

An Exploratory Study of Biometrics using Trajectory Images of Eye Movements collected by Natural Image as Stimuli

Antonio Ricardo Alexandre Brasil¹, Luiz Alberto Pinto¹, Karin Satie Komati²

¹Programa de Pós-graduação em Engenharia de Controle e Automação (ProPECaut)
IFES - Serra - ES - Brasil

²Programa de Pós-graduação em Computação Aplicada (PPComp)
IFES - Serra - ES - Brasil

anribrasil@gmail.com, {luiz.pt, kkomati}@ifes.edu.br

Abstract. *The study of eye movements as biometrics began in 2004 and its typical process is to capture movements using stimulus images, treating these data as signals in time followed by classification. This work proposes two differentials, one is the use of the DOVES database of eye movements collected using natural images as a stimulus, and the other is the treatment of eye movements as images. Images of the eye movement trajectories are generated, which are the input of two types of architectures: one which uses the HOG and LBP techniques for extracting characteristics, followed by the SVM, MLP, and random forest classifiers and other which uses the power of Convolutional Neural Network (CNN) VGG-19 and ResNet-50 for person identification. The highest accuracy was 39.59% with the ResNet-50 architecture, a result comparable with the best of EMVIC 2014 competition, which used face images as a stimulus, of 39.63%.*

Resumo. *O estudo dos movimentos oculares como biometria começou em 2004 e seu processo típico é capturar movimentos usando imagens de estímulo, tratando esses dados como sinais no tempo seguido de classificação. Este trabalho propõe dois diferenciais, um é o uso do banco de dados DOVES de movimentos oculares coletados usando imagens naturais como estímulo e o outro é o tratamento de movimentos oculares como imagens. São geradas imagens das trajetórias de movimentos oculares, que são a entrada de dois tipos de arquiteturas: uma que usa as técnicas HOG e LBP para extrair características, seguida pelos classificadores SVM, MLP e RF e outra que usa o poder do Convolutional Rede Neural (CNN) VGG-19 e ResNet-50 para a identificação de pessoas. A maior precisão foi de 39,59% com a arquitetura ResNet-50, um resultado comparável ao melhor da competição EMVIC 2014, que usou imagens de rosto como estímulo, de 39,63%.*

1. Autores

Informações solicitadas na chamada:

1. Versão online da dissertação <http://biblioteca.ifes.edu.br:8080/pergamumweb/vinculos/00001a/00001a37.pdf>
2. Trabalhos publicados: [Brasil et al. 2020a] e [Brasil et al. 2020b]
3. Outros trabalhos publicados durante o mestrado: [Brasil et al. 2018] [Ceravolo et al. 2019] [Rodrigues et al. 2018] [Brasil et al. 2019]

2. Introdução

Biometria é o estudo da identificação de pessoas baseada em características físicas ou comportamentais, como a face, impressões digitais, voz ou íris [Jain et al. 2007]. Uma das premissas de sistemas biométricos é de que as características físicas ou comportamentais das pessoas são distintas, e que podem ser capturadas por meio de sensores, tornando automático o reconhecimento de um indivíduo.

Os movimentos oculares, que é uma característica comportamental de um indivíduo, foram utilizados na última década para o reconhecimento de pessoas [Kasprowski 2004, Rigas et al. 2012]. As vantagens do uso de movimentos oculares como forma de biometria é que: (i) são extremamente difíceis de serem replicados e (ii) não podem ser roubados de um indivíduo. Esses benefícios são uma consequência da sua característica comportamental, pois os movimentos oculares são involuntários e não podem ser replicados pelos métodos atuais [Komogortsev et al. 2010]. Acrescenta-se ainda como vantagem o uso de autenticação contínua de um usuário, durante todo o tempo de uso do sistema, não só no momento de entrada/início do sistema.

