

A Proposal of a Universal Remote Control System Based on Head Movements

Cassio T. Batista

Laboratório de Visualização,
Interação e Sistemas
Inteligentes
Universidade Federal do Pará
Belém, Brasil
cassiotb@ufpa.br

Erick M. Campos

Laboratório de Visualização,
Interação e Sistemas
Inteligentes
Universidade Federal do Pará
Belém, Brasil
erickcampos@ufpa.br

Nelson C. Sampaio Neto

Laboratório de Visualização,
Interação e Sistemas
Inteligentes
Universidade Federal do Pará
Belém, Brasil
nelsonneto@ufpa.br

ABSTRACT

Technological developments converge to make people interact with electronic devices in an easy way. For people with disabilities, however, that interaction becomes something more than simple: it becomes possible. The current work presents a proposal of an open-source, low cost universal remote control system that translates user's head poses into commands to electronic devices. In addition, a proximity sensor circuit was combined to radio-frequency modules in order to act as a wireless switch. Among the applications found with respect to alternative remote controls, none of them supports head gestures as input, which would certainly make them a viable option to people whose upper limbs are compromised. A mean opinion score questionnaire was applied to volunteers in order to evaluate the system. The results show great interest of the users in this kind of technology.

Author Keywords

Head pose estimation; Assistive Technology; universal remote control; mean opinion score

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces; I.4.9. Image Processing and Computer Vision: Applications; J.7. Computer Applications: Computers in Other Systems; K.4.2. Computers and Society: Social Issues

INTRODUÇÃO

Com o passar dos anos, a interação das pessoas com as máquinas tem-se tornado cada vez mais simples. O que antes era portado somente por empresas e pessoas com poder financeiro diferenciado e conhecimento acima da média, em termos de tecnologia, é hoje muito mais acessível e fácil para usuários domésticos sem profundo entendimento do assunto. Devido à abrangência dos computadores (pessoais e embarcados) e

da Internet, novas oportunidades e expectativas em termos de trabalho, estudo e até lazer são criadas a fim de melhorar ainda mais essa comunicação, de modo que a máquina se aproxime mais de ações típicas do ser humano, como pensar, agir e falar.

Acredita-se que a tecnologia de reconhecimento de gestos (AGR, do inglês *active gesture recognition*) [5] torna a interface humano-computador muito mais prática e natural, de forma que a comunicação de fato assemelha-se àquela estabelecida entre duas pessoas. O AGR refere-se ao sistema que, utilizando técnicas de computação visual para processar *frames* de um vídeo recebido como entrada, é capaz de definir na saída a ação referente ao movimento motor realizado pelo usuário. Dentre os inúmeros sistemas que envolvem reconhecimento de gestos, destacam-se os de automação residencial, os quais promovem comodidade e praticidade às pessoas no controle de equipamentos eletrônicos; e os que visam ajudar pessoas que tenham dificuldades na utilização desses equipamentos. Normalmente enquadrados no conceito de Tecnologia Assistiva, tais sistemas tornam-se umas das poucas soluções acessíveis às pessoas com deficiência motora, cujas limitações concentram-se principalmente na locomoção e manipulação de certos objetos do cotidiano.

A Tecnologia Assistiva é um campo do conhecimento de característica interdisciplinar, dedicado a aumentar a independência e mobilidade de pessoas com deficiência (PCD), englobando produtos, metodologias, práticas e serviços que objetivam promover sua autonomia, qualidade de vida e inclusão social [6, 11]. O termo refere-se à tecnologia que fornece assistência às PCD a fim de reduzir os efeitos das deficiências e permitindo-lhes participar ativamente das suas respectivas rotinas. Segundo a Convenção sobre os Direitos das Pessoas com Deficiência [23], os olhares do século XXI deixam de enxergar as PCD como “objetos” de caridade, tratamento médico e proteção social, e passam a vê-las sob uma nova perspectiva, na qual as PCD são “sujeitos” ativos na sociedade, capazes de reivindicar seus direitos e tomar decisões baseadas em seu próprio consentimento.

A Tecnologia Assistiva busca, então, reduzir a dificuldade vivenciada por pessoas que precisam de soluções que não as deixem à margem da utilização de dispositivos eletrônicos. Visando diminuir a exclusão digital imposta às PCD

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. IHC'17, Proceedings of the 16th Brazilian symposium on human factors in computing systems. October 23-27, 2017, Joinville, SC, Brazil. Copyright 2017 SBC. ISBN 978-85-7669-405-2 (online).

pela dificuldade ou total incapacidade para manipular certos equipamentos, a acessibilidade é vista como elemento fundamental para elevar a autoestima e o grau de independência dessas pessoas. Além disso, as soluções apresentadas também podem ser úteis para as pessoas sem qualquer deficiência, já que o controle de equipamentos torna-se mais prático e confortável [26].

Nesse sentido, este trabalho buscou desenvolver um sistema de controle remoto universal de baixo custo baseado nas plataformas embarcadas Arduino e C.H.I.P., que foram utilizadas em conjunto como um *hub* centralizado para controlar diversos aparelhos eletrônicos. O sistema é capaz de processar, através de técnicas de computação visual, imagens de vídeo em tempo real, as quais são a base para o reconhecimento de gestos da cabeça que servem como interface alternativa para promover autonomia às PCD nas tarefas de controle. Além disso, um sensor de proximidade foi combinado a um módulo de rádio-frequência (RF) para atuar como um acionador externo sem fio, utilizado com o propósito de ativar o sistema de controle.

A ideia é que os usuários-alvo em potencial (pessoas com limitações motoras dos membros superiores, porém sem paralisia total; e que tenham os movimentos do pescoço preservados) consigam realizar seis movimentos básicos com a cabeça para emitir comandos do tipo ligar/desligar, aumentar/diminuir e avançar/recuar em uma grande variedade de equipamentos eletrônicos disponíveis no ambiente residencial. Vale ressaltar que todas as interfaces de programação (APIs, do inglês *application programming interfaces*), plataformas embarcadas e pacotes de *software* utilizados para a criação dos módulos e recursos do sistema estão disponíveis livremente na Internet.

O restante do trabalho está organizado da seguinte forma. A Seção 2 discute sobre trabalhos relacionados a respeito de formas alternativas de controle remoto, bem como sistemas de controle que utilizam movimentos da cabeça como entrada. A Seção 3 apresenta as principais ferramentas utilizadas no sistema, seguida da Seção 4, que fornece detalhes sobre o sistema de controle remoto proposto. Os resultados dos testes realizados com voluntários são relatados na Seção 5 e, finalmente, conclusões e trabalhos futuros são abordados na Seção 6.

TRABALHOS RELACIONADOS

Ajudar PCD a utilizar equipamentos eletrônicos de forma independente não é uma ideia recente. Diversos trabalhos já propuseram soluções para tentar minimizar os mais variados tipos de limitação. O Navigo [19], por exemplo, foi desenvolvido como um sistema multimodal para deduzir informações relevantes extraídas de pessoas com paralisia cerebral através de reconhecimento de fonemas, objetos e movimentos. Em [8], a acessibilidade na interação com aparelhos televisores, dada por um sistema de diálogo, foi abordada com relação às pessoas idosas. Já em [10], o nariz foi utilizado como ponto específico para controlar o cursor de um *mouse* através de técnicas de processamento de imagens, enquanto a função de *click* foi atribuída à transição cerrado-aberto dos lábios.

