

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE COMPUTAÇÃO
BACHARELADO EM CIÊNCIAS DA COMPUTAÇÃO

MARCOS ROBERTO E SOUZA

UMA ANÁLISE DE DIFERENTES ABORDAGENS
PARA O RECONHECIMENTO OFF-LINE DE
ESTILOS DE ESCRITA EM DOCUMENTOS
MANUSCRITOS

TRABALHO DE CONCLUSÃO DE CURSO

CAMPO MOURÃO
2015

MARCOS ROBERTO E SOUZA

UMA ANÁLISE DE DIFERENTES ABORDAGENS
PARA O RECONHECIMENTO OFF-LINE DE
ESTILOS DE ESCRITA EM DOCUMENTOS
MANUSCRITOS

Trabalho de Conclusão de Curso de graduação do
Curso Superior de Bacharelado em Ciências da
Computação da Universidade Tecnológica Fede-
ral do Paraná - UTFPR.

Orientador: Diego Bertolini Gonçalves, Dr.

CAMPO MOURÃO

2015

RESUMO

O reconhecimento de estilos de escrita é uma etapa necessária para automatizar o reconhecimento fim de um sistema multi-estilos. A tarefa de reconhecimento de estilos de escrita é desafiadora devido a similaridade entre alguns estilos. Desta forma, o principal objetivo deste trabalho é realizar uma análise do reconhecimento de estilos de escrita em documentos manuscritos considerando técnicas robustas indiferente do estilo de escrita considerado. Comparamos duas abordagens para o reconhecimento de estilos de escrita, sendo que a primeira consiste na utilização dos descritores de textura GLCM, LBP, LPQ e SURF em conjunto com o classificador SVM, enquanto na segunda abordagem utilizaremos uma técnica de aprendizagem profunda descrita como Rede Neural Convolutiva (CNN). Além da comparação, analisamos o impacto da quantidade de informação presente nos documentos manuscritos, a relação existente entre o número de classes (estilos de escrita) e a taxa de reconhecimento e o desempenho dos descritores de textura para esta abordagem. Avaliaremos ainda a importância do processo de compactação de escrita para esta aplicação. Experimentos foram realizados em uma base de dados com cinco classes, nos quais conseguimos taxas de acerto similares as apresentadas na literatura em ambas abordagens, a melhor taxa média de acerto foi de 98,48% utilizando o descritor de textura SURF a partir do documento original. Por meio dos experimentos realizados na primeira abordagem, concluímos que o ganho da compactação de escrita e divisão em blocos está estritamente relacionado ao descritor utilizado. Enquanto que para a segunda, na qual obtemos a melhor taxa média de 91,43%, chegamos a conclusão de que a utilização da compactação de escrita não possui grande impacto, desde que uma quantidade de informação suficiente seja fornecida. Mesmo que a segunda abordagem atingido taxas inferiores, ambas abordagens conseguem taxas de acerto relevantes para o problema. [Palavras chave: Reconhecimento de estilos de escrita, aprendizagem de máquina, reconhecimento de padrões]

LISTA DE FIGURAS

1	Trechos de Imagens de Documentos com Estilos de Escrita. (a) Bengali, (b) Oriá, (c) Persa e (d) Romano	p. 4
2	Exemplo de Matriz de Co-ocorrência para uma imagem com 4 níveis de cinza com $d = 1$ e $\theta = 0^\circ$	p. 13
3	Exemplo de operação básica do LBP com valor resultante 173. Adaptada de (AMARAL; THOMAZ, 2011).	p. 15
4	Diferentes Valores para P e R . Retirada de (GONÇALVES, 2014).	p. 15
5	Ideia Geral da Técnica SVM com Vetores de Suporte Circulados. Retirada de (GONÇALVES, 2008).	p. 18
6	Visão Geral da Rede Neural Convolucional. Retirada de (CIREŞAN et al., 2011) ..	p. 19
7	Visão Geral do Método Proposto	p. 21
8	Linhas dos Estilos de Escrita Utilizados. (a) Bengali, (b) Oriá, (c) Persa, (d) Romano e (e) Canará	p. 22
9	Etapas de Pré-Processamento	p. 22
10	Blocos 256×256 da Carta Original e da Compactada	p. 23
11	Abordagem I - Descritor de Textura + Classificador SVM.....	p. 24
12	Abordagem II - Aprendizagem Profunda com CNN.	p. 24
13	Método da Divisão de Blocos sem Sobreposição. (a) horizontal (b) vertical	p. 26
14	Método da Divisão de Blocos com Sobreposição. (a) horizontal (b) vertical	p. 26
15	BoxPlot para LBP	p. 28
16	Cartas Originais \times Compactada	p. 29
17	Visão Geral da Arquitetura Utilizada nos Experimentos	p. 30

LISTA DE TABELAS

1	Tabela com Relação entre Idiomas e Estilos de Escrita Indianos. Adaptada de (OBAIDULLAH et al., 2015)	p. 3
2	Resumo da Revisão Bibliográfica	p. 11
3	Base de Dados	p. 22
4	Quantidade de Documentos Manuscritos Utilizados em Cada Estilo de Escrita ..	p. 25
5	Quantidade de Blocos Gerados em Cada Estilo de Escrita	p. 26
6	Taxa de Reconhecimento (%) com Aumento de Classes	p. 27
7	Taxa de Reconhecimento (%) Utilizando Compactação de Escrita	p. 27
8	Taxa de Reconhecimento (%) para Cartas Originais	p. 28
9	Quantidade de Documentos Utilizados em Cada Estilo de Escrita	p. 30
10	Taxa de Reconhecimento (%) Utilizando Compactação.....	p. 31
11	Taxa de Reconhecimento (%) para Carta Originais	p. 31

SUMÁRIO

1 INTRODUÇÃO	p. 1
1.1 Análise de Documentos em Imagens	p. 2
1.2 Problemática	p. 3
1.3 Objetivos e Questões de Pesquisa	p. 4
1.4 Organização	p. 5
2 ESTADO DA ARTE	p. 6
2.1 Abordagens Usando Bases com Textos Manuscritos	p. 6
2.2 Abordagens Usando Bases com Textos Datilografados	p. 8
2.3 Considerações Finais	p. 9
3 FUNDAMENTAÇÃO TEÓRICA	p. 12
3.1 Descritores de Textura	p. 12
3.1.1 Matriz de Co-ocorrência de Níveis de Cinza (GLCM)	p. 13
3.1.2 Padrões Binários Locais (LBP)	p. 15
3.1.3 Quantização Local de Fase (LPQ)	p. 16
3.1.4 SURF	p. 16
3.2 Classificadores Tradicionais	p. 17
3.2.1 Máquina de Vetores de Suporte (SVM)	p. 18
3.3 Aprendizagem Profunda	p. 18
3.3.1 Rede Neural Convolucional (CNN)	p. 19
3.4 Considerações	p. 20
4 MÉTODO	p. 21
4.1 Descrição da Base de Dados	p. 21

4.2	Pré-Processamento	p. 22
4.3	Reconhecimento	p. 23
5	EXPERIMENTOS	p. 25
5.1	Abordagem I	p. 25
5.2	Abordagem II	p. 29
6	CONCLUSÕES	p. 32
6.1	Contribuições	p. 32
6.2	Trabalhos Futuros	p. 33
	Referências	p. 34

1 INTRODUÇÃO

Um padrão, no contexto deste trabalho, pode ser visto como uma base de comparação consagrada como modelo por consenso geral (HOUAISS, 2001). Os padrões se manifestam na natureza de diferentes formas, como em sons, imagens, odores, formas e sabores, os quais são percebidos e interpretados naturalmente pelos seres humanos e por outros animais. Entretanto, a simples tarefa de perceber um padrão pode ser árdua e complexa para uma máquina, que em seu primórdio, foi proposta com o intuito único de realizar cálculos.

Na computação, a área de Reconhecimento de Padrões vem sendo estudada ao longo das últimas décadas a fim de conceder ao computador essa habilidade natural dos seres humanos. Entre as definições existentes na literatura, destacaremos duas delas. Para Theodoridis (THEODORIDIS; KOUTROUMBAS, 1999), Reconhecimento de Padrões é uma área de pesquisa que tem por objetivo a classificação de objetos em um número de categorias ou classes. Enquanto que para Duda (DUDA et al., 2012), é o ato de observar os dados brutos e tomar uma ação baseada na categoria de um padrão.

Dentre as diversas subáreas de computação, o Reconhecimento de Padrões relaciona-se principalmente com a Inteligência Artificial (IA). A IA segundo Luger (LUGER, 2004), pode ser definida como o ramo da Ciência da Computação que se ocupa do comportamento inteligente, já segundo Rich (RICH; KNIGHT, 1994), um estudo de como fazer os computadores realizarem coisas que, atualmente, os humanos fazem melhor.

Dentre os conceitos básicos de Reconhecimento de Padrões apresentamos três:

- **Características:** Do inglês *features*, pode ser definida como os dados extraídos de uma instância por meio de alguma medida, geralmente apresentados como vetor de características. Tais medidas devem maximizar as semelhanças intraclasse e as diferenças extraclasses.
- **Classe:** Conjunto de instâncias que possuem características em comum.
- **Classificação:** Ato de rotular instâncias de acordo com suas características.

No caso de reconhecer padrões em imagens e ou vídeos, o Reconhecimento de Padrões relaciona-se com o Processamento Digital de Imagens (PDI). Para Young, o PDI caracteriza-se por uma série de atividades executadas ordenadamente em que tanto a entrada, quanto a saída são imagens (YOUNG et al., 1998). Tendo como premissa tais definições, é possível definir o Reconhecimento

de Padrões em Imagens como uma subárea específica de IA, que utiliza técnicas de PDI com o objetivo de reconhecer ou identificar padrões. Entretanto devemos salientar que muitos problemas de reconhecimento de padrões podem não utilizar imagens, um exemplo seria identificar padrões em sons.

As técnicas de Reconhecimento de Padrões em imagens podem ser aplicadas nos mais variados tipos de problemas, podendo ter objetivos diferentes, como: reconhecimento de face (AHONEN et al., 2004), rastreamento de objetos (YILMAZ et al., 2006), reconhecimento de objetos (LOWE, 1999) dentre outros. Comumente vemos pesquisas nas quais as imagens referem-se a documentos de texto, nestes casos o exemplo mais comum são sistemas do tipo *Optical Character Recognition* (OCR), entretanto, a utilização de documentos de texto pode ser empregado nos mais variados tipos de sistemas, como: Verificação de assinaturas, reconhecimento de escritores, análise e reconhecimento da autoria de documentos, entre outros (BERTOLINI et al., 2010). Para o reconhecimento de imagens em documentos de texto, utiliza-se especificamente o nome Análise de Documentos em Imagens na qual este trabalho esta inserido.

1.1 Análise de Documentos em Imagens

A Análise de Documentos em Imagens é o processo que realiza a interpretação de imagens de documentos, aplicando técnicas e algoritmos a imagens de documentos (composta por *pixels*) a fim de obter uma descrição legível para o computador (SALVI, 2014). Ao descrever sobre Análise de Documentos em Imagens, duas categorias podem ser definidas: Processamento de Texto e Processamento Gráfico.

O Processamento de Texto, o qual empregaremos neste trabalho, abrange desde determinar a inclinação, encontrar parágrafos, colunas, linhas e palavras, até o reconhecimento do texto em si. Uma tecnologia muito conhecida e que possui diversas aplicações em Análise de Documentos em Imagem é o OCR. Dentre suas aplicações, um exemplo notável é o reconhecimento de caracteres manuscritos (SALVI, 2014). O Processamento Gráfico, por sua vez, lida com imagens que contêm elementos não-textuais como tabelas, símbolos, logos, entre outros.

O reconhecimento de caracteres pode ser dividido em duas grandes áreas, texto datilografado e texto manuscrito. Normalmente, textos manuscritos implicam em uma maior dificuldade no reconhecimento de caracteres por possuírem uma maior variação (SALVI, 2014), devido a diversos fatores, que vão desde o tipo da caneta e do papel até fatores como a habilidade, estilo e estado psicológico do escritor. Os problemas de Reconhecimento de Padrões em textos manuscritos podem ser divididos em *on-line* e *off-line*. Em abordagens empregando técnicas *on-line*, dispositivos especiais são usados para monitorar o movimento da caneta e registrar informações temporais, en-

quanto que no reconhecimento *off-line*, uma imagem do texto manuscrito é digitalizada e gravada.

