

Interface Humano-Computador utilizando abordagens de Gaze Tracking

Human-Computer Interface using Gaze Tracking approaches

Eduardo K. Sakamoto¹, Claudio Roberto M. Mauricio¹, Fabiana F. F. Peres¹

¹Centro de Engenharias e Ciências Exatas
Universidade Estadual do Oeste do Paraná
Foz do Iguaçu, Paraná, Brasil

eduardoksakamoto@outlook.com, {claudio.mauricio, fabiana.peres}@unioeste.br

Abstract. *This article presents the development of a Human-Computer Interface prototype utilizing gaze tracking techniques with an emphasis on open-source technologies. The solution was designed to enable interaction with the computer through eye direction, making it suitable for accessibility applications and task automation in digital environments. The prototype uses libraries such as OpenCV and MediaPipe for facial detection and eye direction calculation. Preliminary tests demonstrated the feasibility of the solution, although additional improvements are suggested to enhance the system's accuracy and usability.*

Resumo. *Este artigo apresenta o desenvolvimento de um protótipo de Interface Humano-Computador utilizando técnicas de gaze tracking com ênfase em tecnologias open-source. A solução foi projetada para possibilitar a interação com o computador através da direção do olhar, sendo adequada para aplicações de acessibilidade e automatização de tarefas em ambientes digitais. O protótipo utiliza bibliotecas como OpenCV e MediaPipe para detecção facial e cálculo da direção do olhar. Testes preliminares demonstraram a viabilidade da solução, embora melhorias adicionais sejam sugeridas para aumentar a precisão e a usabilidade do sistema.*

1. Introdução

O rastreamento do olhar, ou gaze tracking, é uma tecnologia promissora na área de Interfaces Humano-Computador (IHC) [Drewes 2010]. Ao permitir que a direção do olhar controle interações com dispositivos computacionais, essa tecnologia oferece novas possibilidades, especialmente em contextos envolvendo acessibilidade [Esposito et al. 2021] com uso de elementos virtuais. Este trabalho visa desenvolver um protótipo de IHC utilizando gaze tracking com foco em soluções de código aberto, abordando desafios e propondo melhorias para aumentar a precisão e a usabilidade do sistema [Chandra et al. 2015].

2. Metodologia

A metodologia adotada envolveu a construção de um protótipo utilizando as seguintes etapas principais:

- Fundamentação Teórica: Estudo das tecnologias de IHC, com foco em gaze tracking, abordando desde a estrutura do olho humano até as técnicas e algoritmos relevantes para o desenvolvimento [Karray et al. 2008].

- Caracterização da Interface: Escolha de tecnologias e algoritmos, como a video-oculografia para aquisição de imagens e o modelo de MediaPipe para detecção da região ocular e cálculo da direção do olhar. O piscar foi definido como a métrica de interação do usuário.
- Desenvolvimento e Testes: Implementação do protótipo utilizando as bibliotecas OpenCV, MediaPipe e PyAutoGUI, seguidos de testes em cenários práticos como manipulação de documentos e navegação em vídeos [Morimoto and Mimica 2005].

3. Resultados

Os testes realizados demonstraram que o protótipo é capaz de detectar e interpretar a direção do olhar, permitindo a interação com o computador sem o uso das mãos. Três cenários foram analisados: Manipulação de Documentos, Navegação em Vídeos e Clique Universal.

3.1. Manipulação de Documentos (Word)

A interface permitiu a navegação entre textos e a seleção de opções de formatação através do olhar.

A Figura 1(esquerda) demonstra a interface inicial e a devida interação de clique, utilizando a parcela de fechamento de olho (utilizando cálculo de EAR e contagem de período de ‘fechamento’ para fins de voluntariedade de comando) e a verificação de sobreposição de contorno/parte interna do botão com o extremo da linha referente a direção de olhar.

