Algorithms and tools for Big Data Application - Comparison of Tensor and Matrix based Approach in Foreground\Background Separation

Suhail Ahamed

Jianshu Zhang

Responsible Professor: Univ.-Prof. Dr. –Ing. Martin Haardt

Outline

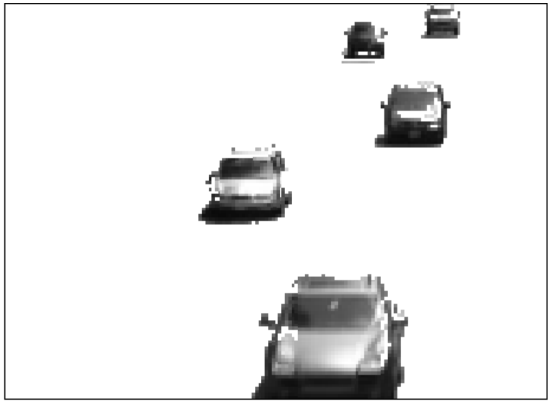
- Motivation and State of the art
- Subspace estimation
  - ⇒ Matrix based
  - ⇒ Tensor based
- Simulation results
- Conclusion

Motivation and State of the Art

- Identifying moving objects is a fundamental and critical task in many video applications. It is a **big data application** since the resolution of each frame might be huge or the video consists a large number of frames

Original video

Background



Foreground

- **Compressed sensing based approaches** [1]-[2] has gained great attraction in recent years ,e.g. the Recursive Projected Compressive Sensing (ReProCS). These methods use matrices (not taking into account the R-D structure)
- Our **task** is to investigate whether and when it is beneficial to use tensor based approach

Foreground and background Separation via Recursive Projected Compressive Sensing

- Assume that the video sequence consists of a slowly changing **low-rank** background and a **sparse** foreground. For each frame, we have

$$\underbrace{\mathbf{x}_t}_{\textbf{Measurement}} = \underbrace{\mathbf{b}_t}_{\textbf{Low dimensional background}} + \underbrace{\mathbf{f}_t}_{\textbf{Sparse foreground}}$$

- Training phase: Estimating the subspace (U_s) of the slowly varying background
- Data processing phase

⇒ **Perpendicular Projection:** project the measurement vector ($\mathbf{x}_t$) into the space orthogonal to the U_s

$$\mathbf{y}_t := \Phi_t \mathbf{x}_t \qquad \Phi_t := (I - U_s U_s')$$

⇒ **Sparse Recovery:** basis pursuit is used to estimate the foreground $\mathbf{f}_t$,solving

$$min_{\mathbf{z}} ||\mathbf{z}||_1 s.t. ||\mathbf{y}_t - \Phi_t \mathbf{z}||_2 \leq \xi \qquad \xi = ||\Phi_t \mathbf{b}_{t-1}||_2$$

⇒ **Recover $\mathbf{b}_t$** as: $\mathbf{b}_t = \mathbf{x}_t - \mathbf{f}_t$

Existing matrix based subspace estimation

- The 3-mode unfolding $X \in \mathbb{R}^{(w \cdot h) \times T}$ of training video volume $\mathcal{X} \in \mathbb{R}^{w \times h \times T}$ is used and we have

$$X = [\mathcal{X}]'_{(3)}$$

- **Low-rank approximation** of X using the truncated **Singular Value Decomposition**

$$X = U \cdot \Sigma \cdot V^H \approx U_s \cdot \Sigma_s \cdot V_s^H$$

- The matrix based **Background Subspace** is decided by

$$U_s = U(:, 1 : \hat{r}) \qquad \hat{r}, \text{s.t.} \sum_{i=1}^{\hat{r}} \lambda_i = \text{b\% of energy}$$

we use b=95%

Proposed tensor based subspace estimation

- Higher-Order Singular Value Decomposition **(HOSVD)** of $\mathcal{X} \in \mathbb{R}^{w \times h \times T}$

$$\mathcal{X} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$$

- **Truncated HOSVD**

$$\mathcal{X} \approx \mathcal{S}^{[s]} \times_1 U_1^{[s]} \times_2 U_2^{[s]} \times_3 U_3^{[s]}$$

- The n-ranks are decided via the n-mode unfoldings
- The tensor based **background subspace**

$$\mathcal{U}^{[s]} = \mathcal{S}^{[s]} \times_1 U_1^{[s]} \times_2 U_2^{[s]} \times_3 \left(\Sigma_3^{[s]}\right)^{-1}$$

- A link between the **matrix** based and the **tensor** based subspace estimates

$$\boxed{[\mathcal{U}^{[s]}]'_{(3)} = (T_1 \otimes T_2) \cdot U_s} \qquad T_i = U_i^{[s]} \cdot U_i^{[s]H}$$

$\Rightarrow$ **do not need** to compute the **core tensor**, only U_1, U_2, U_3.