1.2 Problemática

Um estilo de escrita pode ser definido como um conjunto de notações que representam um idioma, ou um conjunto de idiomas. Em países que possuem mais de um estilo de escrita oficial, como é o caso da Índia, na qual mais de 22 idiomas e 13 diferentes estilos de escrita são oficiais (OBAIDULLAH et al., 2015), o uso de documentos com mais de um estilo de escrita, especialmente dois e três estilos, é bastante comum. Na Tabela 1, é possível verificar a relação entre idiomas e estilos de escrita indianos, demonstrando quais idiomas são representados por qual escrita, além de informar para cada estilo de escrita o número de escritores em milhões.

Tabela 1: Tabela com Relação entre Idiomas e Estilos de Escrita Indianos. Adaptada de (OBAIDULLAH et al., 2015)

No.	Idioma	Estilo de Escrita	Num. Escritores (M)	No.	Idioma	Estilo de Escrita	Num. Escritores (M)	
1	Assamesa	Bengali	211,5	14	Dogri	Dogri	3,8	
2	Bengali			15	Gujarati	Gujarati	46,5	
3	Manipuri			16	Canará	Canará	3,63	
4	Bodo	Devanágari	328,23	17	Caxemira	Caxemira	5,6	
5	Hindi			18	Malaiala	Malaiala	35,9	
6	Concani			19	Oriá	Oriá	31,7	
7	Maithili			20	Punjabi	Punjabi	1,05	
8	Marati			21	Tamil	Tamil	65,7	
9	Nepali			22	Telugu	Telugu	69,8	
10	Sânscrito			23	Urdu	Urdu	60,6	
11	Sindi							
12	Santali			Romano	334,2			
13	Inglês							

O sucesso de uma técnica qualquer para o reconhecimento de caracteres é altamente dependente de um reconhecimento prévio do estilo de escrita/idioma em questão (HOCHBERG et al., 1999). Assim, para um melhor desempenho de um sistema OCR é indispensável que este consiga reconhecer previamente o estilo da escrita. Desta maneira, o reconhecimento prévio do estilo de escrita de determinado trecho ou do documento é uma etapa essencial para uma maior automatização do sistema. Isso torna-se ainda mais importante quando uma quantidade abundante de estilos de escrita são utilizados em um mesmo país ou em uma mesma região, como é o caso da Índia.

Um aspecto que torna esta pesquisa desafiadora deve-se a similaridade entre alguns estilos de escrita. Na Figura 1 apresentamos trechos de diferentes documentos manuscritos com estilos de escrita distintos. Três destes estilos são largamente utilizados na Índia, sendo eles: Bengali, Oriá e Romano. Facilmente podemos distinguir o estilo Romano dos demais, entretanto percebe-se uma maior similaridade entre os outros estilos.

Analisando um grupo maior de diferentes estilos de escrita, percebe-se que geralmente existe similaridades entre estilos de escrita de uma determinada região.

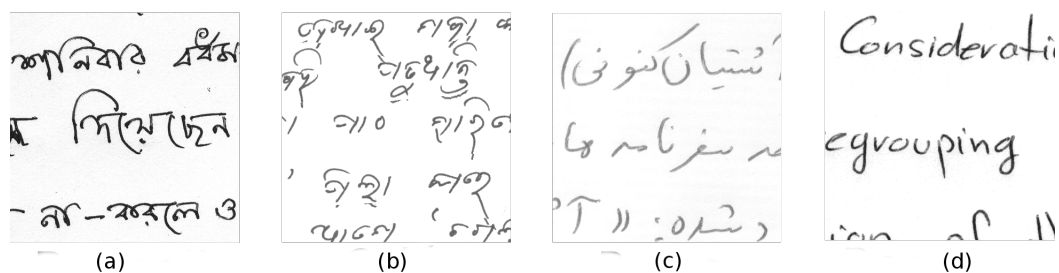


Figura 1: Trechos de Imagens de Documentos com Estilos de Escrita. (a) Bengali, (b) Oriá, (c) Persa e (d) Romano

1.3 Objetivos e Questões de Pesquisa

Este trabalho tem como principal objetivo realizar uma análise para o reconhecimento *off-line* de estilos de escrita utilizando a textura da escrita para representar uma determinada classe. Este objetivo principal está atrelado aos seguintes objetivos secundários.

- Verificar o desempenho de diferentes descritores de textura *Gray Level Co-ocurrence Matrix* (GLCM), *Local Binary Patterns* (LBP), *Local Phase Quantization* (LPQ) e *Speeded-up Robust Features* (SURF) empregando o classificador *Support Vector Machine* (SVM) (Abordagem I);
- Comparar o desempenho apresentado pela Abordagem I com blocos de escrita compactada gerados a partir da abordagem de Hanusiak (HANUSIAK et al., 2012) *versus* blocos da imagem original pré-processada;
- Empregar uma segunda abordagem usando aprendizagem profunda com a *Convolutional Neural Network* (CNN) (Abordagem II);
- Analisar o impacto do tamanho dos blocos e a quantidade de blocos extraídos;
- Comparar o desempenho das duas abordagens.

Desta forma, este trabalho tem como objetivo avaliar o desempenho do reconhecimento de estilos de escrita, além de avaliar o desempenho de duas abordagens para resolução deste problema. Considerando os objetivos apresentados, as questões que este trabalho dispõe-se a responder são apresentadas a seguir.

- A utilização da compactação da escrita como proposto por Hanusiak em (HANUSIAK et al., 2012), contribui para uma melhora no desempenho se comparada a não utilização desta?
- Dividir a imagem em blocos para seu reconhecimento tem forte influência no resultado? Qual tamanho do bloco mostra-se mais adequado? A quantidade de blocos extraídos mostra-se influente?
- A abordagem de aprendizagem profunda usando CNN é razoavelmente boa ou até superior as abordagens descritas na literatura?

Pode-se considerar que a originalidade deste trabalho está no fato de utilizarmos um esquema para a compactação da escrita o qual contribui para uma melhor representação da classe (GONÇALVES, 2014). Esta abordagem torna-se interessante pois não é necessário uma etapa de segmentação da imagem a fim de encontrar detalhes da escrita. Ainda para a Abordagem I, temos a aplicação do descritor de textura SURF, em que não encontrou-se trabalhos na literatura que o aplicam a este problema. Aliado a isto temos a investigação de uma abordagem na qual não necessitamos a etapa de extração de características (Abordagem II) a qual recentemente tem movido diversas pesquisas (RUSSAKOVSKY et al., 2014; KARPATY et al., 2014; KRIZHEVSKY et al., 2012). Embora a problemática concentrou-se mais nos estilos de escrita indiano, os métodos analisados neste trabalho não dependem dos estilos de escrita considerados e devem alcançar resultados equivalentes para os outros estilos de escrita.

1.4 Organização

Este documento está organizado da seguinte forma. O Capítulo 2, apresenta uma visão geral sobre o estado da arte a partir de alguns trabalhos sobre o reconhecimento de estilos de escrita. No Capítulo 3, é feito um estudo sobre importantes tópicos que contribuirão para um melhor entendimento em relação aos métodos computacionais utilizados neste trabalho. No Capítulo 4, o método proposto para o desenvolvimento deste trabalho é explicado em detalhes. Por fim, os experimentos preliminares são exibidos e analisados no Capítulo 5, seguido pelas referências.

2 ESTADO DA ARTE

Neste capítulo, apresenta-se uma revisão da literatura sobre reconhecimento de estilos de escrita. Análises de trabalhos usando textos manuscritos *off-line* e datilografados foram realizadas, os quais são divididos e expostos respectivamente nas Seções 2.1 e 2.2. Para cada trabalho, foi realizado um estudo sobre a base de dados e a abordagem empregada, incluindo principalmente detalhes sobre as características e o método de classificação utilizado. Por fim, são apresentados as melhores taxas de acerto. A não existência de bases de dados bem estabelecida para o problema implica em uma dificuldade na comparação do desempenho obtido pelas técnicas propostas nesses trabalhos. A seguir apresentamos alguns trabalhos relevantes da literatura.

2.1 Abordagens Usando Bases com Textos Manuscritos

Em (HOCHBERG et al., 1999), um sistema para o reconhecimento de estilos de escrita a nível de bloco foi desenvolvido, ou seja, o reconhecimento é aplicado visando identificar uma parte do documento de tamanho pré-definido denominada bloco. Outros trabalhos realizam este reconhecimento a nível de palavra ou linha, nos quais o objetivo é identificar respectivamente a palavra ou a linha. Neste trabalho, considerou-se 496 documentos escritos por 281 escritores e divididos em seis estilos de escrita: Árabe, Chinês, Cirílico, Devanágari, Japonês e Romano. Para cada documento, componentes conexos são extraídos considerando oito vizinhos conectados. Em seguida, cinco características são extraídas em cada componente: Centróide relativo em Y, centróide relativo em X, número de lacunas brancas (nos caracteres), esfericidade e proporção de aspectos. Essas características foram escolhidas por maximizar a similaridade (segundo o olhar dos autores) diferenciando aspectos específicos das classes utilizadas. Para cada par de estilos de escrita possíveis da base, uma função discriminante linear de Fisher foi treinada. Os documentos foram classificados aplicando as funções obtidas ao vetor de características de cada documento. Por fim, o documento é atribuído a classe que receber o maior número de votos. A taxa de reconhecimento foi de 88%.

A aplicação de um pré-processamento nas imagens de documentos de entrada foi proposta em (SINGHAL et al., 2003) utilizando 480 documentos escritos por 360 escritores diferentes, divididos igualmente entre quatro estilos de escrita: Romano, Devanágari, Bengali e Telugu. O pré-processamento deu-se por cinco passos na seguinte ordem: Remoção de ruídos, esqueletização morfológica, *pruning*, a conectividade-m, que encontra as componentes de uma imagem por

meio da adjacência-m e por fim, a normalização no tamanho do texto. Essa técnica visa diminuir as variações existentes em textos manuscritos a fim de aproximar a dificuldade desse problema com o de textos datilografados. O Filtro de Gabor Multicanal foi utilizado como características de textura invariante a rotação. Em seguida, para cada classe realizou-se uma abordagem probabilística para clusterização, a motivação para o uso dessa surgiu devido à grande disparidade intraclasse das quatro classes utilizadas nesse trabalho. Os centroides de cada *cluster* obtidos na etapa anterior foram utilizados como instâncias daquela classe pelo classificador. Os autores reportam uma taxa de acerto de 91,6%.

Dois sistemas para o reconhecimento de estilos de escrita foram propostos em (DHANDRA; HANGARGE, 2007) considerando uma base quem contém 3000 palavras e 400 números escritos por 250 escritores. No primeiro, utiliza-se características globais e locais extraídas por filtros morfológicos e descritores de região para identificar três estilos de escrita: Canará, Romano e Devanágari. Usando uma abordagem a nível de palavra e utilizando o classificador *k Nearest Neighbors* (k-NN), obteve-se a taxa de acerto de 96,05%. Enquanto que no segundo sistema, classificou-se estilos de escrita em documentos que só possuem números e não possuem letras, considerando os estilos de escrita Canará e Romano, com isso obteve-se 99% de acerto.

A textura foi empregada como característica para o reconhecimento de estilos de escrita a nível de blocos em (HANGARGE; DHANDRA, 2010). Neste caso os autores consideraram 150 documentos divididos em 300 blocos, com 100 blocos para cada um dos três estilos de escrita utilizados, sendo eles: Devanágari, Romano e Urdu. No processo de extração de características tentou-se obter a densidade dos traços de tinta e a densidade dos *pixels* como característica. O k-NN foi utilizado para classificação e com isso as taxas obtidas foram de 99,2% para dois estilos e 88,6% para cenários com três estilos de escrita.

