

Sistema de gestão de fluxo clientes em pontos comerciais utilizando visão computacional

Luana Goncalves^{1,2}  orcid.org/0000-0002-7287-2911 2427

Carmelo J. A. Bastos Filho¹  <http://orcid.org/0000-0002-0924-5341>

Victor Mendonça de Azevedo³  <http://orcid.org/0000-0003-2943-4622>

¹ Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

² Pós-graduação em Inteligência Artificial, Escola Politécnica de Pernambuco, Pernambuco, Brasil.

³ Fundação para Inovações Tecnológicas, Pernambuco, Brasil.

E-mail do autor principal: Luana Goncalves luanagoncalves4@live.com

Resumo

Para a administração de ativos com mais precisão de acordo com a demanda e público em pontos comerciais, é cada vez mais comum o uso de técnicas para extrair informações acerca da distribuição dos clientes. O objetivo é minimizar o tempo dos clientes em filas e atendimentos, adequar os produtos e reorganizar o layout do espaço a partir do mercado consumidor. Neste trabalho, é proposto um sistema voltado para otimização de recursos em pontos comerciais, através do processamento de imagens de câmeras de segurança para o monitoramento do fluxo e perfil de clientes. O baixo custo de implementação do sistema aumenta o mercado potencial de compra do serviço, diminui o tempo de retorno sobre o investimento e assim reduz a barreira de entrada em grandes companhias.

Palavras-Chave: Visão computacional; Processamento de imagens; Teoria de filas; Inteligência Computacional; Aprendizado de máquina.

Abstract

The number of tools to extract information regarding distribution in stores has increased in the last years. The goal is to minimize customer time in queues and service, tailor products, and reorganize space using the consumer perspective. This work proposes a system aimed at optimizing resources at stores, through the processing of videos acquired from security cameras to obtain the number of customers. The low cost of system implementation increases the potential market for service purchases, decreases the return on investment time, and reduces the entry barrier in large companies.

Key-words: Computer vision; Image processing; Queuing theory; artificial intelligence; Machine Learning.

1 Introdução

Com o objetivo de aumentar as vendas e se destacar em um cenário de alta competitividade, um dos grandes desafios do varejo físico é conseguir mapear o comportamento do seu público alvo. O monitoramento do fluxo, assim como do perfil, das pessoas que ingressam nos estabelecimentos, por exemplo, traz para as empresas alguns índices estratégicos que antes eram privilégios das lojas virtuais. Sendo assim, vem crescendo o investimento em ferramentas que tornem a coleta, análise e gestão desses dados mais eficiente, permitindo que o varejista faça avaliações específicas para sua realidade, crie melhores ofertas e proporcione experiências agradáveis a quem visita a loja [1].

A transformação digital no setor do varejo pode viabilizar uma gama de possibilidades referentes a análises e inferências dos gestores nas tomadas de decisão. No entanto, a coleta de dados em relação às características da distribuição estatística dos clientes atendidos nem sempre é uma tarefa trivial. Em [2], é relatado que para enfrentar o problema da adequação do número de caixas à demanda as Lojas Americanas S/A designa um funcionário para verificar a cada 30 minutos a quantidade de clientes que aguardam na fila dos caixas. Em [3], estabelecimentos criam filas virtuais com o auxílio de smartphones para gerir sistemas de filas, o que além de coletar dados visa melhorar a vivência dos clientes durante a espera.

Apesar da facilidade de implementação dessas técnicas em uma loja ser alta, quando se trata de uma rede varejista esse processo precisa ser escalado e acaba sendo altamente custoso. Nesse cenário, a proliferação de câmeras de vídeo de segurança em lugares públicos fornece uma infraestrutura propícia para o desenvolvimento de algoritmos de monitoramento e captação de dados. O objetivo é aproveitar a arquitetura já implementada na maioria dos grandes lojistas a fim de minimizar os investimentos iniciais. Outra vantagem de utilização das câmeras de segurança é que a maioria delas já possui uma interface de disponibilização de suas imagens via *stream*, o que permite a centralização do processamento em servidores mais robustos que podem processar diversas câmeras ao mesmo tempo [4].

