0,10	0,13552	0,27718	0,18204

YCBCR 32 e RGB 32), neste caso a precisão chegou a 100% quando utilizado uma GHSOM com características RGB com 32 *bins* por canal.

Tabela 4.2: Valores de precisão e cobertura para algumas classes.

Extratores	Classe	Precisão(%)	Cobertura (%)
HSV 16	futebol	83,93	29,00
YCBCR 32	futebol	78,57	2,75
RGB 32	futebol	100,00	2,75
LBP	motocicletas	78,13	1,75
WAVELETS	aviões	56,25	2,50
HUMOMENTS	folhas	14,16	100,00

Em classes onde a característica mais marcante das imagens é a textura aconteceram resultados similares. Por exemplo, a classe **motocicletas**, quando classificada por meio de

uma rede com características LBP, obteve uma taxa de precisão média de 78,13%, já a classe **aviões** (com menos texturas presentes) obteve uma taxa de precisão média de apenas 56,25% quando utilizado uma rede treinada com características *Wavelets*. As redes treinadas com os Momentos Invariantes de Hu obtiveram destaque na taxa de cobertura. Por exemplo a classe **folhas** o valor da cobertura média foi de 100%.

Os experimentos apresentados anteriormente produziram taxas médias de precisão e cobertura relativamente menores do que as dos resultados até então referenciados na literatura [BOLG05; PGB01; Edv06]. A partir de uma pesquisa mais minuciosa foram detectadas algumas particularidades nos testes até então realizados pelos sistemas existentes: os bancos de testes e treinamento geralmente são pequenos ou subdivididos em poucas classes que, na maioria das vezes são compostas por imagens facilmente diferenciadas (por exemplo, árvores, desertos e veículos) [Edv06].

4.2 Experimento 2 - Combinação de Classificadores

O experimento 2 teve com objetivo aplicar uma técnica de combinação de classificadores que busca uma melhoria do resultado referente à precisão.

Para o experimento foram utilizadas as redes GHSOM treinadas com os valores de τ_2 que obtiveram os melhores resultados de recuperação, baseando-se nas métricas de análise de resultados (Precisão, Cobertura e Medida-F). As redes escolhidas para cada uma das características extraídas foram: RGB com 32 *bins*, HSV com 16 *bins* e YCbCr com 32 *bins* para as redes que utilizam características de cor e para as características de textura e forma foram utilizadas as redes *Wavelets*, LBP e Momentos de Hu. Conforme resultados apresentados na Seção 4.1.

O processo de combinação de classificadores fundamentou-se no método de votação, no qual para cada imagem recuperada pelos classificadores selecionados é calculada um valor de voto. Esse valor é então utilizado para ordenar as imagens ao final da combinação, as mais bem votadas são classificadas como mais similares à imagem de consulta. O método de combinação está descrito em maiores detalhes no Capítulo 3.

Uma vantagem em utilizar o processo de combinação é fazer com que as deficiências de um determinado classificador sejam suprimidas pelo bom desempenho de outros classificadores.

Na execução dos testes foi utilizado o mesmo banco de imagens rotuladas do experimento anterior, desta forma, foi possível fazer uma comparação entre os resultados da classificação

utilizando redes GHSOM com extratores simples e a combinação desses classificadores.

Além do experimento envolvendo a combinação dos classificadores, também foi idealizado um teste de redução da quantidade de imagens recuperadas, esse teste teve como objetivo reduzir gradualmente o número de imagens recuperadas (iniciando com no máximo 400 e reduzindo para 300, 200, 100, 90, 80, 70, 60, 50, 45, 40, 35, 30, 25, 20, 15, 10 e 5 imagens recuperadas), na tentativa de reduzir imagens classificadas de forma incorreta (ou seja, aumentar a Precisão). Esse teste foi realizado tanto para os melhores classificadores simples como também para os classificadores combinados. Entretanto, já era esperado que, com a redução da quantidade de imagens recuperadas para classificadores simples, a precisão tenderia a ser constante, uma vez que as imagens mais relevantes do grupo recuperado estão espalhadas dentro do mesmo grupo e não ordenadas por voto como no método de combinação.

As combinações dos classificadores foram feitas de forma simples, combinando-os dois a dois e reduzindo gradativamente o número de imagens recuperadas ao final, os resultados das métricas de Precisão, Cobertura e Medida-F foram observados. Em seguida, foram realizados experimentos de combinação de classificadores três a três, quatro a quatro, cinco a cinco e seis a seis, também sendo ajustadas as quantidades máximas de imagens de recuperação para cada combinação.

Nos experimentos de combinação foram considerados apenas classificadores que obtiveram os melhores resultados de precisão e cobertura médias quando foram testados individualmente. Os classificadores selecionados foram, RGB com 32 *bins*, HSV com 16 *bins*, YCbCr com 32 *bins*, Momentos de Hu, LPB e *Wavelets*. Os classificadores selecionados tiveram seus resultados combinados e comparados segundo as métricas utilizadas anteriormente, de precisão, cobertura e medida-F. O gráfico apresentado na Figura 4.16 destaca o desempenho das combinações dos classificadores arranjados dois a dois em relação à precisão média, no qual se pode destacar a combinação dos extratores RGB com 32 *bins* e LBP e YCbCr 32 com *bins* e LBP (baseados em Cor e Textura).

As melhorias dos resultados da combinação podem ser observadas também em relação aos classificadores simples, por exemplo, os melhores classificadores quando combinados dois a dois são visivelmente melhores do que os classificadores executados de forma simples. Pode-se observar na Figura 4.17 o gráfico referente à precisão média versus a quantidade de imagens recuperadas, no qual à medida que se reduz o número de imagens recuperadas obteve-se uma melhoria na medida de Precisão.

Essa melhoria também está presente nos resultados referentes à métrica de cobertura

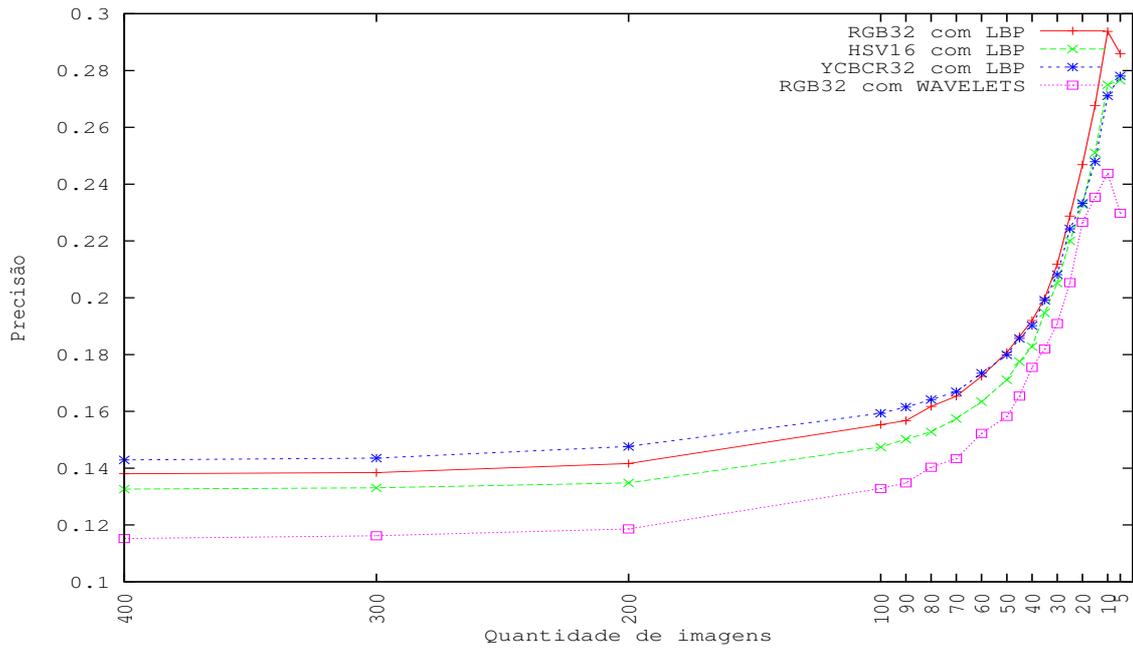


Figura 4.16: Precisões médias para combinação de classificadores dois a dois.

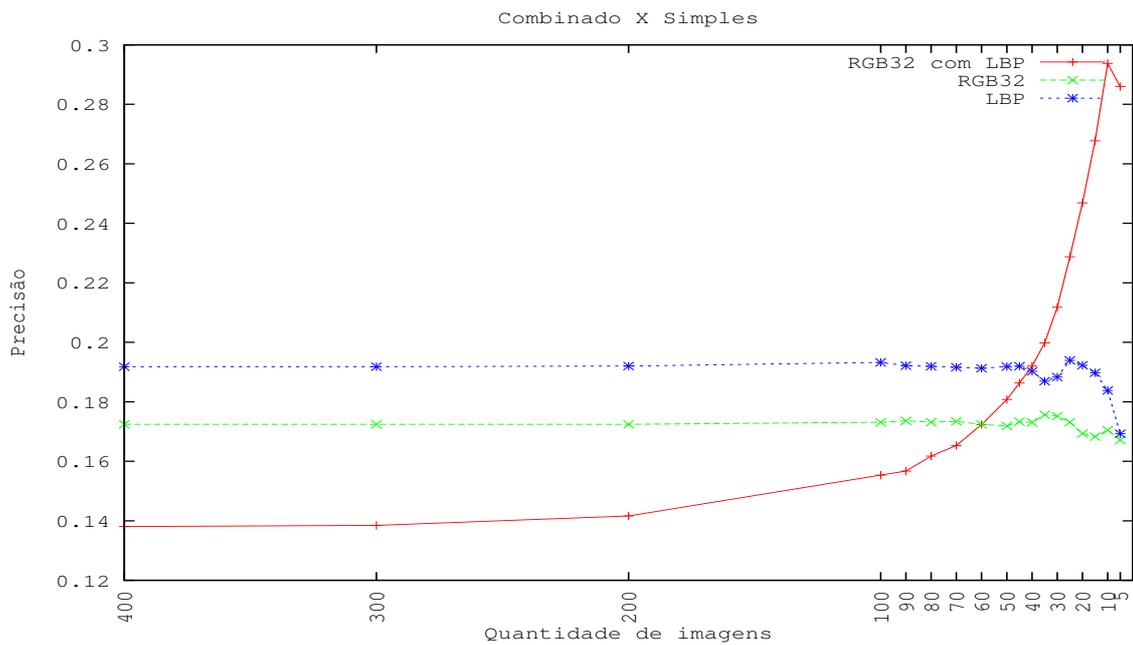


Figura 4.17: Precisão média para combinação de classificadores *versus* classificadores simples.

que decai mais lentamente, quando o número de imagens recuperadas é reduzido, para a combinação de classificação do que para os classificadores quando utilizados de forma individual. O gráfico da Figura 4.18 apresenta o resultado da cobertura para a combinação e os

classificadores individuais.

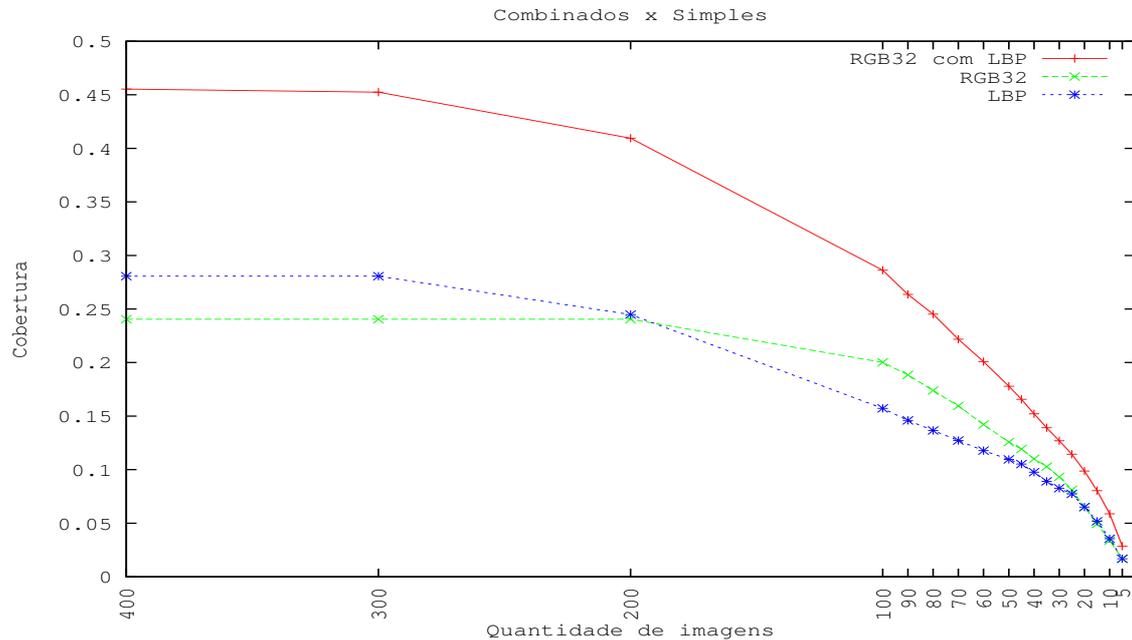


Figura 4.18: Cobertura para combinação de classificadores *versus* classificadores simples.

Para as combinações de classificadores arranjados três a três obteve-se destaque na precisão quando foram combinados os classificados HSV com 16 *bins*, RGB com 32 *bins* e LBP, resultando um valor máximo de 34,06% para a precisão média. Os classificadores HSV com 16 *bins*, YCbCr com 32 *bins* e LBP também obtiveram bons resultados relacionados à precisão média, decaindo apenas quando quantidade de imagens recuperadas foi menor do que 70 imagens. O gráfico da Figura 4.19 apresenta as quatro melhores combinações de resultados de classificadores arranjados três a três.

Em relação à cobertura média para a combinação arranjada três a três, as quatro melhores combinações obtiveram resultados bastante similares mesmo com a redução da quantidade de imagens recuperadas, o gráfico com as coberturas médias das combinações pode ser observado na Figura 4.20.

As combinações de classificadores arranjados quatro a quatro que produziram os melhores resultados foram a combinação dos classificadores HSV com 16 *bins*, RGB com 32 *bins*, LBP e *Wavelets*, com um valor de 37,97% de precisão quando a quantidade de imagens recuperadas foi igual a cinco, o gráfico pode ser observado na Figura 4.21.

Para a cobertura média, três combinações obtiveram resultados bastante similares, a pri-

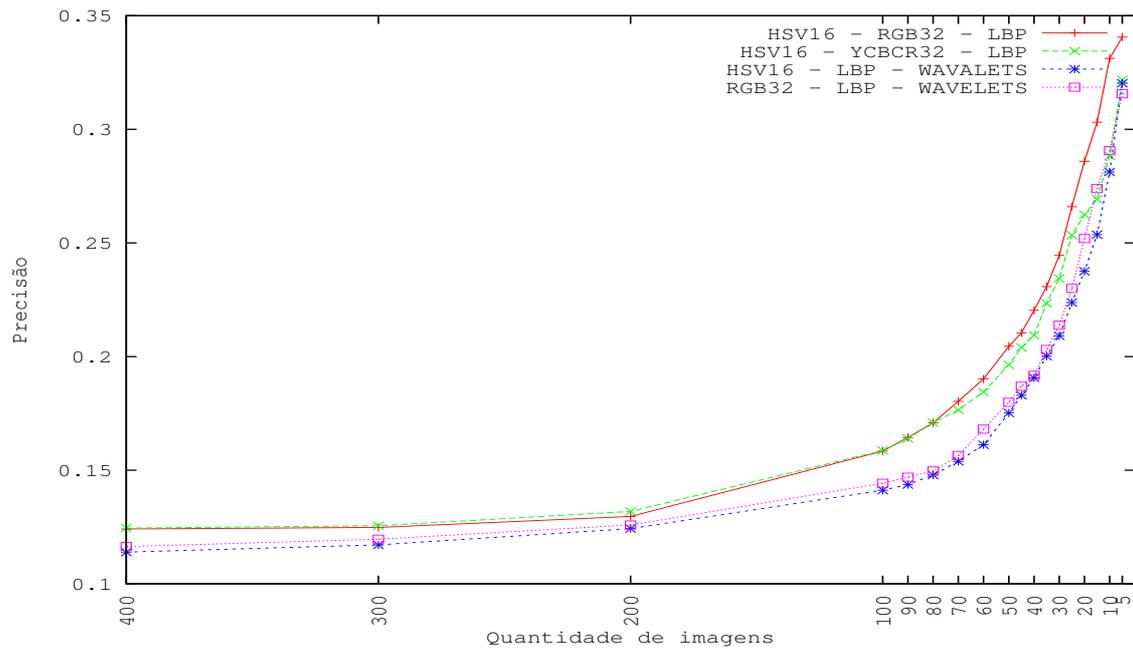


Figura 4.19: Precisões médias para a combinação de classificadores três a três.

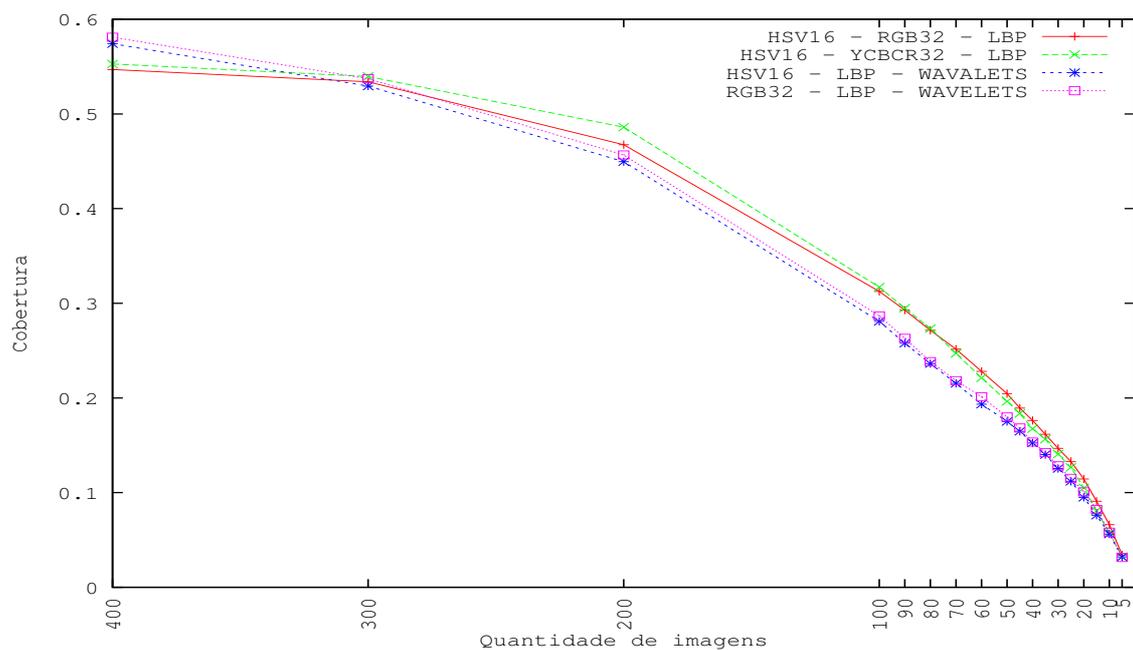


Figura 4.20: Precisões médias para a combinação de classificadores três a três.

meira combinação envolvendo HSV com 16 bins, RGB com 32 bins, LBP e Wavelets, a segunda com HSV com 16 bins, YCbCr com 32 bins, LBP e Wavelets e a última combinando HSV com 16 bins, RGB com 32 bins, YCbCr com 32 bins e LBP. O gráfico para este experimento é apresentado na Figura 4.22.

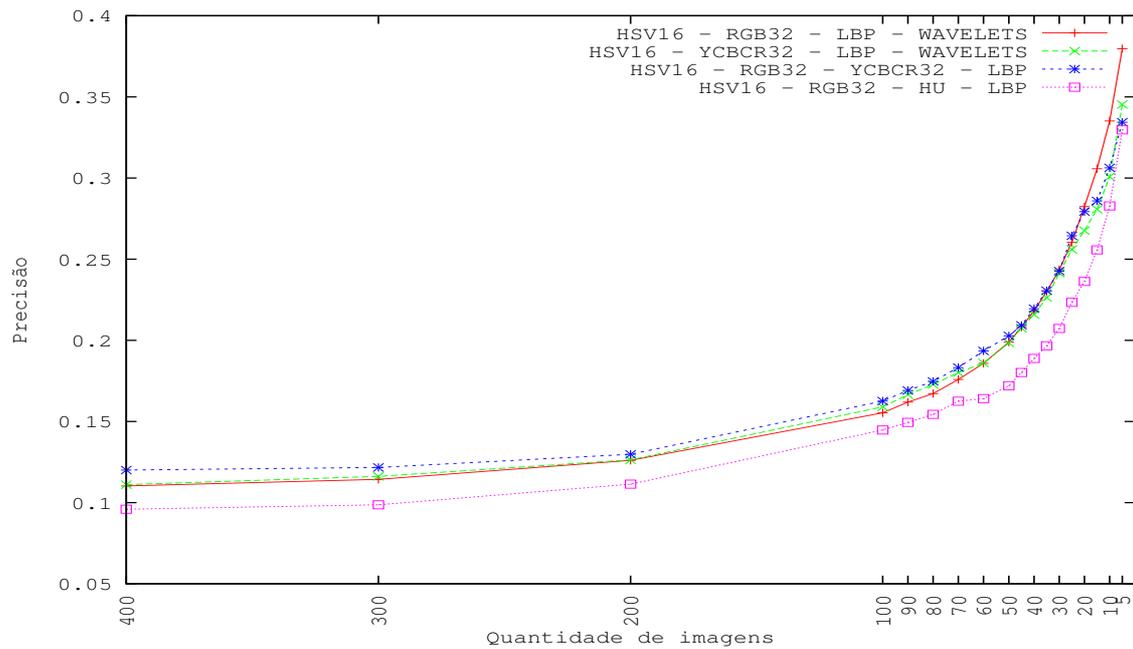


Figura 4.21: Precisões médias para a combinação de classificadores quatro a quatro.

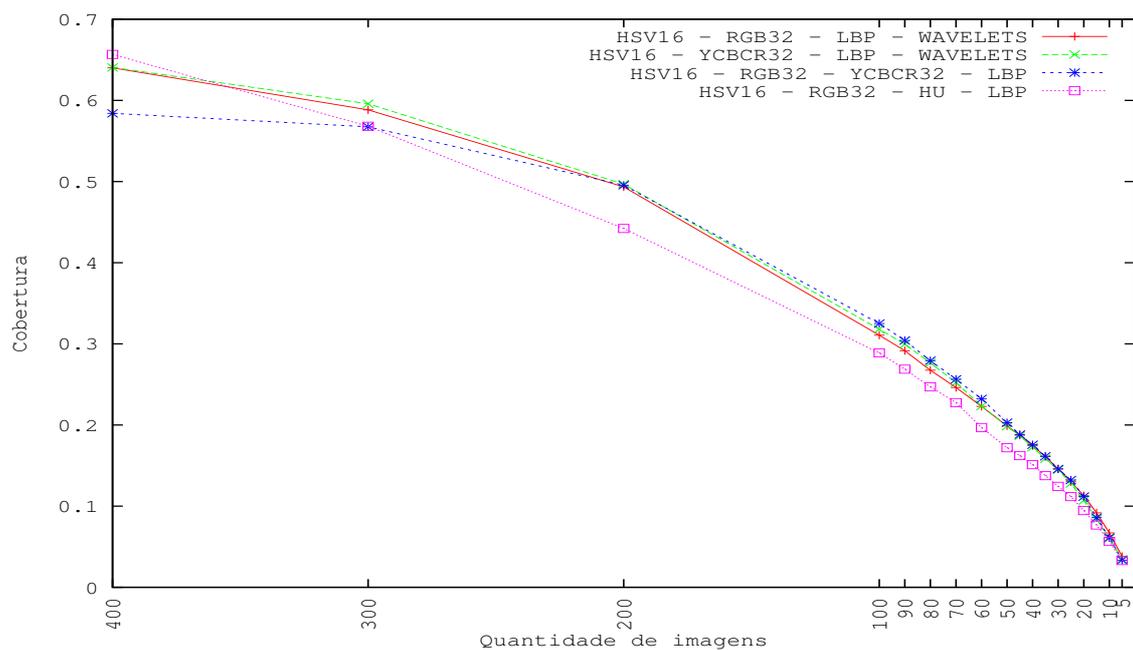


Figura 4.22: Coberturas médias para a combinação de classificadores quatro a quatro.

Nos classificadores combinados cinco a cinco, o destaque entre as melhores precisões médias foram para o arranjo dos classificadores HSV com 16 *bins*, RGB com 32 *bins*, YCbCr com 32 *bins*, LBP e *Wavelets*, que ao longo da variação da quantidade de imagens recuperadas obteve sempre melhores resultados do que as demais combinações cinco a cinco, obtendo a

maior precisão média de 34,69% quando a quantidade de imagens recuperadas era igual a cinco. Os resultados referentes a essa combinação podem ser observados no gráfico da Figura 4.23.

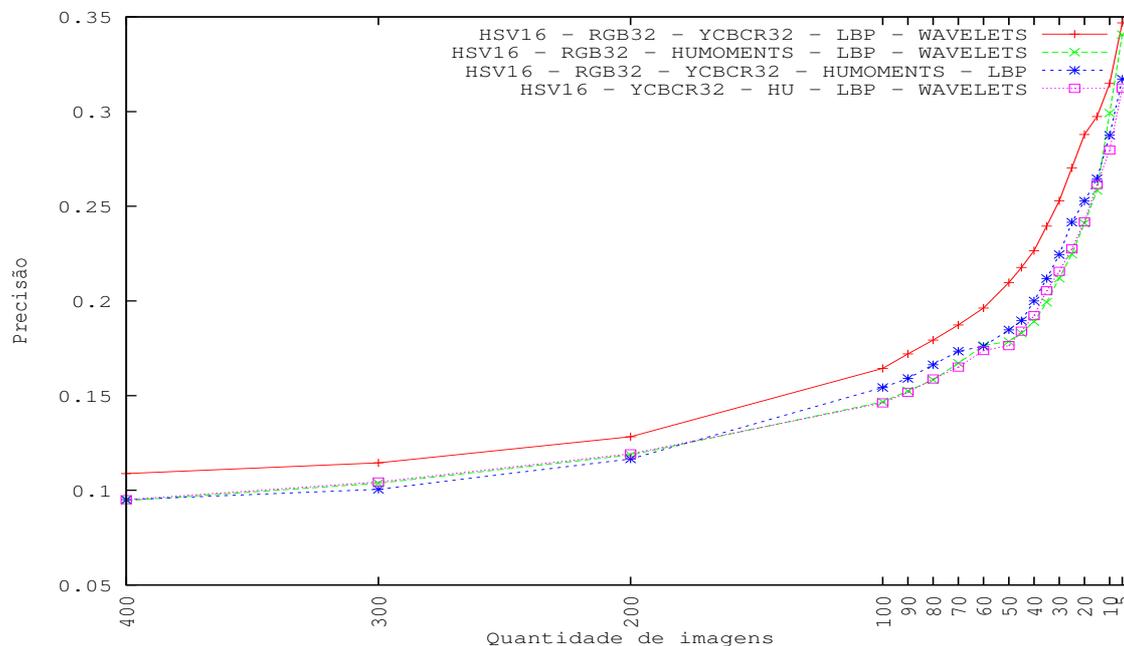


Figura 4.23: Precisões médias para a combinação de classificadores cinco a cinco.

Já para as coberturas médias, as quatro melhores combinações de classificadores arranjados cinco a cinco obtiveram resultados similares, mesmo com a redução da quantidade de imagens recuperadas os resultados entre as combinações decaíram em proporções semelhantes. Os resultados referentes às coberturas médias são apresentados no gráfico da Figura 4.24.

