

Algorithm for decoding visual gestures for an assistive virtual keyboard

Rafael Augusto Da Silva, *Member, IEEE*, and Antônio Cláudio Paschoarelli Veiga, *Member, IEEE*

Abstract—Text production is one of the most frequent activities on a computer, a trivial task that can be limiting for individuals affected by severe neuromotor disorders such as Amyotrophic Lateral Sclerosis (ALS) that can lead to Locked-in syndrome (LIS). These individuals need augmentative and alternative communication tools, since they may have only the eye movements as a form of communication and interaction with the outside world. This work investigates methods of interaction based on eye movement tracking and presents a virtual keyboard that utilizes gaze detection as a text input. It describes the development of the shape detection algorithm for the assistive keyboard, typed word voting from a Brazilian Portuguese lexicon and preliminary results on the decoding algorithm.

Index Terms—Ocular keyboard, Virtual keyboard, Augmentative and alternative communication, Locked-in syndrome, Human computer interaction.

I. INTRODUÇÃO

A maior parte das interações de um indivíduo com o computador é realizada utilizando-se teclado e mouse, dispositivos que requerem utilização das mãos [1]. Isso pode representar obstáculos para indivíduos acometidos por distúrbios neuromotores que em estágios avançados tornam a comunicação extremamente restrita e impossibilitam atividades motoras, como é o caso de pessoas afetadas pela Esclerose Lateral Amiotrófica (ELA) ou por algum motivo acometidos pela Síndrome do Encarceramento (SE).

Novas modalidades de interação com dispositivos de comunicação como entrada por gestos ou objetos tangíveis (canetas em telas sensíveis ao toque por exemplo), comandos de voz e diversos tipos de sensores são cada vez mais comuns. A entrada de informação através do rastreamento ocular é uma modalidade que encontra grande aplicação em sistemas de acessibilidade [2].

Aproveitar a informação implícita no movimento dos olhos pode economizar atividades de interação, de modo que desenvolver técnicas e interfaces que se aproveitam da facilidade e velocidade dos movimentos oculares pode trazer benefícios aos usuários.

Pesquisas recentes relacionadas à Comunicação Aumentativa e Alternativa (CAA) se beneficiam de todas as possibilidades oferecidas pelos dispositivos eletrônicos modernos e procuram diminuir a fadiga visual, aumentar a velocidade de digitação e encontrar formas de entrada de informação mais

eficientes através do movimento dos olhos ou direção do olhar [3]–[6].

Existe atualmente uma grande quantidade e variedade de interfaces assistivas, concebidas para auxiliar indivíduos com restrições de comunicação. Um dos elementos fundamentais de uma interface assistiva é o teclado virtual assistivo, que de modo geral ainda apresenta baixa eficiência de entrada de dados e alta complexidade de utilização quando comparado a métodos tradicionais de digitação [7].

O contínuo desenvolvimento de redes neurais e elementos de Inteligência Artificial (IA) aplicados no processamento da linguagem natural (reconhecimento óptico de caracteres, tradução de texto, reconhecimento de fala, entre outros) tornaram possível o reconhecimento de formas do traçado sobre teclados virtuais e sua decodificação em texto. Destaca-se as produções bibliográficas relacionadas de Kurauchi (2018), Loja (2015), Pedrosa (2015), Alsharif *et al.* (2015), Kristensson e Zhai (2004) [3], [6], [8]–[12].

Busca-se neste trabalho aperfeiçoar o conceito de teclado virtual assistivo controlado pelos olhos de forma a diminuir a quantidade de interações necessárias na operação de entrada de texto e aumentar a velocidade de digitação. São encontradas na literatura diversas implementações de teclados que utilizam como entrada de texto gestos feitos pelos dedos do usuário em um teclado virtual [11], [13], [14]. Há diversas denominações para esses teclados, incluindo 'teclados gestuais', 'teclados *swipe*', 'teclados de escrita através de formas', ou ainda 'teclados por gestos de palavras'.

A principal motivação deste trabalho consiste, portanto, no desenvolvimento de um teclado virtual assistivo que utilize como entrada o traçado do olhar, dispensando a necessidade da digitação letra-a-letra e proporcionando a indivíduos com capacidades motoras reduzidas uma opção mais confortável, eficiente e que ofereça menor carga cognitiva (cansaço mental) durante a operação de entrada de texto. É apresentada uma revisão bibliográfica relacionada, seguida do desenvolvimento e teste do algoritmo de reconhecimento de forma que será futuramente integrado a um sistema de comunicação assistiva através do traçado do olhar.

As publicações mais recentes como [8], [3], [4], [15], [6] demonstram que o tema apresenta interesse à comunidade científica e que diversos aspectos relacionados ao '*dwell-free eye typing*' devem ser desenvolvidos para que se possa obter um teclado ocular por gestos que seja realmente eficaz proporcionando mais conforto e liberdade a indivíduos com mobilidade reduzida.

R. A. Da Silva is with the Department of Electrical Engineering, Federal University of Uberlândia, Uberlândia, MG, 38400-902 BRAZIL e-mail: rafaelaugusto@ufu.br.

A. C. P. Veiga is with Federal University of Uberlândia.

Manuscript received February 15, 2020.

II. REVISÃO BIBLIOGRÁFICA

A. Comunicação Aumentativa e Alternativa

A Comunicação Aumentativa e Alternativa (CAA) representa uma subárea da Tecnologia Assistiva (TA) com foco clínico e de pesquisa no campo das patologias da fala [16]. Como descrito por Loja (2015) a CAA reúne os métodos e as tecnologias projetadas para auxiliar ou substituir a comunicação oral de pessoas com limitação de fala [3].

O campo de estudo e prática da CAA é reconhecido internacionalmente possuindo dois propósitos principais com dois públicos alvo distintos: (a) melhorar a capacidade comunicativa de indivíduos com fala ininteligível ou não-fluente, e (b) fornecer uma forma de comunicação alternativa para pessoas que não podem falar ou não possuem capacidade de desenvolver uma linguagem suficiente para efetuar uma comunicação eficiente [17].

A Tecnologia Assistiva (TA), área que engloba a CAA e outros ramos de pesquisa e desenvolvimento, denomina o conjunto de recursos e serviços que auxiliam na promoção ou ampliação das habilidades funcionais de pessoas com deficiência, promovendo dessa forma a inclusão e possibilidade de uma vida independente [18].