Em (HIREMATH et al., 2010), um sistema para o reconhecimento de estilos de escrita a nível de bloco utilizando características de textura foi proposto. Nesse trabalho, considerou-se 4000 documentos divididos entre oito estilos de escrita: Canará, Tamil, Urdu, Telugu, Bengali, Hindi, Malaiala e Romano. As características foram extraídas baseadas na co-ocorrência de histogramas de imagens decompostas pela *wavelet*, que captura a informação sobre as relações entre cada frequência alta de sub-banda e a baixa frequência de sub-banda da imagem transformada ao nível correspondente. Para a classificação, o k-NN foi utilizado. Experimentos foram realizados variando o número de escritores por estilo de escrita de um a três, obtendo taxas médias de 97,5%, 91,8% e 79,5% respectivamente, o que demonstra uma grande dependência do sistema proposto em relação ao escritor.

Em (OBAIDULLAH et al., 2013), utilizou-se seis estilos de escrita populares da Índia. Foram utilizados 152 documentos no total, divididos entre seis estilos de escrita: Bengali, Devanágari,

Malaiala, Urdu, Oriá e Romano. Sendo 32 documentos Bengali e 24 para cada um dos demais estilos de escrita. Diversas características foram extraídas, as quais foram divididas em características abstratas/matemáticas, características baseadas na estrutura e características dependentes do estilo de escrita. Uma Rede Neural Perceptron Multi-camada foi aplicada a fim de realizar a classificação. No melhor caso o sistema apresenta taxa de reconhecimento de 92,8%.

Uma abordagem empregando diversos classificadores para reconhecimento de estilos de escrita a nível de linha e palavra foi proposta por Ferrer (FERRER et al., 2014), considerando uma base com 1909 linhas e 15481 palavras divididos em três estilos de escrita: Bengali, Persa e Romano. As características foram obtidas a partir da concatenação de histogramas gerados pela técnica LBP obtidos a partir da divisão horizontal da palavra ou da linha. Para a classificação, inicialmente uma medida para estimar a dimensão da escrita foi definida e denominada *Word Information Index* (WII). Em seguida, um classificador para cada dimensão da escrita foi treinado com as palavras que possuem uma quantidade semelhante de informações. Para estimar o estilo de escrita de determinada palavra, a dimensão desta é obtida para então aplicar o classificador treinado para tal dimensão. A melhor taxa obtida ao comparar-se os três estilos de escrita foi de 89,89%, além disso os resultados mostraram a notória relação linear entre a dimensão da escrita e as taxas de acerto.

Recentemente, em (PARDESHI et al., 2014), uma técnica de reconhecimento de estilos de escrita a nível de palavra foi proposta considerando uma base de dados com 28100 palavras. As transformadas de Radon, *wavelet* discreta e discreta do cosseno, além de filtros estatísticos foram aplicados a fim de extrair características espaciais de multi-resolução direcional. Os classificadores SVM e k-NN foram aplicados em onze estilos de escrita indianos: Romano, Devanágari, Urdu, Canará, Oriá, Gujrati, Bengali, Gurumukhi, Tamil, Telugu e Malaiala. Nos experimentos executados, a melhor taxa descrita é de 98% e 96% para cenários com dois e três estilos de escrita respectivamente.

2.2 Abordagens Usando Bases com Textos Datilografados

Um método a nível de linha para o reconhecimento de estilos de escrita em documentos impressos foi proposto em (PAL et al., 2003) considerando 4000 linhas divididas igualmente em onze estilos de escrita: Bengali, Devanágari, Romano, Gurumukhi, Malaiala, Tamil, Telugu, Gujrathi, Canará, Urdu e Oriá. As características utilizadas foram escolhidas usando as seguintes considerações: (a) Presença de caracteres em alguns estilos de escrita e ausência de caracteres em pelo menos um estilo de escrita; (b) Robustez, precisão e simplicidade de detecção; (c) Velocidade de computação; (d) Independência de fontes, tamanho e estilo do texto. No reconhecimento dos estilos de escrita, uma Árvore Binária foi criada, na qual, os nós intermediários representam condições

binárias dependentes de características específicas e pré-definidas, enquanto que os nós folhas representam onze as classes. Desta forma, para testar uma instância, a árvore é percorrida a fim de encontrar um nó folha. Nos experimentos realizados, a taxa obtida foi de 99,2%.

Um sistema para o reconhecimento de estilos de escrita e idioma a nível de bloco foi proposto em (PAN; TANG, 2011) considerando 4500 blocos divididos igualmente em seis classes: Chinês, Inglês, Francês, Coreano, Japonês e Russo. Essa divisão não considera somente o estilo de escrita, mas também o idioma. Utilizou-se a decomposição bidimensional de modo empírico para decompor as imagens em alguns componentes, e então o LBP foi utilizado extrair as características dos componentes. Um classificador baseado no SVM foi utilizado nesse trabalho. Nos experimentos realizados, a taxa de acerto obtida foi de 95,41%, e por meio de uma comparação realizada entre o método proposto, *wavelet* baseada em características de energia e LBPV (VERIFICAR SIGLA), foi possível verificar que o método proposto é mais robusto em relação a rotação.

Em (DAS et al., 2012), foi proposto um modelo para identificar estilos de escrita a nível de palavra, considerando uma base com 1409 palavras divididas em três estilos de escrita: Telugu, Romano e Hindi. Sete características foram extraídas, e a classificação foi realizada por meio de heurísticas pré-definidas para cada estilo de escrita. Nos experimentos realizados para os três estilos de escrita, os autores reportam a taxa de reconhecimento de 93%.

Recentemente, em (FERRER et al., 2013), foi realizado o reconhecimento de estilos de escrita a nível de linha por meio de análise de textura. Como características foi utilizado o descritor de textura LBP, além de uma versão modificada do LBP, denominada *Oriented Local Binary Patterns* (OLBP). *Least Squares Support Vector Machine* (LS-SVM) foi utilizada como classificador. Para gerar a base para treinamento os autores utilizaram a ferramenta Google Tradutor. Desta forma, foi gerado 20 imagens para cada um dos 10 diferentes estilos de escrita, enquanto que para o conjunto de testes jornais e livros foram digitalizados. Nos experimentos realizados, as taxas obtidas foram de 90%.

Uma abordagem baseada em Lógica *fuzzy* para o reconhecimento de estilos de escrita a nível de palavra em imagens de baixa resolução em placas é apresentado por Angadi em (ANGADI; KODABAGI, 2013). Considerando uma base com 1200 palavras, a fim de distinguir cinco estilos de escrita: Hindi, Canará, Romano, Malaiala e Tamil, a taxa de reconhecimento descrita nos experimentos realizados foi de 94,33%.

2.3 Considerações Finais

A Tabela 2 apresenta um resumo dos trabalhos aqui apresentados. A partir dela é possível notar um aumento, ao decorrer dos anos, no uso de técnicas relacionadas às usadas neste trabalho.

Entretanto, como descrito anteriormente, realizar uma comparação entre elas torna-se uma tarefa difícil devido a falta de padronização nas bases de dados utilizadas, as quais possuem quantidade e estilos de escrita distintos, e também ao emprego do reconhecimento em diferentes níveis (palavra, linha e bloco).

Tabela 2: Resumo da Revisão Bibliográfica

Autores	Ano	Estilos de Escrita	Quant. de Dados	Nível	Características	Classificação	Desempenho
Hochberg et al.	1999	6	496 Documentos	Bloco	Centroide relativo em Y, centroide relativo em X, número de lacunas brancas, efericidade e proporção de aspectos	Função Discriminante Linear de Fisher	88%
Singhal et al.	2003	4	480 Documentos	-	Filtro de Gabor Multicanal	Abordagem Probabilística	91,6%
Dhandra e Hangarge	2007	3	3000 Palavras e 400 Números	Palavra	Filtros morfológicos e descritores de região	k-NN	96,05% para palavras e 99% para números
Hangarge e Dhandra	2010	3	150 Documentos	Bloco	Densidade dos traços de tinta e dos pixels	k-NN	99,2%
Hiremath et al.	2010	8	4000 Documentos	Bloco	Co-ocorrência de histogramas compostas pela wavelet	k-NN	97,5%
Obaidullah et al.	2013	6	152 Documentos	-	Abstratas/matemáticas, baseadas na estrutura e dependentes do estilo de escrita	Rede Neural Perceptron Multi-camada	92,8%
Ferrer et al.	2014	3	1909 Linhas e 15481 Palavras	Linha e Palavra	Concatenação de histogramas LBP	Multi-classificadores	89,89%
Pardeshi et al.	2014	11	28100 Palavras	Palavra	Transformada de Radon, transformada wavelet discreta, transformada discreta do cosseno, filtros estatísticos	k-NN e SVM	98% com dois e 96% com três estilos de escrita
Pal et al.	2003	11	4000 Linhas	Linha	Diversas Características	Árvore Binária	99,2%
Pan e Tang	2011	6	4500 Blocos	Bloco	BEMD + LBP	SVM	95,41%
Das et al.	2012	3	1409 Palavras	Palavra	Diversas Características	Heurísticas Pré-defindas	93%
Ferrer et al.	2013	10	200 Documentos	Linha	LBP e OLBP	LS-SVM	90%
Angadi e Kodabagi	2013	5	1200 Palavras	Palavra	-	Lógica Fuzzy	94,33%

3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo as técnicas computacionais utilizadas serão apresentadas de forma a servir de base para o entendimento deste trabalho. Para um maior aprofundamento em qualquer dos pontos abortados, as referências aqui citadas devem ser consultadas. A Seção 3.1 apresenta os descritores de textura empregados neste trabalho, enquanto que na Seção 3.2 são descritos alguns conceitos de Aprendizagem de Máquina referentes a classificadores tradicionais, além da técnica de aprendizado conhecida como Máquina de Vetores de Suporte. Por fim a Seção 3.3 apresenta conceitos da Aprendizagem Profunda e a técnica pertencente a essa abordagem utilizada neste trabalho.

3.1 Descritores de Textura

Segundo Tamura (TAMURA et al., 1978), uma textura é uma constituinte de uma região macroscópica em que sua estrutura é formada pela repetição de padrões com primitivas dispostas conforme uma regra de composição. *Pixels* contíguos formam tais primitivas, dentre as quais podem ocorrer iterações aleatórias ou dependentes. A ocorrência de iterações aleatórias caracterizam texturas finas, enquanto que interações melhor definidas caracterizam texturas ásperas.

A textura encontra-se entre as características utilizadas pelo sistema visual humano na interpretação de informações visuais (PEDRINI; SCHWARTZ, 2008). Mesmo que o sistema visual humano tenha facilidade no reconhecimento de texturas, formalizar sua definição ou desenvolver descritores que possam ser utilizados em diferentes domínios de aplicações é consideravelmente difícil (PEDRINI; SCHWARTZ, 2008).

Segundo Gonzales (GONZALEZ; WOODS, 2008), as principais abordagens para extrair descritores de textura são: Estatística, estrutural e espectral. Emprega-se aqui técnicas embasadas em abordagem estatística e estrutural. Modelos embasados na abordagem estatística têm como objetivo extrair estatísticas de imagens, tais como contar a ocorrência de níveis de cinza ou verificar o modo como *pixels* com diferentes intensidades se relacionam. Modelos embasados em abordagem estrutural por sua vez, descrevem a textura a partir da relação espacial existente entre regiões ou primitivas presentes na imagem (GONZALEZ; WOODS, 2008).

3.1.1 Matriz de Co-ocorrência de Níveis de Cinza (GLCM)

GLCM é um método estatístico proposto por Haralick (HARALICK et al., 1973) para descrever textura. O descriptor de textura GLCM consegue descrever atributos como: suavidade, rugosidade, granularidade, entre outros atributos presentes na imagem. Para descrever tais atributos, é extraída uma matriz quadrada $n \times n$, onde n representa o número de níveis de cinza presente na imagem. Essa matriz representa a probabilidade de que dois valores de intensidade de cinza estejam envolvidos por uma determinada relação espacial.

