

CONTAGEM, DETECÇÃO E RASTREAMENTO DE PESSOAS EM UM AMBIENTE MULTICÂMERAS APLICADOS A UMA ARQUITETURA DE SISTEMAS DISTRIBUÍDOS

Iago José Pattas Bastos Franco¹; Anúcio Menezes Correia²; Tiago Silveira de Andrade Cajahyba³

¹ Bolsista Mestrado; PD&I; pattasiago@gmail.com

² Consultor I; SENAI CIMATEC; Salvador-BA; anusio.correia@fieb.org.br

³ Especialista III; SENAI CIMATEC; Salvador-BA; tiago.cajahyba@fieb.org.br

RESUMO

Realizar rastreamento de pessoas e objetos utilizando visão computacional é uma tarefa complexa requisita por muitas áreas de estudo. O problema é ainda maior quando nos deparamos com um ambiente multicâmeras, pois é de se esperar que o rastreamento de uma pessoa/objeto se mantenha por toda a área monitorada. Com este intuito, este trabalho propõe-se a desenvolver e aplicar métodos para detecção, rastreamento e contagem de pessoas num cenário com múltiplas câmeras num ambiente de sistemas distribuídos. Para chegar a uma solução, fez-se o uso de algoritmos de aprendizado profundo como YOLO e estratégias para re-identificação de pessoas, além de ferramentas já utilizadas para realizar rastreamento, como o filtro de Kalman. Por fim, avaliou-se as vantagens e desvantagens dos métodos e a carga de impacto no consumo de tempo na arquitetura de sistemas distribuídos utilizada. Com isso, é demonstrado que os métodos adotados são satisfatórios para realizar as atividades almejadas, porém com algumas ressalvas no que concerne, hardware adotado, resolução da imagem computada e número de pessoas no ambiente, que influenciam negativamente nos tempos de inferência dos algoritmos e nos tempos de latência do sistema distribuído.

PALAVRAS-CHAVE: visão computacional, rastreamento de pessoas, re-identificação de pessoas, sistemas distribuídos.

1. INTRODUÇÃO

Ambos detecção e rastreamento de pessoas são atividades complexas, especialmente quando aplicadas a cenários no mundo real. Ambientes com muitas pessoas, oclusão e incerteza são alguns dos problemas enfrentados nessas tarefas. Sistemas robóticos, sistemas de vigilância, detecção de pedestres e interação humano-máquina são exemplos onde tais esforços são requisitados.

Diversos são os algoritmos de rastreamento encontrados na literatura^{1 2 3 4}, entretanto, para o contexto deste trabalho, a abordagem utilizada, que tem se tornando popular, denomina-se *'tracking-by-detection'* (rastreamento por detecção), que consiste em rastrear um objeto *frame a frame* utilizando a saída de algum algoritmo de detecção.

Para alcançar o objetivo principal deste projeto, que é desenvolver e aplicar métodos para detecção, rastreamento e contagem de pessoas num cenário com múltiplas câmeras num ambiente de sistemas distribuídos, faz-se necessário o uso de abordagens de rastreamento de pessoas e métodos para re-identificação destas quando saem do escopo de visão de uma câmera para outra.

2. METODOLOGIA

A metodologia deste trabalho consistiu inicialmente em: pesquisas relacionadas às técnicas de detecção e rastreamento de pessoas, assim como a implementação preliminar das mesmas. O algoritmo de detecção utilizado para realizar a estratégia "tracking-by-detection", foi o YOLO⁵, ferramenta voltada a detecção de objetos e pessoas, baseada em redes convolucionais e aprendizado profundo. Neste esquema, o detector retorna a localização das pessoas encontradas e o algoritmo de rastreamento resume-se a tarefa de associar rótulos(ID's) às pessoas, comparando o quadro atual com o anterior. Para para ampliar o alcance do rastreamento e lidar com ambientes multicâmeras, um algoritmo de re-identificação de pessoas foi inserido. Para chegar ao objetivo proposto, adotou-se a seguinte etapas na metodologia:

- Utilizar o YOLO e implementar a heurística de contagem de pessoas num ambiente multicâmeras, neste caso, a heurística utilizada para saber quantas pessoas existem numa sala (num ambiente com câmeras voltadas a um ponto comum) foi baseada no maior número de pessoas vistas em uma câmera única, ou seja, se numa sala com 3 câmeras, a 1ª contar 3, a 2ª contar 1 e a 3ª contar 2, o número final de pessoas contadas será 3;
- Utilizar o YOLO junto a um algoritmo de rastreamento, neste caso, foi inserido um simples rastreamento por centroide utilizando distancia euclidiana⁶. Neste esquema, utiliza-se o centroide das *boundboxes* das

pessoas retornadas pelo YOLO, tentando-se associar o indivíduo do *frame* atual com o *frame* passado através da menor distância euclidiana entre os centroides;

- Melhorar os algoritmos de rastreamento utilizando o Índice de Jaccard⁷ e Filtro de Kalman⁸ e introdução do algoritmo de Re-Identificação⁹ de pessoas utilizando aprendizado profundo;
- Coletar os tempos de inferência dos algoritmos de detecção e rastreamento para avaliar latência total gerada no sistema.

Para executar e avaliar estes algoritmos foram utilizadas máquinas com as seguintes configurações de hardware e software:

- GPU – Quadro P2000 com 5 Gb RAM;
- CPU – Intel Xeon w-2133;
- RAM – 32Gb.
- Sistema Operacional Ubuntu 18.04 LTS

As imagens utilizadas para alimentar o algoritmo YOLO possuem resolução VGA.

3. RESULTADOS E DISCUSSÃO

Para a 1ª etapa, a contagem de pessoas é feita utilizando as detecções retornadas pelo YOLO. A heurística utilizada para contar as pessoas detectadas num espaço comum apresenta-se satisfatória quando o esquema de câmeras é voltado a um único ponto.

As lacunas apresentadas pela 1ª etapa foram:

- A heurística escolhida fornece a contagem de pessoas de forma aproximada, já que uma única câmera pode não ser capaz de capturar todas as pessoas do ambiente.
- Esta heurística funciona bem para um esquema específico de posicionamento das câmeras.

Para a 2ª etapa, utilizando a mínima distância entre os centróides das *boundboxes* das pessoas do *frame* atual e passado, o algoritmo gera um ID Local (por câmera) para identificar cada pessoa.

As lacunas apresentadas pela 2ª etapa foram:

- A abordagem com distância euclidiana não lida muito bem com oclusões e quando há baixa taxa de predições por segundo do YOLO, devido a problemas na rede;
- O ID dos indivíduos podem ser confundidos entre si (trocados), ou considerados como novos (novo ID), devido aos problemas anteriores.

Para a 3ª etapa, visou-se atenuar os problemas apresentados pelas oclusões e pelo uso da distância euclidiana, além de, inserir o método de Re-Identificação de pessoas para prover rastreamento em ambientes multicâmeras. Para tal, foi acrescentado: o filtro de Kalman e o Índice de Jaccard para o rastreamento, e um extrator de características baseado em aprendizado profundo⁹ para a re-identificação. Nesta abordagem, o algoritmo de re-identificação decide se um pessoa já é “conhecida” pelo sistema, é “desconhecida”, ou “não tem certeza”. A rede convolucional deste algoritmo de aprendizado profundo extrai 128 características de cada pessoa no ambiente. Após extraídas as características, a cada vez que uma nova pessoa entra em cena ou ocorra alguma perda no rastreamento, o algoritmo de re-identificação é iniciado e tenta encontrar o indivíduo de maior semelhança para re-associar o ID. O ID dado pela abordagem de re-identificação é chamado de ID global, ou seja, é o mesmo ID de um indivíduo em todas as câmeras, diferente do ID gerado pelo rastreamento, que é um ID local (por câmera).

Para aumentar o poder discriminativo do método de Re-Identificação, utilizou-se um filtro para acumular fotos das pessoas e descartar imagens menos relevantes, como as com: baixo score de detecção (parâmetro retornado pelo YOLO), resolução da imagem do indivíduo, e proporção da imagem.

As lacunas apresentadas pela 3ª etapa foram:

- A troca dos ID's devido a oclusão foi reduzida pelo Filtro de Kalman, mas ainda é uma lacuna;
- Demora no processo de re-identificação e por vezes re-identificação errônea quando há muitas pessoas registradas na memória;
- Dificuldade em re-identificar a mesma pessoa em diferentes posições (frente e trás) às vezes gerando mais de um ID para ela.

