



UNIVERSIDADE FEDERAL DE PERNAMBUCO

GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO  
CENTRO DE INFORMÁTICA

2007.2

---



## DETECÇÃO DE OBJETOS EM IMAGENS

---

TRABALHO DE GRADUAÇÃO

**Aluna:** Máira Carneiro Nascimento (mcn2@cin.ufpe.br)

**Orientador:** George Darmiton da Cunha Cavalcanti (gdcc@cin.ufpe.br)

Recife, janeiro de 2007

### Agradecimentos

Primeiramente, agradeço aos meus pais que sempre me apoiaram, e em quem eu sei que sempre posso confiar. Obrigado por tudo que vocês fizeram por mim. Por sempre quererem o melhor para mim, mesmo quando eu não entendo isso. Amo muito vocês.

Agradeço a minha irmã Lívia, não saberia como viver sem as suas “reclamações”. Eu te amo muito irmã. Obrigada também pela companhia nas longas noites de trabalho, e pelos lanches de madrugada. Você é muito importante na minha vida.

Agradeço a Dalilinha, pela alegria e carinho. Pela companhia durante as noites de trabalho. A casa não seria a mesma sem a sua presença.

Agradeço a Adriano, pelo apoio e estímulo, e pela paciência de ter me agüentado durante os períodos de stress. E também claro por ter me emprestado o notebook quando meu computador quebrou. Se não fosse por você não teria conseguido. Te amo.

Agradeço a George pela oportunidade, pela orientação e apoio. E também pela paciência com os atrasos.

Agradeço a todos os colegas do Cin e a todos que me ajudaram durante os projetos estressantes da graduação e aos que também participaram dos momentos de descontração.

### Resumo

A detecção de objetos vem sendo um problema bastante discutido dentro da área de visão computacional. Soluções para esse tipo de problema possuem muitas aplicações na área de indexação de imagens por conteúdo. O objetivo principal desse projeto de graduação é a modelagem e desenvolvimento de um sistema de visão computacional que seja capaz de detectar automaticamente objetos na imagem, indicando a presença ou ausência do objeto. Para o seu desenvolvimento foram implementadas algumas diferentes técnicas de classificação e comparado os seus resultados. E para extração de características dos objetos foi utilizado histogramas espaciais.

**Palavras-chave:** Reconhecimento de padrões. Histogramas espaciais. Detecção de objetos. Visão computacional.

### **Abstract**

The detection of objects has been discussed quite a problem within the area of computer vision. Solutions for this type of problem have many applications in the area of indexing of images by content. The main objective of this graduate project is modeling and development of a computer vision system which is able to automatically detect objects in the image, indicating the presence or absence of the object. For its development were implemented some different techniques for the classification and compared their results. And for extraction of features of the objects was used spatial histogram.

**Keywords:** Pattern recognition. Spatial histograms. Object detection. Computer vision.

## Sumário

Agradecimentos.....	2
.....	2
Resumo.....	3
Abstract.....	4
Sumário.....	5
Índice de Figuras.....	7
Índice de Gráficos.....	8
Índice de Tabelas.....	9
1.Introdução.....	10
1.1Objetivos.....	11
1.2Estrutura do documento.....	11
2.Estado da Arte.....	12
2.1Estudos desenvolvidos .....	12
3.Técnicas Utilizadas.....	15
3.1Visão Geral de Reconhecimento de Padrões.....	15
3.1.1 Elementos de Análise de Imagens.....	17
3.1.2 Padrões e Classes de Padrões.....	18
3.1.3 Sistema de classificação.....	23
3.2Extração de Características com Histogramas Espaciais.....	26
3.2.1 Local Binary Pattern.....	26
3.2.2 Histogramas Espaciais.....	27
3.2.3 Características dos objetos extraídas dos histogramas espaciais.....	28
3.3Métodos de Decisão Teórica.....	30
3.3.1 Classificador de Distância Mínima.....	31
3.3.2 Classificador Kappa.....	32
3.3.3 Classificador Gamma.....	33
4.Resultados Experimentais.....	35
4.1Estudo de Caso.....	35
4.2Banco de Dados.....	37
4.3Resultados.....	38
4.3.1 Banco do Brasil.....	38
4.3.1.1 -Classificador de Distância Mínima.....	38
4.3.1.2 -Classificador Kappa.....	40
4.3.1.3 -Classificador Gamma.....	42
4.3.2 Banco Bradesco.....	44
4.3.2.1 -Classificador de Distância Mínima.....	44
4.3.2.2 -Classificador Kappa.....	46
4.3.2.3 -Classificador Gamma.....	48
4.3.3 Conclusão .....	50
5.Considerações Finais.....	52
5.1Conclusões.....	52
5.2Trabalhos Futuros.....	52
Referências Bibliográficas.....	53

---



## Índice de Figuras

Figura 1: Elementos de análise de imagens.....	17
Figura 2: Duas medidas realizadas em três tipos de flores Iris.....	21
Figura 3: (a) Estrutura em forma de escada; (b) codificação em termos de a e b.....	22
Figura 4: Gráfico do valor da média versus o desvio padrão para um número de diferentes imagens originadas das classes A e B. Nesse caso, uma linha reta separa as duas classes.	23
Figura 5: Os estágios básicos presentes no desenvolvimento de um sistema de classificação.....	25
Figura 6: Vizinhança 3x3 utilizada para a computação do LPB.....	27
Figura 7: Janela da imagem e o modelo espacial.....	28
Figura 8: Podemos verificar que y pode distinguir melhor entre regiões esparsas e densas do que k.....	34
Figura 9: Exemplo dos logotipos utilizados para testar os algoritmos.....	37

### Índice de Gráficos

Gráfico 1: Curva ROC gerada a partir de um conjunto de treinamento com três logotipos do Banco do Brasil para o classificador de distância mínima.....	40
Gráfico 2: Curva ROC gerada a partir de um conjunto de treinamento com nove logotipos do Banco do Brasil para o classificador Kappa.....	42
Gráfico 3: Curva ROC gerada a partir de um conjunto de treinamento com três logotipos do Banco do Brasil para o classificador Gamma.....	44
Gráfico 4: Curva ROC gerada a partir de um conjunto de treinamento com três logotipos do Banco do Brasil para o classificador de distância mínima.....	46
Gráfico 5: Curva ROC gerada a partir de um conjunto de treinamento com nove logotipos do Banco do Brasil para o classificador Kappa.....	48
Gráfico 6: Curva ROC gerada a partir de um conjunto de treinamento com um logotipo do Banco Bradesco para o classificador Gamma.....	50



## Índice de Tabelas

Tabela 1: Resultado do classificador de distância mínima para o logotipo do Banco do Brasil.....	39
Tabela 2: Resultado do classificador kappa para o logotipo do Banco do Brasil.....	41
Tabela 3: Resultado do classificador Gamma para o logotipo do Banco do Brasil.....	43
Tabela 4: Resultado do classificador de distância mínima para o logotipo do Banco Bradesco.....	44
Tabela 5: Resultado do classificador kappa para o logotipo do Banco Bradesco.....	46
Tabela 6: Resultado do classificador Gamma para o logotipo do Banco Bradesco.....	49

### 1. Introdução

A detecção de objetos vem sendo um problema bastante discutido dentro da área de visão computacional. A importância desse problema deve-se ao fato de muitas imagens serem geradas diariamente, criando a necessidade de uma classificação e organização dessa base de dados. Além disso, é imprescindível a existência de um mecanismo que possibilite o acesso fácil, rápido e eficiente a uma imagem específica [14]. Apesar da grande diversidade de soluções desenvolvidas para detecção de objetos em imagens, ainda falta uma aplicação satisfatória, pois muitos desses algoritmos foram desenvolvidos para detectar apenas um objeto específico não sendo estendível para qualquer classe de objetos [12].

O objetivo principal da detecção de objetos em uma imagem é a verificação da existência desses objetos na imagem, e, além disso, a localização dos mesmos na imagem original. A detecção de objetos em uma imagem pode ser analisada do ponto de vista de um problema de classificação que possui duas classes. A classe que representa o objeto que se deseja encontrar e o restante da imagem. Sendo um problema de classificação, uma abordagem utilizada para resolvê-lo é construir um classificador a partir de uma grande quantidade de imagens do objeto a ser encontrado, essas imagens devem expor as características fundamentais do objeto. Esse classificador deve ser capaz de diferenciar de forma satisfatória os objetos encontrados do fundo da imagem [3].

Um dos fatores de maior dificuldade na construção de um sistema de detecção de objetos em imagem reside na grande variedade de formatos que um mesmo objeto pode possuir em uma imagem. Pois, existem os problemas de rotação, de translação, de escala e de iluminação [3]. Tornando o problema complexo. Dessa maneira, um sistema de detecção de objetos em uma imagem deve ser capaz de detectar automaticamente um objeto levando em consideração as possíveis variações causadas por esses fatores.

### 1.1 Objetivos

O objetivo principal desse projeto de graduação é a modelagem e desenvolvimento de um sistema de visão computacional que seja capaz de detectar automaticamente objetos na imagem, indicando a presença ou ausência do objeto.

Para modelar a solução desenvolvida, foi realizada uma análise detalhada das propostas existentes para esse tipo de problema. Foi implementado os algoritmos propostos por Zhang et al. [3] e Harmeling et al. [4]. A partir dos resultados obtidos, foi realizado um estudo comparativo para avaliar o desempenho dos mesmos.

### 1.2 Estrutura do documento

O capítulo 2 aborda uma análise das soluções atuais desenvolvidas para reconhecimento de padrões. O capítulo 3 será definido todos os conceitos utilizados durante o desenvolvimento do projeto. Desde conceitos básicos de reconhecimento de padrões até a definição das técnicas utilizadas para a implementação da solução. O capítulo 4 conterá os resultados dos testes realizados para avaliação dos algoritmos propostos. As conclusões e trabalhos futuros são apresentados no capítulo 5.

## 2. Estado da Arte

Neste capítulo será descrito as principais aplicações e métodos desenvolvidos para a detecção de imagens. Ou seja, será realizado um aparato do cenário atual de reconhecimento de padrões de objetos.

### 2.1 Estudos desenvolvidos

Muitas abordagens têm sido propostas para detecção objeto em imagens sob fundos complexos. Na maioria das abordagens, o problema de detecção de objeto é resolvido por um framework de aprendizagem estatística. Primeiramente, amostras de imagem são representadas por um conjunto de características e, em seguida, métodos de aprendizagem são usados para identificar objetos de uma classe de interesse. Em geral, estas abordagens podem ser classificadas em duas categorias: abordagens baseadas em aparência global e abordagens baseadas em componentes [3].

Abordagens baseadas em aparência global consideram um objeto como uma única unidade e executam a classificação sobre as características extraídas de todo o objeto. Muitos mecanismos de aprendizagem estatísticos são explorados para caracterizar e identificar padrões de objetos. Rowley et al. [15] e Carcia e Delakis [16] detectam faces utilizando um classificador de redes neurais na intensidade de padrões de uma imagem. Pre-processamento é aplicado nas imagens para corrigir variações de iluminação, contraste e orientações. Baseada em características de wavelet, Osuna et al. [17] e Papageorgiou e Poggio [18] adotam support vector machine para localizar rostos humanos e carros. Recentemente, novos algoritmos são aplicados para detectar faces frontais por Viola e Jones [19] e, em seguida, são alargados para detecção de face por vários ângulos diferentes por Li et al. [20] e para detecção de texto por Chen e Yuille [21].