Em [5], câmeras de segurança são utilizadas como suporte para um sistema de contagem de pessoas, através da segmentação do plano de fundo utilizando k-means. Um modelo estimativo do fluxo de pessoas

a partir de visão computacional é proposto em [6], pelo uso do algoritmo de aprendizado AdaBoost. A limitação das técnicas se dá pelo fato dos algoritmos serem facilmente confundidos quando há grande aglomeração de pessoas. Para a detecção de pessoas em imagens e vídeos, destaca-se o algoritmo You Only Look Once (YOLO), uma abordagem para classificar e localizar pessoas na cena [7].

Neste trabalho, é apresentado um sistema para monitoramento do fluxo e perfil de clientes em pontos comerciais usando visão computacional. O objetivo é fornecer simulações que permitam aos gestores personalizar os sistemas de atendimento, otimizar o uso de recursos e apoiar as tomadas de decisões. Dentre os produtos disponíveis estão: a contagem de pessoas que entram e saem do estabelecimento ao longo do dia; o perfil de gênero e idade dos clientes; o mapa do fluxo de pessoas no espaço físico da loja; e a previsão do número de clientes em filas para dimensionamento de servidores de atendimento.

2 Arquitetura do sistema

Em linhas gerais, a ferramenta pode ser dividida em três módulos: a detecção de pessoas na cena, o rastreamento das pessoas detectadas e os produtos disponíveis, conforme ilustrado na Figura 1. Os produtos oferecidos aos gestores são: a distribuição do número de pessoas que ingressam no estabelecimento, permitindo adequar o quadro de funcionários aos horários de picos e vales de demanda de clientes; a distribuição de gênero dos clientes, cujo objetivo é descobrir o público alvo mais aderente e

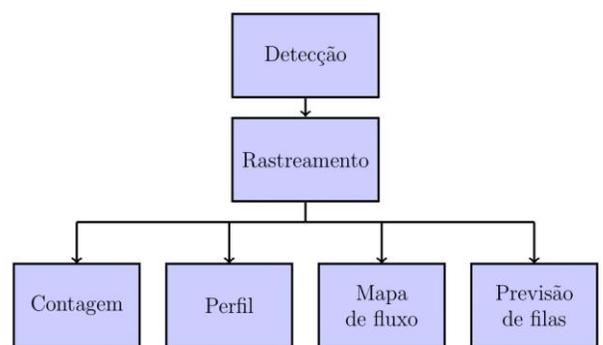


Figura 1: Arquitetura do sistema.
Fonte: Autor (2019).

adequar estratégias de *marketing* para resgatar mercados consumidores; o mapeamento do fluxo de pessoas na distribuição espacial da loja, possibilitando estudar o melhor posicionamento dos artigos de venda e quais campanhas e marcas são mais atrativas; e previsão do tempo de espera em filas para adequação do número de servidores. Tais produtos tem a finalidade de otimizar os recursos disponíveis de forma dinâmica, onde o foco é a análise dos padrões cíclicos e sazonais dos índices estatísticos observados.

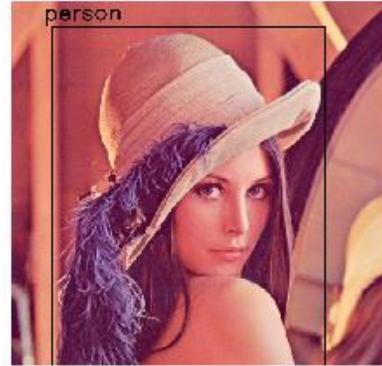


Figura 2: Exemplo de detecção de pessoas em imagens. Fonte: Autor (2019).

2.1 Detecção de pessoas

O sistema visual humano apesar de complexo é capaz de detectar e reconhecer objetos em imagens de forma rápida e eficaz. Automatizar essa tarefa tem sido um tópico bastante discutido. Para a detecção de pessoas na cena, foi utilizado o algoritmo *You Only Look Once*, uma rede neural convolucional treinada para classificar e localizar elementos na cena [7]. Uma de suas principais características é a alta velocidade de predição dado que a imagem é apresentada à rede apenas uma vez, como o nome sugere, técnica chamada *Single Shot Detectors*, e é recomendada para problemas de detecção de objetos em tempo real.

