

Vinícius de Sousa Carvalho
Orientador: Flávio Henrique Duarte de Araújo

Desenvolvimento de uma ferramenta para contagem e monitoramento de pessoas em vídeos

Picos - PI
10 de junho de 2025

Vinícius de Sousa Carvalho
Orientador: Flávio Henrique Duarte de Araújo

Desenvolvimento de uma ferramenta para contagem e monitoramento de pessoas em vídeos

Monografia submetida ao Curso de Bacharelado em Sistemas de Informação como requisito parcial para obtenção de grau de Bacharel em Sistemas de Informação.

Universidade Federal do Piauí
Campus Senador Heuvídio Nunes de Barros
Bacharelado em Sistemas de Informação

Picos - PI
10 de junho de 2025

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca José Albano de Macêdo

C331d Carvalho, Vinícius de Sousa.

Desenvolvimento de uma ferramenta para cotagem e monitoramento de pessoas em vídeos / Vinícius de Sousa Carvalho – 2025.

44 f.

1 Arquivo em PDF.

Indexado no catálogo *online* da biblioteca José Albano de Macêdo-CSHNB
Aberto a pesquisadores, com restrições da Biblioteca.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal do
Piauí, Curso de Bacharelado em Sistemas de Informação, Picos, 2025.

“Orientador: Flávio Henrique Duarte de Araújo”.

1. Sistemas informacionais. 2. Ferramenta computacional. 3. Recuperação
da informação – monitoramento. I. Carvalho, Vinícius de Sousa. II. Araújo,
Flávio Henrique Duarte de. III. Título.

CDD 005.7

Elaborada por Maria Letícia Cristina Alcântara Gomes
Bibliotecária CRB nº 03/1835

DESENVOLVIMENTO DE UMA FERRAMENTA PARA CONTAGEM E RASTREAMENTO DE PESSOAS EM VÍDEOS

VINÍCIUS DE SOUSA CARVALHO

Monografia aprovada como exigência parcial para obtenção do grau de Bacharel em Sistemas
de Informação.

Data de Aprovação

Picos – PI, 27 de Junho de 2025



Documento assinado digitalmente
FLAVIO HENRIQUE DUARTE DE ARAUJO
Data: 27/06/2025 16:19:16-0300
Verifique em <https://validar.iti.gov.br>

Prof. Flávio Henrique Duarte de Araújo



Documento assinado digitalmente
MAURO ANTONIO GUIMARAES CLARK
Data: 27/06/2025 16:50:38-0300
Verifique em <https://validar.iti.gov.br>

Mauro Antônio Guimarães Clark



Documento assinado digitalmente
JEANDERSON GOMES DE SOUSA
Data: 27/06/2025 17:35:37-0300
Verifique em <https://validar.iti.gov.br>

Jeanderson Gomes de Sousa

Agradecimentos

Agradeço a Deus, em primeiro lugar, por todas as bênçãos que me concedeu ao longo da minha jornada. Sua presença constante e sua força me sustentaram nos momentos mais difíceis e me ajudaram a alcançar meus objetivos.

Agradeço também aos meus pais, Francileide Lourença de Sousa e Givaldo João de Carvalho, pelo amor, apoio e educação que me deram. Vocês são a base que me mantém firme e me fazem ser a pessoa que sou hoje.

Aos meus professores, que compartilharam seus conhecimentos e foram fundamentais para o desenvolvimento deste trabalho, em especial, ao meu professor orientador, Flávio Henrique Duarte de Araújo, por ter me guiado e contribuído significativamente para o meu crescimento, tanto na carreira acadêmica quanto na profissional. Sua orientação foi essencial para o meu desenvolvimento pessoal e profissional.

Agradeço, de coração, à minha namorada, Raglicia Sandi Ribeiro de Sousa, por todo o amor, apoio e paciência durante essa jornada. Sua presença constante, suas palavras de incentivo e sua compreensão nos momentos mais difíceis foram essenciais para que eu seguisse firme até aqui. Sou grato por cada gesto de carinho e por acreditar em mim mesmo quando eu duvidei. Este trabalho também carrega um pouco de tudo o que vivemos juntos.

Agradeço aos meus amigos Anderson Geovane, José Filho, Luís Clício, Lucas Rabelo e Maria Luísa, que sempre estiveram ao meu lado, oferecendo apoio, incentivo e amizade ao longo dessa caminhada. Cada um de vocês, de alguma forma, contribuiu para que eu chegasse até aqui, seja com uma palavra de motivação, uma ajuda nos momentos de dificuldade ou simplesmente com a presença acolhedora.

Gostaria de fazer um agradecimento especial a José Wanderlei, que foi meu parceiro de trabalho ao longo de todo o curso. Sua colaboração, comprometimento e companheirismo foram fundamentais para o desenvolvimento de várias atividades e para o sucesso dos nossos projetos em dupla. Juntos, superamos desafios e conquistamos grandes momentos de aprendizado.

Por fim, agradeço a mim mesmo pela dedicação, persistência e resiliência ao longo dessa jornada. Foram muitos desafios, momentos de dúvida e superação, mas, acima de tudo, houve um compromisso inabalável com o meu crescimento pessoal e acadêmico. Reconheço o esforço investido, as noites em claro e as escolhas que fiz para chegar até aqui. Cada obstáculo foi uma oportunidade de aprendizado, e cada vitória, uma confirmação do meu potencial. Agradeço por nunca ter desistido de meus objetivos, mesmo quando o caminho parecia difícil, e por sempre buscar a excelência em tudo o que faço.

Aqueles que são loucos o suficiente para pensar que podem mudar o mundo são os que o fazem.

Steve Jobs

Resumo

A gestão da segurança e do fluxo de pessoas em ambientes públicos representa um desafio contínuo, com implicações significativas para a segurança pública e o planejamento urbano. Em locais como estações de transporte e centros comerciais, onde a concentração de indivíduos é alta, a eficácia na gestão da segurança e do fluxo de pessoas é essencial para garantir a ordem e o bem-estar dos cidadãos. Dessa forma, a implementação de tecnologias de monitoramento e controle, especialmente a vigilância por vídeo, surge como uma solução crucial. Essa estratégia tem um papel relevante no gerenciamento e proteção desses ambientes, permitindo a prevenção de incidentes, além da análise detalhada de padrões de movimento. Apesar da crescente adoção da vigilância por vídeo, persistem desafios significativos na contagem e rastreamento de pessoas em vídeo. A contagem manual, muitas vezes empregada, é suscetível a erros e não consegue lidar eficientemente com o volume de pessoas em áreas urbanas movimentadas. Essa limitação compromete não apenas a segurança, mas também o gerenciamento do fluxo de pedestres, levando a congestionamentos, dificuldades na prevenção de incidentes e prejuízos para a economia local, especialmente em eventos de grande porte. A falta de uma solução automatizada para monitoramento é, portanto, uma lacuna crítica na segurança e no planejamento urbano. Neste trabalho, foi desenvolvida uma solução para a contagem e o monitoramento de pessoas em vídeos, com o objetivo de extrair informações relevantes como o número de indivíduos, sua velocidade e o tempo de permanência nas cenas monitoradas. Para isso, foi utilizado o modelo de detecção YOLOv8 Nano, aliado ao algoritmo de rastreamento BoT-SORT. A ferramenta foi desenvolvida em Python com uso das bibliotecas OpenCV e PyQt5, e permite o processamento de vídeos armazenados ou em tempo real. Foram implementados módulos para extração de métricas como contagem de pessoas, tempo de permanência e velocidade de deslocamento. Os testes foram realizados com vídeos em diferentes cenários, utilizando delimitação manual de área de interesse para avaliação do desempenho do sistema. Os resultados obtidos indicam que o modelo YOLOv8 Nano apresentou desempenho satisfatório na detecção de pessoas, mesmo em cenários variados. Na validação da métrica de tempo de permanência, a ferramenta apresentou um Erro Absoluto Médio (MAE) de 0,1789 segundos e um Erro Quadrático Médio (MSE) de 0,0423 segundos, demonstrando precisão na extração temporal. A ferramenta desenvolvida foi capaz de extrair métricas como contagem de pessoas, tempo de permanência e velocidade de deslocamento com eficácia, contribuindo para a análise do comportamento humano em ambientes monitorados. Além disso, foi implementado um sistema com interface gráfica intuitiva, que permite o monitoramento e contagem de pessoas tanto em vídeos pré-gravados quanto em tempo real.

Palavras-chaves: Contagem de Pessoas. Desenvolvimento de Software. Visão Computacional.

Abstract

The management of safety and pedestrian flow in public environments remains an ongoing challenge with significant implications for public security and urban planning. In locations such as transportation stations and shopping centers, where large crowds gather, efficient safety and flow management is essential to maintain order and ensure citizens' well-being. In this context, the implementation of monitoring and control technologies—particularly video surveillance—emerges as a crucial solution. This strategy plays a key role in managing and securing such environments, enabling both incident prevention and detailed analysis of movement patterns.

Despite the increasing adoption of video surveillance, significant challenges remain in accurately counting and tracking individuals in video footage. Manual counting, still often used, is prone to errors and cannot efficiently handle high volumes of people in busy urban areas. This limitation affects not only safety but also pedestrian flow management, leading to congestion, hindered incident prevention, and economic losses—especially during large-scale events. The lack of an automated monitoring solution represents a critical gap in urban safety and planning.

This work presents a solution for counting and monitoring individuals in video feeds, aiming to extract relevant information such as the number of people, their speed, and the duration of their presence within monitored scenes. The proposed tool leverages the YOLOv8 Nano detection model in conjunction with the BoT-SORT tracking algorithm. Developed in Python using the OpenCV and PyQt5 libraries, the system supports both stored and real-time video processing. Modules were implemented for extracting metrics such as person count, dwell time, and movement speed. Testing was conducted using videos from various scenarios, with manually defined areas of interest for performance evaluation.

The results demonstrate that the YOLOv8 Nano model achieved satisfactory performance in detecting people across different environments. For dwell time validation, the tool achieved a Mean Absolute Error (MAE) of 0.1789 seconds and a Mean Squared Error (MSE) of 0.0423 seconds, indicating precision in temporal analysis. The developed system effectively extracted metrics such as person count, dwell time, and movement speed, contributing to the analysis of human behavior in monitored environments. Additionally, an intuitive graphical user interface was implemented, allowing for both real-time and pre-recorded video monitoring and person counting.

