Revista de Engenharia e Pesquisa Aplicada, Volume 2, Número 1, 2016

Extração de características e classificação de assinaturas manuscritas

Features extraction and classification of handwritten signatures

Danilo Simplício da Rocha
Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001 - Recife, Brasil
dsr@ecomp.poli.br

Byron Leite Dantas Bezerra
Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001 - Recife, Brasil
byronleite@ecomp.poli.br

Resumo *O intuito do presente relatório é descrever as atividades desenvolvidas ao longo d um projeto de iniciação científica, cuja finalidade é extrair características e classificar uma assinatura manuscrita, além de definir alguns conceitos importantes relacionados ao tema. O principal resultado alcançado, foi a elaboração de um software que permite a captura de assinaturas on-line, e a criação de três bases de dados de assinaturas off-line, que serão utilizadas em diferentes abordagens de classificação, podendo levar a um algoritmo que identifique a autenticidade de uma assinatura independente da forma que ela foi obtida (on-line ou off-line)*

Palavras-Chave: *Verificação de assinatura; Reconhecimento de padrões; Biometria; Processamento de Imagens*

Abstract *The report's aim is to describe the activities developed over a research project, whose purpose is to features extraction and sort the handwritten signature and defines some important concepts related to the topic. The major achievement was the development of a software that allows to capture online signatures, and the creation of the three signatures offline databases that will be used in different classification approaches, leading an algorithm which identifies the authenticity of an independent signature the way it was obtained (online or offline)*

Keywords: *Signatures verification, Pattern recognition, Biometry, Image processing*

1 Introdução

A assinatura manuscrita é um método largamente utilizado para autenticação pessoal na sociedade contemporânea, sobretudo em processos administrativos e financeiros [1]. Sua eficácia se dá pelo fato de ser um identificador biométrico.

Um classificador biométrico é essencialmente um sistema de reconhecimento de padrões que opera por meio das características biológicas ou tratamento comportamental [2][3]. Dentro desse entendimento, a assinatura pertence a uma classe única e diferenciada, pois diferentemente da retina, da impressão digital ou mesmo da voz, uma assinatura manuscrita não possui relação anatômica com seu autor. Por conseguinte, poderão existir variações intrapessoal entre subscrições de um mesmo mentor, semelhanças caligráficas entre diferentes sujeitos e tentativas de fraudes que podem levar técnicos forense experientes à indecisão ou ao erro.

Ao passo que um documento pode ser furtado e uma senha copiada, a informação biométrica é muito mais difícil de ser transferida ou reproduzida, sendo assim, um meio de identificação mais confiável e seguro.

A verificação automática de assinaturas desperta um amplo fascínio da comunidade científica, tanto pelos seus desafios tecnológicos, quanto pela importância de sua aplicação prática. Uma vez que, subscrições contrafeitas podem acarretar danos significativos ao indivíduo e em instituições públicas e privadas [7].

Neste relatório os principais conceitos que envolvem a extração de características e classificação de assinaturas são abordados. Além disso, expõe o processo de criação das bases de dados que serão utilizadas nas próximas etapas deste projeto de iniciação científica.

2 Referencial teórico

2.1 Formas de captura

A captura de uma assinatura pode ser obtida de duas maneiras distintas, que são as formas off-line (estática) e on-line (dinâmica)[4].

Na forma off-line, não se tem conhecimento dos movimentos necessários para gerar a assinatura, ou seja, o usuário faz a assinatura em uma folha de papel, que é subsequentemente digitaliza, utilizando por exemplo um scanner ou uma câmera fotográfica. Já na forma on-line, a captura é realizada sobre um dispositivo, geralmente é um tablete, que permite capturar informação dinâmicas da assinatura (coordenadas, pressão em ponto, velocidade, aceleração, etc.).