A distância entre os *pixels* e o ângulo a partir do *pixel* central aos seus vizinhos são dois parâmetros diretamente relacionados ao método. Desta forma, uma matriz extraída tem grande dependência da distância e do ângulo pré-definidos. Em geral, utilizam-se distâncias $d = \{1, 2, 3, 4, 5\}$ e ângulos $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ (GONÇALVES, 2014). A Figura 2 apresenta um exemplo da matriz extraída de uma imagem com quatro níveis de cinza, considerando $d = 1$ e $\theta = 0^\circ$ na qual cada posição $M[i][j]$ contém a soma do número de vezes em que o nível de cinza i aparece logo a esquerda do nível j , com o número de vezes em que o nível j aparece logo a esquerda do nível i . Em seguida, a matriz P é criada a partir da divisão dos elementos da matriz M sobre a somatória destes.

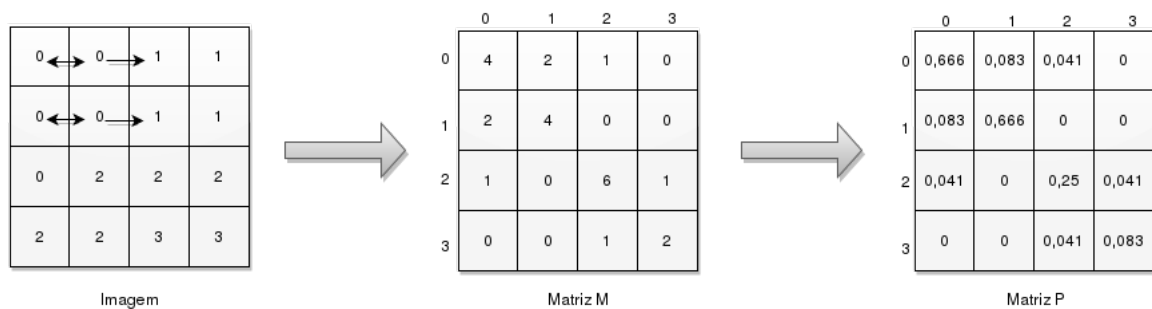


Figura 2: Exemplo de Matriz de Co-ocorrência para uma imagem com 4 níveis de cinza com $d = 1$ e $\theta = 0^\circ$

Quatorze medidas, denominadas características propostas por Haralick, são então extraídas da matriz P , sendo as seis principais, segundo Baraldi (BARALDI; PARMIGGIANI, 1995), descritas a seguir.

- **Energia:** Também chamado de segundo momento angular, avalia a uniformidade da textura em uma imagem. Mostrado na Equação 3.1, em texturas ásperas apresenta valores próximos de um, o máximo possível para tal medida.

$$f_{sma} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} P_{i,j}^2 \quad (3.1)$$

- **Entropia:** Expressa a desordem contida na textura, apresentando valores altos para imagens não uniformes, entretanto não normalizados no intervalo [0, 1]. A Equação 3.2 descreve tal medida.

$$f_{ent} = - \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} P_{i,j} \log(P_{i,j}) \quad (3.2)$$

- **Contraste:** Como pode ser visto na Equação 3.3, mede a presença de grandes transições de níveis de cinza na imagem. Um baixo contraste é dado quando há uma pequena diferença entre níveis de cinza contíguos.

$$f_{con} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} (i-j)^2 P_{i,j} \quad (3.3)$$

- **Heterogeneidade:** Apresenta valores altos quando os tons de cinza desviam do nível de cinza médio. Esta medida independe da localização dos elementos e da frequência espacial da textura. As Equações 3.4 e 3.5 apresentam a variância, sendo μ_i e μ_j o valor médio para i e j respectivamente.

$$f_{var_i} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} (i - \mu_i)^2 P_{i,j} \quad (3.4)$$

$$f_{var_j} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} (j - \mu_j)^2 P_{i,j} \quad (3.5)$$

- **Homogeneidade:** Mede a regularidade presente na imagem. Valores altos indicam pequenas variações de níveis de cinza entre pares de *pixels*. Esta medida esta representada na Equação 3.6.

$$f_{hom} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} \frac{1}{1 + (i-j)^2} P_{i,j} \quad (3.6)$$

- **Correlação:** Apresentada pela Equação 3.7, sendo σ_x e σ_y o desvio padrão calculado com base nas Equações 3.4 e 3.5, mede a dependência no nível de cinza de um pixel em relação aos seus vizinhos. Valores altos indicam a existência de relação entre os pares de níveis de cinza.

$$f_{corr} = \frac{1}{\sigma_x \sigma_y} \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} (i - \mu_i)(j - \mu_j) P_{i,j} \quad (3.7)$$

Mesmo sendo um método bastante antigo, o GLCM ainda é muito utilizado hoje em dia para descrever texturas nos mais diversos problemas, (GONÇALVES, 2014; CHAMPION et al., 2014; KHALILI; DANESH, 2015).

3.1.2 Padrões Binários Locais (LBP)

LBP é um método estrutural invariante à rotação para descrição de textura (OJALA et al., 2002). Esse método baseia-se na premissa de que padrões binários locais e a região da vizinhança de um pixel são características fundamentais na textura da imagem.

O LBP avalia para cada pixel da imagem seus vizinhos e um valor binário é atribuído a cada vizinho v por meio da Equação 3.8.

$$v = \begin{cases} 1, & \text{Se } i^v > i, \\ 0, & \text{Caso Contrário.} \end{cases} \quad (3.8)$$

Onde i^v e i são a intensidade de cinza do vizinho e do pixel central, respectivamente. Após essa etapa o valor do *pixel* central é substituído pelo resultado em decimal equivalente a concatenação dos números binários obtidos pelos vizinhos. A Figura 3 ilustra essa operação, por meio de um exemplo.

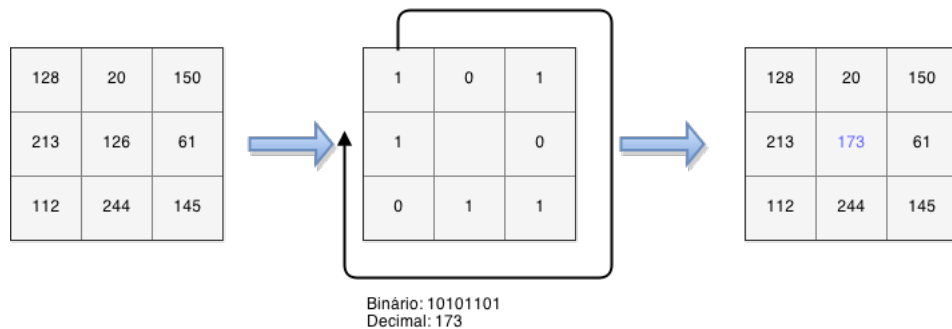


Figura 3: Exemplo de operação básica do LBP com valor resultante 173. Adaptada de (AMARAL; THOMAZ, 2011).

O *pixel* central é associado a um conjunto de amostras de tamanho P uniformemente espaçadas e distribuídas sobre determinada circunferência de raio R , tendo como centro o *pixel* central. Desta forma, dois parâmetros são pertinentes ao método e podem assumir diversos valores; O número de vizinhos (P) e o tamanho do raio (R). A Figura 4 demonstra alguns valores para P e para R .

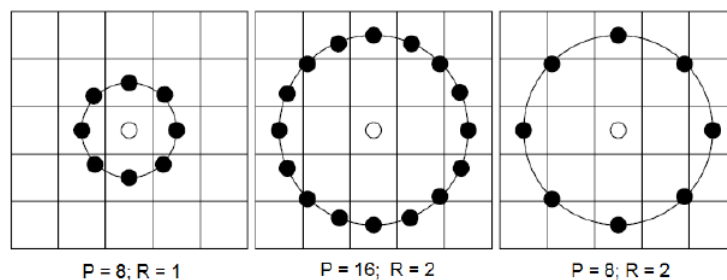


Figura 4: Diferentes Valores para P e R . Retirada de (GONÇALVES, 2014).

A partir da imagem obtida na etapa anterior, gera-se um histograma, o qual mede a frequência dos valores obtidos. Para $P = 8$ o histograma resultante tem 256 valores. Entretanto, somente 58 dos valores atendem a definição de uniformidade definida por Ojala (OJALA et al., 2002), em que a ocorrência de no transições do *bit* um para zero e vice-versa deve ocorrer no máximo duas vezes, esta definição foi estabelecida por ser demonstrado empiricamente que tais ocorrências são as mais importantes para a caracterização da textura. Os 198 valores restantes são contabilizados juntamente, resultando assim 59 características.

3.1.3 Quantização Local de Fase (LPQ)

LPQ é um método local para descrição de textura, descrito por Ojansivu (OJANSIVU; HEIKKILÄ, 2008) como relacionado e complementar ao LBP. Sua principal característica é a robustez para imagens borradas ou afetadas por uma iluminação não uniforme. De forma análoga ao LBP, para cada pixel p um código é calculado a fim de representar a textura em uma vizinhança centrada em p , de tamanho $m \times m$.

O método tem sua base nas propriedades de espectro de fases da *Short-Term Fourier Transform* (STFT), levando em conta apenas quatro coeficientes complexos dos componentes real e imaginários. São gerados, para cada *pixel* da imagem original, oito *bits*, os quais são concatenados a fim de formar um valor inteiro de oito *bits* o qual representa a textura na vizinhança.

Partindo da premissa de que a fase possui a maior parte das informações da STFT, um processo de redução de dimensionalidade com relação aos oito valores obtidos é realizado. Por fim, duas outras etapas denominadas decorrelação e quantização são aplicadas. O vetor de características gerado através do LPQ é um histograma com 256 posições. Em (OJANSIVU; HEIKKILÄ, 2008), o LPQ pode ser visto em detalhes.

3.1.4 SURF

Speeded Up Robust Features (SURF) é um descritor de textura e detector de pontos de interesse inspirado no *Scale-Invariant Feature Transform* (SIFT). Sendo assim, o SURF possui os mesmos princípios e etapas do SIFT, diferindo nas técnicas utilizadas em cada etapa. Segundo Bay (BAY et al., 2008), o SURF é mais rápido e mais robusto que o SIFT. No contexto deste trabalho, o descritor do SURF pode ser dividido em duas etapas principais:

A primeira etapa, denominada Detecção de Pontos de Interesse, determina os pontos de interesse por meio da determinante da Matriz de Hesse. Seja uma posição $f(x, y)$ na imagem e escala σ , o determinante $D(H)$ é definido na equação 3.9, na qual $L(f, \sigma)$ refere-se a derivada de segunda ordem da imagem em escala de cinzas. Os pontos de interesse são definidos onde o determinante

é máximo local.

$$D(H) = \begin{vmatrix} L_{x,x}(f, \sigma) & L_{x,y}(f, \sigma) \\ L_{y,x}(f, \sigma) & L_{y,y}(f, \sigma) \end{vmatrix} \quad (3.9)$$

A etapa de Descrição da Vizinhança Local, tem como objetivo prover uma descrição das características da imagem. Para cada ponto de interesse definido por meio da etapa anterior, o SURF descreve como a intensidade dos *pixels* de sua vizinhança são distribuídos. Para isso, a primeira etapa consiste em fixar uma direção a partir de uma região circular em torno do ponto de interesse, a fim de obter um certo grau de invariância a rotação.

Em seguida uma região quadrada é extraída centrada no ponto de interesse e orientado de acordo com a direção fixada. Com a finalidade de preservar as informações espaciais, esta região é dividida em sub-regiões de tamanho $n \times n$, determinou-se que $n = 4$ obtém os melhores resultados. Para cada sub-região, as respostas da Transformada de Haar na vertical e na horizontal são extraídas. Então, as respostas da Transformada de Haar são resumidas sobre cada sub-região e formam as primeiras características. Além disso, as somas dos valores absolutos das respostas da Transformada de Haar são computadas. Concatenando isto para todas as sub-regiões com $n = 4$, o *SURF size* ou quantidade de características de cada ponto de interesse é igual a 64.

Por fim, o vetor de características é gerado, sendo a primeira característica o número de pontos de interesse encontrados na imagem. Enquanto que as demais referem-se às medidas extraídas em relação as características de cada ponto de interesse. Sendo assim, para cada medida considerada concatena-se *SURF size* características no vetor de características. Considerando *SURF size* = 128 e quatro medidas, gera-se um vetor de características de tamanho 513.