Pode-se dizer então que o objetivo fundamental da TA é proporcionar à pessoa com deficiência maior independência, qualidade de vida e inclusão social, de forma a possibilitar a comunicação, controle do ambiente, mobilidade e habilidades de aprendizado e trabalho [19].

As pesquisas relacionadas à Comunicação Aumentativa e Alternativa buscam desenvolver ferramentas inovadoras e criativas para auxiliar a comunicação de indivíduos com pouca ou nenhuma capacidade de fala. Diversas investigações recentes enfocam o desenvolvimento da linguagem e de habilidades de alfabetização entre crianças com deficiências de desenvolvimento, como a paralisia cerebral [20], [21].

Lancioni et al. (2008) e também Mirenda (2001) exploram a utilização de recursos de CAA, eletrônicos ou não, para comunicação básica de crianças consideradas dentro do espectro do autismo [22]–[24].

B. Direção do Olhar como Fonte de Informação

O estudo e desenvolvimento de interfaces baseadas nos movimentos oculares requer a compreensão sobre a anatomia e funcionamento do olho humano, sendo essenciais também conceitos relacionados aos movimentos oculares.

O movimento dos olhos possui características singulares, servindo por vezes como única forma de comunicação e interação com o mundo exterior por pessoas com deficiências neuromotoras severas como a esclerose lateral amiotrófica (ELA) e síndrome do encarceramento (SE) [3]. A utilização dos movimentos oculares como método de entrada e interação apresenta vantagens e desvantagens.

Utilizar a direção do olhar como fonte de informação não é uma tarefa trivial, uma vez que necessita de equipamento ou software específico para o rastreamento ocular que possua altas taxas de varredura e capacidade de processamento em tempo real da entrada de dados. A imprecisão dos rastreadores oculares também é um desafio encontrado por iniciativas de

pesquisa e desenvolvimento na área de CAA através dos olhos, além dos efeitos de escorregamento (*drifting*) e descalibração que também trazem complexidade extra na decodificação e utilização da informação implícita no olhar do usuário. Desvantagens da utilização do rastreamento ocular incluem ainda a menor capacidade de controle (movimentos involuntários e inconscientes), conflito entre controle e visão livre e fadiga visual [2].

Uma interface controlada pela direção do olhar pode oferecer diversos benefícios e possui algumas das seguintes vantagens: facilidade de uso, maior velocidade de interação, dispensa de manutenção, interface higiênica, controle remoto e coleta de informações sobre a atividade do usuário [25].

Um rastreador ocular fornece dados brutos na forma de coordenadas x e y correspondentes à posição do olhar de um indivíduo na tela do computador. Os pontos observados - da expressão em inglês Point of regard (POR) - podem ser utilizados em tempo real ou registrados para posterior análise ou reprodução, permitindo a identificação dos pontos de interesse do usuário. Os pontos observados são comumente utilizados na caracterização de movimentos mais complexos como fixações e sacadas [26].

O desenvolvimento de dispositivos de medição dos movimentos oculares remonta ao início do século 20 [27], mas nas últimas duas décadas a tecnologia dos rastreadores oculares avançou consideravelmente, tornando-os mais precisos, eficientes e diminuindo seu custo. Antes destinados a experimentos de laboratório e aplicações de ponta, pode-se encontrar atualmente rastreadores voltados a consumidores finais em aplicações reais de interfaces humano-computador (IHC) como Comunicação Aumentativa e Alternativa (CAA) ou jogos eletrônicos.

Os movimentos dos olhos são provavelmente os movimentos mais frequentes, esteja o indivíduo acordado ou dormindo [28]. Estima-se que o ser humano faça cerca de 170.000 movimentos sacádicos por dia [29]. Os movimentos oculares representam uma infindável fonte de informações úteis para investigações científicas e clínicas.

Nos últimos 60 anos as pesquisas relacionadas à mobilidade ocular aumentaram significativamente graças a avanços na tecnologia de rastreamento ocular e na teoria psicológica relacionando o rastreamento ocular e processos cognitivos [30]. Os movimentos oculares têm sido usados desde então para se obter melhor compreensão de patologias envolvendo desde a distrofia muscular ao autismo [23], [24], passando por distúrbios neuromotores [17], [31], distúrbios de atenção [32] e marketing digital [33].

Os seres humanos, assim como a maior parte dos primatas, movimentam os olhos segundo uma combinação de cinco movimentos básicos: sacádicos, de fixação, de perseguição, vergenciais e de reflexo vestibulo-ocular [27].

Uma sacada (*saccade* ou movimento sacádico) é um movimento balístico com duração típica de 30 a 120 ms [29]. É um movimento rápido - podendo alcançar velocidades de até 500 graus por segundo - efetuado para reposicionar a fóvea no centro das áreas de interesse do indivíduo [34].

A fixação corresponde ao momento em que os olhos permanecem parados, apesar do fato de que os olhos nunca

permanecem completamente estáticos. As fixações são caracterizadas, então, pelos períodos onde os movimentos oculares são mínimos, e são denominados movimentos fixionais: os tremores, escorregamentos e microssacadas [35]. Os movimentos fixionais ocorrem entre as sacadas, e são necessários para se estabilizar a imagem de um objeto estacionário na retina.

Acredita-se que mais de 80% do tempo do observador seja dedicado à realização de fixações, e que a maior parte da percepção visual seja efetuada nesses instantes [36]. Os movimentos fixionais possuem pelo menos 150 ms, mas a maioria dura de 230 a 300 ms aproximadamente [37].

C. Digitação Ocular Sem Permanência

Diversas aplicações já possuem suporte para entrada de texto utilizando a direção do olhar [11], [15], [38], [39]. Entretanto, a taxa de entrada de texto em teclados visuais ainda é baixa se comparada a métodos tradicionais de escrita [40].

A escrita através da direção do olhar é comumente realizada olhando-se para os caracteres desejados em uma interface de usuário. Tal processo requer um teclado virtual presente na interface e um dispositivo de rastreamento ocular como ilustrado na Figura 1. A seleção dos caracteres é efetuada fixando o olhar em determinada tecla durante um pequeno intervalo de tempo, chamado de *'dwell-time'* ou simplesmente permanência [28].



