

# Identificação Biométrica Através de Micromovimentos dos Olhos

Lucca A. Martins<sup>1</sup>

<sup>1</sup>Departamento de Computação – Universidade Federal de Ouro Preto  
35400-000 – Minas Gerais – MG – Brazil

lucca.martins@aluno.ufop.edu.br

**Abstract.** *This paper describes the use of a deep CNN for biometric identification. The input data are sample records with duration of at least 1 second, recorded by an eye-tracking device while the users were reading. The goal of the model is to identify to which known user a record belongs, using the angular velocities of the yaw and pitch axis in each milisecond of the sample data.*

**Resumo.** *Este artigo descreve a utilização de uma rede neural convolucional profunda para identificação biométrica. Os dados de entrada da rede são gravações de a partir de 1 segundo de rastreamento ocular de usuários enquanto realizavam leituras. O objetivo da rede é conseguir identificar a que usuário uma dada gravação pertence, a partir das velocidades angulares horizontais e verticais do olho em cada milissegundo de gravação.*

## 1. Introdução

Os movimentos dos olhos são objetos de interesse em diversas áreas de estudo, desde diversos ramos da psicologia, oftalmologia e optometria, até variadas aplicações com auxílio da computação podendo envolver avaliação de sistemas interativos com interface gráfica, jogos virtuais e no nosso caso, identificação biométrica. Com uma série de estudos foi comprovado que assim como as digitais que possuímos nos dedos das mãos são únicas para cada indivíduo, os macromovimentos dos olhos também são [Bargary 2016]. Tais macromovimentos são classificados em 3 classes: fixações, movimentos sacádicos e movimentos suaves (em tradução livre para *smooth pursuit*). Adquirido esse conhecimento, surge então o estudo de aplicações para biometria através dos movimentos dos olhos. Dependendo das intenções práticas para com a biometria dos olhos, como no caso de controle de acesso a dispositivos ou locais de entrada restrita, o tempo demandado gasto na identificação biométrica correta precisa ser curto para que a solução não se torne inviável.

Além dos movimentos citados, existem também os micromovimentos dos olhos, que ocorrem involuntariamente (suas utilidades para a nossa visão também são fontes de estudo), são eles: movimentos microscádicos e, entre cada um desses, deslizos e tremores [Yarbus 1967]. No estado da arte existem evidências de que os movimentos microscádicos também possuem diferenças mensuráveis de um indivíduo para outro [Poynter 2013]. Esses movimentos ocorrem numa frequência maior, e por isso são mais relevantes para os estudos que objetivam desenvolver uma identificação biométrica que necessite de um curto período de rastreamento ocular. Além disso, o fato de serem involuntários garantem a imparcialidade por parte dos indivíduos ao serem submetidos ao rastreamento ocular.

O presente trabalho busca replicar o estudo realizado em [Lena A. Jager and Scheffer 2020], nomeado como DeepEyedentification, cuja motivação se dá pelo fato de que todos os estudos prévios sobre identificação biométrica extraem apenas características mais explícitas dos movimentos dos olhos, como a duração das fixações, padrões de variação entre os movimentos e amplitude dos movimentos sacádicos. Estes estudos analisam apenas os macromovimentos, seja por limitações dadas pela abordagem utilizada para o rastreamento ocular (e.g. métodos baseados em processamento de vídeos) ou por peculiaridades das devidas aplicações. Visando potencializar a identificação biométrica, o estudo utiliza uma rede neural convolucional que processa os sinais do rastreador ocular (o dispositivo propriamente dito) e prediz a qual indivíduo aquele sinal pertence.

## 2. Caracterização do Problema

A estratégia de funcionamento dos sistemas de identificação biométrica a partir de sequências de sinais de rastreamento ocular podem variar conforme demanda a aplicação. Aqui focamos em três dessas variações, e é importante ressaltar algumas diferenças a fim de elucidar as particularidades e limitações de cada possível solução.

