Controle de Mouse Virtual Assistivo por Movimentos da Cabeça

Nataly B. Pereira[1], **José Valentim dos S. Filho**[1]

[1]Centro de Ciências Exatas e Tecnológicas
Universidade Federal do Recôncavo da Bahia (UFRB)

nataly.nath@aluno.ufrb.edu.br, valentim@ufrb.edu.br

Abstract. *Conventional input devices impose barriers on individuals with motor disabilities, and the imprecision caused by involuntary micro-tremors remains a critical obstacle in assistive head-tracking control. A predictive virtual mouse is proposed, integrating computer vision (MediaPipe), a Random Forest classifier, and web-context-aware adaptive magnetism. Trained on 1,980 samples with five physical motion features grounded in Fitts' Law, the model predicts click intention with 70.45% accuracy (*$69.51 \pm 3.20\%$ *cross-validated), applying graduated assistance conditioned on facial confirmation. Operating on conventional hardware without a GPU, the solution establishes itself as a low-cost inclusive alternative.*

Resumo. *Dispositivos convencionais de entrada impõem barreiras a pessoas com deficiência motora, sendo a imprecisão por microtremores involuntários um obstáculo crítico no controle assistivo por movimentos da cabeça. Propõe-se um mouse virtual preditivo integrando visão computacional (MediaPipe), classificador Random Forest e magnetismo adaptativo web. Treinado com 1.980 amostras e cinco features físicas fundamentadas na Lei de Fitts, o modelo prediz intenção de clique com 70,45% de acurácia (*$69{,}51 \pm 3{,}20\%$ *em validação cruzada), aplicando assistência gradual condicionada à confirmação facial. Operando em hardware convencional sem GPU, a solução consolida-se como alternativa inclusiva de baixo custo.*

1. Introdução

O computador tornou-se um recurso imprescindível nas atividades humanas. Contudo, o paradigma dominante de dispositivos de entrada, como o mouse óptico e os *touchpads*, pressupõe plena capacidade motora para a execução de movimentos precisos, impondo barreiras a pessoas com deficiência físico-motora. Para minimizar esse cenário, as tecnologias assistivas combinam computação, engenharia e ciências da saúde para ampliar a autonomia desses indivíduos, conforme estabelece a Lei Brasileira de Inclusão (Lei n.º 13.146/2015) [Brasil 2015].

Nesse contexto, as Interfaces Naturais de Usuário (NUI) [Wigdor and Wixon 2011] viabilizam o rastreamento de movimentos faciais por *webcam* convencional, sem *hardware* especializado e sem contato físico. Entretanto, a imprecisão inerente — microtremores involuntários, variações de iluminação e latência — compromete a seleção de alvos pequenos, isto é, os elementos interativos de destino

(botões, links, ícones) que o usuário busca acionar, criando uma barreira direta à usabilidade.

A fim de solucionar essa lacuna, este trabalho propõe um sistema de mouse virtual que trata o controle não apenas como rastreamento espacial, mas fundamentalmente como um problema de classificação de intenção. O sistema integra rastreamento facial via MediaPipe [Google 2026], um modelo de aprendizado de máquina supervisionado para classificação de intenção binária baseado em Random Forest [Breiman 2001] e automação de interface por meio de magnetismo adaptativo — mecanismo que atrai o cursor em direção ao elemento provável, reduzindo artificialmente a distância a ser percorrida —, fundamentado nos princípios de facilitação de apontamento de cursores de área e alvos expansíveis [Worden et al. 1997, McGuffin and Balakrishnan 2002]. O escopo desta versão do protótipo é a navegação em páginas web no navegador Chrome, cujo modelo estrutural (DOM) fornece em tempo real a geometria dos elementos interativos (posição e largura dos alvos) que alimenta tanto o classificador quanto o magnetismo adaptativo; a extensão a aplicações nativas do sistema operacional é discutida como trabalho futuro na Seção 6.