Para a 4ª e última etapa, avaliou-se as médias dos tempos de inferência dos algoritmos: YOLO, rastreamento e re-identificação na arquitetura utilizada. Estas medidas de tempo podem ser vistas na Tabela 3.1.

Tabela 3.1 Média dos Tempos de Processamento

| | YOLO/(s) | Rastreamento/(s) | Re-Identificação/pessoa (s) |
|-------|----------|------------------|-----------------------------|
| Média | 0.15539 | 0.00149 | 0.08429 |

Estes tempos se mostraram satisfatórios para as máquinas que estavam comportando o sistema, e para um ambiente livre de multidões, ou seja, é necessário um ambiente controlado, rede estável e hardwares robustos para conseguir bom desempenho no sistema distribuído.

4. CONSIDERAÇÕES FINAIS

Diante dos resultados apresentados, é possível notar que, apesar das lacunas na abordagem, a detecção, rastreamento e re-identificação de pessoas num ambiente multicâmeras é possível num sistema distribuído. Um dos pontos fracos que pode surgir neste tipo de arquitetura distribuída é a latência. Na abordagem apresentada, a latência é gerada pelo tempo de processamento dos algoritmos e o envio e recebimento dos dados, isso é piorado a medida que se adiciona câmeras ao ambiente, pois as transferências de imagens entre os processos são custosas mesmo a resolução sendo VGA.

Como necessidades futuras, tem-se o melhoramento das algoritmos apresentados, onde ficam abertos testes para o uso de diferentes técnicas como:

1. Para o rastreamento:
 - Abordagens como "*Skeleton Tracking*" e uso de um filtro de Kalman não linear;
2. Para o algoritmo de re-identificação:
 - Usar fotos da mesma pessoa em diferentes posições para aumentar a taxa de acerto da rede convolucional ou investigar outros modelos de aprendizado profundo para re-identificação de pessoas.

5. REFERÊNCIAS

- ¹ Lukežič, Alan; Vojříř, Tomáš; Čehovin Zajc, Luka; Matas, Jiří, e Kristan, Matej. Discriminative correlation filter tracker with channel and spatial reliability. **International Journal of Computer Vision**, 2018.
- ² Henriques, João F.; Caseiro, Rui; Martins, Pedro e Batista, Jorge. Exploiting the circulant structure of tracking-by-detection with kernels. In proceedings of the **European Conference on Computer Vision**, 2012.
- ³ B. Babenko, M. Yang e S. Belongie, "Visual tracking with online Multiple Instance Learning," **2009 IEEE Conference on Computer Vision and Pattern Recognition**, Miami, FL, 2009, pp. 983-990.
- ⁴ Z. Kalal, K. Mikolajczyk e J. Matas. Forward-backward error: Automatic detection of tracking failures. In **2010 IEEE 20th International Conference on Pattern Recognition (ICPR)**, pages 2756–2759, 2010.
- ⁵ Redmon, et al. "YOLOv3: An Incremental Improvement." **ArXiv.org**, 8 Apr. 2018, Disponível em: <arxiv.org/abs/1804.02767> Acesso em: 23 mai. 2016.
- ⁶ DISTÂNCIA EUCLIDIANA. In: **WIKIPÉDIA, a enciclopédia livre**. Flórida: Wikimedia Foundation, 2016. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Dist%C3%A2ncia_euclidiana&oldid=45689943>. Acesso em: 23 mai. 2016.
- ⁷ Bewley, Alex et al. "Simple Online and Realtime Tracking." **2016 IEEE International Conference on Image Processing (ICIP)** (2016):
- ⁸ S. S. Pathan, A. Al-Hamadi e B. Michaelis, "Intelligent feature-guided multi-object tracking using Kalman filter", **2009 2nd International Conference on Computer, Control and Communication**, Karachi, 2009, pp. 1-6.
- ⁹ N. Wojke and A. Bewley, "Deep Cosine Metric Learning for Person Re-identification", **2018 IEEE Winter Conference on Applications of Computer Vision (WACV)**, Lake Tahoe, NV, 2018, pp. 748-756.