Fig. 1. Exemplo de teclado virtual e dispositivo de rastreamento ocular.

Uma alternativa explorada por alguns pesquisadores [5], [41], [42] baseia-se na eliminação do período de fixação ou da piscada como método de seleção de caracteres individuais, em um método chamado de *'dwell-free eye typing'*, ou digitação ocular sem permanência, em tradução livre.

Uma das primeiras pesquisas a explorar gestos oculares foi apresentada por Isokoski (2004) [14]. Ele desenvolveu e efetuou testes em uma interface de entrada de texto utilizando objetos da interface fora da tela. O trabalho de Qvarfordt e Zhai (2005) descreve o desenvolvimento de um sistema de diálogo, para interação com o usuário através de padrões de olhar [43].

Wobbrock et al. (2007) publicaram em 2007 um trabalho no qual apresentavam a interface apelidada EdgeWrite para entrada de texto através de gestos visuais e testes que indicavam que a entrada de texto através de gestos seria mais lenta em relação à entrada de texto utilizando piscadas ou intervalos de espera em cada caractere. Usuários treinados na interface desenvolvida atingiram a velocidade de digitação de 7,99 WPM (palavras por minuto) [44].

No ano de 2008 foram publicados diferentes trabalhos relacionados a gestos visuais, tanto demonstrando sua aplicação com diferentes métodos de rastreamento ocular [45], [46], apresentando algoritmos de reconhecimento dos gestos efetuados pelo usuário [47], [48], e aplicando-os na entrada de texto [25], [46], [49].

Kristensson e Vertanen (2012) demonstraram que o método de digitação pelo olhar sem utilização de fixações poderia ser muito mais rápido que os métodos tradicionais, obtendo em seus experimentos uma taxa de 46 WPM que representaria o limite teórico de eficiência na digitação ocular [7]. Hoppe, Löchtefeld e Daiber (2013) sugerem em seu trabalho que uma técnica de digitação sem fixações poderia aumentar significativamente a velocidade de digitação em um sistema [50].

Um importante trabalho, desenvolvido no Brasil e conduzido por Pedrosa, Pimentel e Truong, desenvolveu uma interface denominada *Filteryedping*, também com o objetivo de permitir a digitação de palavras através de gestos oculares [6], [41]. Os resultados foram promissores e demonstraram a viabilidade da técnica de decodificação do traçado realizado pelo olhar.

Pode-se observar pelos trabalhos relacionados citados que a temática acerca da digitação ocular sem permanência vem sendo abordada em pesquisas nos últimos 15 anos. Entretanto, as publicações sobre o tema não são numerosas, nenhuma delas considera o modelo linguístico do português e não há até o momento um sistema final para entrada de texto a partir da decodificação de gestos feitos pelo olhar, como proposto neste trabalho.

III. TECLADO VIRTUAL PROPOSTO

Este trabalho propõe o desenvolvimento de um teclado virtual para entrada de palavras através de formas realizadas pelo olhar do usuário. O teclado assistivo emprega um algoritmo de decodificação do traçado do olhar combinando o reconhecimento de forma com um modelo linguístico do português brasileiro, de forma a mapear gestos oculares em palavras de um léxico pré-definido do idioma.

O sistema pretendido não requer a fixação do olhar do usuário uma letra por vez, operando de forma a registrar o caminho percorrido pelos olhos na interface de um teclado virtual ao digitar uma palavra e em seguida retornar uma lista de prováveis palavras que o usuário tenha digitado. Para digitar a palavra *'cor'* na interface pretendida, por exemplo, o usuário deve percorrer o olhar pelas teclas *'c'*, *'o'*, *'r'* como exemplificado na Figura 2. O gesto contínuo efetuado pelo olhar corresponde a uma representação espacial da palavra *'cor'*, a ser reconhecida pelo decodificador do teclado.

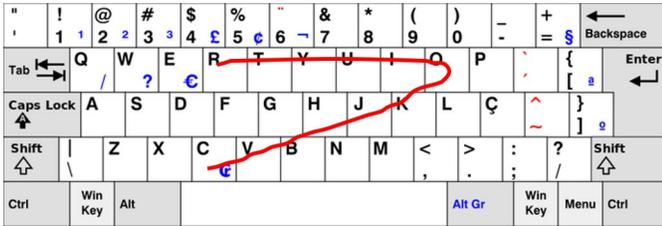


Fig. 2. Exemplo de digitação da palavra 'cor' utilizando um gesto de palavra.

O sistema desenvolvido deve idealmente atender a alguns requisitos básicos, listados a seguir:

- Deve ser capaz de reconhecer todas as palavras do léxico através de traçados oculares unívocos;
- deve operar em 'tempo real', efetuando o reconhecimento de um gesto em uma fração do tempo gasto para se digitar uma palavra: tempo médio de reconhecimento menor que 500ms;
- deve possuir um léxico com uma quantidade de palavras suficiente para a elaboração de textos genéricos em português do Brasil: léxico mínimo de 2500 palavras;
- não deve ocupar mais de 10 megabytes em disco¹;
- deve apresentar relativa robustez com relação a entradas consideradas ruidosas como erros ortográficos e letras repetidas: taxa de erros menor que 5%.

A similaridade de forma por si só não é suficiente para efetuar a desambiguidade entre gestos de palavras que são similares ou muito parecidos. Ao se aumentar o tamanho do léxico a eficiência do reconhecimento através da forma diminui drasticamente, como demonstrado por [12]. É necessário, portanto, um modelo linguístico que leve em consideração os hábitos de comunicação do usuário. A Figura 3 representa o diagrama do sistema de CAA proposto.

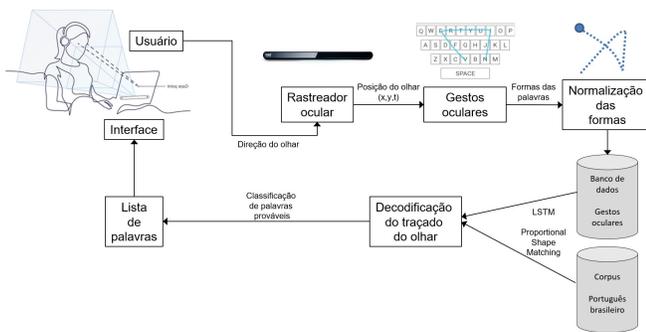


Fig. 3. Arquitetura do teclado ocular assistivo proposto.

