

INTERAÇÃO NATURAL NA DOMÓTICA: CONTROLES E ACIONAMENTOS UTILIZANDO O KINECT E REDES NEUROFUZZY

*Rafael X. L., **R. C. Betini

*Departamento de Informática - UTFPR, Curitiba, Brasil

**Departamento de Eletrotécnica - UTFPR, Curitiba, Brasil

e-mail: rafael.xavier.lima@gmail.com, rcbetini@terra.com.br

Resumo: O princípio da domótica é facilitar as tarefas diárias comuns, como controle de temperatura, iluminação, portas, janelas, etc., situações que podem criar um amplo ambiente assistido, mas falha ao atender pessoas com mobilidade e coordenação avariadas. Para atender essa deficiência o uso dos conceitos da Interação Natural (NI) amplia a aplicabilidade e as possibilidades de um ambiente assistido. Para isso será utilizado o sensor kinect, que disponibiliza um grande conjunto de funcionalidades para a identificação e rastreamento de gestos e expressões faciais e, juntamente com o sistema de redes neurofuzzy, abranger e possibilitar o reconhecimento e interpretação de um conjunto mais diversificado e não lineares de gestos e poses. Desta forma espera-se possibilitar que usuários com dificuldades motoras nos membros superiores, mais especificamente portadores de necessidades especiais, possam alcançar uma maior independência nas atividades do dia-a-dia, pois sistemas baseados em gestos apresentaram uma grande aceitação do público em questão. Testes iniciais, que além de ter obtido uma boa aceitabilidade, apresentaram resultados promissores com uma precisão de reconhecimento de 96% para problemas motores brandos e 67% para problemas mais acentuados, demonstrando que ainda existe a necessidade de aprimoramentos que podem aumentar a taxa de precisão.

Palavras-chave: interação natural, domótica, kinect, redes neurofuzzy, reconhecimento de gestos e poses.

Abstract: *The domotic principle is to facilitate the common daily tasks like temperature control, lighting, doors, windows, etc., situations that can create a wide assisted ambient, but fails to attend disabled people. To attend this deficiency can be used the concepts of Natural Interaction (NI) that extends the applicability and possibilities of an assisted environment. To achieve this the kinect sensor must be used, which provides a great set of features for the identification and tracking of gestures and facial expressions, together with a neurofuzzy network, include and allow the recognition and interpretation of a more diverse set and non-linear gestures and poses. Thus it is expected to enable users with disabilities of the upper limbs, more specifically those with special needs, can achieve a greater independence in activities of day-to-day, for systems based on gestures presented a wide acceptability by the public in question. Initial tests, which besides having*

obtained a good acceptability, showed promising results with recognition accuracy of 96% for mild motor problems and 67% more pronounced problems, demonstrating that there is still need for improvements that can increase the accuracy rate.

Keywords: *natural interaction, domotic, kinect, neurofuzzy network, gestures and poses recognition.*

Introdução

Não é de hoje o conhecimento sobre certas dificuldades que existem para determinados grupos na sociedade, como por exemplo, idosos e principalmente portadores de necessidades especiais [1]. Tarefas simples do dia-a-dia como abrir uma janela, atender um telefonema, ligar ou desligar a luz, televisão, qualquer outro dispositivo ou até mesmo algo tão simples como uma descarga no banheiro tornam-se um enorme transtorno e causam uma grande frustração.

Em meados de 2010 foi criado um dispositivo composto por um conjunto de componentes de forma a torná-lo um sensor capaz de ampliar o conceito da acessibilidade e ajudar no desenvolvimento de aplicações assistivas, criando facilidades para melhorar a qualidade de vida das pessoas portadoras de necessidades especiais. Conhecido como kinect, esse sensor atingiu um novo nível de interação homem-máquina, empregando conceitos de Interação Natural (IN) no reconhecimento de gestos aboliu a necessidade de dispositivos junto ao usuário [2, 3,4].

Outros estudos que abordam o reconhecimento de gestos com o kinect se utilizaram de métodos que seguem um padrão limitado, utilizando painéis (monitores) com pontos fixos e gestos pré-determinados, falhando na premissa de reconhecer gestos ambíguos de pessoas com dificuldades motoras mais graves [12, 13]. Ou utilizam somente como apoio na identificação da intenção do portador, tendo como meio principal de interação outro meio, como a fala [14].

