

Interação com o Ambiente Mediante Classificação por Distância Euclidiana

Werikson F. de O. Alves * Ketia S. Moreira * Alexandre S. Brandão *

** Núcleo de Especialização em Robótica (NERo)
Departamento de Engenharia Elétrica
Programa de Pós-Graduação em Ciência da Computação
Universidade Federal de Viçosa, Viçosa – MG
(e-mail: {werikson.alves, ketia, alexandre.brandao}@ufv.br).*

Abstract: Applications based on gestures or actions are increasingly common in everyday life. This is because it allows the user to activate the equipment without the need to touch it, besides being possible to activate it remotely. Thus, this work proposes a gesture recognition method using Principal Component Analysis (PCA) as a dimensionality reducer of variables, and the Euclidean distance as a classifier. The objective here is to interact with an environment through the actions of a user. To do this, a own database with 14 classes of actions were created. The classifier is validated through an online operation and confusion matrix are used to analyze the results. The classifier achieved a hit rate higher than 80% for 6 of the 14 classes. This shows that the present work can be improved. However, it is worth noting that the main contribution was the formalization of a simplistic classifier, which makes use of a small database.

Resumo: As aplicações baseadas em gestos ou ações são cada vez mais comuns na vida cotidiana. Isto porque permite ao usuário ativar o equipamento sem a necessidade de tocá-lo, além de ser possível ativá-lo remotamente. Assim, este trabalho propõe um método de reconhecimento de gestos utilizando a Análise de Componentes Principais (PCA) como redutor de dimensionalidade de variáveis, e a distância euclidiana como classificador. O objetivo aqui é o interagir com um ambiente através das ações de um usuário. Para isso, foi criado uma base de dados próprio com 14 classes de ações. O classificador é validado através de uma operação online e uma matriz de confusão é utilizada para analisar os resultados. O classificador atingiu uma taxa de acerto superior a 80% para 6 das 14 classes. Isto mostra que o presente trabalho pode ser melhorado. No entanto, vale ressaltar que a principal contribuição foi a formalização de um classificador simplista, que faz uso de um pequeno banco de dados.

Keywords: Action Recognition; Interactive Robotics; PCA; OptiTrack.

Palavras-chaves: Reconhecimento de Ações; Robótica Interativa; PCA; OptiTrack.

1. INTRODUÇÃO

A robótica é uma das áreas de pesquisa sobre tecnologia muito importante e que está sempre em desenvolvimento, sendo muitas vezes usada para auxiliar o ser humano em tarefas que operam de forma automática e de forma repetida (Campos, 2019). Desde o seu surgimento, as aplicações da robótica sempre foram extensas começando na indústria e se espalhando para outras áreas, como por exemplo na saúde, segurança e na educação. Com o passar do tempo, a interação entre pessoas e tecnologias se desenvolveu e ainda continua evoluindo de forma muito rápida, tanto para tarefas simples quanto para tarefas mais complexas. Com a busca por uma técnica simples e que possa atender diversas áreas, várias técnicas de interação foram desenvolvidas e estudadas ao longo do tempo. Entre elas, aplicações baseadas em movimentos do corpo humano (tais como, reconhecimento de gesto, expressões e vozes) têm sido usadas cada vez mais em diversas áreas, por exemplo, jogos, acessibilidade, reabilitações, esportes, treinamentos,

etc. Afinal, esta ferramenta permite uma interação mais humana com a tecnologia, sem a necessidade de cabos ou artefatos semelhantes ligados ao usuário (Viana, 2015) .

Estas formas de controles, cujos comandos são realizados pelo corpo humano através da captação de gesto, podem ser chamadas de Interface Natural do Usuário (INU) (de Souza et al., 2016). Alguns exemplos destas aplicações são: sistema de INU desenvolvido para televisão interativa desenvolvido por Fan et al. (2013), controle de diversos modelos de robôs ou até mesmo o controle de um ambiente de forma remota (iluminação, ar condicionado ou um portão eletrônico, por exemplo).

Dado esta evolução tecnológica, novas formas de processar esses dados foram desenvolvidas ao longo do tempo como, por exemplo, a Análise de Componentes Principais (PCA) e a Análise Discriminante Linear (LDA) (Martinez and Kak, 2001). Com estas ferramentas, pode-se realizar a redução de dimensionalidade, compactação de dados, extração de recursos e visualização de dados.