Métodos baseados em componentes tratam um objeto como uma coleção de partes. Estes métodos primeiro extraem alguns componentes dos objetos e, em seguida,

os detectam utilizando informações geométricas. Mohan et al. [22] propõem uma abordagem para detecção de objetos por componentes. Na sua abordagem, uma pessoa é representada por componentes como a cabeça, braços e pernas e, em seguida, classificadores de support vector machine são utilizados para detectar esses componentes e decidir se uma pessoa está presente. Fergus et al. [23] usam operadores de interesse para extrair partes de objetos e realizar a detecção por representação probabilística e reconhecimento de muitas classes de objetos, como motos, rostos humanos, aviões e carros.

Em oposição à maioria das abordagens descritas anteriormente, o problema da detecção objetos com muitas classes e objetos em muitos ângulos de visões diferentes recentemente ganhou muita atenção na comunidade de visão computacional. Lin e Liu [24] propõem uma abordagem para impulsionar a detecção de muitas classes para reconhecer diretamente faces em muitos cenários, tais como rostos em vários ângulos diferentes, rostos sob diferentes condições de iluminação, e faces parcialmente ocultas. Amit et al. [25] usam uma estratégia de detecção de várias formas de classes utilizando uma aplicação para leitura placas de carros. Existem 37 classes de objetos a serem reconhecidas, incluindo 26 letras, 10 dígitos, e um símbolo especial. Tu et al. [26] propõem um framework de análise de imagem para combinar segmentação de imagem, detecção de objeto, compreensão e reconhecimento de cena.

Uma tarefa visual relacionada com a detecção objeto é reconhecimento de objeto, cujo objetivo é identificar instancias específicas do objeto nas imagens. Métodos baseados em descritores locais são cada vez mais utilizados para reconhecimento objeto. Schiele [27] propõe a utilização de derivadas Gaussianas como características locais para criar um histograma multidimensional como a representação do objeto, e, em seguida, executar a tarefa de reconhecer muitos objetos 3D. Lowe [28] desenvolve um sistema de reconhecimento de objetos que utiliza descritores SIFT baseados em orientação locais dos histogramas. No entanto, esses métodos são projetados para reconhecer um objeto específico, ao invés de classificar qualquer classe de objeto.

Outra área de aplicação importante é a detecção automática de logotipos e o seu reconhecimento. Pois, continua a ser de grande interesse para a comunidade de recuperação de documento, uma vez que permite a identificação efetiva da fonte de um documento. A literatura existente relacionada a logotipos de documentos tem se preocupado principalmente no seu reconhecimento [30]. Dado um logotipo extraído precisamente, o seu reconhecimento tem como objetivo a sua classificação dentro de um conjunto de classes de logotipos ou concluir que o objeto não pertence a nenhuma das classes avaliadas. A maioria dos estudos em reconhecimento de logotipos assume que uma detecção de logotipos e a sua extração estão disponíveis [11].

Extração de características de objeto para representação desempenha um papel importante nos sistemas de detecção automática de objetos. Os métodos anteriores usavam muitas representações para extração de características dos objetos, tais como intensidades dos pixels [15,16], wavelets [17,18], características de retângulos [19-21] e, local binary pattern [29]. No entanto, que tipos de recursos são estáveis e flexíveis para detecção de objetos continuam a ser um problema aberto.

### 3. Técnicas Utilizadas

Neste capítulo será descrito as principais técnicas utilizadas para o desenvolvimento do estudo de caso implementado para encontrar o melhor algoritmo para aplicação de detecção de logos de cheques de bancos. Também serão expostos os conceitos básicos que englobam essas técnicas e a área de reconhecimento de padrões.

#### 3.1 Visão Geral de Reconhecimento de Padrões

Reconhecimento de padrões é uma disciplina científica cujo objetivo principal é a classificação de objetos em uma série de categorias ou classes. Dependendo da aplicação, estes objetos podem ser imagens ou sinais de ondas ou qualquer tipo de medidas que precisam ser classificadas, ou seja, esses objetos podem ser genericamente vistos como padrões. A área de reconhecimento de padrões possui uma longa história, mas antes da década de 1960, foi principalmente a produção de pesquisa teórica na área de estatística. No entanto, com o advento da computação houve um aumento da procura por aplicações práticas de reconhecimento padrão. Dessa forma, está presente na maioria dos sistemas inteligentes construídos para tomada de decisões [2].

A análise de imagens é constituída basicamente de um processo de descobrimento, de identificação e de reconhecimento de padrões relevantes. Possui como objetivo primordial o aumento do desempenho e eficiência de uma tarefa que em alguma de sua etapa a análise de imagens seja essencial. Dessa maneira, uma das principais metas de análise de imagens automaticamente é dotar um computador com a capacidade similar dos seres humanos de reconhecer e compreender uma imagem a partir de um padrão visto anteriormente [1]. Por exemplo, se uma pessoa observar uma letra alfanumérica irá reconhecê-la independente da variação de estilos e caligrafias. Esse procedimento não pode ser implementado de maneira tão direta por uma aplicação para reconhecimento de caracteres automaticamente. Pois, apenas comparar imagens de caracteres previamente conhecidos não será suficiente. Será necessário extrair características dos caracteres analisados e comparar com os padrões dos caracteres alfanuméricos. Portanto, a meta é

atingir com uma precisão satisfatória que se assemelhe a dos seres humanos ou até mesmo superior de reconhecimento de caracteres.

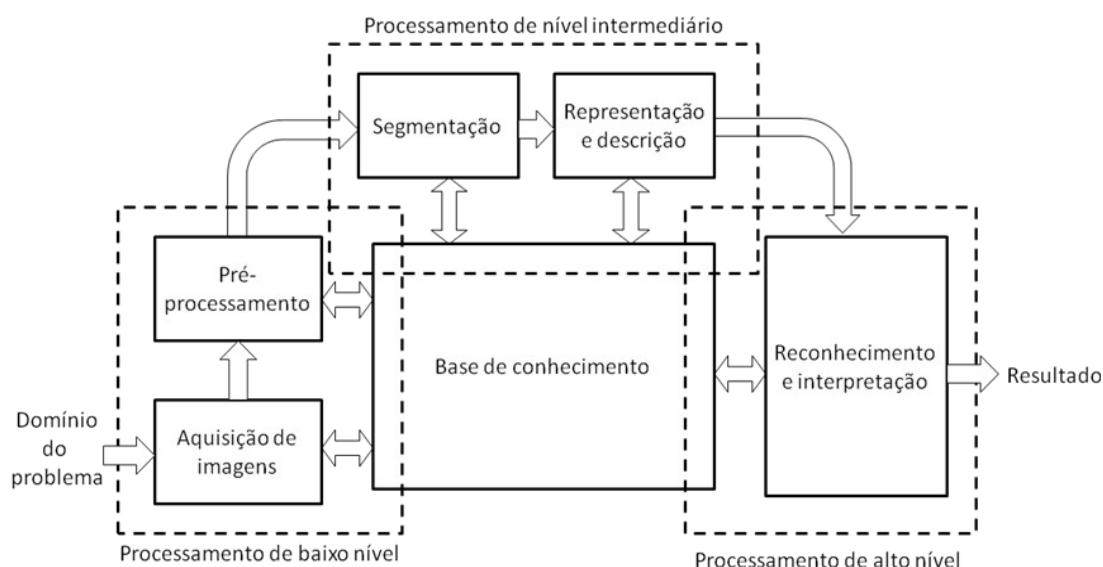
Portanto, um sistema de análise automática de imagens tem que possuir a capacidade de exibir diferentes graus de inteligência. O conceito de uma máquina inteligente é um pouco vago, no entanto, podem-se definir vários tipos de comportamento geralmente associados à inteligência. Como por exemplo, a habilidade de extrair informação pertinente a partir de um fundo de detalhes irrelevantes. Outro exemplo de comportamento seria a capacidade de aprender a partir de exemplos e de generalizar o conhecimento de maneira que ele possa ser aplicado em novas e diferentes circunstâncias. E ainda a habilidade de fazer inferências a partir de informações incompletas [1].

Essas características podem ser obtidas em sistemas de análise de imagens projetados e utilizados em ambientes operacionais limitados. Ou seja, sistemas específicos, desenvolvidos para reconhecimentos de padrões de uma classe restrita de imagens. No entanto, o desenvolvimento de um sistema para reconhecimento de qualquer padrão de objetos em imagens, que se assemelhe a capacidade humana para realizar esta tarefa, ainda não foi eficientemente idealizado. A pesquisa em sistemas biológicos e computacionais está continuamente descobrindo novas e promissoras teorias para explicar a cognição visual humana. No entanto, o estado da arte em análise computadorizadas de imagens está em sua maior parte, baseado em formulações heurísticas projetadas para resolver problemas específicos. Por exemplo, algumas máquinas são capazes de ler documentos impressos apropriadamente formatados em velocidades mais rápidas que a maioria dos humanos. Entretanto, sistemas desse tipo são altamente especializados, possuindo pouco ou nenhuma capacidade de generalização. Ou seja, as limitações atuais teóricas e de implementação na área de análise de imagens acarretam soluções que sejam altamente dependentes do problema [2].



### 3.1.1 Elementos de Análise de Imagens

Para auxiliar definição das técnicas de análise de imagens, será utilizada a divisão adotada por [1], que a separou em três áreas básicas. Essas áreas são *processamento de baixo nível*, *processamento de nível intermediário* e *processamento de alto nível*. Embora essas subdivisões não possuam fronteiras definidas, elas fornecem um esquema de trabalho útil para a categorização dos vários processos que são componentes inerentes de um sistema autônomo de análise de imagens. A Figura 1 ilustra esses conceitos, com as linhas tracejadas sobrepostas indicando a inexistência de fronteiras bem definidas entre os processos. Por exemplo, a limiarização pode ser vista como um melhoramento (pré-processamento) ou como uma ferramenta de segmentação, dependendo da aplicação.



**Figura 1:** Elementos de análise de imagens

O *processamento de baixo nível* trata de funções que podem ser vistas como reações automáticas, dessa forma, esse sistema não precisa exercer nenhum processo inteligente. A aquisição e o pré-processamento de imagens podem ser considerados como funções de baixo nível. Portanto, esse nível inclui atividades desde o próprio processo de formação da imagem até compensações, como redução de ruídos ou borramento da imagem.

A tarefa de extração e caracterização de componentes em uma imagem resultante do processo de baixo nível constitui o *processo de nível intermediário*. Como Figura 1 indica, os processos de nível intermediário incluem a segmentação e a descrição. As tarefas do processo de nível intermediário podem possuir capacidades para um comportamento inteligente, como os procedimentos flexíveis de segmentação. Por exemplo, o preenchimento de pequenos buracos em uma fronteira segmentada envolve elementos mais sofisticados de resolução de problemas que as meras reações automáticas de baixo nível.

