

# An Exploratory Study of Biometrics Using Eye Movement Trajectory Images Collected by Natural Image Stimuli

Antonio R. A. Brasil<sup>1</sup>, Jefferson O. Andrade<sup>2</sup>, Luiz A. Pinto<sup>1</sup>, Karin S. Komati<sup>2</sup>

<sup>1</sup>Programa de Pós-graduação em Engenharia de Controle e Automação (ProPECAut)

<sup>2</sup>Programa de Pós-graduação em Computação Aplicada (PPComp)  
Instituto Federal do Espírito Santo (IFES)

**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. The general objective of this work is to evaluate the use of eye movements as biometrics, by using the images of the eye movements of 29 volunteers, with the use of natural images as a stimulus. 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 the HOG and LBP techniques for extracting characteristics, followed by the SVM, MLP, and random forest classifiers. The highest accuracy was 33.3% with the combination of the HOG and MLP techniques, a result comparable with the best of EMVIC 2014 competition, which used face images as a stimulus, of 39.6%.*

**Resumo.** *O estudo de movimentos oculares como forma de biometria começou em 2004 e tem como processo típico a captura dos movimentos usando imagens estímulos, tratamento destes dados como sinais no tempo seguido de classificação. O objetivo geral deste trabalho é avaliar o uso de movimentos oculares como forma de biometria, a partir das imagens dos movimentos oculares de 29 voluntários, com o uso de imagens naturais como estímulo. Este trabalho propõe dois diferenciais, um é o uso da base de dados DOVES de movimentos oculares coletadas usando-se imagens naturais como estímulo, e o outro é o tratamento dos movimentos oculares como imagens. São geradas imagens das trajetórias dos movimentos oculares, que são a entrada das técnicas HOG e LBP para extração de características, sucedido pelos classificadores SVM, MLP e floresta aleatória. A maior acurácia foi de 33,3% com a combinação das técnicas HOG e MLP, valor comparável com o melhor resultado da competição EMVIC 2014, de 39,6%, que utilizou imagens de faces como estímulo.*

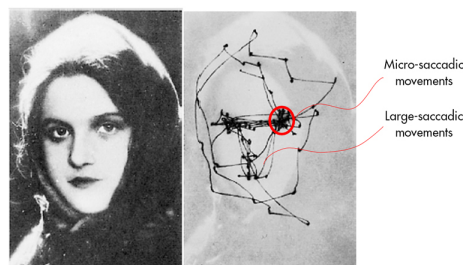
## 1. Introdução

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

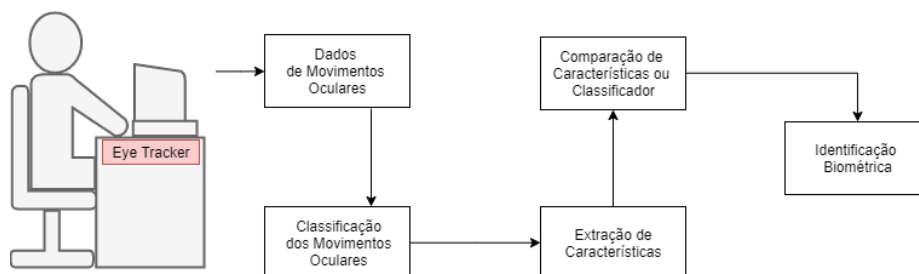
[Kasprowski 2004, Rigas et al. 2012]. As vantagens do uso de movimentos oculares como forma de biometria é que: (i) são extremamente difíceis de serem replicados e (ii) não podem ser roubados de um indivíduo. Esses benefícios são uma consequência da sua característica comportamental, pois os movimentos oculares são involuntários e não podem ser replicados pelos métodos atuais [Komogortsev et al. 2010].

Os principais movimentos oculares são os de fixação e os sacádicos. Os movimentos de fixação ocorrem quando os olhos estabilizam em um ponto por um tempo, e os movimentos sacádicos são movimentos rápidos dos olhos entre pontos de fixação [Holland and Komogortsev 2013]. 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 (micro-sacadas) 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 (Figura 2). 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.



