UM ESTUDO COMPARATIVO ENTRE TÉCNICAS DE RECONHECIMENTO DE GESTOS DA LIBRAS UTILIZANDO DADOS CAPTURADOS PELO MICROSOFT KINECT™

INESSA DINIZ LUERCE[1]; LUCAS MENDES TORTELLI[1], PLÍNIO FINKENAUER JÚNIOR[1], MARILTON SANCHOTENE DE AGUIAR[1]

[1]*Universidade Federal de Pelotas - { idluerce, lmtortelli, pfinkenauer, marilton } @inf.ufpel.edu.br*

1. INTRODUÇÃO

No Brasil, quando uma criança surda entra na escola o primeiro idioma que lhe é ensinado é o da Língua Brasileira de Sinais (Libras). Só então são ensinadas outras disciplinas como Matemática, História, Geografia, Ciências e o próprio Português, o qual é considerado uma segunda língua para os surdos. Visto que a grande maioria dos materiais didáticos disponibilizados no Brasil são em Português, um aluno surdo acaba enfrentando ainda mais dificuldades de compreensão que um aluno sem problemas auditivos.

Embora exista incentivo do Governo Brasileiro para o desenvolvimento de aplicações para inclusão social, as ferramentas disponíveis atualmente cobrem apenas uma pequena parte do público. Neste contexto, as aplicações que mais se destacam são os aplicativos para dispositivos móveis HandTalk e ProDeaf (OSSADA; RODRIGUES, 2015). Entretanto, estas aplicações restringem-se a dicionários e tradutores de Português para Libras, onde o usuário entra com as palavras em Português e um avatar exibe a representação destas em Libras. Embora úteis para ouvintes que gostariam de comunicar-se com surdos, ou até mesmo interessados em aprender Libras, ainda carecem de recursos no que diz respeito ao aprendizado completo da Língua.

Um dos motivos que justifica a abordagem destas aplicações está relacionado à alta complexidade para o desenvolvimento de ferramentas eficientes que façam o caminho contrário, ou seja, traduzam Libras para Português. Isso se deve ao fato de necessitar de aplicações intermediárias que consigam capturar os gestos do usuário, analisá-los, classificá-los e determinar, em tempo real, com a qual palavra aquele gesto está relacionado.

O propósito deste trabalho consiste em avaliar diferentes técnicas de inteligência artificial para o reconhecimento de gestos na Libras, a fim de servir de suporte para o desenvolvimento de aplicações que possam auxiliar o aprendizado da Língua de Sinais.

2. METODOLOGIA

Inicialmente, foi realizada uma pesquisa bibliográfica sobre a Língua Brasileira de Sinais (PEIXOTO, 2006) e sobre as técnicas de Inteligência Artificial mais utilizadas e eficientes para reconhecimento de gestos. Com base nesta pesquisa, foi dado enfoque nos métodos *Support Vector Machine, Decision Tree* e *Stochastic Gradient Descent* (RUSSELL et al., 2003).

Máquinas de Vetores de Suporte (do inglês *Support Vector Machine* - SVM*)* é uma técnica de aprendizado supervisionado que analisa dados em busca de padrões. Ela possui fortes bases teóricas e tem sido aplicada a tarefas como

reconhecimento de manuscritos, reconhecimento de objetos e classificação de texto (TONG; CHANG, 2001).

Uma Árvore de Decisão (do inglês, *Decision Tree* - DT) é um algoritmo de aprendizado supervisionado que possui a capacidade de lidar com problemas complexos de decisão, particionando-os em uma coleção de decisões simples. Isso é feito em programas nos quais as decisões são sequenciais e indeterminadas. Estas árvores descrevem graficamente todas as decisões a serem tomadas, os eventos que podem ocorrer e através dos caminhos gerados pela execução da árvore, os resultados da combinação dos eventos e decisões. (WITTEN; FRANK, 2005).

O *Stochastic Gradient Descent - SGD*, consiste de uma aproximação estocástica do método de *Gradient Descent*, técnica de propósito-geral para otimização que pode ser aplicada em funções diferenciáveis. É utilizado para minimização de funções e determinação de pesos em um Rede Neural (WITTEN; FRANK, 2005).