Por fim, foram combinados todos os classificadores utilizados (seis a seis), o melhor resultado da precisão média foi obtida quando a quantidade de imagens recuperadas era igual a cinco, com um valor de 33,13%. Já a cobertura média obteve o melhor resultado de 73,13% de acerto, quando a quantidade de imagens recuperadas era igual a 400. Os resultados analisados referentes à combinação de classificadores seis a seis pode ser observada no gráfico da Figura 4.25.

Os resultados do Experimento 2 foram de grande importância na validação da investigação sobre a melhoria dos resultados de precisão e cobertura com a utilização de combinação de diferentes tipos de classificadores. A redução gradativa da quantidade de imagens recuperadas na busca também foi de grande importância na qualidade dos resultados, além de reduzir o

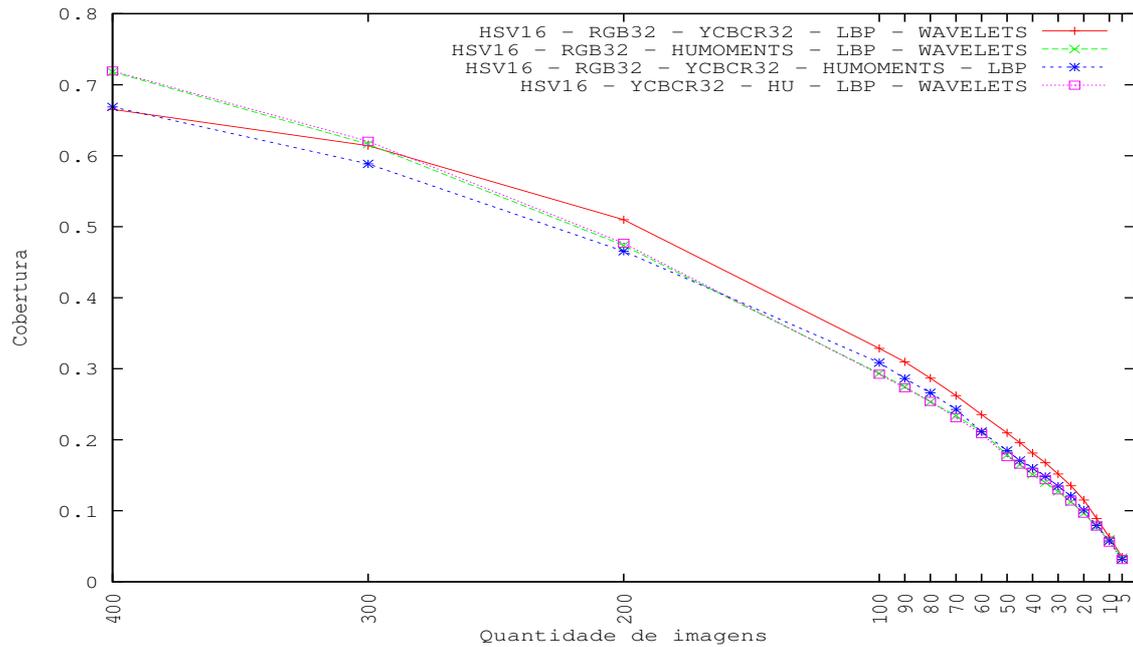


Figura 4.24: Coberturas médias para a combinação de classificadores cinco a cinco.

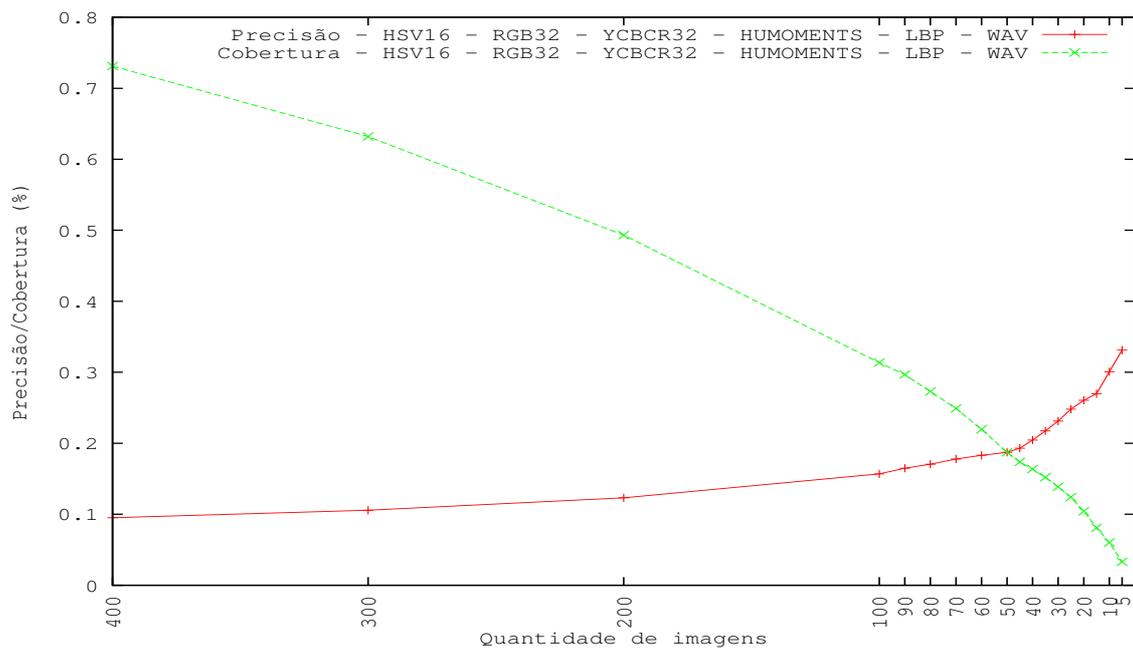


Figura 4.25: Precisões e coberturas médias para a combinação de classificadores seis a seis.

tempo de resposta do sistema para o usuário e fornecer resultados mais concisos.

Em comparação com as técnicas de classificação simples, foi observada uma melhoria considerável tanto para a precisão, como para a cobertura média. A técnica de classificação

que obteve a melhor taxa de precisão média quando de forma individual foi a que utilizava a característica de cor no espaço YCbCr com 32 *bins*. Inicialmente no experimento quando o número de imagens recuperadas foi superior a 70, o classificador simples se mostra mais preciso do que as combinações. Com a redução da quantidade de imagens recuperadas e a eliminação de imagens consideradas não tão similares a imagem de consulta, os resultados dos combinadores se mostram bastante superiores.

O destaque entre os classificadores combinados foi obtido pelo arranjo quatro a quatro, com os classificadores HSV com 16 *bins*, RGB com 32 *bins*, LBP e *Wavelets*, chegando a aumentar a precisão média em 17% quando a quantidade de imagens recuperadas é igual a cinco, em relação aos melhores classificadores simples. Mesmo com a combinação mais simples, os resultados se mostraram superiores quando o número de imagens recuperadas é inferior a 40. O gráfico da Figura 4.26 destaca os resultados das precisões médias comparando o melhor classificador simples com as melhores combinações em diferentes formas de arranjo.

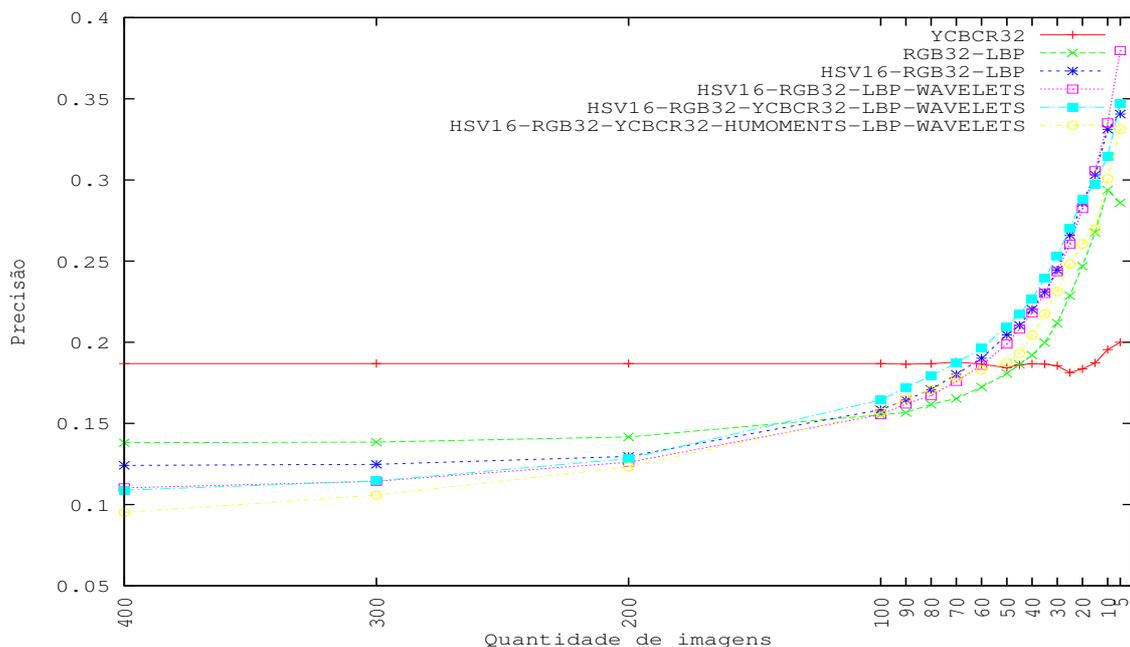


Figura 4.26: Precisões médias para comparar os classificadores e suas combinações.

Em relação à cobertura, a combinação mostrou-se superior à técnica de classificação simples desde o início do experimento, a melhor técnica combinada chegou a ser 25,56% superior do que a técnica de classificação simples. Para a classificação simples foi utilizado o classificador baseado em Momentos Invariantes de Hu (HUMOMENTS) que, entre os classificadores

simples, foi o classificador que obteve maior destaque em relação a cobertura média. Mesmo assim, as técnicas combinadas se mostraram mais precisas e abrangentes do que a classificação simples. Entre as combinações, a de maior destaque foi a que utilizava um arranjo de cinco a cinco e utilização HSV com 16 *bins*, RGB com 32 *bins*, YCbCr com 32 *bins*, LBP e *Wavelets*.

Os resultados podem ser observados no gráfico da Figura 4.27, no qual o melhor classificador simples e a melhor combinação de cada arranjo são comparadas em relação a métrica de cobertura.

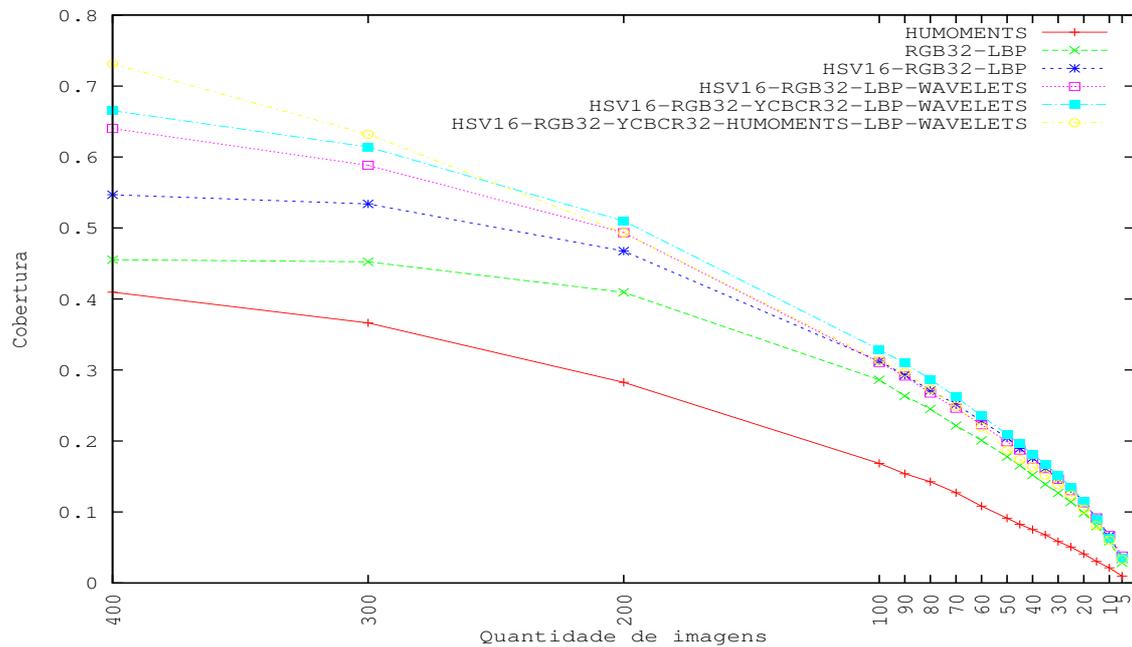


Figura 4.27: Coberturas médias para comparar os classificadores e suas combinações.

4.3 Experimento 3 - Teste com a Interface Web os Usuários potenciais

O Experimento 3 tem como objetivo apresentar os resultados da utilização do sistema proposto por um conjunto de usuários finais. Por intermédio da interface *Web* do sistema os usuários recuperam imagens similares uma imagem de consulta submetida ao sistema.

Para a recuperação, o sistema conta com um banco de imagens da *Web*, com aproximadamente 64000 URLs de imagens. Essas URL's foram obtidas através de um *Web Crawler*, descrito no Apêndice A. Cada URL de imagem encontrada pelo *Crawler* é submetida ao conjunto de regras para a aprovação da imagem. Essas regras verificam algumas especificações como altura, largura e área mínima da imagem. O banco de imagens é formado a partir das informações referentes às imagens encontradas pelo *Crawler*, como a URL em que a imagem se encontra e suas características extraídas utilizando-se das técnicas de extração por cor, forma e textura.

Os parâmetros selecionados para o treinamento das redes GHSOM foram os que obtiveram o melhor resultado de precisão nos testes com o banco rotulado, realizados na Seção 4.1. Na Tabela 4.3 são apresentados os valores de τ_2 selecionados para cada uma das redes GHSOM treinadas.

Os testes com os usuários foram subdivididos em três partes, o primeiro teste os usuários utilizaram o sistema seguindo um roteiro de testes, em que as imagens testadas no sistema foram selecionadas pelos usuários de forma aleatória. No segundo teste os usuários utilizariam o sistema do mesmo modo que o primeiro teste, mas utilizando um conjunto de imagens pré-selecionadas para cada classificador. Para o terceiro e último teste, os usuários utilizariam a mesma imagem em todos os classificadores.

Inicialmente o usuário deve fazer o *download* de uma imagem que será utilizada como imagem de consulta. Essa imagem é de conteúdo livre, desta forma, fica a critério do usuário que tipo de imagem que utilizar como consulta no sistema.

Após o *download* da imagem, o usuário acessa a interface de busca avançada do sistema, no qual ele configura todos os parâmetros dos extratores como é solicitado no roteiro de testes. Na Figura 4.28 é apresentada a tela de busca avançada.

Para executar a consulta, o usuário tem que selecionar o tipo de extrator que irá ser utilizado na recuperação e definir seus parâmetros se necessário. Depois desses passos, o usuário poderá realizar a recuperação das imagens.

Tabela 4.3: Valores de τ_2 selecionados para cada rede GHSOM.

Extratores	τ_2
LBP	0,0005000
WAV	0,0000010
HUMOMENTS	0,0000050
HSV64	0,0000100
HSV32	0,0000001
HSV16	0,0000010
RGB64	0,0000050
RGB32	0,0000010
RGB16	0,0000500
YCBCR64	0,0001000
YCBCR32	0,0000500
YCBCR16	0,0001000

O resultado da recuperação é exibido numa tela, na qual as imagens retornadas são apresentadas de forma decrescente, isto é, das mais similares para as menos similares à imagem de consulta (considerando a ordem de leitura de cima para baixo e da esquerda para a direita). Como forma de se fornecer dados relevantes à análise dos resultados, os usuários finais devem selecionar as imagens mais relevantes retornadas a partir de um *checkbox*. Desta forma, é possível calcular a precisão bem como classificar os melhores extratores de características. A tela de resultados é ilustrada na Figura 4.29.

Cada usuário que utilizou o sistema repetiu o teste de consulta de imagens para todas as redes GHSOM treinadas, são elas: HSV com 16 *bins*, HSV com 32 *bins*, HSV com 64 *bins*, RGB com 16 *bins*, RGB com 32 *bins*, RGB com 64 *bins*, YCbCr com 16 *bins*, YCbCr com 64 *bins*, LBP, Wavelets, Momentos Invariantes de Hu, combinação de todos os extratores de cor (RGB, HSV e YCbCr com 16 *bins*), combinação de todos os extratores de textura (LBP e Wavelets), combinação dos extratores de cor e textura, combinação dos extratores de cor e forma, combinação dos extratores de textura e forma, combinação dos extratores de cor, textura e forma.

É importante salientar que o contexto de similaridade no caso desse experimento é totalmente exclusivo de cada usuário, desta forma o usuário deverá julgar uma imagem como

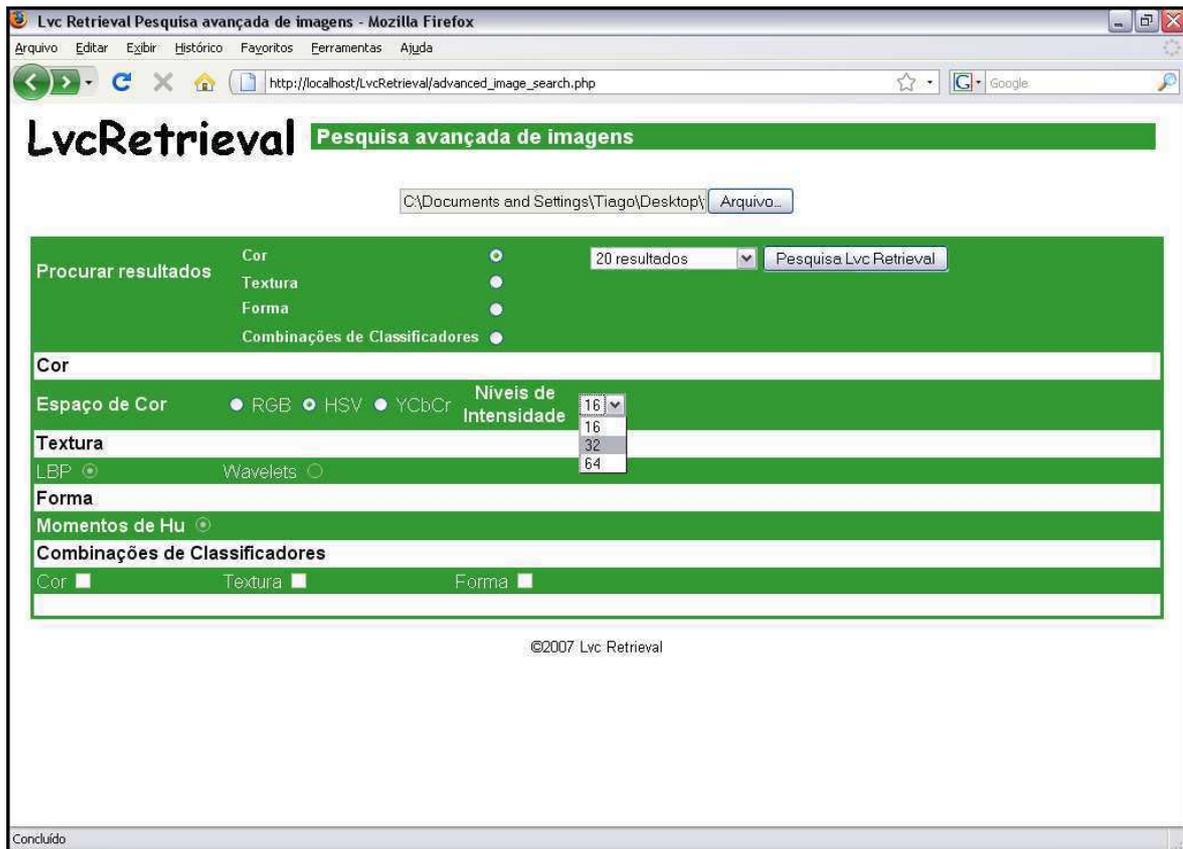


Figura 4.28: Tela de configurações avançadas do sistema de recuperação.

similar a imagem de consulta em decorrência apenas das características de cor, forma, textura ou a partir da combinação dessas características.

4.3.1 Experimento com Usuários - Imagens Aleatórias

Para o experimento com as imagens aleatórias foram selecionados 25 usuários, todos do Departamento de Sistemas e Computação da Universidade Federal de Campina Grande, os quais utilizaram o sistema e realizaram os testes como descrito no roteiro do Apêndice D. O tempo médio de utilização para cada usuário foi de aproximadamente 22 minutos.

Para a obtenção dos dados de testes, foram registrados no banco de dados o horário e a data da utilização do sistema para cada teste realizado, como também o endereço IP (*Internet Protocol*) da máquina do usuário. Além do número de imagens selecionadas pelo o usuário como classificadas corretamente pelo sistema, desta forma, foi possível calcular a precisão média e identificar os classificadores com os melhores desempenhos na recuperação.

No gráfico 4.30 são apresentados os resultados do primeiro teste com os usuários.

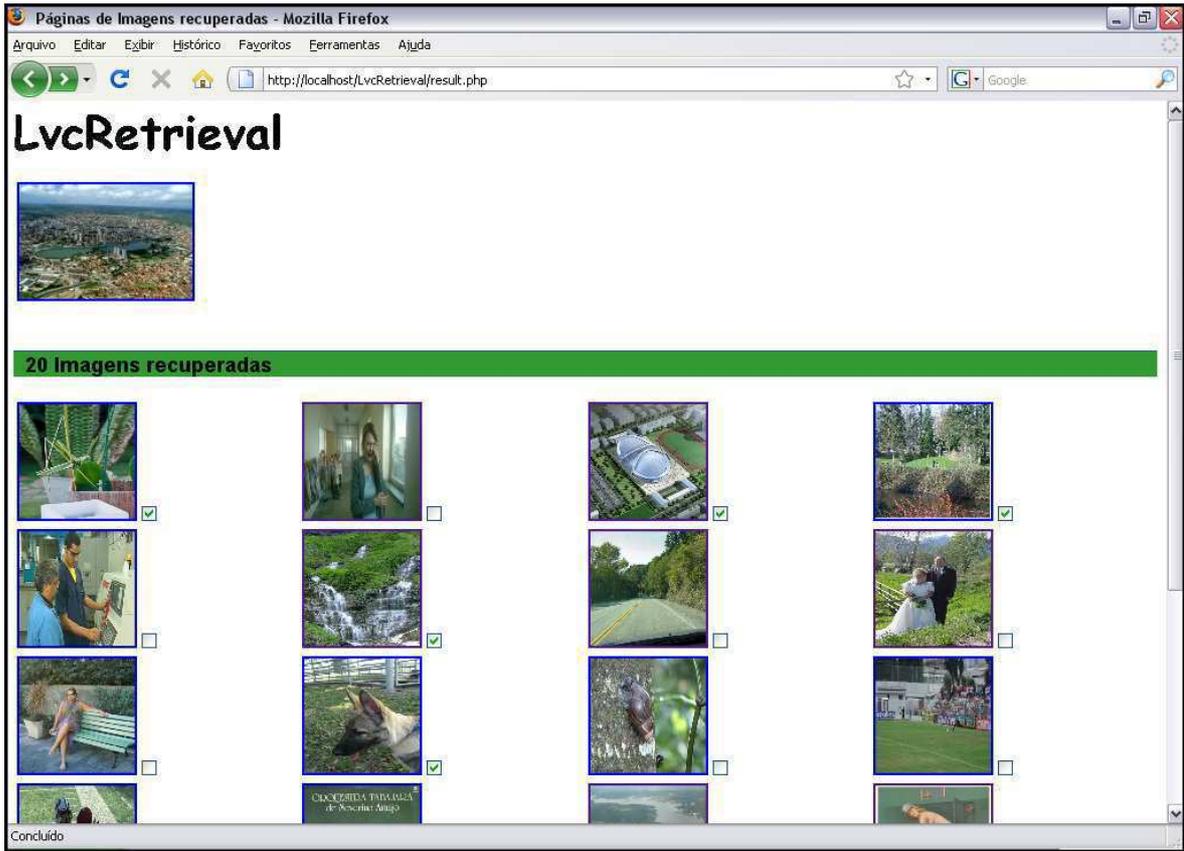


Figura 4.29: Tela de resultados recuperados do sistema de recuperação.

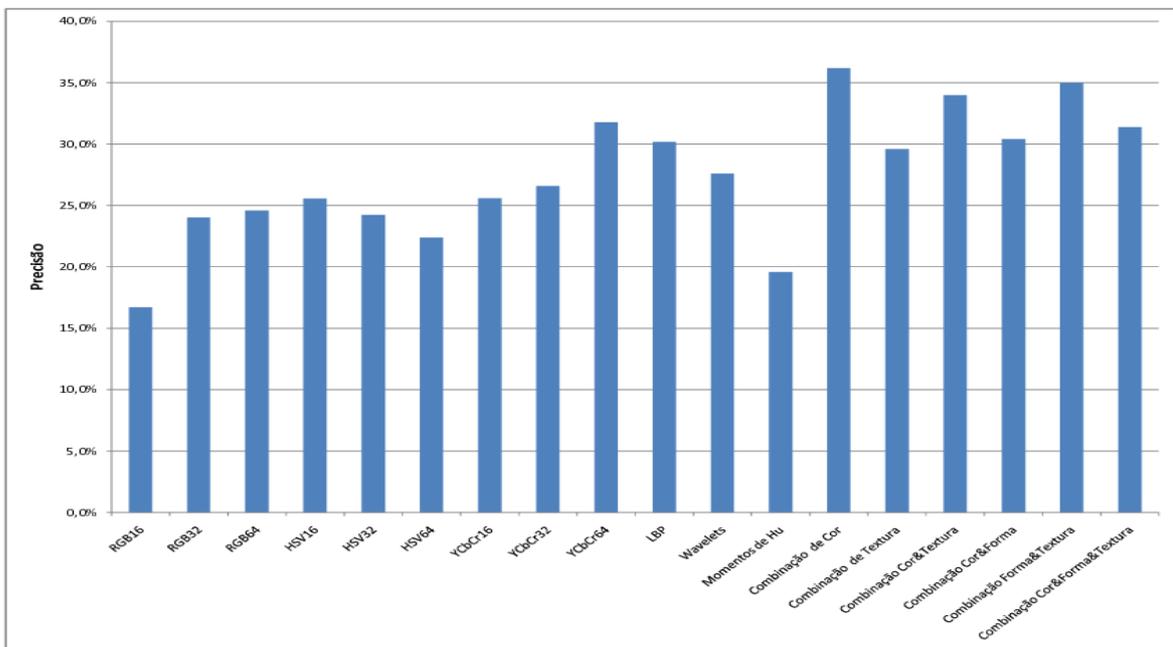


Figura 4.30: Precisões médias para todos os usuário.

De uma forma diferente dos experimentos realizados anteriormente na Seção 4.2, nos quais foram utilizadas imagens rotuladas, os experimentos que envolvem os usuários possui um grau de dificuldade maior para a interpretação dos resultados. Esta dificuldade está relacionada às diferentes formas de interpretação ou julgamento subjetivo que cada usuário pode fazer. Por exemplo, uma imagem recuperada pelo sistema pode ser considerada certa por alguns usuários e por outros não.

O resultado exposto no Gráfico 4.30 ilustra a superioridade de alguns classificadores combinados, como por exemplo as combinações de cor (RGB, HSV e YCbCr) que obtiveram uma precisão média de 36,20%, mesmo assim alguns classificadores simples obtiveram bons resultados como o classificador baseado em cor que utiliza o espaço YCbCr com 64 *bins* por canal, que apresentou uma precisão média de 31,80%.

Um problema encontrado após a execução dos testes foi a identificação de possíveis usuários incomuns durante a realização do experimento. Definimos esses usuários como pessoas que por algum motivo não utilizaram critérios coerentes para a escolha das imagens, podendo dessa forma prejudicar ou ajudar os resultados finais do experimento. Como as imagens de consulta utilizadas eram diferentes e os critérios para a escolha dependeram do julgamento de cada usuário, foi praticamente impossível, neste experimento, identificar os usuários incomuns.