O sistema Handmote [21], por outro lado, foi configurado para processar imagens de gestos das mãos para controlar remotamente, via luz infravermelha, dispositivos domésticos. O

controle de aparelhos eletrônicos também foi implementado no ActiveIris [15], no qual técnicas de rastreamento ocular atuavam como método de entrada em uma interface desenvolvida para plataformas *desktop*. Já em [9], movimentos das mãos e antebraços eram reconhecidos também com o intuito de controlar aparelhos eletrônicos de forma mais natural. Outros trabalhos focam em *designs* limpos e uniformes para controles remotos universais em *smartphones*, como é o caso de [1].

A respeito de sistemas que utilizam os movimentos da cabeça, muitos são utilizados para controlar cadeiras de rodas através de acelerômetros [18, 22] e outros sensores de posição e orientação [12, 13] colocados em contato direto com a cabeça do usuário. Já em [2], movimentos de aprovação e rejeição em ambientes automotivos foram capturados por uma câmera e processados através de técnicas de computação visual.

Como se pode perceber, dentre a literatura consultada, nenhum dos trabalhos encontrados apresentou propostas que utilizam gestos da cabeça — capturados por uma câmera e processados por técnicas de computação visual — para o controle remoto de aparelhos eletrônicos com foco em acessibilidade. A escassez de trabalhos envolvendo acionadores externos também foi notada, bem como a de referências sobre dispositivos equivalentes sem fio e baseados em proximidade, não em contato físico ou pressão.

FERRAMENTAS UTILIZADAS

Esta seção apresenta as duas principais ferramentas de *hardware* utilizadas, descrevendo também algumas bibliotecas e dispositivos de entrada e saída utilizados para construir o sistema proposto. Os componentes, apresentados na Figura 1, serão discutidos com detalhes a seguir.

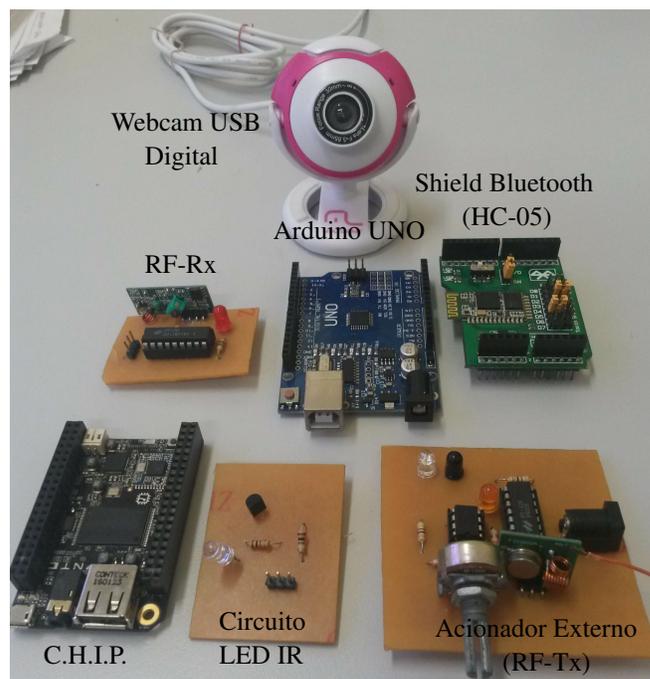


Figura 1. Componentes do sistema.

C.H.I.P.

C.H.I.P. (<https://getchip.com/pages/chip>) é um microcomputador *open-hardware* de baixo custo lançado em 2015 pela Next Thing Co. A plataforma embarcada é baseada na versão Debian Jessie do Linux, o que torna fácil a tarefa de compilar e executar *softwares* que já estão disponíveis para a versão *desktop* do sistema operacional.

A OpenCV (<http://opencv.org>), biblioteca *open-source* de computação visual, foi configurada no C.H.I.P. para ser a base para a realização do reconhecimento de gestos da cabeça. Uma *webcam* digital USB foi utilizada para capturar as imagens do rosto e da cabeça do usuário.

O C.H.I.P. envia dados referentes ao reconhecimento dos gestos para um Arduino via Bluetooth. Isso é feito através da API da biblioteca BlueZ (<http://www.bluez.org>) para a linguagem C (<https://people.csail.mit.edu/albert/bluez-intro>), a qual faz parte da lista de pacotes oficial do Linux.

Arduino

O Arduino (<https://www.arduino.cc>) é uma das plataformas mais famosas no mundo da eletrônica justamente por ser fácil de ser manipulada. Dentre as muitas versões existentes da plataforma, neste trabalho foi utilizado o Arduino UNO equipado com um *shield* Bluetooth baseado no módulo serial HC-05. A biblioteca `SoftwareSerial`, que acompanha o ambiente de desenvolvimento, foi utilizada para receber dados advindos do C.H.I.P. via Bluetooth através do HC-05.

O protocolo de comunicação entre o Arduino e o aparelho a ser controlado é definido por uma modulação por largura de pulso (PWM, do inglês *pulse-width modulation*) na frequência de 38 kHz. Como diferentes aparelhos decodificam informações de diferentes protocolos, a biblioteca `IRremote` (<https://github.com/z3t0/Arduino-IRremote>) foi utilizada para evitar a necessidade de implementar vários protocolos via *software*.

SISTEMA DE CONTROLE REMOTO PROPOSTO

O sistema foi construído sob a ideia do controle remoto universal e acessível às PCD. As plataformas embarcadas C.H.I.P. e Arduino foram configuradas para agir como um *gateway* (ou *hub*) centralizado no ambiente residencial, conforme mostrado no esquemático ilustrado na Figura 2.

Um acionador externo foi utilizado para que o usuário pudesse inicializar o sistema. Uma *webcam* USB digital, conectada ao C.H.I.P., é responsável por capturar, em tempo real, imagens do rosto e cabeça do usuário e repassá-las aos classificadores, os quais tentam traduzir determinados movimentos em comandos para o aparelho. Obedecendo à exigência imposta pela comunicação infravermelha entre o controle remoto e o dispositivo eletrônico, o *gateway* foi posicionado na “linha de visão” do aparelho a ser controlado. Quando a pose é classificada, um sinal é enviado via Bluetooth para um Arduino para que este possa, finalmente, emular o comando de controle correto e enviá-lo para o aparelho eletrônico através de um LED infravermelho (LED IR) acoplado a um circuito amplificador.

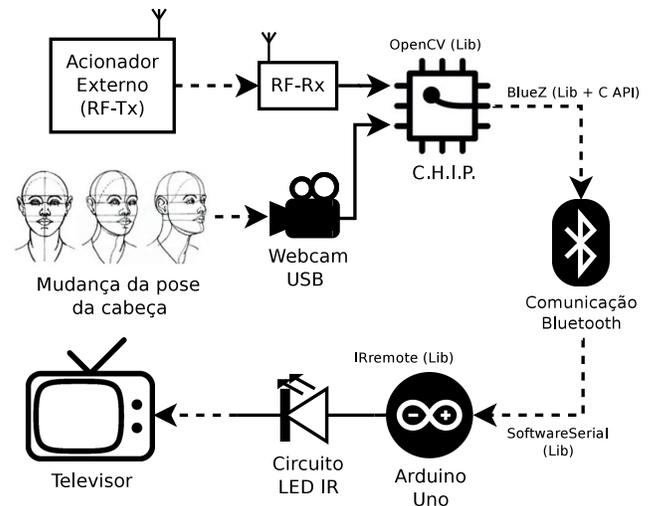


Figura 2. Esquemático do sistema.