A hipótese central é que a integração de um classificador probabilístico (Random Forest) com magnetismo adaptativo contextual reduz a distância efetiva de apontamento, mitigando microtremores na seleção de alvos em interfaces web. O trabalho está estruturado da seguinte forma: a Seção 2 revisa trabalhos relacionados; a Seção 3 apresenta a fundamentação teórica; a Seção 4 descreve a metodologia; a Seção 5 apresenta os resultados fundamentados na Lei de Fitts; e a Seção 6 apresenta as considerações finais e as perspectivas de trabalhos futuros.

2. Trabalhos Relacionados

2.1. Sistemas baseados em movimentos corporais

Ashtankar et al. [Ashtankar et al. 2015] demonstram a viabilidade do controle de cursor por movimento da cabeça em *hardware* convencional, sem uso das mãos, ainda que sem avaliação quantitativa de desempenho. Trabalhos posteriores ampliaram o escopo para além do controle de cursor: Haseeb et al. [Haseeb et al. 2018] aplicaram gestos da cabeça ao comando de robôs assistivos, evidenciando a consolidação do rastreamento cefálico como canal de interação sem contato. Já Sairam, Gowra e Kopparapu [Sairam et al. 2022] mapeiam gestos manuais captados por webcam para ações do cursor, mas dependem de gestos manuais deliberados, o que restringe o uso a pessoas com preservação motora dos membros superiores. Mais recentemente, Goulart e Buiatti [Goulart and Buiatti 2025] atestaram a eficácia de um mouse virtual baseado em MediaPipe e OpenCV, apontando a sensibilidade à iluminação como um desafio prático. Em uma abordagem técnica muito próxima à deste trabalho, Santos et al. [Santos et al. 2025] provaram que o uso conjunto de MediaPipe e PyAutoGUI supera a instabilidade da biblioteca DLIB, resultando em trajetórias mais fluidas e precisas.

2.2. Facilitação de apontamento

Para reduzir a dificuldade de seleção em interfaces, Worden et al. [Worden et al. 1997] introduziram cursores de área com *hot spot* dinâmico, comprovando que ampliar a região de ativação melhora o desempenho sobre alvos pequenos. McGuffin e Balakrishnan [McGuffin and Balakrishnan 2002] investigaram o aumento dinâmico de alvo, demonstrando que a expansão pode ocorrer a 90% do trajeto percorrido e, ainda assim,

produzir um ganho de desempenho equivalente ao de uma expansão precoce. Apesar da viabilidade, ambas as soluções implementam uma atração estritamente reativa, sem a capacidade de inferir o contexto da interface ou modelar antecipadamente a intenção do usuário.

2.3. Aprendizado de máquina e intenção do usuário

Diferentemente da estimação espacial da Seção 2.1, que localiza a cabeça ou o cursor na tela, este eixo trata de um problema distinto: inferir se um movimento em curso constitui uma aproximação deliberada a um elemento clicável ou um deslocamento exploratório. Kang, Lee e Kang [Kang et al. 2017] aplicaram Random Forest à estimação de pose da cabeça, comprovando a robustez do algoritmo a variações de iluminação e oclusão análogas às do rastreamento por webcam. A diferença essencial está na natureza da saída do modelo. No trabalho de Kang et al., a saída é contínua — estima-se a orientação da cabeça, uma tarefa de regressão. No sistema proposto, a saída é categórica — decide-se entre clique e navegação, uma tarefa de classificação binária. É essa reformulação, de regressão de pose para classificação de intenção, que distingue a abordagem proposta.

O preenchimento dessa lacuna é a contribuição central deste trabalho, que subordina o magnetismo de interface [Worden et al. 1997, McGuffin and Balakrishnan 2002] a uma estimativa probabilística de intenção obtida por Random Forest [Breiman 2001]. Diferentemente das abordagens reativas da Seção 2.2, a assistência preditiva manifesta-se apenas como projeção visual: caso a inferência seja incorreta, a continuidade do movimento do usuário se sobrepõe à predição, sem gerar resistências físicas indesejadas.