Para isso, o algoritmo divide a imagem em uma grade de $S \times S$ células idênticas. Cada célula é usada para prever as probabilidades condicionais em relação às classes de treinamento. As saídas são objetos retangulares (*bounding box*), que representam as regiões de maior confiança dos objetos detectados, e as probabilidades de pertinência da classificação. Embora a rede consiga detectar até 80 classes de objetos diferentes, foi adaptada para identificar apenas a classe “pessoa”. Na Figura 2 é apresentado um exemplo de detecção de pessoas utilizando o YOLO.

2.2 Rastreamento de pessoas

A detecção de pessoas, frame após frame, implica em localizar a mesma pessoa repetidamente até não aparecer mais no vídeo. O rastreamento relaciona posições de um mesmo indivíduo ao longo do percurso realizado em relação ao tempo, evitando que seja contado mais de uma vez [8].

Em [9], o algoritmo *Simple Online and Real-time Tracking (SORT)* propõe a combinação entre o filtro de Kalman e o método húngaro de otimização para o rastreamento de pessoas em cenas de vídeo. Entretanto, devido à uniformidade das imagens um método mais simples foi desenvolvido.

Logo, para rastrear pessoas, primeiramente, uma identificação (ID) é fornecida a todas que são detectadas no primeiro frame do vídeo. Nos frames subsequentes, para cada pessoa detectada, o ID é atribuído em relação à menor distância às detecções no frame anterior. A associação é determinada através da distância euclidiana, definida por:

$$d = \sqrt{(X_k - X_{k-1})^2 + (Y_k - Y_{k-1})^2}, \quad (1)$$

onde (X_k, Y_k) e (X_{k-1}, Y_{k-1}) representam os centros de massa da detecção do frame atual e frame anterior, respectivamente.

Exceto quando a distância mínima for maior que um limiar, obtido por testes empíricos, e nos casos em que o número de detecções no frame for maior que o número de detecções no frame anterior, evidenciando que novas pessoas ingressam na cena e, portanto, são necessários novos IDs. Na Figura 3, é apresentado o pseudocódigo do processo de rastreamento.

Após identificar os possíveis percursos realizados faz-se necessário descobrir o sentido do deslocamento de cada trajetória, dado pelas condições abaixo:

$$D = \begin{cases} up, & y_k > y_0 \\ down, & y_k < y_0' \end{cases} \quad (2)$$

DOI: 10.25286/repa.v5i2.1324

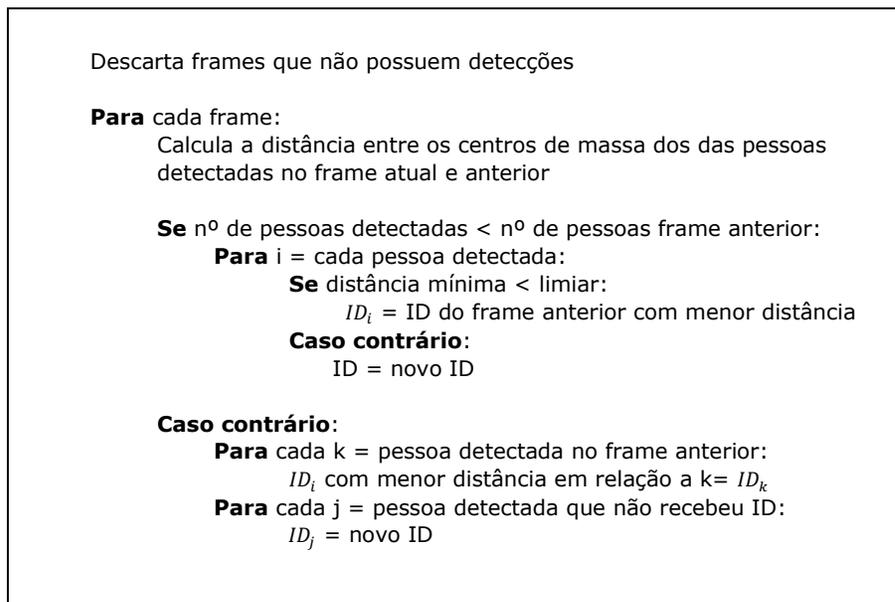


Figura 3: Processo de rastreamento de um indivíduo detectado.
Fonte: Autor (2019).

em que, D é a direção do deslocamento (*up, down*), y_k e y_0 representam as coordenadas verticais finais e iniciais do percurso rastreado, respectivamente.