* Este trabalho teve suporte financeiro da FAPEMIG.

O foco deste trabalho é o uso da técnica PCA, uma vez que sua função principal é encontrar o padrão dos dados analisados e assim encontrar a base de dados mais significativa para então retorná-lo como um subconjunto de informações revelando a estrutura oculta no conjunto de dados e filtrando o ruído (Kurita, 2019). Tal funcionalidade se enquadra à proposta do desenvolvimento de um classificador simples.

Neste contexto, o objetivo deste trabalho é a criação de um método de classificação de ações, que utilize poucos dados de entradas e possua uma técnica de processamento simples, e que o usuário possa realizar gestos predefinidos, naturais e intuitivos, que se relacionam com ações de alto nível. Como objetivos específicos tem-se a) a coleta de informações espaciais do usuário, b) o processamento destes dados e c) a aplicação da técnica desenvolvida.

A sequência deste trabalho está organizada como: A seção 2 apresenta trabalhos relacionados, compondo a base teórica deste trabalho. A seção 3 descreve as etapas do método desenvolvido. A seção 4 faz um estudo dos resultados obtidos. Por fim, a seção 5 apresenta as conclusões deste trabalho e sugestões para sua continuidade.

2. REFERENCIAL TEÓRICO

À medida em que os meios tecnológicos evoluem, também é possível observar uma crescente em relação a Interação Homem-Máquina (IHM), a qual pode ser aplicada em diversos setores, como, por exemplo, controle de robôs, plataformas ou de salas e/ou ambientes fechados de forma remota (Costalonga et al., 2014). Dessa forma, pode-se dizer que a IHM se baseia em um sistema onde uma tarefa é executada pela cooperação do componente humano e de automação (Ramaswamy et al., 2014).

Segundo Pereira et al. (2009), a busca por uma interface de comunicação mais natural ocorre pela tentativa de deixar as pessoas cada vez mais à vontade ao interagir com uma máquina. Com isto em mente, diversas formas de interfaces foram criadas e são estudadas rotineiramente, como, por exemplo, reconhecimento de voz (Hudson et al., 2017), periféricos (Huttenrauch and Eklundh, 2002) e reconhecimento de gestos (Chen et al., 2019). Dentre estas, no reconhecimento de gesto existem diversas de técnicas utilizáveis podendo ser baseadas na forma da mão, movimento do corpo ou de membros específicos do corpo.

De acordo com (de Carvalho et al., 2022a), devido a facilidade e a possibilidade de se trabalhar de forma remota e as mais variadas aplicações, o reconhecimento de gestos vem ganhando mais destaque no meio acadêmico. Dessa forma, a IHM por meio de gestos pode ser realizada através da linguagem corporal, viabilizando que uma pessoa possa controlar um determinado ambiente fechado sem precisar estar presente no local.

Para realizar o reconhecimento de uma ação, inicialmente é necessário realizar o pré-processamento das informações coletadas para, em seguida, realizar o processo de reconhecimento. Existem diversos meios de se fazer este processamento, comumente a captura de informações para a classificação dos gestos é proveniente de câmeras de vídeo, como *OptiTrack*, *Intel RealSense* e *Kinect*. As imagens adquiridas possibilita o rastreamento das informações

relacionadas ao usuário, por meio da correspondência entre consecutivos quadros baseados e das características principais observadas.

Para se ter um reconhecimento de ação eficiente diversos dispositivos podem ser usados e que eles podem variar de situação para situação, o *OptiTrack* se torna uma excelente ferramenta para a coleta de dados, devido a sua alta precisão e pela quantidade de informação disponibilizada. Em seu conjunto, o *OptiTrack* é composto por marcadores reflexivos e câmeras infravermelhas (que fazem a função de sensores de profundidade), capazes de inferir a posição tridimensional desses marcadores.

Este sistema de rastreamento possibilita a calibração e a configurações de diversas funções que auxiliam a captura de informações, o fornecimento de interfaces para captura e o processamento de dados 3D, podendo ser gravado ou transmitido ao vivo para outros programas (Motive, 2020). Além destas funcionalidades, o *OptiTrack* pode ser utilizado em outras aplicações, como, por exemplo, produções visuais, estudo de movimentos de corpos, realidade virtual, animações e na robótica.

