



Pós-Graduação em Ciência da Computação

“Verificação de assinaturas com *One Class Classifier*”

Por

***Renatha do Amaral Accioly
Alves Negreiros***

Dissertação de Mestrado



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE, AGOSTO/2014



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Renatha do Amaral Accioly Alves Negreiros

Verificação de assinaturas com *One Class Classifier*

ESTE TRABALHO FOI APRESENTADO À PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DO CENTRO DE INFORMÁTICA DA UNIVERSIDADE FEDERAL DE PERNAMBUCO COMO REQUISITO PARCIAL PARA OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIA DA COMPUTAÇÃO.

ORIENTADOR: PROF. DR. CLÉBER ZANCHETTIN

RECIFE, AGOSTO/2014

Dissertação de Mestrado apresentada por **Renatha do Amaral Accioly Alves Negreiros** à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título “**Verificação de assinaturas com One Class Classifier**” orientada pelo **Prof. Dr. Cléber Zanchettin** e aprovada pela Banca Examinadora formada pelos professores:

Prof. Leandro Maciel Almeida
Centro de Informática / UFPE

Prof. Byron Leite Bezerra
Engenharia da Computação / UPE

Prof. Cléber Zanchettin
Centro de Informática / UFPE

Visto e permitida a impressão.
Recife, 28 de agosto de 2014

Prof. Nelson Souto Rosa
Coordenador da Pós-Graduação em Ciência da Computação do
Centro de Informática da Universidade Federa de Pernambuco.

Aos meus pais, esposo, irmão e avó, pelo apoio, incentivo e força constantes.

Agradecimentos

Por muitas provações e desafios passei, momentos bons e ruins ocorreram, contudo muitas lições e frutos surgiram destas vivências. A partir de familiares e amigos obtive forças para seguir com os estudos do mestrado. Agradeço primeiramente aos meus pais, Ricardo Accioly e Iris do Amaral, contribuintes diretos de toda a vida, ensinando os bons valores, princípios e caminhos cominante neste trabalho.

Ao meu namorado, noivo e esposo, Adriano Negreiros, pelo apoio emocional, incentivo, paciência, e acima de tudo amor, estando ao meu lado nas boas e más horas, cortando imagens, estudando sobre a área, prospectando esta produção.

Ao meu irmão, Ricardo Accioly Filho, pela grande companhia e incentivo, por compartilhar a mesma garra pelo setor acadêmico e contribuir com conhecimentos do setor jurídico.

A minha avó Maria da Soledade do Amaral pela crença acima de tudo no meu potencial, a dedicação a minha área de estudo, e pelo orgulho como neta.

Ao meu orientador, Cléber Zanchettin, pela paciência, incentivo e contribuição para o desenvolvimento de um estudo contribuinte para a área.

Aos amigos Everton Lacerda e Luma Vitorino pela amizade, incentivo ao desenvolvimento deste trabalho, apoio, aprendizagem sobre a área, produção de texto dissertativo e revisões.

“Veni, vidi, vici.”

(Vim, vi e venci)

Júlio César

Resumo

Diversas soluções foram desenvolvidas para o problema da verificação automática de assinaturas manuscritas *off-line*. Esta dissertação apresenta contribuições para uma característica particular desta problemática, assinaturas genuínas em diferentes tamanhos de aquisição. A assinatura, como característica biométrica comportamental, está sujeita a variações de escrita resultantes de efeitos emocionais, de humor, e de materiais de escrita, resultando em assinaturas genuínas escritas em tamanhos diferentes. Com o intuito de investigar possíveis soluções para a verificação de assinaturas em diferentes tamanhos de aquisição são analisadas características locais e globais das assinaturas e aplicados os classificadores OCC e SVM sobre dois conjuntos de assinaturas. O primeiro produzido neste estudo, com exemplares de tamanhos diferentes e, o segundo fornecido a partir da competição ICDAR 2009.

A partir das características e classificadores selecionados foram construídos quatro conjuntos de características e, através das bases de assinaturas, foram desenvolvidos 22 experimentos para cada grupo de características. Nestes experimentos foram avaliados cenários com assinaturas verdadeiras e forjadas; apenas com assinaturas verdadeiras na fase de treinamento e teste; assinaturas verdadeiras no treinamento e, verdadeiras e forjadas na fase de teste; assinaturas verdadeiras em diferentes tamanhos de aquisição aplicados nas fases de treinamento e teste. Em todos os experimentos foram geradas taxas de erro FAR (taxa de falsa aceitação) e FRR (taxa de falsa rejeição). Para comparação com os resultados obtidos na competição ocorrida no ICDAR 2009 foi gerado para o melhor resultado obtido o valor DET (curva de detecção de erro) da solução. A melhor solução em cenários envolvendo a base desenvolvida foi obtida aplicando o classificador SVM e as características CPDP, atingindo a taxa de falsa aceitação (FAR) de 26.50%, e de falsa rejeição (FRR) de 3.51%. Na base do ICDAR 2009, que contém somente exemplares com mesmo tamanho de aquisição, o melhor resultado foi obtido pelo SVM com as características CPDP, com taxas FAR de 37.50% e 0% para FRR.

Palavras-chave: Verificação de assinaturas *off-line*, Assinaturas de diferentes tamanhos de aquisição, Características locais e globais.

Abstract

Several solutions were developed to the automatic offline handwritten signature verification. In this paper, studies are developed to analyse the problematic true signature with different size acquisition. The signature is a behavior biometric characteristic. Due to emotional, mood effects or handwriting materials, like pen and pencil, writing variations might occur. In order to investigate possible solutions to the handwritten verification with different sizes of acquisition are analyzed, local and global characteristics, OCC and SVM classifiers on two signature databases. The first database is formed by signatures with different sizes of acquisition, and the second is formed by signatures from ICDAR 2009 competition database.

From characteristics and classifiers selected were built four characteristics sets and through the signature databases were developed 22 experiments to each set of characteristics. In this experiments were evaluated approaches of genuine and forgery signature, only genuine signature in training and testing phases, genuine signature in training phase and genuine and forgery applied in testing phase, genuine signature with different size applied in training and testing phases. In all experiments were generated the FAR (false accepted rate) and FRR (false rejected rate) error rates. The DET curve is used to compare the results with ICDAR 2009 competition. The best solution in first database was result of SVM classifier and CPDP characteristics, with false accepted of 26.50% and false rejected of 3.51%. In the ICDAR 2009 database formed by signatures with the same size, the best result was the SVM and CPDP with false accepted of 37.50% and false rejected of 0%.

Key-words: *offline signature verification, signatures in different size, local and global characteristics, CPDP, shadow code*

Sumário

1	INTRODUÇÃO.....	1
1.1	Motivação.....	2
1.2	Objetivo.....	3
1.3	Organização da Dissertação.....	3
2	SISTEMA DE VERIFICAÇÃO DE ASSINATURAS MANUSCRITAS.....	5
2.1	Verificação de Assinaturas.....	5
2.2	Sistema de Verificação de Assinaturas Manuscrita Off-line.....	6
2.2.1	Aquisição e pré-processamento.....	7
2.2.2	Extração de Características.....	9
2.2.3	Classificação.....	11
2.3	Estado da Arte.....	12
2.3.1	Características Periféricas e Diferenciais Periféricas.....	12
2.3.2	<i>Shadow Code</i>	13
2.3.3	Centro de Massa.....	13
2.3.4	Altura.....	14
2.3.5	Largura.....	14
2.3.6	Ângulo de Inclinação (<i>Baseline Slant Angle</i>).....	14
2.3.7	Laços.....	14
2.3.8	Densidade dos pixels.....	15
2.3.9	Centro Geométrico.....	15
3	Extração de Características.....	17
3.1	Características Globais e Locais.....	17
3.1.1	Globais.....	17
3.1.2	Locais.....	20
4	Técnica <i>One Class Classifier</i> e Classificadores.....	25
4.1	Técnica <i>One Class Classifier</i>	25
4.1.1	Modelo Gaussiano.....	26
4.1.2	Janela de Parzen.....	27
4.1.3	kNN.....	28
4.1.4	Descrição de Dados Vetor de Suporte.....	28
4.2	Máquina Vetor de Suporte.....	29

5 Metodologia.....	30
5.1 Base de Imagens.....	30
5.1.1 A Base produzida.....	30
5.1.2 Base de Imagens ICDAR 2009.....	33
5.2 Ferramentas Utilizadas.....	34
5.3 Técnicas, Métricas e Ferramentas de Avaliação.....	35
5.4 Experimentos.....	39
5.5 Configuração dos Classificadores.....	43
5.6 Estimação da taxa de erro – Particionando o conjunto de dados.....	44
5.7 Resultado ICDAR 2009.....	45
6 Conjunto de Características Propostos para a Verificação de Assinaturas.....	47
6.1 Conjunto de Características 1.....	47
6.2 Conjunto de Características 2.....	48
6.3 Conjunto de Características 3.....	48
6.4 Conjunto de Características 4.....	49
7 Resultados.....	51
7.1 Conjunto de Características 1.....	51
7.2 Conjunto de Características 2.....	54
7.3 Conjunto de Características 3.....	56
7.4 Conjunto de Características 4.....	58
7.5 Comparação entre os experimentos.....	60
8 Conclusões e Trabalhos Futuros.....	63
8.1 Trabalhos Futuro.....	64
9 Referências Bibliográficas.....	65

Lista de Figuras

Figura 1 - Processo de Verificação de Assinaturas <i>Off-line</i>	6
Figura 2 - Um sensor de aquisição de imagens [GONZALEZ, 2010].....	7
Figura 3 - Exemplo de digitalizador <i>on-line</i>	8
Figura 4 - Exemplo de assinatura binarizada: assinatura capturada (a), assinatura binarizada (b).	9
Figura 5 - As características e suas categorias segundo [IMPEDOVO, 2008].....	11
Figura 6 - Definição da altura h da imagem. Número de pixels entre o ponto mais alto e o mais baixo da imagem [PATIL, 2011].....	16
Figura 7 - Definição da largura w da imagem. Número de <i>pixels</i> entre os pontos mais a esquerda e o mais a direita [PATIL, 2011].....	17
Figura 8 - Assinatura disposta integralmente na horizontal. Ângulo de inclinação 0.....	17
Figura 9 - Assinatura com ângulo de inclinação igual a 15 graus.....	18
Figura 10 - Os centros de massa de um dos exemplares da base produzida. Os centros de massa estão localizados nas coordenadas $x = 177$ e $y = 138$, e o segundo centro nas posições $x = 398$ e $y = 125$.Figura 10 - Os centros de massa de um dos exemplares da base produzida. Os centros de massa estão localizados nas coordenadas $x = 177$ e $y = 138$, e o segundo centro nas posições $x = 398$ e $y = 125$	19
Figura 11 - A assinatura e os loops.....	21
Figura 12 - A técnica de processamento de imagem erosão identifica os laços da assinatura.....	21
Figura 13 - Resultado da aplicação da técnica erosão com o elemento estruturante disco na assinatura.....	21
Figura 14 - Faixas horizontais (a) e faixas verticais (b) definidas para a imagem da assinatura..	21
Figura 15 - Extração das características periféricas (a) e diferenciais periféricas (b) em algumas das faixas horizontais e verticais da assinatura.....	22
Figura 16 - Técnica <i>shadow code</i> aplicada sobre uma região da imagem e as sombras das projeções [SOUZA, 2009].....	23
Figura 17 - Classificação de uma Classe. Definição do limiar da região de amostras conhecidas.....	25
Figura 18 - Distribuição das amostras no Modelo Gaussiano. Evidenciação das regiões de aceitação e rejeição.....	26
Figura 19 - Técnica kNN aplicada à amostra triangular.....	27

Figura 20 - Imagens de assinaturas produzidas por assinantes do sexo feminino (a - d) e do sexo masculino (e - h). Traços de letras mais elaboradas e bem definidas permitem a nitidez das sílabas e a separação de palavras.....	32
Figura 21 - Imagens de assinaturas produzidas em diferentes tamanhos. Assinatura pequena (a), assinatura de tamanho normal ou padrão (b) e assinatura grande (c). Estes exemplares foram projetados para analisar o comportamento da técnica na verificação de assinaturas em diferentes tamanhos.....	32
Figura 22 - Assinatura em carteira de habilitação (aquisição via tablet) e de identidade (produção no próprio documento). A assinatura na habilitação é produzida em modo manuscrito e armazenada digitalmente, podendo ser redimensionada. Por sua vez, a assinatura na identidade é produzida no próprio documento, em uma área restrita que define o tamanho da assinatura. Nestes dois casos as assinaturas são apresentadas em uma versão menor do que o modo de escrita do assinante, na maioria dos exemplares.....	31
Figura 23 - Assinaturas da base AMODA para a verificação de assinaturas forjadas.....	32
Figura 24 - Imagem verdadeira (a) e forjada (b) presentes na base do ICDAR 2009 [BLANKERS et al, 2009].Figura 24 - Imagem verdadeira (a) e forjada (b) presentes na base do ICDAR 2009 [BLANKERS et al, 2009].....	33
Figura 25 - Gráfico com as curvas pelo FAR e FRR. Estas resultam no ponto EER, local em que ocorre o cruzamento das funções.....	36
Figura 26 - Exemplo de Curva ROC [FAWCELT, 2006].....	37
Figura 27 - Exemplo de medição da área sob a curva ROC.....	38
Figura 28 - Exemplo de aplicação da curva DET. Neste gráfico, o classificador representado pela linha rosa, possui a melhor classificação.....	38
Figura 29 - Exemplo de aplicação do <i>k-fold</i> em conjunto de imagens.....	45
Figura 30- Curva DET dos classificadores <i>off-line</i> do ICDAR 2009. A curva vermelha mais baixa é o classificador 13, apresentando a melhor taxa de classificação [BLANKERS et al, 2009].....	45
Figura 31 – Conjunto de Características 1, composta pelas características globais apresentadas na seção 3.1.1.....	46
Figura 32 – Conjunto de Características 2, composta por as características locais.....	47
Figura 33 – Conjunto de Características 3, projetado para simular a Arquitetura 1 de [SOUZA, 2009].....	48
Figura 34 - Conjunto de Características 4, com a inclusão da característica <i>shadow code</i>	50

Lista de Tabelas

Tabela 1 - Matriz de confusão e a relação entre a saída esperada e a saída obtida.....	34
Tabela 2 - Grupo 1: cenários de verificação de assinaturas envolvendo assinaturas originais e forjadas nas bases ICDAR 2009 e AMODA.....	40
Tabela 3 - Grupo 2: cenários OCC envolvendo as assinaturas originais da base ICDAR 2009 e as assinaturas com diferentes tamanhos de aquisição da base AMODA.....	41
Tabela 4 - Grupo 3: cenários OCC envolvendo assinaturas forjadas como <i>outlier</i>	41
Tabela 5 - Grupo 4: com assinaturas originais com diferentes tamanhos de aquisição e assinaturas forjadas habilidosas e aleatórias.....	42
Tabela 6- Avaliação dos sistemas apresentados no ICDAR 2009 para a análise <i>off-line</i>	45
Tabela 7 - Resultados para o grupo de experimento 1 no conjunto de características 1.....	52
Tabela 8 - Resultados para o grupo de experimento 2 no conjunto de características 1.....	52
Tabela 9 - Resultados para o grupo de experimentos 3 no conjunto de características 1.....	53
Tabela 10 - Resultados para o grupo de cenários 4 com o grupo de características 1.....	54
Tabela 11 - Resultados para o grupo de experimentos 1 no conjunto de características 2.....	55
Tabela 12 - Resultados para o grupo de experimentos 2 no conjunto de características 2.....	55
Tabela 13 - Resultados para o grupo de experimentos 3 no conjunto de características 2.....	55
Tabela 14 - Resultados para o grupo de cenários 4 do conjunto de características 2.....	56
Tabela 15 - Resultados para o grupo de experimentos 1 no conjunto de características 3.....	57
Tabela 16 - Resultados para o grupo de experimentos 2 no conjunto de características 3.....	57
Tabela 17 - Resultados para o grupo de experimentos 3 no conjunto de características 3.....	58
Tabela 18 - Resultados para o grupo de cenários 4 no conjunto de características 3.....	58
Tabela 19 - Resultados para o grupo de experimentos 1 no conjunto de características 4.....	59
Tabela 20 - Resultados para o grupo de experimentos 2 no conjunto de características 4.....	59
Tabela 21 - Resultado para o grupo de experimentos 3 no conjunto de características 4.....	60
Tabela 22 - Resultados para o grupo de cenários 4 no conjunto de características 4.....	60

Lista de Abreviações

AMODA	Assinaturas Manuscritas <i>Off-line</i> de Dimensões Aleatórias - Base de assinaturas produzida contendo assinaturas de diferentes tamanhos de aquisição.
AUC	Área sob a curva ROC
CG	Centro Geométrico
CPDP	Características Periféricas e Diferenciais Periféricas
DET	<i>Detection Error Trade-off curve</i>
DTW	<i>Data Time-Warping</i>
EER	<i>Error Equal Rate</i>
FAR	<i>False Accepted Rate</i> (Taxa de Falsa Aceitação)
FN	<i>False Negative</i> (Falso Negativo), taxa de exemplares negativos classificados incorretamente
FP	<i>False Positive</i> (Falso Positivo), taxa de exemplares positivos classificados incorretamente
FRR	<i>False Rejected Rate</i> (Taxa de Falsa Rejeição)
ICDAR	<i>International Conference on Document Analysis and Recognition</i>
GPDS	<i>Grupo de Procesado Digital de Senales</i>
kNN	<i>k-Nearest Neighbors</i> (k-vizinhos mais próximos)
ME	<i>Multiexpert</i> (Vários experientes)
MLP	<i>MultiLayer Perceptron</i>
NIST	<i>National Institute of Standards and Technology.</i>
NN	<i>Nearest Neighbor</i> (<i>Vizinho mais próximo</i>)
OCC	<i>One Class Classifier</i>
OCON	<i>One Class One Network</i>
RBF	<i>Radial Base Function</i>
RBFFNs	<i>Radial Basis Function Neural Network</i>
RGB	<i>Red-Green-Blue</i>
ROC	<i>Receiver Operation Characteristic</i>
SVDD	<i>Support Vector Description Data</i> (<i>Descriptor de Dados baseado em vetores suporte</i>)

SVM	<i>Support Vector Machine</i> (Máquina de Vetor de Suporte)
TN	<i>True Negative</i> (Verdadeiro Negativo), taxa de exemplares negativos classificados corretamente.
TP	<i>True Positive</i> (Verdadeiro Positivo), taxa de exemplares positivos classificados corretamente.

1 Introdução

A evolução e o crescimento populacional contribuíram para o surgimento de diferentes meios de comunicação [PONTILLE, 2008] e de necessidades sociais, como o registro de fatos, a comprovação da veracidade de um acordo, a liberação de acesso a áreas restritas, entre outros. O conhecimento da identidade de um indivíduo em um grupo tornou-se dependente do tamanho do grupo de pessoas e do envolvimento entre os seus membros. O aumento do contingente social demandou a necessidade da comprovação da identidade, realizada através de documentos compostos por assinaturas (manuscritas ou digitais), e/ou fotografias, e impressões digitais. Como exemplos destes documentos têm-se: carteira de identidade, carteira de motorista e título de eleitor.

A partir da necessidade de reconhecer o outro e suas características individuais em um grupo, cresceu o campo da biometria [JAIN, 2011] e a demanda pela identificação dos indivíduos, realizada de diferentes modos. Existem duas divisões entre as abordagens biométricas: a fisiológica e a comportamental. A biometria fisiológica, como o nome indica, utiliza-se das características fisiológicas do indivíduo para realizar a identificação, como exemplos são encontrados a impressão digital, a palma da mão, a íris, a face, entre outros. Por conseguinte, a biometria comportamental estuda as características individuais mais sujeitas a variações causadas por fatores como condições físicas, emocionais, psicológicas e/ou consequências da idade. Exemplos de biometria comportamental são o padrão de voz, o texto manuscrito e a assinatura manuscrita. A identificação única é uma necessidade atendida por todas as diferentes biometrias, independentemente da origem dos indivíduos, hábitos, condições físicas, entre outros, através da biometria é possível evidenciar a unicidade.

Dentre as diversas biometrias, a assinatura é a mais encontrada em documentos, por ser uma biometria menos intrusiva e apresentar grande aceitação na sociedade e no sistema jurídico [JUSTINO, 2001]. A assinatura manuscrita é utilizada na autenticação de documentos, cheques, cartões de crédito, transações comerciais e bancárias.

Contudo, a importância atribuída à autenticação através de assinaturas e o avanço tecnológico possibilitaram o desenvolvimento das fraudes dos documentos de identificação e das assinaturas. A verificação de assinaturas foi o processo desenvolvido para avaliar a autenticidade da assinatura [IMPEDOVO, 2007]. Inicialmente e durante muito tempo foi realizado manualmente, sofrendo as consequências da falha humana, da necessidade de mão-de-obra especializada (grafoscopia) e do tempo consumido na verificação [SOUZA, 2009]. A grafoscopia é a área responsável pelo estudo do estilo de escrita manuscrita humano, possuindo diversos meios de comparar assinaturas e identificar exemplares forjados. O aumento

populacional resultou no crescimento do número de transações e documentos, consequentemente do número de assinaturas e documentos fraudados. O desafio imposto pelas fraudes incentivou o desenvolvimento dos estudos envolvendo reconhecimento de padrões a partir da década de 60, momento em que as primeiras pesquisas a respeito de reconhecimento e verificação de assinaturas manuscritas foram apresentadas. Com base no campo da Inteligência Computacional e no Reconhecimento de Padrões desenvolveram-se os Sistemas de Verificação de Assinaturas.

Através dos Sistemas de Verificação é determinada a autenticidade da assinatura através de técnicas aplicadas para a discriminação entre uma assinatura genuína (autêntica) e uma falsificada. O processo aplicado aos Sistemas de Verificação é dividido em dois grupos: *on-line* e *off-line* [IMPEDOVO, 2007]. A verificação dinâmica ou verificação *on-line* é realizada durante o processo de escrita da assinatura [YUSOF, 2003]. O modo estático também conhecido como verificação *off-line* é o processo de verificar uma assinatura depois de ser totalmente produzida e capturada [YUSOF, 2003]. Neste estudo serão analisadas assinaturas manuscritas produzidas no modo *off-line*.

1.1 Motivação

Muitos trabalhos foram desenvolvidos na área de verificação de assinaturas *off-line* [IMPEDOVO, 2008] [BHARADI, 2010] [MUSHTAQ, 2013], contudo pelo conhecimento obtido no estudo da literatura de verificação de assinatura, não há trabalhos registrados nos congressos da área de reconhecimento de padrões e processamento de imagens evidenciando as dificuldades ou contribuições da verificação de assinaturas com exemplares genuínos com diferentes tamanhos de aquisição. Devido a grande probabilidade de variações no tamanho da assinatura, o registro do desempenho de características e classificadores aplicados neste contexto de verificação são significativos para a área.

Bases de imagens contendo assinaturas manuscritas com grafia latina existem e são acessíveis [YEUNG et al, 2004] [BLANKERS et al, 2009]. Contudo nenhuma das bases referenciadas na literatura possui exemplares produzidos intencionalmente em tamanhos diferentes. O fato de não serem encontradas bases atendendo a esta característica singular torna a criação de uma base para este campo da verificação de grande contribuição.

A técnica *One Class Classifier* desenvolvida por [TAX, 2001] apresenta um número cada vez maior de trabalhos em todas as áreas inteligência computacional, processamento de imagem, e também fora da computação, com aplicações na medicina [GANESAN et al, 2014] [KRAWCZYK et al, 2014]. Devido ao seu grande emprego e sua principal característica, a classificação binária, é válida a identificação da relevância da aplicação da técnica no contexto estudado.

Outros assuntos da área de verificação de assinaturas apresentam um grande número de trabalhos, como a contribuição dos classificadores *One Class* [TAX, 2001] [SOUZA, 2009], das características locais e globais [IMPEDOVO, 2007], e das redes neurais [ALI et al, 2004], demonstrando o contínuo crescimento do campo, as lacunas não resolvidas e as possibilidades futuras da área.

1.2 Objetivo

Este trabalho objetiva avaliar o desempenho do classificador OCC e de características globais e locais na verificação de assinaturas manuscritas *off-line* com diferentes tamanhos de aquisição. Diferentes combinações de características foram avaliadas almejando verificar as vantagens e desvantagens entre as características estudadas, bem como a análise da contribuição de cada uma destas características e classificadores ao problema investigado.