Finalmente, o *processo de alto nível* envolve o reconhecimento e a interpretação. Esses dois processos fazem parte da área de cognição inteligente. Pode-se observar que a maioria das técnicas utilizadas pelo processamento de baixo nível e de nível intermediário é baseada em teorias razoavelmente bem definidas e formuladas. O mesmo não acontece na área de reconhecimento e especialmente na área de interpretação. Pois, o conhecimento e compreensão dos princípios fundamentais são muito menos precisos e muito mais especulativos. Essa falta relativa de entendimento resulta, em ultimo caso, em uma formulação de restrições e idealizações com a intenção de reduzir a complexidade da tarefa em um nível tratável. O produto final é um sistema com capacidades operacionais altamente especializadas.

### 3.1.2 Padrões e Classes de Padrões

O reconhecimento de padrões existentes em uma imagem é uma habilidade fundamental na análise de imagens. Pode-se definir um padrão para o caso de análise de imagens como sendo uma descrição quantitativa ou estrutural de um objeto ou alguma entidade de interesse em uma imagem. Em geral, um padrão é formado por um ou mais descritores. Em outras palavras, um padrão é um arranjo de descritores, também chamados de características [1]. Uma classe de padrões constitui uma família de padrões que compartilham algumas propriedades em comum. As classes de padrões são denotadas como  $w_1, w_2, \dots, w_M$ , onde  $M$  é o número de classes existentes. O reconhecimento de

---

padrões automaticamente envolve técnicas de atribuições de padrões às suas respectivas classes, sem nenhuma ou com mínima intervenção possível.

Os principais arranjos de padrões utilizados na prática são os vetores, usados para descrições quantitativas, as cadeias e as árvores, estas duas últimas usadas para descrições estruturais [1]. Os vetores de padrões podem ser representados por letras minúsculas em negrito, como por exemplo, **x**, **y** e **z**. O vetor de padrão tem a seguinte estrutura:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

cada componente,  $x_i$ , representa o  $i$ -ésimo descritor e  $n$  é o número totais de tais descritores. Vetores de padrões são representados como colunas, ou seja, matrizes  $n \times 1$ . Portanto, um vetor de padrão tanto pode ser representado como na equação acima ou na forma de uma matriz transposta. Ou seja,  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ , onde T indica a transposição.

A natureza dos componentes de um vetor de padrões **x** depende da técnica de medida usada na descrição do próprio padrão físico. Utilizando o exemplo descrito por [1], o problema descrito consiste na separação de três tipos de flores Iris (*Iris setosa*, *virginica* e *versicolor*) a partir da largura e comprimento de suas pétalas. Nesse caso, estaríamos tratando com vetores de padrões bidimensionais da forma:

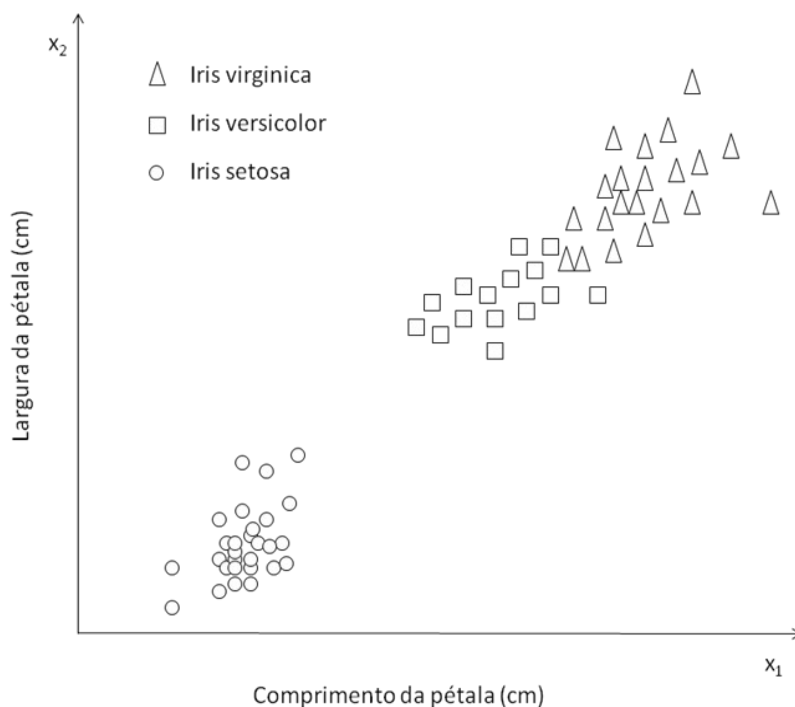
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

em que  $x_1$  e  $x_2$  correspondem ao comprimento e à largura da pétala, respectivamente. As três classes de padrões nesse caso, denotado por  $w_1$ ,  $w_2$  e  $w_3$ , correspondem às variedades *setosa*, *virginica* e *versicolor*, respectivamente.

Uma vez que as pétalas de todas as flores variam um pouco em largura e comprimento, os vetores de padrões que descrevem essas flores também variarão, não

apenas entre as diferentes classes, mas também dentro de cada classe. A Figura 2 mostra as medidas de comprimento e de largura para diversas amostras de cada tipo da flor Iris. Uma vez que um conjunto de medidas tenha sido selecionado, no caso do exemplo a largura e comprimento, um vetor de padrões torna-se a representação completa de cada amostra física. Portanto, cada flor nesse caso torna-se um ponto no espaço euclidiano bidimensional. Vale a pena salientar que apesar da medida de comprimento e largura de uma pétala ter separado adequadamente a classe de *Iris setosa* das outras, o mesmo não aconteceu as classe de *virginica* e *versicolor*. Esse resultado ilustra o clássico problema de seleção de características, em que o grau de separação entre as classes depende fortemente da escolha das medidas dos padrões selecionadas para uma aplicação.

Dessa forma, pode-se verificar que os vetores de padrões podem ser gerados de diversas formas. No entanto, pode-se destacar que a seleção de uma medida ou medidas nas quais se baseiam os componentes do vetor de padrões possui uma influência profunda na eventual desempenho de um sistema de análise de imagens baseado nessa abordagem de vetores de padrões.



**Figura 2:** Duas medidas realizadas em três tipos de flores Iris

A técnica descrita para geração de vetores de padrões utiliza informações quantitativas para separar as classes de padrões. No entanto, nem sempre isso ocorre, em algumas aplicações as características de padrões são melhores descritas por relações estruturais. Por exemplo, o reconhecimento de impressões digitais baseia-se nas relações entre características das impressões chamadas *minutiae*. Essas características são componentes primitivas que descrevem as propriedades dos sulcos das impressões digitais, como terminações abruptas, ramificações, fusões e segmentos desconectados, juntamente com suas posições e tamanhos relativos. Problemas de reconhecimento desse tipo, em que não apenas as medidas quantitativas sobre cada característica, mas também as relações espaciais entre as características determinam as classes, são geralmente melhor resolvidos por abordagens estruturais [1].

Um exemplo de uma descrição estrutural pode ser observado em [1], que destaca um padrão simples em forma de escada como se pode visualizar na Figura 3 (a). Esse padrão poderia ser descrito na forma de um vetor de padrões como mostrado

anteriormente. Entretanto essa abordagem não destacaria a sua estrutura básica, que consiste na repetição simples de dois elementos primitivos. Então, uma descrição mais significativa seria através da definição dos elementos primitivos  $a$  e  $b$ , dessa forma, o padrão poderia ser classificado como uma cadeia desses símbolos, ou seja,  $w = \dots ababab \dots$ , como mostrado na Figura 3(b). A estrutura da classe descrita pode ser capturada a partir da representação em forma de cadeia. Essa construção estrutural é aplicável a escadas de qualquer tamanho, mas exclui outros tipos de estruturas que poderiam ser geradas por outras combinações das primitivas  $a$  e  $b$ .

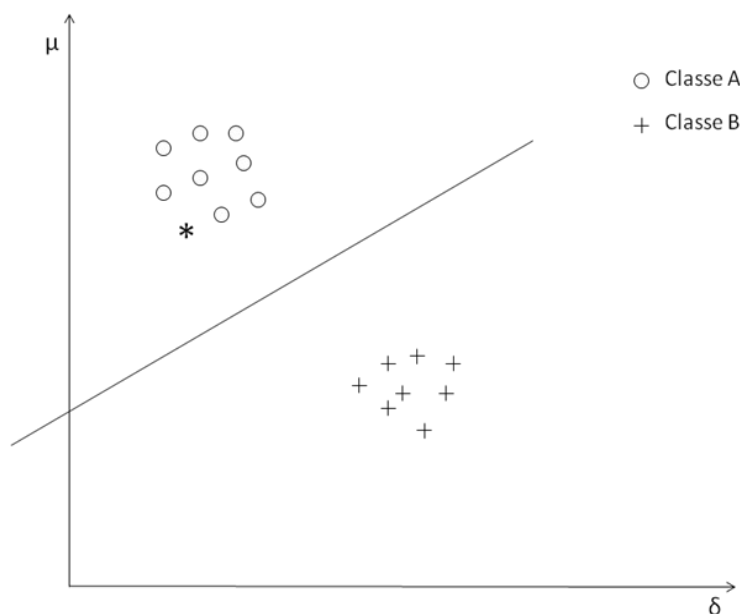


**Figura 3:** (a) Estrutura em forma de escada; (b) codificação em termos de  $a$  e  $b$ .

Representações por cadeias geram adequadamente padrões de objetos e outras entidades cujas estruturas baseiam-se em conectividade relativamente simples de primitivas, usualmente associadas com a fronteira da forma. Uma abordagem mais poderosa para muitas aplicações baseia-se no uso de descrições por árvores. Basicamente, a maior parte dos esquemas de organização hierárquica leva a estruturas em árvore. Todos os arranjos de padrões descritos anteriormente podem ser utilizados no estágio de reconhecimento de sistemas de análise de imagens [1].

### 3.1.3 Sistema de classificação

Nesta seção será visto de uma maneira geral as etapas, descritas por [2], para o desenvolvimento de um sistema de classificação. Por exemplo, para se separar duas regiões pertencentes a duas classes diferentes, primeiramente é necessário identificar as quantidades mensuráveis que tornam essas duas regiões distintas umas das outras. A Figura 4 mostra um gráfico gerado a partir do valor médio da intensidade em cada região de interesse versus o correspondente desvio padrão em torno desta média. Cada ponto corresponde a uma imagem diferente disponível no banco de dados utilizado pelo sistema. Acontece que a classe de padrões A tende a se disseminar em uma área diferente da classe de padrões B. Nesse caso, uma reta pode ser um bom candidato para se separar as duas classes. Se uma nova imagem é fornecida para ser devidamente classificada e não se sabe a que classe a mesma pertence. Dessa forma, seria calculado o seu ponto no gráfico, isso é demonstrado pelo asterisco (\*) na Figura 4. Então é razoável supor que o desconhecido padrão é mais provável pertencer à classe A do que a classe B



**Figura 4:** Gráfico do valor da média versus o desvio padrão para um número de diferentes imagens originadas das classes A e B. Nesse caso, uma linha reta separa as duas classes.