3.2 Classificadores Tradicionais

Podemos definir os algoritmos de Aprendizagem de Máquina segundo três diferentes abordagens: Aprendizagem Supervisionada, Aprendizagem Não Supervisionada e Aprendizagem Por Reforço (DUDA et al., 2012). A primeira consiste em reconhecer, a partir de exemplos, outras instâncias. A segunda abordagem por sua vez, tem como objetivo agrupar as instâncias de alguma forma, sem conhecimento prévio. Enquanto que a terceira usa um esquema no qual recompensas ou punições são dadas ao sistema no lugar da resposta correta. O algoritmo de classificação utilizado neste trabalho pertence a primeira abordagem e será apresentado a seguir.

3.2.1 Máquina de Vetores de Suporte (SVM)

SVM é uma técnica para o treinamento de classificadores proposta por Vapnik (VAPNIK, 1995). Essa técnica tenta separar duas classes por meio de um hiperplano, o qual pode ser considerado ótimo quando separa os dados com máxima margem possível através dos vetores de suporte. Os vetores de suporte são as instâncias mais importantes presentes no conjunto de treinamento, após serem encontrado as demais instâncias podem ser descartadas. Na Figura 5 apresentamos a ideia geral da técnica SVM.

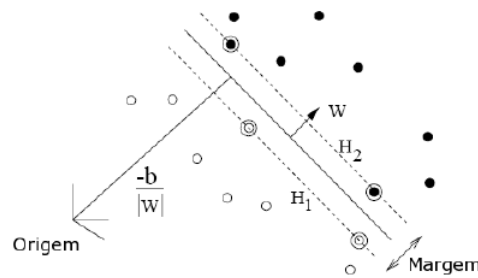


Figura 5: Ideia Geral da Técnica SVM com Vetores de Suporte Circulados. Retirada de (GONÇALVES, 2008).

Considerando que a grande maioria dos problemas reais não são linearmente separáveis, o SVM mapeia os dados de entrada para um espaço de dimensão maior, onde o conjunto, naquela dimensão, passa a ser linearmente separável. A função utilizada para realizar essa projeção é denominada *kernel*.

Outro problema inerente ao SVM é a classificação de múltiplas classes, uma vez que o SVM foi originalmente concebido para lidar com classificações binárias. Desta forma, para problemas multi-classes é necessário a transformação em diversos problemas de classes binárias. Para isso, duas abordagens são utilizadas: um contra todos, em que um classificador é construído para distinguir cada classe de todas as demais e um contra um, na qual um classificador é construído para cada par de classes.

3.3 Aprendizagem Profunda

A Aprendizagem Profunda é uma subárea de Aprendizagem de Máquina que possui um conjunto de técnicas que usam várias camadas para extração de características e classificação ou regressão. O ponto fundamental das técnicas de Aprendizagem Profunda é a descoberta automática de características, descartando a fase de extração de características a qual pode ser custosa em muitas aplicações por necessitar de um especialista no domínio (BENGIO; COURVILLE, 2013).

3.3.1 Rede Neural Convolucional (CNN)

Para Haykin (HAYKIN, 2001), uma Rede Neural é um processador denso e paralelamente distribuído, constituído de unidades de processamento simples, chamadas de neurônios. Ela se assemelha ao cérebro em dois aspectos: O conhecimento é adquirido pela rede através de um processo de aprendizagem. Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

A CNN é uma Rede Neural de Aprendizagem Profunda que vem obtendo bons resultados em diversos desafios de Aprendizagem de Máquina (CIREŞAN et al., 2012; RUSSAKOVSKY et al., 2014). Segundo Simard (SIMARD et al., 2003), sua estratégia geral é extrair características simples com uma maior resolução e convertê-las em características mais complexas com uma baixa resolução, ou seja encontrar várias características por toda imagem original e representá-las em filtros que possuem somente as informações importantes. A Figura 6 apresenta uma visão geral da CNN.

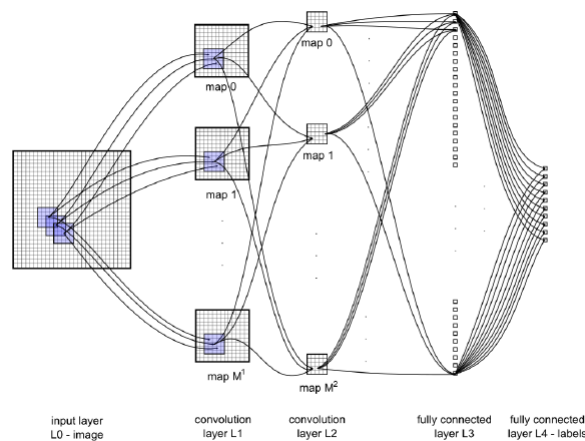


Figura 6: Visão Geral da Rede Neural Convolucional. Retirada de (CIREŞAN et al., 2011)

Segundo (LECUN et al., 1998), as Redes Neurais Convolucionais combinam três ideias arquiteturais: Campos locais receptivos, pesos compartilhados e subamostragem espacial ou temporal. Essas ideias garantem um certo grau de invariância em relação a deslocamentos, escalas e distorções. Na arquitetura de uma CNN, uma divisão é feita em camadas, na qual cada camada é responsável por determinada tarefa. Em cada uma das camadas, existem diversos parâmetros que dependem da funcionalidade da camada e que podem ser previamente configurados. Descrevemos a seguir algumas das principais camadas e suas funcionalidades.

- **Camada Convolucional:** Possui filtros treináveis que são aplicados por toda a entrada (LECUN et al., 1989). Para cada filtro, cada neurônio é somente conectado a um subconjunto de neurônios na camada anterior. No caso de imagens, os filtros definem uma pequena área e cada neurônio é conectado somente aos neurônios mais próximos da camada anterior. Os

pesos são compartilhados, ou seja, iguais entre os neurônios, levando os filtros a aprender padrões frequentes que ocorre em qualquer parte da imagem. A inspiração para esta camada originou-se de modelos do sistema visual dos mamíferos (HAFEMANN, 2014).

- **Camada Localmente Conectada:** Possui os mesmos filtros que a Camada Convolutiva, entretanto conecta neurônios dentro de uma pequena janela para a próxima camada sem pesos compartilhados.
- **Camada de Pooling:** Implementa uma função não linear para a diminuição de resolução, a fim de reduzir a dimensionalidade e capturar pequenas variações de translação, adicionando robustez ao modelo. Existem diversos tipos de Camada de *Pooling*, entre eles o *Max Pooling* é o tipo que apresenta os melhores resultados, segundo (SCHERER et al., 2010).
- **Camada Totalmente Conectada:** É a camada padrão das redes neurais e conecta todos os neurônios de uma camada para outra sem usar pesos compartilhados (HAFEMANN, 2014).

Em resumo, as Camadas Convolutivas e Localmente Conectadas podem ser vistas como um extrator treinável de características, enquanto a Camada Totalmente Conectada como um classificador treinável (SIMARD et al., 2003).

3.4 Considerações

Neste capítulo procuramos apresentar algumas técnicas que devem ser empregadas neste trabalho. No próximo capítulo apresentaremos em detalhes o método proposto neste trabalho o qual irá fazer uso das abordagens descritas.

4 MÉTODO

Neste capítulo apresenta-se o método utilizado no desenvolvimento deste trabalho. A Figura 7 representa a abordagem proposta juntamente com procedimentos a serem realizados. Em seguida, cada etapa é descrita detalhadamente.

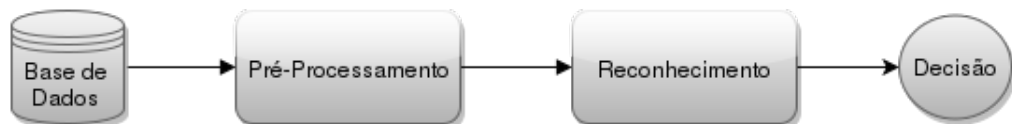


Figura 7: Visão Geral do Método Proposto

Como apresentado na Figura 7, a partir de uma base de dados serão aplicadas técnicas de Processamento de Imagens Digitais com intuito de reduzir os possíveis ruídos existentes, além de gerar uma textura mais densa e representativa a partir da escrita. Essa fase de pré-processamento tem como principal objetivo preparar os documentos para o processo de reconhecimento em si. Em seguida, na etapa de reconhecimento empregaremos a classificação por meio de duas abordagens nas imagens já pré-processadas com o intuito de reconhecer os estilos de escrita de um documento todo. Por fim, realizaremos uma comparação entre os resultados obtidos no processo anterior.

A seguir na Seção 4.1 descreveremos brevemente a base atual a qual estamos realizando nossos experimentos preliminares, enquanto que as Seções 4.2 e 4.3 detalham respectivamente as fases de pré-processamento e do processo de reconhecimento.

4.1 Descrição da Base de Dados

A base de dados utilizada contém 280 documentos manuscritos digitalizados a 150 dpi e divididos em cinco diferentes estilos de escrita: Bengali, Oriá, Canará, Persa e Romano. Os três primeiros são estilos de escrita utilizados em idiomas na Índia, o segundo usado no idioma homônimo falado no Irã, Afeganistão e Tajiquistão, enquanto que o Romano é utilizado em diversos idiomas, como por exemplo, no português e no inglês. Esses estilos de escrita podem ser vistos na Figura 8. A Tabela 3 apresenta a quantidade de documentos por classe, além do número de escritores e o local em que as instâncias de cada classe foram coletadas.

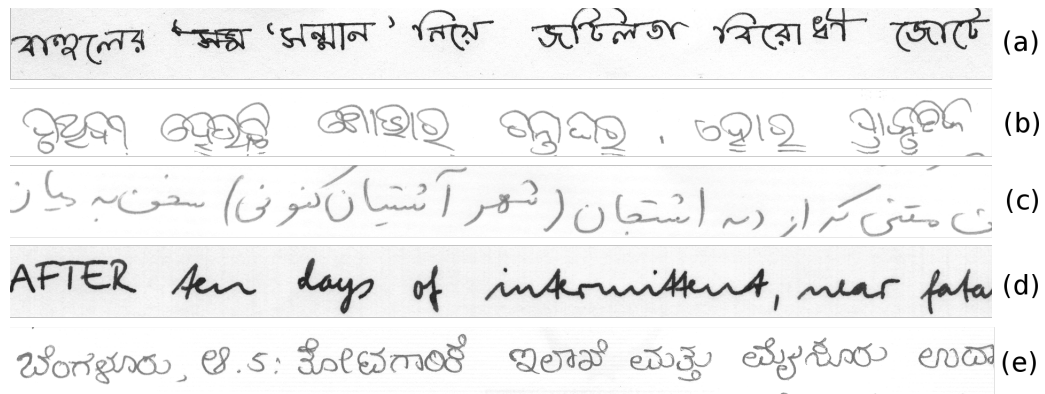


Figura 8: Linhas dos Estilos de Escrita Utilizados. (a) Bengali, (b) Oriá, (c) Persa, (d) Romano e (e) Canará

Tabela 3: Base de Dados

Classe	Quant. Documentos	Obtida No
Bengali	67	Instituto Estatístico da Índia (FERRER et al., 2014)
Oriá	50	Não Informado (FERRER et al., 2014)
Persa	16	Instituto Estatístico da Índia (FERRER et al., 2014)
Romano	90	Banco de Dados IAM (FERRER et al., 2014)
Canará	57	Contato com Autor (ALAEI et al., 2012)

4.2 Pré-Processamento

A Figura 9 apresenta as etapas presentes processo de pré-processamento. Posteriormente, cada uma dessas etapas será descrita.