Quando os dois métodos são comparados, percebe-se, do ponto de vista técnico, que o reconhecimento de assinaturas on-line é mais vantajoso em relação ao off-line [5]. Entre os fatores que interferem neste desempenho, destacam-se:

- Maior riqueza de informações: além das caracterís-ticas visuais da assinatura, é possível obter infor-mações temporais e dinâmicas;
- Pureza da imagem: uma assinatura off-line costuma apresentar distorções provenientes do processo de digitalização, o que dificulta o processo de autenti-cação;
- A forma off-line demanda um tratamento muito mais complexo, que resulta em um desempenho in-ferior.

Todavia, a necessidade de um dispositivo para captura on-line limita sua aplicação, visto que impede a verificação em cheques bancários ou documentos previamente assinados.

2.2 Tipos de Fraude

Uma assinatura falsa pode ser bastante diferente de uma autêntica, tanto em formato como em suas propriedades, ou serem tão idênticas que mesmo um técnico forense experiente pode ter dificuldades em classificá-la corretamente. As assinaturas não fidedignas podem ser classificadas em [6]:

- Imitação exercitada - o falsário treina uma assinatura qualquer, até reproduzi-la sem a necessidade de um modelo (não possui equivalência com a legí-tima);
- Imitação de memória - o falsário memoriza traços da assinatura genuínas;
- Imitação servil - fraude habilidosa, pois o falsifica-dor, possuindo cópias da assinatura autêntica, treina até conseguir reproduzir os traços;
- Sem imitação - o falsário reproduz o nome do autor sem se preocupar com a forma;
- Auto falsificação - o próprio autor gera uma fraude com o objetivo de negar a autoria;

- Decalque - O falsário utiliza uma transparência da assinatura real.

2.3 Etapas do processo

O processo de criação de um sistema de verificação de assinatura, seja on-line ou off-line, é divido normalmente em duas etapas: criação de uma base de dados e processo de verificação.

A base de dados é criada adquirindo versões autênticas de cada autor, que serão empregadas como referência ao longo da etapa de averiguação da legitimidade da assinatura. Feito isso, o próximo passo é o pré-tratamento dos dados, objetivando a eliminação de ruído, em seguida, é realiza a extração de características [1][7].

As características adquiridas das assinaturas autênticas, serão utilizadas quando se desejar verificar a veracidade de uma dada assinatura (Figura 1).

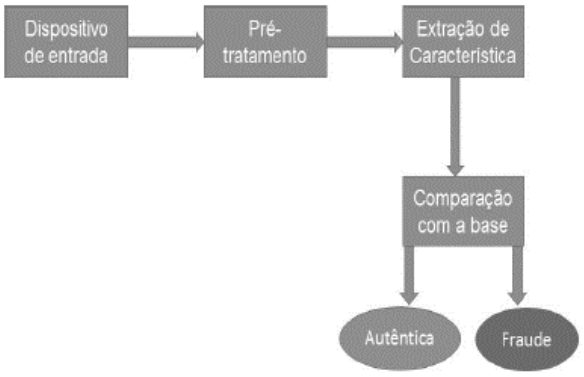

Figura 1: Etapas para verificação de uma assinatura

A Tabela 1 contém as principais técnicas empregadas em cada etapa do processo de classificação.

Etapa	Técnica	
Pré-tratamento	Filtragem	- Filtro morfológico
		- Filtro médio
	Compressão	- Filtro de integração
		- Filtro de subamostragem
		- Filtragem ponderada
	Binarização	- Limiarização de histograma
		- Filtragem
	Diluição	- Algoritmo de Hilditch
Extração de características	Globais	- Função densidade de probabilidade
		- Linha de base global
		- Análise espectral
		- ESC
		- Projeções e momentos das projeções
		- Centros de massa
		- Inclinação global
	Locais	- Descritor de formas
		- Pontos extremos
		- Analise topológica
		- Propriedades geométricas
Classificação	Estocástica	- HMM
		- Quantificação vetorial/máquinas de suporte vetorial
	Redes Neurais	- Rede acíclica (feedforward)
		- ADALINE
		- Perceptron multicamada
		- ART
		- Neocognitron
		- Retro-propagação (backpropagation)
	Medidas de Distância	- kNN
		- Classificador limiarizado
		- Emparelhamento elástico
		- Programação dinâmica
		- Distância Euclidiana