A Figura 1 apresenta os movimentos sacádicos através das linhas e a fixação por pequenos círculos (na figura à direita), enquanto uma pessoa observava a foto da face (figura à esquerda). Nota-se que a pessoa concentrou os movimentos nos olhos, no nariz e na boca da face, o círculo vermelho ressalta a aglomeração de movimentos curtos realizados.

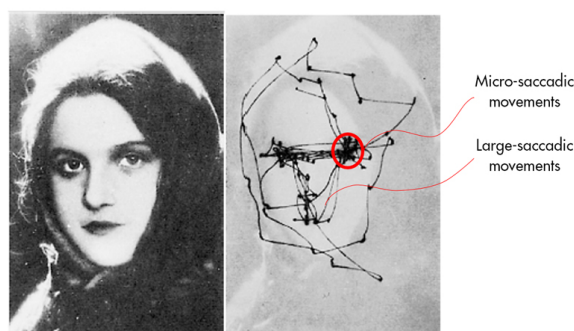


Figura 1. As linhas nesta imagem mostram os movimentos sacádicos do olho de uma pessoa enquanto ela olhava para esse rosto. Imagem sob licença CC BY-SA 4.0.

Na literatura a arquitetura comum de identificação de pessoas por meio do uso de movimentos oculares se inicia com a captura, utilizando um aparelho chamado *Eye Tracker* que coleta todos os movimentos oculares realizados pelo participante do experimento. Estímulos visuais são apresentados na tela de um computador, sendo a mais comum a apresentação de pontos saltitantes [Holland and Komogortsev 2011]. A partir dos movimentos coletados do participante, passa-se à classificação de movimentos oculares, separando-os em tipos de movimentos, tais como os de fixação e sacádicos.

Uma vez realizada a etapa de classificação dos movimentos, segue-se com a etapa de extração das características dos movimentos, onde são obtidas métricas como a velocidade do movimento, duração da fixação, e amplitude da sacada [Rigas et al. 2018] [Ceravolo et al. 2019]. Podem ser usados duas formas de identificação de pessoas com os

vetores de características: comparação das características ou classificação. A comparação usa medidas de similaridade/distância, que buscam encontrar o valor máximo/mínimo entre dois vetores [Rigas and Komogortsev 2014b], já a classificação constrói um modelo de predição de identificação de pessoas, baseado numa fase de treinamento com dados que são conhecidos anteriormente.

Este trabalho propõe dois diferenciais em relação ao método descrito acima:

1. Realizar o reconhecimento biométrico de pessoas utilizando as imagens das trajetórias dos seus movimentos oculares em duas abordagens. Na primeira abordagem, o passo “Classificação de Movimentos Oculares” será removido, e a entrada para o passo de extração de características será a imagem resultante dos movimentos, tais como as linhas da Figura 1. A segunda abordagem consiste em eliminar os passos: “Classificação de Movimentos Oculares”, “Extração de Características” e “Comparação de Características ou Classificador”, sendo estes realizados pelas técnicas de CNN.
2. Os pontos saltitantes na tela são os mais usados como estímulo visual, pois mantém um controle, direcionando para onde o indivíduo deve olhar e com isso, gera uma diferenciação maior dos movimentos dos participantes [Rigas and Komogortsev 2017]. No entanto, acredita-se que para uma aplicação de reconhecimento de uma pessoa sem que ela perceba que está sendo avaliada, um estímulo natural é necessário. Assim, outro diferencial deste trabalho é o uso da base de dados DOVES [Van Der Linde et al. 2009], que usa imagens naturais em escala de cinza como estímulo visual para a captura dos movimentos dos olhos.

