AMATH 482-582 Winter Quarter 2018

Homework 5: Background Subtraction in Video Streams

DUE: Friday, March 15, 2019

Use the Dynamic Mode Decomposition method to take a video clip containing a foreground and background object and separate the video stream to both the foreground video and a background. Make several test videos to try the algorithm on. You can use your smart phone to generate appropriate videos.

The DMD spectrum of frequencies can be used to subtract background modes. Specifically, assume that ω_p, where $p \in \{1, 2, \ldots, \ell\}$, satisfies $\|\omega_p\| \approx 0$, and that $\|\omega_j\| \;\forall\; j \neq p$ is bounded away from zero. Thus,

find minimum omega frequency; this is the background

$$\mathbf{X}_{\text{DMD}} = \underbrace{b_p \boldsymbol{\varphi}_p e^{\omega_p \mathbf{t}}}_{\text{Background Video}} + \underbrace{\sum_{j \neq p} b_j \boldsymbol{\varphi}_j e^{\omega_j \mathbf{t}}}_{\text{Foreground Video}} \tag{1}$$

Assuming that $\mathbf{X} \in \mathbb{R}^{n \times m}$, then a proper DMD reconstruction should also produce $\mathbf{X}_{\text{DMD}} \in \mathbb{R}^{n \times m}$. However, each term of the DMD reconstruction is complex: $b_j \boldsymbol{\varphi}_j \exp(\omega_j \mathbf{t}) \in \mathbb{C}^{n \times m} \;\forall j$, though they sum to a real-valued matrix. This poses a problem when separating the DMD terms into approximate low-rank and sparse reconstructions because real-valued outputs are desired and knowing how to handle the complex elements can make a significant difference in the accuracy of the results. Consider calculating the DMD's approximate low-rank reconstruction according to

$$\mathbf{X}_{\text{DMD}}^{\text{Low-Rank}} = b_p \boldsymbol{\varphi}_p e^{\omega_p \mathbf{t}}.$$

Low-rank DMD; associated to background

Since it should be true that

$$\mathbf{X} = \mathbf{X}_{\text{DMD}}^{\text{Low-Rank}} + \mathbf{X}_{\text{DMD}}^{\text{Sparse}},$$

then the DMD's approximate sparse reconstruction,

$$\mathbf{X}_{\text{DMD}}^{\text{Sparse}} = \sum_{j \neq p} b_j \boldsymbol{\varphi}_j e^{\omega_j \mathbf{t}},$$

can be calculated with real-valued elements only as follows...

$$\mathbf{X}_{\text{DMD}}^{\text{Sparse}} = \mathbf{X} - \left|\mathbf{X}_{\text{DMD}}^{\text{Low-Rank}}\right|,$$

where $|\cdot|$ yields the modulus of each element within the matrix. However, this may result in $\mathbf{X}_{\text{DMD}}^{\text{Sparse}}$ having negative values in some of its elements, which would not make sense in terms of having negative pixel intensities. These residual negative values can be put into a $n \times m$ matrix $\mathbf{R}$ and then be added back into $\mathbf{X}_{\text{DMD}}^{\text{Low-Rank}}$ as follows:

R stores our moving object. Subtract it from our low-rank matrix (background) and add to sparse matrix (foreground)

$$\mathbf{X}_{\text{DMD}}^{\text{Low-Rank}} \leftarrow \mathbf{R} + \left|\mathbf{X}_{\text{DMD}}^{\text{Low-Rank}}\right|$$

$$\mathbf{X}_{\text{DMD}}^{\text{Sparse}} \leftarrow \mathbf{X}_{\text{DMD}}^{\text{Sparse}} - \mathbf{R}$$

This way the magnitudes of the complex values from the DMD reconstruction are accounted for, while maintaining the important constraints that

$$\mathbf{X} = \mathbf{X}_{\text{DMD}}^{\text{Low-Rank}} + \mathbf{X}_{\text{DMD}}^{\text{Sparse}},$$

so that none of the pixel intensities are below zero, and ensuring that the approximate low-rank and sparse DMD reconstructions are real-valued. This method seems to work well empirically.

NOTE: it is pretty easy to produce a video on your smart phone.