

## CRIAÇÃO DE BASE COM AMOSTRAS DE FALA E ESCRITA PARA RECONHECIMENTO BIOMÉTRICO DE PESSOAS

Stany Helberth de Souza Gomes da Silva (PIBIC/CNPq/FA/UEM), Yandre Maldonado e Gomes da Costa (Orientador) e Diego Bertolini (Coorientador), e-mail: ra112683@uem.br, yandre@din.uem.br e diegobertolini@gmail.com.

UEM/CTC/DIN.

**Área: Ciência da Computação.**

**Subárea: Metodologia e Técnicas da Computação.**

**Palavras-chave:** amostras de manuscrito, amostras de fala, reconhecimento de padrões.

### Resumo

Este trabalho tem como objetivo a criação de uma base de dados com amostras de fala e escrita com o objetivo de ajudar no desenvolvimento de técnicas e tarefas referente ao reconhecimento de padrões, em especial, biométrico de pessoas. A base é composta por dez amostras de cada um dos vinte voluntários, sendo cinco referentes à fala e outras cinco à escrita. Para a realização dos testes, inicialmente adotou-se uma abordagem de classificação da base inteira somente por parte das amostras de fala e de escrita isoladamente. Já na segunda abordagem, utilizou-se técnicas de fusão entre ambos tipos de amostras. Os resultados obtidos apontam que o uso da combinação entre os dois tipos de amostras propicia um aumento no desempenho de reconhecimento do indivíduo (autor/locutor). As melhores taxas de acurácia obtidas foram de 97,5% e 80% para as amostras de fala e escrita, respectivamente, e 97,5% com o uso da fusão tardia entre as amostras, ressaltando que, apesar de apresentar a mesma taxa para os resultados da fala, os autores acertados em ambos testes foram diferentes.

### Introdução

O crescimento de estudos referente à identificação de pessoas tem sido notado na área de reconhecimento de padrões, visto o grande número de características, em específico, biométricas, que o ser humano possui com unicidade. Assim, nota-se a necessidade de bases de dados, especialmente gratuitas e de fácil acesso ao público, a fim de apoiar o desenvolvimento de novas técnicas na área de reconhecimento de padrões. Dessa forma, foi desenvolvido neste trabalho uma base de dados ao admitir que tanto a fala quanto a escrita, características biométricas diferentes, podem ser utilizadas efetivamente na identificação de indivíduos e com o objetivo de fornecer uma base de dados contendo dois tipos de características de um mesmo

voluntário que possibilite novas investigações referentes a classificação de áudios e manuscritos.

## Materiais e métodos

**Extração de características** - Para a extração de características das amostras, foi utilizado, para os manuscritos, o extrator *Local Binary Pattern* (LBP), e, para os áudios, o descritor *Statistical Spectrum Descriptor* (SSD), ambos propostos por (Ojala *et al*, 2002) e (Lidy *et al*, 2005), respectivamente. A fim de obter melhores resultados quanto a performance dos manuscritos, utilizamos em nossos experimentos a técnica de divisão de imagens dos manuscritos em setores, como visto no trabalho de (M. Roberto e Souza *et al*), limitados em até três blocos. Após essa divisão, o descritor LBP foi aplicado em cada um dos setores, visto que tal técnica foi empregada apenas nas amostras de manuscrito.

**Classificadores** - Foi utilizado neste trabalho a biblioteca LIBSVM proposta por (CHANG; LIN, 2011), uma implementação do classificador *Support Vector Machine* (SVM), para a classificação das amostras. Em especial, os experimentos foram realizados usando os parâmetros escolhidos pelo *grid search*, com o *kernel Radial Basis Function* (RBF). Com o objetivo de combinar os dois tipos de amostras, realizamos as técnicas de fusão antecipada e tardia, no qual a primeira citada refere-se à combinação dos descritores LBP e SSD antes de realizar a classificação, enquanto que a segunda diz respeito ao uso das regras de Soma, Produto e Máximo.

## Resultados e Discussão

Para a obtenção de resultados referentes à identificação do voluntário, foram realizados três conjuntos de experimentos. Os dois primeiros referem-se a testes de classificação após o uso dos descritores SSD e LBP em cada amostra de áudio e manuscrito, respectivamente. Após isso, foi aplicada a estratégia de aprendizagem supervisionada, separando três amostras para treino e duas para teste em ambos tipos de amostra. Dessa forma, foi possível alcançar taxas satisfatórias, com precisão de 97,5% para os áudios e 80% para os manuscritos.

Já no terceiro conjunto de experimentos, utilizamos novamente o LIBSVM para gerar a matriz de probabilidade de cada tipo de amostra e, assim, aplicar as regras de Soma, Produto e Máximo, obtendo, respectivamente, resultados com 92,5%, 92,5% e 97,5% de acurácia.

Além disso, outros resultados aplicando diferentes técnicas como fusão antecipada e divisão entre setores dos manuscritos, podem ser vistos na Tabela 1.

**Tabela 1** - Resultados das classificações obtidas durante as aplicações de diferentes técnicas

	<i>Early Fusion</i>	<i>Late Fusion</i>	Divisão entre setores	Normal (áudios)	Normal (manuscritos)
Taxa de acurácia	90%	<b>97,5%</b>	77,5%	<b>97,5%</b>	80%

Vale ressaltar que apesar da taxa do Máximo ser igual à taxa do áudio inteiro, as amostras acertadas por ambas classificações são diferentes, mostrando que houve um aumento considerável ao combinarmos as duas fontes de características se comparado ao resultado obtido utilizando apenas o manuscrito.

## Conclusões

Neste trabalho, foi desenvolvido uma base de dados contendo amostras de fala e escrita de vinte diferentes autores, sendo cinco amostras de cada tipo, com o intuito de apoiar pesquisas e estudos na área de reconhecimento de padrões.

Ao término dos experimentos, conseguimos obter resultados satisfatórios e interessantes com a combinação entre ambas fontes de dados, visto que, apesar de apresentar a mesma taxa da classificação normal somente dos áudios, houve diferença na predição correta na amostra e, conseqüentemente, no voluntário, provando que o uso dessa combinação pode resultar em um aumento na performance de identificação de pessoas. Dessa forma, pretendemos em trabalhos futuros a ampliação de voluntários e amostras na base de dados, além de aplicar outros descritores, classificadores e técnicas de identificação como dissimilaridade e também seleção de classificadores a fim de buscar taxas de acerto ainda superiores às obtidas neste trabalho.

## Agradecimentos

Agradecemos a todos os voluntários que se prontificaram a participar da coleta de amostras e a Fundação Araucária pelo apoio financeiro fornecido.

## Referências

- T. Ojala, M. Pietikainen and T. Maenpaa, **Multiresolution gray-scale and rotation invariant texture classification with local binary patterns**, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, July 2002.
- Lidy, T. and A. Rauber. **Evaluation of Feature Extractors and Psycho-Acoustic Transformations for Music Genre Classification**. ISMIR, 2005.

CHANG, Chih-Chung; LIN, Chih-Jen. **LIBSVM: A library for support vector machines**. ACM transactions on intelligent systems and technology (TIST), v. 2, n. 3, p. 1-27, 2011.

M. Roberto e Souza, D. Bertolini, H. Pedrini, and Y. M. G. Costa, **Offline handwritten script recognition based on texture descriptors**, IEEE 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), 2019, pp. 57–62.