PEDRO RODOLFO KALVA

CLASSIFICAÇÃO DE IMAGENS USANDO COMBINAÇÃO DE CLASSIFICADORES E INFORMAÇÃO CONTEXTUAL

Dissertação apresentada ao Programa de Pós-Graduação em Informática Aplicada da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática Aplicada.

Área de Concentração: *Sistemas Inteligentes*

Orientador: Prof. Dr. Alessandro L. Koerich
Co-Orientador: Prof. Dr. Fabrício Enembreck

CURITIBA

2005

Kalva, Pedro Rodolfo

Classificação de Imagens Usando Combinação de Classificadores e Informação Contextual. Curitiba, 2005. 116p.

Dissertação (Mestrado) – Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática Aplicada.

1. Classificação de Imagens 2. Combinação de Classificadores 3. Sistemas Inteligentes. I.Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e de Tecnologia. Programa de Pós-Graduação em Informática Aplicada II-t

Esta página deve ser reservada à ata de defesa e termo de aprovação que serão fornecidos pela secretaria após a defesa da dissertação e efetuadas as correções solicitadas.

A meus pais que sempre
acreditaram em mim.

A minha querida esposa Maristela
que sempre me deu apoio.

A meus queridos filhos
"Fefe" e "Pepe" que
são a minha vida.

Ofereço esta dissertação.

Agradecimentos

Agradeço de coração ao Professor Doutor Alessandro L. Koerich pelas excelentes contribuições e orientações que recebi. Agradeço a todos os professores e colegas que de alguma forma contribuíram para a formação deste trabalho.

Agradeço a Pontifícia Universidade Católica do Paraná, que vem mudando minha vida deste a época do curso de graduação. Agradeço pela minha formação acadêmica e profissional.

Agradeço a minha família, parentes, amigos e colegas de trabalho, que sempre valorizaram este trabalho e encontraram uma forma de colaborar, mesmo que fosse por meio de palavras motivadoras.

Agradeço aos meus colegas do grupo "javapover", Fernando, Marcelo e Vanderlei que sempre estiveram me apoiando neste e em outros trabalhos. A contribuição de vocês fez toda a diferença.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Símbolos

k	*Valor constante*
L.	*Dimensão*
I	*Índice*
RGB	*Modelo de cor RGB – Red, Green e Blue*
HSI	*Hue, Saturation and Intensity, modelo de cor*
HSV	*Hue, Saturation and Value, modelo de cor*
R	*Canal de cor Red do modelo RGB*
G	*Canal de cor Green do modelo RGB*
B	*Canal de cor Blue do modelo RGB*
P	*Nível de Cinza*
Nnu	*Número de níveis usados*
x	*Coordenada indicando a coluna em uma matriz*
y	*Coordenada indicando a linha em uma matriz*
C(x,y)	*Representação de um ponto em uma imagem resultante*
R(x,y)	*Representação de um ponto em uma imagem, no canal Red do modelo RGB*
G(x,y)	*Representação de um ponto em uma imagem, no canal Green do modelo RGB*
B(x,y)	*Representação de um ponto em uma imagem, no canal Blue do modelo RGB*
R	*Valor resultante de uma operação*
Wx	*Posição em uma mascara*
Zx	*Peso atribuído em uma posição da mascara*
g(x,y)	*Função com base em uma posição da matriz*
f(x,y)	*Função com base em uma posição da matriz ou valor de um ponto em uma matriz*
G	*Imagem de saída*
G'	*Imagem de entrada 1*
G''	*Imagem de entrada 2*
Nf	*Número de formas encontradas*
Nt	*Número total de formas comportadas (capacidade)*
VT	*Vetor de características*
a_{11}	*Posição em uma matriz*
c_{ij}	*Matriz de co-ocorrência*
i	*Linha da matriz*

j	*Coluna da matriz*
LH	*Layer Hidden – Número de neurônios na camada escondida*
LI	*Layer Input – Número de neurônios na camada de entrada*
LO	*Layer Output – Número de neurônios na camada de saída*
P(h)	*Probabilidade a priori*
h	*Hipótese*
D	*Dados de treinamento*
P(D)	*Probabilidade a priori dos dados de treinamento D*
P(h|D)	*Probabilidade da hipótese h dado D*
v_j	*Classe i*
$P(v_j)$	*Probabilidade a priori da classe* v_j
w_k	*Palavra*
$P(w_k \mid v_j)$	*Probabilidade da palavra* w_k pertencer a classe v_j
v_{NB}	*Classe resultante – Naïve Bayes*

Lista de Abreviaturas

HTML	*Hyper Text Makup Language*
SNNS	*Stuttgart Neural Network Simulator*
k-NN	*k – Nearest Neightbord*
SVM	*Support Vector Machine*
HTTP	*Hyper Text Transfer Protocol*
WWW	*World Wide Web*
RNA	*Redes Nerais Artificias*
MSE	Erro Médio Quadrático
SNNS2C	*Stuttgart Neural Network Simulator to C (computer language)*
MLP	*Multi Layer Perceptron*

Resumo

Este trabalho apresenta um novo método para a classificação de imagens que combina informações extraídas das próprias imagens e informações extraídas do contexto. A hipótese principal verificada neste trabalho é de que as informações contextuais associadas a uma imagem podem auxiliar no processo de classificação de imagens. Para verificar esta hipótese utilizou-se um ambiente rico em imagens e informação contextual, a Internet. Neste ambiente foram coletadas páginas *web* contendo imagens e textos que foram então, armazenadas de maneira organizada e estruturada para formar uma base de dados. Inicialmente desenvolveram-se classificadores independentes para imagem e texto. Das imagens foram extraídas características de cor, forma e textura que formaram vetores de características. Estes vetores foram utilizados para treinar e testar classificadores baseados em redes neurais artificiais. Por outro lado, as informações textuais foram processadas e posteriormente utilizadas para treinar e testar um classificador estatístico Naïve Bayes. No final, foram combinadas as saídas de ambos classificadores na tentativa de melhorar a taxa de acerto na classificação de imagens através de diferentes regras de classificação. Os resultados experimentais sobre um conjunto de testes mostram que a combinação dos classificadores propicia um aumento significativo (aproximadamente 16%) na taxa de classificação correta de imagens em comparação aos resultados obtidos pelo classificador baseado em redes neurais que não faz uso da informação contextual. Assim, estes resultados confirmam a hipótese de que informações contextuais podem contribuir de maneira relevante para a classificação de imagens.

Palavras-Chave: Classificação de Imagens, Combinação de Classificadores, Sistemas Inteligentes.

Abstract

This work presents a novel method for the classification of images that combines information extracted from the images and contextual information. The main hypothesis verified in this work is that contextual information related to an image can contribute in the image classification process. To verify such a hypothesis we have used an environment rich in images and contextual information: the Internet. From this environment, web pages containing images and text were collected and stored in an organized and structured fashion to build a database. First, independent classifiers were designed to deal with images and text. From the images were extracted several features like color, shape and texture. These features combined form feature vectors which are used to train and test neural network based classifiers. On the other hand, contextual information is processed and further used to train and test a Naïve Bayes classifier. At the end, the outputs of both classifiers are combined through different rules in an attempt to improve the correct image classification rate. Experimental results on a test dataset have shown that the combination of classifiers provides a meaningful improvement (about 16%) in the correct image classification rate relative to the results provided by the neural network based image classifier which does not use contextual information. Therefore, the results validate the hypothesis that the contextual information is relevant for the image classification task.

Keywords: Image Classification, Classifier Combination, Intelligent Systems.

Capítulo 1

Introdução

O cérebro humano é muito eficiente para reconhecer imagens. A imagem visualizada é projetada na retina, onde os receptores (cones e bastonetes) são estimulados e enviam as informações através do nervo óptico para o cérebro, que decodifica e processa os sinais para então concluir a tarefa de reconhecimento [GON00]. Para este processamento o cérebro humano conta com uma enorme quantidade de informações, sejam visuais ou não, que se cruzam para produzir um resultado final ótimo. Em situações novas o cérebro tende a demorar um pouco mais, pois, novas informações devem ser assimiladas e processadas até o reconhecimento completo.

Os sistemas artificiais construídos com a finalidade de reconhecer imagens trabalham com poucas informações se comparados ao sistema biológico humano. As informações são normalmente pontuais, referenciando apenas algumas características principais extraídas de imagens presentes em um conjunto de treinamento, sem levar em conta qualquer outro tipo de informação. Isto acontece devido às limitações de software e hardware presentes nos sistemas de reconhecimento de imagens. Estas limitações existem pela dificuldade de representar a complexidade envolvida em uma imagem, pois a imagem normalmente representa uma entidade que pode estar relacionada a sons e outros sinais. Mesmo que fosse possível capturar muitas informações a respeito de um determinado objeto, ainda sim poderíamos ter limitações de tempo de processamento e espaço para armazenamento.

1.1. Objetivos

Nos últimos anos, diversas técnicas de reconhecimento de padrões e inteligência artificial têm sido utilizadas na resolução de problemas reais com o objetivo de minimizar a interação humana em tarefas meramente repetitivas. No caso de imagens, a idéia é desenvolver sistemas que

façam a interpretação das imagens de forma autônoma. Os sistemas de reconhecimento/classificação de imagens não devem se limitar a apenas reconhecer os componentes incluídos nesta matriz de pontos coloridos que é chamada de imagem, eles podem também ir além e interpretar o significado de todos os componentes que nela aparecem.

Inspirado pelo comportamento do cérebro humano, neste trabalho o objetivo principal é estudar e verificar o impacto da utilização de informações extras, isto é, informações próximas e/ou relacionadas com a própria imagem, no processo de aprendizagem e classificação de imagens. Estas informações extras consistem em informações textuais que estão também presentes no local onde as imagens se encontram, e que ao longo deste trabalho chamaremos de "*informação contextual*". Deste modo, assumimos como ambiente de aplicação, páginas da Internet no formato HTML. Todas as imagens utilizadas neste trabalho são originárias da Internet e fazem parte de páginas HTML que, além das imagens, possuem muitas informações textuais, cuja utilidade na tarefa de classificação de imagens foi objeto de investigação.

Existem diversos tipos de sistemas de busca e recuperação onde a maioria destes sistemas está associada a grandes bancos de imagens comerciais ou a outros mecanismos que trabalham com imagens [KHE04]. No entanto a grande maioria utiliza palavras-chave como método de indexação e busca [KHE04]. Esta abordagem geralmente implica na necessidade da intervenção humana na classificação e indexação das imagens provocando, assim, um alto custo associado a este tipo de sistema. Por outro lado, existem também sistemas dotados de algoritmos de representação e reconhecimento que utilizam informações extraídas da própria imagem armazenada na base de dados. O resultado da busca se dá pela representação da imagem submetida na entrada e seu reconhecimento para uma das classes pré-definidas, retornando as imagens reconhecidas previamente, daquela classe, presentes na base de dados. Em alguns classificadores são usados algoritmos de agrupamento, onde se tem apenas a quantidade de classes, mas sem definí-las previamente. O reconhecimento (agrupamento) é dado pela similaridade da representação extraída nas imagens submetidas ao sistema.

1.2. Desafio

O primeiro grande desafio é conseguir uma base de dados que apresente condições similares às condições reais, ou seja, imagens e textos. Não se encontraram bases de dados comerciais ou não comerciais com as características necessárias: classes definidas; vínculo com texto e imperfeições (ruídos) normalmente encontradas na Internet. Estas imperfeições referem-se a imagens não

preparadas, com desproporções de tamanho, cores e enquadramento da imagem, portanto não padronizadas como normalmente ocorre com bases de dados comerciais.

Neste trabalho uma ferramenta de busca foi criada para esta finalidade, ou seja, capturar imagens e textos da Internet. Também foram desenvolvidas outras ferramentas para auxiliar no processo de seleção e rotulagem de imagens e textos. Estes procedimentos estão descritos no Capítulo 4, onde é apresentado em detalhes como a base de dados está constituída além de estatísticas interessantes sobre esta forma de aquisição de informações através de ferramentas automáticas.

O segundo desafio é a extração de características e desenvolvimento de um classificador de imagens. Diversos experimentos foram realizados a fim de conseguir um resultado razoável e válido para verificação da hipótese. Este passo é importante, pois os resultados obtidos neste estágio serão comparados com os resultados finais obtidos a partir da combinação de classificadores a fim de demonstrar que a classificação de imagens pode ser melhorada ao utilizar-se de informações contextuais. Este passo está detalhado no Capítulo 5.

Finalmente, o terceiro desafio é a criação de um algoritmo que interprete o resultado dos classificadores de imagem e de texto, combine-os de maneira adequada produzindo um resultado mais eficiente do que a simples classificação de imagens.

Dentre as diversas técnicas pesquisadas para a classificação de imagens, as redes neurais ganham grande destaque pelo desempenho conseguido e pelo número de pesquisas já realizadas. Esta técnica exige que as imagens sejam transformadas em representações vetoriais que podem ser calculadas e comparadas computacionalmente com maior facilidade. Estas representações vetoriais são conhecidas como vetores de características (ou *features*), e seu objetivo é o de representar de forma numérica o conteúdo da imagem, baseando-se em algumas características específicas (quantidade e distribuição de cores, formas, etc.). Estes vetores são então submetidos à camada de entrada de uma rede neural devidamente configurada e treinada para esta finalidade. Após o processamento, a rede apresenta na sua camada de saída uma indicação da classe de pertinência mais provável, levando em conta somente as características extraídas.

Em sistemas deste tipo devemos inicialmente definir classes de reconhecimento, que são as interpretações possíveis para cada saída da rede neural (normalmente uma saída, ou combinação delas). Exemplos de classes seriam: pessoas, automóveis, árvores, objetos em geral, etc. Entretanto, uma imagem pode conter várias classes simultaneamente, e o contexto pode apontar para apenas uma classe (ou algumas delas). Como exemplo, observe a imagem exibida na Figura 1.1. Se nossas

classes forem: carro, pessoa e animal. Temos as três classes nesta imagem e o classificador, se estiver funcionando adequadamente, vai apresentar em sua saída um resultado próximo de 33,3% para cada uma destas classes. Supondo agora que esta imagem estivesse em um site de automóveis. Neste caso, provavelmente diríamos que é uma propaganda de um automóvel e a classificaríamos como tal. Por outro lado, se esta foto estivesse em uma página web sobre animais domésticos, provavelmente classificaríamos esta imagem como sendo de um cachorro e consideraríamos que os outros elementos como a automóvel e a pessoa somente fazem parte do cenário. Este simples exemplo demonstra a importância do contexto e mostra que ele pode levar as diferentes interpretações e influenciar de maneira significativa o processo de classificação.

Figura 1.1 – Exemplo de imagem com várias classes

Se a classificação de imagens geralmente necessita do contexto para um bom desempenho, então talvez seja uma boa idéia considerá-lo em um sistema computacional. Para isto, pode-se utilizar a informação presente no local onde a imagem se encontra. Assim, em imagens provenientes da Internet, pode-se usar a informação textual presente na página HTML. Desta forma têm-se duas informações distintas e que devem ser tratadas separadamente: imagens e textos. Uma das possíveis utilizações desta informação contextual seria no caso o resultado de um classificador

de imagens apresentar um nível baixo de certeza, indicado por probabilidades similares nas saídas de uma rede neural, por exemplo. Neste caso, utiliza-se a informação contextual para reforçar ou até mesmo alterar a classe final de classificação do sistema.

Além do classificador de imagens propriamente dito, foi necessária a construção de um classificador para o que chamamos neste trabalho de "informação contextual". Este classificador tem por objetivo classificar o texto que compõe a página *web* de onde a imagem foi extraída. Para a classificação dos textos foi construído um classificador estatístico que está devidamente detalhado no Capítulo 6.

O último passo aborda a necessidade de juntar todos os resultados dos classificadores de imagens e de textos e fazê-los trabalhar em conjunto para a produção de um resultado final e a constituição de um ambiente propício para validação dos resultados obtidos. A junção, ferramentas e técnicas utilizadas na combinação dos classificadores são discutidas no Capítulo 7. Os resultados obtidos são apresentados, avaliados e discutidos no Capítulo 8.

1.3. Contribuições

Para a comunidade científica, diversos aspectos deste trabalho podem ser tomados como interessantes. A contribuição mais significativa deste trabalhado está em mostrar que a informação contextual realmente contribui de maneira significativa no processo de classificação de imagens.

A estratégia de formação da base de dados usada neste trabalho demonstra que é possível usar a Internet para capturar imagens e textos e então formar uma base de dados significativa para trabalhos científicos. A própria base de dados formada neste trabalho, contendo imagens de automóveis, pessoas, animais domésticos, motos e CDs e DVDs pode ser usada para outros trabalhos similares, ou que apenas necessitem de imagens deste tipo. Com algumas modificações no processo de captura de imagens é possível diminuir a interação humana, agilizando e facilitando a criação da base de dados de imagens com base na Internet.

Para a classificação de textos este trabalho usou o algoritmo Naïve Bayes. A principal colaboração neste caso é que este trabalho usou o algoritmo com textos em português e inglês ao mesmo tempo. Tanto o vocabulário quando as amostras foram formadas de palavras nestas línguas sem qualquer tipo de separação. Este classificador teve bom desempenho mesmo com esta característica, demonstrando a eficiência do classificador Naïve Bayes.

Este trabalho abordou uma estratégia de combinação de classificadores baseadas em regras que foram geradas com apoio de algoritmos conhecidos, como o C4.5 [DUD00][MIT97]. Esta

estratégia mostrou-se eficiente, podendo ser tomado como uma opção avançada para problemas de combinação de classificadores.

Para aplicações comerciais e a comunidade em geral, a classificação de imagens em ambientes que possuem informação contextual é muito útil em diversas situações. Poderíamos criar um sistema mais eficiente de procura de informações na Internet. Bastaria utilizar os procedimentos aqui propostos para buscar as imagens na Internet, porém, ao baixar a imagem esta seria imediatamente classificada e somente em caso positivo seria definitivamente armazenada, constituindo um sistema de busca de imagens pela Internet muito eficiente.

Outro exemplo seria um classificador de conteúdo para utilização com navegadores Internet. Atualmente muitos classificadores de conteúdo trabalham analisando os endereços ou meta informações na página. A eficiência fica comprometida se a página não apresentar o conteúdo classificado. Muitos sites pessoais também não usam classificação e podem conter imagens impróprias para crianças. Com o uso de um mecanismo de classificação de imagens como o demonstrado neste trabalho, a probabilidade de barrar uma imagem imprópria não mais dependeria de uma lista atualizada de endereços ou as meta-informação da página, sendo analisados diretamente as imagens e seu conteúdo contextual.

1.4. Hipótese

Dado o problema e suas aplicações, este trabalho pretende comprovar a hipótese de que informações contextuais são relevantes para o processo de classificação de imagens. Para chegar nesta conclusão foi construído um classificador de imagens de forma convencional (rede neural do tipo perceptron multicamadas, extraindo características de formas, cores e textura) e também um classificador que utiliza a informação contextual juntamente com a informação da classificação de imagens. Uma base de dados representativa foi usada para avaliar o resultado final que é gerado a partir da comparação dos resultados dos dois métodos. Um índice de desempenho relativamente alto representa a confirmação desta hipótese.

1.5. Organização do Documento

Este trabalho está dividido em oito capítulos sendo o primeiro esta introdução, que mostrou de forma superficial o problema, o desafio e motivação que este trabalho aborda. O Capítulo 2 descreve o estado da arte, apresentando os principais trabalhos que existem nesta área. O Capítulo 3 aborda o experimento sob um ponto de vista geral. O Capítulo 4 apresenta o procedimento de

formação da base de dados, escolha de amostras válidas, rotulagem e como foi feito a separação em conjuntos para treinamento e testes. Os Capítulos 5 e 6 apresentam os classificadores de imagens e texto em detalhes, como foram construídos e como foram treinados. No Capítulo 7 abordamos detalhadamente a camada de decisão, que provê a combinação dos classificadores, a geração e uso de regras para a produção do resultado final. E finalmente no Capítulo 8 são apresentados os resultados finais, detalhando passo-a-passo o processamento em cada parte deste experimento.

Os Capítulos 5, 6 e 7 utilizam algoritmos com algum tipo de treinamento, desta forma são apresentados resultados de testes feitos para o problema específico de cada classificador e estes resultados diferenciam dos resultados apresentados no Capítulo 8, que contempla o resultado totalmente integrado com todas as partes deste experimento e com um conjunto diferenciado de teste.

Capítulo 2

Estado da Arte

Classificação de imagens é um problema amplamente pesquisado atualmente, isto pode ser observado em razão do grande número de artigos publicados sobre este assunto [FLE96]. Em nenhum dos trabalhos pesquisados encontramos uma taxa de classificação correta de 100%. Isto demonstra a dificuldade desta tarefa que pode ser agravada se a qualidade das amostras de imagens for baixa. Na maioria dos algoritmos pesquisados a única informação utilizada na classificação provém da própria imagem, através da extração de informações relevantes por meio de algoritmos que analisam a imagem sob diferentes aspectos. Poucos trabalhos fazem uso de informações externas à imagem, ou seja, informações extraídas na origem da imagem ou outras informações ligadas à imagem, mas que, porém, não estão contidas na própria imagem.

Destes trabalhos, principalmente quando o ambiente de classificação de imagens é a Internet, encontramos alguns trabalhos que utilizam informações externas à imagem, como o trabalho de Rowe [ROW02]. O conteúdo textual analisado para a classificação de imagens é um parâmetro específico para esta finalidade, especificado pelo HTML (parâmetro *alt* de uma *tag* de imagem). Porém, o problema é que nem todas as imagens em páginas HTML possuem este parâmetro preenchido, pois ele não é obrigatório no padrão HTML. Para tentar contornar este problema, [FEN04] utilizam, juntamente com o parâmetro de descrição da imagem, o conteúdo textual próximo à imagem, o título da página HTML que contém a imagem e até mesmo o nome do arquivo da imagem. Todo este conteúdo é utilizado juntamente com um classificador de imagens a fim de melhorar seu desempenho. Os resultados destes experimentos indicam que as informações presentes no local onde a imagem se encontra podem ser importantes para a classificação de imagens, quando estes ambientes forem ricos em conteúdo, como é o caso das imagens na Internet.

Neste capítulo serão abordados classificadores de imagens, classificadores de textos e algumas formas de combinar estes resultados. Ao final serão abordados alguns trabalhos que utilizaram outros classificadores além dos classificadores de imagens para melhorar o desempenho durante um processo de classificação de imagens.

2.1. O Processo de Classificação de Imagens

O processo de classificação de imagens geralmente segue um modelo onde alguns elementos básicos são comuns para qualquer tipo de classificador, como, por exemplo: aquisição de imagens, o pré-processamento, a segmentação, a representação (extração de características) e classificação propriamente dita [GON00]. Este esquema é apresentado na Figura 2.1.

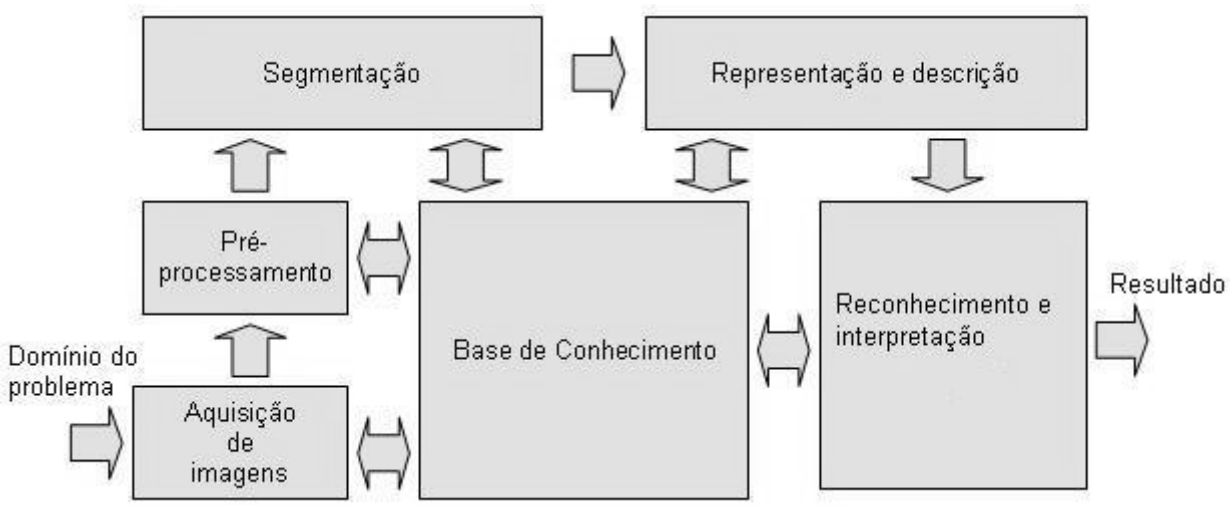


Figura 2.1: Esquema básico para classificação de imagens proposto por Gonzalez e Woods [GON00].

Dependendo do caso, alguns passos intermediários podem ser incluídos ou até mesmo eliminados deste modelo. O problema inicia durante a aquisição da imagem, que pode vir de uma base de dados, de uma câmera de segurança, etc. Esta imagem passa por um pré-processamento, onde ocorre a eliminação de ruídos e pequenos ajustes, como rotações, filtragens, etc. A segmentação trata de extrair ou identificar na imagem alguns elementos pré-determinados, como por exemplo, em um sistema de reconhecimento de faces a segmentação pode identificar o rosto da pessoa. Basicamente esta etapa é realizada com algoritmos de agrupamento [SHI97]. A imagem, então, é submetida a um processo de representação, também conhecido como extração de características (atributos ou *features*) que analisam a imagem sob um determinado ponto de vista e

gera valores numéricos que os representam. De posse da representação, a imagem pode ser submetida ao reconhecimento, classificação ou interpretação através de um determinado algoritmo de classificação. Todos os passos geram informações que podem ser usadas para a melhoria geral do processo. Estas informações são armazenadas formando uma base de conhecimento.

2.2. Aquisição, Seleção e Pré-Processamento de Imagens

Os sistemas classificadores de imagens necessitam de um banco de dados contendo amostras de imagens para treinamento e testes. Uma maneira de adquirir estas informações é através de uma base comercial de imagens como, por exemplo, a CORBIS [COR04] e a NIST [NIS04], e até mesmo coletar na Internet por meio de processos automáticos [OLI02]. O problema deste último caso ocorre pelo excessivo trabalho na aquisição e principalmente na separação de amostras úteis e não úteis, além de ser necessário proceder a uma rotulagem manual das imagens. Alguns tratamentos podem ser efetuados automaticamente, como no trabalho de Oliveira *et al.* [OLI02] onde as imagens passam por etapas básicas de pré-processamento.

Após uma pré-seleção das imagens, estas podem necessitar de pré-processamento [GON00] [ALB00] [GON94]. Este pré-processamento pode incluir redimensionamento das imagens, filtragens, etc. O pré-processamento auxilia a corrigir alguns problemas, como ruídos, aumentando a qualidade do resultado final. Se o processo exigir um formato de arquivo único, então devem ser efetuadas as conversões necessárias das amostras que não estão no formato exigido.

O pré-processamento pode utilizar desde técnicas simples e bem conhecidas ou incorporar algoritmos complexos, como é o caso de Albuquerque *et al*. [ALB00], que empregou redes neurais para problemas simples, como conversão de imagens coloridas para níveis de cinza.

2.3. Extração de Características de Imagens

Devido a grande quantidade de informações presentes em uma imagem, onde muitas delas podem não ser relevantes para o processo de classificação, ela deve ser representada de uma maneira mais sucinta de modo a permitir sua manipulação através de um vetor de características (ou *features*) [NAD93]. O processo de extração e geração de um vetor de características também está intimamente ligado ao tipo de classificador, pois, deve fornecer um resultado compatível com a entrada deste classificador. Normalmente as características apresentam valores numéricos que são posicionadas lado a lado formando um vetor de características [MIC94] [GON00].

As características são extraídas de forma a representar sucintamente, porém significativamente as classes, enfatizando suas diferenças. Park *et al* [PAR99] utilizaram uma estrutura de nós e conexões extraídas após um processo de *esqueletização* das imagens para representar animais marinhos. Os nós, conexão e ordem são as representantes que compõe o vetor de características. Este processo é útil para a representação de formas. Outra abordagem para formas foi usada por Hirata *et al.* [HIR99], onde formas geométricas pré-definidas eram procuradas na imagem.

As cores também são muito utilizadas como características. Uma análise de histograma verificando-se o relacionamento entre as cores pode revelar diferenças entre fotos e desenhos [OLI02]. Características extraídas das cores também são comuns. Shinmoto *et al.* [SHI02] utilizaram estas informações para a classificação de pessoas e paisagens. Jain e Vailaya [JAI95] utilizaram os histogramas calculando a proximidade, num processo de agrupamento. Normalmente quando são utilizadas cores, estas são convertidas para o espaço HSV que se caracteriza por ter maior representatividade do que o modelo RGB, que mantém os três canais independentes.

Além de formas e cores, outra característica muito encontrada nos classificadores de imagens são características baseadas em texturas, como em [LEP03], onde um classificador baseado somente nesta característica apresentou bons resultados na classificação de imagens de rochas. Para outro classificador [LON00], a textura foi analisada sob quatro características: intensidade, contraste, escala e orientação. Além destas características, muitas outras informações podem ser extraídas de imagens. Bittencourt *et al.* [BIT00] utilizou informações sobre vizinhança e bordas para resolver problemas simples no processamento de imagens.

Dependendo do problema, podem-se criar extratores de características especializados, como um sistema de reconhecimento de rostos humanos. Neste caso as características podem ser tão específicas quanto à análise de elementos como os olhos, o nariz e a boca [LAS01]. Outra forma específica de reconhecimento de faces está na transformação da imagem para figuras [TAN02], onde o tratamento de cores pode facilitar a classificação.

2.4. Classificadores Neurais

Neste trabalho estão sendo abordados dois tipos distintos de informações: as informações textuais e as informações contidas na própria imagem. Como o objetivo deste trabalho é verificar a influência da informação contextual no processo de classificação de imagens, foi necessário

construir dois classificadores. Como as informações têm características bastante distintas, serão utilizados classificadores diferentes para trabalhar com cada tipo de informação. Para a classificação de imagens foi utilizada uma rede neural.

A opção por utilizar redes neurais na classificação de imagens se deve a facilidade das redes em trabalhar com problemas complexos [KOV02]. Desde a descoberta de seu potencial atuando em multicamadas, as redes ficaram conhecidas como classificadores universais, podendo atuar com bons resultados em quase qualquer tipo de problema [GON00]. Koerich [KOE03] utilizou uma rede para classificar imagens de caracteres manuscritos, onde se verificou o bom desempenho para este tipo de problema. Ainda relacionado a imagens, foram feitas pesquisas utilizando-se redes em problemas simples de pré-processamento [BIT00], como criação de filtros para eliminação de ruídos.

Embora haja muitos trabalhos na área de redes neurais, existem muitos que apresentam deficiência em seu conteúdo, conforme observado por Flexen [FLE96] que comparou 119 artigos sobre redes neurais e verificou que a maioria apresentava falta de algum tipo de informação, conjunto de dados muito pequeno ou incoerência durante o experimento. Neste trabalho foi seguido as diretrizes apresentadas no artigo citado.

Outras características como cores [MEL97], textura [CHE96], formas (*shapes)* [PAR99] e diversos outros experimentos [OLI02] [KOE03] [BIT00] [MIC94] [GON00] [KOV02] [MEL97] [HAY01] [RIC93] [NAD93], demonstram que este tipo de classificador tem bom desempenho para este tipo de problema. Todos os trabalhos citados também apresentam uma margem de erro, e nenhum dos experimentos obteve 100% de acerto, o que justifica a introdução de um outro classificador auxiliar para tentar melhorar o desempenho. Entretanto uma comparação direta dos resultados obtidos por outros pesquisadores dificilmente será possível devido a diferenças nas bases de dados e protocolos experimentais.

2.5. Classificadores Estatísticos

Dado a margem de erro demonstrada nos diversos trabalhos referentes aos classificadores baseados em redes neurais, uma alternativa seria tentar otimizá-los. Entretanto, na linha adotada neste trabalho, buscam-se a construção de outros classificadores que operem em espaços de representação diferente. Neste caso referimo-nos a textos coletados juntamente com as imagens, e que formam (representam) o contexto das mesmas.

Muitos algoritmos podem ser utilizados para a classificação de textos [GOL02], como redes neurais, Naïve Bayes*, k* vizinhos mais próximos (ou *k-NN*)*,* máquinas de suporte vetorial (ou *SVM*) dentre outros. Goller *et al.* [GOL02] apresentou um estudo fazendo a comparação entre os classificadores de texto e nas condições do teste, o classificador SVM apresentou superior desempenho em relação aos demais. Porém, em muitos casos, desempenho não é o único critério para a escolha de um classificador. Deve-se levar em consideração outras características, de acordo com o problema. Se a interação com outros classificadores for necessária, é interessante utilizar classificadores que apresentem saídas compatíveis de modo a facilitar uma possível combinação.

Outro algoritmo que merece destaque é o *Naïve Bayes* [SUE00]. O classificador *Bayesiano* trabalha com a fórmula de *Bayes* que envolve o cálculo de probabilidades *a priori* e determina a probabilidade de uma determinada amostra pertencer a uma classe [SUE00] [CIR03] [LAN92]. Apesar do algoritmo *Naïve Bayes* se basear no pré-suposto de independência condicional (corrigida com as redes *Bayesianas*), a classificação de textos tem um bom desempenho com o uso deste algoritmo [SUE00].