**Figura 2. Sistema geral de coleta e processamento de movimento ocular para biometria. Adaptado de Holland e Komogortsev [Holland and Komogortsev 2011].**

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].

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:

1. Realizar o reconhecimento biométrico de pessoas utilizando as imagens das trajetórias dos seus movimentos oculares. Assim, o segundo passo “Classificação de Movimentos Oculares” do processo da Figura 2, 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 motivação desta abordagem é que este tipo de imagem captura eficientemente a *layout* espacial dos movimentos oculares.
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. 2020], para a qual foram encontrados 212 artigos, sendo 200 na língua inglesa. Dentre os duzentos artigos encontrou-se duas propostas 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 eram 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 2014 (*Eye Movement Verification and Identification Competition*) [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 será usada a base de dados DOVES [Van Der Linde et al. 2009], que usa imagens naturais como estímulo. Para extração das características das imagens dos movimentos oculares dos participantes foram utilizadas duas técnicas clássicas de processamento digital de imagens: 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, do inglês *Support Vector Machine*), floresta aleatória (RF, do inglês *Random Forest*) e redes neurais perceptron de múltiplas camadas (MLP, do inglês *MultiLayer Perceptron*).

Este ano, o XL CSBC (Congresso da Sociedade Brasileira de Computação) traz o tema central "Artificialmente humano ou Humanamente artificial? Desafios para a sociedade 5.0". Nesse contexto, onde os sistemas inteligentes se tornarão aliados para resolver alguns problemas da população, incluindo a falta segurança, o estudo e melhoria de novas formas de biometria se torna necessário; o que é um enfoque deste trabalho. Tornar o dia a dia das pessoas mais seguro, apoiando-se em tecnologias não invasivas, que sustentem o cotidiano das pessoas é importante para tornar uma sociedade – humanamente – artificial.

Este artigo está dividido da seguinte forma: na Seção 2 apresenta-se uma breve descrição de trabalhos relacionados ao tema; na Seção 3, os materiais e os métodos utilizados para o desenvolvimento do trabalho. A Seção 4 discute os experimentos e os resultados, e por fim, a conclusão e os trabalhos futuros são apresentados na Seção 5.

## 2. TRABALHOS RELACIONADOS

Nesta seção, além dos trabalhos citados na introdução, Bednarik e colegas [Bednarik et al. 2005], soluções do EMVIC 2014 [Kasprowski and Harežlak 2014], Rigas e Komogortsev [Rigas and Komogortsev 2014a, Rigas and Komogortsev 2014b] e Li e colegas [Li et al. 2018], também serão descritos alguns dos trabalhos considerados importantes na evolução do tema.

O primeiro artigo que propôs o uso de movimentos oculares como forma de biometria foi escrito por Kasprowski [Kasprowski 2004]. No artigo precursor de Kasprowski, o autor usou a análise cepstral para o processamento da extração de características dos movimentos oculares. A análise cepstral é o resultado da transformada inversa de Fourier do logaritmo do espectro estimado de um sinal, e é muito usado no estudo da fala humana. Os nove voluntários do experimento foram apresentados a estímulos com pontos saltitantes, para coleta dos seus movimentos oculares. Posteriormente, os autores aplicaram as técnicas de aprendizado de máquina KNN (do inglês *K-Nearest Neighbor*, em português K-vizinhos mais próximos), Redes Bayesianas (RB), Árvores de Decisão e SVM. Como resultado, os autores conseguiram obter os melhores resultados com o KNN com 3 vizinhos mais próximos, resultando em uma Taxa de Aceitação Falsa (FAR, do inglês *False Acceptance Rate*) de 1,48% e uma Taxa de Falsa Rejeição (FRR, do inglês *False Rejection Rate*) de 22,59%.