Acionador Externo

Um acionador externo sem fio foi construído com o intuito de evitar que o sistema permanecesse sempre ativo, o que acarretaria diversos erros de reconhecimento por conta de movimentos involuntários do usuário. Um sensor de proximidade foi configurado para tal, utilizando um emissor e um receptor infravermelho, o qual reage a tensão quando algum objeto se aproxima. Portanto, não há necessidade de contato físico, e sim apenas a aproximação de uma mão, um braço, ou qualquer objeto. Essa mudança de tensão é amplificada pelo componente LM358 e enviada a um dos canais do circuito codificador HT12E, o qual modula uma forma de onda específica para ser enviada pelo módulo emissor (Tx) RF 433 MHz.

Conectado ao C.H.I.P., um circuito detecta o sinal na frequência 433 MHz através do módulo receptor (Rx) RF. O sinal passa pelo circuito decodificador HT12D antes de, finalmente, chegar a um pino digital do microcomputador, quando o processo de inicialização do sistema é devidamente efetuado.

Então, antes de realizar qualquer movimento, o usuário deve necessariamente inicializar o sistema com o acionador. Por esse motivo, espera-se que a movimentação dos membros superiores não seja totalmente comprometida, como no caso dos tetraplégicos, ou mesmo proibitiva, considerando os amputados. O sistema para automaticamente quando um movimento da cabeça é reconhecido ou quando um movimento é classificado como fora do padrão, resultando em uma situação de erro. Se o usuário desejar ligar a TV e aumentar o volume, por exemplo, o acionador deve ser utilizado duas vezes, uma para cada ação de controle.

Reconhecimento de Gestos da Cabeça

A etapa seguinte à captura das imagens é a de reconhecimento de gestos da cabeça. Como o C.H.I.P. ainda é bastante limitado (apesar de bastante poderoso quando comparado ao seu tamanho), a busca por algoritmos mais simples e com baixa complexidade foi fundamental. Assim, o presente sistema foi desenvolvido com base na solução proposta em [3], a qual chama a atenção por ter sido desenvolvida para ser executada

sobre plataformas embarcadas em tempo real, além de atingir taxas de acerto bastante satisfatórias. Em outras palavras, a proposta é uma junção de algoritmos já consolidados (nenhum algoritmo novo foi proposto), e faz uso das técnicas-chave de detecção facial e estimação da pose da cabeça, implementadas em sequência, conforme mostradas no fluxograma da Figura 3. Ambas serão abordadas nas seções seguintes.

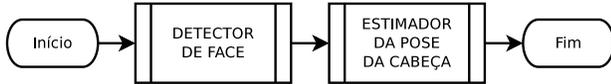


Figura 3. Visão geral do sistema de reconhecimento de gestos da cabeça.

Detecção Facial

A primeira etapa consiste em detectar um rosto através do algoritmo proposto por Viola e Jones [24, 25], implementado no método `detectMultiScale()` da OpenCV. Classificadores em cascata (*haar cascade*) são aplicados sobre as imagens capturadas pela *webcam* para tentar detectar um rosto. Se encontrado, um LED laranja, conectado ao C.H.I.P., acende, indicando que uma face foi detectada e que o algoritmo está se comportando conforme esperado; e a localização dos olhos e do nariz (baseada no modelo geo e antropométrico da face humana e calculada de forma puramente heurística e relativamente precisa) é estimada no plano bidimensional. Alguns resultados do processo de detecção facial e estimação dos três pontos são ilustrados na Figura 4.

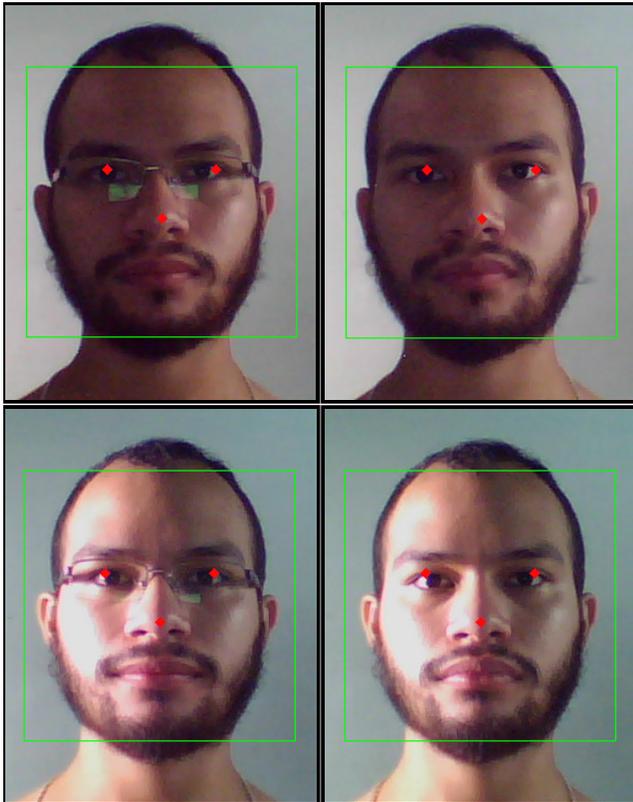


Figura 4. Detecção facial, realizada pelo algoritmo Viola-Jones, e cálculo heurístico da posição dos olhos e do nariz sob duas condições de iluminação diferentes, com e sem o uso de óculos.

O fluxograma do algoritmo de detecção facial descrito é mostrado na Figura 5. Se houver uma face em dez *frames* consecutivos, o 11º *frame* é armazenado, e a etapa seguinte de estimação da pose da cabeça, sinalizada pelo acendimento de um LED verde, é iniciada. Enquanto a condição de haver um rosto em dez quadros consecutivos não for atendida, a captura e classificação dos *frames* continua. Se os classificadores falharem na tarefa de detectar uma face, um LED vermelho sinaliza o erro.

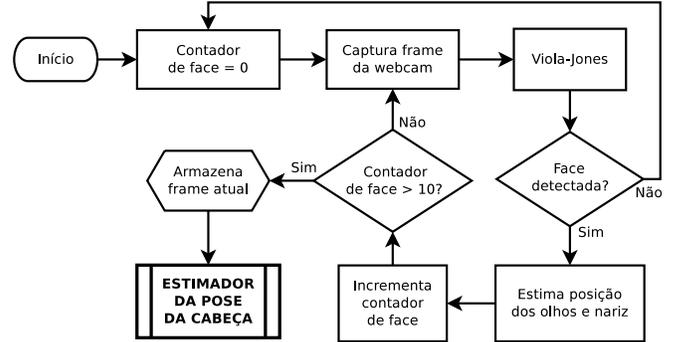


Figura 5. Algoritmo implementado para detecção facial.