O trabalho proposto por Alvarenga, que se utiliza de redes neurais multicamadas para o reconhecimento de gestos capturados pelo kinect, mostrou ser possível identificar gestos básicos em tempo real com altas taxas de precisão com 92% de acerto [15]. O proposto no estudo deste artigo pretende ampliar o que foi apresentado por Alvarenga aumentando a aplicabilidade, a precisão na interpretação e identificação dos gestos

com a Rede Neurofuzzy por apresentaram ótimos resultados de fácil interpretação e, principalmente, por permitirem a representação de uma grande variedade de comportamentos dinâmicos não lineares, como movimentos imprecisos ou que podem gerar ambiguidade na interpretação [5,6].

Portanto, aplicando o reconhecimento de gestos utilizando o sensor kinect com rede neurofuzzy em um ambiente residencial automatizado pode proporcionar um aumento significativo na qualidade de vida de pessoas com dificuldades motoras, auxiliando nas tarefas do dia-a-dia de forma a aumentar a independência na realização de atividades comuns, como ligar um televisor, abrir uma porta ou uma janela, etc.

Interação Natural

Contida na área de Visão Computacional, o Reconhecimento de Gestos constitui-se de um conjunto de técnicas de processamento e análise sequencial e cronológica de imagens de forma a possibilitar uma interpretação computacional de um determinado gesto [7,8].

São interfaces com foco no uso progressivo da ação natural das pessoas, de forma a proporcionar uma integração cada vez melhor sem a exigência de qualquer conhecimento prévio ou treinamento.

Domótica

Termo criado com a junção da palavra latina “Domus” (casa) com a palavra “Robótica” (controle automatizado) onde o principal ponto é a automatização das rotinas e tarefas em um domicílio, como por exemplo: controles de temperatura, iluminação, fluxo de água e segurança [9].

Redes Neurofuzzy

Mesclam as vantagens da capacidade de aprendizagem das redes neurais com manipulação de informação linguística e numérica dos sistemas fuzzy superando as deficiências individuais. Em razão disto esses sistemas apresentam um ótimo desempenho em inúmeras aplicações em contrapartida a sistemas puros. As redes neurofuzzy recorrentes são mais complexas, difíceis e trabalhosas, mas o desempenho para identificar sistemas não lineares, estimar parâmetros, reconhecer padrões, controlar processos, aproximar funções, etc., tem apresentado resultados com taxas de acerto superiores em comparação com outros modelos de processamento de informação [10].

Por estas características citadas e também considerando o comparativo de resultados com outros métodos amplamente utilizados (Tabela 1), a estrutura foi escolhida como componente do estudo em questão.

Microsoft Kinect Sensor

Considerado a maior inovação no mundo do entretenimento. É utilizado principalmente em jogos, onde dispensa qualquer controle ou dispositivo de interação com o usuário. O sensor é composto basicamente por uma câmera RGB, um sensor de luz infravermelho e o CMOS (*Complementary Metal-Oxid Semiconductor*) que atuam juntos para determinar a profundidade, e uma fila de microfones para comandos e localização relativa do usuário.

Ferramenta de desenvolvimento e IDE – A Microsoft disponibiliza gratuitamente o suporte para o desenvolvimento de aplicativos via PC que utilizem o kinect, fornecendo um arcabouço com todas as funcionalidades disponíveis no sensor. Atualmente a SDK (*Standard Development Kit*) está na versão 1.8 e funciona somente com o IDE Microsoft Visual Studio 2010 ou 2012.

Portanto, devido ao imenso suporte da comunidade e do próprio fabricante, as altas taxas de precisão no rastreamento de gestos, assim como a facilidade de aquisição e baixo custo, o sensor foi escolhido para o estudo apresentado neste artigo.

