

T2 - IVC

Cristiano M Martins

27 de junho de 2023

Introdução

Este relatório detalha a segunda atividade prática da disciplina de Introdução a Visão Computacional. Foi desenvolvido um projeto que utiliza o modelo YOLO (versão 4) e a biblioteca OpenCV para detectar objetos em vídeos, extrair suas características e atribuir identificação única a esses objetos, para que seja possível saber quantas pessoas ou carros passaram em determinado horário, por exemplo.

Dados utilizados

Para a realização deste trabalho foram capturados vídeos utilizando uma câmera com resolução de 640x480 pixels a 30 fps. A câmera ficou apontada para uma rua da cidade de Porto Alegre e os vídeos foram gravados durante o dia, quando ocorria alguma movimentação dentro da área visível pela câmera.

Desenvolvimento

O código do projeto foi implementado em um jupyter notebook e foram criadas duas classes para auxiliar no desenvolvimento: ObjectFeatures e ObjectDetector.

A classe ObjectFeatures é responsável por armazenar as características de um objeto detectado, incluindo seu ID, nome da classe, ID da classe e o vetor de características.

A classe ObjectDetector é responsável por carregar o modelo YOLO e as classes de objetos que o modelo pode detectar, bem como realizar o processamento das imagens para realizar a extração e identificação de objetos através da extração de características pelo HOG e cálculo de similaridade do cosseno.

Processo de detecção

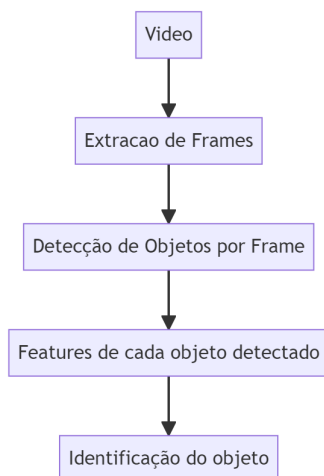


Figura 1: Fluxo de detecção e identificação de objetos

O processo de detecção e identificação funciona da seguinte forma: para cada frame do vídeo, o detector de objetos é executado para identificar os objetos presentes no frame. Para cada objeto detectado, um vetor de características é extraído usando o descritor HOG e com os vetores de características de cada objeto detectado, é realizada a comparação dos vetores de características de cada objeto que já foi detectado nos frames anteriores, conforme a sua própria classe (person - person, car - car).

O processo de detecção de objetos e extração de características é repetido para cada frame do vídeo. O rastreamento de objetos implementado foi pensado para que fosse possível rastrear os objetos ao longo dos frames e determinar se um objeto detectado em um frame é o mesmo objeto detectado em um frame anterior.

A comparação dos vetores de características dos objetos foi feita utilizando a similaridade do cosseno. É realizado o cálculo da similaridade do novo objeto detectado com cada objeto já existente, conforme a classe, e havendo uma diferença maior que 0.3, então o objeto é identificado como um objeto novo no vídeo e é adicionado na lista de objetos detectados.

Resultados

Os resultados obtidos não atingiram um nível de satisfação muito bom. A identificação única dos objetos apresentou desafios significativos, principalmente devido à movimentação dos objetos no vídeo. A simples rotação de um objeto pode alterar a saída do HOG de tal forma que o mesmo objeto pareça diferente. Quando calculada a similaridade do cosseno, é encontrado um ângulo maior do que o esperado entre os dois vetores e isso resulta na identificação do objeto existente como um novo objeto.

Referências

- [1] *YOLO Object Detection (TensorFlow tutorial)*, Available at: <https://youtu.be/4eIBisqx9g>
- [2] *Object Tracking with Opencv and Python*, Available at: <https://youtu.be/O3b8IVF93jU>
- [3] <https://openai.com/research/chatgpt> ChatGPT v4