Há um questionamento na área sobre se um mesmo indivíduo apresentará movimentos característicos para seu reconhecimento em dias diferentes. A base de dados do trabalho de Zhang, Laurikkala e Juhola [Zhang et al. 2015] foi um grande diferencial, pois a maioria das bases de dados usam várias coletas de movimentos do mesmo indivíduo em uma única sessão, no mesmo dia. A base de dados foi criada com os movimentos de 109 voluntários estudantes, sendo coletados 120 sacadas de cada estudante, em média. Em seguida, os autores dividiram os estudantes em 5 diferentes grupos de base de dados: o primeiro, contendo todos os 109 estudantes e todas as sacadas coletadas em um único dia. No segundo, as sacadas foram coletadas por sessão, resultando em 1.200 sacadas por cada estudante; na terceira, os autores dividiram em dois sub-grupos, o primeiro sub-

grupo (i) contendo 14 estudantes com suas sacadas coletadas durante 4 dias, resultando em 960 sacadas, e o segundo (ii) sub-grupo contendo 10 estudantes com as sacadas coletadas em 5 dias (dependendo do participante, foram necessárias entre 4 e 10 e 4 sessões de coleta) resultando em 1.200 sacadas. Na quarta, os autores coletaram 52 sacadas de 10 estudantes em 3 dias de coleta, variando entre 3 e 5 sessões. Os melhores resultados obtidos pelos autores foram na segunda base de dados, com uma taxa FRR de 1,3%, FAR de 3,5%. A última base produziu os piores resultados, com FRR de 45,6% e FAR de 56,7%.

O trabalho inicial de Kasproski [Kasproski 2004] apresentou métricas de FRR de 22,5, assim como no trabalho de Zhang, Laurikkala e Juhola [Zhang et al. 2015]. Após uma década de estudos desde o primeiro artigo do tema, a competição BioEye 2015<sup>1</sup> [Komogortsev and Rigas 2015] contou com 64 inscrições e um total de 200 submissões durante os 26 dias de competição. A métrica final foi uma composição dos resultados de quatro diferentes bases de dados, a primeira com 306 voluntários tendo como estímulo pontos saltitantes, a segunda com 306 voluntários sob estímulo de leitura de textos, a terceira com 74 voluntários do conjunto de 306 voluntários iniciais com estímulo de pontos saltitantes, mas coletados após um intervalo de 8 a 13 meses, e da mesma forma que a terceira base, a quarta base tem 74 voluntários do conjunto de 306 voluntários iniciais com estímulo de leitura, mas coletados após um intervalo de 8 a 13 meses. O primeiro lugar, obtido por Anjith George<sup>2</sup> e Prof. Aurobinda Routray<sup>3</sup> do *Indian Institute*, teve como abordagem a extração de uma gama de informações estatísticas dos movimentos de fixação e das sacadas. Foi efetuado um pré-processamento para remover dados redundantes e ruídos, por meio de técnica de seleção de características e ao final foi usada uma Rede Neural. Alcançou acurácia de 95,78%.

Em 2010, Komogortsev e colegas [Komogortsev et al. 2010] desenvolveram uma nova abordagem para identificação de pessoas usando as propriedades musculares do olho humano. Os autores coletaram os movimentos de 41 participantes, usando pontos saltitantes como estímulos, e extraíram o modelo matemático da planta oculomotor (OPMM, do inglês *Oculomotor Plant Mathematical Model*). Os melhores resultados foram de FAR de 5,4% e FRR de 56,6% para o KNN, e FAR de 80% e FRR de 0% para o classificador C4.5. Este trabalho foi importante dada a sua inovadora proposta de usar um modelo com os parâmetros musculares do globo ocular, esses parâmetros modelam propriedades do olho humano, incluindo sinal de controle neuronal, elasticidade, velocidade, força e tensão. Os altos valores de FAR e FRR não indicam o uso para um sistema comercial.