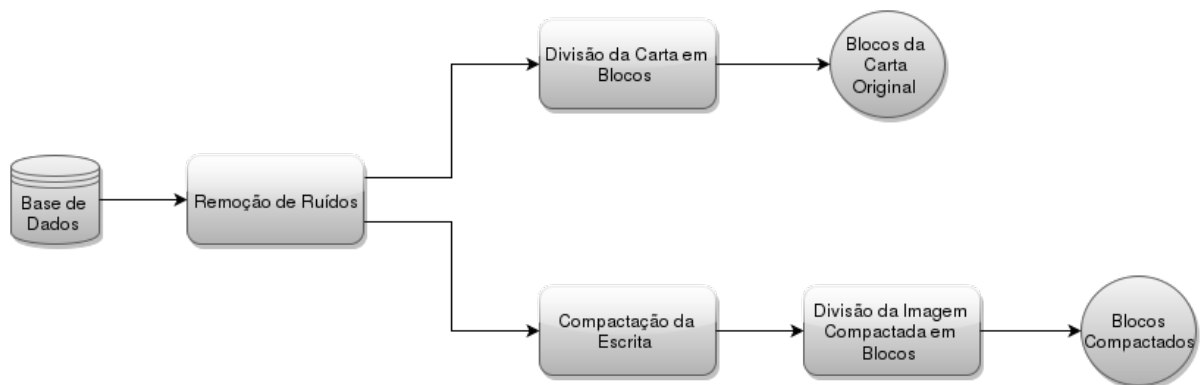


Figura 9: Etapas de Pré-Processamento

Para que seja possível gerar a compactação de escrita a partir de uma imagem de documento I utilizando a abordagem proposta por Hanusiak (HANUSIAK et al., 2012), deve ser realizado um processo de remoção do ruídos oriundos da fase de digitalização, de modo que tudo o que não for considerado caractere em I deve possuir tonalidade totalmente branca. Para resolver tal problema, uma limiarização automática da imagem I por meio do algoritmo de Otsu (OTSU, 1975) foi realizada, obtendo assim uma imagem binária I_b . Em seguida, todo *pixel* da imagem original $I[p]$

torna-se branco caso o *pixel* equivalente $I_b[p]$ for branco, e nada é feito caso o *pixel* $I_b[p]$ for preto. Esse último passo deve ser tomado para manter os tons de cinza da imagem original, sendo que os tons de cinza possuem informações relevantes para os descritores de texturas.

Em seguida, a compactação da escrita realizada por meio da abordagem proposta por Hanusiak (HANUSIAK et al., 2012) obtém uma nova imagem I' para cada imagem I original. Essa abordagem consiste na remoção de espaços entre as palavras e entre as linhas de um documento, de forma a compactar a imagem ao remover tais informações. A motivação da abordagem de compactação de escrita é gerar uma textura mais densa e mais representativa, de forma a descrever melhor um determinado estilo de escrita.

Após estes procedimentos, as instâncias de cada classe são divididas em conjunto de Treinamento e Teste. Essa divisão é aleatória, considerando certa porcentagem para cada conjunto.

Por fim, tanto as imagens pré-processadas, quanto as compactadas são separadas respectivamente em n_1 e n_2 blocos de tamanho $m \times m$. A Figura 10 apresenta um bloco gerado a partir da imagem original, além de um bloco gerado após a aplicação da etapa de compactação de escrita.

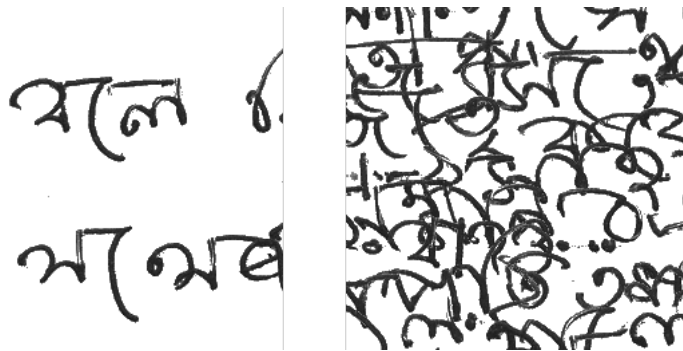


Figura 10: Blocos 256×256 da Carta Original e da Compactada

Na Figura 10 é possível notar o nítido aumento na quantidade de informação útil em um bloco extraído após a compactação.

4.3 Reconhecimento

Dois abordagens serão consideradas e suas taxas de acerto serão comparadas neste trabalho (Abordagem I \times Abordagem II). A primeira consiste na utilização de descritores de textura para a extrair características das imagens e do classificador SVM para o processo de classificação. Enquanto que a segunda abordagem consiste na utilização de aprendizagem profunda, por meio de uma CNN. As Figuras 11 e 12 ilustram as duas abordagens.

A abordagem apresentada pela Figura 11 é constituída de três etapas. Na primeira etapa, empregaremos descritores de textura em cada bloco, sendo o bloco compactado ou referente a carta



Figura 11: Abordagem I - Descriptor de Textura + Classificador SVM.

original de determinado documento. Quatro descritores de textura serão avaliados separadamente, o LBP, LPQ, SURF e GLCM. O vetor de características gerado a partir do processo de extração de características será utilizado para alimentar o classificador SVM. Por fim, uma combinação das saídas dadas pelo classificador será feita considerando o esquema da soma proposto por Kittler (KITTLER et al., 1998). Essa combinação tem o fim de reconhecer o estilo de escrita de um documento, já que o processo de classificação apresenta as predições em relação aos blocos. Por fim, uma decisão final é gerada, rotulando um documento questionado a uma determinada classe.

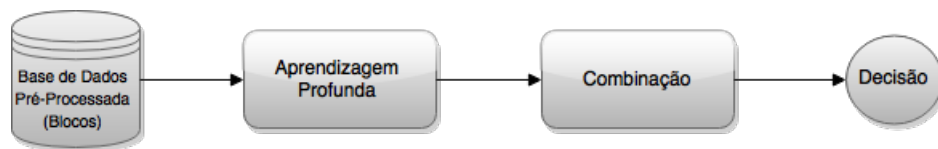


Figura 12: Abordagem II - Aprendizagem Profunda com CNN.

A abordagem apresentada pela Figura 12, por sua vez, possui apenas duas etapas. Na etapa de Aprendizagem Profunda, aplica-se o classificador CNN, sendo que a informação de entrada é o próprio bloco, sendo ele compactado ou referente a carta original e a saída é a classificação daquele bloco. Em outras palavras, essa única etapa desempenha tarefas equivalentes às duas primeiras etapas da abordagem apresentada pela Figura 11. Por fim, uma combinação equivalente a da abordagem anterior será realizada.

5 EXPERIMENTOS

Nos experimentos executados, considerou-se cinco diferentes estilos de escrita: Bengali, Oriá, Persa, Romano e Canará. As Seções 5.1 e 5.2 apresentam os experimentos realizados para a Abordagem I e Abordagem II, respectivamente. Em todos experimentos 60% das amostras foram utilizadas no conjunto de treinamento e 40% no conjunto de testes. A Tabela 4 apresenta a quantidade de documentos por estilo de escrita, além da quantidade de documentos utilizados para treino e teste. Em ambas abordagens dividimos o documento em blocos, aplicamos a classificação nos blocos e por fim, realizamos combinações das predições do classificador a fim de reconhecer o documento.

Tabela 4: Quantidade de Documentos Manuscritos Utilizados em Cada Estilo de Escrita

Estilos de Escrita	Quant. Total	Treino (60%)	Teste (40%)
Bengali	67	41	26
Oriá	50	30	20
Persa	16	10	6
Romano	90	54	36
Canará	57	35	22
Total	280	170	110

5.1 Abordagem I

Inicialmente, a técnica de compactação de escrita proposta por Hanusiak (HANUSIAK et al., 2012) foi aplicada, em seguida as imagens foram divididas em blocos considerando os tamanhos $m = \{128 \times 128, 256 \times 256\}$, com sobreposição de metade de m e sem sobreposição para as cartas compactadas. Para as cartas originais, utilizou-se $m = 256 \times 256$, sem sobreposição. As Figuras 13 e 14 apresentam os métodos de geração de blocos aplicados sem e com sobreposição. Enquanto que a Tabela 5 apresenta a quantidade de blocos gerados tanto para as cartas originais, quanto da imagem obtida após a compactação de escrita.

Considerando a Tabela 5 e as Figuras 13 e 14 nota-se que a quantidade de blocos gerados é consideravelmente superior ao aplicar a sobreposição de blocos. Isto implica em um maior número de informações de uma mesma imagem, entretanto em um maior custo computacional na geração do modelo no SVM. Para a extração de características os parâmetros dos descritores foram fixados em: LBP com $P = 8$ e $R = 2$; LPQ com $winSize = 7$; GLCM com $\theta = 0^\circ$ e $d = 1$, além das medidas:

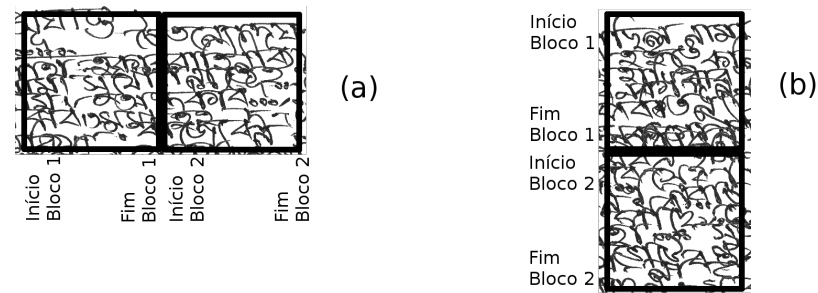


Figura 13: Método da Divisão de Blocos sem Sobreposição. (a) horizontal (b) vertical

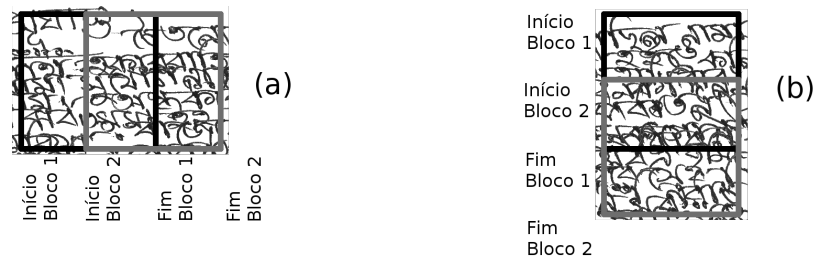


Figura 14: Método da Divisão de Blocos com Sobreposição. (a) horizontal (b) vertical

Tabela 5: Quantidade de Blocos Gerados em Cada Estilo de Escrita

Estilos de Escrita	Número de Blocos Compactados				Número de Blocos Cartas Originais
	Sem Sobreposição		Com Sobreposição		
	128×128	256×256	128×128	256×256	256×256
Bengali	5564	1110	19599	3815	3015
Oriá	4535	998	17057	3397	2250
Persa	787	167	2687	494	706
Romano	2792	730	9466	1730	4044
Canará	4482	990	17100	3264	2565
Total	18160	3995	65909	12700	12580

Energia, contraste, correlação e homogeneidade; SURF com $SURF\ size = 128$, além das medidas: Média, desvio padrão, obliquidade e curtose. A classificação foi realizada usando o classificador SVM, empregando a abordagem um contra todos com *kernel* RBF, adicionalmente utilizamos o método da soma para realizar a combinação das predições das instâncias. Para aplicação do SVM, utilizamos a biblioteca livre LIBSVM desenvolvida na linguagem C++ pela Universidade Nacional de Taiwan.

Cada experimento foi executado três vezes, para diferentes combinações de amostras no conjunto de treinamento e teste. Desta forma, os resultados apresentados referem-se a média e o desvio padrão das execuções. Nos casos que o desvio padrão não foi apresentado, realizou-se somente uma execução. A Tabela 6 apresenta uma relação entre o aumento de classes e a taxa de reconhecimento para blocos compactados, com $m = 256 \times 256$ sem sobreposição utilizando os

descritores LBP e LPQ. Nestes experimentos, as classes foram adicionadas na seguinte ordem: Bengali, Romano, Persa, Oriá e Canará.

Tabela 6: Taxa de Reconhecimento (%) com Aumento de Classes

Descritores	Duas Classes	Três Classes	Quatro Classes	Cinco Classes
LBP	98,92 ± 0,94	98,37 ± 1,41	97,73 ± 1,97	96,06 ± 3,68
LPQ	100,0 ± 0,00	96,34 ± 0,00	96,96 ± 1,31	97,27 ± 0,91

A partir das taxas reportadas na Tabela 6 é possível verificar que existe uma certa relação entre o número de classes presentes na base de dados e as taxas de reconhecimento. Em ambos descritores, nota-se uma queda considerável na taxa de reconhecimento após uma certa quantidade de estilos de escritas avaliados.