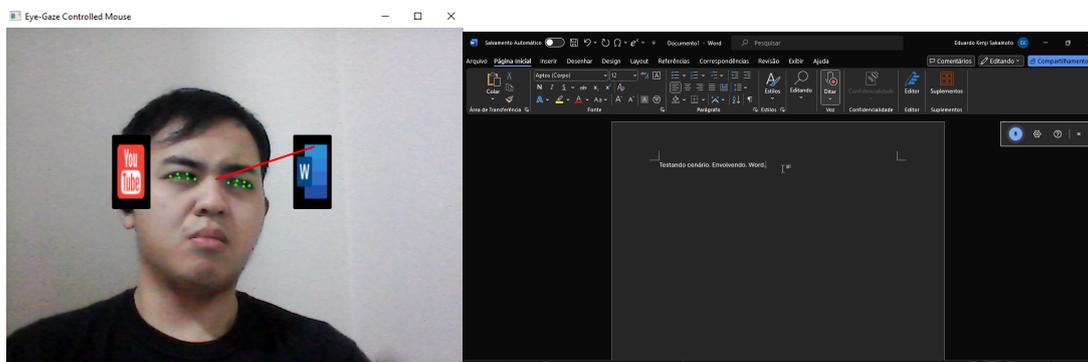


Figura 1. Início de cenário de acesso ao Word (esquerda) e Término de cenário de acesso ao Word (direita).

Nas etapas posteriores, foram utilizadas funções do PyAutoGUI voltadas para locomoção de cursor, clique de cursor, pressionamento de teclas (inserção de texto) e um conjunto de atrasos temporais, que possibilitaram a abertura e renderização de ferramentas, sem que o fluxo se interrompa por sua rapidez de emissão de comando.

Por fim, a Figura 1 (direita) demonstra como ficaria o fim de fluxo de acesso do *Word* e a ferramenta ‘Ditado Office’, o que possibilita o usuário, não somente a prosseguir a navegar por ferramentas dentro do *Word*, mas também utilizar da ferramenta de ditado para redigir documentos/blocos de notas utilizando a voz.

3.2. Navegação em Vídeos (YouTube)

Foi possível controlar a seleção e reprodução de vídeos apenas com o movimento ocular.

A Figura 2 (esquerda) demonstra a inicialização de abertura da plataforma *YouTube*, sendo os procedimentos de reconhecimento de botão, de comando visual e inserção de textos/cliques, similares ao cenário envolvendo a ferramenta *Word*.

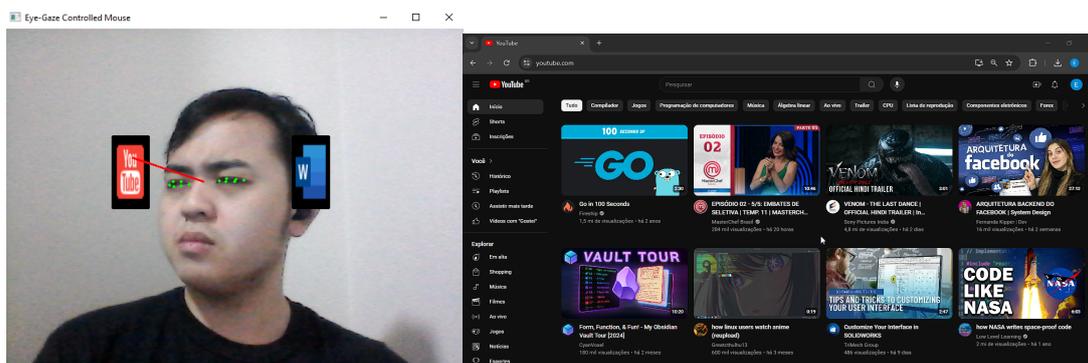


Figura 2. Início de cenário de acesso ao Youtube (esquerda) e Término de cenário de acesso ao Youtube (direita).

Finalizando o procedimento automatizado, a Figura 2 (direita) exemplifica um cenário de acesso da plataforma, liberando o controle de outros comandos ao usuário, possibilitando o acesso de vídeos disponibilizados pelo *Youtube* através de um clique desencadeado pela interface visual.

3.3. Clique Universal

O usuário conseguiu realizar cliques e seleções em interfaces gráficas diversas, destacando a versatilidade da solução.