Como objetivos específicos têm-se:

- Criar uma base de assinaturas manuscritas *off-line* com assinaturas de tamanhos diferentes (exemplares normais e com tamanhos maiores e menores);
- Analisar o desempenho dos classificadores selecionados em assinaturas manuscritas de tamanhos diferentes;
- Avaliar diferentes características aplicadas na verificação de assinaturas manuscritas contendo exemplares genuínos e forjadas;
- Comparar os resultados obtidos na verificação de assinaturas genuínas e forjadas com os resultados da literatura;
- Identificar o conjunto de características com a melhor taxa na verificação de assinaturas de tamanhos diferentes;
- Comparar o melhor resultado obtido com os resultados apresentados no *International Conference on Document Analysis and Recognition* (ICDAR) de 2009.
- Avaliar estatisticamente os resultados dos experimentos.

1.3 Organização da Dissertação

Neste capítulo foram apresentadas as motivações e objetivos deste trabalho, acompanhados da descrição dos conceitos iniciais que embasam o estudo apresentado.

No capítulo 2 é apresentado o Sistema de Verificação de Assinaturas, com uma descrição das etapas constituintes: processo de aquisição de assinaturas, pré-processamento, extração de características e classificação. Seguida da apresentação do estado da arte das características aplicadas a Verificação de Assinaturas Manuscritas *Off-line*, acompanhados das taxas de erro obtidas em cada estudo.

No capítulo 3 são explanadas as características locais e globais selecionadas para este estudo a partir da literatura. Em seguida, o capítulo 4 aborda os classificadores *One Class Classifier* e o classificador SVM.

O capítulo 5 apresenta a metodologia aplicada neste estudo, composta pela descrição da base criada neste trabalho e da base distribuída pela competição de Verificação de Assinaturas do ICDAR no ano de 2009; também composta pelas métricas, técnicas e ferramentas de avaliação aplicadas; pela explanação sobre os *toolboxes* utilizados; e pelos experimentos, técnica de partição da base, e resultados obtidos no ICDAR 2009.

Em seguida, o capítulo 6 divide o grupo de características locais e globais apresentadas no estado da arte em conjuntos de características.

No capítulo 7 os resultados obtidos para os conjuntos de características planejados são analisados, seguidos de uma comparação entre os melhores resultados.

Por último, no oitavo e último capítulo serão feitas as conclusões deste projeto, com a apresentação da contribuição deste para a literatura, e de outros objetivos, em seguida são propostos alguns trabalhos futuros.

2 Sistema de Verificação de Assinaturas Manuscritas

Este capítulo tem como objetivo apresentar o sistema de verificação de assinaturas e as pesquisas que avaliam a classificação de assinaturas originais em diferentes tamanhos. Na Seção 2.1 é apresentada a área de Verificação de Assinaturas, na seção 2.2 são tratados os Sistemas de Verificação de Assinaturas Manuscritas *Off-line* e suas fases. Por último, na seção 2.3, são comentados os trabalhos que avaliam o reconhecimento de assinaturas originais produzidas em diferentes tamanhos.

2.1 Verificação de Assinaturas

O objetivo da verificação de assinaturas é validar o autor de uma assinatura através de exemplares de suas assinaturas originais [IMPEDOVO, 2008]. Para identificar o indivíduo, compara-se uma imagem de assinatura de entrada com uma base composta de outras imagens de assinaturas originais, das quais os autores são conhecidos [WIJESOMA et al, 2001].

A Verificação de Assinaturas Manuscritas é o campo do reconhecimento de padrões que estuda a classificação das assinaturas originais e forjadas [IMPEDOVO, 2007]. Existem dois tipos principais de Verificação de Assinaturas: dinâmica e estática. A verificação dinâmica ou verificação *on-line* é gerada quando o indivíduo escreve sua assinatura utilizando um dispositivo eletrônico, como um *tablet* digital [HAMILTON, 1995], e a verificação da assinatura é realizada durante o processo de escrita da mesma [YUSOF, 2003], pois este tipo de verificação visa à observação dos movimentos realizados durante o ato de assinar. Nessa modalidade, o sistema de análise de assinaturas adquire os dados, gerados durante a produção da assinatura, diretamente da ferramenta usada na digitalização (digitalizador sensível a valores dinâmicos) [IMPEDOVO, 2008]. Assim, valores dinâmicos (valores de coordenadas, tempo, velocidade da escrita, precisão, pressão da caneta) são obtidos além da própria imagem da assinatura. Todos os valores obtidos na aquisição *on-line* são utilizados como características no processo de verificação da assinatura [RUSSELL, 2003].

O modo estático também conhecido como verificação *off-line* é o processo de verificar uma assinatura depois de ser totalmente produzida e capturada [YUSOF, 2003]. Em nosso cotidiano dispomos abertamente de dispositivos de aquisição de imagens que produzem a assinatura *off-line*, porém, nessa situação não se registram as características geradas pelo indivíduo durante o processo de escrita da assinatura. A verificação de assinaturas *off-line* é uma verificação mais prática e barata, visto que os dispositivos envolvidos são de baixo custo, gerando um conjunto de informações disponíveis para a extração de características menor do que as encontradas na verificação *on-line*. Assim, devido às características resultantes do modo *off-*

line serem obtidas a partir da imagem produzida, há uma grande dependência entre as características a serem extraídas e a qualidade da imagem obtida. Mesmo apresentando desvantagens como um menor número de informações sobre a assinatura, e dependências, por exemplo, o conjunto de características a ser extraído depender da qualidade da imagem, o custo, a facilidade de aplicação e as técnicas para contornar os problemas relativos à qualidade das imagens torna o método de verificação *off-line* o mais aplicado nas pesquisas de verificação de assinaturas [IMPEDOVO, 2008].

2.2 Sistema de Verificação de Assinaturas Manuscritas *Off-line*

Um Sistema de Verificação de Assinaturas possui três fases principais: aquisição das imagens e pré-processamento, extração de características e classificação [IMPEDOVO, 2008]. Este fluxo pode ser visto na Figura 1. Na primeira fase as imagens de assinaturas são capturadas e tratadas; ao obter as imagens, as características são extraídas e armazenadas na base de conhecimento. Na etapa de classificação as características extraídas da assinatura são comparadas contra as informações contidas na base de conhecimento com o objetivo de definir a autenticidade (verdadeira ou forjada) da assinatura [IMPEDOVO, 2008]. Para facilitar o entendimento são apresentadas, nas próximas seções, as principais partes que compõem um Sistema de Verificação de Assinaturas. Na seção 2.2.1 é apresentado o processo de aquisição de assinaturas *off-line*, seguido por algumas técnicas de pré-processamento aplicadas por trabalhos da literatura. Em seguida, na seção 2.2.2, são apresentadas as divisões das características aplicadas na verificação *on-line* e *off-line*. Por último, a seção 2.2.3 apresenta a etapa de classificação e alguns grupos de classificadores utilizados na verificação de assinaturas *on-line* e *off-line*.

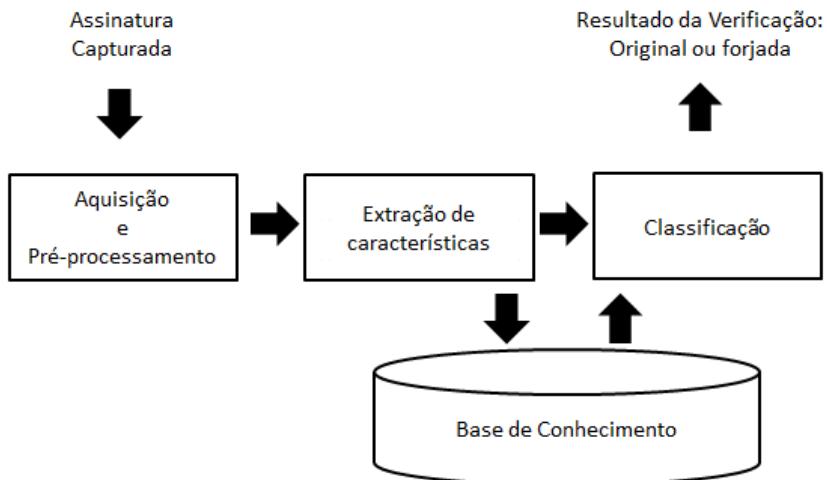


Figura 1 - Processo de Verificação de Assinaturas *Off-line*

2.2.1 Aquisição e pré-processamento

A etapa de aquisição em sistemas estáticos utiliza dispositivos de aquisição de imagens *off-line* que capturam a assinatura após o processo de escrita ter sido concluído [IMPEDOVO, 2008]. A maioria das imagens adquiridas é gerada pela combinação de uma fonte de iluminação e a reflexão ou absorção de energia dessa fonte pelos elementos da “cena” cuja imagem está sendo gerada [GONZALEZ, 2010]. A transformação da energia de iluminação para imagens digitais é efetivada através de sensores, como o apresentado na Figura 2. A ideia da transformação da energia realizada por um sensor é descrita como segue: a energia que entra e é transformada em tensão pela combinação da energia elétrica de entrada e do material do sensor, sensível a um tipo específico de energia que está sendo detectado; a forma de onda da tensão de saída é a resposta do sensor, e uma quantidade digital é obtida de cada sensor por meio da digitalização de sua resposta [GONZALEZ, 2010].

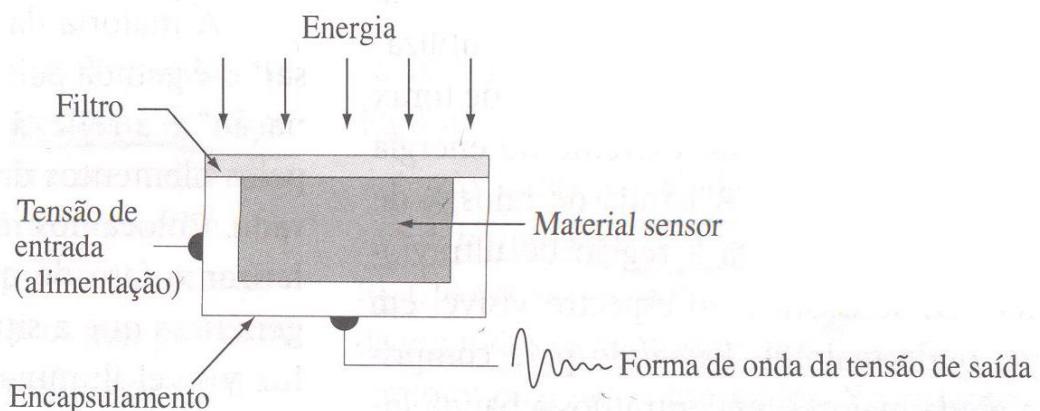


Figura 2 - Um sensor de aquisição de imagens [GONZALEZ, 2010].

Para gerar uma imagem bidimensional utilizando um único sensor, deve haver deslocamentos relativos, tanto na direção *x* quanto na *y* entre o sensor e a área de aquisição da imagem. Dispositivos, conhecidos também como arranjos mecânicos, contendo um único sensor são conhecidos como digitalizadores. Através do alto nível de precisão proporcionado pelo movimento mecânico, esse método representa uma forma acessível de obter imagens de alta resolução. Além do arranjo com um único sensor, demais arranjos com sensores combinados são gerados dependendo da aplicação, como exemplo os sensores de linhas e os sensores de área (matricial) [GONZALEZ, 2010]. A partir dos sensores foram desenvolvidos os digitalizadores *on-line* [SHASTRY, 2011] e *off-line*. Como exemplo de digitalizador a Figura 3 apresenta o *Wacom Intuos4* ilustrando a digitalização *on-line* observado na Figura 3. No modo de captura de assinaturas *on-line* são obtidas informações a respeito da pressão da escrita, velocidade, entre outros, características dinâmicas, que aumentam a eficiência e a rapidez na definição da

autenticidade da assinatura; na captura *off-line*, a imagem é o único resultado do processo de aquisição.



Figura 3 - Exemplo de digitalizador *on-line*.

Na fase de pré-processamento são aplicadas técnicas de processamento de imagens e de sinais sobre o conjunto de dados da imagem. Algumas destas técnicas seriam algoritmos de pré-processamento [IMPEDOVO, 2008], como exemplo: o filtro para conversão de RGB para nível de cinza, filtro de média para remoção de ruído; operações morfológicas; afinamento, binarização (exemplo apresentado na Figura 4) e normalização da área da assinatura.

A técnica de pré-processamento mais empregada pelos sistemas de verificação de assinaturas é a segmentação [IMPEDOVO, 2008] [GONZALEZ, 2010], que consiste em separar os objetos constituintes dos documentos do papel [PARKER, 1997]. Um estudo sobre os diversos tipos de segmentação pode ser encontrado em [MAO, 2003]. A grande dificuldade da área de segmentação de assinaturas é a escrita cursiva das letras, muitas vezes encontradas conectadas. Alguns algoritmos de segmentação de assinaturas são encontrados em [SEVERIN, 2007] [LOULLOUDIS, 2009] [SÁNCHEZ, 2011].

A binarização de imagens é uma operação de segmentação, pois através desta operação separa-se o papel (*background*) do objeto da imagem (*foreground*) [LACERDA, 2012]. Neste estudo, as imagens de entrada aplicadas ao Sistema de Verificação são apresentadas em níveis de cinza. Nos casos em que imagens coloridas são as entradas do Sistema de Verificação (formato RGB) aplica-se um processo de quantização sobre a imagem, visando reduzir o número de níveis de cor da imagem, transformando a imagem para 256 tons de cinza [GONZALEZ, 2010] [LACERDA, 2012]. Portanto, na área de estudo definida no trabalho, no caso ideal, ao final da binarização a assinatura estará separada do papel. Exemplares de assinatura verdadeira e sua versão binarizada são encontrados na Figura 4.



(a)



(b)

Figura 4 - Exemplo de assinatura binarizada: assinatura capturada (a), assinatura binarizada (b).

2.2.2 Extração de Características

De acordo com [IMPEDOVO, 2008] as características podem ser distribuídas em dois tipos: funções e parâmetros. As funções, quando utilizadas, caracterizam a assinatura em termos de uma função temporal da qual os valores constituem um vetor de características. Por sua vez, quando as características paramétricas são utilizadas, a assinatura é caracterizada como um vetor de elementos, com cada um dos elementos representando o valor de uma característica. Ao comparar os dois tipos de características observa-se que as funcionais têm melhor desempenho do que às paramétricas, porém, elas necessitam de um tempo de consumo na coleta das características considerável para realizar a comparação entre as imagens. Devido a desvantagem relativa ao tempo de consumo apresentada pelas características funcionais foram selecionadas neste trabalho somente características paramétricas para a Verificação de Assinaturas *Off-line*.

Algumas das características funcionais mais comuns encontradas na literatura são: duração do tempo de assinatura, contatos da caneta com o papel (*pen-down*, *pen-up*), velocidade, e aceleração; estas características são fortemente usadas na verificação de assinaturas *on-line*. O tempo de consumo é uma desvantagem devido às características funcionais serem mensuradas durante todo o período de escrita da assinatura, pois ao extrair esta característica apenas em parte da assinatura, a caracterização da assinatura não é completa.

As características paramétricas são classificadas em duas categorias: global e local [IMPEDOVO, 2008]. As características globais são descritas ou identificadas analisando-se a assinatura inteira [ARYA, 2010] e são extraídas considerando todos os *pixels* ativos (pontos pretos da imagem binária observada na Figura 4) encontrados na imagem. Na maioria das imagens que contém assinaturas há áreas brancas que não agregam valor à extração das características, com o objetivo de tornar a extração mais rápida, as áreas brancas são removidas, definindo uma versão da imagem conhecida como *bounding box*, contendo apenas a assinatura. Características globais são mais fáceis de serem extraídas e não são sensíveis a ruído, mas dependem do alinhamento da assinatura, e são sensíveis a distorções e variações de estilo [ARYA, 2010]. Alguns exemplos de características globais são a orientação da assinatura, área

da imagem, coeficiente da transformação matemática, número de componentes, entre outros [IMPEDOVO, 2008].

Os valores das características paramétricas locais são extraídos de partes específicas da assinatura; sendo alguns exemplos os parâmetros de componentes orientados e os parâmetros orientados por *pixel*. Os componentes orientados são extraídos no nível de cada componente (razão altura e largura do elemento, posição relativa dos elementos, entre outros). Por sua vez, os parâmetros orientados a *pixels* são extraídos no nível de *pixel* (densidade de pixel, intensidade nível de cinza, textura, entre outros). A divisão das características e suas classificações podem ser observadas na Figura 5.

Independente do conjunto de características extraído, a unicidade das assinaturas é garantida, pois as assinaturas de um mesmo autor possuem valores de características próximos, contudo nunca iguais [JUSTINO, 2000] [JAIN, 2011]. Assinaturas de diferentes autores podem conter ou não o mesmo conjunto de características, como exemplo: uma assinatura apresentar laços e outra não; esta situação resulta na impossibilidade de um conjunto universal de características atribuído a todas as assinaturas analisadas pelo Processo de Verificação de Assinaturas [IMPEDOVO, 2008] [JAIN, 2011]. Além deste motivo, selecionar as características mais representativas é um requisito requerido para o sistema de verificação ser eficiente, apresentar os menores custos e atender aos menores requisitos de memória possíveis [IMPEDOVO, 2008]. Um bom sistema de Verificação de Assinaturas Manuscritas *Off-line* busca possuir o menor vetor de características, ter o maior número possível de autores representados por suas assinaturas no sistema e apresentar a velocidade mais rápida em um processo de verificação.

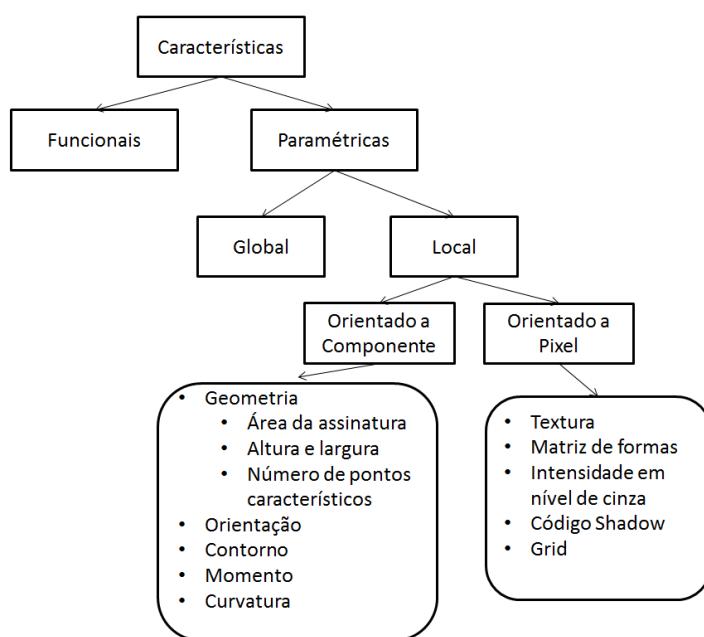


Figura 5 - As características e suas categorias segundo [IMPEDOVO, 2008].

2.2.3 Classificação

A etapa de classificação avalia a autenticidade da assinatura (verdadeira ou forjada) através de comparações entre as características da assinatura contra as características das assinaturas armazenadas na base de conhecimento. A autenticidade da assinatura é expressa por uma saída de valor booleano (sim ou não). A saída sim é atribuída às assinaturas com valores de características similares às assinaturas registradas como originais na base de conhecimento. A saída não indica a não similaridade da assinatura com os vetores de características das assinaturas originais da base. O sucesso da classificação das assinaturas envolve muitos aspectos que vão desde a definição do conjunto de características coletado das imagens, incluindo a técnica de comparação de assinaturas, até a estratégia de criação e desenvolvimento da base de assinaturas de referência (base de conhecimento) [IMPEDOVO, 2008].

A base é composta por assinaturas produzidas por diferentes autores, assinaturas de um mesmo autor possuem grande similaridade, e assinaturas de diferentes autores possuem um grau de diferenciação, por este fato para o campo do reconhecimento de padrões afirma-se que assinaturas similares compõem uma classe. Neste contexto, o objetivo da classificação é atribuir à classe de assinaturas mais semelhante à assinatura analisada. As técnicas conhecidas como classificadores são as responsáveis por aprenderem as características de uma assinatura e definir a sua classe. Os classificadores são agrupados de acordo com abordagem de classificação aplicada, como exemplo: técnicas de comparação de *templates* (*DTW*), estatística (*kNN*, distância Euclidiana, Mahalanobis, entre outros), vetor de suporte (*SVM* e *SVDD*), redes neurais (*MLP*, *RBF*, redes fuzzy). Contudo, nos últimos anos a abordagem *multiexpert* (ME) tem sido aplicada na verificação de assinaturas [IMPEDOVO, 2008]. Nesta abordagem, a classificação não é definida por apenas um classificador, e sim por um conjunto de classificadores que contribuem uns com os outros para definir a classe da assinatura, essa colaboração resulta no aumento da contribuição das características vantajosas de cada um, e na diminuição das desvantagens.

2.3 Estado da Arte

Através da necessidade de identificar a originalidade das informações (letra, assinatura, entre outros) contidas em documentos digitalizados desenvolveu-se a área Verificação de Assinaturas Manuscritas. Muitos trabalhos foram desenvolvidos desde a década de 60 na tentativa de definir uma solução genérica para a Verificação de Assinaturas. Algumas destas propostas foram submetidas em competições como as promovidas pelo ICDAR ocorridas nos anos de 2009 [BLANKERS et al, 2009], 2011 [EL ABED et al, 2011] e 2013 [MALIK et al, 2013]. A Verificação de Assinaturas cresceu com o aumento do número dos casos de fraudes. As

assinaturas forjadas são divididas em três grupos: aleatória, simples e habilidosa [JUSTINO et al, 2001]; neste estudo foram analisadas assinaturas forjadas aleatórias e habilidosas. Diversos trabalhos da literatura analisam os três tipos de assinaturas forjadas [COETZER et al, 2006] [PUSHPALATHA et al 2013]. Outra problemática da verificação de assinaturas são as assinaturas originais com diferentes tamanhos de aquisição, este campo apresenta registros na literatura apenas para a verificação *on-line* [ARAUJO et al, 2007].

Neste estudo, a Verificação de Assinaturas Manuscritas é executada no modo *Off-line*, com aplicação de assinaturas originais dispostas em diferentes tamanhos de aquisição e assinaturas forjadas nos tipos aleatórias e habilidosas. Os experimentos apresentados analisam esta problemática através da aplicação de um grupo de características globais e locais.

2.3.1 Características Periféricas e Diferenciais Periféricas

Souza [SOUZA, 2009] utilizou a extração de Características Periféricas e Diferenciais Periféricas (CPDP) para a verificação de assinaturas forjadas. O número de faixas foi analisado e definido em três. Nos experimentos apresentados em seu trabalho com a Base 1 definida em [GOMES, 1995], composta por assinaturas falsas habilidosas e aleatórias, os melhores resultados com o CPDP foram com o classificador NN atingindo, na verificação de assinaturas aleatória, a taxa de rejeição de assinaturas forjadas de 59,64% para 5% falsos positivos e 71,08% para 10% de falsos positivos. O método CPDP se destacou nos classificadores NN e KNN na Base 2, conjunto de assinaturas disponibilizada no ICDAR 2009 [BLANKERS ET al, 2009] apresentando para a taxa de rejeição de 59,41% para 5% de falsos positivos e de 60,88% para 10% de falsos positivos. A técnica CPDP apresentou altos valores para a taxa de erro de falsa rejeição em ambos os casos, devido a grande similaridade entre as assinaturas originais e forjadas habilidosas proporcionada pela determinação das características periféricas através da distância entre as bordas do *bounding box* e o primeiro *pixel* preto de cada linha, e da característica periférica diferencial. Por esta característica, resultante das altas taxas de assinaturas forjadas rejeitadas, a técnica apresenta boas razões nos cenários de assinaturas com diferentes tamanhos de aquisição.

2.3.2 Shadow Code

Souza [SOUZA, 2009] aplicou em seu estudo também a característica *Shadow Code* no mesmo contexto das características CPDP. Através da variação da dimensão em pixels dos blocos da imagem obteve-se o tamanho de janela 32x32 apresentando a mais alta taxa de verificação. Neste trabalho o método *shadow code* foi considerado o melhor na maioria dos

experimentos apresentando para Base 1 [GOMES, 2006] o melhor resultado para a verificação de assinaturas habilidosas igual a 59,18%, e para as forjadas aleatórias igual a 91,87%, e globalmente a melhor taxa de 75,26%. Na Base 2 [BLANKERS et al, 2009] foi atingida a taxa global de 58,98%.

2.3.3 Centro de Massa

- Pansare e Bhatia

O estudo desenvolvido por [PANSARE e BHATIA, 2012] analisou a verificação de assinaturas originais e forjadas através da extração de características geométricas e treinamento com redes neurais. A base de assinaturas utilizada foi a Grupo de Procesado Digital de Senales (GPDS). Como resultado foi obtido 82,66% de taxa de classificação correta.

- Kapil et al

O trabalho apresentado por [KAPIL, 2014] utiliza o centro de massa como uma das características para a verificação das assinaturas, além desta a área e a razão são calculadas. Através do classificador conhecido como rede neural foi obtida 82,66% de taxa de classificação correta.