O cálculo das palavras prováveis possui duas componentes dependentes (a) da similaridade de forma entre o traçado efetuado pelo usuário e formas ideais pré-calculadas e armazenadas para cada palavra do léxico e (b) da probabilidade de utilização de cada palavra do léxico pelo usuário (calculada a partir do modelo linguístico).

¹No reconhecimento de fala, modelos linguísticos geralmente não devem exceder o espaço em memória de 5 a 10 megabytes, suficiente para permitir a modelagem de até duzentas mil palavras [51].

A modelagem matemática para a decodificação de forma está descrita na equação 1 e pode ser derivada do teorema de Bayes, como demonstrado por [38].

$$W' = \underset{W}{\operatorname{argmax}} \frac{P(G|W).P(W)}{P(G)} \quad (1)$$

W' representa a probabilidade de correspondência entre traçado efetuado e as palavras do léxico, $P(G|W)$ corresponde à probabilidade condicional entre o gesto efetuado pelo usuário (G) e as palavras do léxico (W), $P(W)$ corresponde à probabilidade da palavra W constante no léxico ser utilizada pelo usuário - de acordo com o modelo linguístico - e $P(G)$ corresponde à probabilidade do usuário efetuar determinado gesto de palavra.

A probabilidade $P(G)$ é invariante nesta situação, uma vez que o usuário poderia efetuar qualquer gesto de forma aleatória. Desta forma pode-se resumir o cálculo de probabilidade de um gesto corresponder a determinada palavra à fórmula 2, onde o primeiro fator é calculado através da comparação geométrica entre os traçados e o segundo fator é calculado através do modelo linguístico [52].

$$W' = P(G|W).P(W) \quad (2)$$

A seguir são apresentados detalhes do desenvolvimento dos componentes do teclado virtual que correspondem à interface, modelo linguístico e algoritmo de comparação de forma.

A. Interface

A fase inicial de desenvolvimento e testes apresentados neste trabalho foram baseadas em um teclado virtual com disposição de teclas no formato QWERTY. Esse *layout*, inventado na década de 1860 para minimizar a colisão mecânica em teclados continua sendo o mais utilizado em teclados, sejam mecânicos ou virtuais [53].

A interface apresenta *feedback* visual ao usuário, consistindo no desenho do traçado efetuado pelo cursor correspondente à posição do olhar - ou ponteiro do mouse.

Ao identificar um traçado de palavra completo, a interface exibe ao usuário até cinco palavras com maior probabilidade de entrada, considerando-se a similaridade de forma e também a probabilidade de utilização da palavra no modelo linguístico.

Versões posteriores do teclado assistivo poderão incluir outras disposições de teclado otimizadas, como o *layout* Dvorak, Fitaly ou Opti por exemplo [54]. Pode-se incorporar ainda maneiras de se digitar números ou caracteres de acentuação.

B. Modelo Linguístico do Português Brasileiro

O desenvolvimento de um teclado virtual que efetue predição de palavras, digitação através de gestos de teclado ou que implemente otimizações baseadas no idioma requer a integração do modelo linguístico do idioma desejado. Um modelo linguístico básico pode ser construído estatisticamente a partir do corpus linguístico, que consiste em uma quantidade representativa de texto em um determinado idioma [55].

A decodificação de gestos em palavras do português brasileiro desenvolvida neste trabalho é baseada no cálculo

da probabilidade de digitação de determinada palavra através de seu gesto, cálculo esse que possui um fator de similaridade geométrica e um fator de probabilidade linguística. O modelo linguístico pode ser representado como uma lista de palavras do léxico e sua probabilidade de ocorrência extraídos de uma quantidade representativa de texto.

Para construção do modelo linguístico utilizado neste trabalho foi utilizado o corpus desenvolvido por Finatto (2012) a partir de textos de notícias publicados no jornal online "Massa!" entre 03 de janeiro e 06 de fevereiro de 2013 [56]. A amostra total foi extraída de 70 textos processados, onde pode-se identificar 6.101 palavras que não se repetem. O modelo linguístico baseia-se em uma versão do corpus fornecido pela autora contendo as 3.000 palavras mais frequentes da amostra [56].

Dado o tamanho da amostra do corpus adotado, assume-se que a amostra é representativa, e a probabilidade de utilização de determinada palavra pode ser calculada através da razão entre a frequência de ocorrência da palavra e o quantidade total de palavras do corpus. O modelo linguístico adotado corresponde a um modelo do tipo unigrama, onde a utilização de determinada palavra será considerada um evento independente, não dependendo de palavras anteriores. A probabilidade de palavra pode, portanto, ser calculada de acordo com a equação 3,

$$P(w = W) = \frac{f(W)}{\sum_{i=1}^K f(w_i)} \quad (3)$$

onde k é o número de palavras diferentes presentes no corpus (tamanho do léxico) e $f(w)$ corresponde à frequência de ocorrência da palavra no corpus.

A tabela contendo as palavras do léxico e suas respectivas probabilidades de utilização são utilizadas pelo algoritmo no processo de cálculo das palavras mais prováveis que o usuário tenha desejado digitar. Pode-se observar na Tabela I uma amostra contendo as 20 palavras mais frequentes do corpus, frequência e probabilidade de ocorrência calculada para utilização no teclado virtual. Salienta-se que para os testes iniciais foram suprimidos os elementos de acentuação, caracteres numéricos e de hifenização, de forma a simplificar as condições de avaliação do algoritmo de detecção.

C. Decodificação do Traçado Efetuado

A decodificação do traçado do olhar é semelhante à decodificação de gestos efetuados por canetas ou pelos dedos em interfaces sensíveis ao toque. Para esse tipo de entrada por gestos, a trajetória efetuada é amostrada em intervalos regulares de tempo, retornando dados que contêm as coordenadas de posição correspondentes e instantes de tempo de amostragem. A função do decodificador consiste, então, em encontrar a palavra no léxico existente com melhor correspondência à sequência de pontos de entrada (correspondente a um gesto ou a um traçado do olhar) [42], [44], [50], [57].