No primeiro caso, temos que cada indivíduo a ser identificado já foi submetido a uma ou mais sessões de rastreamento ocular, cujas sequências de sinais gravados são utilizados para o treinamento de um modelo multi-classificador. O modelo tem como objetivo classificar uma nova sequência de sinais obtida em tempo de execução como sendo pertencente a um dos usuários previamente presentes na população utilizada para o treinamento. Portanto, o modelo desta abordagem é limitado a reconhecer apenas indivíduos que já foram inseridos no sistema. Caso um novo usuário precisa ser cadastrado, o modelo precisaria voltar para a fase de treinamento incluindo sequências de sinais pertencentes a este novo usuário.

Numa segunda abordagem, considerando a intenção de incluir novos usuários sem o tormento da necessidade de treinar o modelo novamente, um novo usuário poderia ser inserido simplesmente armazenando-se no sistema sua sequência de sinais gravada instantaneamente. Neste caso, o modelo deve ser treinado de forma a ajustar uma métrica de similaridade que será utilizada para se comparar uma nova sequência de entrada com as sequências armazenadas de cada um dos usuários. Caso a similaridade encontrada não seja satisfatória, a sequência de entrada pode ser considerada como pertencente a um novo usuário e ser utilizada em seguida numa nova identificação.

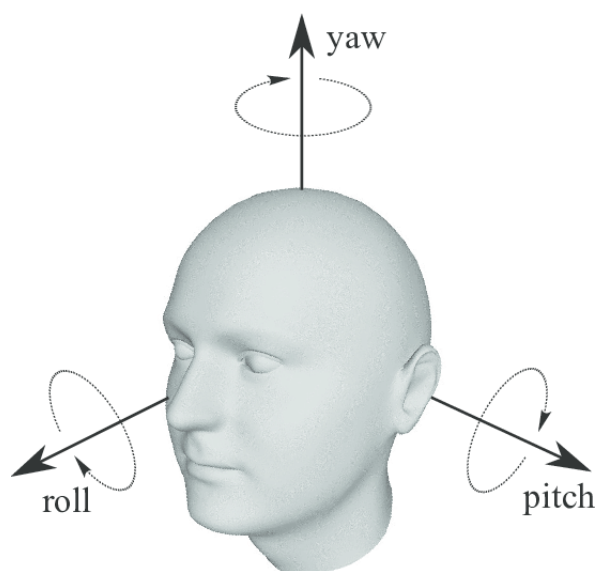
No último caso, temos que o sistema objetiva identificar um único indivíduo. Logo, o modelo é treinado apenas com sequências de sinais que pertencem a tal usuário, e deve se tornar capaz de verificar em tempo de execução se uma nova sequência de entrada pertence ou não ao usuário cadastrado.

O modelo presente neste trabalho utiliza a primeira abordagem descrita, desenvolvendo um modelo multi-classificador que utiliza sequências de sinais de entrada que contêm a angulação dos olhos em relação ao eixo vertical (*yaw*) e ao eixo horizontal no sistema de coordenadas angulares, (*pitch*), vide Figura 1 <sup>1</sup>.

---

<sup>1</sup>Disponível em <https://www.researchgate.net/profile/Alberto-Fernandez-Villan/publication/309543534/>

<https://www.researchgate.net/profile/Alberto-Fernandez-Villan/publication/309543534/>



**Figure 1. Sistema de coordenadas angulares**

### 3. Comparando Tecnologias

Esta seção tem como foco comparar os variados métodos para identificação biométrica e suas diferenças para com a biometria através dos movimentos oculares. A primeira grande vantagem é que métodos como biometria digital ou escaneamento de veias da mão, são relativamente mais intrusivos, pois exigem uma ação específica do usuário (como por exemplo, levar o dedo até o dispositivo de biometria). Diferentemente da biometria através de reconhecimento facial, escaneamento de íris ou rastreamento de micromovimentos, em que o usuário não precisa realizar nenhuma interação explícita com o sistema.