2.6. Combinação de Classificadores

Após os resultados gerados pelos dois classificadores, um baseado em redes neurais e o outro um *Bayesiano*, pretendemos combiná-los e entregar o resultado para a camada de decisão. A combinação envolve técnicas de fusão dos dois classificadores, obtendo uma terceira saída. Esta saída também será submetida à camada de decisão, que fará a avaliação final.

Para combinar os resultados primeiramente devem-se normalizar as saídas dos classificadores para que apresentem saídas compatíveis. No caso da rede neural basta simplesmente normalizar todas as saídas e no classificador *Bayesiano* devem-se avaliar todas as classes para então normalizá-las. Com as saídas compatíveis podem-se utilizar técnicas de combinação convencionais como a votação, listas de *ranks*, multiplicação, somatório ou média das saídas, conforme descritas em *Suem et al.* [SUE00], *Kittler et al*. [KIT98] e *Bahler et al.* [BAH00].

2.7. Segmentação de Áreas da Imagem

Outra forma de extração de características para as imagens consiste em dividí-las em várias sub-regiões e atribuir pesos diferentes para cada parte. Com esta técnica, as partes consideradas mais importantes das imagens podem ser realçadas [GON94] [STR96] [ZHO04]. Dentre os estudos

analisados sobre divisão de regiões, *Gong* [GON94] propôs a divisão em nove áreas. Já *Stricker et al.* [STR96] propôs um modelo em cinco regiões salientando a região central, conforme se pode observar na Figura 2.2.

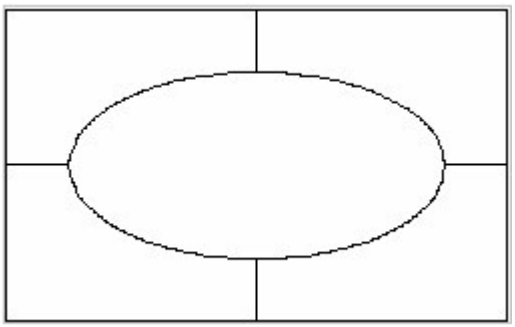


Figura 2.2 – Divisão em regiões proposto *Stricker et al* [STR96]

Aperfeiçoando o modelo anterior, *Zhou* [ZHO04] preocupou-se além da região central e observou que estes modelos apresentam deficiência em relação às regiões não-centrais, pois figuras simétricas podem ser penalizadas pelas diferenças nas regiões não centrais. Este problema pode ser observado na Figura 2.3, onde uma imagem simétrica apresenta diferenças, se considerarmos cada área separadamente.

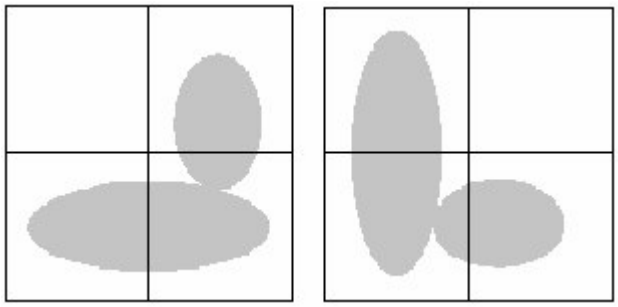


Figura 2.3 – Diferenças apresentadas nas regiões em uma imagem simétrica [ZHO04]

O modelo apresentado por *Zhou* [ZHO04] apresenta uma separação radial, onde a simetria é considerada em qualquer ângulo de rotação. Um exemplo desta técnica pode ser observado na Figura 2.4.

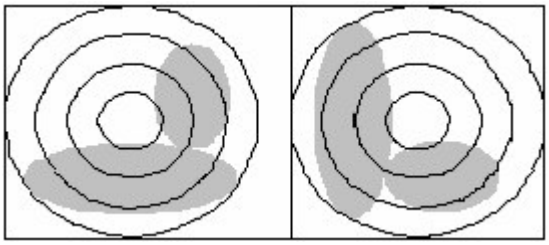


Figura 2.4 – Método de particionamento de regiões radiais [ZHO04]

2.8. Abordagem de Classificação de Imagens usando Combinação de Classificadores

Para imagens oriundas da Internet, existe muita informação contextual na página de origem das imagens. Assim, além da simples classificação convencional de imagens é possível levar em conta também as informações de contexto. Um exemplo de uso destas informações é apresentado por Rowe *et al.* [ROW02] onde foi construído um robô para busca de informações na Internet (*web crawler*), neste caso para buscar imagens. Este trabalho foi baseado em informações textuais coletados em pontos estratégicos da página HTML, como o título, incidências textuais referenciando a imagem, dicas obtidas pelo parâmetro *alt* da *tag* de imagens (*IMG*) entre outros. A Figura 2.5 demonstra um exemplo da captura de informações [ROW02].

```
<title>Sea Others</title>
<h2>The California Sea Others</h2>
<a href="imagens/other.jpeg><img
src="imagens/smallotther.jpeg" alt="Pair of sea others"
/></a>
<center><i>Click on the above to see a larger picture.</i></center>
<hr><a href="home.html">Go to home page</a>
```

Figura 2.5 – Exemplo de captura de textos relacionados com a imagem [ROW02]

Neste exemplo apenas as quatro frases em negrito seriam consideradas. A última frase "*Go to home page*" não é considerada, pois existe uma *tag HR,* que representa uma linha separadora. O título é sempre considerado, além de elementos textuais destacados pelas *tags H2* (texto de nível 2), *I* (texto em itálico), etc. Neste mesmo trabalho é feito uma verificação na imagem para descobrir se

é uma fotografia ou apenas um gráfico, pois os gráficos devem ser descartados. Para isto são verificados alguns atributos como tamanho, número de cores, variação de cores, etc.

Já no trabalho de [HU04] o objetivo é categorizar as imagens de um *site* de notícias e desta forma possibilitar decisões quanto a prioridade no carregamento de imagens, principalmente com dispositivos ou ambientes que tenham restrições de largura de banda como é o caso de alguns dispositivos pessoais portáteis. A categorização procura identificar imagens que constituem propagandas, cabeçalhos, ícones, etc., possibilitando carregar os itens prioritários, que são associados ao texto principal e então, se tiverem largura de banda suficiente, baixar as demais. Um exemplo de elementos das páginas é apresentado na Figura 2.6 [HU04].

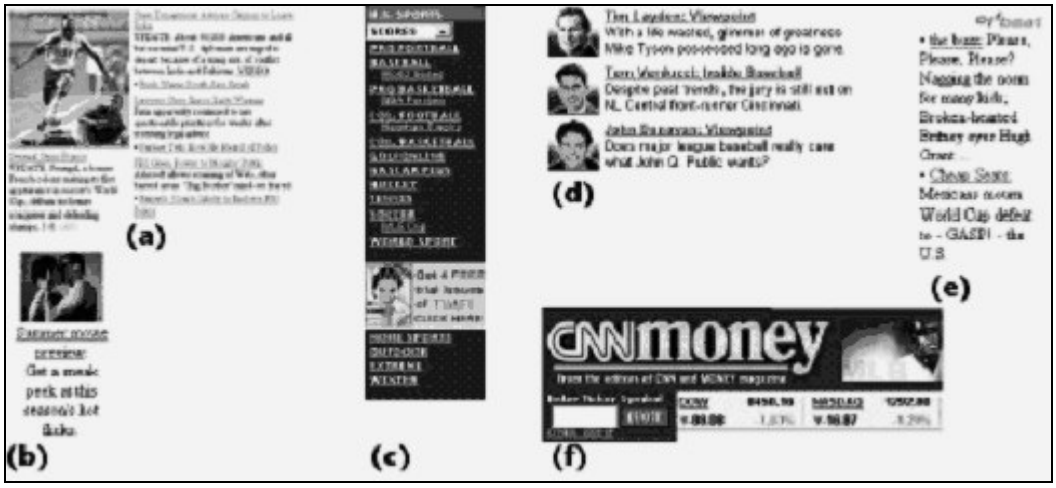


Figura 2.6 – Exemplo de elementos em páginas HTML [HU04]

Para efetuar a categorização são classificados as imagens e textos associados. No caso das imagens, assim como no trabalho de Rowe *et al.* [ROW02], são classificadas como imagens ou fotografias. Ao ser categorizada como fotografia, a imagem é marcada como representante da superclasse "SPA" (*Story*, *Preview* ou *Host* - Representado pela letra A para não confundir com a classe *Heading,* a seguir), caso seja gráfico é marcada como "CIHF" (*Commercial*, I*cons and Logos*, *Heading* ou *Formatting*).

O texto extraído para classificação é semelhante ao trabalho anterior, ou seja, a partir da coleta de informações de locais estratégicos e principalmente do parâmetro *alt* da *tag* de imagens (IMG). Destes textos são coletadas cinco características e formam um vetor de características juntamente com a classificação da imagem ficando com seis dimensões o vetor final. Estes vetores são submetidos para um classificador baseado em SVM (*Support Vector Machine*).

O trabalho de Lu *et al*. [LU01] utilizou-se de classificadores textuais e de imagens, para agrupar (fazer um *clustering*) páginas com conteúdo multimídia, formando uma árvore com as amostras de páginas a serem classificadas, como apresentado na Figura 2.7 [LU01].

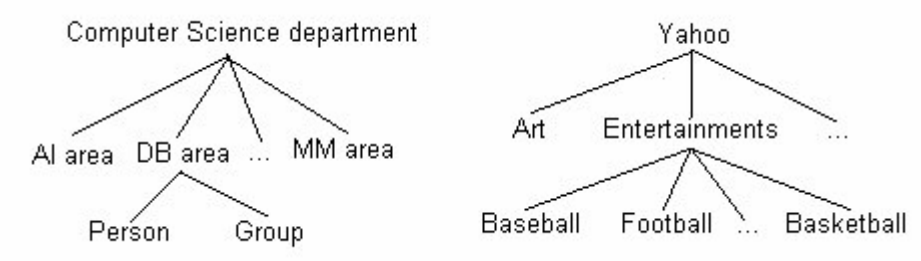


Figura 2.7 – Estrutura hierárquica exemplo [LU01]

Para a classificação textual foi utilizado o classificador *Naïve Bayes* e para a classificação de imagens foi usado um processo baseado no histograma. Um processo recursivo, com base nos classificadores textuais e de imagens define a hierarquia. Para efetuar o agrupamento das imagens foi escolhido o algoritmo *k-Means*, sendo baseado na distância Euclidiana. Este algoritmo é bastante usado em vários problemas relacionados à classificação de imagens, como em Jain *et al.* [JAI95], onde a principal característica extraída através de cores e formas era calculada utilizando PCA (*Principal Component Analysis*).

Por último, o trabalho de Fen *et al.* [FEN04] segue a tendência dos trabalhos anteriores, coletando o texto que supostamente referencia a imagem, praticamente nas mesmas condições que o trabalho de [ROW02], [HU04] e [LU01], ou seja, no título da página HTML, frases em torno de *tags* de formatação, no parâmetro *alt* da *tag* de imagens, etc. Para as imagens são extraídas características baseadas em cores, em formas geométricas (*shapes*) e em texturas. As características extraídas das imagens e os textos são classificadas por um classificador do tipo SVM (*Support Vector Machines*).

2.9. Resumo

Neste capítulo pode-se perceber a dificuldade pertinente a uma tarefa de reconhecimento de imagens. Muitos passos são necessários na tentativa de interpretação e reconhecimento da imagem e todos se baseiam na extração de características, que analisam a imagem sob um ponto de vista fixo. A união destas características pode melhorar o desempenho da classificação. Apresentou-se também que nenhum classificador teve um índice de 100% de reconhecimento, dado o volume de estudos

nesta área [FLE96]. Com base nesta afirmação, é perfeitamente justificável o fato de tentar buscar informações sobre as imagens em locais externos a ela, como no seu local de origem, principalmente quando a origem apresenta muitas informações, como é o caso da Internet. Sendo assim é importante efetuar análises no texto que acompanha a imagem. Os últimos quatro trabalhos apresentados abordam justamente este ponto de vista. O primeiro trabalho abordado [ROW02] utilizou-se apenas das informações textuais, porém os outros trabalhos juntaram os dois tipos de informações (imagem e texto) em níveis diferentes e com formas diferentes de prover a classificação. Os objetivos dos trabalhos também são diferentes, mas todos envolviam imagens no ambiente Internet.

Todos estes trabalhos usaram a informação textual, praticamente efetuando as mesmas seleções, ignorando o restante do texto. Mas textos eliminados não poderiam ser pertinentes para a definição do assunto geral da página HTML? Muitas páginas não apresentam o conteúdo devidamente formatado, prejudicando a coleta de informações nas condições apresentadas nestes trabalhos. Será que toda a informação contextual de uma página HTML não poderia ser relevante para a classificação de imagens, se comparado à classificação de imagens de forma convencional?

Enfim, o objetivo principal deste trabalho é o de comprovar esta hipótese, sem escolher itens específicos do contexto de uma imagem e sim utilizar o máximo possível de informações em classificadores distintos e posteriormente unidos por uma camada de decisão que verificará situações onde é melhor confiar em um dos classificadores, ou nenhum deles e definir a decisão final por meio de combinações entre eles.

Capítulo 3

Método Para a Classificação de Imagens

Este capítulo descreve a metodologia adotada para comprovar a hipótese da melhora da taxa de classificação correta de imagens quando se utilizam informações contextuais ao lugar da simples classificação baseada unicamente no conteúdo da imagem. Para esta comprovação foram gerados resultados com um classificador de imagens baseado em redes neurais e em seguida estas imagens foram submetidas ao experimento completo, ou seja, foram obtidos resultados da rede neural e de um classificador estatístico, que foram combinados e processados por uma camada de decisão. Os resultados foram comparados e a vantagem em se utilizar informação contextual foi comprovada. Detalhes sobre resultados são apresentados no Capítulo 8.

Um dos requisitos deste trabalho é o de efetuar testes com uma base de dados que contenha informações muito próximas da realidade. Assim, seriam necessárias amostras de imagens e textos com ruído. A tarefa de construir uma base de dados adequada faz parte deste trabalho, sendo detalhada no Capítulo 4. Este capítulo apresenta uma visão geral a respeito do sistema construído, sendo que as partes importantes deste sistema serão detalhadas em capítulos específicos.

3.1. Visão Geral do Método Para a Classificação de Imagens

Este trabalho propõe o uso de informações contextuais para auxiliar e melhorar o desempenho de classificadores de imagens. Sendo assim, foi construído um classificador de imagens, cujos detalhes serão apresentados no Capítulo 5. Este classificador será treinado e avaliado e em seguida, com as mesmas amostras submetidas para o classificador de imagens, serão avaliados os resultados da combinação dos classificadores de imagem e texto para verificar a melhora de desempenho que a utilização das informações contextuais pode propiciar.

Para a realização deste trabalho teve-se a preocupação em cuidar para que nenhuma amostra de imagem usada durante as etapas de treinamento fosse reutilizada durante os testes, o que poderia produzir uma "tendência". Outro ponto importante é que para este experimento não foi utilizada nenhuma base de dados de laboratório. A base de dados foi construída a partir de informações extraídas "em campo". A base de dados será detalhada no próximo capítulo.

Podemos dividir este trabalho em duas partes, sendo a primeira abordando a construção e treinamento de cada parte do sistema, que basicamente foi feita de forma separada para cada módulo. Os módulos são: base de dados, classificador de imagens, classificador de texto e camada de decisão. A Figura 3.1 apresenta uma visão geral deste processo.

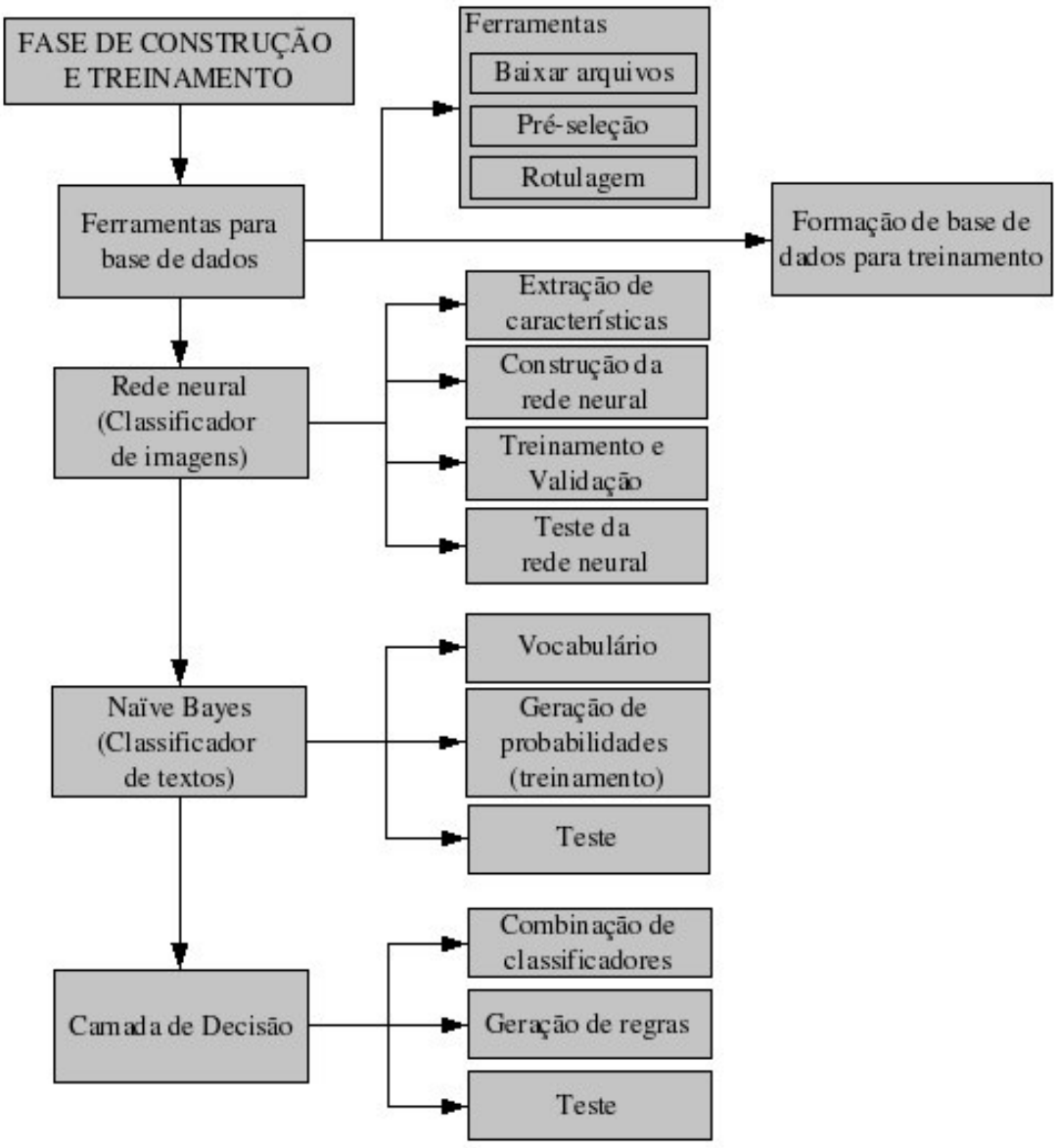


Figura 3.1 – Visão global do processo de classificação de imagens usando informações contextuais.

Durante a fase de construção e treinamento, a primeira atividade foi construir ferramentas para a obtenção das informações na Internet e construir uma base de dados para treinamento e testes. Este processo foi relativamente demorado, sendo então construída uma base de dados preliminar para que testes com imagens e textos pudessem ser feitos antes da conclusão da base de dados. Ao final do processo de construção da base de dados, todos os algoritmos passíveis de treinamento foram treinados novamente.

A base de dados incorpora ferramentas para buscar a informação automaticamente pela Internet a partir de uma lista de endereços iniciais. As imagens coletadas neste processo passam por pré-seleções e rotulagem, que serão apresentadas com detalhes no próximo capítulo.

Logo após a formação de uma base de dados parcial iniciou-se a construção dos extratores de características e da rede neural, sendo construída uma rede para cada tipo de característica e ao final construída uma rede neural única contemplando todas as características. Detalhes sobre este procedimento estão apresentados no Capítulo 5.

De maneira similar, o classificador de textos foi construído, treinado e testado com uma base parcial e posteriormente o treinamento foi repetido com a base de dados final. Detalhes sobre este procedimento estão descritos no Capítulo 6. A camada de decisão foi construída após a base de dados de treinamento e testes estar pronta. O propósito da camada de decisão é integrar os resultados gerados pelo classificador de imagens e o classificador de texto, decidindo a classe final de forma mais “inteligente”.

Com a base de dados final formada e rotulada, o classificador de imagens (rede neural) treinado, o classificador de texto (Naïve Bayes) treinado e a camada de decisão com suas regras definidas, pode-se passar para a segunda etapa, que consiste na avaliação final dos resultados e análise dos mesmos.

Na primeira fase dos trabalhos foram feitos a construção e treinamento dos classificadores, sendo cada parte desenvolvida separadamente. O treinamento foi realizado de forma independente e com a base de dados reduzida na maior parte do tempo. Ao final da primeira fase todos os algoritmos que envolviam algum tipo de treinamento foram novamente treinados e validados com a base de dados de treinamento final.

Ao contrário da primeira fase, a segunda fase trabalha com o conjunto todo integrado e não utiliza as amostras de imagens usadas durante qualquer etapa de treinamento. O esquema da segunda fase está apresentado na Figura 3.2. Com todos os módulos finalizados, cada imagem submetida ao experimento está vinculada a um determinado texto. Os classificadores de imagem e

texto processam a requisição e informam o resultado para a camada de decisão, que efetua alguns processamentos e submete a um conjunto de regras que finalmente informa o resultado final da classificação.

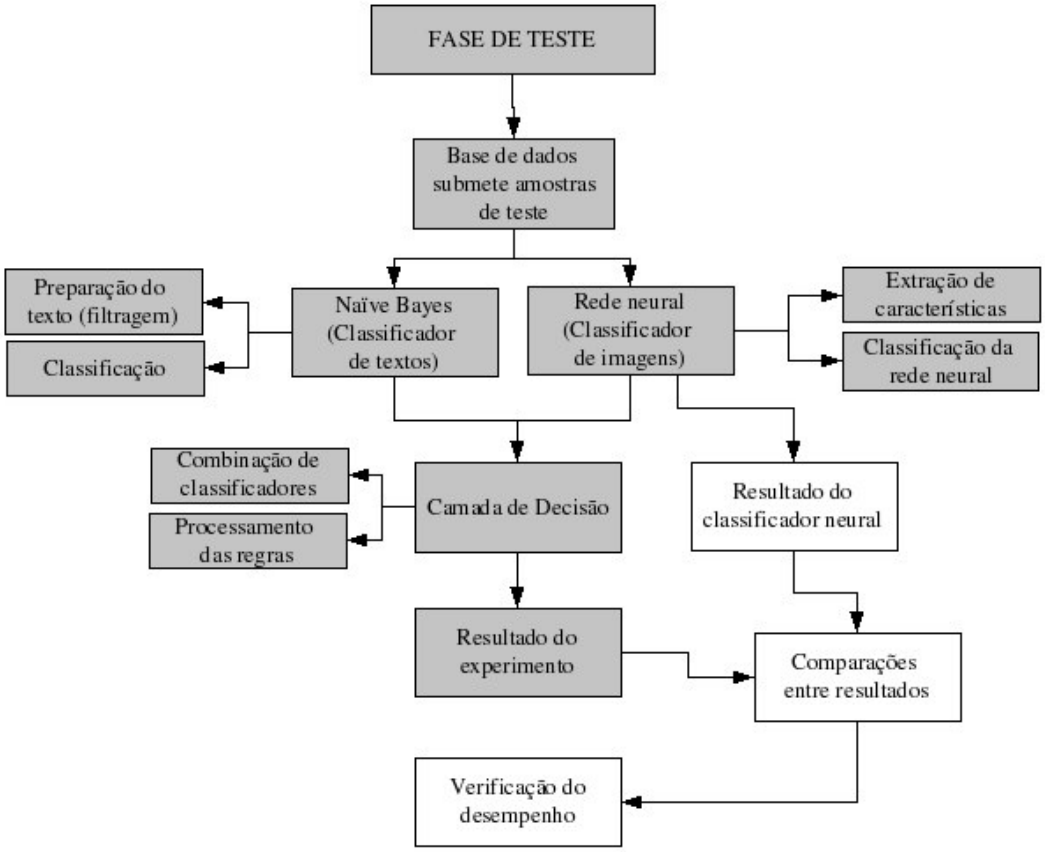


Figura 3.2 – Visão geral da fase final de testes

Para a verificação da eficiência da combinação proposta, o resultado da classificação de imagens realizado somente pelo classificador baseado em redes neurais é comparado ao resultado do experimento total, que combina os resultados do classificador baseado em redes neurais e o classificador baseado em informações contextuais. Um aumento considerável da taxa de classificação correta confirma a hipótese de que informações contextuais contribuem significativamente para a classificação de imagens. Os resultados estão apresentados em detalhes no Capítulo 9.

3.2. Definição das Classes

Este trabalho pretende comprovar o ganho da taxa de classificação correta de imagens através de uma combinação com as informações contextuais. Para esta avaliação é necessário o uso de um classificador de imagens que deve trabalhar de forma convencional para que a comparação seja válida. Para isso definimos algumas classes de imagens, onde o classificador indica a qual classe a amostra processada pertence. A escolha das classes a serem consideradas neste trabalho baseou-se em alguns princípios estabelecidos que caracterizem a natureza do problema abordado. São elas:

- As classes devem apontar algo concreto e visível.
- Podem existir variações de cores e tamanhos entre elementos de mesma classe. Mas não deve haver variações muito grandes, pois dificultaria a obtenção de exemplos de todos os tipos para o correto treinamento dos classificadores.
- Deve permitir a convivência de outros objetos de outras classes na mesma imagem, assim como permitir diferentes planos de fundo.
- Devem ser representativos, com características próprias bem definidas. Uma pessoa comum deve ser capaz de diferenciar os objetos das classes facilmente.
- Devem ser facilmente encontradas na Internet, acompanhado de textos para que ocorra a viabilidade na fase de construção da base de dados.
- A quantidade de classes deve ser compatível com o tempo previsto para conclusão deste trabalho.

Seguindo estas diretrizes foram escolhidas as seguintes classes:

- Automóveis
- Pessoas
- Animais domésticos
- Motos
- CDs/DVDs

As classes escolhidas possuem alguns níveis de ligação, como pessoas dentro de carros ou motos, assim como motos e carros são maquinas e animais domésticos também são observados ao lado de pessoas ou dentro de carros. CDs/DVDs podem conter qualquer combinação anterior fazendo parte de sua imagem final, mas normalmente possuem dimensões sempre iguais,

acompanhado de características idênticas, como algumas letras indicando o título do filme ou música. A escolha pelo total de cinco classes refere-se ao comprometimento do tempo gasto para a aquisição de imagens. Cada classe deve ter um número de amostras suficientes para permitir treinamento, validação e testes.

Outro ponto fundamental é que estas imagens pertencem a um domínio comum e são facilmente encontradas na Internet, viabilizando a construção da base de informações. Na Figura 3.3 pode-se ver um exemplo de cada uma das classes escolhidas.

Figura 3.3. – Exemplo das classes escolhidas: automóveis, motos, CDs/DVDs, pessoas e animais domésticos.

3.3. Base de Dados

As imagens para classificação deverão ser obtidas através da Internet com um conjunto de ferramentas próprias, desenvolvidas para esta finalidade. Tal requisito origina-se da necessidade de analisar a informação textual que também será utilizada para auxiliar na classificação da imagem. No entanto, o treinamento do classificador de imagens pode ser feito com imagens obtidas de diversas fontes, pois ao treiná-lo não será necessário nada mais além das amostras de imagens e etiquetas atribuindo uma das classes.

Outra grande vantagem do uso da Internet se dá pelo sistema de *hyperlinks*, que será utilizado para determinar a próxima página a ser consultada, bastando analisar os comandos HTML e armazenando os *hyperlinks* para as páginas. Este processo é recursivo, permitindo que uma pequena lista de endereços de páginas *web* iniciais seja suficiente para que o sistema adquira uma

grande quantidade de informações e imagens. Trabalhos similares a este podem ser encontrados em [OLI02] e [DON02].

Para a formação da base de dados foi utilizado *hyperlinks* referenciando repositórios conhecidos de páginas de determinado assunto, como *Yahoo!*[1] ou *Altavista*[2], porém não foi limitado a estes repositórios para propiciar maior diversidade nas amostras.

Entretanto, as imagens obtidas devem ser classificadas manualmente, após uma pré-seleção executada pelo sistema. Para o treinamento serão utilizadas somente imagens que contenham apenas uma das classes indicadas. Imagens vindas da Internet podem ter muitas características distintas como formatos, conteúdo e finalidades [GON00] [OLI02]. Grande parte das imagens não está apta a ser classificada, por não apresentar pré-requisitos mínimos, como:

- Ser imagem do tipo foto, e não gráfico.
- Não ser *banner* promocionais.
- Não ser imagem animada.
- Não ser imagem promocional.

Note na Figura 3.4 um exemplo de imagens não-válidas desprezadas pelo sistema de busca e recuperação. Esse é um *banner* que normalmente contém informação promocional e não interessa ao nosso sistema. Este tipo de imagem é facilmente eliminado da base de imagens com a simples verificação da relação entre largura e altura (detalhado no próximo capítulo).

Figura 3.4 – Exemplos de imagens que devem ser descartadas

[1] http://www.yahoo.com
[2] http://www.altavista.com

As imagens alvo de classificação devem ser imagens do tipo foto, com dimensões proporcionais a uma foto, em cores ou tons de cinza que não contenham letreiros significativos na sua face (caracterizando propaganda) e não sejam animadas. Exemplos de imagens válidas podem ser vistos na Figura 3.5.

Uma suposição importante nesta etapa de formação da base de dados é a suposição de que, dado um *hyperlink*, as imagens e informações presentes na página HTML relativas a este *hyperlink* estarão relacionadas com uma das classes pré-estabelecidas, ou seja, um *hyperlink* para uma página de uma fabricante de automóveis conterá informações sobre automóveis.

Quando o sistema extrair as informações da Internet a partir dos *hyperlinks* informados manualmente, serão trazidos diversos tipos de imagens e textos, sendo que muitos deles poderão não conter informações relevantes, sendo necessário, assim, efetuar uma pré-seleção (detalhado no próximo capítulo) antes de sua inclusão na base de dados. Caso não seja efetuado este passo, corre-se o risco de encher o espaço de armazenamento com informações inúteis e, devido à quantidade, comprometer a qualidade do sistema. Exemplos destes casos são *frames* e menus que compõe os sites HTML, *banners* e outras formas de propaganda que aparecem no formato de imagens. Estes tipos de imagens, porém, não são interessantes e podem atrapalhar todo o processo.

Figura 3.5 – Exemplos de imagens válidas

3.4. Resumo

Neste capítulo foi apresentada uma visão geral do sistema nas suas fases de construção/treinamento e testes. Foram definidas as classes que serão utilizadas e que tipos de imagens são consideradas válidas. Também foi definido como será a aquisição e a utilização de

informações (imagens e textos) ao longo deste trabalho. Com todas estas definições podemos agora apresentar com mais detalhes as partes principais deste trabalho e ao final, verificar o conjunto de resultados e as taxas de classificação correta alcançadas para a validação da hipótese.

Capítulo 4

Base de Dados

De extrema importância para sistemas que utilizam aprendizado, a base de dados pode definir a qualidade das respostas produzidas por estes sistemas. Uma base de dados significativa deve conter exemplos coletados em situações reais às que existem em ambiente de produção do sistema. Esta preocupação foi levada em conta na hora da escolha da base de dados para este projeto, e definiu-se criar procedimentos próprios para a obtenção e organização de dados.

Normalmente uma base de dados comercial contém amostras padronizadas, com mesmo tamanho, mesma resolução e mesmo enquadramento. Neste trabalho o objetivo é chegar muito próximo a um ambiente real, que não contém nenhum tipo de padrão e numa mesma classe é possível encontrar amostras com diferentes tamanhos e diferentes aspectos. Como por exemplo, a classe "pessoas" considera imagens que têm apenas o rosto, o corpo inteiro, multidões ou mesmo pessoas fazendo parte do cenário. Além das imagens, o texto também precisa conter ruídos normais deste tipo de ambiente, como propagandas, *hyperlinks* para outras páginas e vários textos numa mesma página referindo-se a assuntos diferentes.

4.1. Informações Necessárias para o Projeto

Este trabalho constitui um mecanismo para reconhecimento de imagens, adotando um classificador baseado em redes neurais, um classificador estatístico para reconhecimento de textos e um mecanismo integrando os resultados destes classificadores. Desta forma, a base de dados deve conter imagens e textos devidamente rotulados nas classes definidas previamente (Capítulo 3). Além disto, é preciso manter também uma ligação entre o texto e a imagem, constituindo-se assim, o que ao longo deste trabalho é chamado de "informações vinculadas".

A proposta deste trabalho refere-se à Internet, onde os textos e imagens não são preparados e contém muitos ruídos e incertezas. Portanto, a melhor base de dados seriam imagens e textos coletados diretamente da Internet, que é o ambiente de produção (testes) deste trabalho.

