

DIEGO STIEHL

**RECONHECIMENTO DE SÍMBOLOS DE CONFIGURAÇÃO
DE MÃO DO SIGNWRITING**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre. Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Luiz Eduardo Soares de Oliveira

CURITIBA

2015

DIEGO STIEHL

**RECONHECIMENTO DE SÍMBOLOS DE CONFIGURAÇÃO
DE MÃO DO SIGNWRITING**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre. Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Luiz Eduardo Soares de Oliveira

CURITIBA

2015

S855r Stiehl, Diego
Reconhecimento de símbolos de configuração de mão do
SignWriting / Diego Stiehl. – Curitiba, 2015.
74f. : il., tab.

Dissertação (mestrado) - Universidade Federal do Paraná,
Setor de Ciências Exatas, Programa de Pós-Graduação em
Informática.

Orientador: Luiz Eduardo Soares de Oliveira

Bibliografia: p. 70-74.

1. Sistemas de reconhecimento de padrões. 2. Sinais e
símbolos. 3. Língua de sinais. I. Oliveira, Luiz Eduardo Soares de.
II. Universidade Federal do Paraná. III. Título.

CDD: 006.4



Ministério da Educação
Universidade Federal do Paraná
Programa de Pós-Graduação em Informática

PARECER

Nós, abaixo assinados, membros da Banca Examinadora da defesa de Dissertação de Mestrado em Informática, do aluno Diego Stiehl, avaliamos o trabalho intitulado, “RECONHECIMENTO DE SÍMBOLOS DE CONFIGURAÇÃO DE MÃO DO SIGNWRITING”, cuja defesa foi realizada no dia 26 de maio de 2015, às 09:00 horas, no Departamento de Informática do Setor de Ciências Exatas da Universidade Federal do Paraná. Após a avaliação, decidimos pela:

aprovação do candidato. reprovação do candidato.

Curitiba, 26 de maio de 2015.

Prof. Dr. Luiz Eduardo Soares de Oliveira
PPGInf/UFPR – Orientador

Prof. Dr. Alceu de Souza Britto
JR/PUCPR - Membro Externo

Prof. Dr. Cayley Guimarães
UTFPR – Membro Externo

Prof. Dr. Daniel Weingaertner
PPGInf/UFPR – Membro Interno



SUMÁRIO

LISTA DE FIGURAS	iv
LISTA DE TABELAS	v
LISTA DE SIGLAS	vi
RESUMO	vii
ABSTRACT	viii
1 INTRODUÇÃO	1
1.1 Motivação	4
1.2 Desafios	5
1.3 Objetivos	5
1.4 Contribuições	5
2 FUNDAMENTAÇÃO TEÓRICA	7
2.1 Representação Gráfica de Línguas de Sinais	7
2.1.1 Stokoe Notation	8
2.1.2 HamNoSys	9
2.1.3 SignWriting	9
2.1.3.1 Utilização e Notação do SignWriting	11
2.2 Técnicas de Classificação	15
2.2.1 Máquinas de Vetor de Suporte	15
2.2.2 Aprendizagem Profunda	18
2.2.3 Combinação de Classificadores	20
2.2.4 Verificação	23
3 ESTADO DA ARTE	25

4	MÉTODO PROPOSTO	33
4.1	Conjunto de Símbolos e Base de Dados	33
4.2	Extração de Características	36
4.2.1	Histogramas de Projeção	36
4.2.2	Histograma de Orientação	37
4.2.3	Histograma de Concavidade	39
4.2.4	Zoneamento	40
4.3	Classificação	41
4.3.1	Taxas de Reconhecimento e Matrizes de Confusão	42
4.3.2	Avaliação de Classificadores e Características	43
4.3.3	Aprendizagem Profunda	45
4.3.4	Combinação de Classificadores	46
4.3.5	Verificação	46
5	AVALIAÇÃO DO MÉTODO PROPOSTO	55
5.1	Avaliação de Classificadores e Características	55
5.2	Aprendizagem Profunda	58
5.3	Combinação de Classificadores	59
5.3.1	Verificação	65
6	CONSIDERAÇÕES FINAIS	68
	REFERÊNCIAS BIBLIOGRÁFICAS	70

LISTA DE FIGURAS

2.1	Perspectiva do emissor	12
2.2	Configurações básicas de mão	12
2.3	Adição de linhas para representação dos dedos	13
2.4	Símbolos de contato	13
2.5	Planos: paralelo à parede e ao chão, respectivamente	14
2.6	Expressões faciais	14
2.7	Exemplos de hiperplanos para separação entre duas classes	17
2.8	Arquitetura de uma rede neural convolucional para o reconhecimento de sinais de trânsito	19
2.9	Processo de verificação de resultado de classificação	24
3.1	Arquivo no formato SWML para a palavra Brasil	26
3.2	Interface do SWEdit	28
3.3	Processo de sistema proposto por Bouzid and Jemni (2013)	31
4.1	Configurações de mão da base de dados	34
4.2	Utilização da aplicação para captura de base	36
4.3	Histogramas de projeção horizontal e vertical	37
4.4	Zoneamento da curva de contorno: (a) divisão 4x4 da imagem; (b) canto superior direito; (c) histograma de orientação para a zona em questão.	38
4.5	Medição de concavidades: (a) concavidades; (b) vetor de características; (c) direções auxiliares; (d) direções na 4-vizinhança.	40
4.6	Zoneamento uniforme utilizado	41
4.7	Processo proposto para a classificação automática de símbolos de confi- guração de mão do SignWriting	42
4.8	Exemplo de matriz de confusão	43
4.9	Classificação utilizando Support Vector Machine	44

4.10	Mapas de características aprendidos pela primeira camada convolucional	45
4.11	Símbolos de configuração de mão: (a) 84 original, (b) 86 original, (c) 84 manuscrito e (d) 86 manuscrito.	48
4.12	Símbolos de configuração de mão: (a) 26 original, (b) 34 original, (c) 26 manuscrito e (d) 34 manuscrito.	50
4.13	Símbolos de configuração de mão: (a) 38 original, (b) 39 original, (c) 38 manuscrito e (d) 39 manuscrito.	50
4.14	Símbolos de configuração de mão: (a) 52 original, (b) 53 original, (c) 52 manuscrito e (d) 53 manuscrito.	51
4.15	Símbolos de configuração de mão: (a) 79 original, (b) 80 original, (c) 79 manuscrito e (d) 80 manuscrito.	53
4.16	Símbolos de configuração de mão: (a) 64 original, (b) 71 original, (c) 72 original, (d) 64 manuscrito, (e) 71 manuscrito e (f) 72 manuscrito.	53
4.17	Símbolos de configuração de mão: (a) 90 original, (b) 91 original, (c) 90 manuscrito e (d) 91 manuscrito.	53
4.18	Símbolos de configuração de mão: (a) 6 original, (b) 7 original, (c) 6 ma- nuscrito e (d) 7 manuscrito.	54
5.1	Matriz de confusão da classificação com histograma de concavidade	57

LISTA DE TABELAS

2.1	Comparação entre os diferentes sistemas de escrita	7
2.2	Grupos de configurações de mão do SignWriting	15
2.3	Particularidades do grupo 2 (dedo indicador e médio) de configurações de mão do SW	16
3.1	Trabalhos relacionados no estado da arte	32
4.1	Total de amostras válidas por classe (símbolo)	35
4.2	Subconjuntos de características utilizados na classificação	44
5.1	Resumo dos resultados dos experimentos com características e classificadores	56
5.2	Tempo de execução dos experimentos	56
5.3	Sumarização das principais confusões com o histograma de concavidade . .	58
5.4	Matriz de confusão resumida da classificação com histograma de concavidade	58
5.5	Resultado do experimento utilizando aprendizagem profunda	59
5.6	Matriz de confusão resumida da classificação com CNN	60
5.7	Resultado analítico da combinação de classificadores	61
5.8	Matriz de confusão resumida da melhor combinação de classificadores . . .	62
5.9	Resultado analítico da combinação de classificadores (top 2)	63
5.10	Resultado analítico de combinação (sem vetores de características conca- tenados - top 1)	64
5.11	Resultado analítico de combinação (sem vetores de características conca- tenados - top 2)	65
5.12	Melhor resultado de combinação por regra	65
5.13	Resultado da verificação de classificadores	66
5.14	Comparativo entre confusões em classificadores antes (Conf.) de depois (Corr.) da verificação	67

LISTA DE SIGLAS

API	Application Programming Interface
ASL	American Sign Language
BAP	Body Animation Parameters
CNN	Convolutional Neural Network
DAC	Deaf Action Committee
FAP	Facial Animation Parameters
FENEIS	Federação Nacional de Educação e Integração de Surdos
FMS	Federação Mundial dos Surdos
GIF	Graphics Interchange Format
GPU	Graphics Processing Unit
IHC	Interação Humano-Computador
LGP	Língua Gestual Portuguesa
LIBRAS	Língua Brasileira de Sinais
MEC	Ministério da Educação e Cultura
OCR	Optical Character Recognition
PNG	Portable Network Graphics
RBF	Radial Basis Function
SASL	South African Sign Language
SML	Sign Modeling Language
SMS	Short Messaging Service
SVG	Scalable Vector Graphics
SVM	Support Vector Machine
SW	SignWriting
SWML	SignWriting Markup Language
TDD	Telecommunication Devices for the Deaf
TIC	Tecnologia da Informação e Comunicação
VRML	Virtual Reality Modeling Language
XML	eXtended Markup Language

RESUMO

Apesar de a surdez ser um fator limitante para a comunicação de um indivíduo, as comunidades surdas têm conseguido manter intra-relações através da utilização das línguas de sinais, no Brasil representadas pela LIBRAS. Apesar de permitirem a troca de informações de forma pessoal, as línguas de sinais apresentam restrições com relação à documentação por meio escrito ou impresso, devido a diferenças estruturais com relação a idiomas baseados em fonética. O SignWriting é uma notação que surgiu com o objetivo de preencher esta lacuna e tem tido aceitação em diversas comunidades surdas, permitindo a representação “textual” de informações produzidas de forma espacial. Atualmente, há elevada carência de estudos e aplicações de informática que visem otimizar a utilização dessa notação por parte de seus conhecedores. Este trabalho objetiva elaborar um método que realize a identificação de símbolos de configuração de mão descritos pelo SignWriting, através de técnicas de reconhecimento de padrões. O método proposto visa identificar símbolos dados como entrada dentre um conjunto de 103 símbolos de configuração de mão válidos, previamente selecionados de acordo com a relevância de suas características estruturais, utilizando diferentes técnicas e estratégias para a classificação, como Máquinas de Vetores de Suporte (SVM), aprendizagem profunda, combinação de classificadores e verificação de resultados de classificação. Experimentos realizados mostraram que a combinação de classificadores via regra do produto obteve os melhores resultados dentre todas as estratégias testadas, atingindo taxas de reconhecimento de 94,65% em top 1 e 98,74% em top 2, e 94,75% em top 1 após realizar a etapa de verificação.

Palavras-Chave: SignWriting, reconhecimento de padrões, combinação de classificadores, aprendizagem profunda, verificação de classificadores.

ABSTRACT

Although deafness be a limiting factor for communication of an individual, the deaf communities have managed to maintain intra-relationships through the use of sign languages, which, in Brazil, are represented by LIBRAS. However, although they allow the exchange of information personally, sign languages have restrictions regarding the documentation by written or printed ways due to structural differences from phonetic based languages. SignWriting is a notation that was meant to fill this gap, which has been accepted in many deaf communities, allowing the “textual” representation information produced in the spatial form. Currently, there is high lack of studies and computer applications aiming the optimization of the use of such notation. This work aims to develop a method that performs the identification of hand configuration symbols described by SignWriting through pattern recognition techniques. The proposed method aims to identify input symbols as being one among 103 valid hand configuration symbols, which were previously selected according to the relevance of their structural characteristics, by using various techniques and strategies for classification, as Support Vector Machines (SVM), deep learning, combination of classifiers and verification of classification results. Experiments have shown that the combination of classifiers using the product rule achieved the best results among all tested strategies, achieving recognition rates of 94.65% on top 1 and 98.74% on top 2, and 94.75% on top 1 after performing the verification step.

Keywords: SignWriting, pattern recognition, combination of classifiers, deep learning, verification of classifiers.

CAPÍTULO 1

INTRODUÇÃO

De acordo com dados da Federação Mundial dos Surdos (FMS), existem cerca de 360 milhões de pessoas surdas ou com surdez incapacitante¹ no mundo (World Health Organization, 2015). Esse número é obtido através da contagem de participantes das comunidades de surdos de 123 países. No Brasil o órgão responsável pela representação dessas comunidades é a Federação Nacional de Educação e Integração de Surdos (FENEIS), a qual conta com cerca de 130 filiadas, entre escolas e associações de surdos, e tem como interesse, assim como federações de outros países, servir às pessoas surdas, tendo caráter educacional, assistencial e sociocultural (Stumpf, 2005).

Comunidade surda é a denominação que se dá ao grupo de pessoas (surdas e ouvintes, porém conhecedoras e utilizadoras da língua de sinais) de determinada região. Cada comunidade tende a desenvolver, naturalmente, sua própria linguagem, com detalhes que são relacionados e influenciados pela cultura na qual está inserida (assim como ocorre com a língua oral, fato que tende a gerar semelhanças entre as duas), normalmente variando de um país para outro. Portanto, pode-se afirmar que a língua de sinais, definitivamente, não é uma língua universal, havendo diferentes “dialetos” (Silva, 2012).

Segundo Stumpf (2005) “o surdo politizado não se considera deficiente e sim membro de uma comunidade cultural e linguística”, tendo como de grande importância a necessidade da divulgação de sua língua. Tal necessidade foi citada no documento resultante da Conferência Mundial de Educação para Todos (realizada na Espanha em 1994), sendo elicitada a necessidade de que surdos sejam educados através da língua de sinais, e exposta a preocupação de que este fato tem sido pouco respeitado (Stumpf, 2005).

A língua de sinais é uma língua completa, não uma pantomia, tendo definições de gramática e caráter visual e espacial. Uma pessoa surda, ao terminar sua educação “tra-

¹Surdez incapacitante refere-se à perda de audição maior do que 40dB no ouvido com melhor audição em adultos e maior que 30dB em crianças (World Health Organization, 2015).

dicional”, acaba por não saber como utilizar a língua escrita aprendida em sua totalidade: como meio de comunicação, para reflexão e enriquecimento do pensamento, como fonte de prazer. Este é um problema que pode ser observado em todo o mundo.

Silva (2012) que tem seu trabalho baseado na Língua Gestual Portuguesa (LGP), afirma que a mesma é “consagrada na Constituição da República Portuguesa desde 1997”, sendo considerada instrumento de ensino, o que é importante para protegê-la e valorizá-la como “expressão cultural e instrumento de acesso à educação de igualdade de oportunidades”.

Na comunidade brasileira existe a LIBRAS (Língua Brasileira de Sinais), que foi oficializada como língua oficial dos surdos brasileiros pela Lei 4.857 de 2002. Fato que, aliado à aceitação desta pelo Ministério da Educação e Cultura (MEC), tende a tornar os aspectos relacionados à vida e a educação de pessoas surdas menos penosos. Assim como as demais línguas de sinais, a LIBRAS é considerada de modalidade visual-espacial, pois usa, como meio de comunicação, movimentos gestuais e expressões faciais. A diferença com relação à língua portuguesa é que, esta outra, é uma língua de modalidade oral-auditiva, pois utiliza a articulação de sons que são percebidos pelo ouvido. Mas a diferença não é apenas com relação ao meio utilizado, havendo, também, especificidades nas estruturas gramaticais, tanto de línguas orais, quanto de línguas de sinais. (Ramos, 2006).

Também é considerada importante a aquisição de uma compreensão bilíngue (língua de sinais e língua escrita da região em questão) por partes das pessoas surdas, porém esta deve ocorrer somente em momento seguinte à introdução da criança surda ao seu idioma natural (língua de sinais), o que tende a facilitar bastante seu acesso ao que é tido como uma segunda língua, em um processo semelhante ao ocorrido com uma criança ouvinte que começa a estudar um idioma diferente.

Tradicionalmente, a escrita tem sido ensinada às crianças surdas como se as mesmas pensassem auditivamente, baseando-se apenas na língua oral escrita da região em questão (no caso do Brasil, o português). Porém, há a necessidade de se levar em consideração todo o conjunto de especificidades das crianças surdas, conhecedoras (ou em fase de aprendizagem) da língua de sinais, no momento do letramento escrito das mesmas, pois, caso

o aluno surdo seja alfabetizado através da LIBRAS e consiga virar um leitor proficiente nesta, o mesmo conseguirá utilizar técnicas equivalentes durante o processo de alfabetização em português, tornando seu aprendizado de leitura e escrita menos dificultoso (Pereira and Fronza, 2006).

Considerando as características previamente mencionadas a respeito da amplitude das comunidades surdas e das próprias línguas de sinais de cada uma, é fato que há a necessidade de uma representação gráfica (escrita) para estas últimas, de tal forma que a mesma não seja uma mera tradução, na forma escrita da linguagem oral, da mensagem que o autor tem a intenção de proferir com a utilização de sinais.

Pereira and Fronza (2006) chegam a propor uma forma escrita visual (gráfica) diretamente baseada nos sinais como uma forma provisória para o suprimento das divergências entre uma língua escrita e sua língua de sinais equivalente, ou seja, a construção de uma ponte metalinguística entre as mesmas, com o objetivo de facilitar o letramento bilíngue das pessoas, bem como facilitar a associação de termos existentes nas duas línguas.

Sabe-se que é essencial que exista uma forma não fotográfica (através de imagens estáticas, vídeos, representações artísticas etc) das línguas de sinais. Porém também se sabe que é inviável a representação de tais sinais através de um alfabeto convencional, pois estes seguem um padrão de parâmetros específicos que se encaixam em um formato não oral, tornando a correspondência entre fonema e grafema impraticável. Descobriu-se que tais sistemas de escrita são viáveis e concluiu-se que então, diferentemente do anteriormente imaginado, as línguas de sinais não são língua ágrafas².

A escrita de uma língua de sinais deve, obrigatoriamente, basear-se fortemente na representação do sinal que é produzido por um emissor e possibilitar a interpretação e reprodução completa deste sinal por qualquer conhecedor da sintaxe escrita da língua em questão, pois os surdos não podem depender de aspectos que se baseiam em sons (que não são percebidos por eles) para representar o que eles têm a necessidade de emitir. Assim, a memória de pessoas utilizadoras da língua de sinais pode deixar de ser a única forma de arquivar e detectar a evolução dos sinais.

²Ágrafa: Que não admite escrita (Michaelis, 2012)

A tarefa de representar as línguas de sinais de forma escrita tem sido um desafio que só foi levado a sério após a realização das primeiras pesquisas linguísticas sobre a American Sign Language (ASL), que é a língua de sinais utilizada nos Estados Unidos da América, as quais conferiram a esta o mesmo status das línguas orais.