Tabela 1: Principais técnicas utilizadas em cada etapa do processo de classificação

3 Sistema Proposto

O projeto possui como objetivo classificar assinaturas independente da forma que foram obtidas. Então, a fim de se criar uma base de dados que contenha as características on-line de uma assinatura foi adquirido uma mesa digitalizadora Wacom modelo STU-530 e, conseguinte, desenvolvida uma ferramenta para o ambiente Windows (utilizando a linguagem Java e as bibliotecas fornecida pela fabricante) que permitisse a captura de assinaturas (on-line) genuínas e falsas. Possibilitando assim, uma análise detalhada da trajetória, bem como da força exercida, no processo de construção da assinatura.

A seguir, uma descrição das funcionalidades encontradas no software desenvolvido

3.1 Configuração

Para se adequar à variados cenários de teste, o sistema permite definir algumas diretrizes de configuração (Figura 12), são elas:

- Registro de usuário – Ao definir como sim, será ne-cessário realizar o cadastro do usuário antes de ini-ciar a captura;
- Salvar tempo – Afirmando essa opção, o sistema irá salvar o tempo (em segundos) em que cada carac-terística da assinatura foi capturada.
- Quantidade de capturas – Permite que o pesquisa-dor indique a quantidade de assinaturas (autênticas e forjadas) que deseja capturar para cada autor.

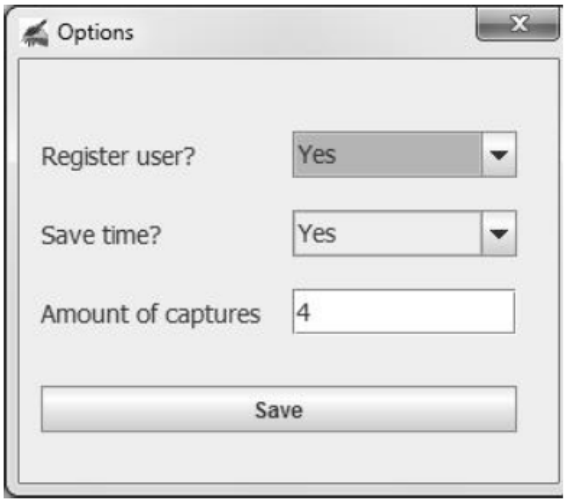

Figura 2: Tela de configuração

3.2 Cadastro de usuário

Se na configuração do sistema for definido que para iniciar a captura da assinatura dever-se-á indicar um usuário, o sistema terá em seu menu a opção para tal fim (Figura 13).

A grande vantagem desta opção, é que permite ao pesquisador realizar capturas de assinatura em diferentes momentos, e assim, compreender as variações das informações obtidas (seja por condições psicofísicas ou idade).

Os dados de cada usuário, bem como seu histórico de capturas, são salvos em arquivos XML. Veja a Figura 14.

