

Detection, Tracking of the Dynamic Foreground in Complex Videos and Background Subtraction using Low Rank

Guru Prasad Rugi
M.tech 4th SEM Dept. of E&IE
DSCE, Bangalore

Dr V S Krushnasamy
Assoc. Prof, Dept. of E&IE
DSCE, Bangalore

Abstract: In the modern computer vision applications many different techniques have been proposed for the enhancement of the performance of compression methods of video. One of the approach is for a given scenario contents motion models are used. Then compensated coder of the classical motion is made use which skips or the some parts of the sequential frames of the videos are not coded. The decoder is used for reconstructing the parts of the video frames which are not coded. This paper demonstrates the method of video coding by making use of special surface or texture models. We build the gesture models by making use of motion insights of the humans. Foreground objects moves differently from background hence we construct a new model known as motion classification model.

Keywords— Dynamic Object, Soft-Impute, PCP, Object Detection, Low-Rank, DECOLOR.

I INTRODUCTION

Dynamic object detection is an important issue in the surveillance of the videos. Now a days the cameras for video surveillance are being used in all the public location's like hotels, malls, offices, ATM's, schools etc. surveillance of the videos is helpful in monitoring a recorded or a live video.

For different vision applications the programmed video analysis is a vital step. The applications are detection of the theft, observing the traffic, investigations of the police etc. the three main steps involved in programmed video analysis are recognising the behaviour of the object, tracking of the object, detecting the object. In detecting the object, the location of the object is traced in a given video. Then the detected objects are tracked from frame to frame, or from border to frame for analysing the behaviour of the object. Hence the phase of detecting the object is an important step.

The task of object detection is accomplished by the background subtractions or by the object detectors. One of the method employed in the object detector is it uses a basic classifier which scans all the input frames by window and each u-image is labelled as either background or foreground. Usually the classifier is constructed online, at the starting of the video a labelled frame is initialised. Usually the method opted in the background subtraction is the comparison of frames with the frames with the model of the background constructed. The assumption made on the background model is that there are no objects. Hence when the models are built on such assumption are limited for only some of the scenarios which cannot be applied to the real time videos or scenarios in the applications of the analysis of the programmed videos. The method proposed to overcome such limitation makes use of motion data for separating the objects from the background.

The complete situation can be presented as follows; given a set of input sequence of frames where the dynamic objects are present. The foreground or the object moves in a completely different way as compared to the background. The issue here is that can we separate the background from the foreground by coding or programming. For example in this paper we need to segment the smoke from the background. In such a situation we can make use of the motion segmentation which classifies the pixels differently for each motion patterns. The methods defined above can calculate the optical flow and the segmentation is accomplished for the large moving cameras also. However the assumption made by these approaches is that the motion is rigid which is not true practically. Hence these approaches will fail for the real time applications. The motions in the background may be of non-rigid type like smoke, illuminations changes due to the weather the ocean waves and waving of the trees etc. The background estimation is another technique used in motion-based approach. When the testing sequences are available the background model is constructed directly which is different from the background subtraction method. This method make san assumption that the background is not dynamic hence it fails to handle the scenarios which include dynamic changes in the background or complex background.

2. OBJECTIVES.

The main objective here is to enhance the efficiency of the video coding and precise segmentation of the foreground involving the complex scenarios like waving of the tree, segmentation of the smoke, segmentation of the objects captured by the dynamic cameras.

Here we are going to use the lo-rank representation which enhances the capability of the incorporating the dynamic cameras variations. Here we use a method low-rank matrix is approximated by for background and outliers are detected as the dynamic objects or foreground objects.

3. PRBOLEM DESCRIPTION

There are many techniques which have been proposed for object detection with the training pulse. We concentrate on foreground detection without the training pulse which is a challenging task in this paper. Many techniques have failed in dealing with the complex scenarios like illumination changes in the background and the when dealing with the non-rigid

motions. Hence we develop an algorithm for dealing with such situations accurately.