Tratando-se de notações (sistemas de escrita) para representação gráfica de línguas de sinais, os de maior influência e difusão, atualmente, são: Stokoe Notation, HamNoSys e SignWriting.

O sistema SignWriting (SW) é uma notação para a representação escrita de sinais criado por Valerie Sutton em 1974, que tem apresentado maior aceitação com relação aos demais. Ele foi concebido com o objetivo de preservar as características gestuais das línguas de sinais, permitindo que qualquer uma destas línguas seja representada de forma escrita (Pereira and Fronza, 2006; Sutton, 2002).

1.1 Motivação

Como atualmente o sistema SignWriting possui bastante aceitação pelas comunidades surdas, sendo utilizado em mais de 20 países para fins de pesquisa e alfabetização/letramento de pessoas surdas (Pereira and Fronza, 2006; Silva, 2012), este foi selecionado como alvo de estudo neste trabalho.

Ainda existem poucas publicações e as ferramentas computacionais criadas na área de SW são precárias. Tais ferramentas não satisfazem por completo as necessidades de usuários da notação, não oferecendo um ambiente gráfico adequado ou otimização no processo de criação e consumo de conhecimento formatado em SW (Guimaraes et al., 2014).

Percebe-se a necessidade da existência de um editor de textos para a inserção de informações formatadas em SW de forma automatizada, que tenha a capacidade de abstrair, identificar e oferecer um resultado, com base em um “dicionário” de dados em SW pré-especificado, e tendo como base informações injetadas manualmente (de forma manuscrita) por usuários.

A criação desta ferramenta exige uma série de etapas que não são tratadas neste

trabalho. O método aqui proposto considera apenas configurações de mão do SignWriting isoladas e selecionadas para representar diversas características presentes nas mesmas, sendo este apenas uma parte de um sistema maior, que teria como objetivo final reconhecer toda uma sentença em SignWriting.

1.2 Desafios

Este trabalho tem como desafios identificados:

- Coletar uma base de dados representativa para viabilizar a pesquisa;
- Encontrar conjunto de características que consiga discriminar os símbolos de configurações de mão do SignWriting.

1.3 Objetivos

Este trabalho tem como objetivo utilizar técnicas de reconhecimento de padrões para elaborar um método que realize a identificação de símbolos de configurações de mão do SignWriting (SW). Os objetivos específicos são descritos a seguir.

- Construir base de dados consistente para reconhecimento dos símbolos de configuração de mão do SignWriting;
- Identificar características que consigam representar de forma relevante e genérica os símbolos de configuração de mão do SignWriting;
- Comparar diferentes estratégias de classificação e características para o reconhecimento dos símbolos;
- Avaliar os resultados e comparar as estratégias aplicadas.

1.4 Contribuições

Espera-se como contribuição deste trabalho um método que realize a identificação de símbolos de configurações de mão do SW, bem como a definição de métodos e estratégias

para identificação de símbolos de configuração de mão do SW, possibilitando, assim, a futura criação de aplicativos utilitários, em diferentes plataformas, para pessoas utilizadoras desta notação.

Também considera-se a base de dados de símbolos de configurações de mão do SW como uma contribuição, pois a mesma pode ser futuramente reutilizada para estudos mais amplos e/ou com o objetivo de melhorar os resultados obtidos neste trabalho. A base de dados, em seu formato original e normalizado, bem como os símbolos-base, podem ser obtidos no endereço <http://web.inf.ufpr.br/vri/signwriting-database>.

Este trabalho também contribui para a valorização do sistema de escrita do SignWriting, conseqüentemente contribuindo com as línguas de sinais e suas respectivas comunidades.

CAPÍTULO 2

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os principais conceitos relacionados ao processo de classificação dos símbolos do SW, a ser desenvolvido na metodologia proposta. Na Seção 2.1, são apresentados conceitos e notações de representação gráfica de línguas de sinais, enquanto a Seção 2.2 apresenta as técnicas de classificação utilizadas neste trabalho.

2.1 Representação Gráfica de Línguas de Sinais

Segundo Pereira and Fronza (2006), atualmente, as notações (sistemas) para representação gráfica de línguas de sinais de maior influência são: Stokoe Notation, HamNoSys e SignWriting. A Tabela 2.1 apresenta, de forma comparativa, exemplos da aplicação da escrita de uma mesma palavra (casa, em ASL) e a descrição das características em cada um destes.

Tabela 2.1: Comparação entre os diferentes sistemas de escrita
Fonte: Pereira and Fronza (2006)

Sistema	Características	Sinal
Stokoe Notation	Estrutura linear. Utiliza elementos do alfabeto latino.	
HamNoSys	Baseada no Stokoe Notation, mas já com símbolos de configuração de mão mais visuais. Conserva estrutura linear. Existem grandes obstáculos à representação de expressões não manuais.	
SignWriting	Estrutura não linear. Não utiliza símbolos do alfabeto latino. Possui elementos próprios para representar todos os parâmetros das línguas de sinais.	

Cada notação tem seu conjunto de símbolos que serve para representar as palavras (sinais) de cada idioma. De acordo com o conhecimento atual sobre línguas de sinais, sabe-se que cada sinal é obtido, em geral, pela combinação de cinco parâmetros (Martin,

2000):

- Movimento: mostra o deslocamento realizado pela mão do emissor;
- Forma da mão: reproduz o formato criado pela mão do emissor;
- Localização: define a posição de uma mão com relação à outra e aos demais elementos;
- Orientação: apresenta o posicionamento da mão no espaço tridimensional;
- Sinais gramaticais não manuais: referencia as características não ligadas às mãos, como face, toques e movimentação da cabeça.

2.1.1 Stokoe Notation

A Stokoe Notation (Notação de Stokoe) é fruto do trabalho do linguista William Stokoe, que publicou trabalhos pioneiros tratando as línguas de sinais como sistemas linguísticos legítimos, sendo conhecida como o primeiro sistema de notação para linguagem de sinais, criado em 1960 com o objetivo de registrar as línguas de sinais para fins de pesquisa. Portanto, esse sistema não tem por objetivo ser utilizado de forma massiva por leigos, e sim, para fins especificamente científicos (Pereira and Fronza, 2006).

A intenção de Stokoe não era escrever em língua de sinais, mas provar que a ASL era, definitivamente, um idioma, pois os cientistas em geral acreditavam que os sinais eram apenas desenhos no ar. Stokoe comprovou isso utilizando métodos tradicionais de linguística para isolar os segmentos do ASL e identificar seus parâmetros (Martin, 2000).

Sua notação não se assemelha em ponto algum com um sistema visual direto, sendo baseado no alfabeto latino. O pesquisador utilizou derivações de caracteres conhecidos desse alfabeto para representar as configurações de mão, acrescidos de símbolos específicos que objetivam representar as articulações dos sinais com relação ao corpo e as formas de toque e movimento realizados. Sua escrita é basicamente linear, como acontece com os sistemas de escrita alfabéticos.

Stokoe determinou três categorias de queremas (configuração da mão, local de articulação e movimento), as quais, juntas, determinam o significado global de um sinal. Caso haja a alteração de alguma delas, o sinal obtido é diferente. Porém, apenas estes três elementos são insuficientes para a representação e distinção completa de sinais (Silva, 2012).

2.1.2 HamNoSys

HamNoSys é uma abreviação para Hamburg Notation System (Sistema de Notação Hamburg). É baseado na notação Stokoe anteriormente citada, porém foi desenvolvido para uma possível utilização em comunidades surdas (melhorando, por exemplo, a forma da representação das configurações de mão, que tendem a parecer com a forma visual), visando suprir a lacuna deixada por Stokoe, que almejava apenas a utilização acadêmica de seu sistema.

Esse sistema continua sendo utilizado, principalmente na Alemanha, porém ainda apresenta restrições de características não manuais, como, por exemplo, a representação das expressões faciais do emissor.

2.1.3 SignWriting

O sistema SignWriting (SW) é uma notação para a representação escrita de sinais criado por Valerie Sutton, uma bailarina dos Estados Unidos, que o desenvolveu inicialmente com o intuito de aplicá-lo na representação de movimentos em coreografias de dança, o qual era inicialmente chamado de DanceWriting. Este só foi adaptado ao que se conhece por SignWriting após a autora ser contatada (em 1974) por pesquisadores da universidade de Copenhague, que visavam criar um sistema para notação da língua de sinais dinamarquesa (Pereira and Fronza, 2006; Silva, 2012; Martin, 2000).

O SW não muda os idiomas que são escritos através dele, e sim, os preserva, permitindo que sinais possam ser representados no papel para que outras pessoas possam lê-los e interpretá-los. Os símbolos do SW mostram como o corpo se apresenta enquanto uma pessoa está emitindo informações sinalizadas, podendo captar as sutilezas de qualquer

língua de sinais existente no mundo, porque ele é focado em registrar movimentos corporais (Sutton, 2002).

Após esse contato, Sutton começou a desenvolver um sistema que tomou como base cinco pressupostos querológicos¹ que compõem um sinal, acrescidos de componentes não manuais, facilmente relacionáveis aos expostos por Martin (2000), que, segundo Silva (2012) são:

- Configuração da mão;
- Movimento;
- Pontos de articulação;
- Localização;
- Orientação da palma da mão;
- Componente não manual ou expressão facial.

SignWriting apresenta, basicamente, uma representação gráfica padronizada da parte superior do corpo humano (do tronco para cima), que é a parte de maior foco durante uma comunicação em língua de sinais, com a devida aplicação de símbolos que mostram de forma direta e icônica as configurações de mão, contatos e movimentos envolvidos no processo da representação do termo desejado, focando, inclusive, na necessidade de demonstrar os elementos especiais simultâneos e quaisquer outras expressões tidas como não manuais. O sistema apresenta refinamento a ponto de apresentar até a posição e movimentos finais realizados pelos dedos.

Atualmente a notação apresenta aceitação em diversas comunidades surdas, sendo utilizada em mais de 20 países para fins de pesquisa e alfabetização/letramento de pessoas surdas. Sua literatura (física e eletrônica) expande-se a cada dia. O sistema e sua notação vêm sendo desenvolvidos, incrementados, refinados e atualizados até os dias atuais, através da parceria de Sutton com uma organização sem fins lucrativos, que é

¹A fonologia das línguas de sinais é representada pela querologia, ciência que estuda o movimento da mão e dos dedos, sendo as unidades mínimas distintivas denominadas por queremas (Fernandes, 2003).

atualmente dirigida por ela, conhecida como Deaf Action Committee (DAC), aderindo, cada vez mais, às necessidades e requisitos dos surdos utilizadores. O DAC mantém um portal online específico para a notação, que pode ser acessado pelo endereço eletrónico <http://www.signwriting.org>.

O SW, por si só, é considerado um sistema de escrita, pois garante a fixação de um sinal, de forma simples e direta, gerando legibilidade para uma grande parte da comunidade surda. Assim, esse sistema pretende ser uma forma prática para a escrita de sinais, o que torna possível a comunicação escrita, de maneira rápida e inequívoca entre os conhecedores de diversas línguas de sinais.

2.1.3.1 Utilização e Notação do SignWriting

A apresentação dos símbolos de SW se dá através de duas possíveis perspectivas: do emissor e do receptor, porém a forma de representação mais comum é a visualizada pelo emissor, sendo esta, inclusive, o padrão utilizado para publicações sobre o tema, sendo a primeira eventualmente usada em momentos específicos, como, por exemplo, durante a transcrição de determinado gesto de um vídeo.

Os símbolos utilizados na transcrição de sinais em SignWriting são independentes de qualquer língua de sinais, tendo carácter internacional, podendo ser aplicados para representar sinais de qualquer língua de sinais existente, considerando que possa haver alguma necessidade de adaptação estrutural da notação. Fato este, que permite o proposto por Silva (2012), que aplica o SW à Língua Gestual Portuguesa, que é o sistema considerado em seu estudo.

Silva (2012), adaptando Sutton (2002), apresenta vários exemplos didáticos da utilização dos elementos básicos componentes de termos em SW. São eles:

- Perspectiva, com base no emissor (Figura 2.1);
- Configurações de mão básicas (Figura 2.2);
- Representação dos dedos (Figura 2.3);

- Símbolos de contato (Figura 2.4);
- Planos (Figura 2.5);
- Expressões faciais (Figura 2.6).



Figura 2.1: Perspectiva do emissor

Fonte: Silva (2012)

		Mão fechada
		Mão circular
		Mão aberta com dedos unidos

Figura 2.2: Configurações básicas de mão

Fonte: Silva (2012)

		Mão - Indicar
		Mão - Seis
		Mão aberta com dedos afastados

Figura 2.3: Adição de linhas para representação dos dedos

Fonte: Silva (2012)

*	Tocar
+	Agarrar
*	Tocar entre
#	Bater
⊕	Escovar
@	Esfregar

Figura 2.4: Símbolos de contato

Fonte: Silva (2012)

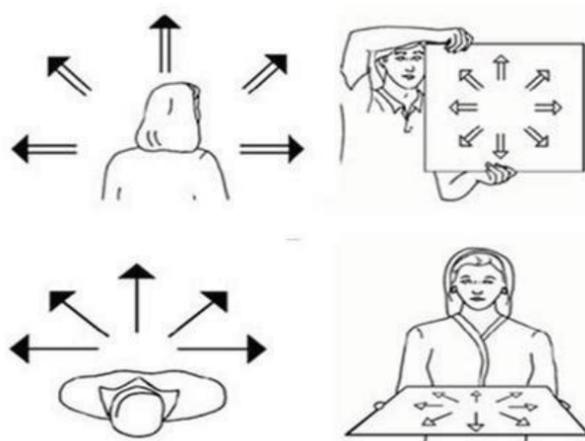


Figura 2.5: Planos: paralelo à parede e ao chão, respectivamente

Fonte: Silva (2012)

	Testa
	Sobrancelhas para cima
	Olhos abertos
	Orelhas
	Por o ar para fora
	Boca com sorriso fechado
	Língua lambe o lábio de cima
	Dentes
	Queixo para cima
	Outras partes: pescoço

Figura 2.6: Expressões faciais

Fonte: Silva (2012)

No SW, existem 10 grupos (conforme Tabela 2.2 apresentada por Stumpf (2002)) de símbolos para representar as configurações de mão, sendo que cada um representa uma diferente utilização dos dedos durante determinada formação. Esses grupos são o início

da sequência de símbolos, que podem ser utilizados como ordenação para dicionários de SW (Sutton, 2002).

Tabela 2.2: Grupos de configurações de mão do SignWriting

Fonte: Stumpf (2002)

Grupo	Símbolo	Descrição	Grupo	Símbolo	Descrição
1		Indicador	6		Mínimo e polegar
2		Indicador e médio	7		Anular e polegar
3		Indicador, médio e polegar	8		Médio e polegar
4		Quatro dedos	9		Indicador e polegar
5		Cinco dedos	10		Polegar

Cada grupo possui particularidades com relação a utilização dos dedos e das mãos, conforme pode ser visualizado na Tabela 2.3: (Sutton, 2002; Stumpf, 2002).

2.2 Técnicas de Classificação

Esta seção apresenta as técnicas de classificação utilizadas neste trabalho. A Seção 2.2.1 apresenta a definição do classificador do tipo Máquina de Vetor de Suporte (Support Vector Machine - SVM). A Seção 2.2.2 apresenta a técnica de aprendizagem profunda. A Seção 2.2.3 descreve o processo de combinação de classificadores. Por fim, na Seção 2.2.4 são apresentadas técnicas de verificação e metaclasses.

2.2.1 Máquinas de Vetor de Suporte

Um problema com apenas duas classes (também chamados de binário) que é linearmente separável pode conter infinitos hiperplanos que dividem (de forma correta) os dados entre estas classes. Diversos classificadores, como os perceptrons, não focam esforços em tentar

Tabela 2.3: Particularidades do grupo 2 (dedo indicador e médio) de configurações de mão do SW

Fonte: Stumpf (2002)

Símbolos	Descrição	Sinal	Significado
	Mão-2 Indicador e médio espalhados		Ver
	Flexionado Mão-2 com dedos flexionados		Aspas
	Mão-U Mão-2 com dedos unidos		Nome
	Curvado-U Mão-U com dedos curvados		Metrô
	Mão-N Mão-U com dedos para frente		Nunca
	Mão-R Mão-U com dedos cruzados		Chefe

encontrar uma fronteira otimizada para o problema, utilizando, assim, a primeira fronteira encontrada. As Máquinas de Vetor de Suporte (Support Vector Machine - SVM) trabalham de forma diferenciada destes, tendo como foco encontrar um hiperplano com a maior distância possível dos pontos de treinamento (Theodoridis and Koutroumbas, 2008).

Considerando um problema binário e com apenas duas características envolvidas, a Figura 2.7 apresenta o conceito de obtenção de margem máxima. Os pontos ressaltados com um círculo extra são chamados de vetores de suporte e são responsáveis por moldar a fronteira de decisão, pois são os exemplos de treino mais próximos à margem máxima de separação. No exemplo em questão, dois hiperplanos ideais de separação são propostos, sendo o primeiro calculado com uma orientação mais horizontal e com uma distância z_1 das linhas obtidas de determinados vetores de suporte e o segundo, uma separação mais diagonal e com uma distância z_2 (maior que a anterior) das linhas obtidas de outros vetores de suporte (Theodoridis and Koutroumbas, 2008; Almeida, 2013).