Lista de ilustrações

Figura 1 – Etapas fundamentais da visão computacional.	14
Figura 2 – funcionamento do YOLO na detecção de objetos. Imagem adaptada do trabalho de (REDMON et al., 2016)	17
Figura 3 – Comparação da forma da caixa delimitadora entre o filtro de Kalman tradicional (em azul tracejado) e o filtro proposto pelo BoT-SORT (em verde). O filtro proposto resulta em melhor ajuste da caixa ao objeto. Os círculos em vermelho destacam a região dos pés, evidenciando que o BoT-SORT consegue acompanhar com mais precisão o contorno do corpo durante a locomoção. Imagem adaptada do trabalho de (AHARON; ORFAIG; BOBROVSKY,).	19
Figura 4 – Fluxograma representando as etapas do método proposto.	25
Figura 5 – Tela inicial do sistema.	31
Figura 6 – Tela de selecionar vídeo do sistema.	31
Figura 7 – Tela de selecionar área de rastreo.	32
Figura 8 – Tela de processamento do vídeo.	33
Figura 9 – Tela de resultados.	34
Figura 10 – Tela de inserção de campos para webcam.	35
Figura 11 – Tela de inserção de campos para câmera RTSP.	35
Figura 12 – Tela de resultados em tempo real.	36

Lista de tabelas

Tabela 1 – Trabalhos relacionados	23
Tabela 2 – Avaliação do tempo de permanência por vídeo	28
Tabela 3 – Requisitos Não Funcionais	29
Tabela 4 – Requisitos Funcionais	30

Lista de abreviaturas e siglas

IBGE	Instituto Brasileiro de Geografia e Estatística
GUI	Interface Gráfica do Usuário (<i>Graphical User Interface</i>)
RTSP	<i>Real Time Streaming Protocol</i>
CCTV	Televisão de Circuito Fechado (<i>Closed-Circuit Television</i>)
YOLO	<i>You Only Look Once</i>
CNN	Rede Neural Convolucional (<i>Convolutional Neural Network</i>)
HOG	<i>Histogram of Oriented Gradients</i>
COCO	<i>Common Objects in Context</i>
SORT	<i>Simple Online and Realtime Tracking</i>
BoT-SORT	<i>Bag of tricks for SORT</i>
ReID	Reidentificação de objetos (<i>Re-Identification</i>)
FPS	Quadros por Segundo (<i>Frames Per Second</i>)
IoU	Intersecção sobre a união (<i>Intersection over Union</i>)
mAP50	Precisão média calculada com um limiar de intersecção sobre união (IoU) de 0,50 (<i>Mean average precision calculated at an intersection over union (IoU) threshold of 0.50</i>)
MOTA	<i>Multiple Object Tracking Accuracy</i>
MOTP	<i>Multiple Object Tracking Precision</i>
NINTEC	Núcleo de Inovação Tecnológica
RF	Requisito Funcional
RNF	Requisito Não Funcional
MAE	Erro Absoluto Médio (<i>Mean Absolute Error</i>)
RMSE	Raiz do Erro Quadrático Médio (<i>Root Mean Square Error</i>)
mAP	<i>Mean Average Precision</i>
mAP50-95	Média da precisão média (<i>mean Average Precision</i>) calculada em diferentes limiares de IoU, variando de 0,50 a 0,95.

Sumário

1	Introdução	12
1.1	Objetivos Gerais e Específicos	13
1.2	Disposição do Trabalho	13
2	Referencial Teórico	14
2.1	Visão Computacional	14
2.2	Aprendizado profundo	15
2.3	Redes Neurais Convolucionais	15
2.4	Redes de detecção de objeto	16
2.4.1	<i>You Only Look Once</i> (YOLO)	16
2.5	Rastreamento de objetos	17
2.5.1	BoT-SORT	18
3	Trabalhos Relacionados	20
4	Metodologia Proposta	24
4.1	Entrada do vídeo	24
4.2	Detecção de Pessoas	25
4.3	Rastreamento de Pessoas	26
4.4	Extração de Métricas	26
4.4.1	Contagem de Pessoas	27
4.4.2	Tempo de permanência	27
4.4.3	Velocidade de Deslocamento	27
5	Resultados e Discussão	28
5.1	Validação das métricas extraídas	28
5.2	Requisitos Funcionais e Não Funcionais	29
5.3	Funcionalidades do sistema	29
6	Conclusão	37
6.1	Trabalhos futuros	37
7	Registros e Publicações	39
	Referências	40

1 Introdução

Garantir a segurança e a circulação adequada de pedestres tornou-se uma prioridade crucial, motivada pela crescente necessidade de lidar com desafios relacionados à segurança pública e ao planejamento urbano. De acordo com dados do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2022, 54,6% das pessoas não se sentiam seguras em suas cidades (IBGE, 2022). Embora a segurança muitas vezes dependa de agentes físicos, como policiais ou seguranças, em certos casos, os custos são altos demais. Assim, a videovigilância surge como uma alternativa viável e eficaz, proporcionando não apenas monitoramento remoto, mas também uma gestão otimizada dos recursos humanos envolvidos (DUQUE, 2009).

Por meio da vigilância por vídeo, é possível extrair informações relevantes para uma variedade de aplicações. O rastreamento e contagem de pessoas desempenha um papel fundamental em muitas aplicações práticas, incluindo a vigilância por vídeo e o gerenciamento de tráfego de pedestres (XU; LV; MENG, 2010). A estimativa do número de pessoas presentes em uma área pode ser uma informação útil por várias razões, dentre elas, segurança e otimização. Uma mudança anômala no número de pessoas pode ser a causa ou o efeito de um evento perigoso. Além disso, essa variação pode ser fundamental para a otimização do cronograma de um sistema de transporte público, permitindo ajustes com base na demanda de passageiros (CORDEIRO et al., 2019).

Apesar dos avanços na vigilância por vídeo, persistem dificuldades na análise manual das imagens capturadas. Seres humanos não são bons em detectar eventos, devido ao fato de não conseguirem manter a concentração em uma atividade durante longos períodos de tempo, acarretando em um aumento de falhas no processamento de vigilância (GARCIA, 2007). Isso pode resultar em dificuldades na identificação de comportamentos suspeitos ou na resposta rápida a incidentes de segurança. Para auxiliar no monitoramento de segurança, é necessário automatizar a tarefa de videovigilância (NASRY; EZZAHOUT; OMARY, 2023). A visão computacional emerge como uma solução promissora para enfrentar esse desafio.

A visão computacional destaca-se na resolução de problemas por meio do reconhecimento de padrões em imagens. Na literatura foram encontrados trabalhos relacionados que utilizam imagens de vídeo para contar a quantidade de pessoas que entram e saem de uma determinada região, como é observado em (VASAVI et al., 2023) e (NASRY; EZZAHOUT; OMARY, 2023). Apesar da disponibilidade de trabalhos na literatura que exploram a detecção de indivíduos em imagens por meio de técnicas de visão computacional, encontrar uma ferramenta gratuita, com contagem e rastreamento de pessoas por meio de uma interface gráfica intuitiva, ainda é uma lacuna significativa.

1.1 Objetivos Gerais e Específicos

Com base no exposto acima, o objetivo deste trabalho é criar um sistema que possibilite a contagem e monitoramento de pessoas em vídeos, por meio de uma Interface Gráfica de Usuário (Graphical User Interface - GUI). Com o intuito de alcançar esse objetivo geral, definiram-se os seguintes objetivos específicos:

1. Implementar técnicas de visão computacional para extrair automaticamente métricas de contagem de pessoas, tempo de permanência e velocidade de deslocamento a partir dos vídeos.
2. Criar uma interface gráfica para a visualização dos resultados e realizar a integração da mesma com os algoritmos de visão computacional desenvolvidos.
3. Validar a precisão e a eficácia do sistema por meio de testes em diferentes cenários.

Com base nisso, as principais contribuições deste trabalho foram: o desenvolvimento de uma ferramenta computacional capaz de realizar a contagem e o rastreamento de pessoas em vídeos; a aplicação de técnicas de visão computacional para extração de métricas, como tempo de permanência e velocidade; e a integração de todos os módulos em uma interface gráfica funcional e de fácil uso. Além disso, o sistema oferece suporte tanto para vídeos armazenados quanto para transmissões em tempo real, demonstrando sua aplicabilidade em diversos contextos de monitoramento.

1.2 Disposição do Trabalho

Este trabalho está estruturado em sete capítulos, além da introdução. O Capítulo 2 aborda os conceitos e técnicas fundamentais que sustentam o desenvolvimento da pesquisa, bem como as métricas utilizadas. No Capítulo 3, realiza-se uma breve revisão da literatura, destacando os principais trabalhos relacionados ao tema. O Capítulo 4 detalha a metodologia empregada, descrevendo as abordagens adotadas em cada etapa do desenvolvimento. Já o Capítulo 5 apresenta os resultados alcançados, acompanhados de informações relevantes para sua análise. O Capítulo 6 reúne as conclusões do estudo e propõe possíveis direções para pesquisas futuras. Por fim, o Capítulo 7 reúne as publicações e registros gerados a partir dos resultados obtidos.

2 Referencial Teórico

Esta seção apresenta os conceitos fundamentais para o desenvolvimento do projeto. Na Subseção 2.1, é introduzida a visão computacional, delineando seu escopo e aplicações. A Subseção 2.2 aborda os princípios do aprendizado profundo, seguido pela Subseção 2.3, que detalha as redes neurais convolucionais, fundamentais para o processamento de imagens. A Subseção 2.4 discute as redes de detecção de objetos, com ênfase no método You Only Look Once (YOLO) na Subseção 2.4.1, destacando sua arquitetura e desempenho. Por fim, a Subseção 2.5 trata do rastreamento de objetos, com foco no método BoT-SORT na Subseção 2.5.1, evidenciando suas técnicas de associação e atualização de trajetórias.

2.1 Visão Computacional

A visão computacional refere-se à capacidade de um sistema computacional entender e interpretar o mundo ao seu redor através de imagens adquiridas por câmeras de vídeo, sensores, *scanners* e outros dispositivos, permitindo a extração de informações relevantes (BALLARD; BROWN, 1982). O objetivo da visão computacional é tomar decisões úteis sobre objetos e cenas reais com base em imagens captadas (STOCKMAN; SHAPIRO, 2001). A visão computacional proporciona ao computador uma vasta gama de informações precisas derivadas de imagens e vídeos, capacitando-o a realizar tarefas inteligentes e a simular, assim, a inteligência humana. (MILANO; HONORATO, 2010). A Figura 1 apresenta as etapas fundamentais da visão computacional.

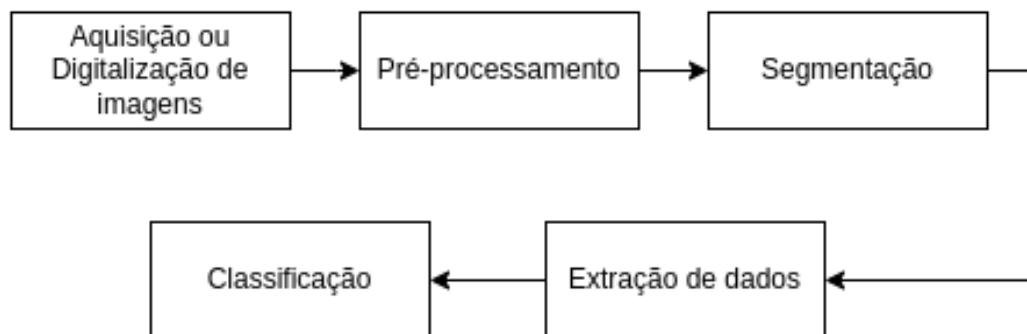


Figura 1 – Etapas fundamentais da visão computacional.

O processo de visão computacional é composto por cinco etapas fundamentais que são: aquisição ou digitalização de imagens, pré-processamento, segmentação, extração de dados e classificação (REHEM; TRINDADE, 2009). A aquisição de imagens é o primeiro passo na visão computacional, envolvendo a captura, armazenamento e transmissão de imagens. O pré-processamento visa melhorar a qualidade da imagem, reduzindo ruídos e

imperfeições. Na segmentação, a imagem é dividida em regiões de interesse, separando a região do objeto e o fundo da imagem. Na extração de dados, informações específicas são extraídas das regiões segmentadas, como forma, cor ou textura. Por fim, na classificação, os dados são categorizados conforme critérios predefinidos (BORTH et al., 2014).

2.2 Aprendizado profundo

O aprendizado profundo é uma forma de explorar várias camadas de processamento de informações não-lineares por meio de técnicas de aprendizado de máquina, de maneira que os algoritmos consigam aprender múltiplos níveis de representação e abstração, dando sentido nos dados (DENG; YU et al., 2014).

Métodos que utilizam aprendizado profundo buscam descobrir um modelo, como regras, parâmetros ou padrões, a partir de um conjunto de dados (exemplos). Utilizam também um método para guiar o aprendizado do modelo com base nesses exemplos. O objetivo é encontrar padrões complexos em diferentes tipos de entradas, como imagens, textos ou áudios. Além disso, busca-se relacionar a saída desses exemplos com padrões presentes no conjunto de entrada. Ao final do processo de aprendizado, obtém-se uma função capaz de receber dados brutos como entrada e fornecer, como saída, uma representação adequada para o problema em questão (PIRES, 2017).

