

Reconhecimento das configurações de mão de LIBRAS baseado na análise de discriminante de Fisher bidimensional, utilizando imagens de profundidade

Jonilson R. Santos, Marly G. F. Costa, Cícero F. F. Costa Filho

Centro de Pesquisa e Desenvolvimento de Tecnologia Eletrônica e da Informação –
Universidade Federal do Amazonas (CETELI/UFAM)
CEP 69077-000 – Manaus – AM – Brasil

joniroque@hotmail.com, {mcosta,ccosta}@ufam.edu.br

***Abstract.** Deaf people communicate using sign language; however, they are just able to communicate with others who have this same knowledge. This communication is limited to other people with knowledge of sign language that, usually, are other deaf people. There are too many people that interact with deaf people in education, health and leisure areas that do not know about sign language. Then the inclusion of deaf people is seriously affected, because they are unable to make themselves understood. This study presents a methodology for automatic gesture recognition; the gestures represent the settings of hands from LIBRAS (Brazilian Language of Signs). The approach consist of constructing an image database by Kinect® sensor. In such images, we applied 2D²LDA technique to reduce their dimension and create new characteristics for classification step. The system is able to segment the hand image and recognize whole 61 settings of Sign Language. The average achieved hit rate was 95.7%. As the capture device is insensitive to light, background and colors of clothes and skin, there are no restrictions about environment.*

***Resumo.** As pessoas surdas comunicam-se com outras pessoas por meio da Língua de Sinais. O fato de existirem muitas pessoas nas principais áreas da sociedade (saúde, educação e lazer) que interagem com os surdos, com pouco ou nenhuma familiaridade com uma língua de sinais, afeta sobremaneira a inclusão social dos mesmos. Este artigo apresenta uma metodologia para o reconhecimento automatizado dos gestos que representam as configurações de mãos da Língua Brasileira de Sinais - LIBRAS. A abordagem inicial consistiu da construção de um banco de imagens de profundidade adquiridas com o sensor Kinect®. Nessas imagens, foi aplicado a técnica 2D²LDA para a redução de dimensionalidade do conjunto de características usado para classificação. O sistema é capaz de segmentar a mão e reconhecer as 61 configurações de mão da LIBRAS. A taxa média de acerto alcançada foi de 95,70%. Como o dispositivo de captura é insensível a luminosidade, fundo e cores das roupas e da pele, a aplicação desenvolvida adapta-se sem modificações a qualquer ambiente de captura das configurações*

1. Introdução

A comunicação entre as pessoas é efetuada de diversas formas: oral, escrita ou gestual. A língua de sinais é uma forma gestual que possibilita as pessoas surdas interagirem com outras pessoas. Assim, em vez de transmitir suas ideias acusticamente, eles fazem uso de sinais. Como diversas pessoas que interagem com os surdos em diferentes áreas da sociedade, como saúde, educação e lazer, não possuem o domínio da referida língua, a inclusão social do surdo é afetada negativamente, pois o mesmo não é capaz de fazer-se entender.

Em línguas de sinais, “fonemas” se referem às suas unidades espaciais que funcionam igualmente aos fonemas das línguas orais. As unidades mínimas distintivas em LIBRAS são: Configuração de Mãos (CM), Ponto de Articulação, Movimento-Orientação e Expressão Facial. A Língua Brasileira de Sinais, conta com 61 possíveis CM, as quais são apresentadas na Figura 1 (Pimenta and De Quadros 2010).



