



Identificação e Estimativa de posição de pessoas baseada em visão computacional para um robô de serviço da UTFPR.

Identification and pose estimation of people, based on computer vision, for a service robot at UTFPR.

David Segalle¹, Gustavo Fardo Armênio², João Alberto Fabro³

RESUMO

Robôs de serviço trazem diversos desafios ligados a visão computacional, entre eles reconhecer pessoas é de interesse desta pesquisa. Este artigo se propõe a utilizar um algoritmo KNN para realizar o processo de reconhecimento facial respeitando as limitações presentes em um sistema portátil de baixo custo energético. Este algoritmo precisa ser capaz de alcançar um treinamento rápido, permitindo o uso do sistema no ambiente de uma casa, onde existe a necessidade de aprender um novo operador e interagir com ele rapidamente para realizar ações como buscar um objeto. As dificuldades presentes limitam a utilização de diversas formas de reconhecimento facial desenvolvidas recentemente, porém, ela tem a vantagem de não necessitar aceitar muitos operadores, dado que o robô opera em um ambiente caseiro. Com isso em mente foi escolhido um algoritmo KNN para realizar uma análise de como melhor utilizá-lo com as restrições apresentadas.

PALAVRAS-CHAVE: algoritmo KNN; reconhecimento de pessoa; reconhecimento facial; robótica de serviço.

ABSTRACT

Service robots bring several challenges related to computer vision, among them recognizing people is the goal attempted by this research. This paper proposes a KNN algorithm to perform the task of recognizing individuals while respecting limitations presented by the usage of portable hardware with low energy usage. This algorithm needs to be capable of quickly performing a training routine, allowing for its use in a home environment, where there may be the necessity to learn a new operator and quickly interact with him in order to perform tasks such as taking an object to the operator. The challenges presented limit the usage of most face recognition technologies developed recently, on the other hand, being a service robot also brings with it the advantage of not requiring many people to be recognizable in the system. With all this in mind we decided to analyze a KNN algorithm in order to figure out how to best utilize it with the limitations present.

KEYWORDS: KNN algorithm; people recognition; facial recognition; service robotics.

INTRODUÇÃO

A robótica de serviço é um campo que vem crescendo nas últimas décadas com a aplicação de robôs móveis em ambientes comerciais e domésticos (como robôs aspiradores de pó). O avanço nos campos de Processamento Computacional, Inteligência Artificial e Robótica traz uma perspectiva de robôs de serviço cada vez mais autônomos, confiáveis e multifuncionais. Muitos desafios ainda se mantêm em aberto para a comunidade científico-tecnológica solucionar, em uma gama extensa de

¹ Bolsista da CNPQ. Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brasil. E-mail: davidsegalle@alunos.utfpr.edu.br. ID Lattes: <http://lattes.cnpq.br/9178633491461017>.

² Bolsista da CNPQ. Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brasil. E-mail: gustavofardoarmenio@alunos.utfpr.edu.br. ID Lattes: <http://lattes.cnpq.br/9811585096151665>.

³ Docente no Departamento Acadêmico de Informática-DAINF-CT, Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brasil. E-mail: fabro@utfpr.edu.br. ID Lattes: <http://lattes.cnpq.br/6841185662777161>.



áreas do conhecimento. A Robocup@Home é a maior iniciativa de fomento do desenvolvimento de robôs de serviço para ambiente residencial que possam auxiliar pessoas em tarefas domésticas de forma versátil e segura (WISSPEINTNER, 2009). Sendo uma categoria da competição Robocup, se propõe a estimular soluções inovadoras através de desafios interdisciplinares que envolvem problemas de Navegação Autônoma, Manipulação de Objetos, Interação Humano-Robô e Reconhecimento de Pessoas e Objetos. Esta pesquisa abordará a última categoria citada. Para os desafios propostos na competição, o processamento de informações visuais das entidades - pessoas, no contexto desta pesquisa - no ambiente pode envolver, a partir da detecção, o reconhecimento (ou a identificação), a segmentação, a localização e diversas outras técnicas de Visão Computacional e Inteligência Artificial em conjunto. A combinação delas objetiva extrair o máximo de informação relevante possível sobre as entidades de interesse para o processamento de decisão do robô. Contudo, muitas das técnicas mais avançadas encontram uma limitação de hardware. O robô deve ser capaz de se movimentar pelo ambiente, o que implica em portabilidade de seus componentes, incluindo os dispositivos de computação. Tais dispositivos devem, portanto, ser compactos e com o menor custo energético possível (já que são alimentados por baterias), o que gera uma capacidade de processamento reduzida. Esta pesquisa, portanto, se propõe ao estudo de métodos e técnicas de detecção e reconhecimento de pessoas aplicados a hardware de computação portátil, para a solução de tarefas domésticas dentro do escopo dos desafios da competição Robocup@Home, dentro do desenvolvimento realizado pela equipe UtBots@Home da UTFPR (PALAR, 2022).