Apenas para fases de treinamento da rede neural e do classificador estatístico (texto) poderiam ser utilizados imagens e textos separados, obtidos de bases de dados prontas ou comerciais, desde que possuam as características de ruídos indicadas. Para o restante do trabalho estas informações precisam estar relacionadas.

4.2. Origem da Base de Dados

Sistemas classificadores de imagens são amplamente estudados pela comunidade científica, desta forma encontram-se diversas bases de dados comerciais como as bases de dados NIST [NIS04], por exemplo. Além destas bases de dados, existem diversas bases disponibilizadas pela comunidade científica, normalmente utilizadas previamente em algum trabalho científico. O pesquisador pode ter construído a base com imagens geradas por ele mesmo, ou estas imagens podem ter vindo de outras fontes, sendo disponibilizadas por ele para outras pesquisas científicas.

No caso deste trabalho, não conseguimos encontrar bases com as características requeridas. Assim, a grande maioria das informações utilizadas na base de dados foi gerada a partir do mecanismo construído para esta finalidade e será detalhado posteriormente neste capítulo. Uma pequena parte de amostras de imagens vieram de navegação normal pela Internet e de bases oriundas de experimentos. Neste caso as imagens não coletadas pelo mecanismo automatizado foram utilizadas somente durante a fase de treinamento e validação individual das redes neurais, pois nas outras fases foram necessárias informações sobre o relacionamento do texto com as imagens.

O processo de formação desta base de dados levou um tempo considerável para a captura e rotulação de imagens e textos. Com todas essas formas de obter as informações, a base de dados ficou constituída como exemplificado na Figura 4.1.

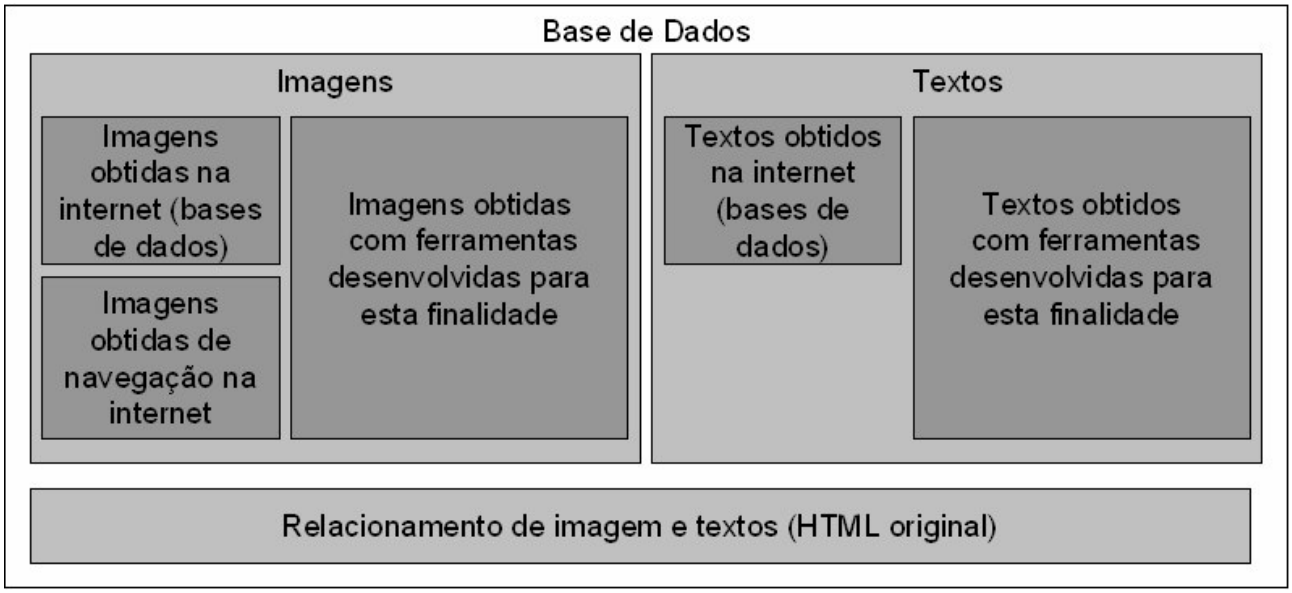


Figura 4.1 - Base de dados construída e utilizada neste trabalho

Além destes três componentes (imagens, textos e relacionamentos entre eles) foram também armazenadas informações de apoio como a árvore de *hyperlinks*, as imagens e textos que foram descartados, assim como o motivo do descarte. As páginas HTML também foram armazenadas e tanto estas páginas quanto o texto e as imagens ficaram armazenadas em disco, ficando sua referência (caminho) na base de dados assim como suas características de recusa ou aprovação, rotulagem e dados de apoio. Nem todas as imagens rotuladas possuem o texto correspondente e os textos podem ou não possuir as imagens de origem. Este assunto será abordado adiante neste capítulo.

4.3. Sistemática de Aquisição, Seleção e Rotulação de Imagens

Para adquirir as informações provindas da internet desenvolveu-se um conjunto de ferramentas responsáveis por coletar imagens, textos e outras informações, como *hyperlinks* para outras páginas e então visitá-las, tornando-se um processo cíclico. Todas as informações coletadas continham muitos exemplos que não podiam ser aproveitados e, portanto, foram criadas outras ferramentas para auxiliar na tarefa de seleção. Na seqüência necessitou-se analisar todos os exemplos restantes um a um rotulando-os ou descartando-os. O processo de coleta de informações para a base de dados está esquematizado na Figura 4.2.

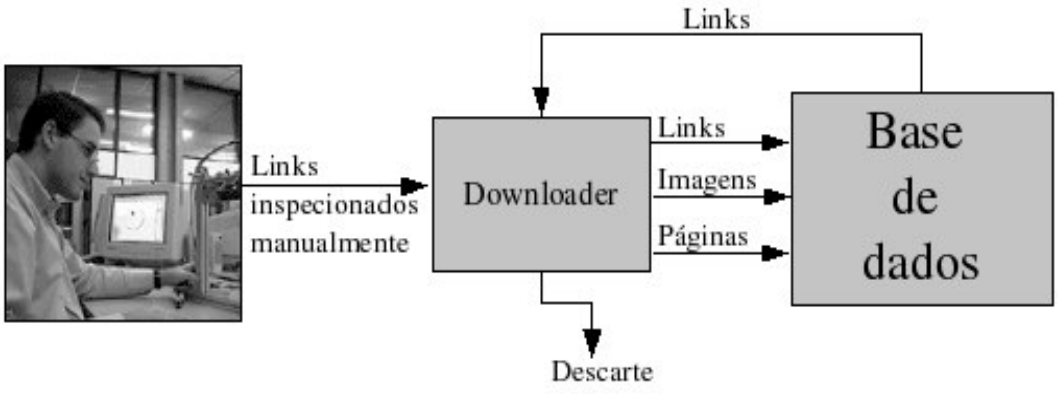


Figura 4.2 - Processo de coleta das informações da base de dados

De forma geral a ferramenta recebe uma pequena lista inicial de *hyperlinks*, escolhidos manualmente, que apontam para páginas que contém imagens das classes pré-definidas. Não procuramos diretamente por textos das classes definidas, o interesse é por imagens, pois o foco do trabalho é melhorar o desempenho de classificadores de imagens dentro das características definidas no Capítulo 3.

Através de uma ferramenta intitulada "*downloader",* que foi construído para esta finalidade, os *hyperlinks* são visitados e o texto em formato HTML é tratado e armazenado. Na seqüência este texto tem seus *hyperlinks* extraídos e armazenados para posterior consulta. Os endereços para imagens presentes nas páginas também são armazenados na base de dados para posterior captura. Finalmente o texto HTML tem suas *tags* removidas e seu conteúdo também é armazenado. Este texto livre de *tags* é que será usado no classificador textual.

Um segundo processo que pode estar rodando em paralelo recupera os endereços para as imagens baixadas durante a captura de páginas e baixa as imagens para a base de dados. Nenhuma classificação é feita neste momento, todas as imagens que puderam ser recuperadas são armazenadas. Este sistema somente trabalha com imagens no formato JPG ou GIF. Existem situações onde os *hyperlinks* apontam para outros objetos como arquivos PDF, imagens de formatos diferentes, arquivos de vídeos ou arquivos binários. Nestas situações o sistema aborta o processo de recuperação e invalida o *hyperlink* assim que identificado o tipo de arquivo.

Após este processo, a grande maioria das imagens baixadas é imprópria, pois são desenhos usados para decorar a página, ícones e outros acessórios visuais normalmente encontrados em páginas HTML. Para facilitar o processo de rotulagem das imagens foi adotado um processo de seleção, como ilustrado na Figura 4.3.

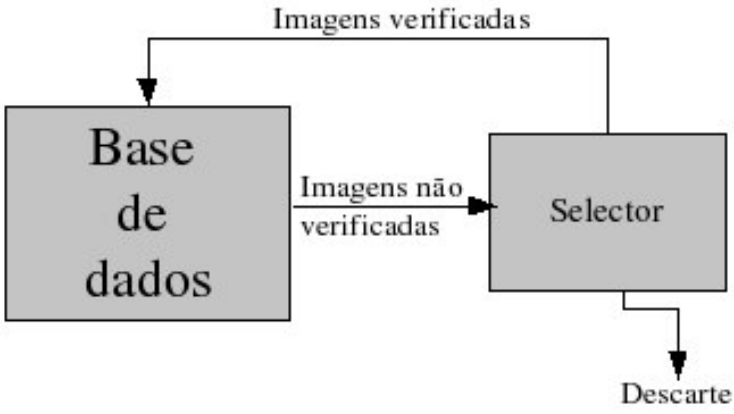


Figura 4.3 - Processo de seleção e rotulagem das imagens

Neste processo, que roda em um momento separado da coleta das informações, as imagens são selecionadas ou descartadas. O primeiro passo consiste em fazer uma seleção automática baseada em características simples, que é detalhado adiante neste capítulo. Esta fase é realizada pelo programa denominado "*selector*" e procura identificar imagens válidas, baseando-se em características como tamanho mínimo, tamanho máximo, proporção e a identificação de uma imagem do tipo fotografia ou imagem do tipo gráfico. Este processo é executado sem intervenção humana, analisando as imagens baixadas pelo passo anterior. Neste processo são analisadas somente as imagens. Os textos não tiveram nenhum analisador semelhante, sendo encaminhados diretamente para a próxima etapa.

O último passo para o tratamento das imagens é a rotulação. Para este processo foi criada uma ferramenta denominada "*rotulator*" e seu funcionamento é apresentado na Figura 4.4.

Este é um processo com interação humana, pois a ferramenta apenas facilita no sentido de organizar esta tarefa. As imagens marcadas como "não rotuladas" são ordenadas e apresentadas uma a uma para o observador, que identifica o rótulo da imagem. As imagens podem ser classificadas para qualquer uma das cinco classes escolhidas ou ser eliminada na base de dados.

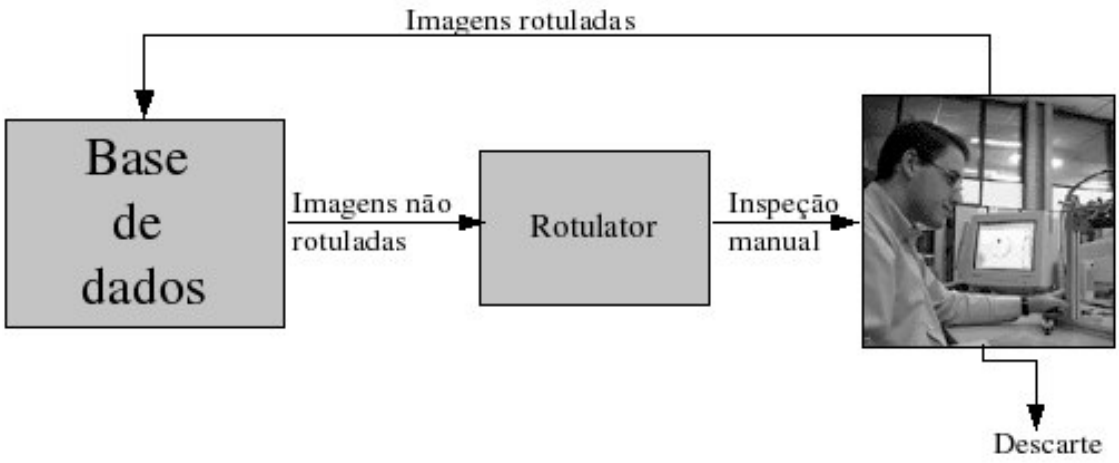


Figura 4.4 – Rotulação de imagens

4.4. Processo de Captura das Informações na Internet

A ferramenta desenvolvida para coletar informações na internet denominada "*downloader*" funciona em conjunto com uma base de dados que contém endereços válidos para páginas Internet. Inicialmente, estes endereços são inseridos manualmente, com uma prévia seleção manual buscando apontar para páginas que contenham imagens relacionados às classes pré-definidas. Após este passo a ferramenta inicia seu trabalho gerando requisições HTTP e coletando o conteúdo HTML de cada endereço cadastrado na base de dados. O conteúdo HTML é analisado e são coletadas todas as referências para imagens. Em um processo separado, as referências de imagens são usadas para coletar a imagem na Internet e gravadas em disco. O texto restante é armazenado de forma original, e então, retiradas todas as *tags* ficando apenas o texto puro da página. Este texto é armazenado em disco para uso posterior pelo classificador textual. Toda informação do endereço de origem, imagem e texto vinculado assim como a rotulação que ocorrerá em um passo posterior é armazenada para preservar os vínculos, pois este será necessário durante a fase de combinação de classificadores. A cada página analisada são identificados todos os *hyperlinks* existentes e estes são realimentados na base de dados, não se admitindo *hyperlinks* repetidos. Esta metodologia gera um agrupamento das informações, conforme mostrado na Figura 4.5.

Este agrupamento se reflete em todos os elementos da base de dados, como as imagens e os textos. Ao navegar pela base de dados nota-se que quanto maior for o identificador do elemento (texto ou imagem), mais amostras de uma mesma classe estão agrupadas, conforme mostrou a Figura 4.5. No início eram apenas alguns *hyperlinks* para cada classe. Ao analisar a página apontada

por cada um destes *hyperlinks*, houve uma realimentação na base de dados de todos os *hyperlinks* disponíveis nesta página. Este é um processo recursivo que se encerra apenas quando se atinge o número necessário de amostras a serem coletadas. Ao todo foram encontrados 119.902 *hyperlinks* para imagens, onde muitas destas não puderam ser recuperadas por motivos de dificuldades na conexão ou formato inadequado. A partir destes *hyperlinks* foram efetivamente recuperadas 84.289 imagens, que foram submetidas à etapa seguinte.

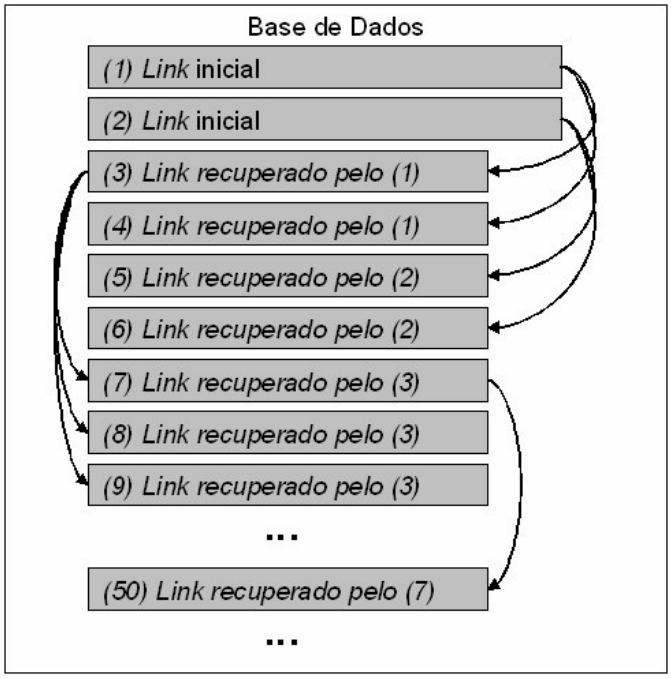


Figura 4.5 - Representação do agrupamento na base de dados

Neste processo de captura das informações, os *hyperlinks* alimentados manualmente são previamente verificados e classificados. Estas alimentações de *hyperlinks* manuais ocorreram no início da formação da base de dados e em outros momentos, definidos de acordo com os resultados da rotulação. Quando as amostras coletadas não indicavam mais o conteúdo procurado, os *hyperlinks* derivados eram cancelados. Desta forma existiu certa limitação na profundidade da busca, porém não era definida anteriormente e sim determinada de forma dinâmica de acordo com os resultados obtidos. Os procedimentos de captura, seleção e rotulagem ocorreram de forma paralela.

4.5. Seleção de Imagens

Após as imagens serem armazenadas em disco é necessário separar as imagens que não

servem ao sistema. Estas imagens provêm de gráficos que enfeitam as páginas como menus, bordas, ícones, etc. Uma outra grande parte são propagandas, como *banners* que também devem ser eliminadas. É importante lembrar que são muitas imagens. Ao todo foram recuperadas 84.289 imagens, sendo que apenas 11.810 realmente passaram para a fase de rotulação, ou seja, apenas 14% das imagens vindas por este processo são realmente imagens de interesse.

São consideradas imagens válidas apenas fotos coloridas ou em tons de cinza que obedeçam a algumas regras gerais como: tamanho, proporção, e que não sejam gráficos. A primeira regra é muito simples e eliminou grande parte das imagens inválidas. Trata-se de uma verificação do tamanho da imagem. Caso a imagem não tenha largura ou altura dentro dos limites definidos como mínimos ou máximos esta é descartada. Para este programa a primeira regra verifica o tamanho mínimo e a segunda regra verifica o tamanho máximo. A terceira regra refere-se à proporção e define um percentual máximo de proporção entre altura e largura a fim de identificar possíveis *banners*. O calculo do índice é feito de acordo com a fórmula abaixo:

$$I = \frac{L_{me}}{L_{ma}} * 100 \tag{4.1}$$

onde L_{ma} indica a dimensão maior de uma imagem e L_{me} indica a dimensão menor de uma imagem e I indica o índice de proporcionalidade. Em nossos experimentos definimos o valor experimental de I = 50% como limiar. A Figura 4.6 apresenta um exemplo de uma imagem que passa no teste da proporção (1) e outro que não passa neste teste (2).

Imagem 1 – 160 x 262 pixels

Imagem 2 – 468 x 60 pixels

Figura 4.6 - Exemplos de imagens submetidas ao teste de proporção

No lado esquerdo da Figura 4.6 temos uma imagem com altura de 262 *pixels* e largura de 160 pixels. Aplicando a fórmula apresentada tem-se o índice de proporção de 61,08%, portanto superior ao índice de 50% que é o mínimo exigido. Nesta mesma figura, para a imagem da direita, o índice de proporção é 12,82%, sendo então corretamente eliminada. O objetivo deste teste é justamente eliminar imagens do tipo *banner* e outras imagens que apresentam conteúdos diferentes das imagens de interesse.

A quarta e última regra analisada por esta ferramenta procura identificar imagens que representam gráficos de imagens que representam fotografias. Para este processo a imagem passa por alguns tratamentos e tem seu histograma extraído para a análise. A Figura 4.7 apresenta exemplos destes dois tipos de imagens e o respectivo histograma.

Figura 4.7 – Exemplos de imagens representando gráfico e fotos respectivamente

Nos exemplos mostrados na Figura 4.7 pode-se notar que gráficos têm características distintas como poucas cores, formas bem definidas, uso intenso das cores, etc. A exemplo do trabalho de Oliveira [OLI02], que fez experimentos para separar gráficos de fotos, neste trabalho é feito algo parecido analisando simplesmente a quantidade de cores que uma imagem possui. Um limiar é definido e se a imagem não atingir este limiar é então descartada.

Como as cores em um formato RGB apresentam-se em três diferentes dimensões, é necessário efetuar a contagem de cores após um procedimento de transformação da imagem para

tons de cinza, pois poderia haver enganos se fosse usado apenas um dos canais de cor ou validar todos os canais. Este problema é ilustrado na Figura 4.8.

Note que a imagem exemplo da Figura 4.8 contém diferentes histogramas para cada um dos canais de cor. A contagem do nível zero para o histograma R desta imagem é 9.386. Já para o canal G e B são respectivamente 11.037 e 16.676. Desta forma não se pode usar apenas um dos canais de cor e sim extrair a média que é feita transformando a imagem de RGB para tons de cinza. A fórmula usada para esta conversão pode ser vista abaixo:

$$P = \frac{R + G + B}{3} \tag{4.2}$$

Esta formula é aplicada na imagem e *P* representa o ponto da imagem transformada após o cálculo aplicado com os valores dos canais R, G e B, transformando-a em uma imagem com tons de cinza, que variam numa escala 0 de 255 níveis. O menor nível (0) é o preto total e o maior nível (255) é o branco total. A imagem na Figura 4.9 apresenta exemplos de imagens que representam gráficos e fotografias já convertidas para níveis de cinza com seus respectivos histogramas.

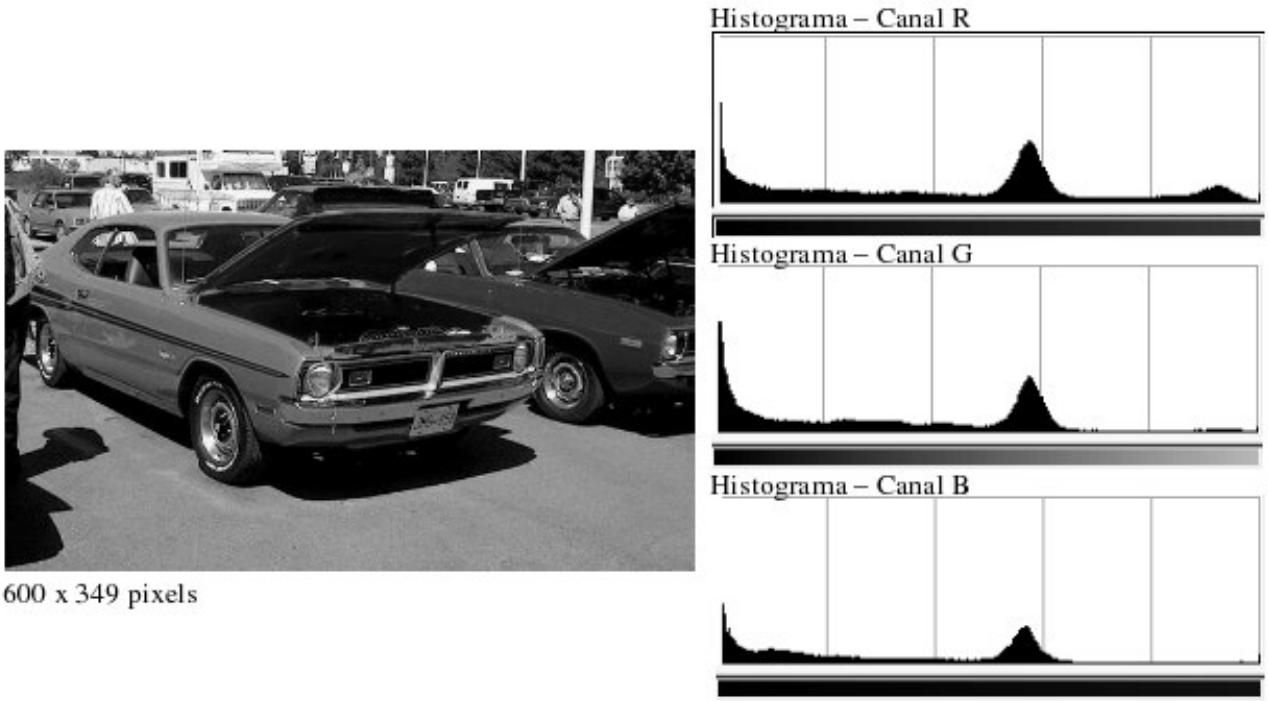


Figura 4.8 – Exemplo de histogramas de cada canal de cor para o padrão RGB

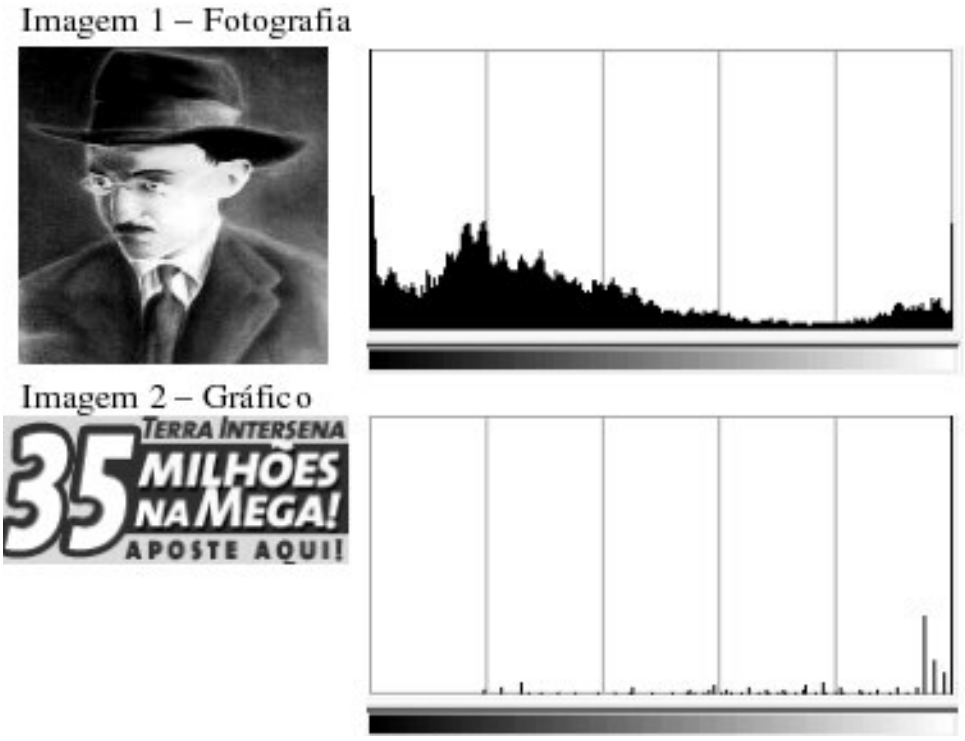


Figura 4.9 – Exemplo de imagens convertidas e seus histogramas

A partir da Figura 4.9 é possível notar a clara diferença no histograma entre uma fotografia e um gráfico. Numa fotografia praticamente todos os níveis de cinza são usados, tendo diferenças entre sua distribuição. No caso de gráficos somente alguns níveis de cinza são usados. Muitas formas de separar fotografias de gráficos poderiam ser usadas com base nesta diferença de histograma, para efeitos de simplicidade, neste programa é feito uma contagem simples de níveis de cinza. Para ser considerada fotografia, o histograma precisa conter pelo menos 150 níveis de cinza, de acordo com a fórmula abaixo:

$$\begin{aligned} &Nnu = \text{Número de níveis usados} \\ &Nnu \geq 150 \end{aligned} \tag{4.3}$$

Caso a imagem apresente mais de 150 níveis de cinza então ela é considerada valida, caso contrário ela é descartada. Todas as 84.289 imagens recuperadas na primeira fase foram submetidas a estas quatro regras e o resultado geral desta fase é apresentado na Tabela 4.1.

Estas quatro regras apresentadas foram derivadas de trabalhos similares que precisaram de pré-seleção de imagens [OLI02]. Neste trabalho foram determinados os valores mínimos e máximos para tamanho das imagens, o cálculo e índice de proporcionalidade e a seleção entre gráfico e fotografia que também teve algumas variações, limitando-se a contagem de cores utilizadas.

Tabela 4.1. - Resultado analítico da fase de seleção

	Quantidade	***Regras***
0	11.810	Aceitas – não eliminadas
1	693	Eliminadas somente por tamanho máximo.
2	23.706	Eliminadas somente por tamanho mínimo.
3	0	Eliminadas somente por tamanho mínimo e tamanho máximo.
4	204	Eliminadas somente por proporção.
5	1	Eliminadas somente por proporção e tamanho máximo.
6	2.268	Eliminadas somente por proporção e tamanho mínimo.
7	3	Eliminadas somente por proporção, tamanho máximo e tamanho mínimo.
8	2.434	Eliminadas somente por histograma.
9	6	Eliminadas somente por histograma e tamanho máximo.
10	27.700	Eliminadas somente por histograma e tamanho mínimo.
11	0	Eliminadas somente por histograma, tamanho máximo e tamanho mínimo.
12	146	Eliminadas somente por histograma e proporção
13	5	Eliminadas somente por histograma, proporção e tamanho máximo.
14	17.540	Eliminadas somente por histograma, proporção e tamanho mínimo.
15	6	Eliminadas por histograma, proporção tamanho máximo e tamanho mínimo (todos).

A importância deste processo está principalmente no fato de que seria praticamente inviável um humano em um tempo limitado analisar 84.289 imagens, ainda mais que apenas 14% seriam validas. Com esta fase o processo de formação da base de dados foi possível, e o trabalho manual foi drasticamente reduzido.

4.6. Rotulagem de Imagens

Nesta fase foram analisadas 11.810 imagens provenientes da seleção detalhada no tópico anterior. Esta fase teve um papel fundamental para a qualidade da base de dados, pois com a

inspeção manual a probabilidade de uma imagem ser rotulada erroneamente é baixa.

As imagens foram rotuladas com auxílio de uma ferramenta intitulada "*rotulator*" que basicamente apresentava a imagem ao humano e este selecionava o número correspondente à classe da imagem ou simplesmente a descartava, quando esta não pertencia a nenhuma das classes. A Figura 4.10 apresenta a interface da ferramenta de rotulagem.

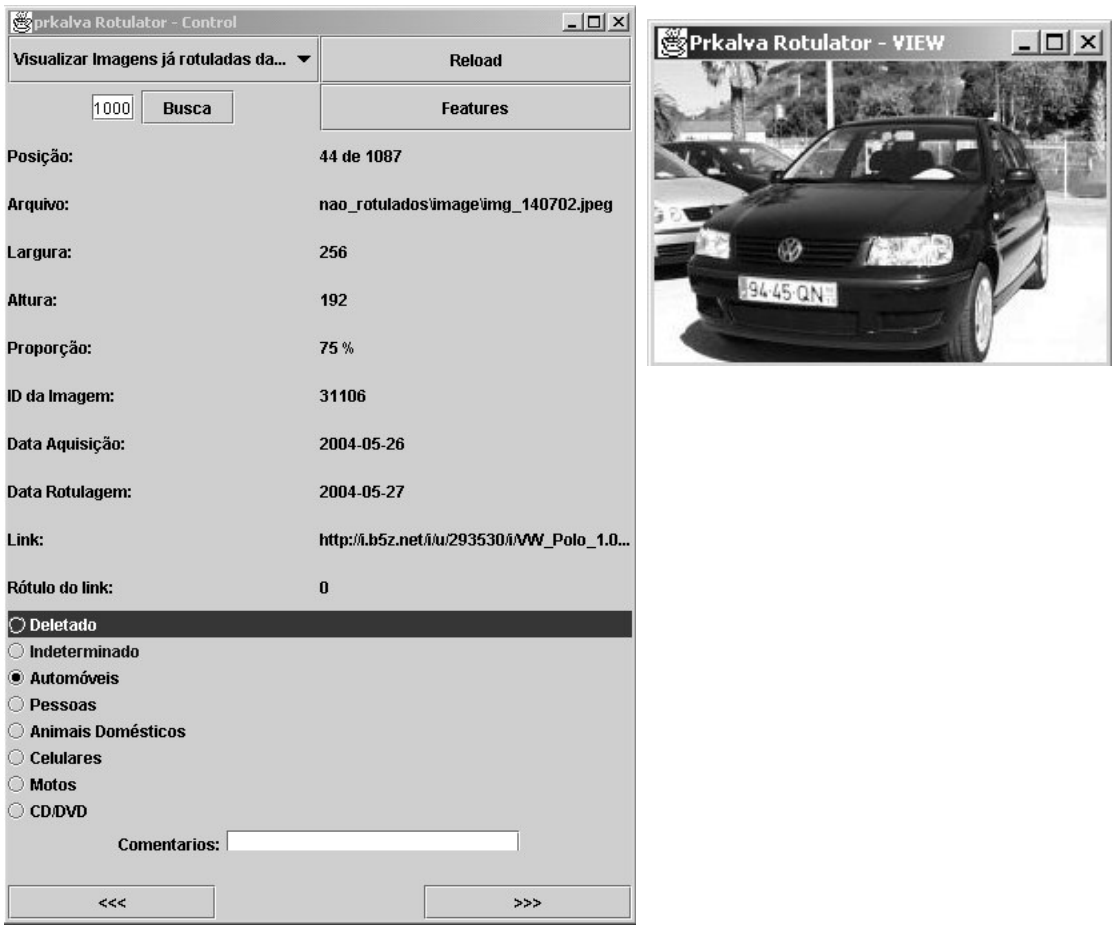


Figura 4.10 – Interface da ferramenta de rotulação de imagens

Das 11.810 imagens que foram inspecionadas visualmente por este processo apenas 5.405 foram consideradas para compor a base de dados, representando 45,8% do total das imagens submetidas a este processo. A Tabela 4.2 apresenta a distribuição das imagens escolhidas em relação a suas classes.