3. Fundamentação Teórica

3.1. Lei de Fitts e Facilitação de Apontamento

A Lei de Fitts [Fitts 1954] é um dos modelos mais consolidados na IHC para descrever o desempenho humano em tarefas de apontamento. O modelo estabelece que o tempo necessário para mover um cursor até um alvo é uma função logarítmica da distância até o alvo e de sua dimensão. Neste trabalho, adota-se a formulação que integra o fator multiplicativo original de Fitts à constante da Formulação de Shannon, abordagem predominante nos estudos de IHC [McGuffin and Balakrishnan 2002, MacKenzie 1992], conforme a equação:

$$T = a + b \cdot \log_2 \left(\frac{2D}{W} + 1 \right) \tag{1}$$

onde T é o tempo de movimento, D a distância inicial, W a largura do alvo, e a e b são constantes empíricas. O termo $2D$ assegura um índice de dificuldade (ID) estritamente positivo e preserva a monotonicidade da variável espacial, otimizando o particionamento do espaço de características pelo classificador.

Neste trabalho, a Lei de Fitts cumpre dois papéis. Primeiro, o índice de dificuldade (ID) é empregado como um dos cinco atributos cinemáticos que alimentam o classificador de intenção (Seção 4). Segundo, a relação entre a largura do alvo e o tempo de apontamento orienta a calibração do magnetismo adaptativo: reduzir artificialmente a distância efetiva D equivale a diminuir o índice de dificuldade e, portanto, o tempo de aquisição previsto pela Equação 1. Os raios de atração são ajustados à dimensão

típica de cada elemento de tela, estendendo o conceito de cursores de área de Worden et al. [Worden et al. 1997] ao contexto web. A contribuição relativa do ID frente aos demais atributos é avaliada empiricamente na Seção 5, evitando atribuir a priori centralidade a qualquer atributo isolado. A hipótese de que o perfil cinemático parcial de um apontamento carrega informação sobre o alvo pretendido possui consolidação estabelecida na literatura de predição de alvo [Lank et al. 2007], o que fundamenta o uso desses atributos para a inferência de intenção.

3.2. Detecção de Marcos Faciais e Mapeamento de Gestos

A identificação de regiões específicas da face em sistemas de visão computacional é comumente realizada por meio da extração de pontos de referência (*landmarks*). Essa técnica permite modelar a geometria facial sem a necessidade de sensores invasivos. Para a análise específica da região ocular, Soukupová e Čech [Soukupová and Čech 2016] propuseram o *Eye Aspect Ratio* (EAR), uma métrica baseada nas distâncias bidimensionais entre os marcos das pálpebras, definida pela equação:

$$\mathrm{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2 \cdot \|p_1 - p_4\|}, \tag{2}$$

onde $p_1, \ldots, p_6$ representam as coordenadas faciais ao redor do olho. O numerador calcula a abertura vertical, enquanto o denominador representa a largura horizontal. A principal vantagem teórica do EAR é ser parcialmente invariante à pose da cabeça e completamente invariante à escala da imagem, garantindo que o algoritmo de detecção funcione independentemente da distância do usuário em relação à câmera.

Adicionalmente, o sistema monitora o índice de sorriso, definido pela razão entre a distância horizontal das comissuras labiais e a largura facial, que atua como um gesto de confirmação voluntária. Ao detectar um sorriso, o sistema interpreta esse estado como uma aprovação do usuário à trajetória predita pelo magnetismo adaptativo, executando a ação de posicionamento de forma determinística. Esse mecanismo oferece uma modalidade de interação proativa e expressiva, sendo especialmente relevante para usuários com amplitude de movimento físico-motora reduzida.