Objetivo e Proposta

Esse trabalho propõe uma análise sobre identificação de gestos utilizando redes neurofuzzy e uma arquitetura para um sistema de automação residencial controlado por gestos utilizando o sensor kinect. Serão utilizados os reconhecimentos de acordo com a necessidade do usuário, isto é, utilizar os movimentos das mãos, dos braços, da face, ou até mesmo dos olhos, de forma simplificada, adaptável e amigável com foco no usuário cadeirante ou com limitações/dificuldades nos membros inferiores, de maneira a auxiliar nas atividades do dia-a-dia proporcionando uma melhor qualidade de vida. Inicialmente com a capacidade de pelo menos reconhecer seis gestos elementares (aumentar, diminuir, para esquerda, para direita, empurrar e puxar), de forma a torná-los comandos para os dispositivos.

A identificação de gestos e o uso de redes neurofuzzy recorrentes aplicados na domótica podem trazer uma grande melhora na qualidade de vida de pessoas portadoras de necessidades especiais, mais necessariamente cadeirantes, ao realizar atividades cotidianas de forma automática sem a necessidade de um controle ou dispositivos junto ao usuário. Devido a grande variedade de movimentos imprecisos que podem gerar ambiguidade na interpretação, foi elaborada uma estrutura mais robusta de rede neurofuzzy contendo recorrência local, que permite a representação de uma grande variedade desses comportamentos dinâmicos e não lineares dos membros superiores de um usuário capturados pelo sensor kinect.

Avaliação dos principais componentes de IA – o desempenho dos principais componentes de inteligência computacional: máquina de vetor (SVM - Support Vector Machine), função de base radial (RBF - Radial

Basis Function), redes neurais multicamadas (MLP - Multilayer Perceptron) e redes neurofuzzy adaptativas (ANFC-LH - Adaptive Neuro-Fuzzy Classifier with Linguistic Hedges), mostrou que a rede neurofuzzy atingiu a maior taxa de acerto no treinamento assim como nos testes com o segundo melhor tempo de processamento conforme apresentado na Tabela 1 [5].

Tabela 1: Taxa de acerto na distinção entre pacientes normais e portadores de Mal de Parkinson (variação de $\pm 2,1\%$).

Técnica	Treinamento	Testes	Tempo(s)
SVM	90,20%	91,40%	5,4
RBF	96,12%	95,23%	4,8
MLP	98,32%	96,54%	3,9
ANFC-LH	99,18%	98,54%	4,1

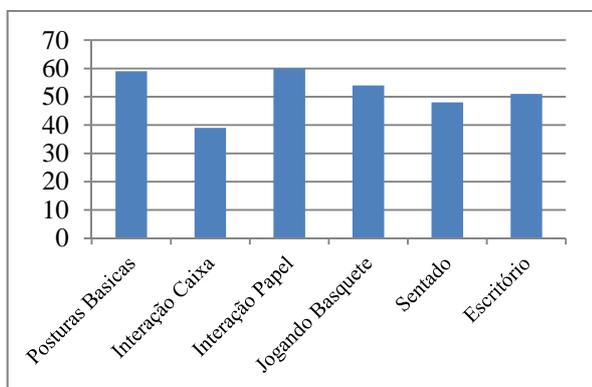


Figura 1: Taxa de acerto na reconstrução de posturas.

Precisão no rastreamento dos gestos – O kinect apresenta um bom nível de precisão no rastreamento dos membros do corpo humano. Nos testes realizados com 21590 posturas gravadas, a rastreada é comparada a um filtro que restringe para uma identificação entre apenas 2574 posturas. É mesmo em situações onde parte do corpo é ocultado por um anteparo, como uma caixa de papelão, a taxa permaneceu por volta dos 40% conforme apresentado na Figura 1 [11].

Conceitualização do protótipo – Idealização do protótipo do objeto de pesquisa.

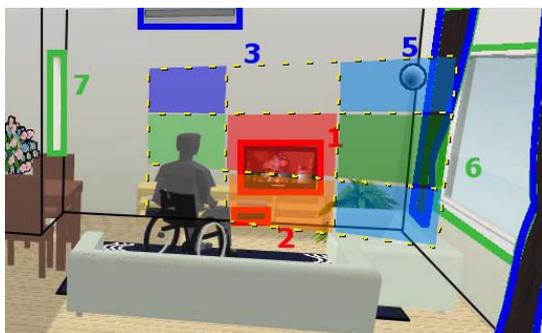


Figura 2: Ilustração dos itens controláveis mapeados em uma sala: TV, VCR, Ar cond., Cortinas, Janelas, Porta.