Tabela 4.2. - Resultado da rotulagem de imagens

Classe	***Quantidade***
Automóveis	1.087
Pessoas	880
Animais domésticos	1.166
Motos	1.425
CD/DVD	847
Descartadas	6.405

É importante também descrever os critérios que foram levados em conta no momento da rotulagem. Para ficar dentro dos requisitos exigidos neste experimento, as imagens tem que apresentar variações diversas. Um exemplo disto pode ser visto na Figura 4.11.

Figura 4.11 – Exemplos rotulados para a classe pessoa

Estes exemplos demonstram que existe muita diversidade na base de dados, pois para a classe pessoa foram aceitas imagens em diversos tamanhos (respeitando os limites mínimos e máximos discutidos no tópico anterior) em diferentes ângulos e conteúdo. Note que foram consideradas imagens de rosto, de corpo inteiro, de mais pessoas compondo a cena, imagens de multidões, etc. Mais exemplos, incluindo exemplos de outras classes são apresentados em anexo.

Devido a este tipo de consideração, é de se esperar que o classificador de imagens não chegue a índices muito elevados de classificação, pois as amostras apresentam muitas diferenças

entre eles. Mas esta diretriz faz com que os resultados apresentados sejam, praticamente resultados em um âmbito real, de produção. Provavelmente se fossem utilizadas amostras padronizadas todos os classificadores iriam apresentar taxas de classificação corretas bem melhores, porém não seria um resultado próximo do ambiente real.

4.7. Processamento dos Textos

Os textos tiveram um tratamento bem mais simplificado que os apresentados para as imagens. Um dos motivos é que o trabalho com textos sob o ponto de vista computacional, de forma geral, é mais simples. O outro motivo é que este trabalho refere-se à classificação de imagens e não necessariamente classificação de textos, embora seja utilizado um classificador de texto, para se chegar a melhores resultados em classificação de imagens. É importante salientar que tanto as imagens como os textos foram rotulados de forma independente.

Para cada página HTML baixada durante a procura de imagens, eram retiradas as *tags* que compõe a página HTML sobrando apenas o texto. Porém, este texto contém inúmeras frases auxiliares do texto principal, como rodapés, propagandas e até mesmo outros textos, pois uma página pode se referir a diversos temas. Além disto, várias línguas foram consideradas. Textos em inglês, português e espanhol que puderam ser lidos, foram rotulados.

Para este trabalho utilizamos a palavra “TEXTO” para nos referirmos ao conteúdo da página HTML que teve suas *tags* retiradas. Referimo-nos a “PÁGINAS” o conteúdo completo (com as *tags*) das páginas baixadas.

4.8. Rotulação dos Textos

Os textos foram rotulados diretamente com uma ferramenta parecida com a ferramenta usada para rotular imagens, mas numa versão adaptada para o texto. Um inspetor humano analisava texto a texto indicando a qual classe ele pertencia. Este processo é um pouco mais trabalhoso que a rotulação das imagens, pois necessita que o responsável pela rotulagem leia o texto, procurando identificar a idéia central. Caso a idéia central fosse entendida e caso esta se referisse de alguma forma a uma das classes pré-definidas, este texto era rotulado ou então descartado. Caso a idéia central do texto (ou predominante) fosse identificada como mais de uma classe, este era descartado. Textos com poucas palavras também foram descartados, porém isto não foi feito automaticamente e sim manualmente. Um exemplo de um texto capturado já livre de *tags* é apresentado na Figura 4.12.

Note que este texto apresenta diversas palavras e frases que não se referem ao conteúdo, mas somente a outras características da página. Assim como as imagens que possuem muita diversidade, estas palavras que não tem a ver com o assunto principal do texto formam uma espécie de ruído nas amostras, tornando-as também aptas a demonstrarem um ambiente real e não padronizado.

```
carros de rua | esse é o nosso mundo| INICIAL| DESTAQUE DO MÊS| INFO E TÉCNICA|
GALERIA| LOJA VIRTUAL| EVENTOS| DOWNLOADS| FORUM| LINKS| CDR NA MÍDIA| ANUNCIE|
CONTATO| MATERIAIS| LOJAS E OFICINAS| FAÇA VOCÊ MESMO| GALERIA| MEMBROS| MINHA
GALERIA| 1° Etapa Camp Paranaense de Arrancada-2003--------------------------
páginas 1 2 3 4 5 6 INTRODUÇÃO Como de costume, a primeira etapa do campeonato
paranaense de arrancada, realizado em Curitiba, foi um enorme sucesso. O público
esteve presente durante os três dias do evento (02 a 04 de maio), com maior
concentração no domingo. O espetáculo contou com mais de 130 pilotos inscritos.
Grande parte veio de fora do estado, demonstrando o interesse nacional pelo
evento. IMAGEM DO EVENTO NO DOMINGO A organização do evento, realizada pela
Força Livre Motorsport, é sem dúvida nenhuma a principal culpada por esse
sucesso. Boxes limpos e organizados, restaurantes dentro dos boxes, banheiros,
seguranças, fiscais de pista, todo o suporte médico e quaisquer outros que
venham a ser necessário, estão lá disponíveis para todos. Os carros e pilotos
estão cada vez mais profissionais, não excluindo o espaço para amadores
obviamente. A pista é um espetáculo a parte, inteira de concreto, com uma reta
de 980 metros é dividida em três partes; alinhamento e aquecimento de pneus, a o
```

Figura 4.12 – Exemplos de um texto já tratado (livre de *tags*)

Naturalmente apareceram muitos *hyperlinks* para textos, pois cada página HTML pode conter inúmeros *hyperlinks* para outras páginas. Ao total houveram 407.758 *hyperlinks* para páginas HTML e destes *hyperlinks* apenas 28.100 foram recuperados (páginas HTML recuperadas). É considerado um *hyperlink* para texto quando está presente em uma página HTML, dentro da *tag* A no parâmetro HREF, como no exemplo a seguir:

<a href="**http://www.pucpr.br/cursos/informatica/apresentação.html**" />

ou então:

<a href="**apresentação.html**" />

Todos os *hyperlinks* são traduzidos para a forma completa (protocolo + servidor/porta + diretórios, se tiver + nome da página) quando não estiver. No segundo caso do exemplo acima são colocados os elementos faltantes antes da armazenagem. Estes *hyperlinks* são armazenados na base de dados e não permitem duplicatas.

Os 28.100 textos não puderam ser classificados integralmente, pois isto seria inviável por questões de tempo e recurso. Assim, somente foram rotulados 5.169 textos, sendo distribuídos de acordo com a Tabela 4.3.

Tabela 4.3 – Distribuição dos Textos Rotulados

Classe	***Quantidade***
Automóveis	1.029
Pessoas	1.010
Animais domésticos	1.069
Motos	1.004
CD/DVD	1.057

Para que não fosse necessário rotular os 28.100 textos, o programa de rotulagem permitia saltos aleatórios entre as amostras de texto não rotuladas, permitindo que o número de amostras rotuladas de cada classe ficasse distribuído de forma semelhante em relação à quantidade e proveniente de qualquer posição entre os 28.100 textos coletados.

4.9. Conjunto Vinculado

Tanto as imagens como os textos foram rotulados de forma independente. Durante o treinamento dos classificadores também existe esta independência, pois cada classificador usa somente suas amostras sem verificar qualquer forma de vínculo entre imagens e textos. Porém, conforme descrito no Capítulo 3, após a fase do uso individual dos classificadores entra em ação a camada de decisão, que utiliza informações dos dois classificadores a fim de determinar padrões de respostas destes classificadores que levam a uma taxa de classificação de imagens melhor quando aplicado. Neste momento precisa-se ter um conjunto especial de informações que possa ser usado durante a fase de treinamento desta camada e um conjunto de informações que possam ser submetidos a todo o processo a fim de validar todo o experimento.

Este conjunto de informações é chamado de “conjunto vinculado”. Este nome dá-se ao relacionamento entre os textos com as imagens. Para efetuar um teste é necessário que uma imagem (apenas uma, mesmo que a página tenha várias imagens, pois o que se quer é classificar a imagem) e o texto correspondente seja submetido a um processo de combinação e este apresente o resultado final da classificação. A melhora é comprovada ao comparar a classificação final da combinação

com a classificação processada apenas no classificador de imagens baseado em redes neurais. Portanto, vínculo é a ligação que está relacionada a responder a seguinte pergunta: “Qual é o texto que existia na página de onde esta imagem foi encontrada?”.

Na base de dados deste trabalho existem diversos casos onde textos que foram rotulados não tiveram nenhuma imagem válida correspondente, textos que contém muitas imagens e de diferentes classes e de imagens que não tem seu texto rotulado. Devido a estas diferenças, as amostras “desvinculadas” foram usadas em procedimentos preliminares como treinamento e validação dos classificadores individuais, reservando-se o conjunto vinculado para treinamento e testes da camada de decisão.

Para o treinamento da camada de decisão foi formado um conjunto vinculado que usou amostras de imagens e textos que já tiveram sido utilizados de forma individual nos passos anteriores (para treinamento e validação da rede neural, por exemplo). Porém, como a camada de decisão trabalha com as saídas dos classificadores individuais, não existe o problema de influência, pois a camada de decisão é treinada para encontrar padrões de respostas, conforme será detalhado mais adiante. Para o conjunto de testes que fornece o resultado final deste trabalho, não existe nenhuma amostra de imagem que foi usado em qualquer fase do treinamento. A composição deste conjunto vinculado está descrita na Tabela 4.4.

Tabela 4.4 – Formação dos conjuntos vinculados

Conjunto	***Quantidade de Amostras***
Treinamento	712
Testes	821

4.10. Separação dos Conjuntos e Formação da Base de Dados

Todas as amostras desta base de dados foram distribuídas em diversos conjuntos de acordo com suas necessidades de treinamento, validação e testes, tomando-se cuidado para que nenhuma amostra pertencente ao conjunto final de testes fosse utilizada durante qualquer treinamento ou validação.

Para o classificador de imagens foram necessários três conjuntos de imagens para formar os conjuntos de treinamento, de validação e de testes. O conjunto de treinamento foi submetido à rede neural e os seus pesos ajustados a partir deste conjunto. O conjunto de validação foi usado para monitorar o grau de generalidade da rede e, quando esta foi considerada pronta, foi utilizado o

conjunto de testes para avaliar o desempenho da rede neural. No classificador textual foi necessário um conjunto de treinamento e um conjunto de teste apenas, não sendo necessário o conjunto de validação. Tanto no classificador textual como no classificador de imagens o conjunto de treinamento é o mais volumoso. Estes procedimentos estão explicados em detalhes nos Capítulos 5 e 6 onde é apresentado o desenvolvimento de cada um dos classificadores. Neste trabalho quando nos referimos aos conjuntos estamos nos referindo às amostras de um tipo (texto ou imagem) a que parte do processamento se refere. São conjuntos: treinamento, validação e teste.

Para que não houvesse a possibilidade de alguns dos conjuntos apresentarem somente amostras de origens similares, um programa de separação dos conjuntos foi desenvolvido. Este programa recebe como parâmetro a composição do conjunto e navega pelas amostras marcando o conjunto a qual deve pertencer. O programa navegou por todas as amostras, ordenadamente, marcando quais seriam pertencentes ao conjunto de treinamento, validação e teste, seguindo estes parâmetros. No caso das imagens os parâmetros foram 3,1,1 que representa que as três primeiras amostras são do conjunto de treinamento, o próximo pertence ao conjunto de validação e o próximo pertence ao conjunto de teste. A seqüência fica se repetindo, de forma que o sexto, sétimo e oitavo elemento seja também pertencente ao conjunto de treinamento, o nono pertencente ao conjunto de validação e assim por diante até percorrer todas as amostras.

Todo o processo de formação da base de dados funcionou em pequenas etapas. Enquanto um conjunto de *hyperlinks* era explorado pelo "*downloader*", um conjunto de imagens recém-baixadas era submetido à pré-seleção e em seguida rotulado. Não poderíamos esperar acabar todo o procedimento da formação da base de dados para iniciar as outras partes do sistema. Assim, com o processo de captura e rotulagem em andamento, uma parte das imagens foi separada para o desenvolvimento do classificador de imagens. Este conjunto foi separado durante todo o desenvolvimento da rede neural e após estar finalizada, ela foi re-treinada com o conjunto de treinamento inteiro.

Este trabalho de coleta de informações mostrou que é possível formar uma base de imagens a partir da Internet, porém é um longo e árduo trabalho, que deve dispor de ferramentas para pré-seleção para a viabilidade do processo. Uma escolha apurada dos endereços iniciais também colaborou muito para o sucesso desta parte do trabalho.

Capítulo 5

Classificação de Imagens

Um dos estágios mais importantes para este trabalho está na classificação de imagens. Esta fase utiliza-se de imagens vindas da base de dados e submetidas a um processo de extração de características que analisa a imagem sob um determinado aspecto e gera alguns valores numéricos que formam um vetor de características. Este vetor denomina-se "vetor de características", e será utilizado para treinar, validar e testar os classificadores baseados em redes neurais. Estas redes efetuam a classificação das imagens e os resultados obtidos demonstram a eficiência deste tipo de classificação. Estas redes, além de apresentarem os resultados da classificação de imagens, compõem a combinação de classificadores final. Os resultados obtidos na classificação de imagens a partir da utilização desta rede será comparado, ao final, com os resultados obtidos a partir da combinação de classificadores, verificando se o uso de informações contextuais é relevante para a classificação de imagens.

5.1. Escolha do Classificador de Imagens

Conforme descrito no Capítulo 2, existem diversas formas de realizar uma classificação de imagens. Neste trabalho foi escolhido um classificador do tipo neural por, principalmente, prover as saídas em forma de probabilidades *a posteriori* do item apresentado à entrada da rede pertencer a cada uma das classes pré-definidas. Com probabilidades sendo fornecidas na saída, a integração com outros classificadores para avaliação do contexto torna-se simplificado. Outro motivo da escolha de redes neurais para a classificação de imagens é que as RNAs (*Redes Neurais Artificiais*) são capazes de classificar, generalizar e aprender funções desconhecidas a partir de um conjunto de exemplos [BIT00].

5.2. Extração de Características

O capítulo 2 abordou alguns princípios da rede neural e de extração de características. Para cada imagem é necessário gerar um vetor de características. A representatividade deste vetor de características é um ponto fundamental para que a rede neural funcione corretamente e de forma eficiente. Muitas características podem ser extraídas de imagens, e neste trabalho foram escolhidas três abordagens que são muito utilizadas em classificação de imagens [KHE04]: formas, cores e textura.

O vetor de características deste trabalho apresenta um total de 120 dimensões sendo 15 para as características baseadas em forma, 90 para as características baseadas em cores e 15 para características baseadas em textura. A análise das imagens e a extração de características foram executadas por três programas criados para esta finalidade, denominados respectivamente: "*SimpleShape"*, "*ColorSimilar"* e "*TextureDetector"* que serão detalhados nos próximos tópicos.

5.2.1 Áreas das Imagens

Conforme descrito no Capítulo 2, geralmente em uma imagem, as áreas centrais são as mais representativas e as mais externas normalmente representam cenários ou outros itens menos importantes. Neste trabalho adotou-se este conceito de separação de áreas, e independente do algoritmo que é usado para a extração das características, é aplicada a separação de áreas das imagens. O algoritmo de extração de característica analisa separadamente cada uma das áreas da imagem e atribui um peso referente à área em questão. Áreas internas recebem pesos maiores enquanto áreas externas recebem pesos menores. A definição de áreas das imagens foi feito dividindo-se a imagem em três áreas proporcionais ao tamanho total da imagem. Um exemplo desta divisão de áreas numa imagem pode ser vista na Figura 5.1.

Figura 5.1. – Exemplo da separação de áreas numa imagem

Nesta imagem pode-se verificar que existem três áreas distintas. A área central é considerada mais representativa, e a mais externa menos representativa. Isto ocorre porque a área próxima à borda faz parte do cenário de fundo e normalmente procura-se centralizar o alvo durante uma fotografia. Assim, a área central conterá o maior peso durante os cálculos para os valores que irão compor o vetor de características.

5.2.2 Transformação para Níveis de Cinza

Das três características utilizadas para compor o vetor de características, apenas a característica baseada de cores necessita da correta representação das cores na imagem, as outras duas características utilizam a transformação da imagem para níveis de cinza, pois analisam formas geométricas e padrões. Para estas duas características é importante transformar a imagem de três canais (RGB) em um único canal, ou em uma imagem em níveis de cinza.

Neste trabalho, este procedimento é feito com a simples soma dos níveis de cores dos *pixels* que representam as cores: vermelho, azul e verde (byte 1, 2 e 3). A Figura 5.2 apresenta um exemplo dos valores de um pixel numa imagem alvo.

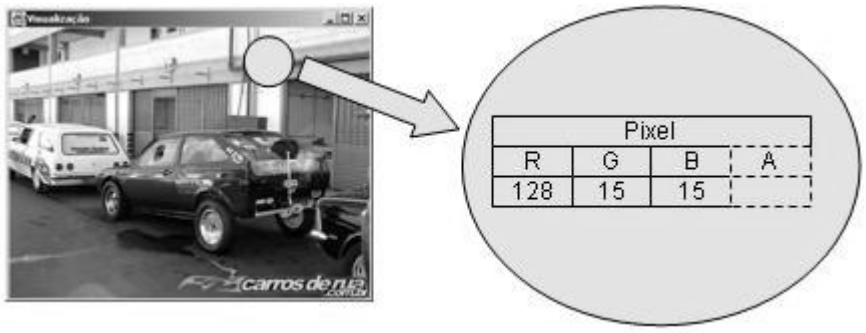


Figura 5.2. – Exemplo da representação de um pixel da imagem

Para transformar a imagem colorida em uma imagem em tons de cinza foi aplicado a fórmula apresentada abaixo, que foi descrita em [GON00]:

$$C(x, y) = \frac{R(x, y) + G(x, y) + B(x, y)}{3} \tag{5.1}$$

onde x e y indicam as coordenadas da imagem. C representa a imagem de saída em tons de cinza. R, G e B representam os canais de cores, respectivamente: *Red*, *Green* e *Blue*.

No exemplo da Figura 5.2, temos um *pixel* com valores RGB 128, 15 e 15 respectivamente, o que daria 128+15+15 = 158/3 = 53 (com o devido arredondamento). Caso exista informação no canal *alpha*, este é ignorado. Este cálculo é feito para todos os *pixels* da imagem, gerando uma nova imagem em níveis de cinza.

5.3. Características Baseadas em Formas

Esta característica preocupa-se em extrair informações baseadas em formas geométricas simples que podem ser representadas em matrizes de 9x9 *pixels*. Este tipo de matriz aplicada a imagens também são chamadas de máscaras [GON00]. No Capítulo 2 foram apresentados vários algoritmos que trabalham com este tipo de máscara. Neste trabalho utilizaram-se máscaras para tratar a imagem alvo e para contar quantas formas pré-definidas existem em cada área da imagem alvo.

O processo inteiro de extração desta característica consiste de vários passos. Primeiramente uma imagem é selecionada da base de dados e submetida aos algoritmos de extração de característica. Para este algoritmo considere a imagem de exemplo apresentada na Figura 5.3.

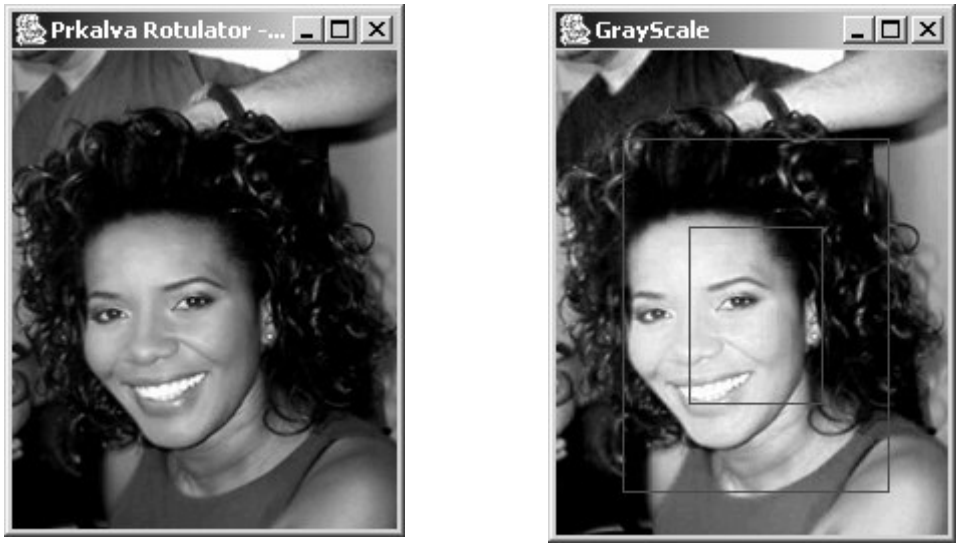


Figura 5.3. – Imagens de exemplo para a extração de características baseado em formas

A imagem original encontra-se à esquerda da Figura 5.3. Esta imagem é transformada em níveis de cinza e dividida em áreas conforme descrito nas seções iniciais deste Capítulo. O passo seguinte é a execução um algoritmo de detecção de bordas. Este algoritmo funciona com a aplicação de máscaras na imagem como filtros. As duas máscaras representadas nas Tabelas 5.1 e 5.2 foram aplicadas separadamente sob a imagem original. Estas máscaras foram retiradas de trabalhos realizados em [GON00].

Tabela 5.1 e 5.2 – Máscaras aplicadas às imagens para detecção de bordas

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Estas máscaras são utilizadas juntamente com algoritmos de detecção de descontinuidades e detectam linhas horizontais (5.1) e verticais (5.2), gerando novas imagens através da aplicação da Equação 5.2 [GON00]:

$$R = w_1 z_1 + w_2 z_2 + \ldots + w_9 z_9 = \sum_{i=1}^{9} w_i z_i \qquad (5.2)$$

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

onde *R* é o resultado da operação, sendo atribuído ao pixel em questão, que coincide com a posição w_5. Os pontos w_1, w_2,...,w_9 representam posicionamentos relativos ao ponto central (w_5). Por exemplo, o ponto w_2 representa um pixel acima do pixel em questão. A posição w_6 representa um pixel à direita, a posição w_4 representa um pixel à esquerda, a posição w_8 representa um pixel abaixo, e assim sucessivamente.

Este cálculo é feito para as duas máscaras indicadas, gerando duas novas imagens com linhas horizontais e verticais detectadas. Nestas duas imagens também é feito o seguinte cálculo:

$$g(x,y) = \begin{cases} 0 & \text{se } f(x,y) < 127 \\ 255 & \text{se } f(x,y) \geq 127 \end{cases} \tag{5.3}$$

Na Equação 5.3, g(x,y) indicam as imagens de saída, f(x,y) indicam as imagens de entrada, ou seja, as imagem processadas pelas máscaras, conforme a Equação 5.2. Os valores 0 e 255 representam as cores preto e branco. A partir deste cálculo têm-se duas imagens com as bordas detectadas compondo-se apenas por pontos de nível mínimo (cor preta) ou nível máximo (cor branca). Uma das imagens representa linhas verticais e a outra imagem apresenta linhas horizontais. O próximo passo é unir estas duas imagens em uma única imagem através de um calculo como uma operação lógica "ou" (OR), conforme a Equação 5.4.

$$G = G' \cup G'' \tag{5.4}$$

onde *G* indica a imagem de saída, *G'*indica a imagem de entrada com uma das imagens processadas pelo cálculo 5.3 (usando mascara para linhas horizontais). *G''* indica a outra imagem também processada pelo cálculo anterior (usando mascara para linhas verticais) e:

$$g(x,y) = \begin{cases} falso(0) & \text{se } f(x,y) = 0 \wedge f'(x,y) = 0 \\ verdadeiro(255) & \text{outros casos} \end{cases} \tag{5.5}$$

onde *x* e *y* indicam as coordenadas da imagem. *g(x,y)* representa a imagem de saída com as bordas detectadas. *f(x,y)* representa uma das imagens de entrada e *f'(x,y)* representa a outra imagem de entrada. Para cada pixel analisado, será atribuído o valor 0 ou 255. Somente será atribuído o valor 0 no caso de nenhuma das imagens apresentar um ponto com valor 255. O valor 255 é atribuído

somente para que a imagem possa ser visualizada em uma interface visual (255=cor branca). Para efeitos de cálculo é considerado somente os valores 0 e 1.

Através deste cálculo, as duas imagens são unidas gerando uma nova imagem composta pela união das imagens anteriores a Equação 5.5. Esta imagem apresenta apenas pontos brancos e pretos representando as bordas detectadas (linhas verticais e horizontais), como ilustra a Figura 5.4.

Figura 5.4 – Imagem gerada com o cálculo 5.5 (bordas detectadas)

O próximo passo deste algoritmo de extração de características baseados em formas geométricas é definir as formas que se quer contar e então verificar se a imagem alvo contém as formas. As formas definidas para este algoritmo devem ser pequenas retas ou pontos que podem ser colocadas em matrizes 9x9. Estas matrizes são apresentadas na Tabela 5.3 [GON00].

Tabela 5.3 – Matrizes 9x9 com formas a serem pesquisadas

-1	-1	-1
-1	8	-1
-1	-1	-1

-1	-1	-1
2	2	2
-1	-1	-1

2	-1	-1
-1	2	-1
-1	-1	2

-1	2	-1
-1	2	-1
-1	2	-1

-1	-1	2
-1	2	-1
2	-1	-1

Estas matrizes são usadas para a detecção de pequenas linhas ou pontos na imagem alvo. Esta matriz é processada em todos os pontos da imagem e é considerado "*encontrado*" quando o cálculo resultar zero, conforme a Equação 5.6.

$$R = w_1 z_1 + w_2 z_2 + ... + w_9 z_9 = \sum_{i=1}^{9} w_i z_i \qquad (5.6)$$

$$Se \;\; R = 0 \quad \text{então} \quad "encontrado"$$

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

onde *R* é o resultado da operação, somente é considerado o status "*encontrado*" quando o resultado do cálculo for 0. Os pontos w_1, w_2,...,w_9 representam posicionamentos relativos ao ponto central (w_5). Por exemplo, o ponto w_2 representa um pixel acima do pixel em questão. A posição w_6 representa um pixel à direita, a posição w_4 representa um pixel à esquerda, a posição w_8 representa um pixel abaixo, e assim sucessivamente. O valor de cada pixel somente pode ser 0 ou 1 (1 = 255, cor branca) e o valor é multiplicado de acordo com o posicionamento e a mascara em questão. Os valores são multiplicados e em caso de resultar em 0 é considerado "encontrado".

Cada matriz é processada de acordo com a Equação 5.6 na imagem alvo com as bordas detectadas. As formas são então contadas de acordo com a área em que se encontra, sendo que informações sobre posicionamento não são consideradas A Figura 5.5. apresenta exemplos deste cálculo, identificando as regiões onde as formas foram encontradas. Para o algoritmo o importante é apenas o número de ocorrências em cada área.

Para a imagem alvo, são processadas as cinco formas em três áreas, totalizando 15 valores. Estes valores passam por um processo denominado "*verificação de capacidade*". Este processo tem o objetivo de normalizar a contagem, pois como se trabalha com imagens de diferentes tamanhos e as matrizes das formas são de tamanho fixo (9x9), a contagem a saída pode ser prejudicada em imagens de tamanho pequeno quando comparado a imagens grandes. Esta verificação consiste em medir a área da imagem para verificar a capacidade e então determinar o índice de ocupação, de acordo com a Equação 5.7.

$$I = \frac{Nf}{Nt} * 100 \tag{5.7}$$

onde *Nf* é o número de formas encontradas e *Nt* é o número de formas que a área pode comportar. O resultado é o índice de ocupação (% de ocupação) que é então atribuído o peso de acordo com a área e então é normalizado. A saída gerada para o exemplo em questão, está ilustrada na Figura 5.6.

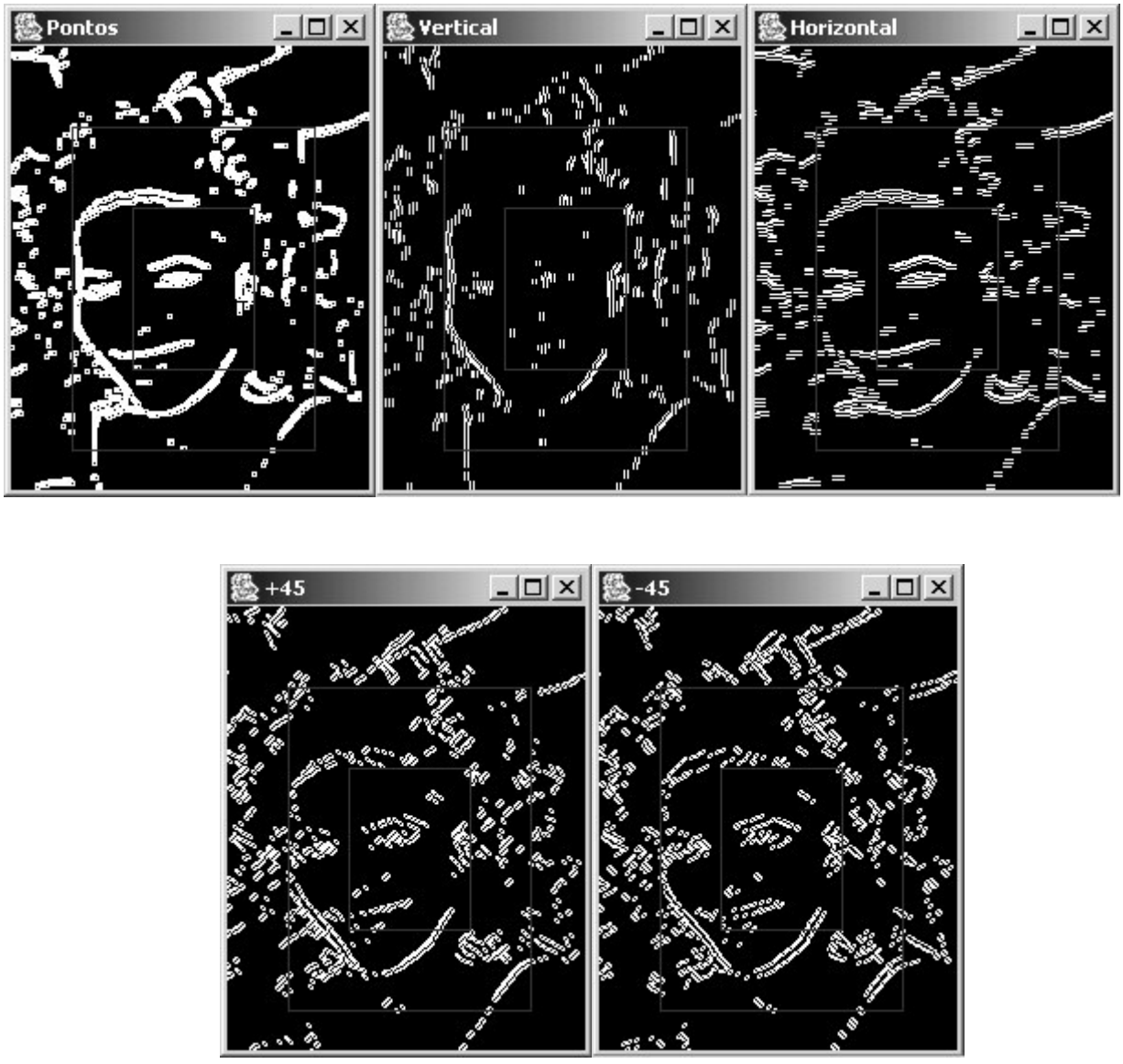


Figura 5.5 – Imagens geradas a partir de identificação de formas na imagem

Status

Resultado SimpleShape		Capacidade	%	Com Peso	Normalizado
Area 0 - Point:	306	5208	5.875576036866359	35.25345622119816	0.17133258678611424
Area 0 - Vertical:	90	5208	1.728110599078341	10.368663594470046	0.0503919372900336
Area 0 - Horizontal:	207	5208	3.974654377880184	23.847926267281103	0.11590145576707726
Area 0 - +45:	139	5208	2.6689708141321042	16.013824884792626	0.07782754759238522
Area 0 - -45:	130	5208	2.49615975422427	14.976958525345621	0.07278835386338187
Area 1 - Point:	1193	15624	7.635688684075781	30.542754736303124	0.14843847206669156
Area 1 - Vertical:	515	15624	3.29621095750128	13.18484383000512	0.06407863630707976
Area 1 - Horizontal:	651	15624	4.166666666666667	16.666666666666668	0.08100037327360957
Area 1 - +45:	630	15624	4.032258064516129	16.129032258064516	0.07838745800671894
Area 1 - -45:	611	15624	3.9106502816180235	15.642601126472094	0.0760233918128655
Area 2 - Point:	989	26040	3.7980030721966207	3.7980030721966207	0.01845837999253453
Area 2 - Vertical:	461	26040	1.7703533026113671	1.7703533026113671	0.008603956700261293
Area 2 - Horizontal:	717	26040	2.7534562211981566	2.7534562211981566	0.013381858902575589
Area 2 - +45:	618	26040	2.3732718894009217	2.3732718894009217	0.011534154535274357
Area 2 - -45:	635	26040	2.4385560675883258	2.4385560675883258	0.011851437103396792

Figura 5.6 – Resultados gerados

Após a obtenção do percentual ocupado (coluna 5) são calculados pesos para indicar a área mais relevante (área 0 – região mais interna da imagem) e então normalizado (última coluna da Figura 5.6). Este resultado é então inserido no vetor de características, ocupando as 15 primeiras posições.