2.3 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNN) são arquiteturas de aprendizado profundo voltadas ao processamento de imagens, capazes de extrair automaticamente padrões espaciais relevantes por meio da aplicação de filtros convolucionais. Diferentemente das redes totalmente conectadas, elas reduzem a complexidade computacional ao aproveitar a estrutura local das imagens, reutilizando pesos e diminuindo o número de parâmetros (SAHA, 2018). Uma vantagem é a possibilidade de visualizar mapas de ativação, tornando o processo mais interpretável que em outras redes neurais (BROWNLEE, 2016).

Sua arquitetura é composta por três camadas principais: convolucionais, pooling e totalmente conectadas. As camadas convolucionais extraem características locais, as de pooling reduzem a dimensão preservando informações importantes, e as totalmente conectadas realizam a classificação final. Cada uma possui hiperparâmetros, como tamanho do filtro e número de neurônios, definidos antes do treinamento (PEŠEK, 2018). A principal aplicação das CNNs está no processamento visual, especialmente imagens, já que a convolução filtra essas informações considerando sua estrutura bidimensional (PONTI; COSTA, 2018).

2.4 Redes de detecção de objeto

Detecção de objetos em imagens é uma tarefa de visão computacional que visa localizar instâncias de objetos de uma determinada classe em uma imagem digitalizada (SILVA et al., 2018). A detecção de objetos envolve dois processos principais: a identificação das instâncias de objetos de uma classe específica (como carros, pessoas, animais) e a determinação de suas posições nas imagens através de caixas delimitadoras. Na localização de objetos, a imagem de entrada é processada pelo algoritmo e a saída é a classe do objeto e a sua localização determinada a partir do centro da caixa de contorno, que é definido por 'x' e 'y' e a largura e a altura definidas por w e h, respectivamente. A detecção de objetos possibilita a classificação e a localização de mais de um objeto por imagem (SILVA, 2021).

Historicamente, técnicas de detecção foram propostas para serem capazes de resolver problemas relacionados a visão computacional, como o método de detecção por seguimento de bordas proposto por (YAKIMOVSKY, 1976), a transformada de Hough (HOUGH,) e características baseadas em gradientes (HOG) (DALAL; TRIGGS, 2005). No entanto, com o advento das redes neurais profundas, houve um avanço significativo na área. Segundo Liu et al. (2016), as redes neurais convolucionais (CNNs) revolucionaram a detecção de objetos, permitindo a extração automática de características visuais hierárquicas diretamente das imagens.

2.4.1 *You Only Look Once* (YOLO)

You Only Look Once (YOLO) (REDMON et al., 2016) é um método de detecção de objetos de fase única que utiliza uma rede neural convolucional (CNN) para extrair características e gerar propostas de regiões de interesse (MARTINS, 2021). O modelo YOLO realiza a detecção em etapa única, o que o diferencia de outros métodos. Essa abordagem proporciona uma velocidade significativamente maior, sem comprometer a precisão em comparação com técnicas concorrentes. Com esse modelo, é possível analisar uma imagem para prever quais objetos estão presentes e onde estão localizados.

A ideia principal do modelo YOLO consiste em dividir uma imagem em quadrículas e utilizar as mesmas como referência para gerar as caixas delimitadoras com valores de confiança e probabilidades de classe associadas (MARTINS, 2021). Uma única rede convolucional prevê simultaneamente múltiplas caixas delimitadoras e probabilidades de classe para essas caixas. O treinamento do YOLO é realizado em imagens completas, visando otimizar diretamente o desempenho da detecção. Essa abordagem simplificada resulta em uma velocidade excepcionalmente rápida, pois trata a detecção de objetos como um problema de regressão, dispensando a necessidade de um pipeline complexo (REDMON et al., 2016). A Figura 2 ilustra o processo de divisão da imagem em grades e a geração de caixas delimitadoras com probabilidades associadas a cada classe.

Além da velocidade, a arquitetura do YOLO traz melhorias na generalização da detec-

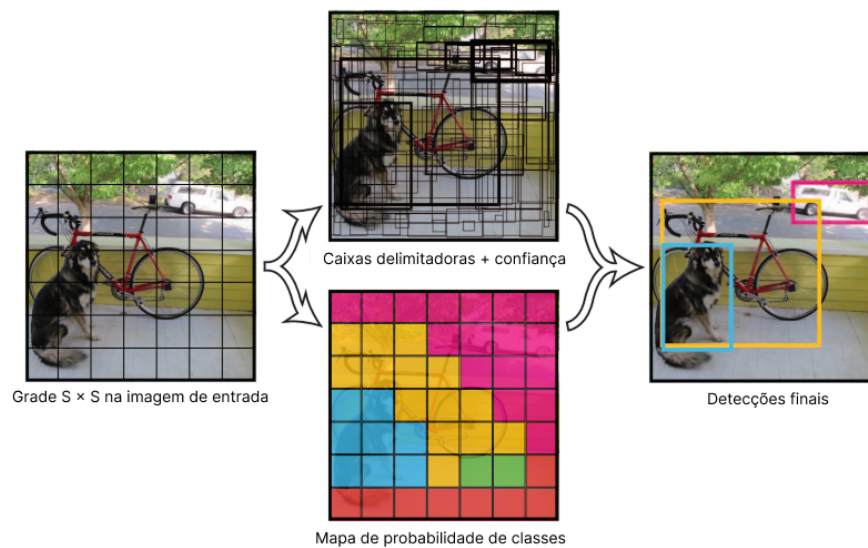


Figura 2 – funcionamento do YOLO na detecção de objetos. Imagem adaptada do trabalho de (REDMON et al., 2016)

ção para novas imagens, sendo eficaz mesmo em contextos onde os objetos aparecem em diferentes escalas ou posições. Contudo, um desafio observado nas primeiras versões da arquitetura é a dificuldade em detectar objetos pequenos, principalmente em cenários onde há muitos objetos próximos entre si. Isso ocorre porque cada célula da grade é responsável por prever, no máximo, duas caixas delimitadoras e apenas uma classe por célula, o que limita a capacidade do modelo em lidar com alta densidade de objetos (REDMON et al., 2016).

2.5 Rastreamento de objetos

O monitoramento visual de objetos consiste em basicamente estabelecer relações temporais entre objetos alvo de quadros consecutivos de vídeos, ou seja, identificar na sequência de quadros de um vídeo, a localização dos objetos alvo detectados no quadro anterior (COSTA et al., 2008). Os métodos de rastreamento tentam modelar a relação entre a aparência do objeto de interesse e os valores dos pontos da imagem onde ele está presente para estimar a localização do objeto em todos os quadros do vídeo (CAMPOS, 2017). O rastreamento de objetos é um componente essencial dos sistemas de vigilância por vídeo, permitindo a monitorização contínua e a análise de movimentos de objetos de interesse ao longo do tempo (MISHRA; SAROHA, 2016).

Os métodos de rastreamento e detecção de objetos podem ser classificados em dois grupos principais: métodos unificados de rastreamento e detecção, e métodos de rastreamento por detecção. Entre os métodos unificados de rastreamento e detecção, destacam-se os métodos baseados em filtros, que utilizam técnicas como o Filtro de Kalman (KALMAN; BUCY, 1961) e Filtros de Partículas para o processo de rastreamento. Os métodos

de rastreamento por detecção incluem métodos baseados em detecção de fundo, que utilizam técnicas como a diferença de quadros e a subtração de fundo para detectar objetos. Esses métodos são simples e apresentam desempenho robusto em aplicações em tempo real ([MIRZAEI et al., 2023](#)).

2.5.1 BoT-SORT

O *BoT-SORT* (Bag of Tricks SORT) é um rastreador de múltiplos objetos que segue a abordagem *tracking-by-detection*, ou seja, de rastreamento por detecção, na qual objetos são detectados quadro a quadro e suas posições são associadas ao longo do tempo. Essa técnica utiliza um filtro de Kalman para prever as posições futuras dos objetos e métricas como IoU (Intersection over Union) ou similaridade de aparência (ReID) para associá-los entre os quadros. O BoT-SORT surge como uma evolução dos métodos SORT e DeepSORT, propondo melhorias na modelagem de movimento, compensação de movimento da câmera e na fusão de pistas visuais ([AHARON; ORFAIG; BOBROVSKY,](#)).

Entre as principais inovações do BoT-SORT está a redefinição do vetor de estado do filtro de Kalman para prever diretamente largura e altura da caixa delimitadora, resultando em melhor ajuste ao objeto. Como mostrado na Figura 3, essa modificação proporciona caixas delimitadoras mais ajustadas à forma real do objeto em comparação ao filtro tradicional utilizado por rastreadores como o DeepSORT. Além disso, o método introduz a Compensação de Movimento da Câmera (CMC), utilizando transformações afins para corrigir deslocamentos causados por movimentações da câmera ([AHARON; ORFAIG; BOBROVSKY,](#)).



Figura 3 – Comparação da forma da caixa delimitadora entre o filtro de Kalman tradicional (em azul tracejado) e o filtro proposto pelo BoT-SORT (em verde). O filtro proposto resulta em melhor ajuste da caixa ao objeto. Os círculos em vermelho destacam a região dos pés, evidenciando que o BoT-SORT consegue acompanhar com mais precisão o contorno do corpo durante a locomoção. Imagem adaptada do trabalho de ([AHARON; ORFAIG; BOBROVSKY](#),).

3 Trabalhos Relacionados

Esta seção apresenta uma revisão da literatura existente, com objetivo de examinar de forma abrangente a produção científica recente na área de contagem e rastreamento de pessoas em vídeos. Para isso, foram utilizadas bases de dados acadêmicas reconhecidas, empregando termos de busca relevantes e critérios de inclusão predefinidos. A revisão destaca as contribuições relevantes para o tema deste projeto, identificando lacunas e oportunidades para pesquisas futuras, proporcionando uma visão abrangente e atualizada do estado da arte nesse campo de estudo. A Tabela 1 apresenta uma visão geral dos trabalhos relacionados analisados neste estudo. Cada trabalho é descrito em termos de 4 critérios: qual o objetivo do estudo, quais técnicas de detecção foram utilizadas, métricas geradas por meio da detecção e presença de interface gráfica.

O trabalho de [Vasavi et al. \(2023\)](#) apresenta a aplicação da rede de detecção YOLO para desenvolver um método de contagem de pessoas em vídeos de circuito fechado de televisão (CCTV). A metodologia envolveu a coleta de um conjunto de dados com 10.000 imagens, seguido do treinamento do modelo YOLOv3. Foram implementados algoritmos de rastreamento e contagem de pessoas. Os resultados apresentaram uma precisão média de 0,92 e *recall* de 0,89. A margem de erro média na contagem foi de apenas 1,2 pessoas por quadro, o que indica a eficácia do sistema na contagem precisa de indivíduos em diversos cenários. A análise visual dos resultados demonstrou que o modelo YOLO identificou e rastreou corretamente as pessoas nas imagens de CCTV, com áreas delimitadas precisas ao redor delas. Esses resultados destacam a viabilidade e eficácia do algoritmo, tornando-o útil para controle de multidões e monitoramento de segurança e análise de varejo.

O estudo de [Xu, Lv e Meng \(2010\)](#) apresenta um sistema de contagem de pessoas em vídeos de vigilância focado na detecção e rastreamento de cabeça-ombro. O sistema realiza contagem de entrada e saída de pessoas em áreas específicas. Composto por quatro módulos (extração de primeiro plano, detecção de componentes cabeça-ombro, rastreamento e análise de trajetória), o método mostrou ser robusto diante de variações de iluminação e útil em ambientes reais, como supermercados, bancos e ônibus. Os resultados experimentais revelaram taxas de detecção de sucesso entre 85% e 95%, com ocorrência de detecções incorretas principalmente devido a oclusões severas. A eficácia do sistema foi validada em várias sequências de vídeo, evidenciando a viabilidade e precisão do método proposto em cenários de curto prazo.