Para realizar a coleta de dados, é necessário fazer o rastreamento através de marcadores ativos ou passivos. Ao receber esse sinal luminoso emitido ou refletido, respectivamente, a câmera consegue identificar os marcadores mais brilhantes e diferencia-los dos objetos menos brilhantes da cena. A diferença entre eles é que os marcadores passivos não necessitam de cabos ou baterias, porém requerem mais pré-processamento para sua identificação (Vital et al., 2015). Seguindo a proposta de trabalhos prévios (de Carvalho et al., 2022b), após a captura das imagens em profundidade pelo *OptiTrack*, o programa identifica o movimento feito pelo usuário, classifica a ação realizada e realiza a interação com o ambiente.

3. METODOLOGIA

3.1 Base de dados

Trabalhar com reconhecimento de gestos por meio de técnicas de aprendizado de máquinas, exige comumente uma grande quantidade de conjuntos de dados. Este trabalho, entretanto, possui a proposta de interagir com um determinado ambiente através de gestos simples realizados por diferentes usuários. Assim, optou-se pela criação de uma base de dados própria, onde cada tarefa está associada a uma classe diferente, de forma a ganhar um pouco mais de liberdade e flexibilidade.

O sistema de rastreamento *OptiTrack* é constituído por oito câmeras infravermelhas posicionadas conforme na Figura 1. Neste trabalho, foram utilizados três marcadores passivos conforme mostrado na Figura 2. O posicionamento das câmeras possibilitou a execução dos gestos de diversas posições. Assim, a fim de tornar o processo de reconhecimento do gesto mais dinâmico e simples, foram utilizadas duas fitas reflexivas no formato de pulseiras, colocadas uma em cada punho do usuário e um terceiro marcador na região da cabeça, atuando como ponto de referência.

Ao realizar uma análise comparativa com bases de dados de trabalho prévios (Basilio et al., 2019; de Carvalho et al.,

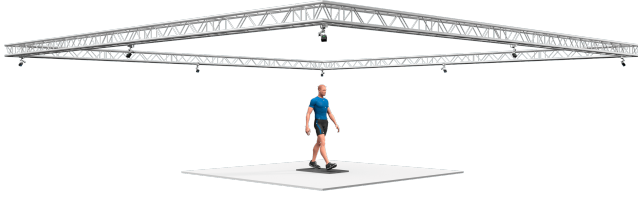


Figura 1. Posicionamento das câmeras no ambiente de simulação. Motive (2020)

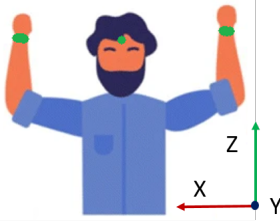


Figura 2. Localização dos marcadores no usuário.

2022b), percebe-se que muitas delas têm uma grande quantidade de amostra para cada gesto. Contudo, a proposta da trabalho é desenvolver um classificador simples, com uma base de dados de apenas 70 amostras iniciais, sendo 14 classes diferentes e 5 amostras para cada classe. A diferença dos trabalhos anteriores é a possibilidade de ampliação do número de amostras, caso o desempenho do classificador não seja satisfatório. Desta forma, utilizando o *software* Motive, foram feitas as gravações dos gestos, cujas marcações de início e o fim de cada ação se deram quando uma pessoa está com os braços para baixo, representando a posição de repouso.

Para se detectar o início e o fim de um gesto, a abordagem escolhida foi a utilização de um gatilho duplo, no qual são avaliados a altura e a velocidade de ambas as pulseiras. Levando-se em conta que todos os movimentos necessitam que ao menos uma das pulseiras seja erguida em relação ao solo, foi escolhido a coordenada z de cada marcador, a qual está posicionada perpendicular ao solo. Daí, analisou-se a posição e velocidade de cada pulseira, em uma janela fixa de f quadros, num período de t milissegundos. Logo, se a velocidade ou a altura do marcador alcançarem um determinado valor mínimo (V_{min} em metros por quadros e H_{min} em metros), serão armazenados os últimos d quadros anteriores ao acionamento de uma das condições, as janelas atuais e até aqueles onde ambas as condições não são mais cumpridas.

Os valores de f e t são baseados na frequência de coleta de informações utilizada no *software* Motive, neste caso, 240 Hz. Portanto, para uma amostragem de 100 milissegundos com intervalos fixos de 24 quadros, o valor de d será sempre um a menos que f , 23 quadros. Os valores de V_{min} e H_{min} foram determinados ao considerar a altura média da posição de repouso dos possíveis usuários, 0,95 metros, e a velocidade média do movimento dada a precisão milimétrica do sistema de rastreamento *OptiTrack*, 0,001 metros por quadros.