A tarefa anterior de classificação mostra o raciocínio por trás de uma grande classe de problemas de reconhecimento de padrão. As medições utilizadas para a

classificação, o valor médio e desvio-padrão neste caso, são conhecidos como características. Cada um dos vetores de características identifica exclusivamente um único padrão (objeto). Pode-se também visualizar as características e os vetores de características como variáveis aleatórias e vetores simplesmente, respectivamente. Pois, como as medições resultantes de diferentes padrões exibem uma variação aleatória. Isso é causado em parte ao ruído natural causado pela falta de precisão dos dispositivos de medição e, em parte, as características distintas de cada padrão.

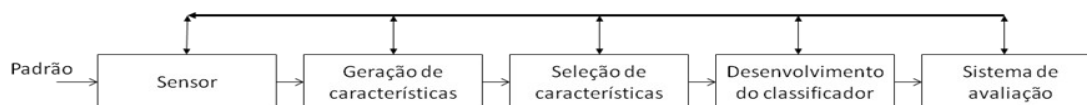
A linha reta na Figura 4 é conhecida como a linha de decisão, e que constitui o classificador cujo papel consiste em dividir o espaço em regiões que correspondem ou a classe A ou classe B. Se um vetor de características  $x$ , correspondendo a um padrão desconhecido, pertence à região A, ele é classificado como a classe A, caso contrário como a classe B. Isso não necessariamente significa que essa decisão de classificação foi correta. Se não foi correto, uma classificação incorreta ocorreu. Para traçar a linha reta na Figura 4 foi explorado o fato de que as classes A e B eram conhecidas para cada ponto da figura. Os padrões (vetores de característica), cuja verdadeira classe é conhecida e que são utilizados para a concepção do classificador são conhecidos como padrões de treinamento (vetores de características de treinamento).

No entanto alguns pontos fundamentais sobre a tarefa de classificação podem ser destacados, como:

- Como são as características são geradas? No exemplo anterior, foi utilizado a média e o desvio-padrão, porque era sabido previamente como as imagens tinham sido geradas. Na prática, isso está longe de ser tão óbvio. Isso depende totalmente do sistema, e diz respeito à fase de geração de características de um sistema de classificação que realiza uma determinada tarefa de reconhecimento de padrão
- Qual o melhor número de características para ser utilizado? Esta é também uma tarefa muito importante e que diz respeito à fase de seleção de características do sistema de classificação. Na prática, um número maior do que o necessário de candidatos de características é gerado e, em seguida, o "melhor" deles é adotado.



- Após ter adotado o adequadamente as características utilizadas, como desenhar um classificador? No exemplo anterior a linha reta foi estabelecida empiricamente. Na prática, isto não pode ser o caso, a linha deve ser traçada otimamente. Além disso, problemas para os quais um classificador linear (linha reta ou hiperplano no espaço 1-dimensional) pode resultar em desempenho aceitável não é a regra. Em geral, as superfícies dividindo o espaço nas diversas classes regiões são não lineares. Que tipo de não-linearidade deve ser adotado e qual o tipo de critério de otimização deve ser utilizado, a fim de localizar uma superfície no lugar certo no espaço de características 1-dimensional? Estas questões dizem respeito à fase de concepção do classificador.
- Finalmente, uma vez que o classificador foi projetado, como se pode avaliar o desempenho deste classificador? Ou seja, qual é a taxa de erro de classificação? Esta é a tarefa do sistema de avaliação.



**Figura 5:** Os estágios básicos presentes no desenvolvimento de um sistema de classificação.

A Figura 5 mostra os vários estágios seguidos para o desenvolvimento de um sistema de classificação. Esses estágios não são independentes, o que pode ser visualizado pelas setas de retorno. Pelo contrário, eles são interligados e, dependendo dos resultados alcançados, pode-se voltar para as fases anteriores para conseguir um desempenho mais satisfatório.

Outra característica que irá influenciar o desenvolvimento de um sistema de classificação é a presença ou não de uma base de treinamento. Ou seja, ao se partir do princípio que um conjunto de dados treinamento está disponível, e o classificador foi desenvolvido ao explorar esta informação conhecida a priori. Isto é conhecido como padrão de reconhecimento supervisionado. No entanto, nem sempre é esse o caso, e existe outro tipo de tarefa de reconhecimento de padrão para a qual os dados de

treinamento não estão disponíveis. Neste tipo de problema, é fornecido um conjunto de vetores de características  $\mathbf{x}$  e o objetivo é revelar as semelhanças, e agrupar vetores "semelhantes" juntos. Isto é conhecido como padrão de reconhecimento não supervisionado ou agrupamento. Essa tarefa pode surgir em muitas aplicações em ciências sociais e de engenharia, como o sensoriamento remoto, segmentação de imagens, reconhecimento de imagem e voz.

Uma questão importante no reconhecimento de padrões não supervisionado é a definição de "semelhança" entre dois vetores de características e escolher uma medida adequada para ele. Outra questão de extrema importância é escolher um esquema de algoritmos que irá agrupar os vetores em função da medida adotada de similaridade. Em geral, os diferentes algoritmos podem levar a resultados diferentes.

### 3.2 Extração de Características com Histogramas Espaciais

Representação de objetos e extração de características é essencial para a detecção de objetos. Nesta seção será descrito a representação de padrões de objetos combinando textura e estruturas espaciais. Essencialmente, objetos são modelados pelos seus histogramas espaciais e extração de características específica de cada classe de objetos para sua detecção.

#### 3.2.1 Local Binary Pattern

Na solução testada, a imagem em tons de cinza é pré-processada utilizando Local Binary Pattern (LPB). LPB é um modelo simples de textura e tem sido provado ser bastante poderoso na classificação de texturas [8, 9]. LPB é invariante em relação a qualquer transformação monotônica na escala de cinza. Como mostra a Figura 6, o operador básico do LPB, como foi definido por [3], utiliza a intensidades dos pixels vizinhos para calcular o valor do pixel central [3].

$g_1$	$g_2$	$g_3$
-------	-------	-------

$g_4$	$g_0$	$g_5$
$g_6$	$g_7$	$g_8$

**Figura 6:** Vizinhança 3x3 utilizada para a computação do LPB

Aos pixels vizinhos contidos na janela 3x3 são atribuídos valores em relação ao pixel central, seguindo a seguinte função:

$$s(g_0, g_i) = \begin{cases} 1, & g_i \geq g_0 \\ 0, & g_i < g_0 \end{cases} \quad 1 \leq i \leq 8$$

Os valores das oito diferenças são codificados em um número de oito bits para obter o valor LPB do pixel central, da seguinte maneira:

$$LPB(g_0) = \sum_{i=1}^8 s(g_0, g_i) 2^{i-1}$$

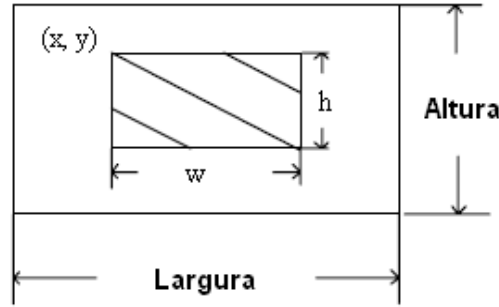
Para cada imagem testada, primeiramente será calculado o resultado LPB da imagem, seguindo o modelo definido anteriormente, para posteriormente ser calculado o seu histograma espacial.

### 3.2.2 Histogramas Espaciais

O histograma é uma representação global do padrão da imagem, o mesmo é invariante a rotação e translação. No entanto, histogramas não são adequados para detecção de objetos, pois, não codificam a distribuição espacial dos mesmos. Para alguma imagem diferente do objeto a ser encontrado, o seu histograma e o do objeto alvo podem ser bastante semelhantes ou até mesmo idênticos. Dessa forma, a comparação de histogramas simples não é bastante para a detecção de objetos.

Para elevar a habilidade discriminatória, serão utilizados os histogramas espaciais, como foi definido por [3]. Nos quais modelos espaciais são utilizados para codificar a distribuição espacial dos padrões dos objetos. Como mostrado na Figura 7, utiliza-se uma

janela com tamanho fixo de largura e altura para adquirir amostras dos padrões dos objetos na imagem. Assim, é codificado o padrão da distribuição espacial utilizando esses modelos espaciais. Cada modelo é uma máscara binária retangular, como mostrada na Figura 7. Cada modelo é denominado como  $rt(x,y,w,h)$ , onde  $(x,y)$  é a localização da posição do topo esquerdo da máscara, enquanto  $w$  e  $h$  são as larguras e altura da máscara, respectivamente.



**Figura 7:** Janela da imagem e o modelo espacial

Para cada modelo espacial  $rt(x,y,w,h)$ , será modelado uma sub-imagem dentro da janela da máscara por um histograma. Esse tipo de histograma é chamado de histograma espacial. Para uma amostra de imagem  $P$ , o seu histograma espacial associado ao modelo  $rt(x,y,w,h)$  é denominado como  $SH^{rt(x,y,w,h)}(P)$ .

### 3.2.3 Características dos objetos extraídas dos histogramas espaciais

Muitos métodos podem ser utilizados para medir similaridade entre dois histogramas, como distância quadrada, distância Chi-square e interseção de histogramas [6]. Interseção de histogramas é bastante estável e computacionalmente eficiente. A medida de similaridade da interseção de dois histogramas é calculada da seguinte maneira [10]:

$$D(H_1, H_2) = \frac{\sum_{i=1}^k \min(H_1^i, H_2^i)}{\sum_{i=1}^k H_1^i}$$

onde  $H_1$  e  $H_2$  são dois histogramas, e  $k$  é o número de tons de cinza dos histogramas.

Suponha uma base de dados com  $n$  amostras de objetos e um modelo espacial, o modelo de histogramas é representado sobre o modelo espacial como a média dos histogramas espaciais das amostras de treinamento, definido por [3] como:

$$SH^{rt(x,y,w,h)} = \frac{1}{n} \sum_{j=1}^n SH^{rt(x,y,w,h)}(P_j)$$

onde  $P_j$  é um amostra de objeto de treinamento, e  $rt(x,y,w,h)$  é um modelo espacial. Para cada amostra  $P$ , a característica do histograma espacial, representada por  $f^{rt(x,y,w,h)}(P)$ , é definida como a sua distancia para a media dos histogramas dos objetos, dado como:

$$f^{rt(x,y,w,h)}(P) = D(SH^{rt(x,y,w,h)}(P), SH^{rt(x,y,w,h)})$$

Um padrão de objeto é codificado por um conjunto de modelos espaciais,  $\{rt(1), \dots, rt(m)\}$ , onde  $m$  é o número de modelos espaciais. Portanto, uma amostra de objeto é representada pelo vetor de características de histogramas espaciais no espaço das características dos histogramas espaciais, ou seja:

$$F = [f^{rt(1)}, \dots, f^{rt(m)}]$$

Como as máscaras podem variar em posições e tamanhos na janela da imagem, o conjunto completo de características de histogramas espaciais é bastante grande. Portanto, o espaço das características dos histogramas espaciais codifica completamente a textura e distribuição espacial dos objetos. Em adição, características dos histogramas espaciais são como se fosse uma classe de características de objetos específica, já que ele codifica similaridade das amostras para modelos de histogramas dos objetos.