2.3 Contagem de pessoas

Para proporcionar uma gestão inteligente da escala de trabalho, as alocações devem considerar os turnos dos funcionários e as demandas dos clientes. Tal processo é comumente chamado de *Workforce Management*, o qual consiste em alocar de forma otimizada os trabalhadores da empresa, utilizando previsões do número de visitantes e a performance dos funcionários. A proposta é entender os padrões de fluxo de pessoas a serem atendidas a fim de adequar a força de trabalho às necessidades; definir regras de trabalho, carga horária, turnos e programação de jornada; além de estabelecer os melhores dias e horários de treinamentos e reuniões [10].

Para isso, a contagem de pessoas é obtida a partir da captação de imagens nos acessos de entrada e saída do estabelecimento. Então, o contador de entrada, ou seja, de clientes que ingressam na loja, é definido como a soma dos IDs que tiveram a direção de rastreamento definidos como *up*, e o contador de saída é a soma de IDs definidos como *down*. Nas Figuras 4 e 5, são apresentados os gráficos da distribuição acumulada dos contadores de entrada e saída, respectivamente, de um vídeo disponível em [11], como exemplo.

2.4 Perfil

Público alvo trata-se de um grupo específico de consumidores que compartilham um perfil semelhante e por isso devem ser o foco das ações de *marketing* e vendas da empresa, uma vez que estão mais dispostos a adquirir os produtos oferecidos, evitando direcionar recursos a públicos que potencialmente não se interessarão pelos serviços disponíveis. O objetivo principal de identificar o perfil dos clientes é garantir uma aproximação mais personalizada com os clientes [12]. Logo, para analisar o público alvo atingido pela empresa, busca-se os padrões de gênero e idade dos clientes em imagens de câmeras de segurança. Para a tarefa de classificação, após a detecção e rastreamento, apenas um frame rastreado é escolhido para análise. Primeiramente, a técnica *Viola Jones* é utilizada para reconhecimento e segmentação facial, que consiste na convolução da imagem com *kernels* de extração de características que serve de entrada para um modelo pré-treinado [13]. Dado que a regia do rosto foi obtida, é reescalada para 256×256 pixels e alimenta as redes neurais convolucionais treinadas para classificação de gênero e idade, as quais estão disponíveis para *download*¹ [14]. Nas Figuras 6 e 7, são apresentados exemplos de classificação de gênero e idade, respectivamente.

¹<https://github.com/GilLevi/AgeGenderDeepLearning>



Figura 4: Exemplo de contador de entrada.
Fonte: Autor (2019).



Figura 6: Exemplo de classificação de gênero.
Fonte: Autor (2019).



Figura 5: Exemplo de contador de saída.
Fonte: Autor (2019).



Figura 7: Exemplo de classificação de idade.
Fonte: Autor (2019).

2.5 Mapa de fluxo

Com o objetivo de descobrir os espaços mais atraentes e diminuir as áreas ociosas, mapas de fluxo da loja, também conhecido como mapas de calor, são indicações das áreas mais e menos visitadas do local, o que permite reformular o *layout* da loja periodicamente. A ferramenta fornece representações visuais de cores que apontam a movimentação dos consumidores. Quanto mais quente for a cor, maior o fluxo de clientes no setor. Já os tons frios indicam os locais que passam despercebidos. Portanto, uma vez que a imagem é adquirida, é realizada a detecção de movimentos usando o *Background subtraction*, uma técnica amplamente usada em câmeras estáticas, a

qual gera uma máscara em primeiro plano, ou seja, uma imagem binária que discrimina os pixels que representam os objetos em movimento na cena [15].

A máscara de primeiro plano é calculada a partir da subtração entre o frame atual e o frame anterior, o que gera um modelo de plano de fundo, contendo a parte estática da cena, dadas as características da cena observada. Posteriormente, é calculado um histograma bidimensional dos modelos obtidos a partir da equação:

$$H(x, y) = \sum_1^n m(x, y), \quad (3)$$

onde, $m(x,y)$ são as máscaras de primeiro plano obtidas a partir dos frames do vídeo e n representa o número de frames. Então, o mapa $H(x,y)$ do vídeo utilizado para contagem de pessoas, o mesmo utilizado nas Figuras 4 e 5, é normalizado e mostrado na Figura 8, onde as regiões onde não houve passagem de pessoas são representadas em azul e em vermelho onde houve maior fluxo de clientes.