O clique num botão do *frame* é realizado através da verificação de sobreposição entre coordenada de olhar e uma área reservada para o acesso do botão. Mas caso não haja essa sobreposição e tem-se a intenção de disparo de um comando, o comando a ser disparado é um simples clique do cursor, sendo representado pela função `pyautogui.click()`.

Embora essa interação seja considerada muito simples, a interação com elementos da tela do computador expande as possibilidades de acesso a qualquer ferramenta disponibilizada pela aplicação ativa que seja acessível pelo cursor.

3.4. Considerações

O cenário de testes envolvendo um único usuário levanta questões de variabilidade de comportamento de interface, onde a execução dos cenários com outros usuários e diferentes configurações de tamanho de tela, implicam a especificidade de ambiente computacional e a ausência de *feedback* externo, o que deixa a desejar do enriquecimento maior de solução ao inserir usuários portadores de mentalidades e/ou necessidades distintas.

A solução possui inconvenientes que interferem na experiência do usuário, sendo eles detalhados no próximo capítulo, mas ainda assim apresenta potencial para melhorias, não somente em quesitos de precisão de modelos, mas também em alcance de solução

ao integrar outros tipos de saídas (interfaces híbridas) e atender usuários com âmbitos e áreas de conhecimento distintos.

Diante de questões de teor teórico que utiliza de diversos ramos de conhecimento, implementação da solução computacional e demonstração prática da interface, verifica-se alguns dos desafios inseridos na construção de uma IHC, onde o vasto grau de acessibilidade e de aplicabilidade dos esteriótipos desta interface podem agregar na adição de camadas de abstração e de complexidade, mas também podem resultar em grandes avanços para a tecnologia contextualizado para uma determinada necessidade.

4. Conclusões e Trabalhos Futuros

O protótipo desenvolvido demonstrou o potencial das tecnologias de gaze tracking aplicadas a IHC, podendo ser idealizado o incremento das interações nos âmbitos real e virtual para fins de imersão e expansão do uso aplicado em soluções de problemas do mundo real. No entanto, foram identificadas limitações relacionadas à precisão do rastreamento e à suavidade da interação.

Sobre possíveis melhorias da interface proposta, pode-se sugerir a otimização através do uso de outros algoritmos e técnicas (em especial, para calibração de câmera, estimação de arranjo de cabeça e cálculo de direção de olhar de usuário), onde a diversidade de abordagens existentes para *gaze tracking* podem exigir do uso de equipamentos auxiliares para fins de suporte de câmera, um *hardware* mais robusto e/ou da adição de sensores (infravermelho, por exemplo), prezando pela precisão da interface e segurança do usuário.

Além disso, a presença de oscilações de precisão (trêmula excessiva) dadas pelo comportamento ‘agitado’ da interface proposta traz à tona a demanda de algumas medidas como o controle de percepção temporal dos *frames*, da revisão de mapeamento e (re)cálculo das coordenadas da tela de computador, uso de configurações e métricas alternativas da engenharia dos olhos e outras abordagens, de forma a suavizar a experiência enfrentada pelo movimento brusco realizado pelo cursor num cenário estável de posição de olhar.

Referências

- Chandra, S., Sharma, G., Malhotra, S., Jha, D., and Mittal, A. P. (2015). Eye tracking based human computer interaction: Applications and their uses. In *2015 International conference on man and machine interfacing (MAMI)*, pages 1–5. IEEE.
- Drewes, H. (2010). *Eye gaze tracking for human computer interaction*. PhD thesis, Imu.
- Esposito, D., Centracchio, J., Andreozzi, E., Gargiulo, G. D., Naik, G. R., and Bifulco, P. (2021). Biosignal-based human–machine interfaces for assistance and rehabilitation: A survey. *Sensors*, 21(20):6863.
- Karray, F., Alemzadeh, M., Abou Saleh, J., and Arab, M. N. (2008). Human-computer interaction: Overview on state of the art. *International journal on smart sensing and intelligent systems*, 1(1):137–159.
- Morimoto, C. H. and Mimica, M. R. (2005). Eye gaze tracking techniques for interactive applications. *Computer vision and image understanding*, 98(1):4–24.