Comparação de histogramas é um método direto para reconhecimento de objetos. Nesse método, um modelo de histograma de um padrão de objeto é primeiramente gerado para um modelo espacial. Se o histograma de uma amostra é próximo do modelo de histograma dentro de certo limite, a amostra é classificada como um padrão do objeto. Se

P for um amostra e sua característica de histograma espacial com o modelo  $rt(x,y,w,h)$  for  $f^{rt(x,y,w,h)}(P)$ , P é classificado como um padrão de objeto se  $f^{rt(x,y,w,h)}(P) \geq \theta$ , se não P é classificado como não padrão do objeto.  $\theta$  é o limite para a classificação.

### 3.3 Métodos de Decisão Teórica

As abordagens de decisão teórica para o reconhecimento baseiam-se na utilização de *funções de decisão* ou *discriminantes*. Seja  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  um vetor de padrões n-dimensional. Para M classes de padrões  $w_1, w_2, \dots, w_M$ , o problema básico em reconhecimento de padrões por decisão teórica é encontrar M funções de decisão  $d_1(\mathbf{x})$ ,  $d_2(\mathbf{x})$ , ...,  $d_M(\mathbf{x})$ , com a propriedade que, se o padrão  $\mathbf{x}$  pertencer à classe  $w_i$ , então:

$$d_i(\mathbf{x}) > d_j(\mathbf{x}) \quad j = 1, 2, \dots, M; j \neq i.$$

Em outras palavras, um padrão desconhecido  $\mathbf{x}$  pertence a i-ésima classe de padrões se a substituição de  $\mathbf{x}$  em todas as funções de decisão fizer com que  $d_i(\mathbf{x})$  tenha o maior valor numérico. Empates são resolvidos arbitrariamente [1].

A fronteira de decisão que separa as classes  $w_i$  e  $w_j$  é dada pelos valores de  $\mathbf{x}$  para os quais  $d_i(\mathbf{x}) = d_j(\mathbf{x})$  ou, equivalentemente, pelos valores de  $\mathbf{x}$  para os quais

$$d_i(\mathbf{x}) - d_j(\mathbf{x}) = 0.$$

É comum identificar-se a fronteira de decisão entre duas classes pela função

$$d_{ij}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x}) = 0 \quad . \text{ Portanto, } d_{ij}(\mathbf{x}) > 0 \quad \text{para os padrões da classe } w_i \text{ e}$$
$$d_{ij}(\mathbf{x}) < 0 \quad \text{para os padrões da classe } w_j \text{ [1].}$$

Existem varias abordagens para encontrar funções decisões que satisfaçam a equação

$$d_i(\mathbf{x}) > d_j(\mathbf{x}) \quad j = 1, 2, \dots, M; j \neq i.$$

Uma delas é utilizar o classificador de distância mínima como função de decisão. Este classificador será descrito na próxima seção.

### 3.3.1 Classificador de Distância Mínima

Suponha que cada classe de padrões seja representada por um vetor *protótipo* ou *médio*, como descrito a seguir:

$$\mathbf{m}_j = \frac{1}{N_j} \sum_{\mathbf{x} \in w_j} \mathbf{x}, \quad j = 1, 2, \dots, M.$$

em que  $N_j$  é o número de vetores de padrões da classe  $w_i$ , e a soma é realizada sobre esses vetores. Uma maneira de definir a pertinência de um vetor  $\mathbf{x}$  de padrões desconhecidos é atribuí-lo à classe de seu protótipo mais próximo. A distância euclidiana pode ser usada para determinar a proximidade, reduzindo o problema à computação das distâncias:

$$D_j(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_j\|, \quad j = 1, 2, \dots, M$$

em que  $\|\mathbf{a}\| = (\mathbf{a}^T \mathbf{a})^{1/2}$  é a norma euclidiana. Atribuímos, então,  $\mathbf{x}$  à classe  $w_i$  se  $D_i(\mathbf{x})$  for a menor distância. Ou seja, a menor distância implica no melhor casamento nessa formulação. Não é difícil mostrar que isso é equivalente a avaliar as funções

$$d_j(\mathbf{x}) = \mathbf{x}^T \mathbf{m}_j - \frac{1}{2} \mathbf{m}_j^T \mathbf{m}_j, \quad j = 1, 2, \dots, M$$

e a atribuir  $\mathbf{x}$  à classe  $w_i$  se  $d_i(\mathbf{x})$  levar ao maior valor numérico. Essa formulação está de acordo com o conceito de função de decisão definido anteriormente.

A partir das equações anteriores, pode-se ver a fronteira de decisão entre as classes  $w_i$  e  $w_j$  para o classificador de distância mínima é

$$d_{i,j}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x})$$

$$= \mathbf{x}^T (\mathbf{m}_i - \mathbf{m}_j) - \frac{1}{2} (\mathbf{m}_i - \mathbf{m}_j)^T (\mathbf{m}_i - \mathbf{m}_j) = 0$$

A superfície dada por esta equação é a bissecção perpendicular do segmento de linha entre  $\mathbf{m}_i$  e  $\mathbf{m}_j$ . Para  $n = 2$ , a bissecção perpendicular é uma linha, para  $n = 3$  é um plano, e para  $n > 3$  é chamada de hiperplano.

Na pratica, o classificador de mínima distância funciona bem quando a distância entre as médias for grande em comparação com a dispersão ou aleatoriedade de cada classe em relação à sua média. O classificador de mínima distância leva um desempenho ótimo, em termos de minimização média de classificações equivocadas, quando a distribuição de cada classe em torno de sua origem for da forma de uma “hiper nuvem” esférica no espaço  $n$ -dimensional de padrões. A ocorrência simultânea de grandes separações entre as médias com relativamente pouca dispersão nas classes ocorre raramente na prática, a menos que o projetista do sistema controle a natureza da entrada de dados.

### 3.3.2 Classificador Kappa

De acordo com [4], o classificador kappa consiste na avaliação da densidade de um determinado ponto. Ou seja, o  $k$ -vizinho mais próximo estimador de densidade avalia a densidade de um determinado ponto, calculando o volume da menor bola centrada nesse ponto que contém seus  $k$  vizinhos mais próximos e relacionando isso ao quociente  $k/n$ . Pode ser provado que este estimador de densidade é consistente-L2 (ver [7]). Infelizmente, a estimativa não é sempre muito precisa se o número de pontos de dados é pequeno ou se a dimensionalidade é elevada. No entanto, detecção de pontos não pertencentes ao padrão pesquisado não exige a densidade real. A fim de decidir se um ponto de dados é pertencente a classe alvo ou não, uma estimativa aproximada é um indicador suficiente. Esse índice representa, portanto, a essência do  $k$  vizinho mais próximo estimador de densidade:  $k(x)$  é o raio da menor bola centrada em  $x$  contendo



seus  $k$  vizinhos mais próximos, ou seja, a distância entre  $x$  e seu  $k$ -ésimo vizinho mais próximo,

$$k(x) = \|x - z_k(x)\|$$

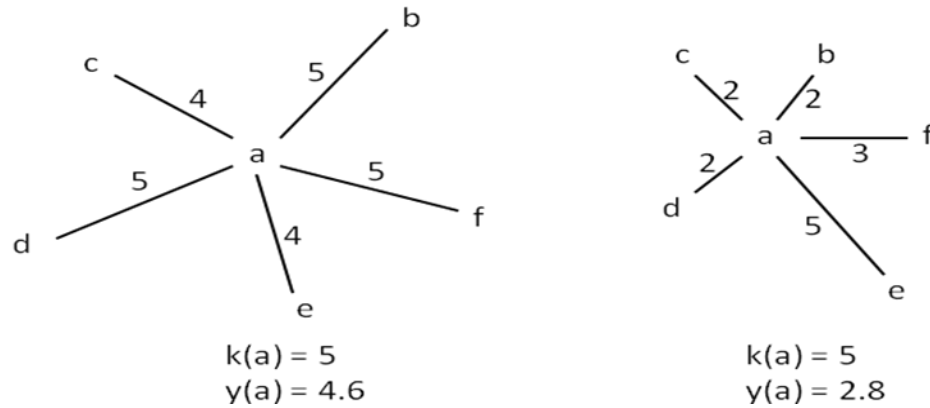
É evidente que, em regiões densas  $k$  é pequena e em regiões esparsas  $k$  é grande, o que faz dele um bom candidato para um índice de detecção de pontos longes do padrão selecionado, uma vez que a lógica é que pontos não pertencentes às classes se encontram em regiões esparsas.

### 3.3.3 Classificador Gamma

O classificador Gamma foi descrito em [4], como um refinamento do classificador Kappa. Pois o índice  $k$  como descrito no classificador kappa parece possuir um tanto de desperdício, ele considera a distância para o  $k$ -ésimo vizinho mais próximo, mas ignora as distâncias para os vizinhos mais próximos. Isto sugere um refinado índice que leva as distâncias de todos os  $k$  vizinhos mais próximos em conta. Dessa forma, o classificador Gamma,  $y(x)$ , é a média da distancia do valor de  $x$  aos seus  $k$  vizinhos,

$$y(x) = \frac{1}{k} \sum_{j=1}^k \|x - z_j(x)\|$$

Este índice possibilita a distinção das duas situações representadas no painel esquerdo da Figura 8. O valor de  $k$  é a mesmo em ambas as situações, pois o ponto  $e$  possui a mesma distância para o ponto  $a$ , embora a região à direita seja mais densa. Explorando todas as distâncias,  $y$  pode distinguir as duas situações.



**Figura 8:** Podemos verificar que  $y$  pode distinguir melhor entre regiões esparsas e densas do que  $k$

Notamos que  $\Gamma$  é delimitada por  $\kappa$ ,  $\kappa(x) \geq y(x) \geq 0$ . O que pode ser facilmente verificado observando que  $2 \max(\|a\|, \|b\|) \geq \|a\| + \|b\| \geq 0$  é verdadeiro. Isso significa que, se  $y(x)$  é grande (o que implica que provavelmente  $x$  não pertence ao padrão procurado)  $k(x)$  também é grande. Pelo contrário, se  $y(x)$  é pequeno, então  $k(x)$ , não precisa necessariamente ser pequeno, já que ele poderia ter ignorado algumas informações relevantes. Portanto, ao contrário de  $y$ ,  $k$  poderia mal interpretar um ponto de uma região esparsa como sendo não pertencente à classe.

### 4. Resultados Experimentais

Neste capítulo serão expostos os resultados encontrados para testar os algoritmos descritos no capítulo anterior. Para isso, foram implementados cada um deles e utilizados os resultados para comparar o que melhor se adequava ao estudo de caso selecionado.

#### 4.1 Estudo de Caso

Para testar os algoritmos estudados foi utilizada uma aplicação de detecção de logotipos de bancos em cheques escaneados. Ou seja, o sistema utiliza vários tipos de algoritmos propostos e compara os seus resultados finais. Para isso, para um determinado logotipo de banco, era gerado um conjunto de treinamento contendo várias amostras dos mesmos, e utilizado isso para detectar se a imagem processada é do mesmo padrão que o conjunto. Caso fosse, seria classificado o cheque como pertencente ao banco do qual o logotipo faz parte.