Cada classe de movimento pode ser usada para controlar uma determinada função dentro de uma sala. As classes utilizadas são apresentadas na Figura 3, sendo elas:

- A) Levantar as duas mãos e depois abaixar a direita;
- B) Girar os dois braços, terminando com eles na frente do peito, com a mão direita ligeiramente acima da esquerda;
- C) Levantar as duas mãos e abaixar a esquerda;
- D) Desenhar um círculo na frente do rosto com a mão direita, enquanto mantém o braço esquerdo ao lado do corpo;
- E) Desenhar um círculo na frente do rosto com a mão direita, enquanto levanta o braço esquerdo até a altura do ombro;
- F) Desenhar um círculo na frente do rosto com a mão direita, enquanto levanta o braço esquerdo até a altura da cabeça;
- G) Desenhar um círculo na frente do rosto com a mão esquerda, enquanto mantém o braço direito ao lado do corpo;
- H) Desenhar um círculo na frente do rosto com a mão esquerda, enquanto levanta o braço direito até a altura do ombro;
- I) Desenhar um círculo na frente do rosto com a mão esquerda, enquanto levanta o braço direito até a altura da cabeça;
- J) Cruzar os dois braços formando um X na frente do peito;
- K) Mover o braço direito para frente e depois deslizar para a esquerda;
- L) Levantar os dois braços até a altura do estômago e depois levantar a mão direita;
- M) Levantar os dois braços até a altura do estômago e depois levantar a mão esquerda;
- N) Mover o braço esquerdo para frente e depois deslizar para a direita.

Cada quadro consiste em um conjunto de pontos de interesse que em sequência compõem a matriz de características. Tais pontos de interesse indicam a posição dos marcadores posicionados na região da cabeça (O), no braço esquerdo (P) e no braço direito (R), conforme mostrado na Figura 2. Matematicamente, a posição tridimensional de cada marcador é expressa por

$$F_q = [O_x \ O_y \ O_z \ P_x \ P_y \ P_z \ R_x \ R_y \ R_z], \quad (1)$$

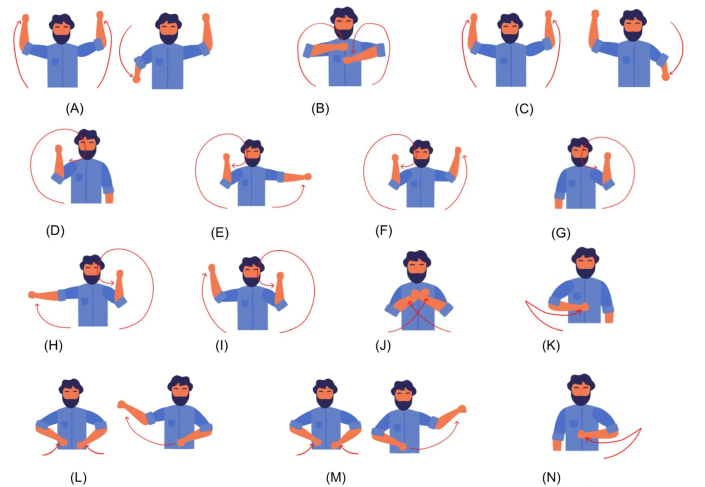


Figura 3. Classes de gestos no conjunto de dados (de Carvalho et al., 2022b).

para $q = 1, 2, \dots, Q$, onde Q informa o quadro atual. As colunas de F_q representam as coordenadas x , y e z de cada marcador, onde z é perpendicular ao solo. Em seguida, a amostra completa é armazenada como a concatenação subsequente da matriz de características,

$$A_{Q,U} = [F_1 \ F_2 \ \dots \ F_Q]^\top, \quad (2)$$

onde U é o número da coluna.

Para realizar o reconhecimento da ação, foram coletadas as informações das duas pulseiras, ao longo do gesto. Já o marcador que está posicionado na região da cabeça, foi utilizado como referência, afinal ao realizar um mesmo gesto em posições diferentes, o movimento relativo será praticamente idêntico. Este marcador também serviu para centralizar todas as características (em seus respectivos eixos) em relação às coordenadas do marcador posicionada na cabeça, corrigindo uma eventual movimentação da cabeça em decorrência do movimento de outra parte do corpo. Para isto, adotou-se

$$B_{Q,U} = A - \bar{O}, \quad (3)$$

onde \bar{O} é a média das coordenadas das informações relacionadas ao marcador O , A é a matriz contendo todas as informações dos três marcadores referentes aquela amostra e B é o novo conjunto de dados gerado.