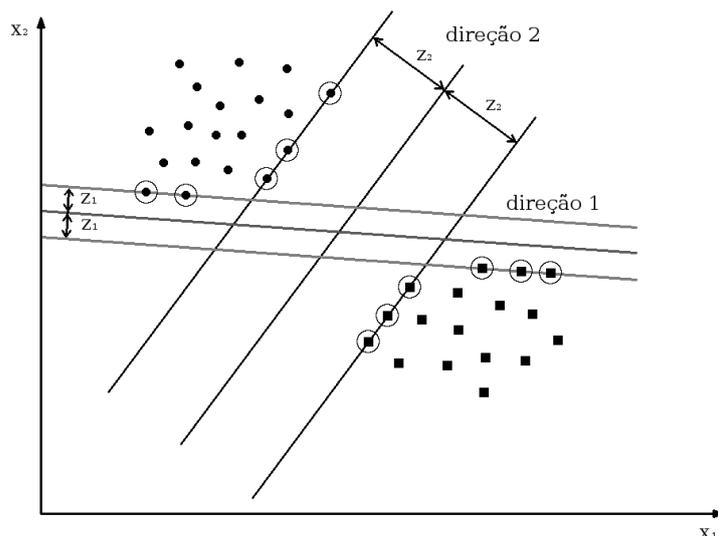


Figura 2.7: Exemplos de hiperplanos para separação entre duas classes
Fonte: Theodoridis and Koutroumbas (2008)

Em muitos casos, os dados não podem ser linearmente separáveis. Para estes casos, é necessário elevar os dados para dimensões mais altas. Nestas dimensões, os dados tendem a ser separáveis linearmente por um hiperplano. Para se realizar esta transformação no conjunto de dados, deve-se lançar mão de funções de kernel, as quais tomam um par de vetores x_j e x_k para computar o produto escalar em determinado espaço de características. As funções de kernel mais utilizadas são:

- Polinomial de grau d

$$K(x_j, x_k) = (1 + x_j \cdot x_k)^d \quad (2.1)$$

- Função de base radial (Radial Basis Function - RBF) de comportamento da exponencial definido por γ

$$K(x_j, x_k) = \exp(-\gamma \|x_j - x_k\|^2) \quad (2.2)$$

- Tangente hiperbólica, para $\beta > 0$ e $\alpha > 0$

$$K(x_j, x_k) = \tanh(\beta x_j \cdot x_k + \alpha) \quad (2.3)$$

Os resultados obtidos após a execução de classificadores SVM são, por padrão, desca-

librados, não gerando probabilidades a posteriori. Porém, Platt (1999) propõe uma forma para converter as saídas do classificador em probabilidades, consistindo, basicamente, na estimação de uma função sigmóide e utilização desta na conversão das saídas dos classificadores em scores entre 0 e 1. Estes scores representam a probabilidade de cada exemplar de teste pertencer à cada uma das classes propostas no problema.

Apesar de os classificadores SVM conseguirem realizar a distinção apenas entre duas classes, eles podem ser adaptados para resolver problemas com mais classes. Pode-se treinar um classificador para cada classe do problema, onde cada um destes diferencie se os exemplares de teste pertencem a classe em questão ou se pertencem à qualquer uma das demais classes do problema (um-contra-todos). Porém, também pode-se tratar o problema no formato um-contra-um, onde treina-se um classificador para cada par de classes existente no problema, o que acaba por totalizar $M(M - 1)/2$ classificadores para um caso de M classes possíveis, sendo dada a decisão final sobre a classe predita do exemplar de teste em questão através de um método de combinação como, por exemplo, uma votação majoritária, incluindo as saídas de todos os classificadores treinados.

2.2.2 Aprendizagem Profunda

Técnicas de aprendizagem profunda (deep learning) tem sido foco de pesquisas na área de reconhecimento de padrões e classificação. Essas técnicas têm aumentado a eficácia de diferentes problemas de reconhecimento de padrões e um de seus principais objetivos é melhorar a descoberta automática de representações multinível (?).

Aprendizagem profunda consiste na utilização de redes neurais profundas (com mais de três camadas), criadas para que trabalhem de forma semelhante ao cérebro de mamíferos, que processa informações através de diversos estágios, como, por exemplo, no caso do sistema visual (Bengio, 2009). Embora o assunto de redes neurais profundas tenha sido estudado por décadas, os resultados obtidos não foram relevantes, havendo problemas com ajustes de parâmetros. Por este motivo os algoritmos ficavam estagnados em mínimas locais quando iniciados com variáveis randômicas. Este cenário mudou com os resultados publicados por Hinton et al. (2006), que utilizou um pré-treinamento para cada camada

com um algoritmo de aprendizagem não supervisionada. Depois desse pré-treino, os pesos aprendidos são utilizados para inicializar a rede neural e então o algoritmo de back-propagation padrão é utilizado para o ajuste fino da rede.

No processo de classificação de imagens, os melhores resultados publicados usam um tipo específico de rede neural, chamada Rede Neural Convolutiva (Convolutional Neural Network - CNN), que combina três ideias arquiteturais: campos receptivos locais, pesos compartilhados e sub-amostragem espacial ou temporal (LeCun et al., 1998). Uma CNN utiliza diversas camadas convolucionais que possuem filtros treináveis com uma pequena área (como 5x5 pixels), que são aplicados por toda a área da imagem de entrada. A troca de informações entre os neurônios dessa estrutura permite o aprendizado de padrões que ocorrem com frequência em qualquer parte da imagem. Apesar de ser plausível, o modelo teórico das CNN ainda não é completamente compreendido (LeCun et al., 1989; Bengio, 2009). A Figura 2.8 apresenta um exemplo de arquitetura de rede neural convolutiva (CNN).

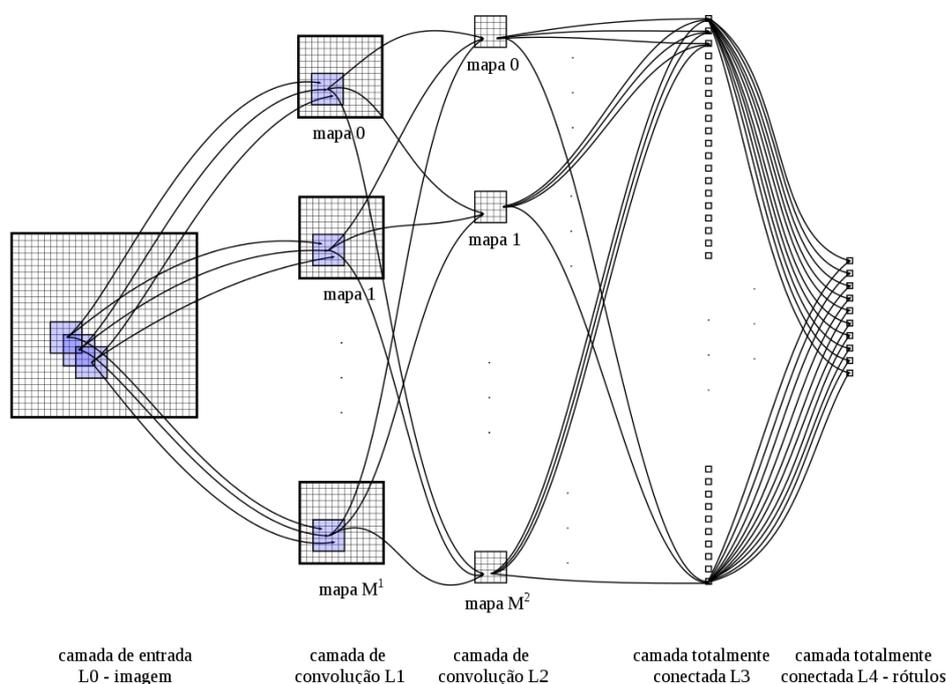


Figura 2.8: Arquitetura de uma rede neural convolutiva para o reconhecimento de sinais de trânsito

Fonte: Cirezan et al. (2011)

Através da CNN, pode-se utilizar dados crus, como, por exemplo, as imagens dos

símbolos de configuração de mão do SW, como entrada para os modelos e deixá-los aprender representações intermediárias, ou seja, deixar os modelos aprenderem a detectar características. Isto é importante em casos onde as características são difíceis de serem determinadas ou formalizadas (Bengio, 2009; ?).

2.2.3 Combinação de Classificadores

Os resultados obtidos para determinados classificadores e/ou conjuntos de características podem não ser, isoladamente, muito satisfatórios, contendo baixas taxas de reconhecimento, especialmente para determinados símbolos. Adicionalmente, algum classificador ou conjunto de características pode conseguir generalizar e classificar de forma mais eficaz alguns símbolos específicos, enquanto outra combinação destes realiza o mesmo, porém para outros símbolos.

Nesses casos, pode-se lançar mão da combinação de classificadores. O objetivo da aplicação desta no problema das configurações de mão do SW é tentar obter um melhor resultado através da fusão das técnicas de classificação apresentadas nas duas seções anteriores.

Mesmo que determinado classificador seja um especialista, unir as informações de diferentes classificadores normalmente resulta em uma decisão final melhor. Porém, para que uma combinação de classificadores seja eficaz, deve-se seguir algumas premissas (Surowiecki, 2005):

- Diversidade de opiniões: todo classificador deve conter ao menos um pedaço da informação, mesmo que este seja fruto de uma interpretação excêntrica;
- Independência: a resposta de um classificador não pode ser afetada pela dos outros;
- Descentralização: cada classificador deve conseguir se especializar e extrair informações do conhecimento local;
- Agregação: deve haver mecanismos para tornar os julgamentos individuais em resultado coletivo.

O resultado de uma classificação pode ser abstrato, baseado em ranking e baseado em probabilidades. Em uma saída abstrata, o classificador em questão informa apenas qual a classe tida como correta. Em uma saída com ranking, o classificador ordena as classes possíveis da mais provável à menos provável. No último caso, utiliza-se probabilidades associadas a cada classe do ranking (Costa, 2013). Quando classificadores geram estimação de probabilidade a posteriori, pode-se utilizá-las para obter a combinação dos resultados de classificação destes. As regras mais comuns para realizar a fusão de classificadores são: voto majoritário, soma, produto, máximo, mínimo e média (Kittler et al., 1998; Salvadeo, 2009; Costa, 2013).

A seguir, são apresentadas, segundo Kittler et al. (1998) e Costa (2013), as regras mais comuns para a realização da combinação dos classificadores, as quais se enquadram na categoria de combinações paralelas mostradas por Jain et al. (2000).

Voto Majoritário

Esta técnica realiza uma votação entre as saídas dos classificadores combinados, assumindo como correto o resultado que mais vezes for citado. É a regra mais simples e popular para realizar a combinação de classificadores, sendo que as estimativas de probabilidade não se fazem necessárias, permitindo combinar até classificadores com resultados abstratos. A equação 2.4, calcula a votação para uma amostra x , onde n é o número de classificadores, y_i a classe de saída do classificador de número i , para os possíveis rótulos de classe $\Omega = \omega_1, \omega_2 \dots \omega_c$.

$$mv(x) = \underset{k=1}{\overset{c}{\operatorname{argmax}}} \sum_{i=1}^n y_{i,k} \quad (2.4)$$

Regra do Produto

Esta regra, assim como as demais, exceto a do voto majoritário, necessita de um classificador que gere saídas de probabilidade estimadas para cada classe do problema. Realiza a combinação pelo cálculo do produto entre as probabilidades associadas às saídas dos classificadores c_i . É uma regra bastante severa, bastando que o resultado de apenas um

dos classificadores não esteja de acordo com a decisão geral para que o resultado final seja penalizado, sendo, portanto, indicada para situações mais críticas, onde erros não são aceitos. Na equação 2.5, x é o padrão a ser classificado, onde n é o número de classificadores, y_i a classe de saída do classificador de número i , para os possíveis rótulos de classe $\Omega = \omega_1, \omega_2, \dots, \omega_c$ e onde $P(\omega_k | y_i(x))$ é a probabilidade de x pertencer à classe ω_k do classificador de número i .

$$pr(x) = \underset{k=1}{\overset{c}{\operatorname{argmax}}} \prod_{i=1}^n P(\omega_k | y_i(x)) \quad (2.5)$$

Regra da Soma

Esta regra realiza a combinação pelo cálculo da soma entre as probabilidades associadas às saídas dos classificadores c_i . É menos severa que o produto, não penalizando tanto o resultado final se algum classificador errar, sendo, portanto, indicada para situações menos críticas. O cálculo dessa regra é representado pela equação 2.6.

$$sr(x) = \underset{k=1}{\overset{c}{\operatorname{argmax}}} \sum_{i=1}^n P(\omega_k | y_i(x)) \quad (2.6)$$

Regra da Média

Esta regra calcula a média entre as probabilidades associadas às saídas dos classificadores, produzindo resultados parecidos com os da soma. O cálculo dessa regra é representado pela equação 2.7.

$$mr(x) = \frac{1}{n} \underset{k=1}{\overset{c}{\operatorname{argmax}}} \sum_{i=1}^n P(\omega_k | y_i(x)) \quad (2.7)$$

Regra do Máximo

Esta regra considera apenas a maior probabilidade dentre as classes, sendo que para cada classe, pega a maior probabilidade informada por todos os classificadores. A regra do máximo é de baixa severidade, pois se uma classe obter um bom desempenho em

qualquer classificador, ela tem uma grande chance de ser escolhida. O cálculo dessa regra é representado pela equação 2.8.

$$\max(x) = \underset{k=1}{\overset{c}{\operatorname{argmax}}} \max_{i=1}^n P(\omega_k | y_i(x)) \quad (2.8)$$

Regra do Mínimo

Esta regra considera o maior mínimo dentre as classes, sendo que para cada classe, pega a menor probabilidade informada por todos os classificadores. A regra do mínimo é de alta severidade. O cálculo dessa regra é representado pela equação 2.9.

$$\min(x) = \underset{k=1}{\overset{c}{\operatorname{argmax}}} \min_{i=1}^n P(\omega_k | y_i(x)) \quad (2.9)$$

2.2.4 Verificação

Para a resolução de confusões pontuais, pode-se lançar mão de técnicas de agrupamento de classes. Pode-se utilizar o conceito de níveis de verificação, o qual considera dois níveis: alto e baixo. A verificação de alto nível considera subconjuntos da base original, a fim de confirmar ou negar a hipótese produzida pelo classificador genérico. A verificação de baixo nível trabalha com a noção de metaclasses, como caracteres ou partes destes, tendo como propósito não reconhecer um caractere específico, e sim, determinar se a hipótese gerada pelo classificador é válida ou não (Oliveira et al., 2003).

A Figura 2.9 apresenta o processo de verificação que ocorre após a rejeição de determinado resultado, devido a este não ter atingido um nível mínimo de confiabilidade (de acordo com os resultados de probabilidade fornecidos pelo processo de classificação aplicado) e possuir confusões clássicas previamente levantadas. Estas especificidades serão tratadas na etapa de verificação apresentada, gerando um novo resultado, que pode ser o mesmo (verificado) do classificador anterior, ou não (não verificado).

Dessa forma, pode-se utilizar o conceito de metaclasses proposto para criar uma verificação de confusões específicas, agrupando os símbolos com altas taxas de confusão, analisando suas semelhanças e diferenças estruturais e criando um classificador ou con-

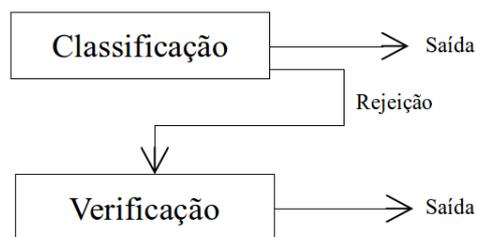


Figura 2.9: Processo de verificação de resultado de classificação

junto de classificação específico para reconhecer essas confusões, podendo-se, inclusive, agregar novas características específicas no vetor de características inicial.

Como a etapa de verificação é um procedimento realizado em momento posterior à classificação de forma genérica, com o objetivo de reforçar o resultado desta, ele não precisa, necessariamente, se utilizar do mesmo classificador previamente criado, tampouco reaproveitar as mesmas características, visto que se faz necessária a detecção de classificadores e características especializadas para cada uma das confusões que se deseja verificar.

CAPÍTULO 3

ESTADO DA ARTE

Este capítulo apresenta uma revisão do estado da arte de SignWriting aplicado à área da computação. O SignWriting ainda é um assunto computacionalmente pouco explorado. A quantidade de trabalhos encontrados é relativamente baixa e de conteúdo não totalmente relacionado com a área efetiva de estudo deste trabalho, sendo os mais relevantes apresentados em ordem cronológica de publicação.

Costa and Dimuro (2001a), conhecendo a crescente tendência de utilização do SW pelas comunidades surdas utilizadoras de línguas de sinais, estudam as características gráficas de símbolos e suas combinações presentes no sistema. A análise de tais símbolos permite a organização dos mesmos em subconjuntos que visam agrupar características e/ou significados semelhantes. Como resultado dessa análise, é proposta a criação da SignWriting Markup Language (SWML), uma linguagem baseada em eXtended Markup Language (XML), a qual visa codificar e permitir a representação dos símbolos e termos existentes no SW através de um arquivo que descreve textualmente (de forma estruturada) os detalhes gráficos de apresentação de cada um destes. A criação dessa sintaxe tem por objetivo permitir que textos em SW sejam computacionalmente representados, possibilitando assim, por exemplo, ações como: entrada e saída de dados de aplicações, armazenamento e recuperação de dados, análise, geração, tradução, correção e busca em documentos e animação.

A linguagem SWML pode permitir a interoperabilidade de diferentes aplicações, desenvolvidas em diferentes linguagens e utilizadoras de diferentes ambientes de execução, de uma forma totalmente padronizada, como prega a própria especificação do XML. Para reforçar tal afirmação, o trabalho cita a utilização da linguagem em algumas aplicações simples desenvolvidas em um projeto conhecido como SignNet Project. Alguns exemplos são:

- Ferramenta para exibição de textos em SW;
- Editor de textos em SW;
- Páginas web mostrando seu conteúdo em SW;
- Dicionário online para línguas de sinais;
- Repositórios de documentos em SW, utilizando busca baseada em pedaços de textos em SW;
- Animação automatizada de sentenças e diálogos.

Um exemplo da sintaxe da linguagem SWML é apresentado na Figura 3.1.

```

<?xml version="1.0" ?>
<swml version="1.0-d2" symbolset="SSS-1995">
  <generator>
    <name>SignWriter</name>
    <version>4.3</version>
  </generator>
  <sw_text>
    <sw_text_defaults>
      <sign_boxes>
        <unit> pt </unit>
        <height> 60 </height>
      </sign_boxes>
      <text_boxes>
        <box_type> graphic_box </box_type>
        <unit> pt </unit>
        <height> 60 </height>
      </text_boxes>
    </sw_text_defaults>
    <new_line />
    <sign_box>
      <!-- the B hand -->
      <symbol x="8" y="13">
        <shape number="21" fill="1" variation="1" />
        <transform flop="0" rotation="0" />
      </symbol>
      <symbol x="7" y="25">
        <!-- the movement -->
        <shape number="108" fill="0" variation="1" />
        <transform flop="1" rotation="4" />
      </symbol>
    </sign_box>
  </sw_text>
</swml>

```

Figura 3.1: Arquivo no formato SWML para a palavra Brasil
Fonte: Costa and Dimuro (2001a)

A proposta dos autores é válida, porém, a mesma abra diversas possibilidades para futuras implementações de aplicações que trabalhem com SW, ela ainda é falha, pois não consegue cobrir de forma eficiente todos os símbolos e formações existentes na notação.

Em trabalho correlato, Costa and Dimuro (2001b) propõem a renderização de símbolos, oriundos de dados em SWML, através da utilização do formato de imagens vetorial Scalable Vector Graphics (SVG), apresentando as vantagens da aplicação deste conjunto (SWML + SVG) em aplicações web. A intenção dessa proposta é substituir a utilização de imagens estáticas em sites com conteúdo voltado para pessoas surdas, os quais, por padrão, utilizam-se do formato Graphics Interchange Format (GIF) para tal, o que viria a criar uma possibilidade muito maior de geração de conteúdo de forma automatizada e padronizada.