Figura 1- Configurações de mão da LIBRAS [Pimenta and De Quadros,2010]

Para atender a demanda de comunicação dos surdos com a sociedade, foram criados sistemas automáticos que têm o intuito de reconhecer os gestos da língua de Sinais. A maioria desses sistemas é baseada em câmeras sensíveis a luz visível e, portanto, sensíveis às condições de luminosidade. Diversas técnicas destinadas ao reconhecimento de gestos da língua de sinais em imagens adquiridas por meio de câmeras digitais têm sido reportadas na literatura (Ribeiro e Gonzaga, 2006; Bragatto *et al.*, 2006; Carneiro *et al.*, 2009; Maraça *et al.*, 2012). Técnicas que usam outros dispositivos de captura sensíveis a profundidade, como por exemplo o *Kinect*[®], são relatadas por Rakun *et al.* (2013), Chao *et al.* (2013) e Porfirio *et al.* (2013). Nas abordagens que empregam câmeras digitais a detecção da mão não é uma tarefa trivial. Carneiro *et al.* (2009) montou um banco de imagens de 6 pessoas num ambiente de fundo e iluminação controlados. Para segmentação da mão aplicaram técnicas de limiar simples, no espaço de cor YCbCr (Luminância, Crominância Azul, Crominância Vermelha). Para segmentação da mão, Ribeiro e Gonzaga (2006) usaram os processos de eliminação do fundo com misturas de Gaussianas e limiar de cor de pele. No pós-processamento utilizaram filtros morfológicos para recuperar as falhas oriundas da segmentação. Em contrapartida, Bragatto *et al.* (2006) e Maraça *et al.* (2012) usaram luvas coloridas que facilitam sobremaneira a segmentação da mão. Além disso, Bragatto *et al.* (2006), na classificação das configurações da mão - CM, extraiu quatro características de cada dedo, provenientes da análise de componentes principais, descritas em Lamar *et al.* (1999). Numa abordagem que utiliza câmera de profundidade, Rakun *et al.* (2013) utiliza o sensor *Kinect*[®]. As seguintes características são extraídas pelos autores para classificação dos gestos: centroide, eixos maior e menor, orientação e menor polígono convexo da imagem segmentada. Por sua vez, Chao *et al.* (2013) utilizaram o Histograma das Orientações do

Gradiente (HOG), pose do corpo, forma e movimentação da mão para a classificação de configurações de mão em vídeo. Para classificação das CM utilizou-se Máquinas de Vetores de Suporte Latente. O método desenvolvido por esses autores é capaz de encontrar quadros discriminativos e representativos em cada conjunto de vídeo.

Porfirio *et al.* (2013) faz o reconhecimento das 61 CM da Língua Brasileira de Sinais - LIBRAS. Na aquisição das imagens, cinco pessoas pousaram para a câmera realizando uma determinada sequência de movimentos. Seleccionando, manualmente, a visão frontal e lateral do gesto constrói-se uma malha 3D de cada CM. Ao final, 610 malhas 3D são reconstruídas. A segmentação da mão foi feita manualmente no *software Gimp*. Na extração de características de cada malha aplicaram-se descritores harmônicos esféricos, que são insensíveis a translação, rotação e escala. Para classificação das configurações são usadas máquinas de vetores de suporte (*Support Vector Machines - SVM*) com *kernel* de funções de base radial (RBF) e *kernel linear*. Os melhores resultados alcançados foram 96,67% para o *kernel* RBF e 96,83% para o *kernel linear*.

Este trabalho apresenta uma contribuição para o reconhecimento automatizado da LIBRAS, centrando inicialmente a sua atenção no reconhecimento de um dos seus “fonemas”, a configuração da mão (CM). Em trabalhos futuros pretende-se reconhecer todos os demais “fonemas”. Esse reconhecimento é feito utilizando-se imagens de profundidade adquiridas pelo sensor *Kinect*[®]. Para obtenção das variáveis de entrada do classificador aplicou-se a técnica 2D²LDA. Através dessa técnica, reduziu-se as dimensões da imagem original para 5x5 pixels, 10x10 pixels, 15x15 pixels e 20x20 pixels, respectivamente. Como classificador, empregou-se o *k-vizinhos mais próximos* (KNN - *K-Nearest Neighbor*).

2. Materiais e Métodos

Nesse trabalho, conforme mostrado na Figura 2, são implementadas todas as fases de um sistema de reconhecimento de padrões. A implantação desse sistema foi realizada no ambiente de simulação Matlab2013[®]. A seguir descrevemos cada uma dessas etapas.

