ADVANTAGES OF SUBSPACE ESTIMATION TECHNIQUES OVER GAUSSIAN MIXTURE MODELS FOR BACKGROUND SUBTRACTION OF NOISY VIDEOS

Pooya Khorrami, Xingqian Xu, Mert Dikmen, Thomas S. Huang

University of Illinois - Urbana Champaign
Department of Electrical and Computer Engineering
Beckman Institute, 405 N. Matthews Avenue, Urbana IL, 61801

ABSTRACT

The abstract should appear at the top of the left-hand column of text, about 0.5 inch (12 mm) below the title area and no more than 3.125 inches (80 mm) in length. Leave a 0.5 inch (12 mm) space between the end of the abstract and the beginning of the main text. The abstract should contain about 100 to 150 words, and should be identical to the abstract text submitted electronically along with the paper cover sheet. All manuscripts must be in English, printed in black ink.

Index Terms— One, two, three, four, five

1. INTRODUCTION

In the field of video surveillance, the user typically wishes to extract meaningful and salient information from a video sequence in a completely automatic fashion. In several instances, the video sequences are captured using stationary cameras leading to a relatively static scene layout. The absence of camera motion implies that the background of the video sequence exhibits very little variation while dynamic changes in the scene represent the objects of interest. In such a case, the most common approach is to perform background subtraction to separate said dynamic regions (i.e. foreground) from the background of each frame.

When performing background subtraction, some naïve methods include frame differencing and approximate median [1]. While these algorithms are simple to use and quite efficient, the most popular technique, by far, is adaptive Gaussian Mixture Models (GMMs) [2, 3]. These works posit that each pixel in the background image can be represented by a probability distribution formed by a mixture of Gaussians. If a pixel greatly deviates from its corresponding model, then the pixel is labeled foreground. While the number of Gaussians used at each pixel is usually fixed, there has been some work [4] that adaptively selects the number of mixture components.

Although background subtraction via Gaussian Mixture Models enjoys widespread use in the computer vision community, it is not without drawbacks. As opposed to the frame differencing and approximate median techniques, GMMs possess several parameters that must be individually tuned. This implies that the algorithm is innately sensitive to different scene configurations. Therefore it should be no surprise that GMMs tend to perform rather poorly on noisy videos where the foreground objects are not immediately distinguishable.

When dealing with noisy video sequences, we advocate the use of low-rank subspaces for background subtraction. Given that each of the video sequences was obtained using a stationary camera, the high level of temporal redundancy between the frames suggests that the backgrounds lie on a low-dimensional subspace. Therefore, foreground activity can be thought of as sparse deviations from said subspace. In this paper, we consider two different subspace estimation algorithms and show empirically how they both achieve superior performance to GMMs on noisy traffic video sequences.

The first method, Robust Principal Component Analysis (RPCA or Robust PCA) [5], attempts to represent a data matrix (or video sequence) as the sum of a low-rank matrix and a sparse matrix via a convex optimization problem. While typically performed in batch on purely intensity videos, we describe how performing Robust PCA on each of the color channels of video and using a reduced number of frames leads to additional performance gains. If a user requires a real-time alternative, we also consider the newly proposed Grassmannian Robust Adaptive Subspace Tracking Algorithm (GRASTA) by He et al. [6] that learns the aforementioned low-rank subspace by subsampling the video frames and proceeds in an online fashion.

The remainder of this paper is organized as follows. Section 2 will describe the improvements made to the batch Robust PCA algorithm. Section 3 will present our experimental setup and findings. Section 4 will describe our conclusions and directions for future work.

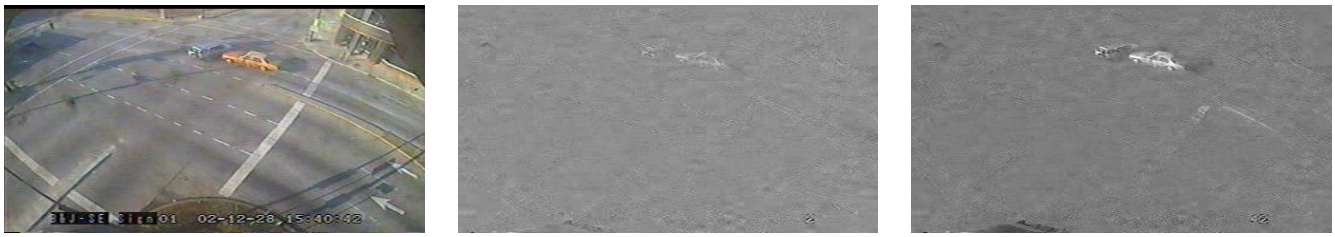
2. METHOD DESCRIPTION

When given a data matrix M, the goal of Robust PCA is to decompose it into a low-rank matrix L and a sparse matrix S.