O trabalho de [Karagulian et al. \(2023\)](#) propõe um algoritmo de visão computacional para analisar imagens de vídeos da Estação Central em Milão, Itália. O objetivo do modelo é estimar a velocidade, direção e densidade de pedestres. Foram analisados vídeos da Pizza Duca d'Aosta, gravados em abril de 2022, com câmeras capturando em MP4 em HD a 20 fps. O modelo utilizado foi a YOLOv3 com a rede neural Darknet-53. Os

resultados mostraram uma precisão média de detecção de pedestres de cerca de 61% medida pelo mAP50, com uma confiança de detecção em torno de 67,6%. O rastreamento de múltiplos objetos apresentou uma média de MOTA (*Multiple Object Tracking Accuracy*) de aproximadamente 56%, indicando uma precisão na detecção de 56%. A média de MOTP (*Multiple Object Tracking Precision*) foi de cerca de 0,6, refletindo uma boa precisão espacial com baixa média de erro entre as posições reais e previstas.

O artigo de [Mokayed et al. \(2022\)](#) apresenta um sistema de detecção e contagem de pessoas em tempo real, utilizando técnicas de visão computacional e aprendizado profundo. O sistema inclui uma interface gráfica que permite aos usuários visualizarem e interagirem com os resultados de maneira intuitiva. A metodologia empregada envolve o modelo YOLOv3 para a detecção e classificação de pessoas e o algoritmo DeepSORT, responsável por rastrear os indivíduos ao longo dos quadros de vídeo. Os testes revelaram uma precisão de 91,07% para o modelo YOLOv3, com uma média de 6,6 quadros por segundo (FPS). Já o modelo YOLOv3-tiny, uma versão mais leve e rápida do YOLOv3, alcançou uma precisão de 76,8%, com uma média de 32 FPS. Este modelo destaca-se pela maior velocidade de processamento, sendo ideal para cenários onde a rapidez é crucial. Os resultados evidenciam a eficácia e eficiência do sistema proposto na detecção e contagem de indivíduos em tempo real.

O artigo de [Telicko e Jakovics \(2023\)](#) apresenta um estudo comparativo entre as redes de detecção YOLOv8 e Mask-RCNN para a contagem de pessoas em imagens de câmeras com lente olho de peixe. A proposta do estudo é avaliar qual abordagem oferece melhor desempenho para a contagem de pessoas em ambientes internos. A metodologia envolveu a geração de um conjunto de dados personalizado a partir de diferentes câmeras de ângulos de visão distintos. O pré-processamento envolveu duas etapas principais: a geração de um conjunto de dados contendo imagens capturadas por câmeras olho de peixe e a correção da distorção da lente dessas câmeras. Os resultados obtidos mostraram que o YOLOv8-N, uma versão mais leve da YOLOv8, foi a melhor solução, com uma precisão de contagem de pessoas de 95% e acurácia de 95,84%. Já o modelo Mask-RCNN, obteve acurácia de 77,2% e precisão de 52,2%. Dessa forma, este trabalho se mostrou eficaz na contagem de pessoas em ambientes internos, evidenciando a maior eficácia do YOLOv8-N em comparação com o Mask-RCNN.

O artigo de [McCarthy et al. \(2025\)](#) apresenta um estudo comparativo entre dois contextos de aplicação de um sistema automático de contagem de pessoas baseado em vídeo: com câmeras instaladas dentro de ônibus (on-bus) e fora dos ônibus, em pontos de embarque (off-bus). O objetivo do estudo é avaliar qual configuração oferece melhor desempenho na contagem de passageiros em serviços de transporte público, especialmente em situações de substituição ferroviária por ônibus. A metodologia envolveu a implementação de um sistema baseado em redes neurais convolucionais (CNNs) para detecção e rastreamento de pessoas utilizando vídeo RGB comum, com detecção por MobileNetv3-SSD, rastrea-

mento e reidentificação. Foram realizados dois testes de campo, totalizando 152 horas de vídeo, com 95 horas utilizadas na análise quantitativa. Os resultados demonstraram que a abordagem off-bus alcançou maior acurácia na contagem de passageiros (96,6%) em comparação à abordagem on-bus (91,8%). Em termos de métricas de erro, o on-bus apresentou Erro Absoluto Médio (MAE) de 2,65 e Raiz do Erro Quadrático Médio (RMSE) de 4,75, enquanto o off-bus apresentou MAE de 4,91 e RMSE de 7,66, refletindo o maior número médio de passageiros por evento no cenário off-bus. Os autores concluíram que o sistema é eficaz em ambos os contextos, mas que a instalação fora do ônibus tende a oferecer maior estabilidade e acurácia, especialmente por reduzir o impacto de oclusões e permitir melhor posicionamento da câmera.

Com base nos trabalhos relacionados, observa-se que, embora existam diversas abordagens voltadas para a detecção e rastreamento de pessoas em vídeo, poucas integram esses recursos a uma interface gráfica de fácil uso. Além disso, são raros os projetos que unem, em uma única solução, a aplicação do modelo YOLOv8 com a geração de métricas detalhadas, como contagem de pessoas, velocidade e tempo de permanência. Diante disso, o objetivo deste projeto é desenvolver uma ferramenta com interface gráfica intuitiva, que reúna essas funcionalidades em um único sistema e que possa ser disponibilizada gratuitamente para diversas aplicações.

Tabela 1 – Trabalhos relacionados

Trabalho	Objetivo	Técnica utilizada	Métrica	Interface Gráfica
Vasavi et al. (2023)	Algoritmo para contagem de pessoas em vídeos de CCTV	YOLOv3	Contagem de pessoas	Não
(XU; LV; MENG, 2010)	Contagem de pessoas para videovigilância	HOG	Contagem de pessoas	Sim
(KARAGULIAN et al., 2023)	Algoritmo para análise de vídeos de pessoas em espaços públicos	YOLOv3	Contagem, velocidade, direção de movimento e densidade de pessoas	Não
Mokayed et al. (2022)	Sistema de contagem de pessoas em tempo real	YOLOv3, DeepSORT	Contagem de pessoas	Sim
Telicko e Jakovics (2023)	Contagem de pessoas em câmeras com lente olho de peixe	YOLOv8, Mask-RCNN	Contagem de pessoas	Não
McCarthy et al. (2025)	Comparar desempenho de contagem automática on-bus e off-bus	MobileNetv3-SSD, ReID	Contagem de passageiros	Não
Este trabalho	Contagem e rastreamento de pessoas em vídeo	YOLOv8	Contagem, velocidade, tempo de permanência no vídeo	Sim

4 Metodologia Proposta

Esta seção apresenta a metodologia empregada no trabalho, bem como conceitos importantes para o entendimento dos processos e resultados. Inicialmente, realizou-se o estudo e compreensão das ferramentas empregadas, visando adquirir o conhecimento necessário sobre as tecnologias e ferramentas que serão utilizadas para conduzir os experimentos e desenvolver a ferramenta proposta. As tecnologias e ferramentas estudadas nesta etapa foram: análise e processamento de vídeo (Python, OpenCV), detecção de objetos (YOLO), métodos de rastreamento de objetos em vídeos (BoT-SORT), ferramentas de desenvolvimento de interface gráfica (PyQt5) e manipulação de dados (NumPy).

Em seguida, procedeu-se com o levantamento bibliográfico. Essa fase compreendeu uma revisão das abordagens e algoritmos voltados à detecção e rastreamento de pessoas em vídeos. Foram analisadas diferentes técnicas baseadas em redes neurais convolucionais para detecção de objetos, com destaque para a família de modelos YOLO (You Only Look Once), bem como algoritmos de rastreamento capazes de associar detecções ao longo dos quadros de vídeo. A partir dessa análise, foram selecionados o modelo YOLOv8 para detecção de pessoas e o algoritmo BoT-SORT para o rastreamento, considerando seu bom desempenho em métricas de precisão, robustez em ambientes variados e viabilidade de implementação, além da ampla documentação e disponibilidade do código.

Após a seleção das ferramentas e técnicas a serem utilizadas, procedeu-se à elaboração do método proposto. Este foi estruturado em quatro etapas distintas: a primeira etapa corresponde à aquisição do vídeo; na sequência, realiza-se a detecção das pessoas presentes nas imagens; posteriormente, efetua-se o rastreamento dos indivíduos identificados ao longo dos quadros; por fim, realiza-se a extração e apresentação das métricas relevantes para a análise. O fluxograma da metodologia encontra-se ilustrado na Figura 4.

4.1 Entrada do vídeo

A etapa de entrada de vídeo é responsável por fornecer ao sistema os dados visuais necessários para a análise. Essa etapa pode ocorrer de duas formas: por meio da seleção de arquivos armazenados localmente ou via transmissão em tempo real, utilizando câmeras conectadas via protocolo RTSP (Real Time Streaming Protocol). Na entrada por arquivos locais, o usuário seleciona um vídeo por meio da interface gráfica do sistema. É possível ajustar parâmetros como resolução e taxa de quadros por segundo (FPS) para otimizar o desempenho durante o processamento. Já na entrada por transmissão ao vivo, o sistema realiza a captura contínua de imagens, a partir de webcams integradas ou câmeras IP conectadas à rede, como as da marca Intelbras.

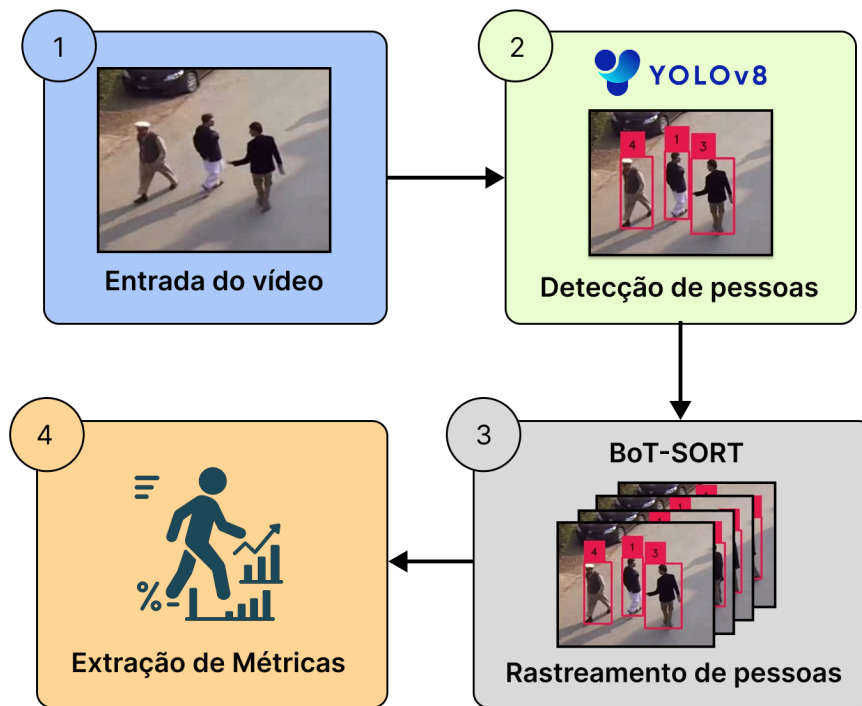


Figura 4 – Fluxograma representando as etapas do método proposto.

Além disso, é possível delimitar uma área de interesse dentro do vídeo. Essa funcionalidade permite ao usuário selecionar manualmente a região da imagem onde a análise será realizada, excluindo áreas irrelevantes. Com isso, o sistema passa a considerar apenas as detecções ocorridas dentro da região selecionada, o que contribui significativamente para a precisão das métricas extraídas, evitando a contagem indevida de pessoas fora do contexto desejado.

