



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

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)

Problem Statement

Human counting using surveillance cameras has become a common task in many real-world applications such as transportation stations, shopping malls, airports, and large public events. These systems are used to estimate how many people are present in a scene and sometimes to monitor their movement over time. This information plays an important role in crowd management, safety monitoring, and resource planning.

Although the idea of counting people looks simple, the problem becomes much more complex in real-world scenarios. One of the main challenges is occlusion, where people overlap with each other, especially in crowded environments. When occlusion occurs, the system may fail to detect some individuals or count them incorrectly. This leads to undercounting or overcounting, both of which reduce the reliability of the system.

Another major issue is lighting variation. In many real-world situations, lighting conditions are not stable. For example, outdoor environments change throughout the day, and indoor environments may include shadows or uneven lighting. Many existing models perform well under controlled lighting conditions but show weaker performance when lighting changes.

Camera perspective is also an important factor. People who are far from the camera appear smaller, which makes detection more difficult. This is known as the scale variation problem. In addition, changes in camera angle can affect how objects are viewed, which impacts the accuracy of both detection and counting.

From the literature review, it is clear that many approaches have been proposed to improve human counting. For example, CNN-based models such as MCNN were designed to handle scale variation by using multiple columns to detect people at different sizes (Zhang et al., 2016). Later, CSRNet used dilated convolution to improve performance in dense crowd scenes (Li et al., 2018). These methods showed strong results in datasets, especially in highly crowded environments.

Other approaches, such as SANet (Cao et al., 2018) and TEDnet (Jiang et al., 2019), focused on improving feature extraction and producing better density maps. These methods improved accuracy but often required higher computational resources, which makes real-time use more difficult.

In addition, perspective-aware methods like PGCNet (Yan et al., 2019) attempted to address the issue of people appearing smaller at different distances from the camera. While these methods improved performance, they depend on additional information such

as perspective maps, which may not always be available in real-world systems.

Tracking-based methods also play an important role. For example, Deep SORT (Wojke et al., 2017) is widely used to track individuals across video frames. This helps reduce double counting and improves consistency over time. However, tracking methods depend heavily on detection quality. If detection fails due to occlusion or poor lighting, tracking will also fail.

Despite these improvements, the same limitations appear across many studies. Most models are evaluated using prepared datasets, where conditions are more controlled than real-world environments. In practice, surveillance systems must handle dynamic conditions, including crowd density changes, lighting variation, and camera movement.

This creates a gap between research results and real-world performance. Many models achieve high accuracy in datasets but do not maintain the same level of performance when deployed in real environments. Therefore, there is a need to develop a more robust human counting approach that can handle these challenges and provide reliable results in real-time applications.

Research Objectives

The main objective of this research is to improve the accuracy of real-time human counting systems in real-world surveillance environments.

- To analyze the main challenges that affect human counting accuracy, including occlusion, lighting variation, and camera perspective
- To study existing crowd counting methods such as detection-based models, density estimation approaches, and tracking algorithms, and identify their limitations
- To compare the performance of different approaches under real-world conditions rather than only controlled datasets
- To design an improved method that combines the strengths of existing techniques to reduce counting errors
- To implement the proposed method using available datasets or real-time surveillance data
- To evaluate the performance of the proposed method using metrics such as counting accuracy, error rate, and robustness under different conditions
- To analyze how the proposed method performs in crowded and dynamic environments compared to existing approaches

Link to Literature Review

The literature review shows that most research focuses on improving counting accuracy using deep learning techniques. CNN-based models such as MCNN and CSRNet address issues related to crowd density and scale variation. These models are effective in many cases but do not include tracking, which limits their use in applications that require movement analysis.

Encoder-decoder models such as SANet and TEDnet improve feature representation and produce better density maps. However, they require more computation and are not always suitable for real-time systems. Perspective-aware models attempt to handle scale differences but depend on additional information that may not be available.

Tracking methods like Deep SORT help improve consistency across video frames by following individuals over time. However, their performance depends on the accuracy of detection, which is still affected by occlusion and lighting conditions.

Overall, the literature confirms that the main challenges remain the same across different studies. These include overlapping people, changing lighting conditions, and the difference between dataset evaluation and real-world deployment. This supports the need for further research to develop more reliable and adaptive human counting systems.

References

- Zhang, C., et al. (2016). Multi-column convolutional neural network for crowd counting.
- Li, Y., et al. (2018). CSRNet: Dilated convolutional neural networks for crowd counting.
- Cao, X., et al. (2018). Scale aggregation network for accurate crowd counting.
- Jiang, X., et al. (2019). TEDnet: Trellis encoder-decoder network for crowd counting.
- Yan, Z., et al. (2019). Perspective-guided crowd counting.
- Wojke, N., et al. (2017). Deep SORT: Simple online and real-time tracking.