4. ANALYSIS OF THE SYSTEM

A. Existing System

One of the existing method is signal sparse recovery where the assumption made is that the cameras are static and there is no dynamic changes in the background. The techniques used for object or foreground detection are background subtraction segmentation of the image, object detectors. These techniques work only when the motion is rigid kind and static backgrounds. Hence the limitations of such techniques are:

- Lower efficiency
- Computational capability is low
- Higher computational complexity

B. Proposed System

We propose a new system where we precisely formulate the all the issues into a single optimization problem. The algorithm designed here can deal with the non-rigid motions, complex backgrounds or the dynamic backgrounds. It can handle the scenarios like sudden changes in the illuminations, dynamic textures etc. The algorithm designed and developed is the DECOLOR.

The input video sequences are correlated linearly, hence video frames are vectorized and a matrix is constructed for estimating the background by representing in low-rank and from the constructed low-rank we can detect the objects as outliers. Hence the assumptions made on the foregrounds can be excluded. The variations in the backgrounds can be accommodated into the low-rank matrix and hence the system designed is more flexible. Here in the proposed system the DECOLOR does not need the training pulses for detecting the object and estimating the background

5. SYSTEM DESIGN

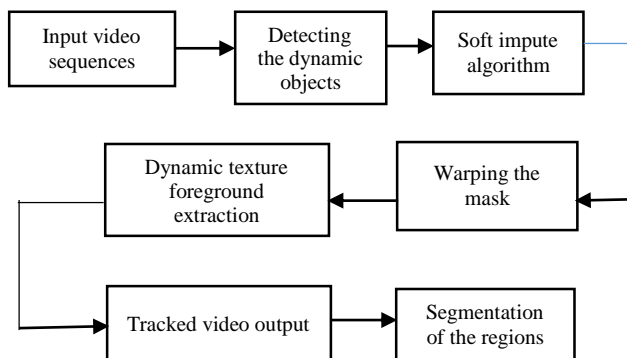


Fig 1 system architecture

A. Background and foreground identification

In the videos captured by the moving or dynamic cameras contains both the background and foreground. We need to identify the background and the foreground for the process of detecting the objects and estimating the background. This can be accomplished by following methodology.

$$S_{ij} = \begin{cases} 0 & \text{if background is } ij \\ 1 & \text{if foreground is } ij \end{cases} \quad (1)$$

The j-th frame's i-th pixels is represented as i

Modelling the background: The intensity of the background is conserved in all the input sequences until and unless there is an illumination change or due to the dynamic textures

Foreground modelling: The object that moves differently from the background is identified as foreground

Signal modelling: For a given S and B we need to formulate the D input matrix. $S_{ij} = 0$ indicated the background region the assumption made here is $D_{ij} = B_{ij} + Q_{ij}$, Q_{ij} represents the Gaussian noise. In the situation of least square B_{ij} fits into the D_{ij} provided $S_{ij} = 0$. For $S_{ij} = 1$ represents the foreground region, foreground occludes the background region. When $S_{ij} = 1$, D_{ij} is not constrained because we don't make any assumptions on the foreground.

B. Low Rank Matrix formulation:

The conservation of the background intensity for a sequence of frames indicates that the sequence of images are linearly correlated results in formation of low-rank matrix except the intensity changes which may arrive due to the changes in the illumination. Whenever the moving object is encountered there is a change in the background model. due to which some of the pixels in the background are lost. For filling the missing pixels we use SOFT-impute algorithm.

By number of iterations we fill the missing pixel values by SOFT-impute. Here the nuclear norm regularisation is used for matrix completion. We have two techniques which can be used for filling the missing values one approach is by making use of soft threshold svds iterations and the second approach is using the least square alternatively. There is a special sparse matrix called "incomplete" which can handle large matrices efficiently.