Uma outra questão são as limitações de aplicabilidade das tecnologias de biometria. Por exemplo, as digitais dos dedos podem ser ofuscadas ou suavizadas quando o indivíduo é um trabalhador braçal e isso claramente influencia negativamente na qualidade da identificação. Em relação à biometria através de reconhecimento de íris, uma das possíveis complicações é quando o indivíduo a ser identificado faz uso de lentes de contato. As lentes de contato são uma interferência negativa na performance dos sistemas de identificação, pois ofuscam a imagem real da íris. Portanto, nesse sentido a biometria através de movimentos oculares é a mais vantajosa, pois satisfaz as condições de identificação para uma base maior de indivíduos.

Outra questão bastante interessante é que a técnica abordada por este estudo é a mais difícil de ser sabotada dentre as demais conhecidas. No reconhecimento facial, por exemplo, um usuário pode facilmente se passar por outro ao utilizar uma fotografia deste outro usuário como entrada para a identificação. No caso de reconhecimento da íris, vários usuários podem ser reconhecidos como um só se estiverem utilizando uma mesma lente de contato. Já no caso dos métodos baseado nos movimentos dos olhos, essa sabotagem é muito mais complexa. O sabotador precisaria exibir um vídeo do usuário pelo qual ele quer se passar, e este vídeo precisaria de 1000fps para que o sistema reconhecesse corretamente.

Existem algumas técnicas para tornar os sistemas seguros quanto a essas tentativas de ataque. Um exemplo seria, no caso do uso de uma fotografia alheia, a falta de movimento poderia ser utilizada para acusar uma inconsistência na entrada de dados, bloqueando o sistema de identificação. Ainda assim, o sabotador poderia utilizar também um vídeo do outro usuário, e essa situação requer um tratamento muito mais complexo, com sensores e um investimento maior em segurança. Portanto, com todas essas possibilidades citadas, podemos notar que a biometria através dos micromovimentos oculares é, por si só, a mais segura dentre as demais técnicas apresentadas.

$$t_s(\delta_i^x, \delta_i^y) = (\tanh(c\delta_i^x), \tanh(c\delta_i^y))$$

Figure 2. Equação para transformação das baixas velocidades

$$t_f(\delta_i^x, \delta_i^y) = \begin{cases} z(0) & \text{if } \sqrt{\delta_i^{x2} + \delta_i^{y2}} < \nu_{min} \\ (z(\delta_i^x), z(\delta_i^y)) & \text{otherwise} \end{cases}$$