Não há até o presente momento muitos trabalhos relacionados à comunicação assistiva através de gestos oculares, entretanto as abordagens existentes empregam diferentes métodos de reconhecimento. Métodos utilizados englobam

TABLE I
20 PALAVRAS MAIS FREQUENTES DO CORPUS LINGUÍSTICO ADOTADO, ORDENADAS POR FREQUÊNCIA E PROBABILIDADE DE OCORRÊNCIA.

Palavra	Freq.	Probabilidade de ocorrência
de	1283	0.055678514
o	992	0.04304995
a	950	0.041227271
e	627	0.027209999
que	601	0.026081673
do	498	0.021611769
da	428	0.01857397
em	360	0.015622966
no	321	0.013930478
com	293	0.012715358
na	286	0.012411578
para	273	0.011847416
um	225	0.009764354
os	219	0.009503971
não	201	0.008722823
é	197	0.008549234
uma	183	0.007941674
foi	155	0.006726555
por	153	0.00663976
as	145	0.006292583

a comparação entre strings [41], aprendizado de máquina [52] e redes neurais recorrentes do tipo *Long Short Term Memory* (LSTM) [9], [58], *Finite State Transducers* [59], [60] e técnicas espaciais como *Dynamic Time Warping* [5], [57] ou ainda *Elastic Shape Matching* [6], [12].

O algoritmo de reconhecimento de forma empregado no teclado assistivo é baseado no cálculo da distância euclidiana e sua utilização no método *Elastic Shape Matching*, muito utilizado no processamento de imagens e que pode também ser utilizado no reconhecimento de gestos de teclado [6], [11]. É um algoritmo inspirado no *dynamic time warping* e opera buscando a distância entre dois conjuntos de pontos calculando o mínimo de deformação necessária para transformar um conjunto de pontos no outro [61]. É um método invariante em escala, translação e rotação [12]. A Figura 4 ilustra o processo de comparação entre um gesto efetuado pelo usuário para digitar a palavra 'cor' e o *template* correspondente, de forma a mensurar a componente de similaridade de forma.

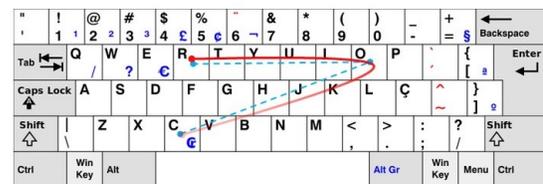


Fig. 4. Exemplo do processo de decodificação de traçado através da comparação com *template*.

O processo de decodificação dos gestos pode ser descrito pelas seguintes etapas:

- cálculo das formas ideais (*templates*) para todas as palavras do léxico, contendo uma quantidade arbitrária

de 100 amostras por traçado para comparação com gestos efetuados;

- captura do traçado correspondente a uma palavra (posição do cursor controlado pelo olhar);
- re-amostragem da série capturada para que tenha mesma dimensão dos *templates* (100 amostras);
- cálculo da similaridade de forma entre *template* e gesto efetuado, retornando um escore;
- correção da probabilidade para o gesto efetuado considerando-se a probabilidade de utilização de palavras do léxico;
- retorno do vetor contendo as palavras mais prováveis (classificadas em ordem decrescente de escore).

Os *templates* - ou gestos ideais de palavra - são gerados sinteticamente produzindo-se um vetor de 100 amostras unindo as coordenadas centrais das teclas correspondentes às letras de cada uma das palavras do léxico. O gesto de palavra efetuado pelo usuário é capturado e re-amostrado para que possua 100 amostras, de forma a possibilitar o cálculo da similaridade entre gesto e *template*.

Para que o algoritmo possa operar em tempo real, o gesto efetuado só é comparado com os *templates* cujos pontos inicial e final estejam distantes até um limiar máximo ajustável. Dessa forma não é necessária a comparação entre o gesto efetuado e todas as palavras do léxico, o que tornaria o algoritmo desnecessariamente complexo.

IV. TESTES E RESULTADOS

O algoritmo de decodificação de gestos em palavras é o elemento central do sistema de comunicação assistiva proposto. Os testes realizados buscam aferir a capacidade de detecção dos gestos de palavras, aumentando-se progressivamente o tamanho do léxico e comparando-se a taxa de acertos de palavras para cada caso.

O desempenho do reconhecimento de gestos foi avaliado utilizando-se léxicos contendo 5, 10, 50 e 100 palavras mais frequentes do corpus, respectivamente. Os gestos das palavras dos léxicos utilizados foram efetuados utilizando-se o mouse do computador - de forma a simular os traçados efetuados pelo olhar e possibilitar o desenvolvimento e ajuste do algoritmo - e em um segundo momento foram realizados testes de digitação utilizando-se um rastreador ocular, do modelo Tobii 4C [62], de forma a avaliar efetivamente a efetividade da decodificação dos gestos oculares em palavras.

Os testes iniciais foram realizados com 3 indivíduos adultos, estudantes de pós-graduação que não apresentavam problemas de visão ou condição de restrição de comunicação².

Os indivíduos que participaram dos testes não possuíam experiência prévia com a utilização de rastreadores oculares ou conhecimento sobre o programa desenvolvido. Eles foram instruídos a digitar as palavras de uma lista fornecida utilizando gestos do olhar na interface apresentada, após a calibração do rastreador.

²Devido à condição de distanciamento social imposta no período, não foi possível até o momento efetuar o recrutamento de indivíduos externos ao departamento para realização dos testes.

A tabela II apresenta a taxa de acertos aferida digitando-se todas as palavras - em uma única tentativa por palavra - de léxicos selecionados de tamanho máximo de 100 palavras, através de gestos efetuados pelo mouse e também pela posição do olhar (média entre os três indivíduos).

TABLE II
TAXA DE ACERTO DE GESTOS DE PALAVRAS EM RELAÇÃO AO TAMANHO DO LÉXICO.

Tam. léxico (k)	% acertos mouse	% acertos rastreador ocular
5	100%	60%
10	100%	60%
50	98%	52%
100	96%	56%

Constata-se que léxicos de tamanho reduzido apresentam melhor precisão de reconhecimento, devido ao menor tamanho das primeiras palavras do léxico que levam a gestos menos complexos e à menor quantidade de gestos a serem comparados - o que evita a ambiguidade de gestos. Como esperado, ao aumentar a quantidade de *templates* com léxicos maiores, as comparações de gestos passam a apresentar desempenho inferior.