No contexto deste trabalho, essa fundamentação é operacionalizada por meio do *framework* MediaPipe Face Mesh [Google 2026], que extrai 478 *landmarks* tridimensionais em tempo real. O mapeamento espacial do cursor é guiado pelo *landmark* central da face (ponta do nariz, índice 1), que é projetado no plano da tela. A partir do monitoramento contínuo do EAR, o sistema implementa uma máquina de estados que distingue os padrões temporais: a piscada rápida simples (duração inferior a 300 ms), mapeada para o clique esquerdo, e o fechamento prolongado dos olhos (superior a 500 ms), mapeado para o *hold click*. Essa distinção temporal reproduz a recomendação de Soukupová e Čech [Soukupová and Čech 2016] de tratar o fechamento ocular prolongado como evento distinto da piscada reflexa, garantindo que apenas comandos intencionais do usuário ativem ações de interface.

Por fim, para corrigir os desvios posturais acumulados no uso prolongado, a abertura da boca — medida pela distância vertical entre os *landmarks* labiais — aciona uma recalibração que redefine a origem das coordenadas de rastreamento para a direção atual da face.

3.3. Random Forest

O algoritmo Random Forest [Breiman 2001] é um método de aprendizado de máquina baseado em *ensemble learning* reconhecido por sua robustez e capacidade de generalização. O modelo constrói múltiplas árvores de decisão independentes, treinadas sobre amostras aleatórias obtidas por *bagging* (*bootstrap aggregating*) para mitigar o sobreajuste (*overfitting*); a predição final resulta da agregação probabilística dos estimadores.

No contexto deste sistema, o Random Forest justifica-se por sua natureza *ensemble*, que suaviza ruídos e microtremores do rastreamento facial; pelas estimativas probabilísticas contínuas, que viabilizam as políticas graduais do magnetismo adaptativo; e pela interpretabilidade por importância de atributos, que permite validar a engenharia de *features* a baixo custo de inferência, compatível com a execução em tempo real em *hardware* doméstico.

4. Metodologia

4.1. Arquitetura do Sistema

A arquitetura do mouse virtual foi projetada para operar em tempo real, adotando processamento assíncrono em duas *threads* para evitar gargalos de latência. O fluxo de dados entre as camadas de entrada, *threads* de processamento, núcleo de decisão preditiva e atuadores de saída é ilustrado na Figura 1.

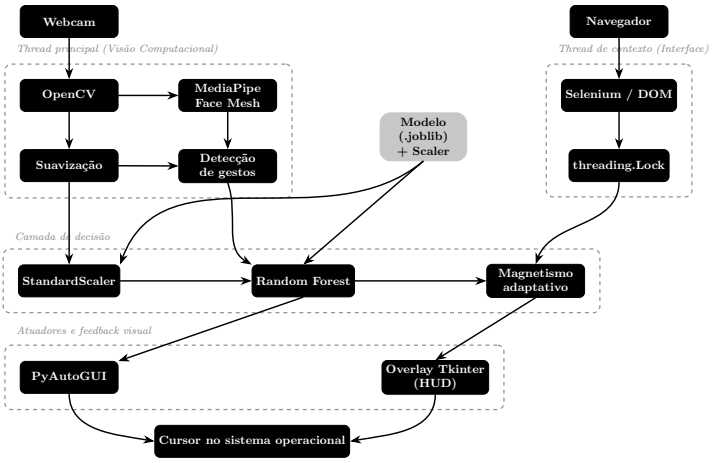


Figura 1. Diagrama do sistema de controle do mouse virtual.

Nessa estrutura, a *thread* principal inicia o processamento com a captura de vídeo via OpenCV, seguida pela extração de 478 *landmarks* faciais por meio do MediaPipe Face Mesh. A partir desses pontos, executa-se a suavização da trajetória por média móvel simultaneamente à extração de gestos e da métrica EAR. Em seguida, as variáveis passam por uma etapa de normalização para alimentar a inferência probabilística do modelo Random Forest.

