Algorithms and tools for Big Data Application - Foreground and background separation

Suhail Ahamed

Jianshu Zhang

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

Different approaches

- Classical statistical approach[5]
  - ⇒ Frame differencing
  - ⇒ Median filter
  - ⇒ Mean filter
- Compressed sensing approach[1-4]
  - ⇒ Usage of robust PCA (Principle Component Analysis) method

Classical statistical approach

- Frame Differencing
  - ⇒ The background for the current frame is assumed to be the previous frame
  - ⇒ The difference between the current frame and the previous frame is calculated
- Median/mean filtering
  - ⇒ Obtaining the temporal median/mean of the training sequence -> obtaining the background
  - ⇒ Compare each frame with this background to obtain this background
- Drawbacks
  - ⇒ Accuracy depends on object speed and frame rate
  - ⇒ Have relatively high memory requirements
  - ⇒ Global threshold (not a function of time)

- **Assumption 1**: Slowly changing low-dimensional subspace (background)
  - $||(\boldsymbol{I} - \boldsymbol{P}_{(t-1)}\boldsymbol{P}'_{(t-1)})\boldsymbol{k}_t||_2 \ll \min(||\boldsymbol{l}_t||_2, ||\boldsymbol{s}_t||_2)$
- **Assumption 2**: Dense background
  - Changes in background are mainly global changes
  - Hence sparse vectors are recoverable from : $(\boldsymbol{I} - \boldsymbol{P}_{(t-1)}\boldsymbol{P}'_{(t-1)})\boldsymbol{m}_t = (\boldsymbol{I} - \boldsymbol{P}_{(t-1)}\boldsymbol{P}'_{(t-1)})\boldsymbol{s}_t$
- **Assumption 3**: Changes in $\boldsymbol{s}_t$ are small

$\boldsymbol{m}_t$ - Measurement vector ($\boldsymbol{m}_t = \boldsymbol{l}_t + \boldsymbol{s}_t$)

$\boldsymbol{l}_t$ - Low-dimensional background vector

$\boldsymbol{s}_t$ - sparse vector (foreground)

$\boldsymbol{P}_t$ – orthonormal basis of low-dimensional subspace

Recursive Projected Compressive Sensing (ReProCS) Approach

- The video is treated to be composed of a slowly changing low-dimensional subspace and a sparse component
- Calculate the projection matrix (***P***$_0$) of the training sequence (***M***$_{train}$)

$$\hat{P}_0 = approx - basis(M_{train}, b\%)$$

- **Perpendicular Projection:** project the measurement vector (***m***$_t$) into the space orthogonal to range(***p***$_t$) to get the projected measurement vector
- **Sparse Recovery:** Multiplying the obtained orthogonal complement to the current frame (***m***$_t$) will eliminate the background. Recover ***s***$_t$ by solving :

$$\min_x ||\boldsymbol{x}||_1 \text{ s.t. } ||\boldsymbol{y}_t - (\boldsymbol{I} - \boldsymbol{P}_{(t-1)}\boldsymbol{P}'_{(t-1)})\boldsymbol{x}||_2 \leq \xi$$

$\Rightarrow$ compressed sensing algorithm, e.g., basis pursuit, is used

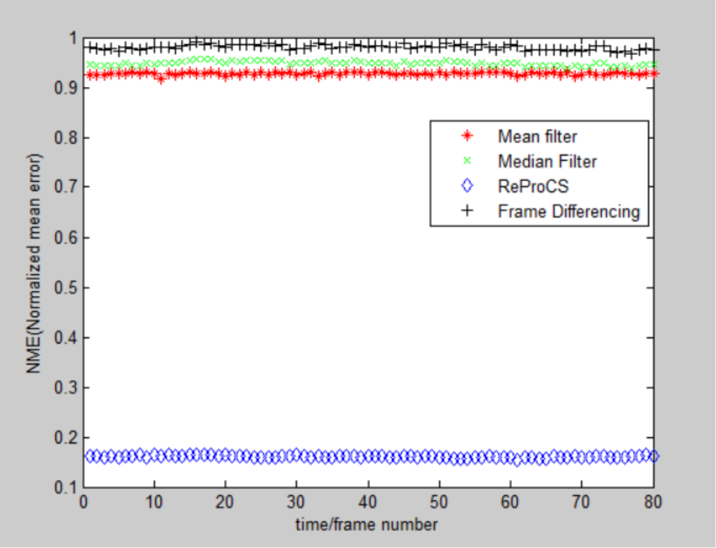
- **Recover l_t** as : $\boldsymbol{l}_t = \boldsymbol{m}_t - \boldsymbol{s}_t$
- **Subspace Update** (update ***P***$_t$) : projection PCA is used to update the subspace every few frames[]

CPU time difference

- Time to execute to training(*1420 frames*) , as well as the real video (*80 frames*) for the lake sequence
- Frame Differencing : 0.593 s
- Median Filter : 341.790 s
- Mean Filter : 8.877 s
- ReProCS : 43.33 s