Porém o trabalho apenas propõe tal utilização, não sendo fornecida qualquer característica relacionada a real implementação, como frameworks, bibliotecas ou especificações.

Os mesmos autores ainda apresentam uma introdução a um possível procedimento de busca automatizada de símbolos em um texto ou dicionário de SW representado em SWML, sendo levantadas as principais dificuldades e vantagens iniciais detectadas para a realização desta tarefa. Ponto, este, que é explorado de forma mais ampla em momento posterior, no qual Costa et al. (2004) propõem uma técnica, baseada na semelhança gráfica, para fazer a comparação entre dois símbolos de SW apresentados sob o formato SWML, com o objetivo de saber se, mesmo sem uma equivalência de 100% nos padrões gráficos, os dois representam o mesmo sinal. Seu funcionamento baseia-se no princípio de que determinados itens da representação do sinal (como posições das mãos e setas de movimento) podem sofrer pequenas variações sem interferir no significado, porém alguns outros (como posição de indicadores de contato e rotação de alguns símbolos) têm grande impacto na leitura geral do mesmo.

Este trabalho também não pode ser considerado completo, pois não oferece um protocolo de como as informações seriam indexadas e comparadas.

Torchelsen et al. (2002) utilizaram a tecnologia do SWML para propor o SWEdit, um sistema para auxiliar usuários surdos na criação de textos em linguagem de sinais, baseados no sistema de representação de sinais em SignWriting. O SWEdit consiste em um editor de texto, que permite a redação através da escolha de símbolos do SW em

menus categorizados de acordo com a natureza dos símbolos desejados. O aplicativo produzido assemelha-se a editores de texto convencionais (para idiomas orais), porém uma característica básica torna sua utilização bastante diferente destes: a necessidade de maior utilização do mouse para a formação dos verbetes, devido a dificuldade de aplicação direta do teclado para a representação dos símbolos do SW. Este fato resulta em elevado tempo de confecção dos sinais, o que acaba por afetar o desempenho do autor durante a produção de um texto. A interface do aplicativo é apresentada na Figura 3.2.

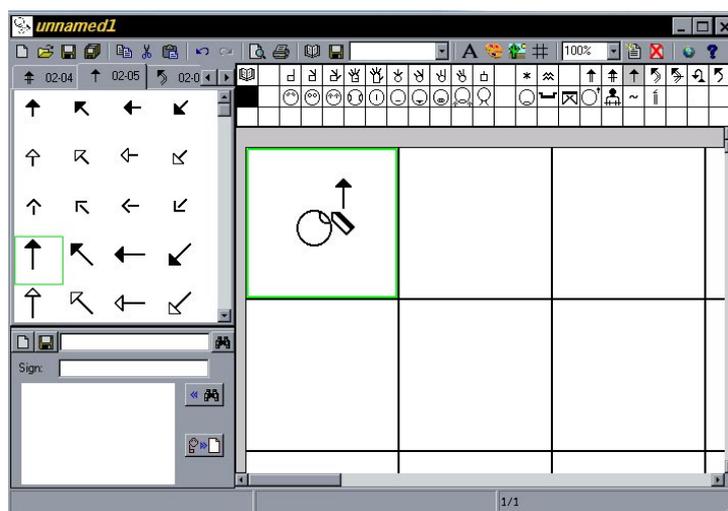


Figura 3.2: Interface do SWEdit
Fonte: SignWriting (2004)

Papadogiorgaki et al. (2004) se utilizaram da linguagem SWML proposta por Costa and Dimuro (2001a) para propor uma abordagem para a geração de uma animação usando uma Linguagem de Modelagem de Realidade Virtual (Virtual Reality Modeling Language - VRML) baseada no formato MPEG-4 Body Animation, onde a representação de cada símbolo em SW é convertida em uma sequência de Body Animation Parameters (BAPs) do padrão MPEG-4, que visa corresponder ao sinal real que este representa. O objetivo de seu trabalho é permitir a animação de um avatar virtual, de tal forma que este consiga apresentar, de forma gestual, determinado sinal criado em SW e estruturalmente especificado através do SWML, com possibilidades de aplicação em páginas web, clientes de e-mail, telas em quiosques de informação e como auxílio para pessoas surdas em noticiários televisivos. Como uma sequência do citado trabalho, Papadogiorgaki et al. (2006) realizaram uma integração com o formato MPEG-4 Facial Animation, convertendo os detalhes

faciais apresentados no símbolo em SWML para uma sequência de Facial Animation Parameters (FAPs), que serve como complemento ao avatar do trabalho anterior. Porém, devido à complexidade envolvida na reprodução de sinais, através da simples utilização de um avatar virtual, não é possível representar com perfeição os sinais que se deseja emitir. Como este trabalho está vinculado ao SWML, ele também está sujeito às restrições anteriormente citadas.

Ahmed and Seong (2006), considerando as dificuldades apresentadas por surdos durante a expressão e comunicação via língua oral escrita, bem como o alto custo de aparelhos específicos para comunicação entre surdos (Telecommunication Devices for the Deaf - TDD), criaram o protótipo de uma aplicação para a leitura e escrita de mensagens SMS (Short Messaging Service) para ser utilizado em aparelhos celulares comuns, em substituição (ou complemento) ao aplicativo padrão destes mesmos dispositivos. Os autores realizaram uma pesquisa de aceitação junto à comunidade surda, tendo como resultado um total de 88,80% de preferência deste novo aplicativo em comparação ao tradicional leitor/editor de SMS do dispositivo. Porém, é importante notar que esse trabalho não leva em consideração nenhum detalhe técnico necessário para a criação de tal aplicativo, nem questões de viabilidade em nenhum ponto, como, por exemplo: hardware, software, meio de comunicação e financeira.

Esta proposta se mostra inviável, pois o SWML produz uma quantidade de dados muito elevada para ser trafegada via SMS. Nos dias atuais, com a popularização dos smartphones e ampliação das conexões de rede, um aplicativo equivalente poderia ser viável.

Os trabalhos de Moemedi and Connan (2010) e Moemedi (2010) propõem uma aplicação semelhante à proposta por Papadogiorgaki et al. (2004), utilizando, inclusive, as mesmas tecnologias, porém sobre sinais da South African Sign Language (SASL), aplicando um estudo de viabilidade que concluiu que 82% dos sinais representados pelos avatares virtuais foram reconhecidos pelas pessoas conhecedoras da SASL consultadas, mesmo com uma coerência de 62% entre o símbolo em SW dado como entrada e sua efetiva representação na forma de sinal virtualizada. Devido às suas naturezas, as conclusões para

estes trabalhos são equivalentes às anteriormente propostas para respeito a manipulação de avatares virtuais.

Lu et al. (2010) apresentam um sistema para interação humano-computador (IHC) desenvolvido para o reconhecimento e visualização de movimentos manuais. Seu trabalho foca em formas de interpretação de sinais (obtidos através da utilização de uma ou mais câmeras e algoritmos de visão computacional) e suas possíveis conversões em símbolos do SW, focando em quatro diferentes formas de identificação:

- Identificação de movimentos repetidos;
- Identificação de planos de movimento;
- Identificação de direções de movimento;
- Identificação de linearidade do trajeto do movimento.

Este trabalho não considera um ambiente real de emissão de mensagens por meio de sinais e sim, apenas um ambiente controlado monitorado com câmeras, que precisam estar bem alinhadas e calibradas.

Bianchini et al. (2012a,b) propõem o SWift (SignWriting improved fast transcriber), uma ferramenta avançada para a edição de símbolos em SW, que permite a colaboração e o intercâmbio de ideias entre utilizadores de línguas de sinais, na tentativa de quebrar as barreiras eletrônicas que mantém as comunidades surdas longe das Tecnologias da Informação e Comunicação (TICs) em geral, porém especificamente do e-learning. A aplicação permite compor e salvar os símbolos desejados usando os componentes básicos do SW. O uso da aplicação é difícil e pouco funcional, pois seu ambiente é ainda mais restrito que o editor anteriormente citado, sendo que os mesmos problemas são válidos para este caso.

Bouزيد and Jemni (2013) também trabalham com a questão da geração de avatares virtuais para a sintetização de movimentos de línguas de sinais, previamente gerados e armazenados em SW através do SWML, porém sua proposta difere da de Papadogiorgaki et al. (2004) na geração dos dados de saída, pois, enquanto o trabalho anterior utiliza-se do formato predefinido MPEG-4 BAP, o trabalho em questão define a utilização de

mais uma metalinguagem intermediária no processo, o SML (Sign Modeling Language), que é uma linguagem desenvolvida para suportar a sucessiva movimentação de um grupo de junções, sendo cada movimento tem seu tempo prefixado. Seu trabalho tem foco na extração e síntese de características nem sempre totalmente visíveis em SWML, gerando uma saída em SML, que proporciona uma saída mais “legível”, com relação à descrição de uma animação. O processo realizado é descrito na Figura 3.3.

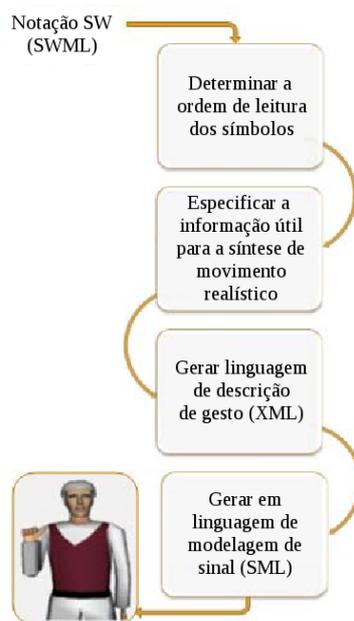


Figura 3.3: Processo de sistema proposto por Bouzid and Jemni (2013)
 Fonte: Bouzid and Jemni (2013)

Guimaraes et al. (2014), considerando a importância do ensino de línguas de sinais, bem como sua representação escrita (através do SW), expõem a necessidade por ferramentas computacionais voltadas para usuários surdos, principalmente na área do SW. Muitas das ferramentas existentes são inadequadas e acabam por não cumprir de forma eficaz seus objetivos, como no citado caso do editor SWEdit (Torchelsen et al., 2002), que, apesar de cumprir seu objetivo, necessita uma série de comandos difíceis e que demandam uma elevada quantidade de tempo para obter um quadro com as primitivas do símbolo SW desejado. Com base nessas necessidades, e reconhecendo a existência de estudos na área de reconhecimento de caracteres manuscritos, os autores propõem uma nova e mais natural abordagem: usando uma tela (tablet) e um método manual de inserção (caneta) para o reconhecimento de símbolos SW manuscritos. Este trabalho divulga uma base de

dados com transcrições manuscritas de 103 símbolos de configuração de mão do SW pré-selecionados, que podem ser futuramente utilizados junto a técnicas de reconhecimento de padrões, a fim de identificar outros símbolos.

Apesar de os autores deste último trabalho apresentarem a base de dados utilizada nesta pesquisa, o trabalho ainda deixa muitas lacunas a respeito da real aplicação e suas possíveis implementações em um sistema de reconhecimento real. Lacunas, estas, que são, em partes, mais amplamente exploradas aqui.

A Tabela 3.1 apresenta uma sumarização dos trabalhos apresentados neste capítulo.

Tabela 3.1: Trabalhos relacionados no estado da arte

Autor(es)	Proposta
Costa and Dimuro (2001a)	SWML
Costa and Dimuro (2001b)	SWML + SVG
Torchelsen et al. (2002)	SWEdit
Costa et al. (2004)	Técnica para comparação de símbolos SW
Papadogiorgaki et al. (2004)	Geração de animação de avatar VRML
Papadogiorgaki et al. (2006)	Geração de animação de avatar VRML
Ahmed and Seong (2006)	Protótipo de aplicação para SMS
Moemedi and Connan (2010); Moemedi (2010)	Geração de animação de avatar VRML
Lu et al. (2010)	Reconhecimento de sinais para SW
Bianchini et al. (2012a,b)	SWift
Bouزيد and Jemni (2013)	SML e geração de animação de avatar
Guimaraes et al. (2014)	Base de dados de configurações de mão

CAPÍTULO 4

MÉTODO PROPOSTO

O método proposto para este trabalho é composto por quatro módulos: formação de base de dados (imagens) do conjunto de símbolos, definição e extração de características relevantes para o reconhecimento dos símbolos, realização de estudo de diferentes estratégias de classificação e análise e comparação de resultados das diversas estratégias utilizadas.

Devido à não existência de trabalhos ou bases de dados prévios para a área de estudo, a avaliação de desempenho se dará pela comparação da taxa de reconhecimento obtida de cada uma das estratégias adotadas durante a classificação das amostras.

4.1 Conjunto de Símbolos e Base de Dados

O conjunto de símbolos utilizado neste trabalho é oriundo de Guimaraes et al. (2014) e foi escolhido de modo que represente, de forma ampla e genérica, diversas características encontradas nas configurações de mão do SW, como quantidade e posição de dedos, orientação da palma da mão e formato básico do sinal. O conjunto é composto por 103 símbolos (classes), os quais estão apresentados na Figura 4.1.

A base de dados construída para este trabalho tem por objetivo mostrar a representação manuscrita de cada um dos 103 símbolos de configuração de mão, sendo esta um desenho à mão livre obtido digitalmente, com o auxílio de um tablet.

O processo de coleta consiste na transcrição de cada um dos 103 símbolos por voluntários de diferentes faixas etárias, surdos, ou não, e conhecedores do SW ou não. Cada pessoa demora em média 30 minutos para concluir todo o processo.

A base de dados é composta por 79 coletas dos 103 símbolos, de forma isolada um do outro, totalizando 8.137 imagens, porém apenas 7.994 são válidas para o processo de extração de características, devido a inserções de forma incorreta por parte dos usuários da aplicação. A Tabela 4.1 apresenta o total de amostras válidas para cada uma das 103

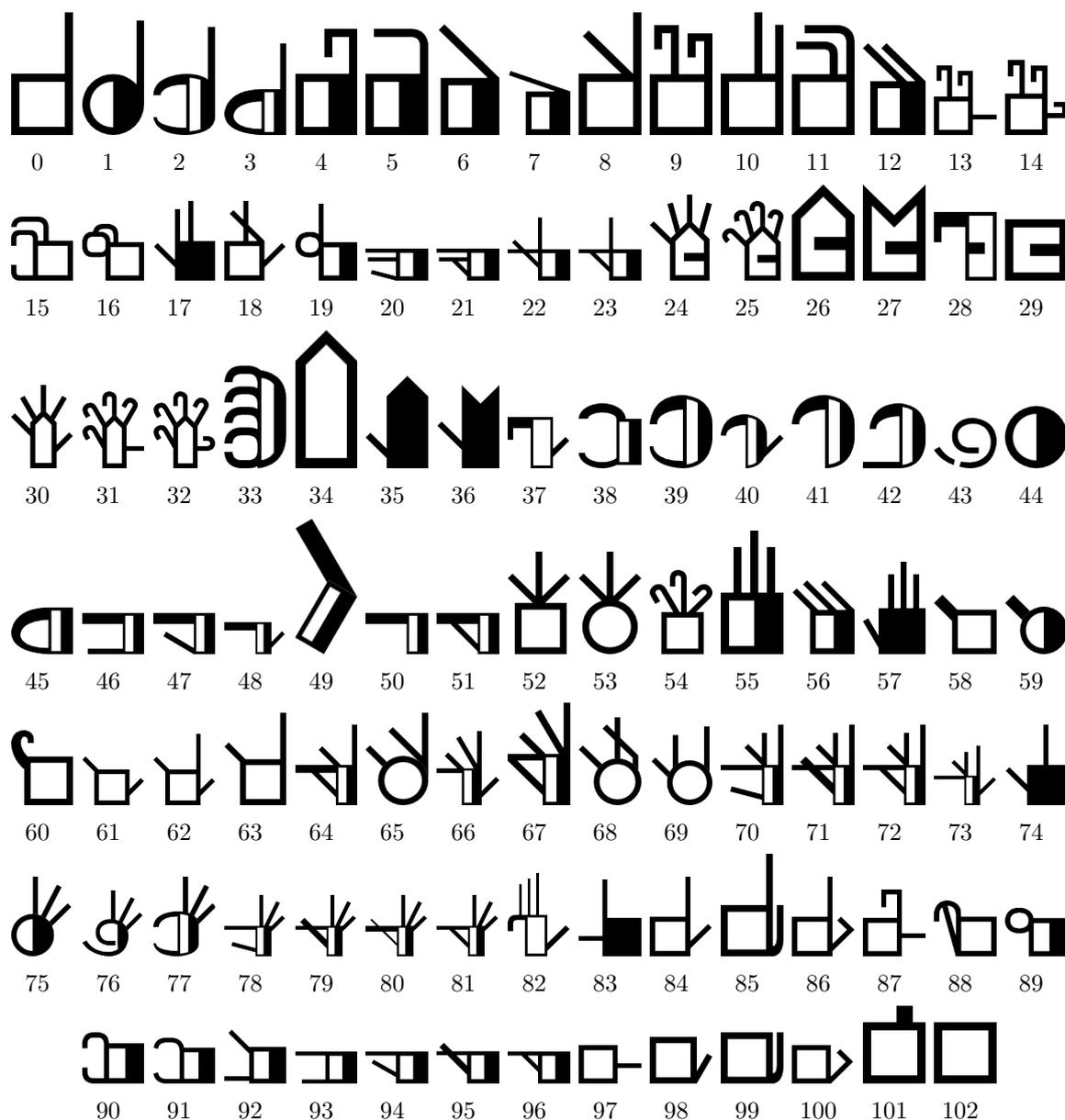


Figura 4.1: Configurações de mão da base de dados
 Fonte: Guimaraes et al. (2014)

classes.

Para realizar a captura de base de dados, foi desenvolvida uma aplicação em Android, que executa em um tablet Samsung Galaxy Tab e utiliza uma Interface de Programação de Aplicações (Application Programming Interface - API) da fabricante (Samsung) para a leitura de inserções em tela (toque) por parte do usuário. A Figura 4.2 apresenta a aplicação para captura de base sendo utilizada.

Todas as imagens geradas possuem resolução de 480×519 pixels e estão armazenadas no formato Portable Network Graphics (PNG) em escala de cinza, devido à suavização

Tabela 4.1: Total de amostras válidas por classe (símbolo)

Clas.	Qtd								
0	79	21	79	42	78	63	76	84	76
1	79	22	79	43	78	64	78	85	76
2	79	23	79	44	78	65	76	86	76
3	79	24	79	45	77	66	76	87	77
4	79	25	78	46	78	67	77	88	77
5	78	26	79	47	78	68	77	89	77
6	79	27	79	48	78	69	78	90	77
7	78	28	77	49	77	70	77	91	77
8	79	29	77	50	78	71	77	92	77
9	78	30	76	51	78	72	78	93	77
10	79	31	76	52	78	73	78	94	77
11	79	32	79	53	78	74	78	95	77
12	79	33	79	54	76	75	78	96	76
13	79	34	79	55	77	76	77	97	77
14	79	35	79	56	78	77	76	98	76
15	79	36	78	57	78	78	77	99	77
16	78	37	78	58	78	79	76	100	77
17	78	38	77	59	78	80	77	101	77
18	78	39	78	60	78	81	73	102	77
19	78	40	78	61	76	82	77		
20	79	41	78	62	77	83	76		

aplicada pela aplicação para captura de base. Paralelamente, uma base normalizada dessas imagens é mantida.