Desta forma, foram propostos mais dois novos testes envolvendo usuários, de forma que fosse possível identificar precisamente possíveis usuários incomuns. Adicionalmente, objetivou-se obter dados mais conclusivos sobre a aplicação da técnica de combinação de classificadores *versus* os classificadores simples.

Os dois novos testes elaborados são melhor detalhados nas subseções a seguir.

4.3.2 Experimento com Usuários - Conjunto de Imagens Fixas

Para a execução do segundo teste com os usuários, foi elaborado um novo roteiro, de forma que os usuários testassem o sistema sem precisar configurar os parâmetros dos extratores. Esse novo roteiro inicialmente tinha como objetivo reduzir ao máximo o tempo em que o usuário iria testar o sistema, mas também serviu como meio de evitar possíveis interferência nos critérios do usuário. Uma interferência foi encontrada no formato do experimento anterior, já que o usuário tinha acesso ao tipo de extrator estava utilizando, dessa forma ele poderia ser influenciado selecionar imagens pela característica usada. Por exemplo, observar apenas as cores das imagens quando o classificador utilizado é baseado em cor e não na semântica

da imagem.

Além do novo roteiro de testes, foi criada uma nova interface de teste, deixando o teste mais simples. A nova interface consiste em apenas um botão para o *upload* da imagem de consulta, ilustrado na Figura 4.31, deixando para o sistema a responsabilidade da seleção do extrator e a configuração dos seus parâmetros. A tela de resultados é igual à do experimento anterior, apresentado na Figura 4.32.

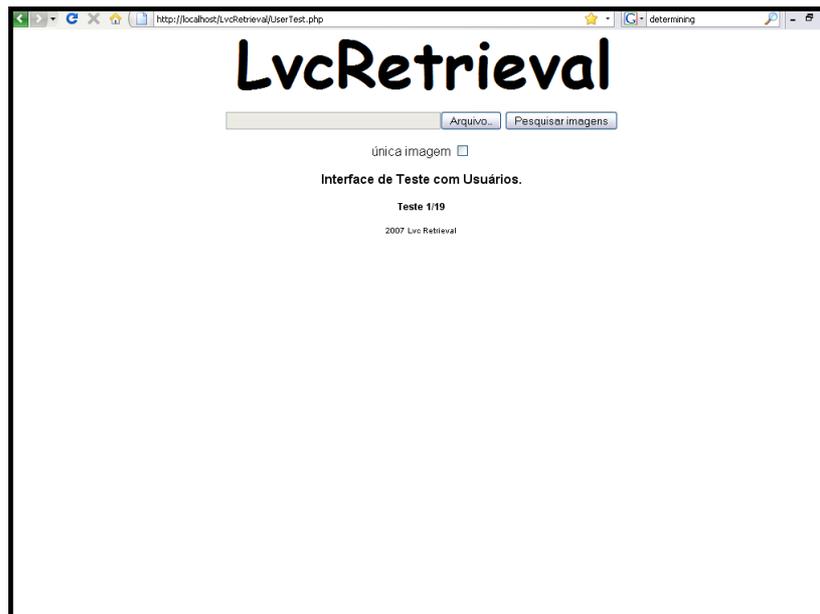


Figura 4.31: Tela inicial dos novos testes com os usuários.

Para a execução do segundo experimento com os usuários foi determinada a seguintes seqüência de classificadores a serem testados: HSV com 16 *bins*, HSV 32 com *bins*, HSV com 64 *bins*, RGB com 16 *bins*, RGB com 32 *bins*, RGB com 64 *bins*, YCbCr com 16 *bins*, YCbCr com 32 *bins*, YCbCr com 64 *bins*, Momentos de Hu, LBP, *Wavelets*, Combinação de Cor, Combinação de Textura, Combinação Cor e Forma, Combinação Cor e Textura, Combinação Forma e Textura, Combinação Cor, Forma e Textura e Melhor Combinação. Foi denominada de Melhor Combinação, a combinação que obteve melhor resultado na Seção 4.2, nesse caso foram os classificadores RGB com 32 *bins*, HSV com 16 *bins*, LBP e *Wavelets*.

Todos os esses testes são executados de forma transparente para o usuário, isto quer dizer, a imagem de consulta é submetida ao sistema, os resultados da recuperação são exibidos e os usuários não sabem que tipo de classificador foi utilizado, desta forma não influenciando o resultado final.

Além disso, foram selecionadas de forma aleatória 19 imagens que foram utilizadas nos

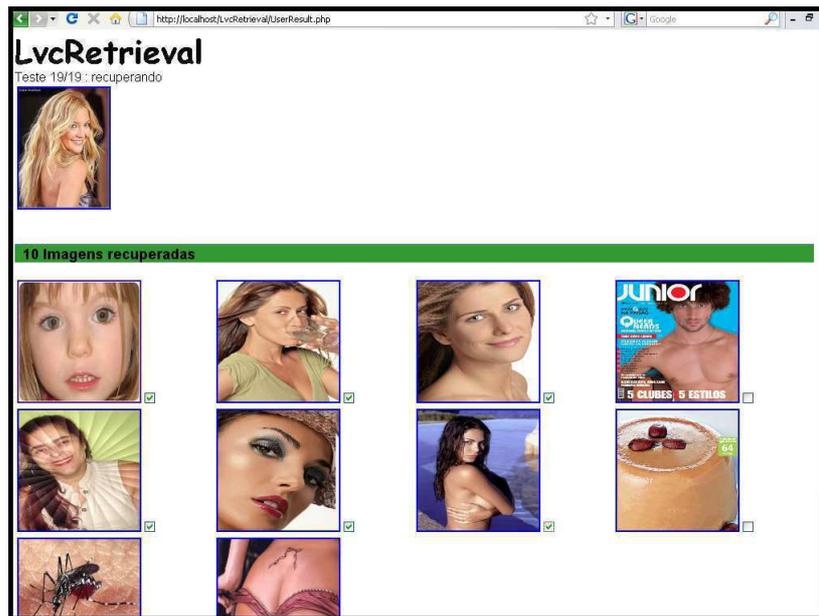


Figura 4.32: Tela de resultados da interface de testes com os usuários.

testes. Cada imagem foi vinculada ao teste (combinação específica de classificadores), desta forma, todos os usuários testam a mesma imagem para cada teste, possibilitando a detecção de possíveis usuários incomuns. Na Figura 4.33 são apresentadas as imagens utilizando o ordenamento esquerda para a direita e de cima para baixo, uma imagem diferente para cada classificador: RGB com 16 *bins*, RGB com 32 *bins*, RGB com 64 *bins*, HSV com 16 *bins*, HSV com 32 *bins*, HSV com 64 *bins*, YCbCr com 16 *bins*, YCbCr com 32 *bins*, YCbCr com 64 *bins*, LBP, Wavelets, Momentos Invariantes de Hu, Combinação de Cor, Combinação de Textura, Combinação Cor e Forma, Combinação Cor e Textura, Combinação Forma e Textura, Combinação Cor, Forma e Textura e Melhor Combinação.

Foram selecionados para o experimento 35 usuários, que utilizaram nos testes as imagens determinadas acima. No Gráfico da Figura 4.34 são apresentadas as precisões médias para cada um dos classificadores utilizados. O Apêndice C apresenta os resultados da classificação para cada imagem utilizada nos testes.

Os três classificadores que mais se destacaram nesse experimentos foram as combinações, com 48% de precisão a Melhor Combinação (RGB32, HSV16, LBP e Wavelets), em seguida a combinação de cor e textura com 39,1% de precisão média e por fim a combinação de cor (RGB, HSV e YCbCr), com 38,9% de precisão média. Mesmo com a predominância dos classificadores combinados, alguns classificadores simples também se destacaram, como o YCbCr com 16 *bins* e o HSV com 16 *bins* com uma precisão média de 35,7% e 32,0%

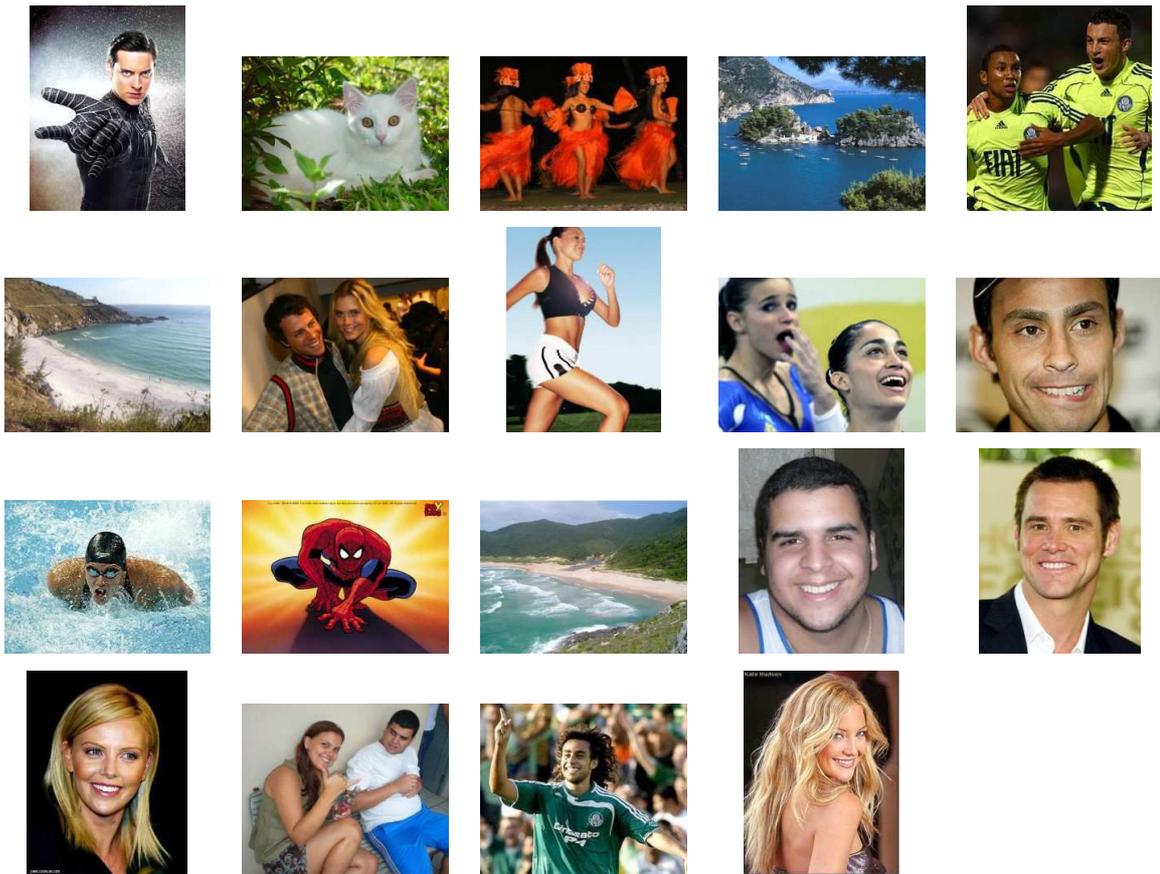


Figura 4.33: Imagens utilizadas nos testes com os usuários.

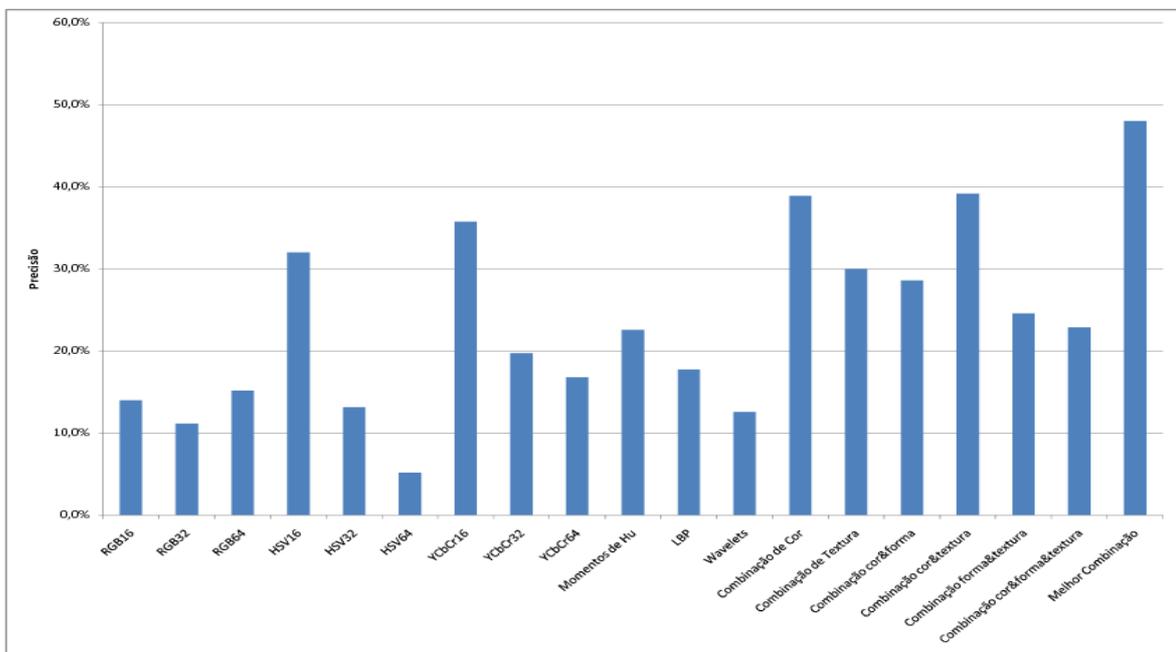


Figura 4.34: Precisões para cada classificador utilizado.

respectivamente.

Para analisar os usuários, foi utilizada a precisão média como medida determinante, assim, os usuários cujas precisões médias estivessem fora de um intervalo, pré-definido em torno da média geral das precisões de todos os usuários, foram classificados como incomuns. Qualquer usuário apresentando uma diferença (em módulo) entre a sua precisão média e a média geral das precisões superior a 23,5%, indicado pela linha vermelha no gráfico da Figura 4.35, foi classificado como incomum. No gráfico da Figura 4.35 são apresentadas as precisões médias para cada usuário.

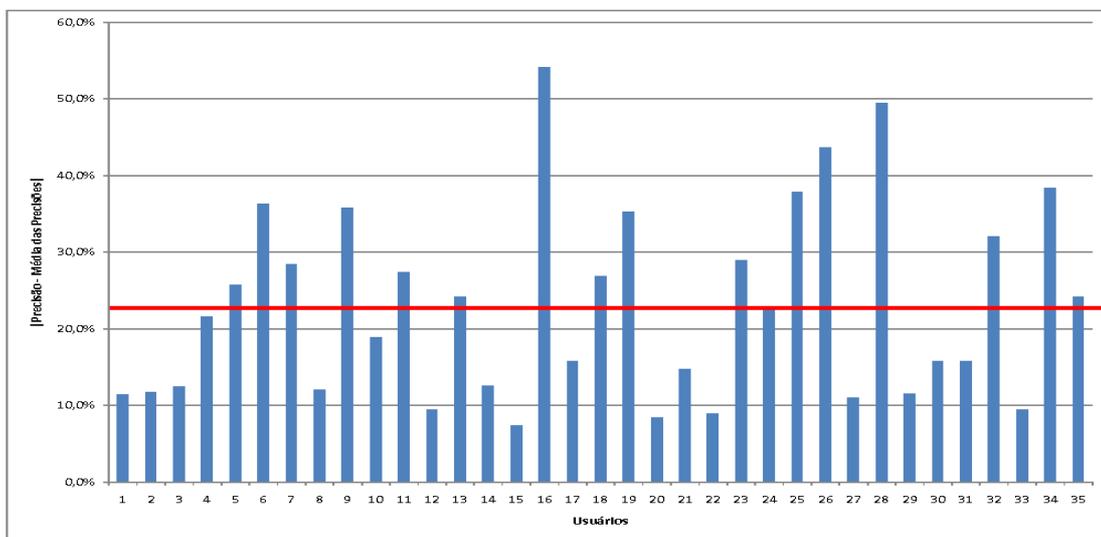


Figura 4.35: Valores de $|\text{Precisão} - \text{Média das Precisões}|$ para cada usuário.

Os usuários identificados como incomuns, são aqueles que possuem as maiores valores para $|\text{Precisão} - \text{Média das Precisões}|$, pode-se destacar alguns destes, como os usuários 16, 26 e 28 que possuem valores de 54,2%, 43,7% e 49,5% respectivamente. Os usuários que possuem menores valores são aqueles classificados como confiáveis, por exemplo os usuários 15, 20 e 22 que possuem valores de 7,4%, 8,4% e 8,9% respectivamente.

No gráfico da Figura 4.36 foi apresentado o desempenho em termos de precisão novamente dos classificadores, dessa vez retirando todos os usuários classificados como incomuns. Os classificadores que possuíam bons desempenhos, antes da retirada dos usuários incomuns, em sua maioria tiveram melhores desempenhos, por exemplo a melhor combinação que antes possuía 48,0% passou a ter 48,8%. Os classificadores que antes possuíam desempenhos ruins tiveram piores desempenhos, como por exemplo HSV com 64 *bins* que possuía 5,1% passou a

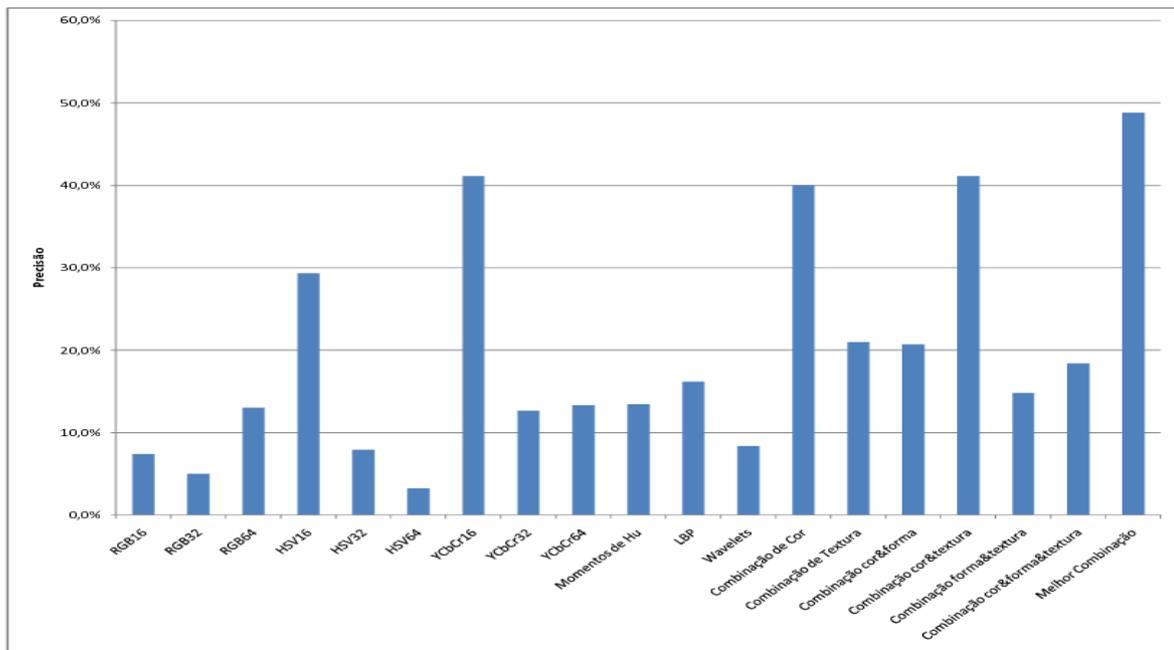


Figura 4.36: Gráfico com as precisões médias para os classificadores sem os usuários incomuns.

ter 3,2%. A retirada dos usuários incomuns do gráfico de desempenho dos classificadores foi de grande importância para a conclusão dos testes, destacando os melhores classificadores.

4.3.3 Experimento com Usuários - Única Imagem

O último experimento com os usuários tem como objetivo identificar o melhor classificador dentre os utilizados. Para isso, diferentemente do experimento anterior, foi utilizada apenas uma imagem de consulta para todos os testes. Para a seleção da imagem padrão foi considerado como critério, encontrar uma imagem que possuísse de forma saliente as três características principais dos extratores utilizados, cor, forma e textura. A imagem selecionada é apresentada na Figura 4.37.

De forma similar ao teste anterior, o experimento utiliza a mesma interface e a mesma sequência de testes. Foram utilizados os mesmos 35 usuários do experimento anterior, que testaram a imagem padrão em 19 diferentes classificadores, sem obter informações sobre qual extrator estavam utilizando. O Apêndice C apresenta os resultados da classificação para cada imagem utilizada nos testes.

No gráfico da Figura 4.38 são apresentados os dados referentes às precisões médias para cada um dos classificadores testados.



Figura 4.37: Imagem padrão selecionada para todos os testes com os usuários.

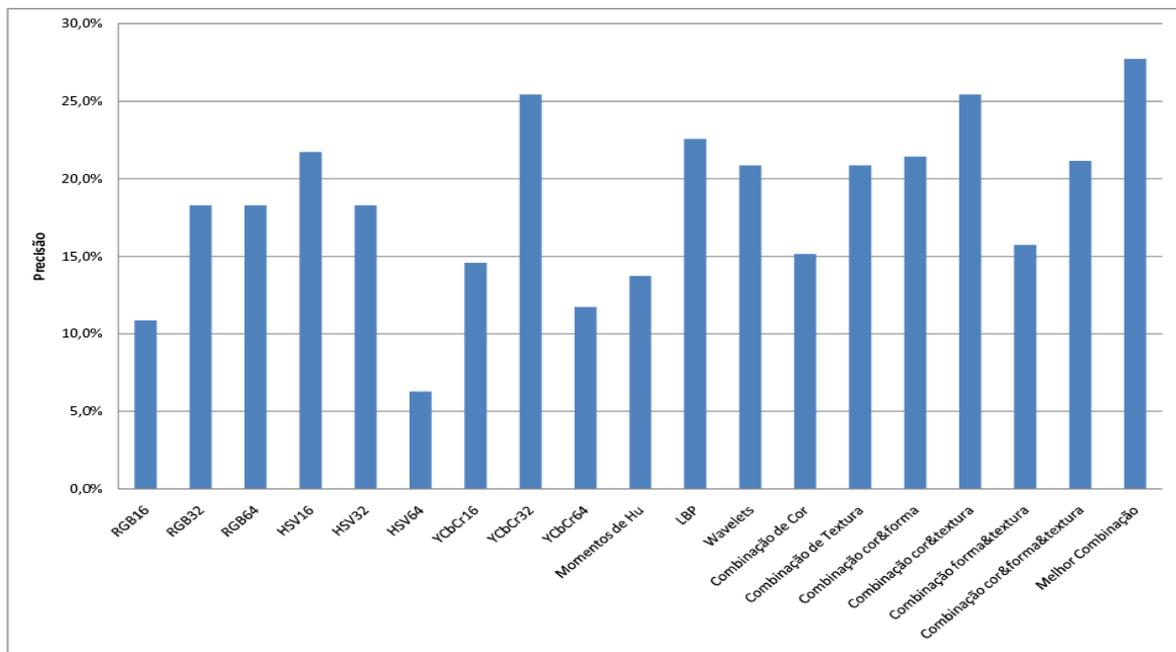


Figura 4.38: Gráfico com as precisões para cada classificador utilizado.

Os três classificadores que mais se destacaram nesse experimento foram três combinações: Melhor Combinação (RGB32, HSV16, LBP e Wavelets), com 27,7% de precisão média entre todos os usuários, Combinação de Cor e Textura com 25,4%, e um classificador simples, YCbCr com 32 *bins* com 25,4% de precisão.

Como no experimento anterior, foram também analisados os resultados de precisão para cada usuário, em busca de possíveis anomalias que pudessem comprometer a integridade dos resultados dos melhores classificadores. No gráfico da Figura 4.39 são apresentados os usuários com suas precisões médias respectivas.

Qualquer usuário apresentando uma diferença (em módulo) entre a sua precisão média e a média geral das precisões superior a 18,5%, indicado pela linha vermelha no gráfico da Figura 4.39, foi classificado como incomum. Entre os usuários incomuns identificados

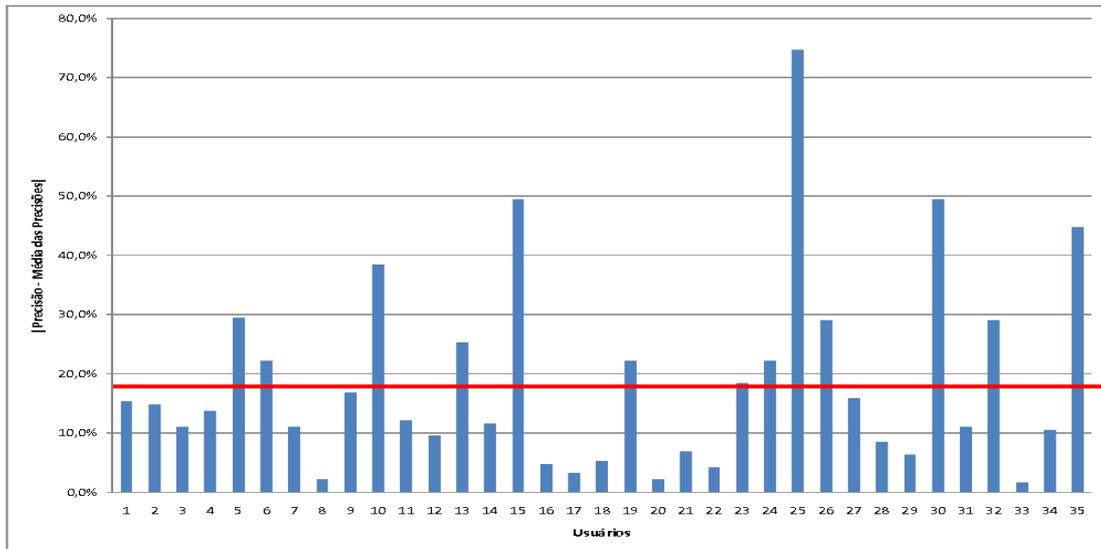


Figura 4.39: Valores de $|\text{Precisão} - \text{Média das Precisões}|$ para cada usuário.

podem-se destacar os usuários 15, 25 e o 30 que possuem os valores das $|\text{Precisão} - \text{Média das Precisões}|$ iguais a 49,5%, 74,7% e 49,5%, respectivamente. Dentre os usuários classificados como confiáveis, podem-se destacar 8, 20 e 33 com os valores de suas $|\text{Precisão} - \text{Média das Precisões}|$ iguais a 2,1%, 2,1% e 1,6%, respectivamente.

De uma forma diferente do experimento anterior, a quantidade de usuários incomuns para o experimento 3 foram aparentemente menores, provavelmente pela facilidade do teste, já que apenas uma imagem era utilizada para todos os testes, diminuindo assim o esforço do usuário.

No gráfico da Figura 4.40 é apresentado o desempenho novamente dos classificadores, dessa vez retirando todos os usuários incomuns como no experimento 2.

Os classificadores, após a retirada dos usuários incomuns, em sua grande maioria obtiveram uma redução em suas precisões, apesar disso, os que possuíam bons desempenhos continuaram sendo ainda os de melhores classificadores. Como no experimento anterior a retirada dos usuários incomuns ajudou na identificação dos melhores resultados de classificadores no gráfico em questão.