5.4. Características Baseadas em Cores

As cores são características importantes para o processo de classificação. Muitos trabalhos estão baseados unicamente nesta característica [KHE04], [SHI02] ou utilizam características de cores em conjunto com outras características [JAI95]. Sem levar em consideração o formato das imagens, as cores podem determinar semelhanças entre imagens, pois ao analisar a variação de cores vamos encontrar cores planas, com pouca variação nos veículos, cores muitas vezes fortes. Letreiros gráficos são bastante evidentes quando se analisa cores, estes letreiros podem compor capas de CDs ou DVDs. As pessoas apresentam tons de peles diversos e poucas vezes estes tons de cores são encontrados em veículos e motos. Animais também podem possuir algumas cores distintas, sendo interessante e relevante o uso desta característica neste classificador de imagens.

Neste processo de extração de características, a imagem não é transformada para níveis de cinza, como na característica baseada em formas, mas necessita de algo mais representativo do que o formato RGB, pois de acordo com os trabalhos [SHI02],[GON00] os canais RGB representam apenas o uso de cada canal em cada pixel, portanto é difícil extrair relações deste tipo de informação. Para que tenhamos informações relevantes a respeito do relacionamento das cores na imagem alvo, convertemos a imagem para o formato HSI. Um esquema de conversão é apresentado na Figura 5.7.

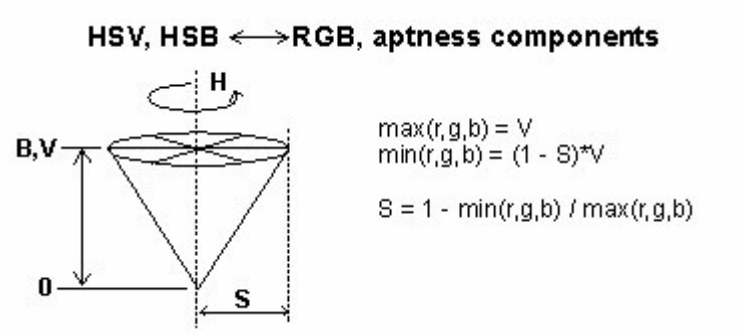


Figura 5.7 – Representação do esquema HSI comparado ao RGB

Desta forma, a figura originalmente no formato RGB é convertida para o formado HSI de onde são extraídos três histogramas com variações entre 0 e 1 (ao contrário do RGB que varia de 0 a 255 por canal), sendo valores de ponto flutuante para cada um dos canais (H, S e I). Para compatibilizar estes valores com o restante do processo, é feito um truncamento de três casas decimais seguidas de uma multiplicação para tornar um número inteiro que passa a variar entre 0 a 99 (100 posições). Um exemplo do histograma gerado é apresentado na Figura 5.8. Neste exemplo não houve divisão por áreas.

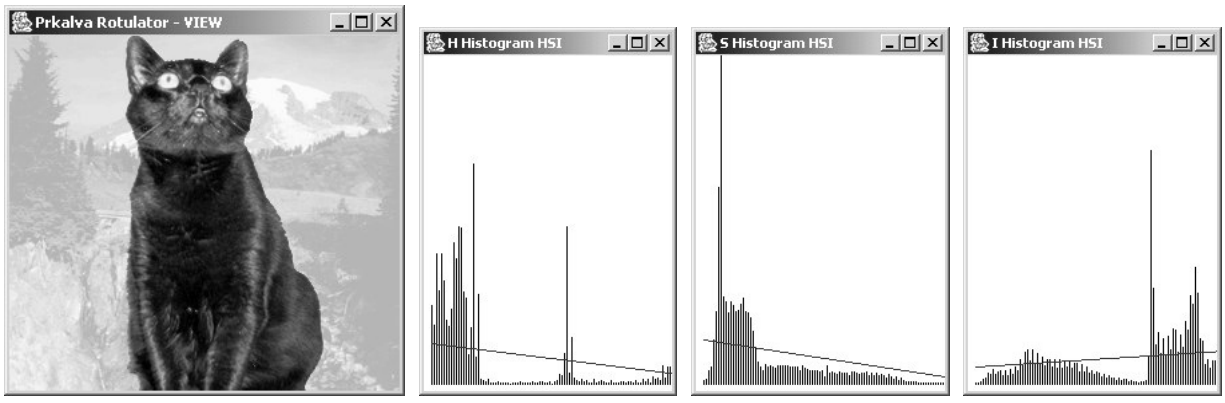


Figura 5.8 – Exemplo do Histograma HSI (100 posições por canal)

Em seguida é feita uma redução para que cada canal comporte até dez variações, uma variação por dezena. Este procedimento é realizado para que haja uma redução em relação à quantidade de entradas para a rede neural. Se considerarmos todos os cem níveis para cada canal, esta característica será representada por 900 posições no vetor de características (considerando as áreas da imagem). Para realizar a redução, cada histograma agrupa seus valores a cada dez posições, sendo que valores de 0 a 9 ficarão na posição 0, valores de 10 a 19 na posição 1 e assim por diante. Os valores agrupados são simplesmente somados, resultando de um valor inteiro.

Os valores de cada histograma, seguidos de agrupamentos, são realizados em cada área da imagem alvo. Em seguida são atribuídos pesos de acordo com a área da imagem. Histogramas gerados a partir da área central da imagem recebem um peso maior. Este valor é normalizado antes de compor o vetor de características.

Para o cálculo de posições no vetor de características são considerados os seguintes termos: três canais (H, S e V), três áreas por canal e dez variações por valor no canal. A ocupação no vetor de características segue o seguinte cálculo:

$$VT = Áreas * Canais * Níveis = 3 * 3 * 10 = 90 \quad (5.8)$$

onde *VT* representa o vetor de características. *Áreas* foi explicado no início deste Capítulo (Seção 5.2.1), *Canais* e *Níveis* foram explicados neste tópico.

Com todos os valores calculados o resultado é então normalizado ficando entre 0 e 1 (ponto flutuante). A característica baseada em cores descrita neste tópico ocupa 90 posições no vetor de característica que é submetido à rede neural. A posição inicia na 16° e termina na posição 105°.

5.5. Características Baseadas em Texturas

A última característica baseia-se na textura da imagem. A textura torna-se importante devido à representatividade que pode oferecer para distinguir as classes existentes. Automóveis têm áreas lisas, animais domésticos tem o corpo com uma textura que segue determinado padrão no caso de pêlos. Esta característica baseia-se em indicadores para obter padrões de textura que podem auxiliar na representação da imagem e melhoria no desempenho do classificador.

São efetuadas análises na imagem para obter-se o resultado de alguns indicadores estatísticos que são, então, calculados e normalizados para a obtenção do valor final que compõe o vetor de características. Estes indicadores foram baseados no trabalho de [GON00], onde a abordagem estatística foi utilizada através de descritores locais.

Inicialmente a imagem é convertida para escalas de cinza, conforme já explicado anteriormente. Em seguida é extraído um histograma com 256 níveis de cinza da imagem alvo e então gerada uma matriz de descrição desta imagem a partir de um determinado operador. Este operador é uma máscara que é utilizada para calcular uma relação de níveis de cinza e posições da imagem em questão. A máscara utilizada neste processamento é apresentada na Figura 5.9.

0	0	0
0	C	1
0	1	0

Figura 5.9 – Máscara usada para o cálculo na geração da matriz

Esta máscara é usada como operador de posição sendo sobreposta em todas as posições da imagem alvo procurando por combinações onde a posição da máscara indicada por “C” (centro) seja de um determinado nível de cinza (que pode variar entre 0 e 255), e as casas indicadas por “1” sejam de outro nível de cinza. Obrigatoriamente todas as posições marcadas com “1” devem ter o mesmo nível de cinza, as posições marcadas com “0” podem ter qualquer valor.

O tamanho da matriz de descrição é dado pela quantidade de níveis de cinza, sendo uma imagem com 255 níveis de cinza apresentará 255 linhas e 255 colunas, que serão representadas pelos níveis de cinza da posição central da máscara e níveis de cinza representados na posição “1” da máscara, respectivamente. Note que quando as posições centrais e posições “1” da máscara são ocupadas pelo mesmo nível de cinza, o posicionamento na matriz de descrição é a diagonal principal. A Figura 5.10 exemplifica a geração da matriz de descrição das posições dos níveis de cinza. Neste exemplo consideramos uma imagem com apenas três níveis de cinza (0, 1 e 2) para efeitos de simplicidade.

0	0	0	1	0
0	1	1	0	2
1	1	2	2	0
1	0	2	0	2
0	2	0	2	0

1	2	3
3	2	0
2	0	1

Figura 5.10 – Exemplo da geração da matriz de descrição (à direita). A imagem exemplo é representada pela matriz à esquerda

Para gerar a matriz de descrição da Figura 5.10 (à direita) foram feitos os seguintes procedimentos: na posição a_{11} da matriz de descrição foram contados quantos pontos de nível “0”

(representando a linha na matriz) que continham pontos de nível “0” (representados pela coluna) no pixel abaixo e no pixel à direita (representados pela mascara da figura 5.9). Na posição a_{12} foram contados quantos pontos de nível “0” que continham pixels de nível “1” abaixo e a direita. Para a posição a_{13} o processo é o mesmo, porém refere-se a contagem de pontos de nível “0” que continham pixels de nível “2” abaixo e a direita. A posição a_{21} usa a mesma regra, porém com nível de cinza “1” para o ponto central (C da figura 5.9) e nível de cinza “0” para as posições marcadas com o nível “1”. O restante da matriz segue a mesma lógica, alterando-se os níveis de acordo com as posições da matriz.

Devido a algumas limitações e representatividade, os níveis de cinza foram reduzidos a 10 posições. Durante o processamento da máscara são considerados como tendo o mesmo valor os níveis que variam entre 0 e 24, sendo atribuído o nível “0” neste caso. Para os níveis entre 25 e 49 é atribuído o valor “1” e assim por diante, perfazendo um total de 10 posições. A matriz de descrição ficou com um tamanho de 10 linhas e 10 colunas (representando os dez níveis de cinza). A matriz da imagem alvo com os tons de cinza já convertidos para os dez níveis é processada da mesma forma exemplificada na Figura 5.10, gerando a matriz de descrição com dez níveis (10x10). O próximo passo foi normalizar esta matriz, somando-se todos os valores da matriz e dividindo cada célula pelo total da soma. Todas as células da matriz ficaram com valores entre 0 e 1 (ponto flutuante). Para o exemplo em questão a Figura 5.11 apresenta a matriz de descrição normalizada, apresentada no exemplo anterior.

1	2	3
3	2	0
2	0	1

1/14	2/14	3/14
3/14	2/14	0
2/14	0	1/14

0,07	0,14	0,21
0,21	0,14	0
0,14	0	0,07

$a_{11}+a_{12}+...+a_{22} = 14$

Figura 5.11 – Matriz de descrição normalizada (Matriz de co-ocorrência)

Resumidamente, a imagem é transformada em tons de cinza e reduzida para dez tonalidades diferentes. Com um operador de posição é gerada uma matriz de descrição com dez linhas e dez colunas, onde na diagonal principal estão ocorrências de mesmo nível de cinza para o operador aplicado. Esta matriz então é normalizada, passando a conter valores entre 0 e 1 de ponto flutuante. Esta matriz é então denominada matriz de *co-ocorrência*. O próximo passo é utilizar os descritores escolhidos para este problema, calculando-os de acordo com a planilha. Estes descritores estão

apresentados nas Equações 5.9 a 5.14 [GON00].

$$\max_{i,j}(c_{ij}) \tag{5.9}$$

onde, c_{ij} representa a matriz de co-ocorrência. i e j indicam a linha e coluna da matriz, respectivamente. Este descritor é denominado *probabilidade máxima* e fornece uma indicação da resposta mais forte.

$$\sum_i \sum_j (i-j)^k c_{ij} \tag{5.10}$$

onde, c_{ij} representa a matriz de co-ocorrência. i e j indicam a linha e coluna da matriz, respectivamente. k representa uma constante. Este descritor é denominado *momento de diferença de elementos de ordem k*. Este descritor possui valor baixo quando os valores altos da matriz de co-ocorrência estiverem próximos a diagonal principal.

$$\sum_i \sum_j c_{ij} /(i-j)^k \qquad i \neq j \tag{5.11}$$

onde, c_{ij} representa a matriz de co-ocorrência. i e j indicam a linha e coluna da matriz, respectivamente. k representa uma constante. Este descritor é denominado *momento inverso de diferença de elementos de ordem k*. Este descritor é o oposto do anterior (5.10).

$$-\sum_i \sum_j c_{ij} \log \vec{c}_{ij} \tag{5.12}$$

onde, c_{ij} representa a matriz de co-ocorrência. i e j indicam a linha e coluna da matriz, respectivamente. Este descritor é denominado *entropia e* é uma medida de aleatoriedade.

$$-\sum_{i}\sum_{j} c_{ij}^{2} \tag{5.12}$$

onde, c_{ij} representa a matriz de co-ocorrência. i e j indicam a linha e coluna da matriz, respectivamente. Este descritor é denominado *uniformidade* e tem o efeito oposto ao anterior (5.11).

São geradas três matrizes como descrito anteriormente, sendo uma para cada área. Os descritores são calculados e a saída é um valor que recebe um peso de acordo com a área que está. Ao final os valores são normalizados. Os cincos descritores extraídos de três áreas totalizam quinze posições no vetor de características, que vão da posição 106ª até a posição 119ª.

Esta característica é a última a ser utilizada no classificador de imagens. Assim, o vetor final é a concatenação dos valores das três características descritas, formando assim um vetor *120-dimensional*. Para cada imagem devem ser processados todos os extratores de características e gerado o vetor que é, então, submetido para a rede neural.

5.6. Classificador Baseado em Redes Neurais

Após a extração de características das imagens, os vetores de características são submetidos à rede neural onde a rede, previamente treinada, irá processar estes valores de entrada e indicará as probabilidades *a posteriori* daquela imagem pertencer a cada uma das cinco classes pré-definidas. Para gerar e treinar a rede neural foi utilizado o simulador SNNS [SNN03].

A rede neural utilizada é do tipo perceptron multicamadas (ou MLP *MultLayer Perceptron*) formada de três camadas: entrada, camada escondida e a camada de saída. A rede é inteiramente conectada, ou seja, todos os neurônios de uma camada estão ligados a todos os neurônios da camada seguinte, como apresentado na Figura 5.12.

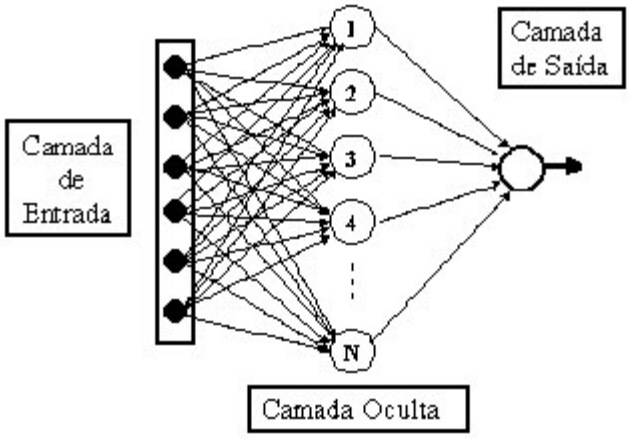


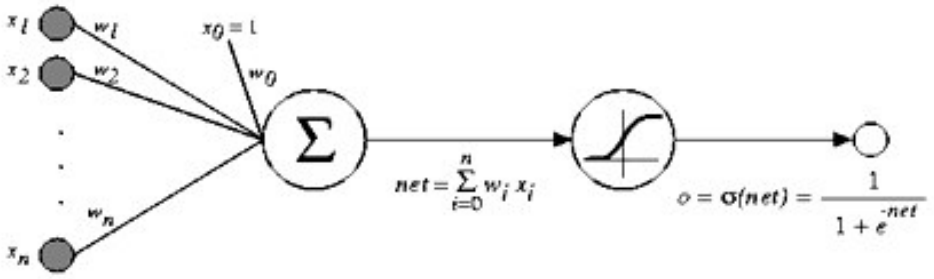
Figura 5.12 – Arquitetura de uma rede neural to tipo MLP

A camada de entrada é composta de 120 neurônios, abrangendo o vetor com os três conjuntos de características descritas nas seções anteriores. A última camada tem cinco neurônios, representando cada uma das classes, sendo respectivamente: automóveis, pessoas, animais domésticos, motos e CD/DVDs. Para determinar a quantidade de neurônios na camada escondida foi utilizado um cálculo heurístico considerando a média do número de neurônios na entrada e saída, conforme a Equação 5.13:

$$LH = \frac{LI + LO}{2} \tag{5.13}$$

onde LH indica o número de neurônios que a camada escondida (*Layer Hidden*) deve ter, LI = quantidade de neurônios na camada de entrada (*Layer Input*) e LO = número de neurônios na camada de saída (*Layer output*). Este cálculo foi baseado na observação de outros trabalhos que usam rede neural [MIC94], [MIT97], [KOV02].

Cada ligação entre neurônios contém um peso associado. Este peso é multiplicado pelo valor do neurônio anterior a qual está ligado, em seguida todos os valores calculados são somados e então submetidos para uma função de ativação. A Figura 5.13 apresenta um exemplo deste esquema.



$\sigma(x)$ is the sigmoid function

$$\frac{1}{1+e^{-x}}$$

Figura 5.13 – Cálculo que ocorre dentro de um neurônio da rede neural

5.6.1. Treinamento da Rede Neural

Os trabalhos de criação e treinamento da rede foram feitos em duas partes: testes gerais da rede e finalização da rede. Esta divisão ocorreu porque a base de dados estava em processo de formação, e não seria possível esperar o encerramento deste processo de coleta de informações para iniciar o desenvolvimento dos classificadores. Desta forma, inicialmente foi coletada uma pequena amostra de imagens e com ela foram desenvolvidos os extratores de características e a rede neural, testando-se cada uma das características isoladamente. Assim foram criadas três redes neurais, uma para cada característica. Ao final foi gerada uma rede única contemplando todo o vetor de características, conforme descrito no tópico anterior.

Nesta fase, uma parte da base de imagens foi separada para ser utilizada em testes posteriores da rede. Três conjuntos de imagens foram criados, onde 1.525 amostras formaram o conjunto de treinamento, 325 amostras formaram o conjunto de validação e 325 amostras formaram o conjunto de testes. Nenhuma imagem foi repetida entre estes três conjuntos. Este procedimento foi feito em duas etapas: resultados usando características individuais e resultados com todas as características.

Na primeira etapa foram gerados os três conjuntos de características para cada uma das três características, formando ao todo nove conjuntos contendo as características já extraídas. Para cada uma delas será apresentado o gráfico de erro médio quadrático (MSE) gerado pelo simulador SNNS, uma matriz de confusão e uma tabela expondo os resultados da classificação.

5.6.2. Características Baseadas em Formas

Para as características baseadas em formas foi inicialmente realizado um treinamento de 30.000 ciclos, onde se pôde verificar uma estabilidade a partir do milésimo ciclo, com uma mínima melhora até o ciclo 10.000. A rede foi reinicializada e re-treinada com 10.000 ciclos. A evolução do erro médio quadrático pode ser visto na Figura 5.14. Nesta figura, a linha inferior representa o erro sobre o conjunto de treinamento e a superior representa o erro sobre o conjunto de validação. O conjunto de testes é usado somente para geração da matriz de confusão. Esta rede foi configurada com 15 neurônios de entrada, 10 neurônios na camada escondida e 5 neurônios de saída.

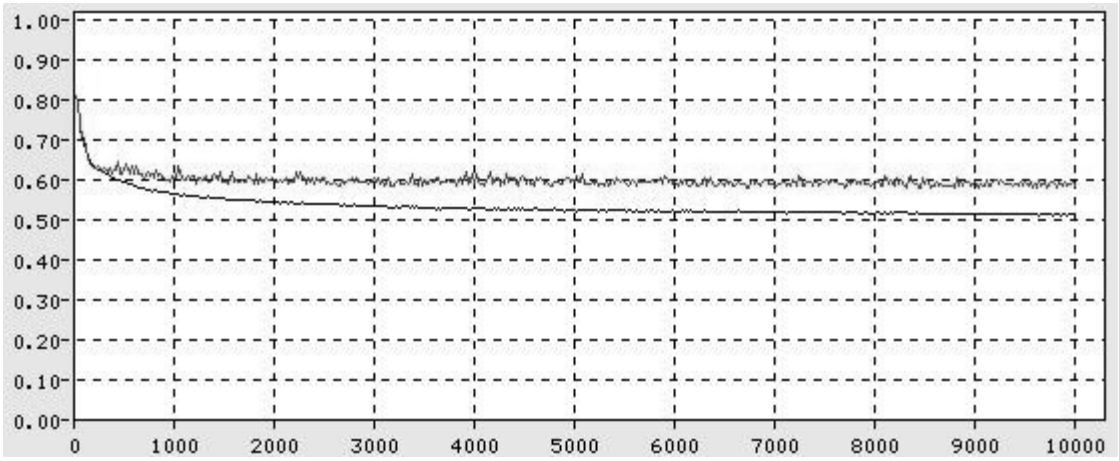


Figura 5.14 – Evolução do erro médio quadrático da característica baseada em formas sobre os conjuntos de treinamento e validação, em função do número de ciclos de treinamento.

Com a rede treinada, foi então utilizado o conjunto de testes para avaliar a *performance*. Cada imagem representada pelo vetor de característica do conjunto de testes foi processada na rede treinada e as saídas foram obtidas. A saída da rede que apresentava a maior probabilidade *a posteriori*, era considerada a classe reconhecida pela rede (*winner take all*). Os resultados obtidos sobre o conjunto de testes na forma de uma matriz de confusão são apresentados na Tabela 5.4.

Tabela 5.4. Matriz de confusão da característica baseada em formas

Classes	Automóveis	Pessoas	Animais	Motos	CD/DVD
Automóveis	33	9	17	2	4
Pessoas	1	55	2	2	5
Animais	7	13	24	10	11
Motos	5	2	14	38	6
CD/DVD	6	9	6	4	40

A partir desta tabela é possível observar alguns resultados interessantes. A classe que obteve melhor classificação foi a classe de “pessoas” e a pior foi a classe “animais”. A maior confusão ocorreu na classe “automóveis”, onde uma grande quantidade de imagens desta classe acabou sendo classificada como “animais domésticos”. A Tabela 5.5 apresenta as taxas de classificação correta das imagens para cada classe individualmente e total.

Tabela 5.5. Resultados obtidos com características baseados em formas

Classe	Acertos	Erros	Total	Taxa de Acerto (%)
Automóveis	33	32	65	50,77
Pessoas	55	10	65	84,62
Animais	24	41	65	36,92
Motos	38	27	65	58,46
CD/DVD	40	25	65	61,54
Total	190	135	325	58,46

5.6.3. Características Baseadas em Cores

As características baseadas em cores ocupam a maior parte do vetor de características, ao todo 90 posições. Os mesmos conjuntos de imagens utilizados para o treinamento da característica anterior foram submetidos ao procedimento de extração e, então, ao treinamento desta nova rede neural. Inicialmente o treinamento desenvolveu-se em 10.000 ciclos e encontrou-se o melhor resultado (erro no treinamento *versus* generalização) próximo a 4.000 ciclos. Da mesma forma que na anterior, a rede foi reiniciada e re-treinada em 4.000 ciclos. A Figura 5.15 apresenta a evolução do erro médio quadrático sobre os conjuntos de treinamento e validação. Esta rede foi configurada para 90 neurônios na camada de entrada, 46 neurônios na camada escondida e 5 neurônios para a camada de saída.

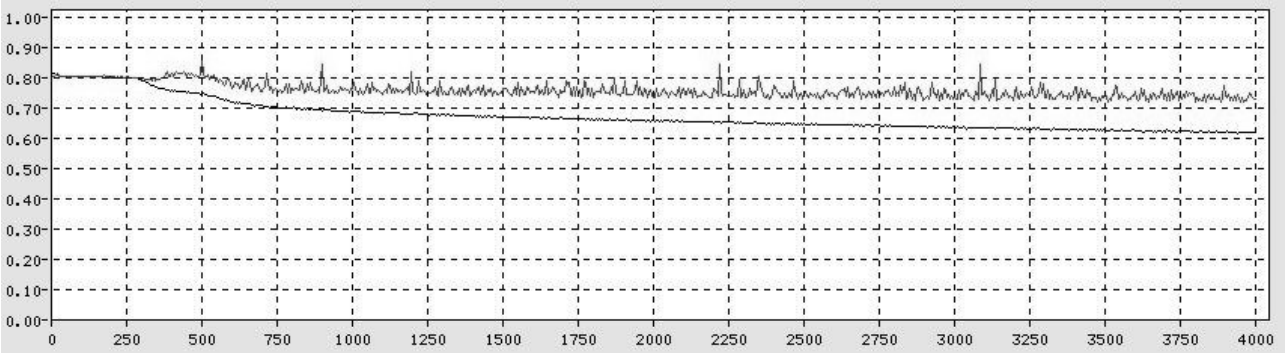


Figura 5.15 – Evolução do erro médio quadrático da característica baseada em cores sobre os

conjuntos de treinamento e validação, em função do número de ciclos de treinamento

Apesar de o erro médio quadrático assumir valores muito elevados, no final do experimento todas as características são usadas juntas, e cada uma delas colabora para a correta classificação. A matriz de confusão obtida a partir do conjunto de testes é apresentada na Tabela 5.6.

Tabela 5.6. Matriz de confusão para característica baseada em cores

Classes	Automóveis	Pessoas	Animais	Motos	CD/DVD
Automóveis	53	0	2	3	7
Pessoas	17	27	2	6	12
Animais	19	24	4	12	6
Motos	26	2	3	29	5
CD/DVD	17	14	1	4	29

A Tabela 5.7 apresenta os resultados sobre o conjunto de testes. Pode-se notar que os resultados são inferiores aos obtidos com as características de forma, porém os acertos ocorrem em classes diferentes da primeira característica, prevendo-se assim um ganho ao unir as duas características, conforme veremos nas seções seguintes.

Tabela 5.7. Resultados obtidos com características baseados em cores

Classe	Acertos	Erros	Total	% de acerto
Automóveis	52	13	65	80,00
Pessoas	27	38	65	41,54
Animais	4	61	65	06,15
Motos	29	36	65	44,62
CD/DVD	29	36	65	44,62
Total	141	184	325	43,38

5.6.4. Características Baseadas em Texturas

Neste conjunto de características houve, a exemplo das anteriores, treinamento inicial em 10.000 ciclos e, então, foi determinado o melhor número de ciclos. Neste caso foram 4.500 ciclos. A rede foi então re-treinada e a evolução do erro médio quadrático pode ser visto na Figura 5.16. Esta rede foi configurada com 15 neurônios de entrada, 10 neurônios na camada escondida e 5 neurônios de saída.

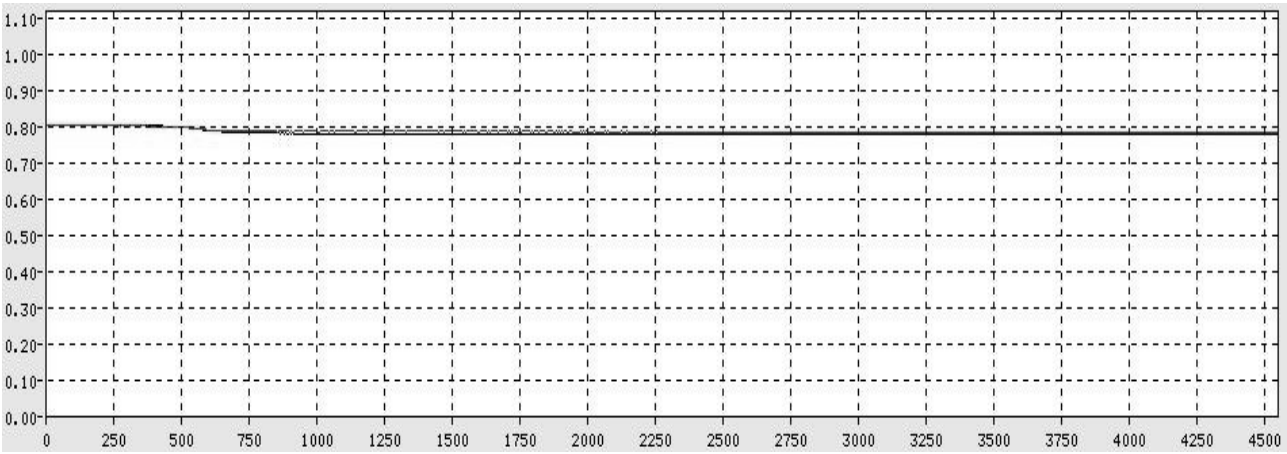


Figura 5.16 - Evolução do erro médio quadrático da característica baseada em texturas sobre os conjuntos de treinamento e validação, em função do número de ciclos de treinamento

Em questões de processamento das características separadamente, esta característica forneceu o pior desempenho, porém com a união das características esperamos um ganho real. A matriz de confusão para característica é apresentada na Tabela 5.8 e os resultados sobre o conjunto de testes são apresentados na Tabela 5.9.

Tabela 5.8. Matriz de confusão para as características baseadas em textura

Classes	Automóveis	Pessoas	Animais	Motos	CD/DVD
Automóveis	11	11	5	8	30
Pessoas	18	13	7	11	16
Animais	10	8	11	29	7
Motos	20	3	14	25	3
CD/DVD	6	17	11	7	24

Tabela 5.9. Resultados obtidos com características baseados em texturas

Classe	Acertos	Erros	Total	% de acerto
Automóveis	11	54	65	16,92
Pessoas	13	52	65	20,00
Animais	11	54	65	16,92
Motos	25	40	65	38,46
CD/DVD	24	41	65	36,92
Total	84	241	325	25,85

5.6.5. Agrupamento das Características: Forma, Cor e Textura

Unificando todas as características, obtivemos um total de 120 elementos no vetor de características, e a rede foi gerada como descrito nas seções anteriores. A evolução do erro médio quadrático até 8.000 ciclos é apresentada na Figura 5.17.

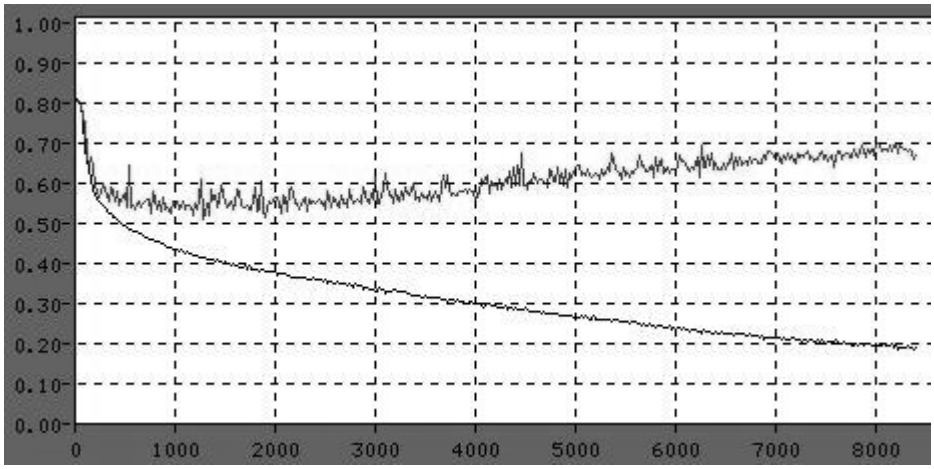


Figura 5.17 – Evolução do erro médio quadrático da característica baseada em todas as características juntas (formas, cores e texturas) sobre os conjuntos de treinamento e validação, em função do número de ciclos de treinamento (8.000 épocas)

Como o erro sobre o conjunto de validação começou a aumentar a partir do 500º ciclo de treinamento provocando a perda da capacidade de generalização, o treinamento foi refeito até 500 ciclos. O resultado é apresentado na Figura 5.18.

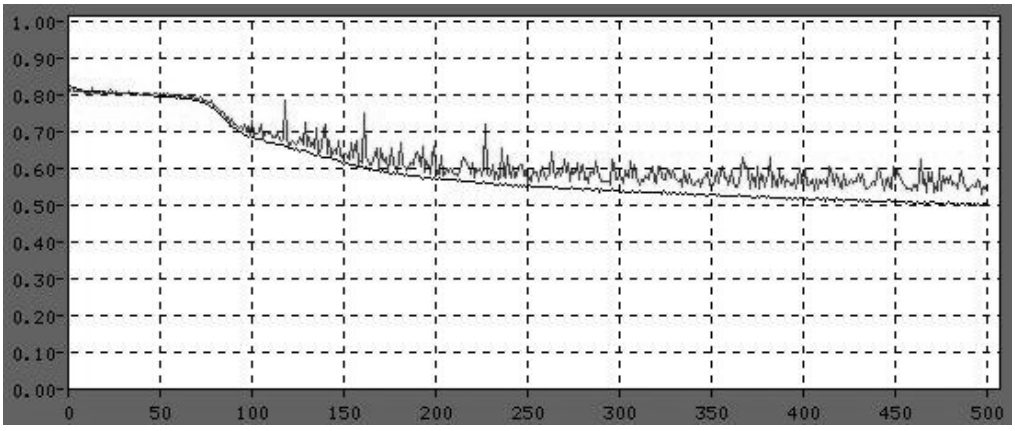


Figura 5.18 – Evolução do erro médio quadrático da característica baseada em todas as características juntas (formas, cores e texturas) sobre os conjuntos de treinamento e validação, em função do número de ciclos de treinamento (500 ciclos)

A matriz de confusão obtida a partir do conjunto de testes para esta rede é apresentada na Tabela 5.10, enquanto a Tabela 5.11 apresenta as taxas de classificação correta para cada classe e total.