As amostras foram normalizadas por

$$C_{Q,U} = \frac{B_{Q,U} - \bar{B}_U}{S_B}, \quad (4)$$

onde S_B é desvio padrão da amostra analisada, \bar{B}_U é a média da coluna do elemento em análise, centralizando a variável e resultando em uma variável com média zero e variância igual a unidade, dessa forma com desvio padrão igual a um. Isto visa evitar ter um conjunto de informações com excessiva desproporcionalidade, evitando redundâncias e tornando-o mais organizado. Assim, as variáveis normalizadas podem ser diretamente comparáveis, devido ao fato de não possuírem dimensão e serem expressas em unidades de desvio padrão.

3.2 Algoritmo de reconhecimento de ações

Agora, tendo esses dados em mãos, foram aplicadas a PCA e método da distância euclidiana para classificar a base de dados, conforme detalhado a seguir.

Para simplificar a classificação de cada amostra, em vez de se analisar todas as informações capturadas pelo sensor, optou-se por aplicar a técnica de PCA, a qual possibilita realizar a redução de dimensão da matriz de características, sendo ela originalmente $A \in R^{Q \times 9}$. Para isto, após a centralização dos dados, são desconsideradas as informações relacionadas ao marcador da cabeça, devido suas informações permanecerem praticamente constante ao longo do tempo em relação as outras informações, gerando $B \in R^{Q \times 6}$, para em seguida, aplicar a normalização, resultando $C \in R^{Q \times 6}$.

Em seguida, aplicando a PCA, obtém-se os autovalores referentes as variáveis de entrada, em ordem decrescente, conforme

$$\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4 > \lambda_5 > \lambda_6, \quad (5)$$

onde λ_i representa a importância de cada variável para aquela amostra. Dado esta informação, ao analisar os

autovalores das amostras, percebe-se que, em quase todos os movimentos, os três primeiros autovalores descrevem mais de 90% das informações sobre o gesto. Logo, optou-se por escolher os três maiores autovalores, diminuindo a quantidade de dados a serem processados nas próximas etapas do trabalho.

Outro resultado que a PCA entrega é o conjunto de dados $D \in R^{6 \times 6}$ que contém as componentes principais. Como foram escolhidas as três componentes mais importantes para cada amostra (λ_1 , λ_2 e λ_3), foi obtido $E \in R^{6 \times 3}$, que contém um conjunto de dados reduzido de D .

O processo de classificação é baseado na comparação de uma ação qualquer com uma outra ação de referência. Para realizar isto, são utilizados os dados contidos na matriz E obtido da amostra de referência e da amostra em análise. Assim, o reconhecimento do gesto pode ser obtido por meio da comparação da série de informações do conjunto de dados reduzidos da ação a ser classificada, com a da ação da referência. Ao alterar a ação de referência para diferentes classes, aquela com menor diferença, ou seja, menor distância Euclidiana, representa a classe da ação a ser classificada. Em outras palavras, a classificação acontece entre a classe corrente e as classes de referência, conforme,

$$Dist = \sum_{j=1}^3 ||D_{jReferencia} - D_{jCorrente}|| \quad (6)$$

A Figura 4 ilustra um fluxograma que resume as etapas do subsistema de classificação realizadas neste trabalho e seus respectivos resultados.

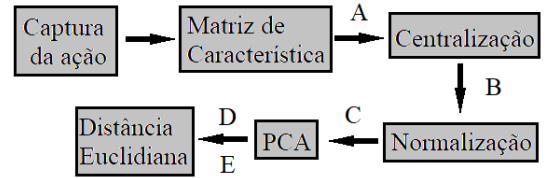


Figura 4. Fluxograma de classificação proposto.