2.3.4 Altura

- Baltazakis e Papamarkos

No trabalho de [BALTAZAKIS e PAPAMARKOS, 2001] a altura é uma das várias características extraídas (características textura, grid e globais) das assinaturas no processo de verificação *off-line*. Aplicando um Perceptron de dois estágios conhecido como OCON (uma-classe-uma-rede) combinado com a distância euclidiana, os resultados alimentam um segundo classificador, o radial base function (RBF) que define a classificação. Os resultados obtidos foram 80,81% de verificação correta e 19,91% de erro na classificação.

2.3.5 Largura

- Patil e Patil

O trabalho de [PATIL e PATIL, 2013] apresenta um método de verificação de assinatura baseado na extração de características globais como: área, altura e largura. Para a classificação

das assinaturas foi utilizada o modelo da distância euclidiana. O sistema apresentou a taxa de reconhecimento de 89%.

2.3.6 Ângulo de Inclinação (*Baseline Slant Angle*)

- Jana

No trabalho de [JANA et al, 2014] a verificação das assinaturas é realizada através das características ângulo de inclinação, razão altura e largura, área e centro de gravidade, como características de textura e topológicas. O processo de verificação das assinaturas neste trabalho utilizou a distância euclidiana como classificador, obtendo resultados em um dos cinco conjuntos de dados analisados a acurácia de 100% de acerto, seguido de 96,15% em uma segunda base.

2.3.7 Laços

- Kovari et al

O trabalho desenvolvido por [KOVARI et al, 2009] foca na etapa de classificação da verificação de assinaturas *off-line*. Diversas características locais, incluindo os laços, são aplicadas e comparadas usando diferentes abordagens de classificação. Através da base de assinaturas da Competição de Verificação do ano de 2004 [YEUNG et al, 2004], devido a esta base ser contida apenas de características de extração *on-line* não dispondo da imagem da assinatura, as informações foram usadas para sintetizar, através de técnicas como a interpolação bicúbica, assinaturas similares para originais. Assinaturas forjadas foram também geradas artificialmente. Os resultados para a verificação de assinaturas verdadeiras contra forjadas apresentaram para as características relacionadas aos laços o valor da taxa *Equal Error Rate* (EER) igual a 0,43.

- Kovari e Charaf

No trabalho de [KOVARI e CHARAF, 2013] foi proposto um modelo de verificação de assinaturas *off-line* probabilístico simplificado. Neste cada etapa da verificação é descrita matematicamente. A extração dos laços foi realizada através da criação do elemento estruturante morfológico 3x3. A base utilizada para os experimentos foi a disponibilizada no site de Kovari, dispondo de assinaturas originais e forjadas, para os testes finais utilizou-se novamente a base da competição de verificação de assinaturas (SVC2004) [YEUNG et al, 2004], composta por características de assinaturas latinas. Através destas Kovari e Charaf afirmam que linhas-base e

laços são boas características para a caracterização de assinaturas escritas em grafias latinas. O melhor resultado incluindo os laços no conjunto de características retornou 26,6% de taxa de erro, maior que a taxa de erro prevista para o mesmo cenário 23,4%.

2.3.8 Densidade dos pixels

No trabalho de [AZZOPARDI, 2008] investigou-se a efetividade da *Radial Basis Function Neural Network* (RBFNNs) nos sistemas de verificação de assinaturas *off-line*. De modo semelhante, ao trabalho de [Baltazakis e Papamarkos, 2001] foram utilizadas características globais, grid e de textura. Neste grupo, encontra-se a característica densidade dos pixels sendo também extraída. O experimento foi realizado sobre assinaturas forjadas randômicas e resultaram em 1.58% como taxa de falsa rejeição (FRR) e 2.5% como taxa de falsa aceitação (FAR).

2.3.9 Centro Geométrico

- Majhi et al

Neste trabalho [NAJHI et al, 2006] um novo método de extração de característica foi sugerido para a verificação *off-line*, neste método utilizou-se a característica centro geométrico. Para a classificação foi aplicada a distância euclidiana, devido a ser uma técnica rápida. Através destas técnicas foram atingidas para as assinaturas forjadas aleatórias, simples e habilidosas as respectivas taxas de erro de falsa aceitação: 2.08%, 9.75% e 16.36%.

- Afsaroodost et al

Com o objetivo de distinguir se uma assinatura verificada é forjada ou verdadeira foi desenvolvido o método apresentado por [AFSARDOOST et al, 2008] baseado na característica centro geométrico. Novamente, como classificador foi aplicada a distância euclidiana. Neste estudo foram obtidos 14% para a taxa de erro falsa rejeição e 10% para a taxa de falsa aceitação das assinaturas originais.

3 Extração de Características

Este capítulo tem como objetivo apresentar as técnicas de extração de características aplicadas neste estudo. Na seção a seguir são descritas as características selecionadas a partir da literatura [IMPEDOVO, 2007]. A etapa de extração de características é de grande importância para o Sistema de Verificação de Assinaturas, visto que um bom vetor de características da imagem em análise resulta em melhores resultados na etapa de classificação [SOUZA, 2009]. Boas características discriminatórias devem ser selecionadas para os classificadores identificarem os objetos. Uma extração com baixa representatividade é realizada com características pouco discriminantes e acarreta em dados não significativos na distinção das classes [IMPEDOVO, 2008].

3.1 Características Globais e Locais

Como apresentado na seção 2.2.2, as características selecionadas neste estudo estão agrupadas no conjunto paramétrico, conjunto detentor das características obtidas na extração *offline*, divido em dois subconjuntos: globais e locais. Estas características são extraídas através de cálculos e projeções realizados sobre os pixels da imagem. As características globais são geradas considerando toda a assinatura. As características locais são extraídas em partes específicas da assinatura. No Sistema de Verificação de Assinaturas Manuscritas deste estudo foram utilizadas as características abaixo.

3.1.1 Globais

Nesta subseção são apresentadas as características que analisam toda a imagem.

- Altura e Largura

A altura da imagem (h) é definida pelo número de pixels entre a posição do ponto mais alto e o ponto mais baixo no eixo y [PATIL, 2011], característica observada na Figura 6. A largura da imagem (w) possui a mesma definição considerando o eixo x [PATIL, 2011], imagem vista na Figura 7.

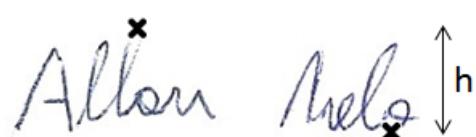


Figura 6 - Definição da altura h da imagem. Número de pixels entre o ponto mais alto e o mais baixo da imagem [PATIL, 2011].

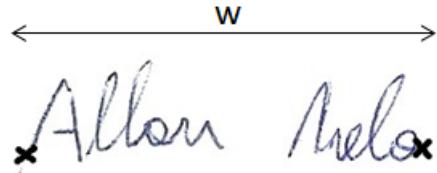


Figura 7 - Definição da largura w da imagem. Número de *pixels* entre os pontos mais a esquerda e o mais a direita [PATIL, 2011].

- Área da Assinatura

A partir das características altura h (imagem observada na Figura 6) e largura w (imagem apresentada na Figura 7) define-se a característica área da assinatura. Para gerar a característica área multiplica-se a largura pela altura, como apresentado pela Equação 1 [PANSARE, 2012].

$$\text{área} = \text{altura} * \text{largura} \quad \text{Equação 1}$$

- Ângulo de Inclinação (*Slant Angle*)

O ângulo de inclinação da assinatura indica o grau de inclinação da assinatura, esta métrica é conhecida na literatura como *slant angle* [MAHANTA, 2013]. Este grau varia de zero nas situações em que a assinatura está definida inteiramente na horizontal (imagem observada na Figura 8) até 90 graus, com assinaturas na vertical (imagem apresentada na Figura 9). A angulação da assinatura é definida a partir do tamanho horizontal e das alturas extraídas do inicio e do fim da assinatura. Através da diferença entre as alturas calculadas para a assinatura define o primeiro cateto para o cálculo da hipotenusa da assinatura, o segundo cateto é o tamanho horizontal da imagem. Com estes dois catetos calcula-se a hipotenusa da assinatura. Por último, calcula-se a razão angular entre a hipotenusa e o tamanho horizontal, definindo o ângulo de inclinação.



Figura 8 - Assinatura disposta integralmente na horizontal. Ângulo de inclinação 0.



Figura 9 - Assinatura com ângulo de inclinação igual a 15 graus.

- Centro de Massa

Para definir o centro de massa de uma imagem de assinatura divide-se a imagem em duas partes e calcula-se o centro de massa das partes individualmente [PANSARE, 2012] [JANA, 2014]. O centro de massa é um par de coordenadas x e y obtido da região analisada, como observado na Figura 10.

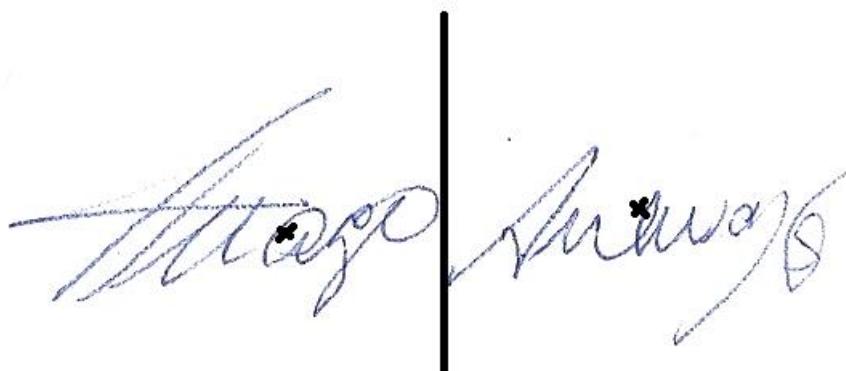


Figura 10 - Os centros de massa de um dos exemplares da base produzida. Os centros de massa estão localizados nas coordenadas x = 177 e y = 138, e o segundo centro nas posições x = 398 e y = 125.

- Centro Geométrico

O centro geométrico [AFSARDOOST, 2008]**Erro! Fonte de referência não encontrada.** é definido através do centro de gravidade, do peso da assinatura e das suas posições inicial e final nos eixos x e y, como visto no Algoritmo 1.

Algoritmo 1: Centro Geométrico

Entrada: Altura e largura da assinatura, valores X e Y da posição inicial da assinatura, valores X e Y da posição final da assinatura.

Saída: centro geométrico da assinatura

1. Cálculo do peso da assinatura através da razão dos centros de massa.
2. Determinação do datum da assinatura (ponto de partida arbitrário).
3. Encontro do centro de gravidade (distância do datum para o ponto central da assinatura).
4. Determinação do momento (multiplicação da distância pelo peso).

3.1.2 Locais

Nesta subseção são apresentadas as características que analisam particularidades de regiões da imagem.

- Laços

Os laços são características marcantes dos idiomas e na língua portuguesa muitas letras possuem laços. Esta característica é definida como um traçado totalmente cruzado, muito semelhante a uma elipse ou um círculo (imagem observada na Figura 11) [IMPEDOVO, 2007]. Os laços podem ser pequenos, médios e grandes, dependendo da letra que está se escrevendo, do modo de escrita do autor, e da região em que se escreve. Para a extração dos laços neste estudo foi aplicado o algoritmo de extração de laços descrito no Algoritmo 2.

Algoritmo 2: Laços

Entrada: imagem da assinatura

Saída: a quantidade dos laços distribuídos nos tipos (laço pequeno localizado na base, laço pequeno localizado no centro da imagem, laço pequeno localizado no topo da imagem, laço de tamanho médio localizado na base, laço médio localizado no topo, total dos laços pequenos, total de laços médios, total de laços grandes e o total de laços).

1. A assinatura binarizada é negada (Figura 11**Erro! Fonte de referência não encontrada.**).
2. Aplica-se a técnica de processamento de imagem do *Matlab imfill* para preenchimento dos buracos (laços) existentes na assinatura (visualizado na Figura 12).
3. Cria-se através do método *Matlab strel* o elemento morfológico estruturante disco com raio igual a 5 para identificar os laços.
4. Aplica-se a técnica de processamento de imagem erosão sobre a imagem através do disco estruturante morfológico (visto na Figura 13**Erro! Fonte de referência não encontrada.**).
5. A imagem apenas com laços é dividida em 3 regiões: a base, região central, e topo. Nestas são identificados os laços pequenos através do método do *Matlab bwareaopen* responsável por remover da imagem binária os objetos menores que um tamanho informado.
6. De modo semelhante a imagem é dividida em duas regiões para a definição dos laços

médios através do aumento do tamanho dos objetos identificados pela função *bwareaopen*.

7. Em seguida são identificados em toda a imagem os laços pequenos e médios ainda não encontrados, e os laços grandes não estudados anteriormente. A variação no tamanho dos laços procurados é definida na função *bwareaopen* por um número inteiro duplicado para todos os laços.

8. Por fim são computados todos os loops existentes na imagem.

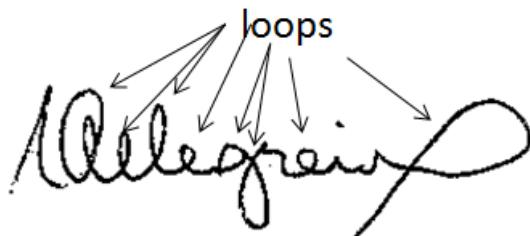


Figura 11 - A assinatura e os loops.

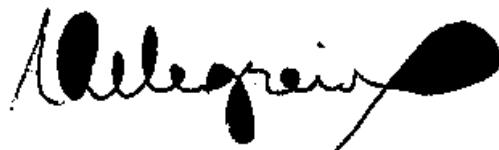


Figura 12 - A técnica de processamento de imagem erosão identifica os laços da assinatura.



Figura 13 - Resultado da aplicação da técnica erosão com o elemento estruturante disco na assinatura.

- Densidade dos pixels

A densidade dos pixels é a quantidade de pixels pretos presente na assinatura dividida pela quantidade de pixels da imagem. [VERMA, 2013].

- Características Periféricas e Diferencias Periféricas

Devido aos bons resultados obtidos ao utilizar estas características na verificação de assinaturas no estudo de [SOUZA, 2009] aplicou-se esta técnica de extração. Este método de extração local subdivide as imagens das assinaturas em faixas horizontais e verticais (imagens

observadas nas letras *a* e *b* da Figura 14), para cada faixa são extraídas a característica periférica e a diferencial periférica [FANG, 2002]. No algoritmo 3 estão descritos os passos para a extração periférica.

Para extrair a característica diferencial periférica altera-se o passo cinco da extração periférica para somar o número de pixels entre um primeiro e um segundo pixel preto no sentido em análise, considerando a existência de pixels brancos entre os pixels pretos referidos (imagem observada na imagem b da Figura 15).

Para cada imagem é extraída um número de características igual a oito vezes a quantidade de faixas, pois o processo de extração ocorre nos sentidos: de cima para baixo, de baixo para cima, na vertical; e da esquerda para a direita, e da direita para a esquerda, na horizontal, para os dois tipos de características (periférica e diferencial periférica).

Algoritmo 3: Extração das características periféricas e diferenciais periféricas.

Entrada: imagem da assinatura

Saída: conjunto de distâncias (número de pixels brancos) periféricas e diferenciais periféricas.

1. Definição de um *bounding box* para cada imagem
2. A imagem resultante é dividida em m faixas horizontais e verticais (Figura 14)
3. Toma-se uma das faixas verticais ou horizontais
4. Percorre-se a faixa em um sentido (horizontal ou vertical). No primeiro caso a imagem pode ser percorrida da direita para a esquerda ou da esquerda para a direita; no segundo, de cima para baixo ou de baixo para cima.
5. Ao percorrer a linha ou coluna soma-se o número de pixels entre a borda de partida e o primeiro pixel preto localizado na direção percorrida (como apresentado na imagem a da Figura 15)
6. Soma-se o número de pixels de cada linha ou coluna de toda a faixa
7. Volta-se ao passo 4 e percorre-se a faixa em outro sentido
8. Define-se uma nova faixa e repete-se todo o processo a partir do item 2

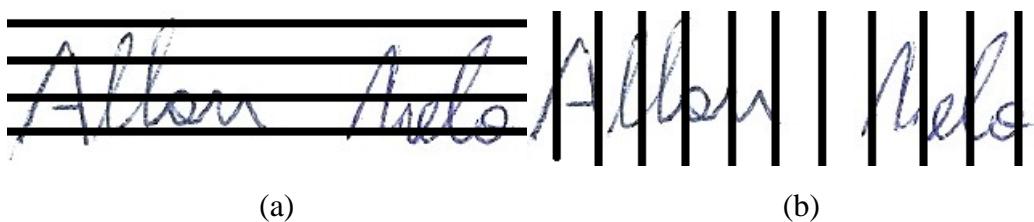


Figura 14 - Faixas horizontais (a) e faixas verticais (b) definidas para a imagem da assinatura.

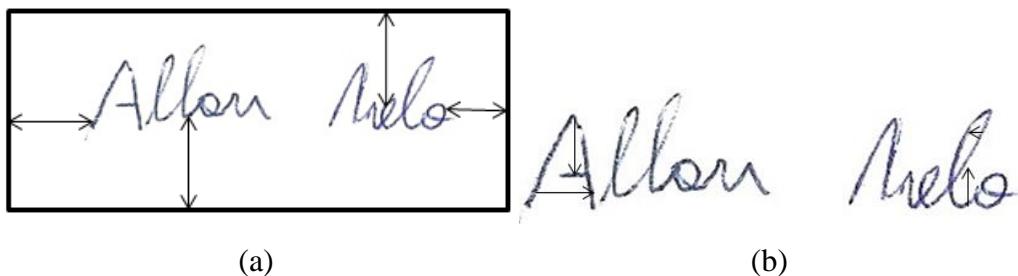


Figura 15 - Extração das características periféricas (a) e diferenciais periféricas (b) em algumas das faixas horizontais e verticais da assinatura.

- Shadow Code

O método de extração de características proposto por Sabourin [SABOURIN, 1993] intitulado *Shadow Code* divide as imagens em uma malha de $n \times n$ pixels. Para cada região definida nesta malha são calculados os valores das características com base no número de pixels pretos projetados sobre as barras (horizontais e verticais) aplicadas na região [SOUZA, 2009]. A projeção dos pixels pretos sobre as barras é denominada como sombra, seis projeções são geradas para cada região: dois horizontais (topo e base), dois verticais (esquerda e direita) e dois diagonais (principal e secundária) [SOUZA, 2009] [SABOURIN, 1993]. O valor de cada característica varia entre 0 e n para as barras horizontais e verticais e entre 0 e $n\sqrt{2}$ para as barras diagonais, o valor definido representa o número de pixels presentes na sombra das barras.

A seguir é apresentada na Figura 16 a imagem de um caractere dividido em regiões compostas por $n \times n$ pixels, cuja sombra encontra-se projetada nas barras. Devido a variação do tamanho das imagens, o número de regiões e o tamanho destas variam de imagem para imagem. Consequentemente, o número total de características também sofre variação. O vetor de características para cada região é definido como $V = \{T, B, E, D, D_p, D_s\}$. T representa o comprimento da sombra no topo da região, B a sombra da base, E a sombra da esquerda, D a sombra da direita, D_p a sombra da diagonal principal e D_s a sombra da diagonal secundária. Considerando a região evidenciada na imagem da Figura 16 definimos o vetor de características $V = \{\frac{n}{2}, \frac{n}{2}, n, n, \frac{n\sqrt{2}}{6}, \frac{2n\sqrt{2}}{3}\}$. Sendo esta região da imagem um bloco 32x32 pixels obtém-se o $V = \{16, 4, 32, 32, 7, 30\}$ [Souza, 2009].

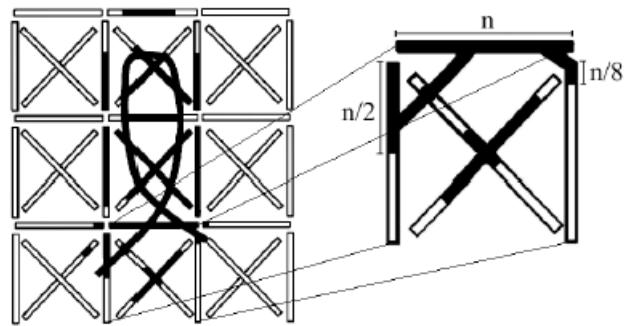


Figura 16 - Técnica *shadow code* aplicada sobre uma região da imagem e as sombras das projeções [SOUZA, 2009].

4 Técnica *One Class Classifier* e Classificadores

Como apresentado na seção 2.4, a etapa de extração de características gera um vetor de características para cada assinatura da base, os vetores gerados são enviados aos classificadores inseridos no Sistema. Na etapa de verificação, os classificadores aprendem as classes estudadas através dos vetores de características, adquirindo com o aprendizado a capacidade de distinguir as classes através de suas características [SOUZA, 2009]. Uma boa verificação das imagens analisadas na fase de testes é resultante de um bom conjunto de características selecionadas e do aprendizado dos classificadores na fase de treinamento. Assim, a verificação permite que um número de objetos tenha a sua classe verificada sem intervenção humana.

Neste capítulo são apresentados os classificadores utilizados nos experimentos de Verificação desenvolvidos. Alguns dos classificadores avaliados compõem o conjunto de classificadores da técnica de classificação com uma classe, conhecida em inglês como *One Class Classifier* [TAX, 2001], e outro classificador é definido pela classificação binária0. Na primeira seção é definida a Técnica *One Class Classifier*, seguida de uma descrição dos classificadores OCC aplicados neste estudo. Em seguida é apresentado o classificador SVM [CORTES et al, 1995], classificador binário com boa utilização no campo da verificação de assinaturas, selecionado para este estudo com o objetivo de estabelecer um comparativo entre os resultados obtidos e os apresentados pela literatura. A escolha do SVM, além do motivo exposto, deve-se também a sua característica binária, cenário ideal da verificação de assinaturas com exemplares originais e forjados.

4.1 Técnica *One Class Classifier*

A Técnica *One Class Classifier* em português é denominada como: classificadores de uma classe, detectores de dados discrepantes ou aprendizes de conceitos; foi desenvolvida na tese de doutorado de Tax [TAX, 2001] e seus classificadores apresentam como diferencial o aprendizado e a classificação/verificação de classes detendo apenas o conhecimento de uma das classes envolvidas no problema. Os classificadores *One Class* geram de modo semelhante à classificação binária uma linha divisória entre os exemplares que pertencem à classe conhecida e os demais exemplares, que não pertencem a esta classe (situação ilustrada na Figura 17). O desafio desta técnica é definir esta fronteira limite possuindo apenas as características (informações) de uma das classes.

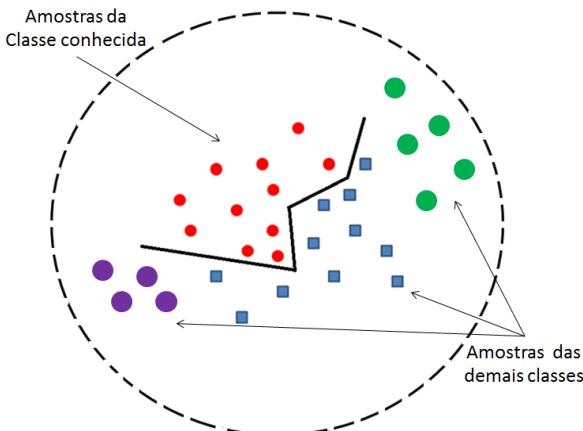


Figura 17 - Classificação de uma Classe. Definição do limiar da região de amostras conhecidas.

Uma das aplicações da Técnica *One Class* ocorre em problemas em que se têm dificuldades em encontrar exemplares de todas (ou da maioria) das classes. A ausência destas no amostras no treinamento das técnicas *One Class* é uma vantagem ao seu emprego. Por terem o comportamento semelhante ao dos classificadores binários, apresentando como saída um valor booleano (sim ou não), as amostras da classe estudada são consideradas positivas (sim), as demais amostras são consideradas negativas (não). Assim, o objetivo dos classificadores de uma classe é aceitar ou rejeitar as amostras com base na similaridade do conjunto de elementos positivos conhecidos [TAX, 2001]. Contudo, o grau de generalização atingido por um conjunto de exemplares de assinaturas e a definição do erro para um cenário em análise são pontos a serem avaliados nos métodos de classificação *One Class*. Estes problemas são motivados pela aplicação exclusiva dos exemplares de uma única classe para definir a superfície de separação das assinaturas em um cenário [TAX, 2001].