A seguir são descritos os trabalhos similares citados na introdução, por ordem cronológica. São estes os trabalhos, cujos resultados serão comparados com os deste trabalho. Em 2005, Bednarik e colegas [Bednarik et al. 2005] utilizaram de características estáticas (tais como, diâmetro da pupila e variação do diâmetro da pupila) e características dinâmicas (tais como, distância entre os movimentos oculares e a velocidade da sacada) para identificação de pessoas. Nos experimentos, os autores coletaram as informações dos movimentos oculares de 12 participantes. Posteriormente, as características extraídas foram treinadas nos classificadores e validados usando o algoritmo KNN. Como resul-

---

<sup>1</sup><https://bioeye.cs.txstate.edu/>

<sup>2</sup><https://scholar.google.com.br/citations?user=4CCrDp0AAAAJ>

<sup>3</sup><https://scholar.google.com.br/citations?user=VYAwEGUAAAAJ>

tado, foi possível obter uma acurácia com as características estáticas de 40% a 50%, e uma acurácia de 90% com as características dinâmicas.

Rigas e Komogortsev [Rigas and Komogortsev 2014b] propuseram um método para extrair características biométricas de padrões espaciais obtidos por meio dos movimentos oculares ao assistir vídeos e rostos como estímulos aos 200 participantes. O método transforma os dados dos movimentos oculares em imagens, denominadas mapa de densidade de fixação (*Fixation Density Map* – FDM). Nestas imagens, representam-se as trajetórias dos movimentos espaciais e os tempos das fixações oculares são transformadas em formas como círculos, quanto maior o tempo, maior é o espaço que essa forma ocupa. Os melhores resultados obtidos foram com a Divergência de Kullback-Leibler (KL), apresentando taxa de identificação *Rank-1* de 51% (acurácia considerando apenas a primeira resposta) e uma melhor EER de 10,8% com 200 indivíduos.

A competição EMVIC 2014<sup>4</sup> [Kasprowski and Harężlak 2014] utilizou imagens de rostos como estímulos, em todas as imagens os olhos estavam posicionados na mesma localização. Foram disponibilizadas cerca de 837 amostras de 34 voluntários para a base de treino e 593 amostras de 22 voluntários para a base de teste. Com cerca de 82 participantes registrados, e 19 com submissões realizadas, a melhor acurácia obtida foi de 39,63% (participante Vinnie Monaco, da Pace University). Supõe-se que um dos fatores para o resultado inferior quando comparado com a primeira edição da competição, foi o uso das imagens de estímulo serem fotos de rostos de pessoas reais. Infelizmente, logo após o término da competição, foi retirado o acesso à base de dados, e não foi possível o uso da mesma para comparações.

Trabalho recente realizado por Li e colegas [Li et al. 2018] utilizaram como estímulos, perguntas objetivas com 5 respostas apresentadas para os participantes, em mandarim. A metodologia foi aplicada em um total de 58 participantes (24 do sexo masculino e 34 do sexo feminino). A trajetória dos movimentos oculares foram plotadas em um gráfico e posteriormente suas informações foram extraídas usando a transformada de Gabor Wavelet de multicanal (GWT). Aplicando essas informações no classificador KNN e no SVM, os autores conseguiram obter uma melhor EER variando entre 11,49% e 0,89%, com uma melhor taxa de identificação (*Rank-1*) de 96,44%.

### 3. MATERIAIS E MÉTODOS

Nesta seção serão apresentados: a base de dados DOVES (*Database Of Visual Eye movementS*) [Van Der Linde et al. 2009], as técnicas de extração de características de imagens, HOG e LBP, e as técnicas de classificação, RF, SVM e MLP.