Estimação da Pose da Cabeça

A segunda etapa realiza, efetivamente, o reconhecimento dos gestos através da estimação da pose da cabeça. O fluxo óptico é calculado através do algoritmo de Lukas-Kanade, apresentado em [16] e implementado no método `calcOpticalFlowPyrLK()` da OpenCV, o qual utiliza como referência os pontos dos olhos e do nariz que foram passados na etapa anterior, a fim de rastreá-los em relação a um próximo *frame* recém capturado pela câmera. Para tornar o algoritmo proposto mais confiável, as dimensões do rosto são verificadas em três condições simples que definem limites heurísticos com base na distância euclidiana. Sendo P_{OE} o ponto de localização do olho esquerdo; P_{OD} a localização do olho direito; e P_N a do nariz, ambos no *frame* atual:

- A razão entre a distância do olho esquerdo para o direito $\|P_{OE}P_{OD}\|$ e a distância de cada um dos olhos para o nariz $\|P_{OE}P_N\|$ deve estar entre 0,5 e 3,5.
- A distância entre o olho esquerdo e o direito $\|P_{OE}P_{OD}\|$ deve estar entre 20 e 160.
- A distância entre qualquer dos olhos e o nariz $\|P_{OE}P_N\|$ deve estar entre 10 e 150.

De posse das coordenadas dos pontos P_{OE} , P_{OD} e P_N no plano bidimensional e assumindo que as dimensões da face atendem às condições impostas acima, é possível calcular os eixos tridimensionais de rotação da cabeça (a saber: *roll*, *yaw*, e *pitch*) através de um conjunto de três equações:

$$roll = \arctan\left(\frac{y_{OE} - y_{OD}}{x_{OE} - x_{OD}}\right) \quad (1)$$

$$yaw = x'_N - x_N \quad (2)$$

$$pitch = y_N - y'_N, \quad (3)$$

onde, na Equação 1, que depende somente da localização dos olhos no *frame* atual, é assumido que (x_{OE}, y_{OE}) são as coordenadas do ponto de localização P_{OE} que representa o olho esquerdo no plano 2D; e (x_{OD}, y_{OD}) são as coordenadas do olho direito, localizado em P_{OD} . Já as Equações 2 e 3, que dependem somente da localização do nariz, precisam das coordenadas dos pontos identificados tanto no *frame* anterior quanto no atual, onde é assumido que $P'_N = (x'_N, y'_N)$ é o ponto de referência do nariz no *frame* anterior e $P_N = (x_N, y_N)$, a localização do nariz no *frame* atual.

Os valores obtidos pelas Equações 1, 2 e 3 foram filtrados com *thresholds* heurísticos baseados em movimentos suaves e pouco amplos, na tentativa de atender às limitações das PCD e, além disso, reduzir ruídos. Os valores são mostrados abaixo.

- *Roll*: os valores devem estar entre -60 e -15 para movimentos para a direita e entre 15 e 60 para a esquerda;
- *Yaw*: os valores devem estar entre 0 e 20 para movimentos para a direita e entre -20 e 0 para a esquerda;
- *Pitch*: os valores devem estar entre 0 e 20 para movimentos para a baixo e entre -20 e 0 para cima;

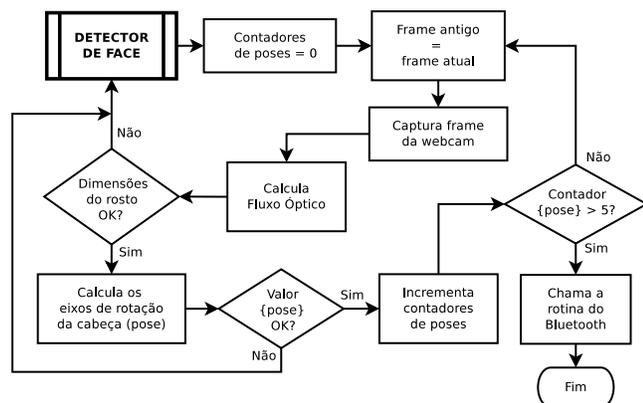


Figura 6. Algoritmo implementado para estimação da pose da cabeça.

O algoritmo para a estimação da pose da cabeça é ilustrado no fluxograma da Figura 6. Se o movimento de uma das três poses for detectado por 5 *frames* consecutivos, o gesto é reconhecido de forma correta e o respectivo sinal é mandado para o Arduino via Bluetooth. A API em C/C++ desenvolvida sobre a biblioteca BlueZ foi utilizada para iniciar a comunicação entre as duas plataformas embarcadas.

Controle Remoto do Televisor

A plataforma Arduino UNO foi utilizada para receber uma *string* diferente para cada respectivo gesto reconhecido no microcomputador e, por fim, traduzi-la em um comando específico para ser enviado a uma TV da Samsung. A informação do reconhecimento do gesto é recebida através de um *socket* serial criado com auxílio da biblioteca `SoftwareSerial`.

Os comandos da TV foram obtidos através da biblioteca `IRremote`. O sensor infravermelho VS 18388 foi conectado a um dos pinos analógicos do Arduino UNO para receber os comandos do controle original do aparelho televisor, exibindo no monitor serial a informação referente aos *bits* do protocolo de comunicação estabelecido entre o controle remoto e a respectiva TV. Esses comandos foram salvos e, então, a mesma biblioteca foi utilizada para enviá-los ao aparelho através de um LED infravermelho, o qual foi conectado a um dos pinos de PWM do Arduino UNO e inserido em um circuito amplificador para garantir um maior alcance do sinal modulado.

TESTES COM USUÁRIOS

Os testes foram realizados com pessoas de diferentes faixas etárias e graus de instrução, com diagnóstico de alguma doença que ocasionou deficiência ou mesmo sem limitação alguma. Vale ressaltar que deficiência motora dos membros superiores não significa amputação ou paralisia total, assim como deficiência visual ou auditiva, por exemplo, não significa cegueira ou surdez completas, respectivamente.

Perfil dos Participantes

O grupo de participantes foi composto por 12 pessoas do sexo masculino e cinco do sexo feminino, totalizando 17 pessoas com idades variando de 12 a 62 anos. Todos foram convidados a preencher um formulário simples, de modo que um perfil de pessoas fosse formado durante a análise para auxiliar na interpretação dos resultados dos testes. Os voluntários foram questionados sobre faixa etária; grau de escolaridade; tipo de deficiência; familiaridade com o uso de tecnologia e aparelhos eletrônicos; e facilidade no manuseio de controles remotos convencionais. O perfil demográfico dos participantes é mostrado na Figura 7.