Na próxima seção são apresentados os classificadores de uma classe (gaussiana, janela de parzen, knn e svdd) selecionados para os experimentos deste trabalho. Os classificadores escolhidos representam diferentes abordagens disponibilizadas por Tax [TAX, 2001] no *toolbox DDT* [TAX, 2013]. O modelo gaussiano, janela de parzen e kNN são diferentes classificadores definidos sob o conceito da estimativa de densidade, estes classificadores foram selecionados por implementarem de diferentes modos esta estimação. O classificador SVDD foi selecionado dentre os classificadores OCC desenvolvidos por Tax por ser o único baseado em vetores de suporte, característica encontrada também no classificador SVM, por este motivo escolhido.

4.1.1 Modelo Gaussiano

A primeira técnica da abordagem *One Class* apresentada é a estimativa da densidade através do descritor de dados gaussiano. Este modelo apresenta como característica a

simplicidade da função gaussiana (gráfico contendo um modelo para a função gaussiana que pode ser visto na Figura 18) na estimativa da densidade dos dados de treinamento. De modo semelhante ao aplicado a todos os classificadores do *toolbox dd_tools*, parâmetros são definidos para a etapa de treinamento da função gaussiana. Os parâmetros conjunto de amostras de treinamento (A), fração do conjunto pertencente à outra classe (fracrej) e taxa de regulação para a matriz de covariança estimada (r) são definidos para a etapa de treinamento. Através das características definidas a partir das assinaturas, o classificador define os parâmetros: média e variância, utilizados no cálculo da função gaussiana, esta função é responsável na fase de teste pela atribuição da classe os exemplares testados.

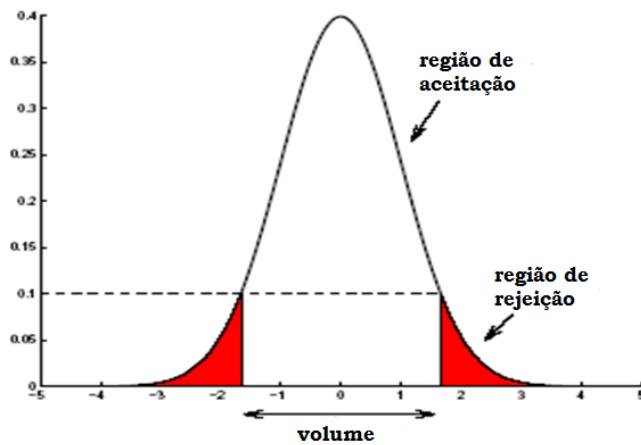


Figura 18 - Distribuição das amostras no Modelo Gaussiano. Evidenciação das regiões de aceitação e rejeição.

4.1.2 Janela de Parzen

Outro método de estimativa de densidade muito utilizado na classificação de assinaturas é a Janela de Parzen. Este método tem suas origens na técnica de representação de imagem conhecida como histograma. A estimativa de densidade é utilizada para estimar a probabilidade de um vetor x que é obtida da função de estimativa de densidade $p(x)$, apresentada na equação 2, cair dentro de uma região R .

$$p(x) = \frac{k}{V} \quad 2$$

Em problemas reais, ao fixar o volume V , o tamanho da região R e determinar o valor de k através dos dados de aprendizagem surge o método denominado de Janela de Parzen. Para estimar a densidade no ponto x do gráfico, centra-se a região R em x e contamos o número de exemplos em R substituindo na equação 2. A quantidade de pontos da região R é a função kernel ou Janela de Parzen. Com o tamanho da Janela de Parzen sendo definido por h . Ao escolher um

valor para h supõe-se a região na qual a densidade é constante. O valor de h deve ser estabelecido através de uma base de validação. Maior h resulta em uma melhor generalização, um h pequeno gera uma classificação perfeita com fronteiras muito especializadas.

4.1.3 kNN

O método k -vizinhos mais próximos (k -Nearest Neighbot – k NN) [FUKUNAGA, 1990] é um método de aprendizagem supervisionada [RUSSELL, 2004] e não paramétrico empregado em tarefas de classificação e regressão, além de ser um dos mais simples algoritmos de aprendizagem de máquina. O classificador k NN é um método que calcula a probabilidade das características de uma amostra X serem semelhantes às características dos pontos na vizinhança de X . Por ilustração, na Figura 19, um classificador recebe como entrada um elemento triangular X e, através de uma quantidade k de vizinhos próximos a amostra X , este informa como saída a classe 1 ou 2, a qual este elemento X pertence.

O k NN é um método sem construção de modelo ou função na fase de treinamento, realizando a verificação através da memorização de todas as amostras do conjunto de treinamento. A verificação é realizada através do cálculo da distância euclidiana entre os valores das características constituintes do elemento X com cada uma das demais amostras do espaço da Figura 19. A partir do voto majoritário dos k vizinhos mais próximos definidos para um problema é definida a classe deste elemento.

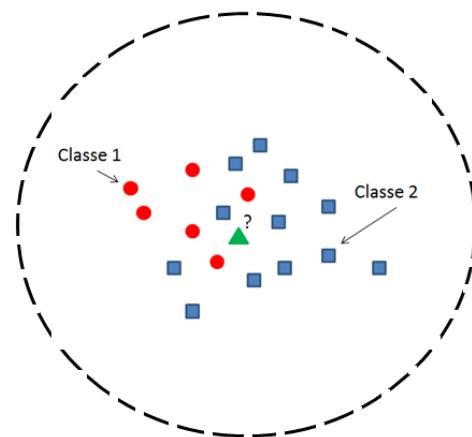


Figura 19 - Técnica kNN aplicada à amostra triangular.

4.1.4 Descrição de Dados Vetor de Suporte

Dentre os métodos *One Class* aplicados neste estudo, o método de descrição de dados vetor de suporte (Support Vector Data Description – SVDD) [TAX, 1999] [TAX, 2004] foi o único desenvolvido por Tax [TAX, 1999] [TAX, 2001] [TAX, 2004], com características muito

similares ao classificador SVM [CORTES e VAPNIK, 1995]. Este método é utilizado para obter as bordas ao redor dos dados do conjunto alvo, os padrões positivos, aplicados ao classificador no treinamento. Para esta tarefa é construída uma hiper-esfera em torno dos dados do conjunto de treinamento [TAX, 1999]. Através dos vetores de suporte são mapeados os dados para um novo espaço de características de maior dimensão, tornando a descrição dos dados mais flexível.

A hiper-esfera, definida na fase de treinamento, é uma estrutura descrita por um modelo $f(x; w)$, e se caracteriza por um centro a e um raio R . A partir da definição matemática da hiper-esfera é possível a classificação de novos objetos na fase de teste (verificando a possibilidade de estes objetos estarem dentro do espaço da hiper-esfera). O teste aplicado aos exemplares da base de teste consiste em mensurar a distância euclidiana da nova amostra ao centro da esfera, sendo maior que o raio da esfera, a amostra não é classificada como um exemplar da classe alvo, caso contrário, a amostra será incluída.

4.1.5 Máquina Vetor de Suporte

A Máquina Vector de Suporte (Support Vector Machine – SVM) é uma técnica estatística de aprendizagem de máquina desenvolvida por Vapnik em 1998 [CORTES, 1995] [VAPNIK, 1998]. Esta técnica apresenta como objetivo a geração de um hiperplano de separação entre as classes visando a maximização da distância euclidiana entre elas [VAPNIK, 2011]. Muitos trabalhos no campo de identificação e reconhecimento de assinaturas aplicam a técnica SVM, com a apresentação de bons resultados para a verificação de assinaturas com exemplares forjados [JUSTINO, 2001].

5 Metodologia

Neste capítulo é descrita a metodologia aplicada aos experimentos. Para o desenvolvimento da proposta são empregadas duas bases de imagens. A primeira desenvolvida durante o estudo e a segunda fornecida à comunidade através do ICDAR 2009 [BLANKERS et al, 2009]. Além disso, técnicas, ferramentas e métricas são descritas e aplicadas para tornar a análise estatística possível.

São detalhados também o ambiente de desenvolvimento *Matlab* [MATHWORKS, 2014], o *toolbox Prtools* [DUIN, 2007] e *dd_tools* [TAX, 2013], a técnica Matriz de Confusão [STORY, 1986] [STEHMAN, 1997] e o algoritmo utilizado para particionar os conjuntos de imagens conhecido como *k-fold* [DUDA et al, 2010]. Por fim, a Competição do ICDAR de 2009 é apresentada, bem como os resultados obtidos pelas técnicas de verificação *off-line* submetidas.

5.1 Base de Imagens

Neste trabalho foi criada uma base de imagens de assinaturas intitulada AMODA (Assinaturas Manuscritas *Off-line* de Dimensões Aleatórias). Este conjunto foi gerado com o objetivo de realizar os experimentos em uma base produzida por pessoas alfabetizadas com o emprego de grafias latinas, mas especificamente o português brasileiro. Esta ação deve-se a inexistência de bases, não encontramos nenhuma disponível, contendo imagens originais com tamanhos diferentes e imagens forjadas dos tipos aleatória e habilidosa (mais detalhes na subseção 5.1.1). Experimentos também são realizados com a base de assinaturas *off-line* disponibilizada pelo ICDAR 2009 [BLANKERS et al, 2009]. Uma breve explicação sobre essa base encontra-se na subseção 5.1.2.

5.1.1 A Base produzida

Uma base de imagens foi criada para ser aplicada aos experimentos deste trabalho. Esta criação foi consequência da dificuldade em encontrar bases de imagens com assinaturas originais em diferentes tamanhos. Para a criação da base contribuiu um grupo de 60 assinantes voluntários. O grupo é formado por pessoas que trabalham 8h por dia utilizando computadores, *tablets* ou celulares, apresentando pouca ou praticamente nenhuma interação diária com instrumentos de escrita, como lápis, caneta e papel. Essas pessoas demonstraram características similares na produção das assinaturas. Diferenciando-se apenas o sexo. Na maioria dos casos, as mulheres continuam escrevendo suas assinaturas utilizando letras cursivas e bem definidas

(imagens observadas nas letras a, b, c e d da Figura 20). Por sua vez, os homens que trabalham nas mesmas condições possuem assinaturas variantes (imagens apresentadas nas letras e, f, g e h da Figura 20). Alguns assinantes masculinos são desleixados com relação à definição da letra e ainda outros utilizam representações ilegíveis ou simplificações do seu nome. Independente do sexo, os assinantes vivenciaram os níveis de construção da escrita [COSTA, 2010] e dominam o sistema alfabético [FERREIRO, 2008] encontrado nas assinaturas. As letras, sílabas, encontros consonantais e dígrafos da língua portuguesa são as características gráficas que tornam a base representativa [SARMENTO, 2005]. Estes atributos são os primeiros indícios da separação entre as assinaturas de dois indivíduos.

Das assinaturas produzidas pelos 60 assinantes foram selecionados, para a produção de assinaturas de diferentes tamanhos, 10 indivíduos através dos seguintes critérios: definição das letras, legibilidade, e alta similaridade entre as assinaturas pertencentes ao mesmo indivíduo. Destes 10 assinantes foram solicitadas 12 assinaturas escritas em um tamanho menor e outras 12 dispostas em um tamanho maior, todos os dois tipos foram gerados considerando as 12 assinaturas informadas inicialmente como sendo a assinatura padrão do voluntário. Exemplos de imagem nos tamanhos solicitados aos assinantes podem ser visualizados na Figura 21

Erro! Fonte de referência não encontrada. As variações de tamanho foram geradas para apoiar o estudo da identificação de assinaturas genuínas de tamanhos diferentes e a rejeição das forjadas, através da simulação de situações como: comparações de assinaturas entre cheques e carteiras de identidade ou contratos e o cartão de autógrafo bancário ou cartorial. Em alguns destes documentos, como na carteira de identidade, a assinatura é escrita em uma área menor que a encontrada na maioria dos documentos. Exemplos de documentos com área de assinatura menor são a carteira de identidade e a de motorista, como dispostas na Figura 22.

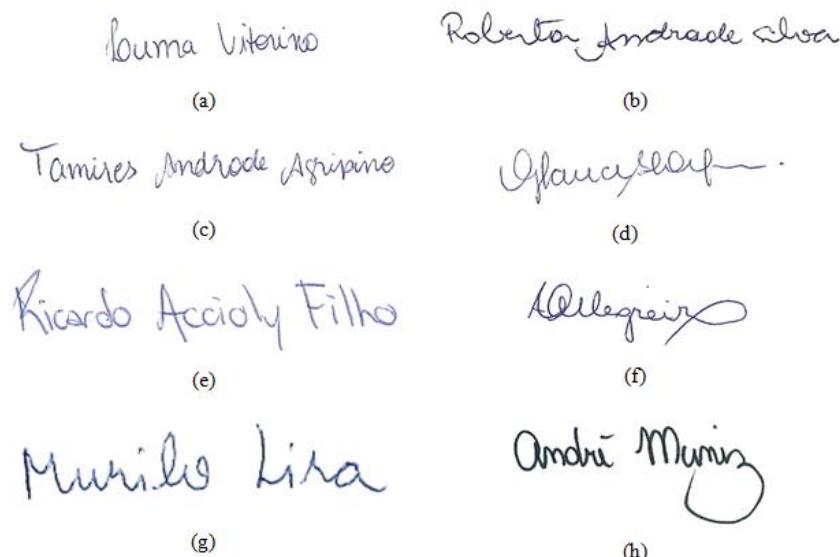


Figura 20 - Imagens de assinaturas produzidas por assinantes do sexo feminino (a - d) e do sexo masculino (e - h). Traços de letras mais elaboradas e bem definidas permitem a nitidez das sílabas e a separação de palavras.

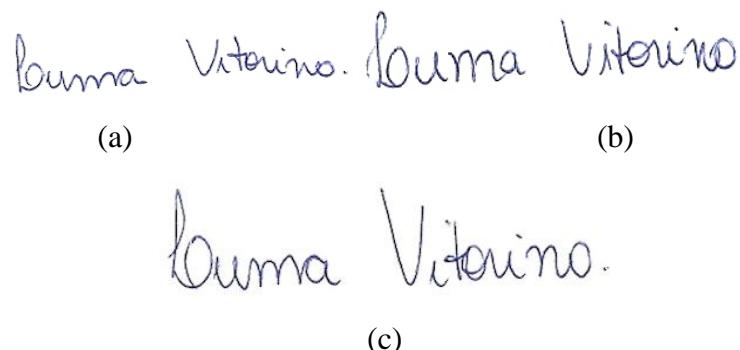


Figura 21 - Imagens de assinaturas produzidas em diferentes tamanhos. Assinatura pequena (a), assinatura de tamanho normal ou padrão (b) e assinatura grande (c). Estes exemplares foram projetados para analisar o comportamento da técnica na verificação de assinaturas em diferentes tamanhos.



Figura 22 - Assinatura em carteira de habilitação (aquisição via tablet) e de identidade (produção no próprio documento). A assinatura na habilitação é produzida em modo manuscrito e armazenada digitalmente, podendo ser redimensionada. Por sua vez, a assinatura na identidade é produzida no próprio documento, em uma área restrita que define o tamanho da assinatura. Nestes dois casos as assinaturas são apresentadas em uma versão menor do que o modo de escrita do assinante, na maioria dos exemplares.

Motivado pelo diferencial da base construída (assinaturas de diferentes tamanhos) e pela necessidade de comparação com outros trabalhos da literatura foram produzidas assinaturas

forjadas. Dos 60 assinantes, foram selecionados 11 para contribuir com 12 assinaturas forjadas habilidosas e 12 assinaturas aleatórias. Exemplos de assinaturas forjadas produzidas para a base podem ser vistas na Figura 23. Uma folha de testes foi definida para a produção da base. Através dessa folha foi possível recolher as 12 assinaturas de cada tipo (padrão, tamanho pequeno e grande, forjada habilidosa e aleatória) dos autores participantes. A base AMODA pode ser obtida no Anexo 1 deste estudo.

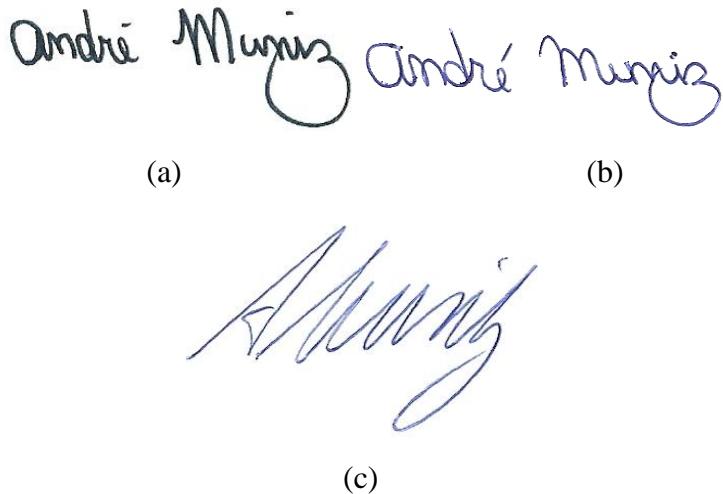


Figura 23 - Assinaturas da base AMODA para a verificação de assinaturas forjadas.

5.1.2 Base de Imagens ICDAR 2009

A base de imagens do *International Conference on Document Analysis and Recognition* (ICDAR) [BLANKERS, 2009] foi utilizada devido a grande representatividade da competição de verificação de assinaturas, proposta a cada dois anos pelo ICDAR; e pelo grande número de referências da base nos trabalhos da área. A utilização desta base foi resultado da necessidade de validar os experimentos em uma base de aplicação ampla e internacional. As imagens utilizadas pelo ICDAR foram produzidas por um grupo de colaboradores do *Netherlands Forensic Institute* (NFI) [BLANKERS et al, 2009], composto por 100 pessoas. Cada membro escreveu sua própria assinatura 12 vezes, tais assinaturas foram rotuladas como legítimas (exemplo na Figura 24). Além disso, cerca de 33 integrantes produziram assinaturas forjadas do tipo habilidosas a partir das assinaturas originais. Estes participantes forjaram seis assinaturas para cada assinatura verdadeira, exemplos de assinatura verdadeira e forjada podem ser observados nas letras *a* e *b*, respectivamente, da Figura 24. Através destes autores, assinaturas *off-line* e *on-line* foram produzidas. Assim, a base de dados utilizada possui um total de 1.898 assinaturas *off-line* para

treinamento. Para testes com assinaturas *off-line*, a base de dados está dividida em dois grupos: 940 imagens genuínas e 624 imagens forjadas [BLANKERS et al, 2009].

Para cada assinatura foi gerado um arquivo em separado. Nas assinaturas *off-line* nenhuma informação além da própria imagem (informação estática) foi considerada na análise [BLANKERS et al, 2009].

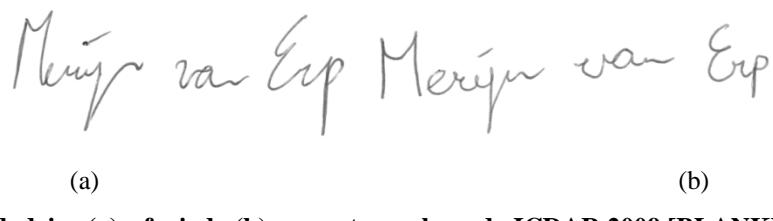


Figura 24 - Imagem verdadeira (a) e forjada (b) presentes na base do ICDAR 2009 [BLANKERS et al, 2009].

5.2 Ferramentas Utilizadas

O *Matlab* é um ambiente de desenvolvimento para aplicações matemáticas e de análise comparativa de valores. O emprego desta ferramenta na análise de imagem e na área de inteligência computacional é bem difundido. A sua escolha foi devido a grande necessidade de processamento matemático, contudo, outros motivos podem ser citados, tais como: alta capacidade de processamento, existência de bibliotecas e funções de processamento de imagem e de aprendizagem de máquina já desenvolvidas e o fato de ser uma ferramenta disponível para estudantes.

Toolboxes foram utilizados para tornar mais rápido o processo de desenvolvimento dos classificadores. O *Prtools* [DUIN, 2007] e o *dd_tools* [TAX, 2013] são bibliotecas que acopladas ao *Matlab* possibilitam a análise dos mais diversos classificadores OCC [TAX, 2001] da literatura, mais detalhes sobre os classificadores OCC no Capítulo 4.

O *Prtools* [DUIN, 2007] é definido como uma biblioteca de funções dedicada à área de reconhecimento de padrões estatísticos. A partir do *Prtools* foi gerado o *toolbox dd_tools* [TAX, 2013] que oferece ferramentas, classificadores e funções de avaliação para pesquisas na área de *One Class Classification*.

5.3 Técnicas, Métricas e Ferramentas de Avaliação

Toda avaliação necessita de métricas para definir a competência da técnica em resolver um dado problema. Na área da aprendizagem de máquina, resultados são aceitos através da aplicação de análises estatísticas sobre os valores resultantes. A utilização da análise estatística deve-se a comparação entre os resultados de pesquisas e as definições registradas na literatura.

Na verificação de assinaturas, técnicas são aplicadas para oferecer suporte à análise de resultados. A matriz de confusão é a mais disseminada, esta técnica é também conhecida como matriz de erros ou matriz de contingência [STEHMAN, 1997]. A matriz de confusão é a técnica que estabelece a relação entre o resultado esperado para um problema avaliado e o resultado obtido ao final do experimento. Para ilustrar a relação é gerada uma tabela (visualizada na Tabela 1).

Tabela 1 - Matriz de confusão e a relação entre a saída esperada e a saída obtida.

		Saída esperada	
		Condição Positiva	Condição Negativa
Saída obtida	Saída de teste Positivo	TP	FN
	Saída de teste Negativo	FP	TN

Na matriz de confusão é gerado, para cada imagem de indivíduo analisado, um dos quatro resultados de análise. As imagens dos indivíduos participantes são conhecidas como amostras. No primeiro bloco da tabela é apresentado o resultado *true positive* (TP). Em português esta sigla significa o verdadeiro positivo, ou seja, a quantidade de vezes que um resultado esperado foi obtido corretamente no experimento (definido pela Equação 3). Oposto ao TP está o *true negative* (TN). O verdadeiro negativo é um resultado positivo para a análise, pois o TN apresenta a quantidade de vezes que amostras rejeitadas para o estudo foram rejeitadas no experimento (definido pela equação 4).

Em oposição às saídas já apresentadas encontram-se os casos de erros da matriz de confusão. Abaixo do TP localiza-se o erro tipo I [SHERMER, 2011], que é conhecido como *false positive* (FP) (definido na Equação 5). Os falsos positivos são contabilizados quando o resultado esperado para uma amostra é a condição positiva e a resposta do experimento é negativa. O erro tipo II [SHERMER, 2011] é intitulado de *false negative* (FN) e neste caso, o experimento informa que os exemplares esperados como negativos são positivos (apresentado na Equação 6). De modo geral, os sistemas que aplicam a matriz de confusão buscam neste estudo: evitar ao máximo o erro do tipo II, porém tolerando o erro do tipo I.

$$TPR = \frac{TP}{TP+FN} \quad 3$$

$$TNR = \frac{TN}{FP+TN} \quad 4$$

$$FPR = \frac{FP}{FP+TN} \quad 5$$

$$FNR = \frac{FN}{TP+FN} \quad 6$$

Considerando os dois tipos de erros são obtidas duas taxas distintas. A primeira, conhecida como taxa de falsa rejeição (em inglês *False Rejection Rate – FRR*) [GIOT, 2012], definida pelo resultado do erro FP da matriz de confusão. O valor desta taxa é resultado da divisão entre a quantidade de amostras rejeitadas pela testadas, como apresentado na Equação 7. A segunda é a taxa de falsa aceitação (em inglês, *False Accept Rate – FAR*) [GIOT, 2012], sendo o resultado do erro FN da matriz. O valor resultante da FAR é a divisão entre a quantidade de amostras aceitas pela testadas, observe a Equação 8. A partir destas duas taxas, o desempenho da maioria dos sistemas de classificação pode ser avaliado.