MATERIAIS E MÉTODOS

As técnicas e métodos abordados neste projeto de pesquisa foram: segmentação de rostos e reconhecimento facial.

Todo o software utilizado nesta pesquisa têm como ambiente de hardware para teste dois dispositivos de computação compactos e portáteis: o computador embarcado Intel Nuc ¹ e a placa de desenvolvimento embarcado NVIDIA Jetson Nano ². Ambos utilizam distribuições do sistema operacional Linux Ubuntu, sendo o primeiro a versão 20.04 LTS e o segundo o Jetson JetPack SDK, baseado na versão 18.04 LTS também do Ubuntu. Além disso, para a extração de imagens, foi utilizado o sensor KinectV1, que fornece imagens RGB-D (red-green-blue-depth), ou seja, fornece imagens RGB coloridas e uma nuvem de pontos 3D ³.

Para desenvolvimento e teste, o uso do ROS ⁴ foi essencial. O ROS estabelece uma estrutura para a comunicação e a execução paralela de aplicações. Comumente, as aplicações implementadas com o ROS são chamadas de pacotes.

O funcionamento do algoritmo de reconhecimento facial utilizado requer uma imagem com apenas um rosto para seu funcionamento, portanto, é de suma importância a utilização de um

¹ Intel Nuc: *Intel Next Unit of Computing* é um dispositivo ou elemento de computação com fator de forma pequeno. Maiores informações em <https://www.intel.com.br/content/www/br/pt/products/docs/boards-kits/nuc/what-is-nuc-article.html>

² NVIDIA Jetson Nano: Sistema de processamento embarcado com aceleração gráfica. Maiores informações em <https://developer.nvidia.com/embedded/jetson-nano-developer-kit>

³ MICROSOFT, *Develop Network: Kinect Sensor*. Disponível em: <http://msdn.microsoft.com/en-us/library/hh438998.aspx>.

⁴ ROS: *Robot Operating System*, é um conjunto de bibliotecas e sub-sistemas que auxiliam o desenvolvimento de aplicações para robôs. Maiores informações em www.ros.org.



algoritmo prévio de segmentação. Para isso, utilizou-se o método *face_locations* da biblioteca *face_recognition* de python, está, por sua vez, utiliza um método da biblioteca *dlib* para realizar a tarefa. A segmentação de rostos de uma imagem utiliza HOG ⁵ para obter as coordenadas.

O processo de reconhecimento facial foi implementado através do algoritmo KNN *K-Nearest Neighbors* (T. M, 1967) e pode ser separado em duas fases principais: a primeira é a aquisição de imagens, e a segunda o reconhecimento. A primeira fase consiste na obtenção de imagens da face de uma pessoa, e para isso foi desenvolvida um rotina onde o usuário interage com o robô e mostra sua face a partir de diversos ângulos, criando uma ampla gama de imagens para o treinamento (SEGALLE, 2023). Já a fase do reconhecimento envolve utilizar os dados obtidos a partir do treinamento para identificar usuários conhecidos e suas posições, criando *bounding boxes* as quais podem ser utilizadas para a realização de diversas tarefas. Foi desenvolvido por este projeto um pacote do ROS que permite a integração destes algoritmos com o restante do sistema.