2.5 Previsão de filas

Uma vez que o tempo de atendimento é um dos maiores fatores de competitividade entre as empresas e a quantidade de servidores disponíveis para garantir que não haverá filas é, muitas das vezes, impraticável [16], a Teoria das Filas tem como objetivo o desenvolvimento de modelos matemáticos para previsão do comportamento de sistemas que envolvam espera a partir da distribuição da população a ser atendida, o que permite a otimização entre a experiência dos clientes e a utilização de recursos [17].

Em sistemas de atendimento com número fixo de servidores, como os processos de chegada, em geral, ocorrem em instantes aleatórios e a duração dos atendimentos também é variável, é comum a flutuação na demanda de clientes ao longo do dia. Logo, a procura pelos serviços pode ser muito intensa durante determinados períodos, o que resulta em uma concentração do volume de pessoas à espera de atendimento. Por outro lado, em outros períodos, é

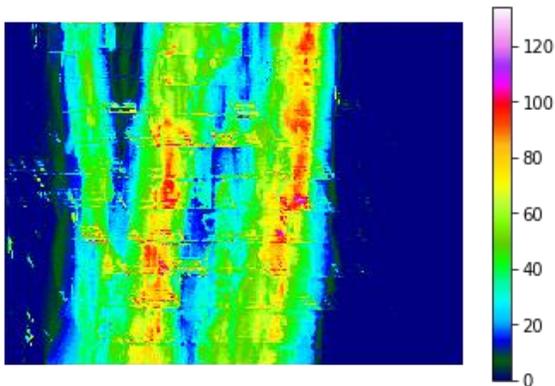


Figura 8: Exemplo de mapa de fluxo.
Fonte: Autor (2019).

possível que a demanda seja tão baixa que os atendentes fiquem ociosos boa parte do tempo. Ambas as situações evidenciam que o número de servidores não está adequadamente dimensionado, o que aponta que a disponibilidade de servidores deve variar em relação à demanda [18].

Em sistemas de atendimento com múltiplos servidores e fila única, o tempo médio de espera, W_q , é dado por:

$$W_q = \frac{L_q}{\lambda}, \quad (4)$$

onde λ é taxa média de chegada e L_q é o número médio de clientes esperando na fila, definido como:

$$L_q = \frac{P_0(\lambda/\mu)^s p}{s!(1-p)^2}, \quad (5)$$

em que s representa o número de servidores, μ a taxa de atendimento, P_0 é a probabilidade de não haver clientes na fila, calculada através de:

$$P_0 = \left[\sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \left(\frac{1}{1-p} \right) \right]^{-1}, \quad (6)$$

onde p é a taxa média de utilização dos servidores dada por:

$$p = \frac{\lambda}{s\mu}. \quad (7)$$

Com isso, simulações do tempo médio de espera e a probabilidade de não haver clientes na fila para todos os possíveis números de servidores $[1,2,\dots,n]$, dependendo da configuração da loja, são realizadas. Para isso, a fila a espera de atendimento é monitorada onde são obtidos a taxa média de chegada e a taxa de atendimento. O objetivo é que o gestor possa escolher o número ótimo de servidores para o atendimento.

Logo, após a delimitação da região da fila e região de atendimento na imagem, a detecção e rastreamento de pessoas são processados. Então, a taxa média de chegada é dada pelo número médio de IDs por hora obtidos na região de fila. A taxa de atendimento é definida como o número médio de IDs por hora obtidos na região de atendimento.

3 Resultados

Para a validação, o modelo de rastreamento foi testado a partir de 2 vídeos disponíveis em [11], que monitoram passagem de pessoas. Primeiramente, foram selecionados apenas os intervalos em que houveram detecções. Na Figura 8, é apresentado um gráfico de dispersão dos centros de massa das detecções referente a um dos vídeos de teste.

Para visualizar a tarefa de rastreamento, na Figura 9, é mostrado um gráfico onde o percurso de cada ID é identificado com uma cor diferente.