A normalização de cada uma das imagens se dá através da aplicação de erosão e dilatação (uma iteração) utilizando, para ambas, um elemento estruturante retangular de 19×19 pixels. Após este processo, descarta-se as quatro bordas em branco que envolvem o símbolo e a imagem resultante é redimensionada para 30×30 pixels.

O diretório onde se encontram as imagens está organizado de tal forma (separado por pessoas e suas capturas) que novas amostras podem ser adicionadas a qualquer momento sem prejuízo à execução da aplicação geradora do arquivo de características e ao processo de classificação.

A base de dados (original e normalizada), bem como os símbolos-base, podem ser adquiridos através do endereço <http://web.inf.ufpr.br/vri/signwriting-database>.



Figura 4.2: Utilização da aplicação para captura de base
 Fonte: Guimaraes et al. (2014)

4.2 Extração de Características

Devido à semelhança entre os símbolos de configuração de mão do SW e caracteres manuscritos, as características foram selecionadas de acordo com o sugerido por trabalhos prévios na área de reconhecimento de caracteres (Rajashekararadhya and Ranjan, 2009; Feng and Manmatha, 2005; Chen et al., 2010).

As seções subsequentes apresentam as características que são empregadas no reconhecimento de caracteres, bem como suas aplicações no reconhecimento de símbolos do SW.

4.2.1 Histogramas de Projeção

Histogramas de projeção são utilizados desde 1956 na área de Optical Character Recognition (OCR), sendo que atualmente essa técnica é mais usada para segmentação de caracteres, palavras e linhas de texto, ou para detectar se uma imagem escaneada encontra-se rotacionada. Os histogramas de projeção são muito sensíveis à rotação e levemente sensíveis à angulação e estilo de escrita. Um histograma horizontal é um vetor de tamanho igual à altura (em pixels) da imagem, no qual cada posição contém a sumarização dos pixels “preenchidos” para a linha relacionada nesta imagem, enquanto um

histograma vertical é um vetor de tamanho igual à largura (em pixels) da imagem, no qual, cada posição contém a sumarização dos pixels “preenchidos” para a coluna relacionada nesta imagem (Trier et al., 1996).

A utilização desse histograma no conjunto de características relevantes para este trabalho se dá através de um vetor de características de 60 posições para cada amostra de imagem normalizada de símbolo existente na base de dados, sendo 30 posições reservadas ao histograma vertical e outras 30 para o horizontal. Ele é relevante para o problema, pois consegue demonstrar de forma eficaz variações na forma e volume das imagens, permitindo, por exemplo, a detecção de picos e áreas altamente preenchidas, sendo que histogramas de mesmos símbolos, quando escritos por diferentes pessoas, tendem a apresentar características semelhantes.

A (Figura 4.3) apresenta os histogramas de projeção verticais e horizontais de duas diferentes configurações de mão do SW reproduzidas pelo mesmo usuário.

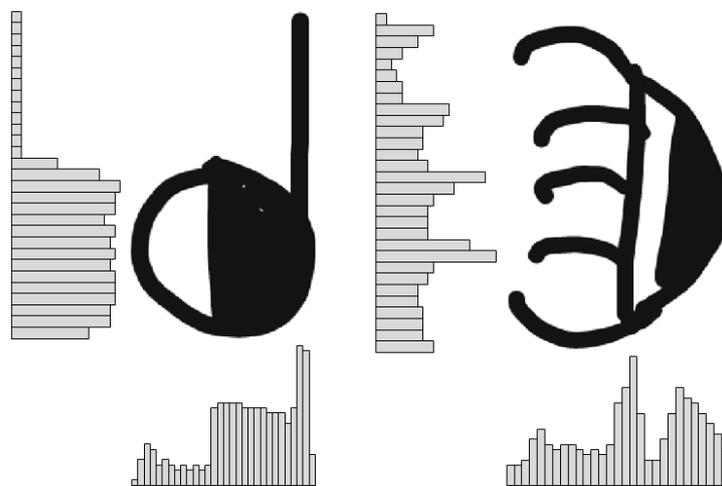


Figura 4.3: Histogramas de projeção horizontal e vertical
Fonte: Autoria própria

4.2.2 Histograma de Orientação

O histograma de orientação (Trier et al., 1996), ou histograma de direção dos contornos (Oliveira et al., 2002), é baseado no contorno externo da imagem.

A curva de contorno externo fechado de um caractere é a curva linear que passa pelo centro de todos os pixels 4-conectados ao plano de fundo exterior da imagem e nenhum

outro pixel. Seguindo o contorno, os pixels são verificados em sentido anti-horário e a curva deve visitar um pixel de borda duas vezes em locais onde o objeto é de largura de apenas um pixel. Cada segmento de linha é uma linha direta entre os centros dos pixels de uma 8-vizinhança (Trier et al., 1996).

Para extrair esta característica, utiliza-se, por motivos de precisão, as imagens não normalizadas, a qual é dividida em seis zonas (duas colunas e três linhas) e os segmentos de linha são agrupados de acordo com sua orientação (angulação): horizontal (0°), vertical (90°) e as duas diagonais (45° e 135°). O número de segmentos de linha é totalizado para cada orientação e em cada uma das zonas. A quantidade de orientações pode ser de quatro, se as duas arestas de mesmo ângulo forem agrupadas, e de oito, caso estas sejam consideradas de maneira individual (Trier et al., 1996; Oliveira et al., 2002).

A Figura 4.4 apresenta um exemplo de geração de histograma de orientação do contorno, com base em um zoneamento de quatro linhas por quatro colunas.

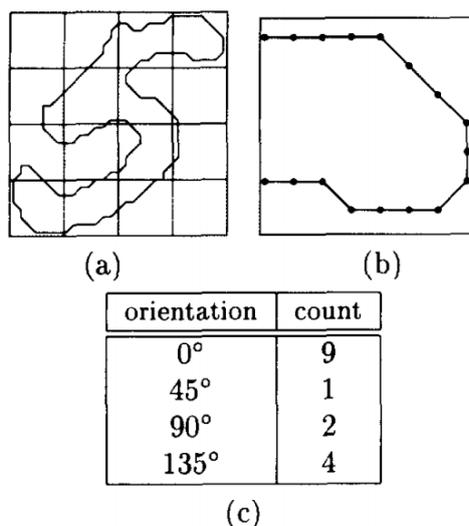


Figura 4.4: Zoneamento da curva de contorno: (a) divisão 4x4 da imagem; (b) canto superior direito; (c) histograma de orientação para a zona em questão.

Fonte: Trier et al. (1996)

A utilização desse histograma no problema das configurações de mão permite uma melhor representação da orientação das linhas presentes em cada uma das zonas definidas da imagem em questão. Sua utilização no conjunto de características relevantes para este trabalho se dá através de um vetor de características de 48 posições para cada amostra de imagem não normalizada de símbolo existente na base de dados, sendo este composto

por 6 zonas, com suas respectivas 8 características relativas à orientação.

4.2.3 Histograma de Concavidade

Na medição de concavidades, na mesma imagem zoneada (duas colunas e três linhas), para cada pixel branco procura-se em uma 4-vizinhança, sequencialmente até as bordas ou até encontrar pixels pretos. Um histograma de 13 posições é criado para cada zona, contendo a sumarização das vizinhanças em diferentes combinações (Figura 4.5b). Pixels com somente um vizinho preenchido não são considerados, bem como pixels com dois vizinhos paralelos. Por exemplo, x_2 na Figura 4.5a é classificado na posição 7 do vetor de característica, por possuir pixels pretos encontrados nas posições 0, 1 e 2 e por não possuir pixels pretos na posição 3 de acordo com a 4-vizinhança (Figura 4.5d) (Oliveira et al., 2002).

Pixels brancos com pixels pretos encontrados em todas as direções (Figura 4.5d) podem estar contidos em contornos fechados, sendo classificados na posição 8 do vetor de características, mas também podem não estar contidos em tais contornos, indicando uma forma aberta, gerando a necessidade da definição de variáveis de direção auxiliares, que indicam a forma de “saída” da concavidade em questão (Figura 4.5c). Por exemplo, x_1 na Figura 4.5a é classificado na posição 9, pois possui pixels pretos nas quatro direções, porém não está contido em uma área totalmente fechada, possuindo uma “saída” da concavidade que o contém à esquerda e acima (Oliveira et al., 2002).

Essa característica é importante para o problema das configurações de mão do SW, pois permite a detecção das formas mais abrangentes em cada imagem, bem como a não presença de determinados formatos de concavidades, possibilitando maior eficácia na distinção de símbolos.

A utilização desse histograma no conjunto de características relevantes para este trabalho se dá através de um vetor de características de 78 posições para cada amostra de imagem normalizada de símbolo existente na base de dados, sendo este composto por 6 zonas, com suas respectivas 13 características relativas à concavidade.

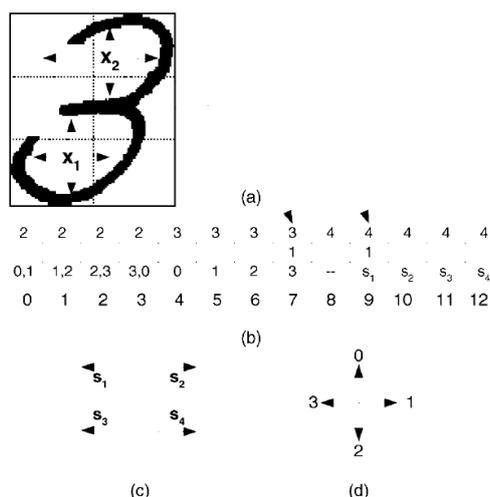


Figura 4.5: Medição de concavidades: (a) concavidades; (b) vetor de características; (c) direções auxiliares; (d) direções na 4-vizinhança.

Fonte: Oliveira et al. (2002)

4.2.4 Zoneamento

O método de zoneamento é utilizado para a extração das características citadas anteriormente, mas também pode ser utilizado de forma isolada para computar o percentual de pixels pretos (preenchidos) em cada zona da imagem, devendo, para tal, a imagem ser subdividida em várias fatias (Trier et al., 1996).

Impedovo and Pirlo (2014) mostram diversas formas para a realização do zoneamento de imagens, focando no reconhecimento de caracteres manuscritos, dividindo as técnicas em dois grandes grupos: estáticas e adaptáveis. Técnicas estáticas levam em consideração que não se tem informações a priori sobre a imagem a ser zoneada, enquanto as adaptáveis resultam de algum tipo de pré-processamento sobre a mesma. Devido à característica segmentada das imagens dos símbolos, a técnica de zoneamento selecionada para o problema em questão é o zoneamento uniforme, do grupo das estáticas, que consiste em dividir a imagem em determinada quantidade de linhas e colunas, criando blocos de tamanho equivalente.

Para a extração de cada uma das características anteriormente citadas, cada imagem é dividida em 3 linhas e 2 colunas, totalizando 6 zonas (Figura 4.6), contendo, cada uma delas, os dados expostos por cada característica em questão.

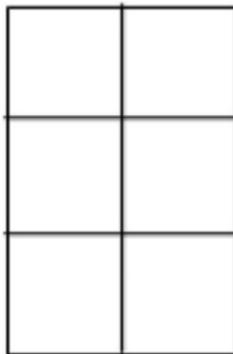


Figura 4.6: Zoneamento uniforme utilizado
Impedovo and Pirlo (2014)

4.3 Classificação

O processo de classificação das configurações de mão contempla quatro estratégias independentes, objetivando melhorar a eficácia na identificação dos símbolos de configuração de mão do SignWriting. As quatro estratégias são: testar classificadores e características individualmente, aplicar técnicas de aprendizagem profunda, realizar a combinação de diversos classificadores e definir metaclasses para verificação.

Considerando-se a base de dados de símbolos, realiza-se um pré-processamento em cada uma das imagens, com o objetivo de equalizar as amostras anteriormente à qualquer operação de reconhecimento de padrões. Como resultado deste, obtêm-se um subconjunto da base de imagens válidas original, com a mesma quantidade de amostras, porém com tamanho reduzido para 30 por 30 pixels e simplificação por binarização (remoção de níveis de cinza). Com as imagens já normalizadas, realiza-se a extração de características de cada uma. Esses dados servem como fonte para treinamento e teste de classificadores na etapa seguinte. Na sequência, de forma paralela, são utilizadas as técnicas de aprendizagem profunda, apresentadas na Seção 2.2.2, diretamente nas imagens normalizadas. Uma terceira etapa visa utilizar diferentes técnicas para realizar a fusão dos resultados das classificações previamente obtidos através dos classificadores e da aprendizagem profunda. Como uma última estratégia, com base na análise de confusões entre classificações de símbolos e em características gráficas de cada um deles, é realizado um agrupamento de caracteres por semelhança e posterior reclassificação, sendo consideradas possibilidades

de aplicações de classificadores para a resolução de confusões entre dígitos ou conjuntos destes específicos.

Essa sequência de processos pode ser verificada na Figura 4.7, os quais são melhores explanados nas seções subseqüentes.

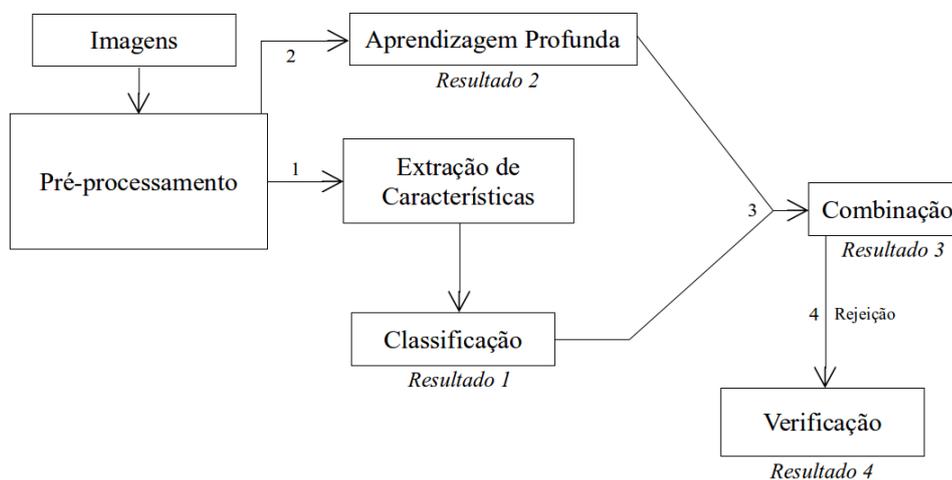


Figura 4.7: Processo proposto para a classificação automática de símbolos de configuração de mão do SignWriting

4.3.1 Taxas de Reconhecimento e Matrizes de Confusão

Ao se trabalhar com um sistema de classificação de objetos, treina-se um modelo com um conjunto específico de dados. Esse modelo deve ser validado, em uma base diferente da de treinamento, com o intuito de verificar seu desempenho de classificação. Como resultado direto da execução, obtém-se uma taxa de reconhecimento, a qual informa o percentual de amostras corretamente classificadas por determinado classificador. O cálculo da taxa de reconhecimento é dado pela equação 4.1, onde B é a base de testes, N_B é o total de amostras desta base e N_{rec} é o total de amostras corretamente classificadas.

$$\text{Taxa de Reconhecimento} = \frac{N_{rec}}{N_B} \times 100 \quad (4.1)$$

Em muitos casos, a taxa de reconhecimento não é suficiente para representar precisamente o desempenho de um classificador, ou quais seus pontos fortes e fracos, portanto, cada resultado de classificação pode ser avaliado com mais atenção através da geração de

uma matriz de confusão. A matriz de confusão de um classificador é uma matriz $n \times n$ que mostra o desempenho de um classificador para cada uma das n classes existentes. Os elementos $A(i, j)$ da matriz indicam a quantidade de exemplares de teste da classe i que foram classificados como sendo da classe j . Considerando essa informação, sabe-se que na diagonal principal da matriz (onde $i = j$), encontram-se os acertos do classificador testado, enquanto o que estiver nos demais elementos representa uma confusão durante a classificação, envolvendo as classes i e j . A Figura 4.8 apresenta um exemplo de matriz de confusão para um problema de 3 classes testado em uma base de 30 exemplares, com rótulos distribuídos igualmente entre elas.

		CLASSE PREDITA		
		CLASSE 1	CLASSE 2	CLASSE 3
CLASSE REAL	CLASSE 1	9	1	0
	CLASSE 2	2	8	0
	CLASSE 3	4	1	5

Figura 4.8: Exemplo de matriz de confusão

4.3.2 Avaliação de Classificadores e Características

Visando a possibilidade de se obter melhores resultados e considerando que a quantidade total de atributos de características é relativamente baixa (186), as três características propostas não são avaliadas somente de forma isolada. Dessa forma, são criados sete subconjuntos da matriz de características, compostos pela reorganização e junção destas, de acordo com a seguinte regra: todas características isoladamente, em seguida usa-se as três combinações válidas de duas características diferentes e, por último, assume-se a união das três como uma única matriz (idêntica à matriz original), conforme pode ser observado na Tabela 4.2.

A função de kernel utilizada para a classificação com SVM é a RBF (Radial Basis Function), pois além de ser indicada para experimentos iniciais, esta é utilizada para

Tabela 4.2: Subconjuntos de características utilizados na classificação

Característica	Classificador						
	1	2	3	4	5	6	7
Hist. de Projeção	•			•	•		•
Hist. de Orientação		•		•		•	•
Hist. de Concavidade			•		•	•	•

classificação em outros trabalhos (Li et al., 2009; Chen et al., 2010), tendo obtido elevada taxa de reconhecimento (Hsu et al., 2003).

Para obter-se um resultado satisfatório com o SVM é preciso um ajuste preciso dos parâmetros de entrada. Esses parâmetros são obtidos através de uma busca gulosa (grid-search) que permite encontrar a melhor combinação de parâmetros através de validação cruzada (Hsu et al., 2003).

A mesma base utilizada para treino é utilizada para teste, porém, visando encontrar o melhor resultado e evitar o overfitting, aplica-se a técnica de divisão conhecida como k-folding, seguida de validação cruzada, a qual consiste em dividir a base em um vários subconjuntos (nesse caso três), tomar uma como conjunto de testes e o restante como de treinamento para a realização da classificação e repetir esse processo até que todos conjuntos tenham sido utilizados como conjunto de testes, totalizando três iterações. Feito isto, os resultados obtidos são agrupados através do cálculo da média das taxas de reconhecimento, bem como é possível agrupar as confusões destas em uma única matriz de confusão (Hsu et al., 2003).