Diferente do experimento anterior com um conjunto de imagens fixas, o experimento atual utilizou apenas uma imagem como padrão para todos os teste, tal imagem selecionada por conter características mais salientes para cor e textura, como se pode observar nos resultados. Os classificadores que mais se destacaram foram exatamente os que envolviam apenas uma cor como YCbCr com 32 *bins* por canal ou envolvia a combinação de cor e textura, como na

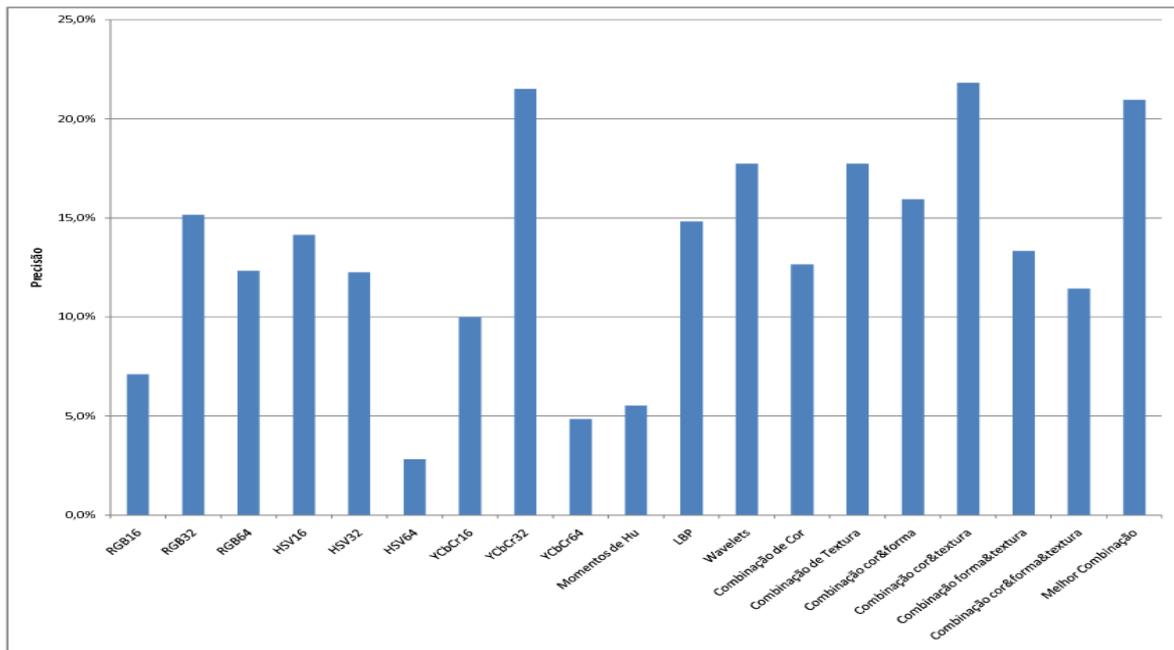


Figura 4.40: Gráfico com as precisões médias para os classificadores sem os usuários incomuns.

Combinação Cor e Textura e no Melhor Combinação, que combina os resultado de RGB32, HSV16, LBP e *Wavelets*.

Os bons resultados da recuperação estão diretamente relacionados com a escolha dos classificadores. Classificadores que utilizam características que são mais salientes na imagem de consulta, provavelmente terão uma taxa de precisão melhor que os demais.

4.4 Considerações Finais

Neste capítulo, foram apresentados três experimentos envolvendo diferentes tipos de extractores de características, como também diferentes combinações de classificadores aplicados ao problema de recuperação de imagens baseado em conteúdo.

No primeiro experimento foram executados testes em busca dos melhores parâmetros para o treinamento de cada uma das redes GHSOM que utilizam diferentes tipos de características extraídas das imagens. Os testes foram realizados de forma automática, buscando parâmetros com os quais as métricas de Precisão, Cobertura e Medida-F atingiram melhores resultados.

No segundo experimento, foram realizados testes com as GHSOM, as quais foram a base para um estudo comparativo entre classificação simples e combinada. Utilizando um banco

de imagens rotuladas, todos os classificadores foram testados, tanto na forma simples como combinada. De uma forma geral, as combinações obtiveram melhores resultados de precisão e cobertura quando o número de imagens recuperadas foi reduzido. O melhor classificador combinado nos experimentos obteve uma precisão média 18% superior ao melhor classificador simples. Em relação à cobertura o melhor classificador combinado obteve uma superioridade de 30% em relação ao melhor simples.

Por último, foram realizados experimentos envolvendo todo o sistema desenvolvido e um conjunto de usuários finais, os quais realizaram testes de recuperação de imagens com uma avaliação subjetiva a respeito da precisão do sistema. O último conjunto de experimentos foi subdividido em três experimentos. O primeiro foi realizado por 25 usuários que escolheram as imagens que seriam utilizadas como consultas, apesar dos resultados apontarem a combinação de classificadores como melhor técnica, o formato do teste não possibilitou a detecção de usuários incomuns. O segundo e o terceiro testes foram realizados por 35 usuários, que utilizaram uma nova interface para testes, de forma que o usuários não possuíam o conhecimento de qual classificador seria utilizado. No segundo teste foram utilizadas 19 imagens fixas para os testes, enquanto que no terceiro apenas uma imagem foi utilizada para todos os testes.

No próximo capítulo, serão apresentados as conclusões da dissertação, as contribuições e os trabalhos futuros.

Capítulo 5

Conclusão

Neste capítulo, apresentam-se um sumário dos principais pontos abordados nesta dissertação, os objetivos iniciais que foram alcançados, as contribuições e as perspectivas para trabalhos futuros.

5.1 Resumo da Dissertação

No Capítulo 1, foi apresentado o problema que serviu de motivação para que o trabalho fosse realizado. A motivação surgiu da observação dos problemas existentes na aplicação de recuperação de imagens da *World Wide Web* a partir de conteúdo. Neste contexto, definimos os principais objetivos desta pesquisa, que foram: (i) Investigar técnicas de extração de características (baseadas em cor, forma e textura); (ii) Investigar a combinação de classificadores neurais do tipo GHSOM (*Growing Hierarchical Self-Organizing maps*); (iii) Aplicar os estudos realizados anteriormente, no desenvolvimento de um Sistema de Recuperação de Imagens Baseado em Conteúdo (SRIBC) e aplicá-lo ao problema de recuperação de imagens na *World Wide Web*, observando as melhores técnicas de classificação com base na precisão.

No Capítulo 2, foi realizada uma revisão bibliográfica sobre sistemas de recuperação de imagens por conteúdo. Foram apresentados diversos sistemas e suas funções, em que se pôde observar uma lacuna ainda existente de trabalhos na área de recuperação de imagens no contexto da *World Wide Web*. Foram também abordados os conceitos fundamentais e trabalhos relacionados para as técnicas que comumente são utilizadas na extração de características de imagens, como também técnicas de indexação e combinação de classificadores. Além disso, foi apresentada a arquitetura típica de um SRIBC, a qual é geralmente constituída por um mó-

dulo de extração de características, um módulo de indexação e um módulo para combinação de resultados e recuperação propriamente dita.

No Capítulo 3, foi apresentada a descrição do projeto e implementação de um Sistema de Recuperação de Imagens Baseada em Conteúdo. Foram também descritos os dois modos de funcionamento do sistema. O modo de busca, que por meio de um robô, procura imagens aptas para extração de características. Todos os dados extraídos, como URL da imagem e suas características são armazenadas em um banco de dados. Os dados armazenados das imagens possibilitam o treinamento das redes GHSOM. O outro modo de funcionamento é o de consulta, que possibilita, a partir de uma interface Web, que o usuário realize a recuperação de imagens por conteúdo.

No Capítulo 4 foram relatados os principais experimentos envolvendo diferentes tipos de extratores de características e tipos de combinação de classificadores aplicados ao problema de recuperação de imagens por conteúdo. O primeiro experimento teve como objetivo encontrar os melhores parâmetros para o treinamento de cada uma das redes GHSOM que utilizam diferentes características extraídas das imagens. No segundo experimento, os melhores parâmetros selecionados no experimento anterior foram utilizados no treinamento das redes GHSOM, as quais serviram de base para um estudo comparativo, utilizando um banco de imagens rotuladas, entre classificações simples e combinadas. Os resultados obtidos no segundo experimento demonstraram uma superioridade da combinação de classificadores em relação aos classificadores simples. Por último, foram realizados experimentos envolvendo todo o sistema e um conjunto de usuários finais, os quais realizaram testes de recuperação de imagens com avaliação subjetiva a respeito da precisão do sistema. De uma forma similar, os resultados foram conclusivos sobre a melhoria na precisão do sistema com a utilização da combinação de classificadores.

Diante dos resultados obtidos ao longo da dissertação, consideramos que o objetivo de investigar técnicas de extração e indexação, bem como, a utilização de um método de combinação de classificadores na melhoria dos resultados da recuperação, foram alcançados com sucesso.

5.2 Contribuições

Como foi apresentado no Capítulo 2, os sistemas de recuperação de imagens baseada em conteúdo atuais [Dow93; HMR96; MM99; NB94; APS96; SC97] geralmente são aplicados a

um domínio específico ou com uma quantidade determinada de imagens, baseando-se nisso o sistema proposto nesta dissertação possui algumas inovações em relação aos sistemas tradicionais:

- Foi desenvolvido um novo método de combinação de classificadores que envolveu as redes GHSOM treinadas com as características extraídas esse método de combinação foi aplicado ao problema de recuperação de imagens e foram avaliados em testes objetivos, utilizando um banco de imagens rotuladas, e subjetivos, a partir de avaliações dos usuários finais.
- Foi desenvolvido um sistema de recuperação de imagens por conteúdo, no qual, diferentemente dos sistemas atuais, foi utilizado um banco de imagens dinâmico, isto quer dizer, o banco que constantemente é alimentado com novas imagens de entradas da Internet por um Robô de Busca. Esta atualização constante permite ao sistema recuperar imagens mais atualizadas e solucionar o problema de possíveis URL inexistentes, já que apenas a URL da imagem e suas características extraídas são armazenadas no banco de dados. O banco criado possui 64000 imagens indexadas por suas características, o qual foi aplicado nos testes subjetivos (com os usuário).

Foram realizadas investigações sobre técnicas que envolvem as principais partes de um SRIBC comum, foram analisados diferentes métodos de extração de características que envolvem as principais características visuais das imagens, cor, forma e textura. Além disso, foram investigadas diferentes técnicas de indexação de imagens, das quais podemos destacar as redes GHSOM, que foram utilizadas nessa dissertação.

5.3 Trabalhos Futuros

Esta seção apresenta algumas sugestões de trabalhos futuros com respeito à obtenção de um melhor desempenho e a novas aplicações do sistema proposto.

Em primeiro lugar, o estudo de novos métodos de extração de características, já que os melhores resultados foram obtidos com a utilização do classificadores baseada em cor e textura. Desta forma, a busca por novos e robustos métodos de extração de características baseado em forma, seria uma das propostas. Além disso, aprimorar as técnicas utilizadas, como a atualização do algoritmo LBP e a utilização de diferentes tipos de *Wavelets*.

Outra proposta de trabalho futuro seria o estudo e a aplicação de novas técnicas de indexação e classificação. Nesta dissertação foi abordada a rede GHSOM como método de indexação, mas a utilização de diferentes métodos seria de grande importância para um estudo comparativo.

Em relação ao aprimoramento do sistema desenvolvido, propõe-se a aplicação de um teste com um maior número de usuários utilizando o sistema de forma simultânea, sendo assim possível avaliar o tempo de resposta para cada usuário. Pode-se citar também como trabalho futuro a utilização processamento distribuído (*grids*) para a extração e classificação de imagens, de forma a possibilitar uma maior escalabilidade do sistema no ambiente da Internet.

Outro aprimoramento do sistema seria o desenvolvimento de uma interface que permitisse que o *feedback* do usuário final influenciasse diretamente os pesos da imagens, desta forma uma imagem que possuíse pesos mais altos seria melhor classificada na recuperação.

Por último, há a proposta de reprogramar partes do robô de busca, de forma a atualizar automaticamente o banco de dados, encontrando novas imagens e excluindo a imagens que deixaram de existir na *World Wide Web*, além disso, existe a necessidade de desenvolver uma aplicação para re-treinar as redes GHSOM de forma automática.

Bibliografia

- [Add02] P. S. Addison. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*. Taylor & Francis, July 2002.
- [AG05] F. F. M. Almeida and H. M. Gomes. Extensões de um mecanismo de busca por conteúdo na internet através de robôs distribuídos e técnicas eficientes de classificação de imagens. In *II Congresso de Iniciação Científica da Universidade Federal de Campina Grande*, pages 45–57, 2005.
- [AMTF07] F. A. Andaló, P. A. V. Miranda, R. S. Torres, and A. X. Falcão. A New Shape Descriptor based on Tensor Scale. In *8th International Symposium on Mathematical Morphology*, pages 141–152, Rio de Janeiro, Brazil, October 2007.
- [APS96] Pentland A., R. W. Picard, and S. Sclaroff. *Photobook: Content-based manipulation of image databases.*, 1996.
- [BOLG05] R. Barcellos, R. S. Oliani, L. T. Lorenzi, and A. Gonzaga. Content based image retrieval using color autocorrelograms in hsv color space. *Brazilian Symposium on Computer Graphics and Image Processing - SIBGRAPI*, pages 1–2, 2005.
- [Car06] F. H. Cardoso. Uma estratégia para detecção e filtragem de conteúdo adulto em páginas da web. Trabalho de Conclusão de Curso - Universidade Tiradentes, June 2006.
- [CT03] C. A. B. Castañón and A. J. M. Traina. Recuperação de imagens por conteúdo através de análise multiresolução por Wavelets. Tese de Doutorado, Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo, 2003.

- [CWK03] Y. Chen, J.Z. Wang, and R. Krovetz. Content-based image retrieval by clustering. *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 193–200, 2003.
- [Dat99] Corel Database. Corel image database. <http://carter.idiap.ch/databases.html>, 1999.
- [DJLW08] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, pages 1–60, 2008.
- [Dow93] J. Dowe. Content-based retrieval in multimedia imaging. *Proceedings of SPIE*, 164(1993):11–16, 1993.
- [Edv06] S. Edvardsen. Classification of images using color, CBIR distance measures and genetic programming: An evolutionary experiment. Dissertação de Mestrado, Norwegian University of Science and Technology, 2006.
- [FMWX03] J. C. French, W. N. Martin, J. V. S. Watson, and J. Xiangyu. Using multiple image representations to improve the quality of content-based image retrieval. In *University of Munich - Tech Report in Dept. of Computer Science*, pages 2–9, 2003.
- [GW02] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (2nd Edition)*. Prentice Hall, January 2002.
- [HE01] S. Haykin and P.M. Engel. *Redes neurais: princípios e prática*. Bookman, 2001.
- [HMR96] T. S. Huang, S. Mehrotra, and K. Ramchandran. Multimedia analysis and retrieval system (mars) project. In *Proc of 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval*, 1996.
- [Hu62] M.K Hu. Visual pattern recognition by moment invariants. *Information Theory, IEEE Transactions on*, 8(2):179–187, 1962.
- [JDM00] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:4–37, 2000.

- [KH98] J. Kittler and A. Hojjatoleslami. A weighted combination of classifiers employing shared and distinct representations. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 924–929, 1998.
- [KHDM98] J. Kittler, M. Hatef, RPW Duin, and J. Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239, 1998.
- [KJ02] A. M. Kulkarni and R. C. Joshi. Content-based Image Retrieval by Spatial Similarity. *Defence Science Journal- New Delhi*, 52(3):285–291, 2002.
- [Koh82] T. Kohonen. Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43(1):59–69, 1982.
- [LKLO00] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja. PicSOM—content-based image retrieval with self-organizing maps. *Pattern Recognition Letters*, 21(13-14):1199–1207, 2000.
- [Mat04] L. N. Matos. Utilização de rede bayesianas como agrupador de classificadores locais e global. Dissertação de Mestrado, Universidade Federal de Campina Grande, Outubro 2004.
- [MM99] W. Y. Ma and B. S. Manjunath. NeTra: A toolbox for navigating large image databases. *Multimedia Systems*, 7(3):184–198, 1999.
- [NA02] M. S. Nixon and A. S. Aguado. *Feature Extraction and Image Processing*. Academic Press, 2002.
- [NB94] W. Niblack and R. Barber. The qbic project: Querying images by content using color, texture and shape. *In Proc. SPIE Storage and Retrieval for Image and Video Databases*, pages 23–32, 1994.
- [OPM02] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [PDM02] E. G. M. Petrakis, A. Diplaros, and E. Milios. Matching and Retrieval of Distorted and Occluded Shapes Using Dynamic Programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:1501–1516, 2002.

- [PGB01] B. G. Prasad, S. K. Gupta, and K. K. Biswas. Color and shape index for region-based image retrieval. *International Workshop on Visual Form*, page 9, 2001.
- [Pra91] W. Pratt. *Digital Image Processing*, volume 2. NY: Wiley-Interscience, 1991.
- [PT08] O. A. B. Penatti and R. S. Torres. Color descriptors for Web image retrieval: a comparative study. *XXI Brazilian Symposium on Computer Graphics and Image Processing*, pages 141–152, October 2008.
- [Ren04] J. D. M. Rennie. Derivation of the F-measure. *In other words*, 1:4, 2004.
- [RHC97] Y. Rui, T. S. Huang, and S. F. Chang. Image retrieval: Past, present, and future. In *International Symposium on Multimedia Information Processing*, pages 2–16, 1997.
- [RMD02] A. Rauber, D. Merkl, and M. Dittenbach. The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. *Neural Networks, IEEE Transactions on*, 13(6):1331–1341, 2002.
- [RN03] S. Russell and P. Norvig. *Inteligência Artificial*. Campus, tradução da 2ª edição, 2003.
- [SC95] J. R. Smith and S. F. Chang. Single color extraction and image query. In *ICIP '95: Proceedings of the 1995 International Conference on Image Processing (Vol. 3)-Volume 3*, pages 528–531, Washington, DC, USA, 1995.
- [SC97] J. R. Smith and S. F. Chang. VisualSEEK: a fully automated content-based image query system. *Proceedings of the fourth ACM international conference on Multimedia*, pages 87–98, 1997.
- [Smi97] J. R. Smith. *Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression*. PhD thesis, Columbia University, 1997.
- [SO95] M. A. Stricker and M. Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 381–392, 1995.
- [Sub98] V. S. Subrahmanian. *Principles of multimedia database systems*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998.

- [SWS⁺00] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pages 2–6, 2000.
- [TAP05] V. Takala, T. Ahonen, and M. Pietikainen. Block-based methods for image retrieval using local binary patterns. *Proc. 14th Scandinavian Conference on Image Analysis (SCIA), Joensuu, Finland*, pages 882–891, 2005.
- [TFa06] R. S. Torres and A. X. Falcão. Content-Based Image Retrieval: Theory and Applications. *Revista de Informática Teórica e Aplicada*, 13(2):161–185, 2006.
- [TJ93] M. Tuceryan and A. K. Jain. Texture analysis, 1993.
- [TMY78] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8:460–473, 1978.
- [WWFXW98] J. Z. Wang, G. Wiederhold, O. Firschein, and S. Xin Wei. Content-based image indexing and searching using Daubechies’ wavelets. *International Journal on Digital Libraries*, 1(4):311–328, 1998.
- [XKS92] L. Xu, A. Krzyzak, and C.Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems Man and Cybernetics*, 22(3):418–435, 1992.
- [ZL04] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37:1–19, 2004.

Apêndice A

Robô de Busca (*Web Crawler*)

Neste apêndice, é apresentada uma descrição detalhada das alterações do WEBINDEX [Car06], que foi utilizado como robô de busca de imagens (*Web Crawler*) para buscar imagens na Internet.

As modificações realizadas no robô de busca o capacitaram a realizar buscar de imagens de tamanho e formato pré-definidos, como também as extrações de características das imagens encontradas. Todas as informações são extraídas das imagens como também a URL de sua localização são armazenadas em um banco de dados. Os principais conceitos e o funcionamento dos módulos adicionais implementados do robô de busca são mais detalhados a seguir.

Os robôs de buscas, também conhecidos como rastejadores (*Crawlers*) ou aranhas (*Spiders*), são softwares que são utilizados regularmente na Internet na tentativa de se obter a maior quantidade de dados sobre documentos *Home Pages*, para integrá-los à sua base de dados.

Atualmente, existem diversas estratégias que são utilizadas pelos robôs para se movimentar de um documento a outro, utilizando-se dos endereços eletrônicos (*link*) encontrados na Internet. Geralmente, eles iniciam a busca em sites conhecidos, especialmente aqueles que referenciam muitos *links*, eles recuperam os dados na *home page* inicial e seguem através dos *link* encontrados. Os robôs utilizam algoritmos próprios para determinar quais *links* devem seguir. Alguns robôs utilizam o algoritmo que recupera os documentos de um grande número de servidores (abordagem *breadth-first* ou busca em largura) dando prioridade a cobertura em largura, enquanto outros capturam todos os documentos em *link* de um mesmo servidor (abordagem *depth-first* ou busca em profundidade) priorizando a busca pelas páginas mais

profundas primeiramente [AG05].

Os motores de busca podem utilizar vários robôs em paralelo para a realização de busca e a armazenagem de documentos na base de dados. No momento da coleta de Páginas *Web* para sua base de dados, uma grande parte dos sistemas de busca permite que os usuários sugiram URLs, assim evitando desperdício de tempo aguardando que os documentos sejam encontrados com a busca realizada regularmente pelos robôs. Esta base de dados contém informações julgadas importantes como URLs ou endereços de páginas HTML e de imagens.

Na macro arquitetura do robô de busca, são apresentados os módulos em que ele é subdividido, a Figura A.1 apresenta os módulos em questão:

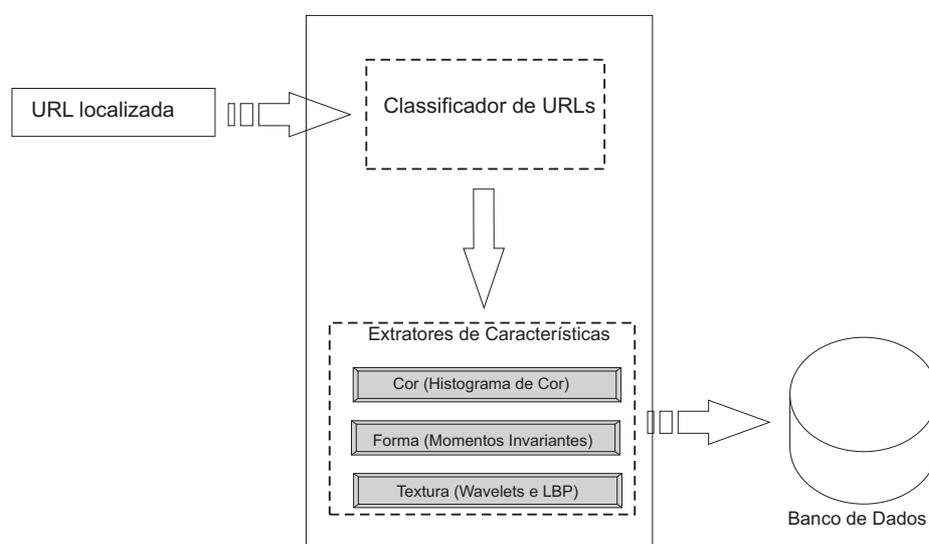


Figura A.1: Arquitetura do Robô de Busca.

A modificação realizada no trabalho de Cardoso [Car06] foi realizada na integração dos módulos de extração de características, permitindo ao robô, além de localizar, extrair e armazenar informações sobre as imagens.

Inicialmente o robô de busca necessita de uma semente (URL inicial) que será seu ponto de partida para a busca de imagens. Ao localizar uma imagem, o módulo Classificador de URL possui um conjunto de regras pré-estabelecidas que avalie a imagem, quanto a altura, largura, área e formato de arquivo. Após a aprovação, a imagem será salva no servidor para que o módulo de extração de características aja. A imagem terá suas características extraídas para cor, forma e textura. O dados obtidos com a extração serão armazenados no banco de dados, juntamente com a URL da imagem.

Apêndice B

Banco de Imagens

Neste apêndice, são apresentadas as imagens utilizadas no banco de imagens para os experimentos do Capítulo 4, especificamente a Seção 4.2, em que foi testada a eficiência da combinação de classificadores quanto a precisão e cobertura médias. O banco possui 800 imagens que são subdivididas em 16 grupos distintos (aviões, vegetação, aves, garrafas, camelos, carros, faces, flores, folhas, futebol, guitarras, casas, motocicletas, outono, planetas e pôr-do-sol).

- A Figura B.1 contém exemplos de imagens do grupo **Aviões**.
- A Figura B.2 contém exemplos de imagens do grupo **Vegetação**.
- A Figura B.3 contém exemplos de imagens do grupo **Aves**.
- A Figura B.4 contém exemplos de imagens do grupo **Garrafas**.
- A Figura B.5 contém exemplos de imagens do grupo **Camelos**.
- A Figura B.6 contém exemplos de imagens do grupo **Carros**.
- A Figura B.7 contém exemplos de imagens do grupo **Faces**.
- A Figura B.8 contém exemplos de imagens do grupo **Flores**.
- A Figura B.9 contém exemplos de imagens do grupo **Folhas**.
- A Figura B.10 contém exemplos de imagens do grupo **Futebol**.
- A Figura B.11 contém exemplos de imagens do grupo **Guitarras**.

- A Figura B.12 contém exemplos de imagens do grupo **Casas**.
- A Figura B.13 contém exemplos de imagens do grupo **Motocicletas**.
- A Figura B.14 contém exemplos de imagens do grupo **Outono**.
- A Figura B.15 contém exemplos de imagens do grupo **Planetas**.
- A Figura B.16 contém exemplos de imagens do grupo **Pôr-do-sol**.

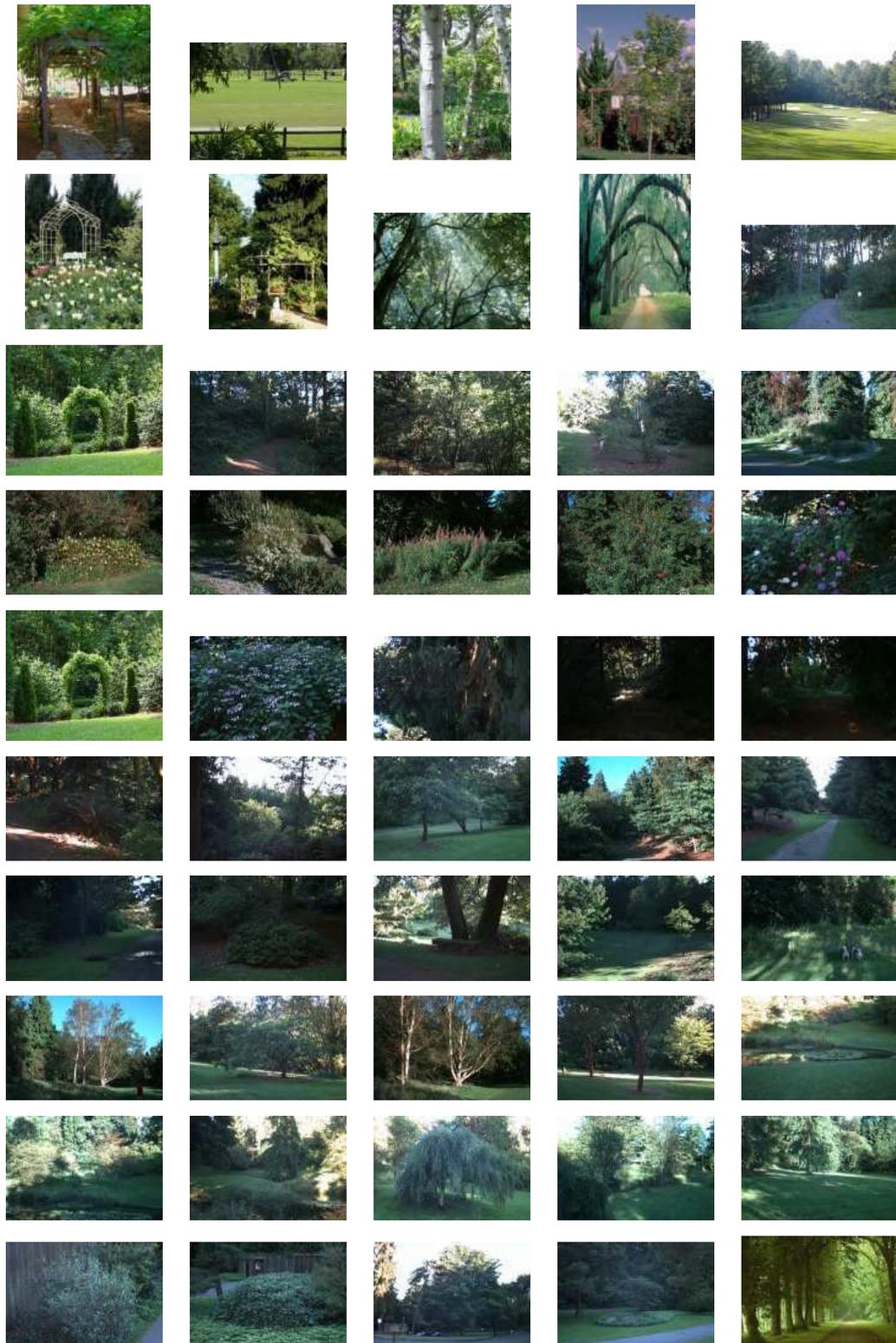


Figura B.2: Grupo de imagens Vegetação.