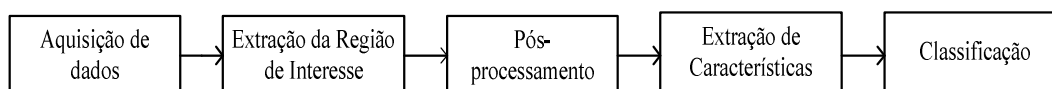


Figura 2- Diagrama em blocos do sistema de visão computacional implementado

2.1. Aquisição de dados

Nesta etapa é utilizado o *Kinect*[®] para obter imagens de profundidade, com resolução de 640x480 pixels. A distância da câmera para o usuário é padronizada em 1,40 metros. O banco de imagens é constituído de 12.200 imagens: 10 indivíduos, 61 configurações de mãos, 20 imagens por configuração, por indivíduo. Dentre os 10 indivíduos, dois são mulheres e oito são homens, sendo que sete indivíduos possuem LIBRAS como primeira língua. Procurou-se seleccionar indivíduos com mãos de diferentes tamanhos: pequenas médias e grandes. Os usuários foram convidados a rotacionarem e transladarem as suas mãos lentamente, num intervalo entre 45° e 135°. As 20 imagens por indivíduo e por configuração foram capturadas durante essa movimentação.

2.2. Extração da Região de interesse

2.2.1. Segmentação da mão

Para segmentação da mão foi utilizado a técnica de agrupamento *k-means*, tendo como variável independente a profundidade z da imagem. Nessa técnica de segmentação, a eliminação do fundo é importante, pelo fato de impossibilitar a formação de agrupamentos diferentes para cada tipo de fundo. Portanto, inicialmente, utilizou-se a técnica de limiar de Otsu (1975) com o objetivo de subtrair o fundo que, presumidamente, é separado do corpo do indivíduo. O algoritmo de segmentação *k-means* utilizado nesse trabalho pode ser encontrado em Theodoridis and Koutroumbas (2008). Nesse algoritmo, a imagem original I é convertida na forma de vetor. Os principais parâmetros utilizados para esse algoritmo foram: 1) O número de grupos a ser formado, k (3, 4 ou 5); 2) Os centros iniciais, c_1, c_2, c_3 e c_4 e c_5 de cada agrupamento, com: $c_1 = 0$, $c_2 = \text{valor mínimo do mapa de profundidade}$, $c_3 = \text{valor máximo do mapa de profundidade}$ e $c_4 = \text{média entre os valores } c_2 \text{ e } c_3$ e $c_5 = c_4/2$. Adotou-se a distância Euclidiana como métrica de distância para formação dos agrupamentos. O grupo referente à região de interesse é selecionado como aquele que tem o centro mais próximo da câmera.

2.3 Pós-processamento

2.3.1. Filtragem

Na maioria das configurações de mão (CM), a mão e o antebraço situam-se à frente do corpo e a segmentação ocorre sem problemas. Entretanto, quando o braço do usuário não está posicionado bem à frente do corpo, e sim lateralmente, a região do antebraço e partes do corpo, como a cabeça, podem ser segmentadas no mesmo grupo. Diz-se, nesses casos, que a segmentação apresenta ruídos. Na Figura 3 apresenta-se um exemplo em que o antebraço é segmentado juntamente com a cabeça. Para filtragem desses ruídos, aplica-se o seguinte procedimento: são medidas as distâncias dos centros dessas regiões segmentadas para o eixo vertical do corpo do usuário. Assumindo que as regiões não estão conectadas entre si, a região contendo o antebraço e a mão (região de interesse - ROI) é aquela que está mais distante do eixo central da imagem. Tal situação é mostrada na Figura 3(a). Na Figura 3(b) apresenta-se o resultado do processo de filtragem.