Foi realizada uma busca no sistema Scopus [Brasil et al. 2020a] onde encontrou-se 200 artigos na língua inglesa. Dentre eles, duas propostas eram similares. A proposta feita por Rigas e Komogortsev [Rigas and Komogortsev 2014a, Rigas and Komogortsev 2014b], em que também usavam imagem dos movimentos dos olhos. No entanto, as imagens foram processadas para um mapa de densidade, similar à um mapa de calor (*heat maps*), onde há maior densidade quanto maior o tempo de fixação. Os trabalhos usaram vídeos e rostos como estímulos. O trabalho de Li e colegas [Li et al. 2018] usaram imagens de rastreamento como entrada do sistema. A diferença foi no estímulo visual, eram apresentadas questões objetivas e coletavam-se os movimentos dos olhos na busca da resposta dentre opções apresentadas, que eram sempre números em caracteres chineses.

Além dos trabalhos de Rigas e Komogortsev [Rigas and Komogortsev 2014a, Rigas and Komogortsev 2014b], outros trabalhos encontrados que usam imagem natural como estímulo foram os que participaram da competição EMVIC 2014¹ [Kasprowski and Harežlak 2014], cuja base de dados é composta de imagens de faces humanas como estímulos. Além dessa competição, o trabalho de Bednarik e colegas [Bednarik et al. 2005] utilizou diferentes estímulos para capturar os movimentos, incluindo imagens naturais em escala de cinza.

Assim, este trabalho propõe o estudo de uma nova abordagem para o reconhecimento biométrico de pessoas utilizando as imagens das trajetórias dos seus movimentos oculares através de estímulos por imagens naturais. Não será feita a coleta de dados, pois

¹Eye Movement Verification and Identification Competition

será usada a base de dados DOVES [Van Der Linde et al. 2009], que usa imagens naturais como estímulo. As duas abordagens serão detalhadas a seguir:

1. Primeira abordagem: Para extração das características das imagens dos movimentos oculares dos participantes foram utilizadas duas técnicas clássicas de PDI: histograma de gradientes orientados (HOG) [Dalal and Triggs 2005] e padrão binário local (LBP) [Ojala et al. 2002]. A partir das características extraídas das imagens foram usadas três técnicas diferentes de aprendizado supervisionado [Gama et al. 2011]: máquina de vetor de suporte (SVM), floresta aleatória (RF) e redes neurais perceptron de múltiplas camadas (MLP).
2. Segunda abordagem: A segunda abordagem consistiu na aplicação da técnica de rede neural convolucional com aprendizado profundo. A diferença entre a abordagem anterior é que não há mais duas etapas (extração de características e AM) e apenas uma única, dada pelos algoritmos de CNN. Foi adotado duas técnicas clássicas da literatura: VGG-19 [Russakovsky et al. 2015] e ResNet-50 [He et al. 2016] com o uso da técnica transferência do conhecimento [Brownlee 2017].

Este artigo está dividido da seguinte forma: na Seção a discute os experimentos e os resultados, e por fim, a conclusão e os trabalhos futuros são apresentados na Seção 4.

3. EXPERIMENTOS E RESULTADOS

Todos os experimentos foram desenvolvidos e avaliados utilizando Python e a biblioteca Scikit-Learn e PyTorch.

3.1. Primeira arquitetura

Como o SVM é um classificador binário, é necessário usar um processamento a mais para ser utilizado em problemas multi-classes. Foi utilizada a abordagem um contra um (*one-to-one*), foram construídos $c * (c - 1) / 2$ classificadores SVMs para treinar todas as possíveis combinações de classes. Onde c é o número total de classes do problema, sendo neste trabalho, o valor de $c = 29$. Para avaliação do resultado do SVM, foram usados dois diferentes *kernels*, o RBF (do inglês *radial basis function*) e o Linear.

A base de dados foi separada em 75% para treino e 25% para teste, sem validação cruzada, tomando-se o cuidado de manter o mesmo conjunto de dados para todos os classificadores. Neste trabalho, para avaliar os resultados, foram consideradas três métricas: taxa falsa de aceitação (FAR), taxa falsa de rejeição (FRR) e acurácia (ACC). A FAR é uma métrica que calcula a porcentagem de indivíduos erroneamente autorizados pelo sistema; já a FRR é uma métrica que calcula a porcentagem de indivíduos erroneamente negados pelo sistema [Saeed 2016].