Foi realizado também um estudo aprofundado sobre o Microsoft Kinect™, sensor inicialmente desenvolvido para a plataforma de jogos Xbox e posteriormente utilizado também para aplicações científicas. A escolha deste equipamento deu-se ao fato de ser um hardware de baixo custo em relação a suas funcionalidades e por ter sido verificado na bibliografia que sua utilização possui bons resultados para captura de gestos relativos a linguagens gestuais (LUN; ZHAO, 2015).

Foi então definida a base de dados para utilização nos testes. Esta consiste dos primeiros dez números cardinais, 0 a 9, em Libras. A escolha dessa base deu-se primeiramente por todos os gestos serem estáticos, ou seja, são independentes de movimentos anteriores e um instante de tempo é o suficiente para representar cada gesto por completo. Outra motivação para a escolha dos números deu-se ao fato de que uma vez definida uma técnica de reconhecimento suficientemente satisfatória, a implementação pode servir de suporte para o desenvolvimento de um software que auxilie o ensino-aprendizagem da disciplina de matemática.

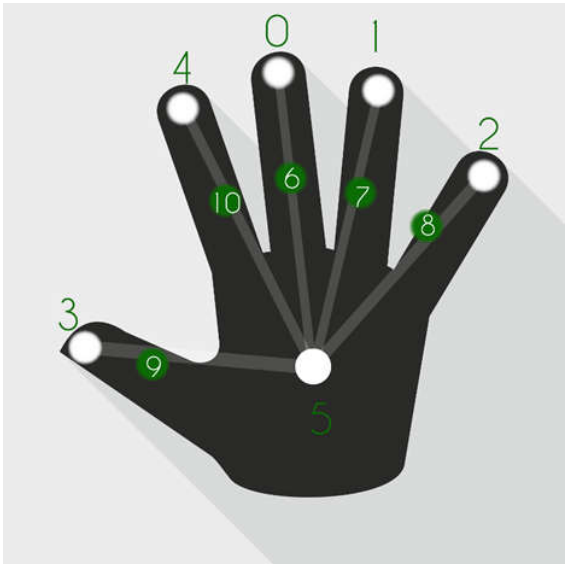

Figura 1: Pontos da mão extraídos através do Microsoft Kinect™.

Os dados contidos nesta base consistem de 180 gestos, obtidos através das coordenadas geográficas de dez pontos da mão, extraídos do Microsoft Kinect™

através de uma aplicação desenvolvida e disponibilizada como código aberto[1]. Para a implementação dos algoritmos, foi utilizada a linguagem Python juntamente com a biblioteca Scikit-Learn (PEDREGOSA, 2011).

Uma vez definida a base de dados e a implementação, foram definidas as especificações sobre os parâmetros a serem avaliados nos testes a fim de verificar qual técnica obteve melhor desempenho utilizando-se dos pontos fornecidos pelo Kinect™.

3. RESULTADOS E DISCUSSÃO

De acordo com (WITTEN; FRANK, 2005), ao utilizar técnicas de Aprendizado de Máquina, para testar o reconhecimento de dados, deve-se sempre utilizar uma base para treinamento da rede e outra para validação. Neste trabalho, todas as técnicas foram avaliadas com as seguintes relações de treinamento e classificação: 70/30, 75/25, 80/20 e 85/15, respectivamente.

O resultado da avaliação das técnicas para reconhecimento das posições dos dedos utilizando coordenadas está representado na Tabela 1.

Tabela 1: Descrição dos resultados obtidos para cada técnica por porcentagem de treinamento/classificação.