4. RESULTADOS E DISCUSSÃO

A partir do *software* Motive, foram realizadas as gravações dos gestos, conforme visto na Figura 5, de forma a construir o banco de dados, no qual cada amostra na base de dados é composta por Q linhas e 9 colunas, conforme explicado na seção anterior. A figura mostra uma das amostras da classe de gesto N utilizadas no trabalho, na qual $Q = 3822$ quadros. Na parte superior da figura é mostrada o espaço de trabalho e o mapeamento dos marcadores em suas respectivas posições, em determinado instante, e na parte inferior é mostrado o gráfico do movimento, em relação ao tempo de simulação. Sendo assim, ao analisar as duas partes simultaneamente, observa-se que a medida que o braço esquerdo do usuário se move, o valor no eixo corresponde, para o marcador *Left*, se altera, indicando o rastreamento daquele marcador.

A validação do método proposto foi dividido em duas etapas, sendo a primeira etapa o teste online e a segunda a classificação final do gesto. O teste online tem como objetivo confirmar se o conjunto de dados não é tendencioso,

ao realizar diversas comparações, na qual cada amostra do banco de dados seria comparada com todas as outras amostras de forma totalmente aleatória. Já a classificação final tem como objetivo realizar a seleção do gesto por meio da maioria das classificações.

Na primeira etapa, cada amostra foi comparada com todas as outras amostras da base de dados. Levando se em conta que no banco de dados existem cinco amostra para cada classe de gesto, ao realizar estas comparações serão obtidas cinco resultados para cada amostra, a partir da distancia Euclidiana. Contudo, um destes resultados será zero devido a comparação da amostra com ela mesma, durante o teste. Com isto, ao desconsiderar este resultado, restaram quatro respostas por amostra, totalizando vinte respostas por classe de gesto. O resultado para esta etapa é apresentado pela matriz de confusão da Figura 6.

Por meio da Figura 6, observa-se que houveram 3 classes que tiveram a classificação das amostras excelente, 100% de acerto (H, L e M), 4 classes com a classificação das amostras boa, acima de 60% de acerto, (A, C, E e K), 2 classes tiveram desempenho razoável (F e J), entre 40 e 60%, enquanto 5 tiveram o desempenho abaixo de 40% (B, D, G, I e N). A classificação entre as amostras, portanto, obteve desempenho médio 59,6%.

Outra informação que se pode extrair é que as classes de amostras mais confundidas são as F e I, as quais foram classificadas indevidamente. Isto pode ser explicado, pelo fato de que várias partes dos movimentos que envolvem

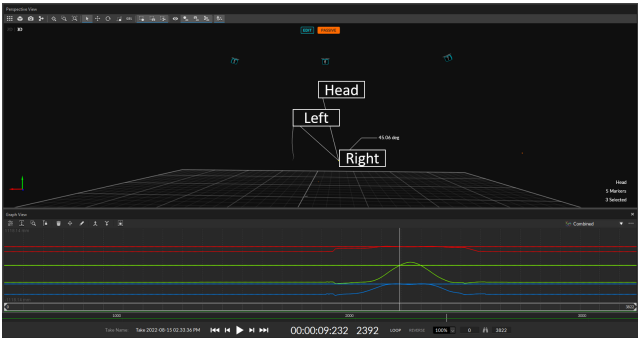


Figura 5. Amostra da classe de gesto N.

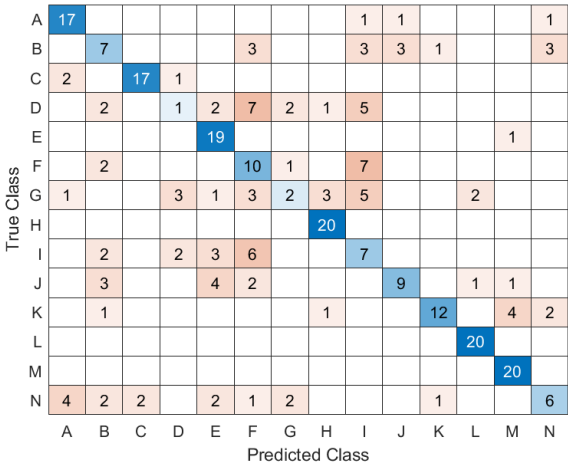


Figura 6. Matriz de confusão da classificação das amostras.

as ações F e I, também fazem parte dos movimentos de outras ações, fazendo com que os dados possuam uma determinada semelhança, causando um erro de classificação.