2.3.2. Rotação e remoção do antebraço

A rotação é utilizada para padronizar todas as ROIs numa mesma direção, a direção vertical. Esse processamento faz-se necessário para a aplicação do algoritmo de extração da mão, conforme será apresentado nas seções seguintes. Para determinação do ângulo de rotação da ROI são utilizados o primeiro e segundo momentos das imagens (Prokop and Reeves, 1992). A matriz de rotação é aplicada na imagem usando a técnica de interpolação do vizinho mais próximo. Na Figura 4(a) apresenta-se um exemplo de ROI rotacionada.

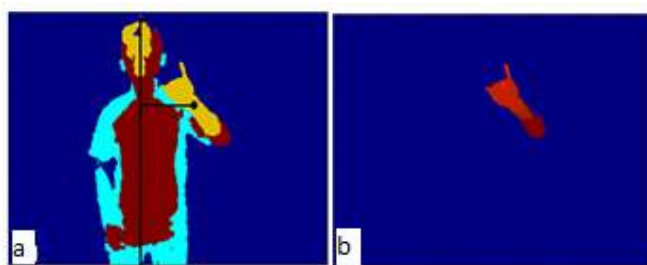


Figura 3- Configuração da mão com antebraço situado lateralmente ao corpo (a) resultado do método *k-means* com 4 grupos mostrando o antebraço segmentado em um mesmo grupo com a cabeça. (b) resultado da filtragem aplicando uma medida de distância.

Após a rotação, a remoção do antebraço é efetuada em duas etapas. O objetivo da primeira etapa é verificar se o cotovelo foi segmentado junto com o antebraço. Nessa etapa os seguintes passos são obedecidos: determina-se o centroide da imagem e traça-se os eixos maior e menor passando por esse centroide (Figura 4(a)); calcula-se a razão entre o eixo maior e o eixo menor. Quando o cotovelo é segmentado junto com o antebraço, a razão entre o eixo maior e o eixo menor é maior do que 3,3. Tal valor foi obtido experimentalmente, a partir da análise do conjunto de imagens do banco. Nos casos dessa razão ser superior ao valor de limiar, elimina-se a parte do antebraço abaixo do centroide (Figura 4(b)) e, posteriormente, aplica-se a transformada de distância. Do contrário, aplica-se simplesmente a transformada de distância.

A segunda etapa objetiva a extração da mão a partir da ROI. Para esse fim, utiliza-se a transformada de distância utilizada por Deimel e Schröter (1998) com o objetivo de determinar o centro e o raio da palma da mão. A transformada de distância consiste em associar a cada pixel p do objeto a menor distância euclidiana, $D(p, q)$, de p para um pixel de borda, q . Gera-se então uma imagem I em que, a intensidade de cada pixel é proporcional a essa distância. As coordenadas do centro, (x_{centro}, y_{centro}) corresponde às coordenadas do pixel de maior intensidade em I . O raio R da palma da mão corresponde ao valor de nível de cinza desse pixel, conforme mostrado na equação (1). A coordenada x_{corte} do ponto de corte (x_{corte}, y_{corte}) utilizado para extração da mão (Figura 4(b)) é encontrada multiplicando-se o valor de R por um escalar igual a 1,53 (Deimel e Schröter, 1998) e somando-se com a coordenada x_{centro} da palma da mão, conforme mostrado na equação (2). A região que estiver abaixo do ponto x_{corte} é eliminada da ROI. Na Figura 4(c) mostra-se o resultado da extração da mão aplicando essa transformada.

$$R = \max(I) \quad (1)$$

$$x_{corte} = x_{centro} + 1.53R \quad (2)$$

2.3.3. Padronização

O conjunto de imagens foi padronizado com tamanho 130x134 pixels, que corresponde as maiores dimensões encontradas nas imagens segmentadas. Imagens menores são preenchidas com intensidade zero à direita e abaixo.



Figura 4 - Seccionamento do braço indivíduo P_1, CM_{29} . (a) cálculo da relação entre os eixos, (b) resultado da primeira da fase, (c) resultado da segunda fase.