Técnica Utilizada	**Relação Treinamento vs. Classificação**							
	70/30		**75/25**		**80/20**		**85/15**	
	Média	**DP**	**Média**	**DP**	**Média**	**DP**	**Média**	**DP**
SVM One-vs-One	64,93	4,67	67,53	3,54	66,00	5,82	68,53	7,30
SVM One-vs-Rest	45,40	7,82	49,60	11,17	45,60	7,10	48,53	11,25
DT	54,46	7,76	56,40	9,87	56,27	4,59	56,67	10,51
SGD	37,87	9,80	39,73	6,36	37,33	9,80	35,80	12,14

Nota: DP = desvio padrão.

Observa-se que, em todas as variantes de conjuntos para treinamento e validação, a técnica *Support Vector Machine One-vs-One* obteve os melhores resultados na classificação dos gestos. A Tabela 2 apresenta a descrição geral para cada técnica utilizada. Diferenças significativas foram encontradas na comparação das médias das quatro técnicas. De acordo com o teste de comparação das médias, os resultados obtidos pela técnica *Support Vector Machine One-vs-One* para a classificação dos gestos, foram superiores aos demais métodos. Quanto à variabilidade, o desempenho foi consideravelmente mais homogêneo na técnica *Support Vector Machine One-vs-One* quando comparado às outras técnicas.

Tabela 2: Descrição do desempenho geral das técnicas utilizadas.

Técnica Utilizada	**Média**	**DP**	**Min**	**Q1**	**Md**	**Q3**	**Max**
SVM One-vs-One	66,75 *a*	5,54	58,00	64,00	67,00	69,50	89,00
SVM One-vs-Rest	47,28 *b*	9,45	22,00	42,00	47,50	55,00	66,00
DT	55,95 *c*	8,33	33,00	52,00	55,00	60,00	81,00
SGD	37,68 *d*	9,60	16,00	29,50	39,00	44,00	60,00

Nota: DP = desvio padrão; Min = menor valor; Q1 = primeiro quartil; Md = mediana, Q3 = terceiro quartil; Max = maior valor. Médias seguidas mesmas letras não diferem entre si pelo teste de de Tukey (α = 0,05).

[1] https://github.com/inessadl/kinect-2-libras

4. CONCLUSÕES

Este trabalho serviu como uma pesquisa exploratória para a buscar a melhor técnica para o reconhecimento de gestos de Libras utilizando pontos da mão extraídos através do Kinect™. Verificou-se que, para a base de dados utilizada, os resultados mais satisfatórios foram utilizando *Support Vector Machine One-vs.One*.

Para trabalhos futuros, pretende-se ampliar a base de dados, capturando juntamente com os pontos, uma imagem que represente o gesto no mesmo instante de tempo que são capturados os pontos. Esta imagem possui o intuito de ser utilizada para classificação utilizando Redes Neurais Convolucionais (CNN). Pretende-se fazer um paralelo entre os resultados obtidos com a técnica mais satisfatória encontrada neste trabalho e CNN. Por fim, uma vez testados e validados, a base de dados será disponibilizada publicamente.

5. REFERÊNCIAS BIBLIOGRÁFICAS

LUN, R.; ZHAO, W. A survey of applications and human motion recognition with microsoft kinect. **International Journal of Pattern Recognition and Artificial Intelligence**, v. 29, n. 05, p. 1555008, 2015.

OSSADA, S. A. R.; RODRIGUES, S. C. M. Uma análise de softwares para inclusão de deficientes auditivos na educação à distância. **Reverte - Revista de Estudos e Reflexões Tecnológicas da Faculdade de Indaiatuba,** n. 13, 2015.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, n. Oct, p. 2825-2830, 2011.

PEIXOTO, R. C. Algumas considerações sobre a interface entre a língua brasileira de sinais (Libras) e a língua portuguesa na construção inicial da escrita pela criança surda. **Cad Cedes**, v. 26, n. 69, p. 205-29, 2006.

RUSSELL, S. J. et al. **Artificial intelligence: a modern approach**. Upper Saddle River: Prentice hall, 2003.

TONG, S.; CHANG, E. Support vector machine active learning for image retrieval. **ACM International Conference on Multimedia**, 2001.

WITTEN, I. H.; FRANK, E. **Data Mining: Pratical machine learning tools and techniques.** Morgan Kaufmann, 2005.