Adicionalmente, o sistema de contagem foi validado em ambos os vídeos. Na Tabela 1, é apresentada a acurácia do modelo.

Tabela 1: Acurácia do sistema de contagem de pessoas.

	Vídeo 1	Vídeo 2
Up	100%	100%
Down	100%	92%

A avaliação dos sistemas de classificação de perfil dos clientes foi realizada a partir de 11 imagens, totalizando 18 pessoas detectadas. Dentre elas, a rede neural para classificação de gênero apresenta acurácia de 83,3% e a rede para classificação de idade 62,5%.

Finalmente, para o sistema de previsão de clientes em filas, a validação foi realizada a partir de um vídeo obtido em [19], o qual monitora um servidor de atendimento. Em que, verificou-se que a taxa média de atendimento corresponde a 2 pessoas/minuto e a taxa média de chegada a 1 pessoa/minuto o que condiz com a realidade do vídeo.

4 Conclusões

Foi mostrado neste trabalho, um sistema para gestão do fluxo de pessoas em pontos comerciais, utilizando visão computacional. Destaca-se o uso de câmeras de segurança para monitoramento da distribuição de pessoas nos estabelecimentos. O algoritmo *You Only Look Once* foi utilizado para detecção de pessoas nas imagens, e foi desenvolvido um método de rastreamento de percurso das pessoas

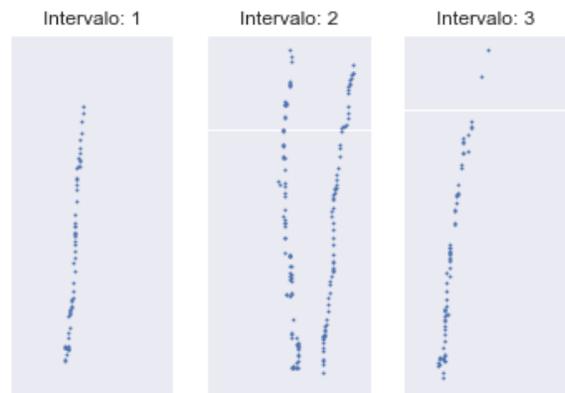


Figura 8: Exemplo de gráfico de dispersão de centro de massa das pessoas detectadas dos três primeiros intervalos do vídeo.

Fonte: Autor (2019).

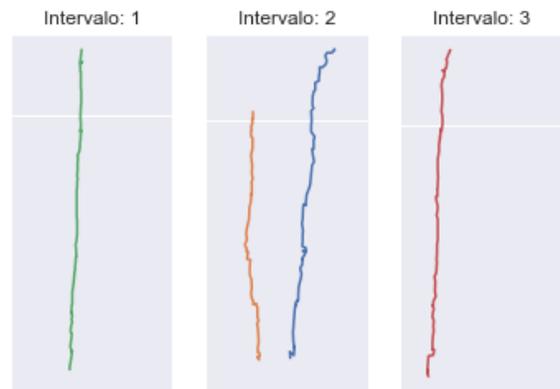


Figura 9: Exemplo de rastreamento a partir de centro de massa das pessoas detectadas dos três primeiros intervalos do vídeo.

Fonte: Autor (2019).

detectadas com notório desempenho, tal modelo serve como base para os produtos desenvolvidos.

Os modelos de contagem de pessoas, mapa de fluxo e previsão de filas apresentam bom desempenho para os vídeos testados, no entanto, o próximo passo é a avaliação com vídeos mais longos e com fluxo maior de pessoas.

Nota-se que o sistema de classificação de gênero apresenta bons resultados, muito embora falha consideravelmente em crianças. Em contrapartida, o modelo de classificação de idade não mostra boa acurácia, portanto, faz-se necessário refazer o treinamento.

Agradecimentos

Os autores agradecem à FITec/SECTI/CMA-Parqtel/UPE/FACEPE pelo incentivo e oportunidade de realizar esta pesquisa.