A Tabela 1 contém os resultados com ambos os métodos de extração de características, HOG e LBP, e com os diferentes classificadores usados neste trabalho. A melhor acurácia usando o extrator HOG, de 33,3%, foi obtida com a técnica de MLP com uma FAR de 59,3% e uma FRR de 71,3%. Já para a técnica LBP, os melhores resultados se mantiveram com o classificador MLP, com uma acurácia de 32,3% e um FAR de 60,2% e um FRR de 72,1%.

Tabela 1. Resultados obtidos neste trabalho utilizando os extratores HOG e LBP e as técnicas de aprendizado de máquina SVM, RF e MLP. Todos os resultados apresentados estão em porcentagem (%). O melhor resultado está em negrito.

Classificador	HOG			LBP		
	ACC	FAR	FRR	ACC	FAR	FRR
SVM (linear)	30,4	61,9	73,8	30,4	61,9	73,8
SVM (RBF)	29,4	62,8	74,6	29,4	62,8	74,6
Random Forest	25,7	66,1	77,8	24,3	67,3	79,0
MLP	33,3	59,3	71,3	32,3	60,2	72,1

Tabela 2. Comparação do melhor resultado obtido pela técnica de MLP com alguns resultados encontrados na literatura, onde SIM é a abreviação do uso de métricas de similaridade e Ind é o número de voluntários.

	técnica	ACC (%)	Rank-1	Ind.
Este trabalho	HOG+MLP	33,3	–	29
EMVIC 2014		39,6	–	56
Bednarik et al. 2005	Estáticas+KNN	40 a 50	–	12
Bednarik et al. 2005	Dinâmicas+KNN	90	–	12
Rigas et al. 2014	FDM+KL	–	51	200
Li et al. 2015	GWT+KNN	–	96,4	58

Não foi possível obter uma taxa de erro abaixo de 59% para a métrica FAR. A acurácia baixa pode ter sido influenciada por diversos fatores, tais como a baixa quantidade de imagens utilizadas para treinamento (101) ou até mesmo os extratores, que não conseguiram identificar os diferentes tipos de movimentos. Os valores de FAR e FRR são considerados ruins para um sistema comercial, mas estão equiparados com alguns trabalhos de abordagens diferenciadas apresentados na Seção 2.

Considera-se que os resultados (acurácia de 33,3%) foram encorajadores quando comparados com os resultados obtidos na EMVIC 2014 [Kasprowski and Harežlak 2014] (acurácia de 39,6%). A Tabela 2 apresenta uma comparação do melhor resultado obtido por este trabalho, pela competição EMVIC 2014 e pelos trabalhos similares, onde a primeira coluna tem a identificação do artigo, a segunda coluna qual a combinação de técnica de melhor resultado, a terceira a acurácia (ACC), seguido do *Rank-1* e por último a quantidade de indivíduos da base de dados (Ind.).

O resultado da acurácia deste trabalho se equipara ao resultado de dados estáticos do trabalho de Bednarik e colegas, mas ficou aquém quando comparado com o resultado da abordagem dinâmica. Relembrando que foram usados diferentes estímulos para captura dos movimentos e não apenas imagens naturais em escala de cinza. Da mesma forma, o resultado foi pior que o de Li e colegas, mas este também usou imagens estímulo diferentes.

O trabalho de Rigas e Komogortsev que utilizou em sua abordagem o uso de mapas de densidade de fixação, trabalhou com três tipos de estímulos: rostos, vídeos e textos. Para melhor comparação, adotamos o melhor resultado do autor usando o estímulo de rostos, que mais se assemelha com os estímulos em imagens naturais. O resultado deste trabalho foi pior que o resultado de Rigas e Komogortsev que apresentou acurácia de 96,4%.