Sete participantes apresentaram diagnóstico de deficiência motora. Um deles é portador de vasculite, uma síndrome que ocasiona desequilíbrio e baixa coordenação motora. Este também informou que utiliza controles remotos com certa dificuldade. Outro participante é portador de quadriplegia espástica, o que causa problemas de mobilidade nos membros inferiores e superiores. Este, por sua vez, declarou utilizar controles remotos sem nenhum problema. Um terceiro voluntário, desta vez portador de paralisia cerebral, disse não ser capaz de utilizar controles remotos devido à atrofia severa nos dedos das mãos. Sua coordenação, no entanto, foi suficiente para a realização de alguns movimentos do pescoço com sucesso. Os demais quatro participantes eram cadeirantes com limitações apenas no movimento das pernas (paraplegia). Três outros voluntários portadores de paralisia cerebral não foram incluídos nos resultados, pois, apesar de terem o cognitivo preservado e entenderem bem os comandos dados, não possuíam controle suficiente sobre os movimentos da cabeça, o que os impediu de executar qualquer um dos movimentos requeridos de forma efetiva.

Ambiente de Testes

Os experimentos com pessoas sem deficiência foram realizados, em sua maioria, no laboratório de teste da universidade, onde a iluminação é favorável e controlada. Já as PCD executaram o teste em uma sala especial das respectivas casas de

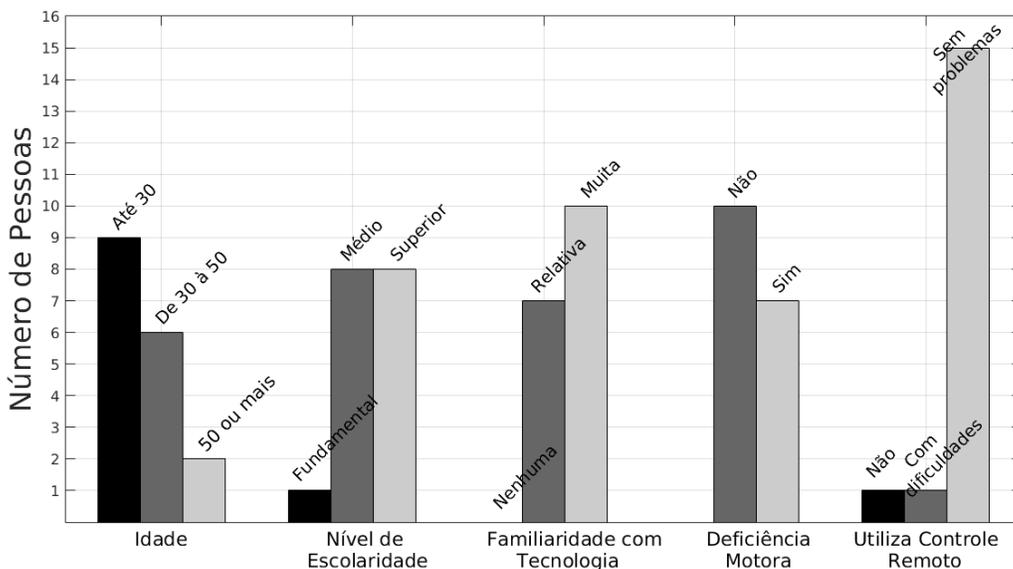


Figura 7. Perfil demográfico dos participantes.

apoio onde fazem tratamento, também com ambiente de iluminação controlado, enquanto uma única PCD realizou o teste em sua própria residência, onde a iluminação era fraca. Como os classificadores utilizados no Viola-Jones não foram treinados com exemplos de rostos em ambientes pouco iluminados, este é um fator fundamental na acurácia do sistema.

Após uma breve introdução sobre o grupo de pesquisa e sobre os objetivos do projeto, o funcionamento do sistema proposto foi explicado de forma verbal a cada um dos participantes. Os movimentos a serem executados em frente à câmera foram ilustrados e apresentados aos voluntários como forma de um manual de consulta, conforme mostrado na Figura 8.

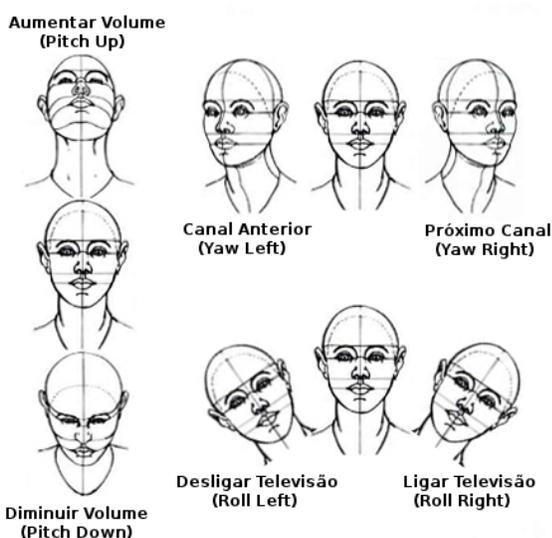


Figura 8. Seis movimentos de rotação da cabeça (ou poses) traduzidos em comandos para um televisor.

Os usuários foram, então, instruídos a seguir um roteiro de tarefas com os seis comandos para controle de um televisor através de gestos da cabeça. As tarefas foram listadas em ordem de execução, juntamente com as respectivas descrições, conforme mostrado na Tabela 1. Os participantes também foram instruídos a desistir após a quarta tentativa de cada comando, caso o sistema não se comportasse conforme desejado.

Movimento	Descrição
<i>Roll right</i>	Encostar a orelha no ombro direito
<i>Yaw right</i>	Encostar o queixo no ombro direito
<i>Pitch up</i>	Olhar para o teto
<i>Yaw left</i>	Encostar o queixo no ombro esquerdo
<i>Pitch down</i>	Olhar para baixo
<i>Roll left</i>	Encostar a orelha no ombro esquerdo

Tabela 1. Roteiro de tarefas a serem completadas pelo usuário.

Resultados

As Figuras 9 e 10 mostram, através de *boxplots*, a distribuição do número de tentativas executadas por cada usuário sem e com deficiência, respectivamente, a fim de realizar cada um dos seis movimentos para controlar a TV. Os *whiskers* são os números mínimo e máximo de tentativas que não são considerados *outliers*. Já as “caixas” (*boxes*), representadas por linhas tracejadas, abrigam a metade central da distribuição, ou seja, 25% dos valores acima da mediana (indicada sob os marcadores quadrados) e 25% dos valores abaixo.

Analisando a Figura 9, pode-se perceber que, para as pessoas sem deficiência, os movimentos de *pitch* e *yaw*, além de terem os limites das caixas entre 1 e 2, ainda obtiveram valor de mediana entre 1 e 1,5, o que significa que a maioria das pessoas precisou de apenas uma ou duas tentativas para realizar os respectivos comandos com sucesso. No entanto, houve um único caso em que ambos os movimentos de *yaw* não foram

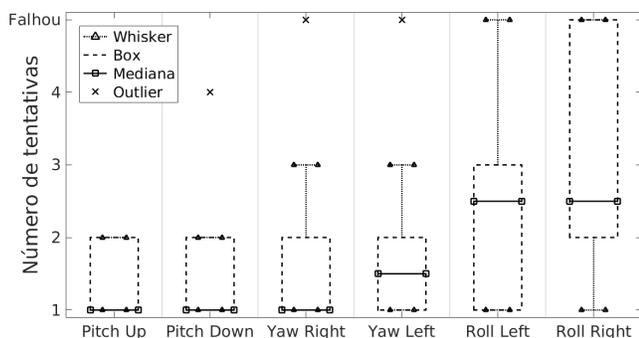


Figura 9. Número de tentativas de realização dos movimentos da cabeça por pessoas sem deficiência.

executados com sucesso, o que explica o valor *outlier* exibido sobre as respectivas caixas. Já a rotação lateral (*roll*) foi o movimento que mais foi executado de forma incorreta, o que tornou os valores mais próximos de 4, que é o limite máximo de tentativas permitidas no roteiro de tarefas. A razão observada para o insucesso do *roll*, relatada também por [3], recaí sobre a dificuldade em realizar o movimento de forma isolada, já que as pessoas acabaram realizando um *yaw* parcial com a cabeça.