Por fim, a inferência probabilística é operacionalizada via biblioteca PyAutoGUI, atuando na interface por meio do magnetismo adaptativo e do *Ghost Cursor* — uma projeção que permite ao usuário validar o alvo inferido antes de qualquer deslocamento físico. Simultaneamente, a *thread* de contexto mapeia a árvore estrutural

Document Object Model (DOM) da página com o *framework* Selenium e a biblioteca `undetected_chromedriver`. O uso dessa variação indetectável do *driver* evita bloqueios *anti-bot*, garantindo a extração dos elementos interativos. O *feedback* visual ao usuário é provido por um *overlay* gráfico (Tkinter), que exibe o *Head-Up Display* (HUD) com os níveis de assistência do magnetismo e a projeção do *Ghost Cursor* sobre o alvo inferido.

Seja P a probabilidade de intenção de clique estimada pelo Random Forest (Seção 3.3) a cada quadro, e d a distância euclidiana entre o cursor e o elemento interativo mais próximo, obtida do mapeamento do DOM. O magnetismo escalona a assistência em quatro modos, combinando esses dois parâmetros de decisão. Quando $P < 0{,}55$, o sistema opera em **Modo Livre**, sem qualquer assistência, e o cursor segue estritamente o movimento da cabeça. Quando $P \geq 0{,}55$, a assistência é habilitada e sua intensidade passa a depender de d: para o alvo ainda distante, aplica-se a **Assistência Suave**, uma atração direcional gradual que reduz progressivamente a distância efetiva; à medida que o cursor se aproxima, ativa-se o **Foco Forte**, que intensifica a atração; e, ao entrar no raio de ancoragem do alvo, o **Snap** captura o alvo de forma determinística.O gatilho de transição é, portanto, duplo: P decide *se* há assistência, e d decide *qual* o grau de atração. Em todos os casos, o deslocamento efetivo até o alvo só se consuma após a confirmação voluntária pelo sorriso (Seção 3.2), preservando o controle do usuário.

4.2. Coleta de Dados e Pré-processamento

Para compor o *dataset* de treinamento, desenvolveu-se um protocolo de navegação supervisionada com gatilhos de intenção em teclado (`F2` para origem, `F9` para destino, `F10` para validação da sequência). O ambiente experimental foi documentado para garantir reprodutibilidade: *webcam* HD integrada (720p, 30 fps), distância focal do usuário entre 55 e 65 cm, e iluminação ambiente ≥ 300 lux (verificada via luxímetro), processados em um notebook Acer Aspire 5 equipado com CPU Intel Core i5-10210U e 8 GB de RAM, sem aceleração gráfica dedicada, simulando fielmente condições reais de *hardware* doméstico.

Na etapa de pré-processamento, o problema de regressão de coordenadas foi convertido em classificação binária de intenção. Amostras cuja distância euclidiana final ao elemento interativo mais próximo resultou inferior a 100 px foram rotuladas como Clique (1); as demais, como Navegação (0).

O vetor de características (*features*) foi reformulado com base na Lei de Fitts em cinco atributos físicos do movimento: distância (D), velocidade (px/s), índice de dificuldade ($ID = \log_2(2D/W + 1)$), tempo (s) e aceleração (px/s²). Amostras acima do percentil 99 de velocidade ou aceleração foram removidas como *outliers* de rastreamento facial. As variáveis foram normalizadas por *z-score* (`StandardScaler`), ajustado sobre o conjunto de treinamento. A base totalizou 2.000 sequências brutas de quatro participantes sem comprometimento motor (criança negra de 12 anos, adulta negra de 20 anos e duas adultas pardas de 22 e 55 anos), cobrindo diversidade de morfologia facial, resultando em 1.980 amostras após o tratamento.