3.2. Segunda abordagem

Para trabalhar com CNN, o ideal é ter uma grande base de dados para o treinamento. Uma vez que a base atual contém apenas 101 registros de imagens para 29 participantes, resultando em uma base de dados com 2.929 imagens, a divisão em treino, teste e validação seguiu 80% para treino, 10% para validação e 10% para testes.

Tabela 3. Experimentos realizados com as arquiteturas VGG-19 e ResNet-50.

CNN	ACC	FAR	FRR
VGG-19	36,51%	2,32%	78,85%
ResNet-50	39,59%	2,29%	77,86%

A Tabela 3 apresenta os resultados obtidos com as CNNs VGG e ResNet, com seus respectivos valores de acurácia, FAR e FRR. Nota-se um valor maior da acurácia comparado com os resultados obtidos pelas técnicas clássicas anteriormente (39,59% da ResNet comparado ao 30,42% do SVM RBF).

Assim, a segunda arquitetura baseada em CNN apresentou melhores resultados que a primeira arquitetura com técnicas clássicas. Em comparação com os resultados encontrados nas competições, eles se mostraram bastante encorajadores. O resultado obtido pela EMVIC 2014 [Kasprowski and Harežlak 2014] que conseguiu uma acurácia de 39,63%, em comparação com a acurácia obtida por este trabalho, de 39,59%, se mostraram comparáveis. Além disso, se mostraram bastante encorajadores, uma vez que foram aplicadas técnicas diferentes às existentes na literatura pelos participantes das competições. Entretanto, é importante ressaltar a diferença de ambas as abordagens, que possuem diferentes voluntários, diferentes técnicas, e que podem, como citado anteriormente, ter afetado os resultados.

4. CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho realizou um estudo exploratório de biometria usando a imagem da trajetória dos movimentos oculares coletados por estímulo visual em imagens naturais. Assim, uma contribuição do trabalho é abordar uma solução baseada em imagens dos movimentos oculares e não nos sinais de variação do ângulo do olho em relação ao tempo.

Muitos dos trabalhos adotam dados coletados via estímulos controlados. Outra contribuição foi o uso de uma base de dados em que se usou imagens naturais como estímulo, sem qualquer tipo de controle. A DOVES contém dados de 29 participantes e 101 imagens naturais.

A primeira arquitetura usou técnicas clássicas de extração de característica das imagens (HOG e LBP) seguido de classificadores (floresta aleatória e SVM). Na segunda arquitetura, o uso de métodos de CNN foram aplicadas. Em ambas, foi feita a comparação da ACC, FAR e FRR para comparação quantitativa. Concluiu-se que a melhor abordagem, em termos de acurácia, foi a da CNN com ResNet-50, que conseguiu resultado maior comparado com a primeira abordagem e comparada com o outro método de CNN. O melhor resultado obteve ACC=39,59%, FAR=2,29% e FRR=77,86%. O valor de FAR são considerados satisfatórios para um estudo exploratório, comparado com os existentes na literatura.

Como trabalhos futuros, é possível calibrar os parâmetros das redes neurais convolucionais, o uso de aumento de dados (*data augmentation*) na base existente. Outra possibilidade seria o desenvolvimento de uma arquitetura CNN nova, sem uso de transferência do conhecimento. Além disso, outras formas de geração de imagens de rastreamento podem ser investigadas, tal como espectogramas dos sinais. Quando as competições disponibilizarem as bases de dados de forma pública, os experimentos desta proposta serão replicados para comparação.