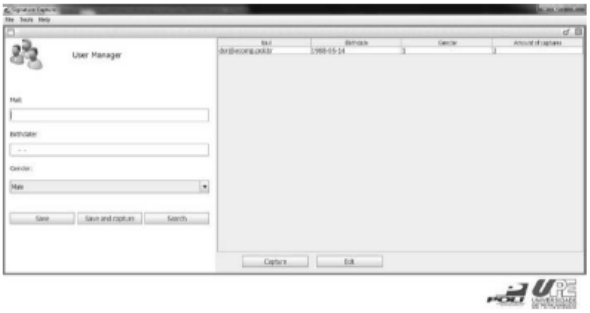

Figura 3: Tela gerenciamento de usuário

```
<?xml version="1.0" encoding="UTF-8" standa-
lone="no"?>
<data>
  <user>
    <emailAddress>dsr@ecomp.poli.br</emai-
lAddress>
    <birthdate>1988-05-14</birthdate>
    <gender>1</gender>
    <captures>
      <capture date="2015-11-01 12:32:30">
        <file>
          <imageFileName>1.png</imageFileName>
          <dataFileName>1.data</dataFileName>
          <filePath>.\signatures\with-
user\dsr@ecomp.poli.br\genuine\1</filePath>
          <genuine>true</genuine>
        </file>
        <file>
          <imageFileName>2.png</imageFileName>
          <dataFileName>2.data</dataFileName>
          <filePath>.\signatures\with-
user\dsr@ecomp.poli.br\genuine\1</filePath>
          <genuine>true</genuine>
        </file>
      </capture>
    </captures>
  </user>
</data>
```

Figura 4: Exemplo de XML com dados de um usuário, contendo o histórico de capturas, neste caso, a configuração indicava que cada captura deveria conter duas assinaturas.

3.3 Tela de captura

Enfim, a tela de captura. Ela funcionará conforme indicado na configuração do sistema. Se não for necessário o registro do usuário, após a captura das assinaturas genuínas, será inicializado automaticamente a obtenção das falsas, pois os registros são salvos em diretórios com numerações sequências, e um novo número é gerado ao atingir a quantidade de assinaturas definida. Diferentemente, quando se requer um usuário, a captura das assinaras autênticas, ou forjadas, poderão acontecer conforme a necessidade do pesquisador (Figuras 15 e 16).

Com ou sem usuário, os dados salvos para cada assinatura on-line são:

- Informações do dispositivo (resolução, pressão má-xima, máximo x e y, largura e altura da tela)
- Coordenada X
- Coordenada Y
- Pressão
- Tempo em segundos (opcionalmente)
- Imagem - gerada através da interpolação dos dados obtidos

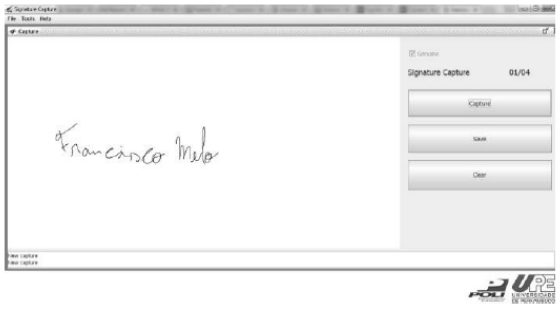

Figura 5: Tela de captura sem o cadastro do usuário

Figura 6: Tela captura com o cadastro do usuário

4 Resultados preliminares

Como descrito anteriormente a criação da base de dados é um dos pilares quando se pretende desenvolver um sistema que verifica a autenticidade de uma assinatura. Por essa razão, os primeiros meses da iniciação científica tiveram enfoque total na confecção deste instrumento, e como efeito três tipos de base off-line estão disponível, visando atender o mais variado cenário.

Nos tópicos subsequentes, são descritos o procedimento de criação de cada base, e sua finalidade.

4.1 Cheques nacionais

O projeto teve à sua disposição um repositório de assinaturas (off-line), contidas em cheques nacionais, com inúmeras subscrições genuínas, pertencentes à cerca de quatro mil pessoas distintas identificadas por um número.

Com o propósito de criar um classificador que aprenda as divergências compreendida entre diferentes assinaturas de um mesmo autor, o repositório foi totalmente inspecionado, de tal modo que as assinaturas destoantes foram separadas.

Para exemplificar, considere que o usuário identificado pelo número 1, possui quatro assinaturas díspar, Figura 2, Figura 3, Figura 4 e Figura 5. É fácil perceber que a Figura 2 e a Figura 4 são semelhantes, bem como a Figura 3 e a Figura 5. Desta forma, as assinaturas do usuário 1 foram separadas em dois subconjuntos, 1-1 (contendo as Figuras 2 e 4) e 1-2 (contendo Figuras 3 e 5).