4.3. Treinamento do Classificador

O conjunto de 1.980 amostras, após o tratamento, foi particionado em 80% para treinamento e 20% para testes (396 amostras), com estratificação pelo identificador de participante. A otimização de hiperparâmetros foi conduzida via *Grid Search* com validação

cruzada de 5 *folds*, explorando `n_estimators` $\in \{50, 100, 200, 300\}$, `max_depth` $\in \{5, 10, 15, \text{None}\}$ e `min_samples_split` $\in \{2, 5, 10\}$. O melhor estimador foi instanciado com semente fixa (`random_state=42`).

5. Resultados e Discussão

O modelo Random Forest otimizado via *Grid Search* atingiu acurácia de 70,45% no conjunto de teste (396 amostras). A validação cruzada com 5 *folds* sobre o conjunto de treinamento confirmou a estabilidade do modelo, com acurácia média de 69,51 ± 3,20% e intervalo empírico de dispersão (média ± 2 desvios-padrão) entre 63,11% e 75,90%, indicando ausência de sobreajuste. Ressalta-se que esse intervalo constitui medida descritiva da variabilidade entre partições, e não um intervalo de confiança inferencial estrito, dado o número reduzido de *folds*. A reformulação do espaço de *features*, de variáveis contextuais estáticas para atributos físicos fundamentados na Lei de Fitts, representa uma mudança de paradigma: o modelo passa a classificar o padrão cinemático do movimento em vez da proximidade espacial ao alvo.

Contudo, em sistemas assistivos, a assimetria do custo dos erros possui impacto ergonômico superior à precisão global. Um falso negativo priva o usuário da assistência necessária, gerando fadiga motora, enquanto um falso positivo projeta o *ghost* indevidamente, mas é imediatamente neutralizado pela ausência do sorriso de confirmação. A Tabela 1 detalha o desempenho por classe sob essa ótica.

Tabela 1. Desempenho do classificador no conjunto de teste (396 amostras).

Classe	Precisão	Recall	F1-score	Suporte
Navegação (0)	0,68	0,79	0,73	203
Clique (1)	0,73	0,62	0,67	193
Média macro	0,71	0,70	0,70	396
Acurácia	–	–	0,70	396

O *recall* de 0,62 na classe Clique indica que em 62% das aproximações intencionais o sistema reconheceu corretamente a intenção, ativando o *Ghost Cursor* e reduzindo o esforço motor do usuário. O *recall* superior para a classe Navegação (0,79) reflete a tendência conservadora do modelo, preferível em termos ergonômicos, pois minimiza falsas ativações do magnetismo. A precisão equilibrada entre as classes (0,68 e 0,73) indica que, quando o sistema ativa a assistência, ela tende a ser pertinente.

5.1. Perfil Cinemático e Dificuldade de Apontamento

A Tabela 2 sintetiza o perfil cinemático das sequências válidas. Enquanto o treinamento e o teste do classificador utilizaram as 1.980 amostras, esta caracterização descritiva aplica um corte adicional de *outliers* na variável tempo (percentil 99), resultando em 1.960 sequências, e evidencia a variabilidade natural do apontamento por rastreamento facial.

A elevada média de tempo (5,80,s) e seu desvio-padrão (31,51,s) refletem pausas naturais da coleta, tornando a mediana (1,72,s) mais representativa. A assimetria da velocidade (mediana 118,3 contra média 231,5,px/s) reflete o padrão típico do apontamento facial: fases de aproximação rápida intercaladas com ajustes finos próximos ao

Tabela 2. Perfil cinemático das sequências válidas após corte adicional de *outliers* temporais — percentil 99 (1.960 amostras).

Variável	Mediana	Média $\pm$ DP
Tempo de movimento (s)	1,72	$5,80 \pm 31,51$
Velocidade (px/s)	118,3	$231,5 \pm 341,3$
Índice de Fitts ID (bits)	2,19	$2,14 \pm 1,39$

alvo, corroborando a Lei de Fitts. O ID médio de 2,14 bits, com amplitude de 0 a 7,36 bits, confirma que o classificador foi exposto a dificuldade variada, não a cenários artificialmente lineares.