Figure 3. Equação para transformação das altas velocidades

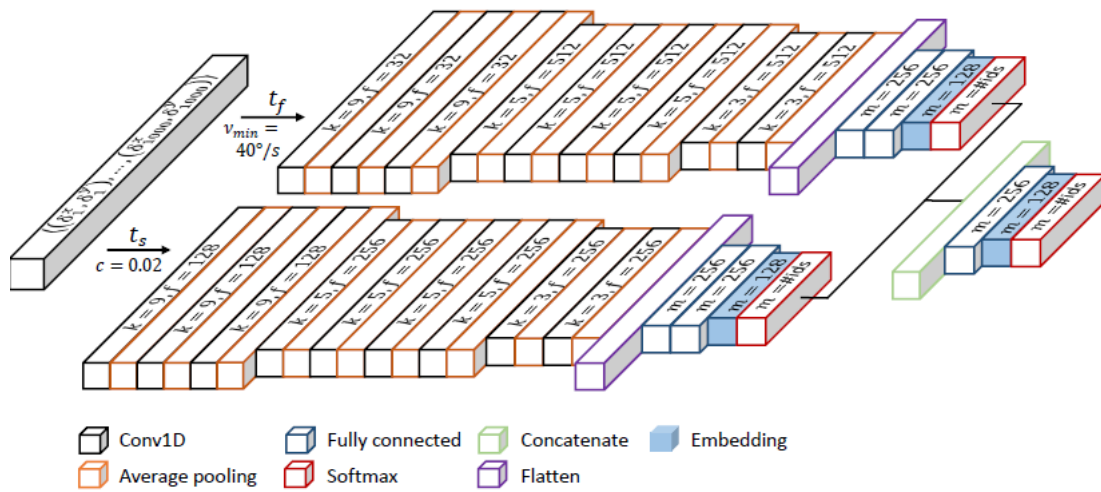
### 3.1. Arquitetura da Rede

O modelo de Rede Neural Convolutacional Profunda a ser apresentado foi batizado de DeepEyedentification. E, como será apontado posteriormente, supera todos os modelos previamente desenvolvidos com a mesma finalidade. As sequências de sinais colhidas pelo eye-tracking que gerou a base de dados consistem em registros dos ângulos absolutos em relação a cada eixo mencionado anteriormente (*yaw* e *pitch*). Porém, para alimentar a rede esses ângulos absolutos são transformados em velocidades angulares em °/s, isso porque através dessas velocidades a diferenciação dos tipos de movimentos se torna mais nítida.

Uma normalização global das velocidades geraria problema em relação às velocidades dos micromovimentos de deslize e tremor, pois estas seriam reduzidas a quase zero e surtiriam muito pouco efeito no treinamento da rede, que seria capaz de extrair apenas características de macromovimentos. Para contornar isso, a arquitetura da rede é composta por duas sub-redes independentes: uma com foco em movimentos rápidos e outras com foco em movimentos mais lentos. Ambas as sub-redes são exatamente iguais em termos de quantidade de camadas, e realizam a predição sobre sequências que correspondem a 1 segundo de gravação do eye-tracking. A diferença é que cada uma delas analisa essas sequências em escalas diferentes, para isso, as sequências passam por uma transformação antes de alimentarem a rede.

Para a rede focada nas baixas velocidades, a equação que transforma a entrada aplica a tangente hiperbólica sobre cada par de velocidades multiplicando pelo parametro  $c$  (definido entre 0.01 e 0.06). A Figura 2 apresenta esta equação. Dessa forma, as velocidades que aparecem em movimentos de deslize e tremor são alocadas em valores entre -0.5 e +0.5, enquanto velocidades de movimentos sácádicos e microssacádicos ficam entre -1 e -0.5 ou 0.5 e 1, dependendo da direção do movimento.

Já a sub-rede que age sobre as altas velocidades utiliza a equação representada na Figura 3. Neste caso,  $z$  corresponde à normalização  $z$ -score, e a equação efetua o truncamento das velocidades que estão abaixo do limite mínimo de velocidade para movimentos considerados sacádicos ou microssacádicos. O valor de  $V_{min}$  é definido entre o intervalo de 10 a 60°/s, conforme prega a literatura de estudos psicológicos.



**Figure 4. Arquitetura da rede utilizada**

Cada sub-rede é composta por 9 camadas convolucionais unidimensionais, seguidas sempre por uma camada de ativação ReLu, uma cada de *BatchNormalization* e uma camada de *AveragePooling*, também unidimensional. Após as camadas convolucionais, encontra-se uma camada de achatamento (*Flatten*), seguidas por três sequências de uma camada densa com uma camada de *BatchNormalization*, e por fim, mais uma camada densa com 75 unidades e ativação Softmax. As saídas de ambas as redes alimentam a rede que as concatena, composta pelas seguintes camadas: uma camada de concatenação, e duas sequências de uma camada densa seguida por uma de *BatchNormalization* e uma de ativação ReLu, e por fim, uma camada densa com 75 unidades e ativação Softmax, que correspondem aos 75 usuários presentes no conjunto de rótulos. A Figura 2 apresenta graficamente o modelo utilizado. As sub-redes são treinadas separadamente, e ao final, seus pesos são concatenados para serem utilizados pela rede como um todo.

