

USO DO KINECT V2 PARA O RECONHECIMENTO DE GESTOS DA LIBRAS

LEROI FLORIANO DE OLIVEIRA¹; MARILTON SANCHONETE DE AGUIAR²; CARLA SIMONE GUEDES PIRES³

¹*Universidade Federal de Pelotas - leroioliveira@gmail.com*

²*Universidade Federal de Pelotas - marilton@inf.ufpel.edu.br*

³*Instituto Federal Sul-Rio-Grandense - carlasmpires@gmail.com*

1. INTRODUÇÃO

Nas últimas décadas tem se observado um grande avanço no que diz respeito a acessibilidade através de tecnologias de informação e comunicação (TIC). A evolução tecnológica vem caminhando no sentido de facilitar da vida das pessoas, trazendo recursos especificamente desenvolvidos para auxiliar e simplificar os dinamismos do cotidiano. A interação humano-computador (IHC) evoluiu muito nas últimas décadas, acompanhando estes avanços tecnológicos. Atualmente vivemos uma nova perspectiva no que se refere a IHC, com a ideia de apoiar o estudo de interfaces adaptativas e adaptáveis, buscando sempre melhorias nas formas de interação.

Neste sentido, é muito comum o emprego de recursos de acessibilidade incorporadas aos mais diversos componentes de um sistemas, promovendo maior interatividade. Exemplos de recursos desta natureza podem ser identificados na utilização de reconhecimento de voz, que permite o envio de comandos ao computador através da fala, contribuindo com a flexibilização das formas de interação (Barbosa 2010). Outra forma notável de interação é a utilização de sensores de movimento aplicados ao reconhecimento de gestos, viabilizando que movimentos do corpo humano sejam interpretados pelo computador, que responde executando comandos no sistema.

Uma aplicação destas interações é o reconhecimento de gestos e inclusive a interpretação de linguagens de sinais. Atualmente muitos estudos vem sendo desenvolvido com este propósito, como pode ser visto em Pugeault & Bowden (2011) que utiliza o Kinect v1 para captura de gestos e como método de classificação utiliza multiclasse de floresta randômica (*multi-class random forest*) e tem como resultado o reconhecimento de 73% dos gestos capturados.

Elons et al. (2012) utiliza câmeras para o reconhecimento da Língua de Sinais Árabe, e obtiveram em seu melhor resultado 96% de reconhecimento utilizando o método de coleção de sistemas de inferência neuro-fuzzi adaptativa (Collection of adaptive neuro-fuzzy inference system (ANFIS)). Outros estudos apontam também bons resultados com o reconhecimento da LIBRAS conforme observamos no trabalho de Almeida et al. (2014) que utilizou o Kinect e também o software “nuiCapture Analyze” para captura e o reconhecimento através do método “Memetic Self-Adaptive Evolution Strategies” o qual registrou diferentes resultados para cada sinal trabalhado no reconhecimento, chegando a 100% no melhor dos casos e a 1% no pior dos casos tendo como resultado médio aproximadamente 80% de reconhecimento dos sinais catalogados.

Este projeto tem como objetivo aprofundar estudos na área de reconhecimento de gestos aplicados a interpretação da Linguagem Brasileira de Sinais (LIBRAS) através da utilização do sensor Kinect v2 (Microsoft, 2015).

2. METODOLOGIA

Para atingir os objetivos propostos serão conduzidas as seguintes atividades:

1. Estudo dos diferentes tipos de sensores presentes no Kinect versão 2 e sua aplicação para o reconhecimento de gestos.
2. Estudo da API do Kinect versão 2 para o reconhecimento de gestos.
3. Estudo da Língua Brasileira de Sinais (LIBRAS), identificando os padrões entre os gestos de forma a conseguir criar um modelo eficiente para realizar a posterior distinção entre os métodos.
4. Estudar Algoritmos de aprendizagem de máquina e aplicações anteriores em reconhecimento de gestos.
5. Criar uma base de dados com gestos coletados pelo Kinect, baseado no modelo definido para realizar o treinamento de algoritmos de aprendizado de máquina.
6. Realizar o treinamento da Inteligência artificial com a base de dados utilizando a ferramenta WEKA.
7. Avaliação dos resultados obtidos no reconhecimento de gestos capturados em pessoas de diferentes características físicas, observar a taxa de reconhecimento para novos gestos utilizando

o mesmo modelo e realizar uma comparação entre diferentes algoritmos de aprendizagem de máquina.

3. RESULTADOS E DISCUSSÃO

Como resultado deste trabalho, foi realizada uma coleta dos dados gerados através da captura de gestos utilizando o sensor kinect v2. Com uma pessoa em frente ao sensor realizando gestos da LIBRAS, o sensor realizou o mapeamento de 25 pontos ao longo do corpo da pessoa conforme a figura 1. Com média de 30 coletas dos pontos por segundo representando cada frame do movimento, foi trabalhando a identificação dos gestos.



Figura 1 - Mapeamento do corpo com as visões de Color, Infrared, Body, e BodyIndex no Microsoft Kinect Studio™

Para a catalogação foram trabalhados apenas gestos estáticos, que são gestos realizados utilizando apenas a região das mãos e sem movimentos ao longo de sua execução.

Utilizando do software WEKA para treinamento de diferentes métodos de aprendizagem de máquina mostrou-se imcompatível para a utilização de dados brutos, por representarem uma grande quantidade de dados. Foi necessário um tratamento dos dados para que fosse possível a utilização no WEKA. Em primeiro momento foram eliminados ruídos de leitura do sensor, removendo pontos muito distantes do valor médio do conjunto de pontos.

Outro tratamento foi identificar o valor médio de cada um dos pontos após a remoção do ruído.

Como resultado observou-se que tal método não foi suficiente para uma identificação clara para gestos estáticos, sendo necessário a exploração dos demais sensores do kinect.

4. CONCLUSÕES

Observou-se que os dados brutos coletados foram de insuficientes para o treinamento e posterior reconhecimento utilizando o software WEKA. Já com um tratamento simples dos dados foi identificado algum resultado e possibilidades de trabalho, apesar da baixa taxa de acerto para gestos estáticos. Identificou-se que a baixa taxa de acerto foi em função dos poucos pontos na mão mapeados pelo sensor kinect. Sendo assim concluiu-se que vale a pena a exploração dos demais sensores, para tentar melhorar a identificação de gestos estáticos e da região da mão.

5. REFERENCIAS BIBLIOGRÁFICAS

Barbosa, S.D.J.; Silva, B.S. **Interação Humano-Computador**. Série SBC, Editora Campus-Elsevier, 2010.

Valente, F.; Rodrigues, C.S. Aspectos Linguísticos Das Libras. IESDE, 2011.

Microsoft, <http://www.microsoft.com/en-us/kinectforwindows/> 01 de janeiro de 2015.

Microsoft, <http://www.microsoft.com/en-us/download/details.aspx?id=44561> 02 de janeiro de 2015.

Pugeault, N.; Bowden, R. Spelling It Out: Real-Time ASL Fingerspelling Recognition. IEEE International Conference on Computer Vision Workshops, 2011.

Elons, A. S.; Abull-elab, M.; Tolbaa, M.F. A proposed PCNN features quality optimization technique for pose-invariant 3D Arabic sign language recognition, Applied Soft Computing 13 (2013) 1646–1660

Almeida, S. G. M.; Guimarães, F. G.; Ramírez, J. A.; Feature extraction in Brazilian Sign Language Recognition based on phonological structure and using RGB-D sensors, Expert Systems with Applications, Volume 41, Issue 16, 15 November 2014