Tabela 3. Matriz de confusão — conjunto de teste (396 amostras).

		Previsão	
		Navegação (0)	**Clique (1)**
Realidade	Navegação (0)	160	43
	Clique (1)	74	119

A Tabela 3 detalha os erros: dos 193 cliques reais, 119 foram identificados (TP) e 74 não (FN); dos 203 movimentos de navegação, 160 foram corretos (TN) e 43 geraram ativações indevidas do *ghost* (FP), neutralizadas pela ausência de confirmação facial. A importância dos atributos, medida pela redução média de impureza no *ensemble*, situou o Índice de Dificuldade como o mais influente (0,262), seguido de tempo (0,208), distância (0,198), aceleração (0,173) e velocidade (0,158). A liderança do ID — que integra a geometria do alvo à amplitude do movimento — corrobora a hipótese central, indicando ser a dificuldade da tarefa o sinal mais discriminativo da intenção; a contribuição equilibrada dos cinco atributos é coerente com a interdependência algébrica entre eles.

5.2. Comportamento em Tempo Real

A Figura 2 ilustra o sistema em um cenário real de navegação, evidenciando a atuação proativa do *Ghost Cursor* sobre a caixa de pesquisa do Google. O *Head-Up Display* (HUD) — interface gráfica de *feedback* sobreposta ao fluxo de vídeo — atesta a ativação do modo *ASSIST_SUAVE* e projeta graficamente a trajetória de atração magnética direcionando o cursor ao alvo predito. Além disso, a interface exibe o índice de sorriso computado em tempo real, que atua como gesto de confirmação da trajetória magnética predita.

6. Considerações Finais

Este trabalho apresentou um sistema de mouse virtual preditivo para acessibilidade motora, integrando visão computacional de baixo custo, mapeamento estrutural da página e aprendizado de máquina. A inovação central está na divisão de etapas: o classificador infere a intenção do usuário, enquanto um módulo independente gerencia a execução física do movimento. Substituir a predição contínua de coordenadas (regressão) por um classificador de intenção binária contorna o ruído dos micromovimentos da cabeça, transferindo o desafio da precisão ao magnetismo adaptativo e ampliando a autonomia do usuário.

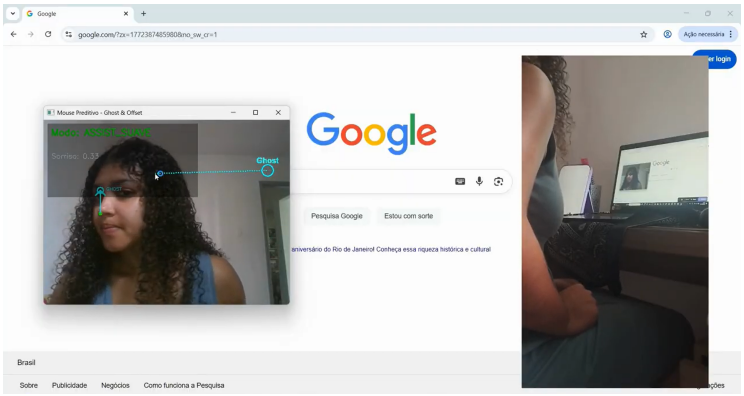


Figura 2. Aplicação do sistema em ambiente real: ativação do *Ghost Cursor* em campo de busca.

Os resultados evidenciam a viabilidade técnica da proposta em *hardware* convencional e sem GPU, com acurácia de 70,45% (Seção 5) e perfil conservador adequado a sistemas assistivos. A liderança do Índice de Dificuldade de Fitts na importância de atributos corrobora a hipótese central. Ademais, a subordinação do deslocamento magnético à confirmação facial, complementada pelo mecanismo de *break-free*, preserva a autonomia do usuário e mitiga riscos de aprisionamento involuntário.