A tabela III exemplifica as matrizes de sugestão de palavras - contendo até cinco palavras classificadas em ordem decrescente de probabilidade - e o tempo de processamento, para cinco palavras do léxico selecionadas de forma aleatória. Testes de digitação de palavras aleatórias indicam que não há uma relação direta entre tamanho de palavra e tempo de processamento, como pode-se pensar a princípio. A palavra 'inconstitucionalidade', por exemplo, apresentou um tempo de processamento médio de 286 milissegundos, menor que o tempo de processamento de palavras menores apresentadas na tabela.

TABLE III
EXEMPLO DE PALAVRAS SUGERIDAS E TEMPO DE PROCESSAMENTO PARA CINCO GESTOS DE PALAVRAS EFETUADOS.

Palavra	Palavras sugeridas	Tempo
de	['de', 'desse', 'desde', 'deste', 'deve']	319 ms
que	['que', 'quase', 'quantidade']	67 ms
com	['com', 'comum']	201 ms
para	['para', 'passada', 'podera', 'praca', 'parada']	1173 ms
segundo	['segundo', 'sendo', 'sido', 'sandro', 'servico']	838 ms

Deve-se salientar o bom desempenho do decodificador obtido simulando-se a digitação de palavras extensas como por exemplo as entradas 'encaminhados', 'administrativamente', 'inconstitucionalidade'. Essas foram corretamente reconhecidas através de seus gestos de palavras. Ao se utilizar o rastreamento ocular, no entanto, a taxa de acertos diminui consideravelmente para palavras extensas.

Os erros de reconhecimento de palavra obtidos são animadores, considerando-se a grande dimensão do léxico adotado. Uma possibilidade de otimização do algoritmo consiste na utilização de um modelo linguístico do tipo bi-grama ou n-grama de maior ordem, que levaria em consideração as palavras anteriormente digitadas durante o cálculo da probabilidade de palavra. Esta técnica poderia diminuir ainda mais as chances de ambiguidade entre gestos de palavras parecidos.

Os testes realizados demonstram que os requisitos definidos para o sistema foram atendidos, com exceção da taxa de erros que ainda não é menor do que 5% na situação real da digitação ocular. Ajustes no programa serão realizados, de forma a ajustar parâmetros e considerar os comentários dos participantes do teste.

O *feedback* dos indivíduos que participaram dos testes indicam que: o teclado deve ser grande, seria melhor se as letras estivessem localizadas no centro das teclas no teclado virtual, digitar sem que o cursor esteja visível evita distrações e deslizamento não intencional para a direita, efetuar gestos mais lentos produz menos erros, a posição ótima da cabeça do usuário é na altura do monitor a uma distância de 50 centímetros da tela.

Constatou-se que a eficiência de digitação dos usuários aumentou após os minutos iniciais de utilização do teclado assistivo. A partir dos resultados obtidos e impressões dos utilizadores é possível aperfeiçoar a primeira versão do programa, de forma a melhorar a usabilidade da interface assistiva para que apresente utilidade real a indivíduos com restrições de comunicação.

V. CONCLUSÃO

A interação entre usuários e computadores através do movimento dos olhos apresenta um enorme potencial ainda não explorado em larga escala, e representa uma solução capaz de transformar a vida de usuários com mobilidade restrita.

Este trabalho propôs um sistema de comunicação baseado em um teclado virtual assistivo controlado a partir de gestos oculares, e explorou a eficácia de um algoritmo de decodificação do traçado ocular em palavras do português do Brasil. Um modelo linguístico foi produzido a partir de um corpus da língua contendo grande quantidade de texto em linguagem natural.

Os testes do algoritmo realizados até então indicam que os traçados de teclado unívocos são detectados com uma eficiência acima de 90% para gestos simulados com o mouse e acima de 50% para gestos do olhar, um resultado bastante promissor que abre caminho para testes mais detalhados e com uma maior quantidade e variedade de indivíduos.

Os próximos passos do desenvolvimento do sistema de comunicação assistiva envolvem o aperfeiçoamento da interface, implementando elementos de *feedback* ao usuário, personalização de elementos de acordo com preferências do usuário, e possibilidade de utilizar diferentes *layouts* de teclado além do QWERTY.