2.4. Normalização e Extração de características

Todas as imagens segmentadas e padronizadas são em seguida normalizadas. Para tal, subtrai-se de todos os pixels do objeto o pixel de menor valor diferente de zero. Através desse procedimento, objetiva-se que uma mesma configuração obtida a diferentes distâncias do *Kinect*[®] correspondam a imagens segmentadas com níveis de cinza semelhantes.

Para extração de características das imagens utilizou-se a técnica $2D^2LDA$, proposta por Nousath et al. (2006). Através dessa técnica consegue-se a redução de tamanho de uma imagem bidimensional em ambas as dimensões. A ideia é projetar as imagens em direções que maximizem um critério de separação entre classes (configurações). Por exemplo, o critério de Fisher (Theodoridis e Koutroumbas, 2008). A seguir descreve-se a operacionalização dessa técnica. Seja uma imagem original I_k , com dimensões $m \times n$. A técnica consiste em se determinar duas matrizes de projeção $X = [x_1, x_2, \dots, x_d]$ e $Z = [z_1, z_2, \dots, z_q]$, em que x_i são vetores com dimensões $n \times 1$ e z_j são vetores com dimensões $m \times 1$. Tanto os vetores x_i como os vetores z_j são associados a máximos valores de autovalores de matrizes de covariância associados as imagens I_k . Assim, X é uma matriz com dimensão $n \times d$ e Z é uma matriz com dimensão $m \times q$. Considerando C_k a matriz com dimensões reduzidas associada a imagem I_k resultante da aplicação de $2D^2LDA$, a mesma é dada pela equação (3):

$$C_k = Z^T I_k X \quad (3)$$

Os elementos da matriz C_k são a entrada do classificador.

Na implementação da técnica, o conjunto original das imagens foi dividido em dois conjuntos de igual tamanho, sendo um de treinamento e outro de teste. As matrizes de projeções X e Z , descritas anteriormente, são construídas utilizando-se apenas as imagens de treinamento. As dimensões utilizadas para C_k são 5×5 , 10×10 , 15×15 e 20×20 .

2.3. Classificação

O classificador utilizado foi o k -vizinhos mais próximos (KNN). A fase de treinamento do algoritmo consiste em construir o conjunto de treinamento, constituído por 61 subconjuntos. Cada subconjunto correspondente a uma CM. Cada subconjunto contém metade das imagens da CM correspondente, ou seja 100 imagens. As outras 100 imagens formam o conjunto de teste. Dado um novo padrão, o classificador calcula a distância desse novo padrão em relação a cada padrão no conjunto de treinamento, criando uma lista ordenada, onde o padrão do conjunto de treinamento mais próximo do novo padrão

encontra-se no topo da lista e o padrão mais longe, na base da lista (Theodoridis e Koutroumbas, 2008). Para o valor de $k=1$, a classe a que pertence o padrão corresponde a do topo da lista. Então, o novo padrão, será classificado como sendo dessa configuração. Quando $k > 1$, o padrão pertencerá a configuração que mais aparecer entre os k primeiros elementos da lista. Foram realizados vários experimentos com os valores de 1 a 10 para o valor de k .

3. Resultados

3.1. Segmentação

As Figuras 5, 6 e 7 ilustram exemplos de segmentação para diferentes configurações, CM_j ($1 \leq j \leq 61$), e indivíduos, P_i ($1 \leq i \leq 10$), ao se utilizar 3, 4 e 5 grupos ($k=3$, $k=4$ e $k=5$), respectivamente, no algoritmo *k-means*. Observando-se as Figura 5(d), 6(d) e 7(d), verifica-se que, quando a mão do usuário está à frente do corpo, a ROI inclui o antebraço e a mão.