Não obstante os resultados, o protótipo atual apresenta limitações de naturezas sensorial, metodológica e contextual. No aspecto sensorial, observa-se a degradação do rastreamento sob iluminação inferior a 200 lux, bem como a necessidade de calibração manual do limiar de piscada (`EAR_THRESHOLD = 0,22`) para adequar o sistema a variações fisiológicas individuais. Sob a ótica metodológica, os testes de verificação funcional foram conduzidos com um grupo de quatro participantes sem comprometimento motor, com o objetivo de validar a robustez do rastreamento facial a diferentes morfologias faciais e condições de uso; testes com usuários com deficiência motora constituem etapa prevista para trabalhos futuros. No plano contextual, a arquitetura atual limita a extração do contexto estrutural ao navegador Chrome por intermédio do Selenium, o que inviabiliza a assistência preditiva em aplicações nativas do sistema operacional e ocasiona latência pontual frente a interfaces com alto volume de elementos interativos.

Diante dessas delimitações, os trabalhos futuros concentram-se em: substituição da arquitetura DOM/Selenium por APIs nativas (*Microsoft UI Automation*) e testes quantitativos com múltiplos perfis motores e extração das constantes a e b da Lei de Fitts mediante regressão linear com grupo controle. Com esses aprimoramentos, a solução consolida-se como um avanço prático no campo das tecnologias assistivas, promovendo a inclusão digital e a autonomia asseguradas pela Lei Brasileira de Inclusão da Pessoa com Deficiência (Lei n.º 13.146/2015)[1]

[1]Declaração de Uso de IA (Portaria CNPq nº 2.664/2026): O modelo de linguagem Google Gemini foi utilizado exclusivamente na fase de editoração para revisão gramatical e adequacão de formatação textual. A autoria, a validação dos dados e a responsabilidade integral pela originalidade do artigo são exclusivas dos autores.

Referências

Ashtankar, H. et al. (2015). Mouse control using head movement. *Journal of Advance Research in Computer Science & Engineering*, 2(3):95–98.

Brasil (2015). Lei nº 13.146, de 6 de julho de 2015. Lei Brasileira de Inclusão da Pessoa com Deficiência. Legislação federal.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391.

Google (2026). MediaPipe Face Landmarker. Disponível em: `https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker`.

Goulart, R. F. and Buiatti, R. C. (2025). Uso da visão computacional para desenvolver mouse virtual e controlar movimentos capturados pela webcam. *Revista DELOS*, 18(67):e5075.

Haseeb, M. A., Kyrarini, M., Jiang, S., Ristic-Durrant, D., and Gräser, A. (2018). Head gesture-based control for assistive robots. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference (PETRA '18)*, pages 379–383. ACM.

Kang, M.-J., Lee, J.-K., and Kang, J.-W. (2017). Combining random forest with multi-block local binary pattern feature selection for multiclass head pose estimation. *PLOS ONE*, 12(7):e0180792.

Lank, E., Cheng, Y.-C. N., and Ruiz, J. (2007). Endpoint prediction using motion kinematics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, pages 637–646.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human–Computer Interaction*, 7(1):91–139.

McGuffin, M. and Balakrishnan, R. (2002). Acquisition of expanding targets. In *CHI Conference on Human Factors in Computing Systems*, pages 57–64.

Sairam, U., Gowra, D. K. R., and Kopparapu, S. C. (2022). Virtual mouse using machine learning and gui automation. In *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 1112–1117. IEEE.

Santos, J. D. M. et al. (2025). Visionhead: alternativa de uso do mouse para pessoas tetraplégicas utilizando visão computacional. In *Anais do LII SEMISH 2025*, pages 549–560. SBC.

Soukupová, T. and Čech, J. (2016). Real-time eye blink detection using facial landmarks. In *Computer Vision Winter Workshop*.

Wigdor, D. and Wixon, D. (2011). *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann.

Worden, A. et al. (1997). Making computers easier for older adults to use: area cursors and sticky icons. In *Proceedings of the ACM SIGCHI Conference*, pages 266–271.