Para simplificar o reconhecimento dos cheques, não foi utilizada uma busca exaustiva de toda a área do mesmo para a detecção do logotipo, pois sua localização é bastante constante em todos eles. Dessa forma, foi atribuída uma área de busca, para evitar o processamento desnecessário.

O sistema foi projetado para receber o local onde se encontra o conjunto de treinamento, de imagens corretas e incorretas. Dessa maneira, a partir das imagens de treinamento fornecidas é calculada a média dos histogramas espaciais de todas elas. Depois, para cada imagem que se deseja reconhecer é calculado o seu histograma espacial, e depois calculado o valor de sua interseção com o histograma espacial médio do conjunto de treinamento. No entanto, a interseção de histogramas é uma medida de similaridade, ou seja, quanto mais próxima maior seu valor. Para utilizar noção de distância padrão para todos os algoritmos, quanto mais próximo menor a distância, foi utilizado o complemento do valor da interseção.

Depois de calculado o valor da interseção do histograma espacial da imagem testada com a média dos histogramas espaciais do conjunto de treinamento, foi implementado três maneiras diferentes de métodos de decisão teórica, o classificador de distancia mínima, kappa e gamma. Dessa forma, para utilizar a aplicação é necessário definir a quantidade de imagens dos conjuntos de treinamento e testes, os parâmetros dos histogramas espaciais utilizados, e o classificador utilizado.

Para realizar a comparação dos resultados dos testes será utilizado a AUC (Area Under Curve) da curva ROC (Receiver operating characteristics). Para a implementação da curva ROC e o calculo da AUC foi utilizado o algoritmo proposto por Fawcett [31]. Os gráficos da curva ROC são úteis para organizar classificadores e visualizar as suas performances. Gráficos ROC são comumente utilizados em pesquisas de aprendizagem de máquina e mineração de dados. O gráfico é bidimensional onde o taxa de acertos, tp (true positive), consiste o eixo y e o taxa de falsos positivos, fp (false positive), corresponde ao eixo x. A curva ROC retrata a relação entre os benefícios (verdadeiros positivos) e os custos (falso positivo).

Informalmente, um ponto no espaço ROC é melhor do que outro se for a noroeste (taxa tp é mais elevada, taxa fp é mais baixa, ou ambos) do primeiro. Classificadores que figuram no lado esquerdo de um gráfico ROC, perto do eixo X, podem ser considerados como “conservadores”: eles fazem classificações positivas apenas com fortes indícios de acerto, portanto, também fazem poucos erros de falsos positivos, mas eles têm, freqüentemente, baixas taxas de verdadeiros positivos também. Classificadores no alto do lado direito de um gráfico ROC pode ser visto como “liberais”: eles fazem classificações positivas com provas fracas, assim, classificam quase todos positivos corretamente, mas normalmente possuem taxas elevadas de falso positivo.

Uma curva ROC é uma imagem bidimensional de classificador de desempenho. Para comparar esses classificadores se reduz o desempenho da curva ROC para um único valor escalar representando o desempenho esperado. Um método comum é o de calcular a área sob a curva ROC, abreviada por AUC. Desde a AUC é uma parte da área do quadrado unitário, o seu valor será sempre entre zero e um. No entanto, porque a

---

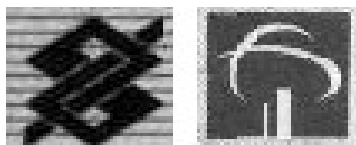
classificação aleatória produz a linha diagonal entre (0, 0) e (1, 1), que tem uma área de 0,5, nenhum classificador realista deverá ter uma AUC inferior a 0,5.

Portanto, foram gerados diferentes testes para cada classificador, diferentes tamanhos de conjunto de treinamento e logotipos de bancos diferentes. E para comparar os resultados a AUC da curva ROC gerada para cada um destes testes.

### 4.2 Banco de Dados

O banco de dados utilizado para testar os algoritmos é formado por imagens de cheques reais de vários tipos de bancos escaneadas em tons de cinza. A utilização de imagens em tons de cinza é devido a sua simplicidade em comparação a imagens coloridas. Dessa forma, podemos demonstrar de maneira mais eficiente o problema.

Para testar o sistema foram seleccionados os bancos do Brasil e Bradesco. E para cada um deles foi gerado um conjunto de imagens contendo os logotipos, como por exemplo, pode se visto na Figura 9, para treinamento e para testar a detecção, e um conjunto de imagens geradas de qualquer parte do cheque não contendo o logotipo e também contendo o logotipo de outros bancos não testados.



**Figura 9:** Exemplo dos logotipos utilizados para testar os algoritmos.

### 4.3 Resultados

Nesta seção serão mostrados os resultados dos testes propostos para avaliar o reconhecimento do logotipo do banco do Brasil e Bradesco. Para os algoritmos propostos foram realizados testes mantendo-se a mesma configuração do ambiente e apenas variando o número de imagens no conjunto de treinamento. Dessa forma, podemos verificar qual conjunto de treinamento se adequa melhor a cada algoritmo e também comparar os resultados dos algoritmos.

Para cada Banco será definido um janela do histograma espacial simples, que englobe o centro do logotipo. E será definido um conjunto de testes formado por noventa imagens de logos e 500 imagens de partes dos cheques que não possuem o logotipo.

E para cada um do melhor resultado testado dos algoritmos será feito uma análise mais detalhada do mesmo. Avaliando qual o melhor Threshold para o mesmo, definido para um falso positivo de 0.05% e observando quais imagens erradas foram consideradas como positivas.

#### 4.3.1 Banco do Brasil

Resultados dos testes utilizando os logotipos do Banco do Brasil.

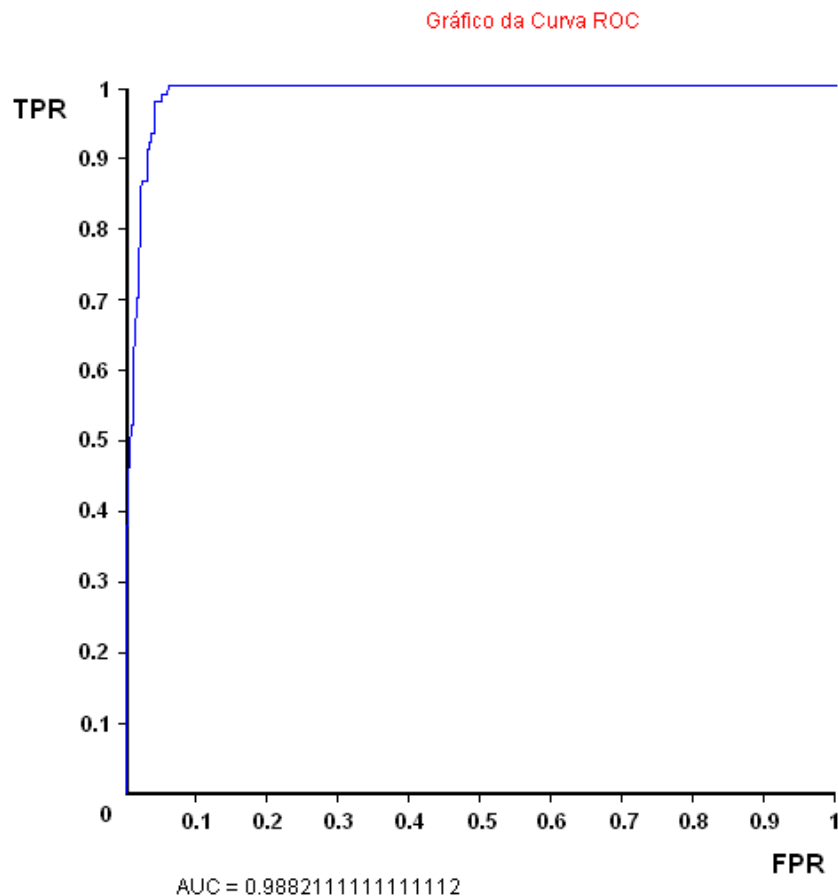
##### 4.3.1.1 - Classificador de Distância Mínima

Os resultados encontrados para o classificador de distância mínima para a detecção do logotipo do Banco do Brasil, variando o numero de imagens do conjunto de treinamento pode ser visualizado na Tabela 1.

**Tabela 1:** Resultado do classificador de distância mínima para o logotipo do Banco do Brasil

Numero do conjunto de treinamento	AUC
1	0.9869
2	0.9872
<b>3</b>	<b>0.9882</b>
4	0.9877
5	0.9878
6	0.9880
7	0.9860
8	0.9862
9	0.9849
10	0.9857

Como se pode observar o melhor resultado encontrado foi para o conjunto de treinamento contendo três imagens de logotipo do Banco do Brasil. A curva ROC do teste utilizando o conjunto de treinamento com três elementos pode ser visualizado no Gráfico 1. Com o threshold de 0.36 obtemos uma taxa de falso positivo de 0.05% e verdadeiro positivo de 98%.



**Gráfico 1:** Curva ROC gerada a partir de um conjunto de treinamento com três logotipos do Banco do Brasil para o classificador de distância mínima.

### 4.3.1.2 - Classificador Kappa

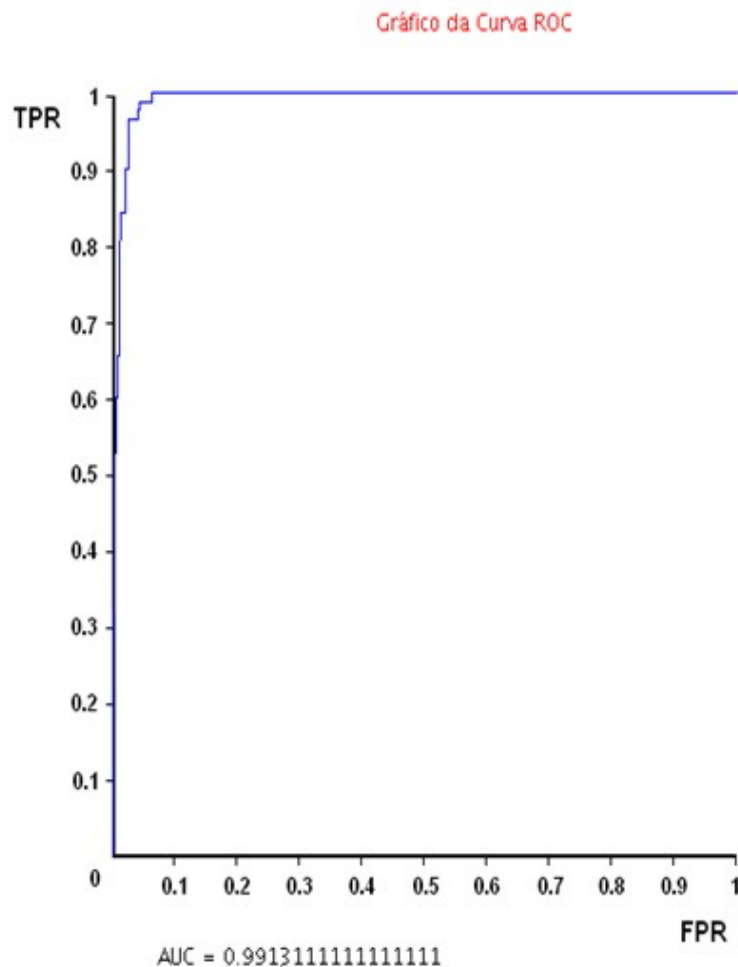
Os resultados encontrados para o classificador kappa para a detecção do logotipo do Banco do Brasil, variando o numero de imagens do conjunto de treinamento pode ser visualizado na Tabela 2.