A taxa *equal error rate* (EER) [GIOT, 2012] é a taxa de erro do sistema quando a taxa de erro FRR é igual à taxa de erro FAR [DEVROYE et al., 1996], ou seja, o ponto do gráfico que a curva dos dois erros sofre cruzamento. Esta taxa é apresentada na Equação 9 e demonstrada graficamente pela Figura 25. Esta métrica é definida como o erro total do sistema em alguns trabalhos0, por representar toda a base analisada com as taxas de erros diferentemente da curva ROC, que trata apenas dos casos positivos (verdadeiro e falso), assim uma das taxas que torna possível a comparação da acurácia dos classificadores.

$$FRR = \frac{\text{rejeitadas}}{\text{testadas}} * 100 \quad 7$$

$$FAR = \frac{\text{aceitas}}{\text{testadas}} * 100 \quad 8$$

$$EER = \begin{cases} \frac{FAR(t_1)+FRR(t_1)}{2} & \text{if } FAR(t_1) - FRR(t_1) \leq FRR(t_2) - FAR(t_2) \\ \frac{FAR(t_2)+FRR(t_2)}{2} & \text{else} \end{cases} \quad 9$$

$$t_1 = \max_{t \in S} \{t \mid FRR(t) \leq FAR(t)\}, \quad t_2 = \min_{t \in S} \{t \mid FRR(t) \geq FAR(t)\},$$

onde t é o instante no tempo máximo e mínimo que permite determinar os momentos t_1 e t_2 respectivamente, sendo t_1 o instante onde FRR é menor ou igual à FAR e t_2 o instante onde FRR é maior ou igual a FAR

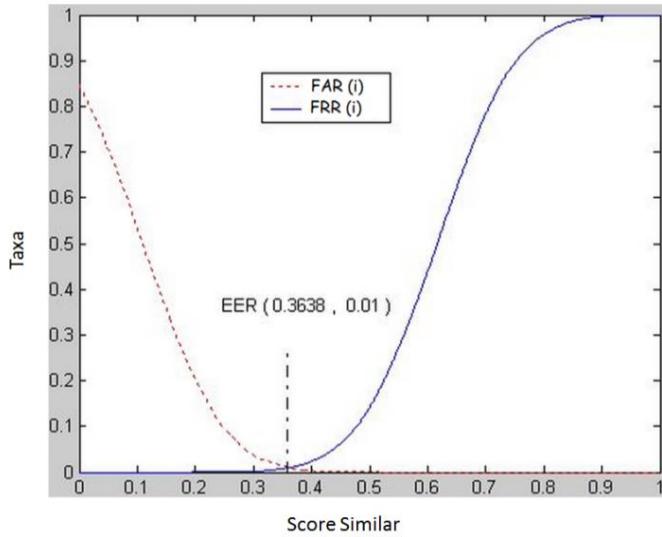


Figura 25 - Gráfico com as curvas pelo FAR e FRR. Estas resultam no ponto EER, local em que ocorre o cruzamento das funções.

Com base nos dados da matriz de confusão, outras métricas podem ser derivadas: acurácia, precisão e eficiência, detalhadas nas equações 10, 11 e 12, respectivamente. No ICDAR 2009, a avaliação dos classificadores estudados foi realizada com a análise do *Error Rate* (EER). Para efeitos de comparação, a EER é a taxa utilizada nos experimentos aplicados utilizando-se esta base.

$$ACC = \frac{(TP+TN)}{(P+N)} \quad 10$$

$$precision = \frac{TP}{TP+FP} \quad 11$$

$$EFF = \frac{(SENS+SPEC)}{2} \quad 12$$

Além da avaliação dos resultados dos experimentos por valores quantitativos, ferramentas são também utilizadas para comparar os resultados através de informação gráfica. A ferramenta mais comum é a curva *ROC* (*Receiver Operation Characteristic*) [FAWCELT, 2006]0. Inicialmente, a curva ROC foi desenvolvida para demonstrar as relações sinal-ruído, interpretando o sinal como verdadeiros positivos (sensibilidade) e o ruído, os falsos positivos (especificidade). Assim, a leitura da curva ROC é de um gráfico de sensibilidade ou taxas de verdadeiros positivos versus taxa de falsos positivos, como apresentado na Figura 26. Os gráficos ROC são bidimensionais, estando o eixo Y com os valores verdadeiros positivos e o eixo X preenchido com os valores da taxa falsos positivo.

O objetivo da ferramenta curva ROC é atingir a representação perfeita do experimento, considerando o conjunto de amostras analisado. O objeto de estudo tem sua avaliação

identificada pela localidade do ponto no gráfico. Quão mais próximo o ponto estiver do eixo Y, melhor será o resultado. Assim, a diagonal traçada do ponto (0,0) ao ponto (1,1) serve de direcionamento para identificar quais são os melhores resultados, estes localizados predominantemente acima da reta [FAWCELT, 2006].

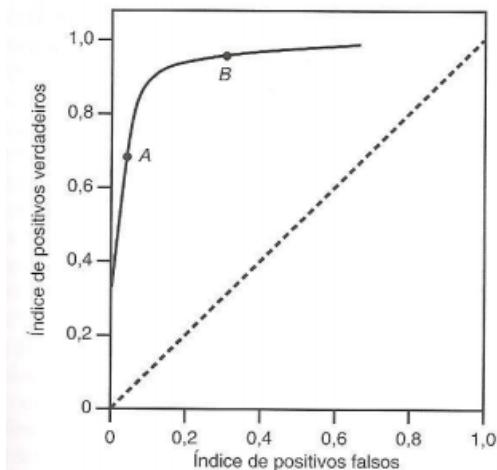


Figura 26 - Exemplo de Curva ROC [FAWCELT, 2006].

Através desta ferramenta, obtém-se a métrica de desempenho denominada área sob a curva ROC (AUC) [FAWCELT, 2006]. Esta área é um indicador bidimensional para comparação de desempenho entre os objetos de estudo. Neste gráfico o objeto que contiver a maior área terá o melhor comportamento médio (objetos representados no gráfico da curva ROC da Figura 27 em rosa e roxo). Porém, mesmo um objeto com menor área poderá ser melhor do que os demais que possuam com maior área (para FP entre 0 e 0.4, o objeto B apresenta melhor desempenho que o objeto A, em FP como 0.5, os dois objetos possuem o mesmo resultado, e de 0.5 em diante, o objeto A apresenta-se como melhor solução, estas definições podem ser visualizadas na Figura 27). Assim, uma análise para indicar os melhores resultados depende do valor desejado para a taxa em avaliação. Além deste, outros gráficos são apresentados pela literatura como, curva de custo, LC index e curva DET [MARTIN, 1997].

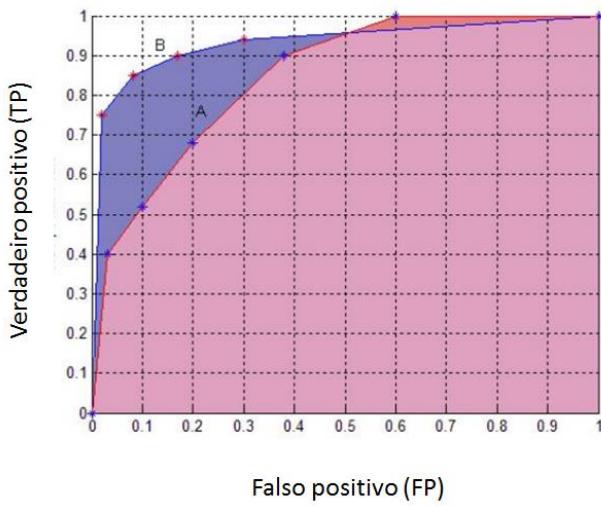


Figura 27 - Exemplo de medição da área sob a curva ROC.

A métrica utilizada no estudo do ICDAR 2009 [BLANKERS et al, 2009] foi a *Detection Error Trade-off curve* (DET) [MARTIN, 1997] (gráfico visualizado na Figura 28). Esta métrica é mais conhecida como a curva DET. Esta curva corresponde a um gráfico que estabelece uma relação entre as taxas de erro diferentemente da curva ROC. A curva é definida como a linha contendo os pontos onde $x = y$, ou seja, posição onde os erros FRR e FAR possuem o mesmo valor. Assim, o gráfico é gerado baseando-se na taxa EER. No gráfico DET o melhor resultado está localizado o mais próximo do eixo inicial (o ponto 0 (zero)). O mais próximo apresenta, assim, o menor erro de classificação. Esta métrica foi utilizada pelo ICDAR para definir o algoritmo de classificação da competição com o melhor EER nos problemas de verificação de assinaturas nos modos *off-line* e *on-line*.

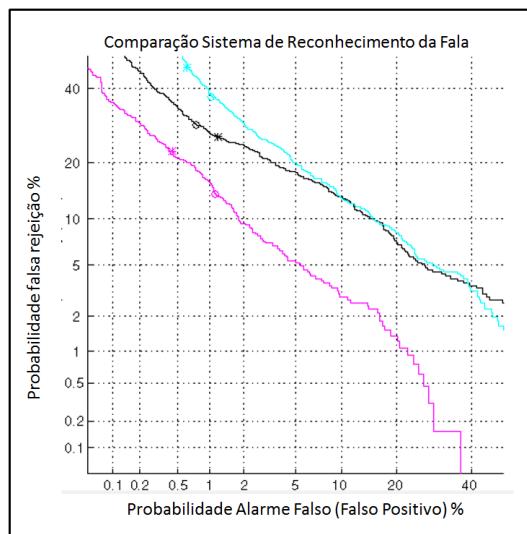


Figura 28 - Exemplo de aplicação da curva DET. Neste gráfico, o classificador representado pela linha rosa, possui a melhor classificação.

5.4 Experimentos

Os experimentos desenvolvidos para a verificação de assinaturas são divididos em quatro grupos. No primeiro estão os cenários de verificação analisados com a base do ICDAR 2009. Os cenários do primeiro grupo são analisados também com a base AMODA, estes experimentos são apresentados na Tabela 2. Cenários OCC (*One Class Classifier*) são apresentados no segundo grupo e na Tabela 3. No terceiro grupo, cenários OCC envolvendo assinaturas forjadas são analisados e apresentadas na Tabela 4. Por último, o grupo 4 apresenta cenários de classificação binária, envolvendo assinaturas originais com diferentes tamanhos de aquisição e forjadas dos tipos habilidosa e aleatória.

O primeiro grupo de experimentos foi desenvolvido para analisar os conjuntos de características e classificadores deste estudo aplicados em cenários de verificação envolvendo assinaturas forjadas nas duas bases (ICDAR 2009 e AMODA). Os experimentos do primeiro grupo analisados com a base ICDAR 2009 foram desenvolvidos com o intuito de estabelecer uma relação entre os resultados obtidos e a literatura. No segundo grupo são encontrados cenários OCC abrangendo as assinaturas verdadeiras com diferentes tamanhos de aquisição, problemática deste estudo. No terceiro conjunto são realizadas verificações em cenários OCC com assinaturas forjadas compondo o conjunto de teste do experimento.

Em todos os experimentos, as assinaturas verdadeiras ou genuínas são intituladas como *target* (em português, alvo) e as forjadas como *outlier* (fora da linha). Para as bases de assinaturas, cada participante contribuiu com assinaturas *target* e *outlier*, gerando um conjunto de assinaturas nomeado de classe da base. Uma classe é um conjunto de exemplares com características em comum, neste estudo temos os seguintes exemplos: a classe autor das assinaturas, a classe de assinaturas originais, a classe de assinaturas forjadas, entre outros. Todo classificador através das características de um conjunto de assinaturas gera uma função de aceitação para a classe do conjunto estudado. Nos classificadores OCC, a função de aceitação é gerada apenas pelos exemplares originais de uma classe, diferentemente do classificador SVM, definido através de duas classes. Na fase de testes, a classificação correta das assinaturas é resultante do grau de representatividade da classe das assinaturas pela função de aceitação. Quão melhor produzida for à função de aceitação, maior será a rejeição aos exemplos que não pertencem à classe e a aceitação aos pertencentes à classe. A qualidade da função de aceitação dependerá da representatividade do conjunto de exemplares da classe, condicionado ao conjunto de características definido. Nos experimentos são encontradas duas situações: a assinatura classificada como da classe (positivo) ou não sendo da classe (negativo). Para aprender e classificar assinaturas é necessário dividir a base em grupos para aplicação destas em cada cenário.

Tabela 2 - Grupo 1: cenários de verificação de assinaturas envolvendo assinaturas originais e forjadas nas bases ICDAR 2009 e AMODA.

Cenários de testes	Treino	Teste	Base
1	<i>Target:</i> Assinaturas Genuínas <i>Outlier:</i> Forjadas	<i>Target:</i> Assinaturas Genuínas <i>Outlier:</i> Forjadas	ICDAR 2009
2	<i>Target:</i> Assinaturas Genuínas <i>Outlier:</i> Forjadas Habilidosas	<i>Target:</i> Assinaturas Genuínas <i>Outlier:</i> Forjadas Habilidosas	AMODA
3	<i>Target:</i> Assinaturas Genuínas <i>Outlier:</i> Forjadas Aleatórias	<i>Target:</i> Assinaturas Genuínas <i>Outlier:</i> Forjadas Aleatórias	AMODA

Tabela 3 - Grupo 2: cenários OCC envolvendo as assinaturas originais da base ICDAR 2009 e as assinaturas com diferentes tamanhos de aquisição da base AMODA.

Cenários de testes	Treino	Teste	Base
4	Assinatura Normal	Assinatura Normal	ICDAR 2009
5	Assinatura Normal	Assinatura Normal	AMODA
6	Assinatura Maior	Assinatura Maior	AMODA
7	Assinatura Menor	Assinatura Menor	AMODA
8	Assinatura Normal	Assinatura Normal, Maior e Menor	AMODA
9	Assinaturas Normal e Maior	Assinatura Normal, Maior e Menor	AMODA
10	Assinaturas Normal e Menor	Assinatura Normal, Maior e Menor	AMODA
11	Assinaturas Normal, Maior e Menor	Assinatura Normal, Maior e Menor	AMODA
12	Assinatura Maior	Assinatura Normal, Maior e Menor	AMODA
13	Assinatura Menor	Assinatura Normal, Maior e Menor	AMODA

Tabela 4 - Grupo 3: cenários OCC envolvendo assinaturas forjadas como *outlier*.

Cenários de testes	Treino	Teste	Base
14	Assinaturas Genuínas	<i>Target:</i> Assinaturas Genuínas <i>Outlier:</i> Forjadas	ICDAR 2009
15	Assinaturas Genuínas	<i>Target:</i> Assinaturas Genuínas <i>Outlier:</i> Forjadas Habilidosas	AMODA
16	Assinaturas Genuínas	<i>Target:</i> Assinaturas Genuínas <i>Outlier:</i> Forjadas Aleatórias	AMODA
17	<i>Target:</i> Assinaturas Normal e Maior	<i>Target:</i> Assinaturas de tamanho Normal, Maior <i>Outlier:</i> Forjadas Habilidosas.	AMODA
18	<i>Target:</i> Assinaturas Normal e Menor	<i>Target:</i> Assinaturas de tamanho Normal, Menor <i>Outlier:</i> Forjadas Habilidosas.	AMODA
19	<i>Target:</i> Assinaturas Normal e Maior	<i>Target:</i> Assinaturas de tamanho Normal, Maior <i>Outlier:</i> Forjadas Aleatórias.	AMODA
20	<i>Target:</i> Assinaturas Normal e Menor	<i>Target:</i> Assinaturas de tamanho Normal, Menor <i>Outlier:</i> Forjadas Aleatórias.	AMODA

Tabela 5 - Grupo 4: com assinaturas originais com diferentes tamanhos de aquisição e assinaturas forjadas habilidosas e aleatórias.

Cenários de testes	Treino	Teste	Base
21	<i>Target:</i> Assinaturas Genuínas com Diferentes tamanhos de aquisição <i>Outlier:</i> Forjadas habilidosas	Assinaturas Genuínas com Diferentes tamanhos de aquisição Forjadas habilidosas	AMODA
22	<i>Target:</i> Assinaturas Genuínas com Diferentes tamanhos de aquisição <i>Outlier:</i> Forjadas Aleatórias	Assinaturas Genuínas com Diferentes tamanhos de aquisição	AMODA

		Forjadas Aleatórias	
--	--	---------------------	--

No grupo 1 (Tabela 2) a verificação das assinaturas é baseada na classificação binária, exemplares *target* e *outlier* são aplicados nas etapas de treinamento e teste. Estudos analisando a verificação das assinaturas com diferentes tamanhos de aquisição não foram encontrados, por este motivo foram projetados alguns dos cenários do grupo 2 (apresentados na Tabela 3) com o objetivo de gerar resultados para os conjuntos de características e classificadores OCC das seções 4 e 5 no campo da verificação de assinaturas. Exemplares com tamanhos de aquisição menores, maiores e normais foram obtidos dos participantes da base. Todo indivíduo possui um tamanho de escrita das letras característico, ao qual intitulamos de normal, com o objetivo de estudar as variações dos tamanhos de aquisição e a separabilidade entre estes exemplares e as assinaturas de tamanho normal foram coletados exemplares produzidos em tamanho menor e maior. Os exemplares com variações nos tamanhos de aquisição são nomeados de assinaturas originais e compõem a base AMODA. A partir da base ICDAR 2009 obtem-se um único tamanho de aquisição das assinaturas, a normal.

Por último, cenários OCC envolvendo assinaturas *outliers* são somadas as assinaturas *target* na fase de teste com o objetivo de averiguar a separabilidade entre as duas classes de assinaturas. Em alguns cenários são incluídas assinaturas com variações de tamanhos de aquisição (AMODA).

Na verificação de assinaturas quanto mais exemplos *target* forem aceitos (reconhecidos corretamente no processo) e mais assinaturas *outlier* forem rejeitadas, melhor é o desempenho do sistema de verificação. O objetivo destes experimentos é identificar a tupla conjunto de características-classificador com a maior taxa na relação entre assinaturas *target* e *outlier*, resultando na maior separabilidade entre as duas classes para a comparação com a literatura e o melhor resultado dentre os experimentos avaliadores dos diferentes tamanhos de aquisição.

5.5 Configuração dos Classificadores

Nesta seção são apresentadas as configurações definidas para os classificadores SVM e OCC aplicadas neste estudo. Os valores aplicados foram definidos empiricamente, apenas com base na descrição dos tutoriais que acompanham os *toolboxes*.

- Classificador SVM

O classificador binário SVM desenvolvido por [CORTES e VAPNIK, 1998] e definido na seção 4.3, foi configurado neste estudo com a função polinomial acompanhado pelo parâmetro valendo 2.

- Função Gaussiana

O classificador fornecido pelo *toolbox dd_tools* permite a aplicação da função de densidade gaussiana em cenários OCC, e aplica a distância Mahalanobis por padrão. Algumas configurações foram experimentadas, os melhores resultados foram encontrados com o parâmetro *fracrej* definido como 0.1.

- Janela de Parzen

O descritor de dados parzen encontrado no *toolbox dd_tools* aplica uma função de parzen em cada característica individual no conjunto de dados, multiplicando os resultados para a estimativa da densidade final. Como parâmetro o *fracrej* apresentou variações nos diferentes cenários avaliados.

- kNN

O kNN é um classificador OCC baseado em descrição dos dados como definido na seção **kNN4.1.3** e calcula a distância de um exemplar para um número k de vizinhos. Para este estudo foi definido para k o valor 3, o método *kappa* (distância euclidiana para os k vizinhos mais próximos), e o parâmetro *fracrej* foi definido diferentemente para cada cenário e cada conjunto de exemplares aplicado. Os valores definidos foram baseados na configuração mais aplicada na literatura.

- SVDD

O classificador SVDD por sua vez, como definido na seção 4.1.4, é o descritor de dados de vetor de suporte. Este classificador utiliza a função gaussiana como método e com o parâmetro *fracrej* variando de modo igual ao classificador kNN.

5.6 Estimação da taxa de erro – Particionando o conjunto de dados

Nos experimentos gerados são utilizados métodos para estimar a taxa de erro favorecendo a capacidade de aprendizagem e verificação dos classificadores. Os métodos de estimação do erro são baseados na partição do conjunto de dados e definem o conjunto de amostras de um experimento. Esta afirmação é devida as características de cada método [RUSSELL and NORVIG, 2003][DEMSAR, 2006]. Neste estudo é aplicado o método de validação cruzada k -fold [DUDA et al, 2010] devido às bases possuírem poucas imagens.

Com k igual a 3, esta abordagem divide a base de imagens em 3 conjuntos de imagens. Através do número k são determinadas k experiências definindo $k-1$ conjuntos como base de treino e 1 conjunto como base de teste. Com o aumento do número de k , os conjuntos de treino e teste sofrem permutações como visto na Figura 29. De cada experiência é definida a taxa de erro por classe da equação 13, ao final das k experiências é definida a taxa de erro global, como visto na equação 14, do sistema de verificação de assinaturas analisado, apresentando a capacidade de classificação das técnicas.

$$E_i = FP + FN \quad 13$$

$$E = \frac{1}{N} \sum_{i=1}^N E_i \quad 14$$

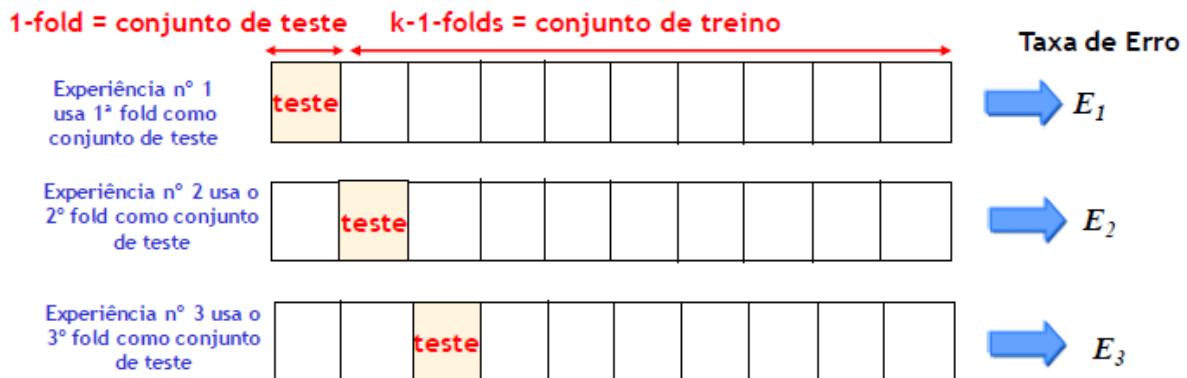


Figura 29 - Exemplo de aplicação do k -fold em conjunto de imagens.

5.7 Resultado ICDAR 2009

O ICDAR 2009 [BLANKERS et al, 2009] gerou um ranking dos algoritmos de classificação avaliados na competição. Esta conferência teve como objetivo identificar o classificador de melhor resultado no campo da verificação de assinaturas nos modos *off-line* e *on-line*. Por este estudo analisar assinaturas off-line serão apresentados apenas os resultados off-line. Dos muitos classificadores submetidos apenas 8 (oito) produziram bons resultados. O vencedor foi o classificador 13. Este classificador foi desenvolvido pelos cientistas A. Hassaine e E. Decencière do Centro de Morfologia Matemática da França. Este classificador atingiu a taxa EER 9,15. Os demais valores EER podem ser vistos na Tabela 6. Com os valores EER foi gerada a curva DET visualizada na Figura 30.

Tabela 6- Avaliação dos sistemas apresentados no ICDAR 2009 para a análise *off-line*.

Número de registro na competição	EER off-line com a base NFI (%)
13	9.15

14	15.50
18b	15.78
18^a	16.10
1	18.27
16	23.0
17	41.12
15	43.02

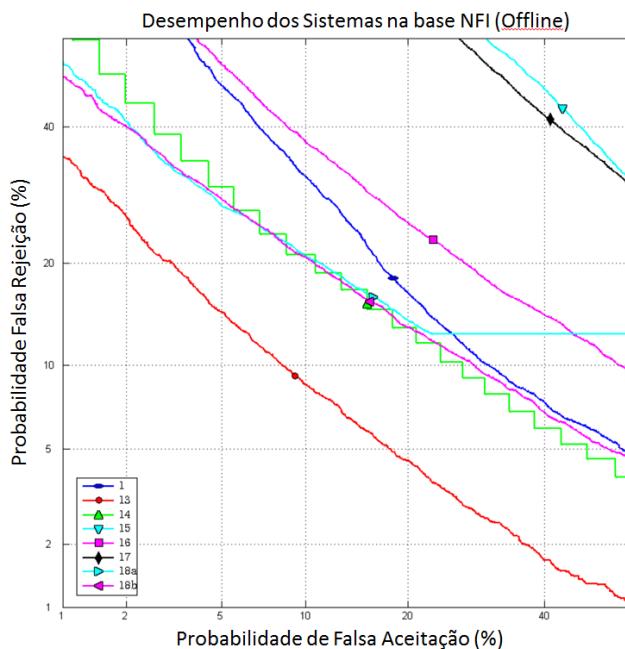


Figura 30- Curva DET dos classificadores off-line do ICDAR 2009. A curva vermelha mais baixa é o classificador 13, apresentando a melhor taxa de classificação [BLANKERS et al, 2009].

6 Conjuntos de Características Propostos para a Verificação de Assinaturas

Este capítulo apresenta os conjuntos de características utilizados na verificação de assinaturas com diferentes tamanhos de aquisição. As técnicas de pré-processamento utilizadas, as características extraídas e os classificadores que foram definidos com base na literatura [IMPEDOVO, 2008] [IMPEDOVO, 2007] [SOUZA, 2009]. Cada conjunto foi projetado com o intuito de analisar a capacidade de aprendizagem de um grupo de características, e a capacidade de generalização dos classificadores selecionados a partir das características definidas através dos experimentos elaborados.