Dentro do algoritmo de reconhecimento facial, o foco foi sempre em buscar a melhor maneira de treinar o algoritmo com as imagens obtidas pelo sistema. Tal escolha se deve a um limite dado pela fase de treinamento, que define que um robô precisa ser capaz de aprender um novo usuário e interagir com ele rapidamente (o que limita a utilização de modelos neurais com longos treinamentos normalmente utilizados para o reconhecimento de faces). Quando implementado sem as limitações dadas o algoritmo pode obter um índice superior a 90% de reconhecimento (ASSIN, 2020)

RESULTADOS E DISCUSSÕES

O processo de análise de duração de treinamentos se deu realizando a rotina de treinamento diversas vezes, é necessário analisar a duração para cada quantidade de sujeitos que possam estar na competição. A quantidade de imagens a ser utilizada pode ser escolhida pelo algoritmo, portanto, é necessário analisar qual a maior quantidade de imagens as quais podem ser utilizadas considerando-se a quantidade de indivíduos presentes na *database* de treinamento. Com isso em mente, para cada comutação de indivíduos e imagens disponíveis, foram realizados 9 treinamentos, sendo a média deles o valor representado na tabela 1.

Os tempos de treinamento são diretamente influenciados pela quantidade de sujeitos a serem analisados 1, assim como a quantidade de imagens fornecidas para cada um deles. No âmbito da competição, não existe a necessidade de reconhecer uma gama muito ampla de indivíduos, portanto, o foco foi buscar a maior quantidade de imagens por pessoa para ser utilizada, permitindo a melhor qualidade do resultado dentro das restrições apresentadas. Nas quantidades de imagens e indivíduos testados, problemas de tempo só começam a aparecer quando 11 indivíduos são apresentados ao sistema, com uma grande quantidade de imagens (15) para cada um.

A qualidade do reconhecimento é diretamente influenciada pela quantidade de imagens e inversamente correlacionada com a quantidade de indivíduos, ou seja, o melhor reconhecimento ocorre quando há o maior número possível de imagens para o menor número de indivíduos. Conforme o número de indivíduos a serem treinados aumenta, 15 imagens se aproxima do limite de 60 segundos

⁵ HOG significa *Histogram of Oriented Gradients*, uma breve explicação do algoritmo pode ser encontrada em <https://towardsdatascience.com/hog-histogram-of-oriented-gradients-67ecd887675f>



XIII Seminário de Extensão e Inovação XXVIII Seminário de Iniciação Científica e Tecnológica da UTFPR

Ciência e Tecnologia na era da Inteligência Artificial: Desdobramentos no Ensino Pesquisa e Extensão
20 a 23 de novembro de 2023 - Campus Ponta Grossa, PR



Tabela 1 – Tabela de duração de treinamento segundo quantidade de indivíduos e de imagens por indivíduo.

peessoas imagens	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0,44	1,84	1,47	2,96	2,72	2,61	3,04	3,49	5,21	4,40	4,80	5,24	5,66
2	1,00	4,01	3,10	6,05	6,59	5,19	6,07	6,98	10,36	8,72	9,58	10,44	11,31
3	3,40	5,24	4,43	6,53	11,06	7,78	9,28	10,46	15,29	13,06	14,36	15,66	16,95
4	4,88	5,30	5,91	8,49	13,15	10,35	14,84	13,96	20,20	17,40	19,14	20,87	22,59
5	6,04	9,09	7,32	10,61	16,06	12,95	15,32	17,44	26,32	21,69	23,90	26,08	28,20
6	7,24	7,54	11,81	13,12	19,50	15,55	22,37	20,92	29,64	26,03	28,64	31,27	33,80
7	8,39	6,87	10,34	16,65	23,86	18,13	28,21	24,40	36,23	30,34	33,40	36,44	39,44
8	9,56	8,27	11,66	16,72	20,27	20,71	31,04	27,85	40,70	34,67	38,18	41,59	45,04
9	10,74	8,82	13,13	18,75	27,00	23,29	36,64	31,37	44,33	39,00	42,96	46,81	50,67
10	12,02	9,79	15,71	20,85	25,81	25,85	41,64	34,85	49,65	43,34	47,71	52,01	56,31
11	13,12	10,93	20,25	22,92	28,43	28,44	45,59	38,30	45,73	47,70	52,50	57,20	61,98
12	13,25	14,72	18,30	24,99	35,43	31,03	51,36	44,74	47,05	52,02	57,29	62,37	67,71
13	13,78	14,03	18,87	27,05	28,63	33,63	47,70	53,86	50,91	56,37	62,06	67,50	73,30
14	16,75	14,02	24,07	29,01	30,13	36,21	42,65	61,89	54,78	60,70	66,74	72,75	79,27
15	19,87	14,59	28,50	36,17	32,27	38,82	45,63	85,47	58,86	64,95	71,58	77,88	84,53

Fonte: Autoria Própria.

e, eventualmente, torna-se necessário diminuir o número de imagens por indivíduo.