#### 3.1. Base de Dados

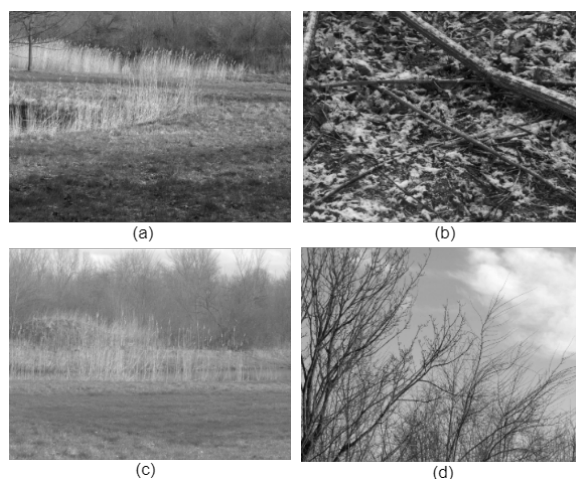
A base de dados utilizada foi a DOVES [Van Der Linde et al. 2009] da Universidade do Texas<sup>5</sup>. Consiste de dados coletados de 29 participantes voluntários, sendo 18 do sexo masculino e 11 do sexo feminino. Todos os indivíduos eram saudáveis e possuíam uma visão normal. A idade média dos participantes é de 27 anos. Foram utilizadas 101 imagens como estímulo de coleta dos movimentos oculares, gerando um total de 2.929 imagens de rastreamento (=101x29).

---

<sup>4</sup><http://www.emvic.org/>

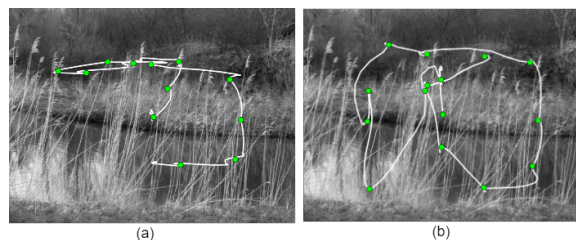
<sup>5</sup><http://live.ece.utexas.edu/research/doves/>

É importante salientar que esta base de dados não foi desenvolvida para o uso biométrico e até o presente momento nenhum trabalho foi encontrado que a utilizasse como dados de entrada para uso biométrico. Na Figura 3 apresentam-se quatro exemplos de imagens que são utilizadas como estímulos para os participantes. Todas as imagens presentes na base de dados da DOVES são de ambientes naturais, tais como árvores, lagos e folhas, todas de dimensões de 1.024 x 768 em escala de cinza.



**Figura 3. Exemplos de imagens naturais usadas como estímulos aos participantes do experimento.**

Os estímulos foram apresentados em um monitor de 21 polegadas, posicionado a uma distância de 134 cm do voluntário. Antes da apresentação de cada imagem-estímulo, uma imagem contendo ruído gaussiano era exibida para eliminar (ou diminuir) a possibilidade de repetição dos movimentos oculares feitos na imagem-estímulo anterior. Além disso, a iluminação na sala de experimentos foi mantida fraca e constante para todos os voluntários. O equipamento de coleta foi um *Generation V dual-Purkinje Eye Tracking*, que tem uma acurácia menor que  $10'$ , precisão de cerca de  $1'$  e tempo de resposta menor que 1 ms. Para cada participante, os autores disponibilizaram os seus movimentos oculares em um arquivo com suas respectivas coordenadas ( $X_p$ ,  $Y_p$ ) e o ângulo da pupila ( $Z_p$ ) com relação à uma reta como se estivesse saindo do nariz da pessoa. Cada imagem foi apresentada para o participante por cinco segundos capturados em 200 Hz. Os autores disponibilizaram as posições dos movimentos de fixação ( $F_x$ ,  $F_y$ ) e o tempo de duração daquela fixação ( $F_t$ ).