6.1 Conjunto de Características 1

O conjunto de características 1 foi definida para tratar as características globais da imagem [IMPEDOVO, 2008]. As imagens após o processo de aquisição estão em tons de cinza, por este fato a técnica de pré-processamento conhecida como binarização é aplicada sobre as imagens, por conseguinte são extraídas das imagens binárias as características globais [GONZALEZ, 2010]. O grupo de características globais definido originou-se das características calculadas através dos pixels extraídos das assinaturas e que são referenciadas pela literatura [PATIL e PATIL, 2013] [JANA et al, 2014] [KAPIL, 2014]. Para cada assinatura analisada um vetor de características é gerado. As características globais analisadas são apresentadas na seção 3.1.1. A técnica de pré-processamento, as características globais e os classificadores que compõem o conjunto de características 1 são vistos na Figura 31.

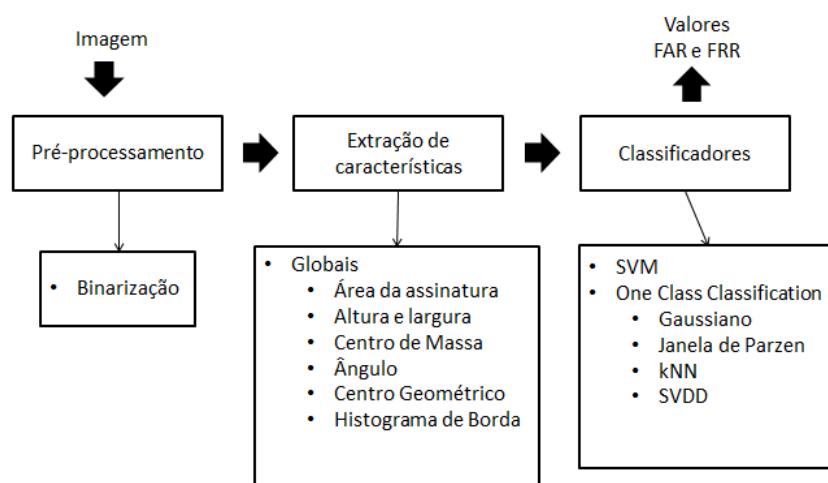


Figura 31 – Conjunto de Características 1, composta pelas características globais apresentadas na seção 3.1.1.

6.2 Conjunto de Características 2

O sistema de Verificação de Assinatura que possui o conjunto de características 2 diferencia-se do conjunto 1 por analisar características locais [IMPEDOVO, 2008] [IMPEDOVO, 2007]. As imagens são pré-processadas e classificadas de modo semelhante as do Conjunto de Características 1. As características locais são geradas também a partir dos pixels da assinatura e definidas de acordo com a literatura apresentada nas seções 2.3 e 3.1.2. Apenas duas características foram selecionadas para o conjunto de características locais com o objetivo de identificar a contribuição dos loops na verificação de assinaturas. A densidade dos pixels foi também adicionada neste conjunto para apoiar a verificação nas situações em que as assinaturas não possuem loops. As técnicas, as características e os classificadores que compõem o conjunto de características 2 são vistos na Figura 32.

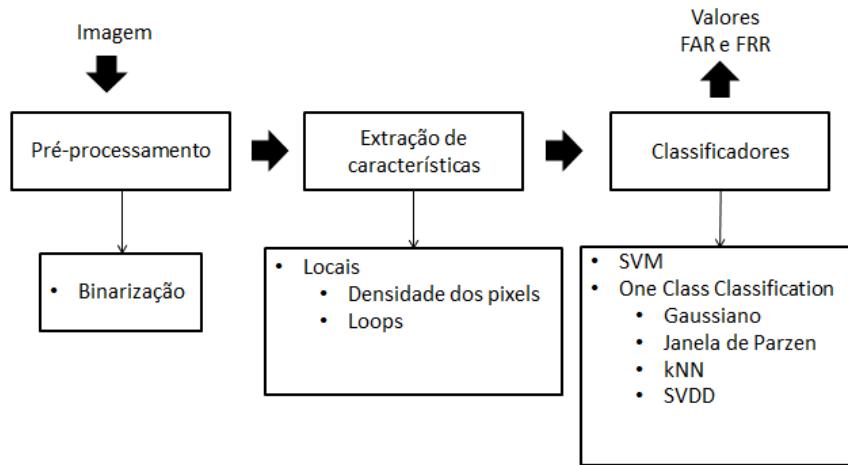


Figura 32 – Conjunto de Características 2, composta por as características locais.

6.3 Conjunto de Características 3

Através do trabalho de [SOUZA, 2009] foram encontrados resultados para o sistema de verificação de assinaturas com a aplicação da extração das características CPDP [FANG, 2002], e com a abordagem do conjunto de distâncias definidos a partir do vetor de características. Para identificar o desempenho do estudo de [SOUZA, 2009] com a aplicação de assinaturas originais de tamanhos diferentes foi criado o conjunto de características 3.

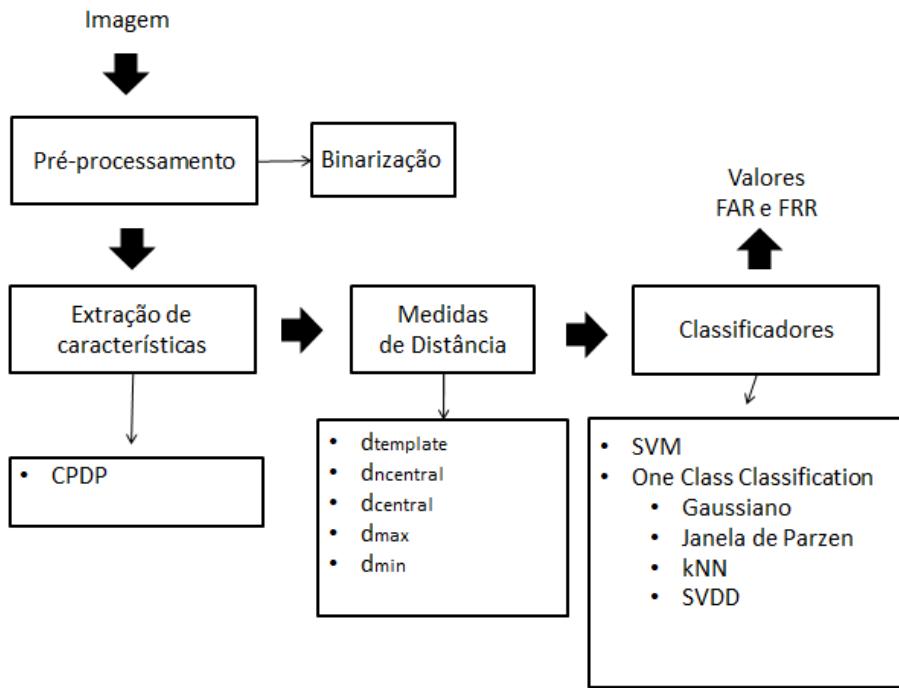


Figura 33 – Conjunto de Características 3, projetado para simular a Arquitetura 1 de [SOUZA, 2009].

As etapas do conjunto de características 3 (Figura 33) podem ser descritas como segue:

1. As imagens adquiridas são binarizada [GONZALEZ, 2010];
2. Na etapa de extração de características são gerados os vetores de características CPDP, apresentados nas seções 2.3.1 e 3.1 [FANG, 2002];
3. Para cada método de extração de características são calculadas as cinco distâncias por imagem;
4. As distâncias geradas são submetidas a cada um dos classificadores.
5. As áreas sob a curva ROC são construídas e calculadas para as distâncias de cada classificador.

6.4 Conjunto de Características 4

No conjunto de características 4 é analisada outra técnica de extração de características estudada por [SOUZA, 2009], a *Shadow Code* [SABOURIN, 1993] [JUSTINO et al, 2001]. De modo similar ao Conjunto de Características 3 é gerado um conjunto de distâncias a partir do vetor de características definido pela técnica *Shadow Code*. As distâncias definidas são aplicadas aos classificadores apresentados na Figura 34.

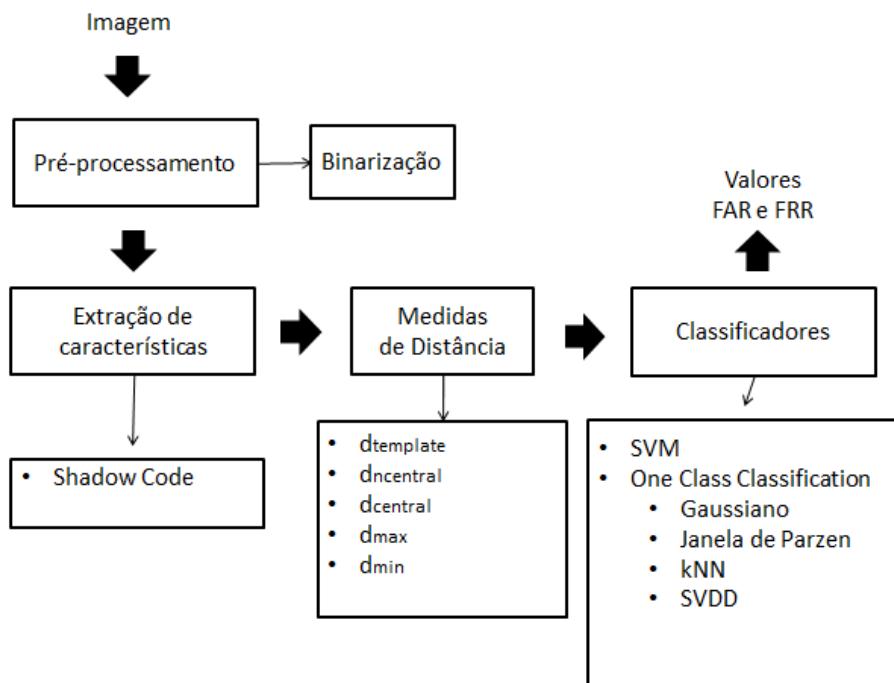


Figura 34 - Conjunto de Características 4, com a inclusão da característica shadow code.

7 Análise dos Resultados

Neste capítulo são apresentados os resultados obtidos para os quatro grupos experimentais avaliados, com assinaturas verdadeiras, com diferentes tamanhos de aquisição, forjadas habilidosas e aleatórias nos quatro conjuntos de características propostos. Os experimentos foram executados com as assinaturas apresentadas no Anexo 1 e com as do ICDAR 2009. As imagens do Anexo 1 são divididas em 11 classes compostas na mesma proporção por assinaturas verdadeiras e forjadas e outras 10 classes compostas por assinaturas nos tamanhos normal, menor e maior, com 12 assinaturas cada. As assinaturas foram organizadas em grupos através do método de partição *k-fold* com o valor de *k* igual a 3. A cada experimento são coletados 3 resultados obtidos da variação do conjunto de treinamento e testes proporcionada pela técnica *k-fold*.

Os grupos 1 e 4 apresentam como principal característica a classificação binária, com exemplares de assinaturas verdadeiras e forjadas nas fases de treinamento e teste. Por este motivo, os classificadores SVM e demais OCC selecionados são submetidos para a análise. A presença de assinaturas forjadas na verificação de assinaturas é considerada rara, devido a um bom percentual de ausência deste tipo de assinatura na prática, pois é possível produzir um padrão de assinatura verdadeira com um grupo pequeno de exemplares genuínos, contudo é impossível o estabelecimento de um padrão de assinaturas forjadas na prática devido a variabilidade deste tipo de exemplar. Nos grupos 2 e 3 encontram-se cenários de classificação OCC, contendo apenas assinaturas verdadeiras na fase de treinamento, por este motivo apenas os classificadores OCC selecionados (gaussiano, parzen, *kNN* e SVDD) foram aplicados nestes cenários. A seguir são apresentados os resultados extraídos de cada grupo de cenários.

7.1 Conjunto de Características 1

Como apresentado no capítulo sobre os Conjuntos de Características, o primeiro Conjunto analisa as características locais das assinaturas. Dependendo do tipo do experimento (com assinaturas genuínas e forjadas / apenas com assinaturas de diferentes tamanhos de aquisição / cenários OCC) são obtidas a taxa de assinaturas genuínas rejeitadas e a de forjadas aceitas. As taxas resultantes são definidas para cada um dos cenários experimentais, bem como para cada um dos classificadores selecionados. Todas as taxas obtidas foram geradas com a base disponibilizada pelo ICDAR 2009 e pelo Anexo 1 e são apresentadas nas Tabela 7, 8, 9 e 10.

A Tabela 7 apresenta o desempenho dos classificadores *One Class Classifier* e SVM nos três cenários do grupo 1. Os três cenários possuem como característica a presença de assinaturas *target* e *outlier* nas fases de treinamento e teste, por este motivo, para cada classificador são

geradas as taxas de erro na verificação de assinaturas *target* (FAR) e as taxas de assinaturas *outliers* aceitas (FRR).

Através das taxas geradas, três resultados podem ser compreendidos em cada cenário analisado: o melhor classificador em cada cenário, o classificador com a melhor taxa de verificação das assinaturas originais e o classificador com o pior rendimento. A primeira análise (objetivo 1) – melhor classificador em cada cenário – almeja definir o classificador com o melhor equilíbrio no *trade-off* entre as taxas de erro falsa aceitação na verificação de assinaturas originais e falsa rejeição na verificação de forjadas. O segundo resultado (objetivo 2) visa apresentar o classificador com a melhor taxa na verificação de assinaturas originais, este tipo de resultado foi definido com o objetivo de identificar o grau de representatividade das assinaturas originais pelas características do conjunto estudado. E por último, o classificador com o pior rendimento (objetivo 3) foi definido para os cenários estudados. Um classificador definido com o pior resultado apresenta alta taxa de falsa rejeição das assinaturas forjadas e alta taxa de falsa aceitação das assinaturas originais, devido aos altos valores o classificador deve ser descartado.

A partir dos resultados da Tabela 7 e dos resultados–objetivos para cada grupo de cenários, o classificador **SVM** apresentou em média as melhores taxas para o objetivo 1. Com a melhor relação para o *trade-off* analisado neste estudo. De modo semelhante, o objetivo 2 também apresentou as menores taxas na verificação de assinaturas originais no classificador **SVM**.

O último objetivo, a identificação do classificador com o pior rendimento, foi atribuído ao classificador **kNN**. Os resultados obtidos com o kNN apresentaram baixa separabilidade entre os dois grupos de assinaturas devido à base representatividade proporcionada pelo conjunto de características 1 não ser contornada pela distância euclidiana. As características analisadas neste conjunto compõem um pequeno grupo e demonstraram ser fracas para uma separação natural – valores de características muito próximos entre as assinaturas do mesmo grupo e muito distantes dos pertencentes aos outros grupos – dos grupos de assinaturas, pois muitas assinaturas produzidas por outros autores possuem valores próximos nestas características. Com o valor do elemento k definido como 3 – considerando a quantidade de assinaturas por classe aplicadas no classificador – sendo compreendido como um bom valor para o número de assinaturas analisadas por vez, e sem a aplicação de outros artifícios para a tomada de decisão na verificação, o resultado do *kNN* apresentou um alto grau de confusão nestes cenários.

Tabela 7 - Resultados para o grupo de experimento 1 no conjunto de características 1.

	Gauss		Parzen		kNN		SVDD		SVM	
Cenário	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR

1	13.02	86.98	49.99	51.31	52.97	61.72	50.00	50.73	48.95	0.26
2	33.87	67.14	49.99	51.62	49.90	60.50	50.00	50.00	21.48	30.20
3	31.07	68.93	49.99	51.62	49.90	60.50	50.00	50.00	20.57	79.42

Na Tabela 8, treze experimentos de classificação OCC foram realizados analisando a influencia do tamanho de aquisição das assinaturas aplicadas sob as mesmas características e classificadores do grupo de experimentos da Tabela 7, exceto o classificador SVM eliminado dos grupos de cenários 2 e 3 devido à fase de treinamento estudar apenas uma classe. Nestes experimentos apenas assinaturas *target* são aplicadas nas fases de treinamento e teste. Considerando as taxas obtidas na Tabela 8 pôde-se definir o melhor classificador médio nos treze cenários através da menor taxa de erro de assinaturas *target* aceitas. Na maioria dos cenários, o classificador gaussiano apresenta o melhor desempenho, ou seja, a mais baixa taxa FAR com as melhores taxas de erro entre 4% e 13%. Considerando a determinação da taxa média entre os resultados obtidos para FAR novamente o classificador **gaussiano** obteve a melhor taxa, de valor igual a 38,51%. Apesar de o classificador SVM apresentar os melhores resultados na Tabela 7, este não foi aplicado nos treze experimentos de classificação OCC da Tabela 8 devido à ausência de assinaturas forjadas nas fases de treinamento e teste nestes cenários, como relatado no início deste capítulo.

Tabela 8 - Resultados para o grupo de experimento 2 no conjunto de características 1.

	Gauss	Parzen	kNN	SVDD
Cenário	FAR	FAR	FAR	FAR
4	22.91	97.39	70.31	99.47
5	5.81	87.30	57.22	100
6	51.41	96.87	47.50	87.50
7	71.72	99.90	86.32	87.50
8	4.00	81.54	48.63	6.25
9	13.81	99.90	41.74	50.00
10	49.56	86.32	70.01	50
11	31.25	98.28	68.99	78.22
12	74.90	96.77	84.47	75
13	59.96	100	71.72	81.44
Média	38.51	94.42	64,69	71,54

Na Tabela 9 são apresentados os resultados com o mesmo conjunto de características e classificadores das Tabela 7 e 8. O conjunto de cenários da Tabela 9 analisa a classificação OCC, considerando as assinaturas forjadas como exemplares *outlier* na fase de teste. Alguns destes cenários experimentais analisam as assinaturas originais com diferentes tamanhos de aquisição. Destes experimentos podemos definir os mesmos três objetivos apresentados na Tabela 7. O classificador com a melhor taxa de verificação de assinaturas, atendendo o *trade-off* anteriormente apresentado foi o classificador **gaussiano**, este mesmo classificador apresentou as melhores taxas de FAR nos experimentos envolvendo as duas bases devido à boa separabilidade da função mesmo com um conjunto de características pouco representativo. Dentre estes classificadores, o **kNN** apresentou o mais baixo desempenho da relação assinaturas *targets* rejeitadas e *outliers* aceitas. O desempenho obtido neste grupo de cenários foi similar ao apresentado pelo classificador na Tabela 7, apresentando como diferencial ausência de assinaturas forjadas na fase de treinamento em alguns cenários OCC, entretanto a aplicação de uma única categoria de assinatura na determinação da função não resultou na boa separabilidade entre os grupos de assinaturas de teste, mais uma vez devido à alta proximidade dos valores numéricos pertencentes às características das assinaturas verdadeiras e forjadas que compõem o conjunto de teste.

Tabela 9 - Resultados para o grupo de experimentos 3 no conjunto de características 1.

Cenário	Gauss		Parzen		kNN		SVDD	
	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
14	13.03	86.97	49.99	51.31	52.97	61.72	50	50.27
15	32.94	67.06	49.99	51.59	49.90	60.48	50	50
16	31.27	68.84	49.99	51.59	49.90	60.48	50	50
17	43.75	56.25	48.43	50.01	57.81	31.35	50	50
18	15.62	84.37	50.0	50.0	37.50	50.12	50	50
19	34.37	65.62	49.99	51.57	39.06	50.10	50	50
20	39.06	60.93	50.0	50.0	43.75	43.87	50	50

Por fim, os cenários do grupo 4 estudam a verificação com assinaturas originais com diferentes tamanhos de aquisição e as forjadas habilidosas e aleatórias aplicadas nas fases de treinamento e teste. Semelhantemente a Tabela 7, os resultados de melhor classificador para o *trade-off*, melhor classificador para a verificação de assinaturas *target* e o pior classificador são determinados na Tabela 10. Para o primeiro objetivo, o classificador **SVM** apresentou o melhor resultado no *trade-off* entre as taxas de erro falsa aceitação de assinaturas originais e falsa

rejeição de assinaturas forjadas habilidosas. O classificador **gaussiano** apresentou o melhor resultado para o *trade-off* envolvendo as assinaturas forjadas aleatórias, pois como relatado anteriormente, rejeitar uma assinatura verdadeira é preferível a aceitar uma assinatura forjada. Para o objetivo: melhor classificador para a verificação de assinaturas *target*, o classificador SVM apresentou resultados entre 8% e 17% similares a resultados da literatura de verificação de assinaturas. O classificador **SVM** apresentou nestes resultados o pior desempenho, com altas taxas de falsa aceitação e falsa rejeição.

Tabela 10 - Resultados para o grupo de cenários 4 com o grupo de características 1.

	Gauss		Parzen		kNN		SVDD		SVM	
Cenário	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
21	44.37	15.62	50.06	42.23	46.44	35.10	50.0	50.0	16.67	8.33
22	36.56	23.43	50.04	45.31	43.98	32.81	50.0	50.0	8.33	58.33

7.2 Conjunto de Características 2

No conjunto de características 2 são avaliados os mesmos classificadores do conjunto 1, contudo aplicados ao conjunto de características globais. As taxas obtidas foram novamente geradas com as bases do Anexo 1 e do ICDAR 2009 e são apresentadas nas tabelas 11, 12, 13 e 14.

Novamente o *trade-off* entre assinaturas *target* e *outlier* pode ser observado nos 3 cenários da Tabela 11. Nestes experimentos novamente o classificador SVM apresentou o melhor resultado para a relação entre os dois tipos de assinaturas (originais e forjadas), devido à busca de uma menor taxa de erro de falsa rejeição ser mais forte do que a menor taxa de falsa aceitação nos sistemas de verificação. Ao analisar apenas a taxa de *targets* rejeitadas a **função gaussiana** apresentou os resultados mais significantes para a separabilidade entre as classes de assinaturas de originais e forjadas. E como classificador de pior desempenho, o **kNN** novamente apresentou os piores resultados para o *trade-off*.

Tabela 11 - Resultados para o grupo de experimentos 1 no conjunto de características 2.

	Gauss		Parzen		kNN		SVDD		SVM	
Cenário	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
1	23.70	76.31	50.0	50.0	85.86	47.66	50.00	50.27	40.10	30.72
2	37.74	62.25	49.94	56.25	56.10	59.04	50.00	50.00	50	0
3	9.93	70.03	43.75	50.06	47.26	37.55	50.0	50.0	58.59	41.40

Na Tabela 12 são apresentados os resultados para os mesmos cenários da Tabela 8 no Conjunto de Características 1. Mais uma vez a **função gaussiana** apresentou a melhor representatividade das assinaturas *targets*, atingindo as melhores taxas entre 2,74% e 19,72%. Estas variações dependem do cenário e conjunto de características analisado. O classificador com o mais baixo rendimento foi da **janela de parzen**, devido ao alto grau de confusão gerado pelas diversas características que resultaram na dificuldade em determinar a janela do método.

Tabela 12 - Resultados para o grupo de experimentos 2 no conjunto de características 2.

	Gauss	Parzen	kNN	SVDD
Cenário	FAR	FAR	FAR	FAR
4	31.77	100	32.29	99.47
5	22.16	91.79	47.70	100
6	41.01	87.50	75.63	87.50
7	47.90	96.87	86.91	87.50
8	6.71	75	28.41	0
9	2.74	87.50	63.08	87.50
10	19.72	90.23	73.24	50
11	50.73	97.65	68.35	75
12	77.92	100	87.25	75
13	42.38	90.62	86.91	75
Média	34.30	91,75	64,97	73.69

Nos resultados do Conjunto de Características 2 para os experimentos do grupo 3 na Tabela 13 são apresentados resultados nas mesmas configurações dos obtidos na Tabela 9. O classificador com a melhor taxa de verificação de assinaturas média foi mais uma vez a **função gaussiana**, os classificadores com os piores desempenhos foram **janela de parzen** e o **SVDD** com baixa diferença entre estes. Todos os cenários da Tabela 13 apresentam apenas a assinatura verdadeira na fase de treinamento e acrescenta assinaturas forjadas na fase de teste.

Tabela 13 - Resultados para o grupo de experimentos 3 no conjunto de características 2.

	Gauss		Parzen		kNN		SVDD	
Cenário	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
14	23.70	36.31	50.0	50.0	85.76	47.34	50.00	50.27
15	37.75	52.26	49.94	56.25	56.10	59.04	50.0	50.0
16	10.01	49.99	49.94	56.25	56.25	52.69	50.0	50.0

17	30.62	59.37	50.0	50.0	32.81	39.34	50.0	50.0
18	31.25	38.75	50.0	50.0	39.06	50.10	50.0	50.0
19	21.87	48.12	50.0	50.0	23.43	48.71	50.0	50.0
20	30.62	39.37	50.0	50.0	42.18	46.98	50.0	50.0

De modo semelhante ao definido a partir da Tabela 10, o classificador SVM apresentou o melhor resultado para a verificação com assinaturas forjadas habilidosas, mesmo com uma alta taxa de erro falsa rejeição, comparada as demais tuplas FAR-FRR, o resultado FAR obtido no classificador SVM equilibra o *trade-off*. O classificador SVM apresenta também os melhores valores na verificação das assinaturas *target*. Por fim, o pior classificador é novamente o SVDD, com altas taxas de erro FAR e FRR. O baixo desempenho do SVDD, resultando em um alto grau de confusão em todos os experimentos, sem exceções, é resultando da baixa qualidade da configuração definida para este descritor de dados neste conjunto de exemplares. A configuração definida foi extraída da literatura pertencente ao *toolbox dd_tools*, e mantida para estudar a versão mais simples do SVDD. Devido aos resultados de baixo desempenho, outras configurações podem ser analisadas, devido a outros estudos como o de [SOUZA, 2009] apresentarem bons resultados.