**Tabela 2:** Resultado do classificador kappa para o logotipo do Banco do Brasil

Numero do conjunto de treinamento	AUC
1	0.9869
2	0.9913
3	0.9943
4	0.9929
5	0.9935
6	0.9952
7	0.9953
8	0.9955
<b>9</b>	<b>0.9957</b>
10	0.9956

Como se pode observar o melhor resultado encontrado foi para o conjunto de treinamento contendo nove imagens de logotipo do Banco do Brasil. A curva ROC do teste utilizando o conjunto de treinamento com nove elementos pode ser visualizado no Gráfico 2. Com o threshold de 0.28 obtemos uma taxa de falso positivo de 0.05% e verdadeiro positivo de 99%.



**Gráfico 2:** Curva ROC gerada a partir de um conjunto de treinamento com nove logotipos do Banco do Brasil para o classificador Kappa.

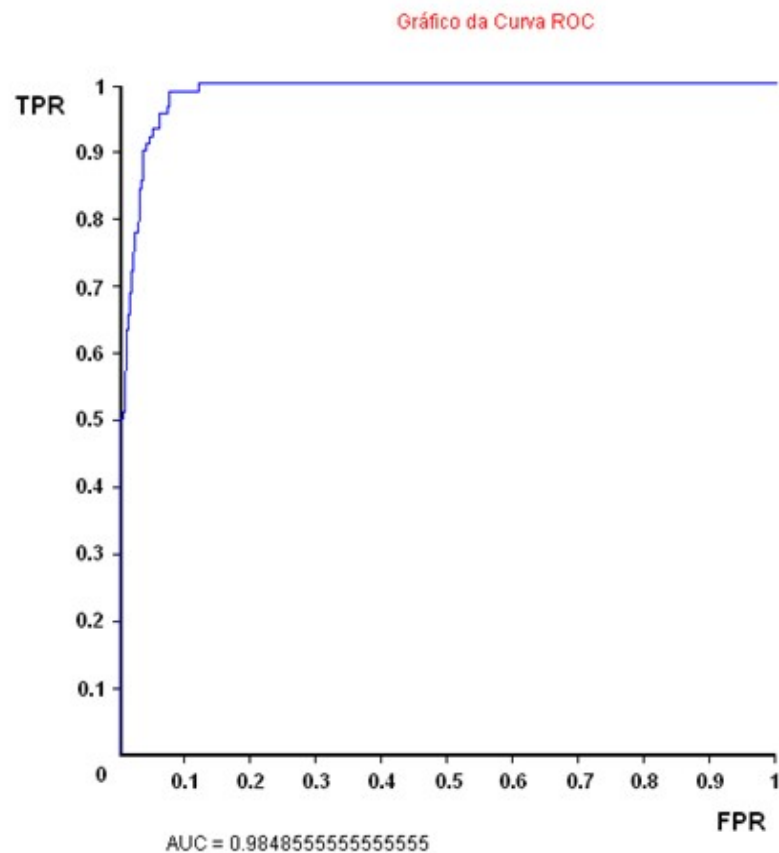
### 4.3.1.3 - Classificador Gamma

Os resultados encontrados para o classificador gamma para a detecção do logotipo do Banco do Brasil, variando o numero de imagens do conjunto de treinamento pode ser visualizado na Tabela 3.

**Tabela 3:** Resultado do classificador Gamma para o logotipo do Banco do Brasil

Numero do conjunto de treinamento	AUC
1	0.9869
2	0.9878
<b>3</b>	<b>0.9883</b>
4	0.9880
5	0.9877
6	0.9875
7	0.9852
8	0.9849
9	0.9836
10	0.9847

Como se pode observar o melhor resultado encontrado foi para o conjunto de treinamento contendo três imagens de logotipo do Banco do Brasil. A curva ROC do teste utilizando o conjunto de treinamento com três elementos pode ser visualizado no Gráfico 3. Com o threshold de 0.36 obtemos uma taxa de falso positivo de 0.05% e verdadeiro positivo de 97%.



**Gráfico 3:** Curva ROC gerada a partir de um conjunto de treinamento com três logotipos do Banco do Brasil para o classificador Gamma.

### 4.3.2 Banco Bradesco

Resultados dos testes utilizando os logotipos do Banco Bradesco.

#### 4.3.2.1 - Classificador de Distância Mínima

Os resultados encontrados para o classificador de distância mínima para a detecção do logotipo do Banco Bradesco, variando o numero de imagens do conjunto de treinamento pode ser visualizado na Tabela 4.

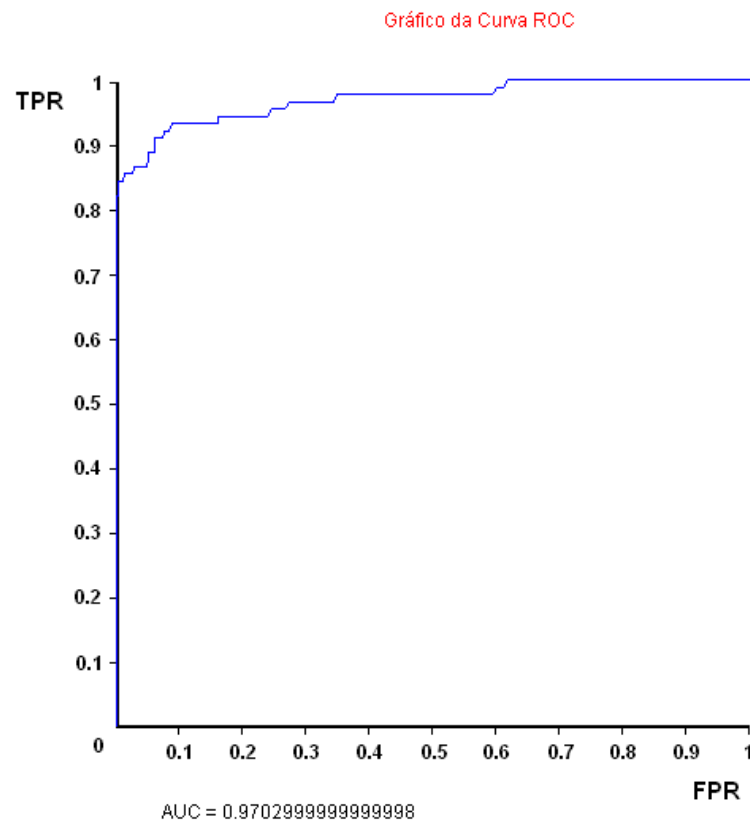
**Tabela 4:** Resultado do classificador de distância mínima para o logotipo do Banco Bradesco

Numero do conjunto de	AUC
-----------------------	-----

---

treinamento	
<b>1</b>	<b>0.97</b>
2	0.9661
3	0.9666
4	0.9545
5	0.9526
6	0.9441
7	0.9347
8	0.9257
9	0.9443
10	0.9339

Como se pode observar o melhor resultado encontrado foi para o conjunto de treinamento contendo uma imagem de logotipo do Banco Bradesco. A curva ROC do teste utilizando o conjunto de treinamento com um elemento pode ser visualizado no Gráfico 4. Com o threshold de 0.57 obtemos uma taxa de falso positivo de 0.05% e verdadeiro positivo de 89%.



**Gráfico 4:** Curva ROC gerada a partir de um conjunto de treinamento com três logotipos do Banco do Brasil para o classificador de distância mínima.

### 4.3.2.2 - Classificador Kappa

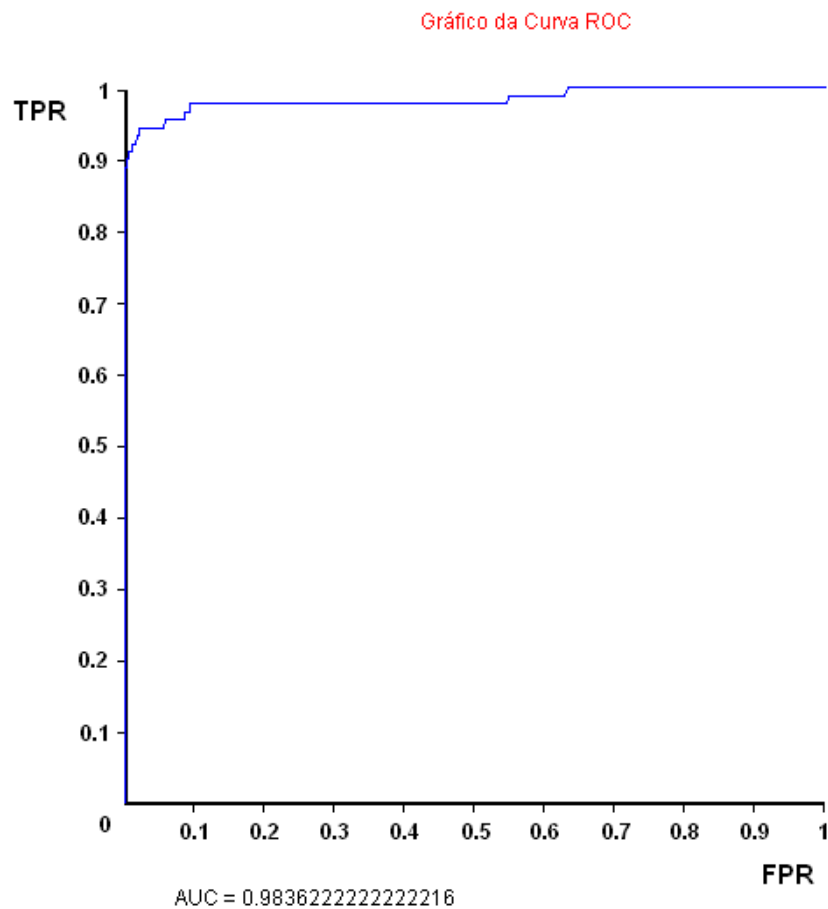
Os resultados encontrados para o classificador kappa para a detecção do logotipo do Banco Bradesco, variando o numero de imagens do conjunto de treinamento pode ser visualizado na Tabela 5.

**Tabela 5:** Resultado do classificador kappa para o logotipo do Banco Bradesco

Numero do conjunto de	AUC
-----------------------	-----

treinamento	
1	0.9732
2	0.9364
3	0.9796
4	0.9835
<b>5</b>	<b>0.9836</b>
6	0.9597
7	0.9435
8	0.946
9	0.9461
10	0.9475

Como se pode observar o melhor resultado encontrado foi para o conjunto de treinamento contendo cinco imagens de logotipo do Banco Bradesco. A curva ROC do teste utilizando o conjunto de treinamento com cinco elementos pode ser visualizada no Gráfico 5. Com o threshold de 0.46 obtemos uma taxa de falso positivo de 0.05% e verdadeiro positivo de 95%.