A análise da Figura 10 confirma que o mesmo problema do *roll* ocorreu com as pessoas com deficiência, porém com uma frequência um pouco maior devido a algumas limitações na rotação da cabeça. No entanto, mais uma vez, *pitch* e *yaw* tiveram bons resultados, o que pode ser observado pelas caixas limitadas pelos valores 1 e 2,75, com medianas entre 1 e 2. Ou seja, as PCD precisaram de no máximo 3 tentativas para realizar os movimentos de forma correta.

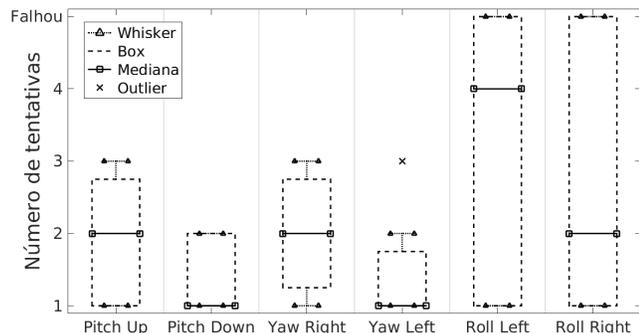


Figura 10. Número de tentativas de realização dos movimentos da cabeça por pessoas com deficiência motora.

Já a Figura 11 mostra a distribuição das tentativas somente para as pessoas com deficiência dos membros superiores. Como suas respectivas deficiências ocasionam comprometimento da maior parte do tronco, a dificuldade em executar o *roll* foi sentida com maior intensidade. Apenas uma das três PCD conseguiu desligar a TV já na última tentativa, enquanto houve falha em todas as outras tentativas de ligar e desligar o aparelho, o que explica o valor de mediana tão alto. Já o desempenho nos movimentos de *yaw* e *pitch* foram novamente bons, mantendo as medianas em 2 e os limites entre 1,25 e 2,75.

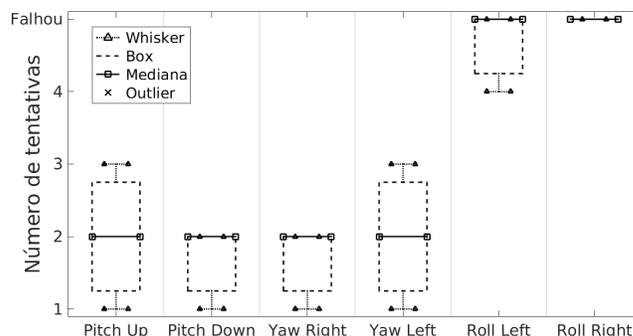


Figura 11. Número de tentativas de realização dos movimentos da cabeça por pessoas com deficiência motora dos membros superiores.

Para avaliar, de forma geral, a qualidade do sistema proposto, solicitou-se que os participantes respondessem a um questionário de duas perguntas ao final do teste. A primeira, objetiva, foi adaptada de [17] e é baseada na medida de qualidade MOS (do inglês *mean opinion score*), contendo cinco alternativas que classificam o sistema em uma escala que varia de “pés-simo” (1) a “excelente” (5). A escala de pontuação é mostrada, de forma similar à que foi exibida aos voluntários, na Tabela 2.

Escola	Descrição
Excelente	Trará(ia) muito conforto e bem-estar às PCD.
Bom	Pode(ria) ser útil para as PCD.
Razoável	É/seria pouco útil para PCD.
Ruim	Não é/não seria útil para PCD.
Péssimo	Causa/causaria desconforto às PCD.

Tabela 2. Escala de avaliação baseada no MOS (adaptada de [14]).

A segunda pergunta, subjetiva, questiona sobre a opinião do usuário, o qual foi encorajado a dar sugestões, criticar e dar sua impressão pessoal sobre o sistema. Caso o participante não fosse capaz de responder de forma independente, o responsável foi solicitado a descrever a impressão pessoal à respeito de como o participante reagiu ao utilizar o sistema.

A distribuição das respostas para a questão objetiva entre pessoas sem deficiência (não PCD) e PCD é mostrada na Tabela 3. Dos 17 entrevistados, seis deram nota 4 (bom) para o sistema e 11 deram nota 5 (excelente), resultando em um escore geral de 4,647 pontos, o que prova que a proposta tem, de fato, um potencial para ajudar pessoas com limitações motoras dos membros superiores e movimentos do pescoço preservados.

Pontuação	Votos (não PCD)	Votos (PCD)	Total
5	6	5	11
4	4	2	6
3	0	0	0
2	0	0	0
1	0	0	0
MOS	4,6	4,7	4,647

Tabela 3. Resultado do questionário MOS.

Discussão

Os participantes que possuem deficiência mostraram-se ansiosos pela eventual disponibilização do protótipo como um produto para ter um módulo “particular em casa”, conforme relatado, e interessaram-se por custos e pela possibilidade de controlar diversos outros aparelhos, incluindo aparelhos de som, ares-condicionados, dentre outros. Os profissionais das casas de apoio, por outro lado, relataram o potencial inovador do protótipo, desde que as PCD se adequassem e acostumassem aos movimentos que devem ser executados, considerando detalhes como velocidade do movimento, postura, etc. Os responsáveis por alguns usuários também apontaram o benefício de não precisar procurar pelo controle remoto de cada aparelho, já que o sistema ficaria centralizado no ambiente. Além disso, a funcionalidade de não precisar pressionar qualquer botão no acionador externo surpreendeu positivamente os voluntários sem comprometimento total dos membros superiores, os quais aprovaram a “nova tecnologia” de apenas aproximar umas das mãos sobre o aparelho para ativar o sistema.

Por outro lado, o sistema também mostrou desvantagens. A limitação na execução de certos movimentos, especialmente o *roll*, frustrou tanto os pesquisadores quanto os usuários, os quais mostraram-se bastante chateados quando não conseguiram concluir a tarefa, o que acabava por fazê-los sentir culpa por tal. Um participante, inclusive, sugeriu que alguma característica facial, como uma piscada longa dos olhos, poderia ser utilizada no lugar do *roll*, movimento que o deixou mais desconfortável. Além disso, movimentos muito rápidos ou bruscos acabavam por tornar verdadeiras as condições definidas para aumentar a confiabilidade do sistema, pois dificultavam o cálculo do fluxo óptico, o que forçava o sistema a reportar uma situação de erro. Do mesmo modo, o fluxo calculado para movimentos muito lentos é tão baixo que não pode ser interpretado como angulação, pois teria a mesma magnitude de um ruído, o que influenciaria diretamente no desempenho do sistema. Em ambos os casos, os participantes foram instruídos a realizar movimentos mais lentos ou rápidos, até, finalmente, conseguirem executar a angulação da cabeça numa velocidade correta.