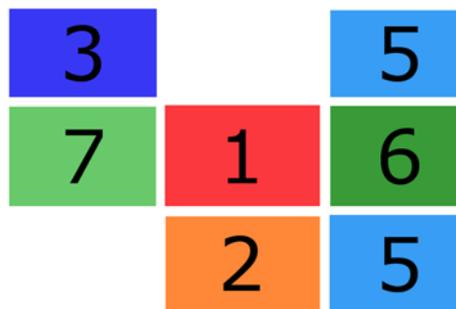


Figura 3: Limite de cada área acessível dos itens mapeados.

A Figura 2 ilustra um ambiente domotizado que tem seus principais itens mapeados para serem controlados por gestos. Na Figura 3 é possível notar melhor a área de limitação para identificar qual dispositivo o gesto irá controlar. A ativação de cada área dar-se-á no momento em que o usuário aponta com uma das mãos por 2s, o gesto seguinte é esperado um de controle por 5s (aumentar, diminuir, esquerda, direita, empurrar e puxar) caso existir. Por exemplo, apontar para a janela implica em abrir ou fechar, não necessita de controles adicionais, diferentemente da TV, que podem ser de aumentar/diminuir o volume ou mudar de canal (esquerda/direita). O gesto de controle independe da área limite ilustrada na Figura 3, que é aplicada somente na escolha do dispositivo.

Resultados

Observamos que ao proporcionar uma nova forma de interação e controle residencial houve uma maior aceitação do público devido à facilidade de uso.

Os testes iniciais tiveram como base um protótipo que se utilizava de uma interface dividida em cinco partes: um televisor ao centro, um vídeo cassete abaixo, um condicionador de ar acima, um abajur na lateral direita e um aparelho de som na lateral esquerda. O usuário deveria ligar e desligar cada um dos dispositivos com o gesto de empurrar com uma das mãos na direção do mesmo e mudar as características deslizando uma das mãos horizontalmente ou verticalmente.

Embora em testes iniciais, obtivemos resultados promissores. Em usuários com dificuldades motoras brandas, que não conseguem esticar e/ou levantar completamente os braços, foi possível reconhecer a intenção de 96% dos gestos com precisão, os 4% foram descartados equivocadamente pelo sistema, talvez indicando a necessidade de um aumento no período de treinamento da rede neurofuzzy desenvolvida. Já usuários com dificuldades um pouco mais acentuadas, problema similar ao mencionado anteriormente, mas com espasmos frequentes nos movimentos, a taxa de acerto caiu para 67%. Usuário comum, com mobilidade total, atingiu 98% de reconhecimento correto da intenção. Comparando com o resultado obtido por Alvarenga [15], que atingiu 92% de precisão em três movimentos (circular, chamar e adeus) utilizando apenas uma rede neural multicamada comum em

contrapartida com os 98% atingidos reconhecendo seis movimentos (aumentar, diminuir, para esquerda, para direita, empurrar e puxar) em uma rede neurofuzzy com recorrência local, embora mais trabalhosa, o ganho é nitidamente considerável. Mesmo para os 67% atingidos com gestos imprecisos e ambíguos, que pode ser sugerido como um caminho para aperfeiçoamento e trabalhos futuros, não foram encontrados relatos de estudos conforme o que foi realizado nesta pesquisa para comparar os achados.

Discussão

Durante os testes, mesmos com os usuários que apresentaram uma baixa taxa de acerto, foi percebido um grande entusiasmo ao interagir com o protótipo. Não foi passado qualquer treinamento ou instrução de como interagir com o ambiente, simplesmente ao realizar o ato de empurrar na direção da televisão e a mesma responder ligando-se foi suficiente para provocar a curiosidade do usuário, que ao levantar a mão observou que o volume aumentara, assim como deslizar para a direita fez com que mudasse de canal, levando-o a testar os mesmos princípios com os outros dispositivos que compunham o ambiente de teste representado.