Tabela 5.10. Matriz de confusão da classificação da rede completa

Classes	Automóveis	Pessoas	Animais	Motos	CD/DVD
Automóveis	32	5	5	2	6
Pessoas	6	32	3	2	7
Animais	12	6	29	2	1
Motos	4	3	18	24	1
CD/DVD	4	6	5	3	32

Tabela 5.11 – Tabela de resultados da classificação da rede completa

Classe	Acertos	Erros	Total	% de acerto
Automóveis	32	18	50	64,00
Pessoas	32	18	50	64,00
Animais	29	21	50	58,00
Motos	24	26	50	48,00
CD/DVD	32	18	50	64,00
Total	149	101	250	59,60

A próxima fase é a geração da rede neural com uma estrutura igual, porém os conjuntos de treinamento, validação e testes são compostos de mais exemplos.

5.7. Resultados Finais da Rede Neural

Depois de concluído a extração de características e a construção da rede neural foi necessário efetuar o treinamento e teste da rede. A rede foi testada com um conjunto reduzido de amostras de imagens, que serviram para testes individuais, como testes da representatividade das características e configurações da rede. Neste tópico é detalhado o treinamento da rede com todas as amostras oficiais de treinamento, validação e testes.

A base de dados apresentou diferenças entre os números de amostra para cada classe, disponíveis para treinamento da rede neural, conforme pode ser visto na Figura 5.11. Isto levantou a seguinte questão: o que seria melhor, treinar a rede com todas as amostras disponíveis ou deixar um

número igual de amostras por classes, descartando as demais amostras? Com base nesta dúvida foram geradas duas redes. A primeira rede foi construída a partir de todas as amostras, perfazendo um total de 1.081 amostras no conjunto de teste. A segunda rede neural utiliza apenas a mesma quantidade de amostras para cada classe, sendo limitada pelo número de exemplos na classe CD/DVD (847 exemplos). Nesta segunda rede as amostras excedentes são descartadas.

Para a realização deste teste, a rede neural construída de acordo com os tópicos anteriores foi previamente zerada e então treinada e testada com o total de amostras indicadas na Tabela 5.12 (coluna central, intitulada "Total de Imagens – Rede neural 1"). Esta rede foi denominada "*rede neural 1*". Os detalhes da evolução do erro é apresentada na Figura 5.19. Depois de treinada com 3.243 amostras (conjunto de treinamento) e validada com 1.081 amostras (conjunto de validação), esta rede foi testada com 1.081 amostras (conjunto de teste). O resultado é apresentado na Tabela 5.13.

Tabela 5.12. Número de amostras para treinamento, por classe para a rede neural 1 e para a rede neural 2

Classe	Total de Imagens – Rede neural 1	Total de Imagens – Rede neural 1
Automóveis	1087	847
Pessoas	880	847
Animais	1.166	847
Motos	1.425	847
CD/DVD	847	847

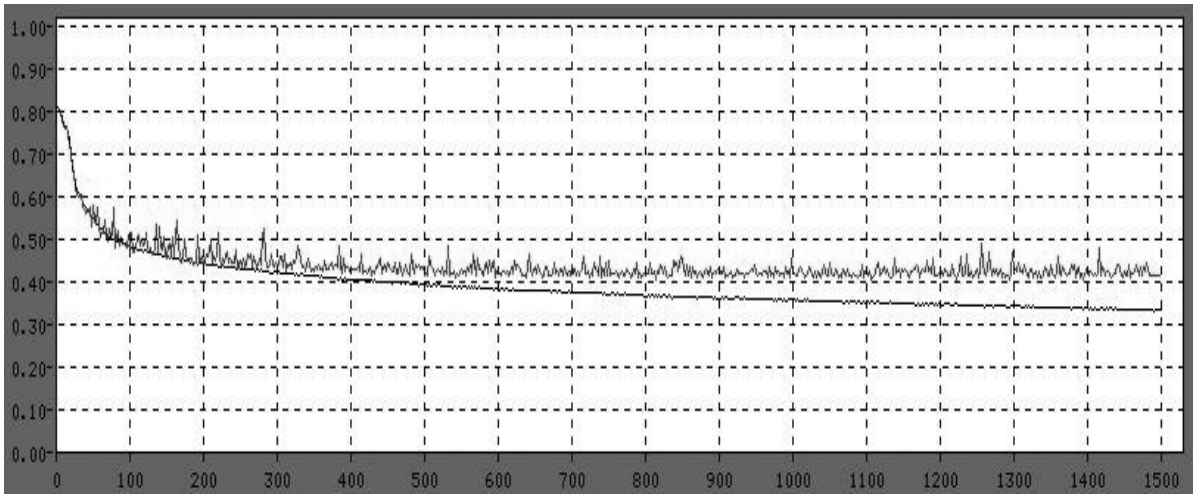


Figura 5.19 - Evolução do erro médio quadrático sobre os conjuntos de treinamento e validação, em função do número de ciclos de treinamento, para a rede neural 1

Tabela 5.13 - Resultados obtidos com todas as características, rede 1

Classes	Automóveis	Pessoas	Animais	Motos	CD/DVD
Automóveis	195	3	2	14	4
Pessoas	12	57	44	25	38
Animais	15	18	156	25	19
Motos	17	1	16	242	9
CD/DVD	7	13	12	9	128

Em seguida, após verificar os resultados da rede neural 1. Esta rede neural foi zerada e treinada com o mesmo conjunto de imagens anteriores, porém com um número igual de amostras por classes, ou seja 847 amostras por classe. Esta rede foi intitulada "*rede neural 2*" e tem como objetivo verificar o desempenho numa situação aparentemente "*ideal*", pois terá o mesmo número de amostras para cada classe nos conjuntos de treinamento, validação e teste. Neste caso a distribuição ficou da seguinte forma: 2.545 amostras para o conjunto de treinamento, 845 para o conjunto de validação e 845 amostras para o conjunto de testes. A evolução do erro para a rede neural 2 é apresentado na Figura 5.20 e a matriz de confusão é apresentado na Tabela 5.14.

Tabela 5.14 - Resultados obtidos com todas as características, rede 2

Classes	Automóveis	Pessoas	Animais	Motos	CD/DVD
Automóveis	132	8	2	20	7
Pessoas	9	89	35	7	29
Animais	5	26	121	9	8
Motos	8	8	24	116	13
CD/DVD	3	23	11	4	128

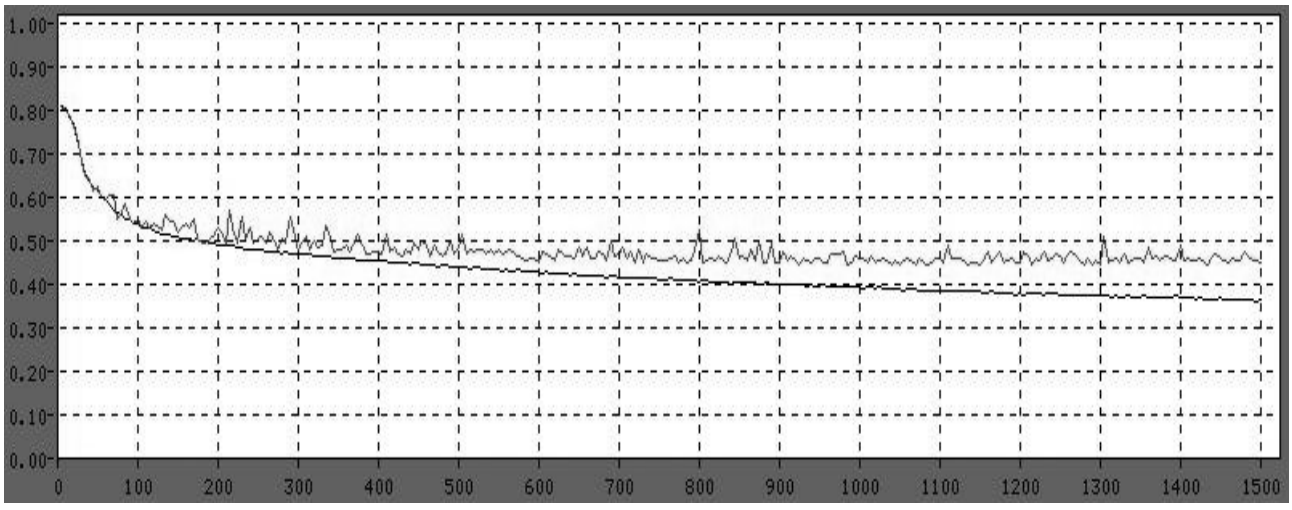


Figura 5.20 - Evolução do erro médio quadrático sobre os conjuntos de treinamento e validação, em

função do número de ciclos de treinamento, para a rede neural 2

A partir destes resultados foi gerada a Tabela 5.15. Nestes resultados percebe-se uma melhora de desempenho da rede neural que é aparentemente ocasionado pelo maior número de amostras. A classe “pessoas” ficou prejudicada provavelmente pelo baixo número de amostras, mas o mesmo não ocorreu com a classe “CD/DVDs”. A classe “motos” que contém a maior parte de amostras foi uma das melhores classificadas na rede neural, enquanto na rede neural 2 ficou na média geral.

Tabela 5.15 – Resultados das redes neurais 1 e 2

	Rede 1			Rede 2			Rede 1	Rede 2
Classe	Acertos	Erros	Total	Acertos	Erros	Total	%	%
Automóveis	195	23	218	132	37	169	89,45	78,11
Pessoas	57	119	176	89	80	169	32,39	52,66
Animais	156	77	233	121	48	169	66,95	71,60
Motos	242	43	285	116	53	169	84,91	68,64
CD/DVD	128	41	169	128	41	169	75,74	75,74
Total	778	303	1081	586	259	845	71,97	69,35

Mesmo com resultados inferiores, a rede escolhida para integrar o experimento foi a rede neural 2, pois está livre de tendências por parte da diferença no número de amostras observados na rede neural 1.

Foi utilizado a ferramenta *SNNS2C* [SNN03] para converter a rede em código na linguagem C, e que na etapa seguinte foi convertido em código *java* e integrado ao restante do experimento. Vale considerar que uma taxa de classificação correta 69,35% pode ser considerada boa para o caso de imagens com muitos ruídos, conforme explicado no Capítulo 3. O objetivo geral deste trabalho é prover uma solução que tenha um ganho quando comparado apenas à classificação de imagens.

Capítulo 6

Classificação de Textos

Para auxiliar a classificação de imagens este trabalho utiliza informações de contexto. No caso de imagens extraídas na Internet, definiu-se que as informações contextuais a serem utilizadas são os textos que estão presentes na página onde a imagem foi capturada. O texto tem um papel importante para a identificação da imagem, pois ao identificar o assunto do texto pode-se utilizar esta informação para direcionar o resultado de uma forma mais inteligente que somente a classificação de imagens. Em muitos casos, os textos podem ser confusos, como o caso de uma página de jornalismo ou um grande portal de assuntos gerais. Nestes casos a classificação do texto acaba sendo irrelevante. Porém, em muitos casos, estes textos podem ser utilizados para limitar o problema a duas ou três classes ou mesmo eliminar uma determinada classe de consideração. Outra colaboração do texto ocorre em casos onde o classificador de imagens apresente resultados com baixas probabilidades. Neste caso o texto pode ser o diferencial para a correta classificação destas amostras.

Neste trabalho, o texto é extraído diretamente das páginas HTML tendo suas *tags* removidas. Porém, todo o texto é submetido ao classificador, inclusive blocos de texto que fazem parte de rodapé e cabeçalho, propagandas, menus, *hyperlinks* e outros blocos de texto presentes na página web.

Os classificadores de textos e imagens enviarão seus resultados em forma de probabilidades para a camada de decisão, que verificará, então, o resultado final com base na combinação destas probabilidades. A camada de decisão será abordada no próximo capítulo.

6.1. Informações Textuais

A exemplo das imagens e já discutido no Capítulo 3, os textos devem ser parecidos aos encontrados em um ambiente real, ou seja, conter ruídos. As páginas HTML encontradas na Internet geralmente têm muito ruído, pois normalmente encontramos páginas como as mostradas na Figura 6.1.

Figura 6.1 – Exemplo de página HTML normalmente encontrado na Internet

Pode-se notar que existem diversos blocos de texto espalhados por várias partes da página. A parte superior da página apresenta propagandas, menus e outros itens estruturais. Ao lado tem-se navegação, lista de caminhos para outras partes do *site,* o conteúdo central e um rodapé (que não aparece na imagem de exemplo). De todo este texto somente o conteúdo central é que aparentemente contém assunto alvo. Porém, teríamos dificuldades em descobrir onde está o conteúdo que de alguma forma referem-se as imagens presente na página.

Para efeitos de simplicidade optamos em não selecionar o conteúdo da página e sim utilizar todo o conjunto textual para ser considerado “contexto”. Em muitos casos pode ocorrer uma quantidade de texto de caráter estrutural que está quase sempre em páginas referentes à CDs ou DVDs ou talvez um conjunto de *hyperlinks* para vários fabricantes de automóveis seja quase sempre visto em páginas de automóveis. Levamos em consideração que o conjunto textual pode apresentar informações sobre as imagens presentes na página, seja um bloco com propaganda, itens estruturais, menus, etc.

As páginas capturadas na Internet por este experimento através do *WWW* (*World Wide Web*) estão em formato HTML e devem receber um tratamento para retirada das *tags*, pois, estes itens podem atrapalhar a tarefa de classificação. Uma parte de código HTML pode ser vista como exemplo na Figura 6.2.

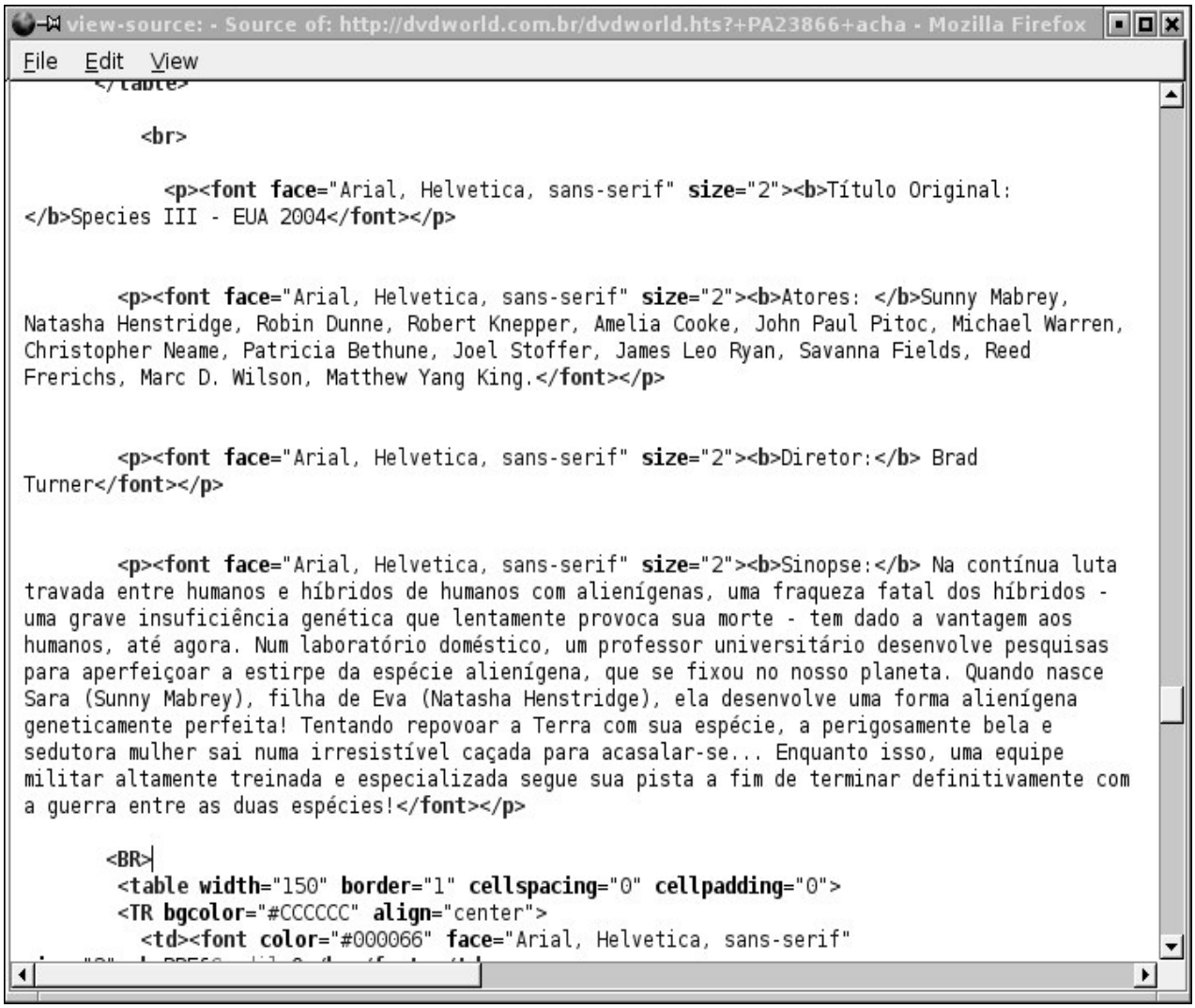


Figura 6.2 – Código da página HTML capturada

As páginas HTML contêm muitas *tags* de formatação, *scripts* de controle e outros elementos que devem ser removidos. Neste processo de classificação de texto, ao receber o texto da base de dados, o texto possui *tags*, *scripts* e outros elementos removidos por um programa criado para esta finalidade. Ao final deste tratamento, o texto fica similar ao apresentado na Figura 6.3.

```
Título Original: Species III - EUA 2004 Atores: Sunny Mabrey, Natasha Henstridge,
Robin Dunne, Robert Knepper, Amelia Cooke, John Paul Pitoc, Michael Warren, Christopher
Neame, Patricia Bethune, Joel Stoffer, James Leo Ryan, Savanna Fields, Reed Frerichs,
Marc D. Wilson, Matthew Yang King. Diretor: Brad Turner Sinopse: Na contínua luta
travada entre humanos e híbridos de humanos com alienígenas, uma fraqueza fatal
dos híbridos - uma grave insuficiência genética que lentamente provoca sua morte -
tem dado a vantagem aos humanos, até agora. Num laboratório doméstico, um professor
 universitário desenvolve pesquisas para aperfeiçoar a estirpe da espécie alienígena,
que se fixou no nosso planeta. Quando nasce Sara (Sunny Mabrey), filha de Eva
(Natasha Henstridge), ela desenvolve uma forma alienígena geneticamente perfeita!
Tentando repovoar a Terra com sua espécie, a perigosamente bela e sedutora mulher
sai numa irresistível caçada para acasalar-se... Enquanto isso, uma equipe militar
altamente treinada e especializada segue sua pista a fim de terminar definitivamente
com a guerra entre as duas espécies!
```

Figura 6.3 – Texto extraído e tratado

Em alguns textos várias palavras ficarão aparentemente sem sentido, pois constituíam menus ou caminhos para outras páginas no *site* ou fora dele. Após este tratamento de retirada de tags, nenhum outro tratamento é efetuado no texto. Palavras soltas, textos de propaganda e/ou estruturais são considerados como “ruídos” e farão parte do texto encaminhado para a classificação.

6.2. Classificador de Textos

Para a classificação de textos foi escolhido o classificador Naïve Bayes, devido a sua simplicidade, bom desempenho para classificação de textos e principalmente posuir as saídas como probabilidades *a posteriori* da entrada pertencer a cada uma das classes sendo compatível com a saída da rede neural, que também apresenta seus resultados como probabilidades.

Este classificador utiliza-se de probabilidades para calcular a qual classe o texto em questão pertence. Isto acontece ao calcular as probabilidades individuais das palavras que compõe o texto com o auxílio de um vocabulário criado antes do treinamento do classificador. O Teorema de Bayes é apresentado na Equação 6.1.

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)} \tag{6.1}$$

onde *P(h)* indica a probabilidade a *priori* da hipótese h. *P(D)* indica a probabilidade a *priori* dos dados de treinamento *D*. *P(h|D)* indica a probabilidade de *h* dado *D* e *P(D|h)* indica a probabilidade de *h* dado *D*.

Este classificador apresenta dois estágios: treinamento e classificação. Durante o treinamento é criado um vocabulário e as probabilidades das palavras deste vocabulário são estimadas com base nas amostras de textos utilizadas no treinamento. Na fase de classificação, as palavras que compõe o texto são analisadas e, somente as palavras coincidentes com as presentes no vocabulário, são consideradas. Neste processamento as probabilidades para cada uma das classes são calculadas. A classificação é indicada pela classe que fornece a probabilidade mais elevada. Estes valores são posteriormente submetidos à camada de decisão, que então analisará juntamente com os resultados da rede neural, fornecendo um resultado final. A camada de decisão é assunto do próximo capítulo.

6.3. Treinamento do Classificador Textual

Este classificador textual utiliza o algoritmo Naïve Bayes. Como este é um classificador muito conhecido, não entraremos em detalhes quanto ao funcionamento do algoritmo, mas apenas os detalhes relativos à aplicação neste trabalho. O algoritmo de treinamento do Naïve Bayes é apresentado na Figura 6.4, conforme descrito em [MIC94].

Aprendizagem_Naïve_Bayes(*Exemplos*,V)

1. Colecionar todas as palavras e outros *tokens* que ocorreram em "*Exemplos*"
   - *Vocabulário* ← Todas as palavras distintas e outros *tokens* em "*Exemplos*"

2. Calcular os termos de probabilidade necessárias $P(v_j)$ e $P(w_k \mid v_j)$

Figura 6.4 – Algoritmo de aprendizagem Naïve Bayes [MIC94]

Neste algoritmo o treinamento ocorre em duas etapas: formação do vocabulário e cálculo das probabilidades. O esquema de treinamento adotado é apresentado na Figura 6.5. Primeiramente os textos HTML da base de dados têm suas *tags* removidas. Em seguida as palavras irrelevantes do texto, tais como: *olá, agora, eu, até, com, também*, etc., são removidas do texto com o auxílio de uma lista previamente constituída (*stopwords*). Símbolos, dígitos e outros caracteres também são removidos. As palavras que restaram formaram o vocabulário após alguns critérios de escolha. Todas as palavras passaram por um processo de remoção de acentuação e transformação para caixa alta, para que não haja confusão durante o processamento.

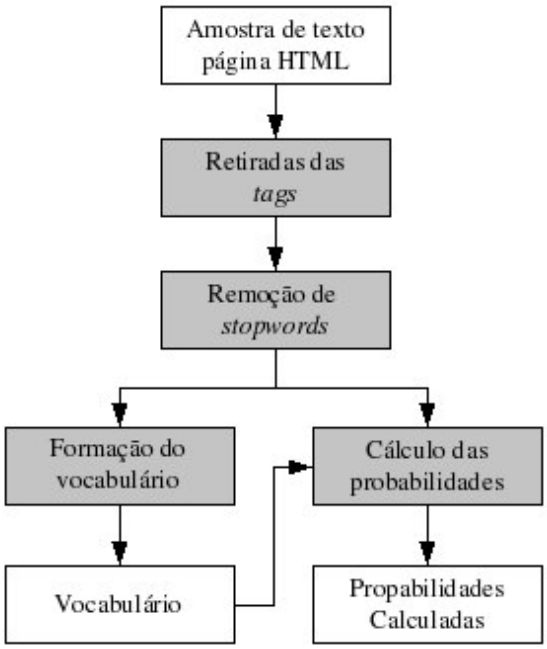


Figura 6.5 – Esquema de treinamento do classificador textual

Com o vocabulário formado, as palavras que o compõe serão avaliadas em relação a todos os textos submetidos para o treinamento. Isto irá resultar em um cálculo de probabilidade por palavra integrante no dicionário e por classe. Este procedimento será detalhado posteriormente.

6.4. Criação do Vocabulário

Na formação do vocabulário são coletadas palavras de todos os textos submetidos ao treinamento. Cada nova palavra encontrada é adicionada ao vocabulário. Palavras repetidas são consideradas uma única vez, porém é computado o número de ocorrências para fins estatísticos. Este conjunto de textos de treinamento foi devidamente separado conforme detalhado no Capítulo 3 e contém 3.104 amostras de textos. Nestes textos foram encontradas 1.375.485 palavras com uma média de 443 palavras por texto. O maior texto encontrado foi em um *website* de venda de DVDs com 48.998 palavras. Como as palavras repetidas não fazem parte do vocabulário, 83.994 palavras foram utilizadas como vocabulário. Porém, com os critérios estabelecidos, 67.812 palavras foram eliminadas, ficando apenas 16.182, ou seja, 19,27% das palavras encontradas. Outros dados estatísticos sobre o processo de formação do vocabulário estão apresentados na Tabela 6.1.

Outro fator que contribuiu para este número de palavras foram os textos estavam em várias línguas (português, inglês e espanhol), sendo mais freqüentes textos em português e inglês. A listagem de palavras irrelevantes, *stopwords,* continha amostras nestas três línguas.

Tabelas 6.1. Estatísticas sobre o processo de formação do vocabulário

Características	***Valor***
Quantidade de amostras de texto	3.104
Tempo de processamento (comp.1)	07:23
Total de Palavras encontradas	1.375.485
Média de palavras por texto	443
Texto com mais palavras	48.998
Total de palavras inseridas no vocabulário (p1)	83.994
Total de palavras eliminadas por baixa freqüência	67.812
Total de vocábulos que compõe o vocabulário final	16.182

Palavras com pouca freqüência no conjunto de treinamento significam palavras incomuns, com uso em textos muito restritivos ou até mesmo erro de digitação. Estas palavras foram eliminadas do vocabulário por um processo denominado "análise de relevância". Este processo verificou a freqüência destas palavras em relação a todos os textos submetidos ao treinamento. Cada palavra teve uma contagem de quantas vezes apareciam em todo o conjunto de treinamento. Entre as palavras com maiores ocorrências ficaram: "*starring*" com 5.523 ocorrências e "*motorcycle*" com 2.409 ocorrências.

As palavras que ocorreram poucas vezes são consideradas irrelevantes, sendo eliminadas do vocabulário, pois, não tem contribuição significativa. A Tabela 6.2 apresenta uma listagem de palavras por ocorrências, que foram consideradas irrelevantes.

Tabela 6.2. Quantidade de palavras irrelevantes

Quantidade de ocorrências	***Quantidade de palavras com a freqüência indicada***
1	36.691
2	12.224
3	5.745
4	4.124
5	2.579
6	2.071
7	1.719
8	1.447
9	1.318
10	1.279
11	946
12	846
13	675
14	589
15	549
16	479
17	441
18	456

Quantidade de ocorrências	*Quantidade de palavras com a freqüência indicada*
19	372
20	347

Esta tabela indica que houve 36.691 palavras inseridas no vocabulário que aparecem somente uma única vez dentre as 1.375.485 dos 3.104 textos submetidos. Analisando o outro extremo desta base de dados, sendo as palavras mais freqüentes, temos o indicado na Tabela 6.3.

Tabela 6.3. Quantidade de ocorrências de palavras com pelo menos a freqüência indicada

Freqüência de pelo menos	*Ocorrências*
10 vezes	15.510
20 vezes	9.810
30 vezes	7.315
40 vezes	5.880
50 vezes	4.917
100 vezes	2.516

Esta tabela indica que houve 15.510 palavras que foram citadas pelo menos 10 vezes durante o processo de formação do vocabulário, assim como houve 2.516 palavras que foram citadas pelo menos 100 vezes. Logicamente, existe uma intersecção entre os conjuntos indicados nesta tabela, sendo que o conjunto com maior freqüência também faz parte do conjunto de menor freqüência. Para saber a quantidade exata de ocorrências, devem-se subtrair as ocorrências do conjunto maior do conjunto atual. Por exemplo, para saber quantas ocorrências contêm entre 20 e 29, basta subtrair o número de ocorrências (9.810) do conjunto maior (7.315), que neste exemplo resulta em 2.495 ocorrências. As palavras cuja ocorrência seja inferior a 10 foram eliminadas do vocabulário, perfazendo-se um total de 16.530 palavras ativas no vocabulário.

6.5. Cálculo das Probabilidades

Com o vocabulário formado, pode-se iniciar a estimação das probabilidades de cada palavra com relação a cada uma das classes a partir do algoritmo da Figura 6.1, no segundo estágio. O treinamento ocorre de forma separada para cada uma das classes. Todos os documentos de

treinamento rotulados com uma classe pré-definida são concatenados. Suas palavras são extraídas e comparadas ao dicionário formado na etapa anterior. Somente as palavras constantes no vocabulário é que são consideradas, eliminando-se todas as outras ocorrências. As palavras presentes no documento concatenado e que aparecem no vocabulário são consideradas palavras ativas, e suas freqüências são computadas (número de vezes que apareceram). Com estas informações extraídas, o cálculo do Naïve Bayes pode ser feito, de acordo com a Figura 6.6.

Calcular os termos de probabilidade necessárias $P(v_j)$ e $P(w_k \mid v_j)$

- Para cada valor alvo v_j faça:
- $docs_j \leftarrow$ subconjunto de "*Exemplos*" para qual o valor alvo é v_j
- $P(v_j) = \dfrac{|docs_j|}{|Exemplos|}$
- $text_j \leftarrow$ um único documento criado pela concatenação de todos os membros de $docs_j$
- n ← número total de palavras em $text_j$ inclusive palavras repetidas múltiplas vezes
- para cada palavra w_k em "*Vocabulário*"
  - $n_k \leftarrow$ número de vezes que a palavra w_k ocorre em $text_j$

$$P(w_k \mid v_j) = \frac{n_k + 1}{n + |Vocabulario|}$$

Figura 6.6 – Algoritmo do calculo dos termos de probabilidade

Ao final dos cálculos teremos a probabilidade a priori - $P(v_j)$ - ou seja, é a probabilidade do texto ser de determinada classe antes de, sequer avaliar qualquer outro parâmetro. No caso da probabilidade a posteriori – $P(w_k \mid v_j)$ - tem-se a probabilidade de uma palavra (constante no vocabulário) ser de determinada classe. Todos os valores calculados são armazenados para serem utilizados durante a classificação. A quantidade de valores de probabilidades *a posteriori* é igual a cinco vezes o número de palavras integrantes no vocabulário (pois temos cinco classes).

6.6. Classificação

Com o classificador treinado (vocabulário formado e probabilidades calculadas e armazenadas), é possível efetuar a classificação de textos. Para este passo foi utilizado um subconjunto de textos com 2.057 amostras de texto não utilizadas durante o procedimento de treinamento. Para o processo de classificação do texto, cada palavra do texto em questão é comparada com as palavras integrantes do vocabulário construído na fase de treinamento. Somente as palavras que aparecem no vocabulário são utilizadas, sendo ignoradas as restantes. O algoritmo de classificação Naïve Bayes está descrito na Figura 6.7 de acordo com [MIC94].

Classificar Naïve Bayes (*DOC*)

- *posições* ← Todas as posições das palavras em DOC que contém *tokens* encontrados em "*Vocabulário*"
- Retornar v_{NB}, onde

$$v_{NB} = \arg\max_{v_j \in V} P(v_j) \prod_{i \in posições} P(a_i \mid v_j)$$

Figura 6.7 – Algoritmo de classificação Naïve Bayes [MIC94]

O uso deste algoritmo neste trabalho pode ser visualizado na Figura 6.8. Primeiramente o texto original, a ser classificado, tem suas *tags* removidas e em seguida é processado um algoritmo que verifica as ocorrências de suas palavras no vocabulário. Com a lista de palavras presentes no texto e também no vocabulário, a probabilidade *condicional*, já calculada na fase de treinamento, é multiplicada sucessivamente (para a classe em questão), até que todas as palavras presentes simultaneamente no texto e no vocabulário sejam processadas, inclusive ocorrências de palavras repetidas (no texto). Após todas as multiplicações, é efetuada a multiplicação da probabilidade *a priori* da classe em questão. Essa operação é feita para cada classe a ser avaliada (cinco classes). Um resultado de ponto flutuante é gerado para cada classe processada e em seguida os cinco resultados são normalizados, ficando entre 0 e 1.

O texto é considerado classificado para a classe que apresentar a probabilidade mais alta. Para efeitos de limitação computacional, as multiplicações foram substituídas pela soma dos logaritmos das probabilidades. Com a normalização dos valores, os resultados ficam compatíveis com a entrada da camada de decisão e semelhante à saída da rede neural.

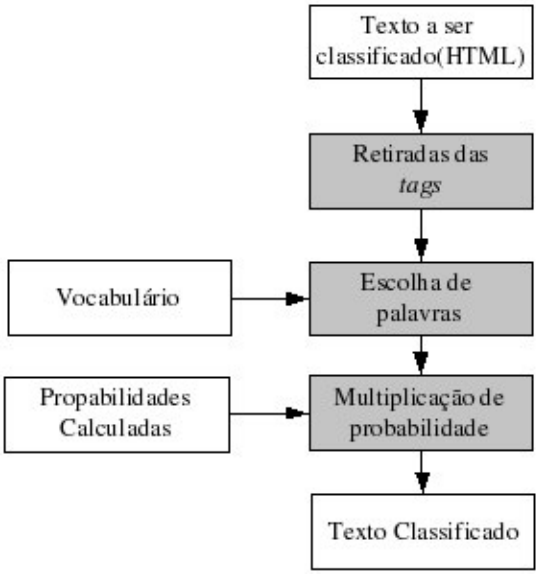


Figura 6.8 – Fluxo da classificação de textos

Após o treinamento do Naïve Bayes com seu respectivo conjunto de treino, foi efetuado um teste utilizando o conjunto de testes. Neste conjunto de teste, nenhuma amostra foi usada durante o treinamento. De acordo com os resultados obtidos, este classificador estatístico se comportou muito bem, sendo os resultados exibidos na Tabela 6.4. A matriz de confusão é apresentada na Tabela 6.5.