Dois problemas podem ser observados nas figuras 5, 6 e 7: ruídos e erros de segmentação. Os ruídos podem ser vistos nas Figuras 5(a), 5(b), 5(f), 6(f), 7(a) e 7(f). De uma forma geral, com $k=3$, 66,5% das imagens apresentaram algum tipo de ruído, enquanto que com $k=4$ e $k=5$ uma porcentagem de 39,2% e 17,7% das imagens, respectivamente, apresentaram ruídos. O procedimento para filtragem dos ruídos foi descrito anteriormente. As 7(d) e 7(e) mostram que para $k=5$, parte da mão foi rotulada em outro grupo (cor marrom). De uma forma geral, com $k=5$ foram encontrados erros de segmentação da mão em 2,12% das imagens, enquanto que, para $k=3$ e $k=4$, esses erros não foram encontrados.

Com base nessas análises conclui-se que para $k=3$ a segmentação da ROI apresenta muitos ruídos, para $k=5$ a segmentação foi errônea em algumas imagens. Portanto, a melhor segmentação foi obtida utilizando-se $k=4$. As análises seguintes são baseadas nas imagens resultantes da segmentação com $k=4$.

3.2. Classificação

A Figura 8 ilustra as variações da taxa média de acerto (taxa obtida considerando a classificação de todas as configurações) do método de classificação em função dos parâmetros utilizadas na etapa de classificação: tamanho da matriz de características, C_k , do algoritmo 2D²LDA e valor de k no algoritmo KNN.

A partir da Figura 8 observa-se que: o maior valor para a taxa média de acerto, considerando todas as dimensões da matriz C_k , é obtida para 1NN e corresponde ao valor de 95,70% com C_k de dimensão 15x15. A Figura 9 mostra valores da taxa de acerto para cada CM e o quão distante as taxas de acerto de cada configuração estão da taxa de acerto média obtida com C_k com dimensão 15x15 e 1NN. As CM 19, 25, e 56 apresentaram as maiores taxas de acerto. As CM 36 e 51 apresentaram as menores taxas de acerto. Na Tabela 1 mostra-se o valor da máxima taxa média de acerto para cada dimensão da matriz C_k e o desvio padrão observado quando se considera os valores das taxas de acerto em cada configuração. Pode-se observar que: a máxima taxa média de acerto é obtida com a matriz de características C_k com dimensão 15x15 e 20x20; o desvio padrão obtido com a matriz de características C_k com dimensão 15x15 é menor do que o desvio padrão obtido com a matriz de características C_k com dimensão 20x20.

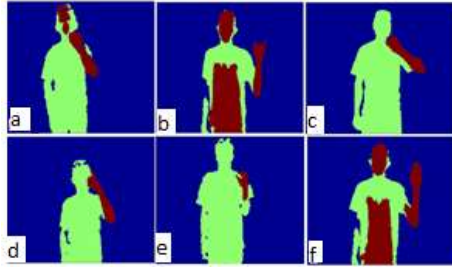


Figura 5 - Segmentação k -means com $k=3$. (a) P_2, CM_1 , (b) P_1, CM_3 , (c) P_5, CM_6 , (d) P_2, CM_{16} , (e) P_3, CM_{28} , (f) P_1, CM_{26} .

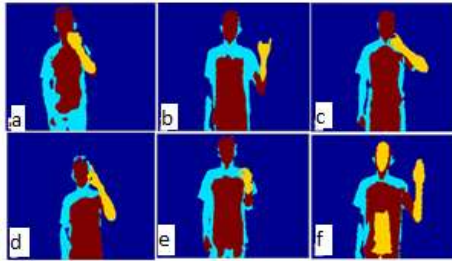


Figura 6 - Segmentação k -means com $k=4$:(a) P_2, CM_1 , (b) P_1, CM_3 (c) P_5, CM_6 , (d) P_2, CM_{16} , (e) P_3, CM_{28} , (f) P_1, CM_{26} .

