



Literature Review

Course: Research Methodology in Computer Science (AI800)

Research Title: Improving the Accuracy of Real-Time Human Counting Using Surveillance Cameras

Student Name: Khalid Abbas Eltijani Elzein (261001961)

1. Introduction

Human counting using surveillance cameras is used almost everywhere now, especially in places like stations, malls, and large events. The main idea is simple, count how many people are there and try to understand how they move.

But once you look closer, the problem is not that simple. In real scenes, people overlap, lighting changes, and sometimes the camera angle itself makes detection harder.

Most recent work depends on deep learning models like CNNs and tracking algorithms. These approaches improved the results quite a bit, but they still don't behave well in all situations.

In this review, I looked at six papers that deal with crowd counting and tracking. The aim here is not just to list what they did, but to see where they work well and where they clearly struggle.

1.1 Zhang et al. (2016) – MCNN

This paper introduced a multi-column CNN model to handle the problem of scale variation in crowd counting. The idea is to detect people at different sizes within the same image, which is common in surveillance cameras.

The results show that the model improves accuracy when the scene contains different crowd densities. However, the model is still relatively simple compared to newer approaches, and it does not include any tracking mechanism. Because of that, it cannot be used directly in systems that require movement analysis.

1.2 Li et al. (2018) – CSRNet

This study proposed CSRNet, which uses dilated convolution to improve counting performance in highly crowded scenes. The model focuses on generating better density maps instead of detecting each person individually.

The results are strong, especially in congested environments where traditional detection methods fail. Still, the system only provides total counts and does not track individuals, which limits its use in real-time monitoring applications.

1.3 Cao et al. (2018) – SANet

This paper introduced an encoder-decoder model with scale aggregation to improve both accuracy and efficiency. The model combines features from different levels to handle variation in crowd density.

The approach shows a good balance between performance and computational cost. However, the evaluation is mostly based on datasets, and there is limited discussion about how the model performs in real-time surveillance systems.

1.4 Jiang et al. (2019) – TEDnet

This research proposed a trellis encoder-decoder network that focuses on improving the quality of density maps. The idea is to preserve more spatial details while still capturing global information.

The results show better density estimation compared to previous models. However, the network is more complex and requires higher computational resources, which makes real-time implementation more difficult.

1.5 Yan et al. (2019) – PGCNet

This study focused on the problem of perspective distortion in surveillance images. People appear smaller when they are far from the camera, which affects counting accuracy.

The proposed model uses perspective-aware techniques to handle this issue and improve performance. The results are promising, but the method depends on perspective information, which may not always be available in real-world setups.

1.6 Wojke et al. (2017) – Deep SORT

This paper introduced Deep SORT, a tracking method that helps follow individuals across video frames. Although it is not a pure counting method, it is very useful in surveillance systems because it reduces double counting.

The approach improves tracking accuracy by combining motion prediction with appearance features. However, its performance still depends on the quality of detection. In crowded scenes, tracking can fail when people overlap heavily.

2. Critical Analysis

Looking at the papers together, you can see a pattern. Most of them try to push accuracy higher, mainly by using deeper models or better feature extraction.

Still, the same problems keep showing up. Occlusion is probably the biggest one. When people overlap, the model starts missing detections or counting wrong.

Lighting is another issue. Some models work fine in good conditions, but once the scene changes, the results drop.

Tracking methods help a bit, especially to avoid counting the same person more than once. But again, if detection is weak, tracking won't fix much.

Honestly, what stands out is that many results look strong in datasets, but when you think about real deployment, things are less stable. So there's still a gap here.

3. Connections Between Papers

All the papers are trying to solve the same core issue, how to count people more accurately.

Most of them rely on deep learning, especially CNN-based models. Even when the methods are different, the base idea is very similar.

Another thing in common is the challenges they deal with. Crowd density, overlapping people, and scale differences appear in almost every study.

Also, most of the evaluation is done on prepared datasets. That probably explains why performance looks better than what you would expect in real environments.

4. Conflicts Between Papers

The main difference is in how each paper approaches the problem.

Some focus on detection, where the model tries to find each person directly. Others use density estimation, which skips detection and just predicts how many people are in the scene.

Tracking papers, like Deep SORT, go in a different direction. They don't focus on counting only, but on following people across frames.

There's also a clear trade-off. Some models are more accurate, but they are heavier and slower. Others are faster but less reliable in crowded scenes.

5. Summary Table

Paper	Method	Result	Limitation
Zhang et al. (2016) – MCNN	Multi-column CNN	Works better with different crowd sizes	No tracking and relatively old model
Li et al. (2018) – CSRNet	Dilated CNN	Good results in dense scenes	No tracking support
Cao et al. (2018) – SANet	Encoder-decoder	Balanced performance	Not clearly tested in real-time
Jiang et al. (2019) – TEDnet	Trellis network	Better density maps	Heavy computation
Yan et al. (2019) – PGCNet	Perspective-aware model	Handles scale better	Needs extra information
Wojke et al. (2017) – Deep SORT	Tracking method	Improves tracking accuracy	Depends on detection quality

6. References

1. Zhang, C., et al. (2016). MCNN. CVPR.
2. Li, Y., et al. (2018). CSRNet. CVPR.
3. Cao, X., et al. (2018). SANet. ECCV.
4. Jiang, X., et al. (2019). TEDnet. CVPR.
5. Yan, Z., et al. (2019). PGCNet. ICCV.
6. Wojke, N., et al. (2017). Deep SORT. ICIP.