#### 4. Treinamento e Resultados

As configurações para o treinamento das redes foram definidas utilizando o Tensorflow na versão 1.12.0 e Keras na versão 2.2.4. Para as sub-redes foi utilizado o otimizador Adam com taxa de aprendizado de 0.001, e para a rede principal (camadas compartilhadas) utilizou-se o mesmo otimizador mas com taxa de aprendizado de 0.0001. Em ambos os casos, o *batch size* definido foi de 64 sequências.

O dataset de treinamento utilizado é chamado de Postdam Textbook Corpus, e se constitui de uma população de 75 indivíduos cujas sequências de rastreamento ocular foram coletadas enquanto estes liam diferentes textos aleatórios. Como dito anteriormente, as sequências são divididas em janelas de 1000ms, e essas janelas é que são usadas como entrada para os modelos. Ainda assim, o modelo também foi alimentado com sequências mais longas e obteve resultados extremamente satisfatórios. Para sequências de 1 segundo a acurácia foi de 91.4%, para 10 segundos a acurácia foi de 99.77%, e para sequências de 40 segundos a acurácia foi de 99.86%. Em comparação com outros estudos, que focavam na análise de macromovimentos, os resultados comprovam que os micromovimentos contêm informações mais relevantes para o dado problema.

Como foi citado, as sub-redes foram treinadas separadamente, e também obtiveram bons resultados, mas que foram facilmente superados pela rede completa. A sub-rede para movimentos rápidos obteve acurácia de 77%, e a sub-rede para movimentos lentos resultou em uma acurácia de 88%. Ambos os valores são para sequências com duração de 1 segundo.

Como experimento adicional, uma variante do modelo foi criada para reproduzir a arquitetura de forma a utilizar uma única rede, sem o pré-processamento das velocidades através das equações representadas na Figura 2 e Figura 3, passando apenas por uma normalização global dos dados. Foi observado que esta rede não alcança sequer a performance da sub-rede de movimentos rápidos (77% de acurácia), isso porque ela extrai apenas os padrões de macromovimentos, o que acaba não gerando uma performance à altura. Além disso, foi observado também que indicar se cada parte da sequência pertence a uma fixação ou um movimento sacádico não influencia no desempenho deste modelo.



## **5. Conclusão**

Com os resultados obtidos, podemos concluir que a pesquisa apresentada comprova - assim como vários outros estudos - que o uso de Redes Neurais Convolucionais Profundas pode trazer enormes vantagens e melhorar a performance dos sistemas de reconhecimento. O modelo apresentado supera todos os modelos previamente conhecidos, tanto em relação à taxa de erro quanto em relação à velocidade de reconhecimento. A divisão da rede em duas sub-redes demonstrou-se como sendo uma excelente estratégia, melhorando significativamente a performance do modelo.

A identificação biométrica baseada nos movimentos dos olhos possui uma vasta possibilidade de aplicação. Trata-se de uma técnica bastante apurada e com diversas vantagens sobre as demais, como reconhecimento facial, reconhecimento de íris e biometria das digitais dos dedos. De alguns anos para cá, essa técnica vem evoluindo com excelentes resultados e deve passar a ser mais comumente utilizada, seja em grandes aplicações de sistemas de segurança ou em dispositivos pessoais.

## References

- Bargary, G., B. J. G. P. L.-O. A. H. R. M. J. (2016). Individual differences in human eye movements: An oculomotor signature? In *Vision Research*.
- Lena A. Jager, Silvia Makowski, P. P.-S. L. M. S. and Scheffer, T. (2020). Deep eyedentification: Biometric identification using micro-movements of the eye.
- Poynter, W., B. M. I. J. W.-C. (2013). Individuals exhibit idiosyncratic eye-movement behavior profiles across tasks. In *Vision Research*.
- Yarbus, A. (1967). Eye movements and vision. In *Plenum, New Your, NY*.