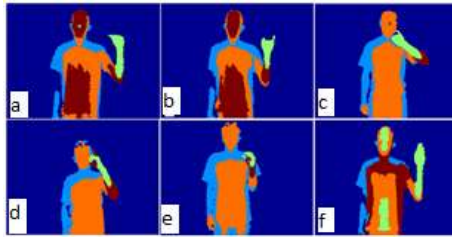


Figura 7 - Segmentação k -means com $k=5$:(a) P_2, CM_1 , (b) P_1, CM_3 , (c) P_5, CM_6 , (d) P_2, CM_{16} , (e) P_3, CM_{28} , (f) P_1, CM_{26} .

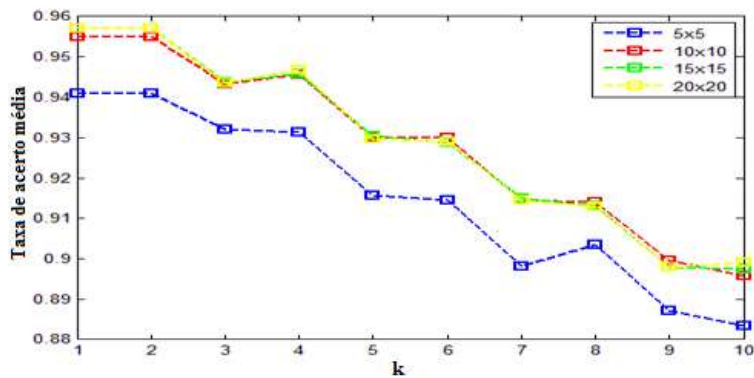


Figura 8 - Taxa de acerto da classificação em função do tamanho da matriz de características, C_K , do algoritmo $2D^2LDA$ e do valor de k do método KNN.

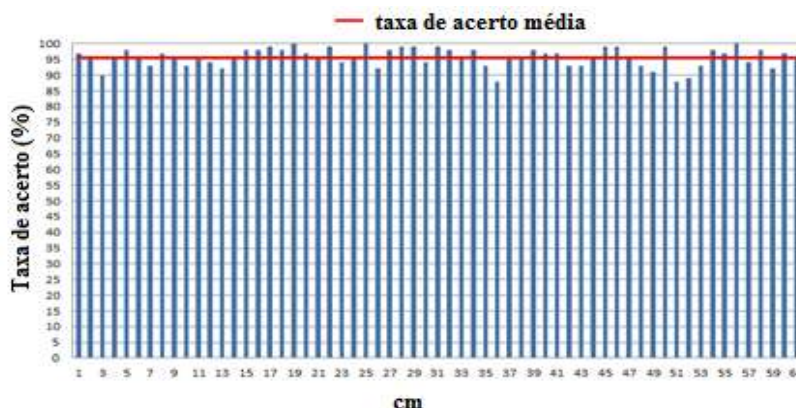


Figura 9- Taxa de acerto para C_k com dimensão 10x10 e 1NN, para cada CM

Tabela 1- Desempenho da classificação para $k=1$ (1NN).

Matriz C_k	Taxa de acerto média(%)	Desvio padrão
5x5	94,10	3,22
10x10	95,51	2,98
15x15	95,70	2,98
20x20	95,70	3,02

4. Conclusão

O estudo apresenta uma metodologia para reconhecimento de gesto baseado em imagens de profundidade. As principais diferenças entre este trabalho e outros já apresentados são as seguintes: uma base robusta com 12100 imagens da língua LIBRAS foi gerada para treinamento e teste; fatores de luminosidade não limitam a aplicação desenvolvida, uma vez que são utilizadas imagens de profundidade; vetores de características com diferentes tamanhos (400, 225, 100 e 25) são gerados utilizando-se a técnica $2D^2LDA$. Vetores contendo 225 e 400 características foram os que resultaram em melhores resultados de classificação com o classificador KNN. Para esse classificador, os melhores resultados foram obtidos com k igual a 1. Os resultados mostram que a máxima taxa média de acerto é de 95,70%, com um desvio padrão de 2,98 para vetores com dimensão 225, e 95,70% com um desvio padrão de 3,02 para um vetor com tamanho 400. Os valores desses desvios padrões podem ser explicados, principalmente, pelas configurações 36 e 51, que devido à baixa resolução do Kinect® acabam gerando confusão com CM 37 e 52, respectivamente, acarretando erros no algoritmo de classificação. O método descrito neste trabalho foi aplicado para o reconhecimento das CM da LIBRAS. Novos esforços serão endereçados futuramente em técnicas que visam reconhecer outros “fonemas” da LIBRAS como a expressão facial.