A segunda etapa do teste, trata-se da classificação final do gesto. Para se realizar a classificação de uma ação (independentemente se ela estiver correta ou errada) foi estabelecida a necessidade de que a maioria das amostras fossem classificadas como uma única classe, ou seja, para se obter uma classificação é necessário ter a “metade mais um” das classificações. Logo, sabendo que para cada amostra simulada foram obtidas 4 classificações, é preciso que no mínimo 3 respostas sejam classificadas como a mesma ação, para que ela receba uma classificação final e, caso essa condição não seja atendida, o gesto será classificado como inconclusivo. O resultado para esta etapa é apresentado na matriz de confusão da Figura 7.

Por meio da Figura 7, pode-se perceber que houve 4 classes tiveram um desempenho excelente (E, H, L e M), 3 classes tiveram um desempenho bom (A, C e K), 1 classe teve um desempenho razoável (F), 3 classes tiveram o desempenho abaixo de 40% (B, G e J) e 3 classes tiveram resultados inconclusivos (D, I e N).

Um ponto importante a se perceber é que 33 ações das 70 foram classificados como inconclusivos enquanto as outras 37 obtiveram uma classificação dentre as classes como resposta, ou seja, 52,86% das ações tiveram uma classificação final. Destas, apenas 3 foram classificados erroneamente, logo, pode-se dizer que o método teve um desempenho de 48,57% em relação ao total de amostras e um desempenho de 91,9% entre as amostras que receberam uma classificação dentre as classes predefinidas.

Essa elevada taxa de acerto se deve ao uso da classe inconclusivo, o qual evita que vários gestos sejam classificados erroneamente ao não serem reconhecidos. Com o uso desta classe Z, ao se realizar um gesto e o mesmo não puder ser classificado, será solicitado ao usuário a repetição do gesto.

Um outro ponto que vale ressaltar é que ao comparar as informações contidas em ambas as figuras, pode-se notar que apenas a ação J, que apesar das amostras terem 45% de acerto entre elas, teve uma classificação final distante do esperado, tendo apenas 20% de acerto. Além disto,

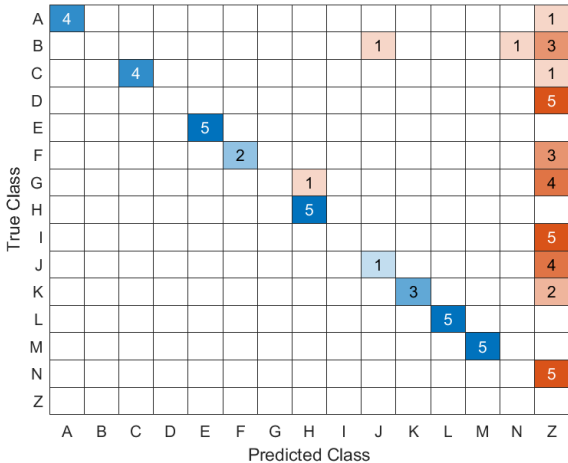


Figura 7. Matriz de confusão da classificação final.

ao se analisar as ações B, I e N, nota-se que apesar de terem índices para ao menos uma classificação final correta, é possível inferir que este resultado está disperso entre as amostras, afinal todas foram classificadas como inconclusivo.

5. CONSIDERAÇÕES FINAIS

Neste trabalho foi proposto um método de classificação de gesto utilizando Distância Euclidiana baseado no rastreamento de pulseiras reflexivas, ao se analisar os dados pré-processados obtidos da matriz de características da PCA, após coleta realizada pelo *Motive/OptiTrack*.

Ao observar as informações relacionadas as posições dos marcadores, foi possível concluir que o *OptiTrack* se mostra ser um bom dispositivo para coleta de dados relacionadas com rastreamento de corpos, devido a precisão e a grande quantidade de informações que se é possível extrair.

A análise de componentes principais se mostrou também ser muito eficaz, ao conseguir reduzir diversos conjuntos de informações, os quais possuíam tamanhos variados, facilitando a análise e comparações das informações.

O classificador final apresentou uma taxa de acerto média de 91,9% o qual pode ser considerado como tendo um excelente desempenho e, juntamente a isto, a classe inconclusiva se provou ser um recurso muito útil, para reduzir a quantidade de erros, reforçando a viabilidade e a eficiência, dada sua simplicidade, quando associado à pequena base de dados.

Desta forma, ao associar as classes com tarefas de alto nível, em algum ambiente fechado, espera-se que a maioria das ações sejam executadas corretamente. Logo, para trabalhos futuros, espera-se que ao se aumentar a base de dados, tenha-se um aumento no desempenho médio e que menos classes sejam classificadas como inconclusivas. Isto possibilitará a realização de missões guiadas por gesto para comandar robôs móveis em uma arena estruturada.