4.2 Detecção de Pessoas

A etapa de detecção de pessoas tem como objetivo identificar, em cada quadro do vídeo, a presença de indivíduos no ambiente monitorado. Essa etapa é fundamental para o funcionamento das fases subsequentes, como rastreamento e extração de métricas. Para essa finalidade, é utilizado o modelo YOLOv8 Nano (YOLOv8n), uma versão leve da arquitetura YOLO (You Only Look Once), que se destaca por seu desempenho em tempo real mesmo em ambientes com restrição de hardware. Esse modelo é capaz de processar cada quadro do vídeo de forma eficiente, detectando instâncias da classe de interesse. A detecção é baseada na geração de caixas delimitadoras ao redor dos indivíduos identificados, que são utilizadas posteriormente nas etapas de rastreamento e análise. A escolha do YOLOv8n foi motivada por sua capacidade de equilibrar acurácia e desempenho computacional. A arquitetura do YOLO divide a imagem em uma grade e gera, para cada célula,

previsões de caixas delimitadoras com scores de confiança e probabilidades associadas às classes detectadas.

A versão YOLOv8n utilizada foi treinada no conjunto de dados COCO (Common Objects in Context) (LIN et al., 2015), um dos mais amplamente utilizados para tarefas de detecção de objetos. Esse *dataset* contém 118.000 imagens de treinamento e 5.000 imagens de validação, abrangendo 80 categorias de objetos, incluindo a classe de interesse deste projeto. Em termos de desempenho, o YOLOv8n apresenta uma média de precisão média (mAP) de 37,3% no intervalo de IoU 0.5 a 0.95 (mAP_{50~95}), segundo os resultados reportados pelos desenvolvedores na avaliação no COCO (JOCHER et al., 2023). Essa métrica reflete a capacidade do modelo em localizar objetos com diferentes níveis de sobreposição entre as predições e as anotações reais. Além disso, a arquitetura da YOLOv8n é otimizada para tarefas em tempo real, o que reforça sua adequação a cenários com restrições de processamento, como o ambiente de execução deste sistema.

4.3 Rastreamento de Pessoas

Após a etapa de detecção dos indivíduos em cada quadro do vídeo, procede-se ao rastreamento, que tem como objetivo associar as detecções ao longo do tempo, preservando a identidade de cada pessoa durante sua permanência na cena. Essa etapa é essencial para garantir a continuidade das informações extraídas, como tempo de permanência e velocidade de deslocamento. Para esse fim, foi adotado o algoritmo BoT-SORT (*Bag of Tricks SORT*), que representa uma evolução dos métodos tradicionais de rastreamento SORT e DeepSORT, incorporando melhorias significativas em desempenho e robustez.

O BoT-SORT adota a abordagem *tracking-by-detection*, em que a cada novo quadro as detecções são comparadas com aquelas realizadas anteriormente. A associação entre objetos é feita por meio de uma combinação entre predições de movimento — realizadas através do Filtro de Kalman — e medidas de similaridade visual baseadas em técnicas de Re-Identificação (ReID). Essa combinação permite que o algoritmo mantenha a identidade de cada indivíduo mesmo em situações de oclusão parcial, movimentação brusca ou sobreposição com outros objetos. Além disso, o algoritmo implementa um módulo de compensação de movimento da câmera, o que é especialmente útil em vídeos instáveis ou oriundos de fontes móveis.

4.4 Extração de Métricas

Uma vez detectadas e rastreadas as pessoas no vídeo, o sistema realiza a extração de métricas específicas que fornecem informações relevantes sobre o comportamento dos indivíduos na cena monitorada. As métricas extraídas incluem a contagem total de pes-

soas, o tempo de permanência na área de interesse e a velocidade de deslocamento dos indivíduos.

4.4.1 Contagem de Pessoas

A contagem de pessoas é realizada a partir da identificação de IDs únicos atribuídos pelo rastreador. Cada vez que um novo ID é detectado dentro da área de interesse. Essa métrica permite estimar a densidade de pessoas ao longo do tempo, bem como identificar fluxos e aglomerações.

4.4.2 Tempo de permanência

O tempo de permanência é calculado a partir do número de quadros (frames) durante os quais um ID permanece visível na cena. Considerando a taxa de FPS do vídeo, esse tempo pode ser convertido em segundos, minutos ou horas. Essa métrica é essencial para entender o comportamento dos indivíduos em determinados pontos do espaço monitorado, como filas, salas de espera ou corredores.

4.4.3 Velocidade de Deslocamento

A velocidade de deslocamento de cada indivíduo é estimada a partir do rastreamento contínuo das suas posições ao longo dos quadros do vídeo. Para isso, considera-se a variação da posição (em pixels) e a quantidade de quadros em que o indivíduo foi detectado. Dessa forma, é possível calcular a velocidade média com que a pessoa se desloca dentro da área de interesse, utilizando a seguinte equação:

$$v = \frac{\Delta s}{\Delta t} \quad (4.1)$$

Em que:

- v representa a velocidade média do indivíduo;
- Δs é a variação total de posição ao longo da trajetória (medida em pixels);
- Δt é o intervalo de tempo correspondente, medido em quadros (frames).

Essa métrica é particularmente útil para analisar o comportamento dinâmico dos indivíduos no ambiente monitorado, como sua movimentação, agilidade ou tempo parado. A unidade utilizada (pixels/frame) foi escolhida propositalmente para evitar a necessidade de calibração espacial do vídeo em metros reais, tornando o sistema mais acessível e aplicável em diferentes contextos, mesmo na ausência de informações exatas sobre dimensões físicas da cena.

5 Resultados e Discussão

Nessa seção, são apresentados os principais resultados obtidos com o desenvolvimento da ferramenta de contagem e rastreamento de pessoas em vídeos, destacando sua eficácia e aplicabilidade em diferentes contextos de monitoramento. A solução proposta alia técnicas modernas de visão computacional a uma interface gráfica intuitiva, resultando em um sistema acessível e funcional. São descritos, a seguir, os resultados da etapa de validação das métricas extraídas automaticamente pela ferramenta, bem como os requisitos funcionais e não funcionais definidos durante o desenvolvimento. Por fim, são detalhadas as funcionalidades das telas que compõem o sistema, evidenciando sua contribuição para a usabilidade e eficiência da aplicação.

5.1 Validação das métricas extraídas

Para avaliar a precisão das métricas extraídas pela ferramenta desenvolvida, foi conduzido um experimento utilizando um conjunto de vídeos da base de dados pública Human Activity Recognition (HAR - Video Dataset) (RAJPUT; BILAL; HABIB, 2023). Foram selecionados nove vídeos contendo a ação de caminhada, a fim de validar especificamente a métrica de tempo de permanência de indivíduos no campo de visão da câmera.

Com base nos dados apresentados na Tabela 2, observa-se que a ferramenta apresentou desempenho consistente na estimativa do tempo de permanência. O Erro Absoluto Médio (MAE – Mean Absolute Error) foi de 0,1789 segundos, enquanto o Erro Quadrático Médio (MSE – Mean Squared Error) foi de 0,0423 segundos. Esses valores indicam um bom alinhamento entre os dados estimados e os valores reais, reforçando a confiabilidade do módulo de extração temporal da aplicação.

Tabela 2 – Avaliação do tempo de permanência por vídeo

Vídeo	Tempo Real (s)	Tempo Estimado (s)	Erro Absoluto (s)	Erro Quadrático (s ²)
V1	10.37	10.100	0.270	0.072900
V2	9.81	9.730	0.080	0.006400
V3	10.29	10.200	0.090	0.008100
V4	10.27	10.100	0.170	0.028900
V5	9.84	9.730	0.110	0.012100
V6	9.23	8.833	0.397	0.157609
V7	9.92	9.700	0.220	0.048400
V8	10.47	10.267	0.203	0.041209
V9	8.73	8.660	0.070	0.004900
Média	–	–	0.1789	0.0423

5.2 Requisitos Funcionais e Não Funcionais

Os Requisitos de software determinam o que o sistema deve fazer e sob quais circunstâncias deve operar (PANOUZZO, 2012). Eles descrevem os serviços, funcionalidades e restrições que o sistema deve atender, servindo como base para o desenvolvimento e garantindo que o produto final esteja de acordo com as necessidades dos usuários e os objetivos do projeto. Esses requisitos podem ser divididos em funcionais, que especificam o que o sistema deve fazer, e não funcionais, que definem características como desempenho, segurança e usabilidade.

Os Requisitos Funcionais (RF) de um programa detalham as funções ou serviços que o software deve oferecer, além de especificar como ele deve reagir a determinadas entradas do usuário. Já os Requisitos Não Funcionais (RNF) delineiam as restrições que o software deve cumprir em relação às suas funções ou serviços, como desempenho, segurança ou usabilidade. A Tabela 3 apresenta as identificações, descrições e categorias dos RNFs, enquanto a Tabela 4 apresenta os identificadores, descrições e dependências dos RFs.

Tabela 3 – Requisitos Não Funcionais

Identificador	Descrição	Categoria
RNF01	O sistema deve processar vídeos em tempo real, garantindo uma resposta rápida e eficiente na contagem e rastreamento de pessoas.	Desempenho
RNF02	A interface gráfica deve ser intuitiva e fácil de usar, permitindo que os usuários naveguem e interpretem os dados sem dificuldades.	Usabilidade
RNF03	O sistema deve manter um alto nível de precisão na contagem e rastreamento de pessoas, minimizando erros e falsos positivos.	Confiabilidade
RNF04	O sistema deve ser compatível com diferentes sistemas operacionais desktops.	Compatibilidade
RNF05	O sistema deve garantir a segurança dos dados capturados e processados, protegendo informações sensíveis contra acessos não autorizados.	Segurança
RNF06	O sistema deve estar disponível sempre que necessário, garantindo que os usuários possam acessá-lo.	Disponibilidade

5.3 Funcionalidades do sistema

Por fim, abordaremos as funcionalidades das telas do sistema. A exposição das funcionalidades seguirá a seguinte abordagem: a imagem será apresentada, seguida de sua

Tabela 4 – Requisitos Funcionais

Identificador	Requisito funcional	Descrição	Dependência
RF01	Selecionar Vídeo	O sistema deve permitir que o usuário selecione um arquivo de vídeo do seu computador.	
RF02	Exibir Informações do Vídeo	Ao selecionar o vídeo, o sistema deve apresentar as informações do vídeo, incluindo nome do arquivo, dimensão e tamanho do arquivo de vídeo.	RF01
RF03	Redução da Dimensionalidade	O sistema deve realizar a redução da dimensionalidade do vídeo.	
RF04	Redução da Taxa de FPS	O sistema deve realizar a redução da taxa de frames por segundo (FPS) do vídeo.	
RF05	Selecionar Área de Rastreamento	O sistema deve permitir que o usuário selecione a área de interesse no vídeo para rastrear as pessoas.	RF01
RF06	Deteção e Rastreamento de Pessoas	O sistema deve detectar e rastrear a movimentação de pessoas nos vídeos.	RF05
RF07	Contagem de Pessoas	O sistema deve contar o número de pessoas em vídeos usando técnicas de visão computacional.	RF06
RF08	Tempo de Permanência	O sistema deve calcular o tempo de permanência das pessoas no vídeo.	RF06
RF09	Velocidade das Pessoas	O sistema deve calcular a velocidade das pessoas em pixels/frame.	RF06
RF10	Análise Frame a Frame	O sistema deve permitir que o usuário analise a contagem atual de pessoas em determinados pontos do vídeo.	RF07
RF11	Visualização do Vídeo	O sistema deve permitir a visualização do vídeo ao final do processamento.	RF01, RF07, RF08, RF09, RF04

descrição; em seguida, serão detalhadas as responsabilidades de cada campo, destacado por uma caixa vermelha e numeração.