Para testar a eficácia suficiente do algoritmo durante a etapa de reconhecimento, outros testes foram realizados. Num primeiro experimento, 18 imagens com um dataset de 6 pessoas e 2 indivíduos em cada imagem, um pertencente ao dataset e um desconhecido, houveram 10 atribuições de sobreajuste para o indivíduo desconhecido e 10 más atribuições, totalizando 36% de erro total. Além deste caso, dois outros testes foram realizados, um com 19 imagens e 12 imagens, tendo elas 2 e 4 pessoas por imagem, respectivamente. Para ambos, o dataset continha apenas um usuário que estava presente em todas as fotos, em apenas um caso o indivíduo foi mal classificado como desconhecido e houveram 2 casos de sobreajuste, gerando um erro total de 3%, mostrando que o algoritmo lida melhor com uma quantidade pequena de sujeitos no dataset.

PERSPECTIVAS FUTURAS

Com o intuito de aumentar a utilidade do sistema desenvolvido, é interessante acoplar o sistema de reconhecimento facial à estimação de gestos e segmentação de pessoas. Os algoritmos de segmentação e pose já foram desenvolvidos ⁶.

Outro desenvolvimento de suma importância seria realizar testes de treinamento com outros algoritmos existentes. Embora modelos de redes neurais, no geral, não consigam tirar proveito efetivo da janela de 1 minuto, existem outros algoritmos os quais potencialmente podem obter sucesso, um exemplo seria técnicas de *one-shot learning* (BELO, 2019).

⁶ MediaPipe Track, desenvolvido por Gustavo Fardo e pode ser encontrado em https://github.com/UtBotsAtHome-UTFPR/mediapipe_track



FINANCIAMENTO

Os autores agradecem o suporte da UTFPR - Universidade Tecnológica Federal do Paraná. O primeiro e segundo autores gostariam de agradecer o suporte financeiro fornecido pela CNPq pela bolsa durante a iniciação científica na UTFPR.

CONCLUSÃO

Os robôs de serviço tem crescido em interesse e aplicabilidade, além de forçarem inovações por necessitarem serem executados em hardware embarcado com limitações de processamento. Os materiais utilizados nesta pesquisa correspondem, atualmente, em um equilíbrio entre custo e desempenho. Portanto, as soluções buscam não ultrapassar os limites estabelecidos pelo sistema, mas, ao mesmo tempo, aproveitar todo o hardware ao máximo. Ambos os objetivos foram alcançados.

O algoritmo de reconhecimento facial funciona com qualidade de reconhecimento razoavelmente alta. O maior desafio para o processo de reconhecimento está no período limitado para que se realize o treinamento, portanto, houve um enfoque em maximizar os resultados dentro de um período de 1 minuto. Os índices de reconhecimento quando poucos operadores do sistema estão na *database* foram bons, porém, conforme mais indivíduos são inseridos os índices de sobreajuste a más atribuições tendem a subir.

Disponibilidade de Código

O código fonte está disponível para acesso em <https://github.com/UtBotsAtHome-UTFPR>.

Conflito de interesse

Não há conflito de interesse.

Contribuições

A solução para reconhecimento facial e a análise de tempos de treino, assim como a escrita deste artigo foram realizadas por David Segalle com ajuda de Gustavo Fardo. O desenvolvimento foi supervisionado por João Alberto Fabro.

REFERÊNCIAS

ASSIN, Kortli. et. al. Face Recognition Systems: A Survey. in Special Issue. **Biometric Systems**, 2020.



BELO, J. P. R. Facial Recognition Experiments on a Robotic System Using One-Shot Learning. **2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) e 2019 Workshop on Robotics in Education (WRE)**, 2019.

PALAR, S. F. et. al. Apollo (based on Pioneer 3-AT) Service Robot - UTFPR. **LARC/CBR**, 2022.

SEGALLE, David. Identifying New Persons in the context of the Robocup@Home competition using KNN. **20th IEEE Latin American Robotics Symposium - LARS 2023**, 2023.

T. M, COVER. et al. Nearest Neighbor Pattern Classification. **EEE TRANSACTIONS ON INFORMATION THEORY, VOL. 15**, 1967.

WISSPEINTNER, T. et. al. scientific competition and benchmarking for domestic service robots. **Interaction Studies. v. 10**, 2009.