Figura B.4: Grupo de imagens Garrafas.

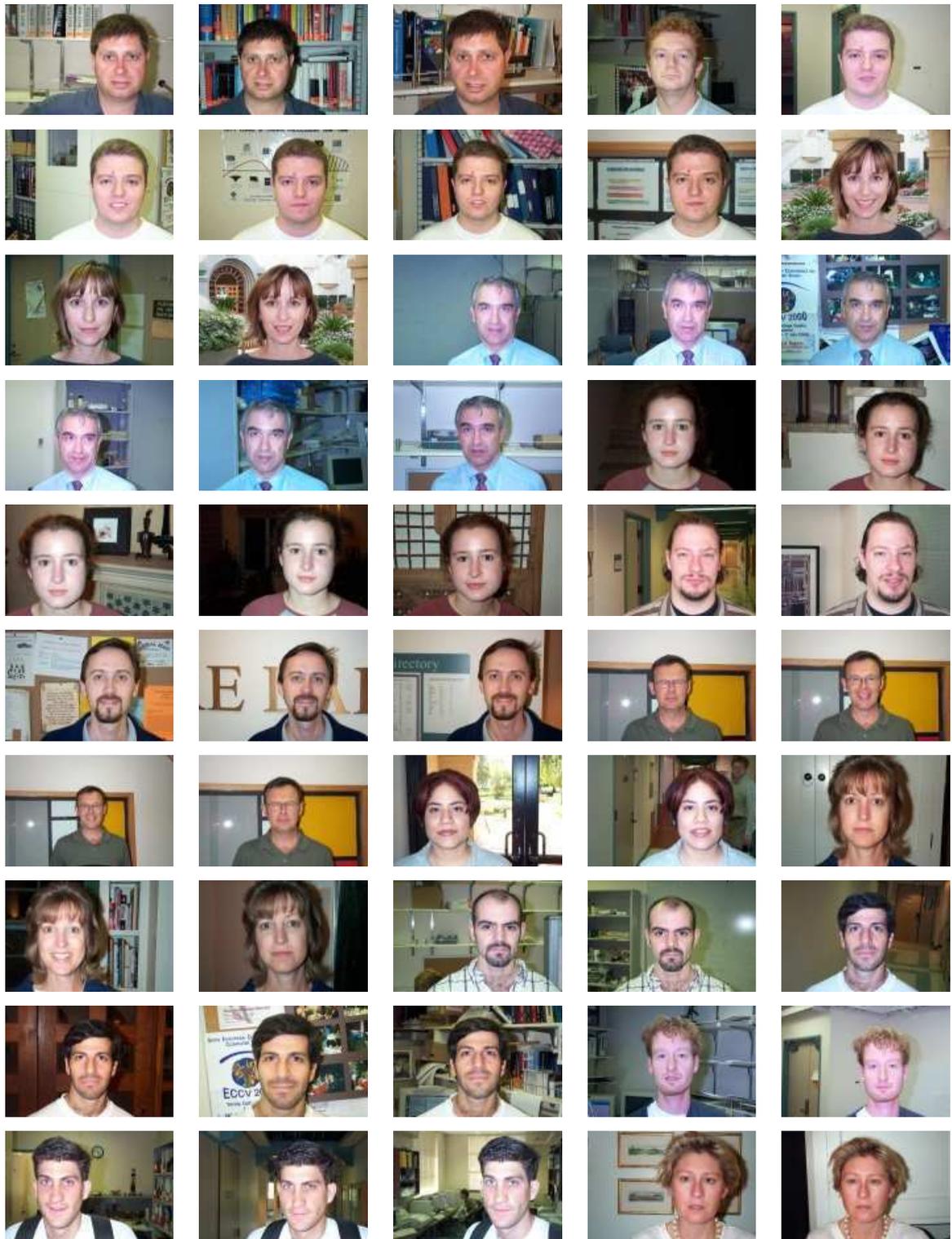


Figura B.7: Grupo de imagens Faces.

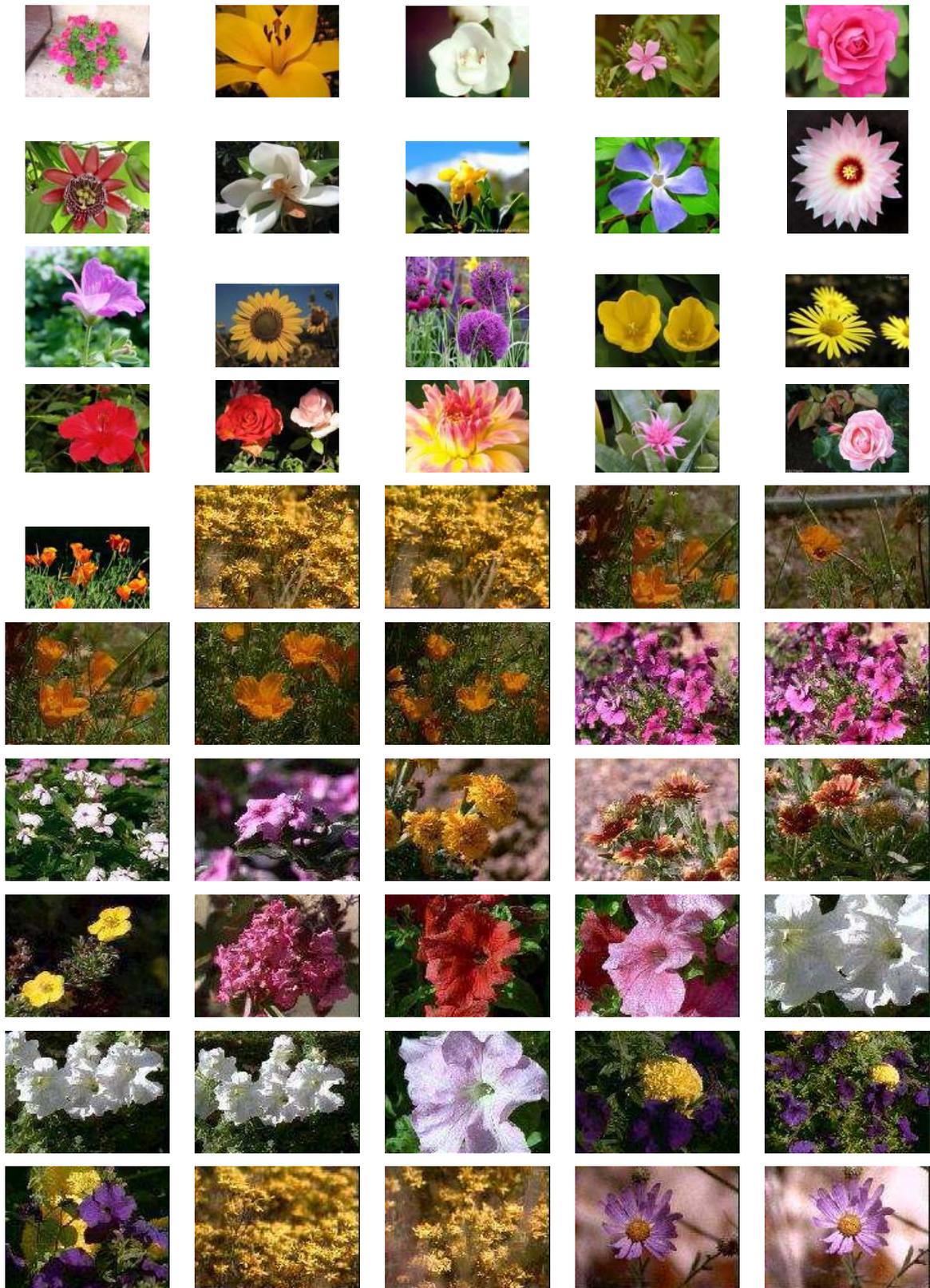


Figura B.8: Grupo de imagens Flores.

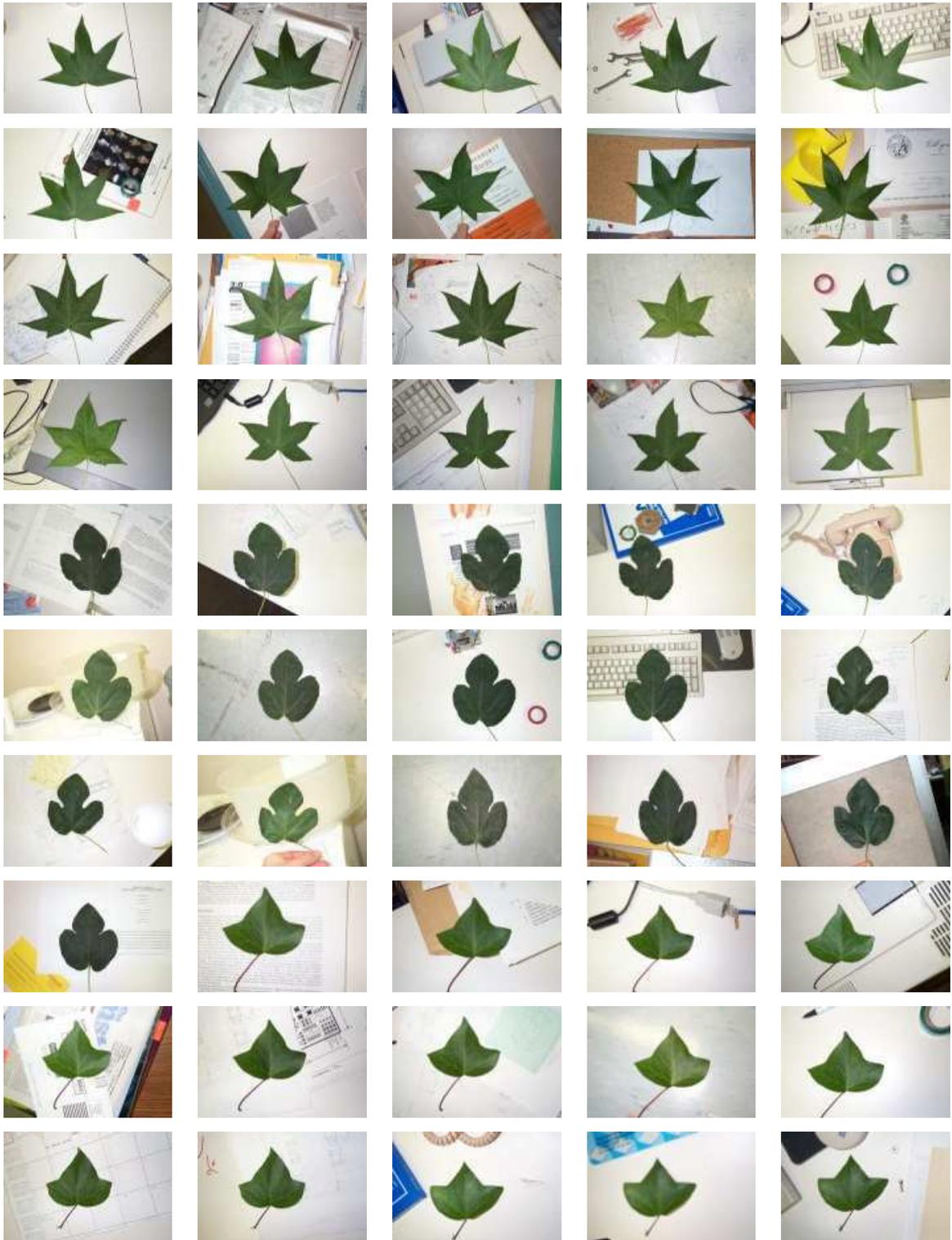


Figura B.9: Grupo de imagens Folhas.

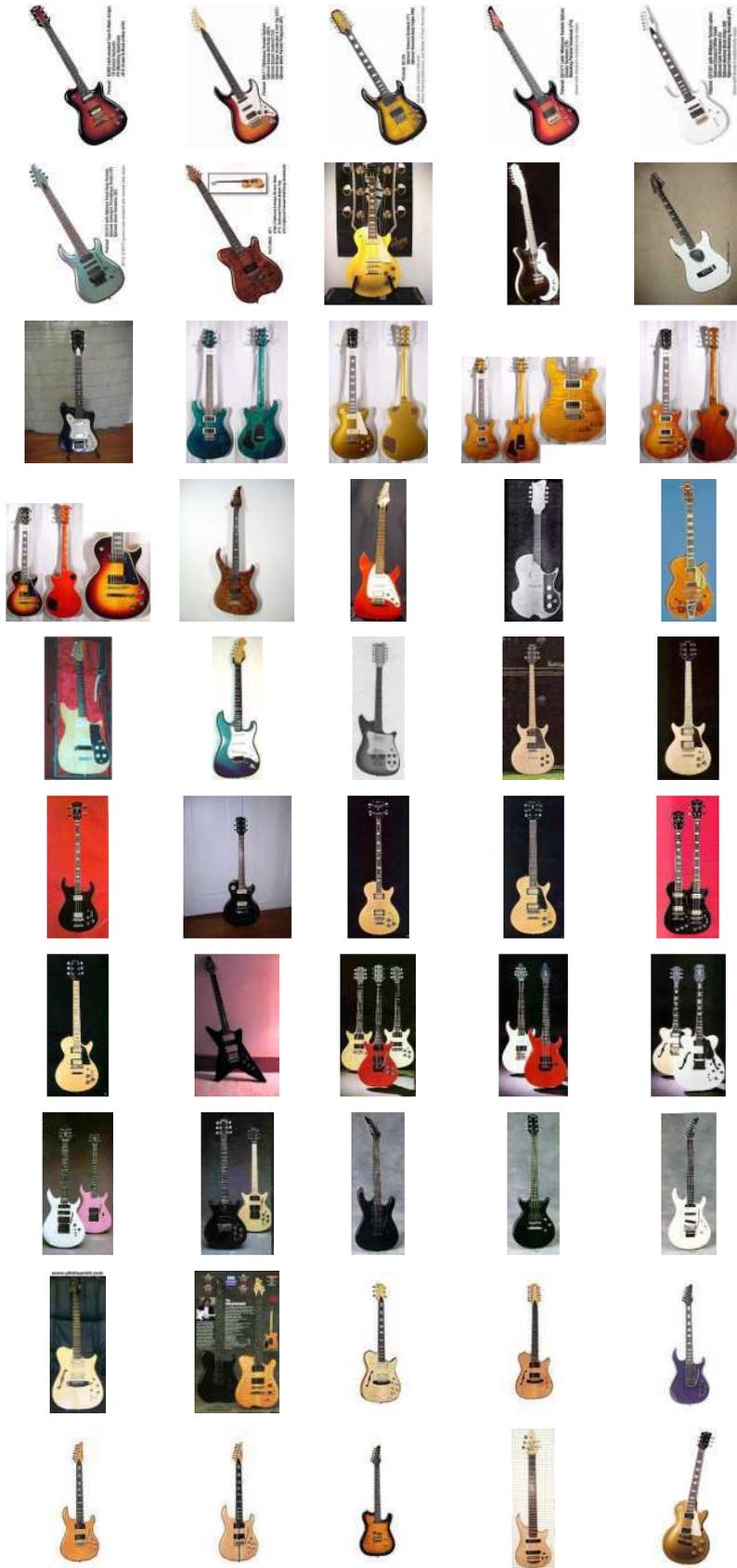


Figura B.11: Grupo de imagens Guitarras.

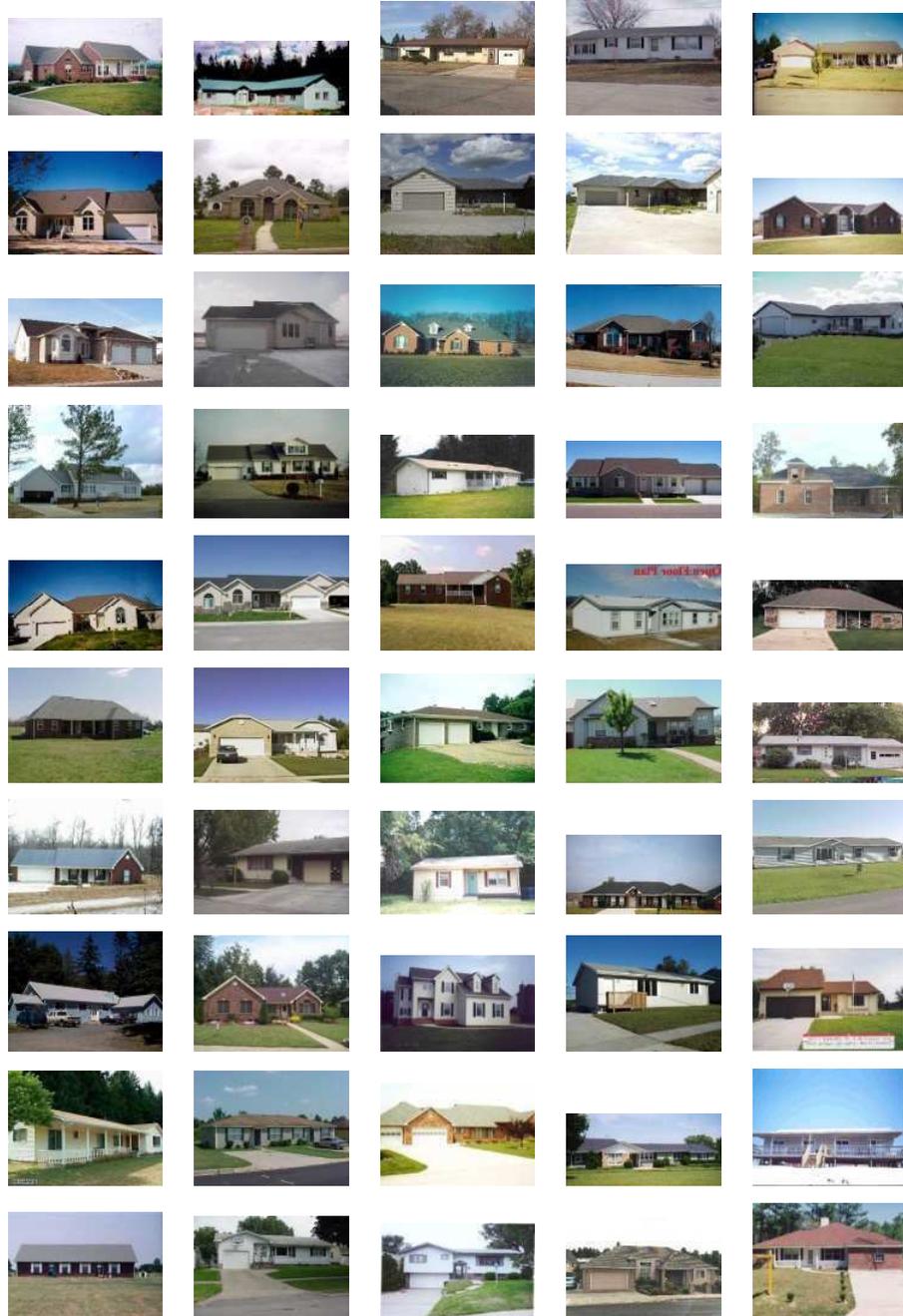


Figura B.12: Grupo de imagens Casas.

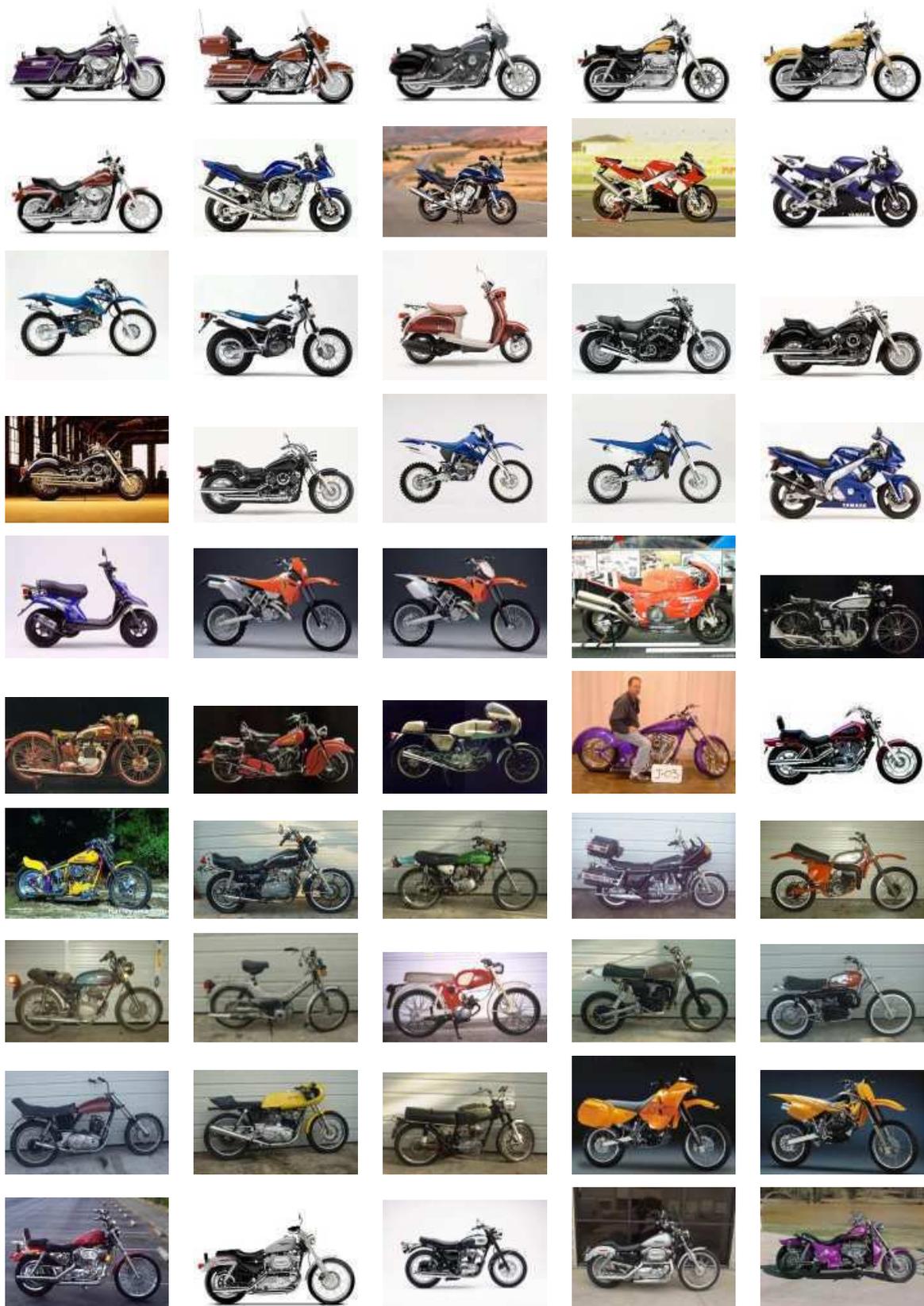


Figura B.13: Grupo de imagens Motocicletas.

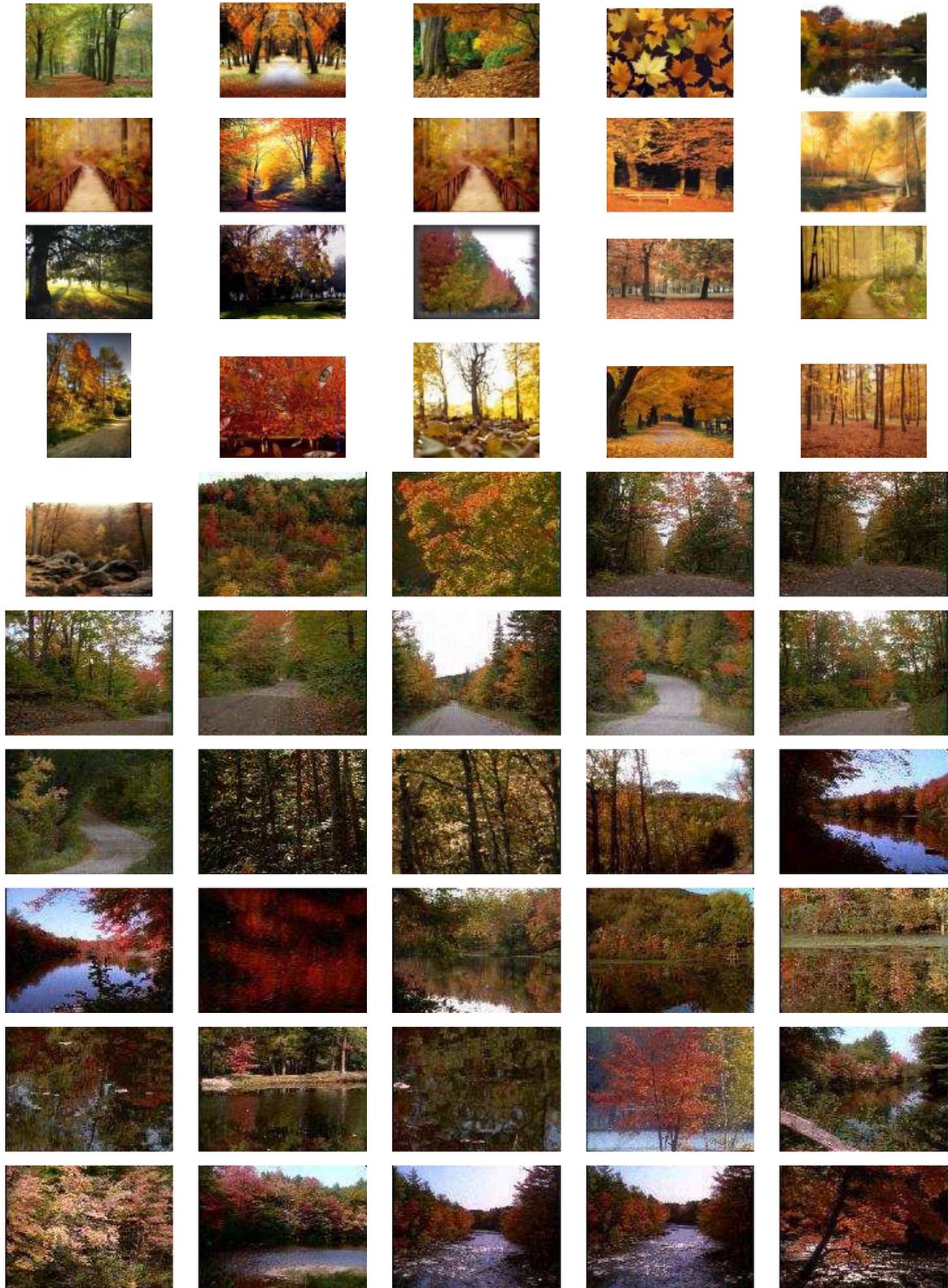


Figura B.14: Grupo de imagens Outono.