Thanks to XYZ agency for funding.

(a) Original Image (b) Intensity Sparse Image (c) Red Channel Sparse Image

Fig. 1. Visual Comparison of Robust PCA on the Intensity Channel and the Red Color Channel



(a) Original Image (b) Robust PCA Sparse Image (c) Reduced Robust PCA Sparse Image

Fig. 2. Reduced Frames Comparison

To do this, each of the video frames is vectorized and stacked as a column in the matrix M. Upon closer inspection, one will notice that the columns are highly correlated. This is expected considering that the video was obtained using a stationary camera. Given that the background of a video lies on a low-dimensional subspace, we will see that the columns of L contain the background of each video frame and S contains the sparse deviations from that background, better known as foreground. Therefore, when doing background subtraction, one simply performs Robust PCA on the entire sequence and does further processing solely on the sparse error frames.

2.1. Robust PCA on Individual Color Channels

While many works acknowledge the merit of using Robust PCA for background subtraction, very few have introduced modifications to improve foreground detection. As such, we consider performing Robust PCA on all three of the color channels of a video sequence rather than just on the intensity channel. Now for each frame we have three sparse responses corresponding to each of the color channels. To ensure that an object detected with high confidence in one color channel is maintained in the final result, we take the maximum response across the three color channels.

Consider the images in figure 1. After Robust PCA is performed on the intensity channel of the video sequence, we see in figure 1b that several of the moving cars can be easily separated from the background just using their intensities. The red car, however, does not seem to stand out from the background intensity. On the other hand, when considering the image in figure 1c, we see that doing Robust PCA on the red color channel makes the red car very distinguishable. Given that we also take the maximum response across all the color channels, this high response will remain intact when extracting a foreground mask.

2.2. Reducing Number of Frames and Post-Processing

Now that we have a protocol in place to better detect colored objects, we turn our attention to detecting objects that have stopped moving. Ordinarily when a object has been stationary for a long period of time, the background subtraction algorithm considers it is considered to be a part of the background. Similarly with Robust PCA, when an object is stationary it is no longer considered a sparse corruption of the background. Rather, it is considered a basis element of the low-rank subspace and enters the frame's background. In figure 2, the cars shown in figure 2a have been stationary for some time, therefore they are no longer considered foreground and appear very faintly in figure 2b.

In order to prevent this from happening, we reduce the number of frames used when computing Robust PCA by taking frames spaced 10 frames apart. By reducing the number of frames in which the stationary object appears, we are making the object look like sparse corruption and thereby making it less likely to enter the low-rank subspace that represents the background. The appearance of the two stationary cars in figure 2c verifies our hypothesis. The reader may have noticed that by using a reduced number of frames in Robust PCA, the remaining frames will not have their own low-rank and

sparse component. This is not a problem given that the scene was captured using a stationary camera. For each frame, we can simply take the nearest background image and subtract it to extract the sparse component.

After obtaining the sparse component of each video frame, we notice that they possess a fair amount of speckled noise. A classical approach to de-noising an image is through shrinking the wavelet coefficients via a hard or soft threshold [7]. We apply a hard threshold to each sparse image to obtain a smoother response to be used during evaluation.

3. EXPERIMENTAL RESULTS

(Cite Toyota Dataset?)

4. CONCLUSIONS

5. REFERENCES

[1] N. J. B. McFarlane and C. P. Schofield, "Segmentation and tracking of piglets in images," *Machine Vision and Applications*, vol. 8, no. 3, pp. 187–193, May 1995.

[2] Nir Friedman and Stuart Russell, "Image segmentation in video sequences: A probabilistic approach," 1997, pp. 175–181.

[3] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, Los Alamitos, CA, USA, Aug. 1999, vol. 2, pp. 246–252 Vol. 2, IEEE.

[4] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition*, aug. 2004, vol. 2, pp. 28 – 31 Vol.2.

[5] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *CoRR*, vol. abs/0912.3599, 2009.

[6] Jun He, Laura Balzano, and Arthur Szlam, "Incremental gradient on the grassmannian for online foreground and background separation in subsampled video," in *CVPR*, 2012, pp. 1568–1575.

[7] David Donoho and Iain M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, pp. 1200–1224, 1995.