O diagrama apresentado na Figura 4.9 resume o processo ocorrido durante a execução do classificador SVM, proposto na Figura 4.7.

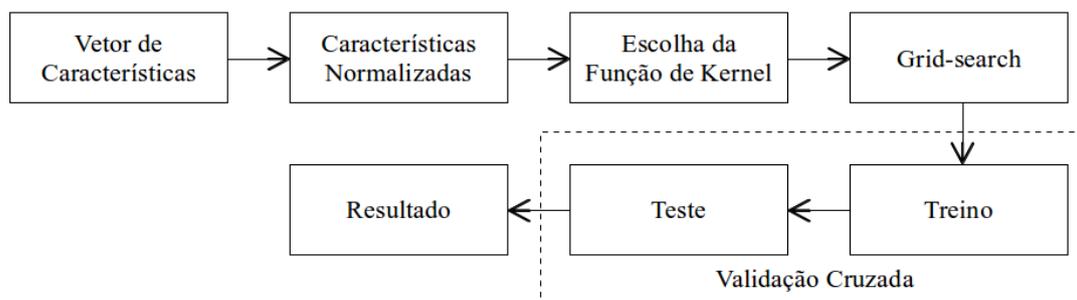


Figura 4.9: Classificação utilizando Support Vector Machine

Ao final dessa etapa, obtém-se os resultados da classificação para cada uma das sete características propostas, que podem ser individualmente visualizados através de suas taxas de reconhecimento. O classificador SVM está habilitado a gerar probabilidades a posteriori, as quais serão utilizadas no processo de combinação de classificadores, bem como na etapa de verificação.

4.3.3 Aprendizagem Profunda

A aplicação das técnicas de aprendizagem profunda apresentadas na Seção 2.2.2 deste trabalho se dá através da implementação de algoritmos utilizando a biblioteca `cuda-convnet`¹ e seu conjunto padrão de parâmetros, os quais foram executados em um computador com uma Graphics Processing Unit (GPU) Tesla C2050.

Além de fornecer resultado com taxa de reconhecimento e matriz de confusão, a biblioteca `cuda-convnet` fornece, com base nas imagens rotuladas dadas como amostra, probabilidades a posteriori, as quais são aproveitadas no processo de combinação de classificadores (definido na Seção 4.3.4).

Essa etapa não utiliza as características propostas na Seção 4.2 deste trabalho, pois o modelo da CNN é responsável por aprender as próprias características. O detector de características que o modelo aprendeu para a primeira camada deste experimento pode ser visualizado através dos pesos do mapa de características. A Figura 4.10 apresenta os 64 mapas de características aprendidos nesta primeira camada convolucional.

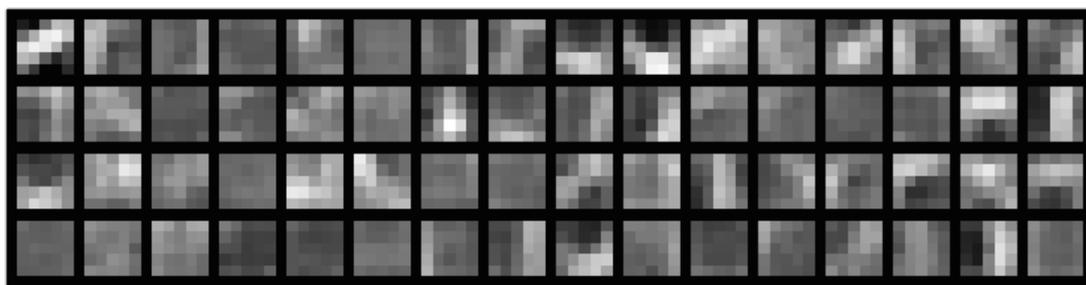


Figura 4.10: Mapas de características aprendidos pela primeira camada convolucional

Esta etapa utiliza o mesmo protocolo experimental proposto para a etapa classificação,

¹<http://code.google.com/p/cuda-convnet>

portanto, é utilizado o classificador SVM e kernel RBF, com busca de parâmetros via grid-search e validação cruzada através de k-folding. Como resultado deste processo, a matriz de probabilidades a posteriori apresenta a mesma estrutura das matrizes dos sete experimentos citados na Seção 4.3.2.

4.3.4 Combinação de Classificadores

Conforme informado nas seções 4.3.2 e 4.3.3, ambas as técnicas empregadas são capazes de gerar probabilidades a posteriori. A combinação dessas probabilidades ocorre através das regras de fusão apresentadas na Seção 2.2.3.

A combinação ocorre de maneira que seja possível testar e comparar os resultados da maior quantidade possível de combinações entre diferentes classificadores, agrupando os mesmos desde em pares até todos os oito em um único grupo. Desta forma obtêm-se a quantidade de combinações definida pela equação 4.2, onde n é a quantidade de classificadores (neste caso, oito) e r é a quantidade de classificadores combinados em cada processo de combinação, que pode variar de dois (não há como combinar somente um classificador) à quantidade de classificadores definida por n .

$$\sum_{r=2}^n \frac{n!}{r! \cdot (n-r)!} \quad (4.2)$$

Considerando o resultado dos oito classificadores testados ($n = 8$), são realizadas, no total, 247 combinações entre classificadores. Como resultado final desse processo de combinação, obtêm-se taxa de reconhecimento e matriz de confusão para cada uma destas. Porém apenas os melhores resultados dentre estes são considerados e comparados neste trabalho.

4.3.5 Verificação

Para realizar o processo de verificação de símbolos, verifica-se a matriz de confusão do melhor resultado obtido (maior taxa de reconhecimento) e extrai-se dela as confusões mais relevantes (que ocorreram em maior quantidade, representando maior volume dentre o

total de erros). Para este trabalho, as verificações foram realizadas sobre:

- Cada um dos classificadores e características isolados;
- Aprendizagem profunda;
- Melhor resultado dentre as combinações de classificadores;
- Melhor resultado para cada uma das regras de combinação, considerando apenas os classificadores não pré-combinados.

Para cada confusão específica, detecta-se possíveis características estruturais que possam diferenciar exclusivamente os símbolos envolvidos nesta.

As novas características derivadas da análise de cada confusão são utilizadas para treinar um novo classificador (SVM com RBF e grid-search), o qual deve ser especialista no reconhecimento entre os símbolos das classes envolvidas nessa confusão. Todos os classificadores especialistas (verificadores) são treinados no começo do processo de classificação, contendo apenas todas as amostras de símbolos das classes envolvidas na confusão, permitindo, assim, que verificações das confusões tratadas por eles possam ser realizadas em qualquer momento seguinte.

Se, durante o processo de classificação, determinado classificador predizer uma amostra como sendo de alguma classe passível de confusão, esse resultado é descartado e um novo processo de classificação ocorre apenas para essa amostra, utilizando o classificador especialista para essa confusão que foi previamente treinado. Os testes são realizados separadamente para cada amostra, com o objetivo de provar a predição em questão. Isso resulta em uma nova predição, que pode confirmar, ou não, o resultado anteriormente fornecido por outro processo de classificação.

Um cuidado que deve-se tomar neste processo é que tanto o classificador original, quanto o classificador de verificação podem tanto acertar, quanto errar durante a predição da classe de determinada amostra, podendo, os resultados de verificação, ser classificados em quatro categorias:

- Confirmado: os dois classificadores acertaram a resposta;

- Corrigido: o classificador genérico errou, porém o resultado foi corrigido pelo classificador especialista;
- Piorado: o classificador genérico respondeu corretamente, mas o especialista acabou errando;
- Igualmente errado: os dois classificadores erraram na predição.

A verificação da melhoria pela implantação de um verificador de confusões é feita através da avaliação do acréscimo de taxas de reconhecimento (em pontos percentuais), bem como pela análise da quantidade de amostras de cada classe enquadradas como corrigidas após o processo de verificação.

As seções subsequentes apresentam as confusões detectadas, e suas respectivas técnicas de verificação, que foram identificadas após o processo de classificação, bem como maiores informações a respeito da definição das classes a serem verificadas nesta etapa.

Confusão entre os símbolos 84 e 86

Os elementos 84 e 86 são dois símbolos de configuração de mão que se distinguem apenas pela presença, ou ausência, de uma curvatura na linha que representa o dedo polegar, conforme pode ser observado na Figura 4.11. Na figura, os itens (a) e (b) apresentam os símbolos originais, conforme a base de símbolos, enquanto os símbolos (c) e (d) apresentam a transcrição manuscrita dos mesmos, já com as devidas normalizações aplicadas.

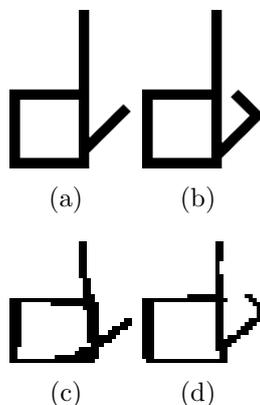


Figura 4.11: Símbolos de configuração de mão: (a) 84 original, (b) 86 original, (c) 84 manuscrito e (d) 86 manuscrito.

Para diferenciar os dois símbolos, uma região de interesse é definida. Considerando as imagens dos símbolos normalizadas e redimensionadas em 30x30 pixels, extrai-se o histograma de concavidade da área entre o pixel 10 e o 29 no eixo da altura e 23 e 29 na largura, com o objetivo de focar apenas na área que apresenta diferenças relevantes para a confusão em questão. O histograma de concavidade não é utilizado por completo, sendo aproveitadas, para esta confusão, apenas as posições 4, 8 e 10 de seu vetor, pois as demais não apresentam relevância suficiente.

Os três atributos de característica são agrupados em uma matriz de características que compreende todas as amostras que forem preditas pelo primeiro classificador como sendo da classe 84 ou da classe 86. Um classificador de duas classes é, então, treinado, com o objetivo específico de tratar as diferenças entre elas.

Confusão entre os símbolos 26 e 34

Os elementos 26 e 34 são dois símbolos de configuração de mão que se distinguem apenas pela presença, ou ausência, de uma linha horizontal sobre o corpo da figura, que representa o dedo polegar sobrando sobre a palma da mão, conforme pode ser observado na Figura 4.12. Na figura, os itens (a) e (b) apresentam os símbolos originais, conforme a base de símbolos, enquanto os símbolos (c) e (d) apresentam a transcrição manuscrita dos mesmos, já com as devidas normalizações aplicadas.

Para diferenciar os dois símbolos, utiliza-se a mesma forma de zoneamento (3x2) especificada na Seção 4.2.4, porém considera-se apenas a zona central direita. Desta zona, extrai-se os histogramas de projeção dos pixels horizontal e vertical. Como a extração é feita sobre uma área de interesse reduzida (10×15 pixels), os histogramas resultantes terão quantidade de atributos equivalentes, sendo 10 valores para o histograma horizontal e 15 para o histograma vertical, totalizando 25 valores.

Esses 25 atributos de característica são agrupados em uma matriz de características que compreende todas as amostras que forem preditas pelo primeiro classificador como sendo da classe 26 ou da classe 34. Um classificador de duas classes é, então, treinado, com o objetivo específico de tratar as diferenças entre elas.

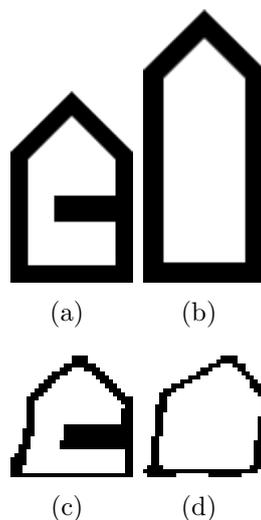


Figura 4.12: Símbolos de configuração de mão: (a) 26 original, (b) 34 original, (c) 26 manuscrito e (d) 34 manuscrito.

Confusão entre os símbolos 38 e 39

Os elementos 38 e 39 são dois símbolos de configuração de mão que se distinguem pela forma lateral da mão. Em diversas transcrições de usuários as diferenças acabam não sendo tão claras devido tanto ao processo de normalização, quanto à interpretações errôneas. A Figura 4.13 apresenta esses símbolos, na qual os itens (a) e (b) apresentam os símbolos originais, conforme a base de símbolos, enquanto os símbolos (c) e (d) apresentam a transcrição manuscrita dos mesmos, já com as devidas normalizações aplicadas.

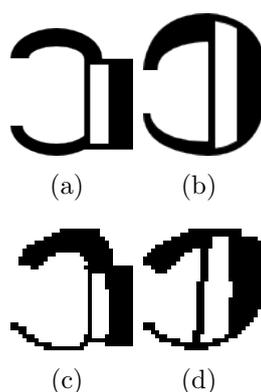


Figura 4.13: Símbolos de configuração de mão: (a) 38 original, (b) 39 original, (c) 38 manuscrito e (d) 39 manuscrito.

Para diferenciar os dois símbolos, utiliza-se uma junção de duas características. Pri-

meiramente, utilizando a mesma forma de zoneamento (3x2) especificada na Seção 4.2.4, é extraído o histograma de orientação dos contornos das zonas superior e central da direita da imagem. Em seguida, é extraído o histograma de projeção dos pixels horizontal da metade superior da imagem. A primeira técnica resulta em 16 atributos, enquanto a segunda resulta em 15, totalizando 31 atributos.

Esses 31 atributos de característica são agrupados em uma matriz de características que compreende todas as amostras que forem preditas pelo primeiro classificador como sendo da classe 38 ou da classe 39. Um classificador de duas classes é, então, treinado, com o objetivo específico de tratar as diferenças entre elas.

Confusão entre os símbolos 52 e 53

Os elementos 52 e 53 são dois símbolos de configuração de mão que se distinguem apenas pela formato da mão, que é ora representado por um quadrado perfeito e ora por um círculo perfeito, conforme pode ser observado na Figura 4.14. Na figura, os itens (a) e (b) apresentam os símbolos originais, conforme a base de símbolos, enquanto os símbolos (c) e (d) apresentam a transcrição manuscrita dos mesmos, já com as devidas normalizações aplicadas.

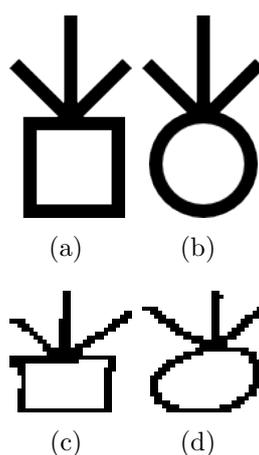


Figura 4.14: Símbolos de configuração de mão: (a) 52 original, (b) 53 original, (c) 52 manuscrito e (d) 53 manuscrito.

Para diferenciar os dois símbolos, utiliza-se uma junção de duas características. Primeiramente, uma região de interesse é definida e, considerando as imagens dos símbolos

normalizadas e redimensionadas em 30x30 pixels, extrai-se o histograma de concavidade da área entre o pixel 10 e o 29 no eixo da altura e 23 e 29 na largura, no mesmo formato da resolução da confusão entre as classes 84 e 86, com o objetivo de focar apenas na área que apresenta diferenças relevantes para a confusão em questão. Da mesma forma que ocorre na outra confusão, o histograma de concavidade não é utilizado por completo, sendo aproveitadas, para esta confusão, apenas as posições 4, 8 e 10 de seu vetor, pois as demais não apresentam relevância suficiente. Em seguida, é extraído o histograma de orientação dos contornos das zonas inferior e central da direita da imagem. A primeira técnica resulta em 3 atributos, enquanto a segunda resulta em 16, totalizando 19 atributos.

Esses 19 atributos de característica são agrupados em uma matriz de características que compreende todas as amostras que forem preditas pelo primeiro classificador como sendo da classe 52 ou da classe 53. Um classificador de duas classes é, então, treinado, com o objetivo específico de tratar as diferenças entre elas.

Outras Confusões

Há outras confusões relevantes resultantes dos processos de classificação, sendo algumas delas até mais relevantes (devido ao número de ocorrências) que as mencionadas nos itens anteriores. Porém as mesma não foram tratadas devido à dificuldade em identificar eventuais diferenças entre os símbolos das mesmas, sendo que, em muitos casos, o processo de transcrição do símbolo pelo usuário gera imagens impassíveis de comparação devido a semelhanças entre elas.

As Figuras 4.15, 4.16, 4.17 e 4.18 abaixo, apresentam essas confusões. Todas as confusões foram descobertas após o processo de classificação utilizando classificadores e características (Seção 4.3.2), sendo as classes aqui demonstradas, extraídas dos dados de geração da Tabela 5.3.

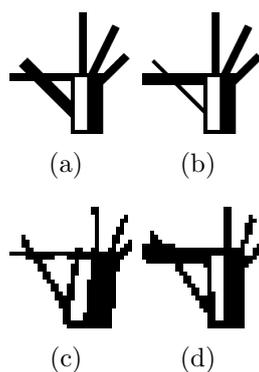


Figura 4.15: Símbolos de configuração de mão: (a) 79 original, (b) 80 original, (c) 79 manuscrito e (d) 80 manuscrito.

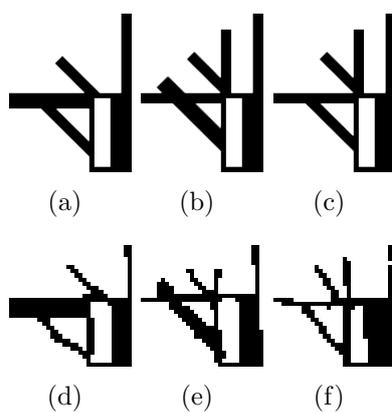


Figura 4.16: Símbolos de configuração de mão: (a) 64 original, (b) 71 original, (c) 72 original, (d) 64 manuscrito, (e) 71 manuscrito e (f) 72 manuscrito.

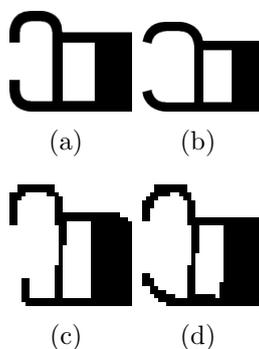


Figura 4.17: Símbolos de configuração de mão: (a) 90 original, (b) 91 original, (c) 90 manuscrito e (d) 91 manuscrito.

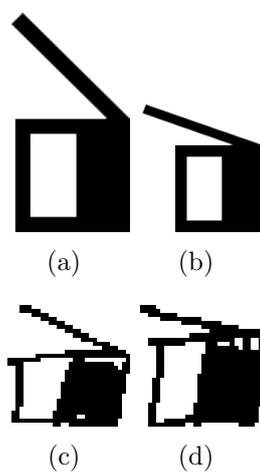


Figura 4.18: Símbolos de configuração de mão: (a) 6 original, (b) 7 original, (c) 6 manuscrito e (d) 7 manuscrito.