Tabela 6.4. Resultado da classificação de textos

Classe	***Acertos***	***Erros***	***Totais***	***% de acerto***
Automóveis	339	72	411	82,48
Pessoas	354	48	402	88,06
Animais	385	36	421	91,45
Motos	390	11	401	97,26
CDs/DVDs	392	30	422	92,89
	1.860	197	2.057	**90,43**

Tabela 6.5. Matriz de confusão da classificação de textos

	Automóvel	***Pessoas***	***Animais***	***Motos***	***Cds/DVDs***
Automóvel	339	34	0	38	0
Pessoas	13	354	4	27	3
Animais	0	29	385	5	1
Motos	3	4	0	390	1
Cds/DVDs	0	13	1	14	392

Com uma taxa de classificação de 90,43%, este resultado demonstra a eficácia do algoritmo Naïve Bayes para este tipo de problema. Deve-se levar em consideração que estes textos contêm muito ruído. O conjunto usado para treinamento foi composto de aproximadamente 60% das amostras de textos disponíveis na base de dados.

Uma página HTML pode conter muitas imagens e quando nos referimos a um *website* com assuntos variáveis, este pode conter imagens de classes diferentes para o mesmo texto. O texto pode ser classificado como, por exemplo, automóveis. Porém, o dono do *website* pode ter colocado a foto de uma pessoa. O classificador de imagens tenderá a classificar esta imagem para a classe "pessoas". O problema ocorre justamente neste ponto, pois o classificador de imagens tenderá a classificar a imagem para a classe pessoa e o texto apresentará a classificação como automóvel. Como o objetivo deste trabalho é melhorar a classificação de imagens, a camada de decisão deve trabalhar com situações deste tipo, descobrindo padrões como o descrito aqui e classificando a imagem de maneira coerente. A camada de decisão é o assunto do próximo capítulo.

Capítulo 7

Camada de Decisão

Nos capítulos anteriores foram apresentados os classificadores de imagens e textos. Apesar destes classificadores funcionarem de forma diferente, ambos fornecem probabilidades *a posteriori* na saída. Estas probabilidades são atribuídas para todas as cinco classes possíveis e os valores variam entre zero e um. Sob o ponto de vista dos classificadores, a classe cuja saída correspondente apresenta o valor mais alto de probabilidade é considerada vencedora (*winner take all*).

O objetivo principal deste trabalho é melhorar a classificação de imagens. Portanto, deve-se levar em conta o resultado obtido pelo classificador de imagens e, através de um processo de decisão, deve levar em conta também o contexto, que neste caso são as informações textuais presentes nas páginas *web*. Somente então teremos um resultado final para a classificação. Todo este processo de avaliação dos classificadores e geração do resultado final fica por conta da camada de decisão. Esta camada utiliza um algoritmo de combinação de classificadores e um conjunto de regras para chegar a uma decisão final.

7.1. Definição da Camada de Decisão

A camada de decisão tem o objetivo de receber os resultados das probabilidades geradas pelos classificadores, analisar estas informações e indicar uma classe que representa a classificação final de uma determinada imagem. No Capítulo 3 foi apresentada uma visão geral deste trabalho onde a camada de decisão aparece como último elemento. Como o nome sugere esta camada decide em qual classificador confiar dada uma entrada ou qual resultado fornecer em determinada situação.

Um exemplo da atuação da camada de decisão é aceitar o resultado dos classificadores quando ambos apresentam altas probabilidades para a mesma classe. Mas e se o classificador estatístico apresentar uma probabilidade alta para uma determinada classe e o classificador neural

apresentar uma probabilidade alta para outra classe? Para resolver estas e outras situações a camada de decisão foi montada com auxílio de alguns algoritmos baseados em regras. Estas regras foram obtidas de várias formas, inclusive através da utilização do algoritmo C4.5 [QUI93] [FER01][FEM01], que é muito utilizado em aprendizagem de maquina e extração do conhecimento. Entretanto, as regras de decisão gerada pelo algoritmo C4.5 teve uma utilização limitada, sendo utilizada apenas como uma ferramenta auxiliar, conforme será detalhado adiante neste capítulo.

7.2. Os Estágios da Camada de Decisão

A camada de decisão opera com base em regras que foram obtidas através de diversos métodos e de ferramentas especialmente criadas para esta finalidade. O algoritmo usado como auxiliar para definir as regras foi o C4.5. Na Figura 7.1 é apresentado o esquema geral da camada de decisão.

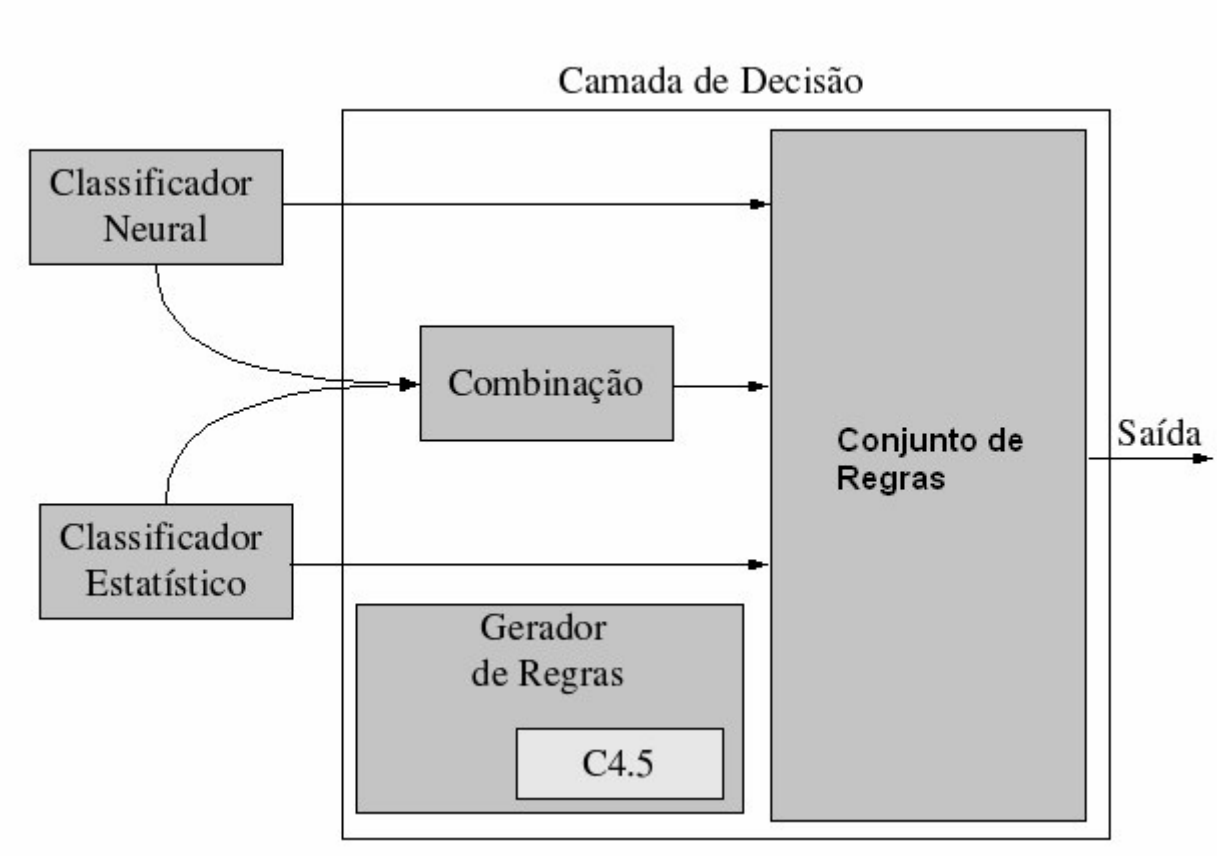


Figura 7.1 – Estrutura da Camada de Decisão

Ambos os resultados dos classificadores são submetidos à entrada da camada de decisão. O primeiro passo é a geração de uma combinação dos resultados dos classificadores, que é feita a partir da multiplicação de probabilidades [KIT98], seguidas de normalização, conforme demonstra

a Figura 7.2. As saídas dos classificadores têm suas probabilidades multiplicadas de forma unitária para cada classe, ou seja, cada classificador apresenta cinco saídas, cada uma representando uma classe distinta. Desta forma a saída que representa, por exemplo, a classe “automóveis” do classificador é multiplicada somente pela saída que também representa a classe “automóveis” do classificador estatístico. Para esta combinação, cinco multiplicações serão feitas (uma para cada classe). Ao final a normalização ocorre para manter os resultados entre zero e um (ponto flutuante).

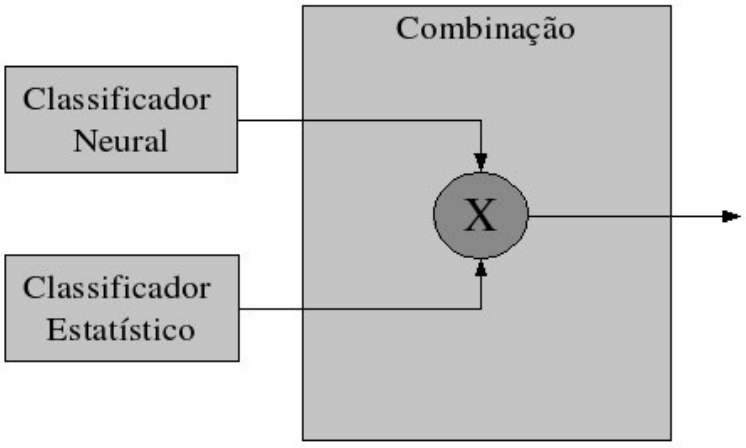


Figura 7.2 – Combinação dos classificadores

Ainda na Figura 7.1, as três saídas (neural, estatístico e combinação) são submetidas às regras, de forma seqüencial. Cada regra utiliza-se de valores, posicionamento no ranking de probabilidades e outras características. Ao testar uma regra é produzido um resultado booleano, e em caso positivo é executada uma ação, que pode ser a submissão do fluxo de processamento para novas regras ou a produção do resultado final. Em caso negativo passa-se para a próxima regra até que seja encontrado um resultado positivo. Ao final de toda a operação, e em caso de nenhuma regra ser positiva, o resultado será o definido pela última ação. Sempre haverá uma resposta, pois a camada de decisão não rejeita a amostra.

7.3. Uso de Regras Simples

Para geração das regras foram usados diversos métodos. Algumas regras foram geradas a partir do uso de alguma ferramenta enquanto outras foram geradas através de heurísticas. O conjunto de dados usado neste estágio é denominado “conjunto vinculado” por conter imagens e textos ligados pela sua origem (texto de conteúdo na origem onde a imagem foi extraída). Este conjunto contém 712 amostras que estão armazenados na base de dados. Este conjunto é

considerado de treinamento, portanto as imagens usadas aqui não farão parte dos resultados finais, que serão apresentados no próximo capítulo. O conjunto vinculado de treinamento foi primeiramente submetido ao classificador de imagens para uma avaliação do método convencional de classificação de imagens. Os resultados da classificação e a matriz de confusão são apresentados nas Tabelas 7.1 e 7.2.

Tabela 7.1. Resultado da classificação de imagens

Acertos	523
Erros	189
Total de Amostras	712
% de Acerto	73,46%

Tabela 7.2. Matriz de confusão para a classificação de imagens

	Automóveis	Pessoas	Animais	Motos	DVD
Automóveis	142	2	4	18	2
Pessoas	12	31	21	25	13
Animais	8	16	175	32	13
Motos	6	1	3	102	4
CD/DVD	0	0	6	3	73

Com os resultados apresentados nesta tabela sabe-se que para o conjunto de dados em questão, a taxa de acerto do classificador de imagens é 73,46%. Deve-se considerar o nível de ruído destas imagens conforme descrito no Capítulo 3. O objetivo é fazer com que a camada de decisão atue no sentido de verificar padrões de resposta vindo dos classificadores e forneça um resultado melhor do que o obtido somente com o uso do classificador de imagens.

Para o funcionamento da camada de decisão, inicialmente fez-se necessário testar alguns métodos de combinação de classificadores e avaliar os resultados. Foram utilizados dois métodos, um que efetua uma soma simples e outro que efetua a multiplicação dos resultados dos classificadores. A combinação é feita aplicando-se o operador em cada uma das saídas que representam a mesma classe entre os classificadores. Por exemplo, o classificador de imagens tem a sua saída que representa a classe “pessoas” combinada com a saída do classificador textual que também representa a classe “pessoas”. A combinação usando o operador soma, ou seja, efetuando uma soma simples entre cada uma das saídas dos classificadores atingiu uma taxa de classificação

de 73,60% e a combinação através do operador multiplicação também forneceu uma taxa de 73,60%. As matrizes de confusão para estas duas combinações são apresentadas nas Tabelas 7.3 e 7.4.

Tabela 7.3. Matriz de confusão da combinação com o operador soma

	Automóveis	Pessoas	Animais	Motos	DVD
Automóveis	142	2	4	18	2
Pessoas	12	31	21	25	13
Animais	8	15	176	32	13
Motos	6	1	3	102	4
CD/DVD	0	0	6	3	73

Tabela 7.4. Matriz de confusão da combinação com o operador multiplicação

	Automóveis	Pessoas	Animais	Motos	DVD
Automóveis	142	2	4	18	2
Pessoas	12	31	21	25	13
Animais	8	15	176	32	13
Motos	6	0	4	102	4
CD/DVD	0	0	6	3	73

Pode-se notar que os resultados são quase idênticos. Somente uma amostra na classe "motos" é que, continuando errada, foi classificada de forma diferente. Se comparado com os resultados do classificador de imagens, obteve-se um pequeno ganho de 0,14% que pode ser considerado desprezível. Assim, o operador escolhido foi o de multiplicação. Nenhum motivo especial levou a esta escolha, pois os resultados do teste efetuado não apresentaram diferenças significativas. Porém, com base no classificador de texto que efetua multiplicações com as probabilidades, este parece ser o mais adequado por ser mais preciso.

Após a definição da combinação de classificadores foi necessário montar uma árvore de regras. Para esta tarefa foram criados diversos mecanismos que geravam algum cálculo que então

era testado e em caso de resultados positivos e relevantes era inserido na árvore final de regras. Para o início desta atividade definiram-se algumas regras através da simples observação e aspecto lógico, são elas:

- Se a maior probabilidade do classificador de imagens é igual a maior probabilidade do classificador de texto. Então assuma esta classe.
- Se a probabilidade do classificador de imagens é maior que 0.8. Então assuma a classe que tem a maior probabilidade do classificador de imagens.
- Se o fluxo de processamento chegou até aqui é porque as regras acima falharam, então assuma a classe indicada pelo classificador de texto.

O fluxo do processamento ocorre avaliando as regras uma a uma na ordem definida e em caso afirmativo, tomam uma ação que leva a classificação e ao encerramento do algoritmo. Em caso negativo passa-se para a regra subseqüente. Estas regras elevaram o índice de classificação para 83,29%, ou seja, um ganho de 9,83% com relação a simples classificação das imagens pela rede neural. A matriz de confusão deste resultado é apresentada na Tabela 7.5.

Tabela 7.5. Matriz de confusão a partir da combinação de classificadores baseada em regras simples

	Automóveis	Pessoas	Animais	Motos	DVD
Automóveis	150	9	2	7	0
Pessoas	19	38	5	38	2
Animais	4	14	212	11	3
Motos	1	0	1	113	1
CD/DVD	0	0	1	1	80

7.4. Uso de Regras Geradas a Partir do Algoritmo C4.5

Para avançar nestes resultados sem perder a generalização foi necessário utilizar algoritmos que buscam padrões baseados em regras. Novas regras foram incluídas no conjunto de regras para conseguirmos um melhor desempenho da camada de decisão.

Primeiramente foi criado um programa de geração de dados a partir dos resultados obtidos pelos classificadores de imagens, de texto e pela combinação de ambos usando a regra da

multiplicação. Em seguida, as informações das saídas dos classificadores alimentaram o algoritmo C4.5 que gerou um conjunto de regras. As regras geradas foram então testadas individualmente sendo selecionados os melhores e os mais relevantes para compor o conjunto de regras final.

Para efetuar o teste individual, um conjunto de amostras já processadas pelos passos anteriores são testados em cada uma das regras. As regras eliminadas eram as que não tinham a maioria de suas respostas corretas. Algumas regras foram combinadas com outras através de um processo empírico com base em observações. No final apenas um conjunto reduzido de regras não foram eliminadas e fizeram parte da árvore de regras.

Para que fosse possível a utilização de regras neste tipo de problema, fez-se necessário a definição de alguns parâmetros com base nas probabilidades e posicionamento. As probabilidades apresentam um número de ponto flutuante que varia de zero a um. O posicionamento é uma indicação de alto nível que representa a classe vencedora sob o ponto de vista da cada classificador individualmente e da combinação de classificadores. Todos estes parâmetros buscam o valor adequado baseado em sua posição no ranking de probabilidades. Na Tabela 7.6 são apresentados os parâmetros usados para formar as regras que fazem parte da árvore de regras nesta camada de decisão.

Tabela 7.6. Parâmetros para uso com as regras

Operador	***Descrição***
txt	Representa o classificador textual. Este operador deve ser usado em conjunto com mais algum atributo, que deve vir separado com "_"(underscore);
img	Representa o classificador neural (imagens) da mesma forma que o classificador textual;
com	Representa a combinação, descrita anteriormente. Funciona da mesma forma que os classificadores;
txt_prob_maior, img_prob_maior ou com_prob_maior	Retorna um número de ponto flutuante que representa a probabilidade campeã normalizada do classificador (a maior probabilidade). Podem ser feitas comparações entre os classificadores, pois os resultados são compatíveis;
txt_prob_pen, img_prob_pen ou com_prob_ pen	Retorna um número de ponto flutuante que representa a penúltima probabilidade campeã do classificador (a segunda maior probabilidade).
txt_prob_ant, img_prob_ant ou com_prob_ ant	Retorna um número de ponto flutuante que representa a antepenúltima probabilidade campeã do classificador (a terceira maior probabilidade).
txt_maior, img_maior ou com_maior	Retorna a classe (número inteiro) que representa a classe campeã da classificação. Ou seja, a que tem a probabilidade maior.
txt_pen, img_pen ou com_pen	Retorna a classe (número inteiro) que representa a penúltima classe campeã (a segunda de maior probabilidade). o sufixo "pen"

Operador	*Descrição*
	representa “penúltimo”.
txt_ant, img_ant ou com_ant	Retorna a classe (número inteiro) que representa a antepenúltima classe campeã (a terceira de maior probabilidade).

Com base nos parâmetros da Tabela 7.6, podem ser construídas diversas regras que permitem o uso de operadores de comparação e operadores lógicos. A Tabela 7.7 apresenta diversos exemplos de como estas regras podem ser criadas com o uso dos parâmetros definidos na Tabela 7.6. Todos os exemplos apresentados são válidos, mas são apenas exemplos e podem não compor a árvore final do experimento.

Tabela 7.7. Exemplo comparações válidas

Regra	*Descrição*
txt_maior == img_maior	Se a classe campeã do classificador textual é a mesma da classe campeã do classificador neural
txt_prob_maior > 0.9	Se a classe campeã do classificador textual é maior que 90%
img_pen == txt_ant	Se a classe penúltima (dai vem a terminação “pen”) da imagem(que tem a segunda maior probabilidade) é igual a antepenúltima classe(terceira maior classificação) do classificador textual.
(img_prob_pen > com_prob_ant) && ((txt_prob_maior > 0.785535) OR (com_maior == txt_maior))	Operadores AND ou OR também são aceitos. Várias combinações são possíveis, mas deve-se usar os atributos comparando sempre de acordo com seu retorno. Não pode ser comparado atributos que retornam a classe com atributos que retornam probabilidades.
(img_maior == 1) AND (txt_ant==3) AND (com_prob_pen < 0.325632)	Constantes podem e foram usadas. O problema no uso de constantes de ponto flutuante está em se estabeler valores para eles.

7.5. Regras que Integram a Camada de Decisão

Para compor o rol de regras de decisão final, utilizaram-se as regras apresentadas, anteriormente (tópico 7.3), regras geradas a partir da observação dos dados e também regras geradas com o auxílio do algoritmo C4.5. Com base na definição dos parâmetros da Tabela 7.6 e através do conjunto vinculado de treinamento, foram gerados alguns conjuntos de dados para que o algoritmo do C4.5 pudesse gerar um conjunto de regras. Todas as regras geradas foram avaliadas de forma

individual e posteriormente de forma integrada. Apenas as regras relevantes que atuaram de forma genérica foram usadas para compor o conjunto de regras final. Também foram feitos testes combinando regras encontradas pelo algoritmo C4.5 e modificadas (as regras) através de esforços heurísticos com base em observação e lógica. O resultado final é um conjunto de regras de decisão que usou como ferramenta auxiliar uma implementação do algoritmo C4.5 com várias modificações nas regras. Regras que não forneceram uma boa taxa de acerto e generalizações foram eliminadas.

A figura 7.3 ilustra o fluxo no processamento das regras. O processamento inicia na regra de número 1 e em caso de resultado positivo o fluxo segue para a direita em caso de resultado negativo segue para baixo. Cada nó contém uma regra especificada na Tabela 7.8.

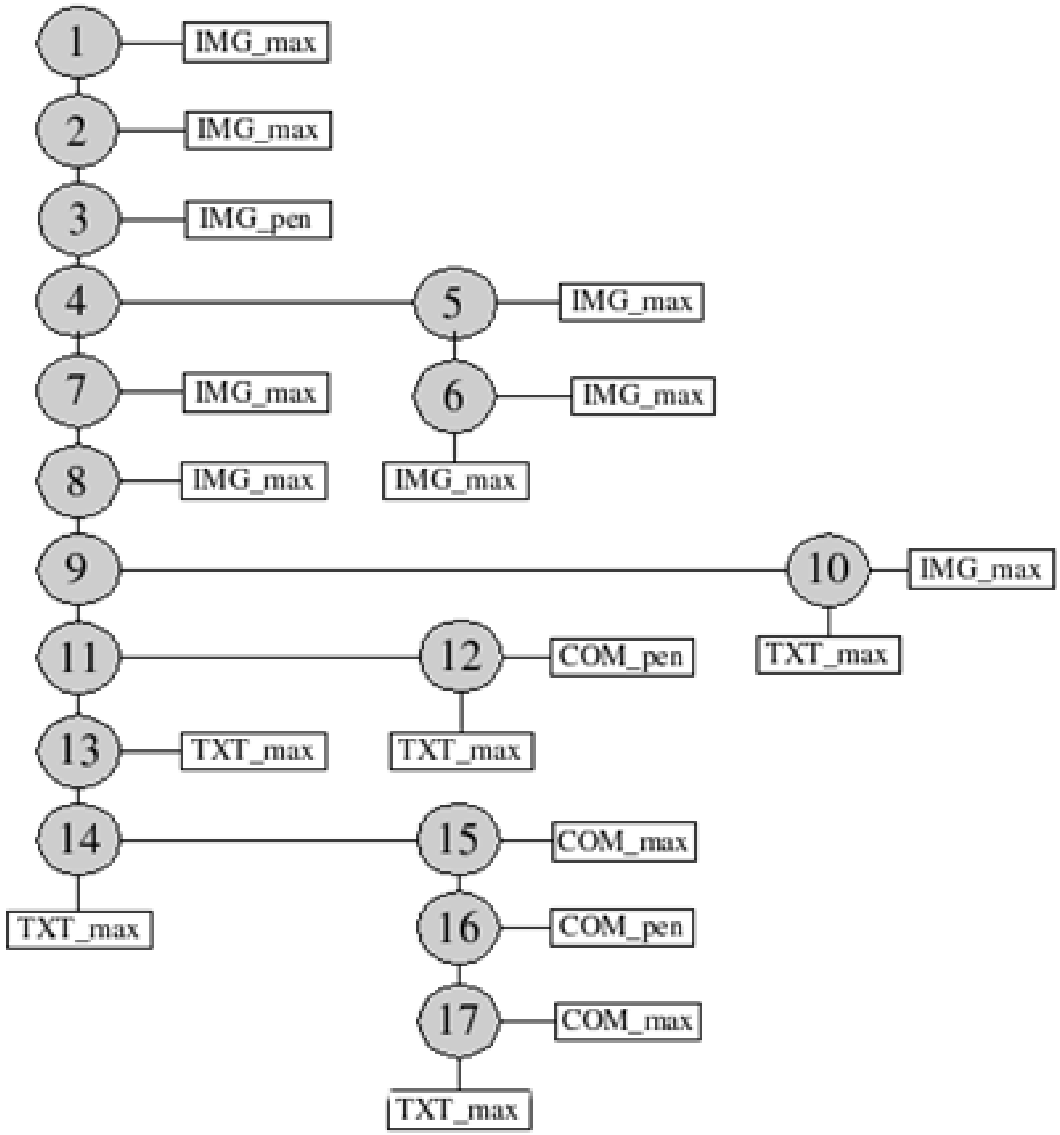


Figura 7.3 – Fluxo do Conjunto Final de Regras

Tabela 7.8. Regras Usadas na Árvore de Decisão

Cod.	*Regra*
1	img_maior != txt_maior && img_maior == com_maior && img_prob_maior > 0.5 && txt_pen == img_maior && txt_prob_maior > 0.203 && com_prob_maior > 0.3 && img_maior == 2
2	img_maior == txt_maior && img_prob_pen <=0.203951 && img_prob_pen <=0.203119)
3	img_maior != txt_maior && img_maior != txt_pen && img_pen == com_pen && img_prob_maior <= 0.935797 && txt_prob_pen > 0.204466 && com_prob_pen > 0.200631
4	img_maior == txt_maior && img_prob_pen > 0.203951 && com_prob_maior <= 0.50439 && com_prob_ant > 0.099557
5	img_maior == 1
6	img_maior == 4
7	img_maior != txt_maior && img_maior == txt_pen && img_prob_maior <= 0.935797 && img_prob_pen <= 0.206093 && txt_prob_pen > 0.202701
8	img_prob_max >= 0.79 && com_prob_max = 0.49 && (txt_pen == img_maior || txt_ant == img_maior)
9	txt_maior == 0
10	img_prob_max > 0.5 && txt_pen == img_maior
11	txt_maior == 1
12	com_pen == 1

Cod.	*Regra*
13	txt_maior == 2
14	txt_maior == 4
15	com_maior == 1
16	com_pen == 1
17	(com_pen == 0 || com_pen==4)

Com a camada de decisão definida e pronta, foram efetuados testes para avaliar o desempenho desta camada. A base de dados utilizada nestes testes foi o conjunto vinculado usado em treinamento, portanto os resultados apresentados abaixo devem ser analisados com todo cuidado para não haver erros de interpretação. Os resultados obtidos a partir deste teste podem ser observados na Tabela 7.9 e sua respectiva matriz de confusão na Tabela 7.10.

Tabela 7.9. Resultado a partir da camada de decisão

Acertos	629
Erros	83
Total de Amostras	712
% de Acerto	88,34%

Tabela 7.10. Matriz de confusão da classificação de imagens utilizando combinação de classificadores e a camada de decisão

	Automóveis	Pessoas	Animais	Motos	DVD
Automóveis	158	7	1	2	0
Pessoas	15	54	10	21	2
Animais	0	13	225	6	0
Motos	0	1	2	112	1
CD/DVD	0	0	2	0	80

Uma comparação deste teste em relação à classificação usando somente o classificador de imagens apresenta um ganho de 14,88%, o que pode ser considerado um ganho significativo. O resultado é apresentado na Tabela 7.11.

Tabela 7.11. Comparação de resultados do classificador de imagens e da combinação de classificadores utilizando a camada de decisão

	Classificador de Imagens	Classificação Usando a Camada de Decisão	Diferença
Acertos	523	629	106
Erros	189	83	-106
Total de Amostras	712	712	
% de Acerto	73,46%	88,34%	14,88%

Com a camada de decisão concluída podemos gerar os resultados finais para avaliar a hipótese de que informações contextuais são realmente relevantes para a classificação de imagens. A geração destes resultados finais será apresentada no próximo capítulo, que ao contrário deste e dos capítulos anteriores, utiliza somente informações referentes às imagens não usadas em nenhuma parte do treinamento, portanto a capacidade de generalização do experimento também será avaliada juntamente com as taxas de classificação obtidas com o uso desta combinação.

Capítulo 8

Resultados Finais

Neste capítulo são apresentados os resultados finais obtidos com um conjunto de amostras de imagens não envolvidas em qualquer fase de treinamento. O experimento tem por objetivo comprovar o ganho de desempenho de um procedimento de classificação de imagens quando se utilizam informações contextuais associadas a origem da imagem. Neste caso, o ambiente utilizado foi a Internet, sendo a base de dados extraída do próprio ambiente original, simulando, assim, uma situação real.

Os testes apresentados neste capítulo foram realizados de maneira integral, incluindo todos os passos descritos neste documento com os algoritmos todos prontos e treinados. A diferença em relação aos testes feitos nos capítulos anteriores, é que estes tiveram uma base própria para treinamento, validação e testes, diferente das informações usadas neste capítulo.

8.1. Conjunto de Teste

O conjunto utilizado para testar o método proposto é composto por 821 amostras de imagens com um mínimo de cem amostras por classe. Estas imagens também contêm vínculos com a página de origem, sendo isto necessário para a realização do processo integral. Nenhuma imagem deste conjunto participou de qualquer fase de treinamento da rede neural para que os resultados não tivessem qualquer tendência.

Este conjunto foi gerado a partir de um processo de escolha na base de dados que definiu qual seria o conjunto pertencente para cada amostra. Para a distribuição das informações na base de dados, este processo navegou através da seqüência de obtenção das imagens originais, atribuindo uma imagem para cada um dos conjuntos, que tinha que ser usado durante as outras fases, e um para

o conjunto que seria usado para o teste final, demonstrado aqui. Foram usadas todas as imagens que continham vínculos com textos também rotulados. Desta forma, este conjunto de imagens possui os mesmos ruídos existentes nas imagens que foram usadas na fase de treinamento e possuem a distribuição também idêntica. Este conjunto de informações contendo imagens, textos e informações vinculadas foi aqui denominado de "conjunto final", por tratar-se de informações usadas somente para os testes finais do método proposto.

8.2. Resultados

Primeiramente as imagens do conjunto final foram submetidas à rede neural para avaliação do desempenho deste conjunto na classificação de imagens convencional. A matriz de confusão é apresentada na Tabela 8.1 enquanto as taxas de acerto são apresentadas na Tabela 8.2.

Tabela 8.1. Matriz de confusão do conjunto de teste final

Classes	Automóveis	Pessoas	Animais	Motos	CD/DVD
Automóveis	143	3	4	24	4
Pessoas	12	34	25	25	23
Animais	20	25	187	23	28
Motos	10	1	14	110	6
CD/DVD	2	3	5	5	85

Tabela 8.2. Resultado da Classificação de Imagens

Classe	Acertos	Erros	Total	% de acerto
Automóveis	143	35	178	80,34%
Pessoas	34	85	119	28,58%
Animais	187	96	283	66,08%
Motos	110	31	141	78,01%
CD/DVD	85	15	100	85,00%
Total	559	262	821	68,09%

O índice de classificação obtido foi de 68,09% para a classificação de imagens pelo modo convencional, ou seja, com uma rede neural do tipo MLP devidamente treinada atuando com base em conjuntos de características (formas, cores e textura), para cinco tipos de classes, com uma base de dados bastante ruidosa.

O próximo passo é submeter este mesmo conjunto de informações para a combinação de

classificadores que utiliza as informações contextuais. Desta forma, juntamente com as imagens, estamos submetendo o conjunto de informações contextuais que o algoritmo de combinação necessita. As amostras de imagens são exatamente as mesmas submetidas ao classificador de imagens baseado em redes neurais.

Este conjunto final foi então processado usando a combinação de classificadores e regras de decisão. A matriz de confusão é apresentada na Tabela 8.3. O resultado do experimento é apresentado na Tabela 8.4.

Tabela 8.3. Matriz de confusão do conjunto de teste final

Classes	Automóveis	Pessoas	Animais	Motos	CD/DVD
Automóveis	160	14	2	2	0
Pessoas	25	55	13	18	8
Animais	0	8	265	7	3
Motos	2	9	7	123	0
CD/DVD	3	1	1	0	95

Tabela 8.4. Resultado da classificação de imagens usando combinação de classificadores e regras de decisão

Classe	Acertos	Erros	Total	% de acerto
Automóveis	160	18	178	89,89%
Pessoas	55	64	119	46,22%
Animais	265	18	283	96,64%
Motos	123	18	141	87,23%
CD/DVD	95	5	100	95,00%
Total	698	123	821	85,02%

Pode-se perceber que o índice de classificação para a combinação de classificadores e regras de decisão ficou em 85,02%, ou seja, um aumento de 16,93% em taxa de classificação correta. Na Tabela 8.5 são detalhados as comparações entre as duas formas de classificação: a convencional (1) e a combinação de classificadores (2).