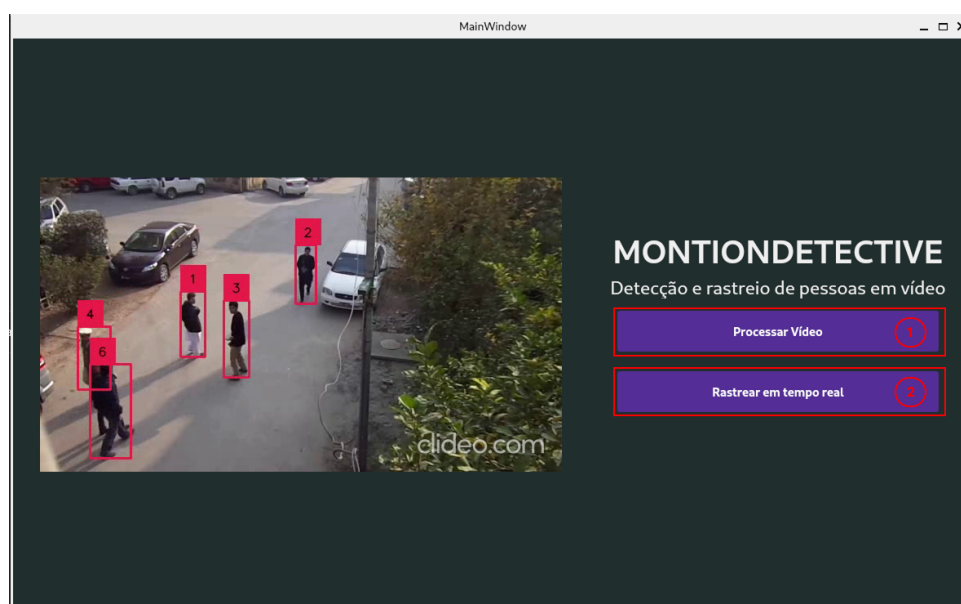


Figura 5 – Tela inicial do sistema.

Figura 5: Tela inicial do sistema, na qual o usuário pode optar entre o processamento de um vídeo previamente gravado ou a contagem de pessoas em tempo real. A interface apresenta os seguintes elementos principais: **1:** Botão que direciona o usuário à tela de carregamento de vídeo, destinado ao processamento de mídias previamente armazenadas; **13:** Botão que direciona à funcionalidade de processamento em tempo real, voltada à análise de vídeos transmitidos ao vivo.

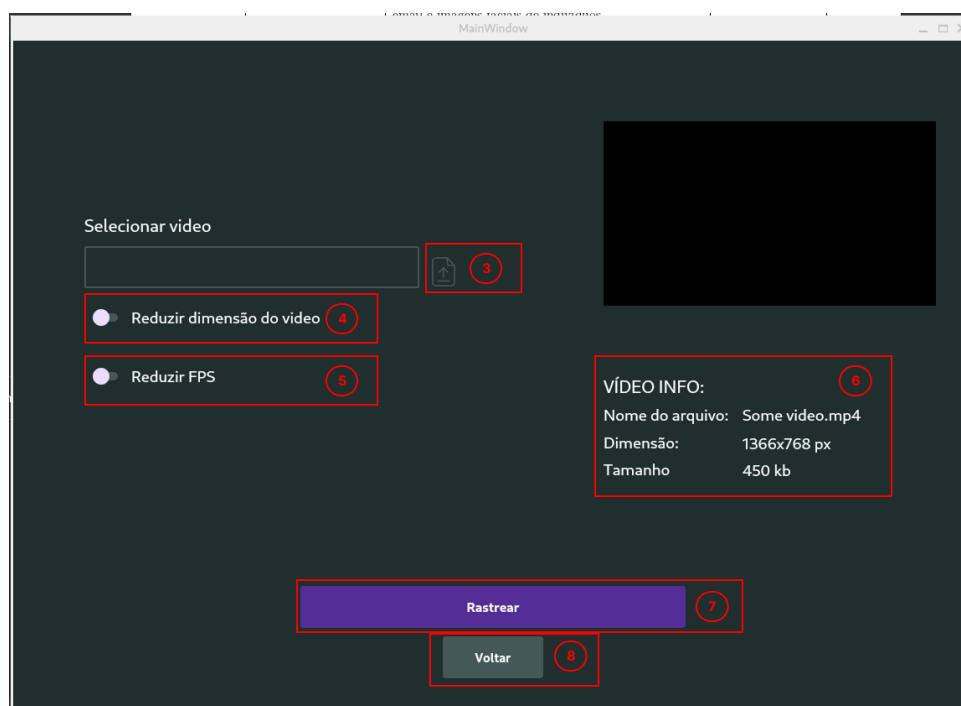


Figura 6 – Tela de selecionar vídeo do sistema.

Figura 6: Tela responsável pelas funcionalidades de carregamento de vídeo, exibida

ao usuário caso essa opção tenha sido previamente selecionada. Nesta interface, é possível importar um vídeo a partir do dispositivo para posterior processamento. Os principais elementos desta tela são: **3**: Botão que abre o gerenciador de arquivos do dispositivo, permitindo a seleção do vídeo a ser processado; **4**: Botão que possibilita a redefinição da resolução do vídeo, permitindo ao usuário especificar uma nova dimensão em pixels, o que pode contribuir para a redução da carga computacional durante o processamento; **5**: Botão que permite a modificação da taxa de quadros por segundo (FPS) do vídeo, possibilitando ao usuário selecionar uma taxa inferior, também com o objetivo de otimizar o desempenho computacional; **6**: Campo de texto que exibe informações relevantes sobre o vídeo carregado; **7**: Botão que direciona o usuário para a próxima etapa do processo, correspondente à seleção da área de rastreamento; **8**: Botão que permite retornar à etapa anterior.

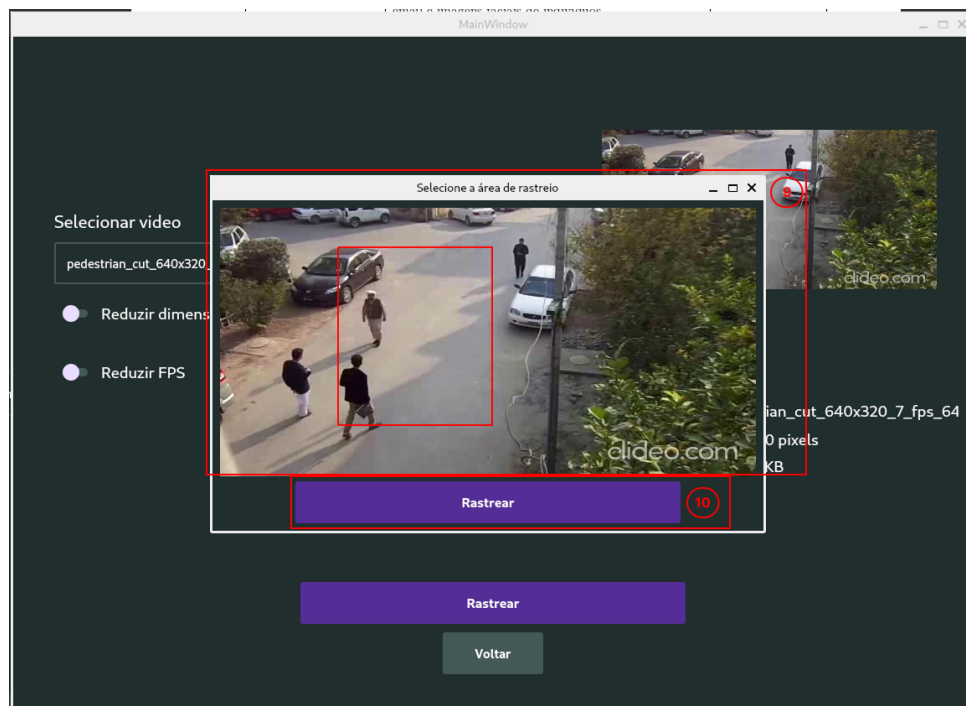


Figura 7 – Tela de selecionar área de rastreamento.

Figura 7: Tela destinada à seleção da área de rastreamento, na qual o usuário pode definir a região do vídeo onde será realizada a detecção e contagem de pessoas. Esta interface é composta pelos seguintes elementos: **9**: Seção interativa que permite ao usuário delimitar manualmente a área do vídeo a ser rastreada; **10**: Botão que inicia o processo de rastreamento na área previamente selecionada.

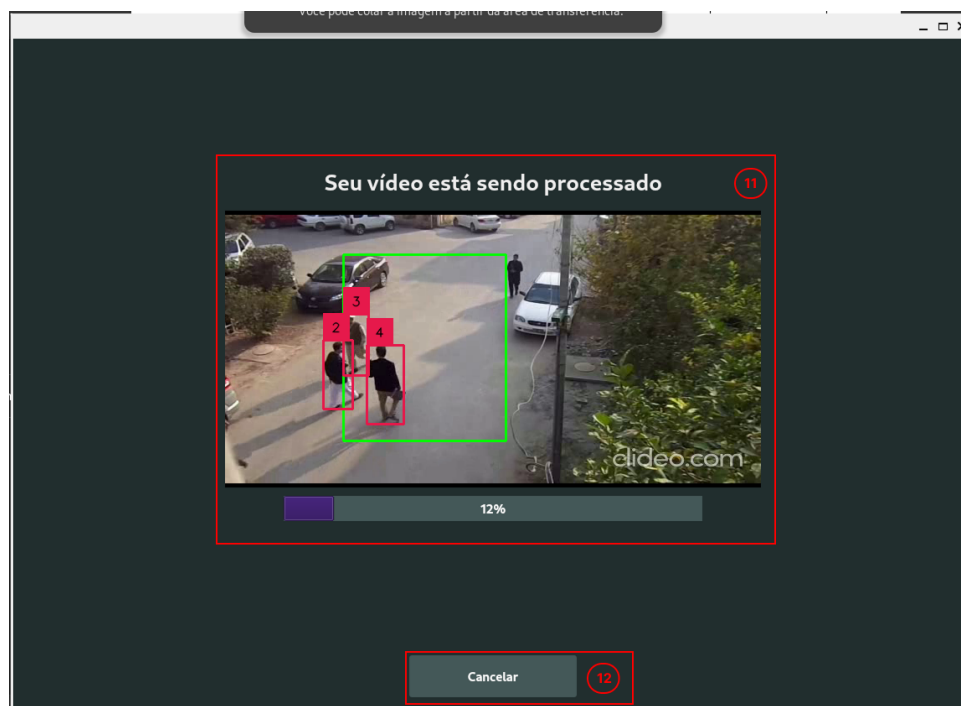


Figura 8 – Tela de processamento do vídeo.

Figura 8: Tela de processamento do vídeo, exibida durante a execução do rastreamento de pessoas na área previamente selecionada. Nessa etapa, o sistema apresenta uma pré-visualização em tempo real do vídeo sendo processado, permitindo ao usuário acompanhar o andamento da análise. Esta interface é composta pelos seguintes elementos: **11:** Área de visualização que exibe, em tempo real, o vídeo sendo processado juntamente com os resultados parciais do rastreamento; **12:** Botão que permite ao usuário cancelar o processamento em andamento, interrompendo a análise do vídeo.