**Figura 4. Movimentos oculares de dois participantes sobrepostos à imagem natural usada como estímulo.**

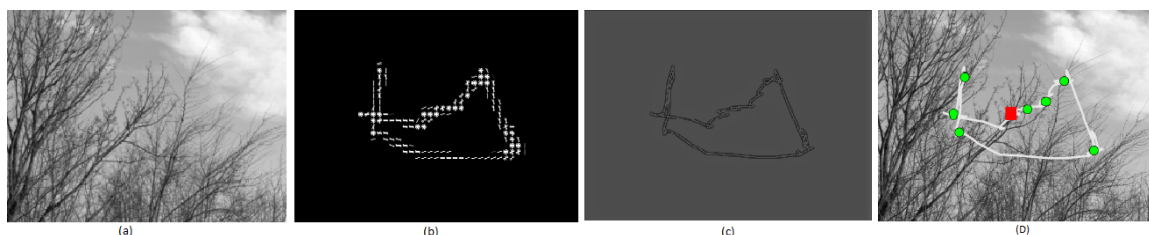
Na Figura 4 apresentam-se os movimentos oculares de dois participantes diferentes, sobrepostos à imagem natural de estímulo. É possível verificar que duas pessoas

diferentes não realizam o mesmo movimento ocular para a mesma imagem, o que evidencia a possibilidade de diferenciação e, aplicar técnicas de aprendizado de máquina para separar as informações de cada participante. Os círculos em verde representam as fixações e as linhas brancas são os movimentos oculares de cada participante.

### 3.2. Extratores de características

Foram realizados dois passos para extrair as características. No primeiro passo, para cada voluntário e para cada imagem, foi criada uma imagem com os movimentos oculares, tendo como resultado uma imagem apenas com a trajetória dos movimentos oculares. No segundo passo, foram utilizadas duas técnicas de extração de características, o Histograma de Gradientes Orientados (HOG) e o padrão binário local (LBP).

O HOG representa a contagem de ocorrências de orientação de gradiente na imagem e foi proposto por Dalal e Triggs [Dalal and Triggs 2005]. A Figura 5 apresenta em (a) a imagem-estímulo e em (d) os movimentos oculares sobrepostos à imagem-estímulo, sendo o ponto inicial que o participante começou a olhar em vermelho. Na imagem resultante do operador HOG (antes da transformação em vetor de características) (Figura 5b) é possível notar os pontos de fixação da imagem de entrada (d). Isto se deve às microsacadas durante a duração da fixação e que são captadas pelo operador HOG, que é um detector de texturas.



**Figura 5. Em (a) tem-se a imagem-estímulo, (b) a imagem resultante do operador, antes da transformação em vetor de características, (c) a imagem resultante do LBP, antes da transformação em vetor de características e em (d) movimentos oculares sobrepostos à imagem-estímulo.**

O LBP é um extrator de características desenvolvido a partir da pesquisa de Ojala e colegas [Ojala et al. 2002]. A ideia do LBP é baseada na hipótese de que cada imagem é composta de vários micro padrões. Sendo assim, o LBP utiliza diversos histogramas locais oriundos desses micro padrões a partir da vizinhança de pixels. Na Figura 5) é possível visualizar a imagem resultante da operação de LBP, similar a Figura 5b). A imagem sofreu uma mudança de contraste apenas para melhorar a visualização neste texto.

### 3.3. Algoritmos de Aprendizado de Máquina

Os seguintes algoritmos foram utilizados para implementação dos classificadores [Gama et al. 2011]: Máquina de Vetor de Suporte (SVM), Floresta Aleatória (RF) e Redes Neurais Perceptron de Múltiplas Camadas (MLP).

A Máquina de Vetor de Suporte é um algoritmo de aprendizado supervisionado que separa todo o espaço amostral em um hiperplano, dividindo-o em duas classes, com o objetivo de minimizar o erro na atribuição das classes. As principais vantagens da técnica



é que consegue lidar bem com grandes conjuntos de exemplos e trata bem dados de alta dimensão.

A RF é um algoritmo que utiliza uma combinação de árvores de decisão em paralelo, constituindo a estrutura de uma floresta. Cada árvore da floresta recebe um conjunto diferente de treinamento, diminuindo o problema de super-ajuste. O seu algoritmo é baseado em votação das várias árvores para escolher a classe de saída para uma determinada amostra desconhecida.

A Rede Neural Perceptron de Múltiplas Camadas (MLP) é uma rede neural que possui uma camada de entrada, uma ou mais camadas intermediárias de neurônios e uma camada de saída. Como vantagem tem boa capacidade de generalização, mesmo diante de dados com ruídos.