Estimation of the low-rank matrix B: For minimization over B we have the estimated support S the problem is stated as matrix completion problem:

$$\min_B \frac{1}{2} \|P_S(D - B)\|_F^2 + \alpha \|B\|_*$$

Lemma I: Z is the matrix given, the solution for the problem is obtained as given below

$$\min_B \frac{1}{2} \|Z - X\|_F^2 + \alpha \|B\|_*$$

C. Contiguous outlier detection:

When the video is captured in an open area, the motion of the camera may go in any direction which results in motion in all the directions, furthermore the object may enter or leave the scenario and the size of the video may vary when it moves towards, away or to any other directions from the camera. Another scenario is that the object may stop for some amount of time in the video. Hence these situations result in the occlusion of the background variations which cannot be predicted. For example two people may stop, shake hands or wave hands etc.

We may solve this issue just by containing the movement of the foreground but this leads to the limiting applications to the few scenarios only. Different type of foregrounds are, in different scenarios here are different type of objects like in parking Lott there will be vehicles, person, birds, dogs. People may leave or pick some object in the scene. These different type of challenges are solved in our implementation.

The motion segmentation methods should be opted for detecting the foreground as outliers. The motion segmentation methods are carried out by implementing the designed algorithm DECOLOR. The scenario in which the foreground appears continuously in the input sequence of images are segmented by the DECOLOR which is a more challenging task. While learning the matrix the background and foreground are estimated simultaneously. As intensity changes are observed due to the foreground motion which cannot fit into the low-rank matrix hence can be detected as outliers. Finally the DECOLOR outputs the foreground, recovers background and masking is carried out accurately

Motion Segmentation: generally the motion segmentation is a process of classifying the pixels with reference to the patterns of the motion.

The steps in DECOLOR algorithm are shown below:

1. Input matrix: $D = [I_1, \dots, I_n] \in R^{m \times n}$
2. Initialising: $\tau^{\wedge}, B^{\wedge} \leftarrow D o \tau^{\wedge}, S^{\wedge} \leftarrow 0, \alpha, \beta$
3. Repeat
4. $\tau^{\wedge} \leftarrow \tau^{\wedge}, \operatorname{argmin}_{\tau} \|P_S(D o \tau^{\wedge} - B^{\wedge} + j_r \tau)\|$
5. Repeat
6. $B^{\wedge} \leftarrow \Theta_{\alpha}(P_{S^{\perp}}(D o \tau^{\wedge}) + P_S(B^{\wedge}))$;
7. Until converge
8. If $\operatorname{rank}(B^{\wedge}) \leq K$ then
9. $\alpha \leftarrow \eta_1 \alpha$;
10. Go to step 5
11. End if
12. Estimate σ^{\wedge}
13. $\beta \leftarrow \max(\eta_2 \beta, 4.5 \sigma^{\wedge 2})$
14. $S^{\wedge} \leftarrow \operatorname{argmin}_S \sum_{ij} \left(\beta - \frac{1}{2} ([D o \tau^{\wedge}]_{ij} - B^{\wedge}_{ij})^2 \right) S_{ij} + \gamma \|A \operatorname{vec}(S)\|_1$
15. Until converge
16. Output: $B^{\wedge}, S^{\wedge}, \tau^{\wedge}$.

II EXPERIMENTAL RESULTS

We will test the implemented method on the number of sequences like office.avi where there is a change in the illumination and the persons are moving, another data or the video is smoke.avi where we are going to segment the smoke from background. Next we carry out the comparison of our proposed method with the principle component analysis.