A Tabela 7 apresenta as taxas de reconhecimento do documento ao se aplicar a compactação de escrita, considerando os diferentes tamanhos de blocos e diferentes descritores.

Tabela 7: Taxa de Reconhecimento (%) Utilizando Compactação de Escrita

Descritores	Sem Sobreposição			Com Sobreposição	
	128 × 128	256 × 256	1 Bloco	128 × 128	256 × 256
LBP	94,84 ± 3,20	96,06 ± 3,68	94,24 ± 2,29	95,45 ± 4,17	95,45 ± 4,17
LPQ	96,66 ± 1,89	97,27 ± 0,91	97,27 ± 0,91	98,18	96,36 ± 0,91
SURF	91,13 ± 1,06	94,54 ± 4,16	94,84 ± 0,53	-	95,75 ± 3,20
GLCM	77,36 ± 1,91	76,36 ± 3,15	78,79 ± 1,05	72,47	74,85 ± 6,58

A partir dos resultados apresentados na Tabela 7, nota-se claramente a inferioridade na utilização do descritor GLCM. Isto já era esperado, pois resultados são apresentados em outros trabalhos que utilizam este descritor de textura para o processo de reconhecimento de escritor (GONÇALVES, 2014).

Além disso, o esquema de divisão em blocos mostra-se interessante somente ao utilizar o descritor LBP. Nos casos em que não há um ganho expressivo, a divisão em blocos se inviabiliza pela necessidade de computação adicional. De qualquer modo, podemos observar a superioridade do LPQ em relação aos demais descritores. Isto possivelmente ocorre por este possuir mais características, entretanto isso também implica em um maior custo computacional para treinar o modelo do SVM. O SURF, por sua vez, obteve resultados consideravelmente bons, entretanto superior somente ao GLCM.

O BoxPlot apresentado na Figura 15 ilustra os experimentos realizados para o descritor LBP, considerando os diferentes tamanhos de blocos, com e sem sobreposição. Nesta visualização, os valores do eixo x estão ordenados de forma crescente em relação a quantidade de blocos gerados, enquanto que as cores estão mapeadas em relação ao uso da sobreposição.

A partir da Figura 15 é possível notar uma discrepância entre as taxas de reconhecimento das

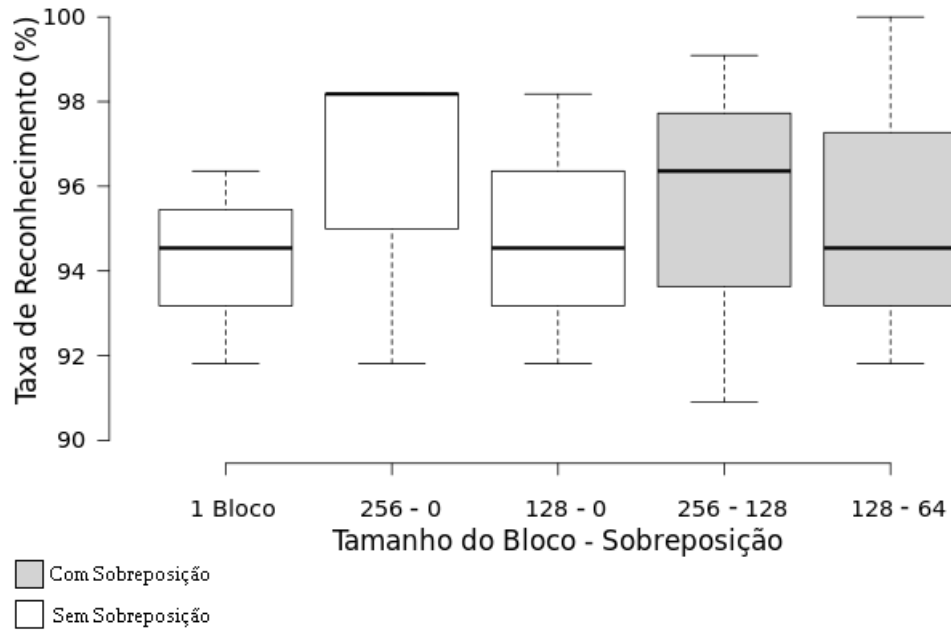


Figura 15: BoxPlot para LBP

execuções com sobreposição. Os experimentos com tamanho de bloco $m = 128 \times 128$ com sobreposição obteve taxas entre 92% e 100%. Embora as taxas médias mostrem que a sobreposição não tem grande influência no resultado, seus melhores casos mostram um desempenho consideravelmente melhor. Ou seja, a utilização da sobreposição de blocos implica em uma maior influência das instâncias escolhidas para o conjunto de treinamento. Desta forma, a utilização de um método sistemático que selecione as instâncias para treinamento possivelmente melhorará o caso médio, reduzindo esta diferença e o aproximando de 100%.

Os resultados referentes a utilização dos blocos das cartas originais são apresentados pela Tabela 8, a qual apresenta as taxas de reconhecimento para blocos de tamanho $m = 256 \times 256$ e para as cartas inteiras.

Tabela 8: Taxa de Reconhecimento (%) para Cartas Originais

Descritores	Taxa de Reconhecimento (%)	
	256 × 256	Carta Inteira
LBP	96,95 ± 1,91	92,72 ± 2,73
LPQ	96,95 ± 1,41	95,75 ± 1,39
SURF	94,53 ± 1,85	98,48 ± 0,53
GLCM	77,43 ± 0,65	77,88 ± 1,89

De acordo com os resultados apresentados na Tabela 8, vemos uma superioridade do descritor SURF para a carta inteira, na qual obteve a melhor taxa média deste trabalho (98,48%). Possivelmente isto é uma consequência da etapa de identificação de pontos de interesse, os quais são

visualmente mais claros nas cartas originais. Acreditamos que para problemas com muitas classes os resultados utilizando o SURF podem não ser muito satisfatórios. Além disso, o esquema de divisão em blocos mostra-se eficiente para o LPQ e principalmente para o LBP.

A utilização do SURF com a carta inteira mostrou-se a mais adequada para esta abordagem, sendo que além de obter a melhor taxa de acerto teve um baixo custo computacional. Além disso, não precisou-se aplicar os procedimentos de compactação de escrita e divisão de blocos, desta forma mesmo que o SURF possua mais características, a não aplicação desses procedimentos compensam o custo computacional na geração do modelo.

O Gráfico de Linhas da Figura 16 apresenta as taxas médias de acerto tanto para cartas originais, quanto para as compactadas, considerando a imagem inteira e a divisão em blocos de $n = 256 \times 256$ sem sobreposição.

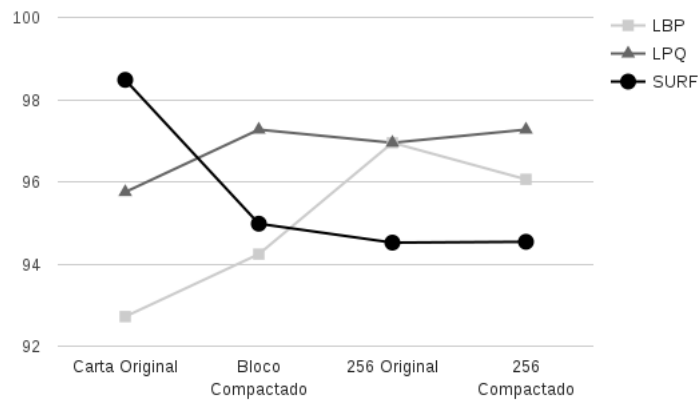


Figura 16: Cartas Originais \times Compactada

Considerando a Figura 16, observamos que o ganho em taxa de reconhecimento obtido por meio da aplicação da compactação de escrita está estritamente relacionado com o descritor de textura utilizado. Dentre os descritores avaliados, tanto o LPQ quanto o LBP recebem uma influência positiva do uso da compactação, enquanto que o SURF uma influência negativa. Além disso, é possível observar um ganho expressivo ao utilizar o esquema de divisão/fusão ao se aplicar o LBP.

5.2 Abordagem II

Nos experimentos referentes a esta abordagem, também aplicou-se a técnica de compactação de escrita proposta por Hanusiak (HANUSIAK et al., 2012). Foram extraídos n blocos aleatórios de tamanho m , sendo $m = \{64 \times 64, 128 \times 128, 256 \times 256\}$ e $n = \{500, 1000\}$ tanto para documentos originais, quanto para compactados. A Tabela 9 apresenta a quantidade de blocos gerados para cada estilo de escrita, considerando os diferentes valores para n . Para os experimentos com a

CNN utilizou-se o *framework* livre *Caffe* (JIA et al., 2014), desenvolvido pelo Centro de Visão e Aprendizagem da Universidade da Califórnia em Berkeley.

Tabela 9: Quantidade de Documentos Utilizados em Cada Estilo de Escrita

Estilos de Escrita	Número de Blocos (n)	
	500	1000
Bengali	33500	67000
Oriá	25000	50000
Persa	8000	16000
Romano	45000	90000
Canará	28500	57000
Total	140000	280000

A Figura 17 apresenta a arquitetura utilizada na CNN. Esta arquitetura foi baseada na rede utilizada no problema MNIST (JIA et al., 2014), tendo sido realizadas pequenas alterações para adaptá-la ao problema em questão.

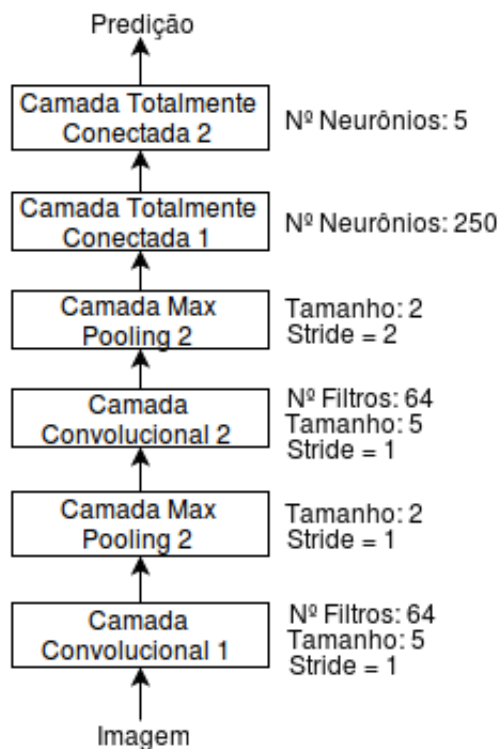


Figura 17: Visão Geral da Arquitetura Utilizada nos Experimentos

Assim como na Abordagem I, os experimentos foram executados três vezes. As Tabelas 10 e 11 demonstram a média da taxa de reconhecimento em percentual (%) e o desvio padrão para os blocos dos documentos originais e compactados, respectivamente.

De acordo com os resultados reportados nas Tabelas 10 e 11 notamos uma grande influência do esquema de compactação de escrita na taxa de reconhecimento para os blocos menores. Isso possivelmente ocorre por gerarmos a mesma quantidade de blocos compactados e originais, ou

Tabela 10: Taxa de Reconhecimento (%) Utilizando Compactação

Tamanho do Bloco	Quantidade de Blocos por Carta	
	500	1000
64 × 64	90,82 ± 2,75	90,86 ± 1,84
128 × 128	90,21 ± 5,83	90,21 ± 5,22
256 × 256	91,43 ± 4,14	-
1 Bloco	78,17 ± 1,82	

Tabela 11: Taxa de Reconhecimento (%) para Carta Originais

Tamanho do Bloco	Quantidade de Blocos por Carta	
	500	1000
64 × 64	80,42 ± 6,11	78,59 ± 2,95
128 × 128	88,68 ± 4,33	90,82 ± 3,31
256 × 256	88,99	-
Carta Inteira	82,11 ± 1,39	

seja, a quantidade de informação total provida à CNN é maior no caso do bloco compactado. Com o aumento do tamanho do bloco e consequentemente da informação total, as taxas começam a subir, até estagnar em cerca de 90%. Nesse caso, a partir de determinada quantidade de informação total, a CNN com a arquitetura utilizada para de convergir. Mesmo que esta abordagem tenha obtido taxas relativamente inferiores a Abordagem I, os resultados obtidos são consideravelmente bons ao serem comparados aos demonstrados na Tabela 2.