Tabela 14 - Resultados para o grupo de cenários 4 do conjunto de características 2.

Cenário	Gauss		Parzen		kNN		SVDD		SVM	
	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
21	78.12	21.87	50.01	48.43	43.95	35.93	50.0	50.0	6.25	37.50
22	81.25	18.75	50.00	49.21	47.10	30.46	50.0	50.0	12.50	66.66

7.3 Conjunto de Características 3

No conjunto de características 3 são avaliados os classificadores aplicados ao conjunto de características CPDP exposto por [FANG, 2002], com resultados apresentados nas tabelas 15, 16, 17 e 18.

Na Tabela 15 são apresentados os resultados para os experimentos do grupo 1. De modo semelhante aos resultados obtidos para o grupo 1 nos Conjuntos de Características 1 e 2, o classificador **SVM** obteve o melhor desempenho médio na verificação de assinaturas. Este aumento no desempenho da verificação de assinaturas forjadas é devido ao conjunto de referência definido na característica CPDP. Ao determinar o conjunto de referência é definida a separabilidade entre os diferentes tipos de assinaturas analisados, tornando a separação controlável. Analisando apenas a taxa de assinaturas *target* rejeitadas o classificador **SVM**

também apresentou a menor taxa média. Dentre todos os classificadores o **kNN** apresentou em média as maiores taxas para assinaturas *outlier* aceitas devido ao pequeno conjunto de referências, contendo apenas três exemplares, limitar a separabilidade das classes das assinaturas através dos vetores de suporte.

Tabela 15 - Resultados para o grupo de experimentos 1 no conjunto de características 3.

Cenário	Gauss		Parzen		kNN		SVDD		SVM	
	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
1	16.15	83.85	49.96	55.17	38.12	63.81	50.0	50.27	37.50	0
2	78.12	21.87	50.0	50.0	43.75	56.25	50.0	50.0	29.50	5.34
3	50.0	50.0	50.0	50.0	37.50	62.50	50.0	50.0	26.50	3.51

Na Tabela 16 são apresentados os resultados para o segundo grupo de experimentos. Considerando o valor médio calculado todos os classificadores apresentaram resultados ruins, acima dos 50% de taxa de erro na verificação de assinaturas verdadeiras, mesmo com a aplicação do conjunto de referência. O aumento desta taxa foi resultante dos cenários contendo os três tipos de variações do tamanho da assinatura verdadeira, gerando um padrão de verificação restrito. Uma possível solução para este caso seria o aumento dos exemplares submetidos ao conjunto de referência e, por conseguinte, o aumento do conjunto de treinamento, com o objetivo de gerar um padrão com maior abrangência. Ao analisar os cenários individualmente, o classificador **gaussiano** apresentou as melhores taxas.

Tabela 16 - Resultados para o grupo de experimentos 2 no conjunto de características 3.

	Gauss	Parzen	kNN	SVDD
Cenário	FAR	FAR	FAR	FAR
4	32.30	91.66	55.72	99.47
5	5.81	87.30	57.22	100
6	51.41	96.87	47.50	87.50
7	29.72	99.90	86.32	87.50
8	26.65	100	62.50	100
9	74.65	100	37.5	100
10	33.49	86.96	65.96	31.25
11	78.36	99.95	77.19	74.95
12	88	100	87.15	100
13	88	100	99.90	100

Média	62.21	96.26	67.69	98.06
-------	--------------	-------	-------	-------

Na Tabela 17 são encontradas as taxas de verificação de assinaturas para o grupo 3 de experimentos. Os melhores resultados foram apresentados novamente para a **função gaussiana** na maioria dos casos.

Tabela 17 - Resultados para o grupo de experimentos 3 no conjunto de características 3.

Cenário	Gauss		Parzen		kNN		SVDD	
	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
14	50.79	49.21	50.00	50.00	40.06	49.74	50.0	50.0
15	36.81	63.18	37.50	62.50	27.92	72.07	50.0	50.0
16	36.13	63.86	37.50	62.50	30.85	69.14	50.0	50.0
17	4.68	95.31	50.0	50.0	31.25	33.17	50.0	50.0
18	20.31	79.68	50.0	50.0	43.75	43.75	50.0	50.0
19	37.50	62.50	40.62	50.09	35.93	43.95	50.0	50.0
20	32.81	67.18	43.75	57.25	56.21	48.56	50.0	50.0

No conjunto de características 3, o classificador **SVM** apresentou os melhores resultados na verificação de assinaturas *target* e novamente o melhor resultado no *trade-off* para as assinaturas forjadas habilidosas. De modo semelhante às duas tabelas anteriores para o grupo de cenários 4, o **SVDD** apresentou-se como pior classificador pelo motivo relatado anteriormente.

Tabela 18 - Resultados para o grupo de cenários 4 no conjunto de características 3.

Cenário	Gauss		Parzen		kNN		SVDD		SVM	
	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
21	53.42	22.10	55.01	38.33	42.95	37.29	50.0	50.0	8.30	29.50
22	44.65	20.00	57.70	50.21	48.20	31.36	48.0	47.0	14.20	40.06

7.4 Conjunto de Características 4

No conjunto de características 4 as assinaturas são avaliadas pelos mesmos classificadores através da característica *Shadow Code* [SABOURIN, 1993]. As taxas geradas para os experimentos expostos na Seção 5.4 são apresentadas na tabela 19, 20, 21 e 22.

Na Tabela 19 são apresentadas as taxas de verificação para o grupo de experimentos 1. Novamente o classificador **SVM** apresenta a melhor relação média entre as taxas de *target*

aceitas e *outlier* rejeitadas, com valores de 0 para a taxa de falsa rejeição com a base do ICDAR, e 0 para a falsa aceitação com assinaturas aleatórias na base AMODA. Ao analisar apenas as taxas de falsa aceitação de assinaturas originais, a **função gaussiana** apresenta novamente em média as menores taxas. Novamente o kNN apresentou os piores resultados, ou seja, baixa separabilidade entre as assinaturas originais de diferentes tamanhos de aquisição e as forjadas através do artifício distância euclidiana.

Tabela 19 - Resultados para o grupo de experimentos 1 no conjunto de características 4.

Cenário	Gauss		Parzen		kNN		SVDD		SVM	
	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
1	11.98	68.03	28.93	70.58	66.38	62.5	50.00	50.27	50	0
2	32.00	50.00	35.06	61.97	49.84	66.31	50.00	50.00	38.80	17.83
3	35.00	40.00	43.73	63.93	49.84	67.00	50.00	50.00	0	100

A Tabela 20 apresenta os resultados da verificação para o conjunto de cenários 2. O classificador **janela de parzen** apresentou dentre todos os resultados médios obtidos para este grupo a melhor taxa de erro, igual a 3%. Esta taxa foi resultante da alta representatividade proporcionada pela característica *shadow code*, e do bom conjunto de janelas proporcionado por estas características.

Tabela 20 - Resultados para o grupo de experimentos 2 no conjunto de características 4.

	Gauss	Parzen	kNN	SVDD
Cenário	FAR	FAR	FAR	FAR
4	23.95	0	41.66	99.47
5	100	7.03	81.29	100
6	87.5	7.95	92.82	87.50
7	100	0.39	96.41	87.50
8	12.69	0	54.93	0
9	51.07	0	69.38	50.0
10	51.07	0	69.38	50
11	87.50	13.47	77.63	75
12	75.97	0.78	96.33	75
13	75.97	0.78	96.33	75
Média	66.50	3.0	77.61	69.90

Na Tabela 21 são apresentados os resultados da verificação com o terceiro grupo de experimentos. Neste contexto o classificador nomeado de **janela de parzen** apresentou novamente os melhores valores na maioria dos cenários, bem como a melhor taxa de verificação de assinaturas *target*. O aumento desta taxa é resultante do aumento da representatividade proporcionada pelos três diferentes tipos de tamanhos de aquisição. O **kNN** apresenta o pior desempenho na verificação de assinaturas originais e forjadas.

Tabela 21 - Resultado para o grupo de experimentos 3 no conjunto de características 4.

	Gauss		Parzen		kNN		SVDD	
Cenário	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
14	11.98	67.03	28.93	30.58	66.38	62.5	50.00	50.27
15	30.00	50.00	15.27	41.85	49.84	65.46	22.43	50.00
16	23.74	50.00	10.92	36.20	49.84	66.46	50.0	50.0
17	45.20	50.00	0.78	24.69	62.36	50.57	30.98	50.0
18	50.34	50.00	0.78	25.00	59.16	62.50	50.0	50.0
19	50.00	50.67	0.78	28.24	59.24	54.69	44.32	50.00
20	30.11	50.00	0.78	44.69	56.04	65.63	50.0	50.0

Por último, o grupo de cenários 4 com os resultados para o *shadow code* definiu as taxas FAR e FRR apresentadas na Tabela 22. Neste contexto, o classificador SVM apresentou o melhor desempenho para o *trade-off* em ambos os cenários. O classificador **janela de parzen**, como nas tabelas anteriores, apresentou os melhores resultados para a verificação de assinaturas originais. Como pior classificador, o kNN possui os mais altos resultados.

Tabela 22 - Resultados para o grupo de cenários 4 no conjunto de características 4.

	Gauss		Parzen		kNN		SVDD		SVM	
Cenário	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
21	45.30	50.20	7.10	24.70	54.22	50.55	30.33	50.56	9.89	8.38
22	33.21	37.20	3.49	22.00	40.00	45.43	30.10	43.54	7.45	5.77

7.5 Comparação entre os experimentos

Com os objetivos: comparar os resultados obtidos neste estudo com os resultados apresentados na competição do ICDAR 2009 e, avaliar a verificação de assinaturas verdadeiras com diferentes tamanhos de aquisição, foram definidos os quatro grupos experimentais. Através dos grupos 1 e 4 foi possível identificar a melhor tupla características-classificador nos cenários

apenas com a base do ICDAR 2009 para comparar com os resultados da competição, o desempenho da base produzida AMODA na mesma configuração aplica ao experimento com a base ICDAR, e o desempenho da base AMODA sem (grupo 1) e com (grupo 4) a aplicação das assinaturas verdadeiras com diferentes tamanhos de aquisição. Ao analisar os resultados obtidos neste estudo e os da literatura, o classificador SVM é utilizado como classificador-ponte entre o campo da verificação de assinaturas composta com assinaturas com diferentes tamanhos de aquisição (proposta deste estudo) e sem este grupo de assinaturas (trabalhos semelhantes aos avaliados no ICDAR 2009).

Para a análise da primeira consideração acima relatada foi utilizada a taxa DET apresentada na seção 5.3. A comparação entre os resultados do grupo um é baseada na curva DET devido à competição promovida pelo ICDAR 2009 comparar os classificadores analisados através desta métrica. A extração da curva DET é realizada apenas sobre o 1º cenário do grupo um devido a apenas este ter a mesma configuração (verificação de assinaturas originais e forjadas, sem discriminação de tipo de forjada específico) aplicada aos classificadores da competição.

Em todos os 22 cenários avaliados, as boas taxas foram definidas com valores abaixo dos 30% na análise individual, e dos 40% na análise em grupo definida a partir do cálculo da média dos resultados para FAR, visto nas tabelas Tabela 8, 13, 17, 21.

Considerando a classificação com o SVM no cenário um, os menores resultados foram obtidos com o conjunto de características três (CPDP + distâncias), com taxas de falsa aceitação de assinaturas originais de 37,50%, e para a falsa rejeição de assinaturas forjadas de 0% na mesma configuração do ICDAR 2009. A literatura apresentou valores para a taxa de rejeição nas diversas características estudadas acima dos resultados obtidos, no trabalho de [SOUZA, 2009] no mesmo contexto com CPDP taxas entre 59% e 61% de rejeição foram obtidas. Em outros trabalhos com grupos diferentes de características, taxas mais próximas as obtidas foram atingidas nos trabalhos de [BALTAZAKIS e PAPAMARKOS, 2001] com a característica altura e a taxa de 19,91% de FAR e [KAPIL, 2014] com centro de massa com taxa de 17,34% de FAR. Contudo, nenhum trabalho pesquisado apresentou rejeição total ao conjunto de assinaturas forjadas, objetivo da verificação de assinaturas. Com a taxa de falsa rejeição sendo zero, a curva DET calculada retornou com valor também zero. Comparando com os resultados do ICDAR 2009, a taxa DET obtida foi inferior a todos os classificadores submetidos na competição (valores apresentados na Figura 30). Este resultado foi obtido através do conceito conjunto de referências gerado para as assinaturas originais e forjadas.

Partindo desta análise, e observando o desempenho do classificador SVM nos demais conjuntos de características pôde-se concluir a sua superioridade na separabilidade dos tipos de

assinaturas perante os classificadores da técnica OCC no grupo um. Em certos classificadores, resultados FAR foram inferiores aos obtidos com o SVM, contudo é preferível uma menor taxa de aceitação de assinaturas forjadas (FRR) em comparação as taxas FAR, como definido pelo *trade-off* da verificação de assinaturas, assim o classificador SVM continua apresentando os melhores resultados.

A respeito dos experimentos com a base AMODA sem exemplares de tamanhos diferentes, os melhores resultados foram obtidos com o conjunto de características CPDP (Tabela 16) e mais uma vez com o classificador SVM. O SVM apresentou a melhor separabilidade entre os dois conjuntos de assinaturas devido à boa combinação entre o conjunto de referência selecionado, as características CPDP e os vetores de suporte gerados na fase de treinamento.

Analizando apenas os classificadores OCC e suas aplicações nos grupos de cenários dois e três (cenários OCC devido a fase de treinamento utilizar apenas uma classe na verificação das assinaturas) em todos os conjuntos de características são feitas as interpretações a seguir. O classificador janela de parzen no conjunto de características quatro apresentou o melhor resultado médio extraído dos cenários do grupo dois na taxa FAR igual a 3,00%. Neste grupo de cenários o objetivo foi identificar o grau de representatividade proporcionada pelo conjunto de características em um grupo de assinaturas verdadeiras. Em certos cenários envolvendo assinaturas com diferentes tamanhos de aquisição no grupo dois do conjunto de características três (Tabela 16), era esperada uma taxa de falsa aceitação baixa proporcionada pelo conceito conjunto de referências. Contudo os classificadores não produziram, com as características extraídas, funções de qualidade para o conjunto de exemplares selecionado. No conjunto de características dois, neste mesmo grupo de cenários, o classificador gaussiano apresentou ótimos resultados devido à função gaussiana definida e as características coletadas, em média a função gaussiana apresenta maior representatividade nas variações dos diferentes tamanhos de aquisição.

Os resultados obtidos com o classificador SVDD na configuração definida não foram relevantes devido aos altos valores obtidos nas taxas FAR e FRR em todos os conjuntos de características. Conclui-se o SVDD como o classificador de menor desempenho para a verificação de assinaturas neste estudo, em cenários com e sem assinaturas de diferentes tamanhos de aquisição.

Por fim, retomando a avaliação aplicada sobre os grupos de cenários um e quatro, o grupo de cenários quatro avaliou a verificação de assinaturas contendo assinaturas com diferentes tamanhos de aquisição e forjadas habilidosas e aleatórias com o objetivo de comparar estes resultados com os dos cenários dois e três. É possível observar uma variação (com valores

maiores ou menores) nas taxas de falsa aceitação e falsa rejeição nos diferentes conjuntos de características. Observando os resultados do grupo quatro nos diferentes conjuntos, a característica CPDP apresentou valores mais altos para a taxa de erro falsa rejeição de assinaturas forjadas de 29%, aceitável na verificação de assinaturas. Contudo, neste mesmo experimento realizado com o conjunto de características quatro, os melhores resultados para a taxa de erro falsa aceitação foram obtidas com o classificador Parzen, devido à boa relação entre estes como apresentado anteriormente. Contudo, o classificador SVM novamente apresenta o melhor entre todos os resultados neste escopo ao analisar o *trade-off* apresentado. Este bom desempenho foi obtido a partir da maior representatividade proporcionada pelas assinaturas de diferentes tamanhos de aquisição e pela característica *shadow code*.

8 Conclusões e Trabalhos Futuros

Neste estudo foi apresentada uma análise comparativa entre conjuntos de características e classificadores aplicados na verificação de assinaturas manuscritas, com cenários experimentais contendo assinaturas originais de diferentes tamanhos de aquisição e forjadas habilidosas/aleatórias. Os resultados gerados foram obtidos através da ferramenta *Matlab* [MATHWORKS, 2014], do *toolbox dd_tools* [TAX, 2013] e *prtools* [DUIN, 2007], mais detalhes nas seções 5.2, com a utilização das bases ICDAR 2009 [BLANKERS et al, 2009] e a base AMODA, produzida neste estudo para análise dos diferentes tamanhos de aquisição. O agrupamento das características selecionadas possibilitou uma visão clara das contribuições das características e dos classificadores selecionados em cada um dos experimentos realizados na verificação de assinaturas. Classificadores de uma classe foram selecionados para os experimentos estudados devido aos cenários de verificação apresentarem, em muitos casos, poucos exemplares forjados.

Nos quatro conjuntos de características foram avaliados 22 cenários experimentais, com variações dos tipos de assinaturas verdadeiras e forjadas. A análise dos resultados nos quatro conjuntos de características é mediada através do classificador-ponte SVM, com aplicações nos três primeiros cenários compostos por assinaturas originais e forjadas nas fases de treinamento e teste (grupo um); com nova aplicação nos cenários 21 e 22 (grupo quatro), similares aos três primeiros com o acréscimo de assinaturas originais com diferentes tamanhos de aquisição.

O primeiro conjunto de característica definido com um grupo de características locais apresentou ótimos resultados para a taxa FRR na verificação com a base do ICDAR 2009 com o classificador SVM. O segundo conjunto composto por características globais apresentou melhores resultados nas duas taxas (FAR e FRR), com a taxa FRR com valores inferiores aos apresentados na seção 2.3. O terceiro conjunto formado pela característica CPDP e pela definição das distâncias a partir do conjunto de referência resultou nos melhores resultados para o classificador SVM com bons comparativos, nas duas taxas de erro, com a literatura referenciada. Cenários similares ao um foram definidos para a base AMODA com forjadas habilidosas e aleatórias, os melhores resultados para as duas taxas foram obtidos com o conjunto de características três (CPDP) novamente.

A partir dos bons resultados obtidos na classificação binária, os demais cenários definidos conceitualmente como experimentos OCC, por apenas uma classe ser informada na fase de treinamento, foram gerados e avaliados. Bons resultados foram obtidos com a função gaussiana apresentando taxas de erro FAR mais baixas do que as apresentadas pela literatura para a configuração utilizada. Os resultados para a janela de parzen com o conjunto de características

shadow code surpreenderam em desempenho nos cenários OCC, tornando-se o melhor resultado OCC.

A função gaussiana e o conjunto de características dois apresentaram o melhor resultado médio nos cenários propostos para a verificação de assinaturas com diferentes tamanhos de aquisição (grupos dois e três). Analisando os resultados do grupo quatro, o classificador binário SVM novamente apresenta os melhores resultados (forjadas habilidosas e aleatórias) com o classificador *shadow code*.

Os resultados com os classificadores kNN e SVDD apresentam-se ineficientes nas configurações utilizadas neste estudo, contudo permitem uma margem para outras configurações possíveis nestes classificadores.

8.1 Trabalhos Futuros

Os resultados dos classificadores apresentaram variações nos diferentes conjuntos de características, o classificador parzen registrou os melhores resultados para os cenários OCC no conjunto de características quatro, superiores aos resultados gerados pelo classificador gaussiano em conjunto com as características CPDP, definidos como melhor resultado deste estudo. A tupla parzen-*shadow code* não foi definida como solução deste estudo devido à configuração aplicada ao classificador SVM apresentar resultados superiores em termos de desempenho aos do conjunto de características quatro. Por estes motivos, avaliar outras configurações para o classificador parzen é relevante devido às características *shadow code* apresentarem ótimos resultados no trabalho de [SOUZA, 2009] para a verificação de assinaturas sem exemplares de tamanhos diferentes de aquisição.

Os resultados obtidos na competição realizada pelo ICDAR 2009 apresentaram classificadores com altas taxas para o contexto clássico (verdadeiras e forjadas) de verificação de assinaturas. Contudo, devido ao pouco tempo destinado ao estudo, a implementação dos melhores classificadores tornou-se inviável, assim seria interessante implementar os classificadores com os melhores resultados visando comparar as taxas obtidas a partir da problemática de verificação de assinaturas com diferentes tamanhos de aquisição contra os resultados apresentados pelo ICDAR 2009.

Aumentar a base construída com a inclusão de mais exemplares de assinaturas em todas as classes também é vantajoso, visando aumentar a acurácia dos resultados.

Referências Bibliográficas

- AFSARDOOST, S., YOUSEFI, S., KHORSHIDI, M. **Offline Signature Verification Using Geometric Center Features.** *International Conference on Signal Processing*, 1491-1494, 2008
- ARAUJO, R. S. A., CAVALCANTI, G. D. C., FILHO, E. C. B. C. **An Approach to Improve Accuracy Rate of On-line Signature Verification Systems of Different Sizes.** *International Conference on Document Analysis and Recognition*, 332-336, 2007.
- ARYA, M. S.; INAMDAR, V. S. **A Preliminary Study on Various Off-line Hand Written Signature Verification Approaches.** International Journal of Computer Applications, vol. 1, no. 9, 2010.
- AZZOPARDI, G., CAMILLERI, K. P. **Offline handwritten signature verification using Radial Baseis Function Neural Networks,** *WICT Malta*, 2008.
- BHARADI, V. A.; KEKRE, H. B. **Off-line Signature Recognition Systems.** *International Journal of Computer Applications*, vol. 1, no. 27, 2010.
- BLANKERS, V. L.; VAN DEN HEUVEL, C. E.; FRANKE, K. Y.; VUURPIJL, L. G. **The ICDAR 2009 Signature Verification Competition.** *International Conference on Document Analysis and Recognition*, 2009.
- COETZER, J., HERBST, B. M., DU PREEZ, J. A. **Off-line Signature Verification: A Comparison between Human and Machine Performance.** *Tenth International Workshop on Frontiers in Handwriting Recognition*, 2006.
- CORTES, C.; VAPNIK, V. **Support-Vector Networks.** *Machine Learning*, vol. 20, 273-297, 1995.
- COSTA, P. C. **Níveis de Construção da Escrita: Como Identificar e Intervir. Concepções e Experiências na Alfabetização de Jovens, Adultos e Idosos.** Cadernos Temáticos da Diversidade. Secretaria de Estado da Educação, Governo do Paraná, 47-53. 2010.

DEMSAR, J. **Statistical Comparisons of Classifiers over Multiple Data Sets**. *Journal of Machine Learning Research*. 2006.

DEVROYE, L.; GYORFI, L.; LUGOSI, G. **A Probabilistic Theory of Pattern Recognition**. Springer. 1996.

DUDA, R. O.; HART, P. **Pattern Classification**, 2nd edition, Wiley, 2010.

DUIN, R. P. W. et al. **PRTTools4 – A Matlab Toolbox for Pattern Recognition**. 2007.

EL ABED, H., WENYIN, L., MARGNER, V. **International Conference on Document Analysis and Recognition (ICDAR 2011) – Competitions Overview**. *International Conference on Document Analysis and Recognition*, 2011.

FAWCELT, T. **An Introduction to ROC Analysis**. Pattern Recognition Letters, 861-874. 2006.

FERREIRO, E.; TEBEROSKY, A. **Psicogênese da Língua Escrita**. Porto Alegre: ArtMed. 2008.

FUKUNAGA, K. **Introduction to Statistical Pattern Recognition**. Academic Press, 1990.

GANESAN, K. ACHARYA, U. R., CHUA, C. K., LIM, C. M., ABRAHAM, K. T. **One-Class Classification of Mammograms Using Trace Transform Functionals**. IEEE Transactions on Instrumentation and Measurement, vol. 63, issue 2, 304-311, 2014.

GIOT, R.; EL-ABED, M.; ROSENBERGER, C. **Fast computation of the performance evaluation of biometric systems: application to multibiometric**. Cornell University Library. 2012.