REFERENCES

- [1] K. Hill, "Advances in Augmentative and Alternative Communication as Quality-of-Life Technology," *Physical Medicine and Rehabilitation Clinics of North America*, vol. 21, no. 1, pp. 43–58, 2010.
- [2] H. Drewes, "Eye Gaze Tracking for Human Computer Interaction," Ph.D. dissertation, LMU München, 2010. [Online]. Available: <https://edoc.ub.uni-muenchen.de/11591>
- [3] L. f. B. Loja, "Tecnologia Assistiva: Um teclado virtual evolutivo para aplicação em sistemas de comunicação alternativa e aumentativa," 2015.
- [4] Y. Liu, B. S. Lee, M. J. McKeown, and C. Lee, "A robust recognition approach in eye-based dwell-free typing," *Proceedings of 2015 IEEE International Conference on Progress in Informatics and Computing, PIC 2015*, pp. 5–9, 2016.
- [5] A. Kurauchi, W. Feng, A. Joshi, C. Morimoto, and M. Betke, "EyeSwipe: Dwell-free Text Entry Using Gaze Paths," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1952–1956, 2016. [Online]. Available: <http://doi.acm.org/10.1145/2858036.2858335>
- [6] D. D. C. Pedrosa, "Data input and content exploration in scenarios with restrictions," Ph.D. dissertation, Universidade de São Paulo, São Carlos, 12 2014. [Online]. Available: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-13042015-144651/>
- [7] P. Kristensson and K. Vertanen, "The potential of dwell-free eye-typing for fast assistive gaze communication," *Proceedings of the Symposium on Eye ...*, vol. 1, no. 212, pp. 241–244, 2012. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2168605>
- [8] A. T. N. Kurauchi, "EyeSwipe: text entry using gaze paths," Ph.D. dissertation, University of São Paulo, 2018.
- [9] O. Alsharif, T. Ouyang, F. Beaufays, S. Zhai, T. Breuel, and J. Schalkwyk, "Long short term memory neural network for keyboard gesture decoding," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2015-Augus. IEEE, 4 2015, pp. 2076–2080. [Online]. Available: <http://ieeexplore.ieee.org/document/7178336/>
- [10] S. Zhai and P. O. Kristensson, "Introduction to Shape Writing," *Text Entry Systems*, no. September, pp. 139–158, 2007.
- [11] P. O. Kristensson, "Discrete and Continuous Shape Writing for Text Entry and Control," Ph.D. dissertation, Linköping University, 2007.
- [12] P.-O. Kristensson and S. Zhai, "SHARK2: A Large Vocabulary Shorthand Writing System for Pen-Based Computers," *Proceedings of the 17th annual ACM symposium on User interface software and technology - UIST '04*, vol. 6, no. 2, pp. 43–52, 2004. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1029632.1029640>
- [13] S. Zhai and P.-O. Kristensson, "Shorthand writing on stylus keyboard," *Proceedings of the conference on Human factors in computing systems - CHI '03*, no. 5, p. 97, 2003. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=642611.642630>
- [14] P. Isokoski, "Text input methods for eye trackers using off-screen targets," 2004.
- [15] H. Cecotti, "A Multimodal Gaze-Controlled Virtual Keyboard," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 4, pp. 601–606, 2016.
- [16] C. Zangari, K. Kangas, and L. Lloyd, "Augmentative and alternative communication: A field in transition," *Augmentative and Alternative Communication*, vol. 4, no. 2, pp. 60–65, 1 1988. [Online]. Available: <https://doi.org/10.1080/07434618812331274647>
- [17] D. R. Beukelman, S. Fager, L. Ball, and A. Dietz, "AAC for adults with acquired neurological conditions: A review," *AAC: Augmentative and Alternative Communication*, vol. 23, no. 3, pp. 230–242, 2007.
- [18] R. Bersch, "Introdução à Tecnologia Assistiva," *Porto Alegre: CEDI*, p. 21, 2008.
- [19] T. G. Filho, "a Tecnologia Assistiva : De Que Se Trata ?" *Porto Alegre: Redes Editora*, pp. 207–235, 2009.
- [20] C. Soto, Gloria; Zangari, *Practically speaking: Language, literacy, and academic development for students with AAC needs*. Paul H. Brookes, 2009.
- [21] E. J. Manzini, "Tecnologia assistiva para educação: recursos pedagógicos adaptados," *Ensaio pedagógicos: construindo escolas inclusivas. Brasília: SEESP/MEC*, pp. 82–86, 2005.
- [22] G. E. Lancioni, M. F. O'reilly, N. N. Singh, J. Sigafos, D. Oliva, and L. Severini, "Three persons with multiple disabilities accessing environmental stimuli and asking for social contact through microswitch and VOCA technology," *Journal of Intellectual Disability Research*, 2008.
- [23] P. Miranda, "Autism, Augmentative Communication, and Assistive Technology: What Do We Really Know?" *Focus on Autism and Other Developmental Disabilities*, 2001.
- [24] X. Fan, J. H. Miles, N. Takahashi, and G. Yao, "Abnormal transient pupillary light reflex in individuals with autism spectrum disorders," *Journal of Autism and Developmental Disorders*, 2009.
- [25] N. Bee and E. André, "Writing with your eye: A dwell time free writing system adapted to the nature of human eye gaze;" in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008.
- [26] M. F. Costa, "Movimentos oculares no bebê: o que eles nos indicam sobre o status oftalmológico e neurológico," *Psicologia USP*, 2017.
- [27] A. T. Duchowski, *Eye Tracking Methodology*. Cham: Springer International Publishing, 2017. [Online]. Available: <http://link.springer.com/10.1007/978-3-319-57883-5>