**Gráfico 5:** Curva ROC gerada a partir de um conjunto de treinamento com nove logotipos do Banco do Brasil para o classificador Kappa.

### 4.3.2.3 - Classificador Gamma

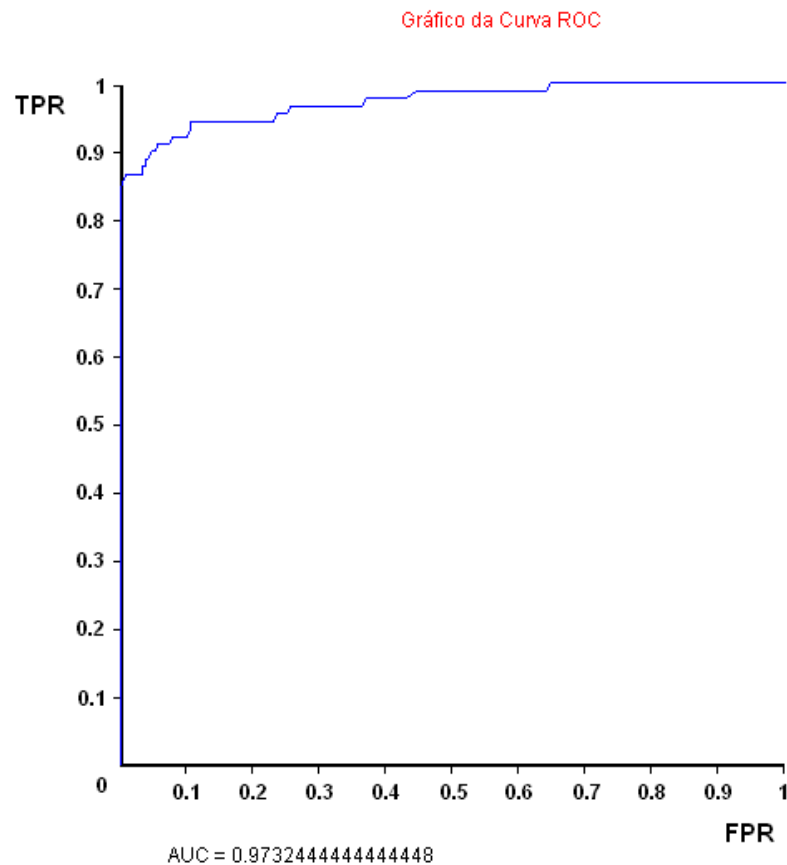
Os resultados encontrados para o classificador gamma para a detecção do logotipo do Banco Bradesco, variando o numero de imagens do conjunto de treinamento pode ser visualizado na Tabela 6.



**Tabela 6:** Resultado do classificador Gamma para o logotipo do Banco Bradesco

Numero do conjunto de treinamento	AUC
<b>1</b>	<b>0.9732</b>
2	0.9268
3	0.9629
4	0.9508
5	0.9565
6	0.9372
7	0.9258
8	0.9167
9	0.9329
10	0.9281

Como se pode observar o melhor resultado encontrado foi para o conjunto de treinamento contendo uma imagem de logotipo do Banco do Brasil. A curva ROC do teste utilizando o conjunto de treinamento com um elemento pode ser visualizado no Gráfico 6. Com o threshold de 0.58 obtemos uma taxa de falso positivo de 0.05% e verdadeiro positivo de 90%.



**Gráfico 6:** Curva ROC gerada a partir de um conjunto de treinamento com um logotipo do Banco Bradesco para o classificador Gamma.

### 4.3.3 Conclusão

A partir dos resultados obtidos podemos concluir que o número ideal de imagens do conjunto de treinamento muda de acordo como o objeto que está sendo detectado e também com o algoritmo de classificação que esta sendo utilizado.

Os melhores resultados encontrados para o logotipos utilizados foram:

- Banco do Brasil: Algoritmo de classificação Kappa com um conjunto de treinamento com nove imagens. Nesse caso, a AUC foi igual a 0.9957.

- Banco Bradesco: Algoritmo de classificação Kappa com um conjunto de treinamento com cinco imagens. Nesse caso, a AUC foi igual a 0.9836.

Portanto, podemos concluir que a utilização do algoritmo de classificação Kappa para a detecção de logotipos de bancos parece ser a melhor maneira. No entanto, não foi encontrado um padrão para o tamanho do conjunto de treinamento.

### 5. Considerações Finais

#### 5.1 Conclusões

A partir da implementação dos algoritmos propostos por Zhang et al. [3] e Harmeling et al. [4], foi possível obter um sistema de reconhecimento de logotipos bastante preciso e com uma taxa de falso positivo de 0.05% e de positivos verdadeiros acima de 90%. Ou seja, para o problema de detecção de logotipos de bancos os algoritmos propostos são suficientes.

O sistema pode ser estendido para detecção de outros objetos, pois a sua implementação não foi desenvolvida especificamente para o problema abordado. A utilização de histogramas espaciais pode ser utilizada para extração de características de diferentes tipos de objetos. No entanto, objetos mais complexos não devem obter a alta taxa de acertos conseguida com o estudo de caso selecionado.

#### 5.2 Trabalhos Futuros

Algumas extensões do estudo realizado podem ser visualizadas. Primeiramente, pode-se estender o problema não apenas para detecção de imagens em tons de cinza, e sim, para englobar imagens coloridas também. Para isto terá que ser feito algumas mudanças no processo de extração de características, pois, os histogramas espaciais como foram propostos nesse projeto são apenas para imagens em tons de cinza.

Outro fator a ser melhorado é a comparação de outros algoritmos de extração de características e métodos de decisões. Dessa maneira, poder gerar uma combinação que seja mais adequada e eficiente. Outro fator é a utilização de detecção de objetos mais complexos para testar a eficiência dos algoritmos.

## Referências Bibliográficas

- [1] GONZALEZ, R. C.; WOODS, R. E. **Processamento de Imagens Digitais**. 1. ed. São Paulo: E. Blücher, 2005. 509 p.
- [2] THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 3. ed. San Diego: Elsevier, 2006. 837 p.
- [3] ZHANG, H. et al. *Object detection using spatial histogram features*. **Image and Vision Computing**, v. 24, p. 327–341, 2006.
- [4] HARMELINGA, S. et al. *From outliers to prototypes: Ordering data*. **Neurocomputing**, v. 69, p. 1608-1618, 2006.
- [5] ZHANG, H.; ZHAO, D. *Spatial histogram features for face detection in color images*. **5<sup>th</sup> Pacific Rim Conference on Multimedia**, computer Science 3331 p. 377-384, 2004.
- [6] SCHIELE, B.; CROWLEY, J. L. *Object recognition using multidimensional receptive field histograms and its robustness to view point changes*. **ECCV**, v. 1, p. 610-619, 1996.
- [7] LOFTSGAARDEN, D.; QUESENBERY, C. *A nonparametric estimate of a multivariate density function*, **Ann. Math**, v. 36, p. 1049-1051, 1965.
- [8] PIETIKÄNEN, M.; OJALA, T.; XU, Z. *Rotation-invariant texture classification using feature distributions*, **Pattern Recognition**, v. 33, p. 43-52, 2000.
- [9] PIETIKÄNEN, M.; OJALA, T.; MÄENPÄ, T. *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, p. 971-987, 2002.

- [10] SWAIN, M.; BALLARD, D. *Color indexing*, **International Journal of Computer Vision**, v. 7, p. 11-32, 1991.
- [11] ZHU, G.; DOERMANN, D. *Automatic document logo detection*. **International Conference on Document Analysis and Recognition**, p. 864-868, 2007.
- [12] CRANDALL, D.; LUO, J. *Robust color object detection using spatial-color joint probability functions*. **CVPR**, v. 1, p. 379-385, 2004.
- [13] BERGBOER, N.H; POSTMA, E.O., HERIK, H.J. *Context-based object detection in still images*. **Image Vision Comput**, v. 24, p. 987-1000, 2006.
- [14] BOSH, A.; MUÑOZ, X.; MARTÍ, R. *which is the Best way to organize/classify images by content?* **Image Vision Comput**, v. 25, p. 778-791, 2007.
- [15] ROWLEY, H.A; BALUJA, S; KANADE, T. *Neural network-based face detection*. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 20, p. 29–38. 1998.
- [16] GARCIA, C.; DELAKIS, M. *Convolutional face finder: a neural architecture for fast and robust face detection*. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 26, p. 1408–1423, 2004.
- [17] OSUNA, E.; FREUND, R.; GIROSI, F. *Training support vector machines: an application to face detection*. **IEEE Conference on Computer Vision and Pattern Recognition**, p. 130–136, 1997.
- [18] PAPAGEPRGIOU, C.P.; POGGIO, T. *A training object system: car detection in static images*. **MIT AI Memo**, v. 180, 1999.
- [19] VIOLA, P.; JONES, M. *Rapid object detection using a boosted cascade of simple features*. **IEEE Conference on Computer Vision and Pattern Recognition**, p. 511–518, 2001.

- [20] LI, S.Z. et al. *Statistical learning of multi-view face detection*. **European Conference on Computer Vision**, v. 4, p. 67–81, 2002.
- [21] CHEN, X.R.; YUILLE, A. *Detecting and reading text in natural scenes*. **IEEE International Conference on Computer Vision and Pattern Recognition**, p. 366–373, 2004.
- [22] MOHAN, A.; PAPAGEORGIOU, C.; POGGIO, T. *Example-based object detection in images by components*. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 23, p. 349–361, 2001.
- [23] FERGUS, R.; PERONA, P.; ZISSERMAN, A. *Object class recognition by unsupervised scale-invariant learning*. **International Conference on Computer Vision and Pattern Recognition**, p. 264–271, 2003.
- [24] LIN, Y.Y.; LIU, T.L. *Robust face detection with multi-class boosting*, **IEEE International Conference on Computer Vision and Pattern Recognition**, p. 680–687, 2005.
- [25] AMIT, Y.; GEMAN, D.; FAN, X.D. *A coarse-to-fine strategy for multiclass shape detection*. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 26, p. 1606–1621, 2004.
- [26] TU, Z.W. et al. *Image parsing: unifying segmentation, detection and recognition*. **International Journal of Computer Vision**, v. 63, p. 113–140, 2005.
- [27] SCHIELE, B. **Object recognition using multidimensional receptive field histograms**. Tese de PhD, I.N.P. Grenoble, 1997.
- [28] LOWE, D.G. *Distinctive image features from scale-invariant keypoints*. **International Journal of Computer Vision**, v. 60, p. 91–110, 2004.

- [29] HADID, A.; PIETIKÄINEN, M.; AHONEN, T. *A discriminative feature space for detecting and recognizing faces*, **IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, p. 797–804, 2004.
- [30] ZYGA, K.; SCHROEDER, J.; PRICE, R. *Logo recognition using retinal coding*. **In Proc. 38th Asilomar Conf. Signals, Systems and Computers**, v. 2, p. 1549–1553, 2004.
- [31] FAWCETT, T. *An introduction to ROC analysis*. **Pattern Recognition Letters**, v. 27, p. 861-874, 2006.