CAPÍTULO 5

AVALIAÇÃO DO MÉTODO PROPOSTO

Esta seção descreve a metodologia de avaliação do método proposto, bem como os resultados atingidos com a execução de seus experimentos.

Os resultados reportados pela etapa de classificadores e características são inicialmente baseados na escolha do melhor resultado de classificação (top 1), sendo que as confusões encontradas são identificadas em matrizes de confusão adequadas. Após análise destas, são aplicadas as demais estratégias, que visam melhorar o desempenho dos classificadores, sendo considerados os resultados de: top 2 (segundo resultado de classificação), aprendizagem profunda, fusão de classificadores e verificação.

5.1 Avaliação de Classificadores e Características

Durante a etapa de avaliação de classificadores e características, sete classificadores SVM foram treinados e testados com subconjuntos de três características estruturais dos símbolos de configuração de mão de SignWriting, com a utilização de grid-search para encontrar os parâmetros com melhor performance, conforme descrito na Seção 4.3.2 sobre o método proposto. Os resultados dessa etapa estão descritos na Tabela 5.1, onde $h_{projecao}$, $h_{orientacao}$ e $h_{concavidade}$ representam, respectivamente, os histogramas de projeção, orientação e concavidade, extraídos de todas as imagens, no mesmo modelo fornecido pela Tabela 4.2. Na tabela em questão, nota-se que os itens denotados pelos índices d , e , f e g não apresentam vetores de características puros e próprios e sim, concatenações dos vetores previamente apresentados pelos itens dos índices a , b e c .

A análise da Tabela 5.1 permite a avaliação comparativa dos resultados dos diversos processos de classificação ocorridos. O melhor resultado reportado com relação à taxa de reconhecimento é referente ao histograma de concavidade ($h_{concavidade}$), tanto em top 1 (91,69%), quanto em top 2 (97,44%). Ainda observando a tabela, pode-se notar

Tabela 5.1: Resumo dos resultados dos experimentos com características e classificadores

Índice	Característica	Taxa de Reconhecimento (%)				Tam. Vetor
		TOP 1	Desvio	TOP 2	Desvio	
(a)	$h_{projecao}$	81,05	1,87	91,09	0,55	60
(b)	$h_{orientacao}$	75,19	2,04	87,84	0,52	48
(c)	$h_{concavidade}$	91,69	0,68	97,44	0,08	78
(d)	$h_{projecao}, h_{orientacao}$	86,35	1,27	94,57	0,23	108
(e)	$h_{projecao}, h_{concavidade}$	91,34	0,85	97,35	0,08	138
(f)	$h_{orientacao}, h_{concavidade}$	91,56	0,70	97,41	0,06	126
(g)	$h_{projecao}, h_{orientacao}, h_{concavidade}$	91,56	0,89	97,42	0,08	186

que o resultado para essa característica foi melhor, inclusive, que todos os resultados de classificações envolvendo vetores de características concatenados. Os resultados desses processos de classificação ficaram bastante próximos, porém ainda inferiores, aos percentuais apresentados para o histograma de concavidade, o que indica que as características referentes a ele apresentam dominância sobre as demais, havendo pouca (ou nenhuma) complementaridade entre as três (considerando esse esquema de concatenação de vetores de características). Se considerado esse processo de avaliação de classificadores e características como único na estratégia de classificação dos símbolos de configuração de mão do SignWriting, o não acréscimo na taxa de reconhecimento através dessa concatenação acaba por tornar seu trabalho de treinamento e teste desnecessário e inviável, em termos de custo computacional, conforme pode ser visualizado no comparativo de tempo de execução dos experimentos apresentados na Tabela 5.2. Para tal, os experimentos foram executados em um computador com processador Intel Xeon quad-core de 3.3GHz e 8GB de memória RAM.

Tabela 5.2: Tempo de execução dos experimentos

Índice	Característica	Tempo de Execução
a	$h_{projecao}$	43m58s
b	$h_{orientacao}$	36m31s
c	$h_{concavidade}$	39m25s
d	$h_{projecao}, h_{orientacao}$	59m22s
e	$h_{projecao}, h_{concavidade}$	1h8m1s
f	$h_{orientacao}, h_{concavidade}$	57m29s
g	$h_{projecao}, h_{orientacao}, h_{concavidade}$	1h24m49s

Adicionalmente, ainda através da análise da mesma Tabela 5.1, percebe-se que, inde-

pendentemente da característica avaliada, houve considerável aumento na taxa de reconhecimento quando aplicada a técnica de top 2, o que pode ser verificado, por exemplo, na taxa do melhor resultado ($h_{concauidade}$): 91,69% \rightarrow 97,44%. Esse resultado indica que boa parte das confusões ocorrem entre símbolos muito parecidos, que podem causar confusão, inclusive, para avaliadores humanos.

Para uma melhor visualização das confusões, a Figura 5.1 apresenta a matriz de confusão para o melhor processo de classificação isolado. As confusões em destaque são as mais relevantes para esse classificador, sendo todas consideradas de forma recíproca, ou seja, onde a confusão ocorre para os dois lados.

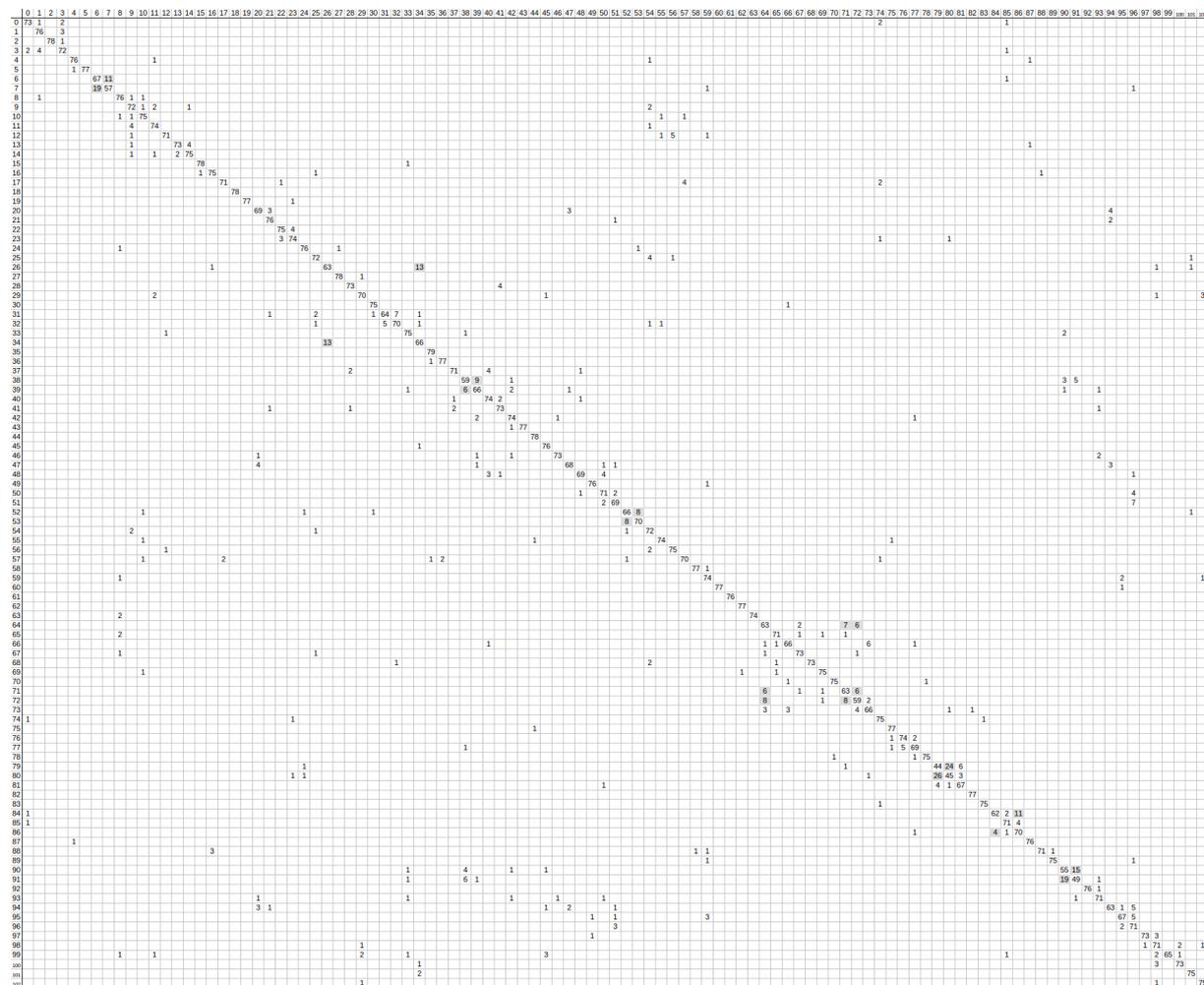


Figura 5.1: Matriz de confusão da classificação com histograma de concauidade

Devido à grande quantidade de classes, o conteúdo da matriz de confusão exposta pela Figura 5.1 é de difícil visualização. Portanto, a Tabela 5.3 totaliza os números das

confusões destacadas, bem como a Tabela 5.4 reinterpreta os dados da mesma figura, porém com foco específico nas confusões citadas.

Tabela 5.3: Sumarização das principais confusões com o histograma de concavidade

Símbolos		Confusões
79	80	50
64	71 72	41
90	91	34
6	7	30
26	34	26
84	86	15
38	39	15
52	53	16

Tabela 5.4: Matriz de confusão resumida da classificação com histograma de concavidade

	6	7	26	34	38	39	52	53	64	71	72	79	80	84	86	90	91
6	67	11															
7	19	57															
26			63	13													
34			13	66													
38					59	9										3	5
39					6	66										1	
52							66	8									
53							8	70									
64									63	7	6						
71									6	63	6						
72									8	8	59						
79										1		44	24				
80												26	45				
84														62	11		
86														4	70		
90					4											55	15
91					6	1										19	49

5.2 Aprendizagem Profunda

A execução do experimento de classificação dos símbolos de configuração de mão do Sign-Writing através da utilização de técnicas de aprendizagem profunda, que se deu através do treinamento de uma rede neural convolucional implementada com a biblioteca `cuda-convnet`, obteve os resultados apresentados na Tabela 5.5.

Tabela 5.5: Resultado do experimento utilizando aprendizagem profunda

Característica	Taxa de Reconhecimento (%)	
	TOP 1	TOP 2
<i>cnn</i>	90,84	96,55

O resultado final da execução desse experimento retornou uma taxa de reconhecimento maior que os histogramas de projeção e orientação dos pixels apresentados na Tabela 5.1, porém, o melhor resultado com relação à taxa de classificação continua sendo o histograma de concavidade, tanto se considerado apenas o melhor resultado (top1), quanto no top 2.

A Tabela 5.6 apresenta a matriz de confusão, resumida apenas às principais confusões, para o processo de classificação utilizando aprendizagem profunda. As mesmas confusões previamente destacadas e sumarizadas Seção 5.1 são mostradas nessa tabela. Com a análise destas, pode-se perceber que várias confusões que ocorrem no primeiro caso, são rigorosamente reduzidas (ou até zeradas) com a utilização da aprendizagem profunda. Porém, com uma breve análise nas demais confusões, percebe-se, de forma global, que diversas novas confusões foram criadas, o que explica o fato de sua taxa de reconhecimento final ter sido inferior, bem como justifica a utilização de técnicas de combinação de classificadores, pois os dados tendem a apresentar complementaridade entre si.

A execução de um experimento utilizando aprendizagem profunda com a biblioteca *cuda-convnet* demanda tempo, hardware especializado, elevado custo computacional, bem como pré-processamento específico nos dados de entrada. Portanto, se esse resultado for considerado apenas de forma isolada, acaba tornando-se inviável.

5.3 Combinação de Classificadores

A etapa de combinação de classificadores depende de recursos obtidos nas duas etapas anteriores: as matrizes de probabilidade de cada um dos classificadores que se deseja combinar.

A fim de testar a maior quantidade de combinações entre diferentes classificadores possível, os experimentos foram realizados utilizando as 247 combinações apresentadas na Seção 4.3.4. A Tabela 5.7 apresenta os resultados de combinação de classificadores de

Tabela 5.6: Matriz de confusão resumida da classificação com CNN

	6	7	26	34	38	39	52	53	64	71	72	79	80	84	86	90	91
6	63	15															
7	16	52															2
26			77	2													
34				78													
38					61	8										1	5
39		1			2	63										1	
52							76	1									
53							2	72									
64									74								
71										65	5		1				
72									2	6	65						
79										2		45	15				
80										2		16	50				
84														62	6		
86														3	66		
90					2											56	16
91					3											19	50

forma analítica e não agrupada, com relação à combinação e regra utilizadas, podendo um mesmo conjunto de combinação aparecer mais de uma vez, porém com regras distintas. Devido à elevada quantidade de resultados, apenas os 15 melhores são reportados.

Os dados na Tabela 5.7 estão dispostos de forma que seja possível identificar os classificadores envolvidos em cada processo de combinação. Os títulos denotados de a até g fazem referência, respectivamente, às probabilidades dos sete classificadores apresentados na Tabela 5.1, mais as probabilidades resultantes do processo de aprendizagem profunda apresentado na Tabela 5.5. Portanto, a , b e c representam os dados dos classificadores isolados, d , e , f e g a utilização de vetores de características concatenados e h a CNN. Os itens marcados com \bullet indicam a presença do respectivo classificador naquele conjunto de classificadores combinado apresentado na linha em questão. Seguindo esse mesmo formato, a Tabela 5.9 apresenta os resultados equivalentes, porém em top 2.

A Tabela 5.8 apresenta a matriz de confusão, resumida apenas às principais confusões, obtida da combinação de classificadores com melhor taxa de reconhecimento, apresentada na Tabela 5.7. A análise desta mostra a resolução uma diminuição de confusões, se comparada às matrizes de confusão apresentadas nas Tabelas 5.4 e 5.6.

Tabela 5.7: Resultado analítico da combinação de classificadores

Classificadores								TOP 1	
<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	Taxa	Regra
	•	•		•	•		•	94,65	Produto
	•	•			•	•	•	94,58	Produto
		•		•	•	•	•	94,50	Produto
	•	•		•	•	•	•	94,47	Produto
•	•	•		•	•		•	94,43	Produto
•	•	•			•		•	94,42	Produto
	•	•	•		•		•	94,40	Produto
•	•	•			•	•	•	94,38	Produto
•	•	•		•	•	•	•	94,38	Produto
	•	•			•		•	94,35	Produto
	•	•		•		•	•	94,35	Produto
	•	•	•	•	•		•	94,35	Produto
	•	•		•			•	94,33	Produto
		•			•	•	•	94,33	Produto
		•	•	•	•		•	94,32	Produto

Analisando essas duas tabelas, pode-se perceber que a única regra que apareceu dentre os 15 melhores resultados foi a do produto, portanto pode-se concluir que, para o problema de configurações de mão do SW, a melhor técnica de combinação de classificadores se dá utilizando essa regra. Também pode-se notar melhora considerável na taxa de reconhecimento em si, sendo elevada de 91,69% no melhor classificador isolado ($h_{concauidade}$) para 94,65% em top 1 e de 97,44% para 98,74% em top 2.

Também deve ser levada em consideração a representatividade desse classificador com melhor resultado anterior ($h_{concauidade}$, aqui representado por c) durante o processo de combinação de classificadores, sendo que o mesmo está presente em todos os 15 melhores resultados de combinação em top 1 e em 13 dos 15 resultados em top 2. O mesmo não ocorre para os outros dois histogramas, se considerados isoladamente. O histograma de projeção dos pixels ($h_{projecao}$, aqui representado por a) apresentava taxa de reconhecimento superior ao histograma de orientação dos pixels ($h_{orientacao}$, aqui representado por b), porém a representatividade do segundo histograma durante o processo de combinação de classificadores foi bem maior que a do primeiro, aparecendo em 12 das 15 combinações em top 1 e em 14 das 15 em top 2, contra um número de 2 e 5 do primeiro, respectivamente. Isto indica que as informações relativas à orientação dos pixels apresentam

Tabela 5.8: Matriz de confusão resumida da melhor combinação de classificadores

	6	7	26	34	38	39	52	53	64	71	72	79	80	84	86	90	91
6	67	12															
7	16	60															
26			77	2													
34				79													
38					64	6										1	6
39					2	73											
52							76	1									
53							2	75									
64									75	1							
71										65	9						
72									1	7	67		1				
79										1		43	22				
80										1		16	54				
84														71	3		
86														4	70		
90					3											58	16
91					3											17	54

complementaridade com as relacionadas à concavidade da imagem.

Outro ponto que deve ser avaliado para esses resultados de combinação é que as probabilidades geradas pelo experimento relacionado à aprendizagem profunda (representado por h na tabela) está presente entre todas os 15 melhores resultados de combinação, tanto em top 1, quanto em top 2. Como essa técnica isolada não havia atingido o melhor resultado, entende-se que seus dados são diretamente complementares aos do melhor classificador durante um processo de combinação. Isso ocorre devido ao fato de a CNN utilizar uma abordagem diferenciada, não diretamente estrutural, para a extração de características de cada símbolo.

Os resultados dos classificadores que utilizaram vetores de características concatenados apresentados nas tabelas, seguem a mesma regra apresentada para os resultados dos três histogramas testados de forma isolada. Os classificadores d e e , que derivam de a possuem menos representatividade nos resultados de combinação, enquanto f , que deriva de c (melhor resultado isolado) está presente em 13 dos 15 resultados em top 1 e em todos os resultados em top 2. Esses resultados são outro ponto indicativo de que a característica do histograma de projeção dos pixels não possui uma boa complementaridade quando

Tabela 5.9: Resultado analítico da combinação de classificadores (top 2)

Classificadores								TOP 2	
<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	Taxa	Regra
	•	•		•	•		•	98,74	Produto
	•	•			•	•	•	98,71	Produto
	•	•		•	•	•	•	98,71	Produto
		•		•	•	•	•	98,65	Produto
	•	•	•		•	•	•	98,65	Produto
	•	•	•		•		•	98,64	Produto
	•	•	•	•	•		•	98,64	Produto
	•			•	•	•	•	98,62	Produto
•	•	•			•	•	•	98,62	Produto
•	•	•		•	•	•	•	98,62	Produto
	•	•	•	•	•	•	•	98,62	Produto
	•	•			•		•	98,61	Produto
•	•			•	•	•	•	98,60	Produto
•	•	•			•		•	98,59	Produto
•	•	•	•		•	•	•	98,59	Produto

combinada, tendendo a piorar os resultados de outros classificadores. O classificador g também não apresentou boa performance devido à existência do mesmo histograma em sua composição.