AGRADECIMENTOS

Os autores agradecem a FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais pelo suporte financeiro dado a esta pesquisa. Sr. Alves também agradece a esta agência pela bolsa de Iniciação Científica concedida, para sua dedicação ao trabalho.

REFERÊNCIAS

- Basilio, V.T., Carvalho, K.B., and Brandao, A. (2019). Reconhecimento de ações por rna em aplicações de robotica social. *XIV Simpósio Brasileiro de Automação Inteligente, Ouro Preto/MG*.
- Campos, F.R. (2019). *A robótica para uso educacional*. Editora Senac São Paulo.
- Chen, F., Lv, H., Pang, Z., Zhang, J., Hou, Y., Gu, Y., Yang, H., and Yang, G. (2019). Wristcam: A wearable sensor for hand trajectory gesture recognition and intelligent human-robot interaction. *IEEE Sensors Journal*, 19(19), 8441–8451. doi:10.1109/JSEN.2018.2877978.
- Costalonga, T.O., Ávila, L.M., Muniz, L., and Brandao, A.S. (2014). Gesture-based controllers to guide a quadrotor using kinect sensor. In *2014 Joint Conference on Robotics: SBR-LARS Robotics Symposium and Robocontrol*, 109–112. IEEE.
- de Carvalho, K.B., Basílio, V.T., and Brandão, A.S. (2022a). Action recognition for educational proposals applying concepts of social assistive robotics. *Cognitive Systems Research*, 71, 1–8.
- de Carvalho, K.B., Villa, D.K.D., Sarcinelli-Filho, M., and Brandao, A.S. (2022b). Gestures-teleoperation of a heterogeneous multi-robot system. *The International Journal of Advanced Manufacturing Technology*, 118(5), 1999–2015.
- de Souza, P.H.R., de Jesus Melo, W.A.L., de Santana, I.M.F., Doria, N.S.F., and de Matos, L.B.C. (2016). Controle de modelos robóticos por meio de gestos naturais. In *Anais do III Encontro Nacional de Computação dos Institutos Federais*, 732–735. SBC.
- Fan, Y.C., Chen, C.L., and Huang, S.S. (2013). Natural user interface for interactive television. In *2013 IEEE International Symposium on Consumer Electronics (ISCE)*, 189–190.
- Hudson, C., Bethel, C.L., Carruth, D.W., Pleva, M., Juhar, J., and Ondas, S. (2017). A training tool for speech driven human-robot interaction applications. In *2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 1–6. doi:10.1109/ICETA.2017.8102488.
- Huttenrauch, H. and Eklundh, K. (2002). Fetch-and-carry with cero: observations from a long-term user study with a service robot. In *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication*, 158–163. doi:10.1109/ROMAN.2002.1045615.
- Kurita, T. (2019). *Principal Component Analysis (PCA)*, 1–4. Springer International Publishing, Cham. doi: 10.1007/978-3-030-03243-2_649-1. URL https://doi.org/10.1007/978-3-030-03243-2_649-1.
- Martinez, A. and Kak, A. (2001). Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228–233. doi:10.1109/34.908974.
- Motive (2020). Motive Documentation - naturalpoint products. URL https://v22.wiki.optitrack.com/index.php?title=Motive_Documentation. Accessed: 2022-04-18.
- Pereira, F.G., Schmitz, N., Vassallo, R.F., and Berns, K. (2009). Cooperação entre homens e robôs baseada em reconhecimento de gestos. *IX Simpósio Brasileiro de Automação Inteligente, Brasília/DF*.
- Ramaswamy, A., Monsuez, B., and Tapus, A. (2014). Model driven software development for human-machine interaction systems. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 270–271.
- Viana, D.L. (2015). *Uma linguagem de domínio específico para descrição e reconhecimento de gestos usando sensores de profundidade*. Master's thesis, Universidade Federal de Pernambuco.
- Vital, J., Couceiro, M.S., Dias, G., and Ferreira, N. (2015). Tecnologias para a análise do movimento humano. In *R. Ruben, M. Vieira, C. Campos, H. Almeida, J. Siopa, P. Bártole, & J. Folgado (Edits.), 6º Congresso Nacional de Biomecânica*, 1–6.