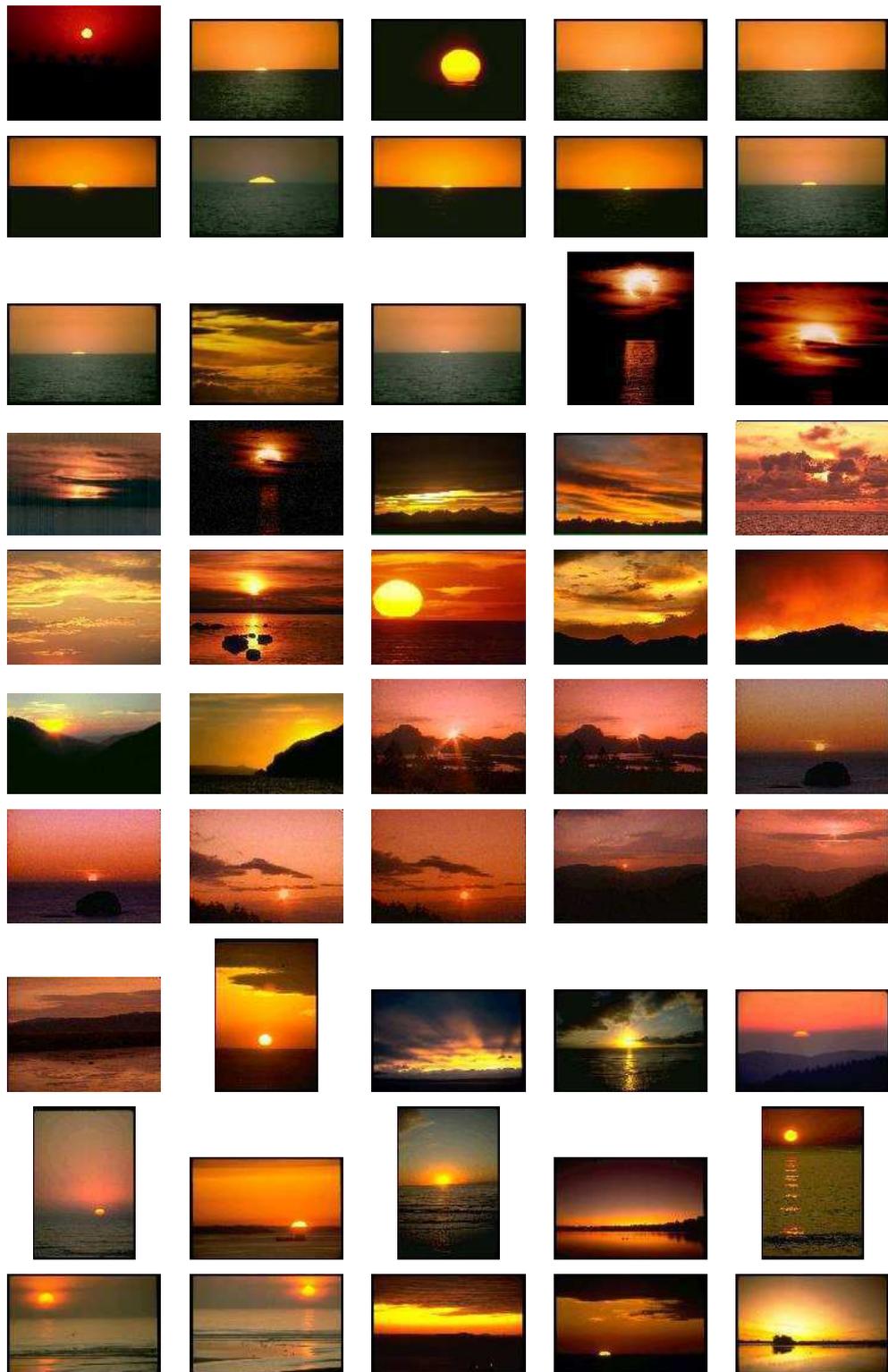


Figura B.16: Grupo de imagens Pôr-do-sol.

Apêndice C

Respostas da Recuperação

Neste apêndice, são apresentados os resultados das recuperações para os experimentos com os usuários (Seções 4.3.2 e 4.3.3). Para todos os testes a cobertura definida foi de 10 imagens, onde o usuário interpretou segundo seus critérios quais imagens eram similares às imagens de consulta. A apresentação desses resultados foi subdividida em duas seções, a primeira referente ao experimento com um conjunto de imagens fixas e a segunda referente ao experimento com uma única imagem de consulta para todos os testes.

Os classificadores são testados na seguinte ordem: RGB com 16 *bins*, RGB com 32 *bins*, RGB com 64 *bins*, HSV com 16 *bins*, HSV com 32 *bins*, HSV com 64 *bins*, YCbCr com 16 *bins*, YCbCr com 32 *bins*, YCbCr com 64 *bins*, LBP, Wavelets, Momentos Invariantes de Hu, Combinação de Cor, Combinação de Textura, Combinação Cor e Forma, Combinação Cor e Textura, Combinação Forma e Textura, Combinação Cor, Forma e Textura e Melhor Combinação.

C.1 Resultado de Classificação - Conjunto de Imagens Fixas

Nesta seção foram apresentados os resultados das classificações dos usuários para as 19 imagens fixas, cada imagem para seu teste específico.



Figura C.1: Resultado da classificação da imagem 1 para RGB com 16 bins.



Figura C.2: Resultado da classificação da imagem 2 para RGB com 32 bins.



Figura C.3: Resultado da classificação da imagem 3 para RGB com 64 bins.

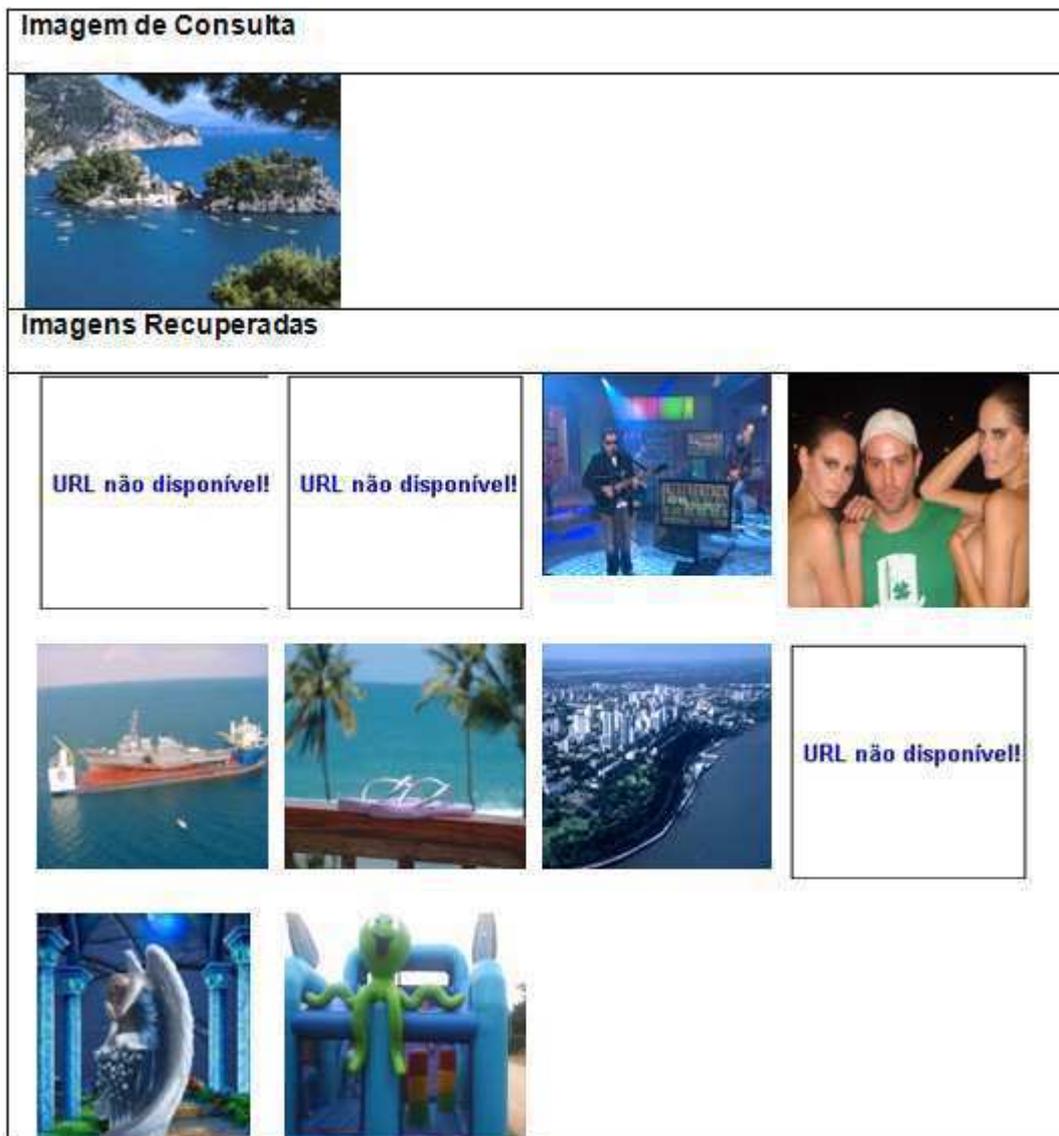


Figura C.4: Resultado da classificação da imagem 4 para HSV com 16 bins.



Figura C.5: Resultado da classificação da imagem 5 para HSV com 32 bins.

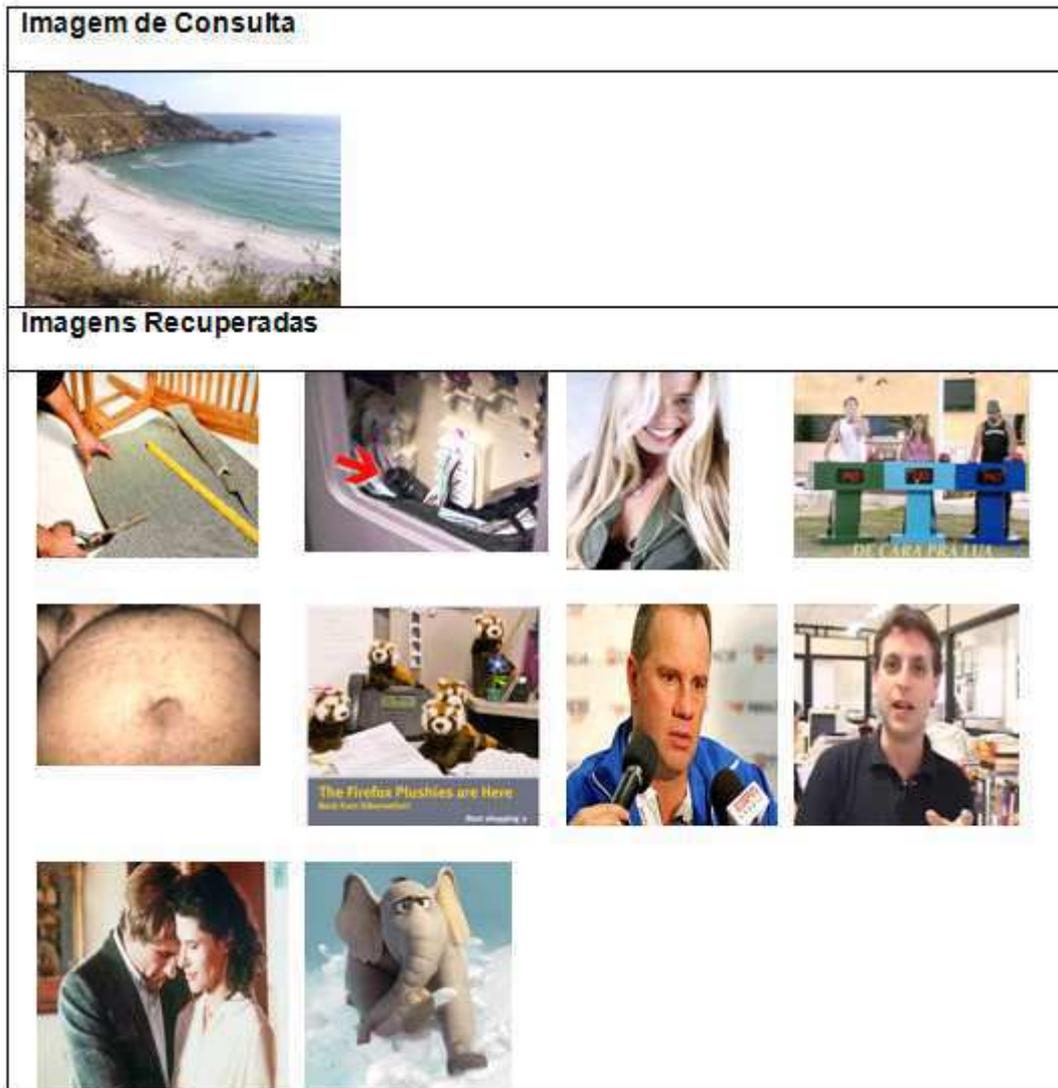


Figura C.6: Resultado da classificação da imagem 6 para HSV com 64 bins.

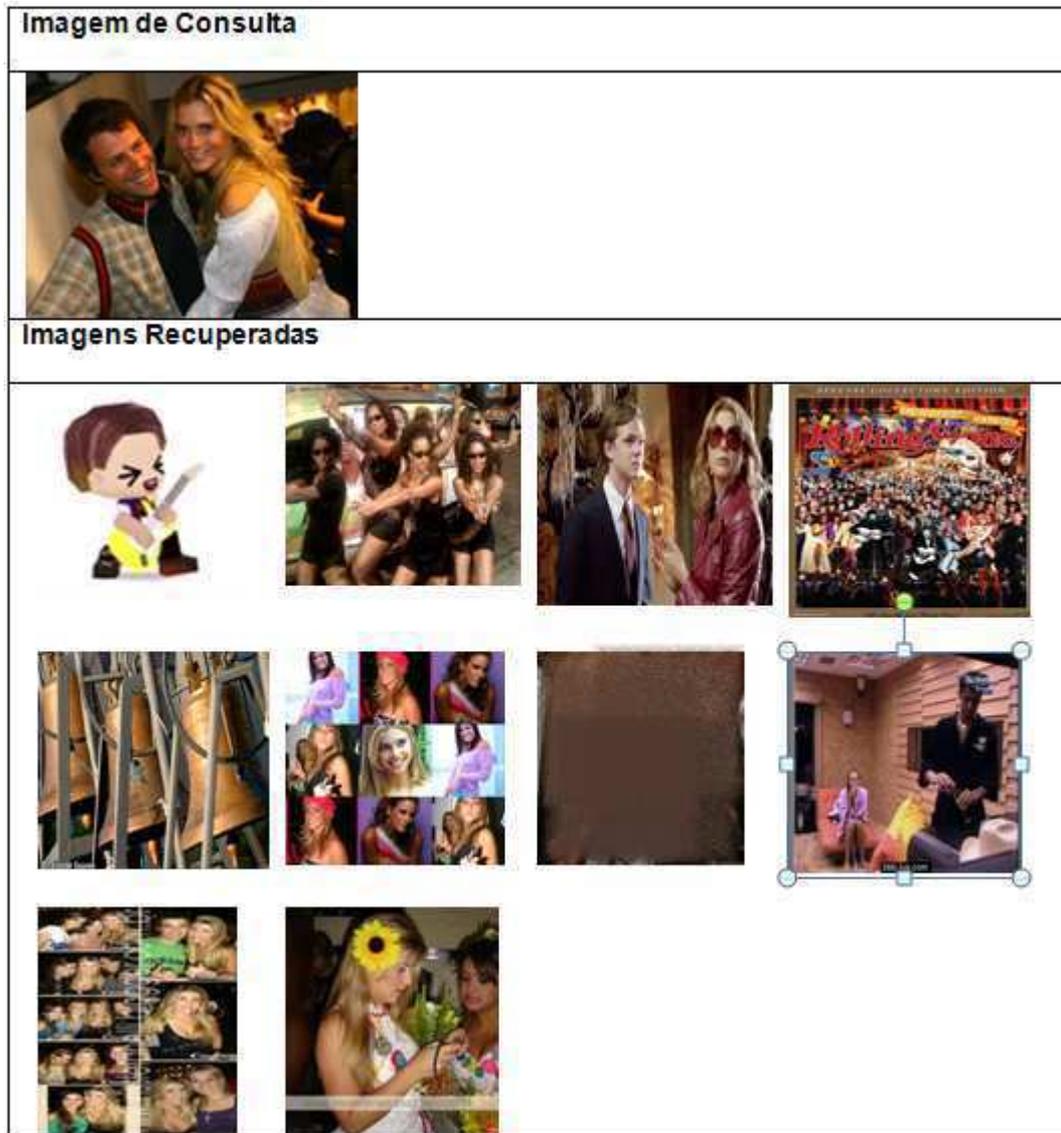


Figura C.7: Resultado da classificação da imagem 7 para YCbCr com 16 bins.



Figura C.8: Resultado da classificação da imagem 8 para YCbCr com 32 bins.

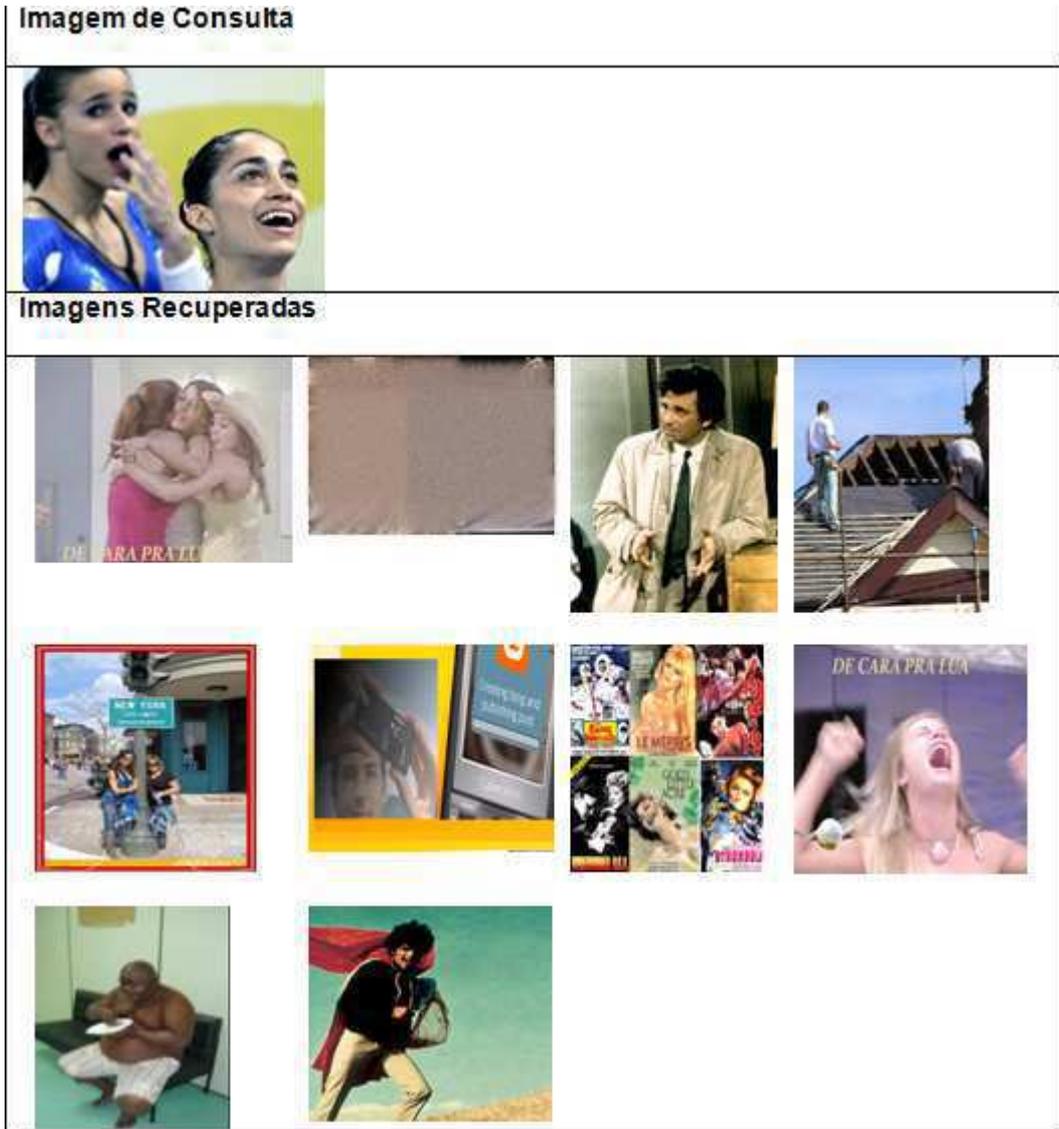


Figura C.9: Resultado da classificação da imagem 9 para YCbCr com 64 bins.

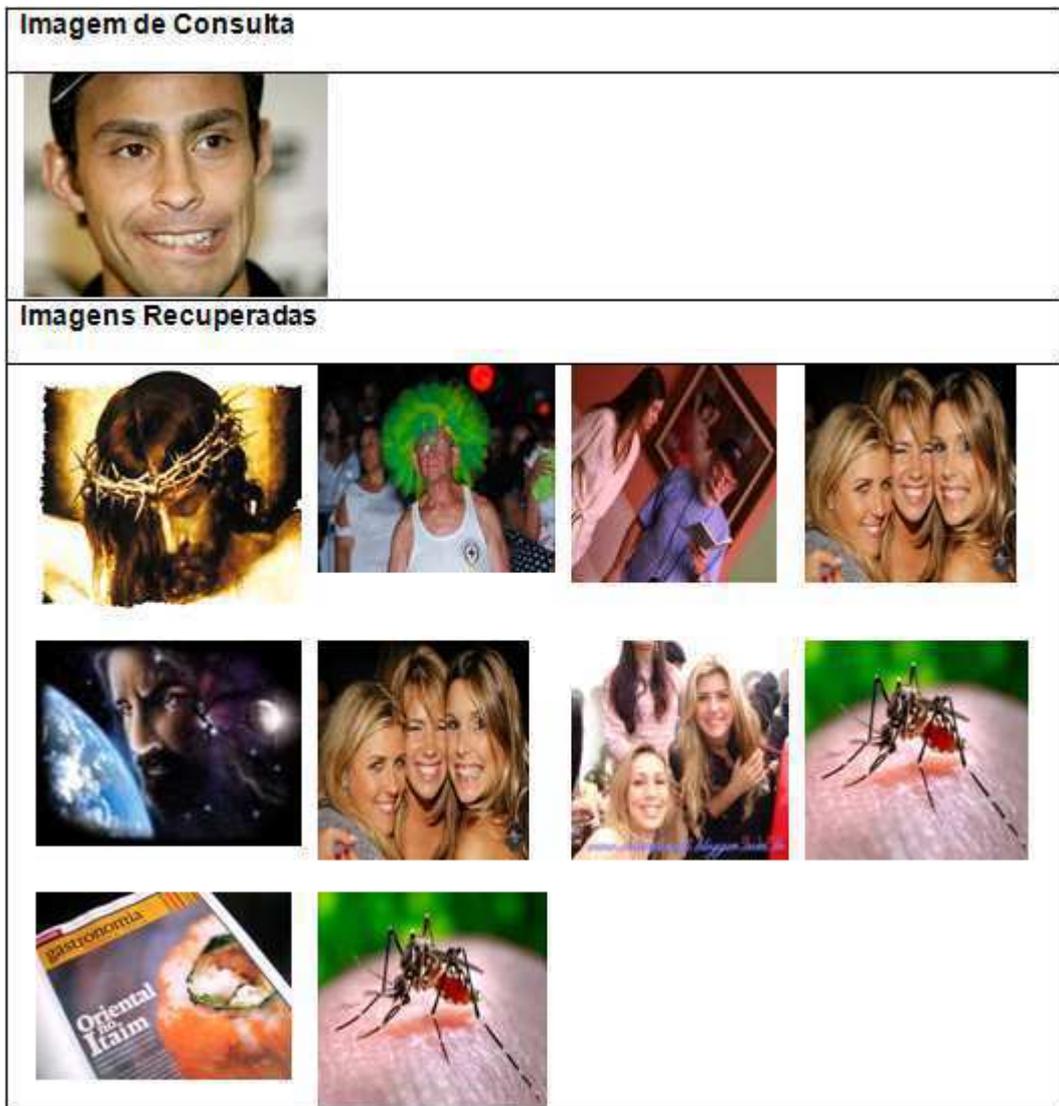


Figura C.10: Resultado da classificação da imagem 10 para LBP.

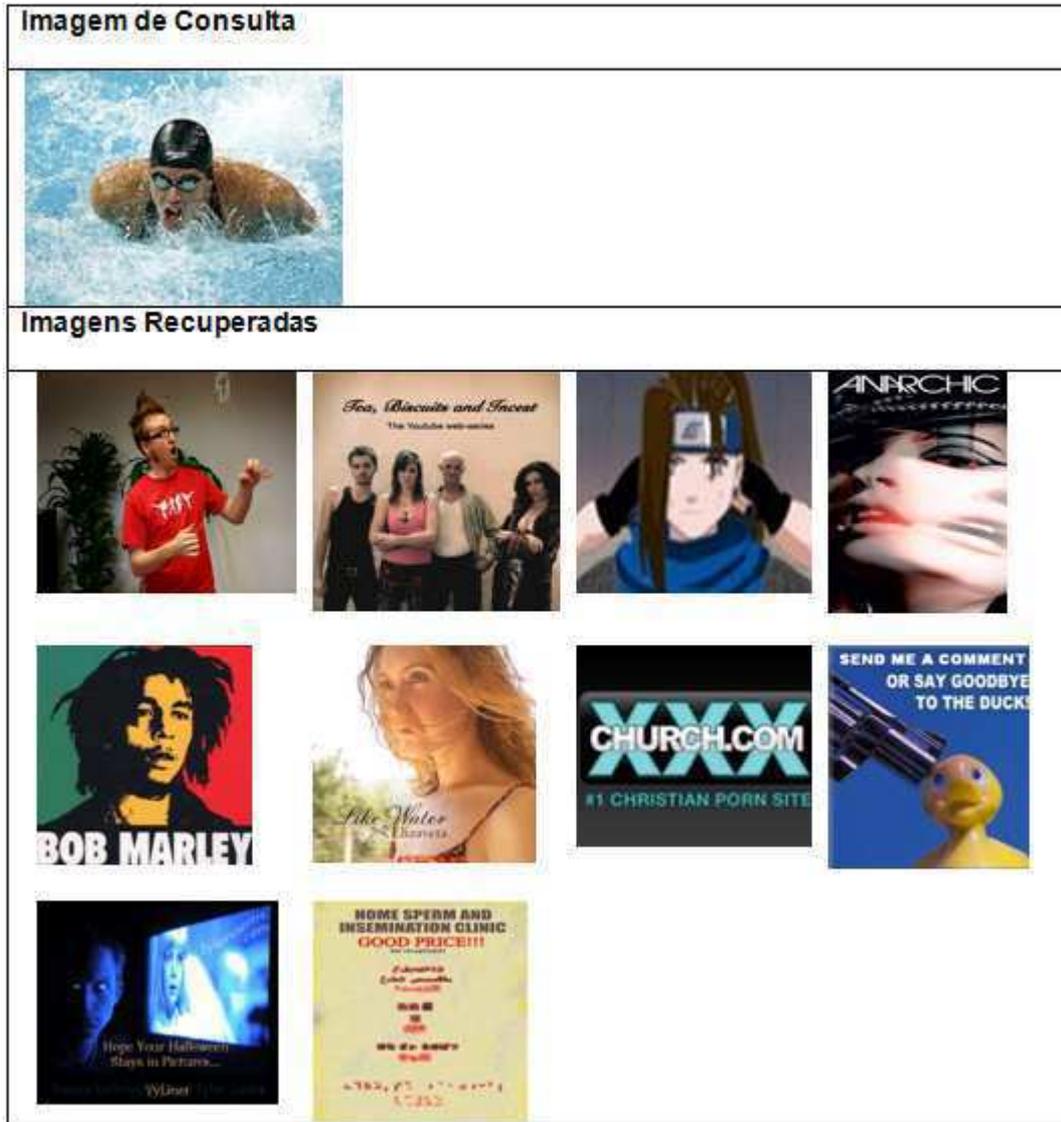


Figura C.11: Resultado da classificação da imagem 11 para *Waveletes*.