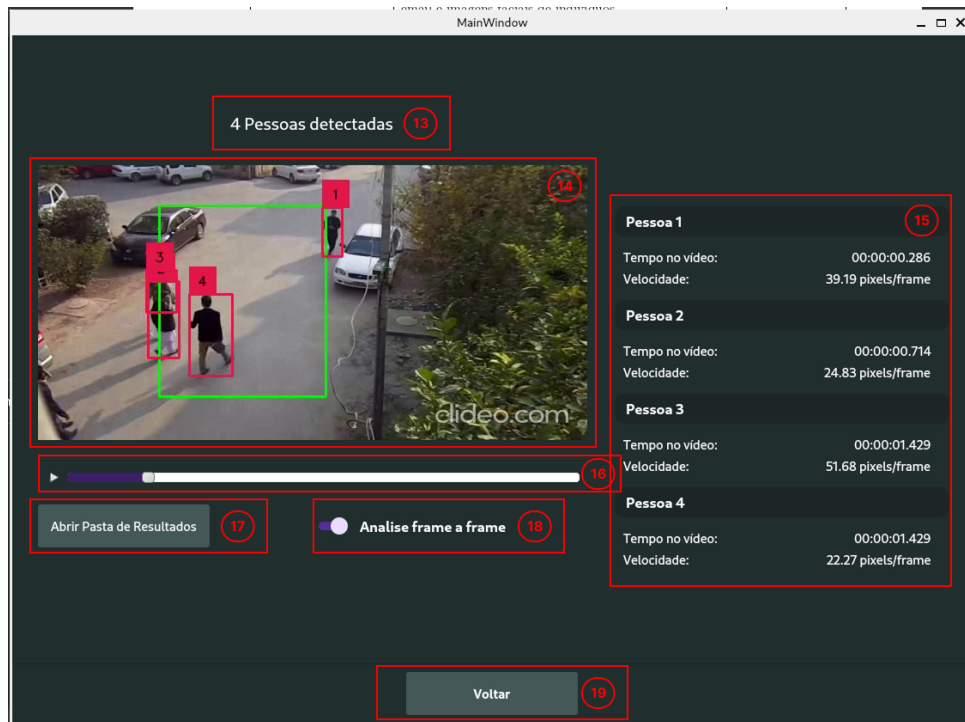


Figura 9 – Tela de resultados.

Figura 9: Tela de exibição dos resultados do processamento, na qual são apresentadas as informações obtidas a partir da análise do vídeo. Essa interface permite ao usuário visualizar o número total de pessoas detectadas, acompanhar o rastreamento individual de cada pessoa no vídeo e acessar os dados quantitativos extraídos. A tela é composta pelos seguintes elementos: **13:** Indicador do número total de pessoas detectadas no vídeo; **14:** Área de visualização contendo o vídeo processado, com sobreposição dos rastros e identificadores únicos atribuídos a cada pessoa; **15:** Painel informativo individual, exibindo para cada pessoa detectada o tempo de permanência no vídeo e sua velocidade estimada (em pixels por quadro); **16:** Barra de progresso do vídeo, que permite ao usuário navegar por diferentes quadros do vídeo processado; **17:** Botão para abrir a pasta de resultados, onde são armazenados os arquivos gerados durante o processamento; **18:** Alternador de análise "frame a frame", que permite ao usuário visualizar o processamento quadro a quadro para análise detalhada; **19:** Botão de retorno à tela anterior.

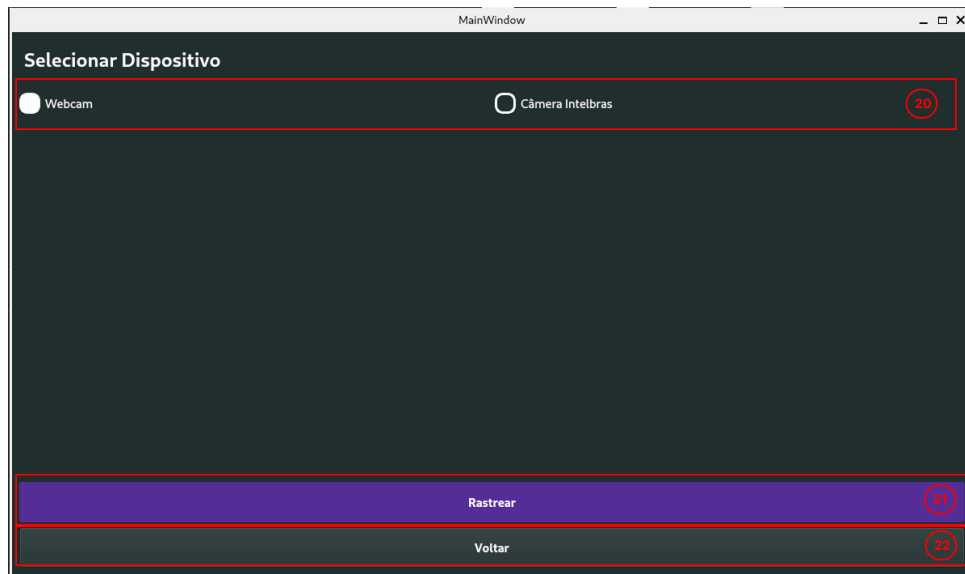


Figura 10 – Tela de inserção de campos para webcam.

Figura 10: Tela de seleção do dispositivo de entrada para rastreamento em tempo real. Esta interface é exibida após o usuário optar por realizar rastreamento em tempo real na tela inicial. O objetivo desta tela é permitir a escolha da fonte de vídeo que será utilizada no processamento ao vivo. Os elementos principais são: **20:** Seletor de dispositivo de vídeo, no qual o usuário pode escolher entre diferentes fontes, como a webcam do sistema ou uma câmera externa intelbras; **21:** Botão “Rastrear”, que inicia o rastreamento em tempo real com base no dispositivo selecionado; **22:** Botão “Voltar”, que permite retornar à tela anterior sem iniciar o rastreamento.

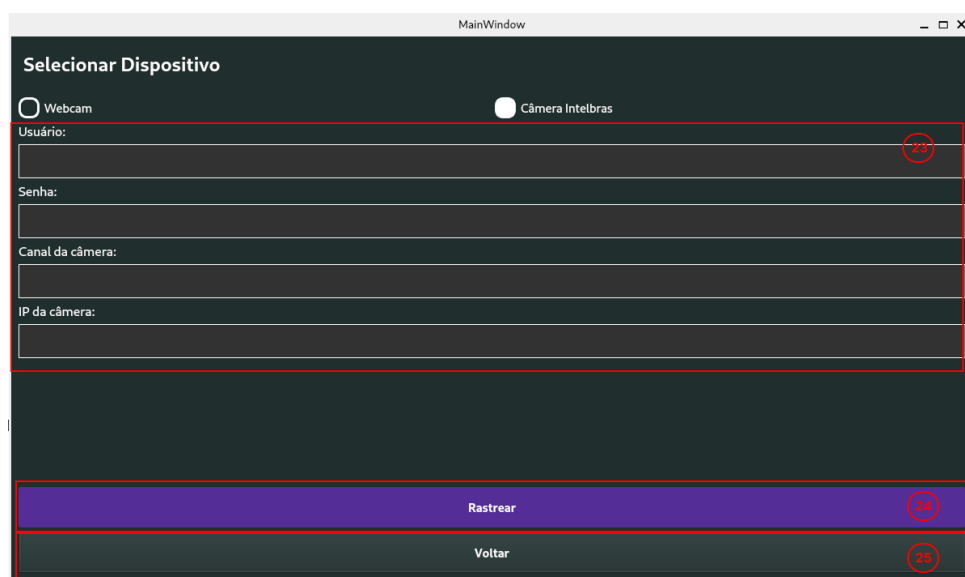


Figura 11 – Tela de inserção de campos para câmera RTSP.

Figura 11: Continuação da tela de seleção do dispositivo para rastreamento em tempo real, exibindo os campos adicionais necessários para conexão com uma câmera In-

telbras. Estes campos são apresentados dinamicamente quando a opção Câmera Intelbras é selecionada. Os componentes incluídos são: **23**: Conjunto de campos de entrada de texto utilizados para configurar a conexão com o dispositivo Intelbras. **24**: Botão “Rastrear”, que inicia o rastreamento em tempo real após o preenchimento e validação dos dados inseridos; **25**: Botão “Voltar”, utilizado para retornar à tela anterior, permitindo ao usuário cancelar a operação sem iniciar o rastreamento.

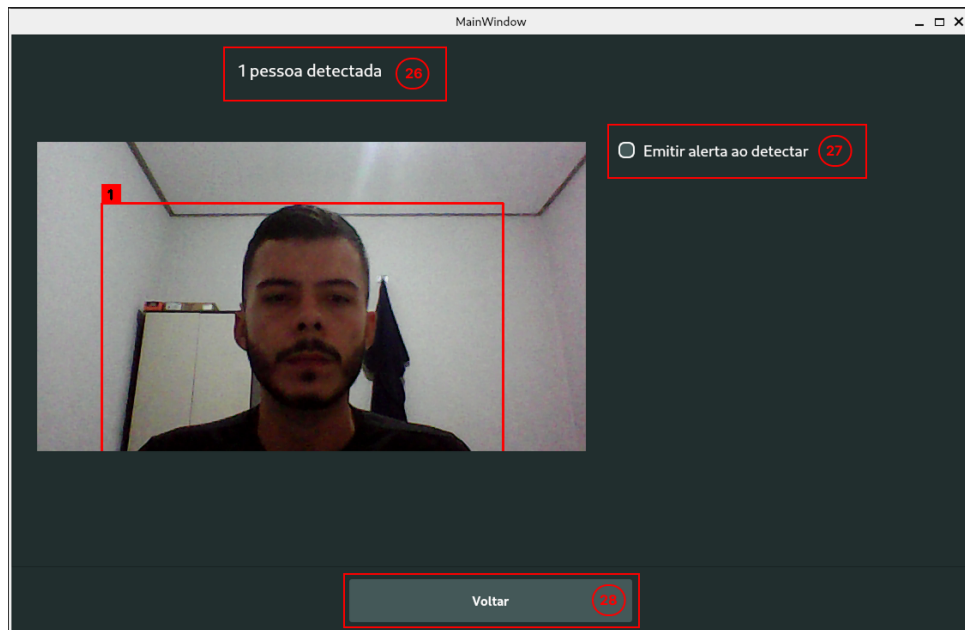


Figura 12 – Tela de resultados em tempo real.

Figura 12: Tela de exibição dos resultados durante a contagem de pessoas em tempo real. Essa interface é apresentada após a seleção e inicialização do dispositivo de rastreamento. Esta tela é composta pelos seguintes elementos: **26** Indicador de contagem de pessoas detectadas na cena. O valor é atualizado dinamicamente à medida que novas pessoas são identificadas pelo sistema de detecção baseado em visão computacional; **27**: Caixa de seleção intitulada “Emitir alerta ao detectar”. Quando ativada, esta opção habilita a emissão de alertas visuais ou sonoros sempre que uma nova pessoa for identificada; **28**: Botão “Voltar”, que encerra o rastreamento atual e retorna à tela anterior para possível reconfiguração ou troca de dispositivo.

6 Conclusão

Neste trabalho, foi desenvolvido um sistema para contagem e rastreamento de pessoas em vídeos, utilizando técnicas de visão computacional. A partir dos resultados obtidos nos testes, foi possível verificar que o modelo YOLOv8 Nano apresentou desempenho satisfatório na detecção de pessoas em diferentes cenários, com baixo custo computacional. Além disso, a extração de métricas como contagem de indivíduos, tempo de permanência e velocidade de movimentação demonstrou ser eficaz para a análise do comportamento humano em ambientes monitorados, mesmo sem a necessidade de calibração espacial.

Durante a etapa de validação, a métrica de tempo de permanência apresentou um Erro Absoluto Médio (MAE) de 0,1789 segundos e um Erro Quadrático Médio (MSE) de 0,0423 segundos, evidenciando a precisão e a consistência do sistema na análise temporal. Esses resultados reforçam a confiabilidade da ferramenta na extração automática de informações comportamentais a partir de vídeos.

Após a escolha e validação do modelo de detecção, foi construída uma ferramenta funcional, com interface gráfica desenvolvida em PyQt5, que integra todas as funcionalidades propostas de forma intuitiva. A linguagem Python mostrou-se adequada para o desenvolvimento dos módulos do sistema, dada sua flexibilidade, vasto conjunto de bibliotecas e grande apoio da comunidade. O design da interface foi elaborado com base em ferramentas de prototipagem e implementação, garantindo uma experiência amigável ao usuário final.

Ao término do desenvolvimento, obteve-se um sistema de software completo, com suporte tanto para análise de vídeos gravados quanto para rastreamento em tempo real por meio de webcam ou câmeras IP. O sistema também permite delimitação de áreas de interesse, emissão de alertas em tempo real e visualização detalhada dos resultados da análise. Além disso, o projeto também resultou na obtenção do registro de software, o que reforça a relevância prática e a originalidade da solução proposta. Essa ferramenta se apresenta como uma alternativa acessível e eficiente frente a soluções comerciais, contribuindo para a democratização do uso de tecnologias de videomonitoramento inteligente.

