Low-Rank Matrix Factorization for Moving Object Detection in Video Data

Cristian Pena[1], Xiaopeng Li[2], Cédric Josz[2]

[1] Department of Ocean and Mechanical Engineering, Florida Atlantic University, Boca Raton

[2] Department of Industrial Engineering and Operations Research, Columbia University, New York

Introduction: This study investigates the application of low-rank matrix factorization for moving object detection in video data. The primary goal is to enhance the extraction of moving objects from a sequence of video frames by approximating the data matrix representing the video as a rank-one matrix plus noise. This approach addresses the challenge of isolating moving elements within a video stream, contributing significantly to fields such as surveillance, traffic monitoring, and video analytics [1, 2].

Methods: The video is encoded as a data matrix $M \in \mathbb{R}^{mxn}$, where each column n corresponds to a video frame and each row m to a pixel. The objective is to find the closest rank-one matrix xy^T to the data matrix M. This problem is formulated as minimizing the sum of squared differences (L2 norm) between xy^T and M: $\min_{x,y} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(x_i y_j - M_{ij}\right)^2$. We also implemented an L1 norm version, which minimizes the sum of absolute differences: $\min_{x,y} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \left|x_i y_j - M_{ij}\right|$. To solve these problems, we use Stochastic Gradient Descent (SGD), an iterative optimization technique commonly used in machine learning to minimize complex functions. The update rules derived from the objective function are applied in Python.

Results: The implementation effectively isolated moving objects from the static background. The results were validated using both synthetic and real video data. Empirical analysis showed that the objective function behaves well, being absent of any spurious local minima. Comparative analysis revealed that the L1 norm was significantly more effective, improving accuracy by approximately 131% compared to the L2 norm.

Conclusions: This research confirms that low-rank matrix factorization, combined with SGD, is a viable method for moving object detection in videos. In particular, the L1 norm can effectively separate moving objects and a background frame from a video with minimal error. Future work will explore extending this approach to more complex scenarios, such as dynamic backgrounds, and implementing different optimization techniques such as momentum SGD and adaptive momentum SGD. The findings have significant implications for improving automated video analysis technologies, potentially leading to advancements in security, traffic management, and entertainment industries.

References:

1. Bouwmans, T. et al. Comput. Sci. Rev., 2017;122:80-93
2. Li, J. et al. Sensors, 2022;22(15):58-77

Acknowledgements:

NSF Grant (#2023032)