Com base na comparação individual pode-se notar que a maior diferença ocorreu na classe referente a “animais domésticos”, com 30,56% de diferença entre o método convencional e o método que considera a informação contextual. A segunda maior diferença ficou por conta da classe “pessoas” que foi de 17,64%. O método convencional, nestas circunstâncias, se mostrou muito ruim

para a classe “pessoas”, pois somente 28,58% foram classificados corretamente. A evolução da taxa de classificação demonstra que o método proposto foi realmente relevante para resolver este problema de classificação de imagens.

Tabela 8.5. Comparação dos resultados

Classes	*Acerto -1*	*Acerto-2*	*Acerto-Dif.*	*% - 1*	*% - 2*	*% dif.*
Automóveis	143	160	17	80,34%	89,89%	9,55%
Pessoas	34	55	21	28,58%	46,22%	17,64%
Animais	187	265	78	66,08%	96,64%	30,56%
Motos	110	123	13	78,01%	87,23%	9,22%
CD/DVD	85	95	10	85,00%	95,00%	10,00%
Total	559	698	139	68,09%	85,02%	16,93%

8.3. Análise dos Resultados

Dado os resultados e comparados com o método tradicional, este tópico fará o detalhamento de cada etapa do método proposto e seus resultados individuais para o conjunto final de testes.

Durante o teste final, as amostras de imagens foram submetidas ao classificador neural cujos resultados já foram mostrados no tópico anterior. Os textos que estavam na origem da imagem foram submetidos ao classificador textual. A matriz de confusão do classificador de textos está apresentada na Tabela 8.6.

Tabela 8.6. Matriz de Confusão do classificador estatístico

Classes	Automóveis	Pessoas	Animais	Motos	CD/DVD
Automóveis	156	4	0	6	0
Pessoas	12	73	13	4	0
Animais	0	11	262	6	0
Motos	0	1	0	164	0
CD/DVD	4	3	0	2	100

No caso dos textos, a classificação é dada pelo assunto do texto e não pelas imagens que a compõe. Para o procedimento de rotulagem foram removidas as imagens e apresentado somente o texto extraído da página. O responsável pela rotulagem leu o texto e definiu a classe baseando-se

unicamente no conteúdo textual. Os detalhes deste procedimento foram detalhados no Capítulo 3.

Uma análise detalhada do classificador de textos é apresentada na Tabela 8.7. Pode-se notar um bom desempenho do classificador estatístico para classificação de textos, com 91,96% das amostras classificadas corretamente, apesar do alto nível de ruído destas amostras. O melhor índice de classificação ficou com a classe "motos" com 99,39% de suas amostras classificadas corretamente e o pior índice ficou com os textos referentes a "pessoas" com 71,57% de suas amostras classificadas corretamente. A classe "pessoas" obteve 29 erros em suas 102 amostras de texto rotuladas com esta classe. Uma análise na matriz de confusão da Tabela 8.6 demonstra que o erro é cometido da confusão com "automóveis" e de "animais domésticos". Isto demonstra que existem elementos comuns nestas classes que levam a esta confusão. Outro ponto é que a classe "pessoas" possui o menor número de amostras avaliadas, se comparado às outras classes. É importante salientar que a fase de treinamento foi executada com quantidades iguais para cada classe, conforme detalhado no Capítulo 3 (base de dados) e Capítulo 7 (classificador estatístico).

Tabela 8.7. Resultado da classificação de textos

Classe	Acertos	Erros	Total	% de acerto
Automóveis	156	10	166	93,98%
Pessoas	73	29	102	71,57%
Animais	262	17	279	93,91%
Motos	164	1	165	99,39%
CD/DVD	100	9	109	91,74%
Total	755	66	821	91,96%

Após o processamento do classificador neural e do classificador estatístico, os resultados são submetidos à camada de decisão, detalhada no Capítulo anterior. Primeiramente é efetuada uma união dos resultados dos classificadores através de um operador de multiplicação seguido de normalização, gerando um terceiro resultado que representa a combinação dos classificadores de imagens e texto. O resultado da combinação e os resultados dos classificadores são usados nas regras de decisão, detalhada no capítulo anterior.

As regras foram processadas neste conjunto final de testes e a forma de uso e acerto de cada regra está detalhado na Tabela 8.8. As regras apresentadas nesta tabela estão definidas no capítulo anterior (camada de decisão). Cada regra apresentada utilizou-se de informações dos classificadores de imagens e textuais e também da combinação, para testar uma condição lógica de resposta booleana.

Tabela 8.8. Uso e erro das regras da camada de decisão

Regra	*Itens processados*	*Erros*	*% de uso da regra*	*% de acerto*
1	17	13	2,07%	23,53%
2	379	12	46,16%	96,83%
3	7	4	0,85%	42,86%
4 (nó)	16 + 25 + 112	2 + 6 + 7		
5	16	2	1,95%	87,50%
6	25	6	3,05%	76,00%
6 (falso)	112	7	13,64%	93,75%
7	8	3	0,97%	62,50%
8	29	14	3,53%	51,72%
9 (nó)	3 + 40	1 + 16		
10	3	1	0,37%	66,67%
10 (falso)	40	16	4,87%	60,00%
11 (nó)	13 + 26	2 + 17		
12	13	2	1,58%	84,62%
12 (falso)	26	17	3,17%	34,62%
13	85	4	10,35%	95,29%
14 (nó)	6 + 11 + 18 + 26	1 + 6 + 9 + 6		
15	6	1	0,73%	83,33%
16	11	6	1,34%	45,45%
17	18	9	2,19%	50,00%
17 (falso)	26	6	3,17%	76,92%

A Tabela 8.8 apresenta a relevância de cada regra para o conjunto de informações e a taxa de acerto por regra. Na primeira coluna está o número da regra, sendo que algumas regras apenas direcionam para outras regras, como é o caso das regras 4 e 9, por exemplo. A maioria das regras que não direcionam para outras regras apontam para alguma classificação caso seja positivo, e outras regras apontam para alguma classificação em caso negativo, a exemplo das regras 6 e 10 com a marcação "falso". A coluna intitulada "Itens Processados" apresenta a quantidade de amostras que a regra definiu como verdadeiro (ou como falso no caso da marcação "falso" na coluna regra). A coluna erro apresenta quantas amostras foram erroneamente classificadas pela regra em questão.

Nas últimas colunas são calculados o percentual do uso da regra, que demonstra a relevância da regra no conjunto de amostras de teste. A última coluna apresenta o percentual de acerto com base no cálculo de itens processados e erros.

Esta tabela deixa claro que a regra dois foi a mais usada e é a mais eficiente deste conjunto de regras, classificando 46,16% do total de amostras submetidas à camada de decisão com uma taxa de acerto de 96,83%.

Capítulo 9

Conclusão

Este trabalho confirmou a hipótese de que informações contextuais são relevantes para melhorar a taxa de classificação correta de imagens, pois as imagens que foram objeto de classificação foram extraídas de um ambiente que contém outras informações sobre o possível conteúdo destas imagens. Em particular, o ambiente considerado foram páginas *web* e assim, utilizou-se o próprio texto das páginas *web* de origem.

Este trabalho apresentou pontos fortes como a confirmação da hipótese, a criação da base de dados, o bom desempenho do classificador de textos e o significativo ganho de desempenho final para as imagens. A base de dados foi construída com interação humana reduzida e com um número de amostras suficiente para os conjuntos usados nos classificadores. Estes classificadores foram treinados e testados individualmente e posteriormente testados em conjunto, sem o uso das imagens usadas no treinamento.

A rede neural, individualmente, apresentou um desempenho relativamente baixo, porém compatível com a desigualdade das amostras de imagens, pois estas representavam um ambiente real, com imagens não preparadas. O gasto computacional com os algoritmos de extração de características consumiu muito tempo, e a rede neural com apenas a característica baseada em textura apresentou um desempenho baixo. As características extraídas poderiam ser melhoradas, ou mesmo incluído outras características para a melhora do classificador neural.

Comparando com outros trabalhos de classificação de imagens, este trabalho necessita de maiores esforços e tempo, porém possibilita um melhor desempenho. O número de amostras envolvidas neste trabalho é relativamente maior comparado à maioria dos trabalhos de classificação de imagens. Também existem poucos trabalhos que formaram sua própria base de dados, com um número relativamente alto de amostras. A evolução deste trabalho pode facilitar ainda mais a

formação de bases de dados de maneira semi-automática para outras pesquisas.

9.1. Resumo dos Resultados

Em todas as fases deste trabalho foram feitos testes individuais demonstrando o desempenho de cada parte. Os resultados individuais foram detalhados nos seus respectivos capítulos e no caso dos testes gerais/finais, os resultados foram detalhados no Capítulo 8. O objetivo deste trabalho foi comparar a classificação de imagens usando um método convencional, neste caso um classificador baseado em rede neural, com um método que utiliza informações contextuais, envolvendo vários classificadores, combinação destes e uma camada de decisão, que neste caso utiliza um conjunto de regras.

O classificador baseado em redes neurais classificou corretamente 68,09% das amostras. Isto corresponde a 559 amostras de um total de 821. Estes números correspondem a apenas amostras de imagens do conjunto final de testes, sendo que muitas outras amostras foram usadas em fases de treinamento e, por isso, não fazem parte do conjunto final.

A classificação utilizando combinação de classificadores elevou a taxa de classificação para 85,02%, correspondendo a 698 amostras de um total de 821, o que representa um aumento de 16,93% comparado ao classificador baseado em redes neurais. Este resultado demonstra a relevância das informações contextuais.

Com relação a base de dados, foram encontrados 407.758 *hyperlinks* para páginas *web* a partir de uma pequena lista de *hyperlinks* iniciais. Foram baixadas e processadas 119.276 imagens, sendo que apenas 11.809 passaram pela fase de pré-seleção e, finalmente, 5.169 puderam ser rotuladas para uma das classes alvo.

9.2. Contribuições

Este trabalho demonstrou como as informações contextuais, obtidas na origem da imagem, são importantes e contribuem significativamente para a classificação de imagens. A partir desta afirmação é possível pensar em estratégias que utilizam informações diversas ao lugar de analisar apenas o objeto alvo.

Aplicações como um supervisor de conteúdo para páginas *web* poderia implementar um

algoritmo similar ao proposto neste trabalho, proporcionando uma boa eficiência, pois não está limitado a verificação de meta-informação nas páginas HTML, restringindo assim a visualização de imagens impróprias, devendo ser acoplado ao navegador Internet ou leitor de *e-mail*.

Em trabalhos científicos é muito comum o uso de bases de dados de imagens. Porém, muitas vezes isto pode representar um grande problema, pois as bases comerciais têm diversas limitações, sendo muitas vezes necessário que seja formado a própria base de dados, o que pode levar um tempo considerável por parte do pesquisador. Em alguns casos, os trabalhos podem ter seus resultados comprometidos por não haver quantidade de amostras suficientes para comprovar os resultados. Com o método apresentado neste trabalho ficou comprovado que é possível formar bases de dados de imagens a partir da Internet. Foi um processo com custo relativamente menor que bases de dados comerciais, possibilitando formar bases de dados de imagens não encontradas em bases de dados comerciais.

Também são contribuições deste trabalho as formas como as características foram extraídas para formar o vetor de características. As três características foram testadas isoladamente, o que pode ser usado para analisar o desempenho de cada uma delas. A rede neural foi também testada em situações onde o número de amostras foi igual para cada classe e também em situações onde variou o número de amostras por classe. Esta diferença ocasionou respostas diferentes que podem ser usadas num trabalho que analise comportamentos de redes neurais.

O classificador Naïve Bayes também foi testado para a classificação de textos numa situação diferente do convencional, que é a classificação com textos de línguas diferentes. Neste trabalho ficou comprovado que mesmo com esta característica, o classificador teve um bom desempenho.

9.3. Trabalhos Futuros

Para atingir melhorias consideráveis neste tipo de problema, poderiam ser melhoradas as extrações de características das imagens, modificando o vetor de características na entrada da rede neural, o que proporcionaria um desempenho melhor no conjunto. Outra forma seria acoplar mais classificadores distintos analisando a imagem ou as informações contextuais, desde que a saída seja compatível com os outros classificadores ou então a camada de decisão interpretasse os resultados destes classificadores como novas regras, sem a necessidade da combinação direta. Estas novas regras seriam usadas verificando-se apenas a qual classe o novo classificador atribui a imagem (ou às outras informações).

Outro ponto de melhora seria na extração de informações contextuais, juntamente com o texto, poderia haver alguns itens com maior poder de influência, como os trechos de texto que referenciam a imagem, caso estes consigam ser detectados. Ao lugar da página HTML poderia também ser classificado o *site* (conjunto de páginas HTML), através de análise dos *hyperlinks.*

Uma mudança na camada de decisão poderia incluir um grau de certeza na classificação final da imagem, com base no caminho percorrido nas regras, a classificação seria acompanhada de um percentual de certeza da classificação.

As melhorias também poderiam ser focadas no processo de obtenção de amostras para a base de dados, agilizando o processo e diminuindo a interação humana. A partir de um processo de classificação como o abordado neste trabalho, poderia ser iniciada a captura de imagens na internet e as imagens extraídas seriam submetidas ao classificador no momento da captura. Desta forma as amostras seriam agrupadas por classes, facilitando o trabalho da inspeção humana durante a rotulagem. Para os textos poderiam ser inseridos processos de pré-seleção como foi feito nas imagens, este processo poderia efetuar a pré-seleção estabelecendo um mínimo de palavras encontradas no vocabulário preliminar. O idioma também poderia ser reconhecido e agrupado, facilitando no momento da rotulagem.

Inúmeros outros experimentos seriam possíveis com base neste trabalho, que afirma a questão de que devemos olhar em volta tentando integrar soluções conhecidas e não apenas tentando criar algoritmos milagrosos analisando sempre as mesmas coisas. Enfim são necessários mais do que ciência para ajudar nestes problemas, conforme afirma Gonzalez *et. al.* [GON00], que finaliza sua obra, dizendo: "*Para o futuro próximo, o projeto de sistemas de análise de imagens continuará a requerer uma mistura de arte e ciência*".

Referências Bibliográficas

[AGA95] AGARWAL, A.; GRANOWETTER, L.; HUSSEIN, K.; GUPTA, A. *Detection of Courtesy Amount Block on Bank Checks*; Montreal; IEEE ICDAR'95, p.748-751; 1995.

[ALB00] ALBUQUERQUE, MÁRCIO PONTES; ALBUQUERQUE, MARCELO PONTES; *Processamento de Imagens: Métodos e Análises*; Revista de Ciência e Tecnologia; ISSN 1519-8022; vol1 – n.1 – pg.10-22; 2000.

[BAH00] BAHLER, DENNIS; NAVARRO, LAURA; *Methods for Combining Heterogeneous Sets of Classifiers*; *17th Natl. Conf. on Artificial Intelligence (AAAI 2000),* Workshop on New Research Problems for Machine Learning; Tecnical report; 2000.

[BIT00] BITTERNCOURT, JOÃO RICARDO; OSÓRIO, FERNANDO SANTOS; *Aplicações de Técnicas de Inteligência Artificial no Processamento de Imagens*; Centro de Ciências Exatas e Tecnológicas; São Leopoldo; 2000.

[CHE96] CHENG, ISAAC K.; EDWARDS, DOUGLAS D.; *Content-Based Image Classification: Recognizing Wildlife and Natural Scenes by Color and Texture*; 1996.

[CIR03] CIRELO, MARCELO C.; COZMAN, FABIO G.; *Aprendizado de Semi-Supervisionado de Classificadores Bayesianos Utilizando Testes de Independência*; XXIII Congresso da Sociedade Brasileira de Computação; 2003.

[COR04] CORBIS; *Stock photography and digital pictures*. http://www.corbis.com.br. Acesso em 15 de fevereiro de 2004.

[DON02] DONG, SHOU-BIN; YANG, YI-MING; *Hierarchical Web Image Classification by Multi-Level Features;* First International Conference on Machine Learning and Cybernetics, Beijing; 4-5 November; Cap. 2; pag. 663-668; 2002.

[DUD00] DUDA, RICHARD O.; HART, PETER E.; STORK, DAVID G.; *Pattern Classification (2nd Edition)*; Wiley-Interscience; 654 pages; ISBN 0471056693; October; 2000.

[FEN04] FENG, HUAMIN; SHI, RUI; CHUA, TAT-SENG; *A Bootstrapping Framework for Annotating and Retrieving WWW Images*; International Multimedia Conference; Proceedings of the 12th annual ACM international conference on Multimedia; ISBN 1581138938; Pages 960-967; New York, NY, USA; 2004;

[FEM01] FERRO, M; LEE, H. D; CHUNG, W. F. *Aplicação de técnicas de aprendizado de máquina para extração de conhecimento e construção de classificadores: estudo de caso de bases de dados médicas*; Primeira Jornada Científica da Unioeste; 24-26 Outubro; 2001.

[FER01] FERRO, MARIZA; LEE, HUEI DIANA; *O processo de KDD – Knowledge Discovery in Database para Aplicações na Medicina*; Primeira Jornada Científica da Unioeste; 24-26 Outubro; 2001.

[FLE96] FLEXER, ARTHUR; *Statical Evaluation of Neural Network Experiments: Minimum Requirements and Current Pratice*; Cybernetics and Systems '96, Proceedings of the 13th European Meeting on Cybernetics and Systems Research. Austrian Society for Cybernetic Studies, Vienna, 2 vols., pp.1005- 1008, 1996.

[GOL02] GOLLER, CHRISTOPH; LÖNING, JOACHIM; WILL, THILO; WOLF, WENER; *Automatic Document Classification: A thorough Evaluation of varios Methods;* SAIL-LABS, Speech and Artificial Intelligence Labs, Germany; iXEC, Executive Information Systems GmbH, Germany; 2002.

[GON00] GONZALEZ, RAFAEL C.; WOODS, RICHARD E.; *Processamento de Imagens Digitais*, Editora Edgard Blucher Ltda; 509 páginas; ISBN 8521202644; 2000.

[GON94] GONG, Y., ZHANG, H., CHUAN, H.C., SAKAUCHI, M. *An image database system with content capturing and fast image indexing abilitie*s. In: Proc. IEEE International Conference on Multimedia Computing and Systems; Boston, MA, USA; pages 121-130; ISBN 0818655305; 1994.

[HAY01] HAYKIN, SIMON; *Redes Neurais, Princípios e prática*; 2ª Edição; ISBN 8573077182; 900 páginas; Editora Bookman; 2001.

[HIR99] HIRATA, KYOJI; MUKHERJEA, SOUGATA; LI, WEN-SYAN; HARA, YOSHINORI; *Integrating Image Matching and Classification for Multimedia Retrieval on the Web;* C&C Research Laboratory, NEC USA, California, IEEE; 0-7695-0253-9/99; 1999.

[HU04] HU, JIANYING; BAGGA, AMIT; *Categorizing Images in Web Documents*; IBM T.J. Watson Res. Center, Hawthorne, NY, USA; IEEE Multimedia; Volume: 11, Issue: 1; Pages 22- 30; 2004.

[JAI95] JAIN, ANIL K.; VAILAYA, ADITYA; *Image Retrieval using Color and Shape*; Pattern Recognition; vol. 29, no. 8, pp. 1,233-1,244; 1996.

[KHE04] KHERFI, M. L.; ZIOU, D.; BERNARDI, A.; *Image Retrieval From the World Wide Web: Issues, Tecniques, and Systems;* ACM Computing Surveys; Vol. 36; No.1; March 2004; pp. 35-67; 2004.

[KIT98] KITTLER, JOSEF; HATEF, MOHAMAD; DUIN, ROBERT P.W.; MATAS, JIRI; *On Combining Classifiers*; IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 20, No 3, March 1998.

[KOE03] KOERICH, ALESSANDRO L.;*Unconstrained Handwritten Character Recognition Using Different Classification Strategies*; IAPR-TC3 International Workshop on Artificial Neural Networks in Pattern Recognition; Florence, Italy; 5 pages; September 12-13, 2003.

[KOV02] KOVÁCS, ZSOLT L.; *Redes Neurais Artificiais – Fundamentos e aplicações*; 3ª. Edição, Editora Livraria da Física; 174 páginas; ISBN 8588325144; 2002.

[KUY99] KUYEL, TURKEY; GEISLER, WILSON; GHOSH, JOYDEEP; *Fast Image Classification Using a Sequence of Visual Fixations;* IEEE transactions on systems, Man, and cybernetics. Part B: Cybernetics, Vol. 29, No. 2, April, 1999.

[LAN92] LANGLEY, PAT; IBA, WAYNE; THOMPSON, KEVIN; *An Analysis of Baysian Classifiers*; Proceedings of the Tenth National Conference on Artificial Intelligence (pp. 223--228). San Jose, CA: AAAI Press; 1992.

[LAS01] LASCA, VIVIAN BELLINI; BORGES, DÍBIO LEANDRO; *Recognizing faces with minimum information;* Laboratory for Image and Vision Science; Pontifical Catholic University of Parana; 2001.

[LEP03] LEPISTÖ, LEENA; KUNTTU, LIVARI; AUTIO, JORMA; VISA, ARI; *Rock Image Classification Using Non-Homogenous Textures and Spectral Imaging;* Tampere University of Tecnology; Institute of Signal Processing, Finland; Saanio & Riekkola Consulting Engineers;Laulukuja, Finland; WSCG'2003, February 3-7, 2003, Plzen, Czech Republic. 2003;

[LON00] LONG, HUI ZHONG; LEOW, WEE KHENG; *Perceptual Texture Space for Content-Based Image Retrieval;* School of Computing, National University of Singapore; MMM2000; World Scientific, August 28, 2000.

[LU01] LU, CHENG; DREW, MARK S.; *Construction of a Hierarchical Classifier Schema using a Combination of Text-Based and Image-Based Approaches*; SIGIR'01, September 09-12, 2001, New Orleans, Lousiana, USA; ACM 1-58113-331-6/01/0009; 2001.

[LUO00] LUO, XULING; MIRCHANDANI, GAGAN; *An Integrated Framework For Image Classification*; Department of Eletrical and Computer engineering; The University of Vermount; IEEE; 2000.

[MEL97] MELLISH, CHRIS*; Machine Learning;* University of Edinburgh ;Department of Artificial Inteligence;Outline Lecture Notes, Spring Term 1997.

[MIC94] MICHIE, D; SPIEGEELHATER, D. J; TAYOR, C.C; *Machine Learning, Neural and Statistical Classification*; Englewood Cliffs, N.J. : Prentice Hall, 1994.

[MIT97] MITCHELL , T; *Machine Learning;* McGraw-Hill Science/Engineering/Math (March 1, 1997); ISBN 0070428077; 432 pages; 1997.

[NAD93] NADLER, MORTON; SMITH, ERIC P.;Pattern Recognition Engineering; John Wiley & sons inc. New York ,ISBN 0471622931; 1993.

[NIS04] NIST; *Scientific and Technical Databases*; National Institute of Standards and Technology; http://www.nist.gov. Acesso em 02 de fevereiro de 2004.

[OLI02] Oliveira, Camillo Jorge Santos; Araujo, Arnaldo de Albuquerque; Severiano Jr.,Carlos Alberto; Gomes, Daniel Ribeiro; *Classifying Images Collected on the World Wide Web*; p. 327, XV 2002.

[PAR99] PARK, JOUG-SEUNG; OH HWANG-SEOK;CHANG DUK-HO; Shape-matching approach to content-based image retrieval; Proc. SPIE Vol. 3846, p. 256-266, Multimedia Storage and Archiving Systems IV, Sethuraman Panchanathan; Shih-Fu Chang; C.-C. J. Kuo; Eds; 1999.

[PRA02] PRABHAKAR, SALIL; CHENG, HUI; HANDLEY, JOHN C.; FAN, ZHIGANG; LIN, YING-WEI; *Picture-Graphics Color Image Classification*; Digital Persona Inc.; Sarnoff Corp.; Xerox Corp.; IEEE II-785; IEEE ICIP 2002.

[QUI93] QUINLAN, J. ROSS; C4.5: *Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*; 302 pages; Morgan Kaufmann; ISBN 1558602380; January 15, 1993.

[RIC93] RICK, ELAINE; KNIGHT KEVIN; *Inteligência Artificial*; São Paulo; Makron Books, 1993.

[ROW02] ROWE, NEIL C.; *Marie-4: A Hight-Recall, Self-Improving Web Crawler That Finds Images Using Captions;* IEEE Intelligent Systems 17 (4); 1094-7167/02; pag 8-14; 2002.

[SHA01] LI, YI; SHAPIRO, LINDA G.; *Consistent Line Clusters for Building Recognition in CBIR*; ICPR02; 16th International Conference on Pattern Recognition; University of Washigton; 2001.

[SHI02] SHINMOTO, MASAYUKI; MITSUKURA, YASUE; FUKUMI, MINORU; AKAMATSU, NORIO; *Color Image Classification Using Neural Networks;* SICE02-0241; TEA11-7; SICE 2002 Aug. 5-7, 2002, Osaka; 2002.

[SHI97] SHI, JIANBO; MALIK, JITENDRA; *Normalized Cuts an Image Segmentation;* Computer Science Division; University of California at Berkeley, CA; Proc. Of the IEEE Conf. On Comp. Vision and Pattern Recognition, San Juan, Pueto Rico, June 1997.

[SIN02] SINOARA, ROBERTA A.; PUGLIESI, JAQUELINE B.; REZENDE, SOLANGE O.; *Combinação de classificadores no processo data mining*; Universidade de São Paulo; Revista de iniciação científica da Sociedade Brasileira de Computação; Edição de Março de 2002; Ano II Volume II Número I; ISSN 1519-8219; 2002.

[SNN03] SNNS; *Sturrgart Neural Network Simulator*; http://www-ra.informatik.uni-tuebingen.de/SNNS/. Acesso em 02 de novembro de 2003.

[STR96] STRICKER, M., A.DIMAI**.** *Color Indexing With Weak Spatial Constraints*, SPIE Proceedings 2670: 29-40, 1996.

[SUE00] SUEN, CHIN Y.;LAM, LOUISA; *Multiple Classifier Combination Methodologies for Different Output Levels*; 1st Int Workshop on Multiple Classifier Systems (MCS2000) Italy, June 21-23; Springer-Verlag; Pages: 52 – 66; ISBN 3540677046; 2000.

[TAN02] TANG, XIAOOU; WANG XIAOGANG; *Face Photo Recognition Using Sketch;* Department of Information Engineering; The Chinese University of Hong Kong; Shatin, Hong Kong; IEEE ICIP 2002; 0-7803-7622-6/02; I-257; IEEE; 2002.

[VAL98] VAILAYA, ADITYA; JAIN, ANIL; ZHANG, HONG JIANG; *On Image Classification: City vs. Landscape*; cbaivl; p. 3; IEEE; 1998.

[ZHO04] ZHOU BING, SHEN JUNYI, PENG QINKE. *A Novel Image Matching Algorithm based on Symmetrical Color-Spatial Feature.* Mini-Micro Systems.

[ZUB00] ZUBEN, FERNANDO J. VON; *Computação Evolutiva: Uma Abordagem Pragmática*; I Jornada de Estudos em Computação de Piracicaba e Região (1a JECOMP), 2000, Piracicaba.

Apêndice A

Estatísticas da Base de Dados

A Tabela A.1 apresenta o número de amostras que compõem a base de dados separadas por classe e considerando textos e imagens.

Tabela A.1 – Número de amostras por classe que compõe a base de dados

	Imagens	Textos
Automóveis	1.087	1.029
Pessoas	880	1.010
Animais	1.166	1.069
Motos	1.425	1.004
CD/DVD	847	1.057

Para conseguir chegar a estes números de amostras, foram capturados 407.758 *hyperlinks* para novas páginas, a partir da listagem de *hyperlinks* iniciais, que gerou 119.902 endereços de imagens, onde apenas 626 não foram baixadas. Destas imagens apenas 11.809 passaram pela pré-seleção, sendo que ao final, somente 5.169 foram rotuladas.

Apêndice B

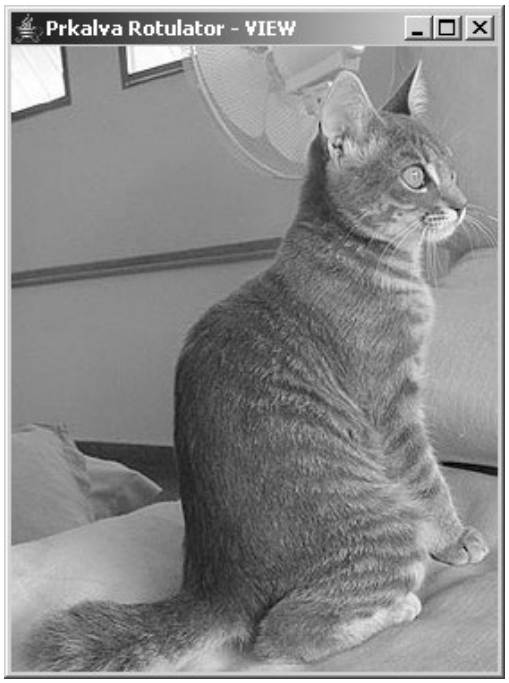
Amostras da Base de Dados - Automóveis

Apêndice C

Amostras da Base de Dados – Pessoas

Apêndice D

Amostras da Base de Dados - Animais



Apêndice E

Amostras da Base de Dados – Motos

Apêndice F

Amostras da Base de Dados – CD's e DVD's

Apêndice G

Amostras Corretamente Classificadas (todas as classes)


Prkalva Rotula...
1759
1730
ALHAMBRA


Prkalva Rotulator - VIEW


Prkalva Rotulator - VIEW


Prkalva Rotulator - VIEW
INCA
INCA


Prkalva Rotulator -...
Vodafone


Prkalva Rotulator - VIEW
GOLDEN STAIRS RD
(MAUNGATUROTO 4km)
AUSER RD
Y.W.A.M.
GOLDEN STAIRS RD
(TAPUHA 12km)
MARERETU
9km
PAPAMOA GOLF CLUB
2km


Prkalva Rotulator...


Prkalva Rotulator...
MICHAEL DOUGLAS
KIRK DOUGLAS
ACONTECE NAS MELHORES FAMÍLIAS


Prkalva Rotulator -...
霊気
Reiki


Prkalva Rotulator -...


Prkalva Rotulator -...


Prkalva Rotulator - VIE...


Prkalva Rotulator - VIEW


Prkalva Rotulator - V...


Prkalva R...
EDIÇÃO ESPECIAL
O PLANETA DOS MACACOS


Prkalva ...
OS TRAPALHÕES E O MÁGICO DE OROZ


Prkalva Rotulator - VIEW


Prkalva Rotul...


Prkalva Rotul...
Fisioterapia

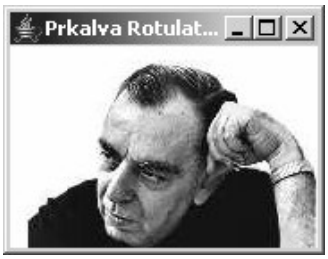

Prkalva ...

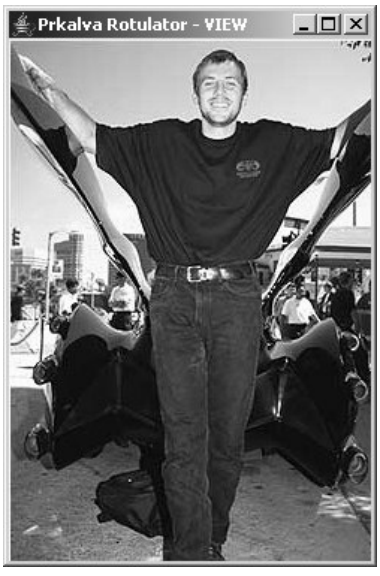

Prkalva Rotulator - VIEW

Apêndice H

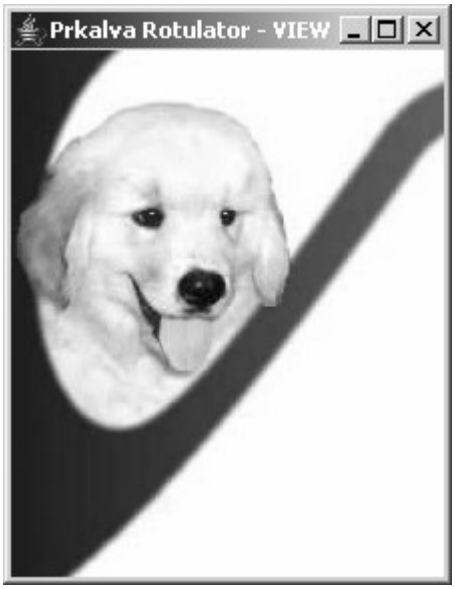
Amostras Incorretamente Classificadas

H1. Classificadas como Automóveis



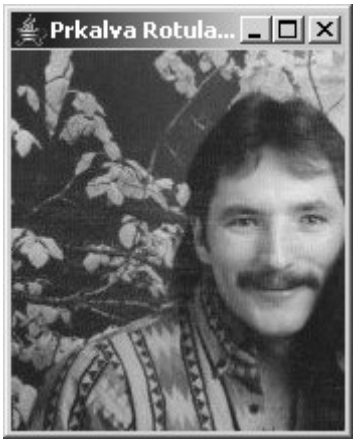


H2. Classificadas como Pessoas



H3. Classificadas como Animais

H4. Classificadas como Motos



H5. Classificadas como CD's/DVD's