Porém, a aparição dos resultados envolvendo vetores de características concatenados dentre os melhores resultados de combinação indica que os mesmos exercem alguma complementaridade aos dados de classificadores isolados, mesmo quando esses vetores são combinados em conjuntos com classificadores isolados que pertencem aos seus subconjuntos, como, por exemplo, o melhor resultado de combinação em top 1, que contém o $h_{orientacao}$ e $h_{concauidade}$, mas também contém e , que possui $h_{orientacao}$ em sua formação e f que possui tanto $h_{orientacao}$, quanto $h_{concauidade}$ em sua formação. Essa informação não pode ser descartada, pois os melhores resultados finais de combinação acabam por utilizar tanto o classificador puro, quanto os classificadores com vetores de características concatenados, o que indica que o processo de concatenação, apesar de mais custoso e repetitivo, acaba por gerar probabilidades com características de complementaridade com relação aos demais, não devendo ser ignorado.

Caso os resultados utilizando vetores de características concatenados sejam excluídos do processo de combinação de classificadores, as taxas de reconhecimento caem, sendo que

o melhores resultados são 93,91% e 98,39%, para top 1 e 2, respectivamente. O resultados completos, com as 15 melhores taxas, são apresentados na Tabela 5.10 e Tabela 5.11, para top 1 e 2, respectivamente. Nesses resultados, pode ser observada a continuidade da dominância de resultados envolvendo o histograma de concavidade, bem como a CNN. Por outro lado, as combinações com melhores resultados não foram somente as que utilizaram a regra do produto, havendo maior diversificação, estando listados nas tabelas: soma, produto, média, mínimo e máximo.

Tabela 5.10: Resultado analítico de combinação (sem vetores de características concatenados - top 1)

Classificadores				TOP 1	
<i>a</i>	<i>b</i>	<i>c</i>	<i>h</i>	Taxa	Regra
•	•	•	•	93,91	Produto
		•	•	93,80	Produto
			•	93,42	Mínimo
	•	•	•	93,26	Mínimo
•		•	•	93,12	Produto
		•	•	92,99	Produto
•		•	•	92,97	Mínimo
•	•	•	•	92,97	Soma
•	•	•	•	92,97	Média
•	•	•	•	92,92	Mínimo
•	•		•	92,71	Produto
	•	•	•	92,71	Soma
	•	•	•	92,71	Média
	•		•	92,57	Produto
	•		•	92,53	Mínimo

Outra informação importante que deve ser descrita nos resultados deste trabalho é a melhor taxa de reconhecimento reportada para cada uma das regras aplicadas. A Tabela 5.12 apresenta esse resultado, contendo apenas as combinações de classificadores sem dados contendo vetores de características concatenados. A análise da tabela permite identificar que, independentemente da regra aplicada, o melhor resultado de classificação sempre envolve a característica do histograma de concavidade e os resultados obtidos pela execução do algoritmo de aprendizagem profunda.

Tabela 5.11: Resultado analítico de combinação (sem vetores de características concatenados - top 2)

Classificadores				TOP 2	
<i>a</i>	<i>b</i>	<i>c</i>	<i>h</i>	Taxa	Regra
•	•	•	•	98,39	Produto
	•	•	•	98,32	Produto
•		•	•	98,12	Produto
•	•	•	•	98,10	Soma
•	•	•	•	98,10	Média
	•	•	•	97,90	Soma
	•	•	•	97,90	Média
		•	•	97,86	Produto
•	•	•		97,85	Produto
•	•		•	97,85	Produto
•		•	•	97,85	Soma
•		•	•	97,85	Média
		•	•	97,84	Soma
		•	•	97,84	Média
		•	•	97,67	Máximo

Tabela 5.12: Melhor resultado de combinação por regra

Regra	TOP 1					TOP 2				
	Taxa (%)	<i>a</i>	<i>b</i>	<i>c</i>	<i>f</i>	Taxa (%)	<i>a</i>	<i>b</i>	<i>c</i>	<i>f</i>
Produto	93,91	•	•	•	•	98,39	•	•	•	•
Soma	92,97	•	•	•	•	98,10	•	•	•	•
Média	92,97	•	•	•	•	98,10	•	•	•	•
Mínimo	93,42			•	•	97,54			•	•
Máximo	91,27		•	•	•	97,67		•	•	•
Voto	92,39	•	•	•	•	96,30			•	•

5.3.1 Verificação

O procedimento da etapa de verificação foi realizado sobre os 11 resultados definidos na Seção 4.3.5. Como pode ser observado na Tabela 5.13, apesar de as confusões mais relevantes não terem sido verificadas devido às dificuldades mencionadas na citada seção, todos os resultados tiveram diferenças positivas na taxa de reconhecimento. O aumento mais considerável foi na classificação usando a característica do histograma de orientação (b), porém esse resultado se deve ao seu baixo desempenho inicial. O histograma de concavidade, que contém o resultado de classificação isolado com melhor taxa de reconhecimento, apresenta elevação da taxa de reconhecimento de 0.75 pontos percentuais, o que se deve ao fato de as confusões terem sido escolhidas com base na matriz de confusão

do mesmo. Esse resultado não se repete com a mesma força nos demais resultados, inclusive no da aprendizagem profunda e melhor combinação de classificadores, devido ao fato de que os classificadores podem não apresentar as confusões previstas, bem como, no caso das combinações, as mesmas podem ter sido resolvidas (ou drasticamente reduzidas). O fato de a taxa não incrementar de forma tão significativa nos resultados de combinação, também se deve ao fato de as taxas de reconhecimento terem sido mais elevadas, havendo menos confusões envolvidas no resultado final.

A Tabela 5.14 apresenta as oito principais confusões recíprocas levantadas pelas Tabela 5.3 e suas respectivas quantidades para cada um dos três processo de classificação. Nela é possível notar que as quatro confusões de maior relevância, para todos os resultados, não apresentam variação com relação à quantidade de amostras corrigidas. Isso se deve ao fato de que as mesmas não puderam ser inseridas no processo de verificação devido aos motivos apontados no item “Outras Confusões” da Seção 4.3.5. Para realizar essas verificações, um estudo mais específico deve ser feito sobre as características de cada uma das classes, chegando a uma característica capaz de representar esses dados, porém deve-se tomar cuidado para não acontecer um overfitting nesse processo. Por ora, a técnica que pode ser aplicada para resolver essas confusões, é a utilização do top 2, pois a grande quantidade de confusões está ligada à elevada semelhança entre os dígitos, o que tende a classificar o símbolo correto como segunda opção.

Tabela 5.13: Resultado da verificação de classificadores

Classificadores								Regra	Taxa de Reconhecimento (%)		Diferença Pts. Perc.
<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>		Original	Verificada	
•								-	81,05	81,27	+0,22
	•							-	75,19	76,22	+1,03
		•						-	91,69	92,44	+0,75
							•	-	90,84	90,91	+0,07
	•	•		•	•		•	Produto	94,65	94,75	+0,10
•	•	•					•	Produto	93,91	94,00	+0,09
•	•	•					•	Soma	92,97	93,04	+0,07
•	•	•					•	Média	92,97	93,04	+0,07
							•	Mínimo	93,42	93,58	+0,16
	•	•					•	Máximo	91,27	91,31	+0,04
•	•	•					•	Voto	92,39	92,61	+0,22

Tabela 5.14: Comparativo entre confusões em classificadores antes (Conf.) de depois (Corr.) da verificação

Símbolos	Histograma		CNN		Melhor Combinação	
	Conf.	Corr.	Conf.	Corr.	Conf.	Corr.
79 80	50	50	31	31	38	38
64 71 72	41	41	41	41	18	18
90 91	34	34	26	26	33	33
6 7	30	30	31	31	28	28
26 34	26	8	2	4	2	7
84 86	15	4	9	5	7	4
38 39	15	3	10	4	8	1
52 53	16	0	3	0	3	0

CAPÍTULO 6

CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentado um método para realização da identificação de símbolos de configuração de mão do SignWriting, uma sistema de escrita para representação de línguas de sinais para surdos, baseado em técnicas de reconhecimento de padrões.

Foi estabelecida uma base de dados com 7994 imagens, sendo distribuídas entre 103 símbolos de configuração de mão, selecionados de forma que representem as diferentes características estruturais existentes na notação.

Foram elencadas e extraídas três características estruturais de cada uma das imagens, sendo essas: histograma de projeção dos pixels vertical e horizontal, histograma de orientação e histograma de concavidade.

Foi estabelecida uma estratégia de classificação, onde, primeiramente, foi realizada a classificação envolvendo os vetores de características isolados e concatenados, utilizando um classificador do tipo SVM. Paralelamente, foram utilizadas técnicas de aprendizagem profunda sobre o mesmo conjunto de dados. Em momento seguinte, houve a combinação de resultados de classificadores, utilizando as saídas de probabilidades geradas pelas duas etapas anteriores. Sobre os melhores resultados obtidos, foi aplicada a técnica de verificação, com o objetivo de melhorar o resultado com base no tratamento de confusões específicas, as quais foram detectadas no momento da primeira classificação.

Como resultado, isoladamente, a melhor taxa de reconhecimento obtida foi com a característica relacionada ao histograma de concavidade, atingindo 91,69% em top 1 e 97,44% em top 2. Os experimentos executados com aprendizagem profunda apresentaram taxas de reconhecimento de 90,84% e 96,55%, em top 1 e 2, respectivamente. Todos os melhores resultados de combinação foram obtidos utilizando a regra do produto, sendo que a melhor taxa de reconhecimento obtida nesta etapa foi de 94,65% em top 1 e 98,74% em top 2.

Após processo de verificação, aplicado sobre confusões detectadas na primeira etapa, foi detectado sensível aumento nas taxas de reconhecimento, sendo este de 0,07 pontos percentuais para o melhor resultado de classificação isolado e de 0,10 pontos percentuais para o melhor resultado de combinação.

Trabalhos futuros relacionados à área de SignWriting e reconhecimento de padrões podem envolver a detecção de características e técnicas relevantes para o tratamento das confusões que não puderam ser resolvidas neste trabalho. Outros trabalhos podem estar relacionados ao reconhecimento de diferentes estruturas da notação, como símbolos de contato, expressões faciais e movimento, bem como a identificação de termos envolvendo conjunto dessas estruturas e um dicionário pré-definido de palavras em SW, onde podem ser utilizadas como auxílio à predição, por exemplo, informações relacionadas ao contexto. A junção de técnicas e estratégias propostas por tais trabalhos pode servir de base para outro, que consiste na criação de uma ferramenta de software para usuários finais, que auxilia a tomada de decisão durante a escrita de textos utilizando o SignWriting (editor de textos).

REFERÊNCIAS BIBLIOGRÁFICAS

- Ahmed, A. S. and Seong, D. S. K. (2006). Signwriting on mobile phones for the deaf. In *Proceedings of the 3rd international conference on Mobile technology, applications & systems*, page 28. ACM.
- Almeida, P. (2013). Combinação de características texturais para a classificação automática de vagas de estacionamento. Master's thesis, Universidade Federal do Paraná (UFPR) - Brasil.
- Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends in Machine Learning*, 2(1):1–127.
- Bianchini, C. S., Borgia, F., Bottoni, P., and De Marsico, M. (2012a). Swift: A signwriting improved fast transcriber. In *Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI '12*, pages 390–393, New York, NY, USA. ACM.
- Bianchini, C. S., Borgia, F., and De Marsico, M. (2012b). Swift-a signwriting editor to bridge between deaf world and e-learning. In *12th International Conference on Advanced Learning Technologies*, pages 526–530. IEEE.
- Bouزيد, Y. and Jemni, M. (2013). An avatar based approach for automatically interpreting a sign language notation. In *13th International Conference on Advanced Learning Technologies*, pages 92–94. IEEE.
- Chen, J., Cao, H., Prasad, R., Bhardwaj, A., and Natarajan, P. (2010). Gabor features for offline arabic handwriting recognition. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, DAS '10*, pages 53–58, New York, NY, USA. ACM.
- Ciresan, D., Meier, U., Masci, J., and Schmidhuber, J. (2011). A committee of neural networks for traffic sign classification. In *The 2011 International Joint Conference on Neural Networks*, pages 1918–1921. IEEE.

- Costa, A. C. d. R. and Dimuro, G. P. (2001a). A signwriting-based approach to sign language processing. In *Proceedings from GW 2001: Gesture Workshop*, pages 202–212. Citeseer.
- Costa, A. C. d. R. and Dimuro, G. P. (2001b). Supporting deaf sign languages in written form on the web. In *10th World Wide Web Conference, Hong Kong. Processing, Written Form on the Web*.
- Costa, A. C. d. R., Dimuro, G. P., and Freitas, J. (2004). A sign matching technique to support searches in sign language texts. In *Proceedings from LREC*, volume 2004, page 4th.
- Costa, Y. M. (2013). *Reconhecimento de gêneros musicais utilizando espectrogramas com combinação de classificadores*. PhD thesis, Universidade Federal do Paraná.
- Feng, S. and Manmatha, R. (2005). Classification models for historical manuscript recognition. *The Selected Works of R. Manmatha*.
- Fernandes, E. (2003). *Linguagem e surdez*. Editora Artmed, Porto Alegre.
- Guimaraes, C., Guardazi, J. F., Oliveira, L. E., and Fernandes, S. (2014). Deaf culture and sign language writing system - a database for a new approach to writing system recognition technology. *47th Hawaii International Conference on System Science*, pages 3368–3377.
- Hinton, G., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554.
- Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al. (2003). A practical guide to support vector classification. Department of Computer Science and Information Engineering, National Taiwan University.
- Impedovo, D. and Pirlo, G. (2014). Zoning methods for handwritten character recognition: A survey. *Pattern Recognition*, 47(3):969 – 981.

- Jain, A., Duin, R. P. W., and Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37.
- Kittler, J., Hatef, M., Duin, R. P. W., and Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Li, J., Zhang, C., and Li, Z. (2009). Battlefield target identification based on improved grid-search svm classifier. In *International Conference on Computational Intelligence and Software Engineering*, pages 1–4.
- Lu, G., Shark, L.-K., Hall, G., and Zeshan, U. (2010). Hand motion recognition and visualisation for direct sign writing. In *14th International Conference Information Visualisation*, pages 467–472. IEEE.
- Martin, J. (2000). A linguistic comparison two notation systems for signed languages: Stokoe notation & sutton signwriting. *Unpublished manuscript, Western Washington University*.
- Michaelis (2012). Michaelis, dicionário de português online. Disponível em <http://michaelis.uol.com.br/moderno/portugues>. Acessado em 18/02/2014.
- Moemedi, K. and Connan, J. (2010). Rendering an animated avatar from signwriting notation. *Unpublished manuscript, University of the Western Cape*.
- Moemedi, K. A. (2010). *Rendering an Avatar from Sign Writing Notation for Sign Language Animation*. PhD thesis, University of the Western Cape.

- Oliveira, L., Sabourin, R., Bortolozzi, F., and Suen, C. (2002). Automatic recognition of handwritten numerical strings: a recognition and verification strategy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(11):1438–1454.
- Oliveira, L. S., Sabourin, R., Bortolozzi, F., and Suen, C. Y. (2003). Impacts of verification on a numeral string recognition system. *Pattern Recognition Letters*, 24(7):1023–1031.
- Papadogiorgaki, M., Grammalidis, N., Makris, L., and Strintzis, M. G. (2006). Gesture synthesis from sign language notation using mpeg-4 humanoid animation parameters and inverse kinematics. In *2nd IET International Conference on Intelligent Environments*, volume 1, pages 151–160. IET.
- Papadogiorgaki, M., Grammalidis, N., Sarris, N., and Strintzis, M. G. (2004). Synthesis of virtual reality animations from swml using mpeg-4 body animation parameters. In *Workshop on the Representation and Processing of Sign Languages, 4th International Conference on Language Resources and Evaluation LREC 2004*. Citeseer.
- Pereira, M. C. P. and Fronza, C. d. A. (2006). Sistema signwriting como uma possibilidade na alfabetização de pessoas surdas. *Encontro do Círculo Lingüístico do Sul (CELSUL). Pelotas: UCEPEL e UFPEL*.
- Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, pages 61–74. MIT Press.
- Rajashekararadhya, S. V. and Ranjan, P. (2009). Support vector machine based handwritten numeral recognition of kannada script. In *Advance Computing Conference, 2009. IACC 2009. IEEE International*, pages 381–386.
- Ramos, C. R. (2006). Libras: a língua de sinais dos surdos brasileiros. *Petrópolis-RJ, Editora Arara Azul*.
- Salvadeo, D. H. P. (2009). *Combinação de múltiplos classificadores para reconhecimento*

- de face humana*. PhD thesis, Universidade Federal de São Carlos. Centro de Ciências Exatas e de Tecnologia. Programa de pós-graduação em Ciência da Computação.
- SignWriting (2004). How to use sw-edit. Disponível em <http://www.signwriting.org/archive/docs2/sw0166-SW-Edit-Manual.pdf>. Acessado em 27/02/2014.
- Silva, R. C. d. (2012). Signwriting: um sistema de escrita das línguas gestuais. aplicação à língua gestual portuguesa. *Exedra: Revista Científica*, (6):387–399.
- Stumpf, M. R. (2002). *Lições sobre o SignWriting. Tradução Parcial e Adaptação do Inglês/ASL para Português LIBRAS do livro 'Lessons in SignWriting'*. Publicado originalmente pelo DAC, Deaf Action Committe for SignWriting.
- Stumpf, M. R. (2005). Aprendizagem de escrita de língua de sinais pelo sistema signwriting: língua de sinais no papel e no computador.
- Surowiecki, J. (2005). *The wisdom of crowds*. Random House LLC.
- Sutton, V. (2002). *Lessons in Sign Writing: Textbook & Workbook*. Deaf Action Committee for SignWriting, third edition edition.
- Theodoridis, S. and Koutroumbas, K. (2008). Pattern recognition. *IEEE Transactions on Neural Networks*, 19(2):376.
- Torchelsen, R. P., Costa, A. C. d. R., and Dimuro, G. (2002). Editor para textos em língua de sinais escritos em signwriting. *XXVIII Latin-American Conference on Informatics (CLEI 2002)*, pages p. 1–12. Uruguay.
- Trier, Ø. D., Jain, A. K., and Taxt, T. (1996). Feature extraction methods for character recognition-a survey. *Pattern Recognition*, 29(4):641 – 662.
- World Health Organization (2015). World health organization: Deafness and hearing loss. Disponível em <http://www.who.int/mediacentre/factsheets/fs300/en/index.html>. Acessado em 20/04/2015.

DIEGO STIEHL

**RECONHECIMENTO DE SÍMBOLOS DE CONFIGURAÇÃO
DE MÃO DO SIGNWRITING**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre. Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Luiz Eduardo Soares de Oliveira

CURITIBA

2015