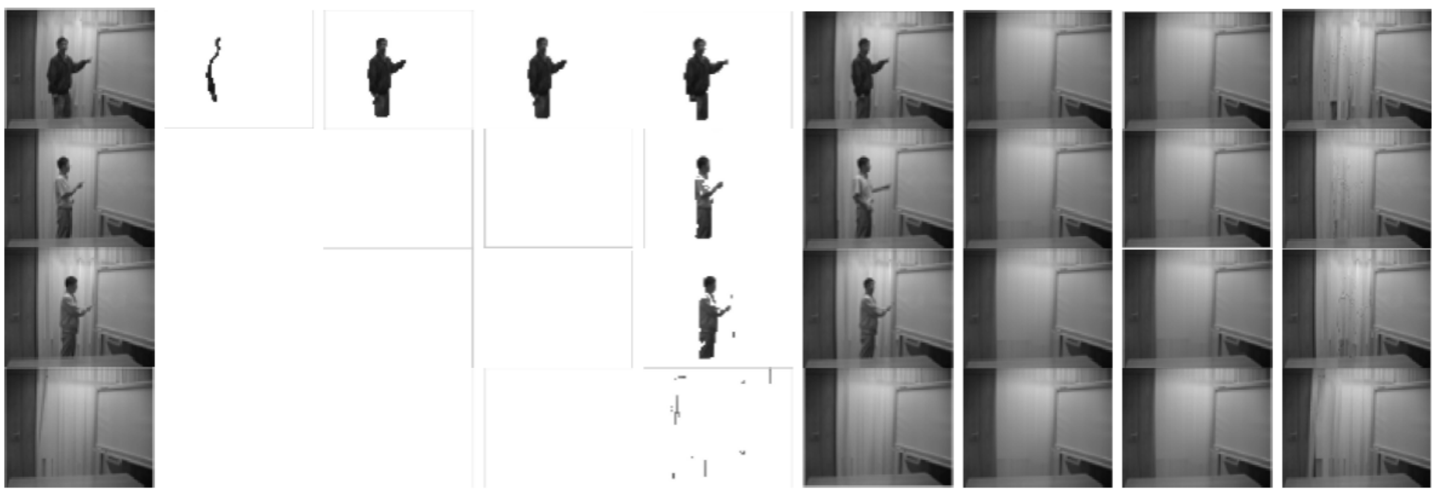
Simulation results: Simulated video sequence

- The classical statistical approaches fail to distinguish the foreground/background
- The compressed sensing based approach performs much better with negligible error



Original frame	Mean Filter (fg)	Median Filter (fg)	ReProCS(fg)	Mean Filter (bg)	Median Filter (bg)	ReProCS (bg)

- *1755* training frames and *1209* frames in real video sequence
- All the methods except for frame differencing perform well when the person is wearing a *black shirt* (white background)
- The statistical approach fail to distinguish when the person is wearing a *white shirt*



Original frame | Frame Diff.(fg) | Mean Filter (fg) | Median Filter(fg) | ReProCS (fg) | Frame Diff.(bg) | Mean Filter (bg) | Median Filter(bg) | ReProCS (bg)

Fig. 4: Curtain video sequence at $t = t_{train} + 74, 477, 1084, 1029$ and its foreground(fg) and background(bg) result using frame differencing, mean filter, median filter and ReProCS algorithms

Simulation results: Real video sequence (2)

- *200* training frames and *52* frames in real video sequence

Original frame | Frame Diff.(fg) | Mean Filter (fg) | Median Filter(fg) | ReProCS (fg) | Frame Diff.(bg) | Mean Filter (bg) | Median Filter(bg) | ReProCS (bg)

Fig. 2: Person video sequence at $t = t_{train} + 42, 45, 52$ and its foreground(fg) and background(bg) result using frame differencing, mean filter, median filter and ReProCS algorithms

Conclusion

- Compressed sensing based approach outperforms the classical statistical approach for the simulated data
- Median filter based background estimation required relatively a lot of memory and takes a lot of time
- The Frame differencing method yields poor results when either the object speed or the frame rate is low
- Since the global threshold is not a function of time in any of the statistical approaches, it is highly susceptible to global changes in the background
- Compressed sensing based approach performs better when there are more training

Future plans

- None of these methods use the temporal component of the video sequence
- Improve upon these methods methods by using the three dimensional nature of video sequences
- Replace the matrix based subspace estimation using the tensor based subspace estimation

References

[1] Han Guo, Chenlu Qiu, Namrata Vaswani, ”An Online Algorithm for Separating Sparse and Low-dimensional Signal Sequences from their Sum,” IEEE Transactions Signal Processing, vol. 62, pp. 4284-4297, Jun. 2014

[2] C. Qiu, N. Vaswani, B. Lois, and L. Hogben, ”Recursive robust PCA or recursive sparse recovery in large but structured noise,” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5954-5958, May 2013

[3] J. Grosek and J. Nathan Kutz, ”Dynamic Mode Decomposition for Real-Time Background/Foreground Separation in Video,” IEEE Transactions on Pattern Analysis and Machine Intelligence, arXiv:1404.7592, Apr. 2014

[4] Andrew E. Waters, Aswin C. Sankaranarayanan, Richard G. Baraniuk, ”SpaRCS: Recovering Low-Rank and Sparse Matrices from Compressive Measurements,” Neural Information Processing Systems (NIPS), Dec. 2011

[5] Birgi Tamersoy, ”Background Subtraction”, Sep. 2009