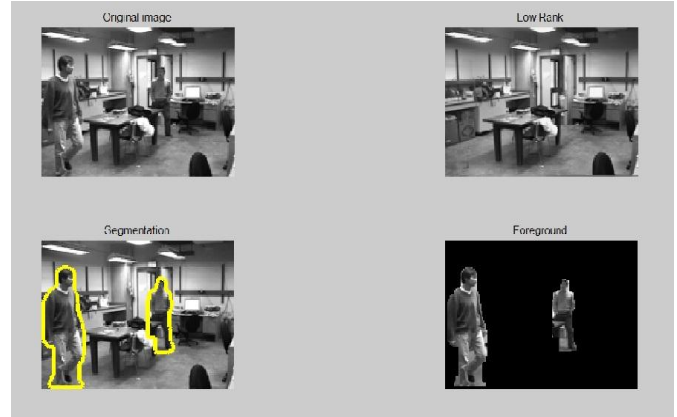


Fig 2. Segmented and background subtracted video of office.avi

The scenario where there is the illumination change and the people are moving is shown in the fig (2). Due to the illumination changes the background may be corrupted due to the noise induced by the illumination changes. Our implementation estimates the background was estimated precisely and segmentation accuracy is achieved.

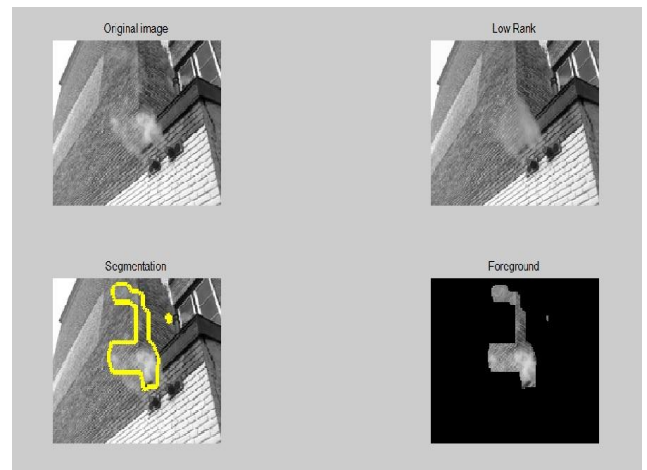


Fig 3 Segmentation and background subtraction of smoke in video smoke.avi

Now we estimate the scenario of dynamic background where we need to segment the smoke as shown in fig (3). The methods well estimates the background when considering the solid or liquid materials, they fail to estimate the backgrounds involving the gaseous materials. Our implemented algorithm can effectively estimate the background and can segment it precisely.

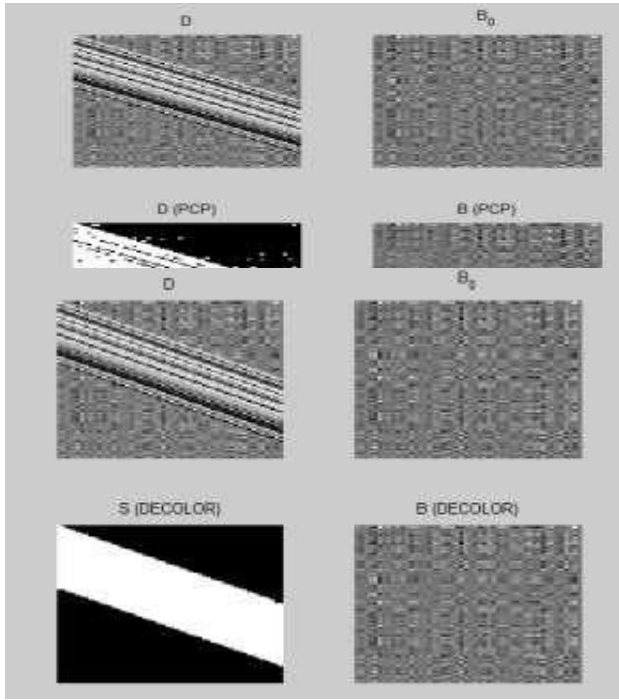


Fig 5. DECOLOR

Here we compare the results obtained from previous method that is principle component analysis and our implemented algorithm as shown in the fig (4) and fig (5). We have generated the results by generating an input matrix of $m=100$ and $n=50$, we have kept the width object ($W=40$) and signal to noise ratio ($SNR=10$) and we have simulated the results.