Conclusão

Ainda em fase muito inicial, a rede precisa de um maior período de treinamento não supervisionado e, talvez, mais uma etapa de treinamento assistido ou por reforço para minimizar os equívocos e falsos positivos.

Os testes foram realizados considerando a captura ótima dos gestos, isto é, sem anteparos ou posições desfavoráveis ao rastreamento, portanto será necessário aplicar novamente os testes em um ambiente mais factível que irá gerar uma maior dificuldade ao reconhecimento dos gestos. Também existe a necessidade de diversificar o ambiente representado, adicionando banheiro, cozinha e quarto por exemplo.

Os gestos com os braços foram testados considerando os três tipos de usuários mencionados nos resultados. Já para usuários com necessidades mais graves e complexas seria necessário o segundo modo de rastreamento mencionado, o da face, que ainda está em fase de aperfeiçoamento e não entrou em fase de testes com usuários diversificados.

Referências

- [1] Galvão Filho, T. A. (2009). A tecnologia assistiva: de que se trata. Conexões: educação, comunicação, inclusão e interculturalidade. Porto Alegre: Redes, pages 207–235.
- [2] Garber, L. (2013). Gestural technology: Moving interfaces in a new direction [technology news]. *Computer*, 46(10):22–25.
- [3] Metcalf, C., Robinson, R., Malpass, A., Bogle, T., Dell, T., Harris, C., and Demain, S. (2013). Markerless motion capture and measurement of hand kinematics: Validation and application to home-based upper limb rehabilitation. *Biomedical Engineering, IEEE Transactions on*, 60(8):2184–2192.
- [4] Ren, G., Li, C., O'Neill, E., and Willis, P. (2013). 3d freehand gestural navigation for interactive public displays. *Computer Graphics and Applications, IEEE*, 33(2):47–55.
- [5] Geman, O. (2013). Nonlinear dynamics, artificial neural networks and neuro-fuzzy classifier for automatic assessing of tremor severity. In *E-Health and Bioengineering Conference (EHB)*, 2013, pages 1–4.
- [6] Hafiz, A., Amin, M., and Murase, K. (2011). Real-time hand gesture recognition using complex-valued neural network (cvnn). In Lu, B.-L., Zhang, L., and Kwok, J., editors, *Neural Information Processing*, volume 7062 of *Lecture Notes in Computer Science*, pages 541–549. Springer Berlin Heidelberg.
- [7] Karam, M. and Schraefel, M. C. (2005). A taxonomy of gestures in human computer interactions. Technical report, University of Southampton.
- [8] Deshayes, R. (2013). A domain-specific modeling approach for gestural interaction. In *Visual Languages and Human-Centric Computing (VL/HCC)*, 2013 IEEE Symposium on, pages 181–182.
- [9] Gerhart, J. (1999). Home automation and wiring. McGraw-hill, United States, New York.
- [10] Huamaní, I. R. L. (2003). Redes neurais fuzzy aplicadas em identificação e controle de sistemas. Tese de Mestrado, Universidade Estadual de Campinas.
- [11] Shum, H., Ho, E., Jiang, Y., and Takagi, S. (2013). Real-time posture reconstruction for microsoft kinect. *Cybernetics, IEEE Transactions on*, 43(5):1357–1369.
- [12] Posada-Gomez, R., Rodriguez-Bernardo, C. O., Luna-Bravo, P. S., Alor-Hernandez, G., Martinez-Sibaja, A., & González, A. R. (2012). Development of a natural interaction interface for people with disabilities in a home automation control room. In *Intelligent Environments (Workshops)* (pp. 353-361).
- [13] Wu, C.-H. and Lin, C. H. (2013). Depth-based hand gesture recognition for home appliance control. In *Consumer Electronics (ISCE)*, 2013 IEEE 17th International Symposium on, pages 279–280.
- [14] Dimitra Anastasiou, Cui Jian, and Desislava Zhekova. 2012. Speech and gesture interaction in an Ambient assisted living lab. In *Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments (SMIAE '12)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 18-27.
- [15] Alvarenga, M. L. T., Correa, D. S. O., and Osório, F. S. (2012). Redes neurais artificiais aplicadas no reconhecimento de gestos usando o kinect. *Universidade de São Paulo*.