A iluminação não controlada no ambiente residencial de um dos voluntários também foi um problema, já que o algoritmo Viola-Jones teve dificuldades em identificar um rosto, o que gerou certo desconforto e impaciência. Alguns participantes também possuíam uma postura incorreta, de modo que o algoritmo de detecção facial não funcionava por conta de a posição inicial do rosto não ser exatamente frontal e ereta. Além disso, a ausência de um *feedback* sonoro foi sentida durante os movimentos *pitch* e *yaw*, quando o participante perdia o contato visual com a tela e não sabia, por consequência, se o comando havia sido corretamente enviado à TV.

Sobre os participantes portadores de paralisia cerebral, todas as suas formas de interação são, geralmente, bastante limitadas, o que torna a interface com aparelhos eletrônicos um desafio para a comunidade científica [19]. Além disso, as limitações de duas pessoas com o mesmo diagnóstico de paralisia cerebral normalmente não são parecidas, tampouco iguais, já que tal condição pode resultar em déficit cognitivo, perda significativa

da coordenação motora e problemas de grau variado na fala devido ao comprometimento do aparelho oromotor [20].

CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma proposta de sistema de controle remoto universal cujo método de entrada é baseado nos movimentos ou poses da cabeça do usuário. Baseado na opinião dos próprios usuários (tanto deficientes motores quanto sem deficiência alguma) durante os testes, pode-se concluir que o protótipo do sistema mostrou um potencial inovador para as PCD em termos de tecnologia assistiva, visto que, dentre os trabalhos correlatos encontrados na literatura, nenhum propôs o controle de aparelhos eletrônicos através de gestos da cabeça.

A experiência com as PCD, a reação e os comentários dos responsáveis, profissionais da área da saúde e demais usuários foram fundamentais na avaliação do sistema como um produto que pode ser disponibilizado e que muito provavelmente trará, de fato, mais conforto às pessoas com deficiência dos membros superiores com cognitivo e movimentos do pescoço preservados que se adaptarem ao seu uso. O contato com o público alvo também fomentou novas ideias e complementou as já existentes para o futuro do projeto.

Trabalhos Futuros

Apesar de o sistema ter sido configurado para controlar apenas uma marca de TV, o projeto foi concebido com base no modelo de controle remoto universal. Portanto, é fundamental que um dos próximos passos seja expandir o controle para outros equipamentos dentro do ambiente doméstico. Segundo [1], dispositivos como televisores, ares-condicionados, DVD *players*, *home theaters*, aparelhos de som, dentre outros, possuem três pares de funções semelhantes: i) ligar/desligar ou começar/terminar; ii) aumentar/diminuir ou subir/descer; e iii) avançar/recuar ou mudar de estado. Isso significa que é possível realizar os mesmos seis movimentos da cabeça para realizar funções semelhantes em dispositivos diferentes.

Além da acessibilidade, um segundo ponto chave no *design* do projeto é o baixo custo. Atualmente, um Arduino UNO é utilizado para mandar informação de controle para a TV. Isso acontece porque o acesso direto aos pinos de PWM do C.H.I.P. ainda não é bem documentado e possivelmente não atende à velocidade exigida pela frequência dos protocolos de comunicação entre controles e aparelhos. Nesse sentido, a meta é estudar a fundo a documentação sobre o mapeamento dos pinos do microcomputador para tentar utilizá-lo como única plataforma embarcada do sistema, eliminando o atual uso do Arduino. A biblioteca LIRC (<http://www.lirc.org>), que faz parte do repositório oficial do Linux, poderia ser, então, utilizada no próprio C.H.I.P. para substituir a *IRremote*.

Atualmente, o controle remoto funciona somente com o módulo de reconhecimento de gestos da cabeça como entrada única. A fim de flexibilizar a utilização e atender uma maior diversidade de problemas físicos, o módulo de reconhecimento de fala deve ser inserido para que o usuário possa acessar as mesmas funcionalidades do atual e dos futuros aparelhos através de diferentes comandos de voz. Além disso, o acionador

Dificuldades	Possíveis Soluções
Iluminação do ambiente, posição inicial da cabeça e distância do usuário para a câmera	Substituição do Viola-Jones por outro algoritmo de detecção facial ou treino de um classificador com imagens capturadas no próprio ambiente particular de utilização do sistema
Movimento <i>roll</i>	Interação por comandos de voz (reconhecimento de voz) ou por outro movimento motor que o usuário em questão consiga realizar corretamente
Ausência de <i>feedback</i> sonoro	Síntese de voz
Accionador externo para PCD com total incapacidade de movimentação dos membros superiores	Inicialização pela voz (<i>keyword spotting</i>)
Limitação na fala e nos movimentos do pescoço	Detecção da orientação do movimento dos braços (<i>background subtraction</i>) e interface cérebro-computador (BCI)

Tabela 4. Dificuldades encontradas, acompanhadas de suas respectivas possíveis soluções.

externo poderia ser melhor utilizado por pessoas com total incapacidade de movimentação dos membros superiores através da técnica de reconhecimento de palavras-chave (do inglês *keyword spotting*) [4], onde um comando de voz específico seria dado para inicializar o sistema.

O módulo de síntese de voz, após também inserido, seria útil na tarefa de prover *feedbacks* audíveis das ações executadas nos aparelhos. Tanto o reconhecimento quanto a síntese de fala poderiam ser úteis para que pessoas com deficiências visuais e audição preservada, por exemplo, pudessem interagir de forma mais natural com o sistema proposto.

Para pessoas que não possuem a fala e os movimentos do pescoço preservados, outros métodos de entrada podem ser implementados e integrados ao sistema de controle como formas alternativas ao reconhecimento de fala e de gestos da cabeça. A detecção da orientação do movimento dos braços (direita-esquerda, cima-baixo e vice-versa, por exemplo) poderia ser implementada também utilizando técnicas de computação visual, como a subtração de fundo (do inglês *background subtraction*) [7]. Outra alternativa seria o “reconhecimento do pensamento”, alcançado através da interface cérebro-computador (BCI, do inglês *brain-computer interface*) [14].

Sobre os testes, a dificuldade maior foi encontrar PCD que se adequassem ao perfil do público-alvo. A meta é formar parcerias com mais casas de apoio às PCD e encontrar mais voluntários com deficiências motoras dos membros superiores e limitações no uso de controles remotos convencionais, atendidos pelo módulo AGR; e visuais, sendo estes beneficiados pelo reconhecimento e síntese de voz. O *feedback* de usuários impactados de forma mais direta pela tecnologia assistiva, tanto em forma de críticas quanto em forma de sugestões e elogios, é essencial para a identificação de pontos fracos e posterior melhoria do sistema.

Através dos testes, foi percebido que o sistema apresenta dificuldades na etapa de detecção da face que, dependendo de fatores como iluminação, posição inicial da cabeça e distância entre o usuário e a câmera, não ocorre de maneira correta. Para tentar solucionar este problema, soluções alternativas ao algoritmo Viola-Jones precisam ser investigadas, sempre man-

tendo o foco no baixo custo computacional devido à utilização de um microcomputador embarcado.