Root mean square error (RMSE): The root means square error value for the proposed DECOLOR and existing PCP are calculated. RMSE is calculated as follows

$$RMSE = \frac{\|B^* - B_0\|_F}{\|B_0\|_F} \quad (1)$$

Table 1. Comparison of root means square value of DECOLOR and PCP of different sequences

SEQUENCE	Proposed system	Existing system
Office.avi	0.0389	0.1095
Pedestrian.avi	0.320	0.757
Hall.avi	0.360	0.1254
Airport.avi	0.0261	0.0607
Watersurface.avi	0.0338	0.935

F-Measure: Next we calculate the F-measure of all the sequences we have used in the implementation. F-measure quantifies the accuracy of the segmentation of the objects or the tracked regions in the videos. The following table gives the total information of the data sets we have used throughout the project.

Table 2. Information of the sequences used

Data sequence	size	description
Office.avi	[160 120]	Complex scenario
Pedestrian.avi	[238 158]	Complex scenario
Hall.avi	[160 128]	Complex scenario
Airport.avi	[160 128]	Dynamic background
Watersurface.avi	[160 128]	Dynamic background
People1.avi	[320 240]	Dynamic camera
Peple2.avi	[320 240]	Dynamic camera
Cars6.avi	[320 240]	Dynamic camera
Cars7.avi	[320 240]	Dynamic camera
Smoke.avi	[180 144]	Dynamic background

Next we calculate the F-measure and variance for some of the data sets given in the table 2. The results are generated by keeping the signal to noise ratio and width of the object constant. The F-Measure is calculated by using the following equation and the abselon (noise) is calculated online.

$$F - \text{measure} = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (2)$$

Table 3 simulated values of the F-measure and sigma

Input data sequence	F-measure (%)	Abselon(€)
Office.avi	97.55	0.1673
Pedestrian.avi	95.81	0.1982
Watersurface.avi	95.10	0.1697
People1.avi	97.42	0.1757
Smoke.avi	96.64	0.1713

CONCLUSION

From above table we can observe that the root mean square error is reduce in the proposed method. Hence the video Coding efficiency is enhanced and segmentation accuracy is increased. In our work we have developed the efficient algorithms which are based on motion segmentation. We have precisely extracted the dynamic objects from the given

input frames of a video. We have segmented the foreground from both static and dynamic backgrounds precisely where other methods have failed. The computational complexity involved is very less the efficiency is higher compared with the PCP.

IV ACKNOWLEDGEMENT

My sincere thanks to Dr V S krushnasamy Assoc.Professor Department of E&IE for his innovative ideas and suggestions throughout my project for successful completion of the project

V REFERENCES

- [1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM computing surveys*, vol. 38, no. 4, pp. 1–45, 2006.
- [2] T. Moeslund, A. Hilton, and V. Kruger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Image Und.*, vol. 104, no. 2-3, pp. 90–126, 2006.
- [3] C. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Proc. of IEEE Int. Conf. Comput. Vis.*, 1998, p. 555.
- [4] P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *Int. J. Comput. Vis.*, vol. 63, no. 2, pp. 153–161, 2005.
- [5] H. Grabner and H. Bischof, "On-line boosting and vision," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Recogn.*, 2006, pp. 260–267.
- [6] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1619 –1632, 2011.
- [7] M. Piccardi, "Background subtraction techniques: a review," in *IEEE Int. Conf. on Systems, Man and Cybernetics*, 2004.
- [8] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. of IEEE Int. Conf. Comput. Vis.*, 1999.
- [9] R. Vidal and Y. Ma, "A unified algebraic approach to 2-d and 3-d motion segmentation," in *Proc. of Eur. Conf. Comput. Vis.*, 2004.
- [10] D. Cremers and S. Soatto, "Motion competition: A variational approach to piecewise parametric motion segmentation," *Int. J. Comput. Vis.*, vol. 62, no. 3, pp. 249–265, 2005.