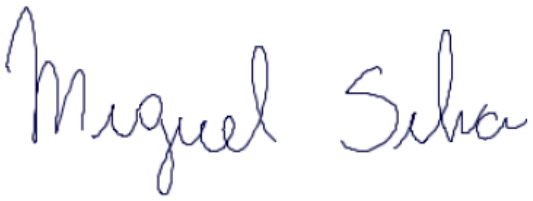

Figura 7: Primeira assinatura do usuário 1

Figura 8: Segunda assinatura do usuário 1

Figura 9: Terceira assinatura do usuário 1

Figura 10: Quarta assinatura do usuário 1

4.2 Assinaturas forjadas do dataset Sigwicomp (2005)

A Sigwicomp é uma competição internacional de algoritmos de classificação de assinaturas (on-line e off-line). Para o desafio de 2015, foram disponibilizadas duas bases de assinaturas off-line, italiana e bengalês.

Com as bases em mãos, foram criadas quatro assinaturas falsas para cada autor.

Abaixo, segue uma amostra deste processo, onde a Figura 6 e a Figura 7 correspondem a subscrições autêntica e forjada respectivamente.

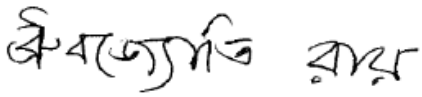

Figura 11: Assinatura bengalês autêntica

Figura 12: Assinatura bengalês forjada

4.3 Documentos oficiais e cheques

Naturalmente, aguçou-se o interesse de simular uma situação ainda mais próxima de um ambiente real. Para isso, foi confeccionado réplicas de documentos oficias brasileiros (CNH, RG e título de eleitor) e cheques, que serão utilizadas em futuros algoritmos de localização e segmentação de assinaturas.

A produção dessa base foi dada utilizando, mais uma vez, o dataset disponibilizado pela Sigwicomp (edição 2015), e o resultado desta produção é ilustrado nas Figuras 8, 9, 10 e 11.

Figura 13: Réplica CNH

Figura 14: Réplica RG

Figura 15: Réplica título de eleitor

Figura 16: Réplica cheque

5 Conclusão

As atividades realizadas até o momento, são cruciais para os próximos passos do projeto. Utilizando a ferramenta desenvolvida, uma nova base de dados (on-line) será preparada. Em seguida, juntamente com os repositórios off-line já elaborados, diferentes métodos de classificação serão estudados e aplicados. Assim, os resultados obtidos nas distintas formas de captura (on-line e off-line) pertencerão a um quadro comparativo, permitindo uma compreensão minuciosa das propriedades contidas em uma assinatura. Por fim, pretende-se desenvolver um software que possibilite, de forma satisfatória, a verificação da autenticidade de uma assinatura, independente da sua origem.

Referências

[1] G. Pirlo, D. Impedovo. A Cosine similarity for analysis and verification of static signatures. *In: IET Biometrics, www.ietdl.org, 2013*

[2] A. Kholmatov. A Biometric Identity Verification Using On- Line & Off-Line Signature Verification. Dissertação de Mestrado - Sabanci University, 2003.

[3] G. Pirlo, D. Impedovo. Verification of Static Signatures by Optical Flow Analysis. *In: IEEE Transactions On Human-Machine Systems*, VOL. 43, NO. 5, Setembro 2013.

[4] L. Ravi Kumar, A.Sudhir Babu. Genuine and Forged Offline Signature Verification Using Back Propagation Neural Networks. *L. Ravi Ku-mar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 2 (4) , 2011, 1618-1624.

[5] M.P. Heinen. Autenticação On-line de assinaturas utilizando Redes Neurais. Trabalho de Conclusão - UNISINOS. 92p, 2002. http://www.bminds.com.br/~milton/.

[6] R. Sabourin e G. Genest, Définition et Évaluation d'une Famille de Répresentations pour la Vérification hors Ligne des Signatures, *Traitement du Signal*, Vol. 12, No. 6, pp. 587-596, 1995.

[7] A. Zimmer e L. L. Ling. A Hybrid On/Off-line Handwritten Signature Verification System. *Proceedings of the ICDAR*, Vol. 1, pp. 424-429, 2003..