6 CONCLUSÕES

O principal objetivo deste trabalho foi analisar o reconhecimento *off-line* de estilos de escrita utilizando a textura da escrita para representar uma determinada classe em duas abordagens distintas. A primeira abordagem consistiu na aplicação de descritores de textura em conjunto com o classificador SVM, enquanto que a segunda na aplicação da técnica de Aprendizagem Profunda, denominada CNN. Em ambas abordagens utilizou-se um esquema de compactação de escrita, além da divisão em blocos e combinação pela regra da soma. A originalidade deste trabalho está na utilização do esquema de compactação da escrita, na aplicação do SURF e na aplicação da CNN.

A partir dos resultados obtidos, é possível verificar que a utilização de ambas abordagens podem contribuir para ótimas taxas de acerto no problema de reconhecimento de estilos de escrita. Comparando os resultados alcançados nestes experimentos com a Tabela 2, a qual apresenta o desempenho de trabalhos relacionados, podemos notar que o desempenho do método proposto é em alguns casos superior aos apresentados em literatura.

As Seções 6.1 e 6.2 apresentam respectivamente as contribuições deste trabalho e os trabalhos a serem realizados futuramente.

6.1 Contribuições

Dentre as contribuições do presente trabalho, podemos destacar:

- Análise consistente de métodos que não dependem das classes para o reconhecimento de estilos de escrita;
- Avaliação da influência de um processo de compactação de textura para o problema;
- Estudo do impacto do uso de diferentes descritores de textura;
- Verificação da relação entre o número de classes e a dificuldade do problema;
- Análise de uma abordagem de aprendizagem profunda para o problema;
- Estudo da influência de um esquema de divisão em blocos e fusão das predições dos classificadores.

6.2 Trabalhos Futuros

Durante a realização deste trabalho, observamos algumas questões relacionadas, que mesmo não fazendo parte do escopo deste trabalho, são relevantes e devem ser realizadas futuramente.

- **Aumentar base de dados:** Contamos atualmente com cinco estilos de escrita. Possuímos motivações em acrescentar novas classes a nossa base, de modo a realizar uma análise mais abrangente.
- **Seleção de instâncias para conjunto de treinamento:** Ao utilizarmos o esquema de sobreposição de blocos na Abordagem I, notamos uma maior dependência do conjunto de treinamento para o sucesso do reconhecimento. Isto sustenta a aplicação métodos para a seleção de instâncias a serem utilizadas no conjunto de treinamento.
- **Diferentes níveis de reconhecimento:** O reconhecimento de estilos de escrita neste trabalho é aplicado à nível de documento. Entretanto, algumas aplicações podem requerer o reconhecimento a nível de linha e/ou palavra. Desta forma, existe uma questão pertinente a ser respondida futuramente: Os modelos do SVM gerados a partir da Abordagem I utilizando blocos compactados ou originais conseguem boas taxas se aplicados aos demais níveis de reconhecimento?
- **Dissimilaridade:** Acreditamos que a verificação por meio da dissimilaridade pode prover bons resultados se aplicada a este problema.

Referências

- AHONEN, T.; HADID, A.; PIETIKÄINEN, M. Face recognition with local binary patterns. In: *Computer vision-eccv 2004*. [S.l.]: Springer, 2004. p. 469–481.
- ALAEI, A.; PAL, U.; NAGABHUSHAN, P. Dataset and ground truth for handwritten text in four different scripts. *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific, v. 26, n. 04, p. 1253001, 2012.
- AMARAL, V. do; THOMAZ, C. E. *Extração e Comparação de Características Locais e Globais para o Reconhecimento Automático de Imagens de Faces*. Tese (Doutorado) — Dissertação de Mestrado, Centro Universitário da FEI, SP, Brasil, 2011.
- ANGADI, S.; KODABAGI, M. A fuzzy approach for word level script identification of text in low resolution display board images using wavelet features. In: IEEE. *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*. [S.l.], 2013. p. 1804–1811.
- BARALDI, A.; PARMIGGIANI, F. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *Geoscience and Remote Sensing, IEEE Transactions on*, IEEE, v. 33, n. 2, p. 293–304, 1995.
- BAY, H. et al. Speeded-up robust features (surf). *Computer vision and image understanding*, Elsevier, v. 110, n. 3, p. 346–359, 2008.
- BENGIO, Y.; COURVILLE, A. Deep learning of representations. In: *Handbook on Neural Information Processing*. [S.l.]: Springer, 2013. p. 1–28.
- BERTOLINI, D. et al. Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers. *Pattern Recognition*, Elsevier, v. 43, n. 1, p. 387–396, 2010.
- CHAMPION, I. et al. Retrieval of forest stand age from sar image texture for varying distance and orientation values of the gray level co-occurrence matrix. *Geoscience and Remote Sensing Letters, IEEE*, IEEE, v. 11, n. 1, p. 5–9, 2014.
- CIREŞAN, D. et al. A committee of neural networks for traffic sign classification. In: IEEE. *Neural Networks (IJCNN), The 2011 International Joint Conference on*. [S.l.], 2011. p. 1918–1921.
- CIREŞAN, D.; MEIER, U.; SCHMIDHUBER, J. Multi-column deep neural networks for image classification. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.], 2012. p. 3642–3649.
- DAS, M. S.; RANI, D. S.; REDDY, C. Heuristic based script identification from multilingual text documents. In: IEEE. *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on*. [S.l.], 2012. p. 487–492.
- DHANDRA, B.; HANGARGE, M. Global and local features based handwritten text words and numerals script identification. In: IEEE. *Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on*. [S.l.], 2007. v. 2, p. 471–475.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012.

- FERRER, M. A.; MORALES, A.; PAL, U. Lbp based line-wise script identification. In: IEEE. *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. [S.l.], 2013. p. 369–373.
- FERRER, M. A. et al. Multiplettraining—one test methodology for handwritten word-script identification. 2014.
- GONÇALVES, D. B. *Agrupamento de Classificadores na Verificação de Assinaturas off-line*. [S.l.]: Setembro de, 2008.
- GONÇALVES, D. B. *Identificação e Verificação de Escritores Usando Características Texturais e Dissimilaridade*. Tese (Doutorado) — Universidade Federal do Paraná, 2014.
- GONZALEZ, R.; WOODS, R. Digital image processing: Pearson prentice hall. *Upper Saddle River, NJ*, 2008.
- HAFEMANN, L. G. An analysis of deep neural networks for texture classification. 2014.
- HANGARGE, M.; DHANDRA, B. Offline handwritten script identification in document images. *International Journal of Computer Applications*, International Journal of Computer Applications, 244 5 th Avenue,# 1526, New York, NY 10001, USA India, v. 4, n. 6, p. 6–10, 2010.
- HANUSIAK, R. et al. Writer verification using texture-based features. *International Journal on Document Analysis and Recognition (IJDA)*, Springer, v. 15, n. 3, p. 213–226, 2012.
- HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. H. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, IEEE, n. 6, p. 610–621, 1973.
- HAYKIN, S. S. *Redes neurais*. [S.l.]: Bookman, 2001.
- HIREMATH, P. et al. Script identification in a handwritten document image using texture features. In: IEEE. *Advance Computing Conference (IACC), 2010 IEEE 2nd International*. [S.l.], 2010. p. 110–114.
- HOCHBERG, J. et al. Script and language identification for handwritten document images. *International Journal on Document Analysis and Recognition*, Springer, v. 2, n. 2-3, p. 45–52, 1999.
- HOUAISS, A. Houaiss: Dicionário eletrônico da língua portuguesa. *São Paulo: Editora Objetiva Ltda*, v. 1, 2001.
- JIA, Y. et al. Caffe: Convolutional architecture for fast feature embedding. In: ACM. *Proceedings of the ACM International Conference on Multimedia*. [S.l.], 2014. p. 675–678.
- KARPATHY, A. et al. Large-scale video classification with convolutional neural networks. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. [S.l.], 2014. p. 1725–1732.
- KHALILI, K.; DANESH, M. Identification of vibration level in metal cutting using undecimated wavelet transform and gray-level co-occurrence matrix texture features. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, SAGE Publications, v. 229, n. 2, p. 205–213, 2015.
- KITTLER, J. et al. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 20, n. 3, p. 226–239, 1998.

- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.
- LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural computation*, MIT Press, v. 1, n. 4, p. 541–551, 1989.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998.
- LOWE, D. G. Object recognition from local scale-invariant features. In: IEEE. *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. [S.l.], 1999. v. 2, p. 1150–1157.
- LUGER, G. F. *Inteligência Artificial-: Estruturas e estratégias para a solução de problemas complexos*. [S.l.]: Bookman, 2004.
- OBAIDULLAH, S. M.; DAS, S. K.; ROY, K. A system for handwritten script identification from indian document. *Journal of Pattern Recognition Research*, v. 8, n. 1, p. 1–12, 2013.
- OBAIDULLAH, S. M. et al. Development of document image database for handwritten indic script-a state-of-the-art. 2015.
- OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 24, n. 7, p. 971–987, 2002.
- OJANSIVU, V.; HEIKKILÄ, J. Blur insensitive texture classification using local phase quantization. In: *Image and signal processing*. [S.l.]: Springer, 2008. p. 236–243.
- OTSU, N. A threshold selection method from gray-level histograms. *Automatica*, v. 11, n. 285-296, p. 23–27, 1975.
- PAL, U.; SINHA, S.; CHAUDHURI, B. Multi-script line identification from indian documents. In: IEEE COMPUTER SOCIETY. *2013 12th International Conference on Document Analysis and Recognition*. [S.l.], 2003. v. 2, p. 880–880.
- PAN, J.; TANG, Y. A rotation-robust script identification based on bemd and lbp. In: IEEE. *Wavelet Analysis and Pattern Recognition (ICWAPR), 2011 International Conference on*. [S.l.], 2011. p. 165–170.
- PARDESHI, R. et al. Automatic handwritten indian scripts identification. In: *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*. [S.l.: s.n.], 2014. p. 375–380. ISSN 2167-6445.
- PEDRINI, H.; SCHWARTZ, W. R. *Análise de imagens digitais: princípios, algoritmos e aplicações*. [S.l.]: Thomson Learning, 2008.
- RICH, E.; KNIGHT, K. *Inteligência artificial*. [S.l.: s.n.], 1994.
- RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *arXiv preprint arXiv:1409.0575*, 2014.
- SALVI, D. Document image analysis techniques for handwritten text segmentation, document image rectification and digital collation. 2014.

- SCHERER, D.; MÜLLER, A.; BEHNKE, S. Evaluation of pooling operations in convolutional architectures for object recognition. In: *Artificial Neural Networks–ICANN 2010*. [S.l.]: Springer, 2010. p. 92–101.
- SIMARD, P. Y.; STEINKRAUS, D.; PLATT, J. C. Best practices for convolutional neural networks applied to visual document analysis. In: IEEE COMPUTER SOCIETY. *2013 12th International Conference on Document Analysis and Recognition*. [S.l.], 2003. v. 2, p. 958–958.
- SINGHAL, V.; NAVIN, N.; GHOSH, D. Script-based classification of hand-written text documents in a multilingual environment. In: IEEE. *Research Issues in Data Engineering: Multi-lingual Information Management, 2003. RIDE-MLIM 2003. Proceedings. 13th International Workshop on*. [S.l.], 2003. p. 47–54.
- TAMURA, H.; MORI, S.; YAMAWAKI, T. Textural features corresponding to visual perception. *Systems, Man and Cybernetics, IEEE Transactions on*, IEEE, v. 8, n. 6, p. 460–473, 1978.
- THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern recognition*academic press. *New York*, 1999.
- VAPNIK, V. N. *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995. ISBN 0-387-94559-8.
- YILMAZ, A.; JAVED, O.; SHAH, M. Object tracking: A survey. *Acm computing surveys (CSUR)*, *Acm*, v. 38, n. 4, p. 13, 2006.
- YOUNG, I. T.; GERBRANDS, J. J.; VLIET, L. J. V. *Fundamentals of image processing*. [S.l.]: Delft University of Technology Delft, The Netherlands, 1998.