- [28] P. Majaranta, U.-K. Ahola, and O. Špakov, "Fast gaze typing with an adjustable dwell time," *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, p. 357, 2009. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1518701.1518758>
- [29] A. L. Yarbus, *Eye Movements and Vision*. Boston, MA: Springer US, 1967, vol. 6, no. 4. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/0028393268900122http://link.springer.com/10.1007/978-1-4899-5379-7>
- [30] R. J. Jacob and K. S. Karn, "Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises," *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, no. January 2003, pp. 531–553, 2003.
- [31] L. J. Ball, S. Fager, and M. Fried-Oken, "Augmentative and alternative communication for people with progressive neuromuscular disease," *Physical Medicine and Rehabilitation Clinics of North America*, vol. 23, no. 3, pp. 689–699, 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.pmr.2012.06.003>
- [32] M. J. Tsai, H. T. Hou, M. L. Lai, W. Y. Liu, and F. Y. Yang, "Visual attention for solving multiple-choice science problem: An eye-tracking analysis," *Computers and Education*, 2012.
- [33] Q. Wang, S. Yang, M. Liu, Z. Cao, and Q. Ma, "An eye-tracking study of website complexity from cognitive load perspective," *Decision Support Systems*, 2014.
- [34] C. Clifton Jr, A. Staub, and K. Rayner, "Eye movements in reading words and sentences," in *Eye movements*. Elsevier, 2007, pp. 341–371.
- [35] L. M. Chalupa and J. S. Werner, Eds., *The visual neurosciences, Vols. 1 & 2*. Cambridge, MA, US: MIT Press, 2004.
- [36] M. F. Land, "Eye movements and the control of actions in everyday life," vol. 25, pp. 296–324, 2006.
- [37] A. M. Melchiades, "PARÂMETROS OCULARES NO RASTREAMENTO VISUAL DE CENAS COM CONTEÚDO EMOCIONAL," Ph.D. dissertation, Universidade de Brasília, 2014.
- [38] S. Zhai and P. O. Kristensson, "The word-gesture keyboard," *Communications of the ACM*, vol. 55, no. 9, p. 91, 9 2012. [Online]. Available: <http://dl.acm.org/citation.cfm?doi=2330667.2330689>
- [39] D. J. Ward and D. J. C. MacKay, "Fast Hands-free Writing by Gaze Direction," *Nature*, vol. 418, no. August, 8 2002. [Online]. Available: <http://arxiv.org/abs/cs/0204030http://dx.doi.org/10.1038/418838ahttp://www.nature.com/articles/418838ahttp://arxiv.org/abs/cs/0204030%0Ahttp://dx.doi.org/10.1038/418838a>
- [40] P. Majaranta and K.-J. Rähkä, "Twenty Years of Eye Typing: Systems and Design Issues," *Eye Tracking Research & Applications (ETRA) Symposium*, pp. 15–22, 2002.
- [41] D. Pedrosa, M. d. G. Pimentel, and K. N. Truong, "Filteryedping: A Dwell-Free Eye Typing Technique," *Extended Abstracts of the ACM CHI'15 Conference on Human Factors in Computing Systems*, vol. 2, pp. 303–306, 2015. [Online]. Available: <http://dx.doi.org/10.1145/2702613.2725458>
- [42] Y. Liu, C. Zhang, C. Lee, B.-s. Lee, and A. Q. Chen, "GazeTry : Swipe Text Typing Using Gaze," *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, pp. 192–196, 2015. [Online]. Available: <http://doi.acm.org/10.1145/2838739.2838804>
- [43] P. Qvarfordt and S. Zhai, "Conversing with the user based on eye-gaze patterns," 2005.
- [44] J. O. Wobbrock, J. Rubinstein, M. Sawyer, and A. T. Duchowski, "Not Typing but Writing: Eye-based Text Entry Using Letter-like Gestures," *The 3rd Conference on Communication by Gaze Interaction*, no. Figure 1, pp. 4–6, 2007.
- [45] A. Bulling, D. Roggen, and G. Tröster, "It's in your eyes: towards context-awareness and mobile HCI using wearable EOG goggles," ... *of the 10th international conference on ...*, 2008.
- [46] M. Porta and M. Turina, "Eye-S: a full-screen input modality for pure eye-based communication," *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 2008.
- [47] M. Maia, "Processos bottom-up e top-down no rastreamento ocular de imagens," *Revista de Estudos Linguísticos Veredas online*, vol. 12, no. 2, pp. 8–23, 2008.
- [48] I. S. MacKenzie and X. Zhang, "Eye typing using word and letter prediction and a fixation algorithm," *Proceedings of the 2008 symposium on Eye tracking research & applications - ETRA '08*, vol. 1, no. 212, p. 55, 2008. [Online]. Available: <http://portal.acm.org/citation.cfm?doi=1344471.1344484>
- [49] J. O. Wobbrock, J. Rubinstein, M. W. Sawyer, and A. T. Duchowski, "Longitudinal evaluation of discrete consecutive gaze gestures for text entry," 2008.
- [50] S. Hoppe, M. Löchtfeld, and F. Daiber, "Eype - Using eye-traces for eye-typing," in *Workshop on grand challenges in text entry (chi 2013)*, 2013.
- [51] T. Ouyang, D. Rybach, F. Beaufays, and M. Riley, "Mobile Keyboard Input Decoding with Finite-State Transducers," 2017. [Online]. Available: <http://arxiv.org/abs/1704.03987>
- [52] J. Stendahl and J. Arnör, "Gesture Keyboard Using Machine Learning," Dissertation, 2014.
- [53] J. Rick, "Performance optimizations of virtual keyboards for stroke-based text entry on a touch-based tabletop," in *Proceedings of the 23rd annual ACM symposium on User interface software and technology*, 2010, pp. 77–86.
- [54] S. Sarcar, P. Panwar, and T. Chakraborty, "EyeK," in *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction - APCHI '13*. New York, New York, USA: ACM Press, 2013, pp. 215–220. [Online]. Available: <http://dl.acm.org/citation.cfm?doi=2525194.2525288>
- [55] Michaelis, *Moderno dicionário da língua portuguesa*. Melhoramentos, 1999. [Online]. Available: michaelis.uol.com.br/
- [56] M. J. B. Finatto, "Projeto PorPopular, frequência de verbos em português e no jornal popular brasileiro," *As Ciências do Léxico: lexicografia, terminologia. Campo Grande: Ed. da UFMS*, vol. 6, pp. 227–244, 2012.
- [57] R. d. Zoeten, B. Opleiding, K. Intelligentie, and J. Zuidema, "Recognizing input for swipe based keyboards," 2013.
- [58] R. Johnson and T. Zhang, "Supervised and Semi-Supervised Text Categorization using LSTM for Region Embeddings," vol. 48, 2016. [Online]. Available: <http://arxiv.org/abs/1602.02373>
- [59] D. Anderson, C. Bailey, and M. Skubic, "Hidden Markov Model Symbol Recognition for Sketch-Based Interfaces," *AAAI Fall Symposium*, p. 15–21, 2004. [Online]. Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Hidden+Markov+Model+Symbol+Recognition+for+Sketch+Based+Interfaces#0>
- [60] M. Elmezain, A. Al-Hamadi, and B. Michaelis, "Hand trajectory-based gesture spotting and recognition using HMM," in *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 11 2009, pp. 3577–3580. [Online]. Available: <http://ieeexplore.ieee.org/document/5414322/>
- [61] C. C. Tappert, "Cursive Script Recognition by Elastic iy/latcliing," vol. 26, no. 6, pp. 765–771, 1982.
- [62] Tobii, "Tobii Eye Tracker 4C," 2016. [Online]. Available: <https://gaming.tobii.com/product/tobii-eye-tracker-4c>



Rafael Augusto da Silva is a PhD student in Electrical Engineering (emphasis in Digital Signal Processing) at the Federal University of Uberlândia (UFU). He holds master's and bachelor's degrees in Electrical Engineering by the same institution. He participated in a joint program at the Université Lille 1 (France), with research in 'Micro et Nanotechnologies: Systèmes de Communication Radio-fréquence' with CAPES funding. <http://lattes.cnpq.br/7332279341824131>



Antônio Cláudio Paschoarelli Veiga holds a PhD in Electrical Engineering from the State University of Campinas (2002). He is currently a professor at the Federal University of Uberlândia. He has experience in the area of Electrical Engineering, with emphasis on Image Processing and digital communications. He has received several Masters and PhD students in the postgraduate program of the Faculty of Electrical Engineering of UFU. He is currently the coordinator of the Electronic Engineering and Telecommunications course at the Federal University of Uberlândia. <http://lattes.cnpq.br/2835416571685218>