Um resumo das dificuldades encontradas é mostrado, juntamente com suas respectivas, futuras e possíveis soluções, na Tabela 4. Em breve, espera-se que a solução proposta possa atender a uma gama mais diversa de limitações físicas, ao passo que o sistema torna-se, de fato, multimodal.

AGRADECIMENTOS

Este trabalho foi patrocinado pelo Conselho Nacional de Pesquisa (CNPq); pela Universidade Federal do Pará (UFPA), Brasil, Edital n° 03/2016 – PROPESP; e pelo Programa de Apoio à Pós-Graduação e à Pesquisa Científica e Tecnológica em Tecnologia Assistiva no Brasil – CAPES/PGPTA, Edital 59/2014. Os autores agradecem a todos os voluntários, especialmente aos do Abrigo Especial Calabriano e da Associação Paraense das Pessoas com Deficiência (APPD), bem como a todos os profissionais da área da saúde que os auxiliaram durante esta pesquisa.

REFERÊNCIAS

1. Kamran Ahsan, Sarwat Iqbal, Adnam Nadeem, and Muhammad Sarim. 2014. UniCon Remote Control Model – A Mobile System for Assistive Technology. *Research Journal of Recent Sciences* 3, 4 (2014), 95–102.
2. Frank Althoff, Rudi Lindl, and Leonhard Walchshäusl. 2005. Robust Multimodal Hand and Head Gesture Recognition for Controlling Automotive Infotainment Systems. *VDI-Tagung Der Fahrer im 21. Jahrhundert 1919* (2005), 187.
3. Euclides N. Arcoverde Neto, Rafael M. Duarte, Rafael M. Barreto, João Paulo Magalhães, Carlos C. M. Bastos, Tsang I. Ren, and George D. C. Cavalcanti. 2014. Enhanced Real-Time Head Pose Estimation System For Mobile Device. *Integrated Computer-Aided Engineering* 21, 3 (2014), 281–293.
4. Guoguo Chen, Carolina Parada, and Georg Heigold. 2014. Small-footprint Keyword Spotting Using Deep Neural Networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 4087–4091.

5. Trevor Darrell and Alex Pentland. 1996. Active Gesture Recognition using Learned Visual Attention. *Advances in Neural Information Processing Systems* (1996), 858–864.
6. Brasil. Subsecretaria Nacional de Promoção dos Direitos da Pessoa com Deficiência. Comitê de Ajudas Técnicas. 2009. Tecnologia Assistiva. *Brasília: CORDE* (2009), 138p.
7. B. Dhananjaya, B. Ramamurthy, and P. Thimmaiah. 2015. Moving Object Tracking with OpenCV on ARM Cortex-A8 in surveillance Applications. *International Journal of Current Engineering and Technology* 5, 2 (2015), 843–848.
8. Gorka Epelde, Xabier Valencia, Julio Abascal, Unai Díaz, Ingo Zinnikus, and Christian Husodo-Schulz. 2011. TV as a Human Interface for Ambient Intelligence Environments. In *2011 IEEE International Conference on Multimedia and Expo*. 1–6.
9. Fatih Erden and Ahmet E. Çetin. 2014. Hand Gesture Based Remote Control System Using Infrared Sensors and a Camera. *IEEE Transactions on Consumer Electronics* 60, 4 (2014), 675–680.
10. Paulo A. A. Esquef and Rogério Caetano. 2007. ROCC – A Webcam-Mouse for People with Disabilities. *Proc. XXV Simpósio Brasileiro de Telecomunicações* (2007).
11. Gert J. Gelderblom and Luc P. de Witte. 2002. The Assessment of Assistive Technology: Outcomes, Effects and Costs. *Technology and Disability* 14, 3 (2002), 91–94.
12. Satish Kumar, Dheeraj, Neeraj, and Sandeep Kumar. 2015. Design and Development of Head Motion Controlled Wheelchair. *International Journal of Advances in Engineering & Technology* 8, 5 (2015), 816–822.
13. Geonhee Lee, Keunhyun Kim, and Jaehyo Kim. 2016. Development of Hands-Free Wheelchair Device Based on Head Movement and Bio-Signal For Quadriplegic Patients. *International Journal of Precision Engineering and Manufacturing* 17, 3 (2016), 363–369.
14. Wei T. Lee, Humaira Nisar, Aamir S. Malik, and Kim H. Yeap. 2013. A Brain Computer Interface for Smart Home Control. In *2013 IEEE International Symposium on Consumer Electronics (ISCE)*. 35–36.
15. Pamela C. Levy, Nirvana S. Antonio, Thales R. B. Souza, Rogério Caetano, and Priscila G. Souza. 2013. ActiveIris: Uma Solução para Comunicação Alternativa e Autonomia de Pessoas com Deficiência Motora Severa. In *Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems (IHC '13)*. Brazilian Computer Society, 42–51.
16. Burce D. Lucas and Takeo Kanade. 1981. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2 (IJCAI'81)*. Morgan Kaufmann Publishers Inc., 674–679.
17. Yang-Yen Ou, Po-Yi Shih, Ta-Wen Kuan, Shao-Hsien Shih, Jhing-Fa Wang, and Jaw-Shyang Wu. 2014. A Happiness-Oriented Home Care System for Elderly Daily Living. In *2014 International Conference on Orange Technologies*. 193–196.
18. Aleksandar Pajkanović and Branko Dokić. 2013. Wheelchair Control by Head Motion. *Serbian Journal of electrical engineering* 10, 1 (2013), 135–151.
19. Himanshu Pokhariya, Parag Kulkarni, Vasudhara Kantroo, and Tanushree Jindal. 2006. Navigo–Accessibility Solutions for Cerebral Palsy Affected. In *2006 International Conference on Computational Intelligence for Modelling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce (CIMCA'06)*. 143–143.
20. Chitra Sankar and Nandini Mundkur. 2005. Cerebral Palsy–Definition, Classification, Etiology and Early Diagnosis. *The Indian Journal of Pediatrics* 72, 10 (2005), 865–868.
21. Utpal V. Solanki and Nilesh H. Desai. 2011. Hand Gesture Based Remote Control for Home Appliances: Handmote. In *2011 World Congress on Information and Communication Technologies*. 419–423.
22. Preeti Srivastava, S. Chatterjee, and Ritula Thakur. 2014. A Novel Head Gesture Recognition Based Control For Intelligent Wheelchairs. *International Journal of Research in Electrical & Electronics Engineering* 2, 2 (2014), 10–17.
23. The United Nations. 2006. Convention on the Rights of Persons with Disabilities. *Treaty Series* 2515 (2006).
24. Paul Viola and Michael Jones. 2001. Rapid Object Detection Using a Boosted Cascade of Simple Features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Vol. 1. 1–511–I–518 vol.1.
25. Paul Viola and Michael Jones. 2004. Robust Real-Time Face Detection. *International Journal of Computer Vision* 57, 2 (2004), 137–154.
26. Ina Wechsung and Anja B. Naumann. 2009. Evaluating a Multimodal Remote Control: The Interplay Between User Experience and Usability. In *Quality of Multimedia Experience, 2009. QoMEX 2009. International Workshop on*. 19–22.