Referências

- [1] Catalini, J.R; Papa, M.C.O; Ferro, R; Helleno, A.L. Avaliação do processo de atendimento em uma loja de telefonia celular por meio da aplicação dos conceitos de teoria das filas e simulação. IV Congresso de Sistemas LEAN, Porto Alegre, 2014.
- [2] FIGUEIREDO, D.D; ROCHA, S.H. aplicação da teoria das filas na otimização do número de caixas: um estudo de caso. Iniciação Científica CESUMAR - jul./dez. 2010, v. 12, n. 2, p. 175-182 - ISSN 1518-1243.
- [3] SANCHIS, Joseph et al. Virtual queue management system. Depositante: Joseph SANCHIS.. Titular: Airometric Wireless Inc. US20140089075A1. Depósito: 23 set. 2013. Concessão: 16 dez. 2016..
- [4] NOGUEIRA, A.B. Análise de viabilidade no uso de Deep Learning para contagem de pessoas com câmeras de segurança. Orientador: Aldo von Wangenheim. 2018. 78 f. Trabalho de Conclusão de Curso. Universidade Federal de Santa Catarina, Florianópolis, 2018.
- [5] ALMEIDA, S. S.; MELO, V. H. C.; MENDES, J. C. Uma Implementação de um Sistema de Contagem de Pessoas Baseado em Vídeo Processamento. Orientador: David Menotti. 2010. Trabalho de disciplina. Universidade Federal de Ouro Preto, Florianópolis, 2010.
- [6] SOUZA, W.G; FONSECA, I.E; BELO, F.A. Modelo estimativo do fluxo de embarque e desembarque de passageiros em sistemas de monitoramento metroferroviário com base em técnicas de visão computacional. XXXV Simpósio Brasileiro De Telecomunicações, 2017, São Pedro.
- [7] REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You Only Look Once: Unified, Real-Time Object Detection, arXiv, 2015, Cornell University.
- [8] AMARAL, W.M. Rastreamento de pessoas em vídeo. Orientador Clésio Luis Tozzi. 2013. Dissertação de Mestrado. Universidade de Campinas, Campinas, 2013.
- [9] Bewley, A; Ge, Z; Ott, L; Ramos, F; Upcroft, B. Simple Online and Realtime Tracking. 2016 IEEE International Conference on Image Processing (ICIP), sep. 2016, ISBN 9781467399616.
- [10] THOMAS, H.R.; HORMAN, M.J. Fundamental Principles of Workforce Management. Journal of Construction Engineering and Management/Volume 132 Issue 1 - January 2006
- [11] ROSEBROCK, A. OpenCV People Counter. [S. l.], 2017. Disponível em: https://www.pyimagesearch.com/2018/08/13/opencv-people-counter/#post_downloads. Acesso em: 18 nov. 2019.
- [12] Público-alvo: o que é e como dialogar com quem você precisa!. [S. l.], 2015. Disponível em: <https://rockcontent.com/blog/publico-alvo/>. Acesso em: 18 nov. 2019.
- [13] Viola, P; Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE Conf Comput Vis Pattern Recognit. 1. I-511. 10.1109/CVPR.2001.990517.
- [14] Levi, G.; Hassner, T. Age and gender classification using convolutional neural networks, 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, 2015, pp. 34-42. CVPRW.2015.7301352.
- [15] How to Use Background Subtraction Methods. [S. l.], 2015. Disponível em: https://docs.opencv.org/3.4/d1/dc5/tutorial_background_subtraction.html. Acesso em: 18 nov. 2019.

[16] IGLESIAS, F. Comportamentos em filas de espera: uma abordagem multimétodos. Orientador Clésio Hartmut Gunter. 2007. Dissertação de Mestrado. Universidade de Brasília, Brasília, 2007.

[17] Marins, Fernando Augusto Silva Introdução à Pesquisa Operacional / Fernando Augusto Silva Marins. - São Paulo: Cultura Acadêmica: Universidade Estadual Paulista, Pró-Reitoria de Graduação, 2011.

[18] BLAKE, Herbert Blake. Waiting Line Models. [S. l.], 2017. Disponível em: <https://www.csus.edu/indiv/b/blakeh/mgmt/documents/OPM101SupplC.pdf>. Acesso em: 18 nov. 2019.

[19] Queue Length Monitoring with Searchlight for Retail. [S. l.], 2015. Disponível em: <https://www.youtube.com/watch?v=5rkwqp6nnr4>. Acesso em: 18 nov. 20.