6.1 Trabalhos futuros

Em trabalhos futuros, recomenda-se a implementação de funcionalidades complementares que ampliem a capacidade analítica da ferramenta desenvolvida. A adição de um sistema de análise da direção de movimento das pessoas pode oferecer maior detalhamento sobre os padrões de deslocamento nos ambientes monitorados. Além disso, sugere-se a criação de mapas de calor que identifiquem as áreas mais movimentadas dentro do vídeo, proporcionando uma visualização intuitiva da concentração de pessoas ao longo do

tempo. Outra possibilidade é a integração de alertas inteligentes baseados em regras personalizáveis, permitindo a geração automática de notificações diante de comportamentos suspeitos. Também se propõe permitir a configuração do modelo de detecção, possibilitando o uso de versões mais leves em máquinas com menor capacidade de processamento e modelos mais precisos em equipamentos mais robustos. Tais aprimoramentos visam não apenas enriquecer as funcionalidades da ferramenta, mas também torná-la ainda mais eficaz para aplicações em segurança, gestão de espaços públicos e análise comportamental em ambientes diversos.

7 Registros e Publicações

Ao final deste trabalho, superadas todas as etapas anteriores, foi solicitado o registro do sistema junto aos órgãos competentes e por intermédio do Núcleo de Inovação Tecnológica (NINTEC) da UFPI. A seguir são fornecidas as referências das publicações derivadas dos resultados deste trabalho.

CARVALHO, V. de S.; ARAÚJO, F. H; GOMES, J. de S; SILVA, R. R. V.; FILHO, A. O. de C. MontionDetective: Uma ferramenta para rastreio e contagem de pessoas em vídeo. Processo Nº: BR512025001921-1, 2025.

Referências

- AHARON, N.; ORFAIG, R.; BOBROVSKY, B. Bot-sort: Robust associations multi-pedestrian tracking. arxiv 2022. *arXiv preprint arXiv:2206.14651*. Citado 3 vezes nas páginas 8, 18 e 19.
- BALLARD, D.; BROWN, C. *Computer Vision*. Prentice-Hall, 1982. ISBN 9780131653160. Disponível em: <<https://books.google.com.br/books?id=EfRRAAAAMAAJ>>. Citado na página 14.
- BORTH, M. R. et al. A visão computacional no agronegócio: Aplicações e direcionamentos. *7º Encontro Científico de Administração, Economia e Contabilidade (ECAECO)*, 2014. Citado na página 15.
- BROWNLEE, J. *Overfitting and Underfitting With Machine Learning Algorithms*. [S.l.]: Machine Learning Algorithms, 2016. <https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/>. Acessado em: 17 mar. 2021. Citado na página 15.
- CAMPOS, P. S. d. Combinação de rastreadores utilizando filtro de kalman. IESTI-Instituto de Engenharia de Sistemas e Tecnologia da Informação, 2017. Citado na página 17.
- CORDEIRO, A. et al. Rastreamento e contagem de pedestre em tempo real por meio de imagens digitais. In: *Anais do XVI Congresso Latino-Americano de Software Livre e Tecnologias Abertas*. Porto Alegre, RS, Brasil: SBC, 2019. p. 146–149. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/latinoware/article/view/10350>>. Citado na página 12.
- COSTA, B. A. D. d. et al. Segmentação, rastreamento de objetos e detecção de eventos primitivos com aplicação no monitoramento automático de ações humanas em vídeo. Universidade Federal de Campina Grande, 2008. Citado na página 17.
- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. [S.l.: s.n.], 2005. v. 1, p. 886–893 vol. 1. Citado na página 16.
- DENG, L.; YU, D. et al. Deep learning: methods and applications. *Foundations and trends® in signal processing*, Now Publishers, Inc., v. 7, n. 3–4, p. 197–387, 2014. Citado na página 15.
- DUQUE, D. F. O. *Previsão e identificação de eventos de quebra de segurança em vídeo-vigilância*. Tese (Doutorado) — Universidade do Minho (Portugal), 2009. Citado na página 12.
- GARCIA, M. L. *Design and Evaluation of Physical Protection Systems*. 2nd. ed. USA: Butterworth-Heinemann, 2007. ISBN 075068352X. Citado na página 12.
- HOUGH, P. V. Method and means for recognizing complex patterns. Disponível em: <<https://www.osti.gov/biblio/4746348>>. Citado na página 16.

- IBGE. *40% da população acredita ter chance média ou alta de ser roubada na rua*. 2022. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/35760-40-da-populacao-acredita-ter-chance-media-ou-alta-de-ser-roubada-na-rua:~:text=Em%202021%2C%2089%2C5%25,a%20menor%20na%20regi%C3%A3o%20Norte-.>>. Citado na página 12.
- JOCHER, G. et al. *Ultralytics YOLOv8*. 2023. [urlhttps://github.com/ultralytics/ultralytics](https://github.com/ultralytics/ultralytics). Version 8.0.0, AGPL-3.0 license. Citado na página 26.
- KALMAN, R. E.; BUCY, R. S. New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, v. 83, n. 1, p. 95–108, 03 1961. ISSN 0021-9223. Disponível em: <<https://doi.org/10.1115/1.3658902>>. Citado na página 17.
- KARAGULIAN, F. et al. Pedestrian flows characterization and estimation with computer vision techniques. *Urban Science*, v. 7, n. 2, 2023. ISSN 2413-8851. Disponível em: <<https://www.mdpi.com/2413-8851/7/2/65>>. Citado 2 vezes nas páginas 20 e 23.
- LIN, T. et al. *Microsoft COCO: Common Objects in Context*. 2015. Citado na página 26.
- LIU, W. et al. Ssd: Single shot multibox detector. In: LEIBE, B. et al. (Ed.). *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016. p. 21–37. ISBN 978-3-319-46448-0. Citado na página 16.
- MARTINS, R. A. S. *Sistema de contagem de pessoas na biblioteca da FEUP utilizando visão computacional*. Tese (Doutorado) — Universidade do Porto (Portugal), 2021. Citado na página 16.
- MCCARTHY, C. et al. Video-based automatic people counting for public transport: On-bus versus off-bus deployment. *Computers in Industry*, v. 164, p. 104195, 2025. ISSN 0166-3615. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0166361524001234>>. Citado 2 vezes nas páginas 21 e 23.
- MILANO, D. de; HONORATO, L. B. Visão computacional. *Faculdade de Tecnologia, Universidade Estadual de Campinas*, 2010. Citado na página 14.
- MIRZAEI, B. et al. Small object detection and tracking: A comprehensive review. *Sensors*, v. 23, n. 15, 2023. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/23/15/6887>>. Citado na página 18.
- MISHRA, P. K.; SAROHA, G. P. A study on video surveillance system for object detection and tracking. In: *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. [S.l.: s.n.], 2016. p. 221–226. Citado na página 17.
- MOKAYED, H. et al. Real-time human detection and counting system using deep learning computer vision techniques. *Artificial Intelligence and Applications*, v. 1, n. 4, p. 221–229, Oct. 2022. Disponível em: <<https://ojs.bonviewpress.com/index.php/AIA/article/view/391>>. Citado 2 vezes nas páginas 21 e 23.

- NASRY, A.; EZZAHOUT, A.; OMARY, F. People tracking in video surveillance systems based on artificial intelligence. *Journal of Automation, Mobile Robotics and Intelligent Systems*, v. 17, n. 1, p. 59–68, Dec. 2023. Disponível em: <<https://www.jamris.org/index.php/JAMRIS/article/view/858>>. Citado na página 12.
- PANOZZO, R. P. A dimensão tácita do conhecimento na definição dos requisitos em uma fábrica de software da serra gaúcha. 2012. Citado na página 29.
- PEŠEK, O. *Mask R-CNN in GRASS GIS*. Dissertação (Tese de Mestrado) — Université Libre de Bruxelles, jun. 2018. Acessado em: 15 mar. 2021. Disponível em: <<https://github.com/ctu-geoforall-lab-projects/dp-pesek-2018/blob/master/text/ondrej-pesek-dp-2018.pdf>>. Citado na página 15.
- PIRES, J. M. N. G. *Aprendizagem profunda: estudo e aplicações*. Dissertação (Mestrado) — Universidade de Evora (Portugal), 2017. Citado na página 15.
- PONTI, M. A.; COSTA, G. B. P. D. Como funciona o deep learning. *arXiv preprint arXiv:1806.07908*, 2018. Citado na página 15.
- RAJPUT, S. M.; BILAL, M.; HABIB, A. *Human Activity Recognition (HAR - Video Dataset)*. Kaggle, 2023. Disponível em: <<https://www.kaggle.com/dsv/5722068>>. Citado na página 28.
- REDMON, J. et al. You only look once: Unified, real-time object detection. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 779–788. Citado 3 vezes nas páginas 8, 16 e 17.
- REHEM, A.; TRINDADE, F. H. V. Técnicas de visão computacional para rastreamento de olhar em vídeos. *Publicado em*, v. 3, n. 02, 2009. Citado na página 14.
- SAHA, S. *A Comprehensive Guide to Convolutional Neural Networks, the ELI5 way*. [S.l.]: Towards Data Science, 2018. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>. Acesso em: 20 nov. 2020. Citado na página 15.
- SILVA, G. R. et al. Detecção de objetos em imagens utilizando técnicas de aprendizagem profunda. Florianópolis, SC, 2018. Citado na página 16.
- SILVA, W. S. d. Estimacão de densidade de multidões: proposta de arquitetura. 2021. Citado na página 16.
- STOCKMAN, G.; SHAPIRO, L. G. *Computer Vision*. 1st. ed. USA: Prentice Hall PTR, 2001. ISBN 0130307963. Citado na página 14.
- TELICKO, J.; JAKOVICS, A. Comparative analysis of yolov8 and mack-rcnn for people counting on fish-eye images. In: *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. [S.l.: s.n.], 2023. p. 1–6. Citado 2 vezes nas páginas 21 e 23.
- VASAVI, G. et al. People counting based on yolo. In: *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*. [S.l.: s.n.], 2023. p. 1–6. Citado 3 vezes nas páginas 12, 20 e 23.

XU, H.; LV, P.; MENG, L. A people counting system based on head-shoulder detection and tracking in surveillance video. In: *2010 International Conference On Computer Design and Applications*. [S.l.: s.n.], 2010. v. 1, p. V1-394-V1-398. Citado 3 vezes nas páginas 12, 20 e 23.

YAKIMOVSKY, Y. Boundary and object detection in real world images. *J. ACM*, Association for Computing Machinery, New York, NY, USA, v. 23, n. 4, p. 599-618, oct 1976. ISSN 0004-5411. Disponível em: <<https://doi.org/10.1145/321978.321981>>. Citado na página 16.



TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA “JOSÉ ALBANO DE MACEDO”

Identificação do Tipo de Documento

- () Tese
- () Dissertação
- (X) Monografia
- () Artigo

Eu, Vinícius de Sousa Carvalho, autorizo com base na Lei Federal nº 9.610 de 19 de Fevereiro de 1998 e na Lei nº 10.973 de 02 de dezembro de 2004, a biblioteca da Universidade Federal do Piauí a divulgar, gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação **Desenvolvimento de uma ferramenta para contagem e monitoramento de pessoas em vídeos** de minha autoria, em formato PDF, para fins de leitura e/ou impressão, pela internet a título de divulgação da produção científica gerada pela Universidade.

Picos-PI 05 de julho de 2025.



Documento assinado digitalmente

VINICIUS DE SOUSA CARVALHO

Data: 05/07/2025 19:34:43-0300

Verifique em <https://validar.iti.gov.br>

Assinatura