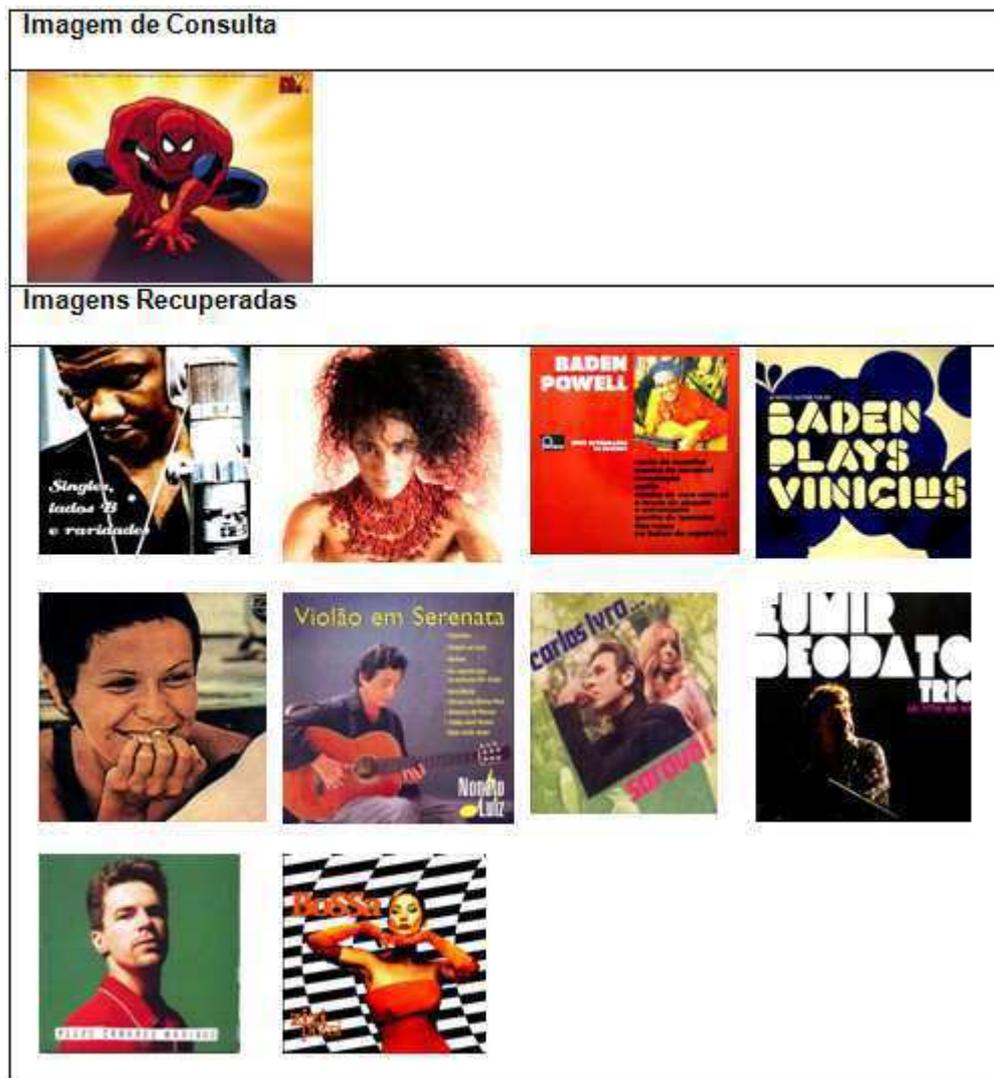


Figura C.12: Resultado da classificação da imagem 12 para Momentos Invariantes de Hu.

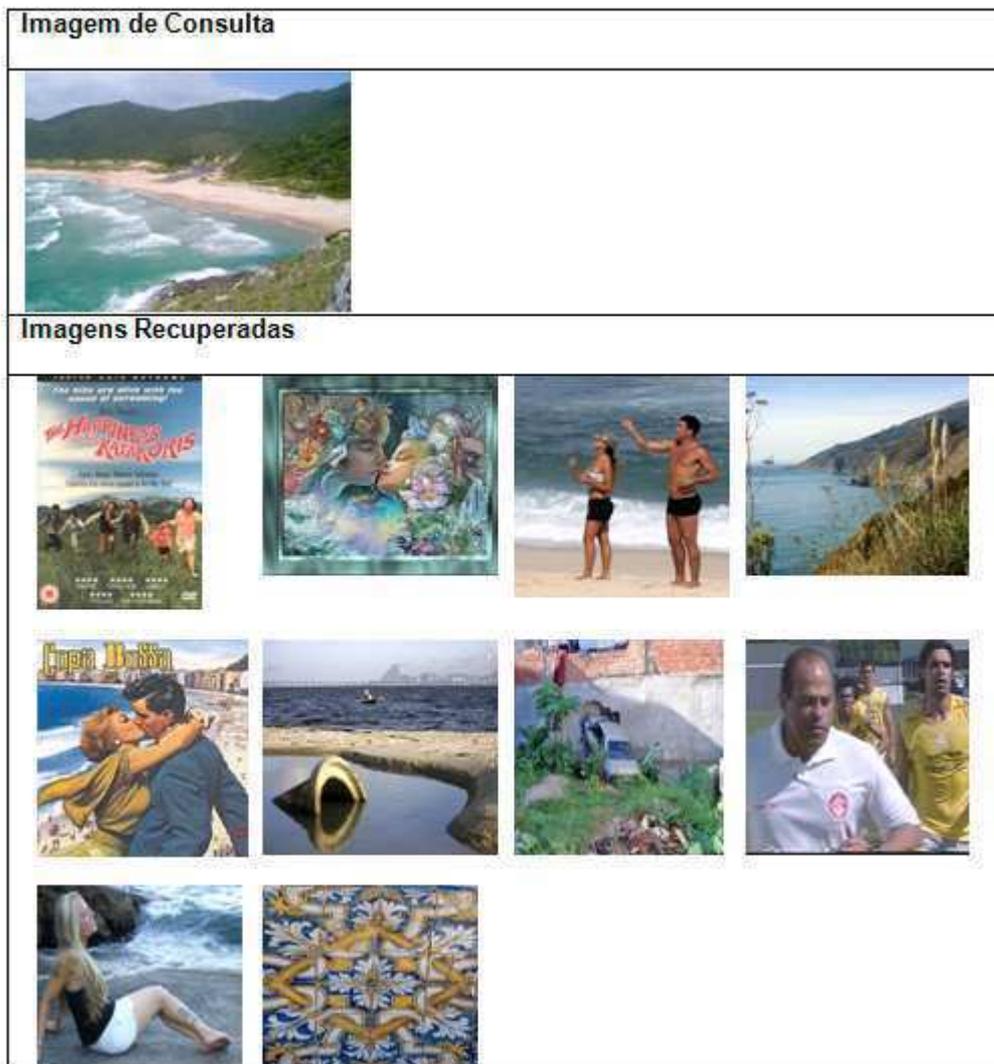


Figura C.13: Resultado da classificação da imagem 13 para combinação de cor.

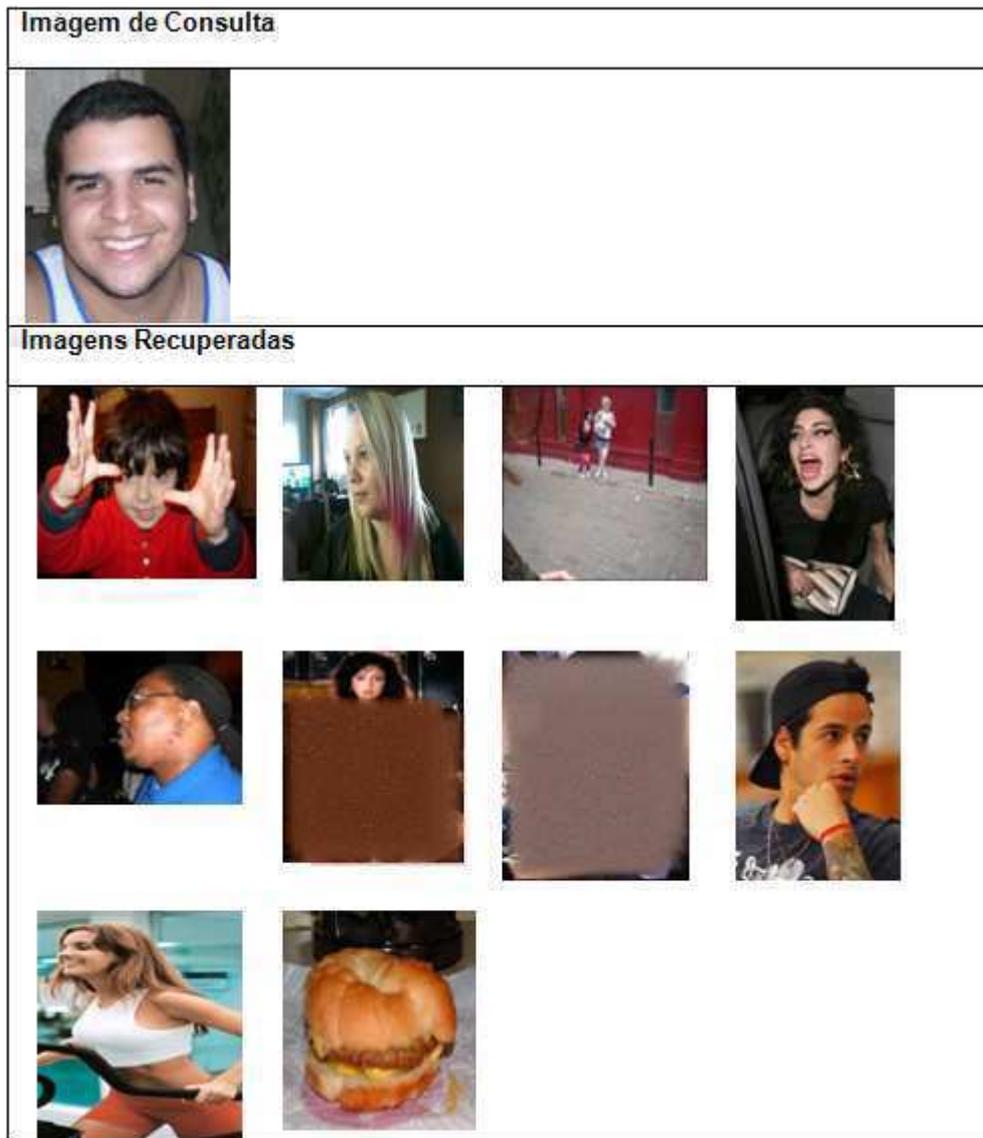


Figura C.14: Resultado da classificação da imagem 14 para combinação de textura.

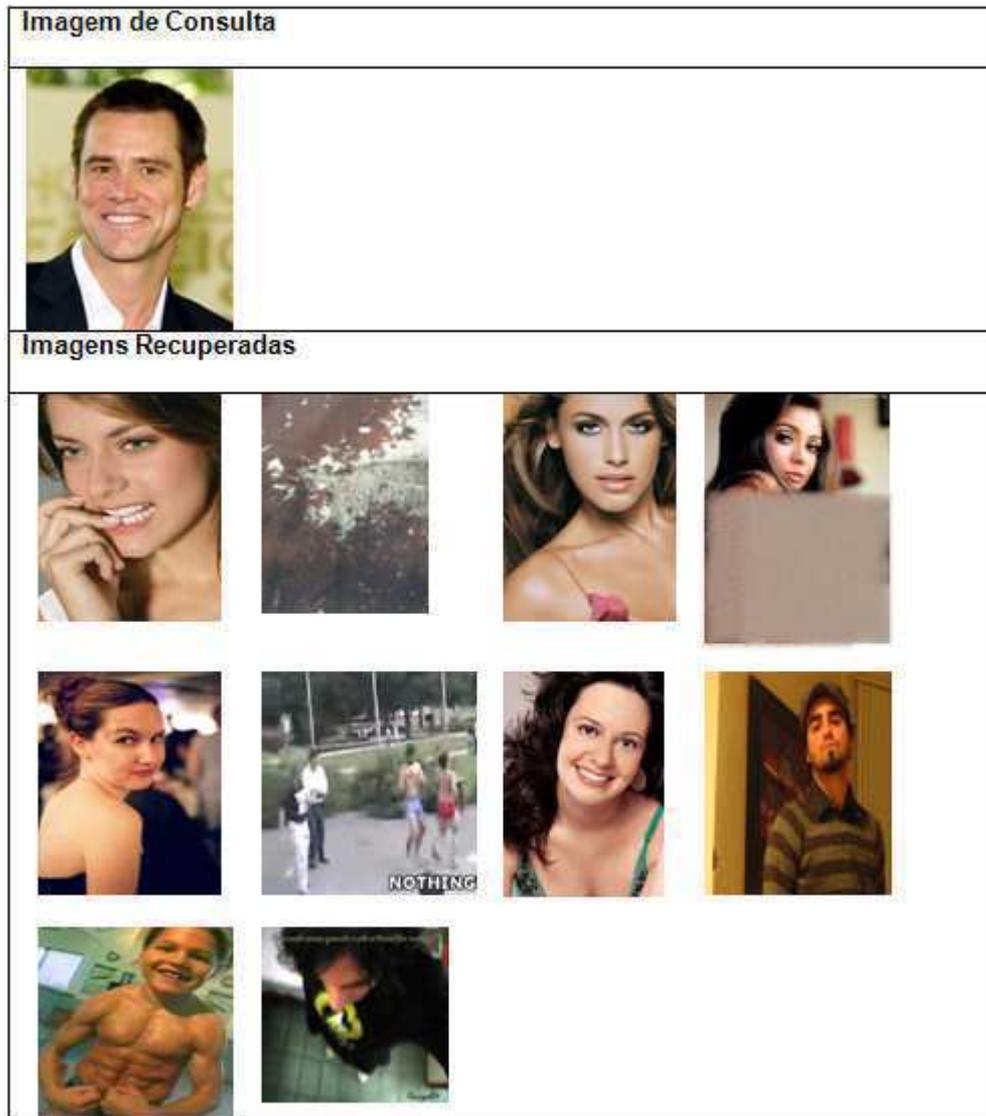


Figura C.15: Resultado da classificação da imagem 15 para combinação de cor e textura.

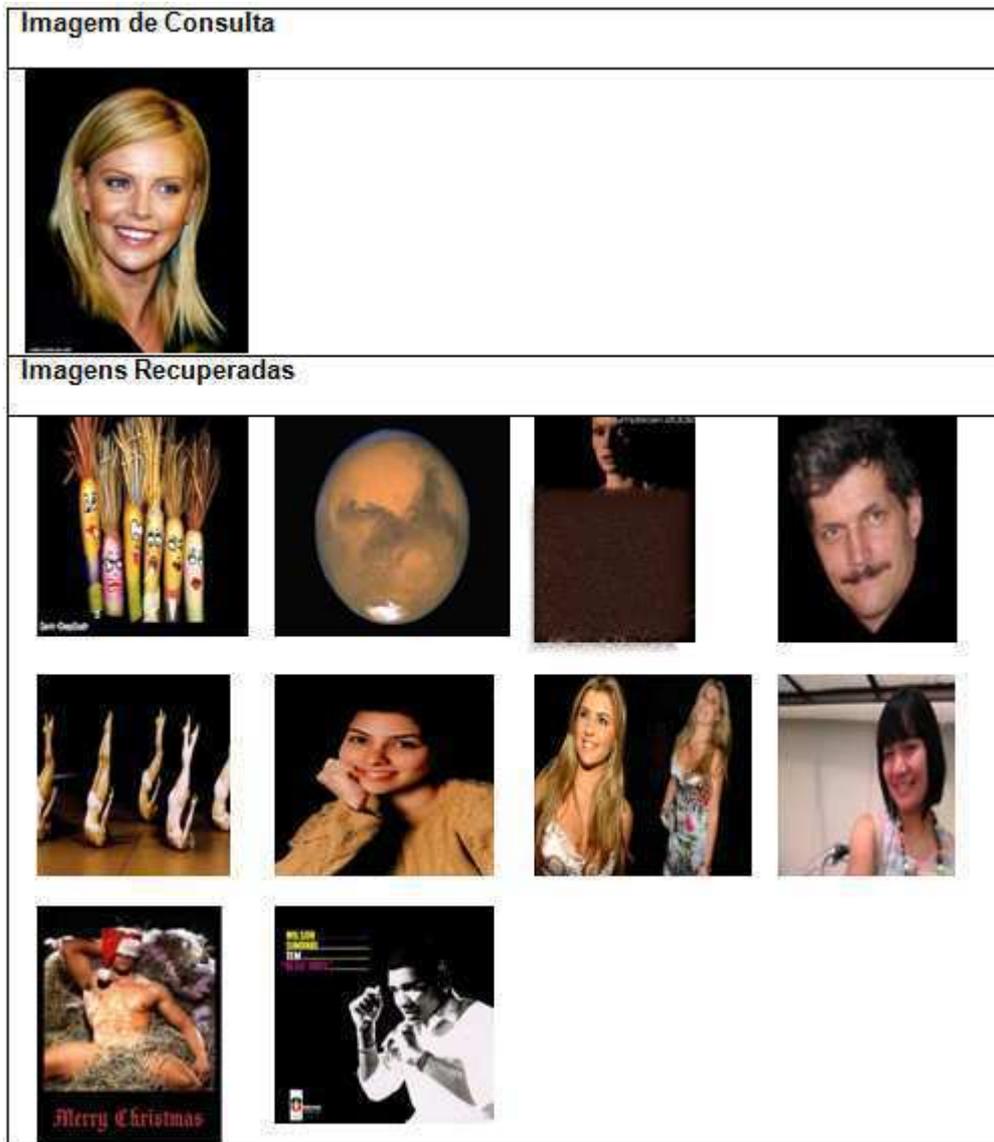


Figura C.16: Resultado da classificação da imagem 16 para combinação de cor e forma.

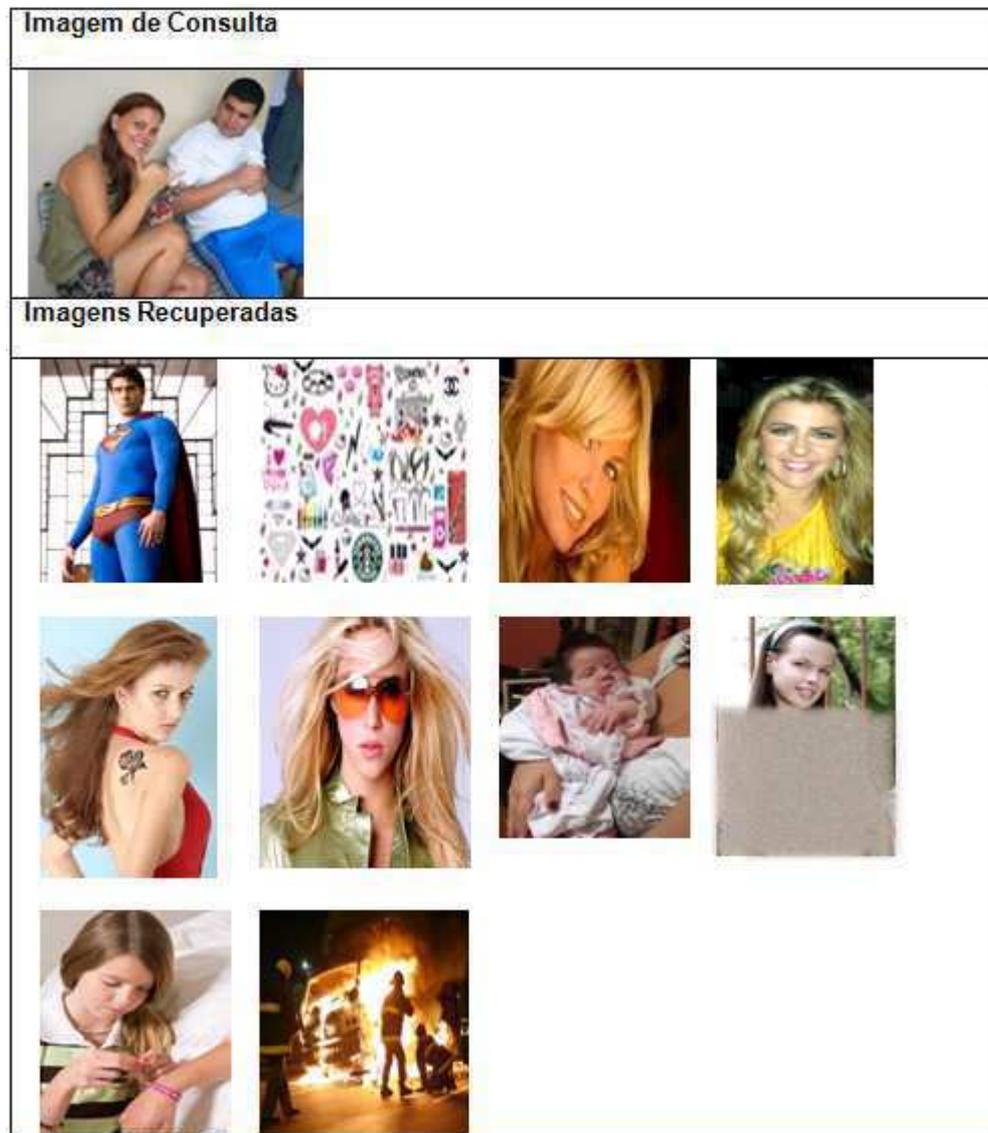


Figura C.17: Resultado da classificação da imagem 17 para combinação de textura e forma.

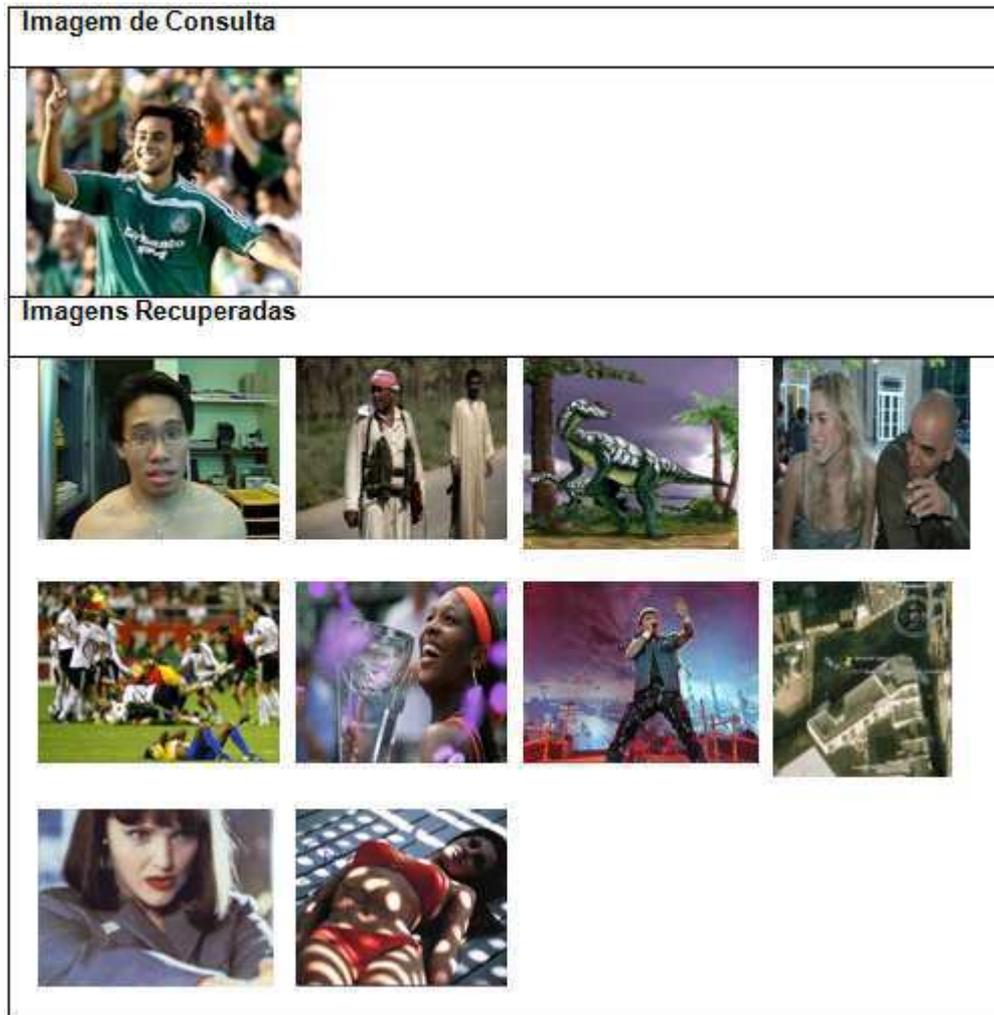


Figura C.18: Resultado da classificação da imagem 18 para combinação cor, textura e forma.

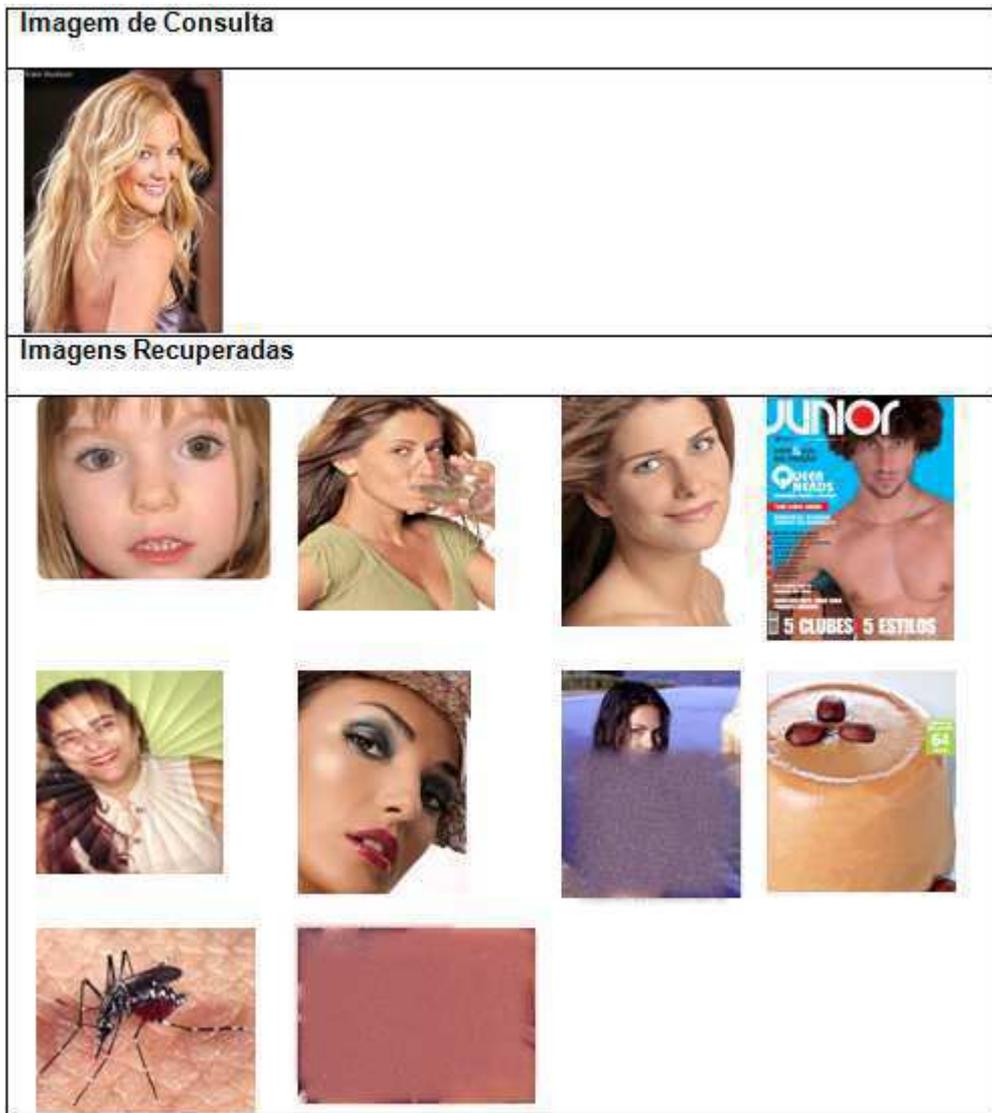


Figura C.19: Resultado da classificação da imagem 19 para a melhor combinação.