Referências

- Bednarik, R., Kinnunen, T., Mihaila, A., and Fränti, P. (2005). Eye-movements as a biometric. *Image analysis*, pages 16–26.
- Brasil, A. R. A., Andrade, J. O., and Komati, K. S. (2020a). Eye movements biometrics: A bibliometric analysis from 2004 to 2019. *International Journal of Computer Applications*, 176(24):1–9.
- Brasil, A. R. A., Andrade, J. O., Pinto, L. A., and Komati, K. S. (2020b). An exploratory study of biometrics using eye movement trajectory images collected by natural image stimuli. In *Anais da XLVII edição do Seminário Integrado de Software e Hardware*, pages 25–36, Porto Alegre, RS, Brasil. SBC.
- Brasil, A. R. A., Pinto, L. A., and Komati, K. S. (2018). Predição da tendência de valorização do bitcoin usando técnicas de aprendizado de máquina. In *Anais Estendidos do XIV Simpósio Brasileiro de Sistemas de Informação*, pages 126–128. SBC.
- Brasil, A. R. A., Pinto, L. A., and Komati, K. S. (2019). Comparação de técnicas de Aprendizado de máquina na Predição da Tendência de Valorização da BITCOIN. In *Information Systems and Technology Management*, pages 244–254. Antonella Carvalho de Oliveira.
- Brownlee, J. (2017). A gentle introduction to transfer learning for deep learning.
- Ceravolo, I. A., Brasil, A. R. A., and Komati, K. S. (2019). Classificação de dislexia a partir de movimentos oculares durante a leitura usando aprendizado de máquina e wavelets. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 880–891. SBC.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.
- Gama, J., Faceli, K., Lorena, A. C., and De Carvalho, A. C. P. L. F. (2011). *Inteligência artificial*. LTC, Rio de Janeiro.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Holland, C. and Komogortsev, O. V. (2011). Biometric identification via eye movement scanpaths in reading. In *2011 International joint conference on biometrics (IJCB)*, pages 1–8. IEEE.
- Jain, A. K., Flynn, P., and Ross, A. A. (2007). *Handbook of biometrics*. Springer Science & Business Media.

- Kasprowski, P. (2004). Human identification using eye movements. *Institute of Computer Science*.
- Kasprowski, P. and Harezlak, K. (2014). The second eye movements verification and identification competition. In *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, pages 1–6. IEEE.
- Komogortsev, O. V., Jayarathna, S., Aragon, C. R., and Mahmoud, M. (2010). Biometric identification via an oculomotor plant mathematical model. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 57–60. ACM.
- Li, C., Xue, J., Quan, C., Yue, J., and Zhang, C. (2018). Biometric recognition via texture features of eye movement trajectories in a visual searching task. *PloS one*, 13(4):e0194475.
- Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987.
- Rigas, I., Economou, G., and Fotopoulos, S. (2012). Biometric identification based on the eye movements and graph matching techniques. *Pattern Recognition Letters*, 33(6):786–792.
- Rigas, I., Friedman, L., and Komogortsev, O. (2018). Study of an extensive set of eye movement features: Extraction methods and statistical analysis. *Journal of Eye Movement Research*, 11(1):3.
- Rigas, I. and Komogortsev, O. V. (2014a). Biometric recognition via fixation density maps. In *Biometric and Surveillance Technology for Human and Activity Identification XI*, volume 9075, page 90750M. International Society for Optics and Photonics.
- Rigas, I. and Komogortsev, O. V. (2014b). Biometric recognition via probabilistic spatial projection of eye movement trajectories in dynamic visual environments. *IEEE Transactions on Information Forensics and Security*, 9(10):1743–1754.
- Rigas, I. and Komogortsev, O. V. (2017). Current research in eye movement biometrics: An analysis based on BioEye 2015 competition. *Image and Vision Computing*, 58:129–141.
- Rodrigues, D. S., Brasil, A. R. A., Costa, M. B., Komati, K. S., and Pinto, L. A. (2018). A comparative analysis of loan requests classification algorithms in a peer-to-peer lending platform. In *Proceedings of the XIV Brazilian Symposium on Information Systems*, page 42. ACM.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252.
- Saeed, K. (2016). *New directions in behavioral biometrics*. CRC Press.
- Van Der Linde, I., Rajashekar, U., Bovik, A. C., and Cormack, L. K. (2009). Doves: a database of visual eye movements. *Spatial vision*, 22(2):161–177.