GONZALEZ, R. C.; WOODS, R. E. **Segmentação de Imagens**. Processamento Digital de Imagens, 3 ed., p. 454-522, Pearson Prentice Hall, 2010.

HAMILTON, D. J.; WHELAN, J., McLAREM, A.; MACINTYRE, I.; TIZZARD, A. **Low cost dynamic signature verification system**. Proc. Eur. Convention Secur. Detection, no. 408, 202-206. 1995.

IMPEDOVO, S.; PIRLO, G. **Verification of Handwritten Signatures: an Overview.** *14th International Conference on Image Analysis and Processing*, 191-196. 2007.

IMPEDOVO, S.; PIRLO G. **Automatic Signature Verification: The State of the Art.** *IEEE Systems, Man, and Cybernetics Society*, 609-635. 2008.

JANA, R., SAHA, R., DATTA, D. **Offline Signature Verification using Euclidian Distance.** *International Journal of Computer Science and Information Technologies*, vol. 5, 707-710, 2014.

JAIN, A. K.; ROSS, A. A., NANDAKUMAR, K. **Introduction to Biometrics.** *Springer*. 2011.

JUSTINO, E. J R.; YACOUBI, A. E.; BORTOLOZZI, F.; SABOURIN R. **An Off-Line Signature Verification System Using HMM and Graphometric Features.** *4th IAPR International Workshop on Document Analysis Systems*, 211-222. 2000.

JUSTINO, E. J. R.; BORTOLOZZI, F.; SABOURIN, R. **Off-line Signature Verification Using HMM for Random, Simple and Skilled Forgeries.** *6th International Conference on Document Analysis and Recognition*, 1031-1034. 2001.

KAPIL, A., SINGH, J., SRIVASTAVA, V. **A Hybrid Approach for Offline Signature Verification using Artificial Neural Network.** *International Journal of Information & Computation Technology*, vol. 4, n° 12, 1113-1121, 2014.

KOVARI, B., TOTH, B., CHARAF, H. **Classification approaches in off-line handwritten signature verification.** *WSEAS Trans. Math.*, vol. 9, 500-509, 2009.

KOVARI, B.; CHARAF, H. **A study on the consistency and significance of local features in off-line signature verification.** *Pattern Recognition Letters*, vol. 34, 247 – 255, 2013.

KRAWCZYK, B., JELEN, L., KRYZAK, A., FEVENS, T. **One-Class Classification Decomposition for Imbalanced Classification of Breast Cancer Malignancy Data.** *Artificial Intelligence and Soft Computing*, vol. 8427, 539-550, 2014.

LACERDA, E. B. . Centro de Informática. Universidade Federal de Pernambuco. 2009.

LOULoudis, G.; GATOS, B.; PRATIKAKIS, I.; HALATSIS, C. **Text line and word segmentation of handwritten documents.** *Pattern Recognition*, v. 42, n. 12, 3169-3183, 2009.

MALIK, M. I.; LIWICKI, M.; ALEWINJSE, L.; OHYAMA, W.; BLUMENSTEIN, M.; FOUND, B. **The ICDAR 2013 Signature Verification Competition.** *12th International Conference on Document Analysis and Recognition (ICDAR)*, 1477-1483. 2013.

MAO, S.; ROSENFIELD, A.; KANUNGO, T. **Document structure analysis: a literature survey.** In: *Document Recognition and Retrieval, International Society for Optics and Photonics, Santa Clara, United States*, v. 5010, 197-207, 2003.

MARTIN, A. et. al. **The DET Curve in assessment of detection task performance.** 1997.

MAJHI, B., REDDY, Y. S., BABU, D. P. **Novel Features for Off-line Signature Verification.** *International Journal of Computers, Communications & Control*, vol. I, nº 1, 17-24, 2006.

MUSHTAQ, S.; MIR, A. H. **Signature Verification: a study.** *4th International Conference on Computer and Communication Technology*, 258-263. 2013.

PANSARE, A.; BHATIA, S. **Handwritten Signature Verification using Neural Network.** *International Journal of Applied Information Systems*, vol. 1, no. 2, 2012.

PARKER, J. R. **Algorithms for Image Processing and Computer Vision**, New York: John Wiley and Sons, 1997.

PATIL P., PATIL, A. **Offline Signature Recognition Using Global Features.** *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, issue 1, 2013.

PONTILLE, D. **Écologies de la signature en Science. Sociétés & Représentations: Ce que signer veut dire**, 135-156. Publications de la Sorbonne. 2008.

PUSHPALATHA, K. N., GAUTHAM, A. K., SHASHIKUMAR, D. R., SHIVAKUMAR, K. B., DAS, R. **Offline Signature Verification with Random and Skilled Forgery Detection Using Polar Domain Features and Multi Stage Classification-Regression Model.** *International Journal of Advanced Science & Technology*, vol. 5, 27, 2013.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence, A Modern Approach**. Second Edition, 2003.

SARMENTO, L. L. **Gramática em Textos**. São Paulo. Moderna. 2005.

SÁNCHEZ, A.; MELLO, C. A. B.; SUAREZ, P. D.; LOPES FILHO, A. N. G. **Automatic line and word segmentation applied to densely line-skewed historical handwritten document images**. Integrated Computer-Aided Engineering, v. 11, n. 2, 125-142, 2011.

SEVERIN, P. D. S. **Separación de Líneas y Palabras em Imágenes de Documentos Históricos Manuscritos**. Trabalho de Conclusão de Curso (*Grado en Ingeniería Informática*) – Universidad Rey Juan Carlos, Madri, Espanha. 2007.

SHASTRY, A.; BURCHFIELD, R.; VENKATESAN, S. **Dynamic Signature Verification Using Embedded Sensors**. *International Conference on Body Sensor Networks*, 168-173, 2011.

SHERMER, M. **The Skeptic Encyclopedia of Pseudoscience**, vol 2, pp. 455. 2011.

SOUZA, M. R. P. **Verificação de Assinaturas Off-line: Uma abordagem baseada na combinação de distâncias e em classificadores de uma classe**. Centro de Informática. Universidade Federal de Pernambuco. 2009.

STEHMAN, S. V. **Selecting and interpreting measures of thematic classification accuracy**. Remote Sensing of Environment, 62, pp. 77-89. 1997.

STORY, M.; CONGRALLTON, R. G. **Accuracy assessment: A user's perspective**. Photogrammetric Engineering and Remote Sensing, 52, pp. 397-399. 1986.

TAX, D. M. J.; DUIN, R. P. W. **Data domain description using support vectors**. Proceedings of the European Symposium on Artificial Neural Network, 251-256, 1999.

TAX, D. M. J. **One Class Classification: Concept-learning in the absence of counter-examples**. Doctoral Dissertation, University of Delft, The Netherlands. 2001.

TAX, D. M. J.; DUIN, R. P. W. **Support Vector Data Description.** Machine Learning, 45 – 66, 2004.

TAX, D. M. J. **Data description toolbox, dd_tools 2.0.0.** 2013.

MATHWORKS. **Matlab.** 2014.

VAPNIK, V. **Statistical Learning Theory.** Wiley-Interscience, 1998.

YUSOF, M. H. M.; MADASU, V. K. **Signature Verification and Forgery Detection System.** Student Conference on Research and Development, 9-14. 2003.

WIJESOMA, W. S.; YUE, K. W., CHIEN, K. L.; CHOW, T. K. **Online handwritten signature verification for electronic commerce over the internet.** Springer-Verlag, 227-236. 2001.

ANEXO 1

Neste anexo estão dispostas as imagens da base composta de assinaturas de tamanhos diferentes produzida para os experimentos analisados. 16 autores de assinaturas participaram da coleta de assinaturas, foram produzidas por cada autor 6 assinaturas em padrão considerado o habitual, 6 assinaturas em tamanho maior e 6 em tamanho menor.

Assinaturas habituais

Buma Vitorino Buma Vitorino

Buma Vitorino Buma Vitorino.

Buma Vitorino Buma Vitorino.

Buma Vitorino Buma Vitorino

Buma Vitorino Buma Vitorino.

Buma Vitorino Buma Vitorino

Buma Vitorino Buma Vitorino

Buma Vitorino Buma Vitorino.

Ricardo Accioly Filho

Allegreix Allegreix

Allegreix Allegreix

Allegreix Allegreix

Allegreix Allegreix

Allegreix Allegreix

Allegreix Allegreix

Roberto Andrade silva

Roberto Andrade silve

Roberto Andrade silve

Roberto Andrade Silva

~~Hag Snow~~

~~Hag Snow~~

~~Hag Snow~~

~~Hag Snow~~

~~Hag Snow~~

~~Hag Snow~~

~~fragKnows~~

~~fragKnows~~

~~fragKnows~~

~~fragKnows~~

~~fragKnows~~

Tamires Andrade Syripine

Tamires Andrade Agripino

Glaucyshafr - Glaucyshafr -

Glaucyshafr - Glaucyshafr -

Glaucyshafr - Glaucyshafr -

Glaucophan · Glaucophan ·

Glaucophan · Glaucophan ·

Glaucophan · Glaucophan ·

Murilo lira Murilo lira

Ronaths Arcidy

Ronaths Arcidy

Ronaths Arcidy

Ronaths Arcidy

Ronatha Accidy

Brunfennigus Gacilis

Brunfennigus Gacilis

Brunfennigkorska ~~skola~~ skola

Bryf en han gav heta

Leontine

Leontine

Leontine

Leontine

Leontine

Leontine

Leontine

Leontine

Leontine

Jeandine

Jeandine

Jeandine

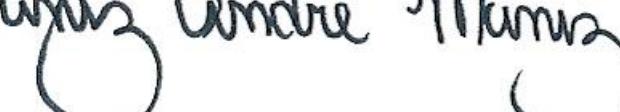
André Muniz André Muniz



André Muniz André Muniz



André Muniz André Muniz



André Muniz André Muniz



André Muniz André Muniz



André Muniz André Muniz



Am CA-

Am DA-

Am DA - Am DA

Am DA

Am C Adam

Clare P. Adam

Clare P. Adam

1 day . 1 day . 1 day

Raphael Jones Raphael Jones

Raphael Gene

Raphael Gene

Raphael Gene

Raphael Gene

Raphael Gene

Raphael Gene

Raphael Gonçalves

Raphael Gonçalves

Raphael Gonçalves

Allon Melo Allon Melo Allon Melo

Vanessa Christine G. de Lima

Wanessa Christine g. de Lima
Wanessa Christine g. de Lima
Wanessa Christine g. de Lima
Wanessa Christine g. de Lima
Wanessa Christine g. de Lima
Wanessa Christine g. de Lima
Wanessa Christine g. de Lima
Wanessa Christine g. de Lima
Wanessa Christine g. de Lima
Wanessa Christine g. de Lima
Wanessa Christine g. de Lima

Assinaturas em tamanho menor

Buma Vitorino Buma Vitorino.

Buma Vitorino. Buma Vitorino.

Buma Vitorino Buma Vitorino.

Buma Vitorino. Buma Vitorino.

Buma Vitorino. Buma Vitorino.

Buma Vitorino. Buma Vitorino.

Ricardo Accioly Filho

Ricardo Accioly Filho

Ricardo Accioly Filho

Ricardo Accioly Filho

Ricardo Accioly Filho Ricardo Accioly Filho

Ricardo Accioly Filho

Ricardo Accioly Filho Ricardo Accioly Filho

Ricardo Accioly Filho Ricardo Accioly Filho

Ricardo Aceidy Filho

Ricardo Aceidy Filho

Allegrej Allegrej

Allegrej Allegrej

Allegrej Allegrej

Allegrej Allegrej Allegrej

Allegrej Allegrej

Allegrej Allegrej

Roberto Andrade Silva

Roberto Andrade Silva

Roberto Andrade Silve

Roberto Andrade Shva

Roberto Andrade Silva

They Anayel

They Anayel

things Anwajh

things Anwajh

things Anwajh

things Anwajh

things Anwajh

things Anwajh things Anwajh

things Anwajh

things Anwajh

things Anwajh

Tamires Andrade Aguiar Tamires Andrade Aguiar

Tamires Andrade Aguiar

Tamires Andrade Aguiar

Glaucylophr. Glaucylophr.

Glaucylophr. Glaucylophr. Glaucylophr.

Glaucylophr. Glaucylophr.

Glaucylophr. Glaucylophr.

Glaucylophr. Muilo hira

Muilo hira Muilo hira Muilo hira

Muilo hira Muilo hira Muilo hira

Munilo hira Munilo hira Munilo hira
Munilo hira Munilo hira

Ronathra Accidly

Ronathra Accidly

Ronathra Accidly

Ronathra Accidly

Ronathra Accidly

Ronathra Accidly

Ronathra Accidly Ronathra Accidly

Ronathra Accidly Ronathra Accidly

Ronathra Accidly Ronathra Accidly

Brunfbergforskningslita Brunfbergforskningslita

Brunfbergforskningslita Brunfbergforskningslita

Brunfbergforskningslita Brunfbergforskningslita

Brunfbergforskningslita Brunfbergforskningslita

Brunfbergforskningslita Brunfbergforskningslita

Brunfbergforskningslita Brunfbergforskningslita

Assinaturas em tamanho maior

Buma Vitorino.

Buma Vitorino.

Buma Vitorino

Buma Vitorino

Buma Vitorino.

Buma Vitorino.

Buma Vitorino

Buma Vitorino.

Buma Vitorino.

Buma Vitorino.

Buma Vitorino.

Buma Vitorino.

Ricardo Accioly Filho

Allegreia

Allegreia

Allegri

Allegri

Allegri

Allegri

Allegri

Allegri

Allegrein

Allegrein

Allesgren

Allegrein

Roberto Andrade Silva

Roberto Andrade Shva

Roberto Andrade Shva

Thago Mawjib

Thago Mawjib

Thago Mawjib

~~Thag Mawb~~

~~Thag Mawg~~

~~Thag Mawb~~

~~Trege Manx~~

~~Flag Manx~~

~~Flag Hawk~~

~~Flag Manx~~

~~Frag. Shows~~

~~Frag. Shows~~

Tamires Andrade Aguiarino

Tamires Andreode Aguiar

Tamires Andreode Aguiar

Tamires Andreode Aguiar

Tamires Andreode Aguiar

Glaucifolia · Glaucifolia ·

Murilo hira Murilo hira

Ronathra Accidly

Ronathra Accidly

Ronathra Accidly

Ronatha Accidly

Ronatha Accidly

Ronatha Accidly

Rontha Accidly

Rontha Accidly

Rontha Accidly

Romitha Accidly

Romitha Accidly

Romitha Accidly

Bryfenniglas Dafita

Bryfenniglas Kefronshita

Bryfenniglas Kefronshita

Bryfenniglas Kefronshita

Bryophyllum pinnatum

Bryophyllum pinnatum

Bryophyllum pinnatum

Bryophyllum pinnatum

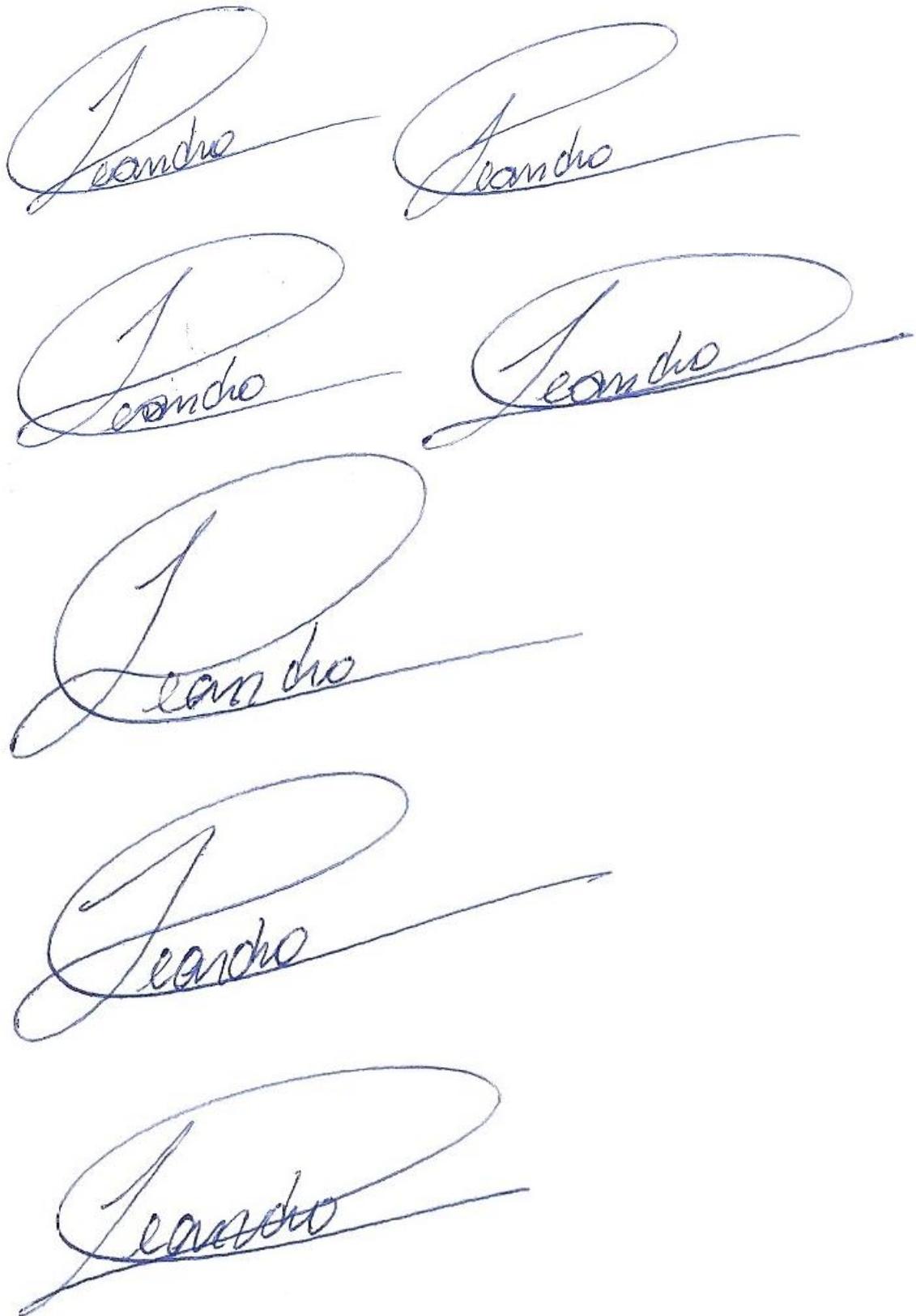
Bryophyllum pinnatum

Bryophyllum pinnatum

Bryggen
Bryggen Bryggen

Bryggen
Bryggen Bryggen

Assinaturas Forjadas Habilidosas



Leonardo

Leonardo

Leonardo

Leonardo

Leonardo

Allegreij Allegreij

Allegreij Allegreij

Allegreij Allegreij

Allegreij Allegreij

Allegreij Allegreij

Allegreij Allegreij

André Muniz

André Muniz

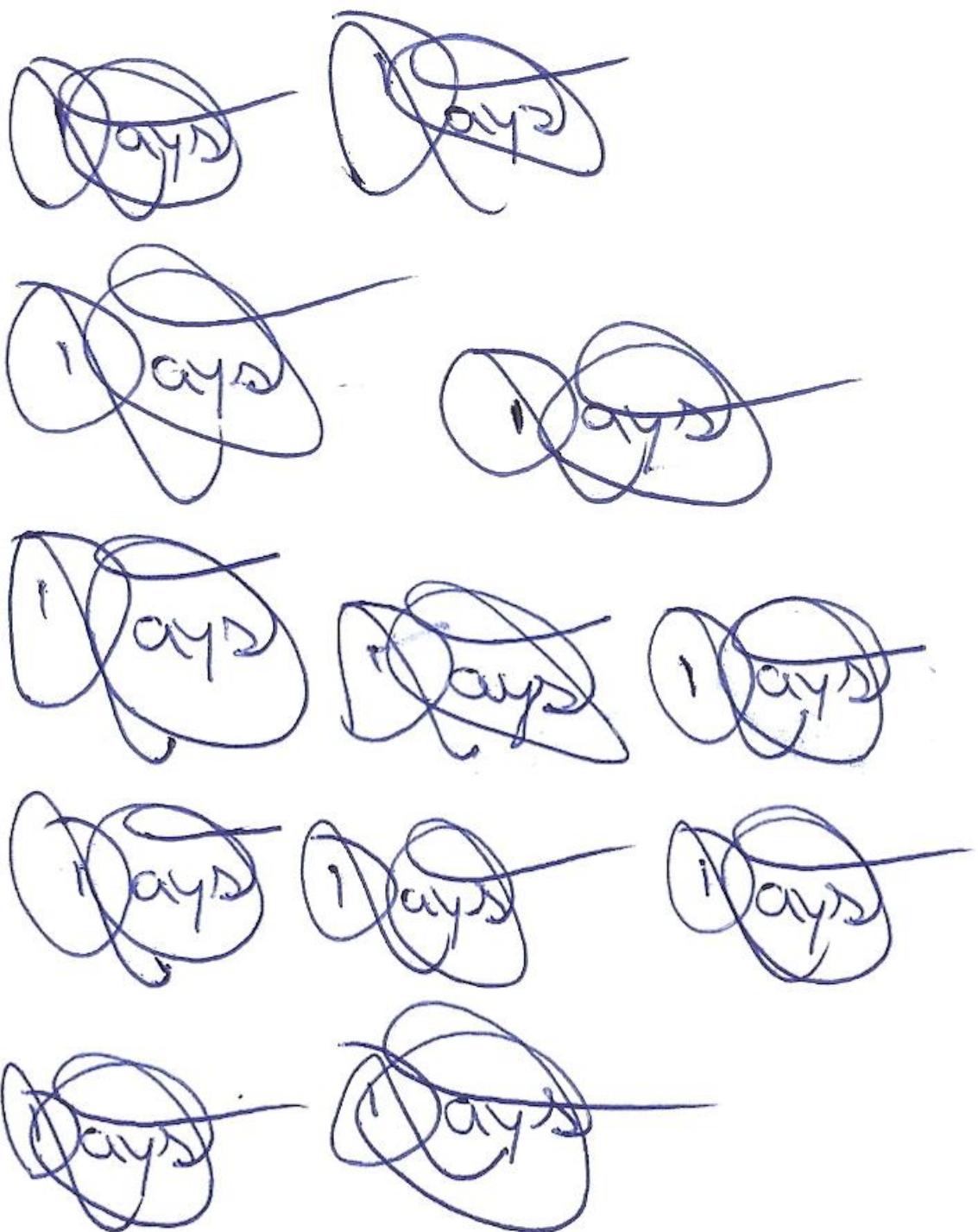
André Muniz

André Muniz

andre' Mumiz

Clin C Lin

Clin C An



Raphael Gomes Raphael Gomes

Raphael Gomes Raphael Gomes

Raphael Gomes

Raphael Gomes

Raphael Gomes Raphael Gomes

Raphael Gomes Raphael Gomes

Raphael Gomes Raphael Gomes

Allan Melo Allan Melo

Vanessa Christine G. de Lima

Roberta Andrade Silva

~~Thug Anangas~~

~~Thug Anangas~~

~~Thug Anangas~~

~~Thug Anangas~~

~~Thug Anangas~~

~~Thiago Ananay~~

~~Thiago Ananay~~

~~Thiago Ananay~~

~~Thiago Ananay~~

~~Thiago Ananay~~

~~thug thong~~

~~thug thong~~

Burna Vitorino

Burna Vitorino

Burna Vitorino

Burna Vitorino

Burna Vitorino

Burna Vitorino

buma Vitorino

Assinaturas Forjadas Aleatórias

Leandro Leandro
ADNegrários
ADNegrários
ADNegrários
ADNegrários

ADNegriros

ADNegriros

ADNegriros

ADNegritos

ADNegriros

ADNegriros

ADNegriros

ADNegriros

Mumiz

Mumiz

Mumiz

Mumiz

Mumiz

Alvin Alvin

Alvin

Alvin

Alvin

Alvin Alvin

Glossy Amorim *Glossy Amorim*

Glossy Amorim *Glossy Amorim*

Glossy Amorim *Glossy Amorim*

Glossy Amorim *Glossy Amorim*

GlossyCamomim GleissorCamomim

GleissorCamomim GleissorCamomina

Dayre Dayse

Dayre Dayse

Dayre Daeyes

Dayre Dayse

Dayre Dayse

Dayre Dayse

Raphael Gomes

Allan Melo Allan Melo Allan Melo

Allan Melo Allan Melo

Allan Melo Allan Melo

Allan Melo Allan Melo

WanMelo WanMelo

WanMelo

Wanessa Christine G. de Lima

Roberta Andrade Silva

Roberto Andrade Silva
Thiago Amorim Jás

Thiago Amorim

Thiago Amorim

Thiago Amorim

Thiago Amorim

Luma Vitorino

Luma Vitorino