Agradecimentos

Parte dos resultados apresentados neste trabalho foram obtidos através do Projeto de Pesquisa e formação de recursos humanos, em nível de graduação e pós-graduação, nas áreas de automação industrial, softwares para dispositivos móveis e TV Digital, financiado pela Samsung Eletrônica da Amazônia Ltda., no âmbito da Lei no. 8.387 (art. 2º)/91.

Referências

- Bragatto, T. A. C., G. I. S. Ruas, and M. V. Lamar. (2006), Real-time video based finger spelling recognition system using low computational complexity Artificial Neural Networks: Telecommunications Symposium, 2006 International, p. 393-397.
- Carneiro, A., P. Cortez, and R. Costa. (2009), Reconhecimento de Gestos da LIBRAS com Classificadores Neurais a partir dos Momentos Invariantes de Hu: Interaction, p. 190-195.
- Chao, S., Z. Tianzhu, B. Bing-Kun, X. Changsheng, and M. Tao. (2013), Discriminative Exemplar Coding for Sign Language Recognition With Kinect: Cybernetics, IEEE Transactions on, v. 43, p. 1418-1428.
- Deimel, B., and S. Schröter. (1998), Improving Hand Gesture Recognition Via Video Based Methods for the Separation of the Forearm from the Human Hand, Dekanat Informatik, Univ.
- Lamar, M. V., M. S. Bhuiyan, and A. Iwata. (1999), Hand gesture recognition using morphological principal component analysis and an improved CombNET-II: Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on, p. 57-62 vol.4.
- Maraqa, M., F. Al-Zboun, M. Dhyabat, and R. A. Zitar. (2012), Recognition of Arabic Sign Language (ArSL) using recurrent neural networks: Journal of Intelligent Learning Systems and Applications, v. 4, p. 41.
- Noushath, S., G. Hemantha Kumar, and P. Shivakumara. (2006), (2D)2LDA: An efficient approach for face recognition: Pattern Recognition, v. 39, p. 1396-1400.
- Otsu, N. (1975), A threshold selection method from gray-level histograms: Automatica, v. 11, p. 23-27.
- Pimenta, N., and R. M. de Quadros. (2010), Curso de LIBRAS 1: iniciante, LSB Vídeo.
- Porfírio, A. J., K. Lais Wiggers, L. E. S. Oliveira, and D. Weingaertner. (2013), LIBRAS Sign Language Hand Configuration Recognition Based on 3D Meshes: Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, p. 1588-1593.
- Prokop, R. J., and A. P. Reeves. (1992), A survey of moment-based techniques for unoccluded object representation and recognition: CVGIP: Graphical Models and Image Processing, v. 54, p. 438-460.
- Rakun, E., M. Andriani, I. W. Wiprayoga, K. Danniswara, and A. Tjandra. (2013), Combining depth image and skeleton data from Kinect for recognizing words in the sign system for Indonesian language (SIBI [Sistem Isyarat Bahasa Indonesia]): Advanced Computer Science and Information Systems (ICACISIS), 2013 International Conference on, p. 387-392.
- Ribeiro, H. L., and A. Gonzaga. (2006), Reconhecimento de gestos de mão usando o algoritmo GMM e vetor de características de momentos de imagem.
- Theodoridis, S., and K. Koutroumbas. (2008), Pattern Recognition, Fourth Edition, Academic Press, 900 p.