C.2 Resultado de Classificação - Única Imagem

Nesta seção foram apresentados os resultados das classificações dos usuários para uma única imagem testada com os classificadores.



Figura C.20: Resultado da classificação da única imagem para RGB com 16 bins.



Figura C.21: Resultado da classificação da única imagem para RGB com 32 bins.

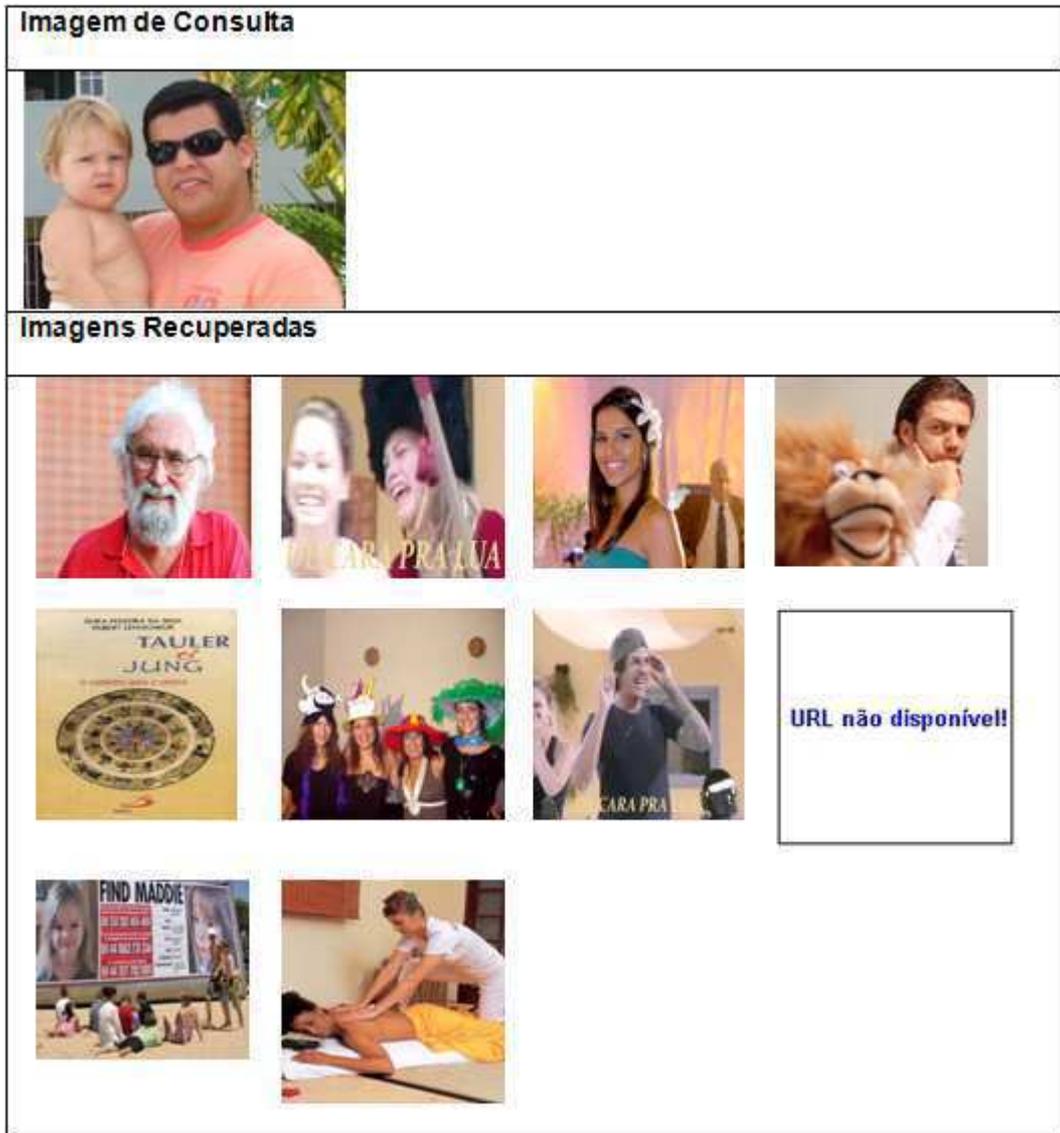


Figura C.22: Resultado da classificação da única imagem para RGB com 64 bins.



Figura C.23: Resultado da classificação da única imagem para HSV com 16 bins.

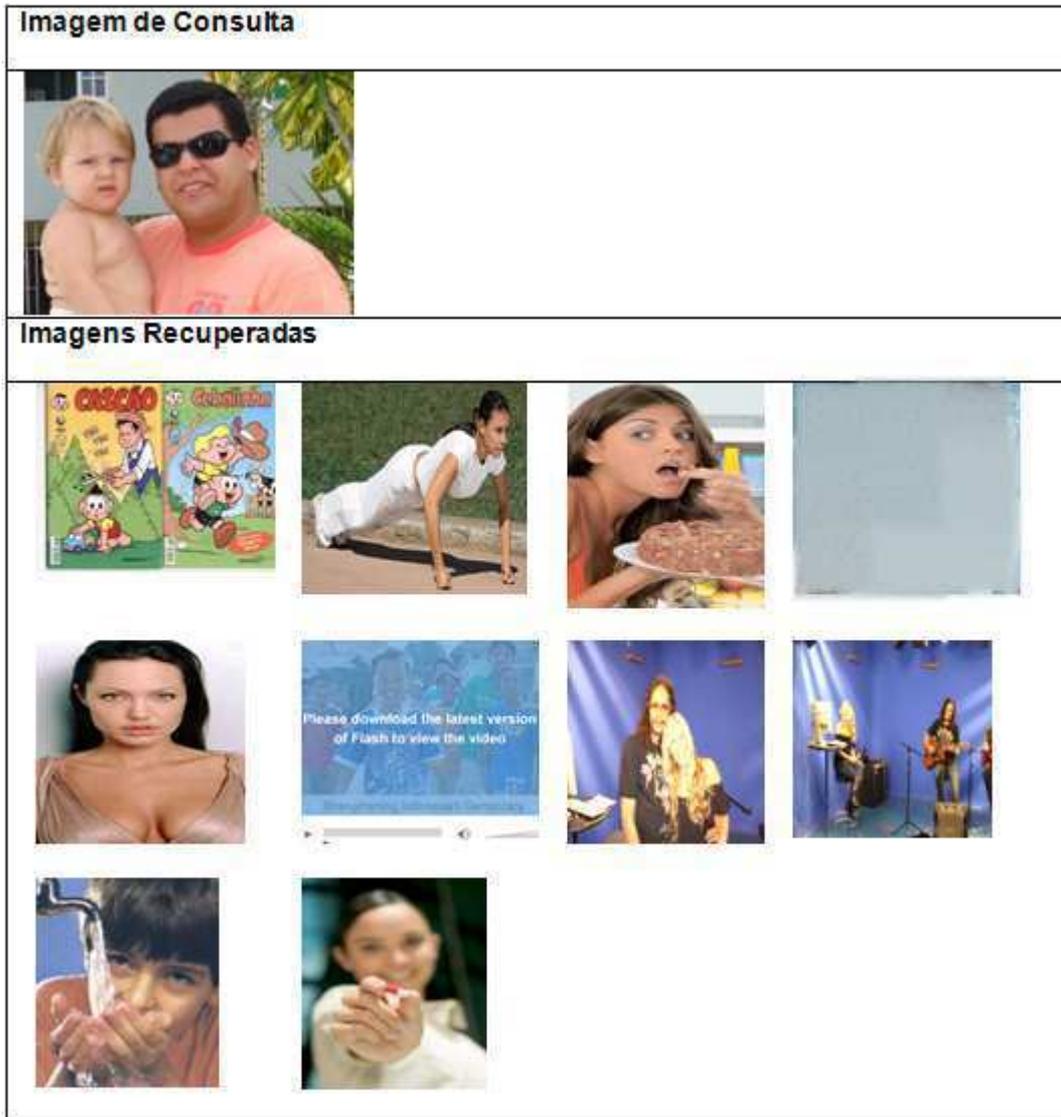


Figura C.24: Resultado da classificação da única imagem para HSV com 32 bins.

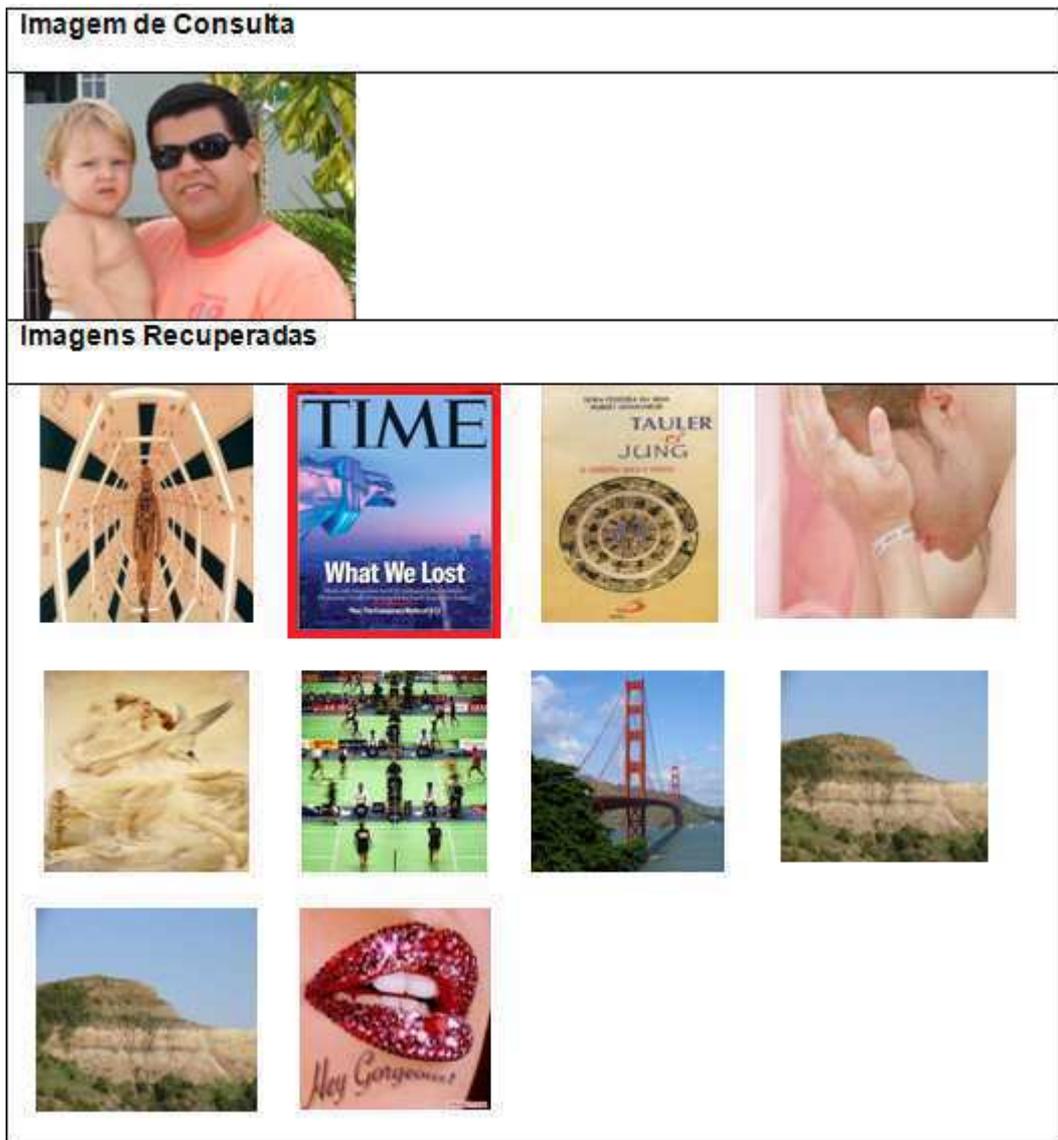


Figura C.25: Resultado da classificação da única imagem para HSV com 64 bins.

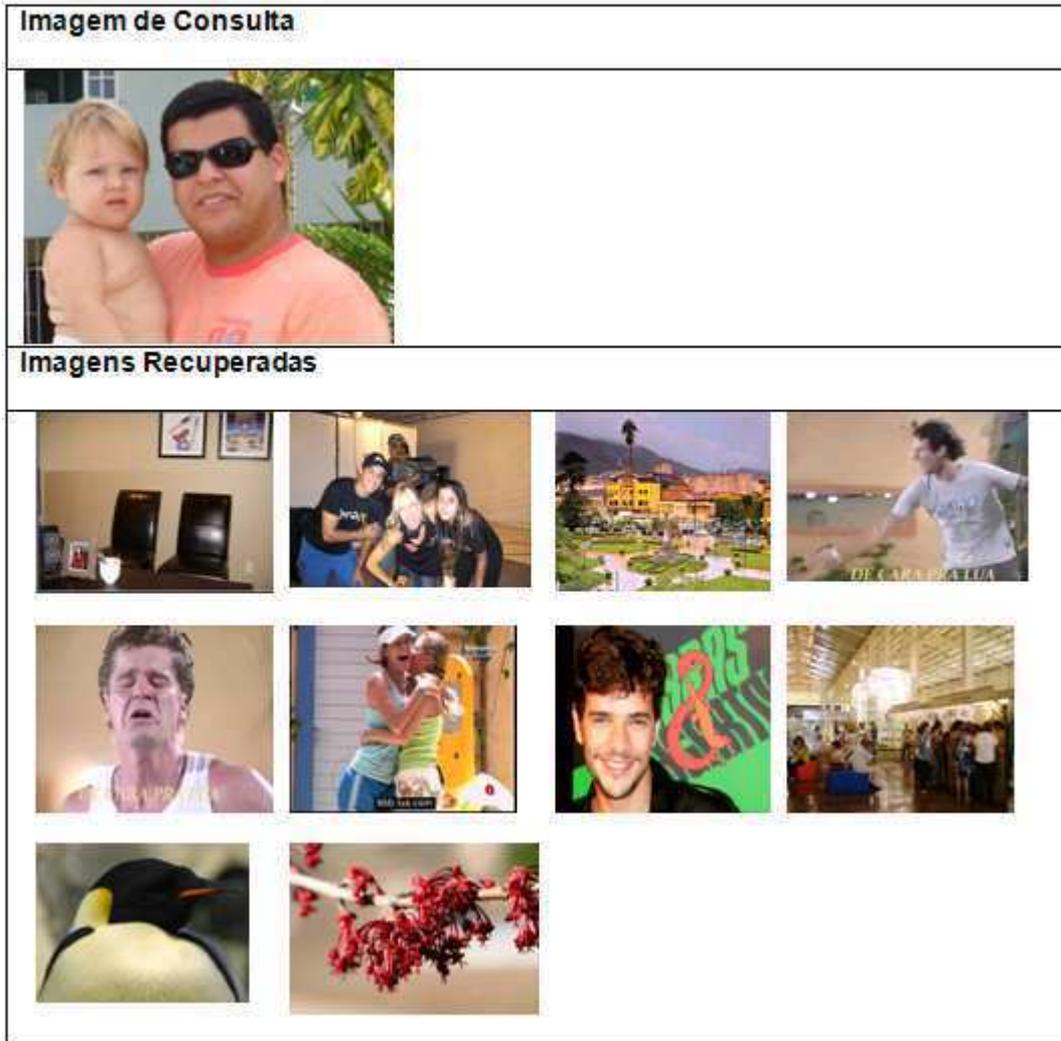


Figura C.26: Resultado da classificação da única imagem para YCbCr com 16 bins.



Figura C.27: Resultado da classificação da única imagem para YCbCr com 32 bins.

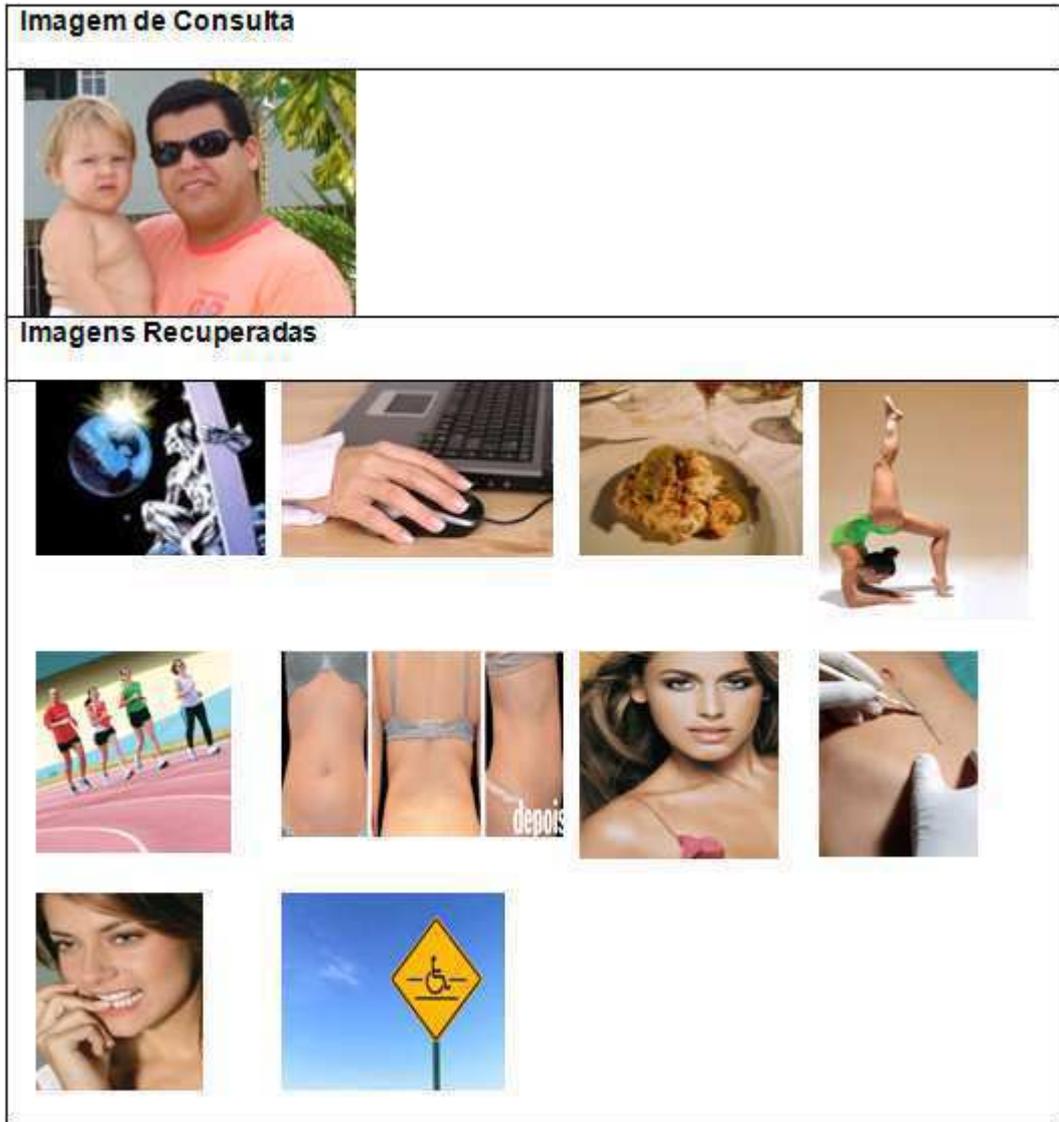


Figura C.28: Resultado da classificação da única imagem para YCbCr com 64 bins.

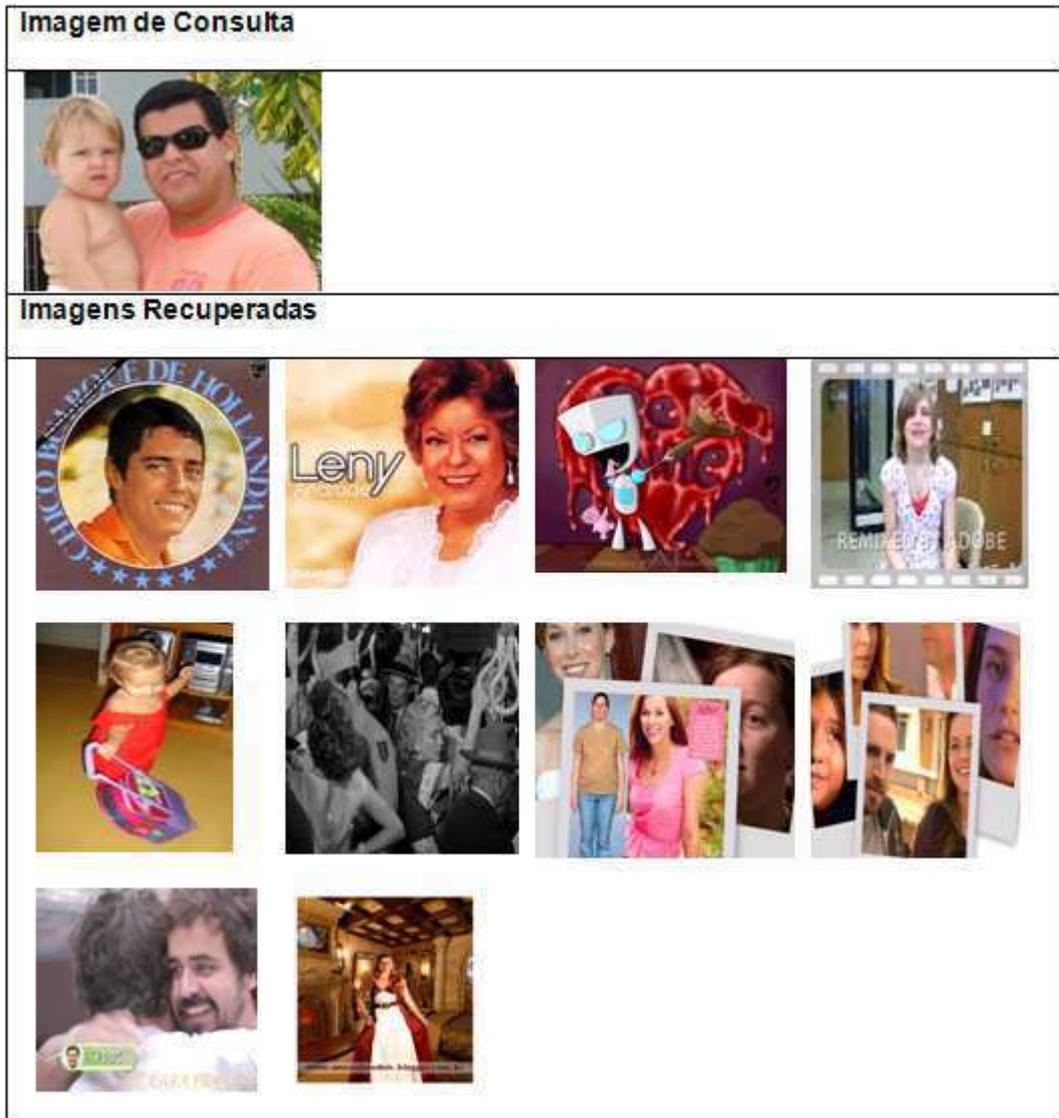


Figura C.29: Resultado da classificação da única imagem para LBP.



Figura C.30: Resultado da classificação da única imagem para *Waveletes*.



Figura C.31: Resultado da classificação da única imagem para Momentos Invariantes de Hu.



Figura C.32: Resultado da classificação da única imagem para combinação de cor.



Figura C.33: Resultado da classificação da única imagem para combinação de textura.



Figura C.34: Resultado da classificação da única imagem para combinação de cor e textura.



Figura C.35: Resultado da classificação da única imagem para combinação de cor e forma.



Figura C.36: Resultado da classificação da única imagem para combinação de textura e forma.



Figura C.37: Resultado da classificação da única imagem para combinação cor, textura e forma.



Figura C.38: Resultado da classificação da única imagem para a melhor combinação.

Apêndice D

Roteiro do Experimento com os Usuários

Instruções

Algumas precauções são de grande importância durante a execução dos testes.

- O usuário não deve estar realizando downloads ou fazendo streams durante os testes (downloads em geral e streams de vídeo ou áudio).
- As imagens de consulta devem estar no formato JPEG (.jpg e .jpeg) e não ocupar mais de 1Mbyte de espaço em disco.
- O usuário deve utilizar os Browsers IE (Internet Explorer, versão 6.0 ou superior) ou Firefox (versão 2.0 ou superior).
- Deve-se seguir todos os passos das instruções, caso algum dê errado, continuar os testes a partir do próximo passo.
- Algumas imagens que não aparecerem, podem ser clicadas e atualizadas para serem visualizadas, e, portanto votadas.

Aviso importante: Todas as imagens utilizadas na recuperação foram obtidas de forma automática pelo sistema, podendo haver imagens de conteúdos politicamente incorretos (por exemplo: racista, erótico e pornográfico).

1. Faça o download de uma imagem da internet genérica (qualquer conteúdo. Ex: praia, futebol, floresta). Essa imagem será sua fonte de teste, não é necessário utilizar várias imagens, com uma imagem você pode realizar todos os testes.

2. Abra a interface Web do LVCRetrieval ([HTTP://150.165.75.116/LvcRetrieval/](http://150.165.75.116/LvcRetrieval/)) e clique em Pesquisa Avançada de Imagens. Uma próxima tela será aberta.

- (a) Observação: Para a votação será necessário apenas selecionar as imagens que você classifica como similar, através dos checkboxes (o contexto de similaridade nesse caso é totalmente individual, já que você poderá julgar a imagem similar devido à cor, forma, textura ou as através das combinação das características. Por isso é importante o usuário observe quais características são mais aparentes na imagem de consulta (Por exemplo: uma imagem com céu e nuvem a característica de cor será a mais importante).

3. Consulta.

- (a) Antes de executar cada teste, é necessário primeiro selecionar uma imagem de consulta.
- (b) Após a seleção da imagem, será necessário selecionar que tipo de extrator irá ser utilizado na recuperação.
- (c) Após a seleção do tipo de extrator a consulta poderá ser executada, através do botão Pesquisa LvcRetrieval.

4. Recuperação.

- (a) A tela de recuperação exibirá a imagem de consulta e logo a baixo as imagens classificadas pelo sistema como as mais similares.
- (b) Selecione os checkboxes das imagens mais similares de acordo com seus critérios (Observação de critérios de votação).
- (c) Clique no botão Submeter Imagens Relevantes.

Os passos 3 e 4 serão repetidos para cada um dos extratores selecionados, totalizando 13 execuções (testes).

Detalhamento dos Testes.

1. Teste com Cor (Selecione a opção Cor)

- (a) Selecionar o espaço de cor HSV com 16 níveis.
- (b) Selecionar o espaço de cor HSV com 32 níveis.

- (c) Selecionar o espaço de cor HSV com 64 níveis.
 - (d) Selecionar o espaço de cor RGB com 16 níveis.
 - (e) Selecionar o espaço de cor RGB com 32 níveis.
 - (f) Selecionar o espaço de cor RGB com 64 níveis.
 - (g) Selecionar o espaço de cor YCbCr com 16 níveis.
 - (h) Selecionar o espaço de cor YCbCr com 32 níveis.
 - (i) Selecionar o espaço de cor YCbCr com 64 níveis.
2. Teste com Texturas (Selecionar a opção Textura)
- (a) Selecionar o extrator LBP.
 - (b) Selecionar o extrator Wavelets.
3. Teste com Formas (Selecionar a opção Formas)
- (a) Selecionar o extrator Momentos de Hu.
4. Combinação 1 (Selecionar a opção Combinação)
- (a) Selecionar os campos Cor e Textura.
5. Combinação 2 (Selecionar a opção Combinação)
- (a) Selecionar os campos Cor e Forma.
6. Combinação 3 (Selecionar a opção Combinação)
- (a) Selecionar os campos Forma e Textura.
7. Combinação 4 (Selecionar a opção Combinação)
- (a) Selecionar os campos Cor, Forma e Textura.