#### 4. EXPERIMENTOS E RESULTADOS

Todos os experimentos foram desenvolvidos e avaliados utilizando Python e a biblioteca Scikit-Learn. Nos experimentos deste trabalho, considerou-se a utilização de uma célula 16 x 16 para a técnica HOG. A partir das imagens das trajetórias oculares, foi obtido um vetor de características de tamanho 130.626 para o LBP e para o HOG foi gerado um vetor de tamanho 76.320.

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%. De forma geral, no comparativo das técnicas de extração de característica propostas, praticamente não houve diferença entre elas. Com relação aos classificadores, o melhor classificador foi o MLP e o pior foi a RF. É possível que a boa tolerância a ruídos do MLP tenha sido a causa do melhor resultado. Essa conjectura poderá ser investigada em trabalhos futuros incluindo-se um passo de pré-processamento para retirada de ruídos ou através da diminuição do tamanho do vetor de características. Não houve diferença pelo diferente *kernel* do SVM, em que é possível se conjecturar que a separação entre classes não é bem definida, seja observando os dados de forma linear ou radial.

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 Hareźlak 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.).

**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	<b>33,3</b>	<b>59,3</b>	<b>71,3</b>	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 (%)	<i>Rank-1</i>	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

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 escalada 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%.

## 5. CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho propôs uma abordagem para identificação biométrica usando imagens de trajetórias dos movimentos oculares obtidas com imagens naturais como estímulo. Uma possível aplicação seria a verificação do correntista no uso de caixas eletrônicas. Nesse caso, seria mostrada uma imagem natural, do cotidiano do usuário, para que seja autenticado no sistema. Além de ser autenticado no sistema, uma das vantagens seria a autenticação contínua do usuário, durante todo o tempo de uso do sistema.

Os resultados iniciais obtidos foram satisfatórios, e espera-se que seja um ponto de partida para futuros estudos utilizando a imagem gerada dos movimentos oculares dos voluntários. Nota-se, no entanto, a necessidade da criação e/ou utilização de uma nova base de dados com imagens naturais e movimentos coletados dos participantes. A base de dados utilizada nesse trabalho foi uma base criada em 2007 para outros fins, sem ser de caráter biométrico. Acredita-se, no entanto, que o aumento do tempo de visualização de determinada imagem natural (de 5s para 15s, por exemplo) ou a utilização de imagens estímulo coloridas possam aumentar a acurácia e diminuir a taxa de erro dos algoritmos estudados no presente artigo. Além disso, novas técnicas podem ser aplicadas, como é o caso da *Deep Learning*. A técnica de *Deep Learning* é uma técnica já madura na literatura e aplicada a diversos problemas relacionados a imagens. Devido a capacidade de abstração e de extrair informações das imagens, ela pode conseguir extrair as informações dos movimentos oculares e aumentar ainda mais a acurácia obtida nos experimentos. Outra possível melhoria dos resultados seria a calibração dos classificadores ou mesmo fusão entre eles. O uso de FDM a partir dos movimentos oculares dos participantes da DOVES também pode ser realizada.

## 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. (2020). Eye movements biometrics: A bibliometric analysis from 2004 to 2019. *International Journal of Computer Applications*, 176(24):1–9.
- 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.
- 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.
- Holland, C. D. and Komogortsev, O. V. (2013). Complex eye movement pattern biometrics: Analyzing fixations and saccades. In *Biometrics (ICB), 2013 International Conference on*, 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 Hareźlak, 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.
- Komogortsev, O. V. and Rigas, I. (2015). BioEye 2015: Competition on biometrics via eye movements. In *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*, pages 1–8. IEEE.
- 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.
- 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.
- Zhang, Y., Laurikkala, J., and Juhola, M. (2015). Biometric verification with eye movements: results from a long-term recording series. *IET Biometrics*, 4(3):162–168.